FAKULTÄT FÜR !NFORMATIK
Faculty of Informatics

Sustainability Risk Monitoring in Supply Chains

Ranking Suppliers using Text Mining and Bayesian Networks with a Focus on Child Labor

DISSERTATION

zur Erlangung des akademischen Grades

Doktor der Sozial- und Wirtschaftswissenschaften

eingereicht von

Mag. Dipl.-Ing. Andreas Thöni, BSc
Matrikelnummer 0426645

an der
Fakultät für Informatik der Technischen Universität Wien

Betreuung:
o.Univ.Prof. Dr. A Min Tjoa
o.Univ.Prof. Dr. Alfred Taudes

Diese Dissertation haben begutachtet:

(o.Univ.Prof. Dr. A Min Tjoa) (o.Univ.Prof. Dr. Alfred Taudes)

Wien, 05.06.2015

(Andreas Thöni)

Erklärung zur Verfassung der Arbeit

Andreas Thöni
Schulweg 744, A-6100 Seefeld in Tirol

Hiermit erkläre ich, dass ich diese Arbeit selbständig verfasst habe, dass ich die verwendeten Quellen und Hilfsmittel vollständig angegeben habe und dass ich die Stellen der Arbeit - einschließlich Tabellen, Karten und Abbildungen -, die anderen Werken oder dem Internet im Wortlaut oder dem Sinn nach entnommen sind, auf jeden Fall unter Angabe der Quelle als Entlehnung kenntlich gemacht habe.

(Ort, Datum) (Unterschrift Verfasser)

Acknowledgements

I am grateful to many supporters who have accompanied my work:

First, to my supervisors Prof. Dr. A Min Tjoa and Prof. Dr. Alfred Taudes for their time, their encouragement, and their constant constructive feedback during the course of this work.

To Dr. Jack L. King, Prof. Dr. Andreas Rauber, Dr. Alexander Schatten, and Arnavaz Schatten for the insightful discussions and suggestions.

To APA, Austria Presse Agentur, for allowing me to use the APA Basisdienst Corpus.

To my colleagues at university and extramural for the exchange of ideas, of motivation, and of laughters.

To my dear girlfriend, Marie-Theres, for her patience and constant encouragement.

Last but not least, to my parents without their support over time this thesis would not have been possible.

Andreas Thöni, 1.6.2015
Vienna, Austria

Abstract

This thesis presents components of a software system designed to monitor social sustainability compliance in supply chains. The components allow to ongoingly rank suppliers based on their risk of breaching sustainability standards. It focuses on child labor given the extension of social sustainability and child labor's imminent risk level. It proposes a novel risk model using a Bayesian network (BN) that allows the integration of public and private data sources. The risk level is modelled with an individual BN for each supplier location. The main focus of the work is on the model and the data integration. In order to integrate public data sources, a text mining approach is suggested to identify and extract social sustainability incident events in unstructured text. These events can be integrated into the BNs. An expert questionnaire together with sensitivity tests confirms the usefulness of the model. For the text mining task, tests on a newly developed gold standard suggest high F1 scores for incident event detection (83.0%). Incident event extraction performs better for the geography and company event attributes and less for the sector attributes. This implies that manual input revision appears necessary, also recommended by experts interviewed. Finally, an analysis of online retrievable empirical input data reveils the availability of interesting and detailed events either published by news agencies or posted by non-governmental organizations.

Kurzfassung

Diese Dissertation präsentiert Komponenten für ein Softwaresystem zur Überwachung von sozialer Nachhaltigkeit in Lieferantenketten (Supply Chains). Es ermöglicht die kontinuierliche Reihung von Lieferanten auf Basis des Risikos, dass diese soziale Nachhaltigkeitsstandards brechen. Aufgrund der Breite von sozialer Nachhaltigkeit und des imminent gegebenen Risikos geht diese Arbeit insbesondere auf Kinderarbeit ein. Dabei wird ein neuartiges Risikomodell basierend auf einem Bayesianischen Netzwerk vorgeschlagen, welches die Integration von privaten und öffentlichen Datenquellen erlaubt. Das Risikoniveau wird für jeden Lieferanten mit Hilfe eines dezitierten Bayesianischen Netzwerks modelliert. Der Fokus der Arbeit liegt auf dem Modell und der Datenintegration. Für die Einbindung von öffentlichen Datenquellen wird ein Text Mining-Ansatz vorgeschlagen, um Vorkommnisse in Bezug auf soziale Nachhaltigkeit automatisch zu identifizieren und zu extrahieren. Diese Vorkommnisse können in das Baysianische Netzwerk integriert werden. Eine Umfrage unter Experten und Untersuchungen der Modell-Sensitivität bestätigen die Nützlichkeit des Modells. Tests des Text Minings unter Nutzung eines neu entwickelten Gold-Standards bestätigen zudem hohe F1 Werte des Ansatzes (83.0%). Die Extraktion von Vorkommnissen funktioniert besser für die Geographie- und die Firmen-Dimension als für das Erkennen von Sektor-Bezeichnungen. Aus diesem Grund ist für eine praktische Anwendung ein manueller Bearbeitungsschritt nötig, welcher auch von befragten Experten empfohlen wird. Schließlich zeigte eine Analyse von Online verfügbaren Datenquellen, dass relevante empirische Daten verfügbar sind, die entweder von klassischen Zeitungen oder von Nichtregierungsorganisationen bereitgestellt werden.

Content

Abbreviations

ACE	Automatic Content Extraction
AHP	Analytical Hierarchy Process
ANNIE	A Nearly-New IE system
ANP	Analytical Network Process
APA	Austrian Press Agency
API	Application Programming Interface
APR	Augmented Precision & Recall
B2B	Business-to-business
BBC	British Broadcasting Corporation
BDM	Balanced Distance Metric
BN	Bayesian Network
BSC	Balanced Scorecard
BSCI	Business Social Compliance Initiative
CAGR	Compound Annual Growth Rate
CL	Child Labor
CLM	Child Labor Monitoring
CO	Co-reference Resolution
COSO	Committee of Sponsoring Organizations of the Treadway Commission
CPC	Central Product Classification
CPV	Common Procurement Vocabulary
CSR	Corporate Social Responsibility
CSRD	Candidate set reduction
DEA	Data Envelop Analysis
EC	Economic Census
ED	Event Detection
EE	Event Extraction
EMM	Europe Media Monitor
ERM	Enterprise Risk Management
ERP	Enterprise-Resource-Planning
ERTMS	European Rail Transport Management System
EU	European Union
GATE	General Architecture for Text Engineering
GDP	Gross Domestic Product
GHG	Greenhouse gas
GIS	Geographical Information Systems
GRI	Global Reporting Initiative
GS	Gold Standard
GSCP	Global Social Compliance Program
HRCA	Human Rights Compliance Assessment
HTTP	Hypertext Transfer Protocol
ICT	Information and Communication Technology
IE	Information Extraction
IIO	Independent Incident Observations
ILO	International Labor Organization
IO	Incident Observation

IPEC	International Programme on the Elimination of Child Labour
IR	Information Retrieval
IS	Information System
ISIC	International Standard Industrial Classification of all Economic Activities
ISO	International Organization of Standardization
IT	Information Technology
ITIL	IT Infrastructure Library
JAPE	Java Annotation Patterns Engine
JSON	JavaScript Object Notation
KNN	K-nearest neighbor
KPI	Key performance indicator
LA	Learning accuracy
LOD	Linked Open Data
LSI	Latent Semantic Indexing
LVA	Lifecycle Assessment
MIS	Management Information System
ML	Machine Learning
NACE	Statistical Classification of Economic Activities in the European Community
NB	Naïve Bayes
NE	Named Entity
NED	New Event Detection
NER	Named Entity Recognition
NGO	Non-Governmental Organization
NIST	National Institute of Standards and Technology
NLP	Natural Language Processing
OSH	Occupational Safety and Health
OWL	Web Ontology Language
PAUM	Perceptrons with uneven margins
PHP	Hypertext Preprocessor
POS	Part-of-speech
PR	Processing resource
QFD	Quality Function Deployment
RDF	Resource Description Framework
RE	Relation Extraction
REACH	Registration, Evaluation, Authorisation and Restriction of Chemical Substances
RFID	Radio Frequency Identification Technology
RoHS	Restriction of the Use of Certain Hazardous Substances
RQ	Research question
RSS	Really Simple Syndication
SA	Sentiment Analysis
SA	Social Accountability
SBSC	Sustainability Balanced Scorecard
SCI	Science Citation Index
SCM	Supply Chain Management
SCRM	Supply Chain Risk Management
SLCA	Social Lifecycle Assessment
SME	Small and Medium Enterprise
SPARQL	SPARQL Protocol And RDF Query Language
SRMS	Sustainability Risk Management System
SRP	Socially Responsible Purchasing
SSCI	Social Science Citation Index
SSCM	Sustainable Supply Chain Management

SVM	Support Vector Machine
TAC	Text Analysis Conference
TAM	Technology Acceptance Model
TDT	Topic Detection and Tracking
TM	Text Mining
TOPSIS	Technique for Order Preference by Similarity to Ideal Solution
TPR	Taxonomic Precision & Recall
TRA	Theory of Reasoned Action
U.S.	United States of America
UCW	Understanding Children's Work
UML	Unified modeling language
URI	Uniform Resource Identifier
URL	Uniform Resource Locator

1 Introduction and motivation

> *A terrible fire had broken out in the forest. All the animals were running away, including the lion, king of the forest. Suddenly, the lion saw a tiny bird rushing towards the fire. He asked the bird, "what are you doing?" To the lion's surprise, the bird replied "I am on my way to extinguish the fire." He laughed and said, "how can you kill the fire with just one drop of water, in your beak?" The bird was adamant, and said, "But I am doing my bit."*
>
> - Kailash Satyarthi, Recipient of the 2014 Nobel Peace Prize for his fight against child labor; childhood story
> (Nobelprize.org 2014a)

> *It takes 20 years to build a reputation and five minutes to ruin it. If you think about that, you'll do things differently*
>
> - Warren Buffett, Investor
> (The Telegraph 2013)

This thesis suggests components for an information technology (IT) system to support companies in managing the social sustainability risk in their supply chains.[1] The societal importance of global supply chains means pressure to comply with elementary social standards. In countries with a particularly high level of companies concerned with social responsibility, also factors beyond sole stakeholder pressure can drive responsibility measures (Julius Raab Stiftung 2014, 30). Hence, this thesis proposes a novel system using Bayesian networks (BN)[2] to combine evidence from public and private information sources, including news, in order to continuously monitor suppliers' compliance with social sustainability norms, specifically focusing on the data and model integration task. Additionally, it suggests a method for extracting social sustainability events from unstructured text that may be related and integrated into the BNs. Finally, the availability of data sources is analyzed. The approach is tailored to child labor, which is a major component of social sustainability.

[1] The introduction is in small parts based on Thöni, King, Tjoa (2014).
[2] BNs are probabilistic networks based on conditional probabilities that allow for rapid updates.

1.1 Supply chain management

Supply chain management (SCM) has become a key discipline for international enterprises (Ashby, Leat, and Hudson-Smith 2012), driven by a multitude of modern market factors such as reduced transportation costs, stronger competition, and shorter lifecycles (Simchi-Levi, Kaminsky, and Simchi-Levi 2003, 1). Multiple definitions of SCM have been proposed. Mentzer et al. (2001) analyzed these and coined the two definitions for supply chains and SCM used in this thesis. Supply chains are defined as "[…] a set of three or more entities (organizations or individuals) directly involved in the upstream and downstream flows of products, services, finances, and/or information from a source to a customer." (Mentzer et al. 2001, 4). Building on this, SCM is defined as:

> *[…] the systemic, strategic coordination of the traditional business functions and the tactics across these business functions within a particular company and across businesses within the supply chain, for the purposes of improving the long-term performance of the individual companies and the supply chain as a whole. (Mentzer et al. 2001, 18)*

Consequently, SCM is concerned with the overall performance of the members of a supply chain, both individually and combined. The behavior of each entity in a supply chain can affect its own performance as well as that of interconnected companies. At the same time these management activities span multiple fields (Mentzer et al. 2001).

1.2 Sustainability and supply chain management sustainability

Over the years, SCM has advanced and new topics have been emphasized or added. In recent years, sustainability has not only become a major topic of public and media discussion (Barkemeyer et al. 2009), but also in SCM. Sustainability in SCM has attracted a broadening community of scientific researchers (Hassini, Surti, and Searcy 2012; Seuring and Müller 2008b), and academic interest in sustainable in SCM is continually increasing (Ashby, Leat, and Hudson-Smith 2012). Pagell and Shevchenko (2014) even argue that all SCM research should be sustainable SCM (SSCM) research. Similarly, a steadily increasing number of companies are trying to make their businesses more sustainable (e.g. Ciliberti, Pontrandolfo, and Scozzi 2008a) and are adapting their behavior (Berns et al. 2009).

The discussion surrounding sustainability has been gaining momentum for decades. Historically, it originates from the field of agriculture (Grober 2009; von Carlowitz 1713). A recent driver has been the debate on planetary boundaries, that fuels the debate on sustainability by drawing consumption limits for the earth's population given the available resources (Meadows et al. 1972; Rockström et al. 2009). This idea is already present in the first widely adopted definition of "sustainable development", coined by the Brundtland Commission in 1987: "Sustainable development is development that meets the needs of the present without compromising the ability of future generations to meet their own needs." (World Commission on Environment and Development 1987, 43).

Elkington (1998) extends this thought and describes sustainability as a triple bottom line with an economic, an environmental, and a social dimension. Although this concept of the triple bottom line is criticized (W. Norman and MacDonald 2004)[3], the three components of sustainability appear to be broadly accepted.

In a business context, and expanding on the Brundtland Commission's definition, "corporate sustainability" can be defined as corporate behavior that meets "[…] the needs of […] stakeholders […] without comprising its ability to meet the needs of future stakeholders." (Thöni, Madlberger, and Schatten 2013a, 2). Underlining the general society as a stakeholder, a company's sustainability contributions are two-fold: it has to secure sustainability for both its own long-term existence and for the general public (Dyllick and Hockerts 2002). Within this environment, different stakeholders such as the media, governments (through regulation), or non-governmental organizations (NGOs) ask for companies to show increased consideration for sustainability (Gopalakrishnan et al. 2012; Mair 2011). Consequently, unsustainable behavior may result in reputational damage (Forstmoser and Herger 2006; Schiebel and Pöchtrager 2003; World Economic Forum 2009). From a business perspective, the human consequences of unsustainable behavior also need to be considered (Dreyer, Hauschild, and Schierbeck 2005). In contrast, sustainable behavior may ameliorate market positioning and drive competitive advantage (Carter and Jennings 2004; Faris et al. 2013; Schaltegger, Windolph, and Harms 2010; Vermeulen and Seuring 2009).

Sustainability is often seen in strong connection with corporate social responsibility (CSR), and the terms are sometimes even used interchangeably (Hutchins and Sutherland 2008). Thereby, CSR has a strong relation to companies' ethical values and stakeholder inclusion (Sarkis et al 2010).[4] Questions have been raised about the connection between CSR and repu-

[3] Norman and MacDonald criticize primarily the "bottom line" used in the concept of Elkington. They state that the "[…] concept of a 'Triple Bottom Line' in fact turns out to be a 'Good old-fashioned Single Bottom Line plus Vague Commitments to Social and Environmental Concerns.'" (W. Norman and MacDonald 2004, 256).

[4] Building on earlier work, Wood summarizes three noteworthy principles of CSR (Wood 1991, 696): legitimacy, public responsibility, and managerial discretion.

tational management, as well as whether CSR might to a certain extent be explained by reputational management (Bebbington, Larrinaga, and Moneva 2008). Indeed, some authors describe CSR as a form of reputational management (Meng and Zhao 2010), as it aims to manage stakeholder relationships that may affect a company's reputation (e.g. media, customers, investors; Fombrun, Gardberg, and Barnett 2000). Therefore, companies need to manage all three areas of sustainability while at least understanding its influence on their company's reputation.

Consequently, sustainability needs to be included in every step of the production process (Cachon and Terwiesch 2013). Sustainability starts with product design and ends with recycling activities (Ramani et al. 2010; H.-F. Wang and Gupta 2011). Given that a "supply chain is only as strong as its weakest link" (Lubber 2012) and that supply chains compete with each other (Kogg and Mont 2012; Dyer and Singh 1998), sustainability management has to go beyond the company itself and include elements along the whole supply chain (Miemczyk, Johnsen, and Macquet 2012; Soosay, Fearne, and Varsei 2014). Hence, compliance with defined sustainability standards needs to be achieved along the whole supply chain and not only by first tier suppliers (Mares 2010). Spence and Bourlakis frame this as a development towards "supply chain responsibility" (Spence and Bourlakis 2009, 294). In fact, companies are increasingly held responsible for their suppliers' behavior (Klassen and Vereecke 2012).

In light of the above and in the supply chain context sustainability has been defined "[…] as the extent to which supply management incorporates environmental, social, and economic value into the selection, evaluation and management of its supply base." (Giunipero, Hooker, and Denslow 2012, 260). Adopting this practice can have several benefits apart from achieving a good reputation, including "[…] lower consumption of natural resources, reduced costs, optimized operating processes, improved business relationships, decreased risks, and higher employee motivation." (Dey, LaGuardia, and Srinivasan 2011, 1250) To also witness increased economic performance, a strategic approach to SSCM is needed (Carter and Rogers 2008).

Nevertheless, competing definitions of SSCM exist. After reviewing multiple definitions of SSCM, Ahi and Searcy suggest a combined version which breaks down elements of SCM and stresses a long-term perspective:

> *[SSCM is the] creation of coordinated supply chains through the voluntary integration of economic, environmental, and social considerations with key inter-organizational business systems designed to efficiently and effectively manage the material, information, and*

capital flows associated with the procurement, production, and distribution of products or services in order to meet stakeholder requirements and improve the profitability, competitiveness, and resilience of the organization over the short- and long-term. (Ahi and Searcy 2013, 339)

"The sustainable supply chain" is a very broad concept (Koh et al. 2013). While the economic perspective is an integrated part of business research, the "green" side of sustainability in SCM (i.e., environmental sustainability) has also received particular attention in recent years (Seuring and Müller 2008b). However, this is not the case for social sustainability, which has historically received only limited attention (Carter and Easton 2011; Seuring and Müller 2008b). Comparable results can be seen in the metrics literature (Hassini, Surti, and Searcy 2012). Nevertheless, recent findings suggest a growing focus on this topic (Winter and Knemeyer 2013), and researchers have explicitly called for an increased focus on social risk (Miemczyk, Johnsen, and Macquet 2012).

A core problem of social sustainability is that it is very hard to grasp (W. Norman and MacDonald 2004). The concept encompasses a multitude of factors (see e.g. Global Reporting Initiative 2011). Actually, social performance can consist of a large set of different parameters, including health, wages, justice for all, or the creation of social capital (Mani, Agrawal, and Sharma 2014). Benoît and Vickery-Niederman (2011) especially stress three issues when detailing social sustainability: First, no fully agreed standard is available for measuring its impact. Second, combining indicators is difficult due to the partly qualitative nature of results. Finally, gathering local social data is constrained for global supply chains. Multiple competing frameworks have been proposed to assess overall social sustainability. While similar to other standards, it is also often influenced by various local cultural and normative considerations (Sarkis, Helms, and Hervani 2010). Issues with elements of social sustainability such as human rights or forced labor can have a significant effect on a company's image (M. B. Taylor, Zandvliet, and Forouhar 2009), even if they occur elsewhere in the supply chain (Lemke and Petersen 2013). Although the responsibility for sustainability is shared along the whole supply chain (Vermeulen and Seuring 2009), the focal, dominant company may be particularly affected. This can be seen in the cases of Nike and Gap in 2000 (Kenyon, Campbell, and Hawkey 2000) or of Foxconn and Apple in 2012 (Tsukayama 2012), which involved accusations of unethical working conditions and show that current risk management practices at least partly fail. Triggered boycotts also pose a significant economic risk (Anderson and Anderson 2009). Hence, in contrast to other perspectives which depict sustainability as a long-term systemic concept, we believe sustainability risks have short-term or mid-term effects on organizations (Gray and Wiedemann 1999). Consequently, any risk management needs to cover the whole supply chain (Seuring and Müller 2008a).

Thus social sustainability is a double-edged sword for companies. While a positive "sustainable" image provides several beneficial effects such as an increased customer base, higher employee motivation, or increased attractiveness to potential employees or partners (U. Hansen and Schrader 2005), sustainability issues – particularly issues with social sustainability – can have damaging results.

Given that potential issues with social sustainability in a supply chain as outlined are often associated with a company's reputation, their effects fall into a different risk category than "classical", resource-disruptive risks which can be mitigated with safety stock (Lemke and Petersen 2013). Hofmann et al. (2014) suggest differentiating at the incident level; whereas "classical" risks materialize in operational disruption, sustainability risks trigger stakeholder reactions. A key problem of reputational risk is that it is hard to measure in financial terms, making rankings more difficult (Lemke and Petersen 2013; Norrman and Jansson 2004). In addition, the public perception after an incident is often independent from the legal point of view (Mares 2010). In this sense, reputational issues require a different management approach than "classical" risks (Giannakis and Louis 2011). Reputational spillovers in the supply chain are particularly important (Lemke and Petersen 2013). Webby describes this from a CSR perspective that can be transferred to social sustainability:

> *CSR risk is unique in that intermediate suppliers often do not provide effective insulation against harm. Many other supply-related risks (e.g., financial insolvency, late delivery) that occur with so-called "lower-tier" suppliers will likely be absorbed to a great extent by the middle-tier, and thus the impact on customers further down in the value chain is often minimized. CSR risks like poor working conditions, unfair labor practices, or environmentally destructive operating practices are different: they can impact any company anywhere along the value chain – but they pose the greatest risk to well-known, highly-branded companies or public institutions. No single supply organization has sufficient resources to adequately manage CSR risks, or take action to address shortfalls, in 100% of its supply chain. Yet the organization may still be held responsible for those shortfalls […] (Webby 2006, 1).*

Consequently, investments in CSR have been strongly connected with the risk associated with social issues in the supply chain (Fombrun, Gardberg, and Barnett 2000), and enforcing CSR practices in the supply chain is seen as a mitigation measure for reputational risks (Lemke and Petersen 2013). Indeed, managing social sustainability is sometimes seen more as risk minimization than a profit enhancing strategy, as the latter in particular is not necessarily guaranteed (Carter and Rogers 2008). It is therefore related to supply chain risk management (SCRM) in the context of sustainability, which may be defined "[…] as the ability of a firm to understand and manage its economic, environmental, and social risks in the supply chain." (Carter and Rogers 2008, 366).

1.3 Management of social sustainability risk in supply chains

Kogg and Mont (2012), building on Kogg (2009) and Rogers (2011), suggest three generic approaches to tackle environmental and social issues in supply chains: de-selecting suppliers, working indirectly via compensation or philanthropy, or adopting direct approaches. The latter may be implemented via criteria in the supplier selection processes or by implementing inter-organizational management systems that exercise influence or monitor, test, and evaluate suppliers. Related to this, Harms, Hansen, and Schaltegger (2013) differentiate between risk- and opportunity-oriented strategies for managing sustainability in supply chains. While supplier collaboration and development offer one strategic option, minimizing risks is still seen as important. Thus companies often use a combination of the two approaches (Ciliberti, Pontrandolfo, and Scozzi 2008b; Harms, Hansen, and Schaltegger 2013). Even though both strategies may lead to improved sustainability performance, an assessment alone might not be completely successful (Gimenez and Tachizawa 2012). Risk assessments can be understood as a preceding activity before advancing to deeper collaboration (Klassen and Vereecke 2012). In this context, Foerstl et al. (2010) identify a first-mover advantage for companies which determine the sustainability of their suppliers as early as possible.

As a result, social sustainability risk needs to be managed as one sub-element of SCRM. Especially Lemke and Petersen (2013) build a strong case for dealing with reputational damage from a risk management perspective. Stakeholders demand a timely response to any incidents in the supply chain (Foerstl et al. 2010). Nevertheless, although present in visions, missions, and goals, strategic sustainability considerations are rarely integrated into operational management (Petrini and Pozzebon 2009a). Schaltegger and Burrit name these types of predominantly reputational risks as being risks with an "'either-or' *nominal* measurement" (Schaltegger and Burritt 2014, 234) given that even a single instance might be too much. This makes them very hard to track; nevertheless, leading indicators may exist that can be used. For this purpose, standards are needed to provide a baseline for what to measure, and approaches might be very specific to a certain sustainability risk.

"Codes of conduct" are acknowledged as a primary instrument for this concern. They define what standards need to be followed, thereby regulating sustainability aspects in supply chains and guiding suppliers (Ciliberti, Pontrandolfo, and Scozzi 2008b; Egels-Zandén 2007; Pedersen and Andersen 2006).[5] Their effects are seen as at least partly positive (Egels-Zandén 2014), although complementary measures may be necessary (Locke et al. 2007). Nevertheless, codes of conduct do not solve the problem of information asymmetry in supply chains (Sarkis, Zhu, and Lai 2011). Consequently, "codes of conduct" must be monitored and enforced to guarantee compliance (Pedersen and Andersen 2006). This has also been empirically seen (Egels-Zandén 2014). Egels-Zandén recently stated that buyers need to "[…] increase their code of conduct demands and auditing, coordinate their code of conduct activities throughout the industry, and value transparency more highly […]" (Egels-Zandén 2014, 71) to improve compliance with their "codes of conduct". In this context, monitoring is different from pure information sharing as the information needs deeper evaluation.[6] The monitoring from an ethical perspective has to cover the entire supply chain, "[…] from the point-of-origin to the point-of-consumption […]" (Svensson 2009, 262).

Supplier assessments are seen as a particularly important tool to safeguard compliance with pre-defined standards (Keating et al. 2008; Miemczyk, Johnsen, and Macquet 2012). For this purpose auditing is an important approach to ensure that pre-defined standard are met (Klassen and Vereecke 2012). On-site auditing is often required to gain insights into social sustainability (Kogg and Mont 2012). Using standards, a varying number of suppliers are subsequently reviewed and audited at regular intervals (Klassen and Vereecke 2012). This practice has become commonplace in companies in recent years (Benoît and Vickery-Niederman 2011). Another important instrument often associated with audits are certifications, defined at an inter-company level (Ashby, Leat, and Hudson-Smith 2012), which help reduce control costs through sector-specific or cross-sector initiatives (Vermeulen and Seuring 2009). Typically, they standardize audit details to a certain extent (Kogg and Mont 2012). These approaches to measuring social sustainability often require a significant amount of resources together with third party input (Vermeulen and Seuring 2009).[7] Consequently, one approach has been to form coalitions to share insights from monitoring and reduce costs (Bremer and Udovich 2001).

[5] Nevertheless, also the quality of a "code of conduct" (in the case of the cited study: CSR) affects a company's sustainability performance (Erwin 2011). Moreover, "codes of conduct" vary in type and number of issues addressed (Emmelhainz and Adams 1999).

[6] Supply chain transparency and visibility are often seen as drivers for improved economic performance within a supply chain (Bartlett, Julien, and Baines 2007).

[7] For example, a complete social lifecycle assessment requires a significant amount of input from all levels of the supply chain (Dreyer, Hauschild, and Schierbeck 2005; Dreyer, Hauschild, and Schierbeck 2010a).

However, monitoring and assessment approaches based on audits and certifications are constrained, for a number of reasons: First, supplier monitoring and assessment in modern supply chains is complex. Today’s supply chains have a large number of suppliers, often globally dispersed; this complexity and the associated costs (Kogg and Mont 2012) makes effective ongoing verification of compliance with standards difficult. It seems impossible for companies to evaluate all factors in depth and let alone first-hand by an internal company employee, although the need practically would exist (Kogg and Mont 2012). Even more resources are required if deeper levels of the supply chain are considered in case these suppliers are known (Svensson 2009; Grimm, Hofstetter, and Sarkis 2014). In 2012, a survey of 600 major U.S. companies showed only around a quarter monitored the “codes of conduct” in place to ensure sustainability of their suppliers (this number increased to 34% in 2014) (Ceres and Sustainalytics 2014, 49). All in all, practices vary: from auditing all suppliers to focusing only on a few, to abstaining from audits altogether due to resource constraints (Leire and Mont 2010).

Second, timeliness is a problem. Given that internal supplier information is mostly only available through audits, certifications, or supplier communication (e.g. Klassen and Vereecke 2012), there is an associated lag in compliance verification. Infrequent certifications only provide limited defense against issues, as, for example, the child labor revelations at the OTTO group showed in 2007 (Aktiv gegen Kinderarbeit 2014; McDougall and Schmitz 2007). Moreover, suppliers nowadays tend to be overloaded with requests for certifications and audits (Ceres and Sustainalytics 2014, 56), which is referred to as “audit fatigue” (Kogg and Mont 2012, 162). Others even describe monitoring as a way to “[…] convey an adversarial stance […] more like a supply chain ‘bully’ than a CSR ‘champion’ […]” (Boyd et al. 2007, 342).

Third, there is a lack of objectivity. Data gathered through monitoring systems may not reflect the truth due to the potential effects of bribery, corruption, and culture or standards differences, particularly if third parties are used to perform the ground work (Leire and Mont 2010). Locke, Amengual, and Mangla declare the information gathered through audits to be “[…] often inaccurate, biased, and incomplete.” (Locke, Amengual, and Mangla 2009, 21). Some companies (e.g. British Telecommunications 2012) use supplier questionnaires to identify the companies to follow up with. However, these selections are based on the quality of data that suppliers themselves provide (Leire and Mont 2010).

Finally, the calculation can take considerable amounts of time. Usually, managers prefer indicators that are easy and fast to calculate (Mcintyre et al. 1998).

1.4 IT for socially sustainable supply chain management

At a more general level, IT has been identified as an important driver of sustainability in supply chains (Dao, Langella, and Carbo 2011; Teuteberg and Wittstruck 2010). IT can be used for SSCM in various capacities (Kurnia, Mahbubur, and Gloet 2012). One important advantage of IT is its ability to support transparency in supply chains (Trienekens et al. 2012). Historically, as with general SSCM, the literature on IT for SSCM mainly focused on the environmental rather than the social dimension of sustainability (Jenkin, Webster, and McShane 2011; Melville 2010).

Nevertheless, also in the specific context of IT for socially sustainable supply chains, IT can provide automated solutions tailored to tackling the constraints of audits and certifications outlined above. IT systems can be set up to reduce resource intensity and ongoingly control selected aspects of hundreds of suppliers. Related risk monitoring activities are seen as a particularly important field for IT (D. D. Wu, Chen, and Olson 2014). Data from diverse sources can be combined to provide an adequate overview of the entire supply chain. In the context of supply chain social sustainability risk monitoring, a combination of company-internal and external data can be useful to overcome time lags and provide additional facts (Hofman 2011; Kogg and Mont 2012). This may also enhance the objectivity of monitoring results.

Because the data available to an IT monitoring system might be limited, the IT system can work as a prioritization step for subsequent, more directed activities. IT can be a preliminary instrument for establishing an elaborated risk-based ranking (e.g. Australian Government 2010; Cahill 2011), reducing the need for other expensive data gathering approaches such as auditing or improving their effectiveness.[8] Given that many approaches to supplier prioritization rely on supplier self-assessment (Grimm, Stölzle, and Hofstetter 2013), improvement seems possible. Even if the relationship with suppliers is viewed as a more "commitment-oriented approach" (Locke, Amengual, and Mangla 2009, 337) with more interaction and exchange instead of strict compliance rules, having a prioritized list of suppliers can still guide resource usage and allow for a better assessment of the remaining suppliers (Foerstl et al. 2010; Locke, Amengual, and Mangla 2009). Finally, it may help managers cope with the

[8] In this thesis, a ranking is understood as "a position in a hierarchy or scale" (Oxford Dictionary 2013).

overabundance of data and information they have been inundated with in recent decades (Ackoff 1967).

External public information may be used to overcome the time lag between supplier reviews and also enhance data availability. Ongoing input from news sources or social networks may also help to identify risk-relevant events in the supply chain. These events can be gathered based on geographic-, sector-, and production-specific relations (see e.g. UNEP 2009, 60), or on other cause and effect relationships derived from literature related to a specific social sustainability issue. Integrating news information can also be seen as a form of external responsiveness required for successful sustainability risk management (Foerstl et al. 2010). In particular, web-based publication channels have led to an increase in the amount of directly available news from around the world (Leetaru 2011). However, as will be shown in the related work section (Chapter 2), current quantitative frameworks for social sustainability risk monitoring do not explicitly use this type of information. Nevertheless, unstructured text has already been promisingly integrated into risk management approaches in non-sustainability domains such as financial risks (e.g. Groth and Muntermann 2011), general business risks (Leidner and Schilder 2010), tracking society-related sustainability indicators (Rivera et al. 2014), or employee fraud (Holton 2009). External risk-related information collection has also been reported in the sustainability domain; however, automated, IT-supported risk modeling is not specified (Koplin, Seuring, and Mesterharm 2007).

From an IT perspective, text mining (TM) is the key technology, enabling automated processing of unstructured text and saving time by reducing the manual processes needed (L. Huang et al. 2010). Text mining can be described as a combination of complex data extraction and data mining. Data mining itself is a combination of information retrieval, information extraction, and information analysis for the purpose of discovering previously unknown information (Bilder 2010). Mining risk-related events is of particular interest for the social sustainability domain, for example incidents related to supplier locations. However, related work on event extraction, the relevant sub-field of TM, does not currently deal with extracting such events. This is particularly relevant as the performance of event extraction approaches depends on the type of event extracted (Arendarenko and Kakkonen 2012; Coppola et al. 2009; IJntema et al. 2012; Van Landeghem et al. 2011).

1.5 Research questions

This thesis aims partly to close the gaps in research into IT support for social sustainability risk management in supply chains. While the usefulness of unstructured text has been acknowledged for general supply chain management, the questions of how exactly TM can be used in the context of social SSCM and how the available data may be usefully integrated to form a prioritized view of suppliers remain. This leads to the overarching research question for this thesis:

Can social sustainability monitoring of supply chains be better supported with an automated risk management system that combines private and public data sources in order to prioritize suppliers based on their risk level?

Establishing a prioritization of suppliers requires a model that outputs at least an ordinarily ranked list of suppliers. Consequently, this thesis suggests a quantitative risk model to manage the social sustainability risk. To the author's knowledge, no risk model for social sustainability has been presented which deals explicitly with the integration of internal and external sources, including news. To ongoingly update the model, this thesis takes the approach of finding evidence that conflicts with the initial hypothesis that all suppliers conform with given, pre-defined standards (e.g., a "code of conduct") – otherwise they would not have been selected in the first place (Ciliberti, Pontrandolfo, and Scozzi 2008a). Consequently, evidence needs to be gathered that allows an assessment to be made of the likelihood that this hypothesis does not hold true.

As suppliers may be based in very diverse locations, the prioritization should, if possible, be at the level of supplier locations (Schaltegger and Burritt 2014). Because companies need to pursue strategies that let them develop a competitive advantage (Dao, Langella, and Carbo 2011), this quantitative model will often need to be tailored to each specific company. For example, indicators may be weighted differently by different companies (Zeydan, Çolpan, and Çobanoğlu 2011). Nevertheless, the amount of configuration needed should still be manageable for decision makers (W. Ho, Xu, and Dey 2010). This leads to the first research question:

RQ1: What would an integrative risk model that ongoingly integrates external and internal information, including news, to calculate the likelihood of a social sustainability incident at a supplier location look like?

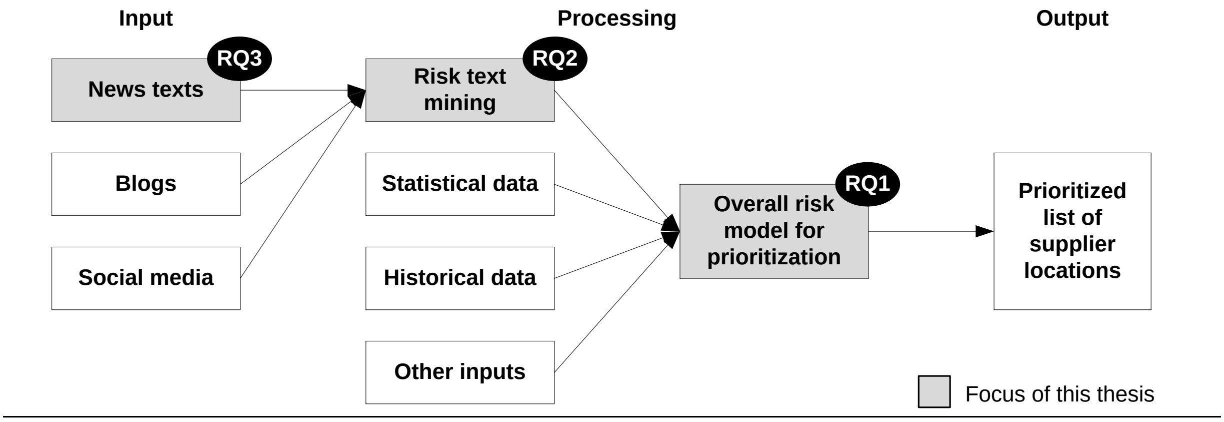
Additionally, extracting useful information from text requires a TM approach. Given that no tailored TM system for extracting risk-relevant social sustainability events has been suggested, this thesis investigates a second research question:

RQ2: What would a text mining approach for the automatic inclusion of publicly reported observations on social sustainability incidents look like?

Finally, the usability of the overall system strongly depends on the availability of practical data. Consequently, a third research question investigates this data availability empirically:

RQ3: Are there online public data sources that publish information related to child labor incidents?

Summing up, this thesis introduces a risk model for ongoing monitoring of social sustainability risks at supplier locations integrated into an IT system which includes text mining of sources outside a company and focuses on the risk of social sustainability issues.[9] Figure 1 shows an overview of this risk model together with the IT system's components and associated research questions. The system analyses online public sources and suggests a text mining approach to identify and categorize social sustainability incidents. The information is integrated in a model which aims to rank supplier locations[10] based on the risk of breaching predefined standards. A prototype of the complete system is implemented as a proof of concept of the individually tested components.



Note: News information is highlighted as the focus of this thesis as this is the major concern of the TM approach. However, links in social media will be used to gather additional text for analysis.

Figure 1: Overview of risk model and system with research questions (author's representation)

[9] By focusing on supplier locations, this thesis excludes transport links between supplier locations.

[10] Supplier locations are production or other locations of a supplier company that have several predefined attributes. They are related to one or more companies, situated at a specific geographic location, produce one or multiple goods or services, and thus can be related to one or several industry sectors.

It should be noted that monitoring is only one step in socially sustainable SCM. Or, as Teuscher, Grüninger, and Ferdinand put it, "[A]udits and codes of conduct alone do not improve working conditions" (Teuscher, Grüninger, and Ferdinand 2006, 7). Consequently, this thesis should be seen as one building block in the work suggested by Leire and Mont with regards sustainable purchasing:

> *Regardless of choice of auditor, the problems that arise from verification of information provided by suppliers seem to constitute the major stumbling block in the process of SRP [socially responsible purchasing, A.T.], mainly due to practical difficulties relating to the suppliers' operations and cultural contexts. Potentially, a different approach to ensuring supplier compliance needs to be developed. (Leire and Mont 2010, 38)*

This thesis assumes that the structure of the supply chain is known, as this poses an extra problem in SCM (Madlberger et al. 2014). Moreover, the focus of TM is on the extraction of events needed for the risk model. Technical aspects of stream-processing news sources (e.g. Weiler, Mansmann, and Scholl 2012) are not discussed. Finally, initial work will partly be restricted to certain domains or geographical areas. To be able to build on available technology and to have enough content and context for further processing (Sankaranarayanan et al. 2009), this work focuses on text mining from news sources. This thesis uses English texts, because a large amount of international news is broadcast in English, international organizations operate in English, English language tools are advanced, and organizations that want to be internationally recognized use English. Given the difficulty of quantifying reputational damage, the supplier ranking will not include the estimated impact of supplier-dependent incidents (Foerstl et al. 2010). All in all, the final system should suggest a system for prioritizing suppliers based on their risk level to facilitate a structured sustainability risk management approach (Foerstl et al. 2010; Harland, Brenchley, and Walker 2003).

1.6 Child labor risk in supply chains

One major element of social sustainability present in multiple sustainability standards is child labor. Child labor is the focus of two important conventions (Nos. 182 and 138) issued by the International Labour Organization (ILO) (International Labour Organization 1973a; International Labour Organization 1999a), and is therefore more strictly defined and acknowledged than other topics of social sustainability (e.g., corruption). Some governments now require companies to report on their measures against practices such as child labor in their supply chains (Travis 2014). The 2014 Nobel prize for Kailash Satyarthi and Malala Yousafzay recently focused global attention on the topic (Nobelprize.org 2014b). Overall, child labor is a social sustainability issue which attracts widespread media attention (Giblin 2012; Vandenberg, Nippierd, and Gros-Louis 2007a, 51). Media coverage involves the publication of company child labor cases (Barboza 2014; Vandenberg, Nippierd, and Gros-Louis 2007a, 51) as well as calls for action on a more generalized level (Weisbrot 2014). Nevertheless, child labor is not universally condemned, as indicated by its legalization in Bolivia for children over 10 years old (Zeit Online 2014).

According to the ILO definition, child labor is still common in many fields of work. In 2012, the number of children working in any form totaled 168 million, with 85 million of these performing hazardous work involving physical abuse or the handling of dangerous machinery (International Programme on the Elimination of Child Labour 2013, 13).

Due to the fact that each social sustainability risk may have different causes, this work will focus on child labor as a single element of social sustainability where issues may result in severe reputational damage for a company and which is usually included in "codes of conduct" or other standards. By reducing the domain and TM work to child labor incidents, we also follow the call of Grimmer and Stewart for a problem-specific evaluation of approaches (Grimmer and Stewart 2013). Nevertheless, the suggested approach can also be transferred to other areas of SSCM apart from child labor risk.

1.7 Methodology and structure

We begin by developing a risk model prototype. This model is primarily evaluated using sensitivity tests and an expert questionnaire. The suggested text mining approach is experimentally tested based on a reference corpus. Finally, empirical data is gathered and analyzed to discuss the availability of input data for practical use. The focus for these tests (and partly also other tests) is India and Indonesia, as these are both countries affected by child labor.

This thesis is split into nine chapters, as shown in Figure 2. Dots below each chapter box indicate which research questions each chapter deals with, as stated on the left of the figure. Chapter 2 highlights the related work necessary for all research questions. Chapter 3 provides initial empirical insight into the field, based on several interviews on sustainability monitoring in supply chains. These are helpful for understanding the risk model methodology outlined in Chapter 4. Chapter 5 focuses on the main input for the risk model and details the TM methodology. Although RQ1 and RQ2 are evaluated separately in Chapter 6, the prototype provides suggestions for the integration of the two steps and may help with practical understanding. Chapter 7 deals with RQ3 and the analysis of potential empirical input data gathered online. Chapter 8 contains a discussion of the results. Chapter 9 closes with the conclusion. The prototypical implementation is presented in “Appendix prototype implementation”.

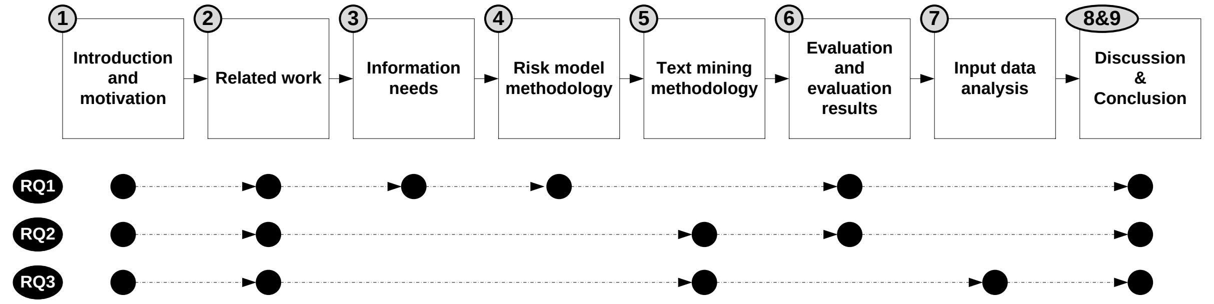


Figure 2: Overview of dissertation chapters and corresponding discussions of research questions, indicated by black dots (author’s representation)

2 Related work

This thesis suggests a risk model for ongoing social sustainability monitoring in supply chains integrating private and public data sources, including news. Therefore, the focus is on a quantitative risk management model that allows suppliers to be ranked. Consequently, the first part of this chapter (2.1) focuses on quantitative sustainability risk management in supply chains, particularly heavily related techniques with a focus on social sustainability. A more general analysis of IT for SSCM can be found in “Appendix A: IT for sustainable supply chain management”.

On a more granular level, this thesis suggests the integration of evidence from textual sources. The extraction of information from unstructured text will be detailed for the domain specific discussion (2.2). This provides insights into missing research in TM for social sustainability risk management in supply chains. Finally, Section 2.3 discusses child labor as an application domain.

For a broader and more general review of SSCM, we refer to several academic papers presented in recent years (Seuring and Müller 2008b; Hassini, Surti, and Searcy 2012; Srivastava 2007; Carter and Rogers 2008; Miemczyk, Johnsen, and Macquet 2012).

2.1 Sustainability risk management in supply chains

Generally speaking, a supplier’s sustainability is managed in one of two phases of its lifecycle (see Figure 3): before they become suppliers to a company or while they are suppliers to a company. Supplier risk management may or may not be seen independent of these two management processes.

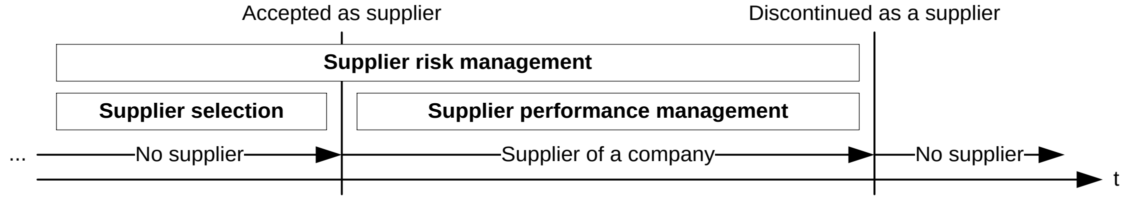


Figure 3: Management approaches for assessing supplier risks over the lifecycle of a supplier (author's representation)

Therefore, this section first introduces supply chain risk management before detailing the understanding of sustainability risk management. A final section surveys social sustainability-related approaches from literature on supplier selection and supplier performance management that have not been directly discussed in the context of risk management.

2.1.1 Introduction

Risk management tries to identify potential issues in the supply chain before they occur in order to mitigate or reduce their impact (Carter and Rogers 2008). In a complex field such as SCM, this can be a time-consuming task, with low confidence that risks have been exhaustively covered (Huang et al. 2010). Nevertheless, managing risks in a supply chain is seen as a cornerstone of a resilient supply chain (Christopher and Peck 2004). The proactive management of risks has become particularly important in recent years, stressing the need to continuously monitor risks over time (Blackhurst, Scheibe, and Johnson 2008; Sodhi and Tang 2009). Partly, SCRM can be included in general proactive SCM as a "secondary benefit" after economic consequences (Zsidisin et al. 2004). Thereby, SCRM is strongly connected with managing exceptions in logistic services (Lingzhe Liu, Daniels, and Hofman 2014). Or, as Liu, Daniels, and Hofman put it with regard to general risk management: "[R]isk management is closely related to the management of business exceptions. The exceptional incidents, themselves albeit not necessarily risky, are early risk indicators and will possibly escalade to severe accident if they are not properly detected and corrected in time." (Lingzhe Liu, Daniels, and Hofman 2014, 190). Typically, exceptions can be detected by comparing an actual value with the norm (Lingzhe Liu, Daniels, and Hofman 2014). However, this can be different for supply chain sustainability risks; for example, the norm of child labor should be "zero cases of child labor". Consequently, any event representing or leading to child labor may already be considered an exception.

2.1.2 Approaches to sustainability risk management in supply chains

In the past, sustainability risk management has mainly been discussed from an environmental perspective. This has been strongly driven by regulatory frameworks and laws that have been established over recent decades (Anderson and Anderson 2009). However, the social dimension of sustainability and its risk implications have also been acknowledged (Klassen and Vereecke 2012; Carter and Rogers 2008; Ghadge, Samir, and Kalawsky 2012). Sustainability risk management may be understood differently depending on which entity is at the core of the analysis. For example, one broad method of understanding sustainability risks is to consider sustainable economic development as being affected by potential risks (Krysiak 2009) or social risks being mitigated by general social protection (Holzmann and Jørgensen 2001; Holzmann and Jørgensen 1999). Conversely, sustainability risks may be understood on a small scale as securing the sustainability of one's personal life(style). Finally, as investigated in this thesis, economic entities like companies may be at the core of the analysis and affected by sustainability risks.

The definition of general "sustainability risks" can be derived from the ISO 31000 standard, an internationally recognized general risk management approach particularly used in a business context. It defines risk management as "[…] coordinated activities to direct and control an organization with regard to risk […]", with risk defined as an "[…] effect of uncertainty on objectives […]" (International Organization for Standardization 2009). "Sustainability risks" can therefore be seen as an extension of this initial concept, replacing "risk" with "sustainability risk" and understanding them as the "effects of uncertainty driven by sustainability effects on objectives," where sustainability is framed by the frequently used "triple bottom line," splitting sustainability into economic, environmental, and social elements (Elkington 1998). This broad definition of sustainability risks is in line with the literature (Anderson and Anderson 2009). Hofmann et al. (2014) provide a new definition that stresses the effects on stakeholders. They "[…] define a sustainability-related risk (in short, sustainability risk) as a condition or a potentially occurring event that may provoke harmful stakeholder reactions." (Hofmann et al. 2014, 168).

The literature contains several suggestions detailing what should be understood as a sustainability risk. Two ideas dominate the social dimension of sustainability. On the one hand, there are different overarching or industry-specific standards that have tried to define the social dimension. These include the Global Reporting Initiative (GRI; Global Reporting Initiative 2011), the SA8000 standard (Social Accountability International 2008a), the Business Social Compliance standard (Business Social Compliance Initiative 2009), the Global Social Compliance Program (Global Social Compliance Programme 2010), the Global Compact assessment (UN Global Compact Nordic Network 2013), the Human Rights Compliance Assessment (The Danish Institute for Human Rights 2006), and the Human Rights Impact Assessment (The International Business Leaders Forum and The International Finance Corporation 2010). Other standards also exist in particular domains, such as the wood (Forest Stewardship Council 2012) or cotton (Cotton Connect 2012) industries. On the other hand, multiple authors in academic literature have contributed their thoughts on the social dimension and introduced additional frameworks and categorizations for social sustainability (Labuschagne, Brent, and van Erck 2005; Carter and Jennings 2002; Closs, Speier, and Meacham 2011; Faris et al. 2013).

The breadth of GRI in the social dimension (see Table 1) showcases the multitude of considerations related to the social dimension of sustainability. Issues in different subtopics represent risks with different potential severity.

GRI category	Element/subtopics
Labor practices and decent work	Employment; labor/management relations; occupational health and safety; training and education; diversity and equal opportunity; equal remuneration for women and men
Human rights	Investment and procurement practices; non-discrimination; freedom of association and collective bargaining; child labor; forced and compulsory labor; security practices; indigenous rights; assessment; remediation
Society	Local communities; corruption; public policy; anti-competitive behavior; compliance
Product responsibility	Customer health and safety; product and service labeling; marketing communications; customer privacy; compliance

Table 1: Social GRI categories (Global Reporting Initiative 2011)

Building on this, sustainability risk management in supply chains extends this view to the overall supply chain (Foerstl et al. 2010). Consequently, Carter and Rogers reframe the basic SCRM definition and integrate sustainability directly, defining SCRM as "[…] the ability of a firm to understand and manage its economic, environmental, and social risks in the supply chain." (2008, 366). Managing sustainability risks in supply chains needs to integrate suppliers in revolving improvement cycles (Teuscher, Grüninger, and Ferdinand 2006). Social sustainability risk management is often associated with reputational risk management, because negative media coverage may lead to serious reputational issues (Anderson and Anderson 2009). Thus, managing sustainability and its attendant risks is at least partly connected with managing reputational risk (Faris et al. 2013; Forstmoser and Herger 2006; Bebbington, Larrinaga, and Moneva 2008; Schiebel and Pöchtrager 2003; World Economic Forum 2012). Still, social sustainability covers very diverse topics related to human dimensions (Dreyer, Hauschild, and Schierbeck 2005) which need to be analyzed in particular ways. One example is tailored auditing approaches with specific assessment schemes (e.g., questionnaires, governmental standards) (Keating et al. 2008; Miemczyk, Johnsen, and Macquet 2012).

Nevertheless, quantitative models for supply chain risk management that reflect social sustainability have rarely been presented in academic discourse. To the best of the author's knowledge, none of the related quantitative approaches deals with ongoing monitoring combining internal and external data sources, including news.[11] Rabenasolo and Zeng (2012) suggest an approach specific to textile supply chains which uses linguistic variables and relies on a combination of performance indicators. Weber, Scholz, and Michalik (2010) include sustainability risks into a credit risk indicator filled by credit officers. Badurdeen et al. (2014) outline a BN-based approach to combine various risk categories into a final probabilistic score. Nevertheless, social or environmental concerns are just one input out of several, and the analysis is based on expert input. Some overview papers on metrics (Clift 2004; Hassini, Surti, and Searcy 2012; Hutchins and Sutherland 2008) provide measures for sustainability risks. They suggest social factors such as "magnitude and nature of penalties for non-compliance," "number of regulatory violations by type," "community complaints," "risk," or "corruption risk" as separate measures without a framework (Ahi and Searcy 2015, 371). Finally, Hadiguna (2012) suggests, without mathematical details, a decision support framework for risk management that addresses social elements such as labor strikes, demonstrations, and local customs to a limited extent.

Different risk management processes lie at the core of previous, more general risk management methodologies. These processes provide the starting point for the workflows that need to be covered by a sustainability risk management system (SRMS). The prevalent ISO 31000 process covers seven main steps with the risk assessment at its core (The Association of Insurance and Risk Managers, The Public Risk Management Association, and The Institute of Risk Management 2010): After establishing the context, the risk assessment takes place in three steps. Risks are identified, analyzed, and evaluated before being treated. All five process elements are supported by monitoring and review tasks that create a circular six-item process. Finally, communication and consultation is introduced to support the previous steps. This process is outlined in Table 2, together with two other processes described below. A broader, also well-known framework has been defined with the enterprise risk management (ERM) system outlined in the COSO standard (Committee of Sponsoring Organizations of the Treadway Commission 2004). It includes an eight-step process model which has already been augmented with suggestions for how sustainability risk management can be included in the different steps (Faris et al. 2013). The eight steps are: internal environment setting, objective

[11] Appendix E: Bayesian networks for supply chain risk management details three quantitative risk approaches based on BNs for environmental sustainability (Joseph Amundson et al. 2014; Maleki, Bashkite, and Machado 2012; Yen and Zeng 2010).

setting, event (risk) identification, risk assessment, risk response, control activities, information and communication, and finally monitoring.

Both frameworks have strongly overlapping elements, especially in risk identification, analysis, and handling. Although not specifically tailored to sustainability risks, their generic approaches could be adapted to comply with the requirements of the research field of this thesis. Some authors have already adapted existing approaches to include sustainability. Anderson and Anderson (2009) suggest a process tailored to environmental sustainability consisting of five steps: identify environmental loss exposure, analyze environmental loss exposure, control environmental losses (pre- and post-loss), implement environmental risk financing alternatives, and then restart the cycle by evaluating the process and implementing improvements. Nevertheless, since it focuses on environmental sustainability, it strongly highlights “loss exposure,” which is almost impossible to estimate in the context of social or reputational risk. As Reuter et al. put it: “Stakeholders will condemn any kind of misconduct irrespective of spend volume with a particular supplier or strategic relevance to the buyer.” (C. Reuter et al. 2010, 57).

More related to social sustainability risks, Grimm, Stölzle, and Hofstetter (2013) provide a recent overview of suggested practices from the supply chain sustainability risk management literature. These practices will not be repeated here; instead, we build on their summarization of related work that resulted in a maturity model for supply chain sustainability risk management. In particular, they provide evaluation questions that may be seen as a requirement for a risk management system and that may also serve as the basis for any IT approach being developed in this context. They differentiate three phases: risk identification, assessment, and control.[12] Finally, Yilmaz and Flouris (2010) outline a sustainability extension of the COSO ERM process model while Knott and Fox (2010) propose a model to improve sustainability risk management, linking risk management techniques to steps towards improving sustainability.

[12] This is related to the approach presented by Foerstl et al. (2010), who roughly split the risk control phase into consequences and response.

ISO 31000	COSO ERM	Anderson and Anderson	Grimm, Stölzle, and Hofstetter
1. Establish context	1. Set environment	1. Identify loss exposure	1. Identification
2. Identify risks	2. Set objectives	2. Analyze loss exposure	2. Assessment
3. Analyze risks	3. Identify events/risks	3. Control losses (pre-/post-loss)	3. Control
4. Evaluate risks	4. Assess risks	4. Implement risk financing alternatives	-
5. Treat risks	5. Response to risks	5. Evaluate and improve process	
Overarching: monitoring and review	6. Perform control activities	-	
Overarching: information and communication	7. communicate and consult/inform		
-	8. Monitor		

Table 2: Risk management processes based on propositions by ISO 31000, COSO ERM, and Anderson and Anderson (Anderson and Anderson 2009; Committee of Sponsoring Organizations of the Treadway Commission 2004; The Association of Insurance and Risk Managers, The Public Risk Management Association, and The Institute of Risk Management 2010; Grimm, Stölzle, and Hofstetter 2013)

A complete, integrative IT-based SRMS would need to cover all or at least the most important of these steps. In particular, the steps of the processes outlined above that are focused on risk assessment (identify, analyze, evaluate) can be seen as important elements of any practical risk management implementation, as they provide the necessary input for further steps. Being relevant for social SCRM, they have been summarized by Grimm, Stölzle, and Hofstetter (2013): In risk identification a detailed understanding of the complete supply chain is necessary (Boyd et al. 2007; Gardner and Cooper 2003; Grimm, Stölzle, and Hofstetter 2013). Risk identification requires the integration of external information including NGO and press news (Foerstl et al. 2010; Grimm, Stölzle, and Hofstetter 2013).[13] Furthermore, a supplier database containing all risk-relevant information (e.g., audit data, products, etc.) is needed for active monitoring (Foerstl et al. 2010; Grimm, Stölzle, and Hofstetter 2013; Cousins, Lamming, and Bowen 2004). Sustainability risk assessments require a supplier prioritization based on the risk of non-compliance while considering different factors such as geography, products, and production (Foerstl et al. 2010; Grimm, Stölzle, and Hofstetter 2013; C. Reuter et al. 2010; Carter and Rogers 2008; Spence and Bourlakis 2009).[14] Especially the latter is often lacking for companies (Anderson and Anderson 2009).[15]

[13] This monitoring of external factors in the context of risk may also be seen as a form of environmental scanning concerned with monitoring a company's environment (also on a supply chain level) in order to detect decision-relevant factors. While the importance of scanning the environment has been discussed before, the sustainability aspect has only been added very recently (Fabbe-Costes, Roussat, and Colin 2011). Fabbe-Costes et al. (2011; 2014), however, define a broad coverage of sustainability-related factors in the environment that may influence sustainability decisions and that may also have risk implications but remain relatively high level (e.g., regulatory changes or technologies).

[14] This paragraph provides the summarized source provided by Grimm, Stölzle, and Hofstetter as well as the original sources for further reference.

[15] This is in line with a different definition of risk management, according to which it "[…] is fundamentally about making decisions – decisions about which risk issues are more critical (prioritization), which risk issues

Finally and connected with social sustainability, risk approaches have also been proposed in combination with auditing efforts. Cahill (2011) suggests to time environmental, health, and safety audits in situations with multiple locations based on the risk associated with a specific location. Similarly, the Australian government (2010) proposes using risk levels to determine audit frequency. Rönninger and Holmes (2009) adopt a risk-based approach to the frequency and depth of audit plans. Hence, a risk- and context-aware approach to auditing frequency can be an effective technique for optimizing auditing activities.

Altogether, the novel model integrated into an IT system discussed in this thesis may provide help in important steps of the supply chain risk management cycle by improving supplier prioritization while including external information. It could help with the quantitative assessment of social sustainability risks in supply chains.

2.1.3 Related approaches to managing social sustainability risks in supply chains

While some researchers have addressed sustainability risk management directly, others have discussed the topic in the context of monitoring suppliers for supplier selection and supply chain performance management.[16] Both can be seen as multiple-criteria decision making tasks (Bhutta and Huq 2002) in a sustainability context (Öztürk and Özçelik 2014). Indeed, sourcing decisions should already include risk management thinking (Harwood and Humby 2008). Therefore, and given that the aim of this work is to construct a quantitative model for social sustainability risks in supply chains, this brief survey will concentrate on academically published quantitative models that focus on social criteria and evaluate individual suppliers rather than supply chains from a macroeconomic perspective. The focus on supply chains seems appropriate, as including external input specifically makes sense if the context of an organization is unknown.

The survey of the literature[17] produced a list of 41 papers (see Table 3). For more general overviews of sustainable supply chain supplier selection and performance management without this particular focus, we refer to other authors with recent surveys (Anthony Alexander,

are not worth worrying about (risk acceptance), and how much to spend on the risk issues to be dealt with (budgeting).” (Dobson and Hietala 2011, 7)

[16] For the general domain of supplier selection, a multitude of frameworks have been proposed (see W. Ho, Xu, and Dey 2010 for an overview).

[17] The survey is based on searches in ScienceDirect (http://www.sciencedirect.com/), ProQuest (http://search.proquest.com), Web of Knowledge (SCI-Expanded, SSCI; http://apps.webofknowledge.com), Ebsco (http://web.ebscohost.com/), and expanded research in cross-references.

Helen Walker, and Mohamed Naim 2014; M. Brandenburg et al. 2014; Taticchi, Tonelli, and Pasqualino 2013; Taticchi et al. 2014; Seuring 2013).

Author	Used for	LCA	Bayes	Delphi	AHP	ANP	TOP-SIS	DEA	QFD	Rough set	Grey system	Other	News*	Probabil-istic**	Special focus
Ahi and Searcy (2014)	Performance											Stochastic modeling***	No	Yes	-
Amindoust et al. (2012)	Supplier selection											Fuzzy inference	No	No	-
Azadi et al. (2015)	Performance							F					No	No	Efficiency/effectiveness
Azadnia et al. (2012)	Supplier selection				F		X					Neural network (self-organizing map)	No	No	Intermediate clustering of suppliers
Azadnia, Saman, and Wong (2015)	Supplier selection				F							Optimization model	No	No	Incl. order lot-sizing
Bai and Sarkis (2010)	Supplier selection									X	X		No	No	-
Bai and Sarkis (2012)	Performance							X		X			No	No	-
Bai and Sarkis (2014)	Performance							X		X			No	No	Key performance indica-tors
Bai, Sarkis, and Wei (2010)	Supplier selection									X			No	No	-
Baskaran, Nachiappan, and Rahman (2012)	Performance										X		No	No	Indian textile suppliers
Büyüküzkan and Çifçi (2011)	Supplier selection					F							No	No	Incomplete preference relations
Chiou, Chou, and Yeh (2011)	Supplier selection				F								No	No	Electronics industry
Dai and Blackhurst (2012)	Performance				X				X				No	No	Incl. voice of company stakeholders
Ding (2005)	Performance											Pair wise evaluation	No	No	Construction
Dou and Sarkis (2008)	Performance					X							No	No	Outsourcing and offshor-ing
Erol, Sencer, and Sari (2011)	Performance											Fuzzy entropy, fuzzy multi-attribute utility	No	No	Incl. an alert system
Foran et al. (2005)	Performance	(x)										Input-output analysis	No	No	High-level
Fuge, McKinstry, and Ninomiya (2013)	Performance											Uncertainty propagation	No	Yes	-
Godfrey and Manikas (2012)	Supplier selection				X								No	No	In-class exercise
Gopal and Thakkar (2014)	Performance				F							Liberatore score, sig-nal-to-noise ratio	No	No	Automobile industry
Govindan et al. (2012)	Performance						F						No	No	-
Hsu, Ou, and Ou (2015)	Performance						X				X		No	No	-
Hutchins and Sutherland (2008)	Performance	(x)										Input-output analysis	No	No	-
Jakhar (2015)	Performance				F							Structural equation modeling (SEM), fuzzy multi-objective linear programming (MOLP)	No	No	Apparel industry
Krajnc and Glavič (2005)	Performance				X								No	No	-
Liu, Wang, and Liu (2012)	Performance	(x)				X							No	No	-
Mani, Agrawal, and Sharma (2014)	Supplier selection				X								No	No	India
Nikolaou, Evangelinos, and Allan (2013)	Performance											Weighted sum	No	No	Reverse logistics
Öztürk and Özçelik (2014)	Supplier selection						F						No	No	-
Reefke, Sundaram, and Ahmed	Performance											Maturity model	No	No	-

(2010)															
Sarkis and Dhavale (2014)	Supplier selection		X									Monte Carlo Markov Chain, Gibbs sampler	No	Yes	-
Schmidt et al. (2013)	Performance	(x)										Weighted indicator	No	No	-
Shokravi and Kurnia (2014)	Performance											Weighted sum	No	No****	-
Tajbakhsh and Hassini (2014)	Performance							X					No	No	Incl. duality theory
Verdecho, Alfaro-Saiz, and Ro-driguez-Rodriguez (2014)	Performance				X								No	No	-
Wen, Xu, and Wang (2013)	Performance						X					Intuitionistic fuzzy sets	No	No	-
Wittstruck and Teuteberg (2012)	Supplier selection				F		X						No	No	Recycling partners
Wu, Hsieh, and Chang (2013)	Supplier selection			F		X	X						No	No	-
Xu et al. (2013)	Supplier selection				X								No	No	-
Yakovleva, Sarkis and Sloan (2012)	Performance				X								No	No	Food industry, industry level

Abbreviations: AHP (analytical hierarchy process), ANP (analytical network process), TOPSIS (technique for order preference by similarity to ideal solution), DEA (data envelop analysis), QFD (quality function deployment); “X” indicates that the approach has been used in the paper and “F” stands for a fuzzy approach. Items in parenthesis are only broadly covered.

* No direct integration of news cases in models considered/shown; ** Probabilistic approach to ranking; *** No weighting technique suggested; **** For social perspective only.

Table 3: Overview of publications on sustainable supplier selection and performance assessment models with a focus on social sustainability (author’s representation)

Different approaches have been identified for ranking suppliers for selection (de Boer, Labro, and Morlacchi 2001; C. S. Tang 2006), including models using weighting, total cost of ownership (TCO), mathematical programming, statistics, simulation, and artificial intelligence (AI). In the context of social sustainability, 9 of the 14 papers identified build on either the analytical hierarchy process (AHP) or the derived analytical network process (ANP), which can be used to combine qualitative and quantitative measures. The majority employ fuzzy approaches (Azadnia et al. 2012; Azadnia, Saman, and Wong 2015; Büyüközkan and Çifçi 2011; Chiou, Chou, and Yeh 2011; Wittstruck and Teuteberg 2012; Amindoust et al. 2012) while the rest do not (Godfrey and Manikas 2012; Mani, Agrawal, and Sharma 2014; C.-M. Wu, Hsieh, and Chang 2013; Xu et al. 2013). Other approaches use rough sets and the Technique for Order Preference by Similarity to Ideal Solution (TOPSIS) as their major contribution (Öztürk and Özçelik 2014; Bai and Sarkis 2010; Bai, Sarkis, and Wei 2010). Only one paper considers a probabilistic approach. Sarkis and Dhavale (2014) introduce an initial approach for sustainable supplier selection using Bayesian inference (without a network). They combine a series of criteria reflecting sustainability, norm them, and apply Gibbs sampling.

A number of papers focus on performance management or at least explicitly consider it. Overall, less using fuzzy approaches, most papers again employ AHP or ANP in their suggestions (Dai and Blackhurst 2012; Gopal and Thakkar 2014; Jakhar 2015; Krajnc and Glavič 2005; Yakovleva, Sarkis, and Sloan 2012; Verdecho, Alfaro-Saiz, and Rodriguez–Rodriguez 2014; Dou and Sarkis 2008). Two authors combine them with further, more complex techniques (Jakhar 2015; Gopal and Thakkar 2014). Another two suggestions rely entirely on data envelopment analysis (DEA) (Azadi et al. 2015; Tajbakhsh and Hassini 2014), while Bai and Sarkis (2012; 2014) combine it with grey theory, which is also used by others (Baskaran, Nachiappan, and Rahman 2012; Hsu, Ou, and Ou 2015). Hsu, Ou, and Ou combine grey theory with TOPSIS which is also included by two further authors (Govindan, Khodaverdi, and Jafarian 2012; Wen, Xu, and Wang 2013). The remaining approaches include weighting (Nikolaou, Evangelinos, and Allan 2013), pairwise comparison (Ding 2005), and a maturity model (Reefke, Sundaram, and Ahmed 2010).

Some papers are particularly relevant in the context of this work. However, they do not discuss the continuous integration of public information from news sources. In addition to an overall indicator representing sustainability performance, Erol, Sencer, and Sari (2011) present a yearly sustainability alert system employing thresholds. Fuge, McKinstry, and Ninomiya (2013) include probabilistic measures of sustainability integrating different private and public indicators via a weighting. They replace unknown specific indicators with more general indicators while respecting the additional variance this brings. In a comparable sense Ahi and Searcy (2014) suggest a stochastic model to measure the sustainability performance of a supply chain. They build on factors that support and hinder sustainability in supply chains and quantify a supply chain's progress towards sustainability. Factors are combined using weighting. Shokravi and Kurnia (2014) include an importance parameter, which is based on analyzing texts in order to derive a weight for each measure. This is different from using texts as an input for the model.

Other performance related papers specifically integrate data from a technique called social lifecycle assessment (SLCA), but still introduce a model for a combined indicator (Foran et al. 2005; Hutchins and Sutherland 2008; S. Liu, Wang, and Liu 2012; Schmidt et al. 2004). SLCA originates from environmental lifecycle assessments. Lifecycle assessments are a specific form of assessment that try to calculate a product's impact on the environment over its entire lifetime. An SLCA is concerned with "[…] the assessment of social and socio-economic impacts of products life cycle." (UNEP 2009, 5). Multiple authors have contributed to this topic (e.g. Dreyer, Hauschild, and Schierbeck 2005; Dreyer, Hauschild, and Schierbeck 2010a; Dreyer, Hauschild, and Schierbeck 2010b; Dreyer and Hauschild 2006; D. Hunkeler and Rebitzer 2005; David Hunkeler 2006; Klöpffer 2008; Jørgensen et al. 2008; Jørgensen et al. 2008; Gauthier 2005; Shuaib et al. 2011). A publication from the United Nations Environmental Program combines these ideas and gives practical advice in line with ISO standards (UNEP 2009). One important remit of an SLCA is the identification of social hotspots that may pose a significant risk to a company. These hotspots could be used to choose locations for on-site visits and where secondary data is sufficient (UNEP 2009, 40, 50, 59–68). They are strongly driven by statistical and manually derived outside-in data. In general, SLCA goals share many similarities with a risk management approach. However, due to the duration of an assessment, revolving updates seem difficult. Moreover, its orientation towards general "endpoints" (i.e., societal effects such as human health) is focused on the sustainability of the overall supply chain rather than the comparative value of a single supplier (Hutchins and Sutherland 2008). The necessary causal relationships for SLCS are often difficult to identify and, hence, SCLA output values at the current stage of research are of limited use for supplier management. In any case, they are not risk and probability oriented. The Life Cycle Index can also be categorized as an SLCA tool (F. I. Khan, Sadiq, and Veitch 2004).

One approach needs to be particularly mentioned even though it is concerned with single economic entities. Balanced scorecards (BSC) can be used to depict all three areas of the triple bottom line of sustainability (Elkington 1998) together. Multiple authors (see E. G. Hansen and Schaltegger 2014 for a rigorous review) have suggested sustainability BSCs (SBSCs).[18] A diverse number of structures and integration patterns have been proposed for integrating sustainability into classical BSCs. SBSCs provide an overview of targets and indicators from different perspectives, including sustainability, with cause-and-effect links between them (E. G. Hansen and Schaltegger 2014). They are not necessarily supplier related, but measuring the indicators included in the sustainability view may require involving the supply chain or may be tailored to the supply chain (Bhagwat and Sharma 2007; Hendrik Reefke and Mattia Trocchi 2013). However, SBSCs are not further investigated in this thesis as they neither discuss an aggregation technique of evidence sources for sustainability measures required for ranking suppliers, nor are tailored to the evaluation of suppliers.

2.1.4 Summary

This thesis addresses the problem of how to better monitor suppliers with regard to social sustainability risks while integrating external news. This requires tracking of risk-related events or measuring the performance of indicators related to this dimension. Although multiple papers have been published dealing with supply chain sustainability risk management and socially sustainable supplier selection and performance management, to the author's knowledge no paper has addressed the question of how to model the continuous data integration of evidence from public data sources such as news articles explicitly in order to monitor sustainability risks. The suggested models typically assume the availability of verbal or quantitative indicators. While evidence from news might be represented in such a way, it is not explicitly included. In general, probabilistic approaches are rare. Fuge, McKinstry, and Ninomiya (2013) depict an initial Bayesian-based approach to measure social performance. More recently, Sarkis and Dhavale (2014) have considered a Bayesian model for supplier selection.

2.2 Text mining for risk event detection and extraction

Text mining (TM) "[…] is concerned with information learning from preprocessed text […]" (F. Hogenboom et al. 2011, 48) and hence covers a wide array of different academic fields not only related to IT. It is increasingly seen as an important tool for data analysis (Ghadge, Samir, and Kalawsky 2012, 12). Unstructured text mining has already been identified as an opportunity to assess and sense general supplier risks (Cecere 2011). This thesis draws from several of these research fields. The following paragraphs will argue for the selection of the relevant fields and provide an overview without a focus on risk event detection.

[18] Bieker and Waxenburger (2002) even suggest extending this view by calling for an "Integrity Scorecard" to foster stakeholder integration.

The general aim of TM as applied in this thesis is to find child labor incidents related to supplier locations (or other company locations) from online news. Several tasks that need to be fulfilled by the system can be derived from this specification. As discussed in the introduction, the main inputs are news texts in English. Extracting child labor incidents from news sources is linked to several tasks. On the one hand, the process of extraction refers to event detection (ED) and particularly new event detection (NED) as "[…] an event should identify something (non-trivial) happening in a certain place at a certain time." (Yang, Pierce, and Carbonell 1998, 29). In ED, additional metadata is enriched in order to be able to state "[W]ho did what to whom, when, with what methods (instruments), where and possibly why." (Tanev, Piskorski, and Atkinson 2008, 207 italics removed). Here, events need to be connected (i.e., related) with supplier locations carrying attributes such as sector or geographic location. For this purpose the "frame" of an event needs to be extracted from a text referring to a child labor incident. The "frame" contains the attributes of an event and includes, for example, the geographical reference of the event. Consequently, this task also relates to information extraction (IE) and more specifically to event extraction (EE). Each event may be related to a supplier location. Additionally, detecting new events is important in order to be able to differentiate the boundaries between the depictions of multiple events. Both fields (ED and EE) are important sub-fields of IE, which will be discussed below. While we understand child labor incidents as events, they may also be understood as "stories". These are a core element of the tasks used in topic detection and tracking (TDT), which will also be described below. On the other hand, extracting child labor incidents can be seen as a classification task with known classes where a child labor incident is one class. Consequently, we will also refer to text classification. The final section will additionally cover other related approaches that are only partly related to or not used in this thesis.

2.2.1 General introduction to information extraction

IE is the relevant field of TM for this thesis. Once concerned with "template filling," IE has a long history going back into the 1960s (Gaizauskas and Wilks 1998, 73). A key series for IE was the Message Understanding Conferences (MUC), held until 1995, which provided a standardized test setup for general approaches to IE (Association for Computational Linguistics 1995).[19] For Hobbs, an IE system is "[…] a cascade of transducers or modules that at each step add structure and often lose information, hopefully irrelevant, by applying rules that are acquired manually and/or automatically." (Hobbs 1993, 87). Thus, in this view the core of IE

[19] These conferences were followed by other regulated evaluation tasks such as the Automatic Content Extraction (ACE) Evaluation (National Institute of Standards and Technology 2014a) or the Text Analysis Conference (National Institute of Standards and Technology 2014b).

is deriving structured information from text. Another definition of IE by Piskorski and Yangarber underlines the importance of domain tailoring in IE: “The task of Information Extraction (IE) is to identify a predefined set of concepts in a specific domain, ignoring other irrelevant information, where a domain consists of a corpus of texts together with a clearly specified information need.” (Piskorski and Yangarber 2013, 24). As a special case, ontology-based IE is concerned with linking extracted items to ontological concepts, allowing cross referencing and reasoning. It thus adds more semantic information to a text (D. Maynard et al. 2005). For a detailed introduction to IE, we refer to the surveys by Appelt (1999), Kaiser and Miksch (2005) as well as Piskorski and Yangarber (2013).[20]

In detail, the aim of IE “[…] is to identify instances of a particular prespecified class of entities, relationships and events in natural language texts, and the extraction of the relevant properties (arguments) of the identified entities, relationships or events.” (Piskorski and Yangarber 2013, 26). Early approaches to IE used rule- and expert-based approaches to perform the extraction as they built on earlier knowledge engineering. Subsequent systems often relied on finite-state approaches, which can be combined with the analysis of deeper linguistic frameworks (Piskorski and Yangarber 2013). Particularly modularization, including the abstraction of knowledge bases, emerged and strengthened over time (Piskorski and Yangarber 2013, 35). As a result, IE systems typically exhibit a special structure. Appel (1999, 11) summarized this structure as a four-step process, shown in Figure 4.

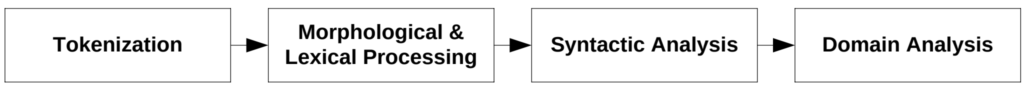


Figure 4: Typical structure of IE systems (Appelt 1999, 11)

The 1990s witnessed an increase in approaches building on “shallow” instead of “deep” parsing (Gaizauskas and Wilks 1998, 92–94). Typically, “deep” approaches include a higher amount of linguistic features that are computationally more intensive to gather. Efforts have been made to shift rule generation to computational resources employing machine learning (ML) approaches already applied in natural language processing (NLP). Further enhancements include multi-language approaches and adaptations to reduce ML tagging requirements (Piskorski and Yangarber 2013). In recent years, IE systems based on ML have achieved similar results to expert-defined extraction systems (Y. Li, Bontcheva, and Cunningham 2005a).

[20] A particularly broad array of work is related to stock trading and employing text mining for fast trading decisions (see Mittermayer and Knolmayer 2007 for a survey).

Pikorski and Yangarber (2013) differentiate four related IE tasks. Named entity recognition (NER) recognizes different types of entities within an unstructured text (e.g., organizations, persons, etc.) and disambiguates them. Coreference resolution (CO) reuses these named entities and tries to relate different “mentions” of the same entity. This includes acronyms, references by pronouns, nominal phrases, and implicit references. Relation extraction (RE) uses NER and CO to extract predefined relationships between entities. Finally, event extraction (EE) is a meta-task combining the different techniques to resolve the extraction of events from texts. EE is usually the most difficult of the four tasks. We will discuss this final approach in more detail below after presenting specific IE approaches to sustainability risk management.

2.2.2 Information extraction for sustainability risk management

Text mining approaches for longer texts tailored to sustainability are rare and can only be related to risk management to a limited extent. Typically, they are concerned with volume, i.e., the amount of news published, tracking the movement of global, regional, or local news on a specific topic. Barkemeyer et al. (2009) present a system which tracks different sustainability topics over a wide array of news outlets. Subsequent studies provided even more details on the behavior of news on sustainability (Holt and Barkemeyer 2012; Barkemeyer, Figge, and Holt 2013). Similarly, the Carbon Capture Report aggregates news on several globally relevant sustainability topics, such as solar power, in order to provide an overview of the current discourse on carbon capture (University of Illinois 2012). Another related toolset is provided by the Media Watch on Climate Change, depicting developments in climate change (Scharl et al. 2013) with a focus on dominant events. Rivera et al. (2014) extend this work by applying text mining to monitor sustainability indicators for particular regions, including societal values and underlying causes while approaching multiple sustainability areas. They do not include a risk model.

More generically, Zheng and Zhou (2012) suggest using text mining to assess the corporate governance of a company based on a predefined set of questions related to typical answer phrases in company filings. Other authors apply text mining to analyze corporate sustainability reports, which can also be used in a competitive environment (Modapothala and Issac 2009; Modapothala, Issac, and Jayamani 2010). On a meta-level, text mining can be used to assess academic literature to draw conclusions about sustainability. Approaches like this have been published for environmentally unhealthy chemicals (Yong Chen et al. 2014) or energy efficiency, supporting environmental sustainability (N. Reuter et al. 2014). A similar direction may be attributed to the work of Ji et al. (2010), who annotate scientific research on carbon sequestration to support research into the extraction of scientific knowledge about the reduction of carbon emissions. Finally, and connected with the former, Ghadge, Samir, and Kalawsky (2012) employed text mining in their review of SCRM.

Several authors have connected text mining and risk management. While none of them focus on sustainability directly, some illustrative, closely related approaches (i.e. techniques) have been presented. Aggour, Interrante, and Gokcen (2006) describe an IE system that can be used to monitor business risk. Capet et al. (2008) suggest an IE system called ADAC to monitor news streams for obtaining possible risk signals. Monitoring is based on expert scenarios of how a risk situation typically evolves, based on templates depicting them. They feed events into the recognition engine to update a “recognition degree” (Capet et al. 2008, 4) that serves as a measure of confirmation for a scenario hypothesis. EE is performed with syntactic analysis (Nakamura et al. 2007). Sándor (2009) focuses on the strategic early warning possibilities of text risk event mining. Raskin et al. (2008) depict an approach for integrating text mining into risk management at the level of national security and intelligence, focused on making text information accessible.

Related risk detection can also be performed at an inter-country level, for example when detecting serious conflicts (Stoll and Subramanian 2006) or estimating country risk (Lianghui, Wang, and Jia 2012). Fraud can pose similar reputational risks to a social sustainability scandal. LaComb, Interrante, and Aggour (2007) suggest using deliberative learning to monitor and potentially detect such events, Holton (2009) analyzes disgruntled communication and Humpherys et al. (2011) deal with affected financial statements. Leidner and Schilder (2010) suggest an approach for mining business risks from text using an IT tool. They do not classify the risk in a more detailed way and the risk identification is not considered in a supply chain specific context or integrated into a risk model. Finally, some authors ignore the question of identifying risks according to risk types, instead focusing on the extraction of risk types themselves. For example Moriizumi et al. (2011) have developed such a system for the supply chain domain.

In summary, none of these approaches concentrate on uncovering sustainability-related low frequency events from the perspective of a company facing risk in its supply chain. Moreover, the text mining approaches related to sustainability do not integrate findings into an overarching risk model with incremental updates. The closest related approach, by Capet et al. (2008), can produce a risk scenario in environments with a significant amount of news and very direct event chains. However, as social sustainability issues display very loose event chains (as depicted in Section 2.3), Capet et al.'s approach appears not or only partly applicable.

2.2.3 Event detection and event extraction

Event text mining or EE is concerned with extracting events from text, more precisely with the differentiation of unique events out of a set of articles where more than one article may relate to a single event. EE is defined as "[…] identifying events in free text and deriving detailed and structured information about them, ideally identifying who did what to whom, when, where, through what methods (instruments), and why." (Piskorski and Yangarber 2013, 27). This goes beyond a pure classification of a text and aims at extracting event types with their properties from a given textual source. Some authors see EE as a successor to text summarization, which is concerned with condensing a text to its most important message(s) (Lam and Ho 2001).

2.2.3.1 Introduction

An event can be seen as a special form of the data structures introduced in the last section, where the events' attributes represent the facts and a single event is an object (previously "schema" has been used; McKeown 1992, 26). Deriving events from news is also a way to make news more accessible for different purposes (W. Chen et al. 2010) – be it for manual inspection or automated mining application (e.g. Lin and Liang 2008; Maybury 1995).

Before an event can be extracted it needs to be detected. ED in its narrowest form translates to spotting events referenced with event anchors. EE often implicitly includes ED. ED can either be performed online as new information arrives (e.g. Chih-Ping and Yen-Hsien 2004) or retrospectively on a fixed set of entries (e.g. Zhiwei Li et al. 2005). The task of detecting an event-like item is also an important part of topic detection and tracking (TDT).

Multiple approaches to mining events have been proposed. They may be classified into data-driven approaches, knowledge-driven approaches, and hybrid approaches combining the two (F. Hogenboom et al. 2011). Data-driven approaches apply statistical techniques like counting frequencies, computing clusters, or machine learning in order to extract events (e.g. Piskorski et al. 2011; Smith 2002), while knowledge-driven approaches rely mostly on expert-defined patterns and rules (e.g. Atkinson et al. 2008). In some cases rules or patterns are learned. Machine-learning based approaches need a training and a test set, which is less and less feasible as the size of the different categories increases (Steinberger, Pouliquen, and van der Goot

2009). When transferring a pattern-based approach to other languages, limitations may arise due to differences in word orders and the general complexity of descriptions (Piskorski and Yangarber 2013). Hybrid approaches combine both techniques. For example, high-level features may be processed based on rules while low-level features are processed using machine-learning (Botsis et al. 2011). Thereby, event extraction can be domain-specific (Piskorski and Yangarber 2013). Furthermore, semantics can be used to support event detection in multiple ways (Xie, Sundaram, and Campbell 2008): to capture multilevel semantics and event polysemy, to deal with implicit semantics and hidden facets, to include context, and to deal with the sensory gap mentioned above.

In summary, there are multiple approaches to ED and EE. The next subsection gives an overview of currently proposed systems combining the two domains.

2.2.3.2 Combined systems for new event detection and extraction from news texts

A significant number of systems have been proposed in academic literature that jointly detect and extract events from news texts. The survey presented here extends the one by Hogenboom et al. (2011) to newer articles and presents a different view of the approaches which is more relevant for this thesis. Papers specifically dealing with the geographic or time dimensions are excluded, as well as those which do not focus on ED or EE. Relevant multilingual approaches presented by Danilova, Alexandrov, and Blanco (2014) are also integrated. The focus of this overview is more recent EE literature (from 2000 onwards) in order to underline the growing importance of the field and to include work that covers the period of high and growing interest in SSCM (Seuring and Müller 2008b).[21] In contrast to earlier work, this survey limits the approaches to those applied to English texts composed of multiple full sentences. Moreover, papers are only included if they combine ED and EE and are not solely retrospective. We also skip papers which use very domain-specific text as input.[22] Finally, we do not include systems that focus on specific, narrow IE and EE topics such as template structure. Given that academically published systems are typically covered in multiple papers, one representative publication has been chosen as reference for each system. All systems combine detection and extraction approaches. This extensive survey identifies 37 papers that meet these criteria. These are presented in Table 4, which includes several detailing categories (up to December 2014).

[21] Of 191 papers analyzed by Seuring and Müller (2008b, 1701), 156 (82%) were published in or after 2000.
[22] This is particularly the case for medicine (e.g. Botsis et al. 2012; Melton and Hripcsak 2005; Zuofeng Li et al. 2010; Kang et al. 2014) and texts dealing with biomolecular topics (Van Landeghem et al. 2011; Z. Lu 2011; e.g. Gerner et al. 2012). They also often have a different definition of an event (Gerner et al. 2012, 2154).

As mentioned above, Hogenboom et al. (2011) distinguish between data-driven, knowledge-driven, and hybrid forms of event detection. This distinction has been retained for this survey, and we indicate whether the authors use the full text or not and whether they extract events from entire documents or on a sub-document basis. The event detection approach differentiates between rules, learned rules, and learned classes, additionally indicating whether the classes are pre-determined before the detection is applied. The extraction approach differentiates along the dimensions suggested by Sarawagi (2007), i.e., rule-based and statistical extraction as well as hand-coded and learned extraction. Finally, the dependency parsing column indicates whether dependency parsing is applied for EE.

				General		Event detection & extraction							
Authors	Year	Hogen-boom	Dani-lova	Domain	Name	Technique	Text used	Level of event	Event detection approach	Known classes	Rule-based vs. statisti-cal extract.	Hand-coded vs. learned extraction	Depen-dency parsing
Aggour, Interrante, and Gokcen	2006			Financial	E-BIG	Hybrid	Full	Sub-document	Learned classification	Yes	Rules	Coded	Yes
Ahn	2006			General news	-	Data	Full	Sub-document	Learned classification	Yes	Statistical	Learned	No
Aone and Ramos-Santacruz	2000	x		General news	REES	Knowledge	Full	Sub-document	Rules	Yes	Rules	Coded	No
Arendarenko and Kakkonen	2012			Business events	BEECON	Knowledge	Full	Sub-document	Rules	Yes	Rules	Coded	No
Avornicului, Popa, and Avornicului	2010			General news	-	Hybrid	Full	Sub-document	Rules	Yes	Statistical	Learned	No
Black et al.	2005			General	CAFETIERE	Knowledge	Full	Sub-document	Rules	Yes	Rules	Coded	No
Capet et al.	2008	x		Early warning/ Nuclear Proliferation	EventSpotter/ ADAC	Knowledge	Full	Sub-document	Rules	Yes	Rules	Coded	Yes
Castellanos et al.	2012			General	SIE-OBI	Hybrid	Full	Sub-document	Multiple	Yes	Multiple	Multiple	No
Coppola et al.	2009			General	-	Hybrid	Full	Sub-document	Learned rules	Yes	Rules	Learned	No
Draicchio et al.	2013			General	FRED	Knowledge	Full	Sub-document	Rules	Yes	Rules	Coded	Yes
Dutkiewicz, Nowak, and Jedrzejek	2014			General	R2E	Knowledge	Full	Sub-document	Rules	Yes	Rules	Coded	Yes
Elloumi et al.	2013			Financial (manage-ment change)	-	Hybrid	Full	Sub-document	Learned rules	Yes	Rules	Learned	No
Freifeld et al.	2008			Disease outbreaks	HealthMap	Knowledge	Full	Document	Rules/Alert-input	Yes	Rules	Coded	No
Glavas and Snajder	2014			General	Event Graphs	Hybrid	Full	Sub-document	Learned classification	Yes	Rules	Coded	No
Grishman, Hutunnen, and Yangarber	2002		x	Diseases	PULS	Knowledge^	Full	Document	Rules	Yes	Rules	Coded	No
Grishman, Hutunnen, and Yangarber	2002			Disease outbreaks	Proteus-BIO/ExDisco	Hybrid	Full	Sub-document	Learned rules	Yes	Rules	Learned	No
Hayes and Nardulli	2011			Societal stability	SPEED	Hybrid	Full	Sub-document	Learned classification	Yes	Manually supported		No
Hecking and Sarmina-Baneviciene	2010		x	Intelligence	ZENON	Knowledge	Full	Sub-document	Rules	Yes	Rules	Coded	No
Hogenboom et al.	2013			Financial	SPEED	Knowledge	Full	Sub-document	Rules	Yes	Rules	Coded	No
Hung, Lin, Hong	2010	x		Commonsense knowledge	-	Knowledge	Full	Sub-document	Rules	Yes	Rules	Coded*	-
Ijntema et al.	2012			General news	HIEE	Knowledge	Full	Sub-document	Rules	Yes	Rules	Coded	No
Inyaem et al.	2009			Terrorist events	-	Data	Full	Document	Learned classification	Yes	Statistical	Learned	No
Ji and Grishman	2008			General news	-	Data	Full	Sub-document	Learned classification	Yes	Statistical	Learned	No
Jungermann and Morik	2008	x		Parliament	-	Hybrid	Full	Sub-document	Learned classification	Yes	Statistical	Learned	No
Lam and Ho	2001			Financial	FIDS	Knowledge	Full	Sub-document	Rules	Yes	Rules	Coded	No
Leetaru and Schrodt	2013			General	TABARI/GDELT	Knowledge	Full	Sub-document	Rules	Yes	Rules	Coded	No
Liao and Grishman	2011			General news	-	Data	Full	Sub-document	Learned classification	Yes	Statistical	Learned	No
Liu et al.	2008	x		General news	-	Data	Full	Document	Clustering	No	Statistical	Learned**	No
Liu et al.	2009			General news	-	Data	Full	Document	Clustering	No	Statistical	Learned***	No
Nishihara, Sato, and Sunayama	2009	x		Personal experiences	-	Knowledge	Full	Sub-document	Rules	Yes	Rules	Coded****	No
Padró et al.	2014		x	General	Xlike	Knowledge	Full	Sub-document	Rules	Yes	Rules	Coded	Yes
Patwardhan and Riloff	2009			General news	GLACIER	Data	Full	Sub-document	Learned classification	Yes	Statistical	Learned	No
Piskorski et al.	2010	x		Security	Frontex system (EUROSUR)	Hybrid	Full	Document	Clustering	Yes	See NEXUS/PULS		
Piskorski et al.	2011			Crisis	-	Hybrid	Full	Document	Clustering	Yes	See NEXUS/PULS		
Ploeger et al.	2013			Activists	Extractivism	Knowledge	Full	Sub-document	Rules	Yes	Rules	Coded	Yes
Radinsky and Horvitz	2013			General news	-	Data	Full	Document	Clustering	No	Statistical	Learned	No
Serrano et al.	2013			General news	-	Hybrid	Full	Sub-document	Learned rules	Yes	Rules	Learned	Yes (cust.)
Tanev, Piskorski, and Atkinson	2008	x		Violence, disasters	NEXUS, Ex-PRESS	Data	Title, top sent.	Sub-document	Clustering	No	Rules	Learned	No
Vargas-Vera and Celjuska	2004	x		General	-	Knowledge	Full	Sub-document	Learned rules	Yes	Rules	Learned	No
Yangarber, Grishman, and Tapanainen	2000			General	ExDisco	Hybrid	Full	Sub-document	Learned rules	Yes	Rules	Learned	No

* Generic, based on semantic role labeling; ** Online learning; *** Association rules; **** Partly based on weighting formula; ^ Bootstrapping partly possible

Table 4: Overview of approaches to combined event detection and extraction (author's representation)

Although apparently less sophisticated, rule-based EE approaches are still very common. Ten of twelve approaches published in 2012, 2013, and 2014 were rule-based (Arendarenko and Kakkonen 2012; Draicchio et al. 2013; Dutkiewicz, Nowak, and Jedrzejek 2014; Elloumi et al. 2013; Glavaš and Šnajder 2014; A. Hogenboom et al. 2013; IJntema et al. 2012; Leetaru and Schrodt 2013; Padró et al. 2014; Ploeger et al. 2013). One aspect that is particularly highlighted is processing speed (F. Hogenboom et al. 2011). Some rule-based systems learn patterns or rules based on a seed or other training set (Coppola et al. 2009; Elloumi et al. 2013; Grishman, Huttunen, and Yangarber 2002a; Tanev, Piskorski, and Atkinson 2008; Vargas-Vera and Celjuska 2004; Yangarber, Grishman, and Tapanainen 2000; Serrano et al. 2013). Similarly, dependency-parsing, which analyzes deeper phrase structures (going back to Tesnière 1959), is only used by seven systems (Aggour, Interrante, and Gokcen 2006; Capet et al. 2008; Draicchio et al. 2013; Dutkiewicz, Nowak, and Jedrzejek 2014; Padró et al. 2014; Ploeger et al. 2013; Serrano et al. 2013). In contrast, for ED most systems employ learning or clustering based approaches. In particular, learned classification is applied in ten cases (Aggour, Interrante, and Gokcen 2006; Ahn 2006; Castellanos et al. 2012; Glavaš and Šnajder 2014; Hayes and Nardulli 2011; Inyaem et al. 2009; H. Ji and Grishman 2008; Jungermann and Morik 2008; Liao and Grishman 2011; Patwardhan and Riloff 2009). Eight systems detect events on a document level (Freifeld et al. 2008; Grishman, Huttunen, and Yangarber 2002b; Inyaem et al. 2009; M. Liu et al. 2008; Piskorski et al. 2010; Radinsky and Horvitz 2013; Piskorski et al. 2011; D.-R. Liu et al. 2009). Several other systems utilize fine-grained rules that can detect a broader array of different events within one document.

Several papers must be analyzed in detail in light of the focus of this thesis. Aggour, Interrante, and Gokcen (2006) see extracting events as a means for business intelligence, an approach with is also adopted by Arendarenko and Kakkonen (2012), Hogenboom et al. (2013), and Lam and Ho (2001). Slightly differently, Castellanos et al. (2012) correlate data and events from external texts with internal reports for decision support. They argue that fusing data from different sources helps to make more informed and timely decisions. Capet et al. (2008; see also above) depict a system tailored to risk management that updates a score between zero and one depending on the evidence that arrives from news events. In partly comparable fields, several authors suggest using ED and EE for detecting and tracing diseases (Freifeld et al. 2008; Grishman, Huttunen, and Yangarber 2002a; Grishman, Huttunen, and Yangarber 2002b) or terrorist events (Inyaem et al. 2009).
However, none of the systems[23] particularly addresses relevant social sustainability events.

---

23 The six systems not mentioned in the following (but included in Table 4) are mainly (except the approach by Avornicului et al.) knowledge-based systems employing hand-coded event extraction rules (Aone and Ramos-

2.2.4 *Related approaches to event detection and extraction*

As well as ED and EE, several further research areas are at least partly related to this thesis and will be touched on to different degrees in the following chapter. TDT has a special definition of “events” and is often used to analyze broader topical evolutions over time. Text classification assigns texts to categories which might also be event categories. Separately, IR deals with the query-based detection of information. Further, less-detailed areas include surveillance systems, information fusion, and weak signal detection.

2.2.4.1 Topic detection and tacking

Formal rounds of TDT evaluation were sponsored by NIST starting in 1998. They lasted officially until 2004 (National Institute of Standards and Technology 2014c). These rounds were succeeded by the ACE program (National Institute of Standards and Technology 2014a), which was more strongly integrated into the field of IE as depicted above. Nevertheless, TDT shows strong relations to IE and especially ED, as it is concerned with the identification of topics consisting of multiple events.

A topic is “[…] a seminal event or activity along with all directly related events and activities.” (Fiscus and Doddington 2002, 2). Consequently, a topic also covers events that are not necessarily the incidental event. Allan (2002, 2) requires that topics relate to a specific place and time; however, this is not necessarily implied by the definition above. Topics are a grouping level for stories, initially defined as “[…] a topically cohesive segment of news that includes two or more declarative independent clauses about a single event.” (Fiscus and Doddington 2002, 2). Events in TDT are seen as “[…] something that happens at some specific time and place along with all necessary preconditions and unavoidable consequences.” (Fiscus and Doddington 2002, 3). Events are typically identified by two areas of text, one including the textual anchor (also called a “trigger”) of an event and one describing it (called a “mention”) (M. Naughton, Stokes, and Carthy 2010).[24] Given these definitions, two different interpretations can be derived when transferring them to our research area of child labor risks, as presented in Table 5.

Santacruz 2000; Avornicului, Popa, and Avornicului 2010; Black et al. 2005; Hecking and Sarmina-Baneviciene 2010; Hung, Lin, and Hong 2010; Nishihara, Sato, and Sunayama 2009).

[24] Several sub-topics have been defined for TDT that are of special interest (Allan et al. 1998; Fiscus and Doddington 2002). Segmentation aims to split a text stream into multiple independent segments (Fiscus and Doddington 2002). Tracking is concerned with identifying stories that relate to a predefined, known topic. Story detection aims to relate a text to an existing topic or start a new topic. Classes are not previously known (Fiscus and Doddington 2002). Finally, linking is concerned with identifying whether several stories are related to a similar topic or not (Fiscus and Doddington 2002).

Element	Broad interpretation	Narrow interpretation
Topic	Child labor incidents in general and related events/activities	One particular child labor incident and related events/activities
Story	Stories about any child labor incident	Stories about a particular child labor incident
Event	Child labor incident and related events in general	The particular child labor incident or the related events

Table 5: Broad and narrow interpretation of definitions of TDT when transferred to child labor detection (author's representation)

The narrower interpretation of TDT depicted in Table 5 is more closely related to the final aim of this thesis. However, most TDT approaches are designed to be useful especially if events cover multiple stories. This might not be true in the case of risk-related events. Conversely, applying the broader interpretation may conflict with the definition of Allan et al. (1998, 1), who would designate this a class of events.

Altogether, TDT's tracking approaches may prove useful for our task and will be referenced when needed in the subsequent chapters. However, TDT is based on unknown topics (Fiscus and Doddington 2002) and thus does not fully apply to our task, where risk classes are known.

2.2.4.2 Text classification

The study of automated text classification is not new.[25] Classified text items are often used as input elements for other models, some of which have already been discussed in the contexts of IE, EE, and TDT. Detecting specific events or topics is also a classification task. Text classification is a widely studied research field with multiple contributions over the years (A. Khan et al. 2010). One of the earliest works on text classification was by Maron in 1961 (Maron 1961). Different sub-areas have sparked intense discussions which will not be presented here in detail. Research is broadly split between document representation and the approach used for the classification task itself (A. Khan et al. 2010).

According to Hagenau (2013, 686), three elements of text classification need to be differentiated: the dataset, feature processing, and the machine learning element. As with event detection, text classification can be implemented using either rule-based or machine learning-based techniques. The latter in particular has been widely discussed in the literature, with a significant array of different methods available (Ikonomakis, Kotsiantis, and Tampakas 2005; A. Khan et al. 2010). There are two important steps to machine learning: selecting the right features as the representation of the document to be classified and choosing the most appropriate machine learning methodology. Among methodologies, Support Vector Machines (SVM) and

[25] For example, a review was conducted by Ikonomakis, Kotsiantis, and Tampakas (2005).

the *k*-nearest neighbors algorithm (KNN) are repeatedly named as highly successful for classification tasks (A. Khan et al. 2010; Yang and Liu 1999). Semantic technologies are also seeing increased use. For example, Frasincar, Borsje, Levering (2009) introduce the use of Semantic Web standards such as OWL and SPARQL together with semantic word sense and a semantic dictionary for text classification.

In summary, current research provides a significant number of tools for classifying documents. Nevertheless, these need to be tailored to the specific classification question and domain, and thus a specific evaluation for the classification of child labor related news appears necessary.

2.2.4.3 Information Retrieval

Information retrieval (IR) can be seen as one field with important links to ED which can also be clearly distinguished from IE and TDT. In contrast to IE, IR aims to select documents from a collection (Jusoh and Alfawareh 2012). Gaizauskas and Wilks describe the contrast as follows: “IR retrieves relevant documents from collections, IE extracts relevant information from documents.” (Gaizauskas and Wilks 1998). Consequently, IR is only partly relevant in the context of our discussion as it is concerned with establishing a ranked list of documents given a query, while IE deals with processing documents to extract facts for further use in other application scenarios. Nevertheless, both technological streams are intertwined and may support each other in different scenarios (Piskorski and Yangarber 2013).

Comparable to IE, TDT is also different from IR. Petrović (2012) identifies four differences which also partly apply to the tasks approached in this thesis. First, relevance is not used as a concept in TDT, as it is replaced by being either related to a topic or not. Moreover, no query is defined, but the set of stories to integrate is implicitly given through a set of predetermined stories, as in topic tracking. Second, instead of classical IR systems, TDT is applied in systems with professional users that require a stronger focus on recall. Third, processing needs to be online, incorporating each new story incrementally. Finally, TDT is often concerned with multi-source and multi-language heterogeneous input. Summing up, IR is only slightly related to this thesis.

2.2.4.4 Surveillance systems

The research on surveillance systems is also related to our topic. Surveillance systems usually output alerts to users, who may then perform an action (Denecke 2012, 48). In the case of this thesis the output of the analysis is not alerts, but the events detected are incorporated into a risk model for supplier locations. Therefore, alerts are generated but they refer to a supplier location rather than an individual event. Multiple events and further background knowledge (e.g., prior knowledge or audits) might be the trigger. “In surveillance, we are normally interested in getting an alert when a high number of counts appears in the data, i.e. when the upper control limit is exceeded.” (Denecke 2012, 48). Therefore, surveillance systems are often related to stronger, better detectable events than might be the case with risks.

2.2.4.5 Information fusion

Information fusion is defined as a way to “[…] remove redundancy, resolve contradictions and uncertainties by multiple information providers and design a general framework for the veracity analysis problem.” (H. Ji 2010). In the context discussed here, cross-document information fusion is of particular interest. Information gained from multiple texts and different pictures on a similar or identical event need to be combined for further processing. This fusion of facts has been studied in multiple domains and is not new (H. Ji 2010). Therefore, this work will draw from some of these relations in order to determine the best approach for the current context.

2.2.4.6 Weak signal detection

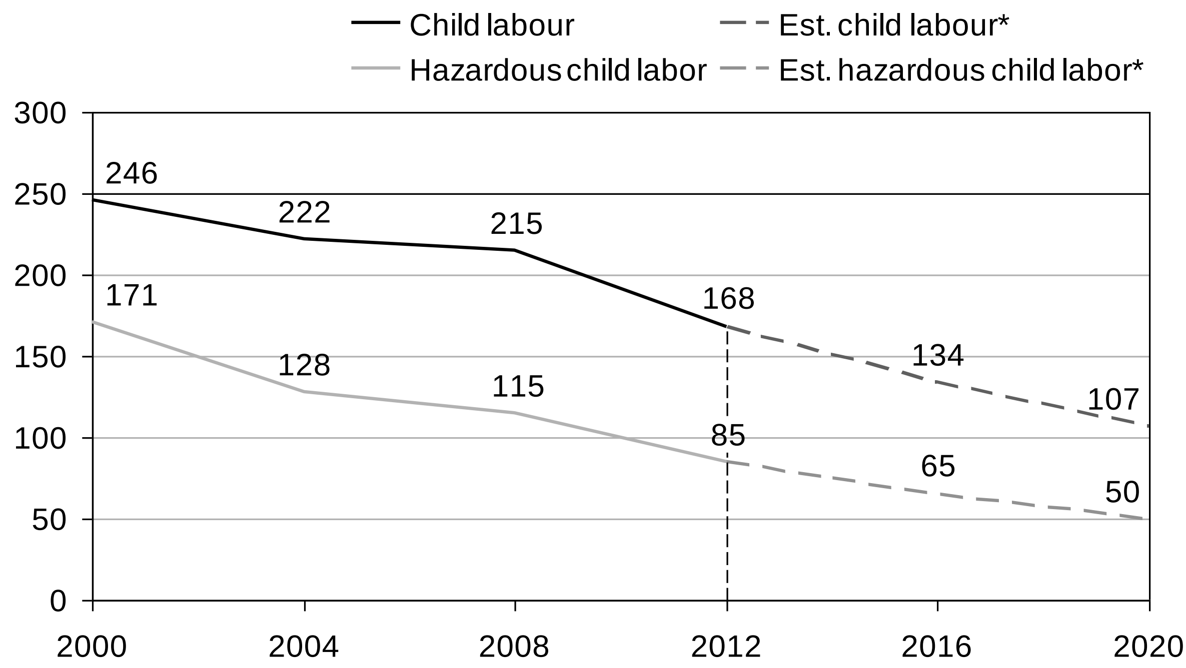
Finding child labor incident reports relating an entity to child labor can also be viewed as a weak signal detection task. Weak signals are defined as “[…] currently existing small and seemingly insignificant issues that can tell us about the changes in the future. In other words, they are today's clues and signs providing us with hints of the possible events and trends in the future.” (Hiltunen 2006, 62). Weak signal detection was introduced by Ansoff in 1975 (Ansoff 1975). In the best case, weak signals of child labor are detected before they are broadcast as “strong signals” and easily detectable. The task is to find the small number of documents that potentially contain these signals (Thorleuchter, Scheja, and Van den Poel 2014). For this purpose, researchers have proposed methods that employ text mining to detect documents carrying a predefined set of weak signal clues (i.e., related keywords) (Thorleuchter, Scheja, and Van den Poel 2014; Thorleuchter and Van den Poel 2013; Yoon 2012). Here, we will mainly focus on events. Given that a unique news event can also be relevant to risk ratings, this might be considered as weak signal as well.

2.2.5 Summary

A multitude of combined approaches to ED and EE have been presented. However, none of them are tailored to the research field of this thesis from a more generalized view of sustainable IE. Moreover, the performance of an EE system strongly depends on the type of event and the specific implementation, which can be at varying degrees of precision and recall rates for different types of events (Arendarenko and Kakkonen 2012; Coppola et al. 2009; IJntema et al. 2012; Van Landeghem et al. 2011; M. Naughton, Stokes, and Carthy 2010). This is further underlined by other authors stressing the relevance of domain specifics in ED and EE (Piskorski et al. 2010). As a result, dedicated research on social sustainability is suggested. This requires in parts a specific tailoring to the domain, particularly for ED. For EE the event frame needs to be adapted to the domain. For this, semantic differentiations are helpful. Consequently, chapter 5 will depict the approach used in this thesis, and will argue in detail about the use of different methods. This will specifically include the use of semantic approaches.

2.3 Child labor as a special case of social sustainability risks

Child labor is a historically relevant phenomenon which has been discussed for decades (Willoughby 1890). Child labor still exists. Although the measurement of an exact number of children working is hardly possible (Bourdillon et al. 2010, 19), a widely reported study was conducted by the International Labour Organization (ILO) and its International Programme on the Elimination of Child Labour (IPEC). According to these, in 2012, the total number of children working in any form of child labor was still 168 million worldwide. About 85 million of these children performed work that can be considered hazardous, e.g., work that includes physical abuse or the handling of dangerous machinery (International Programme on the Elimination of Child Labour 2013, 13). The total number of children working declined by roughly 32% between 2000 and 2012 and is expected to decline by another 36% by 2020. Nevertheless, child labor is still an immense problem. Figure 5 shows this overview and the expected reduction in upcoming years. Child labor is estimated to remain a major part of the work force, particularly in some regions. The predominant regions are Asia and the Pacific with 78 million children working and Sub-Saharan Africa with 59 million children working (International Programme on the Elimination of Child Labour 2013, 15). Nevertheless, surveys on child labor must be treated with caution, as they may have significantly different outcomes depending on the survey instrument and definition used (Guarcello et al. 2010).



* Evolution based on pace of progress during 2008-2012

Figure 5: Number of children in child labor and hazardous work, actual 2000-2012 levels and 2016 and 2020 levels assuming the pace of progress seen during 2008-2012 (based on International Programme on the Elimination of Child Labour 2013, 13)

Child labor is also associated with lower income countries, as has been statistically shown by several researchers. The level of child labor (below 14 years of age) and GDP per capita are negatively correlated (R^2 of 0.51) in over 84 countries (International Labour Office 2013, 12). Another earlier study showed a comparable result, with GDP per capita explaining 73 percent of the variation in the percentage of economically active children between 10 and 14 (Edmonds and Pavcnik 2005). While generally in line with these studies, a third study indicates a slight rise in child labor for high-income countries (Fares and Raju 2007). However, this effect is based on few data points only.

Assessing child labor may be particularly difficult if public perception diverges from the scientific reality. Specifically, in some areas and forms child labor may be considered acceptable, making it more difficult for stakeholders in areas where child labor often is considered unacceptable (Jørgensen, Lai, and Hauschild 2010). Consequently, defining child labor is also not an easy task.

2.3.1 Child labor definition

As there is no internationally agreed upon definition of child labor, some authors (e.g. Edmonds 2008) suggest basing a general definition on the conventions of the ILO. The basic definition of child labor was given in ILO Convention 138 in 1973 (International Labour Organization 1973a).[26] Called the Minimum Age Convention, it set the minimum working age at 15, lowered to 14 for certain exceptions such as apprentices and developing countries. This standard has two exceptions: Light work is already allowed 2 years earlier and hazardous work is in general only allowed from age 18 onwards (International Labour Office 2004, 21). Another important ILO contract is Convention 182 from 1999 (International Labour Organization 1999a). It defines the worst forms of child labor and particularly includes, among others, two elements that are also relevant in a business context: forced or compulsory labor and hazardous work which is expected to be dangerous to a child's health, safety, or morals (International Labour Office 2004, 46). A further international treaty, The United Nations Convention on the Rights of the Child from 1989, describes the rights of the child more generally (Ethical Trading Initiative 2012, 2).

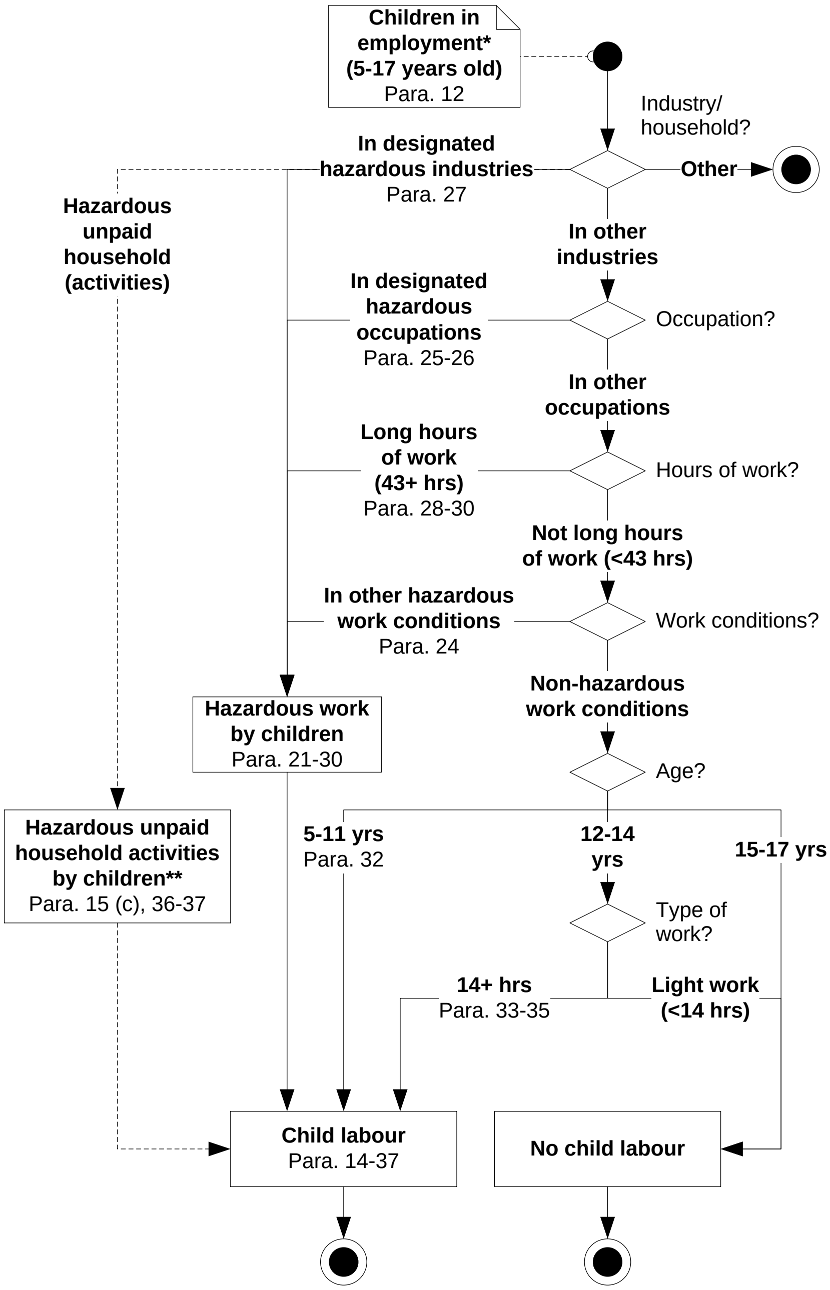
Hazardous work for children is regulated in detail in the ILO standards. ILO recommendation 190 (International Labour Organization 1999b) provides guidance on some of the occupational safety and health (OSH) factors which determine hazardous work for children and adolescents. According to the recommendation, the following types of work should be considered as hazardous (International Labour Organization 1999b, II.3.(a)–(e)):

- Work which exposes children to physical, emotional, or sexual abuse
- Work underground, underwater, at dangerous heights, or in confined spaces
- Work with dangerous machinery, equipment and tools, or which involves the manual handling or transport of heavy loads
- Work in an unhealthy environment which may, for example, expose children to hazardous substances, agents or processes, or to temperatures, noise levels, or vibrations damaging to their health
- Work under particularly difficult conditions such as work for long hours, during the night, or work which does not allow for the possibility of returning home each day

Figure 6 presents an overview of the ILO's child labor definition as an adapted UML state diagram[27] (Diallo et al. 2010, 19). It is based on a long discussion about what determines child labor and reflects a consensus from a multitude of definitions reviewed for this purpose (Edmonds 2008). The context-based definition of child labor is clearly visible, along with the absolute restrictions for children below 12 years of age.

[26] The convention is further detailed in the Minimum Age Recommendation, 1973 #146 (International Labour Organization 1973b).

[27] UML refers to the Unified Modeling Language version 2 (Object Management Group 2011). Some diagrams in this thesis will employ the logic of this modeling language without further detailing the standard.



* i.e., in economic activity in line with national accounts. Excludes non-productive activities and unpaid household services.

** Not included in ILO statistics

Note: Paragraphs reference ILO resolution on statistics of child labor (International Labour Organization 2008b)

Figure 6: Overview of ILO child labor definition (adapted from Hagemann et al. 2006, 40; International Programme on the Elimination of Child Labour 2013, 47)

For a child labor survey, two further differentiations need to be considered when interpreting the results. On the one hand a distinction is often made based on the economic attributes of the work. Hence, an activity can either be economic or non-economic depending on whether or not a productive activity, i.e., an activity “whose performance can be delegated to another person with the same desired result” (Understanding Children’s Work 2007, 3), falls under the definition of the system of national accounts (Understanding Children’s Work 2007, 3). ILO’s child labor definition includes all children performing economic activities, whether entirely market-related or not (International Programme on the Elimination of Child Labour 2013, 45). On the other hand, a distinction into family and non-family work may be possible depending on the location where the work is done. Consequently, work can either be economic family work, non-economic family work, or economic non-family work. The residual fourth category is not available (Understanding Children's Work 2007, 2–4). While one might assume that the family setting might be more important for the negative influence of child labor, the literature identifies work type, and hence the distinction into economic and non-economic work, as potentially more important and is also the prevailing form (Understanding Children's Work 2007, 22). All of these problems lead to the fact that country-level statistics are often very difficult to compare and therefore need additional investigation before being included in further work (Hagemann et al. 2007, 5–6).

The necessary measurement and monitoring of child labor in a country is typically a well-prepared process involving long preparatory phases (International Programme on the Elimination of Child Labour, Statistical Information and Monitoring Programme on Child Labour, and International Labour Office 2004). Child labor measurement often only takes place yearly or at longer intervals (Diallo et al. 2010, 47–49). The monitoring typically involves field officers that actively check on-site factors such as school participation. Hence, it is seen as linked to all levels of authority, especially local government (International Programme on the Elimination of Child Labour 2005, 3). In short, "[T]he simple act of observation and reporting is an active tool for eliminating child labour. CLM [child labor monitoring, A.T.] is thus more than just inspection, it sets in place a long-term process for documenting and following up abuses that will carry on long after particular child labour programmes and projects have been completed." (International Programme on the Elimination of Child Labour 2005, 11).

2.3.2 Child labor effects

Child labor may or may not have harsh effects on children. Some reports are very explicit in their evaluation of the effects: "Child labour deprives children of their dignity and childhood, and can be harmful to a child's physical, mental or social development. Child labour interferes with the child's education and possibilities to acquire basic knowledge and skills needed to pursue a decent life. Child labour is therefore seen as a violation of human rights." (Sustainable Trade Initiative 2012, 10). In the recent Brasilia Declaration on Child Labor, roughly 150 countries state: "[…] Convinced that the goal of eradicating child labour unites all countries since child labour impairs the realisation of children's rights and its eradication constitutes an important issue for development and human rights […]" (III Global Conference on Child Labour 2013, 2).

Indeed, the most evident effect of child labor is on school attendance. Research shows a negative correlation with the level of school attendance (Allais, Hagemann, and International Programme on the Elimination of Child Labour 2008, 4). A combined conclusion from multiple studies, also including test scores, show significant performance gaps between (also only partly) working and non-working children (Understanding Children's Work and International Labour Office 2010, 46). However, the effect on education is still under discussion and may depend on further contextual factors such as age and the type or intensity of the work (Dorman 2008, 28). Some researchers suggest that school and work may be complementary or school and work might both be unattractive, leading to idleness (Fors 2012). Zutshi, Creed, and Sohal relate the "vocational imperative" (Zutshi, Creed, and Sohal 2009, 50) of schooling to the need for an educated labor supply in developed countries. They contrast this with the low living standards in developing countries. Other authors state that combining work and school may eventually lead children to drop school completely and start full-time employment (International Labour Office 2004, 134).

Another frequently discussed issue is health, which is even more complex (Dorman 2008, 46). Here, recent developments suggest a long-term negative health effect of child labor (Understanding Children's Work and International Labour Office 2010, 70). Jorgensen, Lai, and Hauschild (2010) try to build impact pathways of the effects of child labor on domains that should be protected via an SLCA, particularly "overall well-being" and "development of the society." While they are able to establish some relationships via intermediary variables (called "impact pathways" and "endpoints" in an SLCA), many effects in their summary are still unclear or strongly context dependent. Furthermore, they see well-documented, but unpredictable effects of child labor on health and schooling and fuzzier, undocumented effects on the wage level or potentially positive outcomes. If the relationships can be proved, effects on "overall well-being" are caused through effects on health and longevity, equality, standard of living, and human development, while effects on the "development of the society" are caused through changes in human or produced capital (Jørgensen, Lai, and Hauschild 2010).

Given these complexities, and in harsher contrast to the ILO definition of child labor, Bourdillon et al. present a social sciences view "[…] that children's participation in safe, non-exploitative work can be a fundamental mechanism of personal and social development." (Bourdillon et al. 2010, 106). Overall, there is no clear, significant, and commonly supported scientific basis for the effects of child labor on children. Although some researchers and organizations agree on the negative effects on health and school-related outcomes, others question the educational imperative of schooling and see some benefits from work on a similar relevant level.

2.3.3 Importance of child labor for companies

Although discussions on the effects of child labor prevail, international companies have to explicitly deal with the question of child labor. This is especially true for companies based or with customers in the “western world.”[28] The importance of child labor incidents for companies is manifold. From a legal standpoint, one important example is the Ruggie Companion Report for the United Nations Human Rights Council. This report sees a chance of legal action in case of knowingly contributing to child labor (i.e., human rights) issues: “In short, both operating in contexts where abuses occur and the appearance of benefiting from such abuses should serve as red flags for companies to ensure that they exercise due diligence, adapted for the specific context of their operations.” (Ruggie 2008, 21). For US-based companies this is fostered by the Alien Tort Claims Act from 1789 (Collingsworth 2003). In other countries it depends on the local law and will not be detailed here. Suffice to say, more and more governments require companies to report on their measures against child labor in their supply chains (Travis 2014).

Multiple standards and guidelines include the prohibition of child labor and more and more companies acknowledge these, be it voluntarily or due to stakeholder pressure. Examples include:

- GRI – Global Reporting Initiative 3.1 (Global Reporting Initiative 2011)
- SA8000 – Social Accountability 8000 (Social Accountability International 2008b)
- BSCI – Business Social Compliance Initiative (Business Social Compliance Initiative 2009)
- GSCP – Global Social Compliance Programme (Global Social Compliance Programme 2010)
- HRCA – Human Rights Compliance Assessment (The Danish Institute for Human Rights 2006)
- Global Compact (UN Global Compact Nordic Network 2013)
- Fair Labor Association Workplace Code of Conduct (Fair Labor Association 2011)

[28] By “western world” we mean countries within the European Union and/or OECD.

Refraining from child labor is also included in guidelines published in academic journals (Closs, Speier, and Meacham 2011; Carter and Jennings 2002). Companies often include child labor requirements – often in line with the ILO definitions – into their “codes of conduct” based on these guidelines and relevant further sources. Earlier studies on the topic support this (Kolk and van Tulder 2002b; Kolk and van Tulder 2002a) and an even stronger emphasis may be expected today.

It is also necessary to differentiate the reaction if a case of child labor case is detected. While the guidelines suggest stopping working with a supplier as the “last resort”, working together to improve the condition is the preferred option. This goes hand in hand with the suggestion from other standards to, for example, financially support the reintegration of children found working into school life (Social Accountability International 2008b, 5; Business Social Compliance Initiative 2009, 3).

2.3.4 Child labor incidents

For companies, child labor incidents are of particular interest as they constitute clear violations of their “codes of conduct” (assuming child labor according to the ILO definition is part of the code). However, the exact understanding of a child labor “incident” needs to be defined. A basic understanding can be derived from the ILO definition of child labor outlined above. An “incident” may be conceptualized based on the understanding of events and incidents in literature. Events are real-world occurrences defined by a duration, an associated place, and a change in a connected state (Xie, Sundaram, and Campbell 2008). Incidents have also been specified in several contexts. In IT, ITIL[29] defines an incident as “[A]n unplanned interruption to an IT service or reduction in the quality of an IT service.” (ITILFoundations 2014). More generally, in ISO 20000 the International Organization for Standardization defines an incident as “[an] unplanned interruption to a service, a reduction in the quality of a service or an event that has not yet impacted the service to the customer.” (International Organization for Standardization 2011) Combining these definitions, in the context of this work a “child labor incident” is understood as “an event that happens at a particular point in time at a specific place and that changes the state of one or several entities (e.g., companies) towards being associated with child labor.” Examples of such events could include spotting a child in a company’s production facility or releasing a report on child labor in a particular province. Child labor incident news, then, is any type of text (e.g., a news story) covering an event of this type or one that includes a suggestion of child labor. Events that suggest that a certain

[29] The IT Infrastructure Library (ITIL) combines best practices for service management in IT (Greiner 2007).

entity is aiming to reduce child labor are also seen as relating the entity with child labor – otherwise no action would be necessary. This thesis concentrates on business-related child labor incidents.[30] Moreover, it is not necessary for the text to focus on the child labor incident; it is sufficient for it to be included as a side-topic in any news text. However, "the world" in general is not seen as an entity.

2.3.5 Summary

While at first glance child labor may seem to be one of the better defined categories of social sustainability when compared with others such as corruption or diversity, the details make it significantly more complicated. In summary, there is no universally accepted definition of child labor. Depending on the definition, the number of children classified as working will vary. Nevertheless, child labor is likely to prevail in many countries under a wide variety of definitions.[31] Given that the ILO standards exist and are broadly acknowledged, these can be seen as a baseline. Moreover, given the problems with age determination in developing countries, any credible report on child labor calls for a further investigation.[32]

Although the debate surrounding the "rights and wrongs of children's work" (Bourdillon et al. 2010) exists at a higher level, international companies (i.e., companies with international supply chains) face challenges due to regulations, laws, and particularly protests driven by customer perception of child labor incidents at nearly any stage of their supply chain. As a result, while the effects of child labor on children appear not entirely clear, the reputational ramifications for companies detecting child labor in their supply chains can be harsh (Lemke and Petersen 2013). Therefore, they need to react bearing the cultural context of child labor (Bourdillon et al. 2010, 8–11) in mind.[33]

2.4 Summary of related work

In summarizing related work, there is a clear lack of research into models for the ongoing monitoring of social sustainability risks which integrate public news evidence. To the best of

[30] This excludes reports on child labor specifically dealing with child soldiers, sexual exploitation, begging, child criminals, child slaves in non-business contexts (e.g., related to sexual exploitation), or sports.
[31] The discussions surrounding child labor definitions have another consequence. If child labor is discussed in news articles, a definition of child labor is formed dependent on the author of the article. Either the author states the definition of child labor used by a news article explicitly or he does not. Typically the latter is the case, leaving some room for interpretation.
[32] Given that this thesis is not concerned with determining the likelihood that a child labor issue becomes a public, reputational scandal, but with the (true) probability of a compliance breach, a simple assumption like "any report on child labor may be harmful for a company's reputation and, hence, needs to be included into a risk model" appears insufficient.
[33] Reasons and indicators for child labor will be discussed in a later chapter.

the author’s knowledge, no quantitative risk model has been presented in the sustainability domain that explicitly includes ongoing updates from news sources.

Moreover, known currently proposed systems for ED and EE do not specifically support the risk types and the configuration necessary for integration into a risk model focused on social sustainability risks. Given their often depicted domain-dependence, additional research is required. The novelty of this thesis in terms of text mining will be to apply, compare, and evaluate relevant concepts for this specific domain. Thereby, this thesis will build on previously presented work and may offer ideas for other applicable scenarios.

Due to its inherent ambiguities, child labor needs to be integrated into any risk model with great care. In addition, the need for child labor monitoring needs to be checked against the norms within the context of a (focal) company. Depending on cultural background, child labor might be viewed differently. Nevertheless, the need for action within companies with relations to the "western world" appears apparent.

3 Information needs

One important criterion for the success of a piece of software is that it fulfills its requirements. Hence, requirements can be used to evaluate the software after design. A supply chain risk management system focused on child labor can be seen as a specialized version of a system dealing with general sustainability. Consequently, the requirements for the more specialized system should not violate those of the more general one. While requirements defined in academic papers allow for a certain degree of argument, no paper, to the author's knowledge, has provided an analysis of the requirements that need to be fulfilled by a sustainability risk management system for supply chains. Consequently, this chapter uses the input from qualitative interviews with five experts in order to extract information needs which can be used to formulate such requirements.[34]

Providing the right information is key to managing sustainability risks appropriately. However, sustainability incidents in supply chains show that in many cases company information needs with regards to sustainability are not fulfilled. For example, managers often lack knowledge about second or lower tier suppliers. Important data might be distributed over a large variety of data silos and ERP systems. Scientifically, this problem can be explained with the information subset model developed by Szyperski (1980). Szyperski defines different information sets: information supply (the information provided by a system), objective information need (the information required for a task), and subjective information need (the information a particular person requires for a task). In practice, these sets only partially intersect, i.e., the information provided rarely complies fully with the information needed, either subjectively or objectively. However, in the ideal case the information subsets would completely overlap, i.e., the supplied information should also cover the users' theoretical information need. As a consequence, developing an information system has to be based on a profound understanding of information needs. Only then can the information system be designed to match the combined information need as closely as possible with a corresponding information supply (Becker et al. 2004).

[34] The research presented in this chapter was conducted together with two other researchers, Lisa Madlberger and Alexander Schatten, and a prior version of this paper was published in Thöni, Madlberger, and Schatten (2013a).

3.1 Research design

Information needs are inherently context specific and may depend on the business, role, or individual. Consequently, we will aim to derive the requirements suggested by potential users of an SRMS. The resulting analysis serves as a starting point to design the information resources necessary to fulfill the identified needs. The results of the analysis characterize the four core tasks of such a system, as depicted in Figure 7: data provision, data processing, data analysis, and data presentation (Gluchowski and Gabriel 2008, 109).

Figure 7: System structure analyzed (author's representation based on Gluchowski and Gabriel 2008, 109)

3.1.1 Questionnaire

An empirical approach was taken to deriving information needs, and thus requirements, from a company perspective. For this purpose, we re-evaluated the responses to a more general questionnaire designed to understand sustainability management processes and their technological support across multiple companies. However, the scope of the questions matches the goal of this particular study. The questionnaire is split into three sections: The first considers the monitoring of sustainability in companies. The second focuses on understanding the ranking of suppliers. The final section considers the integration of stakeholders and their respective reports. Each section of the questionnaire is divided into three parts. The first part contains questions designed to understand existing processes. The subsequent two parts ask for an evaluation of the existing processes and potential for improvement. All questions can be found in full in "Appendix D: Overview of questionnaire to derive information needs".

The first section of the questionnaire addresses the identification and analysis of sustainability risks. As highlighted above, ranking of suppliers is a key task during an assessment of sustainability risks, which is highlighted in the questionnaire's second section. Finally, the questions in the third section relate to stakeholder communication and integration, which helps to explore information needs that go beyond the supply chain. Thus, the questionnaire reflects major topics identified in the earlier discussion of sustainability risk management. Given the focus of the questionnaire, most information needs derived from the responses refer to risk assessment topics, which are also the focus of this thesis. The information needs are identified from the answers and categorized into predefined groups. This extracted list of information needs is not exhaustive; instead it represents a collection of important elements that should be considered when designing an SRMS.

3.1.2 Case selection and data collection

The dataset from the expert interviews consists of responses from five individuals from four companies from three different industries and three countries (headquarters locations).[35] Table 6 summarizes the companies and managerial roles interviewed. In order to facilitate the discussion and understanding as much as possible, either the English or the German version of the questionnaire was used as a guideline for the interviews.

No	**Company sector** (acc. to NACE Rev. 2; Eurostat 2008)	**Managerial roles interviewed**
1	C11 - Extraction of crude petroleum and natural gas	Head of Corporate Sustainability, Manager of Corporate Sustainability Controlling & Reporting
2	G47 - Retail trade, except of motor vehicles and motorcycles	Head of Quality Management
3	G47 - Retail trade, except of motor vehicles and motorcycles	Senior Sustainability Manager
4	C17 - Manufacture of paper and paper products	Sustainability Manager[36]

Table 6: Companies and managerial roles interviewed (author's representation)

The condensed summaries provide a total of around 3,200 words of content. All interviewees are experts in the field of sustainability.

3.1.3 Data analysis

Methodologically, we extracted information needs from the responses by cross-checking them with each other. Information needs which were mentioned multiple times for the same company were combined. We derived information needs in a technology-independent way and subsequently categorized them in five steps:

[35] The interviews lasted between 30 and 60 minutes. In each interview notes were taken and a summary of the responses in the questionnaire's format was sent to the interviewees for review. Upon correction, these summaries were analyzed and coded. For one company, the manager directly responded to our questionnaire without a formal interview. Not all questions were completed by all respondents.
[36] Response to questionnaire instead of interview in person.

Step	Name	Analysis
1	**Defining the risk management context**	We linked information needs to the processes defined in two risk management frameworks in order to identify information needs which focused on risk assessment as specified in the respective framework. Given their broad acceptance, we applied the ISO 31000 risk management process (The Association of Insurance and Risk Managers, The Public Risk Management Association, and The Institute of Risk Management 2010) along with the specific process suggested by Anderson and Anderson (2009) depicted in chapter 2.
2	**Defining the information scope**	As second step, we matched information needs to the appropriate level in the company hierarchy, indicating whether they related to the whole supply chain, the whole company, or sub-units in the company.
3	**Identifying the user group**	For each information need we identified the stakeholders that would express the specific information need in the first place, e.g., management, employees, suppliers, or partners.
4	**Assigning a risk management strategy**	Here we tried to understand why a particular information need might be expressed beyond identifying, analyzing, or evaluating risks. This was done because it might help with cross-checking the degree of completeness of the information needs. For this purpose we used the framework proposed by the World Economic Forum which suggests five strategies to manage risk (World Economic Forum 2012): (1) scenario planning, (2) data and information sharing, (3) risk quantification metrics, (4) shaping new legislation, and (5) trusted networks across business and government. This step also helped to identify requirement areas which were potentially missing from the analysis.
5	**Determining the technical architecture**	We clustered the information needs and translated them into either functional or non-functional requirements for a potential IT-based sustainability management system. Finally, we designated the technical system part (see Figure 7) which is affected by the requirement.

Table 7: Overview of categorization steps performed (author's representation)

Apart from this categorization, we analyzed the content of the information needs, grouped similar needs, and assessed how many interviews each need was mentioned in, showing the degree of overlap between them. Taken as a whole, the categorization and content analysis provide an initial understanding of the information needs expressed during the interviews. Consequently, the discussion provides propositions for further examination.

3.2 Results

In the following, the results will be presented in two categories: a categorization analysis to identify information needs in the responses followed by a content-driven examination.

3.2.1 Categorization analysis

A total of 110 information needs were identified, which translated to 63 unique information needs with multiple responses. Table 3 shows the categories used in the 5-step analysis, as described in the previous section, and provides an overview of the distribution. The total count of the first column indicates the number of unique information needs assigned to each category. In contrast, the second column multiplies the number of unique needs with the number of interviews from which this information need has been derived. Therefore, this column also reflects information needs mentioned by multiple independent interview partners. As the focus of this study is on risk assessment, it is important to note that these figures relate to the 61 information needs assigned to the risk assessment steps (risk identification, risk analysis, risk evaluation) from ISO 31000 or the analysis or identification of loss exposure in the process by Anderson and Anderson.

Categorization step	Analysis dimension	Assigned category	Total count of unique information needs	Total count weighted by no. of interviews mentioning the need
Step 1 Risk management context	All	Focus: Risk assessment	61	106
	ISO process (most relevant)	Establish context	2	2
		Risk identification	20	35
		Risk analysis	57	101
		Risk evaluation	21	33
		Risk treatment	14	22
		Monitoring and review	5	8
		Communication and consultation	0	0
	Anderson & Anderson process (most relevant)	Identify loss exposure	20	35
		Analyze loss exposure	57	101
		Control loss (pre- and post loss)	8	10
		Implement risk financing alternatives	14	22
		Evaluate process and implement improvements	4	5
Step 2 Information scope	Related level of corporate hierarchy	Industry level	3	3
		Supply chain level	19	36
		Company level	15	26
		Sub-company level (region, unit, etc.)	9	16
Step 3 User group	Internal Stakeholders	Employees and labor unions	3	6
		Management	58	99
	External Stakeholders	Governments	2	5
		Public (incl. NGOs, investors, customers)	2	5
		Suppliers and partners	7	10
Step 4 Risk management strategies	Strategy variants	Scenario planning	3	5
		Shaping new legislation	1	1
		Data and information sharing	23	43
		Risk quantification metrics	34	58
		Trusted networks across business/governments	5	11
Step 5 Technical architecture	Requirements type	Functional requirement	58	103
		Non-functional requirement	3	3
	Technical system part affected mostly	Data provision	23	47
		Data processing	13	15
		Data analysis	17	27
		Data presentation/dissemination	8	13

Table 8: Topics covered by the identified information needs (author's representation)

3.2.1.1 Step 1 – Risk management context

As this analysis is focused on risk assessment (risk identification, risk analysis, risk evaluation), all of the information needs have been assigned to one of these phases of the risk management process. Indeed, with 61 of 63 unique information needs being related to risk assessment, the relevance of the responses to the paper's focus is clear. Additionally, most information needs (57 of 61) relate to risk analysis. The phases of risk identification (20) or evaluation (21) are less covered.

3.2.1.2 Step 2 – Information scope

In step 2 the goal is to determine the scope of the identified information needs. About half of the information needs relate to the company or sub-company level (24 of 46), while more than 40% (19 out of 46) relate to information from the supply chain.

3.2.1.3 Step 3 – User groups

Given that the questionnaire addresses managers, nearly all the identified information needs can be related to managers as the beneficiary stakeholders (58 of 61). Only some information needs can be related to other stakeholders (14 of 61). However, also in this case there is an elevated supply chain context.

3.2.1.4 Step 4 – Risk management strategies

The information needs can be strongly related to two risk management strategies: data sharing and risk quantification (57 of 61). While other strategies were also mentioned, they seem only narrowly covered by this analysis.

3.2.1.5 Step 5 – Technical architecture

Finally, the information needs can be associated with requirements for all system parts with a particular focus on data provision (23 of 61) and to a lesser extent for data presentation (8 of 61). Consequently, the input side needs particular focus when designing an SRMS.

3.2.2 Content analysis

Analyzing the different information needs related to risk assessment from a quantitative perspective, nearly half of the needs identified were only mentioned by a single company. In contrast, ten information needs, or around 16%, were named by at least three companies. Three needs were named by every company: (1) the need for country-/region-level risk metadata (2) the need for support for audit frameworks for data input and (3) the need for a way to aggregate information from different source systems. These relate to different elements of the SRMS, but all three are related to the topic of “data provision,” underlining its importance. For the seven information needs mentioned by three companies (as shown in Table 9), it stands out that five are related to audits. While this might to some extent be triggered by the questions asked, the need to cope with audits and their qualitative elements when dealing with an SRMS seems clear.

Inclusion of data from multiple supply chain levels/tiers
Allow for general and detailed external reporting
Support for audit prioritization
Cope with manual and automated data input
Support for audit frameworks for data input
Handle input from qualitative interviews/quantitative assessments during audits
Support for input from external sources performing social or environmental audits

Table 9: Information needs mentioned by three companies (author's representation)

We clustered similar information needs into groups. These are discussed in more detail in the following sections.

3.2.2.1 Data provision

Requirements group with information needs	Count
Inclusion of external data (external blacklists, external platforms, social media, news media)	5
Inclusion of risk metadata (country/region data, supplier data, legal data, components/products, certifications)	5
Integration of heterogeneous data (e.g., quality, complaints, and security management)	4
Support for standardized data input (audit frameworks, assessments)	3
Support for reports (country managers, walkarounds)	2
Data quality	1
Integration of local sustainability event data	1
Easy integration of new users and partners	1
Support for manual and automated input	1

Table 10: Information need groups for data provision, incl. number of information needs per group (author's representation)

Data inclusion and data integration are two topics that have been strongly emphasized for data provision. Data can either come from internal or external systems and needs to be integrated into the SRMS even if it is provided in heterogeneous forms. Moreover, the data input may be automated or manual. Internally, particularly systems like the quality, complaints, or security management system may provide input that facilitates the derivation of sustainability risks and their status. Externally, multiple approaches can be taken. The data can be provided directly from, for example, other platforms like SEDEX[37] (e.g., providing blacklists; see (Kogg and Mont 2012)) or it can also be derived from unstructured sources such as social media or news. The latter would allow the monitoring of current events and potential issues. If possible, this data should cover at least the whole supply chain across multiple tiers, and if feasible this should be extended to the local level. The input from audits and assessments is a combination of internal and external data and is not necessarily performed by internal staff. In any case, the system has to provide paths to include this (potentially qualitative) data in a standardized way and therefore has to be able to deal with audit and assessment frameworks. Additional qualitative data may be provided by other reports such as results from walkarounds or inputs from sustainability managers. To be able to assess risks properly, further metadata is

[37] SEDEX is a platform for the exchange of supply chain sustainability information between companies; http://www.sedexglobal.com/.

needed. This might include information about countries, legal standards and requirements, suppliers (including their certifications and products). Within this process, data quality can be seen as a major concern affecting the entire assessment. Finally, the SRMS must have procedures to easily integrate new partners and users. New partners might be new supply chain members or freshly acquired companies. Reasonable assessments of sustainability are only possible if the SMRS offers a comprehensive view.

3.2.2.2 Data processing

Requirements group with information needs	Count
Definition of KPI structure and aggregation logic	5
Integration of KPI standards and schemata	4
Scalability and security	2
Support for industry standardized KPIs	1
Policy management	1

Table 11: Information need groups for data processing, incl. number of information needs per group (author's representation)

A key part of the data processing architecture is the handling of predefined KPIs. These KPIs can be seen as key elements for assessing sustainability risks and should be based on existing standards that can be used in a clean or mixed form. Consequently, data needs to be integrated into the structure according to the definition of these standards and, if possible, also into a dashboard logic. Conforming to standards may also be required to allow data exchange and comparison at an industry level based on the data of the SRMS. Besides the processing of data to conform to these standards, company- and even stakeholder-specific adaptive aggregation logics should be allowed to structure KPIs. Given that different sub-organizations of a company may use and need different logics, a consolidated view at the head office level must also be possible. In general, particularly when processing data from other organizations, high security standards need to be met. Additionally, the SRMS should be easily adaptable to new standards and also easily scalable in terms of processing when new requirements need to be met, for example with regard to real-time handling.

3.2.2.3 Data analysis

Requirements group with information needs	Count
Allow for supplier selection, supplier ranking and audit triggering/prioritization	5
Integration of assessments (risk in general, environmental and social impacts, codes of conduct)	3
Definition of targets, thresholds, and binary criteria	2
Implementation of automated alerts and incident tracking	2
Coverage of three dimensions of sustainability	1
Implementation of information workflows	1
Support for regulatory processes	1
Allow for continuous performance management of suppliers	1
Tracking of stakeholder contacts	1

Table 12: Information need groups for data analysis, incl. number of information needs per group (author's representation)

The analysis of sustainability risks should primarily be based on a pre-structured list of risks. An exhaustive analysis would need to cover every sustainability dimension. This might include assessment frameworks, which can be based, for example, on predefined impact assessments or "codes of conduct". Regulatory requirements also provide guidance for risk assessment. For analysis this should include the ability to set targets and easily check the fulfillment of KO criteria (i.e., binary criteria). Given that risk assessment is not restricted to a single point in time, continuous processes and integrated information workflows are needed. Integrating continuous performance management processes and tracking keeps the risk analysis up to date. This tracking can be automated through the implementation of thresholds, triggers, and incident-based alerts. Contact with stakeholders should also be included. An SRMS particularly needs to include an analysis of suppliers based on their sustainability risk level. The support needs to start by implementing supplier selection due diligence and end with a continuously updated risk-based ranking of suppliers. This ranking should be adaptable to specific topics or localities. The prioritization of audits can be based on these rankings which could be triggered from the SRMS.

3.2.2.4 Data presentation

Requirements group with information needs	Count
Support for specific views (stakeholder, plant) and teams	3
Sharing of tracked data across companies	2
Detailed and general reporting	1
Map-based visualization	1
Multi-business unit real time perspective	1

Table 13: Information need groups for data presentation, incl. number of information needs per group (own representation)

As an integrative SRMS will likely have a diverse group of users, supporting specific views appears to be a key feature of the system. This includes support for teams and result filtering, e.g., at the factory level. The data should be presented in real time, even across business units.

The aggregation step needed for understanding can either be very detailed or at a more general level. Hence, drill-downs could be an option for coping with this requirement. If applicable, map-based visualizations can help to better understand risks. Finally, given that risk management often requires involving other stakeholders, data sharing should be possible and simple. However, in order to safeguard internal company data, the sharing process should include tracking procedures for the data elements included (see e.g. Barrett, Strunjaš-Yoshikawa, and H Bell 2007).

3.2.3 Discussion

Based on the insights gained from both interviews with experts and literature, the goal of an SRMS can be seen in providing up-to-date, processable information about the environmental, social, and economic risks faced by a company and its supply chain. Based on the information needs identified in the previous sections, seven basic propositions were derived to define a set of major requirements for an SRMS.

Given that the majority of identified information needs relate to data provision, an SRMS must implement critical features for collecting and integrating sustainability data. Thus, an SRMS can be seen as a hub connecting information about a company's sustainability performance and its social and ecological environment to inform decision makers about sustainability risks. Nevertheless, the overall goal of a complete risk management system – to be aware of all risks at all times – may be constrained in multiple ways. Two dimensions stand out here. Content completeness is strongly influenced by the complexity that companies, especially multinational companies, face (World Economic Forum 2012). To achieve a comprehensive view, all steps in the production process need to be considered as sustainability needs to be captured in all product process steps (Ramani et al. 2010; H.-F. Wang and Gupta 2011). This process does not stop with the company itself – data from lower steps in the value and supply chains are needed. Time completeness implies having data from the past to the present. This is particularly difficult for historic and real-time data. Consequently, an important step is to identify relevant sources contributing to the overall risk awareness in a company. This leads us to the first two propositions:

> **Proposition 1:** Identifying the most important data sources is a key task in the development of an SRMS.

> **Proposition 2:** An SRMS has to provide the means to integrate both internal sustainability data, collected manually or automated in various source systems, as well as external sustainability data, derived from external platforms, media channels, or from suppliers and partners.

From a management point of view, the possibility to retrieve quantified metrics is a major purpose of an SRMS. Current practice, along with ideas for improvement from sustainability monitoring, supplier ranking, and stakeholder integration perspectives assessed in the interviews, strongly highlight information needs focused on risk quantification metrics rather than on other risk management strategies (like scenario planning, legislation shaping, or building trusted networks). Indicator schemes can include measures and rules developed specifically for a particular company but which may also rely on standardized calculation procedures provided by industry-wide standards or risk assessment frameworks. Unified legal regulations and reporting standards are still under development at both national and international levels, particularly in the field of sustainability risk management; therefore an SRMS must be able to dynamically adapt to new standards.

> **Proposition 3:** An SRMS has to be able to integrate and dynamically adapt risk assessment and sustainability reporting frameworks.

> **Proposition 4:** An SRMS should allow for the definition and continuous monitoring of sustainability indicators for company-specific risk assessment. The system should support customized aggregation rules, calculation procedures, and thresholds in order to account for individual risk analysis and evaluation.

Comparing the information needs to the ISO 31000 risk management process shows the information needs focused on and derived in this chapter relate well to the phases of risk assessment (i.e., identification, analysis, and evaluation). One major potential of an SRMS is its ability to automatically analyze the presence of potential risks at a specific company location or supplier site. Going beyond the calculation of sustainability indicators, risk analysis should support more sophisticated functions such as sustainability risk ratings for suppliers (e.g., during supplier selection) or rankings of existing company sites or suppliers in order to prioritize audit activities. Based on defined rules, the system should take over continuous monitoring functions and inform decision makers if a critical risk or incident has been detected.

> **Proposition 5:** An SRMS should support risk ranking and assessment procedures which can be integrated into existing business processes.

> **Proposition 6:** An SRMS should detect and track sustainability incidents and alert decision makers.

As major drivers for sustainability can exist outside the company (e.g., investors, customers, and governments), external communication with regard to sustainability appears important to all interviewees. Information has to be prepared at different scopes, levels of granularity, and in different forms for different stakeholder groups, leading to the final proposition:

> **Proposition 7:** An SRMS should be able to generate reports for both internal and external use, allowing for stakeholder-specific integration into the risk management process.

Figure 8 presents an overview of the resulting propositions, matched to the system's elements.

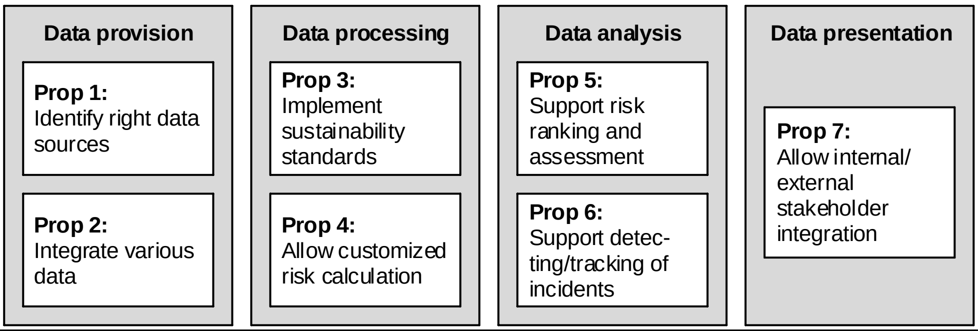


Figure 8: Overview of propositions (author's representation)

Overall, the information needs, derived requirements, and propositions showcase a broad array of potential requirement seeds that need to be fulfilled, particularly from a management perspective. While the information needs capture normative "wishes" for data, practical implementation can be hindered by multiple influences. Nevertheless, they should provide an essential step towards defining all requirements for an overarching and integrative SRMS.

3.3 Summary and conclusion

In addition to economic considerations, companies are increasingly pushed by society and governments to assume responsibility for the social and environmental impact of their corporate actions. Information systems could help companies to identify and analyze sustainability related risks along their supply chain.

In order to explore the requirements of such a sustainability risk management system (SRMS), we identified information needs for sustainability risk management based on expert interviews and literature. These are reflected in the upcoming parts of this thesis, given that the child labor risk management system discussed may be seen as covering one detailed aspect of the overall logic.

We identified that an integrative SRMS has to meet particular data integration requirements since the analysis of sustainability risks requires the integration of diverse sources of both internal and external information (Propositions 1 and 2). Furthermore, both standardized as and customized aggregation procedures for sustainability indicators should be supported (Propositions 3 and 4). Extensive risk analysis, especially during supplier identification and evaluation, should form part of an SRMS, together with automated risk alert functions (Propositions 5 and 6). Finally, stakeholder-specific visualization functions are particularly important for an SRMS, as the system provides information to both internal and external stakeholders (Proposition 7).

The results of the questionnaire revealed details about the requirements for a general SRMS. However, while the responses covered a wide array of elements, particularly within the field of risk identification, analysis, and evaluation, they should not be seen as exhaustive. A more comprehensive questionnaire could help to research other areas in more detail.

4 Child labor evaluation methodology and risk model

This chapter suggests a risk model to evaluate child labor risk in supply chains. The first section describes the overall structure of the risk model. The second section details child labor related evidence which is important for this work. Finally, the third section outlines the suggested risk model in multiple steps.[38]

4.1 General structure

Bayesian networks (BNs) have previously been successfully applied to risk management due to their understandability and ease of information integration (Duespohl, Frank, and Doell 2012; Koks and Challa 2005; Wooldridge 2003). BNs are a well known probabilistic modeling technique introduced by Pearl in 1988 (Pearl 1988). Nevertheless, despite their applicability to the domain of this thesis, BNs have not been suggested for ongoing supply chain social sustainability risk management which includes news (Appendix E: Bayesian networks for supply chain risk management).

BNs exhibit several particularly helpful properties for this domain. The key advantage of BNs is their explicit treatment of uncertain information supporting decision making (Reckhow 1999) and the possibility to include different types of sources into a single consistent model (Uusitalo 2007). Multiple information sources represented with different variables can be integrated into the network (e.g. Duespohl, Frank, and Doell 2012). This includes subjective and quantitative information (Wooldridge 2003) and might be used to include evidence from news. A process of updates allows the inclusion of new news continuously triggering updates of the likelihoods (Neapolitan 2003, 12–29). BNs tend to be easily communicable, fostering a common understanding (Duespohl, Frank, and Doell 2012; Correa, Bielza, and Pamies-Teixeira 2009). These features of BNs are of special interest for a quantitative risk model, as decision-relevant information system must be understandable for company executives, who

[38] Parts of this chapter have been published in Thöni, King, and Tjoa (2014).

have to make and defend their decisions based on the input (Hubbard 2009). An example can be seen in the possible uselessness of data generated through big data analytics if contexts are not properly understood (Amerding 2013).

This thesis suggests an IT system that implements a BN for each supplier location. By calculating the likelihood of breaching social sustainability compliance standards, a supplier location ranking can be derived. The system focuses on detecting true child labor cases at a supplier location (as defined in the Chapter 2) and not the probability of a child labor case going public or viral.[39] Although related, quantifying the latter would require estimates, e.g., about the probabilities that certain media cover a child labor case. Nevertheless, the value provided by the model can be viewed as an initial estimate.

We propose to start with the initial hypothesis that a supplier conforms to given, pre-defined social sustainability standards (e.g., a “code of conduct”). Then we calculate a relative measure for the likelihood of this hypothesis being false. An IT system is developed that can be used to gather and combine evidence on the likelihood of a compliance breach. Hence, computing the likelihoods for individual suppliers and relating them can provide a relative risk ranking. Figure 9 presents an overview of the system.

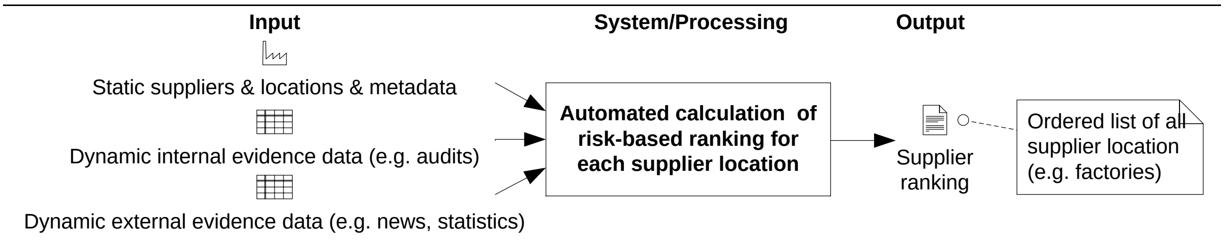


Figure 9: Generic system overview (author’s representation)

The IT system should be able to use data from multiple sources and combine it using a BN. Comparing the compliance risk for different supplier locations can help to establish an ordered list of suppliers to be used for further investigation, review and auditing.

4.1.1 Input

Previously, rather static factors about a supplier location such as its geography or products have been combined with more dynamic factors such as audit scores to determine sustainability risks (Foerstl et al. 2010). Static data could come from internal or public statistical sources

[39] Viral refers to cases where a certain news item is repeatedly shared by other news media, resulting in snowballing coverage.

such as Worldbank, Eurostat, or Transparency International data sets[40]. As with earlier suggestions, static factors are used to derive a static risk factor.

Since a key feature of the system is the ongoing validation of compliance for suppliers, a continuous stream of input data is required. Because input from suppliers, certification processes, and audits (mainly internal) can only be updated in defined or irregular intervals, important external evidence data may be derived from frequently updated sources such as news or other dynamic public information like referenced content in social media feeds. Integrating external data has, for example, already been referenced in the context of situational business intelligence (Rizzi 2012). Here, a text mining system can extract relevant data from news or other items to relate a certain news text to a specific supplier location. Nevertheless, the information is limited to online retrievable sources. Additional channels (e.g., manual input by local NGO representatives) could be a suggestion for future research. The specific extraction process and the usefulness of news data sources will be discussed in chapter 5. Here we focus on how such evidence can be integrated into the risk model.

4.1.2 System/Inference

To process the data, a BN is developed to combine the evidence for each supplier. The BN will produce a likelihood level for a social sustainability compliance issue at a specific supplier location that can be then used for ranking. The initial BN will assume that a supplier has not breached their contract or "code of conduct". Each piece of evidence collected will lead to an updated graph and a recalculated likelihood for the location.

As introduced above, the inference will be based on the static context of a supplier location. This will be integrated into the BN through a context-based "prior" likelihood and complemented with results from internal audits and external observations, e.g. from text sources. Through a process of updates, new news observations or audit evidence can be included continuously and the likelihoods are revised. The specifics of these four central model elements will be discussed in a later section (4.3). Figure 10 shows this conceptual model.

[40] http://data.worldbank.org/; http://ec.europa.eu/eurostat/data/database; http://www.transparency.org/research/.

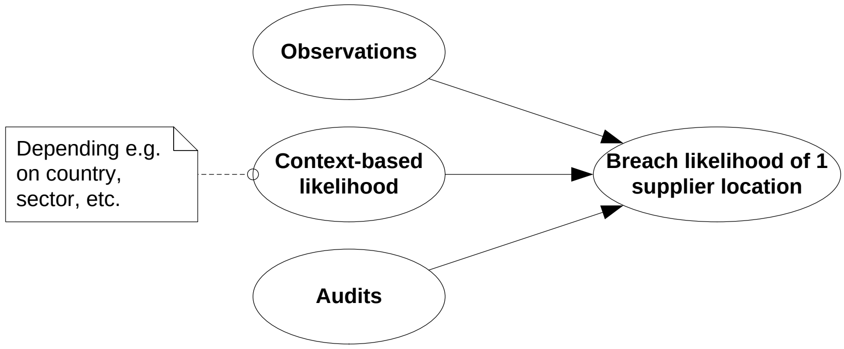


Figure 10: Conceptual inference model of a BN risk model (author's representation)

4.1.3 *Output*

The final result of the approach will be an ordered list of all suppliers based on their risk of conflicting with social sustainability requirements. This list can be used in order to improve the efficiency and effectiveness of deploying auditing resources. Additionally, it could function as an alerting service in day-to-day operations, provided data input is ongoing and timely.

4.1.4 *Excursus: Introduction to Bayesian networks*

BNs are based on Bayes' Theorem and are "[…] a directed acyclic graph (DAG) in which nodes correspond to random variables of interest and directed arcs represent direct causal or influential relation between nodes. The uncertainty of the interdependence of the variables is represented locally by the conditional probability table (CPT) […]" (Watthayu and Peng 2004, 8). A BN has to fulfill the Markov condition, which requires conditional independence for each node from its non-descendents given its parents (Neapolitan 2003, 31–40).

BNs rely on classical probability theory, allowing one to explicitly account for uncertainty (Uusitalo 2007), and their structure and results can be easily understood by non-domain experts (Reckhow 1999). In general, BNs have become especially popular over the last decade as recent practical implementations can now handle a significant number of variables (Pai et al. 2003; Wooldridge 2003). BNs can also handle situations where data is missing (Patcha and Park 2007, 3456). BNs are modeled with a directed acyclic graph where nodes represent variables and arcs between nodes depict conditional dependencies between them (Charniak 1991; Uusitalo 2007). Figure 11 shows an example BN with five variables.

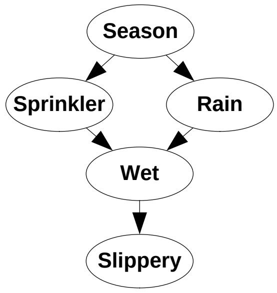


Figure 11: Example Bayesian Network (Pearl 1997, 2)

In this simple example of a Bayesian network, we model our understanding of the connection between the season, the status of a sprinkler, the presence of rain, the wetness of the pavement and whether the floor is slippery as a result. The network includes our belief that the effect of the season on the slippery variable can be expressed through intermediary variables. Thus, the BN represents a model of a certain environment in which, e.g., the wetness depends on the status of the sprinkler and the evidence of rain. Mathematically this is expressed via a conditional probability that only depends on the predecessors of a certain node. In this specific example this would be $P(Wet|Sprinkler, Rain)$. The joint probability distribution is the multiplication of all nodes' conditional probabilities (Pearl 1997). It may be calculated using the following formula, with the nodes $x_1, \dots, x_n$ and the parent nodes of x_i named pa_i (Pearl 1997, 2):

$$P(x_1, \dots, x_n) = \prod_i P(x_i|pa_i)$$

The CPT of each node includes the conditional probabilities. These may be ascribed based on expert input or also learned using data (Wooldridge 2003; Neapolitan 2003). Spiegelhalter and Lauritzen have demonstrated a fast Bayesian approach using Dirichlet random variables for cases where data for a node and its parent nodes are available (Spiegelhalter and Lauritzen 1990; Lauritzen and Spiegelhalter 1988). The distribution over a node's states is seen as a Dirichlet distribution $\sim Dir\left(\alpha_1, \dots, \alpha_j + 1, \dots, \alpha_n\right)$ if state j of the node is observed. The conditional probability of the node's state can then be calculated using $\alpha_j / \sum_i \alpha_i$. This follows under the assumption of independence of parameters (Spiegelhalter et al. 1993).

In the BN software used in this thesis, Netica, this CPT learning approach has been implemented as "counting-learning" (Norsys 2013a). Netica uses the concept of experience to update a BN after a finding (or a case) has been entered to recalculate CPTs. The new value of the experience is derived based on the old value. Here, the new value is depicted with a stroke ('). It should be noted that Bayesian updating in Netica, in contrast to learning CPTs, is performed using the conditional probabilities in the CPTs.

Experience: $exp' = exp + d$

Updated node state probability for node with finding: $prob' = (prob * exp + d)/exp'$

Updated node state probability for node without finding: $prob' = (prob * exp)/exp'$

BNs have previously been employed in various domains, including the detection of fraud or uncollectible debt (Ezawa and Schuermann 1995; Maes et al. 1993), facility location (Dogan 2012), or more related risk of water mains failure (Kabir et al. 2015), project risk management (E. Lee, Park, and Shin 2009), and the decision whether farms should remain organic (Gambelli and Bruschi 2010). BNs have also been widely applied in monitoring and detection systems, including applications such as treaty verification (King 1996) or intrusion detection (Sebyala, Olukemi, and Sacks 2002).

4.2 Approaches to observing child labor

In order to know which continuous evidence to integrate into the BN, potential indicators for child labor need to be better understood. In general, two fundamental approaches can be differentiated: directly measuring child labor or measuring drivers and disablers for child labor. Before discussing this trade-off and the strategy used in this thesis, it is necessary to discuss the drivers and disablers. These affect child labor in any part of a country and, consequently, also induce child labor in companies within a country. They are present at different levels and are differentiated below. Here, only child labor which meets the ILO definition is considered.

4.2.1 Child labor indicators

The discussion of indicators for child labor builds on key ILO and IPEC resources (International Labour Office 2013; International Labour Office 2004; International Programme on the Elimination of Child Labour, Statistical Information and Monitoring Programme on Child Labour, and International Labour Office 2004; International Programme on the Elimination of Child Labour and International Labour Office 2011) that should reflect the most important findings leading to child labor. Several multinational standards have been used to identify indicators based on questions and statements (Business Social Compliance Initiative 2009; Fair Labor Association 2011; UN Global Compact Nordic Network 2013; Global Reporting Initiative 2011; Global Social Compliance Programme 2010; Social Accountability International 2008b; The Danish Institute for Human Rights 2006).[41] These have been extended us-

[41] Several standards form the regulations that apply to child labor (i.e., the abolition of child labor) without detailing measures or other policies. Here we use the outline of standards and guidelines presented in the related work.

ing additional literature referenced in ILO publications, along with additional sources. The aim of the research is to present an overview of major indicators of child labor with no claim to completeness. It is assumed that by covering well-known standards and a broader range of literature the most important items will be included.

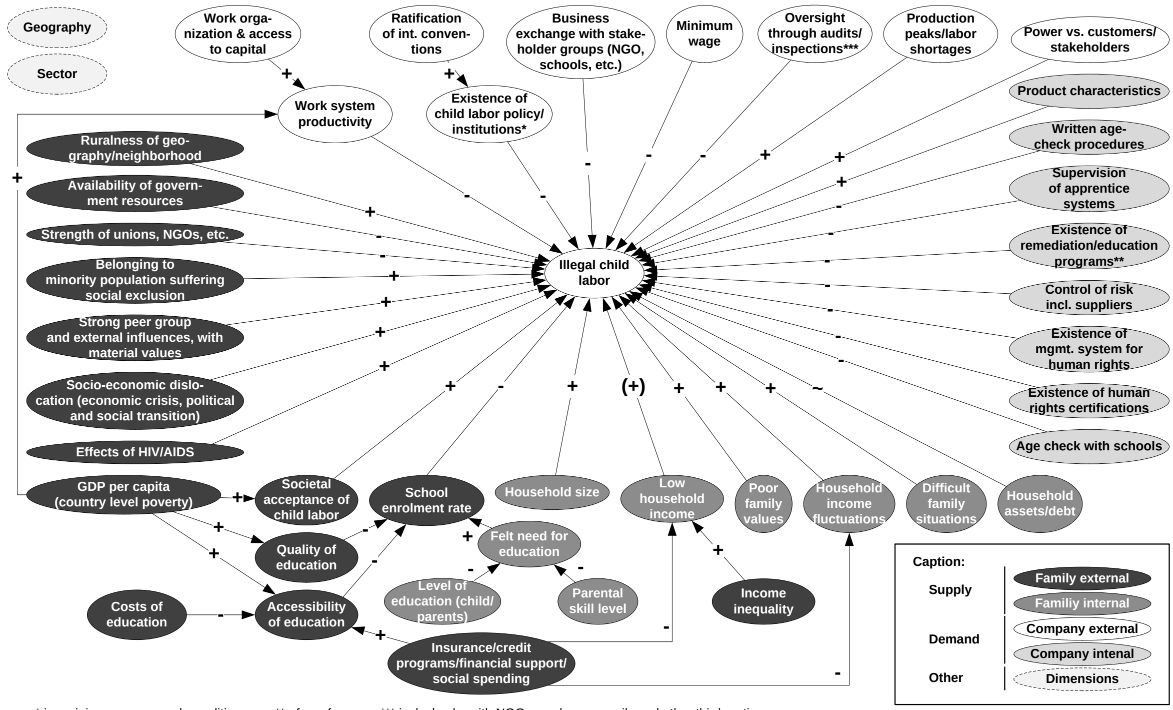
The identified indicators are shown in Table 14 and Figure 12. “Appendix F: Sources for child labor indicators” lists the sources used to create this list.

Primary Category	Secondary Category	Reason (indicator)
Demand[42]	**Company external**	Existence of child labor policy/institutions, Business exchange with stakeholder groups (NGO, schools, etc.), Minimum wage, Oversight through audits/inspections, Power vs. customers/stakeholders, Production peaks/labor shortages, Ratified international conventions, Work organization & access to capital, Works system productivity,
	Company internal	Age check with schools; Control of risk incl. suppliers, Existence of human rights certifications, Existence of management system for human rights, Existence of remediation/education programs, Product characteristics, Supervision of apprentice systems, Written age-check/identification procedures
Supply	**Family external**	Accessibility of education, Availability of government resources, Belonging to minority population suffering social exclusion, Costs of education, Effects of HIV/AIDS, Income inequality, Insurance/credit programs/financial support/social spending, National income (partly) / GDP per capita (country level poverty), Net school enrolment rates, Quality of education, , Societal acceptance of child labor, Socio-economic dislocation (economic crisis, political and social transition), Strength of unions, NGOs, etc., Strong peer group and external influences, with material values, Ruralness of geography/neighborhood
	Family internal	Difficult family situations, Felt need for education, Household assets/debt, Household size, Household income fluctuations, Level of education (child/parents), Low household income, Parental skill level, Poor family values

Table 14: Potential indicators for child labor in a country (author’s representation; sources according to Appendix F: Sources for child labor indicators)

Here we differentiate between the demand for child labor and the supply of child labor (i.e., the availability of children on the labor market), although both forces are intertwined. The former is further differentiated into company internal and company external indicators. The latter is split into family internal and family external indicators. The indicators have been put into the most appropriate category, as judged by the author. It is important to note that Figure 12 draws a direct effect linkage between an indicator and child labor if it is described as a cause without constraining other intermediary factors. Nevertheless, intermediate and other influencing factors might exist.

[42] Supplier size has been excluded based on the discussion in section 4.3.1.4 below.



* i.e. minimum wage, work conditions   ** of any form   *** incl. checks with NGOs, workers councils and other third parties

Distinction adopted partly also used by International Labour Organization (2004); Based on overarching studies of causes and reports of specialized organizations; Only relationships have been drawn that are explicitly mentioned and clearly underlined in the respective sources. Assumption: Existence of compulsory schooling; If direct effect on child labor rate stated then only that one shown – otherwise indirect relationship (e.g. school enrolment rate)

Figure 12: Indicators for child labor (author's representation; sources according to Appendix F: Sources for child labor indicators)

Altogether, a multitude of factors influence the level of child labor in a country and on a company level. In the context of this work, factors are of particular interest if they facilitate the prediction of child labor incidents at a company and can be monitored and integrated into the BN. It seems three criteria must be met for these factors to be of interest:

First, they should be monitorable based on information that is typically accessible through external (news) sources. If gathering the information requires asking suppliers or households for data, then external online monitoring would be contradicted. This excludes all family internal and nearly all company internal factors. Product characteristics can sometimes be gathered without help from the supplier; however, this is often strongly dependent on the sector.

Second, the factors need to show an effect within a reasonable amount of time. In this context "reasonable" may for example be interpreted as "shorter than the average time" between audits for an average supplier. From an external perspective it makes sense to only monitor these factors, as other factors can be better seen from within a company during an audit. Otherwise, as many factors influence the general child labor risk level of a country, the time between two child labor estimates may be another reasonable time limit. In both cases the upper limit would be in the order of months to one or two years and not more. Consequently, this excludes some factors from the list of potentially interesting factors.

Third, a change in the factor itself needs to be large enough and fast enough to provoke a measureable change in child labor at supplier locations. The timescales here are similar to the second point, for similar reasons.

Six factors meet all three of these criteria and are thus potentially relevant for external monitoring. Two of these (marked with *) arguably have the effects with the shortest effect times and which occur the most frequently:

- *Socio-economic dislocation (economic crisis, political and social transition)
- *Production peaks/labor shortages
- Oversight through audits/inspections (i.e., their existence e.g. carried out by third parties)
- Insurance/credit programs/financial support/social spending
- Minimum wage (e.g., newly introduced)
- Existence of child labor policy/institutions (e.g., newly introduced)

4.2.2 Child labor risk detection approaches

Section 4.2.1 discussed several indicators of child labor as stated in the literature. All of these are reasons or indicators for the existence of child labor, and changes in these factors, especially significant ones, can lead to an increase in child labor that is potentially also relevant for a specific supplier. As well as monitoring these indicators in news reports, another option could be to directly monitor child labor incidents that relate specific entities to child labor (as defined above).

Incident events that directly relate to a supplier factory may be rare. However, the predicted risk at a supplier location may increase if similar events occur in close proximity (e.g., through geography or industry) (Dreyer, Hauschild, and Schierbeck 2010a). Foerstl et al. indicate that both the geographic location and the production process (here understood as sector) influence "[…] probability of supplier non-adherence to sustainability standards […]" (Foerstl et al. 2010, 127). Building on this, we can differentiate four options for event-based child labor monitoring. The observed events will be either of the same or different type as the target event (here child labor incidents). Additionally, these events may directly refer to a supplier location or be only related. Figure 13 shows these differentiations.

Relatedness	Other event class	Same event class
Direct ref.	**Demand peak due to significant order of other buyer**	**Child labor incident at supplier factory**
Related to loc.	**Tsunami damaging large parts of the supplier's region**	**Child labor incident at competing supplier in same city**

Indicator type

Caption: Focus of this work

Figure 13: Example events to detect child labor at a supplier location (author's representation)

Non-child labor incident events such as socio-economic dislocations, production peaks, or labor shortages cover a wide array of happenings, including earthquakes, volcanic eruptions, strikes, or demand surges. Moreover, while the literature identifies connections between these events and child labor, the propensity of the effect clearly varies by context. Furthermore, these events are often only temporary.

While the effect of related child labor incidents on a supplier may be subtle, events related to child labor incidents in different news outlets are seen to be more directly related to child labor hotspots. This is partly due to the assumption that more investigation into child labor (e.g., by NGOs) will happen where more child labor is present. Furthermore, the non-direct events of other classes will typically have other consequences apart from child labor and will be investigated and detected without the need for specific child labor-focused systems (e.g., a volcanic eruption would directly disrupt supply chains and will also be discussed from a general risk management perspective). Therefore, this thesis focuses on observing events from "the same event class" as shown in Figure 13. Nevertheless, monitoring of the other events can make sense and may be used as additional input in other work.

4.2.3 Properties of child labor incidents

Child labor incidents have several properties that by themselves are not unique, but are characteristic for the domain. Although they are in some ways more related to the TM approach described in the next chapter, we will present them here jointly. The first primary characteristic of child labor incidents in news is their low frequency of occurrence. As will be explained in section 6.1.1, child labor events are rare in typical news streams. Within the Reuters corpus used for the English evaluation, 117 out of of 1.8 million articles were identified as depicting a child labor incident. Narrowing the initial dataset to articles that contain the character sequence “child” in any form produces a ratio lower than 1%, with 117 out of 16,948 articles.[43] Therefore, relatively few articles within Reuters refer to the same incident. Even for a “high attention” child labor incident, like the Primark incident in 2008 (Hopkins 2008), fewer than 40 articles can be found on Factiva[44].

Second, child labor incidents are not necessarily the main theme of a news article. News stories may take a broader outlook, or deal with a different or multiple topics, so a child labor incident may be mentioned within the overall text as a sub-statement. Table 15 provides statements or phrases from different news articles in the Reuters TRC2 dataset that feature child labor incidents relating to different layers of entity abstraction. The titles suggest that child labor is not the main theme of the articles and the value in the position column indicates that both instances occur close to the middle of the article.[45]

Article title	Child labor statement/phrase (excerpt)	Position*
UPDATE 1-US official decries shrimp industry 'slavery'	In Thailand, workers in shrimp-processing factories earned about $4.60 a day for a six-day work week. Child labor and forced labor were often the norm, the Solidarity Center said.	53%
Indian orphans weave award-winning movie magic	Victims end up in child labour or in the sex industry.	35%

* Percent of article characters before first character of statement/phrase

Table 15: Example of child labor statements in news articles from the Reuters data set (author’s representation)

Third, as is also slightly indicated in the first statement in Table 15, the attribute values of an incident may not be in the same sentence as the anchor for a child labor incident. Other authors also suggest that relevant facts for identifying an event may be spread over several sentences and their expression might vary (Huttunen, Yangarber, and Grishman 2002a; Piskorski et al. 2011). Consequently, approaches building on sentence structure may not be sufficient to determine the maximum number of attribute values available in an article.

Fourth, in contrast to broad event detection, risk monitoring of sustainability related events allows us to predefine the classes of interest. Consequently, filtering out unneeded events may also be based on expert knowledge. Nevertheless, the differentiation of individual events within the class is unknown. Finally, news texts often have a limited amount of available metadata. The only attribute assumed as given for all news reports is the date and time of pub-

[43] Filtering articles that remention an incident as well as summary reports affects roughly half of the 117 articles.
[44] https://global.factiva.com/ - a news aggregation service. Filtering is performed using metadata constraints.
[45] Additionally, this requires that parsing for child labor incidents includes the whole text of a news report. Approaches such as assuming that articles follow the pyramid principle in their structure and the main information is captured in the beginning (Piskorski et al. 2011, 184) appear to be inappropriate.

location. Additional attributes, particularly in the context of social media, could include author and/or an author's social network (Sankaranarayanan et al. 2009).

To sum up, child labor incidents show several characteristics not necessarily given for other types of events. Other events may be more "bursty" (W. Chen et al. 2010), generating attention through the volume of articles referring to it, or be more generic, for example reflecting a "meeting" or, even more generically, a "relationship" between entities. By referring to these domain-specific characteristics we will be able to tailor the intended system to them.

4.3 Child labor risk model elements

This section explains the elements of the suggested BN. Each supplier location is situated at a geographically unique location that may be exposed to different risk levels with regard to child labor (Schaltegger and Burritt 2014). Therefore, to create a sorted list of supplier locations based on their level of child labor compliance, a separate risk value is required for each supplier location. As a consequence, this thesis suggests creating a single template BN that is instantiated (i.e., copied) for each supplier location l allowing the calculation of a different breach probability for each location. A supplier location is understood as "[A] building or group of buildings where goods are manufactured or assembled chiefly by machine […]" (Oxford Dictionary 2014a) and which belong to one economic entity.

Consequently, it presents a novel approach to organize suppliers according to the risk of breaching child labor compliance standards. This is particularly necessary because the effects of child labor on reputation can be seen as being independent of purchasing volumes (compare e.g. with Foerstl et al. 2010, 127). Hence, ranking suppliers based on the probability of a child labor breach avoids the necessity of quantifying the impact of reputational damage. As the model's main purpose is to combine the inputs to produce a "reasonable" output, the inference approaches for the main inputs are presented in the following, detailing the contextual prior, audits, and observations.

4.3.1 Contextual prior

The prior specifies the likelihood of child labor in a company given the contextual attributes of a specific supplier location based on public statistical data from available international sources. Here, we require that these statistics should reference the lowest level of granularity provided while being produced to acceptable standards and available for a substantial number of locations worldwide. This is the only way the prior can be calculated for a sufficient number of sourcing locations and be sufficiently trustworthy. The key indicator suggested for this purpose is the child labor rate in different subsections and subcategories of a country.

A supplier location is defined by a specific geographic place, determined by the country and the "ruralness" (i.e., the distinction between urban and rural as used hereafter), and also by a sector, determined by the product sourced from that location (ISIC definition).[46] If multiple sectors can be attributed to a single supplier location, the predominant product produced at the location determines the sector. If necessary, the location can also be duplicated to model multiple sectors at the same location (i.e., multiple BNs could be used).

Understanding Children's Work (UCW; 2014), the joint project of ILO, Worldbank, and UNICEF, reports child labor statistics for many countries with a level of granularity that allows the calculation of separate shares for country subsections defined by the country, the sector, and the ruralness. Statistics are available at this level of granularity, and supplier locations can also be referenced with this scheme (at a certain level of granularity). It is rare for more detailed regional statistics to be available (particularly for large countries such as India and China). If found they may be integrated to further improve the model at a later stage. Nevertheless, due to frequently missing data and differences as well as ambiguities in the underlying surveys, one must be careful about which data is integrated in which form (Bourdillon et al. 2010, 19). In summary, the suggested (contextual) prior depends on the country, the sector, and the ruralness (see Figure 14).

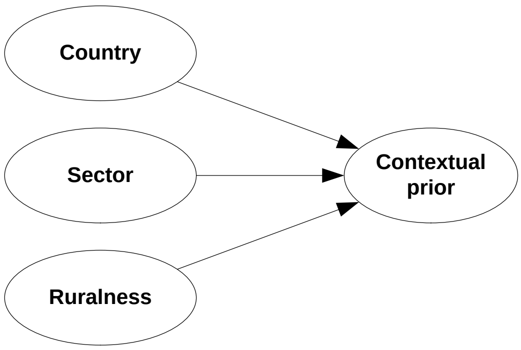


Figure 14: BN structure for the contextual prior (author's representation)

The contextual prior is assumed to be normally distributed, with the mean representing an estimate for the likelihood P_{prior} of child labor at a supplier location. The variance σ_{prior} also needs to be determined. Both the mean and the variance are defined via an external calculation discussed in the following paragraph.

[46] For more details on the ISIC classification of industries, see the UN homepage (United Nations Statistics Division 2014).

4.3.1.1 Calculation approach

The value of the prior is determined using a function that takes three arguments as input and returns the probability value that at least one child is employed at a company in a specific sector in the geographic area, narrowed by the country and the ruralness attribute. It is assumed that all potential company locations can be assigned to exactly one of these geographic areas. The main data for the calculation is taken from UCW (Understanding Children's Work 2014). Nevertheless, conversions and year adjustments of the data are sometimes necessary. The calculation is structured so that it is possible to derive a value with only limited input from country-level statistics from national bureaus. However, if better country level statistics are available these values can be used to eliminate some approximations. For example, some countries may report the absolute number of children working in a certain industry and ruralness directly.

UCW provides the following absolute and relative numbers based on national statistics: children are defined as people between 5 and 14 years of age; all calculations are assumed to be country specific (Understanding Children's Work 2010).

$NC_C\ldots$ Number of children in country C

$NCL_x\ldots$ Number of children working in a specific context x

$PCL_C\ldots$ % of children working in country C

$PCL_{C,R}\ldots$ % of children working per ruralness R (urban or rural areas)

$RPCL_{C,R,S}\ldots$ % share of sector of total children working in a sector S per ruralness

$where\ C \in \{Indonesia, India, \ldots\}; R \in \{urban, rural\}; S \in \{Agriculture, Mining, \ldots\}$

Based on the first value, which gives an absolute number of children in the relevant age group together with the relative sizes of the population given through values 2 and 3, the number of children per ruralness can be calculated with the following formula (see "Appendix G: Child labor per ruralness" for the formulation).

$$NC_{C,R_1} = \frac{NC_C * PCL_C - NC_C * PCL_{C,R_2}}{PCL_{C,R_1} - PCL_{C,R_2}}; R = \{R_1, R_2\}; R_1 \neq R_2\ ; R_1, R_2 \in \{urban, rural\}$$

Multiplied by the sector shares (i.e., the sector's relative share of all working children) for child labor, this allows the absolute total number of children working in a specific country, sector and ruralness to be calculated:

$$NCL_{C,R,S} = RPCL_{C,R,S} * \left(PCL_{C,R} * NC_{C,R}\right)$$

To be able to relate the values used in the model with a supplier and its sector, they should all ideally refer to the same revision of the industry classification applied. For the prior, the industry classification used also depends on the timing of the underlying child labor study. For example, the Indonesian Child Labor Survey from 2009 refers to International Standard Industrial Classification of All Economic Activities (ISIC) in its 3rd revision (Understanding Children's Work 2012, 23).

Another issue of definition arises with the understanding of urban and rural. The model makes no differentiation between possible different usages. Consequently, numbers might change if a different definition of urban and rural is used in different child labor statistics. Nevertheless, by taking a similar statistical base for all input parameters, the effect of slight differences in definitions of urban and rural can be reduced when considering only the ultimate goal of ranking supplier locations. Additionally, the current model assumes that by focusing on a particular sector's share of child labor, the effects of purely domestic work (e.g., household chores) without economic output to a third party are not taken into account. Otherwise, additional data would be necessary that would further extend the one presented in the next table. In top-level country statistics this usually appears to hold, as can be seen in Table 16 where household chores appear to not be included in the values given by UCW:

Residence	Agriculture	Mining	Manufacturing	Construction	Commerce	Hotels, restaurants	Transports	Real estate	Public administration	Education	Other com. services	Private house holds
Urban	6.6	0.0	34.2	8.6	23.6	11.2	3.4	0.3	0.8	0.9	6.1	4.3
Rural	79.1	0.6	9.7	3.3	2.8	1.0	0.7	0.0	0.0	0.1	2.1	0.5
Total	69.5	0.5	13.0	4.0	5.6	2.4	1.1	0.0	0.1	0.2	2.7	1.0

Table 16: Sector of child economic activity by residence in India as a percentage of children in economic activity from age 5 to 14 (India, National Sample Survey Round 66 (NSS-R66) 2009-2010; Understanding Children's Work 2010)

Residence	Employee	Self employed		Unpaid family worker
Urban	27.0	7.4	36.9	28.8
Rural	3.6	5.3	62.8	28.3
Total	6.7	5.6	59.3	28.3

Table 17: Modality of child economic activity by residence in India as percentage of children in economic activity from age 5 to 14 (India, National Sample Survey Round 66 (NSS-R66) 2009-2010; Understanding Children's Work 2010) – self-employed as used in source

The example of India suggests additional details for discussion. Based on Table 17, a significant number of children are self-employed, i.e. work outside the context of a specific supplier. Self-employed people "[…] have *autonomy* (decide how, where and when to produce) and *economic independence* (in respect of choice of market, scale of operation and finance) for carrying out their operation." (National Sample Survey Office 2011, 15). According to the definition of the underlying NSS-R66 survey, unpaid family workers fall into the category of "casual labor" and particularly under "worked as casual wage labour in other types of works" (National

Sample Survey Office 2011, 13). Therefore, this thesis excludes (only) self-employed children from the calculation, yielding the adjusted amount of child labor in absolute numbers in the special case of India (it is assumed that the ratio of self-employed is similar across all sectors). Nevertheless, this might also be appropriate for other countries. Specifically, unpaid family work is unrelated to non-market economic activities that do not lead to market-related production activities, which would therefore be irrelevant for supply chains (see e.g. Hagemann et al. 2007, 22 for a distinction of market and non-market work). Reports such as the child labor report on Indonesia specifically focus on market-related working activities (Badan Pusat Statistik and International Labour Organisation 2010, 14) and, hence, children who are engaged in potentially supply-relevant activities through their unpaid family activities. However, this cannot be guaranteed in all cases and a distinction has to be made with regard to context and specific input statistics. These children might also be working in the context of a supplier. It is important to determine whether a specific child labor study includes only economic child labor as defined in statistical guidelines and consequently all family work is therefore economically relevant (for a discussion see e.g. Understanding Children's Work 2007):

$PSECL_{C,R}$ … % of children working and categorized as self-employed per ruralness

$$ANCL_{C,R,S} = NCL_{C,R,S} * \left(1 - PSECL_{C,R}\right)$$

In order to be able to develop an understanding of the probability of one company employing children, additional data is required. Here, two options are discussed. In both cases, a very basic assumption is that child labor is a perfect substitute for adult labor. This assumption has also been used by other authors (Dinopoulos and Laixun Zhao 2007):

It is assumed that either (1) all child workers are randomly assigned to companies in the specific geographic area or (2) the child labor rate depends on a company's size. Given the number of employees at a supplier location, x employees are randomly selected from the "extended workforce" that includes the absolute number of children in the area and sector. For approach (2) a more detailed discussion of the relation between the size of a company and child labor risk is necessary (See Section 4.3.1.4). In short, a relationship between a company's size and the probability of employing children cannot be scientifically supported yet. As a result, we suggest using the first approach. The second will be shown for reference only.

For approach (1) the number of companies is required as input given the ruralness, sector, and country. For India this value can be derived from the 2005 national census. Here, all companies except those working on their own account are considered because these do not usually recruit any further employees. In some cases adaptations to the number of companies are necessary – a brief discussion can be found in Section 4.3.1.5:

$NCOMP_{C,R,S}$ ... # of not own account companies per ruralness and sector

Therefore, the probability that a child is not linked to a specific company is then:

$$1 - \frac{1}{NCOMP_{C,R,S}}$$

Consequently, if $ANCL_{C,R,S}$ children are working then the probability that all children are not working at the specific company is:

$$\left(1 - \frac{1}{NCOMP_{C,R,S}}\right)^{ANCL_{C,R,S}}$$

The probability of having at least one child working at the company follows with the complementary event:[47]

$$P_{CL} = 1 - \left(1 - \frac{1}{NCOMP_{C,R,S}}\right)^{ANCL_{C,R,S}}$$

One assumption that must hold is that each child in the work force is only working at one company.

Approach (2) would require knowing the number of employees per geographic region (country, ruralness) and sector $NEMP_{C,R,S}$[48]. As above, this number can often be derived from national statistics.

$NEMP_{C,R,S}$ ... # of employees per country, ruralness and sector

Given this number, the odds that there are children in the extended employed workforce can be calculated, yielding the probability of randomly selecting a child:

$$PCEMP_{C,R,S} = \frac{NCL_{C,R,S}}{(NEMP_{C,R,S} + NCL_{C,R,S})}$$

---

[47] An alternative is to use combinatorial counting: determine the probability that no child is assigned to a company if the children working in the same geographic area and sector are randomly assigned to the companies in the area and sector. Consequently, the probability of the complementary events yields the probability that one or more children are assigned to a company, i.e., that child labor is present in a random company in the geographic area and sector. This gives the probability of child labor, P_{CL}:

$$P_{CL} = 1 - \frac{\binom{NCOMP_{C,R,S} + ANCL_{C,R,S} - 2}{ANCL_{C,R,S}}}{\binom{NCOMP_{C,R,S} + ANCL_{C,R,S} - 1}{ANCL_{C,R,S}}}$$

[48] Another way to calculate this value would be to use the employment ratios and percentage employment per sector provided by ILO and population provided by Worldbank

Finally, drawing without replacement for randomly selecting children from the extended employed workforce is assumed. Thus, the probability of picking at least one child is the probability of the complementary event of picking no children when selecting a certain number of employees, x. The probability of selecting no child can be calculated using a hypergeometric distribution:

$$P_{CL}(x) = 1 - \frac{\binom{NCL_{C,R,S}}{k=0} * \binom{(NEMP_{C,R,S}+NCL_{C,R,S})-NCL_{C,R,S}}{x-(k=0)}}{\binom{(NEMP_{C,R,S}+NCL_{C,R,S}}{x}} = 1 - \frac{\binom{(NEMP_{C,R,S}+NCL_{C,R,S})-NCL_{C,R,S}}{x}}{\binom{(NEMP_{C,R,S}+NCL_{C,R,S})}{x}}$$

The descriptions above hold when the data is available for the exact year needed in the model. However, under practical circumstances life data and particularly survey data on child labor is only published at very infrequent intervals. Consequently, it is necessary to apply growth rates to different input data items in order to calculate a prior value for a particular year. Examples of the adjustments needed are given below for the case of India.

4.3.1.2 Standard deviation of contextual prior

The value of the standard deviation of the contextual prior is assumed to be independent of a particular country, sector, or ruralness. If they are not based on the same measurement instruments, child labor estimates may show differences in results (Guarcello et al. 2010).[49] It seems to be particular important to factor for measurement error, as measures for school attendance show a lower amount of variation (Guarcello et al. 2010). Guarcello et al. (2010) suggest that the survey design and the season in which a particular survey is completed can have a strong impact on the survey's results. Consequently, different reasons might drive the differences in results and need to be accounted for in any model using child labor rates in the prior.

To derive a standard error, it is assumed that the errors made in calculating the probability of a child labor incident can be compared those made when estimating child labor in a country, given the comparable difficulties. As the child labor rate is used to calculate the contextual prior, the empirical standard error will be a lower (thus optimistic) estimator for the true error, given that additional empirical variables are used in the calculation.

The suggested way to calculate the variance is to analyze the accuracy of other attempts to measure child labor. Accordingly, the estimated child labor rates from different statistical sources are taken as input and adjusted by their mean to a mean of zero. We used the nine

[49] Guarcello et al. (2010, 7) suggest three different reasons for errors: sampling and non-sampling error as well as the design of the questionnaire/survey.

countries in the analysis by Guarcello (2010, 10). The 18 child labor estimates for these countries all refer to the same reference period. The resulting standard deviation is then taken as an input value for the prior.[50] Consequently, the prior for the standard error of the prior σ_{prior} has been fixed at 13.32.

Country	Absolute difference	Adj. absolute deviation from mean per country
Bangladesh	14.9	7.45
Bolivia	9.1	4.55
Brazil	0.3	0.15
Cameroon	48.3	24.15
Ghana	26.5	13.25
Kenya	36	18
Lesotho	30.9	15.45
Sao Tome e Principe	16.7	8.35
Senegal	12.9	6.45
Average	21.73	
Standard error		**13.32**

Table 18: Estimation of standard error of child labor rate calculations based on country comparisons by Guarcello (2010, 10; author's representation)

4.3.1.3 Excursus: Example calculation for two sectors in India and Indonesia

Table 19 shows an initial calculation for India as an example for the manufacturing sector, divided into rural and urban.

[50] Calculated based on absolute deviation from assumed mean between data points per country.

Data	Variable description (each per country)	Variable	Unit	2005	2010	2012
UCW	# of children	NC - C	#		221,737,447	227,450,359
	% of children working per ruralness	PCL - C,U	%		1.10%	0.91%
		PCL - C,R	%		2.20%	1.82%
	% of children working - total	PCL - C	%		2.00%	1.65%
	# of children per ruralness	NC - C,U[51]	#		40,315,899	41,267,422
		NC - C,R	#		181,421,548	186,182,936
	# of children working per ruralness	NCL - C,U	#		443,475	374,999
		NCL - C,R	#		3,991,274	3,383,704
	% of children working relative to other sectors per ruralness	RPCL - C,S,U	%		34.20%	34.20%
		RPCL - C,S,R	%		9.70%	9.70%
	% of children working self-employed per ruralness	PSECL - C,S,U	%		44.30%	44.30%
		PSECL - C,S,R	%		68.10%	68.10%
	# of children working relative per sector & ruralness excl. self-employed	ANCL - C,S,U	#		84,479	71,435
		ANCL - C,S,R	#		123,502	104,702
The World Bank	# of population	POP - C	#	1,127,143,548	1,205,624,648	1,236,686,732
	population growth in %	POPGR - C	%		2010-2012:	1.03
	# of population per ruralness	POP - C,U	#		372,899,704	391,535,019
		POP - C,R	#		832,724,944	845,151,713
	% of population per ruralness	PPOP - C,U	%		31%	32%
		PPOP - C,R	%		69%	68%
CL Trend	CAGR of % of children working in Asia (2008-12)[52]	PCCL - Asia CAGR				-9.1%
5th econ. census India	# of companies per sector & ruralness excl. own-account enterprises	NCOMP - C,S,U	#	1,648,187	1,762,947	1,808,369
		NCOMP - C,S,R	#	1,580,964	1,691,044	1,734,612
	# of workers per sector & ruralness - incl. children	NEMP - C,S,U	#	9,779,584	10,460,520	10,730,028
		NEMP - C,S,R	#	8,169,936	8,738,795	8,963,944
	Avg. # of workers per company per sector & ruralness (urban)	AEMP - C,S,U	#		6	6
		AEMP - C,S,R	#		5	5
Own calculation	Probability of avg. manufacturing company employing children per sector & ruralness	PCL – C,S,U (approach 2)	%		3.9%	3.3%
		PCL – C,S,R (approach 2)	%		6.8%	5.6%
	Probability of a manufacturing company employing children per sector & ruralness	**PCL – C,S,U (approach 1)**	**%**		**4.7%**	**3.9%**
		PCL – C,S,R (approach 1)	**%**		**7.0%**	**5.9%**

Note on abbreviations: C…Country, S…Sector, U…Urban, R…Rural, #...Number, CAGR…Compound annual growth rate

Table 19: Example calculation for India (author's representation) – numbers based on multiple authors (Diallo et al. 2013; International Programme on the Elimination of Child Labour 2013; Ministry of Statistics and Programme and Implementation of India 2006; The World Bank Group 2013; Understanding Children's Work 2010)

This calculation was repeated for another sector for comparison and also transferred to Indonesia. The results are presented in Table 20.

[51] The increase in urban children is based on the increase of people living in cities in relation to those living in rural areas, and not on the direct population increase (assumption: if people do not move to cities then children grow up in rural areas – if cities grow over-proportionally people tend to move together with their children). The latter approach should be taken for future calculations. In total, differences in the final child labor probability amount to only 0.1 percent points for the example provided here.

[52] Corresponds roughly to the mean of the CAGRs for the 2008-2012 children working values published in two related ILO publications (Diallo et al. 2013; International Programme on the Elimination of Child Labour 2013). Reflects CAGR of percent points.

2012 values			Sector	
Method	**Country**	**Ruralness**	**Manufacturing**	**Construction**
Nr. of CL company independent (approach 1)	**India**	Urban	3.9%	23.6%
		Rural	5.9%	44.7%
	Indonesia	Urban	4.9%	0.0%
		Rural	2.6%	1.5%
Nr. of CL company dependent (approach 2)	**India**	Urban	3.3%	19.2%
		Rural	5.6%	34.3%
	Indonesia	Urban	4.3%	0.0%
		Rural	2.5%	1.5%

Table 20: Comparison of example child labor incident priors for different approaches, countries, ruralness, and sectors (author's representation) – numbers based on multiple authors (BPS Statistic Indonesia 2008; Diallo et al. 2013; International Programme on the Elimination of Child Labour 2013; Ministry of Statistics and Programme and Implementation of India 2006; The World Bank Group 2013; Understanding Children's Work 2010)

As shown in Table 20, the values for the mean of the prior probability distribution for a child labor incident at a supplier location vary significantly between countries and sectors. The type of calculation approach has a limited influence, with approach 1 producing slightly higher probability values. Consequently, the selection of the appropriate calculation mechanism for the final probability value might be of limited importance.

4.3.1.4 Excursus: Company size and child labor probability

One approach described above uses the assumption that an increase in employees leads to a higher probability of employing children. Literature and statistical reports provide mixed support for this.

Studies on the audit performance of labor standards suggest that larger factories have worse sustainability performance than smaller one (Locke, Qin, and Brause 2007). Applying these results to child labor requires the plausible assumption that performance on the child labor dimension of labor rights is correlated with general performance. This is partly supported by the "code of conduct" tested by Locke, Qin, and Brause, which includes child labor explicitly (Locke, Qin, and Brause 2007). Three further surveys conducted on carpet factories in India, Pakistan, and Nepal and funded by the U.S. Department of Labor and ICF International also provide comparable evidence (A. Hansen and Rosell 2012a; A. Hansen and Rosell 2012b; Rosell and Hansen 2012). The following tables present the results for the three countries.

Factory size (Number of employees)	**Carpet factories in India**		**Child carpet workers in India**	
	Total		**Total**	
	N	**%**	**N**	**%**
Small (<= 5 workers)	3,738	50.2	135	17.9
Medium (6-10 employees)	2,971	39.9	410	54.2
Large (>= 11 employees)	741	9.9	211	27.9
Total	**7,449**	**100**	**757**	**100**

Table 21: Estimated population of children working in carpet factories by factory size in India (adapted from Rosell and Hansen 2012, 55)

Factory size (Number of employees)	Factories		Child carpet workers	
	N	%	N	%
Small (<= 5 workers)	256	39.7	17	1.4
Medium (6-10 employees)	201	31.2	235	19.1
Large (>= 11 employees)	188	29.1	980	79.6
Total	**646**	**100**	**1,232**	**100**

Table 22: Estimated population of children working in carpet factories by factory size in Pakistan (adapted from A. Hansen and Rosell 2012a, 47)

Factory size (Number of employees)	Factories		Child carpet workers	
	N	%	N	%
Small (<= 5 workers)	40	5.6	13	0.6
Medium (6-10 employees)	127	17.7	76	3.5
Large (>= 11 employees)	547	76.6	2,070	95.8
Total	**714**	**100**	**2,160**	**100**

Table 23: Estimated population of children working in carpet factories by factory size in Nepal (adopted from A. Hansen and Rosell 2012b, 45)

Although all three tables indicate an increase in children per factory with an increase in factory size, the factories analyzed in these studies are generally rather small (the largest category is defined as having eleven or more employees). Consequently, a generalization that similar logic also holds for companies exceeding 100 employees is not supported.

A diametrically opposite conclusion can be drawn from a study performed by the Asia Monitor Resource Center in 2008. In a survey of 37 companies, companies with less than 200 workers were roughly four times more likely to employ children than companies with more than 2,000 workers (Asia Monitor Resource Centre 2008). Additionally, Williams (2003) states that smaller companies are more likely to face constraints in their social initiatives, fostering child labor.

In summary, there is evidence for and against the theory that the probability of child labor increases with factory size. Although support for an increase appears to be stronger, caution should be exercised when building on this assumption. Consequently, approach (2) would need to be thoroughly discussed before being implemented. In addition, it would also require a change to the model by adding an additional size node to the prior.

4.3.1.5 Excursus: Capturing the informal sector with a focus on India and Indonesia

Additional scrutiny must be given to the informal sector, i.e., roughly unregistered economic activities (OECD Glossary 2015). Calculating the size of the informal sector, especially in developing countries, is challenging. For example, India has no unique registration procedures for companies, making the definition of an appropriate split between the formal and informal sectors difficult (Naik 2009, 3). Nevertheless, in India the 5th economic census used a broad definition of enterprise, stating that an "[…] enterprise is an undertaking engaged in production and / or distribution of goods and / or services not for the sole purpose of own consumption." (Ministry of Statistics and Programme and Implementation of India 2006, 17). This includes informal production if it "[…] consists of all unincorporated private enterprises owned by individuals or households engaged in the sale and production of goods and services operated on a proprietary or partnership basis and with less than ten total workers […]" (Naik 2009, 3). Therefore, the value was not adjusted any further in the example for India below.

The same line of argument has been followed for Indonesia, for which the economic census used defines its coverage as "[…] all economic activities (establishment/enterprise/business) in both permanent and non-permanent places and conducted in Indonesia." (BPS Statistic Indonesia 2008, 4). Similarly, the ILO concept of the economically active population is broad:

> *The economically active population comprises all persons of either sex who furnish the supply of labour for the production of economic goods and services as defined by the United Nations systems of national accounts and balances during a specified time-reference period. According to these systems the production of economic goods and services includes all production and processing of primary products whether for the market for barter or for own consumption, the production of all other goods and services for the market and, in the case of households which produce such goods and services for the market, the corresponding production for own consumption. (International Labour Organization 1982, 2)*

Nevertheless, for some countries it might be necessary to make adjustments to include the informal sector. One option would be to calculate an adjusted number of companies based on the number of people working in the informal sector. Assuming a similar distribution of "companies" in the informal sector, an approximate value for the size of the informal sector can be derived by comparing the number of employees in the formal and informal economies. Consistently, the number of companies can be fictively adjusted to account proportionally for the scenario that a child works in the informal economy. The probability of children working in the formal and informal economy is assumed to be equal. The adjusted number of companies $ANCOMP_{C,R,S}$ is then given by:

$NEMP_C$ … # of employees per country

$INEMP_C$ … # of people working in the informal economy

$$ANCOMP_{C,R,S} = NCOMP_{C,R,S} * \frac{NEMP_C + INEMP_C}{NEMP_C}$$

4.3.2 Audits

Audits are handled in a similar way to observations in the sense that they are used to revise the initial prior probability to determine the breach probability[53]. A higher number of compliance audits does not suggest a supplier is better. Often the compliance level of the supplier stays the same and sometimes also worsens (Locke, Qin, and Brause 2007). Audits are limited in what they can measure and can only be conducted within a defined timeframe, leaving the suppliers alone before and after this timeframe (Locke, Amengual, and Mangla 2009). Consequently, the model in this thesis captures the last audit result and includes the time since the last audit as an approximation of the variance.

The audit score is assumed to be a continuous stochastic variable A with a minimum audit score $a_{min} \geq 0$ and a maximum audit score a_{max}. For reasons of simplicity, in the basic implementation it is assumed that the user has configured the audit score such that the score reflects the likelihood that child labor will occur at a supplier location. This is modeled through a normal distribution for the audit likelihood P_{audit}, i.e., the likelihood of a child labor incident assuming the audit results are normally distributed, with the mean probability being directly dependent on the expected audit score $E(a)$. For this, a linear relationship is suggested where an audit score of a_{max} indicates that the expected value of P_{audit} is m and an audit score of zero means a value of n is expected. The mean of P_{audit} is derived with the following formula:

$$E(P_{audit}) = m - E(a) * \frac{m - n}{a_{max}}$$

Both m and n are later derived via a questionnaire, as discussed in chapter 6. The variance σ_{audit} is determined through a preconfigured value depending on the number of months since the last audit. Alternatively, the variance can also be modeled as a function of the months since the last audit. This is left for future work.

To calculate the prior for A, two options seem promising: the database-derived empirical distribution of earlier audits or the expected distribution of audit results from a normative perspective. This may be an audit of the mother company or another related supplier location.

Figure 15 shows the overall structure of the BN for the audit probability.

---

[53] Audits can also be understood as a form of certification if the audit is not performed internally. Instead of adding an additional node to the model, the last certification score could replace or update the internal audit score.

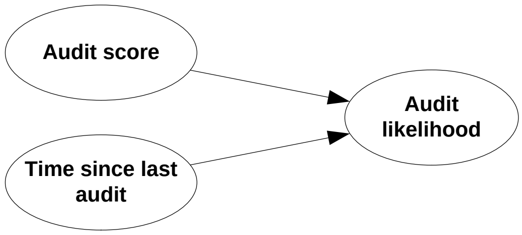


Figure 15: BN structure for audit likelihood (author's representation)

4.3.3 Observations

This thesis suggests using observations as an external input into the risk model for supply chain child labor risk management. In this thesis we assume that a news report can only depict a single child labor incident, as defined in the related work chapter (Chapter 2). As the attributes of an incident event may include several values, a single incident event can be split into multiple incident observations that relate multiple entities to child labor. For example, a single child labor incident event may reference two supplier names. This would result in two distinct incident observations. Nevertheless, each observation may still be related to several supplier locations (e.g., the supplier name refers to all factories associated with this supplier). The number of observations is defined as the number of possible incident event combinations. These observations have a unique combination of corresponding attributes. The connections are shown in Figure 16 in the form of a UML class diagram. The frame (i.e., the set of dimensions/attributes) is similar for child labor incident events and child labor incident observations, with the only difference being that observations have single-valued attributes.

Figure 16: Child labor incident domain model for events and observations (author's representation)

While this assumption appears very restrictive, the manual tagging of the training corpus used in the TM evaluation showed no conflicts.[54] The risk model needs independent child labor incident events as input. In literature, "independent" is upon other defined as "[N]ot connected with another or with each other; separate" (Oxford Dictionary 2014b). Only in this case each independent report may be seen as potentially providing additional evidence for a child labor incident. In other words, as it will be used in this thesis, independent observations should not refer to the same child labor case or (at least) be independently verified and investigated. In contrast if a news report has been copied verbatim, for example, then this additional report does not provide further evidence. Therefore, the software component that provides

[54] Conflicts could arise if a news report contains more than one incident observation, with one attribute having a single possible value that should be present for one incident observation but not the other.

the inputs to the risk model needs to provide these independent observations. In general, it is assumed that the more independent reports that can be gathered, the more likely a child labor incident is evident at a supplier location.

The following begins by specifying dimensions that allow supplier locations to be linked with observations, allowing these observations to be integrated into the risk model. Second, the process of finding the right observations will be discussed. Third, the use of credibility and relevance will be explained to show how observations will be integrated. The final section explains the standard deviation for observational likelihood and how this is updated.

4.3.3.1 Dimensions

Supplier locations and their relation to the risk model have already been defined above. In brief, they are production or other locations of a supplier company that have several predefined attributes. They are related to one or more companies through ownership structures (it is assumed that a company owns a supplier location), they are situated at a specific geographic location (inherently necessary due to their physical structure), and they produce or offer one or multiple goods or services (see also Oxford Dictionary 2014a for factory). Thus, supplier locations can also be related to one or several industry sectors, given that they produce or offer something (i.e., they have a purpose, even if it is only inward-oriented). Several other attributes could be used; however, they would not help in relating observations to BNs as they rarely appear in news text. Subsequent extensions to the model might use them.[55]

A child labor observation relates one or more entities with child labor. The closest and most granular pairing of a child labor incident event to a supplier location is given when they directly share location, sector, and company attribute values. Figure 17 shows the different dimensions which are used for this matching.

[55] Alibaba (http://www.alibaba.com), a B2B online retail platform, suggests several additional attributes as the main introductory attributes for manufacturers. The first is revenue. However, revenue is rarely available at a supplier location level. Based on the discussions provided above and its irrelevance to reputational damage (Foerstl et al. 2010), size has been excluded from the risk model. Similarly, attributes like the age of the location and main markets are not considered when aiming to link an incident with a supplier location. In addition, several attributes can be derived from the supplier relationship between a company and its supplier, such as order volume or audit scores. However, these are attributes of the relationship and not of the supplier location itself.

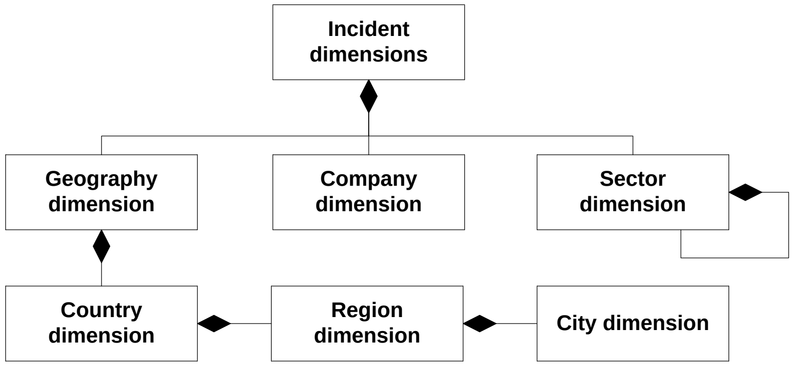


Figure 17: Dimensions used for event extraction (author's representation)

The geographic reference is an inherently hierarchical attribute. While other hierarchical relationships have the same depth for each leaf value, the geographic hierarchy depends on the political structure of the governing administrative region. Consequently, three important elements of this hierarchy have been selected to represent this dimension: country, region, and city. For discriminative specification, the GeoNames[56] ontology and its attributes is suggested due to its broad usage. GeoNames uses feature codes to define specific geographic entities. We use these codes in this thesis and suggest a differentiation of the three concepts as presented in Table 24.

Type	GeoNames feature code(s)	GeoNames definition
Country	A.PCLI, A.PCLS	"Country, state, region, [...]": independent political entity, semi-independent political entity
Region	A.*\{A.PCLI, A.PCLS}	All other "countries, states, regions, [...]" except historical
City	P.*	All "cities, villages, [...]"

Table 24: Differentiation of geographic concepts according to GeoNames - * indicates placeholder (based on GeoNames 2014)

The company dimension of an incident event covers all references within an incident to a company or another organization. Finally, the sector dimension covers references to a sector which should separate from the product level. Nevertheless, mentions of products may be converted to sector references. Again, the sector attribute has a hierarchical structure as industry classifications are typically built in several layers (e.g. the already mentioned ISIC classification of industries shows this structure; United Nations Statistics Division 2014).

4.3.3.2 Retrieval

Next, the process of retrieving relevant independent incident observations (IIO) for a specific supplier location needs to be specified. This retrieval task defines which observations are used for the risk calculation for each supplier location (the quality of the information will be dis-

[56] http://www.geonames.org/

cussed below). This builds on the ideas regarding the relatedness of child labor incidents depicted in Section 4.2.2. The IR task is described as follows:

BN_l… Bayesian network of supplier location l
IIO_i… Independent incident observation i
Dim_l^{loc}… Set of dimensions that specify a supplier location l
Dim_i^{iio}… Set of dimensions that specify an IIO i
$geo, comp, sec \in Dim$… Variables for geography, company, sector
$geo \in Countries \vee geo \in Regions \vee geo \in Cities$… Geography is a country, a region, or a city
G… Set of geographical references that form the hierarchy of a location (incl. city)
S… Set of sectoral references that form the hierarchy of the sector
$\mathrm{Dim} = \mathrm{G} \cup \{comp\} \cup \mathrm{S}$

We retrieve the set of relevant IIOs that need to be included in the BN of a location l to update the risk value of the location. In the following, only one supplier location will be considered and thus the index l is omitted. The sets of filters F_2 and F_3 for process steps 2 and 3 below are user-defined and should depend on the amount of influence a slightly related IIO should have on a location. In total, the retrieval is performed in three process steps:

Process step 1 – load initial list of maximum relevant IIOs for a location:

First, all IIOs are retrieved into set R for which the geography or sector of the observation (IIO) is in the appropriate hierarchy of the location or the company is similar:

$$R = \left\{IIO_i \middle| geo_i^{iio} \in Geo^{loc} \vee comp_i^{iio} = comp^{loc} \vee sec_i^{iio} \in Sec^{loc}\right\}$$

Process step 2 – filter all IIOs that do not conflict with any location dimensions:

The second step filters all IIOs from R that fulfill the criteria defined with filter F_2. F_2 specifies the dimension combinations an IIO can have that would lead to omitting that observation from the risk calculation, despite having no particular conflict with the dimensions of the supplier location. A dimension combination of a filter f is a triple composed of the possible dimension combinations including the empty value, e.g., ‹city, company, empty› or ‹country, empty, sector A02›. It is applied to those observations in set R that have no conflicting dimension with the location (i.e., the dimensions of the observation are either in the hierarchy of the location, similar, or empty):

$$F_2 = \{f_1 \dots f_n\}, n \geq 1 \dots Set\ of\ filters\ f$$

$$f_j = \langle geo^f, comp^f, sec^f \rangle \dots Filter\ specified\ by\ three\ dimension\ states$$

$$geo^f, comp^f, sec^f \in \{0,1\}$$

$$D_2 = \left\{ IIO_i \in R \middle| \begin{array}{c} \forall f_j \in F_2 where \\ \begin{cases} geo^f = 1: geo_i^{iio} \in Geo^{loc} \\ geo^f = 0: geo_i^{iio} \in \emptyset \end{cases} \wedge \\ \begin{cases} comp^f = 1: comp_i^{iio} = comp^{loc} \\ comp^f = 0: comp_i^{iio} = \emptyset \end{cases} \wedge \\ \begin{cases} sec^f = 1: sec_i^{iio} \in Sec^{loc} \\ sec^f = 0: sec_i^{iio} \in \emptyset \end{cases} \end{array} \right\}$$

$$R_{F2} = R \backslash D_2$$

Process step 3 – filter all IIOs that conflict with one or more location dimensions:
The third step filters all IIOs from R that do not fulfill the criteria defined with filter F_3. The filter F_3 defines the dimension combinations of an observation (IIO) for which the overlap with the location dimensions is insufficient to warrant inclusion in the risk calculation. Only dimensions with at least one non-empty dimension value that are not in the hierarchy of the location and not similar are considered. As the filtering is indirectly based on the set R, at least one overlapping dimension always exists:

$$F_3 = \{f_1 \dots f_n\}, n \geq 1 \dots Set\ of\ filters\ f$$

$$f_j = \langle geo^f, comp^f, sec^f \rangle \dots Filter\ specified\ by\ three\ dimension\ states$$

$$geo^f, comp^f, sec^f \in \{0,1\}$$

$$D_3 = \left\{ IIO_i \in R_{F2} \middle| \begin{array}{c} \forall f_j \in F_3 where \\ \begin{cases} geo^f = 1: geo_i^{iio} \in Geo^{loc} \\ geo^f = 0: geo_i^{iio} \in \emptyset \vee geo_i^{iio} \notin Geo^{loc} \end{cases} \wedge \\ \begin{cases} comp^f = 1: comp_i^{iio} = comp^{loc} \\ comp^f = 0: comp_i^{iio} = \emptyset \vee comp_i^{iio} \neq comp^{loc} \end{cases} \wedge \\ \begin{cases} sec^f = 1: sec_i^{iio} \in Sec^{loc} \\ sec^f = 0: sec_i^{iio} \in \emptyset \vee sec_i^{iio} \notin Sec^{loc} \end{cases} \wedge \\ \left|\left\{geo_i^{iio}, comp_i^{iio}, sec_i^{iio}\right\}\right| > geo^f + comp^f + sec^f \end{array} \right\}$$

$$R_{F3} = R_{F2} \backslash D_3$$

In other words, the initial set of the maximum relevant IIOs R is reduced in two steps in order to retrieve the set R_{F3} that is then included into the risk calculation for a supplier location. Process step 2 (set R_{F2}) aims to remove all IIOs that are relevant but apply to predefined filter criteria and do not conflict with a defined dimension. These are those messages that deal with the supplier location directly or in a more generalized form. For example, if the supplier location is in London (United Kingdom) then texts about the United Kingdom are still relevant. However, the user might decide that news only referencing the country (and not the city London) should be filtered from the set of relevant IIOs.

The second filter has a comparable processing logic to step 2; however, step 3 (set R_{F3}) focuses on IIOs that conflict with the location for at least one dimension while still overlapping. For example, the IIO might refer to a different sector while still naming the same city. This second set of filters reduces the amount of overlap (or spillover) configured in the system. While a news report talking about another sector in the same city could still be relevant for a supplier location in the same city, this might be different on the regional level. A report that deals with a different sector in the same region might be irrelevant, while a report that only refers to the region could be seen as relevant. The latter does not conflict with the former as the sector of the supplier location might be involved in the latter. Figure 18 represents this logic visually.

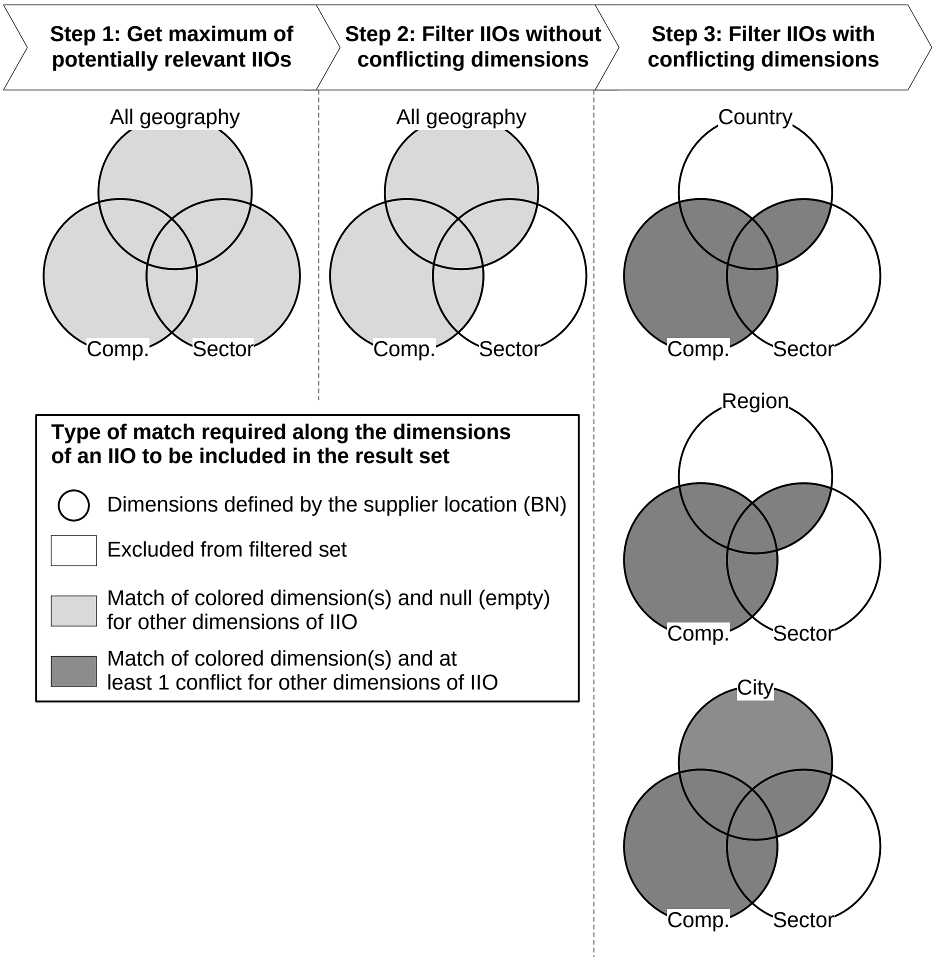


Figure 18: Logic of IIO-reduction followed by the system in order to retrieve relevant IIOs (author's representation)

Combined, these process steps result in a set of IIOs that may then be included into the BN for the supplier location.

4.3.3.3 Credibility and relevance

Not every news report will be of the same quality and, hence, not every IIO should have the same effect on the likelihood of a breach. Particularly when using data based on social feeds, confidence about the data needs to be interpreted. Consequently, we suggest considering two variables, credibility and relevance, to represent the content quality of an IIO. These criteria have been used in other studies as attributes or evaluation criteria for reports (e.g. Beach et al. 1978; Keller 2009; Sarkki et al. 2013).[57] The credibility c is defined as comprising "[…] the content of evidence captured by a sensor which includes veracity, objectivity, observational sensitivity, and self-confidence […]" (Blasch et al. 2013, 1601), while relevance r "[…] assesses how a given uncertainty representation is able to capture whether a given input is related to the problem that was the source of the data request […]" (Blasch et al. 2013, 1600). In other words, the model understands relevance as capturing how closely the messages used as input for a certain supplier location are in fact related to the supplier location. In order to derive a relevance measure, the availability of dimensional attributes in the news articles is used as an approximate indicator. The more an IIO can be linked to a certain IIO in a granular and specific way, the more relevant it is. Thus, "[A] piece of evidence (POE) is relevant if it impacts previous beliefs." (Blasch et al. 2013, 1603; emphasis removed).[58] If IIOs with partly conflicting dimensional information are included, the relevance can only be derived based on the non-conflicting dimensional information. Credibility is suggested as being defined either at an input channels or source level in order to cover different media types as completely as possible. Attaching a specific credibility tag to every possible text source would require significant time and a continuous effort to keep the list up-to-date. Both the credibility and relevance attributes are added to an IIO at the time of inclusion into a BN.

4.3.3.4 Updating

A BN is initialized at a particular point in time t which can serve as a basic reference point. Until this time, zero or more IIOs may have been persisted and a set of IIOs can be retrieved as discussed above. In general, when revising the probability based on evidence from textual media sources, two options may be considered. Either only the latest IIO is entered as a single

[57] Another important attribute often considered is reliability, which is attributed to the consistency of the inputs of a sensor, e.g., over time (Blasch et al. 2013, 1601). It has not been included given that only a rare number of events is expected to be handled. If necessary, it could be included in a later version of the system. Similar holds true for completeness "[…] which reflects whether the totality of evidence is sufficient to address the question of interest […]" (Blasch et al. 2013, 1600) as IIOs are incrementally added and completeness is unknown and cannot be assumed.

[58] Another way to define relevance is to rank news outlets based on the attention they receive (University of Illinois 2012). However, because a website is frequently visited it does not necessarily mean that it is credible. Moreover, taking the frequency as a measure of relevance would allow copy-cat reports to influence the overall relevance.

finding or the network is continuously updated with the evidence from new IIOs. As IIOs are assumed to be independent, each IIO is likely to include valuable information. Consequently, the BN will be modeled using the latter option, allowing the inclusion of evidence from multiple reports. Relevance and credibility can be evaluated for each IIO as described above.

As mentioned, each IIO is interpreted as an additional, independent report about a potential child labor incident at a supplier location. Consequently, even IIOs with low credibility or relevance are understood to increase the overall observational probability. For example, if a single IIO with high credibility is entered into the model, a second IIO should still increase the expected value of the observational probability, even if its credibility and relevance is very low. Given these assumptions, the expected value of the observational likelihood needs to be a monotonically increasing function of the number of IIOs included, as detailed in "Appendix H: Proof: Monotonically increasing expected value of observational likelihood":

$$\forall f \geq 0: E(P_{obs}(f_1, c_1, r_1)) \leq E(P_{obs}((f_2, c_2, r_2)); f_2 \geq f_1; 0 \leq p, c_{1,2}, r_{1,2} \leq 1; f_{1,2} \in \mathbb{N}$$

Both the credibility and the relevance will be continuously updated with new evidence from IIOs. The evidence will then be entered into the BN. This leads to a model containing the frequency f, the credibility c, and the relevance r as variables (see Figure 19). These are combined via the equation $x = f * (c + r)$.

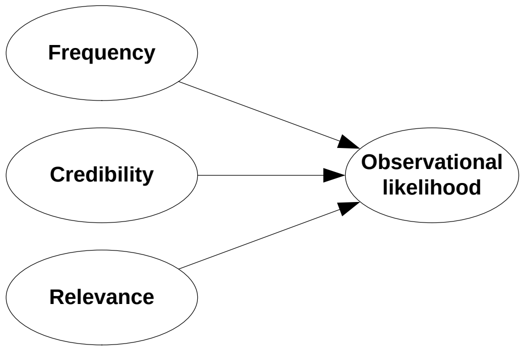


Figure 19: BN structure for observational likelihood (author's representation)

As already stated, the monotonically increasing function x has no defined upper bound. Therefore, a scaling function is needed to return a value between 0 and 1 for the mean of the normal distribution of P_{obs}. For this purpose, a monotonically increasing function with the asymptotic limes at 1 is suggested. This can be achieved with an inverted, shifted hyperbola. The frequency score function $s(x)$ shown below is suggested for this (the function can be parameterized through the parameter τ, which is initially set to 5). Figure 20 shows a plot of the scaling function with its starting value at 0:

$$s(x) = 1 - \left(\frac{1}{1 + x/\tau}\right)$$

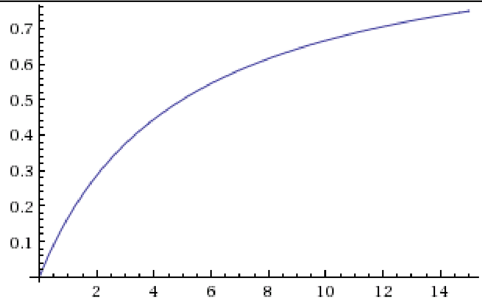


Figure 20: Plot of the scaling function $s(x)$ with $\tau = 5$ (author's representation; plotted with Wolfram Alpha[59])

4.3.3.5 Standard deviation of observational likelihood

The node of the observational likelihood P_{obs} is represented with a normal distribution given the relevance r, the credibility c, and the frequency f. Hence, knowing about f IIOs for a specific supplier location, the calculated likelihood of a child labor incident should be within a predefined confidence interval. This confidence interval should be smaller if a high number of IIRs with high reliability and credibility are received.

For a known standard distribution, 95% of its probability mass lies within the mean μ plus/minus 1.95994 times the standard deviation. If, as defined by the user, the area covered by the 95% interval is p' percent points if no IIO has been received and p'' percent points if ten fully credible and relevant IIOs have been received, then the respective standard deviations in percent points can be calculated with $\sigma(p) = \frac{p/2}{1.959964}$. Setting p' to 40 and p'' to 10 percent points yields $\sigma' = 10.204$ and $\sigma'' = 2.551$. σ is seen dependent on the values of f, r, c and a linear functional connection is assumed.

$$\sigma(f, r, c) = \alpha - \beta * f * r * c$$

Using σ' and σ'', the values for α and β can be determined. This leads to the following function, which will be used in the model and can be adapted to individual needs. A floor value for the deviation can also be defined.

$$\sigma(f, r, c) = \alpha - \beta * f * r * c = 10.204 - 0.7653 * f * r * c$$

4.3.4 *Breach likelihood*

The final node of the BN models the likelihood of a breach of child labor compliance standards, P_{breach}. It combines the contextual prior with the audit and observational likelihoods, thus revising its prior. Hence, it captures the likelihood of a breach and not the likelihood that

[59] http://www.wolframalpha.com/.

the breach is publicly recognized. Focusing on the direct likelihood of a breach stresses the link to supplier "codes of conduct" and terms of trade, where a breach is evident even if no media echo results. Consequently, a company would erode its own standards if potential breaches are treated differently depending on their "media potential". This might eventually cause further suppliers to ignore "codes of conduct" if they felt they were not in the spotlight.

Overall, the proposed model implements a positive connection between the child labor probability and the supplier signing a "code of conduct".[60] The extent can be calibrated by defining the factor φ by which suppliers with signed "codes of conduct" outperform other companies in the same geographic area and sector.[61] This allows the contextual prior to be adjusted for companies that have signed a "code of conduct". In summary, the breach probability is determined by the prior likelihood P_{prior}, the audit likelihood P_{audit}, the observational likelihood P_{obs}, and the existence of a "code of conduct". This results in the overall model structure shown in Figure 21. The mean of the resulting final node can be used to establish the prioritization of the supplier locations.

[60] "Codes of conduct" have a strong influence on supplier operations. Suppliers typically need to make major changes to comply with the compliance levels defined (e.g. Egels-Zandén 2007, 46). However, there is often a lag in implementation, especially in developing countries. Even suppliers known to be strongly compliant often fail on several points of the "code of conduct" (Egels-Zandén 2007, 46, 52). Of particularly interest is the methodology used by Egels-Zandén, who bases his conclusions on unannounced and unofficial interviews, thus reducing potential biases and influences on the interviews' outcomes. He focused on pre-known positive examples to allow for a generalized conclusion if even "the good ones fail". Therefore, "codes of conduct" can have an initial effect on suppliers, but this effect should not be exaggerated. This is distorted if time is also taken into account. Egels-Zandén (2014) shows a positive development over time for suppliers who meet buyers' requirements that can at least be partly attributed to the effect of code of conducts. Consequently, the effect of "codes of conduct" improves over time. Other researchers also argue for complementary measures to "codes of conduct" that also change the management system in order to be fully effective (Locke et al. 2007, 35).

[61] The influence of the "code of conduct" might be higher given a higher prior probability.

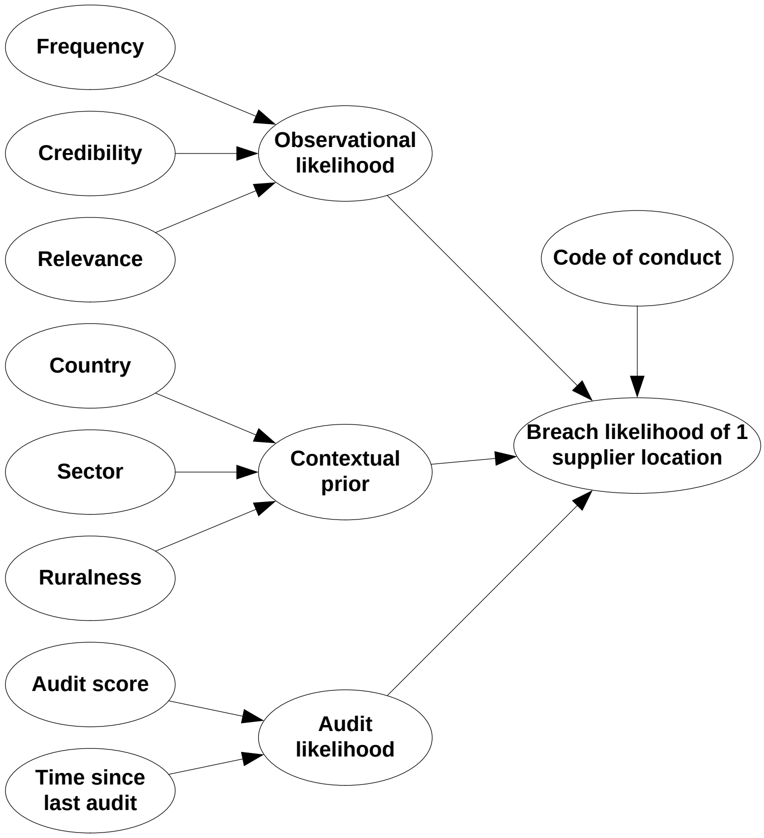


Figure 21: BN structure for overall risk model and breach probability (author's representation)

4.4 Summary

This chapter presented the method for estimating the likelihood of a child labor compliance breach at a supplier location. Different drivers and disablers of child labor were discussed, leading to a focus on child labor incidents. In the main section, a risk model based on a BN was suggested that allows observations to be integrated via continuous updates. A location-specific prior is calculated given the country, sector, and ruralness of a supplier location.

5 Text mining methodology

Text mining is used to derive independent incident observations (IIOs), i.e., independent event descriptions of child labor incidents that refer to a single incident (but may be related to multiple suppliers) which are used as an input for the risk model. The subsequent sections of this chapter will outline the methodological approach used for extracting IIOs from news.

5.1 Suggested text mining architecture

The TM system needs to handle a continuous stream of separate text documents. These can be provided by either a pull-based retrieval or a push-based receiver system. In the context of this thesis, and assuming a perfectly working text mining approach, the more news that can be analyzed, the better overall system performance will be. Consequently, the text mining framework needs to be able to handle a high volume of text documents for processing. Given continuously increasing hardware capabilities (Intel 2015; G. E. Moore 1965) and this thesis' focus on a prototypical assessment of the possibility of incident extraction, processing performance questions will be left for future work.

Based on the related work and the goal of using text mining to process news texts and derive IIOs, several tasks may be identified that need to be fulfilled by the system (Piskorski and Yangarber 2013):

1. *Event detection* to identify different events
2. *Event classification* to focus only on child labor incident events
3. *New event detection* to differentiate a new event from existing events
4. *Event extraction* (incl. cleaning of non-unique elements) to derive event templates

Two approaches seem particularly appealing:

(1) To detect new events and then classify them in order to identify child labor incidents. A variation of this approach would be to detect the new events only after having classified child labor incidents.

(2) To identify texts that potentially contain child labor incidents and then perform event detection, including identifying new events. In this second approach parts of event detection

may already be included in the classification task if the classification is understood (and defined) as resulting in a list of texts that all contain a child labor incident. In this case only extraction and new event detection need to be subsequently performed.

The second approach has been chosen for the specific case discussed here, as a large number of news texts may contain a vast number of events, and detecting only lightly referenced child labor incidents may be sufficiently difficult. Moreover, given that classification classes are known and not dynamically learned, this step can be performed more easily to reduce the set of texts requiring further processing.

Before the domain-specific processing outlined above, a preparation step is performed for each news report that includes steps such as tokenization, morphological and lexical processing, and syntactic analysis as outlined by Appel (1999) and detailed below. Thus, the pipeline corresponds to one of the classical IE systems and can be split into a domain-independent and domain-dependent part (Piskorski and Yangarber 2013). If risk model updating is also included, this results in the four-step process depicted in Figure 22.

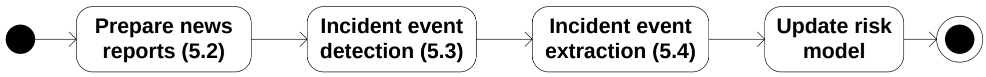


Figure 22: Processing steps for text mining architecture (author's representation)

The following sections will deal with these steps. We will start with a discussion of child labor incident characteristics and the event frame which is the end product needed and used for the risk model update.

5.2 Prepare news reports

As outlined above, several preprocessing steps are taken before the domain-specific handling of news reports begins. It is assumed that the texts received do not include meta-data fields added by content aggregators or providers which already provide preprocessing. Only the date and time of publication is included as a meta-field for each news report. This process is shown in Figure 23.

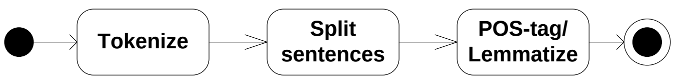


Figure 23: Processing steps for preparing news reports (author's representation)

First, a tokenizer splits the words and other characters before a sentence splitter detects sentence boundaries in a second step. Both are typical processing steps in text mining applications and allow words and sentences to be differentiated (A. Khan et al. 2010).[62] Third, a POS-tagger is used to differentiate word lemmas. A POS-tagger annotates each token of a text with its part of speech (e.g. verb, noun, or adjective; Manning and Schütze 1999, 341). In contrast, a word lemma refers to a word stem after removing affixes (Manning and Schütze 1999, 132). Word lemmas may be more easily compared with each other given that string matches (i.e. a direct comparison of textual characters) on surface forms (i.e. a word including declinations) may fail. After preparing the news report, a domain-specific analysis can be applied.

5.3 Incident event detection

Child labor incidents were defined in the introduction and child labor incident events and observations were defined in the risk model methodology. The task of the incident event detection step is to filter those news reports that contain a child labor incident event. A news text does not necessarily contain a new event. Often a large number of texts will be clustered in order to detect new events (e.g. Sankaranarayanan et al. 2009; Tanev, Piskorski, and Atkinson 2008). However, given the low frequency of child labor events, clustering appears less relevant here. Moreover, the incident class is known and, as defined, one news report contains at maximum one incident event. Hence, the task can also be interpreted as a classification task on the document level.[63]

Two ideas need to be differentiated for the classification task. As discussed, either a knowledge-driven or data-driven procedure can be applied (F. Hogenboom et al. 2011). The former can apply textual patterns to texts in order to identify events (Tanev, Piskorski, and Atkinson 2008). Authors have shown that an ML approach can be helpful in settings where the definition of an event is wide (and therefore possibly referenced in many ways) (M. Naughton, Stokes, and Carthy 2010). Otherwise, rule-based extraction systems appear more successful, partly because configuring rule-based systems may also be cheaper (Appelt 1999). Nevertheless, several reasons support the use of an ML-based approach here. In general, the more statistical an approach is, the less affected it is by noise data and the more it can be transferred to different data sources (Denecke 2012, 37). Furthermore, in the domain consid-

[62] Tokens resemble textual units "[…] where each is either a *word* or something else like a number or a punctuation mark." (Manning and Schütze 1999, 124).

[63] Given that the differentiation will be into two classes ("incident" or "no incident"), the task could also be called "filtering" (Manning and Schütze 1999, 530).

ered here, company experts might not be able to define extraction rules but would be able to specify whether articles contain child labor incidents or not for the training data.

Nevertheless, additional domain knowledge will be used to further combine a data-driven and a knowledge-driven approach. Domain knowledge is applied to improve the ML approach applied for classification (F. Hogenboom et al. 2011). The following sections outline this ML approach. The first section suggests a specific “candidate set reduction” technique. The second section deals with the important extraction of features for ML (A. Khan et al. 2010). Finally, the last section outlines the ML methodology.

5.3.1 Candidate set reduction

Two basic approaches appear applicable in order to developing an overarching classification process: (1) Develop an approach that takes an article as an input and directly classifies the article. (2) Split the task into steps, first reducing the candidate set that needs to be classified and then performing the classification on the reduced set. Here, candidate set reduction is understood as a task of information pre-filtering.

Given the low frequency of child labor incident articles, introducing a candidate set reduction process seems promising to be able to more specifically tailor the classification task. Therefore, having a viable automatic approach that reduces the candidate set further than for example just searching for occurrences of child without producing a large number of false negatives (i.e., articles that would be excluded although they should belong to the candidate set) appears to have several benefits.

- It would allow the classification task to be more specifically tailored to a candidate set that excludes a very large share of non child labor texts. Classes should be well sampled to be sufficient for machine learning (Pimentel et al. 2014).
- It would enhance the speed of processing articles by reducing the calculation of the classification task to a reduced candidate set (Denecke 2012, 33; Fernando and Stevenson 2012; Sankaranarayanan et al. 2009).
- It would improve the precision of information extraction (Thorleuchter and Van den Poel 2013).

The first argument is often stressed. In the comparable area of event tracking, Lei et al. (2005) suggest pruning negative events to increase the ratio between positive and negative events. To cope with an imbalanced dataset with a low number of true positives, Van Landeghem et al.

(2011) employ a filter that reduces the size of the candidate set, given the distance of relevant attributes in a parsed dependency sub-graph and the length of a sentence. The latter is possible as events in their domain are typically found within one sentence (Van Landeghem et al. 2011). Nevertheless, a version of the former may be applicable for the current domain.

Here we propose to make use of the domain's specialty that "child labor" is composed of two words which depict the way child labor is presented in text. At this point it is assumed that texts on child labor will have a combination of a word representing "child" and a modifier stressing "work" in the text. This assumption will be analyzed in the evaluation. Other sustainability-related topics such as "forced labor", "government corruption" or "environmental pollution" display a similar combination of a noun and a modifier.

As sentences close together in a text tend to be on the same topic (Zha 2002; Martina Naughton, Kushmerick, and Carthy 2006) and a direct mention of "child labor" may be seen as the most obvious trigger of a child labor incident event, a distance based approach using a cut-off distance is suggested to prune negative cases. This can be compared to the technique used by Van Landeghem et al. (2011) presented above. The distance is measured as the number of characters between "child" and a word indicating "labor." "Child labor" would have a distance of one (the space character only). Indicative words may be synonyms, hypernyms, or other related word sets. If multiple relations are found in a document, the minimum distance within a news report can be used as the one representing the text. The resulting relations can be used as input for further processing steps, as will be analyzed later. Figure 24 depicts this two-step process.

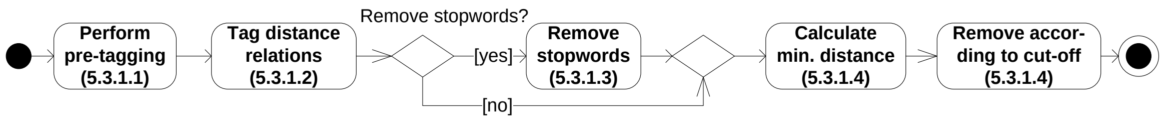


Figure 24: Processing steps for candidate set reduction (author's representation)

This candidate set reduction approach relates to IR as the documents are ranked and then a subset is selected based on a predefined threshold. Like other information systems (see e.g. Piskorski and Yangarber 2013) the IR approach is used to bootstrap the subsequent IE system behind, with the eventual aim of extracting child labor incident events. Combined performance is expected to be better than performing a single classification task without upfront IR.

5.3.1.1 Domain-specific pre-tagging

The basic assumption for the filtering is that relevant news reports on child labor contain a word with “Kind” (German) or “child” without case sensitivity and for any token position. This assumption can be partly proven by the data available during the development of the gold standard. In the 32 (relevant) child labor related articles identified in the APA dataset (see chapter on evaluation), 30 proved this assumption.[64]

To determine the character distance between a word indicating “child” and a word indicating “labor”, gazetteers[65] for each word are defined. Given that the same meaning may be represented by different words, lists of words can be compiled to perform the tagging. Different semantic relationships between words can exist. The next two sections will outline the different approaches taken for English and German. German is used to partially test transferability to other languages.

5.3.1.1.1 English specifics

WordNet is an English semantic lexicon used in a large number of scientific approaches (Fellbaum 1998; Miller 1995; Princeton University 2010). Figure 25 shows different important relationships used in WordNet that can be used to derive related words in English. Definitions are given in the legend.

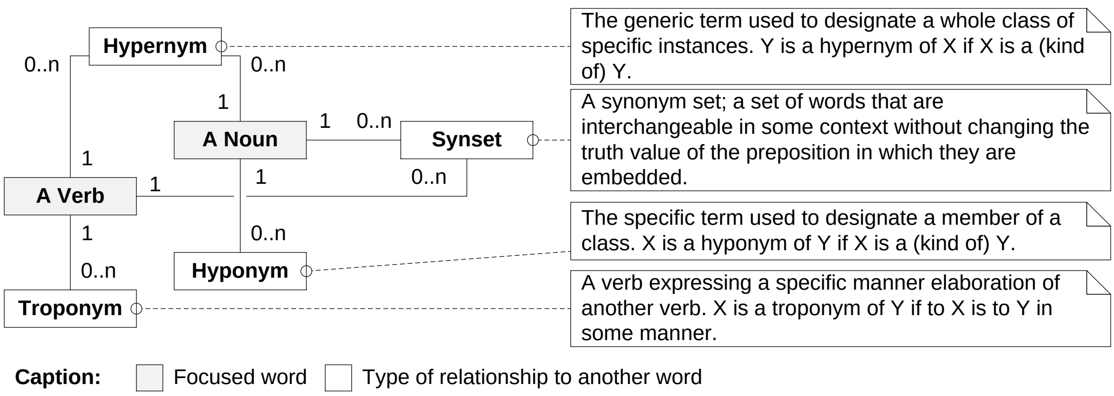


Figure 25: Semantic relationships between words in WordNet (WordNet 2014; author’s representation)

It is important to decide whether to only use relevant synsets covering important[66] synonyms or whether further relationships should also be included. Including more related words can

[64] The single document not fulfilling the assumption was actually close to being excluded from the set of 41 articles, and deals with children that fight bulls. Moreover, the string “Kind” appeared in the article’s keywords.
[65] Gazetteers tag words with an annotation based on syntactic (and in an extended version also semantic) matches.
[66] “Important” in this context means excluding all synonyms that have a different semantic, that are marked as colloquially linked or are too formal, that do not fit into the context, are not used anymore (deprecated), or belong to a special subset of the language.

increase recall at the expense of precision. Given that the distance should be calculated, weighting approaches including 2nd and 3rd order hyponyms or troponyms have not been included (Carpineto and Romano 2012, 14).

Consequently, we suggest building a normal list of related words containing only relevant synsets for the nouns and verbs of "labor". As "child labor" is often also called "child work," this has also been done for "work". Under the assumption that each article contains "child" as a word, "child" is used as the direct anchor for the other side of the distance relationship. The effectiveness of this assumption can be tested, given that each document also embodying a word for "labor" or "work" will allow a distance measure to be derived. The matching needs to consider that words may start or end with different pre- or postfixes. Consequently, word part matches and word lemmas are suggested.

5.3.1.1.2 German specifics

Instead of using WordNet, the German approach is based on the well-known Duden dictionary (Bibliographisches Institut GmbH 2013). After excluding all unimportant (as in English) synonyms, the remaining words are traced until no new seed synonyms can be derived. Additionally, adjectives are formed for nouns where possible and related synonyms are included analogously. Similarly, the synonyms for "work" ("Arbeit") are derived. Here the semantic understanding of "Arbeit" is "work" and not "workplace" (this may be a further extension in the future). This is done for nouns as well as related verbs (e.g. "arbeiten").

The synonyms are used in what will be referred to as the "standard approach". The "extended approach" also includes verb synonyms related to compulsion ("Zwang/zwingen"), as child labor can also appear in the context of forced work, and those related to "earning" ("Verdienst/verdienen"), as working is often related to earning something. Synonyms can also be derived using an automated dictionary that technically links synonyms[67], but this has been left for future work. Summing up, the two different approaches include the synonyms as defined in Table 25. Additional seeds are derived from the main seeds. All synonyms can be found in "Appendix I: Synonym sets used for document classification".

[67] Germanet, a German version of WordNet, is one such tool (Hamp and Feldweg 1997).

Approach	Main seed	Additional seeds
Standard	Kind	Adjectives: klein, jung, halbwüchsig, heranwachsend, jugendlich Nouns: Minderjährige(r), Unmündige(r)
	Arbeit	Verbs: arbeiten, beschäftigen, betätigen, hantieren, tätig sein, ausüben, handeln, verkaufen, bedienen, handhaben, anfertigen, erzeugen, herstellen, produzieren, fertigen, bauen, fertigstellen
Extended	Zwangsarbeit, Ausbeutung	Verbs: ausbeuten, beknechten, versklaven Nouns: Missbrauch
	Verdienst	Verbs: verdienen

Table 25: Seeds for synonyms as used by the processing pipeline for the two approaches (author's representation)

5.3.1.2 Relation tagging

In this context, relation tagging aims to find all relations between subsequent mentions of terms for "child" and "labor" and calculates the number of characters between them. The order of the tags does not matter. As the tagging focuses on subsequent terms, no overlaps between relations can occur. It is important to note that this tagging can cross sentence boundaries. In the subsequent chapters we name such a relation a "child labor distance relation".

5.3.1.3 Child labor specific stopword removal

Typically, stopwords are words that are so frequent and unspecific in a language that they do not provide any differentiating value. Here, we distinguish child labor specific stopwords in a slightly different manner. They are defined as words which contain a word that relates to "child" and one that relates to "labor" within the same token. Eliminating these is particularly helpful in German, which employs many compound nouns; however, English only has a limited number of such words. In the context of this thesis, only "childcare" was found for English, but more for German. The specific stopwords are retrieved from the corpus by extracting all tokens which fulfill the criteria separate for positive and negative child labor incident event texts. Then, the stopwords are manually inspected to remove potentially ambiguous stopwords. Finally, the combined list of stopwords is used to remove all child labor distance relations that contain such a word.[68] [69]

5.3.1.4 Cut-off based filtering

Once the distances of all child labor distance relations are known, a text document is automatically tagged with the minimum distance of all relations. Several distance relations may share the same minimum distance. The area of text covered by minimum distance relations is perceived as the area of the text most directly related to a child labor incident. This understanding will be used for further TM processing.

[68] Instead of a manual inspection, this approach can also be incorporated with a simple rule-based processing step.
[69] The list of stopwords found and used in this thesis is shown in "Appendix N: Stopword lists".

Once all documents in a corpus are tagged with the distance of the minimum child labor relation, they can be ordered in ascending order. This allows a cut-off distance to be defined which reflects a trade-off between precision and recall (both metrics will be discussed in more detail later). As the candidate set reduction process is as a filtering step before classification takes place, it should achieve a reasonable recall-rate while providing a significant reduction in the data skewness for ML input. The cut-off may show significant differences depending on whether it is applied to the approach containing the standard or the extended child and labor tagging sets.

5.3.2 Feature selection

As Kahn et al. (2010) state, deciding on the appropriate data representation is as important as deciding on the right classifier. Therefore, feature selection aims to represent the text features and select the right ones for the ML task, which then tries to classify the news into categories.[70] Given that potentially only a small part of each article refers to a child labor incident, approaches to feature selection are limited. ML classification approaches need a reasonably sized feature vector to perform properly. Consequently, different strategies are tested that aim to restrict the features to the part of the text most relevant to the domain. Moreover, as only a small part of a document may refer to a child labor incident, a content focused approach to feature selection is proposed, one which is not normalized across all documents but document specific. Hence, we neither tune the overall number of features nor restrict the analysis to a number of document-dependent top-scoring features.[71]

Moreover, while some authors argue that semantic features can improve classification results in later steps (Conway et al. 2009; Y. Zhang and Liu 2007); however, these have not been used here and leave room for additional improvement. Nevertheless, including domain-specific features has the chance to significantly increase precision and recall (Zhao and Wang 2012). Furthermore, including more features may result in a better performance of the ML algorithm (Y. Li, Bontcheva, and Cunningham 2005a). Therefore, we suggest testing different numbers of features and leveraging pre-tagging to identify child labor distance and minimum

[70] Multiple feature selection techniques have been published and several overviews have been presented (Forman 2003; Ikonomakis, Kotsiantis, and Tampakas 2005; A. Khan et al. 2010).

[71] Most other approaches that reduce a text to a smaller number of words have only focused on an article's title (Wunderwald 2011). However, Thorleuchter and Van den Poel (2013) and Ikonomakis (2005) suggest using latent semantic indexing (LSI) to reduce the number of words representing a text. Kolcz, Prabakarmurthi, and Kalita (2001) introduce an approach that utilizes features selected by text summarization for classification. However, these approaches assume that child labor is either referenced in the title or across the whole article, contradicting the characteristics listed above. There are also problems in that child labor incidents reported by only one document are already relevant.

child labor distance relations, which should indicate the potentially relevant area of a news report.

Two general selection strategies are proposed. Both focus on the most relevant area of a text (Manning, Raghavan, and Schütze 2008, 339). On the one side many child labor distance relations span a text between a word indicating "child" and a word indicating "labor". These words and the two neighboring ones can be used as a feature input for the algorithm. On the other side, these distance relations often span several sentences within a news report, so, as an alternative approach, considering the other tokens from these sentences could provide additional context. Finally, two additional sub-strategies may be employed when these feature selection tasks are restricted to only the minimum distance relations, leaving all others unused. Some authors argue for including more context in the detection of events, going beyond the sentence level (Liao and Grishman 2010; Patwardhan and Riloff 2009). Figure 26 presents the different cases for an example paragraph. The most restrictive case only takes the minimum distance relation as the input for ML, focusing on a token sequence of "child labor." For all approaches it is suggested to only use word lemmas as input. Additionally, the POS-tag is available and can be used. A further restriction differentiates two additional modes – either taking all minimum distance relations (i.e., if the minimum distance applies to multiple relations) or only using the randomly first one found.[72]

In porttitor. Mauris eget neque at sem venenatis eleifend ut **labor** nonummy. Fusce aliquet pede non pede. Suspendisse **child** dapibus lorem pellentesque **work** magna. Integer nulla. Donec blandit feugiat ligula. Donec hendrerit, felis et imperdiet euismod, purus ipsum pretium metus.	In porttitor. Mauris eget neque at sem venenatis eleifend ut **labor** nonummy. Fusce aliquet pede non pede. Suspendisse **child** dapibus lorem pellentesque **work** magna. Integer nulla. Donec blandit feugiat ligula. Donec hendrerit, felis et imperdiet euismod, purus ipsum pretium metus.
Child labor distance relations	**Minimum child labor distance relations**
In porttitor. Mauris eget neque at sem venenatis eleifend ut **labor** nonummy. Fusce aliquet pede non pede. Suspendisse **child** dapibus lorem pellentesque **work** magna. Integer nulla. Donec blandit feugiat ligula. Donec hendrerit, felis et imperdiet euismod, purus ipsum pretium metus.	In porttitor. Mauris eget neque at sem venenatis eleifend ut **labor** nonummy. Fusce aliquet pede non pede. Suspendisse **child** dapibus lorem pellentesque **work** magna. Integer nulla. Donec blandit feugiat ligula. Donec hendrerit, felis et imperdiet euismod, purus ipsum pretium metus.
Child labor distance relation + sentences	**Minimum child labor distance relations + sentences**

Figure 26: Suggested contexts for tokens selected organized by approach (author's representation)

Another potentially helpful feature can be derived through sentiment analysis (B. Liu and Zhang 2012). Sentiment analysis (SA) aims to extract the specific tone used in an article and can be expressed in scales or categories such as positive and negative. SA can help decision makers understand the underlying opinion of texts without reading all of them (Cao et al. 2012). The tokens in a feature set can express a tone that can be used as an additional input

[72] In most cases only one minimum distance relation will be present.

for the ML classification task. Resources for SA are language specific and several have been published.[73]

Finally, given that the relative frequency of words in the documents and the document length are unclear, the term frequency multiplied by the inverse document frequency (going back to Jones 1972; Luhn 1957) is used to weight the lemmas of the documents as input for ML (Forman 2003). This widely used approach is called $tf-idf$, where tf stands for term frequency and idf for inverse document frequency.

5.3.3 Machine learning approach

5.3.3.1 Introduction

ML is used to classify a set of documents into those containing a child labor incident and those that do not. This is a binary classification task where a document is either in the predefined class or not. One advantage of such a binary classifier is the ability to add a new classifier for a different event in parallel if different incident types should be classified in the system (Van Landeghem et al. 2011). Applying ML requires a training corpus. If no corpus exists for the topic at hand, manual tagging must be performed (for this thesis this will be explained in chapter 6). Furthermore, ML approaches strongly rely on the inputted features. The greater the complexity, the more training data is needed (Piskorski and Yangarber 2013). In this particular case, complexity is limited by the narrow definition of a child labor incident. Consequently, this relaxes the requirements for the training data and "classical" algorithms appear useful.

Aggarwal (2014) distinguishes two types of text classification problems when text streams are involved: Either the application or the training data is the one that is streamed. For this thesis it is the former which is relevant, and all classification approaches typically show reasonable performance in this setting (Aggarwal 2014). Consequently, the propositions made for this purpose may be applicable to direct stream usage. Nevertheless, the literature discusses several ML techniques for text classification, the results of which provide valuable input for the current task. Das (2010, 5) reviews multiple frequently used algorithms such as the Bayes classifier, SVM (Cortes and Vapnik 1995), or word count classifiers which can all work with a broad range of inputs. All of these methods seem to be relevant here. On a sentence-level,

[73] For English we refer to SentiWordNet (Baccianella, Esuli, and Sebastiani 2010; Esuli and Sebastiani 2006) and for German to GermanPolarityClues (Waltinger 2010a; Waltinger 2010b). Both use sentiment categories that can be directly used as input features if added to tokens during parsing.

SVM has seen particular success with event classification (M. Naughton, Stokes, and Carthy 2010). Figure 27 provides an example of how a token vector can be derived based on parts of a text and can serve as basic input for an SVM.

1 document with childLabor = true/false

In porttitor. Mauris eget neque at lorem venenatis eleifend ut **labor** nonummy. Fusce nulla lorem non neque. Suspendisse **child** dapibus lorem pellentesque **work** magna. Integer nulla. Donec blandit feugiat ligula. Donec hendrerit, felis et imperdiet euismod, purus ipsum pretium metus.

Child labor distance relation + sentences

Word	Count
lorem	3
neque	2
mauris	1
eget	1
at	1
venenatis	1
eleifend	1
ut	1
labor	1
nonummy	1
fusce	1
nulla	1
non	1
suspendisse	1
child	1
dapibus	1
pellentesque	1
work	1
magna	1

Token vector

Note: Token vector assuming equivalence of word strings and lemmas and without “classical” stopword removal; td-idf weighting not shown in this example

Figure 27: Example input for SVM when selecting sentences based on child labor distance relations (author’s representation)

In general, SVM is repeatedly highlighted as one of the most effective techniques to classify texts (A. Khan et al. 2010; Mladenić et al. 2004; Yang and Liu 1999). In addition, the *k*-nearest neighbor (KNN) and Naïve Bayes (NB) techniques are suggested to be potentially interesting for classification depending on the feature input, particularly for binary classification tasks (Colas and Brazdil 2006; A. Khan et al. 2010). Torii et al. (2011) suggest that SVM may be more robust than NB when it comes to classification tasks. If the ratio of true positives to true negatives is small, an SVM with uneven margins has been proposed (Y. Li, Bontcheva, and Cunningham 2005a). Nevertheless, because an uneven margin only helps if the genuine frequency of positive elements is higher than in the training dataset, an improvement through uneven margins should not necessarily be expected if both show similar frequencies (Y. Li, Bontcheva, and Cunningham 2005a; Y. Li and Shawe-Taylor 2003). Finally, Li et al. (2005a) propose using perceptrons[74] with uneven margins (PAUM). These can also be configured to work with even margins.

In summary, SVM (with and without uneven margins), KNN, NB, and PAUM seem useful for the classification task at hand and have successfully been tested in earlier works. C4.5 (Quinlan 1993), a decision tree-based algorithm, has also been added due to its availability in the technology stack used in this thesis. Therefore, we plan to test the different classifiers on two training and test sets of tagged news articles based on the standard and extended candi-

[74] Perceptrons are multi-layered classification algorithms (e.g. Y. Li, Bontcheva, and Cunningham 2005b).

date set reduction (with a specific cut-off value defined during evaluation). Different input parameters for the ML techniques will also be tested.

5.3.3.2 Parameters

The parameters used for ML depend on the algorithm. Here the parameter configurations based on the ML implementation detailed in “Appendix prototype implementation” are shown. While SVM and PAUM provide a larger variety of possible parameters, the other chosen ML algorithms rely on fewer. Parameters are partly dependent on the implementation used, as specified in the prototype section. In the following the different parameters will be discussed for each ML algorithm. Most permutations of the parameters have been tested – if filtering partly did not yield results, it has been omitted in some cases to reduce the computational costs. The parameters should cover a wide range of the possible (or reasonable) parameter ranges. Moreover, additional sensitivity tests will be applied later.

Parameter	Values tested
Filtering	None; near (0.2)
Classification threshold	0.1; 0.5

Table 26: Overarching parameters[75] (author’s representation)

Two overarching parameters were configured and tested, independent of the underlying ML algorithm. Filtering removes negative examples from the training set in order to improve the skewness of the training data (Y. Li and Bontcheva 2008). Meanwhile, the classification threshold specifies the level of confidence needed for classification (Cunningham, Bončeva, and Maynard 2011).

Parameter	Values tested
Kernel type	Linear; polynomial
Degree (for polynomial)	If polynomial, then cubic
Cost	1; 10
Tau	0.3; 1

Table 27: Parameters for SVM (own representation)

The primary differentiation for SVM is the kernel used. Either a linear or a polynomial kernel is applied. For a polynomial kernel, a cubic function is applied. The influence of the cost parameter was tested on the best performing approach (based on a cost of one) and the resulting lightest variation of the parameter (10) was applied to all other tests. The uneven margin parameter tau (τ) was set at 1 and 0.3, comparable values to those suggested by Li, Bontcheva, and Cunningham in their initial presentation paper (Y. Li, Bontcheva, and Cunningham 2005a, X). A τ smaller than 1 indicates the use of an uneven margin.

[75] Additional parameters have been set for optimal classification performance: Surround=false

Parameter	Values tested
Positive margin	5; 50
Negative margin	1; 5; 50
Bias term	0; 0.3; 1

Table 28: Parameters for PAUM (author's representation)

For PAUM three parameters need to be chosen, and the positive and negative margins need to be considered together. Consequently, all possible combinations were evaluated to create different types of even and uneven margins. A variation of filtering was once again partly omitted, given that it had no effect on the result. The uneven margin parameter (bias term) was used with three different values (Y. Li, Bontcheva, and Cunningham 2005b; Y. Li et al. 2002).

Parameter	Values tested
Neighbors	1; 5

Table 29: Parameters for KNN (author's representation)

The implementation of KNN only provides the number of neighbors as a parameter for evaluation. Two values were tested for this parameter. No algorithm-specific options were specified for Naïve Bayes or C4.5.

5.4 Incident event extraction

The preceding section detailed how news reports on child labor incidents are detected, using a classification approach. The result is a list of news reports of which each should (and in the case of 100% precision actually does) contain at least one child labor incident observation referring to a child labor incident event. The goal of the next step is to extract these incident observations together with the corresponding attribute values from the text. Building on the process described above, first the child labor incident events are extracted with their frame (i.e., the attribute values) and then the observations are derived using the values, based on a procedure comparable to cross-product.

The technique used for the EE task in this thesis reuses existing work and embeds it into the domain-specific extraction task. For this purpose, and particularly for the extraction of named entities (concepts detected within a text; NE)[76], existing taggers are reused. NER performs differently for different types of NE (Rizzo et al. 2012a). As a result, depending on the frame used for a child labor incident event, different NER approaches need to be implemented. Given that different existing taggers produce syntactically and also partly semantically different

[76] Named entity recognition (NER) has been defined as "[T]he task […] to categorize all proper nouns in a document into predefined classes like person, organization, location, etc." (Gupta and Kaur 2010, 239).

tags, they have to be aligned to a common tagset that then allows the incident events to be extracted with their attributes. This can be achieved via a domain ontology (Gangemi 2013; Rizzo et al. 2012a). "An ontology is a set of concepts relevant to a particular area of interest, thus, it is extremely domain-dependent. It aims at describing terms and relations of a domain." (Denecke 2012, 50). Within the Semantic Web, a machine readable version of the Web, ontologies are used to overcome semantic differences caused by source heterogeneity. Hierarchical relationships are a particularly important element of ontologies (Madlberger, Thöni, and Wetz 2013). Finally, given that more than one news report can depict the same incident, incident events also need to be aligned to find independent ones. Furthermore, the incident observations need to be determined based on the incident events. The output of these steps is a list of independent incident observations with fully or partly filled attributes linked to a domain ontology representing this frame. This can then be incorporated into the risk model. Figure 28 outlines this process. The following sections will describe the methods used for the four steps of this process.

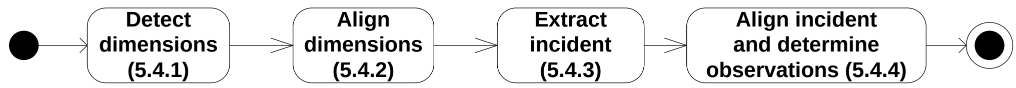


Figure 28: Process for incident event extraction (author's representation)

5.4.1 Detect dimensions

This section deals with filling the attributes of a child labor incident event to be able to derive child labor incident observations.

5.4.1.1 Child labor incident "event" template

A key step of any IE/EE task is defining the structure of the output (Appelt 1999). This extraction refers to adding additional metadata to a single event. The data are typically filled into an event template, which is also called a "frame" (Piskorski et al. 2011). For this thesis, the template has to include all necessary attributes that allow an incident event to be linked with one or multiple supplier locations represented by risk BNs. In the following, the attributes of the frame or template will also be called "dimensions". The dimensions worth including in the risk model are shown in the context of the model above: geography, company, and sector. Covering them allows the referencing of events to BNs reflecting supplier locations on a very granular level. If less data is extracted, only broad references are possible. While more dimensions could be possible, they are rarely available in news texts and will not be detailed here.

An commonly mentioned structure for an event's attributes is 4W1H (e.g. Chang-Shing Lee, Zhi-Wei Jian, and Lin-Kai Huang 2005). This stands for what, where, when, who, how. "What" is covered by the event type, "where" by the geographic dimension, "when" by the publication timestamp, and "who" relates to the company dimension and partly also to the sector dimension; however, the "how" is left unfilled. Using the publication date for a news report as the date for the event is a technique also applied in other event mining systems (e.g. Radinsky and Horvitz 2013).[77] In this way, each slot (i.e., attribute/dimension) in the event template can be seen as a one-to-many relationship, because each dimension can refer to several individual entities. This event template structure is slightly more complex than traditional, but less complex than object orientated templates applied in earlier approaches (Huttunen, Yangarber, and Grishman 2002b).

One important aspect that needs to be handled is the unique referencing of attribute values, as each value refers to an individual of the reality (e.g., a specific company). The Semantic Web (Berners-Lee, Hendler, and Lassila 2001) uses the concept of uniform resource identifiers (URIs) to uniquely identify object instances. Thus, we suggest employing URIs as the values for each dimension. References between individuals can state that they refer to the same real entity. This allows the leveraging of the Semantic Web and also reduces the risk of mismatches. Using URIs can be seen as a way to address named entity normalization, even across stories (Allan et al. 1999, 24).

Summing up, the process from Figure 28 can be expanded to the one in Figure 29. The steps for the expanded "tagging approach" section will be detailed in the following paragraphs.

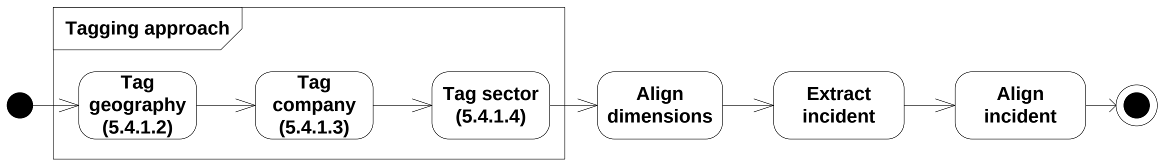


Figure 29: Process for incident event extraction including dimension tagging (author's representation)

[77] As the different attributes are related to a child labor incident they could also be interpreted as "role-based named entities" (Castellanos et al. 2012, 871). These take a specific role such as a contract end date instead of "just" being semantically a date. Similarly, the sector is an incidents related sector and not "just" a sector.

5.4.1.2 Geography tagging

Input data	Text only
Processing	NE tagging
Output	One to several tags per location provided by NE taggers' response, with URIs referencing the specific individual country

Table 30: Overview of geography tagging approach (author's representation)

The need to attribute geographical information to texts for information retrieval is well known (e.g. Pouliquen et al. 2004). Several authors have dealt with extracting geographical information from text, and it is one of the basic concepts offered by NE taggers. Multiple NERs are available on the web from different providers. These produce very different results depending on the goal. Several also include URIs for most of their responses. DBpedia Spotlight[78] and Open Calais[79] by Reuters are two of them. DBpedia Spotlight is a tagging service that can be locally installed and which uses the Semantic Web version of Wikipedia, DBpedia, as input. Open Calais is a service provided by Reuters which is provided through a web interface.

In its introductory paper, DBpedia Spotlight was presented as outperforming a list of earlier approaches on a general level (P. N. Mendes et al. 2011). However, most of these are different from the ones mentioned by Gangemi (2013), who report DBpedia Spotlight and Open Calais to be inferior to other available services for NE tagging. Nevertheless, only one recent evaluation approach is available that compares different NER approaches at the entity type level in order to distinguish them according to their usefulness for the task at hand (Rizzo, van Erp, and Troncy 2014). For news texts, Rizzo, van Erp, and Troncy present a comparison of multiple NER systems. The approaches that perform better than DBpedia for geography tagging (F1 score) (NERD[80], Lupedia[81], Wikimeta[82], Cicero[83], and AlchemyAPI[84]) are not available for local installation. Moreover, they did not use the latest version of DBpedia Spotlight in their paper (they used version 0.6) – Daiber et al. (2013) have recently significantly enhanced DBpedia Spotlight in terms of speed and accuracy (i.e., precision and recall metrics). As this thesis uses an English corpus from Reuters with restrictive confidentiality and usage constraints for evaluation, Open Calais and a local installation of DBpedia Spotlight are used. Changing the NE tagger could result in even better performance. Consequently, the evaluation values presented in this thesis should be understood as lower bounds.

[78] http://spotlight.dbpedia.org/.
[79] http://www.opencalais.com/.
[80] http://nerd.eurecom.fr/.
[81] http://www.old.ontotext.com/lupedia/.
[82] http://wikimeta.com/.
[83] http://www.azavea.com/products/cicero/api/.
[84] http://www.alchemyapi.com/.

Altogether, both services will be used to tag the input text and create tags, including the geographical tags that will be used by subsequent process steps. Relations to the definitions suggested by GeoNames, including hierarchical links, are provided through the dimension alignment process step explained below.

5.4.1.3 Company tagging

Input data	Text only
Processing	NE tagging
Output	One to several tags per location provided by NE taggers' response, with URIs referencing individual companies

Table 31: Overview of company tagging approach (author's representation)

Rizzo, van Erp, and Troncy (2014) also presented an evaluation of organization tags, with a comparable result to that presented for geography above. In addition to the one above, also TextRazor[85] performs better for organizations. Here, the same restrictions apply as above. However, in an earlier paper Rizzo et al. (2012a) show that OpenCalais is especially useful for detecting organizations. If available, Open Calais provides a URI that can be used to uniquely identify a company. The namespace of this URI is provided by Reuters and Open Calais.[86] Therefore, we suggest reusing the Open Calais tags (i.e., annotations), selecting all the relevant ones.

5.4.1.4 Sector tagging

Input data	Text (incl. lemma) and industry classifications/training data
Processing	NE tagging including industry classification codes
Output	One tag per text area or document referencing a specific industry code

Table 32: Overview of sector tagging approach (author's representation)

The final aim of sector tagging in the context of this thesis is to identify all sectors a child labor incident is referring to. One incident can refer to multiple sectors.[87] Hence, the task of extracting the sector to an event also is a special indexing case, where keywords are assigned to a text from a controlled vocabulary (Pouliquen, Steinberger, and Ignat 2003). Two approaches appear relevant here: sectors can be tagged as in the approaches presented above, or texts referring to child labor incidents can be classified with one or multiple sectors.

In both cases a list of sectors is needed as input. Given that the statistical data used in the model refers to the United Nations ISIC industry classification (e.g. United Nations Statistics

[85] https://www.textrazor.com/.

[86] It is slightly problematic that not all companies and organizations have a URI included in their tag. Consequently, some tags are later augmented with a generated URI based on the surface string.

[87] In the gold standard described below, 75 of 117 documents dealing with child labor incidents referred to a sector. Of these, 14 (19%) referenced more than one sector.

Division 2008a), ISIC or a comparable classification structure seems appropriate here. This also allows an existing classification to be reused and eases understanding.

In the literature, the NERD project (Rizzo et al. 2012a) compared ten different named entity extractors and also provided its own meta tagger. None of the eight reachable named entity extractors[88] provides direct links to specific concepts of a standard industry classification. A similar check was performed for Rizzo, van Erp, and Troncy (2014). Tagging product terms and automatically relating them to a specific industry using the tagsets themselves or the corresponding URIs is also not possible.

Most recently, Santos et al. (2013) presented an approach for tagging economic activities. They suggest leveraging the composition of the text and other contextual information.[89] However, their overall objective was to find a broad range of economic activities and company variants, not to classify the company's activity with regard to a standard. In contrast, Paradis, Nie, and Tajarobi (2005) depict a system for classifying text based on a standardized sector. Their approach derives the sector based on the first sentence of tenders. Paradis, Nie, and Tajarobi build on a training set of pre-classified tenders used to learn a classifier; however, this is not possible here. They also do not use an ontology. Finally, Du et al. (2013) present a system that employs event-based sector tagging based on ML classifiers for 40 different sectors trained by an individual training set for each classifier. Nevertheless, their approach involves manual tagging for each sector and they focus on industry classification mainly used in banking (GICS; Global Industry Classification Standard; MSCI 2014).

Consequently, we apply two different approaches that will be presented below – a rule based gazetteer and an ML-based approach which are both based on details available from industry classification standards. The latter has connections with the approach presented by Paradis, Nie, and Tajarobi (2005). For both cases we suggest using the United Nations ISIC V4 industry classification (United Nations Statistics Division 2008a), given that this classification is also widely used for statistical purposes. Statistics presented by major organizations related to child labor also use this categorization (e.g., UCW). It has a related Central Product Classifi-

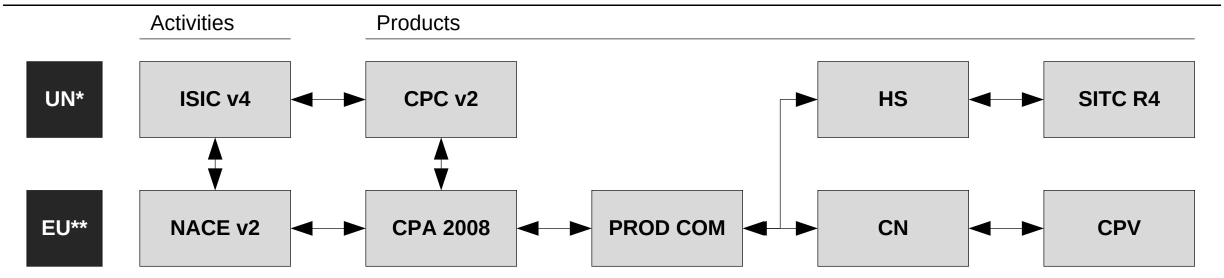
[88] Three services (Evri, Extractive, and YCA) were not reachable at the time of testing (November 2014).

[89] They differentiate between sector, product, and actions and see economic activity strings as composed of one or several of these elements. For example, "production of computers" would be split into the action "production" and the product "computers." Actions include activities that can be done by a company. In the end, their use case is to identify companies' economic activities (mainly from company descriptions) and not to understand the sector context in which an incident occurred. Nevertheless, their contextual clues use keywords to identify words that describe sectors or products. Although this can only help to tag a word as potentially relevant, this technique can be used to differentiate related sectors, as will be discussed below. Finally, their system uses Portuguese and includes language-native product classification standards for basic tag names.

cation V2 (CPC, United Nations Statistics Division 2008b) scheme that is linked to the ISIC industry classification and which will also be partly leveraged, particularly for the rule-based tagger. Figure 30 shows the link between these and other standards. Both classification schemes provide descriptive labels for identifying the elements (e.g., products or services) included in a specific classification element. Table 33 provides an overview of the two classification schemes and the number of categories in them. Both systems are also available in multiple languages. The product tags allow for a much more finely grained tagging procedure, as is apparent from Table 37.

Attribute	ISIC V4	CPC V2
Focus	Classification of industries	Classification of products
No. of hierarchy levels	4	5
No. of elements in 1st hierarchical level	21	9
Total no. of elements in classification scheme (line number)	766	4,409

Table 33: Overview of ISIC V4 and CPC 2 classification systems (author's representation; United Nations Statistics Division 2008b; United Nations Statistics Division 2008a)



* UN … United Nations; ** EU … European Union

Figure 30: Connection between ISIC v4 and CPC v2 and other standards (author's representation; partly based on Statistik Austria 2010)

5.4.1.4.1 Rule-based gazetteer

The rule-based gazetteer tries to tag the text, much like other taggers (used for, e.g., "child" as shown above). Here, we suggest using an adapted and extended gazetteer that includes a reference ontology based on the ISIC and CPC industry classifications. In a sense it is a specialized and adapted form of ontology matching (e.g. Anjewierden and Kabel 2001) also employed in the biomedical domain (Botsis et al. 2012). In order to comply with the approaches for geographic and company tagging, URIs need to be used. Both ISIC V4 and CPC V2 classification systems have been published as structured lists (United Nations Statistics Division 2008b; United Nations Statistics Division 2014), and a toolset to transform the CPC scheme to an ontology has been published by Stolz et al. (2014) which is based on the work of Hepp and de Bruijn (2007), among others.[90] The transformation preserves hierarchical relationships

[90] An alternative RDF version of both the ISIC and CPC schemes was presented by Gayo and Rodríguez (2011). Their approach transforms the classifications to members of an ontology consisting of two classes. In this ap-

and includes all metadata in the classifications as labels. The toolset has been slightly adapted to also include the links to the ISIC V4 classification scheme. As a result, two RDF ontologies, one for ISIC V4 and one for CPC V2, have been created that allow for unique referencing of industries and products.

The ontologies provide URIs for the different sectors and products mentioned and modeled as classes. The majority of CPC classes can be linked to one or more ISIC codes and classes. A tagset can be created by deriving keywords for a gazetteer based on the labels from the ISIC and CPC codes, which allows each gazetteer value to be referenced with one or more ISIC codes. This is a three-step process, as outlined in Figure 31.

Figure 31: Process for initializing sector tagging (author's representation)

A key step in this process is transforming the labels into useful keywords. Example labels are shown in Table 34. Given that the description labels follow very similar patterns, rules can be used for this purpose. An automated rule-based approach allows the tagging framework to be reloaded each time a new version of the classification scheme is released. This is specifically relevant because classifications and products are continuously evolving (see e.g. the example provided by Stolz et al. 2014, 2). Nevertheless, given that ISIC and CPC labels have been manually compiled, rules may not be able to cover all ambiguities.

Classification	Code	Description label
ISIC	A	Agriculture, forestry and fishing
ISIC	0113	Growing of vegetables and melons, roots and tubers
CPC	0192	Fiber crops
CPC	01962 (ISIC: 0119)	Cut flowers and flower buds including bouquets, wreaths, floral baskets and similar articles

Table 34: Example ISIC and CPC classification elements (author's representation; United Nations Statistics Division 2008b; United Nations Statistics Division 2008a)

To derive keywords from the labels, each label is subsequently split into segments and parts (see Figure 31). Given that a large number of labels contains phrases and words that detail and specify the meaning of more general class-related terms, an additional differentiation is used. "Keywords" refer to surface terms (i.e., the actual text) that primarily link to the sector class, while "specification words" refer to surface terms that provide a constraint when a single

proach, hierarchy is only retrievable through an interpretation of textual information in textual attributes to individuals. Moreover, it does not include specific ISIC codes. It could be an alternative option, but due to the direct availability of hierarchical relationships, the other approach was chosen.

keyword refers to multiple classes. Here, this specification is seen as directly related to one label. This may be further extended to include a specification for each label part. Figure 33 shows the detailed process of deriving keywords.

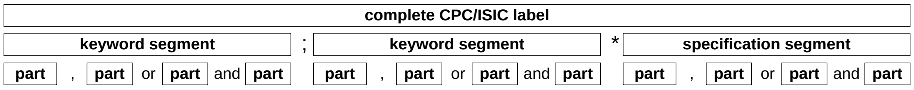


Figure 32: ISIC and CPC label structure for main part and specification (author's representation)

This set of rules and the terms used in each process step (see "Appendix prototype implementation" and "Appendix J: Word lists used for rule-based sector tagging" for details) have been manually defined. First, a manual list of keywords and specifications was produced for each ISIC and CPC class. The results of the rules were then compared with the manual set for improvement options. In general, the automatic, rule-based approach presents a trade-off between the number of rules and the comparison with the manual approach. The manual list of keywords and specifications has been designed to include the most important differentiating terms between classes while also respecting hierarchical relationships.

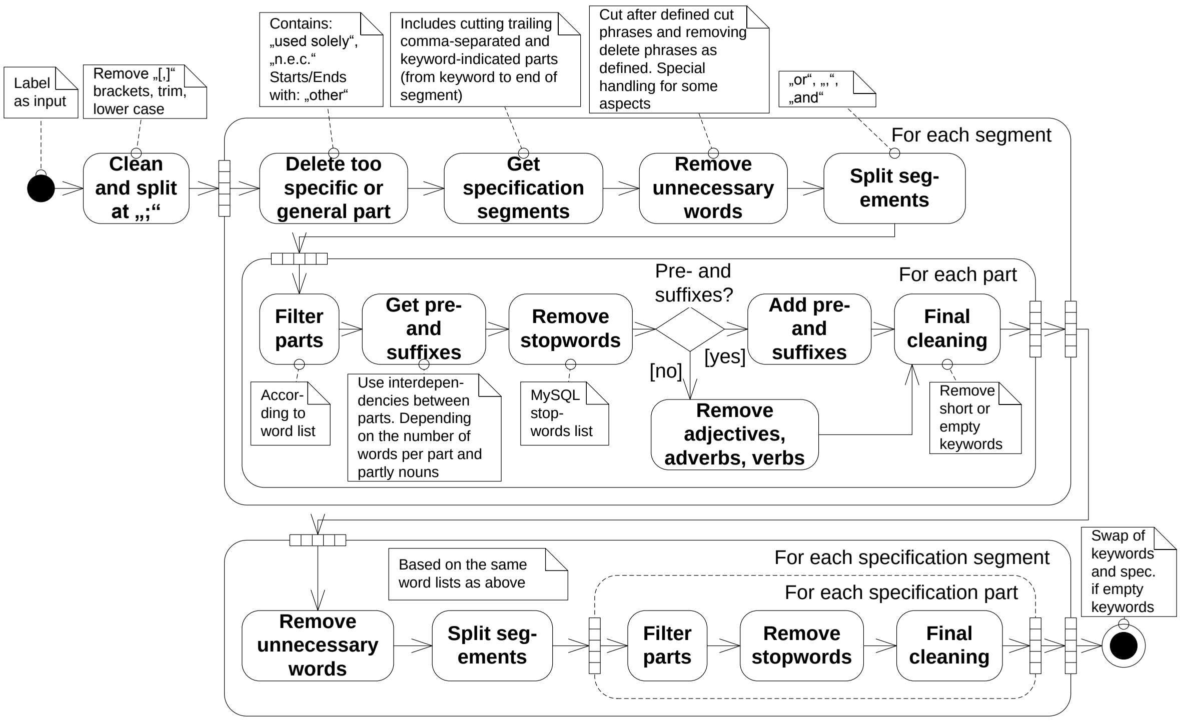


Figure 33: Process of deriving keywords and specifications from a single label (author's representation)

Figure 33 shows the process of generating keywords and specifications from a single input label using the segments and parts of the label.[91] This procedure performed for all ISIC and CPC labels forms the tagset, consisting of keywords and specifications.

---

[91] After cleaning the initial label (e.g., removing brackets), segments are determined by splitting at semicolons. Before specifications are extracted, specific or general labels (e.g., referring to "all other") are deleted. Next,

Finally, the tagging can be performed based on case-insensitive string matching by considering all tokens in a document, one after the other, in ascending order. First, the match strings are created by considering two different types: The first simply adds the different surface strings and the other includes the word's lemma if available. For each token position $2 * n$ match strings are created given that a user may enter a predefined number of words to be considered for matching (n). For example, if the user sets n to four, then one string with the current word, one with the last and the current word, one with the last two words and the current word, and finally one with the last three words and the current word are created, both for normal string matching and for lemma matching. This is shown in more detail in Figure 34.[92]

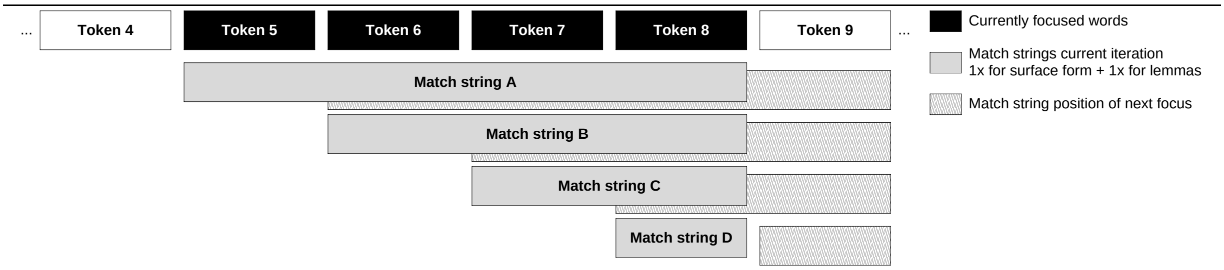


Figure 34: Matching process for string matches (maximum string match constant set to four; author's representation)

When a match with a keyword is found, a tag is added. If the tagging is performed with specifications then the sentences touched by the match string are checked for the specification term. The tag is only added if another match is found.

5.4.1.4.2 Machine-learning classification

In contrast to the rule-based gazetteer, the ML approach aims to classify the text with an ISIC code. Here, we restrict the work to classifying each incident with only one sector, as incidents in their most granular form (i.e., seeing a child performing a specific task) can be directly related to one sector.

Given that a sufficiently large tagged training set for sector tagging would be required covering all ISIC classes, a different approach has been chosen. The labels and descriptions from the ISIC classes can be used to train a classifier based on the lemmas of the respective tokens.

---

specifications primarily triggered through keywords are transferred to a buffer and unnecessary words are removed either by cut-off or space-replace. The subsequent process steps work on the part level, which is achieved by splitting at "and," "or," and commas. Once again, unnecessary parts are directly deleted before prefixes and suffixes are determined. These are then added to the other words. Parts containing only adjectives, adverbs, and verbs are removed if no prefix or suffix is available. After final cleaning (e.g., trimming), specifications are handled. Specifications are edited in a similar way to keywords, apart from not detecting prefixes and suffixes or removing adjectives, adverbs, and verbs.

[92] This matching approach may be compared to the one suggested by Willis and Losee (Willis and Losee 2013, 1333) who also matched cleaned terms from a thesaurus with different *n*-grams from a document.

The descriptions include explanatory notes with additional details.[93] Three different strategies are proposed: First, one can take the full label or sub-parts. Second, one can take each non-empty line of a label and the description as a separate learning document. Third, one may generate parts comparable to the approach depicted above by also splitting at semicolons, commas, points, and “and”. Additionally, the ISIC labels and descriptions have been extended with CPC labels and descriptions that include a one-to-one ISIC code reference and tested with the per-line splitting. For feature selection, only features also available in the training set are chosen as input (incl. stopword removal which is also used in the training phase), given that training is also restricted to them.[94] Either a sentence-level or a minimum sentence-level approach is tested (also due to the results obtained from the document classification).

After the training, the testing can be performed with the same test data used in the general dimension evaluation. To use the created classification they need to be added to the document as also performed for the other dimensions above.

5.4.2 Align dimensions

Input data	Tags from geography, company, and sector tagging
Processing	Detecting tags matching to relevant dimensions using mapping ontology; Geography: Align tags of DBpedia Spotlight and Open Calais that refer to OpenGeoNames; Company: Reuse Open Calais URI; Sector: Reuse Sector Tagging URI
Output	One or several tags per word(s) containing the dimension information of one dimension with aligned URIs. Multiple URIs per tag are possible.

Table 35: Overview of dimension alignment approach (author’s representation)

The alignment approach aims to combine the information generated by multiple taggers to create a single consistent tagset referring to an integrative ontology. This output is comparable to the approach used in the NERD ontology (Rizzo et al. 2012b; Rizzo and Troncy 2012). NERD suggests a top-level ontology to combine the results from the multiple extractors, including DBpedia Spotlight and Open Calais (Rizzo et al. 2012a). Here, we suggest a domain-tailored ontology to structure the dimensions. Whereas Denecke (2012, 49) uses equivalentClass relationships to specify semantic relations between concepts of multiple ontologies, we vote for super class relationships using a shared super class.[95] Moreover,

[93] These explanatory notes might be considered in a further version of the rule-based gazetteer.

[94] If an input document is empty when focusing on export words only, then this document is removed from the overall set. The export words are created using the lemmas from all input features longer than one character in ISIC labels and including explanatory nodes, excluding MySQL stopwords (MySQL 2009).

[95] To include hierarchical information, Castellanos et al. (2012) employ the concept of hierarchical similarity within a hybrid SQL syntax. Using hierarchical relationships from linked data has also been applied before (Radinsky and Horvitz 2013). Other systems encode their calculation of administrative hierarchical relationships

Denecke's approach is broader than needed in the specific context here, and it builds on the general and not news-specific classification proposed in Quaero (Galibert et al. 2011).[96]

We reuse multiple ontologies that have already been mentioned in the previous processing steps. Both DBpedia Spotlight and Open Calais use ontologies to categorize their tags. Each tag provides the respective ontology class whose individual is referenced with the tag based on the surface form. For DBpedia Spotlight this is the standard ontology of DBpedia. Figure 35 provides an overview of the four external ontologies that are combined for this processing step.

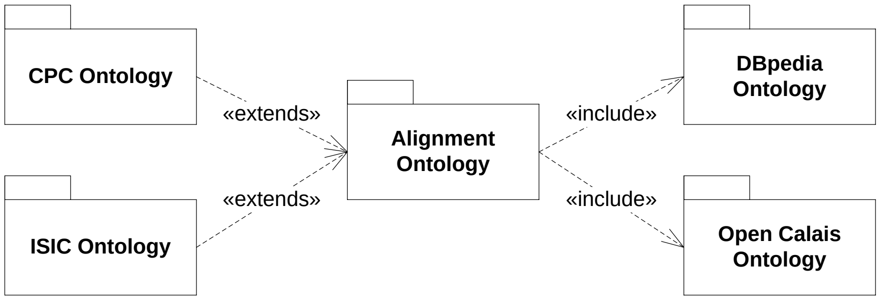


Figure 35: Ontologies used for the alignment processing step (author's representation)

Conceptually, the alignment approach is based on one upper level ontology called "alignment ontology" that is used to integrate the different other ontologies used during the tagging. This upper ontology conceptualizes the dimensions of interest for the application at hand (as depicted above in Figure 35). All other concepts are then integrated into this scheme with subClassOf relationships. Hence, the matching in a hierarchical sense has been performed manually (Euzenat and Shvaiko 2013, 41). The left hand side of Figure 36 shows the structure of the ontology and its hierarchical, taxonomic relationships. In general, ontology alignment has seen deep research involvement, especially also in the domain automated matching not performed here (see e.g. Euzenat and Shvaiko 2013 for a broader overview of the topic).

based on ID-encoded relationships. Thus, no lookups are necessary, but IDs need to be changed if political relations change (Piskorski et al. 2011).

[96] Elements are included in the NERD ontology if they are a concept in two domain ontologies/taxonomies of the APIs referenced by NERD (Rizzo et al. 2012a, 4). However, the NERD ontology does not provide the granularity necessary for differentiating the frame and the included hierarchy of frames used here.

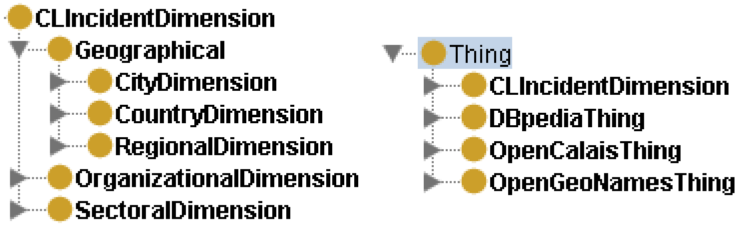


Figure 36: Structure with sub class relationships of alignment ontology and integration of DBpedia and Open Calais ontologies (author's representation)

To generate the alignment ontology, the DBpedia and the Open Calais ontologies are imported (referenced in OWL) and included under an overarching structural superclass element called DBpediaThing and OpenCalaisThing respectively (see right hand side of Figure 36). OpenGeoNames[97] is used by DBpedia and OpenCalais to reference geographical entities. To relate classes from these imported ontologies with the alignment ontology classes subsumed under CLIncidentDimension, the property that a class can be the sub-class of multiple classes in OWL is used. These relationships have been manually formed and include only a subset of the ontologies, because only relevant tags have been used.[98]

The labels of the tags created using the alignment approach refer to two conceptual layers: a class and an individual layer. The attributes of the class layer are used to identify when a specific input tag from the alignment algorithm contains one of the dimensions under consideration. In contrast, the attributes of the individual layer represent the information necessary to identify the individual (i.e., instance; Techopedia 2014) tagged by the NE extractor. While a specific instance exists for geography and company, this is not the case for sectors, where instances would be one specific product or service. Consequently, no instances are used for these dimensions, and only the class URIs are passed on to further processing steps. This design is shown in Figure 37.

[97] http://www.geonames.org/

[98] "Appendix K: Turtle code of alignment ontology" lists the full structure of the alignment ontology. Differently to DBpedia and Open Calais, the ISIC and the CPC ontologies are not directly included, but imported and kept in a separate file.

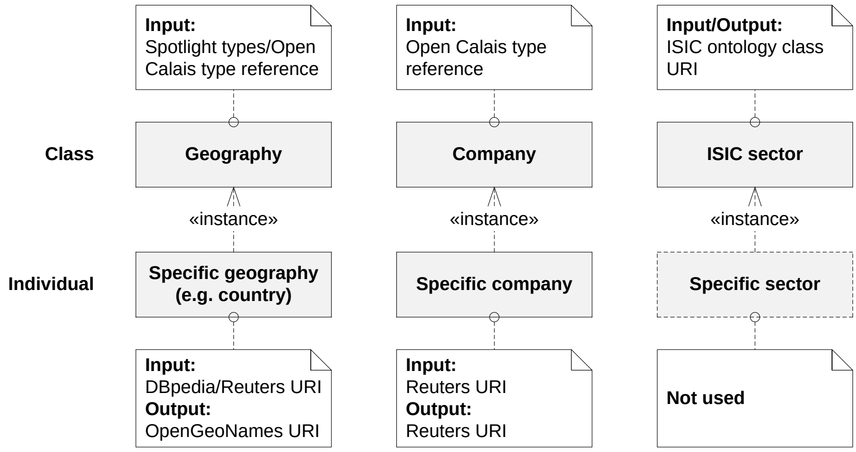


Figure 37: Layers used to map/align existing tags to consistent tagsets (author's representation)

As well as the differentiation into classes and individual layers, Figure 37 also includes the inputs used for establishing the aligned tagset. These inputs are the features of the tags produced by the multiple engines. Altogether, a multi-step process (Figure 38) allows the creation of the aligned tagset.

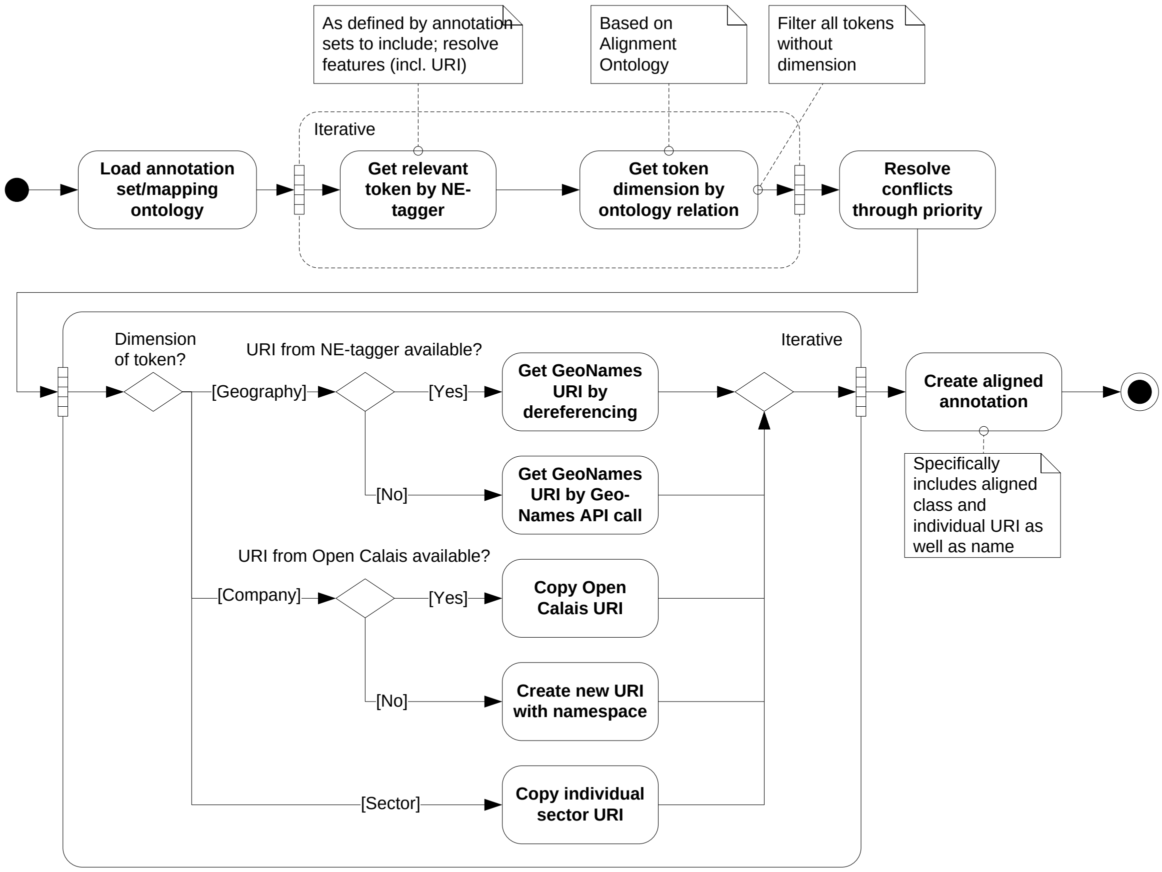


Figure 38: Tag alignment process (author's representation)

First the alignment ontology is loaded. Then a conflict-free set of basic aligned tags is produced for each dimension by parsing the different types of input annotations and respecting the user-specified prioritization of the taggers used as input. A tag is only added to the resolved set if it is linked to one of the dimensions of interest, as specified by the subClassOf relationships in the alignment ontologies. This is achieved by processing the class URIs and resolving the respective relationships in the ontology. Depending on the dimensions attributed to a tag in the cleaned list, the aligned tags are produced reusing the data available (for company and sector dimensions) or by querying linked open data (LOD).[99] For the geographic dimension, the URIs are dereferenced to be able to identify the entity encoded in the tag with a URI from OpenGeoNames. This is possible for most of the original tags that include URIs pointing to either DBpedia[100] or a specification provided by Reuters[101]. If not available, a full text web service can be used to query using an individual's name. In summary, the data is used to enrich the aligned tag which is finally added to the text.

5.4.3 Extract incidents

Input data	Aligned tags referring to alignment ontology classes
Processing	Detect child labor incident event attributes (i.e., the incident frame) by applying different heuristics to extract tags related to the incident
Output	One incident event per document which includes all attribute values available in a document for a child labor incident

Table 36: Overview of incident extraction approach (author's representation)

Incident extraction refers to the task of applying EE to extract incidents from the text. Based on the assumption that one news report contains one child labor incident event that may be composed of several child labor incident observations (see previous chapter), a specific approach is necessary. Attribute information for the incident event may be scattered across the whole text, making extraction more difficult (Huttunen, Yangarber, and Grishman 2002b). Naughton, Kushmerick, and Carthy (2006) also deal with the question of how to identify spans within an article dealing with a specific event. They also list the two main problems faced by this task: that a single article may refer to several events at the same time or a single event may be referenced by multiple parts of the same article. Even a single sentence can contain multiple events. Consequently, the main goal of this process step is to relate event attribute values to the event.

Several approaches appear possible. The approaches already mentioned often apply patterns. A pattern searches for a character-sequence or semantic type when applied to a text, called "match elements" in the following. To find a relation between two match elements, three

[99] LOD allows for easy publication and integration of a diverse set of different data based on a semantic context. By fostering the amount of web-enabled data, LOD as a special field has nourished the content necessary for a semantic and machine-readable form of the web (Bizer, Heath, and Berners-Lee 2009). Semantic representations combined with online and interlinked accessibility are important features of LOD (Heath 2008).

[100] http://dbpedia.org/resource/.

[101] http://s.opencalais.com/1/type/er/.

types can be generically differentiated: First, two or more match elements can be separated by only a space or another punctuation mark. Second, the number of words between two match elements can be defined. Both cases seem very restrictive and require designing patterns based on a sufficiently large training or test set. Third, the words in between two match elements may not be defined. This can be compared to the situation where an “anchor” defines the child labor mention and another pattern is used to identify the attribute mention. This approach will be one of the suggested approaches and several variations of it will be suggested.

Heuristics to associate dimensional information with an event have been used before.[102] In the case of IE, shallow-parsing approaches are often sufficient (Appelt 1999). Therefore, we restrict ourselves to approaches that employ shallow-parsing. As briefly discussed, in controlled domains rule-based approaches can still be seen as up-to-date and relevant (Elloumi et al. 2013).[103] The approach applied here includes features often used to represent statistical connections (relations known from causality detection; Cao et al. 2014). Shared objects may indicate a connection, temporal relation may indicate a connection, and appearing together in a local context with limited distance may indicate a connection. The last heuristic is particularly useful here. For the anchor features indicating a text part deals with child labor, we suggest either using the child labor relations suggested in section 5.3.1 or the word mentions indicating “child” or “labor” introduced in the same section. The number of potential child labor relations can be reduced by using the minimum number, as defined above[104].

Because the frame of a child labor incident event allows multiple values, simply selecting and incorporating one value into the frame of the incident may not be sufficient. Nevertheless, to allow a comparison, a strategy of selecting only the closest attribute found for each dimension is also incorporated. Other authors have used the character distance as a heuristic (Aone and Ramos-Santacruz 2000; Leetaru and Schrodt 2013). To limit the search space, an extension with a limited maximum character distance is added (this may also have a positive effect on precision). To select multiple attributes, we suggest a variant of this idea. By not only select-

[102] Jin, Myaeng, and Jung (2007), for example, suggest a heuristic to link geographical place information with an event by ascribing more weight to names appearing in the first section of an article. Other authors include dependency parsing in order to derive features of an event that relate to each other (see approaches in section 2.2). However, whereas dependency parsing is focused on sentences, attribute information for events can also be found across sentence boundaries (this might partly be overcome with co-reference resolution). Moreover, dependency parsing involves partly deriving complicated textual relations, which may result in a significant increase in possible sources of error (e.g., co-reference resolution). A similar argumentation holds for semantic role labeling (Gildea and Jurafsky 2002), which also requires a semantic connection between two entities to determine them as related. Given the computational cost of deep-parsing, including dependencies only makes sense if the gains outweigh the additional processing cost.
[103] See section 2.2.
[104] The “minimum” word mentions indicating “child” or “labor” are those mentioned within the minimum distance relations.

ing the closest, but all attributes within a pre-defined character-distance, multiple attributes can be identified. Using a shallow-parsing heuristic instead of dependency parsing or semantic role labeling, the sentence structures can be used by selecting all attributes within sentences touched by either the distance relations or the indicative words also leading to a multi-selection. Finally, we suggest taking the attribute most frequently mentioned for each dimension. Figure 39 outlines the resulting extraction process.

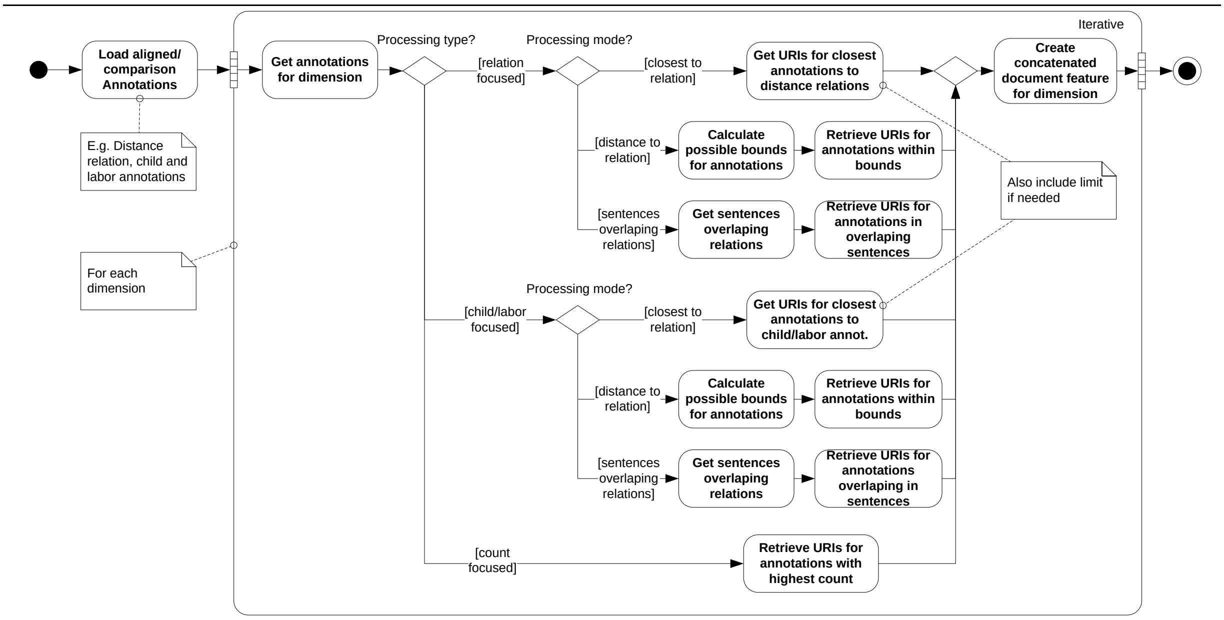


Figure 39: Process for incident extraction (author's representation)

5.4.4 Align incidents and determine incident observations (folding)

Input data	Extracted incident events per document
Processing	Derive incident observations from events and determine overlaps between incidents.
Output	Independent child labor incident observations

Table 37: Overview of incident observation folding approach (author's representation)

The general aim of this process step is to derive independent observations. The initial observations are derived from the incident events via a process comparable to a mathematical cross-product. In the following, this process will be referred to as "observation extraction". This set then needs to be reduced to a set that only contains observations that independently refer to child labor incidents. If several observations are seen to be not independent, they are folded (i.e., merged). Merging is a common way to align similar event mentions where co-reference needs to be included (Hobbs et al. 1997).[105] As defined above, "independence" could mean a new event or an independent verification of a known event. Given that both but

[105] Consequently, the approach first derives the new events before removing non-independent ones (for a related approach see e.g. Piskorski et al. 2010).

particularly the latter are hard to detect, heuristics appear needed. This is also because the background in which a certain story has been crafted remains unclear and cannot be objectively assessed based on the text alone. We suggest building on new event detection.

New event detection approaches often focus on high volume events that are particularly prevalent in news media (i.e., "bursty" events; W. Chen et al. 2010). If large amounts of news are processed and more broadly discussed events are important, then clustering can be employed to group text referring to an event (Atkinson et al. 2008; Steinberger, Pouliquen, and van der Goot 2009). Given that multiple documents are available in clusters, data may be blended across clusters (Atkinson et al. 2008). However, as child labor events may be described within other, longer texts, this approach is limited in the current context. When folding observations, we assume three of the four assumptions given by Hunter and Summerton for their event merging calculus (Hunter and Summerton 2006, 6–7): An observation depicts exactly one child labor incident; an observation does not necessarily contain all information about the event; an observation is correct. The last assumption is only used for the process of folding. This allows an event template to be updated with additional data while concluding that conflicting information must be from different events. We do not assume that no temporal information is present.

In general, new events can be detected by either exploring textual features or relying on the attributes of the event (e.g., location, time, etc.; Denecke 2012, 43). The former is called "feature co-occurrence" and aims to detect events by comparing word frequencies or word bursts.[106]. The latter is called the "event-property based approach" and relies on first performing IE to then be able to compare the extracted attributes of a potentially new event (Denecke 2012, 42–43; Wei and Lee 2004). This can for example be done by comparing the attributes of events and detecting new events if certain thresholds of dissimilarity are reached (Piskorski et al. 2011). Employing rules (Aone and Ramos-Santacruz 2000) or keywords (Atkinson et al. 2008) is also possible. Here we suggest mainly following the latter approach. In a broader case, Allan et al. state that the "[…] assumption that every named entity is unique to a topic generally holds true." (Allan et al. 1999, 22). Yet, this also depends on the context of the categorization (Kumaran and Allan 2004). Here, this statement can be transferred to event

[106] This may also be performed at the sentence level (Martina Naughton, Kushmerick, and Carthy 2006; M. Naughton, Stokes, and Carthy 2010).

frames. Hence, changes in named entities may be used to detect and differentiate old from new events and, in the current context, observations.[107]

Given the limited number of dimensions available to distinguish different observations for comparison, we suggest directly comparing the dimensions of a new observation with existing observations in order to determine overlaps. Additionally, this approach can accommodate for hierarchical relationships between dimensional attributes of geographic links and sectors (see Jin, Myaeng, and Jung 2007 for geographic hierarchies). More sophisticated methods could be built on cosine similarity measures and SVM classifiers (Kumaran and Allan 2005). However, these seem more relevant in settings where the frequency of events is higher. Therefore geography, sector, and company are used as dimensions, along with time. Organizations, people, and locations have also been identified before as potentially helpful features for detecting new stories (Allan et al. 1999, 20).

As stated, from the TM perspective, the independence of reports on child labor is not fully provable. Consequently, it has to be assumed when two incident observations are independent even if their attributes (partly) match. The loosest solution is to view every report as independent. This would represent the view that any media report has a certain value and will increase the belief that a child labor event will be more likely. At the other extreme one might use only the first report matching one of the attributes for a supplier location, skipping all others. This would suggest that every matching article is a copy of the first.

Alternatively, there are several possibilities in between. First, it can be argued that articles that are written differently while sharing the basic attributes of an incident (e.g., city, sector, and approximate timing) are independent.[108] This requires differentiating the writing in two articles, which is possible using text comparison. For example, the generalized Levenshtein distance could be used, which is based on the amount of edit operations needed to transform one string into another (Levenshtein 1966). Second, the relationship between the authors or at least the sources of two reports could be retrieved in order to evaluate how likely they are to be independent. Multiple data sources such as websites or social networks (e.g., checking links between them) might be used for this purpose. Third, child labor incident reports written

[107] Building on this, Kumaran and Allan (2005) also propose a hybrid approach that successfully includes named entities and topic terms for new topic detection. Topic terms are all terms from a text that are not named entities. While this is reported as useful, topic terms in the context of this thesis provide contextual reference which is already partly aligned due to the preceding classification and restriction to child labor incident related news. Consequently, the approach in this thesis uses the named entities in the observation's attributes, but does not suggest topic terms.

[108] This would mean a combination of the two general approaches.

within a predefined timeframe can be considered to not be independent. This view assumes that a news report that hypes quickly includes one initial message that is repeated multiple times. Two options can be considered: Two messages are considered not independent if and only if they were created in the same timeframe and they have (1) exactly the same attributes, or if they have (2) at least non-conflicting attributes. In the latter case it is possible that non-initial messages might include additional information which can be used in conjunction with the primary one. The add-ons would then only be incremental and not seen as an independent child labor incident observation. This final technique more exactly specifies the meaning of "folding" in this thesis. For example, if two texts both reference a potential incident with India/Tamil Nadu (geography) and textiles (sector) then these two items are folded to form one incident. If no time-based thresholds are defined, then an event from this year can be merged with one with more general properties from last year (Allan et al. 1999; Kumaran and Allan 2005).[109]

Altogether, we see two stories referring to the same event if they have non-conflicting dimensional attribute values and are published within a certain timeframe. At the lowest level of granularity, when a specific supplier in a specific city in a specific sector is described, the likelihood of two independent child labor incidents within a short amount of time is considered to be very low. Nevertheless, with decreasing detailed information given in a news report on the three dimensions used in this thesis, the probability of an overlap increases. For example, when a news report refers only to a region and a sector, two incidents might be present when news spanning several days is folded. Each observation needs to be compared with all preceding observations with overlapping attributes.[110]

The close time-relations between articles referring to the same incident can readily be seen when investigating news sources. When Primark was connected with child labor in 2008, most international news reacted within a few days (Figure 40 – area 1). Nevertheless, at some remote period another article referred to child labor in relation to Primark (area 2). Depending on the threshold defined, this could then be marked as an independent incident. For each incident, the media messages used as input for the calculation are tracked.

[109] Timeframes have also been suggested by "classical" new event detection based on the assumption that two news stories with the same attributes are more likely to be about the same event if they appear closer together. Allan et al. (1999, 27) have experimented with an exponential decay function on a two month corpus, detecting no effect. However, they stated that the limited time span of the corpus might have affected the result. Moreover, the special case where there are only a few stories on a potential social sustainability event such as a child labor incident does not apply. Other authors have also successfully employed time-based constraints (Leetaru and Schrodt 2013; Lei et al. 2005, 756; Sankaranarayanan et al. 2009; Smith 2002).

[110] Other approaches might use full document data and not only event data to decide if a new incident is present. Such approaches are related to TDT techniques, e.g., those presented by Papka (1999).

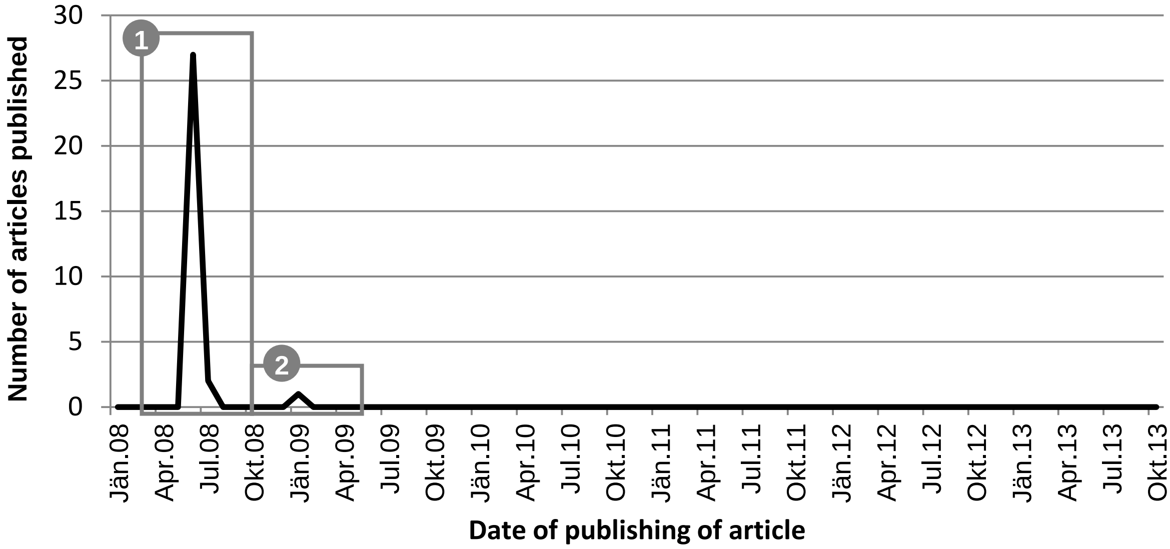


Figure 40: Number of results for "child lab*r" articles for Primark related to India found in Factiva (author's representation)

In this thesis we propose defining a threshold that limits the maximum amount of time for an observation to be considered independent. A folded observation retains the timestamp of the old observation, given that this one needs to be closer to the original event. The resulting process is shown in Figure 41. Figure 42 shows an example workflow that first splits an incident event into two observations before they are folded again to form two IIOs.

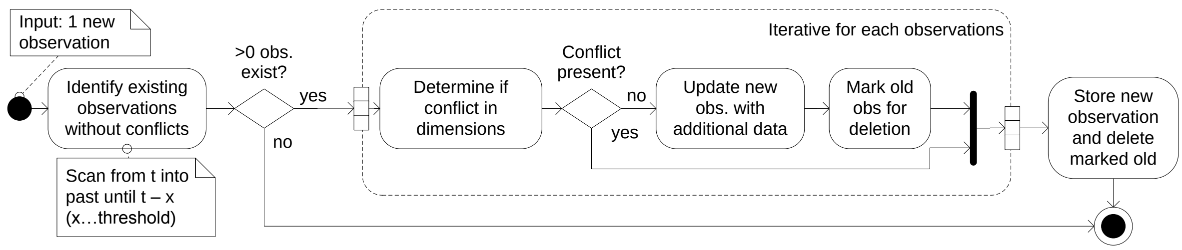
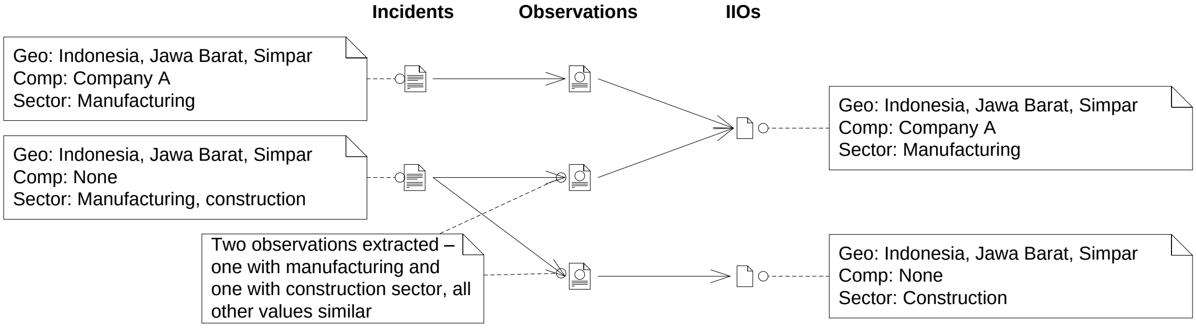


Figure 41: Process for folding observations (author's representation)



Note: Both news items are received within the threshold that would let them be considered as separate observations

Figure 42: Example for incident extraction and folding (author's representation)

5.5 Domain/Data model

To be able to combine the information gained from text mining with the risk model, and to allow persistent storage of the data, a domain and data model are needed. Because the TM application strongly builds on URIs to define individual entities from the Semantic Web, the data model underlying the risk and text mining model is suggested for building on the same standards. Figure 43 provides an overview of the "RDF-logic" [111] based model.[112]

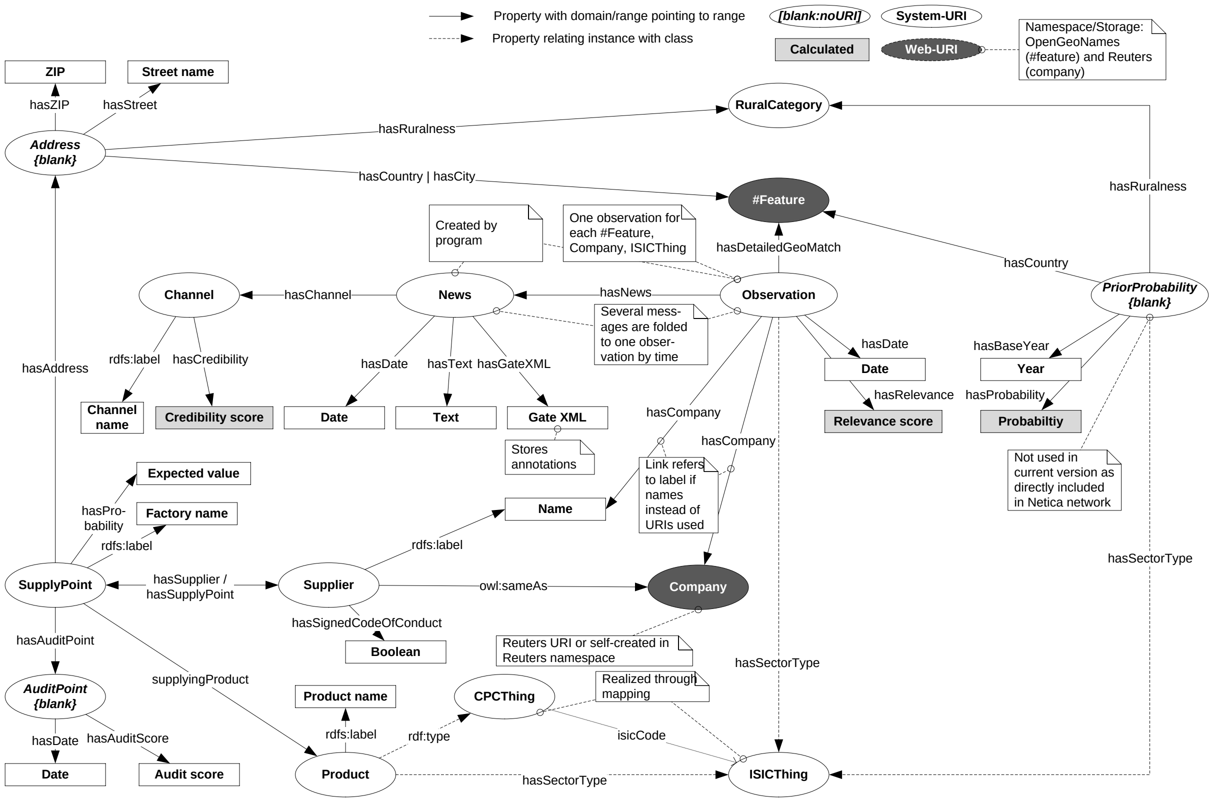


Figure 43: Domain/Data model (author's representation)

One key node in the data model is the supply point representing the lowest granularity of the supply chain model. Being a physical place, the supply point has an address, which can then be associated with ruralness (rural category), city, and country. Furthermore, each supply point is linked with a supplier which corresponds to a company, as in the Reuters data model. A supply point is also a source of products which can be related to CPC and ISIC codes. The other key node is the prior probability, which refers to a sector type (ISIC), country, and ruralness category, and which provides a probability value for each of these constellations. Finally, observations are also stored which refer to a sector (ISIC) and a geographic entity. Observations can also relate to a company. Each observation references the news (incl. the

[111] Resource Description Framework (W3C 2014).

[112] A data model based on LOD for compliance monitoring has also been suggested by Hofman (2013).

text) from which it was derived. This news is related to a channel (i.e., the outlet) providing a credibility score.

5.6 Summary

In summary, the TM approach mainly reuses text classification and IE techniques to derive child labor incident observations from plain texts tagged with only a creation date. It proposes an application that is tailored to the domain of child labor incident detection. It particularly builds on the two-worded property that can anchor a child labor incident in a text. Nevertheless, it may also be applicable to similar social sustainability incidents (e.g., forced labor, corruption) which can be defined in an analogous way. Table 38 includes a list of suggested two word mappings to other social sustainability issues. Particular emphasis is placed on EE and the special feature that child labor incidents can be referenced in one or multiple parts of a text while spanning across one or multiple sentences.

GRI 3.1 social categories	**Example 2-worded mappings for incidents**
Social: Labor Practices and Decent Work	
Employment	unfair contracts
Labor/management relations	management conflict
Occupational health and safety	safety accident
Training and education	training shortage, education shortage
Diversity and equal opportunity	gender discrimination
Equal remuneration for women and men	wage gap
Social: Human Rights	
Investment and procurement practices	illegal investments
Non-discrimination	work discrimination, racial discrimination
Freedom of association and collective bargaining	forbid association
Child labor	child labor
Prevention of forced and compulsory labor	forced labor, compulsory labor
Security practices	security issues
Indigenous rights	right conflict
Assessment	ineffective assessments
Remediation	stopped remediation
Social: Society	
Local communities	community conflict
Corruption	corruption case
Public policy	public intervention, lobbying affair
Anti-competitive behavior	anti-competitive behavior
Compliance	compliance issue
Social: Product Responsibility	
Customer health and safety	health scandal
Product and service labeling	labeling issue
Marketing communications	discriminatory advertising
Customer privacy	privacy concern
Compliance	compliance breach

Table 38: Example two word incident and problem suggestions for GRI categories (author's representation; Global Reporting Initiative 2011)

The frame of the events and observations is adapted to the SC context to allow a clear reference to a supplier location, including geography, company, and sector attributes. For the first two, the use of existing named entity taggers is suggested. For the sector attribute, no existing viable approaches could be identified. Consequently, a sector gazetteer has been proposed that uses ISIC and CPC and preserves hierarchical relationships. All tagging uses URIs included in basic ontologies to identify entities and relations which can later be directly integrated into the overall system's data model. Finally, different recall-prioritizing heuristics have been suggested for selecting the dimensions relating to an incident in a news text from the named entity-sets available.

Given that better performing NE taggers may be available, evaluations of the EE may not indicate the best approach. Instead they should be considered to provide a lower boundary. However, relative statements between different heuristics are also possible, e.g., those suggested for the extraction.

6 Evaluation and evaluation results

This chapter describes the evaluation of the risk model and the text mining framework, which are both related to both the first and second research questions. It builds on the previous chapters, which proposed multiple elements for a child labor risk management system for supply chains. The major elements are the data sources, the text mining approach, and the risk model (and ranked suppliers). These are covered by the different evaluation steps, as shown in Figure 44.

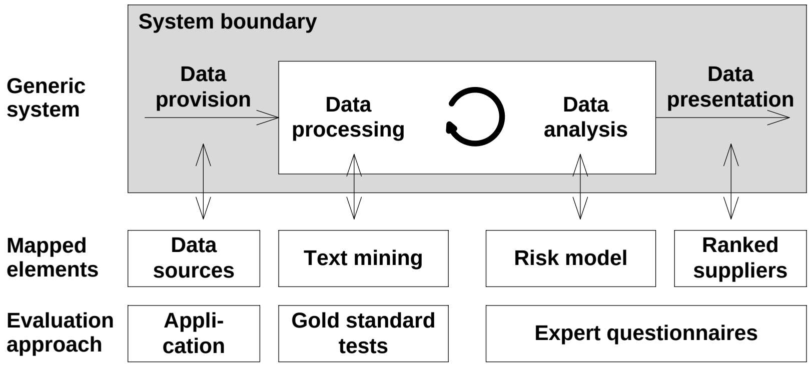


Figure 44: Evaluation approach (author's representation; generic system based on Gluchowski, Gabriel, and Chamoni 2005, 109)

The availability of data sources and their quality are discussed in the next chapter; this chapter particularly deals with TM and the risk model. For TM, the evaluation uses a gold standard. For the risk model, we base the evaluation on expert questionnaires. Each part contains a detailed discussion of the approach, the data (if necessary), and the results.

6.1 Text mining

The evaluation of the TM approaches needs to consider the two elements discussed in the methodology chapter. The evaluation must analyze the suggested approach for the domain-specific ED based on the chosen definition of a child labor incident. Separately, the EE approach needs to be tested for each dimension based on the input taggers used and the alignment and filtering approach. The sector dimension tagging has been specifically designed for this purpose. The following sections present the evaluation corpuses before discussing the evaluation approach and the results for ED and EE independently.

6.1.1 *Evaluation corpuses*

The evaluation of the TM approach is based on two gold standard corpuses, one in English and one in German. Given that the learned classifier is later applied to texts independent of the training set, the training set should reflect the later texts as well as possible. As longer texts (with multiple sentences), particularly news reports, will be used as input, two academically available news corpuses are employed: the Reuters TRC2 corpus[113] containing 1,800,370 news articles (Reuters and National Institute of Standards and Technology 2009) and the Austrian Press Agency (APA) corpus (APA – Austria Presse Agentur 2013).[114] The TRC2 corpus covers the period between 1.1.2008 and 28.2.2009. Therefore, the year 2008 has also been used for the APA corpus, for a total of 209,518 articles.

To establish a gold standard, both corpuses were manually tagged. This process will be outlined below. The main focus of the work was on the English corpus. While the results for the APA corpus are also provided, the broader analysis, along with additional steps to increase reliability, was only performed for the English corpus. Consequently, the results on the German corpus may be seen as indicative of the multilingual efficiency of the suggested methods. Both language corpuses will be described in more detail below. Given that English was used to test EE, it will only be described for the TRC2 corpus.

6.1.1.1 English

The basic assumption for the preparation of the gold standard is that articles on child labor will contain a word with the characters “child,” as has already been argued above.[115] Filtering all articles according to these criteria results in a significantly reduced corpus containing 16,948 articles. These articles were manually tagged based on whether they contained a child labor incident or not. This process resulted in 121 articles containing a child labor incident, based on the assumption that a single article can only contain one incident, but multiple observations (for a definition, see above).

In order to increase the validity of the English gold standard, an expert (a Sustainability Manager at a stock-traded international company) was asked to verify all positive articles from the

---

[113] http://trec.nist.gov/data/reuters/reuters.html; articles in the dataset are partly cropped after a certain length. While this limits demographical statements about the corpus, it should not affect the evaluation as also the gold standard preparation uses the cropped versions.

[114] http://www.apa.at/; provided per individual agreement.

[115] In the context of the gold standard preparation this limitation was also due to the large set of possible false positives of synonyms such as “kid” and the resulting need to manually classify an event larger dataset. Moreover, as argued above, the word “Kind” (child in German) has been the most useful in the German corpus discussed below.

first iteration of manual tagging. This helped to reduce the number of false positives in the gold standard. Nevertheless, it is limited by the fact that false negatives cannot be identified in this way. In the first round the expert excluded 14 articles from the set of child labor incident articles. After clarifying the definition and emphasizing that articles referencing child labor as a side topic also count, eight articles originally excluded were reincluded into the set of true positives (five of these articles are doubles and re-mention incidents). Two other articles, queried by the expert were re-included, because they featured parents pulling their children out of education in order to get jobs. Without these last two articles the agreement rate with the true positives was 95%. The final set of true positives consequently includes 117 articles. Figure 45 depicts this process. The articles' timestamps are presented in "Appendix P: Gold standard items including dimension tags".

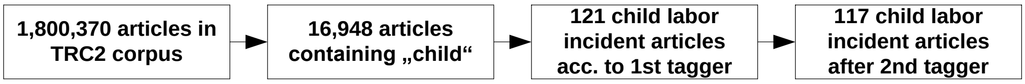


Figure 45: English gold standard creation process (author's representation)

In order to test the machine learning approach, the 117 articles were marked with the child labor incident event class "positive" and the other 16,831 articles in the sub-corpus were marked as "negative." These flags were then used by the further processing steps.

To be able to test the EE methodology, the 117 articles were also manually tagged with the following attributes for the child labor incident in the text:[116]

- Geographic dimensions (city, region, or country)
- Sectoral dimension
- Company dimension
- Tag describing whether the article is a "double"
- Tag describing whether the article's content is "positive"

All tags were included at the article level and, thus, added as class labels during implementation. The resulting features were checked by the expert along with the child labor categorizations of the articles. For the different attributes, the guidelines outlined in Table 39 were applied.[117]

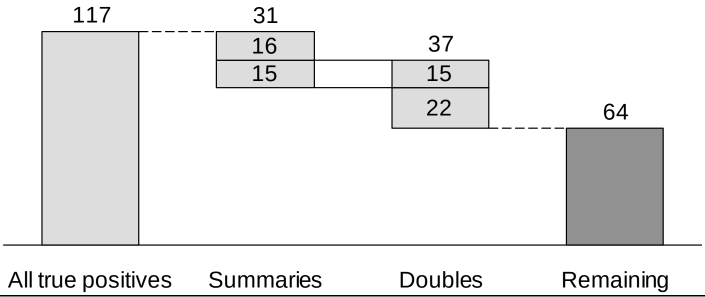
[116] For the definitions of "double" and "positive" see Table 39.
[117] Two slight adaptations were made after the expert reviewed the dimensions: in one article Guangdong province is added to the gold standard and in another one it is removed.

Attribute	Tag(s)	Guidelines/Values
Geography	city/region/ country	The geography tag marks the lowest level of geographic hierarchical detail (based on three levels: country/region/city) related to the child labor issue described in the article. If multiple geographic entities are named with the same hierarchical level and all associated with the child labor issue, then all entities are included. The name of the tag is chosen based on the hierarchical level. For the geographic dimension, the URIs of the individuals from the aligned ontology, OpenGeoNames, are taken.
Sector	sector	The sector tag marks the sector of the economic activity associated with child labor in the article. It is chosen based on the ISIC classification while taking the lowest matching level. For each mentioned activity in the article, only the most fitting ISIC code is chosen. The value added is the URI of the ontology created based on the ISIC specification described above.
Company	company	The company tag includes the stated name of a company mentioned as being linked to child labor in the article. The link can be positive or negative (i.e., the company may be aiming to reduce child labor) given that there is also a need for action in positive cases.
Double	double	The double tag marks articles that have already been published in a slightly changed version. Three values are applicable: - *True*: Only slight textual changes; same dimensional values for geography, sector, and company attributes. - *Partly*: Only slight textual changes, but values added or removed from the geographic, sector, or company attributes. - *False*: Significant textual changes (at least more than half the text changed; particularly related to child labor) with or without changes in values for dimensional attributes. Summaries count as significant textual changes.
Positive	positive	The positive tag marks articles which describe efforts to improve child labor in one of the dimensions (geographic location, sector, or company) mentioned in the article without also having a different, negative, focus on child labor (i.e., an incident according to the definition provided). For example, when children are rescued or a supplier is suspended this is not seen as an effort towards improving child labor but as an incident in contrast to when an anti-child labor program is described in general. A specific measure needs to be taken and not just asked for.

Table 39: Guidelines for manual tagging for dimensional attributes (author's representation)

Tagging at the article level limits the ability to evaluate the performance of individual named entity extractors. Consequently, no individual results for the suggested sector tagging approach at the entity level is given. Nevertheless, the main purpose of the IE approach suggested in the previous chapters is to be able to relate incidents and derived observations with supplier locations. It also provides a lower bound for the dimension extraction. In addition to the tags provided above, articles have also been marked with the tag "summary." This tag indicates whether an article is only a summary of other articles and not a focused report.[118] The summary format strongly lowers the available textual context that can be used for EE. Consequently, additional tests were performed on subsets of the 117 articles with and without doubles and summaries to test the EE approaches (see Table 41).

[118] This includes texts with the following titles in the TRC2 dataset: REUTERS ASIA FEEDBACK/OUTLOOK, Reuters South Asia News Highlights, DIARY, FACTBOX, PRESS DIGEST, *TOP NEWS*, BRIEF.



Note: Doubles exclude “partly” doubles – they are included in the remaining

Figure 46: Overview of texts per type in set of true positives (author’s representation)

6.1.1.2 German

The German gold standard is based on the APA Basisdienst Corpus 2008 and includes 209,518 German language news documents. In German, the basic assumption is that child labor related news contains the word “Kind” as a subpart of one of the article’s words or keywords (case insensitive). With this assumption, the possible candidate set can be drastically reduced to enable a manual classification. Altogether, 14,137 articles were manually tagged, finding 32 articles dealing with child labor.[119] Figure 47 outlines this process.[120]

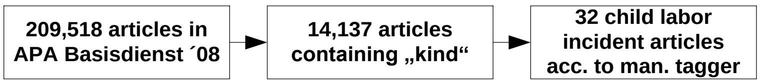


Figure 47: German gold standard creation process (author’s representation)

6.1.2 Evaluation approach

For activist events, Ploeger et al. state that the “[…] diversity in event representations and extraction methods makes it inappropriate to make direct comparisons (e.g. in terms of performance between our work and that of others […].” (Ploeger et al. 2013, 3). A similar situation prevails for this thesis. Nevertheless, we will provide comparisons with baseline approaches to be able to discuss the effects of the suggested methodologies.

6.1.2.1 Incident event detection

The evaluation metrics for ED are based on the GATE batch learning PR (Cunningham et al. 2013; Cunningham et al. 2002; Cunningham et al. 2011) and are detailed below. The GATE Java API was used for the batch-processing.

[119] Only proper news articles were included into the set of incidents and not special reports on “historical events on the same date” (i.e. “Kalenderblatt”), books, or other cultural reviews.

[120] The interpretation of a child labor incident was stricter for the German gold standard –three more neutral (rather than negative or positive) articles were excluded from the gold standard that would have been included in the English set. This may reduce the comparability with the English dataset.

We followed a “breadth-first” approach: First, we analyzed different parameter constellations for different ML techniques for the two sets before analyzing the effect of other variations on the best performing ML algorithm and parameter set only based on the lemmas. Variations included varying the cut-off distance before generating the ML features, for example. Thus, this procedure should provide a reasonable trade-off between computational complexity and usefulness of the result. The aim of the ML evaluation was to test the applicability of machine learning to the problem. Understanding the result of the analysis as a lower bound, it may be sufficient for the use case here.

The evaluation results for ED are calculated using *k*-fold cross validation (Refaeilzadeh, Tang, and Liu 2009). Different parameter constellations for the ML approach are tested using *k*-fold evaluation for each constellation (i.e., several splits of the data into training and testing data). The results for the best parameter constellation will be shown. In this way, the parameters are slightly adapted to the training corpus. Still, evaluating parameters for ML configurations based on *k*-fold cross-validation is considered a standard technique (Refaeilzadeh, Tang, and Liu 2009). An additional split into a second set of training and testing data does not seem possible, based on the limited amount of true positives in the Reuters dataset. Nevertheless, in the next chapter a second validation will be performed by applying the classifier to the data gathered from the internet.

6.1.2.1.1 Baseline

As well as comparing different variations of the suggested methodologies with each other, results for primary steps and intermediate results are also included. The following baseline approaches are suggested for comparing candidate set reduction and classification steps. For testing, the most successful ML configuration can be used:

- **Random:** A mathematical baseline is calculated by assuming a tagger that randomly (i.e., with a probability of 50%) annotates a news text as containing a child labor incident or not.
- **Candidate set reduction only:** The result obtained through candidate set reduction only.
- **Lightweight rule-based approach**: Given that rule-based approaches have been successfully implemented in the past (see related work section) and that child labor can be detected when only using the words “child” and “labor”, a keyword-matching technique has been included in the baseline. If a text contains the patterns “child labor” or “child labour” directly, then it has been classified as a child labor sentence.

- **Direct classification with all tokens**: One very simple baseline approach for the classification of the news reports is be a direct classification of the news reports. All tokens from a text could be used for classification. More complicated approaches should yield improved results.
- **Direct classification with preparation**: Another simple approach is to restrict the processed set to documents containing a child labor relation without implementing a cut-off distance. This results in a child labor incident ED pipeline without candidate set reduction and differentiated by the word set used (small or large).
- **Best approach without feature selection**: Finally, instead of selecting features for ML, all available tokens can be used, much like the direct classification tests but for the best cut-off distance.

6.1.2.1.2 Metrics

Because the evaluation corpuses are closed datasets, we used the classic evaluation metrics precision and recall (Clough and Sanderson 2013). As with skewed datasets, overall precision and recall measures have little comparative value (M. Naughton, Stokes, and Carthy 2010), so class-specific results for true incident events were reported. These can be calculated using the following formula (M. Naughton, Stokes, and Carthy 2010, 141–142; van Rijsbergen 1975, 115):

c … Number of texts correctly classified as containing a child labor incident event by the system
n … Total number of texts classified as containing a child labor incident event by the system
m … Number of texts classified as containing a child labor incident event in the gold standard

$$Precision = \frac{c}{n};\ Recall = \frac{c}{m};\ F1 = \frac{2 * Precision * Recall}{Precision + Recall}$$

Candidate set reduction leads to a trade-off between precision and recall, depending on the cut-off distance selected. In general, the larger the distance, the higher the recall but the lower the precision. As a result, the evaluation will discuss this trade-off in more detail. Additionally, this requires including the number of documents that result as input for the classification. Precision, recall, and F1 measures can be directly calculated.

As stated, a k-fold cross validation technique is used. Cross validation splits the set of tagged instances into a training set and a test set in multiple folds. The number of folds n defines the number of documents included in each set. Per the definition, $1/n$ documents are included in the test and the rest in the training set. The training and testing operation is performed n times and the results are averaged over the runs. Here, 8-fold cross validation was used as the standard technique. Moreover, the effects of changes to this technique were tested for the best performing algorithm configuration.

6.1.2.2 Incident event extraction

Several variables are used for the further argumentation. The geographic dimension consists of the sub-dimensions country, region, and city and is combined for the calculation of precision and recall.[121]

d … Dimensions as specified above $d \in \{geography, sector, company\}$
t … Total number of articles in evaluation set
$a, a \in \mathbb{N}, a < t$… Article index
$m_{a,d}$ … Number of elements in gold standard for article a and dimension d
$n_{a,d}$… Number of elements retrieved for article a and dimension d
$c_{a,d}$ … Number of elements correctly retrieved for article a and dimension d
$c_{a,d} = |\{elements\ in\ gold\ standard\} \cap \{elements\ retrieved\}|$
$P_{a,d}$ … Precision for article a and dimension d
$R_{a,d}$ … Recall for article a and dimension d

The evaluation of the EE relies on the gold standard tagging each document with the dimensions of the incident event depicted. For each dimension, the evaluation measure provides a number for the combined performance of the different process steps.

The dimensions reflect different characteristics. Each attribute can have one or multiple values that can also be in a hierarchical relationship with other values. This has two consequences: Each dimension may require a different type of measure and adjusted measures are needed for the hierarchical dimensions, as classical precision and recall cannot be applied. Consequently, we differentiate between direct and hierarchical metrics, which will be detailed in the following. The "classical" metrics for precision and recall described above are used as direct metrics.[122] Because this would only work for direct matches, we suggest also integrating a

[121] For Open Geonames, the LOD used during processing is based on Yago version 2.5.3 with a Wikipedia dump from 2012-12-01 and accessed via sparql endpoint at http://lod2.openlinksw.com/sparql.
[122] Adjusted to the syntax used in this paragraph, these are $P\_(a,d) = c\_(a,d)/n\_(a,d)$; $R\_(a,d) = c\_(a,d)/m\_(a,d)$

lexical similarity measure, which will be detailed in the next subsection.[123] Another metric is suggested for the hierarchical geographic and sector dimensions, given that hierarchical relationships are more complex than can be captured with a direct match.

6.1.2.2.1 Lexical similarity

The "lexical similarity measure" (Maedche and Staab 2002, 254) for lexical comparison uses the Levenshtein edit distance (Levenshtein 1966) to calculate a degree of similarity between strings in the interval $[0,1]$. This is used for the company dimension, as the gold standard only provides company names and not URIs, and the tagger also returns a string with the company name. Therefore, this lexical approach has been used in the program. Consequently, a combination of the lexical precision and recall measure with the lexical similarity measure is used to allow for slight textual differences in the spelling of company names. This can be reflected in a threshold value in the program. Nevertheless, the prototype assumes the availability of URIs or identical names for companies for identification. This may also lead to a discrepancy between the results obtained here and those obtained if an URI-based gold standard were available. The lexical similarity measure is defined as follows (Maedche and Staab 2002, 254):

$ed(x, y)\dots$ Levenshtein edit distance
$L\dots$ Lexical string

$$SM(L_i, L_j) \coloneqq max\left(0, \frac{min(|L_i|, |L_j|) - ed(L_i, L_j)}{min(|L_i|, |L_j|)}\right) \in [0,1]$$

To calculate a lexical precision LP and lexical recall LR, the elements in the gold standard and retrieved set can be understood as concepts in an ontology:

$O_G, O_R\dots$ Ontologies ($G\cdots$Gold standard, $R\dots$Retrieved)
$C_G, C_R\dots$ Concepts in the ontology ($G\cdots$Gold standard, $R\dots$Retrieved)

$$LP(O_R, O_{GS}) = \frac{|C_R \cap C_{GS}|}{|C_R|};\ LR(O_R, O_{GS}) = \frac{|C_R \cap C_{GS}|}{|C_{GS}|}$$

Two strings intersect if the SM value for both strings exceeds a predefined threshold which can be set by the user. Here, a threshold of 0.9 is applied for comparing company names.

6.1.2.2.2 Hierarchical metrics

The values for precision and recall described above only hold for a direct comparison of values for a single dimension. However, the geographic and sector dimensions have hierarchical relationships that should be taken into account when calculating precision and recall values

[123] Term-matching is a form of a lexical precision and recall that has been explained for ontologies in the past, for example, by Dellschaft and Staab (2006, 232) going back to Sabou et al. (2005, 194) and others (Maedche and Staab 2002; Cimiano, Staab, and Tane 2003).

(see e.g. Dellschaft and Staab 2006). Figure 48 summarizes the hierarchical structures that need to be respected. Either the gold standard or the tags from the tagger can be associated with one of the hierarchical layers. Given that the ISIC codes allow implicit extraction of the hierarchical structure of a tag, no differentiation of the tagging feature name has been made. In contrast, country, region, and city layers are differentiated by different sub-dimension names in the gold standard and the tagger's response.

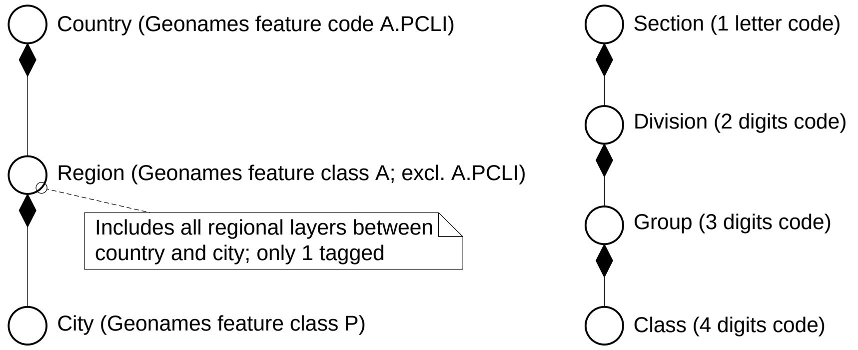


Figure 48: Hierarchical structures in dimensions (geographic according to Geonames; sector according to ISIC; author's representation)

For the gold standard, geography and sector were tagged with the lowest level of detail available that relates to a child labor incident. Similarly, the software used for EE outputs the lowest hierarchical level identified. For example, if the tagger identifies a city and two regions, only the city is kept in the result set. As a result, three cases can be differentiated:

- **Case (1):** The same level of geographic detail is available for the gold standard and the tagging output. In this case the formulas for the calculation of precision and recall could be directly applied if hierarchical depth is ignored.
- **Case (2):** The level of geographic detail provided in the gold standard is more granular than the information returned by the tagger. This occurs when the gold standard specifies a city whereas the tagger only returns information on a regional or sector level. If the entities specified in these two types of information can be semantically connected, for example if the city is truly in the region specified, then the precision and recall should be adjusted to account for this.
- **Case (3):** The information gained from the tagger is more granular than the information included in the gold standard. For example, this can happen when the gold standard depicts one ISIC division while the tagger returns an underlying group.

Cases (2) and (3) require an extension to the formula for lexical precision and recall. The problem can be redefined to one which already exists in the literature by rephrasing the conceptual model. The gold standard tag and the tagger's response can both be understood as implicit hierarchical ontologies that need to be compared. This is outlined in Figure 49. Consequently, an adapted approach for recall and precision needs to return reasonable measures for the comparison of two hierarchical ontologies. By comparing the full hierarchy, differences in the number of responses by hierarchical level can automatically be included in the calculus.

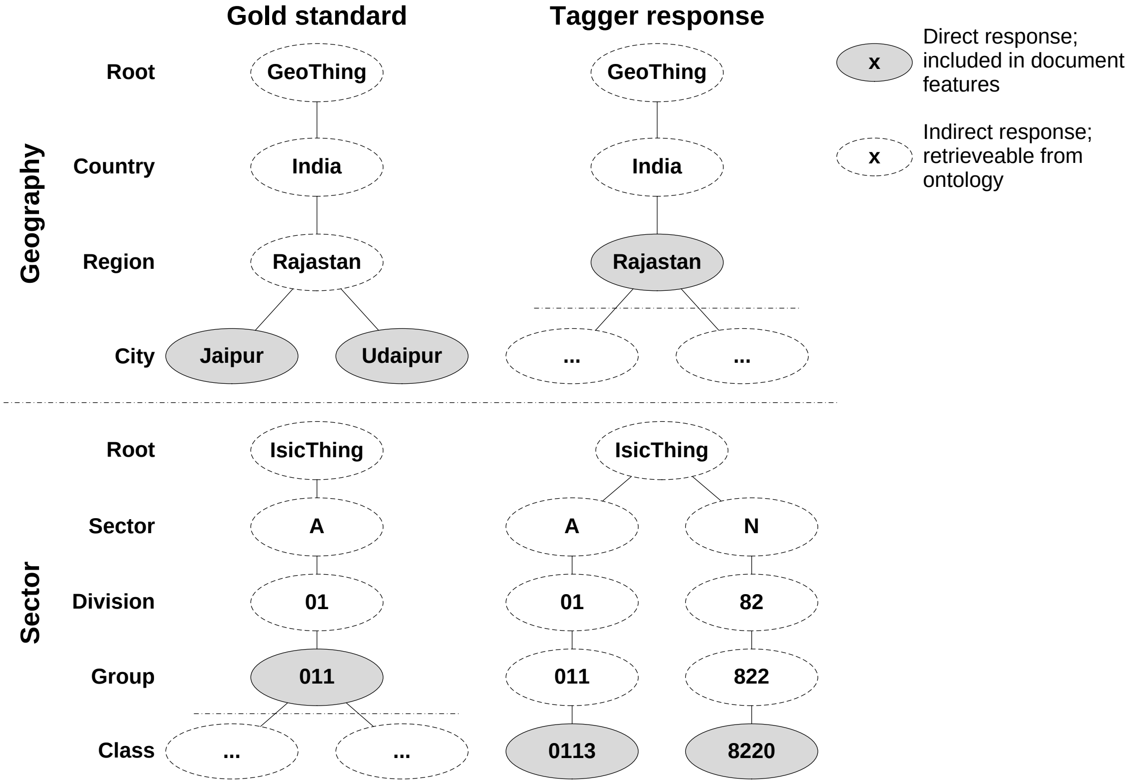


Figure 49: Examples of hierarchical ontologies included in gold standard and tagger response (author's representation)

Multiple authors have already discussed the problem of comparing hierarchies or ontologies. Early approaches that build on including hierarchical information into the precision and recall formula (i.e., granularity in this case) were presented by Langlais, Simard, and Véronis (1998). Later, semantic distance was introduced by Maynard and Ananiadou (1999; 2000) in order to calculate a degree of similarity based on nodes in common (i.e., commonality com) and the position in the hierarchy (i.e., position pos) (D. Maynard and Ananiadou 2000, 532):

$$sim(\omega_1 \dots \omega_n) = \frac{com(\omega_1 \dots \omega_n)}{pos(\omega_1 \dots \omega_n)}$$

Still later, Maynard and several others (D. Maynard, Peters, and Li 2008; D. Maynard, Peters, and Li 2006) provided further adapted metrics, presented below. Several recent approaches have been proposed. Dellschaft and Staab (2006) summarize several of these (including the recent one from Maynard, Peters, and Li) in an overview of measures for concept hierarchies, presented in Table 40. The table includes three quality criteria. Multi-dimensionality allows for a user-defined weighting of different kinds of error. The second variable ensures that the rating reflects the distance between the gold standard and the retrieved result. Finally, "using the interval" indicates whether the amplitude of changes in the retrieved ontology (i.e., errors) are reflected in the usage of the value over the predefined possible interval used by the measure.

	Multi dimensionality	Proportional error effect	Usage of interval
Taxonomic overlap	-	+	?
Learning accuracy	-	O	?
Augmented precision and recall	O	+	?
OntoRand	-	+ / -	+ / -

Note: - … low, O … medium, + … good fulfillment of quality criteria. ? … unknown/unclear;
Better values are for the "common ancestor" approach as indicated in the source (Brank, Mladenić, and Grobelnik 2006).

Table 40: Rating of concept hierarchy measures (based on Dellschaft and Staab 2006, 230)

The taxonomic overlap presented in Table 40 is a measure to allow the comparison of two different hierarchies based on superconcepts and subconcepts (Maedche and Staab 2002; Cimiano, Staab, and Tane 2003). Subsequently, an augmented precision and recall (APR) measure was proposed that also includes the balanced distance metric (BDM) (D. Maynard, Peters, and Li 2006; D. Maynard, Peters, and Li 2008). APR incorporates the older learning accuracy (LA) technique, also mentioned in the table (Hahn and Schnattinger 1998), and additionally respects hierarchical density (D. Maynard, Peters, and Li 2006). While it builds on the classic precision and recall it was primarily designed for matching processes working with one ontology and not for comparing two. Brank, Mladenić, and Grobelnik (2006) propose another approach they call OntoRand, building on the rand index. For the similarity calculation they suggest two options: one is distance-based and comparable to LA or BDM (see also Dellschaft and Staab 2006), and uses common ancestors for the calculation (Brank, Mladenić, and Grobelnik 2006). The latter approach fulfills two of the three quality criteria in Table 40 (Dellschaft and Staab 2006, 230–231). Finally, Dellschaft and Staab (2006) suggest using taxonomic precision and recall (TPR), which fulfills all three quality criteria while extending classical precision and recall. In one version this metric fulfills all the criteria, however, Dellschaft and Staab (2006) also evaluate a second version of the metric that does not satisfy the multidimensionality criterion and where the interval usage is strongly influenced by lexical properties. Nevertheless, this metric is still of interest for the use case discussed here, as argued below.[124]

[124] In addition to the approaches in Table 40 and the one from Dellschaft and Staab, Sun and Lim (2001) suggest a measure for hierarchical text classification evaluation. However, it is not an extension of classical precision

When comparing the metrics described above APR and TPR directly extend classical precision and recall metrics. Building on classical precision and recall enhances understandability (D. Maynard, Peters, and Li 2006). Given that TPR was specifically designed to compare two hierarchies, this approach will be adopted. In the special case discussed in this thesis, the nodes in the compared hierarchies are always related to the correct super-elements and sub-elements. Therefore, because the better performing approach from Dellschaft and Staab uses the intersection of hierarchies for the calculation, relevant hierarchy-depth related information is not used in the calculation. Table 41 compares the values of the main metrics suggested by Dellschaft and Staab (2006) (taxonomic precision, TP, and recall, TR) in their use case based on the notion of semantic cotopy (i.e., better performing common semantic cotopy, csc, as well as lower performing semantic cotopy, sc, together with the lexical precision, LP, and recall, LR). The values are based on the example hierarchies adapted from Figure 49 and shown in Figure 50. Common semantic cotopy and semantic cotopy both define sets of nodes that can be used to calculate precision and recall metrics. "Common" indicates a focus on the intersection of two sets.[125]

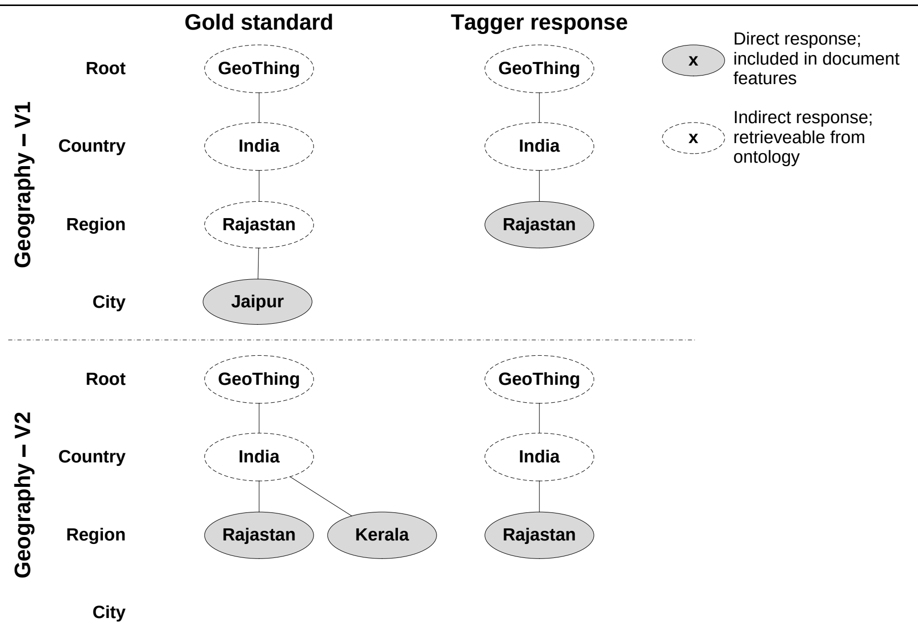


Figure 50: Example graphs for measure comparison based on Figure 49 (author's representation)

and recall, as argued by Ehrig and Euzenat (2005). Therefore, Ehrig and Euzenat themselves (2005) propose an approach for ontology matching which focuses on developing updated measures for precision and recall for relational matches between ontologies. In a later work, Euzenat (2007) builds on this idea and includes semantics. Given that the ideas aim to evaluate the matching of two different ontologies, they will not be further explored here. Li et al. (2008) suggest different weights for calculating a degree of similarity between properties in hierarchies based on earlier work by several authors (Paolucci et al. 2002; Caceres et al. 2006). They explicitly differentiate between sub- and super-class relationships, but do not take the exact position in the hierarchy into account and do not extend classic precision and recall.

[125] Given three ontologies O, O_1, O_2, semantic cotopy is defined as $sc(c, O) := \{c_i \mid c_i \in C \wedge (c_i \leq c \vee c \leq c_i)\}$ and common semantic cotopy as $csc(c, O_1, O_2) := \{c_i \mid c_i \in C_1 \cap C_2 \wedge (c_i <_1 c \vee c <_1 c_i)\}$ (Dellschaft and Staab 2006).

Dimension	LP	LR	TP_{sc}	TR_{sc}	TP_{csc}	TR_{csc}
Geo V1	100%	75%	100%	56,3%	100%	100%
Geo V2	100%	75%	100%	62,5%	100%	100%

Table 41: Comparison of metrics suggested by Dellschaft and Staab (2006) based on examples in Figure 50 (author's representation)

The csc metric cannot differentiate the hierarchical structures. Moreover, lexical precision and recall do not differentiate between version 1 (V1) and version 2 (V2). Only the taxonomic precision and recall based on sc return different values for the two versions. The result that the value for version 1 is lower than the value for version 2 is in line with the expected behavior – the retrieval results for version 1 are punished for not including a greater level of detail compared to version 2, where only less granular information is lost. Given that more dimensional detail allows for better relation between a piece of information and a specific supplier, in a risk management context the increased level of detail is more helpful than a higher degree of more generic information. If a city instead of a region is missed while detecting the other hierarchical level, the remaining greater level of detail is honored with a higher recall. Consequently, although it is influenced by a higher sensitivity to lexical differences (Dellschaft and Staab 2006),[126] APR with semantic cotopy will be used due to its respect for hierarchical depth. Given their close relation, it needs to be used independently from other lexical performance metrics (Dellschaft and Staab 2006). The approach as described in Dellschaft and Staab (2006) is summarized in detail in "Appendix Q: Augmented precision and recall".[127]

6.1.2.2.3 Overall calculation

LP and APR metrics are calculated for each dimension and document – for each article a and over all articles t. Consequently, in order to be able to compare the overall results for a given dimension. They follow a macro-averaging approach, calculating each dimension for each document independently as each dimension should have the same weight in the overall performance (Manning, Raghavan, and Schütze 2008, 280–282):

$$P_a = \frac{\sum_d c_{a,d}}{\sum_d n_{a,d}};\ R_a = \frac{\sum_d c_{a,d}}{\sum_d m_{a,d}}$$

$$P = \frac{\sum_a \sum_d c_{a,d}}{\sum_a \sum_d n_{a,d}};\ R = \frac{\sum_a \sum_d c_{a,d}}{\sum_a \sum_d m_{a,d}}$$

---

[126] Losing the benefits of "multi-dimensionality" is not particularly relevant in the case discussed here, as only one final quality measure is needed.

[127] One slight limitation remains: Due to the inclusion of a common root element, a certain positive precision and recall is given even if the two hierarchies only overlap at the root. Thus, they nothing else (content-wise) in common. Given that this element is always a common feature, all further positive matches will result in an increase in the overall scores, and therefore the aim of having a measurement that increases if a better matching is created is still fulfilled. Nevertheless, measures have also been calculated excluding this "root effect."

One final decision needs to be taken: how to handle empty fields in the gold standard. Not every article will have a value for every dimension of an event and, hence, the gold standard might not provide a value for the calculation of precision and recall value. Two approaches are differentiated during the evaluation to be able to uncover this effect. On the one hand, recall can be evaluated independent of the result and precision if the result of the tagger is also empty. This is shown in Table 42. On the other hand, precision and recall metrics can be ignored for the overall average if no gold standard is available (we refer to this case as "gold standard restriction"). While this approach may more strongly reflect the performance of the NE taggers, it does not honor the correct blanks in the result set.

Gold standard	Retrieved result	Precision	Recall
n/a	n/a	1	1
n/a	available	0	1
available	n/a	0	0
available	available	calculated	calculated

Table 42: Recall and precision values depending on the availability of gold standard and retrieval results (author's representation)

Another suggested comparison is based on the hierarchical relationships in the gold standard and result set. While detailed and granular results are important, being able to get the most general classification (i.e., one level below the root) correct may sometimes be sufficient. Additionally, news reports often only state a general sector and do not provide more detailed references. Consequently, we will provide measures which only consider the top level of the hierarchies (excluding the introduced root) to calculate the metrics.

6.1.2.2.4 Parameters

The EE methodology is based on several processing steps. Different inputs, approaches, and parameters can be used to generate child labor incident events and observations. Consequently, different constellations were tested in order to be able to evaluate their effects.

Area	Parameter	Values tested
Calculation	Root	Include; not include
Input	Spotlight type	Lucerne-based; statistical
	Use specification disambiguation for sectors	True; false
	Allow part matches for sectors	True; false
Filtering	Type of filter	Child labor relation based (closest – with and without limit; sentence-based; distance-based); Child and labor words based (closest – with and without limit; sentence-based; distance-based); Highest count
	Maximum distance*	25 characters; 50 characters; 100 characters
	Limit for closest**	Only one used: 50 characters
	Use child labor minimum relation only***	True; false
	Filter to lowest geography & filter to lowest sector****	True; false

* Only for distance-based filtering approaches; ** Only for filtering using closest with limit; *** Only applicable for relation-based techniques; **** Only in combination

Table 43: Parameters for incident event extraction (author's representation)

Table 43 outlines the different parameters in three categories. For the calculation of the hierarchical dimensions, the root has been excluded for all cases. Nevertheless, we will provide the effect of including the root in the calculation in order to show the sensitivity. Because two types of DBpedia Spotlight are implemented, we differentiate between the two. Four options are offered for the sector, either using an additional disambiguation procedure and/or allowing word part matches. Finally, multiple parameters allow the EE filtering to be customized. Primarily the type of filter needs to be defined and is expected to have the biggest effect on the result. Depending on the type of filter, specific parameters may be required. Distance-based filtering techniques require a maximum distance to be set, and the limit for the "closest with limit" approach and relation-based techniques can be restricted to only consider child labor minimum relations. Additional heuristics allow filtering to the lowest geography or sector hierarchy, removing more upper-level constructs. Altogether, 608 possible combinations were tested.

6.1.3 Evaluation results

The following sections present the TM results for candidate set reduction, ML, and incident event extraction. Each section provides different elasticity discussions for the parameters involved.

6.1.3.1 Candidate set reduction

The effects of candidate set reduction are detailed separately for English and German. The effectiveness of candidate set reduction in general is discussed as part of the comparison with baseline techniques in section 6.1.3.2.2 below. Two important differentiations are reflected in each discussion. The decision to use stopword removal or not is significant, particularly for German, and the size of the synonyms set strongly affects the results. Consequently, the results distinguish these effects.[128]

6.1.3.1.1 English

Table 44 shows the results of applying the candidate set reduction methodology to the English corpus. The following elements are shown: the cut-off distances (i.e., the maximum numbers of characters between a word indicating "child" and one indicating "labor" within a child labor distance relation), the corresponding results for the number of documents that contain a child labor distance relation with a distance less than or equal to the cut-off, and the precision and recall metrics for the total number of positive documents included in the result set. No stopwords were found for the small set of synonyms.

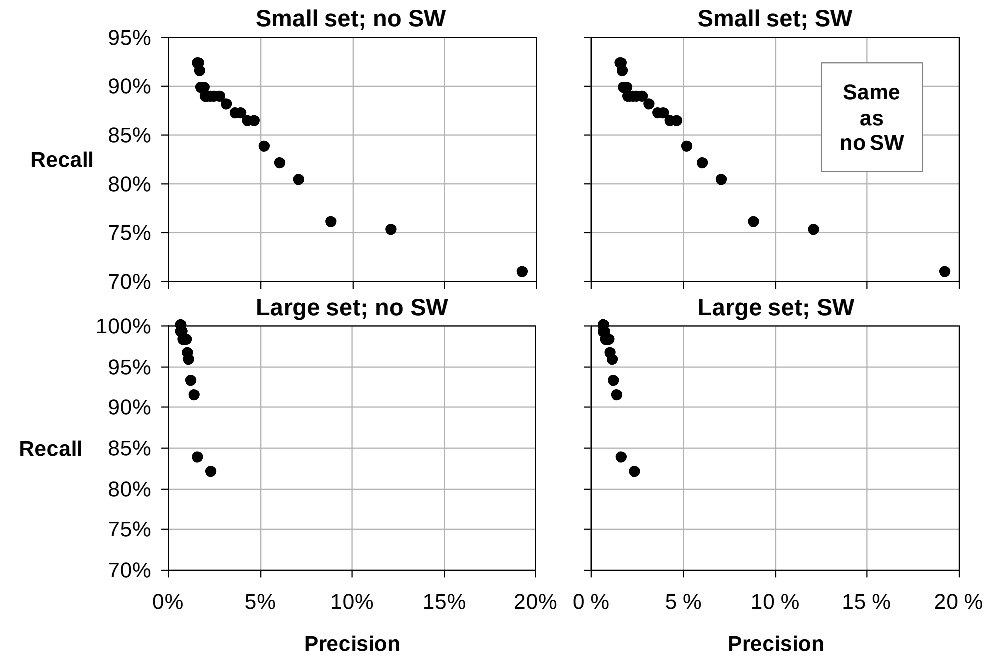
Cut-off distance	Small set; SW			Small set; no SW			Large set; SW			Large set; no SW		
	# of docs	Precision	Recall	# of docs	Precision	Recall	# of docs	Precision	Recall	# of docs	Precision	Recall
10	430	19.3%	70.9%	Same as small set with stopwords (SW) as no stopwords found			3933	2.4%	82.1%	4101	2.3%	82.1%
20	723	12.2%	75.2%				5839	1.7%	83.8%	5978	1.6%	83.8%
30	1003	8.9%	76.1%				**7293**	**1.5%**	**91.5%**	7411	1.4%	91.5%
40	**1317**	**7.1%**	**80.3%**				8602	1.3%	93.2%	8712	1.3%	93.2%
50	1570	6.1%	82.1%				9481	1.2%	95.7%	9587	1.2%	95.7%
60	1865	5.3%	83.8%				10481	1.1%	96.6%	10581	1.1%	96.6%
70	2135	4.7%	86.3%				11187	1.0%	98.3%	11279	1.0%	98.3%
80	2335	4.3%	86.3%				11808	1.0%	98.3%	11891	1.0%	98.3%
90	2557	4.0%	87.2%				12334	0.9%	98.3%	12405	0.9%	98.3%
100	2760	3.7%	87.2%				12775	0.9%	98.3%	12842	0.9%	98.3%
125	3218	3.2%	88.0%				13751	0.8%	98.3%	13810	0.8%	98.3%
150	3656	2.8%	88.9%				14453	0.8%	99.1%	14508	0.8%	99.1%
175	4080	2.5%	88.9%				14880	0.8%	99.1%	14930	0.8%	99.1%
200	4438	2.3%	88.9%				15239	0.8%	99.1%	15287	0.8%	99.1%
225	4737	2.2%	88.9%				15522	0.7%	99.1%	15566	0.7%	99.1%
250	5048	2.1%	88.9%				15763	0.7%	99.1%	15799	0.7%	99.1%
275	5330	2.0%	89.7%				15961	0.7%	99.1%	15992	0.7%	99.1%
300	5595	1.9%	89.7%				16098	0.7%	100.0%	16126	0.7%	100.0%
325	5878	1.8%	89.7%				16205	0.7%	100.0%	16232	0.7%	100.0%
350	6119	1.7%	91.5%				16304	0.7%	100.0%	16331	0.7%	100.0%
375	6383	1.7%	92.3%				16385	0.7%	100.0%	16411	0.7%	100.0%
400	6591	1.6%	92.3%				16461	0.7%	100.0%	16486	0.7%	100.0%

Table 44: Results for candidate set reduction depending on cut-off distance and approach; English corpus (author's representation; emphasis added)

The trade-off between precision and recall is clearly visible in Figure 51. As the classification follows the CSRD, the procedure aims to minimize the number of documents included while maximizing recall. Therefore, no approach appears clearly preferable. While a larger set of labor tagging words including hyponyms and troponyms yields higher recall rates, the number of documents included is significantly lower when only using the synonyms. The maximum

[128] The scores necessary for calculation different measures of the processing steps have been derived by using Mímir (Valentin Tablan and Roberts 2013), an indexing service for GATE documents that allows searching for annotations and features.

recall for the small set is 94.9%, which is lower than the recall for the large set with stopwords at a cut-off of 50. This is also shown in Figure 52.



Note: SW = Stopwords

Figure 51: Precision and recall trade-offs for all approaches (author's representation)

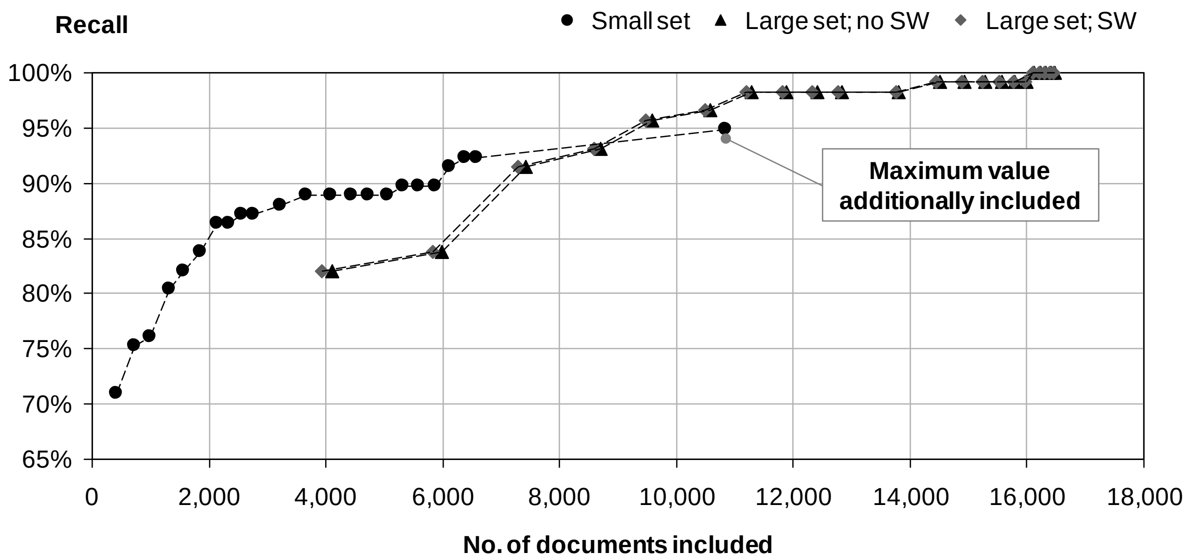


Figure 52: Number of documents and recall trade-off for all approaches (author's representation)

Therefore, the further ML evaluation uses two input sets of different sizes to test the difference between having fewer documents and higher recall as input for the ML. To balance the cut-off distance with recall and number of documents, a minimum of 80% recall was chosen for the smaller set and a minimum of 90% recall for the larger set. As can be seen in Table 44, 90% recall is also possible for the small set of synonyms; however, this requires a very large cut-off distance of between 325 and 350. Consequently, the risk of random connections between a word referring to “child” and one referring to “labor” is increased, which would not be supported by the content.

The candidate set reduction allows a significant reduction in the number of documents (at the respective cut-off points): 92% for the small set and 56/57% for the large set. Moreover, the share of true positive news reports (i.e., those tagged with child labor incident events) increases from 0.7% for the overall corpus based on filtering with “child” to 7.1% for the small set and 1.4/1.5% for the large set. As a consequence, the performance of the ML classification should be positively affected.

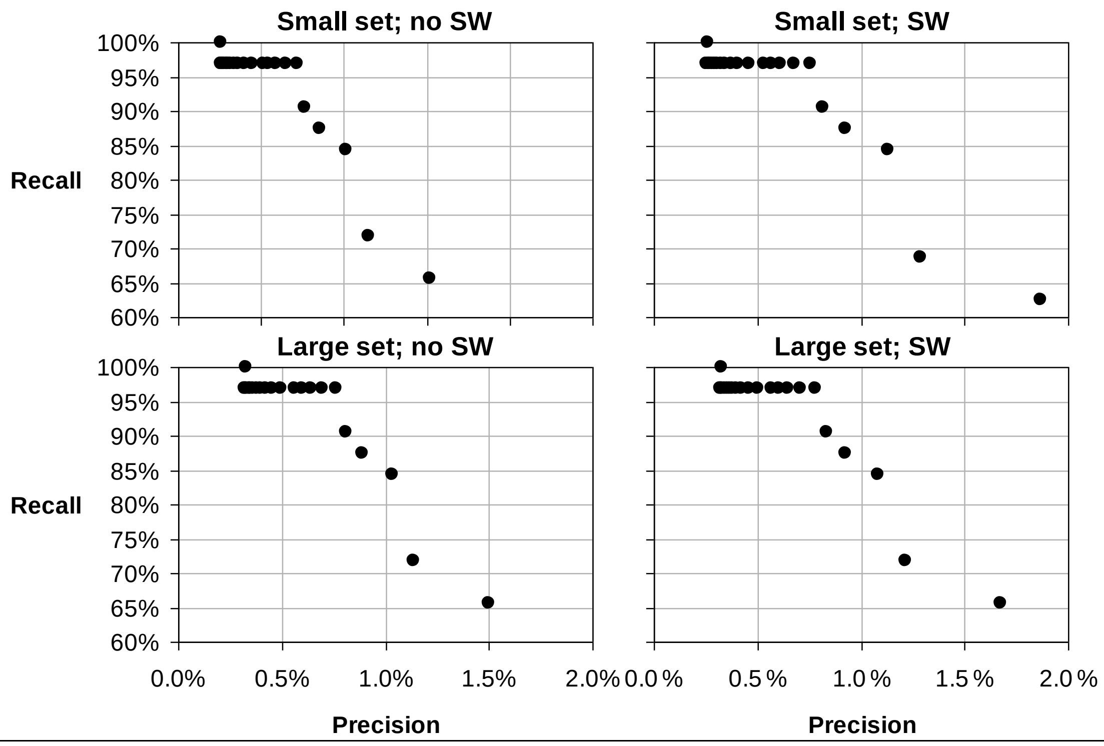
6.1.3.1.2 German

The German approach was tested in a similar manner as the English one. The results are depicted as above and structural differences are highlighted. General results for different cut-off distances are presented in Table 45.

	Small set; SW			Small set; no SW			Large set; SW			Large set; no SW		
Cut-off distance	# of docs	Precision	Recall	# of docs	Precision	Recall	# of docs	Precision	Recall	# of docs	Precision	Recall
10	535	3.7%	62.5%	693	3.0%	65.6%	1255	1.7%	65.6%	1399	1.5%	65.6%
20	854	2.6%	68.8%	1002	2.3%	71.9%	1894	1.2%	71.9%	2024	1.1%	71.9%
30	1195	2.3%	84.4%	1331	2.0%	84.4%	2502	1.1%	84.4%	2619	1.0%	84.4%
40	1514	1.8%	87.5%	1641	1.7%	87.5%	3035	0.9%	87.5%	3146	0.9%	87.5%
50	1780	1.6%	90.6%	1901	1.5%	90.6%	3488	0.8%	90.6%	3594	0.8%	90.6%
60	**2052**	**1.5%**	**96.9%**	2170	1.4%	96.9%	3966	0.8%	96.9%	4066	0.8%	96.9%
70	2289	1.4%	96.9%	2397	1.3%	96.9%	4384	0.7%	96.9%	4472	0.7%	96.9%
80	2539	1.2%	96.9%	2640	1.2%	96.9%	4779	0.6%	96.9%	4861	0.6%	96.9%
90	2741	1.1%	96.9%	2838	1.1%	96.9%	5123	0.6%	96.9%	5204	0.6%	96.9%
100	2921	1.1%	96.9%	3014	1.0%	96.9%	5430	0.6%	96.9%	5509	0.6%	96.9%
125	3378	0.9%	96.9%	3466	0.9%	96.9%	6147	0.5%	96.9%	6220	0.5%	96.9%
150	3826	0.8%	96.9%	3905	0.8%	96.9%	6791	0.5%	96.9%	6854	0.5%	96.9%
175	4165	0.7%	96.9%	4237	0.7%	96.9%	7284	0.4%	96.9%	7344	0.4%	96.9%
200	4505	0.7%	96.9%	4575	0.7%	96.9%	7733	0.4%	96.9%	7792	0.4%	96.9%
225	4793	0.6%	96.9%	4855	0.6%	96.9%	8123	0.4%	96.9%	8178	0.4%	96.9%
250	5028	0.6%	96.9%	5087	0.6%	96.9%	8465	0.4%	96.9%	8516	0.4%	96.9%
275	5254	0.6%	96.9%	5309	0.6%	96.9%	8749	0.4%	96.9%	8799	0.4%	96.9%
300	5442	0.6%	96.9%	5494	0.6%	96.9%	8994	0.3%	96.9%	9039	0.3%	96.9%
325	5662	0.5%	96.9%	5709	0.5%	96.9%	9247	0.3%	96.9%	9290	0.3%	96.9%
350	5860	0.5%	96.9%	5906	0.5%	96.9%	9468	0.3%	96.9%	9510	0.3%	96.9%
375	6043	0.5%	96.9%	6088	0.5%	96.9%	9645	0.3%	96.9%	9685	0.3%	96.9%
400	6205	0.5%	100.0%	6250	0.5%	100%	9815	0.3%	100.0%	9853	0.3%	100.0%

Table 45: Results for candidate set reduction depending on cut-off distance and approach; German corpus (own representation; selected approach in bold)

In the German corpus, a significantly lower number of documents are reachable without diminishing precision. This may be due to the overall low number of true positives (32 documents). The same maximum recall is achieved for both the small and large sets, with a lower number of documents for the small sets, as can be seen in Figure 54. Consequently, the evaluation of the ML will use a cut-off distance of 60, reaching a recall of more than 95%.



Note: SW = Stopwords

Figure 53: Precision and recall trade-offs for all approaches (author's representation)

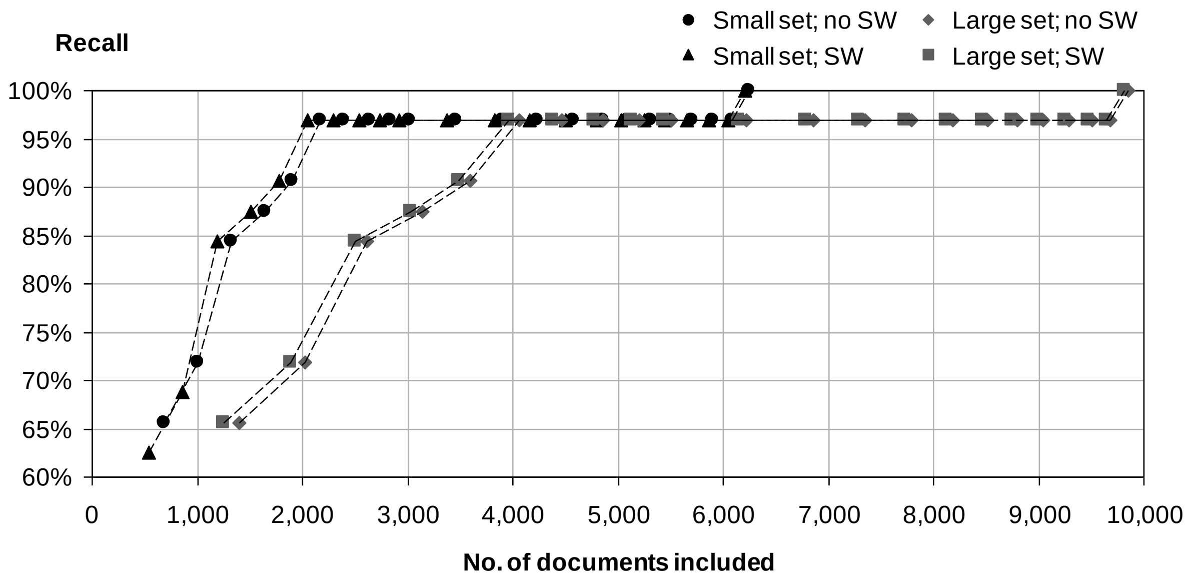


Figure 54: Number of documents and recall trade-off for all approaches (author's representation)

As can be seen in Figure 54, a 100% recall would be possible in the German corpus. However, significantly fewer documents can be included when only requiring 95% recall. And a cut-off at 60 allows the number of documents to be reduced by 85%, which is significantly more than in the English dataset. The ratio of positive documents increases from 0.2% to 1.5%. Consequently, the candidate set reduction procedures appear not only viable for English, they seem to be also effective for German. This may be due to the lower number of words used in German to describe child labor.

6.1.3.2 Machine learning

The ML techniques classify the news reports into those containing a child labor incident event and those that do not. A preprocessing step is used to create the features, which are then used by the ML. Different feature creation approaches have been proposed beyond using the distance relations introduced earlier. Each feature contains a primary lemma. The approaches will be differentiated below, together with the ML algorithms applied. Each ML algorithm has parameters that strongly influence its performance. Consequently, different parameters values were tested (see above); only the best results will be shown. Unless otherwise stated, results are only presented for the “true” label, as the overall scores for true and false are diluted due to the low frequency of true positives in the corpus. The subsequent sections outline the results for the English and the German corpuses.

6.1.3.2.1 English

For English, two result sets were tested as described above: one based on the smaller synonyms tag set and one based on the larger set including hyponyms and troponyms. For the smaller standard set, an initial cut-off of 40 characters was chosen; the initial cut-off for the larger set was 30 characters. Table 46 shows the results for the evaluation of both sets using the algorithms and parameters shown in Chapter 5. The table only contains the top results for the adjusted F1 measure for each approach, feature set, and method. Precision, recall, and F1 (if shown) results are the (macro-)averages of the *k*-fold runs. The adjusted F1 measure multiplies the recall of the ML result (macro-average) with the maximum possible recall due to the cut-off based reduction of the training set and recalculates the F1 value using the precision of the ML step. For each cell, several parameter values may have the same result, which is not differentiated here.

	MinRelation		MinAll Relation		MinSentence		MinAll Sentence		Relation		Sentence	
	F1	Adj. F1	F1	Adj. F1	F1	Adj. F1	F1	Adj. F1	F1	Adj. F1	F1	Adj. F1
Synonyms only (small set; cut-off at 40 characters distance)												
SVM*	75.6%	68.2%	73.7%	65.3%	66.6%	59.1%	73.7%	65.3%	76.3%	67.1%	81.9%	72.2%
PAUM*	73.6%	66.1%	76.3%	68.8%	62.8%	54.8%	72.5%	63.4%	65.4%	56.9%	70.2%	61.3%
KNN	68.4%	62.4%	70.9%	65.0%	53.6%	47.4%	64.3%	56.2%	63.4%	56.5%	64.9%	56.6%
NB	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
C4.5	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
Synonyms, hypo- and troponyms (large set; cut-off at 30 characters distance)												
SVM*	66.4%	63.8%	67.1%	64,0.%	70.5%	68.5%	73.2%	71.5%	73.5%	70.4%	79.3%	76.0%
PAUM*	62.2%	64.2%	69.0%	66.8%	71.7%	72.3%	73.1%	74.3%	72.8%	73.2%	74.0%	74.3%
KNN	65.5%	63.1%	67.2%	64.5%	59.5%	58.2%	69.4%	67.6%	67.1%	66.1%	63.3%	61.6%
NB	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
C4.5	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%

* Due to limited effect, the filtering variation is not included in all combinations

Note: Best configuration for SVM and sentence: linear SVM with cost = 10 – no difference between tau = 0.3/1 and classification threshold = 0.1/0.5

Table 46: Best result for different ML methods for the two sets tested by approach and features used (author's representation)

With a maximum adjusted F1 score of 76.0% (and a F1 score of 79.3%) the SVM using the tokens within the sentences touched by child labor distance relations produced the best result. This top result for the adjusted F1 score (excluding other options for SVM/Sentence) is followed by PAUM/Sentence and PAUM/Relation. The best approach for the smaller set is a combination of SVM and Sentence. Table 47 provides a more detailed view of the precision and recall values for the best overarching approach based on sentences. For comparison, the best result for the same feature set has been included for the smaller synonym set. Using the larger set, SVM achieves particularly high precision values, leading to a larger F1 score.

	Sentence							
	Synonyms only; Cut-off 40				Synonyms, hyponyms, troponyms; Cut-off 30			
	Prec.	Rec.	F1	Adj. F1	Prec.	Rec.	F1	Adj. F1
SVM	96.0%	72.1%	81.9%	72.2%	95.1%	69.1%	79.3%	76.0%
PAUM	97.0%	55.8%	70.2%	61.3%	83.8%	72.9%	74.0%	74.3%
KNN	92.0%	50.9%	64.9%	56.6%	64.3%	64.6%	63.3%	61.6%

Table 47: Detailed overview of results for SVM, PAUM, and KNN (author's representation)

Until now, only variations for the feature sets, ML techniques, and their parameters have been considered. In the following, the results of several additional options are shown, reusing the best parameter constellation identified in Table 46.[129] First, the effect of the cut-off distance is studied, given that the ones chosen for the two test sets were trade-offs. Second, the methodology section introduced the possibility of using sentiments from a lookup dictionary to enhance machine-learning. Their effect on the best-performing approach is discussed. The same is also performed for the POS tag. Third, the result of removing all stopwords in the "classi-

[129] The first, best parameter constellation was chosen based on yielding the best adjusted F1 score: Filtering with a ratio of 0.2 (no impact of filtering seen during calculations), a threshold for probability classification of 0.1, and tau set to 1.

cal" sense is detailed. Finally, the effect of the *k*-fold testing approach is shown. The latter was used to study its effect on the result rather than to find a better-performing ML technique (the best-performing approach from the first three tests was applied).

							Sentence			
							Synonyms, hyponyms, troponyms			
Nr.	**Technique**	**Cut-off dist.**	**Senti-ments**	**POS-tags**	**"Classic" stopwords removed**	**K-fold cross-validation**	**Prec.**	**Rec.**	**F1**	**Adj. F1**
1	**Cut-off distance variation**	10	No	No	No	8	90.5%	66.5%	76.2%	68.1%
		20	No	No	No	8	94.6%	64.4%	74.7%	68.7%
		30 (orig.)	No	No	No	8	95.1%	69.1%	79.3%	76.0%
		40	No	No	No	8	91.9%	68.5%	77.8%	75.3%
		50	No	No	No	8	95.5%	70.3%	80.6%	78.9%
		60	No	No	No	8	96.1%	70.8%	81.0%	79.9%
		70	No	No	No	8	90.9%	73.2%	80.3%	80.3%
		80	**No**	**No**	**No**	**8**	**95.9%**	**73.9%**	**83.0%**	**82.7%**
		90	No	No	No	8	94.2%	71.7%	81.0%	80.6%
		100	No	No	No	8	92.4%	68.3%	77.8%	77.8%
		unlimited	No	No	No	8	88.9%	68.3%	76.1%	77.3%
2	**Include senti-ments/ POS-tags**	30	Yes	No	No	8	95.9%	70.1%	80.2%	76.9%
		30	No	Yes	No	8	88.4%	67.8%	74.9%	72.9%
		30	Yes	Yes	No	8	94.0%	66.2%	77.0%	73.7%
		80	Yes	No	No	8	95.9%	73.0%	82.2%	82.1%
		80	No	Yes	No	8	94.7%	69.9%	79.8%	79.6%
		80	Yes	Yes	No	8	94.7%	69.9%	79.8%	79.6%
3	**Removing stopwords**	30	No	No	Yes	8	94.9%	69.3%	79.6%	76.0%
		80	**No**	**No**	**Yes**	**8**	**97.1%**	**73.7%**	**83.4%**	**83.0%**
4	**K-fold cross validation varia-tion**	80	No	No	Yes	2	90.2%	50.7%	64.7%	64.2%
		80	No	No	Yes	4	95.4%	71.5%	81.6%	80.9%
		80	No	No	Yes	6	97.5%	71.1%	81.7%	81.4%
		80	No	No	Yes	10	95.2%	72.4%	81.8%	81.5%

Table 48: Overview of results for variations of the best-performing ML parameter constellation (author's representation)

Altogether, two areas of improvement can be derived from the values in Table 48. First, adjusting the cut-off distance to 80 showed the best performance when tested in 10-character steps (see Figure 55 for the trade-off). Second, removing stopwords in the classical sense using the MySQL stopword list[130] leads to another improvement in the adjusted F1 value (when applied to the cut-off 80 dataset). However, including sentiments based on the SentiWordNet resource or integrating POS-tags into the ML did not yield any additional improvements.[131] Consequently, a cut-off at 80 together with classical stopword removal was applied in the additional analysis steps. Variations in the *k*-fold cross validation indicate a relatively stable result. The F1 value peaked at a *k* of eight for the best performing parameter constellation; however, this needs to be judged relative to the other results.

[130] Available from http://dev.mysql.com/doc/refman/5.5/en/fulltext-stopwords.html (543 stopwords in total; accessed 29.12.2014).

[131] Additional tests were performed using a cubic kernel, employing 2-grams rather than 1-grams and not using stemming. There were no improvements in the results.

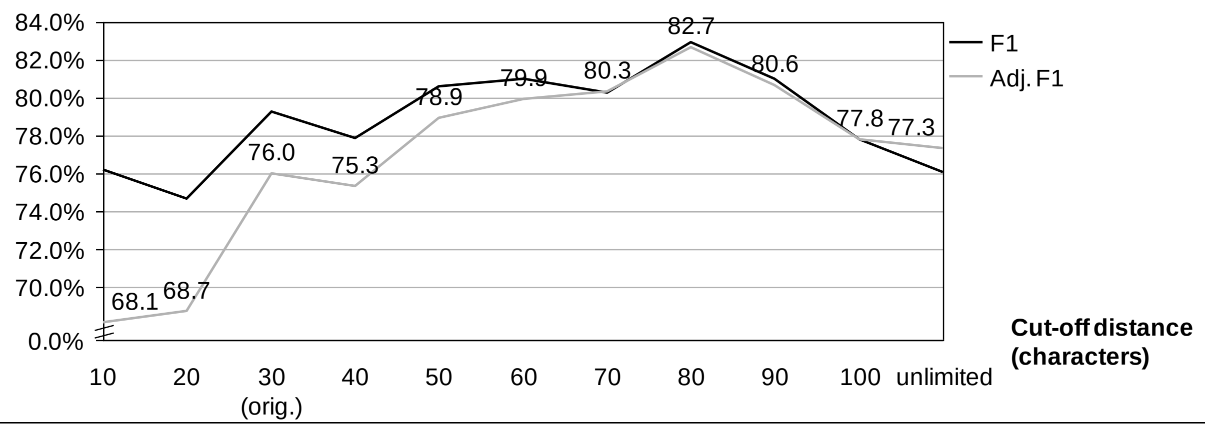


Figure 55: F1 and adjusted F1 measure for different cut-off distances for best-performing configuration of Table 47 (author's representation)

The details of the best result are depicted in Table 49. Instead of only showing the result for the child labor documents (true set), other results are also shown. The strong result of 99.7% for the overarching F1 measure is clearly visible.

True			False			Overarching			Adjusted (true)	
Precision	Recall	F1	Precision	Recall	F1	Precision	Recall	F1	Adj. Recall	Adj. F1
97.1%	73.7%	83.4%	99.7%	100.0%	99.9%	99.7%	99.7%	99.7%	72.4%	83.0%

Table 49: Detailed performance figures of best performing approach on cut-off 80 dataset (author's representation)

In summary, both SVM and PAUM appeared useful in the tests performed here. First, a broader picture was presented based on the cut-off 30 and 40 datasets and mulitple different categories for each set, defined by the ML-approach and feature set used for ML. The best approach was then sensitivity tested to identify even better results. This led to the best ML approach, using SVM to identify nearly three quarters of child labor incident documents (only a few were wrongly tagged) at an F1 of 83%.

6.1.3.2.2 Baseline comparison for English

In the following, the best performing approaches for the English corpus will be compared with other baseline approaches, as described in section 6.1.2.1.1. The best ML applied is the one outlined in section 6.1.3.2.1 with "classic" stopword reduction. No further parameter differentiations were performed. While this might have missed even better ML results for the baseline comparison values, the best approach found in Table 49 uses the largest feature set considered in the comparison from Table 46 (i.e., sentence level feature selection). Table 50 presents the combined results of the different tests. The columns for candidate set reduction and ML show the precision and recall values for the two steps independently. The last three columns indicate the performance of ED when combining the two steps. The precision value is for ML, while the recall value is the multiplication of candidate set reduction and ML. The F1 value represents the adjusted F1 score.

Baselines	**Configuration**						**Cand. set reduction**		**Machine learning**		**Combined**		
Approach	**Input data**	**Candi-dates**	**Cut-off distance**	**ML features**	**ML appr.**	**Classic SW**	**Prec.**	**Rec.**	**Prec.**	**Rec.**	**Prec.***	**Rec.**	**F1**
Random[132]	None	-	-	-	-	-	-	-	0.7	50.0	0.7	50.0	1.4
Child labor filtering**	Candidate reduced	Small set	unlimi-ted	-	-	-	1.0	94.9	-	-	1.0	94.9	2.0
	Candidate reduced	Large set	unlimi-ted	-	-	-	0.7	100.0	-	-	0.7	100.0	1.4
Lightweight rule-based***	All	Search CL	-	-	-	-	74.2	61.5	-	-	74.2	61.5	67.3
All tokens	All	-	-	All tokens	Best	Removed	-	-	63.1	15.2	63.1	15.2	24.5
Direct classi-fication with preparation	Candidate reduced	Small set	unlimi-ted	All tokens	Best	Removed	1.0	94.9	84.5	62.6	84.5	59.4	69.8
	Candidate reduced	Large set	unlimi-ted	All tokens	Best	Removed	0.7	100.0	90.3	69.2	90.3	69.2	78.3
Direct classi-fication	Candidate reduced	Large set	unlimi-ted	Sen-tence	Best	Removed	0.7	100.0	93.1	72.3	93.1	72.3	81.4
Best perform-ing; Cut-off 80	Candidate reduced	Large set	80	All tokens	Best	Removed	1.0	98.3	89.4	68.3	89.4	67.2	76.7
	Candidate reduced	**Large set**	**80**	**Sen-tence**	**Best**	**Removed**	**1.0**	**98.3**	**97.1**	**73.7**	**97.1**	**72.4**	**83.0**

* Precision of last process step (ML); ** Equivalent to all results found with candidate set reduction without distance-based filtering; *** Direct search of “child labor”/”child labour”

Table 50: Comparison of best-performing processes on English corpus with baseline approaches (in percent; author’s representation)

The last line of Table 50 depicts the best result described above. As can be seen, it is superior to all other baseline approaches but to different extents. The naïve approach of only searching for “child labo(u)r” performed significantly worse, with an adjusted F1 value of 67.3%. ML with all tokens achieved a result of 78.3%, roughly five percent points below the best performance. Adding child labor distance relations and restricting features to sentences while refraining from candidate set reduction with a cut-off distance achieved an adjusted F1 value of 81.4%. Hence, introducing the cut-off distance relations for the ML input dataset and feature selection leads to improved adjusted F1 measures for the case discussed in this thesis.

6.1.3.2.3 German

For German, a reduced analysis was performed to make it comparable with the English dataset above. The main results for different approaches are shown in Table 51.

[132] Rounding up, a random tagger would correctly classify 59 articles of the true positives and 8,416 of the false positives. Consequently, it would achieve a precision of $58.5/(58.5+8415.5)=0.69\%$ while reaching a recall of $58.5/117=50\%$.

	MinRelation				Sentence			
	Prec.	Rec.	F1	Adj. F1	Prec.	Rec.	F1	Adj. F1
SVM*	62.9%	58.7%	58.5%	59.7%	56.3%	23.0%	29.4%	32.0%
PAUM*	75.0%	64.0%	61.4%	67.9%	25.0%	8.8%	12.5%	12.7%
KNN	49.3%	64.0%	51.5%	54.9%	0.0%	0.0%	0.0%	0.0%
NB	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
C4.5	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%

* Due to limited effect, the filtering variation is not included in all combinations

Table 51: Best results for different ML methods tested by approach and features used (author's representation)

Overall, as can be seen in Table 51, PAUM, only using information from the minimum relations, performs best with a F1 score of 61.4%. This is significantly worse than the results obtained for the English corpus. Moreover, using more features for learning does not yield better results as was the case for the English dataset. Still, SVM was superior to PAUM when utilizing the larger set of features. KNN performs particularly well, while Naïve Bayes and C4.5 are not helpful for the classification, similar to the results for the English corpus.

6.1.3.3 Incident event extraction

Incident EE evaluation aims to test the different heuristics suggested for EE as well as providing an indication and lower limit for the usefulness of the techniques suggested. For the sector dimension, the result also provides a first indication of the performance of the suggested tagging approach. In order to allow a general statement of its performance, an independent evaluation without the focus on EE would be necessary, as only a selective performance is analyzed. Given that the rule-based tagger can be directly included into the filtering approach, this approach was used as the standard. The differences between the rule-based and ML-based versions are presented below. EE was tested with English; however, several of the heuristics suggested are independent of English and can be used in different languages too.

Evaluation set		Overall			Geography			Organization*			Sector (rules)		
Double	Summary	Prec.	Rec.	F1	Prec.	Rec.	F1	Prec.	Rec.	F1	Prec.	Rec.	F1
Incl.	**Incl.**	32.5%	62.6%	42.8%	55.6%	70.5%	62.2%	32.5%	43.7%	37.3%	9.0%	31.0%	13.9%
Excl.	**Excl.**	35.6%	63.0%	45.5%	59.5%	71.4%	64.9%	35.5%	59.1%	44.3%	10.7%	23.2%	14.6%

* The results for the organization dimension do not provide an evaluation of Open Calais as a service and can only be interpreted in the context of the EE performed

Table 52: Best results based on F1 measure per evaluation set category excluding root for hierarchical metric (with GS restriction; full hierarchy; author's representation)

Table 52 presents the best filtering approaches based on the F1 measure for the overarching result and each dimension per evaluation set. The evaluation was only performed for fields with a gold standard value (i.e., with GS restriction) and using the hierarchical measure. Unless otherwise stated, this configuration will also be used below. Excluding doubles and summaries from the evaluation set provides purer and more realistic values while also improving scores for all fields. However, only the geographic dimension shows F1 scores over

60%. The next highest score is the organization dimension at 44.3%. The worst performing is the sector dimension at 14.6%. Figure 56 gives an overview of the tradeoff between precision and recall for the dataset excluding doubles and summaries. The best results for each technique and dimension are presented in Table 53.

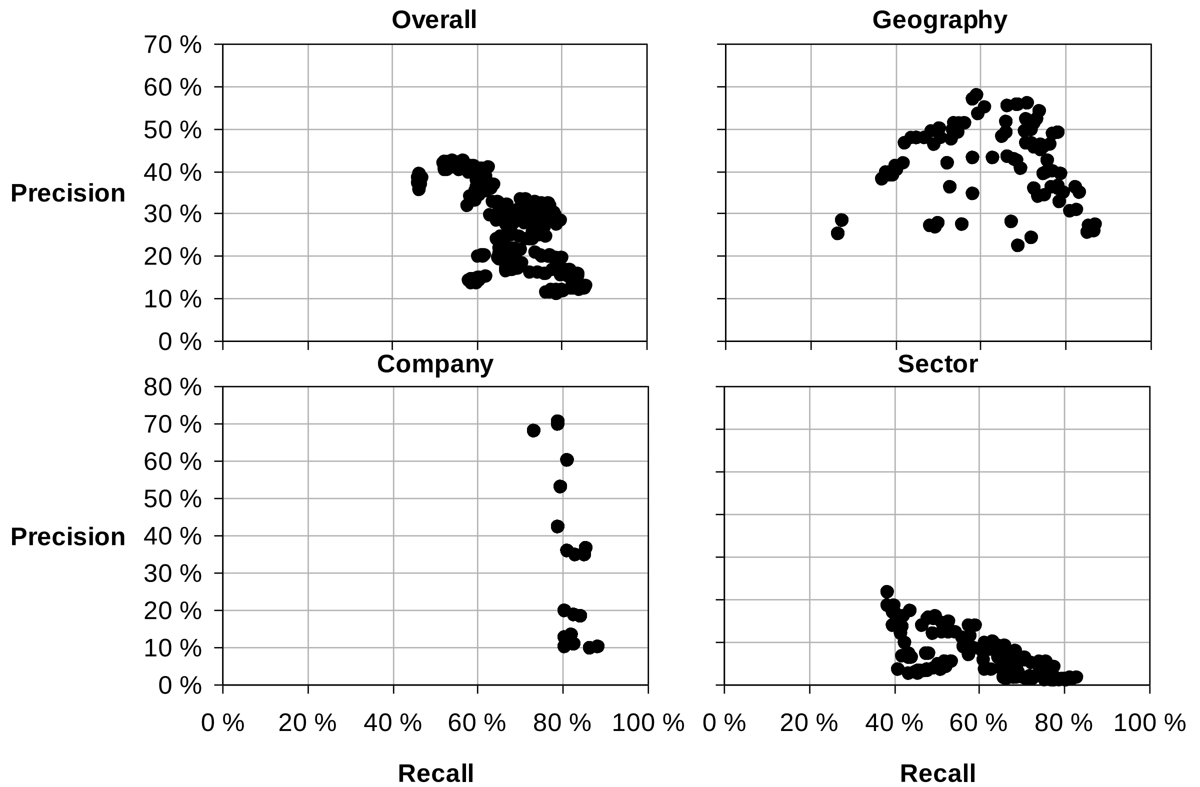


Figure 56: Comparison of precision and recall for different filtering techniques excluding root (excluding doubles and summaries; without GS restriction; full hierarchy; author's representation)

Filtering approach		Overall			Geography			Organization*			Sector (rules)		
Basis	**Technique**	**Prec.**	**Rec.**	**F1**	**Prec.**	**Rec.**	**F1**	**Prec.**	**Rec.**	**F1**	**Prec.**	**Rec.**	**F1**
Child labor relation	**Closest – with limit**	34.7%	36.5%	35.6%	56.8%	55.1%	55.9%	38.9%	25.0%	30.4%	5.0%	14.1%	7.4%
	Closest – no limit	34.9%	49.8%	41.0%	54.0%	76.1%	63.2%	47.2%	36.1%	40.9%	6.2%	28.9%	10.2%
	Sentence	**35.6%**	**63.0%**	**45.5%**	58.2%	68.3%	62.8%	40.4%	48.9%	44.2%	5.3%	25.5%	8.8%
	Distance	33.6%	65.0%	44.3%	54.0%	70.6%	61.2%	39.4%	47.8%	43.2%	5.4%	30.2%	9.2%
Child and labor words	**Closest – with limit**	35.4%	49.0%	41.1%	56.5%	69.8%	62.4%	36.1%	30.6%	33.1%	6.8%	28.7%	11.0%
	Closest – no limit	34.5%	52.3%	41.6%	53.7%	74.8%	62.5%	36.1%	30.6%	33.1%	8.0%	31.1%	12.7%
	Sentence	25.9%	71.7%	38.0%	39.8%	80.6%	53.3%	**35.5%**	**59.1%**	**44.3%**	3.3%	61.5%	6.2%
	Distance	33.8%	62.1%	43.8%	**59.5%**	**71.4%**	**64.9%**	33.8%	52.3%	41.0%	3.3%	63.9%	6.3%
Count	**Highest**	32.5%	50.1%	39.4%	46.8%	73.5%	57.2%	38.6%	38.7%	38.7%	**10.7%**	**23.2%**	**14.6%**

* The results for the organization dimension do not provide an evaluation of Open Calais as a service and can only be interpreted in the context of the EE performed

Table 53: Best results based on F1 measure per filtering heuristic excluding root for hierarchical metric (excluding doubles and summaries; with GS restriction; full hierarchy; best approaches per dimension in bold; author's representation)

No single technique performs equally well across all dimensions. Based on the overall score, an approach using sentences and child labor distance relation performs best, as already highlighted for the ML algorithm. For the geography and organization dimensions, "child" and "labor" word-based techniques perform best. For the sector dimension a count-based technique shows the best results.

Calculation technique		Overall			Geography			Organization*			Sector (rules)		
GS restriction	Hierarchy	Prec.	Rec.	F1	Prec.	Rec.	F1	Prec.	Rec.	F1	Prec.	Rec.	F1
With	Full	35.6%	63.0%	45.5%	59.5%	71.4%	64.9%	35.5%	59.1%	44.3%	10.7%	23.2%	14.6%
	Top-Level	64.8%	64.1%	64.5%	88.8%	90.5%	89.6%	35.5%	59.1%	44.3%	40.7%	46.5%	43.4%
Without	Full	40.9%	62.9%	49.6%	55.8%	71.2%	62.6%	70.3%	78.9%	74.4%	21.5%	38.7%	27.7%
	Top-Level	47.2%	70.4%	56.5%	80.5%	91.4%	85.6%	70.3%	78.9%	74.4%	31.1%	72.9%	43.6%

* The results for the organization dimension do not provide an evaluation of Open Calais as a service and can only be interpreted in the context of the EE performed

Table 54: Best results based on F1 measure per calculation technique excluding root for hierarchical metric (excluding doubles and summaries; author's representation)

The method of calculating precision and recall has a strong impact on the results. In the tables above, the most restrictive version has been applied. When also including situations where no gold standard value is given (i.e., no should be set/no gold standard restriction), the overall F1 score (when analyzing the full hierarchy) increases by 4.1 percent points. Constraining the analysis to only the top categories in the evaluation hierarchies (i.e., countries or top level sectors) leads to an even stronger boost in the F1 score, with 64.5% when including the gold standard restriction and 56.5% without it. This is strongly driven by the sector score, which performs much better when only the top-level industries are analyzed. Depending on the use case, this may be sufficient. For this thesis, the closer the connection to a supplier location the better.

Calculation technique		Incl. doubles	Excl. doubles
GS restriction	Hierarchy	Incl. summaries	Excl. summaries
With	Full	42.8%	45.5%
	Top-Level	63.7%	64.5%
Without	Full	51.4%	49.6%
	Top-Level	60.1%	56.5%

Table 55: Best overall results based on F1 measure per calculation technique excluding root for hierarchical metric (excluding doubles and summaries; author's representation)

Finally, Table 55 presents a comparison of the overall F1 scores for the different calculation techniques and datasets considered. Independent of the calculation technique used, including or excluding doubles and summaries does not strongly affect the result. The increasing scores when not using the gold standard restriction seem to indicate that missing dimension values can be identified by some techniques. Moreover, only aiming for a correct top-level classification (i.e., country or top-level sector according to ISIC) strongly improves F1 scores.

The metrics in the preceding tables have been calculated using the hierarchical measure for geography and sector. For comparison, a second calculation using classic, flat measures without a Levensthein comparison was performed. Table 56 shows the results of a similar split to that in Table 52 for the “flat” calculation of precision, recall, and F1 for each dimension. Table 57 shows the effect of performing the calculation without restricting to the gold standard for the dataset, excluding doubles and summaries.

Evaluation set		Overall			Geography			Organization*			Sector (rules)		
Double	Summary	Prec.	Rec.	F1	Prec.	Rec.	F1	Prec.	Rec.	F1	Prec.	Rec.	F1
Incl.	Incl.	32.6%	44.4%	37.6%	51.0%	70.8%	59.3%	32.5%	43.7%	37.3%	7.8%	18.7%	11.0%
Excl.	Excl.	34.3%	43.3%	38.3%	53.4%	68.1%	59.9%	35.5%	59.1%	44.3%	9.8%	11.6%	10.6%

* The results for the organization dimension do not provide an evaluation of Open Calais as a service and can only be interpreted in the context of the EE performed

Table 56: Best results based on F1 measure per evaluation set category for flat metric (with GS restriction; author’s representation)

Evaluation set	Overall			Geography			Organization*			Sector (rules)		
GS restriction	Prec.	Rec.	F1	Prec.	Rec.	F1	Prec.	Rec.	F1	Prec.	Rec.	F1
With	34.3%	43.3%	38.3%	53.4%	68.1%	59.9%	35.5%	59.1%	44.3%	9.8%	11.6%	10.6%
Without	36.4%	57.8%	44.7%	48.4%	71.1%	57.6%	70.3%	78.9%	74.4%	21.4%	37.5%	27.2%

* The results for the organization dimension do not provide an evaluation of Open Calais as a service and can only be interpreted in the context of the EE performed

Table 57: Best results based on F1 measure with and without gold standard restriction for flat metric (excluding doubles and summaries; author’s representation)

The comparison with the “flat” measures shows comparable results to the hierarchical measures. Overall, F1 measures increase slightly thanks to an increase in both precision and recall for the best-performing geography and sector results. However, the metrics cannot be directly compared in absolute terms (the differences between the two calculation approaches appear in line with, e.g., Table 41). Nevertheless, when comparing the top ten approaches for the calculation without doubles and summaries in terms of F1 measure value (with GS restriction and full hierarchies), half of them overlap.

Up to this point, this section has used the precision, recall, and F1 results produced by the rule-based approach for sector tagging. In the following, we provide the results for the alternative ML-based sector classification approach.[133] The root was excluded for all calculations. Furthermore, the calculation included doubles and summaries. Given the only slight differences between including and excluding doubles and summaries for the rule-based approach (for the F1 measure), a similar result might be assumed here. Table 58, Table 59, and Table 60 present the results for different views.

[133] Given that 14 of the 75 incidents with tagged sectors have more than one sector, the maximum recall rate the single-dimensional classifier can achieve is 84%.

	Sentences			Minimum sentences		
Algorithm	**Prec.**	**Rec.**	**F1**	**Prec.**	**Rec.**	**F1**
SVM	4.1%	12.1%	6.1%	7.5%	14.5%	9.9%
PAUM	5.5%	12.7%	7.7%	4.1%	7.2%	5.3%
KNN	16.5%	14.4%	15.4%	18.3%	22.4%	20.1%
NB	3.6%	16.8%	5.9%	3.6%	16.8%	5.9%
C4.5	3.6%	16.8%	5.9%	3.6%	16.8%	5.9%

Table 58: Results for ML-based sector tagging for different classifiers excluding root for hierarchical metric (including doubles and summaries; restricted to gold standard; full hierarchy; author's representation)

	Sentences			Minimum sentences		
Algorithm[134]	**Prec.**	**Rec.**	**F1**	**Prec.**	**Rec.**	**F1**
Per ISIC code	5.5%	12.7%	7.7%	5.7%	6.8%	6.2%
Per line per ISIC code	16.5%	14.4%	15.4%	18.3%	22.4%	20.1%

Table 59: Results for ML-based sector tagging for different learning approaches excluding root for hierarchical metric (including doubles and summaries; restricted to gold standard; full hierarchy; author's representation)

Calculation technique		Sentences			Minimum sentences		
GS restriction	**Hierarchy**	**Prec.**	**Rec.**	**F1**	**Prec.**	**Rec.**	**F1**
With	**Full**	16.5%	14.4%	15.4%	18.3%	22.4%	20.1%
	Top-Level	57.3%	52.0%	54.5%	57.3%	52.0%	54.5%
Without	**Full**	36.0%	36.2%	36.1%	35.7%	37.1%	36.4%
	Top-Level	36.8%	69.2%	48.0%	36.8%	69.2%	48.0%

Table 60: Results for ML-based sector tagging per gold standard restriction and calculation hierarchy excluding root for hierarchical metric (including doubles and summaries; author's representation)

For the sector classification task, KNN performs best (Table 58). However, even the best performing approach only leads to a F1 score of 20.1 %. This score is reached if the ML classifier uses the minimum sentences as features. Training the classifier for each code yields a worse performance (Table 59). However, if only countries or top-level sector tagging are needed, the results in Table 60 indicate a performance of around 50% for the F1 score in a calculation with and without gold standard restriction. Hence, for these specific tasks the approach may be viable.

6.1.4 Text mining summary

Summing up, the results for the TM are twofold. First, the ED approach produces relatively high F1 measures; however, improvements in the recall value in particular still appear important for practical use. The proposed two word child labor distance relations and the candidate set reduction technique both increase F1 scores. Sentiments and POS tags did not improve performance for the configuration tested. Second, the EE shows weaker results. This is particularly true for the sector dimension (for both approaches suggested). Consequently, given the EE results, a manual cleaning step appears necessary before values are finally entered into a risk management system in a productive scenario.

[134] "Per ISIC code" uses the whole label at once, while for "per line per ISIC code" the label is separated into several parts (Appendix prototype implementation).

6.2 Risk model

Due to the absence of a gold standard with the “true” rating of supplier locations with regard to their child labor breach risk, the evaluation of the risk model is strongly based on simulation and expert input. The following sections present the evaluation approach and the respective results.

6.2.1 Evaluation approach

The evaluation of the overall suggested system configuration and model is based on three steps. As the model is designed to be used by experts, matching the experts’ needs can be seen as an important evaluation step (Shanteau 1988). Therefore, the configuration will first be critically examined with respect to fulfilling the overarching requirements of a sustainability risk management system. Second, a system-specific evaluation will be conducted. This evaluation builds on the technology acceptance model (TAM), which determines the adoption of an information system within an organization. Third, sensitivity tests will be conducted to analyze the model’s reaction to input – the motivation behind these will be explained in the TAM section.

6.2.1.1 General requirements-based evaluation

Chapter 3 provides a discussion of requirements for a general sustainability risk management system based on interviews with five industry experts. The discussion resulted in a set of seven propositions that should be fulfilled by such a general information system (see Table 65 for a summary). As explained above, a child labor supply chain risk management system can be viewed as a specialized sustainability risk management system focusing on one element of sustainability, particularly social sustainability, that features in many standards. Consequently, we will briefly discuss how the propositions are reflected in the system, and particularly in the risk model.

6.2.1.2 System-specific evaluation

The system-specific evaluation aims to empirically validate several assumptions. Furthermore, it is also used to integrate experts’ opinions into the system. The key concern evaluated in this thesis is whether the suggested prototype is likely to be adopted in practice. It is assumed that if experts are willing to use the proposed IT system for their daily work, the system would have a positive impact on a company’s performance. As the system was not used by the respondents in practice, the tests are based on hypothetical questions.

6.2.1.2.1 Introduction

Criteria for predicting the adaption of IT systems have been discussed for decades. One key model in this discussion is the technology acceptance model (TAM), which is based on the theory of reasoned action (TRA) and has been proposed by Davis and other scientists (Davis 1989; Davis, Bagozzi, and Warshaw 1989). Over the years it has become a well-documented and useful model for predicting a user's acceptance of an information system (Legris, Ingham, and Collerette 2003; Venkatesh and Davis 2000). The key feature of the TAM is that perceived usefulness and perceived ease of use mediate the effects of external variables on two further variables, namely the attitude towards using an information system and the behavioral intention to use an information system (Davis, Bagozzi, and Warshaw 1989). In general, it presents a way to forecast how likely it is that a system will be used (i.e., adopted) in a practical context; therefore, it is an appropriate model for evaluating the information system for child labor risks proposed in this thesis.

In extending the original TAM to the TAM2 model, Venkatesh and Davis (2000) particularly focus on the factors influencing the perceived usefulness of a system. They differentiate between social influence and cognitive instrumental processes. Especially the later processes do not depend on an individual's social context. TAM2 suggests "[…] that people use a mental representation for assessing the match between important work goals and the consequences of performing the act of using a system as a basis for forming judgments about the use-performance contingency (i.e. perceived usefulness)." (Venkatesh and Davis 2000, 191).[135]

Well-known competing models have been developed with the theory of planned behavior and the decomposed theory of planned behavior. Both theories have advantages for a fuller understanding of behavioral intentions (S. Taylor and Todd 1995); however, functionality considerations with regard to perceived usefulness may not be derived in comparable detail. This is possible with the extensions applied in TAM2. Consequently, TAM2 has been selected for further discussion. It will be argued that the proposed system's technical design is expected to positively influence the system's use in an organizational context when building on the rela-

[135] Studies that build on TAM and show significant effects have, among other things, been criticized for strong student involvement, a diminishing focus on business process applications, and a high use of self-reported system use (Legris, Ingham, and Collerette 2003). Moreover, TAM's explicative power is limited by only explaining roughly 40% of total variance (Legris, Ingham, and Collerette 2003). While this appears a challenge, Legris, Ingham, and Collerette (2003, 202) mainly suggest including TAM into a broader model acknowledging organizational context and other social factors. Thus, missing functionality variables are not seen as the explanation for the wide variance. Consequently, building on TAM and particularly TAM2 provides a theoretical basis when stressing the influence of application functionality on its subsequent use and determining the degree to which an application's functionality supports this.

tionships of TAM2. Figure 57 depicts the TAM2 model used in the following. Cognitive instrumental processes are included with the following variables: job relevance, output quality, result demonstrability, and perceived ease of use. Social influence processes are included through the following factors: subjective norm, image, experience, and voluntariness. TAM2 was later adapted into TAM3, but no changes were made to the perceived usefulness variable (Venkatesh and Bala 2008, 280). Given that the perceived ease of use will be ignored for the discussion, as argued below, TAM3 does not provide any additional benefits.

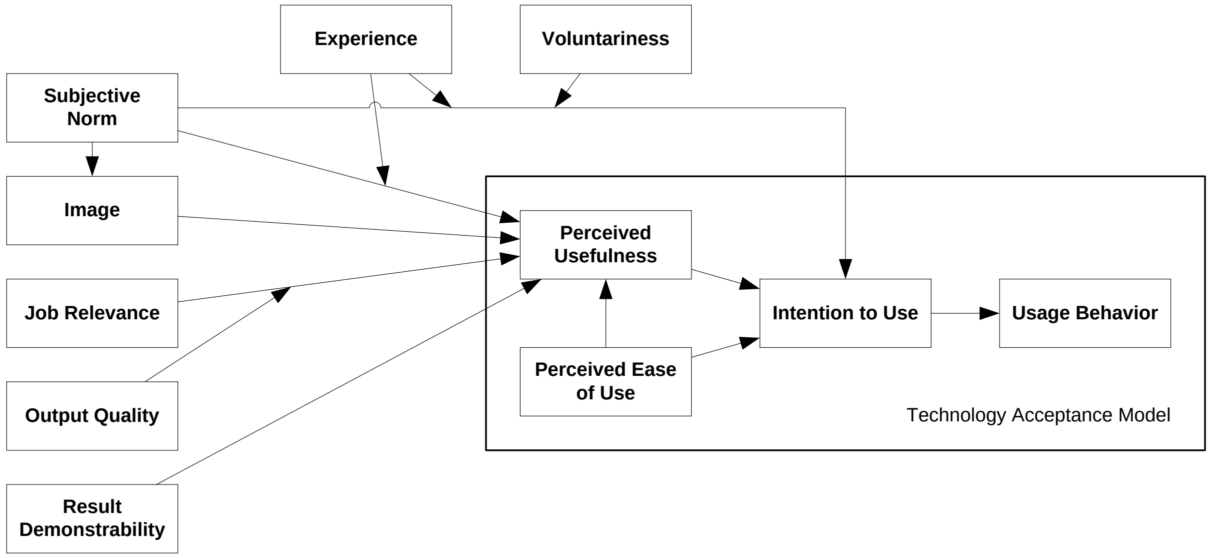


Figure 57: TAM2 with original TAM core highlighted (based on Venkatesh and Davis 2000, 197).

The key feature of the proposed system and model combination is calculating a ranking of supplier locations based on the child labor compliance breach risk. Thus, no user-interaction in the classical sense is needed, as would be the case with a text editor, for example. Hence, the focus of the presented design is on the system's functionality and not user interaction (see for example the discussion on how to evaluate ease of use in Davis, Bagozzi, and Warshaw 1989). Therefore, the system's ease of use will not be part of the analysis. Moreover, social influence processes can only be tested by integrating the system into a user environment. Consequently, these variables have also been omitted from this analysis, although they might provide an additional argument for the system's usefulness in the future. In TAM2 (Venkatesh and Davis 2000, 197), job relevance and output quality have a combined effect on the perceived usefulness, while result demonstrability is seen independent of this. The definitions of these four items are presented in Table 61.

Item	Definition
Perceived usefulness	The extent to which a person believes that using an IT system will enhance his or her job performance (Venkatesh and Bala 2008, 275)
Job relevance	The degree to which an individual believes that the target system is applicable to his or her job (Venkatesh and Davis 2000, 191)
Output quality	The degree to which an individual believes that the system performs his or her job tasks well (Venkatesh and Davis 2000, 191)
Result demonstrability	The degree to which an individual believes that the results of using a system are tangible, observable, and communicable (G. C. Moore and Benbasat 1991; Venkatesh and Bala 2008, 277)

Table 61: Item definitions according to TAM2 (author's representation)

Consequently, the system-specific evaluation focuses on understanding the system's effect on these three variables. For this purpose, several hypotheses have been defined that are depicted in the next section.

6.2.1.2.2 Evaluation Hypotheses

Job relevance and output quality have the most significant combined influence on perceived usefulness within TAM2. Both can be seen as significantly influenced by the technical design of a system. In TAM2, job relevance is determined by asking two questions: whether the usage of the system is important and whether the usage of the system is relevant (Venkatesh and Davis 2000). For both, the system has to implement core functionality of an expert's activity portfolio in order to be able to support them. To evaluate how much the proposed model and system fulfills the criterion of job relevance, two tests are performed: whether the general system idea would support the item of job relevance and how changes in system functionality influences the job relevance for the specific use case. This can be compared with an extension of the job relevance item in the particular context of this thesis and results in the following hypotheses:

- H1a: If the child labor supply chain risk management system should include a specific requirement, then the job relevance of a system combining all requirements is seen as high.
- H1b: All experts significantly agree on including all suggested requirements.
- H1c: The general system containing all mentioned requirements has job relevance.

If H1c is fulfilled, then a system's job relevance will be considered high if it implements the requirements given for H1b. Moreover, given that job relevance as used in TAM2 is formulated in a very generic way, two additional system specific questions are suggested. These two questions focus on making more informed supply chain decisions with reduced biases. This "extended" job relevance will also be tested in a similar manner to the more general.

Just as job relevance relates to the functionality of the system, output quality is linked to "how well the system performs those tasks" (Venkatesh and Davis 2000, 191). Direct measurement of this factor would only be possible if a prototype was embedded into a specific workplace setting. While this is left for future work, two measures for this item are suggested to reflect the implementation of the requirements.

First, we suggest decomposing the overall system into its main components and determining how experts would approach the key decisions included in the system based on example data. As no gold standard for an overall ranking is available, this can be seen as a strategy to reduce the number of hidden factors that influence the experts' decisions. Although experts are subject to at least not fully rational (bounded) behavior (see for an early discussion e.g. W. Edwards 1954; March 1978), experts decide using a set of decision strategies that employ a significant amount of heuristics and learning from the past (March 1994, 12; Shanteau 1988, 207). Therefore, comparing with expert output might provide an indication of any overlap between the system's approaches and these strategies. Moreover, it will help evaluate whether the experts' responses can be covered by the system's design.

Second, given that the output is a ranking based on a quantitative model, it should adhere to applicable quality measures. Statistical organizations have compiled well-established criteria for the quality of quantitative data that can be applied to the case at hand. Within the European Union (EU), the Quality Assurance Framework of the European Statistical System prevails. It has five core dimensions (European Statistical System 2012, 21): (1) relevance, (2) accuracy and reliability, (3) timeliness and punctuality, (4) accessibility and clarity, and (5) coherence and comparability. While (1) has already been discussed for TAM as part of job relevance, the other four criteria should be fulfilled by the system's output. Nevertheless, accuracy and reliability testing would require a true value dataset of either the probability for a single location or the ranking of supplier locations. As knowing of a child labor incident is equivalent to 100% risk probability, and the system should provide results beyond pure auditing scores, a direct gold standard that states the child labor probability location for suppliers for at least a specific point in time appears infeasible. Therefore, criterion (2) is excluded from this evaluation. Similarly, given that the implementation focuses on a prototypical system and model, processing speed (i.e., criterion (3)) is also omitted. Criterion (4) is strongly related to result demonstrability and will be discussed in the next paragraph. Combined, this leads to the following hypotheses, based on the remaining criterion (5):

- H2a: The output of key functionality of the system may correspond to expert strategies for determining them.
- H2b: The system's output is coherent and comparable.

Finally, given the one-dimensionality of the system's output with an ordered list of suppliers, the result demonstrability will be tested with the following underlying hypothesis:

- H3: The general system described has high result demonstrability.

Altogether, the five hypotheses aim to provide a picture of the system and model's perceived usefulness as an indicator of its use according to TAM. Although not all factors influencing the perceived usefulness in TAM2 will be tested, the evaluation provides a basic understanding of the potential practical impact of the system with regard to its use. However, no predictions about the effect on a specific organization may be made.

6.2.1.2.3 Methodology

Methodologically, the hypotheses will be approached in two ways. On the one hand, an empirical survey was used with the goal of eliminating hypotheses via falsification. The results of this approach were incorporated if expert input provided value for testing the hypothesis or if expert input could be used to assess their opinion on a statement in a population-descriptive way. This was done for hypotheses H1, H2a, and H3. On the other hand, hypothesis H2b was qualitatively assessed along with quantitative sensitivity testing. The following sections focus on the discussion of the expert questionnaire. The description and discussions for hypothesis H2b will be provided in section 6.2.1.3.

6.2.1.2.4 Research design

The research design is "quasi-experimental" as experts were selected from natural settings, and differences between experts are not controlled (Bortz and Döring 2006, 54). However, as the aim is to show common links across all participants instead of differences between groups, this is not seen as a limitation. The aim of the research is not to cover differences between companies, but to cover expert opinions on different hypotheses. Consequently, parts of the questions require the design of a fictive setting to detach the experts from their company background as far as possible. In particularly the part dealing with H2a is designed in a laboratorial way to allow for a standardized context (Bortz and Döring 2006, 57). While this diminishes the external validity of the research, it lets us confront each participant with the same situation. The basic population of the questionnaire includes experts who work in one of the following functions directly or in functions that cover aspects of them:

- Supply Chain Management
- Sustainability Management
- Risk Management
- Quality Management

Experts from companies sourcing from outside the EU are very likely to be more concerned with an international sourcing perspective than those who source from within the EU. Still, the latter will also apply standards that need to be fulfilled by suppliers. Furthermore, experts who deal with customers within the EU would be useful. This could stabilize the effect of operationally relevant factors driven by cultural differences with customers that can influence management's behavior with regard to child labor. Consequently, these two variables were traced so they would be available for discussion. Finally, and most importantly, the questionnaire was directly aimed at topical experts. As it is very unlikely that companies will have a dedicated staff member concerned with child labor supply chain risk management, the four closest related company functions were targeted. Therefore, these experts were a particular focus of this survey. This includes the assumption that if the hypotheses are not falsified by topical experts, the system will be useful for a larger amount of companies.

6.2.1.2.5 Questionnaire and evaluation design

The following paragraphs present the operationalization of the different variables. In general, the questionnaire is designed in four parts while covering the hypotheses discussed in section 6.2.1.2.2. A detailed listing can be found in the "Appendix R: Structure of expert survey". In general, the design of the survey follows the "high hurdle technique" (Reips 2002, 249) which can be used to ensure that responses meet a higher level of seriousness. For this purpose the introduction offers a realistic description of the content and duration of the survey. In addition longer, more complicated pages were put before easier ones (Thielsch and Weltzin 2009).

The evaluation of H2a can be compared to an experimental two-group plan where the outcome is compared after a treatment. One group is represented by the experts, who are confronted with different input data to trigger responses which can be compared with the other group, represented by the finished model. To assess the prior and the reaction to observations, a fictive company-context is provided that depicts an international supply chain consisting of four supplier locations. While such a case would not require the assistance of an IT system in real life, this laboratorial setting puts each expert into the same workplace setting and standardizes the environment (Bortz and Döring 2006, 57). The experts were provided with the same data used as input for the IT system.

The first part of the questionnaire covers basic demographic questions to filter whether the response was from an expert who fulfils the criteria set out for the survey's target population. The second part includes questions on H2a, which is based on a comparison between expert-input and the strategies implemented in the software. It was included before questions on H1, because these had the potential to influence the expert in his answers on the questions. It also takes more time to answer these questions. In order to determine the expert approach to different key parts of the system's functionality, the model was split into three parts and questions were focused on the specific context: determining of the prior, the inclusion of external observations, and the analysis of the audit results.

As the system's output is a ranked list of supplier locations, answers about the prior required the experts to provide a similar ranking. This was only used for the prior in order to keep the questionnaire a reasonable length. Moreover, only four fictive supplier locations (see Table 62) were chosen (Bortz and Döring 2006, 156) and were presented to the expert together with additional data (also used by the model) for his answer.[136] The instruction was: "Please rank the suppliers according to their risk of having a child labor incident based on the introductory data provided." Here, the comparison uses ordinal-scaled rankings, given that asking for intervals representing the riskiness of different supplier locations would require a prohibitive amount of expert input. For further simplification, each supplier has only one supplier location.

[136] These are the estimated child labor rates per country and ruralness, the company's sector's share of overall child labor per country and ruralness, absolute estimated number of children working per country, sector, and ruralness, and the estimated number of companies per country, sector, and ruralness.

Naming category	Supplier	Country	Region	City/ Village	Ruralness	Sector
Encoded names	Supplier A	Blueland	Blueregion 1	Bluevillage 1	rural	Construction
	Supplier B	Pinkcountry	Pinkregion	Pinkcity	urban	Manufacturing
	Supplier C	Blueland	Blueregion 1	Bluecity 1	urban	Construction
	Supplier D	Blueland	Blueregion 2	Bluecity 2	urban	Manufacturing
Real names	Supplier A	Indonesia	Provinsi Jawa Tengah	Simpar	rural	Construction
	Supplier B	India	State of Andhra Pradesh	Tirupati	urban	Manufacturing
	Supplier C	Indonesia	Provinsi Jawa Tengah	Purwokerto	urban	Construction
	Supplier D	Indonesia	Jawa Barat	Badung	urban	Manufacturing
OpenGeo-Names identifiers/ ISIC codes	Supplier A	1643084	1642669	6372706	rural	F4312
	Supplier B	1269750	1278629	1254360	urban	C1311
	Supplier C	1643084	1642669	1630328	urban	F4100
	Supplier D	1643084	1642672	1650357	urban	C2511

Table 62: Overview of fictive supplier locations used in the questionnaire (author's representation)

To be able to compare the system's functionality with an expert's reaction to a specific observation, four different news items were provided and the experts were asked to determine the influence the observation would have on the risk evaluation of a supplier location with regard to breaching the child labor compliance code. Again for simplification and reduction in the questionnaire's size, only one of the four suppliers introduced above was used. For this purpose the question focused on fictive supplier B and was formulated as follows "How much does the news report influence the perceived probability of a child labor incident at Supplier B?" For this purpose a five-step Likert influence scale was used (extremely influential-not at all influential; Wigas 2006). The information revealed contains four pieces of information, as depicted in Table 63.

No.	Type	Description
1	News item	Incident report on one supplier with high credibility and relevance (from news report)
2	News item	Incident report on country level from news report
3	News item	Incident report on sector and region level from news report
4	Social item	Incident report on different sector and same city level from social media post

Table 63: Pieces of information revealed during the questions for testing hypothesis H2a (author's representation)

The different news reports and their content were selected to cover significant contextual connections with supplier B. The news reports had different levels of overlap with fictive supplier B, situated in the urban city of "Pinkcity" in "Pinkland," operating in the manufacturing sector. Figure 58 shows these overlaps.

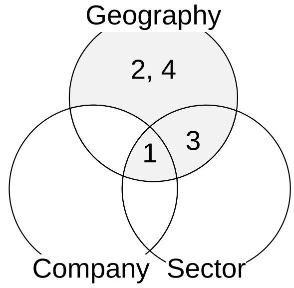


Figure 58: Overlapping areas between news items 1-4 and the dimensional attributes of supplier B (author's representation)

The first news item focused on the effect of a news report with maximum overlap and thus relevance. In contrast, input number two was neglected by the default system configuration and a corresponding action by the experts was tested as it only relates to the minimum overlap of country on the geography dimension. Item three described an incident with overlapping region and sector. Finally, the last news item aims at the effects of a message focusing on a different sector in the same city. Different real news items were adapted to correspond to the fictive suppliers and settings. For this purpose only the dimensions mentioned were changed to fictive names of countries, regions, and cities.

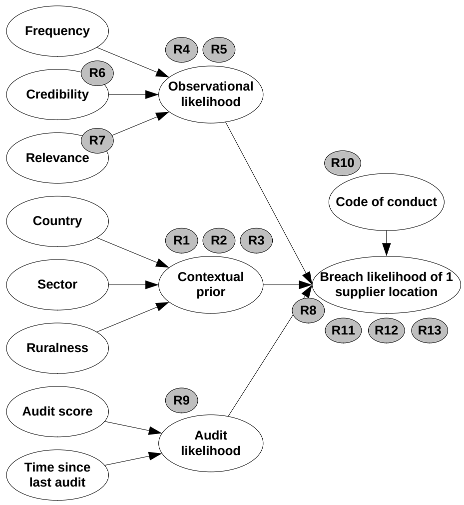
One question was used to determine the probabilities experts attach to standard audit scores by asking "Which probability of having a child labor incident only based on audit data would you associate with a supplier reaching three different types of audit scores?" This probability allows the mean of the audit probability to be determined. A final question then aimed to define the relative weight experts would attach to the different probabilities of the key features.

The third part of the survey investigates job relevance and result demonstrability. For H1a a question was developed that lists the main functionality of the system and asks the expert (for each of them the similar question):

> Do you agree that the following requirements should be fulfilled by a supply chain child labor risk management system that ranks suppliers with regard to their risk level of breaching child labor standards?

The functionalities were derived based on the information system model developed by Gluchowski and Gabriel, which differentiates data provision, data processing, data analysis, and data presentation (2008, 109). Each data manipulation step was investigated for the core parts of the system (prior calculation, observation integration, audit integration and output) and the most critical that have not been argued elsewhere were included in the listing. Table

64 presents an overview of the requirements included as well as a mapping to the corresponding parts of the risk model (BN).



No.	Requirement
1	Use of public statistical data sources (e.g. child labor statistics) in order to calculate an initial incident probability for a supplier location
2	Differentiate initial probability by country, sector and rural/urban
3	Automated calculation of the initial probability (i.e. prior probability) of having at least one child working at a random company
4	Automated use of publicly available news reports on child labor incidents (e.g. via internet) in order to update the initial probability
5	Automated use of publicly available non-news reports on child labor incidents (e.g. from social media sites as Twitter) in order to update the initial probability
6	Adjustment of initial probability depending on the specifity in which an incident report refers to a specific supplier location (i.e. relevance)
7	Adjustment of initial probability depending on the credibility of the channel through which an incident report has been received (e.g. news or social media)
8	Automated update of initially calculated probability without user-involvement (i.e. unbiased)
9	Use of last auditing result to update the initial probability
10	Use of whether Code of Conduct has been signed to adjust the initial probability values
11	Risk model calculates the probability for the risk of a breach of child labor standards for each supplier locations
12	System output (risk rank/score) should be integrateable into further supplier evaluation processes
13	System output (risk rank/score) should be unbiased with regard to different suppliers

Table 64: Overview of requirements for H1a and mapping to risk BN (author's representation)

H1a aims to identify a correlation between the need for a specific functionality and the overall job relevance of the general system. Thus, H1b is based on the analysis of a single variable and asks for the job relevance of the overall system, consisting of all functionalities listed for H1a. Venkatesh and Davis developed measurement items for job relevance (2000, 201) with a Cronbach's alpha above 0.8. Therefore, these items were slightly adapted[137] and two questions added (separately analyzed as "extended job relevance") to the survey, with one overarching question specifying the context. The seven-point Likert scale used in TAM2 was re-used.

> Please answer the questions below. "Such a system" refers to a supply chain child labor risk management system fulfilling all requirements listed above:

Building on these literature-tested scales should minimize the limitations inherently attached and discussed in the literature (Bortz and Döring 2006, 181). Similarly, the measurement items for result demonstrability representing H3 were added to the same question. All details of the questions can be found in "Appendix R: Structure of expert survey". A final question was used to determine the experts' opinion on whether maximizing recall or precision would be more important from the overall perspective of a child labor risk management system for

[137] "the system" was replaced with "such a system"

supply chains. Nevertheless, despite the care taken in constructing the questionnaire, the approach is limited by the coverage of system requirements and whether the measurements used truly capture the intended effects (Bortz and Döring 2006, 19).

The survey was developed using the open source tool LimeSurvey[138], which is based on PHP and MySQL and allows the development of fully-customized online surveys that include token management directly generating customized and personalized invitations and other emails. For the survey, a customized template was created that reflects the corporate design of the "Faculty of Informatics of the Vienna University of Technology." This was used to increase the level of trust in the survey.

6.2.1.2.6 Sample construction

The questionnaire focuses on experts in the fields of supply chain management, sustainability management, risk management, and quality management as well as functions covering those areas of responsibility, expecting that these management groups most likely have expert knowledge on sustainable supply chain risk management. As well as their background and function, industry might be seen as one factor that could affect an expert's responses. As a result, while no specific limitation was set for this survey, experts were invited from a broad range of different companies and sectors in order to achieve a minimum amount of randomization. As a different way of mitigating this influence, no industry-specific questions were asked during the survey. This was partly achieved by using fictional cases.

Experts were invited to participate either through a direct mail from the survey tool or through an invitation forwarded to a contact in a company. They participated in the questionnaire through their own free will after invitation.[139] To acquire participants for the survey, a "snowball-technique" was applied (Thielsch and Weltzin 2009, 74). The survey was first sent to a number of contacts in different companies who were either in the target group for the survey or could very probably forward the invitation email to experts in the target group. This technique appeared the most beneficial given the significant amount of time needed for experts to fill in the survey (an estimated 20 to 30 minutes), as well as the rather specific requirements when compared to more psychological, management-oriented surveys. In addition, the experts were asked to forward the survey to further experts if possible. In order to improve the relia-

[138] http://www.limesurvey.org/.

[139] While in psychological research, free participants are typically better educated, have a perceived higher social status, have higher intelligence, are in need of social recognition, are more likely female, are more social, are less authoritarian, and less confirmatory (Bortz and Döring 2006, 73), these attributes are not judged to have influenced the result in the expert questionnaire at hand.

bility of the response set, several standards were followed in accordance with Bortz and Döring (2006, 84):

- By using an online questionnaire, interaction with the respondents was standardized and consistent across all questions.
- If there were difficulties with the questionnaire, a contact address including a telephone number was provided.
- Pretests were performed with several individuals (scientific experts also from other domains than information technology or social sustainability) to reduce possible misunderstandings.

6.2.1.2.7 Execution

The survey was executed in four steps. First, the survey was prepared before sending the invitations, setting an initial deadline of two weeks. Subsequently, personal reminders were sent, extending the final deadline by two weeks. The tool support of LimeSurvey together with privately formulated email add-ons were used to send the invitations. LimeSurvey provided tokens that were used (included in the survey links) for the initial contacts. Each contact could forward a token to multiple respondents. Token management did not allow securing that a token could not be used twice by the same person. However, the level of token reuse was low, at 1.6 completed surveys per used token. To secure anonymity, LimeSurvey does not store tokens and responses (or response times) together. The invitation email can be found in “Appendix S: Invitation email for expert questionnaire”.

6.2.1.2.8 Data analysis

Data analysis was performed using SPSS (IBM Corp. 2013) and partly augmented with additional, manually performed calculations. As there is significant scientific dispute over whether Likert scales produce interval or ordinal data and thus whether they support parametric tests or not (e.g. Allen and Seaman 2007; Jamieson 2004; G. Norman 2010), the analysis relies on tests that are independent of any interval-scaled property of Likert scales. Also no normal distribution of the responses is required (Bühl and Zöfel 2002, 287).

Consequently, the analyses of H1 and H3 are all based on non-parametric tests. For H1a, the Spearman and Kendall correlation between each requirement and the mean of the items for the job relevance was used (Bühl and Zöfel 2002, 320). For the tests on agreement, given a median above four (i.e., a tendency to overall agreement), a Wilcoxon test for significant differences comparing the variable at hand with a fictively generated variable with all answers at the mean (i.e., four on the Likert scale) was performed (Bühl and Zöfel 2002, 294). This method was used for H1b, H1c, and H3. In contrast, for H2a, different comparisons of the answer frequencies with system and model characteristics were performed. The analyses are more strongly based on averages, standard deviation, and ranks.

6.2.1.3 Sensitivity testing

The sensitivity testing focused on the coherence and comparability of the model, one of the quality criteria defined above under H2b. Coherence is defined as "[T]he quality of being logical and consistent." (Oxford Dictionary 2014c) In this context "logical" can be understood as "[C]haracterized by or capable of clear, sound reasoning" (Oxford Dictionary 2014d) while "consistent behavior" is defined as "[T]he quality of achieving a level of performance which does not vary greatly in quality over time" (Oxford Dictionary 2014e). To be "comparable features", they need to be "[A]ble to be likened to another" (Oxford Dictionary 2014f). While comparability is mainly driven by having a ranked list as the result and using probability values as the network's output, the coherence can be analyzed by comparing the model's expected behavior with its actual behavior. For this purpose, different sensitivity checks are suggested for the main parameters, based on several example cases. When testing one parameter, all other parameters are left at a similar level (ceteris paribus). Each result will be compared with the expected outcome.

Overall, the parameters for the model can be clustered into five different groups. First, eight parameters need to be set before the model can be initialized via a sampling and conversion of the Netica equations into the conditional probability tables. The remaining eleven parameters can be set within the initialized model. Three of them are used to set the location, i.e., the context, for the supplier location currently being analyzed. A third group defines the most recent audit taken at the supplier location using two parameters. A fourth group of parameters is used to incorporate observations. Finally, a single parameter specifies whether the supplier has signed a "code of conduct" and another the degree to which the prior is influenced by a signed "code of conduct".

6.2.2 Evaluation results

The evaluation results will be presented in three steps. First, general requirements are discussed before the results from the system-specific questionnaire are presented. Finally, the results for the sensitivity tests are detailed.

6.2.2.1 General requirements

In the following, the degree of fulfilling the general propositions for a sustainability risk management system proposed above will be discussed. Table 65 provides a summary of these propositions, which will be referenced below.

No.	Proposition
1	Identifying the most important data sources is a key task in the development of an SRMS.
2	An SRMS has to provide the means to integrate both internal sustainability data, collected manually or automated in various source systems, as well as external sustainability data, derived from external platforms, media channels, or from suppliers and partners.
3	An SRMS has to be able to integrate and dynamically adapt risk assessment and sustainability reporting frameworks.
4	An SRMS should allow for the definition and continuous monitoring of sustainability indicators for company-specific risk assessment. The system should support customized aggregation rules, calculation procedures, and thresholds in order to account for individual risk analysis and evaluation.
5	An SRMS should support risk ranking and assessment procedures which can be integrated into existing business processes.
6	An SRMS should detect and track sustainability incidents and alert decision makers.
7	An SRMS should be able to generate reports for both internal and external use allowing for stakeholder-specific integration into the risk management process.

Table 65: Summary of propositions for requirements of a sustainability risk management system (author's representation)

Proposition (1) was discussed in a dedicated part of this thesis. It was shown that identifying appropriate sources is indeed a problem, but one that can be overcome with tailored approaches (see chapter 7). Proposition (2) is reflected in the model's structure of including both internal audit data and incidents derived from external sources. This may be extended to further sources in future work. Proposition (3) is particularly relevant for a system that aims to generally capture the current notion of the shared understanding of the term "sustainability". While the system described here assumes that child labor is a sustainability problem, it predominantly includes data into the risk model that is pre-filtered based on its adherence to the definition of a child labor incident included here.[140] A user controlling the system can delete reports that do not fit the current perception of child labor if it is possible to detect this from a message's context. Consequently, the proposition can be seen as partly fulfilled. Proposition (4) asks for a continuous monitoring of sustainability indicators, which is indirectly fulfilled by this system, as supplier location rankings are ongoingly produced. Aggregation rules allow combined sustainability indicators to be displayed. Given that this system aims to produce an ordering of supplier locations and focuses on one aspect of social sustainability, this proposition is less relevant at the specialized level concerned here. Proposition (5) is directly approached by the system's basic aim. By providing a risk-based ranking of supplier locations,

[140] Child labor statistics are based on a predefined understanding of sustainability. Audit standards typically build on these standards.

the resulting ordering can be used to prioritize auditing activities or ad hoc auditing considerations, among others. Similarly, proposition (6) is a core aim of the system, via the integration of external and public news sources. The primary report generated by the system is a ranking of supplier locations. Nevertheless, features of the system can be used to create further management reports. This could be a report stating the absolute risk levels of supplier locations, summaries for suppliers, geographic regions, or sectors, as well as extracts from the included external evidence itself in the form of a briefing. Consequently, more detailed reports could be a further step to enhance the system and fulfill proposition (7) even more. Overall, the system fulfills all seven procedures to a certain extent, focused on the specialized case of child labor risk.

6.2.2.2 Model questionnaire

6.2.2.2.1 Response rate and final sample

In total 59 tokens (i.e., invitations) were sent to different contacts. On average, the survey was answered by 1.6 contacts per token used. In total, 30 complete responses were received (see Figure 59). The average completion time was 24 minutes – this includes waiting times if users were performing other tasks while filling in the survey (the maximum time taken was 54 minutes).[141]

The 30 responses displayed demographic characteristics that led to the exclusion of two datasets before starting the final analysis. The respondents of these two datasets did not state that they had experience in supply chain, risk, or sustainability management. Consequently, they were not considered experts. Of the remaining 28 answers, five were from experts at companies that do not source from outside the EU. While this may have an effect, the key is the experts and their knowledge. Only one expert in the sample of 28 works in a company without customers in the EU. Figure 60 shows this differentiation. In the following, all demographic statistics will be based on the 28 remaining responses.

[141] For raw data on the usefulness dimension see “Appendix T: Raw data of expert questionnaire (usefulness)”.

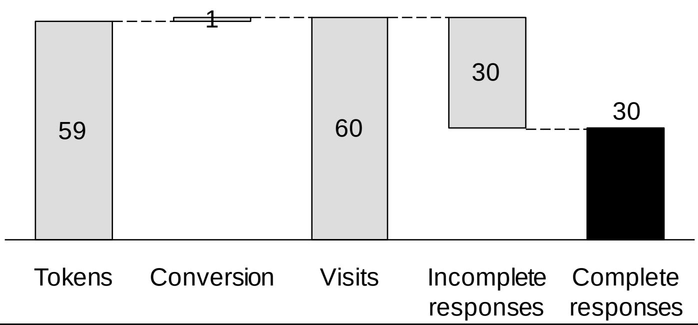


Figure 59: Tokens, visits, and responses for expert survey (author's representation)

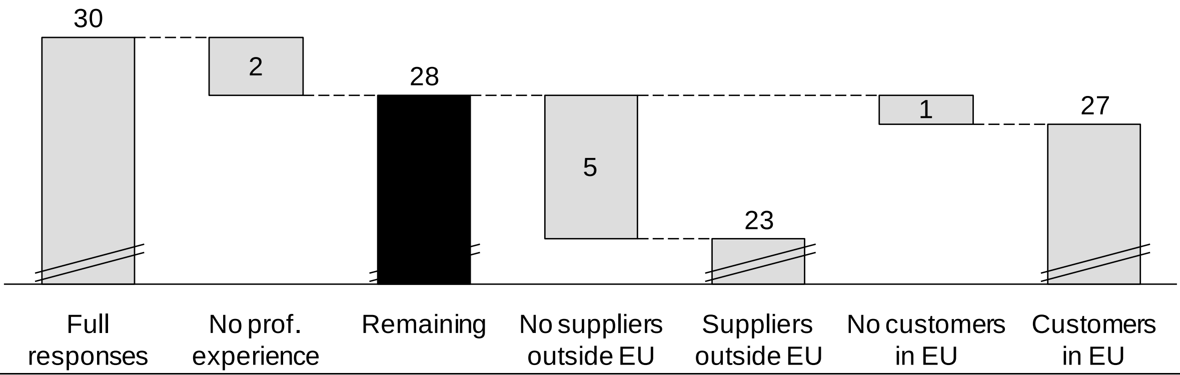
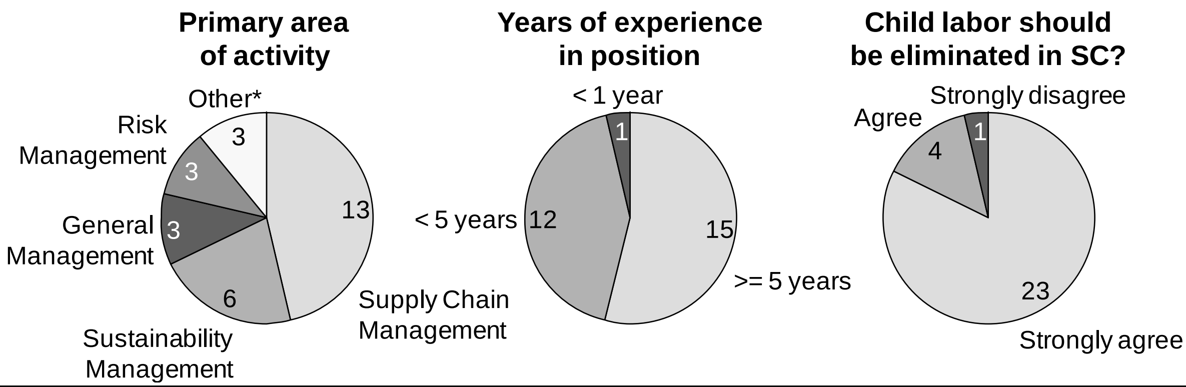


Figure 60: Demographics of responses to expert survey (author's representation)

As depicted in Figure 61, most of the experts have an SCM background and to a lesser extent are from sustainability, general, and risk management. Most (15 responses) of the experts have more than five years of experience in their position while another twelve have between one and five years. Consequently, the responses are based on a significant amount of expertise. Nearly all experts agree that child labor should be eliminated from the supply chain (SC; 23 strongly agree and another four agree). Only one expert disagrees with this – his company is the only company without customers in the European Union. Most experts have their workplace in Austria or Germany. Nevertheless, respondents from other countries are included in the result set (Figure 62).



* Compliance Management, Human Resource Management, Procurement

Figure 61: Demographics of experts responding to survey (n=28; author's representation)

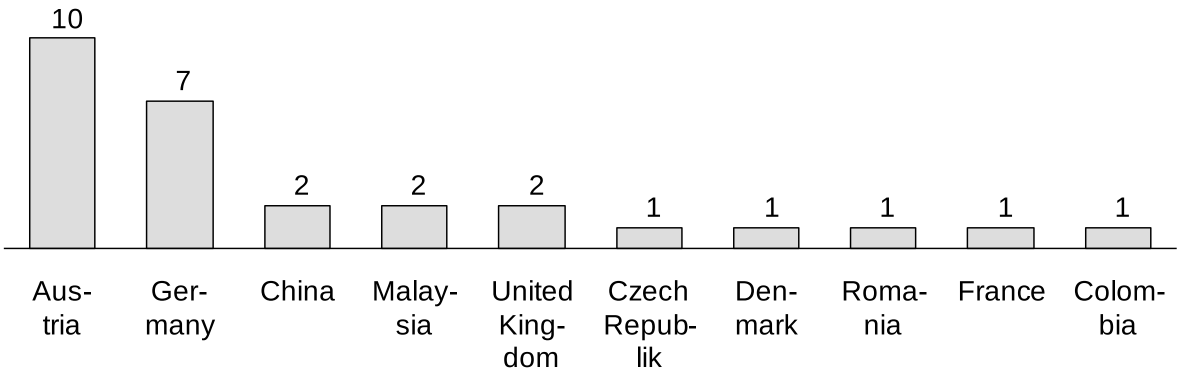


Figure 62: Number of responses per main location of experts' workplaces (n=28; author's representation)

The companies the experts work for are from a larger variety of industries, albeit with a significant spike in manufacturing and wholesale/retail trade (18 responses). These are industries with a potentially higher exposure to child labor. Most companies (23 responses) have more than 1,000 employees (Figure 63).

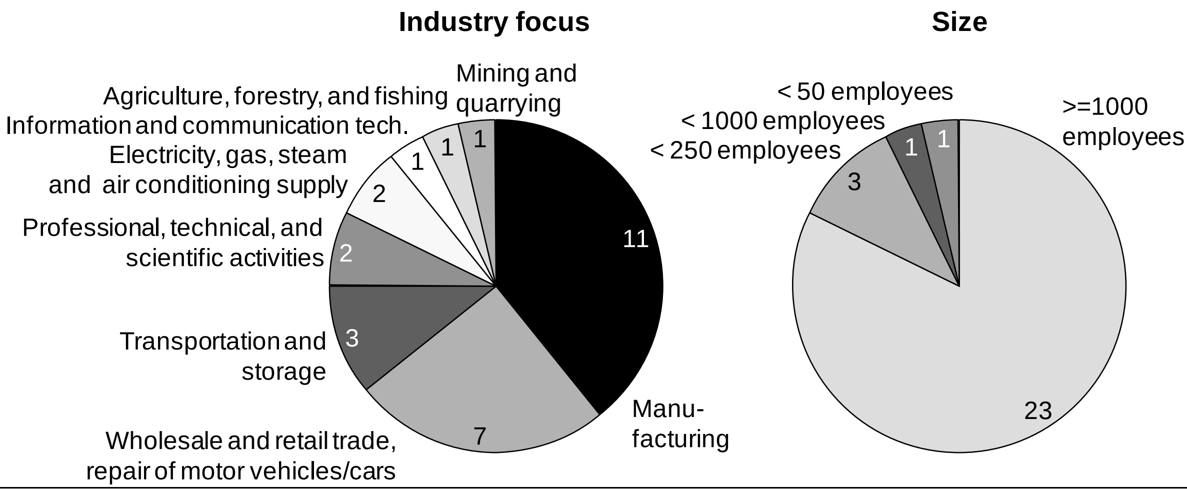


Figure 63: Demographics of companies that experts work for (n=28; author's representation)

When asked whether experts preferred precision or recall[142], two third preferred precision.

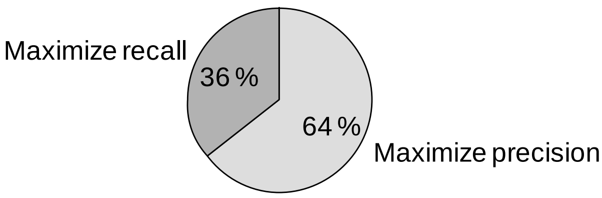


Figure 64: Results for precision and recall preference of experts responding to survey (n=28; own representation)

[142] The following descriptions were used. Precision: focus on reducing the risk of including irrelevant data with the risk of missing news about an incident. Recall: do not miss any potential news about a possibly relevant sustainability incident with the risk of including irrelevant data.

6.2.2.2.2 Job relevance and result demonstrability results

As can be seen in Table 70, with regard to hypothesis H1b, experts agree with all except one requirement at a 0.05 significance level when tested with the Wilcoxon test.[143] The remaining requirement shows agreement for the median, and no significant difference from a neutral position can be seen. However, the test for differing from a disagreeing value of 3 is significant.

Nr.	Requirement	Mode (group)	Median	Z*	Sig. (2-tailed)*
1	Use of public statistical data sources (e.g. child labor statistics)in order to calculate an initial incident probability for a supplier location	5.8	6	-3.774	.000
2	Differentiate initial probability by country, sector and rural/urban	5.8	6	-4.258	.000
3	Automated calculation of the initial probability (i.e. prior probability) of having at least one child working at a random company	5.1	5	-3.434	.001
4	Automated use of publicly available news reports on child labor incidents (e.g. via internet) in order to update the initial probability	5.6	6	-4.222	.000
5	Automated use of publicly available non-news reports on child labor incidents (e.g. from social media sites as Twitter)in order to update the initial probability	4.8	5	-2.614	.009
6	Adjustment of initial probability depending on the specifity in which an incident report refers to a specific supplier location (i.e. relevance)	5.3	5	-3.078	.002
7	Adjustment of initial probability depending on the credibility of the channel through which an incident report has been received (e.g. news or social media)	5.7	6	-3.873	.000
8	Automated update of initially calculated probability without user-involvement (i.e. unbiased)	4.6	5	-1.574	.115
9	Use of last auditing result to update the initial probability	6.2	6	-4.192	.000
10	Use of whether Code of Conduct has been signed to adjust the initial probability values	4.9	5	-2.241	.025
11	Risk model calculates the probability for the risk of a breach of child labor standards for each supplier locations	5.2	5	-2.688	.007
12	System output (risk rank/score) should be integrateable into further supplier evaluation processes	6.1	6	-3.527	.000
13	System output (risk rank/score) should be unbiased with regard to different suppliers	5.7	6	-3.61	.000

* Test of significant difference from 4 (neutral) - Wilcoxon-Test

Table 66: Average agreement levels for requirements stated in questionnaire (n=28; author's representation)

Hypothesis H1a was tested for each requirement against the job relevance using Spearman's correlation (Spearman Rho) for non-parametrical correlations. A significant correlation was assumed for all requirements. If an expert attributes job relevance to the system, the requirements are also seen as important (vice versa).[144] However, as shown in Table 67, this is only true for six of the 13 requirements.

[143] The Kolmogorov-Smirnov test shows that most responses to the different requirements (apart from two at 0.05 sig. level) do not follow a normal distribution and, thus, non-parametrical tests need to be performed.

[144] In the unlikely case that all experts fully agree with the requirements and the job relevance, then no correlation could be calculated. Nevertheless, all variables show variance, with a minimum response of maximum three (only for one requirement – others with one or two) and a maximum of seven.

		Job relevance		Extended job relevance	
Nr.	Requirement	r	Sig. (2-tailed)	r	Sig. (2-tailed)
1	Use of public statistical data sources (e.g. child labor statistics) in order to calculate an initial incident probability for a supplier location	.426*	.024	.372	.051
2	Differentiate initial probability by country, sector and rural/urban	.614**	.001	.599**	.001
3	Automated calculation of the initial probability (i.e. prior probability) of having at least one child working at a random company	.400*	.035	.449*	.017
4	Automated use of publicly available news reports on child labor incidents (e.g. via internet) in order to update the initial probability	.375*	.049	.376*	.049
5	Automated use of publicly available non-news reports on child labor incidents (e.g. from social media sites as Twitter) in order to update the initial probability	.339	.078	.314	.104
6	Adjustment of initial probability depending on the specifity in which an incident report refers to a specific supplier location (i.e. relevance)	.299	.122	.193	.326
7	Adjustment of initial probability depending on the credibility of the channel through which an incident report has been received (e.g. news or social media)	.241	.216	.122	.536
8	Automated update of initially calculated probability without user-involvement (i.e. unbiased)	.611**	.001	.480**	.010
9	Use of last auditing result to update the initial probability	.038	.849	.212	.279
10	Use of whether Code of Conduct has been signed to adjust the initial probability values	0.346	.071	.393*	.038
11	Risk model calculates the probability for the risk of a breach of child labor standards for each supplier locations	.310	.109	.251	.198
12	System output (risk rank/score) should be integrateable into further supplier evaluation processes	.236	.227	.263	.177
13	System output (risk rank/score) should be unbiased with regard to different suppliers	.456*	.015	.307	.112

Note: * Significant at 5%-level; ** Significant at 1%-level

Table 67: Spearman correlations with job relevance for requirements (n=28; author's representation)

Given that a correlation is not supported for all requirements, a differentiation between three cases is suggested: (1) requirements with significant agreement and no significant correlation, (2) requirements with significant agreement and significant correlation, and (3) requirements with no significant agreement. Table 68 provides an overview of these cases.

Agreement	Correlation	Requirement
Significant	Significant	Use of public statistical data sources (e.g. child labor statistics) in order to calculate an initial incident probability for a supplier location
		Differentiate initial probability by country, sector and rural/urban
		Automated calculation of the initial probability (i.e. prior probability) of having at least one child working at a random company
		Automated use of publicly available news reports on child labor incidents (e.g. via internet) in order to update the initial probability
		System output (risk rank/score) should be unbiased with regard to different suppliers
	Not significant	Automated use of publicly available non-news reports on child labor incidents (e.g. from social media sites as Twitter) in order to update the initial probability
		Adjustment of initial probability depending on the specifity in which an incident report refers to a specific supplier location (i.e. relevance)
		Adjustment of initial probability depending on the credibility of the channel through which an incident report has been received (e.g. news or social media)
		Use of last auditing result to update the initial probability
		Use of whether Code of Conduct has been signed to adjust the initial probability values
		Risk model calculates the probability for the risk of a breach of child labor standards for each supplier locations
		System output (risk rank/score) should be integrateable into further supplier evaluation processes
Not significant	-	Automated update of initially calculated probability without user-involvement (i.e. unbiased)

Table 68: Categorization of requirements as seen by experts with regard to job relevance (author's representation)

Requirements falling into the first category can be seen as particularly important for those experts who also agree on the system's job relevance. These are the key differentiating requirements that have a strong connection with positive job relevance. Therefore, they could be called "promising requirements," given that only some of the experts see them as particularly useful, and they might not be necessarily needed in all contexts. Requirements in the second category are agreed upon as relevant for a child labor supply chain risk management system, but no connection with job relevance is visible. If such a system is implemented, they are probably seen as "hygiene" factors.

Table 69 shows the agreement levels for the items representing job relevance (hypothesis H1c). The first two reflect the "classical" job relevance measure and the four combined the extended version.[145] The test for difference from a neutral position is significant for all items except the first. The reduced significance for importance may be due to the specific focus on child labor. A broader system scope could reach a different level. The combined "classical" measure (using the mean) shows a significance level of 0.052 when compared to neutral and the extended measure of 0.002 for the Wilcoxon test. Taking only the SCM managers, it would be significant at the 0.049 level (n=13) and taking only companies with over 1,000 employees would also lead to a significance of 0.048 (n=23).

[145] Cronbach's alpha for the classical items is very high at 0.97. The extended measure still has an alpha of 0.89.

Question	Mode	Median	Z*	Sig. (2-tailed)*
In my job, usage of such a system is important	5	5	-1.734	.083
In my job, usage of such a system is relevant	5	5	-2.237	.025
In my job, usage of such a system would reduce the risk of supply chain surprises	5	5	-3.339	.001
In my job, usage of such a system would reduce the risk of biased supply chain decisions	6	5.5	-3.756	.000

* Test of significant difference from 4 (neutral) – Wilcoxon test

Table 69: Agreement levels for job relevance (n=28; author's representation)

Finally, result demonstrability (H3) was tested using a measure of four items (Table 70). For each individual item as well as for the mean (sig. of 0.000) the difference to neutral is highly significant.

Question	Mode	Median	Z*	Sig. (2-tailed)*
I have no difficulty telling others about the results of using such a system	7	6	-4.045	.000
I believe I could communicate to others the consequences of using such a system	6	6	-4.248	.000
The results of using such a system are apparent to me	6	6	-3.827	.000
I would have no difficulty explaining why using such a system may or may not be beneficial[146]	6	6	-3.071	.002

* Test of significant difference from 4 (neutral) – Wilcoxon test

Table 70: Agreement levels for result demonstrability (n=28; author's representation)

6.2.2.2.3 Expert strategies results

The comparison with expert strategies is performed in several steps. First the prior calculation is described, then judgments on observations are discussed followed by audits. Finally, the combination of all three is detailed.

Based on the data provided, the experts suggested an initial ranking that is in many ways comparable to the one created using the model. In order to better compare the two approaches, Figure 65 introduces a scaled measure. This measure was derived by norming the best-rated supplier to zero and the worst rated supplier to one. For the model, relative distances were then calculated using the mean prior values. The model was set up as related in the methodology chapter, using the values for the prior that correspond to the ones used in the questionnaire. The adjustments for the expert responses were calculated using a weighted score (response frequency times weight of one for best rank, four for worst rank) and taking the average. It must be noted that the experts' responses have not been interval scaled, so the ordering is important. Experts tend to view the location of supplier B as worse than the one for supplier

[146] Inverse of original question for comparison.

D when compared with the model under these assumptions. Nevertheless, in general the model prior and expert responses appear to have a comparable pattern.

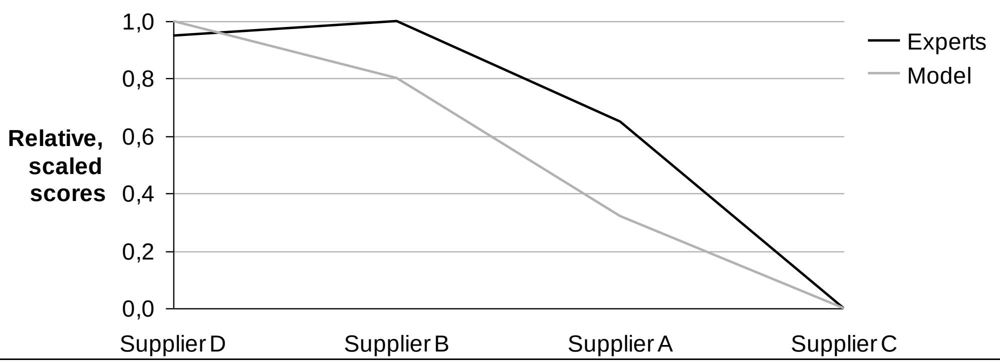


Figure 65: Consolidated ranking comparison between all experts and model (n=28; author's representation)

A more granular analysis of the result, however, shows that experts often strongly disagree in their judgment of the riskiness of the different suppliers (Figure 66). The rank selected by the majority of experts is only equivalent to the one calculated by the model for ranks three and four. Suppliers B and D show a particularly large spread of answers.[147]

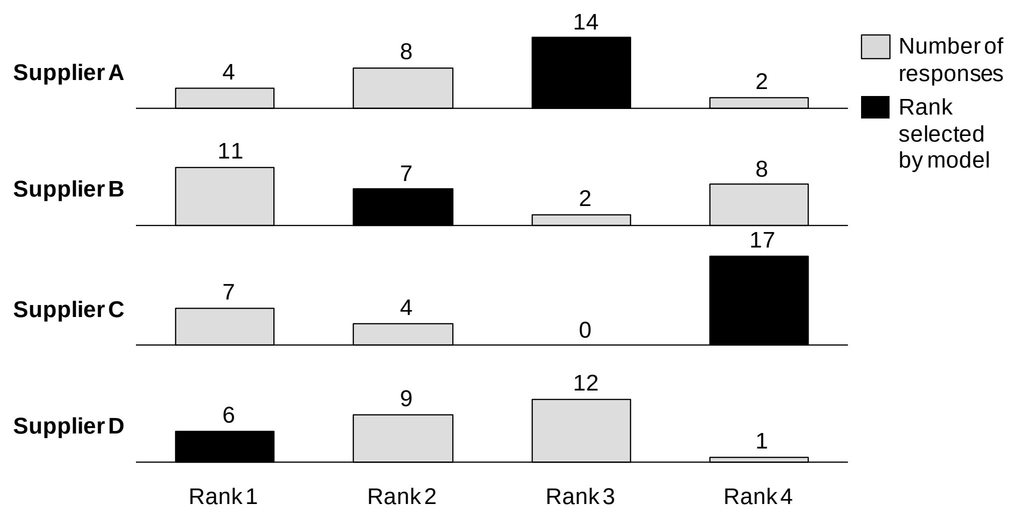
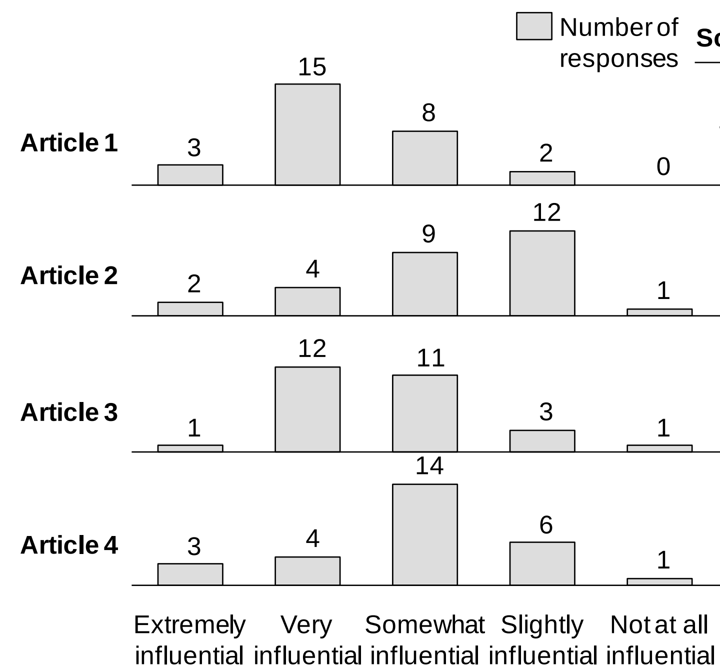


Figure 66: Number of responses per supplier and rank. Black shaded bars indicate rank selected by model (n=28; author's representation)

The most general result from the answers on the influence of articles on ranking decisions is that all the articles provided had at least a slight influence on the experts' decisions (see Figure 67). Building a score based on article frequencies and points for the different categories presents article one as the most influential, followed by articles three, four, and two. Comparing articles one and two, the additional geographic detail (region) is nearly as influential as

[147] One possibility could be that the question "rank according to risk" with the instruction of "highest on top" has been misunderstood by some respondents, thinking of "highest" as "best" and not "highest probability". Even if this misunderstanding occurred, differences in rankings would still be possible.

the explicit mentioning of the company. Hence, closely related geographic proximity drives relevance. This is not the case for the article obtained through social media (the worsening could also be due to the reference to a different sector).



	Score*	Geo-graphy	Com-pany	Sector	Channel
Article 1	103	Same-country	Same	Same	Classic news
Article 2	78	Same-country	None	None	Classic news
Article 3	93	Same-region	None	Same	Classic news
Article 4	86	Same-city	None	Other	Social media

* Calculation of score: Extremely influential – 5 points; very influential – 4 points; somewhat influential – 3 points; slightly influential – 2 points; not at all influential – 1 point.

Figure 67: Number of responses per article and influence selection (n=28; author's representation)

The relation between audit scores and average probability of an incident is judged to be very ambiguous, as can be seen in the high standard deviations of the figures provided (Figure 68 and Table 71). While some experts put a lot of trust in audit scores, others see only limited value.[148] Two suggestions can be derived from the extreme cases. Even if an audit attributes the best score to a supplier, experts tend to still see a certain probability of an incident (0.000 significance for different from zero). Similarly, the worst audit score does not necessarily indicate that child labor is present (0.000 significance for difference of 100% for Wilcoxon test).[149]

[148] This could also be due to different definitions of auditing or dependent on the organization performing an audit. Also the cases including audits in the calculation require a fine-grained definition of what is considered an influential audit and which effects the scores trigger.

[149] The Kolmogorov-Smirnov test shows that one part of the data does not follow a normal distribution. Consequently, the non-parametrical Wilcoxon-test has been applied.

Value	Worst score	Medium score	Best score
Average [percent]	57.4	29.3	13.0
Std. deviation	30.6	22.4	17.9

Table 71: Average estimated probability values (incl. standard deviation) of incident for different audit scores (n=25; author's representation)[150]

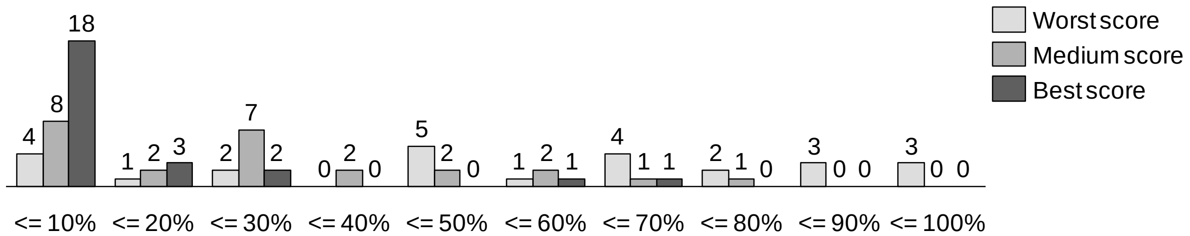


Figure 68: Number of responses for incident probability given a supplier audit score grouped by percentage categories (n=25; author's representation)[151]

Finally, an analysis of the values suggested for combining the probabilities derived from three data sources (i.e., weights; see Table 72) shows that audits are still seen as providing the most important source of information, being most frequently weighted highest (Figure 69). In contrast, statistics are seen as the least important for a ranking. All three are significantly different from zero (0.000 level).

Value	Probability based on statistical data (e.g., child labor statistics)	Probability based on audit data (e.g., last audit result/time)	Probability based on evidence from external reports (e.g., news)
Average [percent]	23.8	42.9	33.3
Std. Deviation	15.7	18.4	14.7

Table 72: Average estimated relative importance of independently found probability values for overall supplier risk judgment (n=28; author's representation)

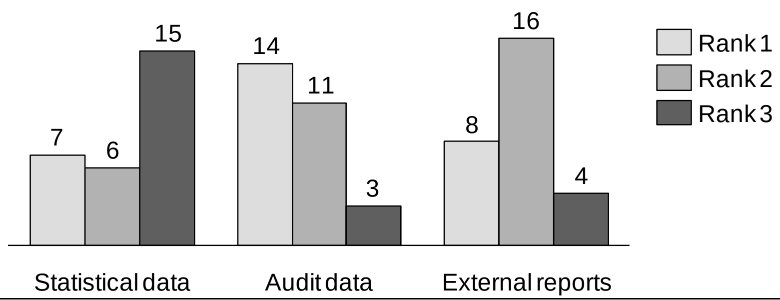


Figure 69: Frequency of ranks calculated from weights attributed to different sources of evidence (n=28; author's representation)[152]

150 For eight responses, the probability for the best audit score is worse than for the worst audit score. For these responses a mistake in understanding has been assumed, and the probability values for best and worst score have been swapped. Three responses have been excluded (one assigned all values the same number, one had the same score for the worst and the best, but a lower for the medium, and another had only the medium level probability high).

151 See footnote 150.

152 Two similar values have been counted as two times the better of the ranks.

6.2.2.2.4 Discussion and summary

The survey was answered by a broad range of experts in relevant fields. Nearly all requirements except full automation were seen as relevant for a child labor risk management system. Only the requirement stating that the risk model should be updated automatically was not significantly agreed upon. However, both mode and mean are above neutral in favor of agreement. For six of the 13 requirements only those experts who saw high job relevance for the system thought they were necessary. Consequently, they have been labeled as “promising requirements,” as they are only seen as partly useful. For the requirements without this correlation, implementing them is seen as a rather hygiene factor without a significant connection to job relevance. Overall, the job relevance of the system is slightly not significant at a 5% level (sig. of 0.052, while both median and mode are above neutral). Nevertheless, the system is seen as reducing supply chain surprises and biased supply chain decisions. Results demonstrability is seen as significantly met.

Furthermore, expert strategies can be reflected in the model. The mathematically calculated prior ranking is comparable to the average of expert users in the general tendency. Here, the mathematical score provides a more strict interpretation of the facts – experts on the individual level show strongly diverging responses. This variation needs to be explained through a follow-up study. All four articles checked were seen by experts as influential for determining a perceived probability value. They were also seen as influential by the model, provided the parameters were set accordingly. The system's assumption that articles more strongly related to a supplier location (e.g., by geographic proximity and specifity) as more important for the risk ranking appears in line with expert strategies. Additionally, experts see audits as not 100% trustworthy – a best or a worst score for an audit does not necessarily indicate 0% or 100% probability of an incident and consequently does not necessarily have to overrule all other facts known about a supplier. This is further strengthened by the combination strategies suggested by experts that rank audits as most important but also see external reports and statistics as necessary. The aggregated combination weights may be included in the model. Furthermore, score-dependent probabilities are realized in the model.

6.2.2.3 Sensitivity testing

The following subsections showcase the different parameters used throughout the sensitivity testing. For each tested sensitivity, the expected result before the test is presented at the beginning. Several model parameters were used, as defined in the methodology section. These “core parameters” have not been analyzed for their sensitivity. They are outlined in Table 73.

Prior sigma σ_prior	Audit gradient ϑ	Obs. scaling function pram τ	α for Obs sigma σ_obs	β for Obs sigma σ_obs	P_breach comb. param ϕ	P_breach comb. param χ	P_breach comb. param ψ
13.32	20	5	10.204	0.7653	0.6	0.2	0.2

Table 73: Overview of initialization parameters used for sensitivity tests (re-sampling of model necessary after change; author's representation)

An initial section explains the effect of the prior on the final result. For all other sensitivity tests, the same location parameters have been used in order to imitate the same supplier location (Table 74). The parameter for a signed "code of conduct" has been set to "true" for all test instances. Unless otherwise stated, variables are used with their prior value without findings.

Location Parameters		
Country	**Sector**	**Ruralness**
India	Manufacturing	Urban

Table 74: Overview of location parameters used for sensitivity tests (author's representation)

6.2.2.3.1 Prior

Expected result: Direct correlation between prior and breach likelihood

Table 75 shows the effect of different prior locations and the resulting prior likelihood on the model's breach likelihood.

Prior location-specific parameters			Mean likelihood	
Country	**Sector**	**Ruralness**	**Prior**	**Breach**
India	Manufacturing	Urban	3.9	14.4
		Rural	5.9	14.4
	Construction	Urban	23.6	15.1
		Rural	44.7	16.3
Indonesia	Manufacturing	Urban	4.9	14.4
		Rural	2.6	14.3
	Construction	Urban	0.0	14.3
		Rural	1.5	14.3

Table 75: Prior and breach likelihoods (percent) for different prior locations (author's representation)

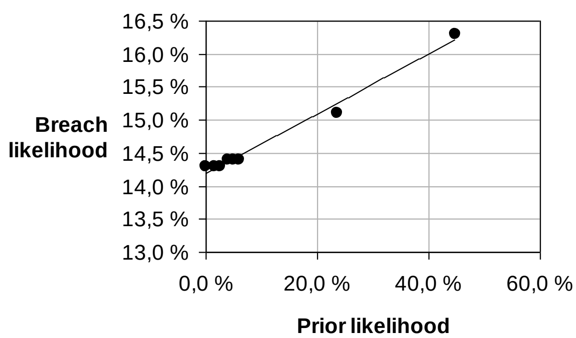


Figure 70: Correlation between breach and prior likelihoods on eight selected cases (author's representation)

Figure 70 gives the correlation between breach and prior probability for the selected cases (R^2=0.985). Due to the model's structure, it is assumed that the correlation also works for other locations.

6.2.2.3.2 Audit parameters

Expected results: The worse the audit score, the higher the breach likelihood; the longer since *the last audit, the higher the standard error of the breach likelihood*

The audit parameters were tested for their sensitivity with regard to the final breach probability in all different currently possible model constellations. Table 76 and Table 77 show the respective sensitivities of the mean and the standard error of the breach likelihood. The effects on the mean can be compared with the expected results.

Mean		Last audit score					
		4.5	3.5	2.5	1.5	0.5	Average
Time since last audit	3	11	14.8	18.5	22.3	26.1	18.5
	15	11.4	14.8	18.5	22.3	26.1	18.6
	33	12.5	15.4	18.8	22.3	26.1	19.0
	Average	11.6	15.0	18.6	22.3	26.1	

Table 76: Sensitivity of breach likelihood mean by audit parameters (author's representation)

Standard error		Last audit score					
		4.5	3.5	2.5	1.5	0.5	Average
Time since last audit	3	3.9	3.9	3.9	3.9	3.9	3.9
	15	5.1	5.3	5.4	5.4	5.4	5.3
	33	6.2	6.7	7.0	7.2	7.2	6.9
	Average	5.1	5.3	5.4	5.5	5.5	

Table 77: Sensitivity of breach likelihood variance by audit parameters (author's representation)

6.2.2.3.3 Observations

Expected result: Overall increasing mean breach likelihood with increasing number of child labor incident observations; stronger effect with higher credibility or relevance.

In order to showcase the effect of including news evidence in the model, two steps of tests were performed. First, the sensitivity was tested for only including one news item. Then, the sensitivity was tested for including two news items for selected cases where the first news item had already been included. News items were included by incorporating their respective credibility and relevance values with Netica. Table 78 and Table 79 show the sensitivity of the breach probability to entering a single observation with a certain credibility and relevance.

Mean		Credibility 0.1	0.3	0.5	0.7	0.9	Average
Relevance	0.1	12.8	13.1	13.5	13.8	14.2	13.5
	0.3	13.1	13.4	13.8	14.2	14.5	13.8
	0.5	13.5	13.8	14.1	14.5	14.9	14.2
	0.7	13.8	14.2	14.5	14.9	15.2	14.5
	0.9	14.2	14.5	14.9	15.2	15.6	14.9
	Average	13.5	13.8	14.2	14.5	14.9	

Table 78: Sensitivity of breach likelihood mean for entering one observation (author's representation)

Standard error		Credibility 0.1	0.3	0.5	0.7	0.9	Average
Relevance	0.1	4.5	4.5	4.6	4.6	4.7	4.6
	0.3	4.5	4.6	4.6	4.6	4.7	4.6
	0.5	4.6	4.6	4.6	4.7	4.7	4.6
	0.7	4.6	4.6	4.7	4.7	4.8	4.7
	0.9	4.7	4.7	4.7	4.8	4.8	4.7
	Average	4.6	4.6	4.6	4.7	4.7	

Table 79: Sensitivity of breach likelihood standard error for entering one observation (author's representation)

The tables indicate the exchangeable behavior of credibility and relevance in the current implementation and the reduced effect on the standard deviation. The negative effect on breach probability can be seen in both tables.

Table 80 and Table 81 build on a case that first incorporates an observation with the highest credibility and relevance set. Table 82 and Table 83 show the sensitivity for building on a first observation with the highest credibility of 0.9 and the lowest relevance of 0.1. In both cases, the means and standard errors shown in the table have been reached by incorporating the first observations. This value increases for all possible cases, as expected.

Mean		Credibility					
After 1st:	15.60	0.1	0.3	0.5	0.7	0.9	Average
Relevance	0.1	17.5	17.9	18.4	18.8	19.2	18.4
	0.3	17.9	18.4	18.8	19.2	19.6	18.8
	0.5	18.4	18.8	19.2	19.6	20.0	19.2
	0.7	18.8	19.2	19.6	20.0	20.4	19.6
	0.9	19.2	19.6	20.0	20.4	20.7	20.0
	Average	18.4	18.8	19.2	19.6	20.0	

Table 80: Sensitivity of breach likelihood mean for entering second observation (1st: Credibility=0.9; Relevance=0.9; author's representation)

Standard error		Credibility					
After 1st:	4.80	0.1	0.3	0.5	0.7	0.9	Average
Relevance	0.1	5.6	5.5	5.4	5.3	5.3	5.4
	0.3	5.5	5.3	5.2	5.2	5.2	5.3
	0.5	5.4	5.2	5.1	5.1	5.1	5.2
	0.7	5.3	5.2	5.1	5.0	5.0	5.1
	0.9	5.3	5.2	5.1	5.0	5.0	5.1
	Average	5.4	5.3	5.2	5.1	5.1	

Table 81: Sensitivity of breach likelihood standard error for entering second observation (1st: Credibility=0.9; Relevance=0.9; author's representation)

Mean		Credibility					
After 1st:	14.20	0.1	0.3	0.5	0.7	0.9	Average
Relevance	0.1	15.8	16.2	16.7	17.2	17.6	16.7
	0.3	16.2	16.7	17.2	17.6	18.0	17.1
	0.5	16.7	17.1	17.6	18.0	18.4	17.6
	0.7	17.1	17.5	18.0	18.4	18.8	18.0
	0.9	17.5	17.9	18.4	18.8	19.2	18.4
	Average	16.7	17.1	17.6	18.0	18.4	

Table 82: Sensitivity of breach likelihood mean for entering second observation (1st: Credibility=0.9; Relevance=0.1; author's representation)

Standard error		Credibility					
After 1st:	4.70	0.1	0.3	0.5	0.7	0.9	Average
Relevance	0.1	5.4	5.3	5.2	5.2	5.2	5.3
	0.3	5.4	5.3	5.2	5.1	5.2	5.2
	0.5	5.5	5.3	5.2	5.2	5.2	5.3
	0.7	5.5	5.4	15.3	5.2	5.2	7.3
	0.9	5.6	5.5	5.4	5.3	5.3	5.4
	Average	5.5	5.4	7.3	5.2	5.2	

Table 83: Sensitivity of breach likelihood standard error for entering second observation (1st: Credibility=0.9; Relevance=0.1; author's representation)

In both versions an increase in the overall breach likelihood can be seen when compared to the results in Table 78 and Table 79. Moreover, the higher relevance in the first case results in increased levels of overall breach probability. Finally, Figure 71 presents the effect of entering several observations with the same credibility and relevance level over the different steps. One line states the effects of entering observations with the lowest credibility and relevance level, one the effects of medium level observations, and one the results of highly relevant observations. As expected, the overall breach likelihood rises with the number of observations in all three cases.

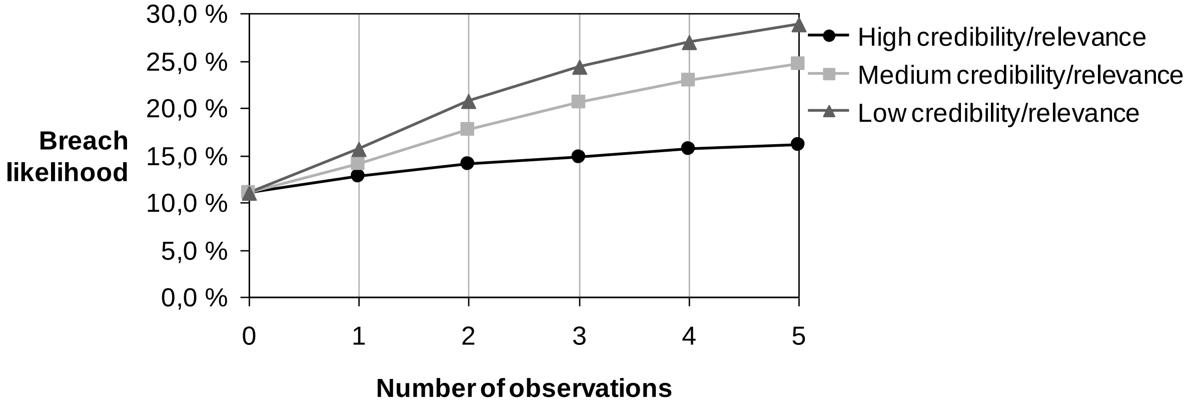


Figure 71: Effect of entering observations into the model (different credibility/relevance; high=0.9, medium=0.5, low=0.1; author's representation)

Overall, the effect of the observations on the breach likelihood are in line with the expected results stated above.

6.2.3 Risk model summary

In summary, the risk model was evaluated by combining an expert questionnaire with sensitivity tests. The expert survey showed that cognitive instrumental processes identified by the TAM2 are positively influenced by a system based on the suggested model. Nearly all key requirements outlining the model were seen as statistically relevant by the experts. Moreover, the job relevance of a system employing the model has only slightly been not "confirmed" at the 5% significance level (it would have been significant at the 6% level) in contrast to high result demonstrability. Expert strategies can also be reflected with the model. Finally, coherence and comparability were acknowledged via the sensitivity tests. Consequently, given positive social instrumental processes in an organizational context, the system's model appears useful and, thus, is likely to be adopted by experts if embedded into a final system.

6.3 Overall system

Finally, two additional evaluation steps were performed to test the two main parts of the system with practical use cases. These should showcase the effect of input on the system's behavior. On the one hand, a first test will use the same fictive four supplier locations of the expert questionnaire and depict the effects of four news messages on the ranking of these locations. The second test will show the observations generated by inputting all news texts related to either India or Indonesia from the child labor messages from the English Reuters corpus. For both tests, the overarching system implementation according to the specifications is used.[153] Furthermore, the ranking test uses a processing resource that automatically attaches pre-specified dimension values to fictive news input.

6.3.1 Ranking test

The ranking test reuses the four supplier locations as used in the expert questionnaire discussed above (Table 62). Table 84 depicts the four news items used to showcase the system's behavior. Furthermore, fictional last audit results were inputted as stated in Table 85.

News ID	Timestamp	Max. relevance	Credibility	Lowest geography	Supplier	Sector
News 1	01.09.2014	0.7	0.9	India	Supplier B	Manufacturing
News 2	01.09.2014	0.1	0.9	Indonesia		
News 3	01.09.2014	0.5	0.9	Provinsi Jawa Tengah		Construction
News 4	04.09.2014	0.7	0.3	Tirupati		Manufacturing

Table 84: Dimensions of news sources used as ranking test input (author's representation)

Supplier	Audit score (best score: 5.0)	Time since audit (categories acc. to BN)
Supplier A	4.5	6-24 months
Supplier B	4.5	> 24 months
Supplier C	3.5	6-24 months
Supplier D	4.5	> 24 months

Table 85: Audit scores used as ranking test input (author's representation)

The combination of the static and dynamic news data leads to the observations in the system shown in Table 86. Given that the third news item details the second, the update leads to an increase in its maximum relevance. Although they have a good audit result, inputting the new news sources causes the rank of supplier B to worsen twofold (as can be seen in Figure 72). At the same time, the ranking of supplier D improves to first place after the third news item.[154] Before the audit results and news are entered, the initial differences between supplier locations are very small.[155] Overall, the rankings generated by the configuration of the BN are strongly affected by audit scores and news input. Consequently, in organizational contexts, the strength of the effect needs further discussion.

---

[153] For Open Geonames, the LOD used during processing is based on Yago version 2.5.3 with a Wikipedia dump from 2012-12-01 and accessed via the Sparql endpoint at http://lod2.openlinksw.com/sparql.
[154] This is also due to the configuration that no same-country-only spillover effects are included.
[155] Small differences between breach likelihoods result from the discretization used in the BN to model probability distributions. Thus, changes that should only affect the variance can also affect the mean of the final breach likelihood. This effect may be reduced by a finer grained discretization of continuous nodes.

Observation ID	Timestamp	Max. relevance	Credibility	Lowest geography	Supplier	Sector
Obs_2015-01-03T23:01:32.594	01.09.2014 - 15:38:56	0.7	0.9	Indonesia	SupplierB	Manufacturing
Obs_2015-01-03T23:01:40.044	01.09.2014 - 16:38:56	0.5	0.9	Provinsi Jawa Tengah		Construction
Obs_2015-01-03T23:01:45.477	04.09.2014 - 18:38:56	0.7	0.3	Tirupati		Manufacturing

Table 86: Observations derived by system based on news input (author's representation)

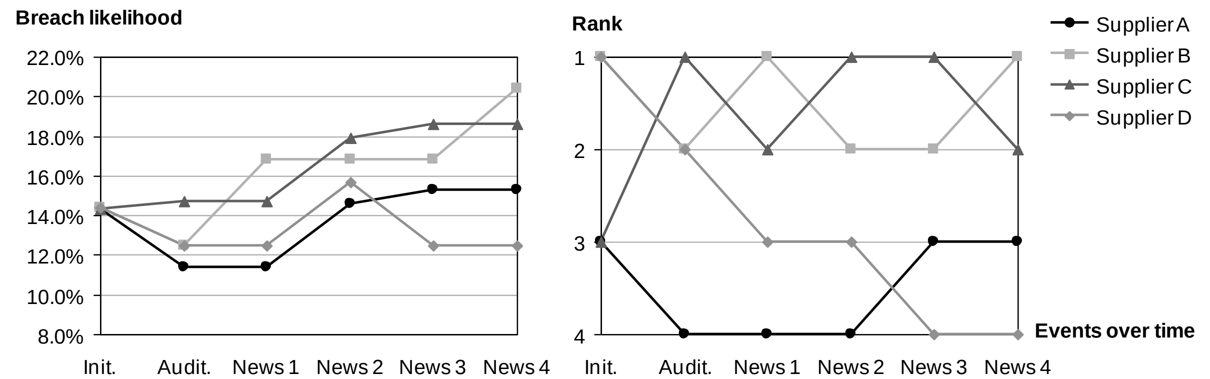


Figure 72: Development of breach likelihood and supplier rank after audit and news inputs (author's representation)

6.3.2 Observations test

The second test uses the 19 news articles from the set of 117 true positive child labor articles in the Reuters dataset (already applied for the TM evaluation) related to either India or Indonesia. Given the low precision and recall results for the sector dimension, this dimension has been excluded from the discussion. In the following, several tables outline the showcase and the results. First, the expected set of observations from the manual annotation is shown in Table 87. Second, the results for two different system cases are presented. The best performing EE approach (Case 1) has been chosen based on the F1 score with doubles and summaries (calculated with GS restriction and full hierarchy).[156] And given the large number of observations in the first run, the EE approach with the best precision (Case 2) also acknowledging empty sets (without GS restriction) with doubles and summaries has been used. For both, the score is based on the best performing approach for the geography and company dimensions.[157] Finally, Table 90 compares the gold standard with the results from the two cases.

[156] This approach produces similar results to the best F1 approach (with gold standard restriction and full hierarchy) without doubles and without summaries.

[157] Parameter configuration for the PRs for the two approaches:
Case (1): Sector Tagging PR [nrOfWordsForMatches=4;useSpecification=false;wordPartMatches=false]; Ontology Alignment PR [maxTypesConsidered=2;Spotlight=Statistical]; Filter Context PR [filterMode=SentenceToCLRelation;minRelationOnly=false;filterAllCapitalSurfaceForms=true;filterToLowestGeography=false;filterToLowestSector=false;];
Case (2): Sector Tagging PR [nrOfWordsForMatches=4;useSpecification=false;wordPartMatches=false]; Ontology Alignment PR [maxTypesConsidered=2;Spotlight=Statistical]; Filter Context PR [filterMode=DistanceToCLRelation;minRelationOnly=false;filterAllCapitalSurfaceForms=true;filterToLowestGeography=false;filterToLowestSector=false;maximumDistance=100]

Observation ID	Timestamp	Max. relevance	Credibility	Lowest geography	Company
Observation 1	02.06.2008 - 06:46:37	0.1	0.9	India	
Observation 2	17.06.2008 - 02:16:42	0.9	0.9	Tirupur	Primark
Observation 3	17.06.2008 - 02:16:42	0.9	0.9	Tirupur	Associated British Foods
Observation 4	09.09.2008 - 10:26:57	0.7	0.9	India	Monsanto
Observation 5	16.10.2008 - 06:09:04	0.3	0.9	Belitung	
Observation 6	23.10.2008 - 04:07:40	0.5	0.9	Deeg	
Observation 7	23.10.2008 - 09:15:02	0.5	0.9	New Delhi	
Observation 8	10.02.2009 - 02:49:12	0.1	0.9	Indonesia	
Observation 9	23.02.2009 - 17:41:50	0.7	0.9	India	Ikea

Table 87: Manual gold standard of observations based on processing rules and dimensions gold standard (shaded results are correctly or partly correct identified observations in test case (2); author's representation)

Observation ID	Timestamp	Max. relevance	Credibility	Lowest geography	Company
Observation 1	02.06.2008- 06:46:37	0.9	0.9	Kolkata	International Labour Organization
Observation 2	02.06.2008 - 06:46:37	0.9	0.9	Kolkata	Centre for Communication and Development
Observation 3	17.06.2008 - 02:16:42	0.7	0.9	India	The Daily Telegraph
Observation 4	09.09.2008 - 17:15:34	0.7	0.9	Norway	Rio Tinto
Observation 5	09.09.2008 - 17:15:34	0.7	0.9	India	Rio Tinto
Observation 6	09.09.2008 - 17:15:34	0.7	0.9	Norway	Ethics Council
Observation 7	09.09.2008 - 17:15:34	0.7	0.9	India	Ethics Council
Observation 8	09.09.2008 - 17:15:34	0.7	0.9	Norway	Monsanto
Observation 9	09.09.2008 - 17:15:34	0.7	0.9	India	Monsanto
Observation 10	09.09.2008 - 17:15:34	0.7	0.9	Norway	FUND KEEPS MONSANTO Norway
Observation 11	09.09.2008 - 17:15:34	0.7	0.9	India	FUND KEEPS MONSANTO Norway
Observation 12	09.09.2008 - 17:15:34	0.7	0.9	Norway	Reuters
Observation 13	09.09.2008 - 17:15:34	0.7	0.9	India	Reuters
Observation 14	09.09.2008 - 10:26:57	0.7	0.9	India	Monsanto Co
Observation 15	16.10.2008 - 06:09:04	0.7	0.9	Indonesia	National Education Department
Observation 16	23.10.2008 - 04:07:40	0.9	0.9	Dīg	Reuters
Observation 17	23.10.2008 - 09:15:02	0.1	0.9	Burma	
Observation 18	23.10.2008 - 10:09:36	0.1	0.9	China	
Observation 19	23.02.2009 - 17:41:50	0.7	0.9	India	Ikea

Table 88: Observations derived by system based on news input; case (1): best F1 for geography and company dimension with GS restriction (author's representation)

Observation ID	Timestamp	Max. relevance	Credibility	Lowest geography	Company
Observation 1	17.06.2008 - 02:16:42	0.5	0.9		Associated British Foods
Observation 2	09.09.2008 - 10:26:57	0.7	0.9	India	Monsanto Co
Observation 3	09.09.2008 - 17:15:34	0.7	0.9	India	Monsanto
Observation 4	23.10.2008 - 04:07:40	0.1	0.9	India	
Observation 5	23.02.2009 - 17:41:50	0.7	0.9	India	Ikea

Table 89: Observations derived by system based on news input; case (2): best precision for geography and company dimension without GS restriction (author's representation)

Observation ID (Table 87)	Case (1)		Case (2)	
	Eval.	Comparison comment	Eval.	Comparison comment
Observation 1	Not similar	Use of too detailed dimensions not related to child labor incident	Not similar	Observation not identified
Observation 2 Observation 3	Not similar	Wrong dimensions	Partly similar	Only one of two companies without geography identified
Observation 4	Partly similar	Observation correctly created, but also additional wrong dimensions found in subsequent message (therefore not folded)	Partly similar	Company once correctly and once only partly identified, however, wrong spelling due to name-based and not URI-based approach (therefore not folded)
Observation 5	Partly similar	Correctly indentified country without detecting city and wrongly integrating organization	Not similar	Observation not identified
Observation 6	Not similar	Wrong dimensions	Partly similar	Only country identified
Observation 7	Not similar	Wrong dimensions	Not similar	Observation not identified/folded
Observation 8	Not similar	Not identified, however, other observation wrongly integrated	Not similar	Observation not identified
Observation 9	Similar	Observation correctly created	Similar	Observation correctly created

Table 90: Comparison of observations identified with manual gold standard (author's representation)

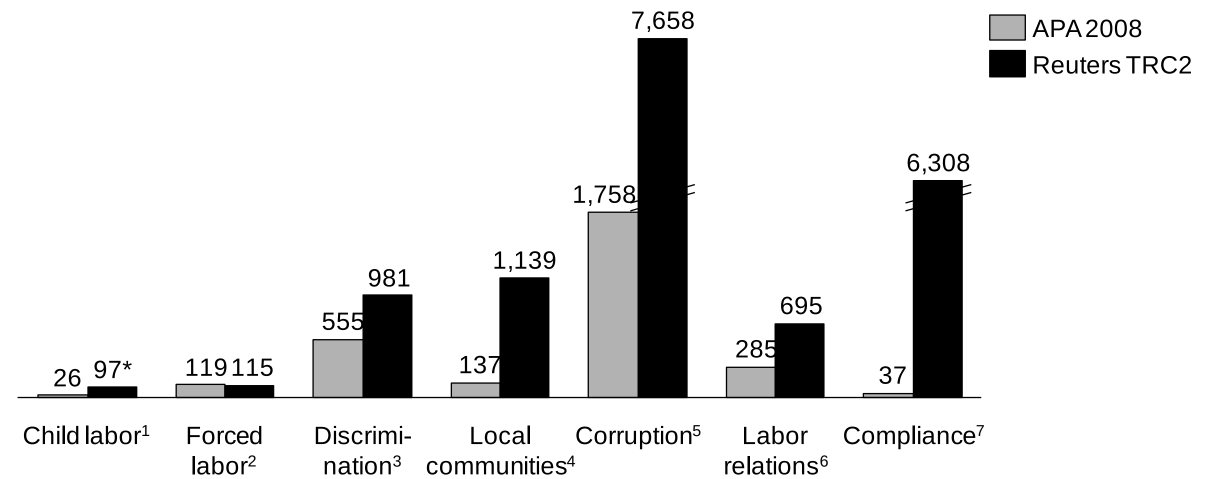
As can be seen in Table 90, both approaches have strengths and weaknesses. Whereas Case 1 produces a large number of wrong observations for the third observation, some observations are missed by the approach used for Case 2. Comparing the two, Case 2 appears more useful, given that the similar cases are found in nearly the same quality as in Case 1, while not including too many incorrect ones. This is also in line with the expert statement discussed in the risk model evaluation, which favored precision over recall (see Figure 64).

6.4 Summary

Altogether, this chapter presented an evaluation of the TM approach and the risk model. Two showcases were then detailed to show the effect of the components working together in the overall system. For the TM, classification performance can be improved above baseline level by applying the domain-specific approaches to newly developed test corpuses. While these performance values for ED suggest a significant reduction in manual work if applied within a company's workflow, the results for EE are less promising. The heuristics show high recall values, whereas precision values appear strongly inferior to the ones in the ED step. This is especially true for sector tagging. The risk model was evaluated using an expert questionnaire. Results appear in line with the assumptions used to design the risk model, leading to the conclusion that, embedded in a fully-working system, it would be practically adopted. Problems in the showcases can be traced back to the limited performance of EE. The next chapter will discuss these results in more detail.

7 Input data evaluation

A key criterion for the usefulness of the system is the availability of sufficient input data. For this thesis this implies having data sources that publish child labor incidents. The sources are better the more granular incident events are reported. The evaluation is particularly essential because child labor is one of the less reported topics within social sustainability (Figure 73).



Note: Results in general based on Mimir indexing; Exemption: * Result based on GATE tagging
APA search strings (in German): 1: Kinderarbeit+Kinder arbeiten; 2: Zwangsarbeit+Sweatshop; 3: Diskriminierung+diskriminieren; 4: lokale communities+Eingebohrene+ Ureinwohner; 5: Korruption; 6: Gewerkschaftsbeziehungen+gerechter Lohn+gerechte Entlohnung+Mindestgehalt/-bezhalung/-entlohnung; 7: Compliance+ Gesetzestreue
Reuters search strings: 1: child labo(u)r; incl. child work for APA2008; 2: forced/compulsory labo(u)r+sweatshop; 3: discrimination; 4: local communities+indigenous; 5: corruption; 6: labo(u)r relations+Minimum/fair wage/compensation; 7: compliance

Figure 73: Number of document hits when searching for different social sustainability categories in closed corpuses used for evaluation (author's representation)

As a consequence and given the scarcity of child labor issues described in a single news source visible in the TRC2 corpus, two additional sources were tested reusing the algorithms developed. These sources showcase the public availability of data that can be fed into the system. Both are online news sources that generally have three advantages (Atkinson and Piskorski 2011, 749):

- They are often the only source offering this level of information granularity
- Online news is faster than "traditional" or "official" channels
- Online news can provide a means to double-check other texts from different channels

This analysis uses more classic online news directly and also builds on texts derived from social media pages. Social media has been acknowledged as a source with high publication speed and a channel which eschews classical news publication paths (Anastasi et al. 2013; Sakaki, Okazaki, and Matsuo 2010). Twitter[158] has repeatedly been highlighted in this context (Du et al. 2013; Krstajic et al. 2012; Sankaranarayanan et al. 2009). The following sections provide details on the datasets and the results from applying the suggested algorithms on them.

7.1 Datasets creation and descriptive statistics

A full collection of all potentially relevant data for the purpose of analyzing child labor incidents is impossible for economic reasons. Consequently, a selection must be chosen based on reasonable heuristics. Similar restrictions have already been identified in earlier studies on risk management and in other contexts (LaComb, Interrante, and Aggour 2007). Hence, both strategies presented here offer heuristics for tackling the immense load of information available on the web.

The two datasets under discussion can be differentiated along two dimensions. First, one dataset is composed primarily from news sources while the other dataset is based on posts by NGOs. Second, most articles in the news dataset were gathered through a search-based (IR) approach (i.e., more "bottom-up") while the NGO dataset is composed of links provided by a pre-defined set of organizations (i.e., more "top-down"). In the following we will refer to the two datasets as the "news dataset" and "NGO dataset."

A detailed description of how the datasets were created will be presented for scientific rigor. This should help practitioners to follow a similar pattern in their attempts to monitor the web. To present the results and overcome time-consuming manual annotation (Y. Li, Bontcheva, and Cunningham 2005a, I), we reuse the approaches outlined in the chapter on text mining and process the original articles using the most appropriate algorithms for each category.

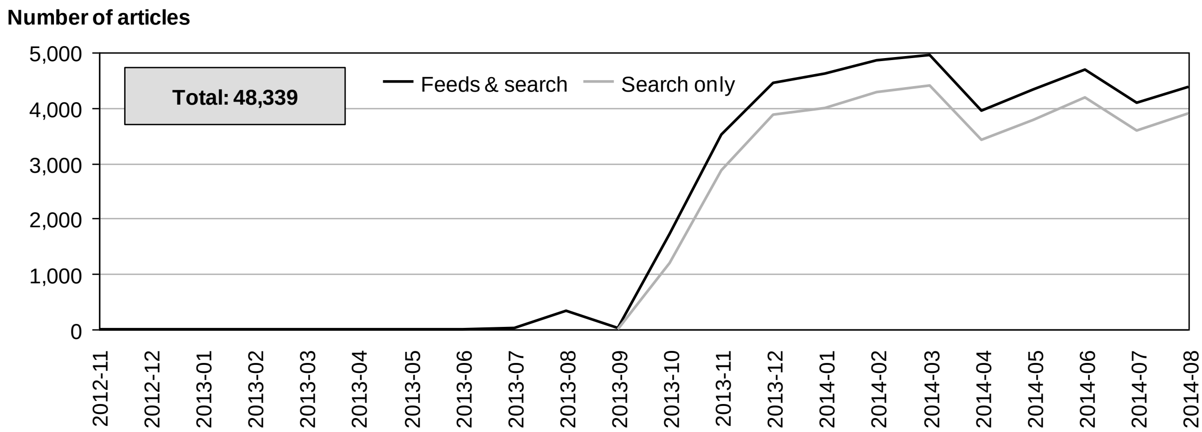
7.1.1 News dataset

News is the classic medium for transmitting current information on relevant topics. In recent year, online news has become important, especially in terms of news consumption (Ofcom

[158] http://www.twitter.com, an online social networking service.

2014, 2; Lomas 2014). This success is probably driven by the increased availability of online news. Online news has the additional advantage that it can be processed automatically.

Consequently, the first approach in this chapter uses online news channels to gather news on child labor incidents. To avoid relying on a self-developed list of news sources, aggregated news searches which retrieve online news from a wide variety of channels are used. Hence, news was gathered through searches on Google News[159], the European Media Monitor[160], and two selected RSS[161] feeds. Using these sources resulted in a very broad coverage of news and allowed for a more sophisticated statement about the availability of reports than would be possible using only closed datasets, which are often restricted to one source. Each of these channels provides an RSS-formatted response that was automatically collected on a daily basis.[162] Only incrementally new articles were added each day.[163] Additional search terms were added at the end of October 2013, leading to an increase in articles from November onwards. Figure 74 shows the distribution of articles in the dataset over the years, differentiating by search only and search including the feeds. Altogether, the dataset contains 48,339 news articles published between 15th March 2011 and 16th September 2014.



Note: Excluding 2,372 tweets from March 2014 to have full month as last month

Figure 74: Number of news articles per year and month in dataset (author's representation)

As shown in Figure 75, most articles were retrieved from the British Broadcasting Corporation (BBC). This is because their RSS feed was used. The same is true for the Times of India.

[159] https://news.google.com/.

[160] http://emm.newsbrief.eu/; The Europe Media Monitor (EMM) covers at least 2,000 news sources (Steinberger, Pouliquen, and van der Goot 2009, 1) and makes them searchable through an RSS interface.

[161] RSS stands for "Really Simple Syndication" and is a data format enabling the exchange of news-related data.

[162] The following search terms have been used: "child labor", "child labour", and "child work". The latter two were additional terms.

[163] While double entries were checked for each retrieval channel separately, it may be possible that duplicate entries exist if, for example, EMM and Google News reference the same article.

Three other important resources include The Daily Mail, The Guardian, and The Hindu.[164] This data also show the importance of English news media in a given country. Moreover, it may be stated that country-specific newspaper sources such as The Hindu in India engage in specific child labor reporting. Otherwise, the news aggregators gathered a wide variety of different news sources.

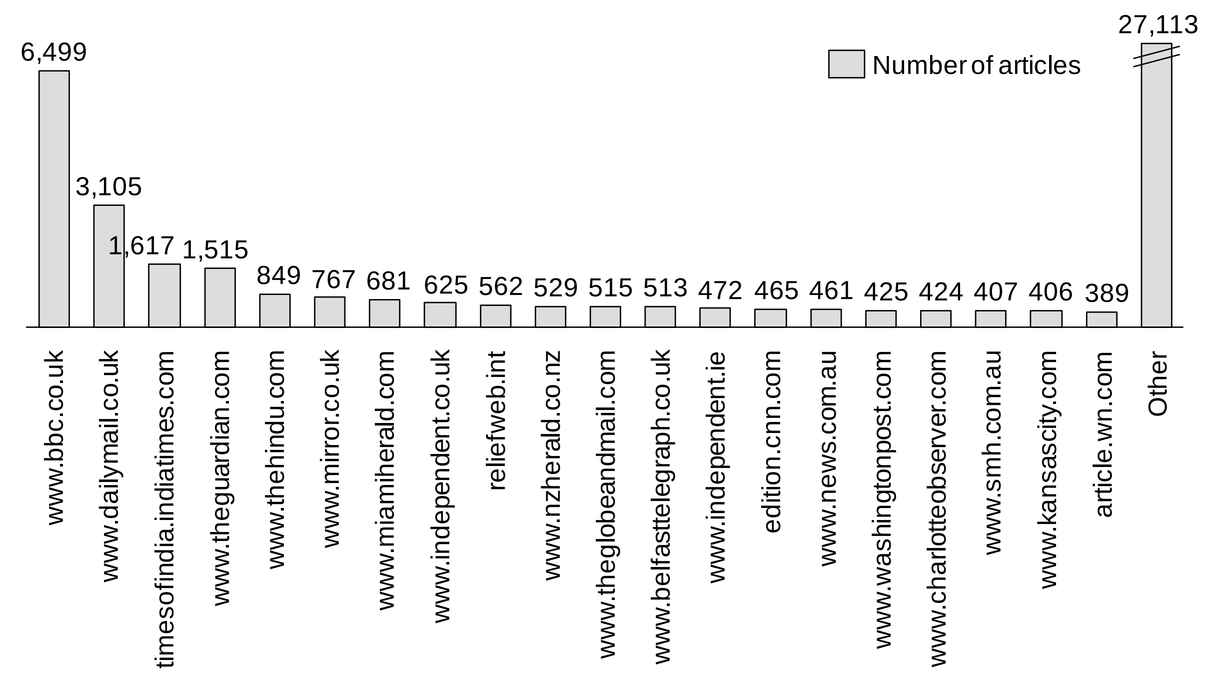


Figure 75: Top 20 news domains in the dataset (author's representation)

Altogether, we retrieved significant number of news articles that roughly covers a year of reporting on child labor. The set of articles has been saved in a serialized GATE corpus for further processing.[165] [166] [167]

7.1.2 NGO dataset

A social media based approach was taken for the second dataset. Twitter is one of the largest and most popular social media channels. It allows users to publish messages ("tweets") with a maximum of 140 characters, which may also include links and other cross-references (Twitter 2014). It is designed with speed in mind, and 2013 saw an average of 500 million tweets per day (Krikorian 2013). Organizations are also prominent on Twitter (Du et al. 2013).

[164] For referenced online domains, see Figure 75.
[165] Additionally, the RSS descriptions (summaries of the original articles) have been included in the article text before the headline if they were available as they provide additional content.
[166] No further reduction focusing on texts containing "child" has been performed as this has been one of the search strings. Although BBC and Times of India articles might not contain the word, these would be filtered out in the subsequent distance-based processing. As shown in Figure 74 their share in the dataset is limited.
[167] Three documents have been removed either due to large size or a high number of special characters.

It has been suggested that "classical" EE can be improved by fusing classical news with information gained from Twitter (Piskorski and Yangarber 2013). The following presents an approach which may be seen as an intermediate but promising step for certain domains. The limited content available in a tweet often leads users to include links to additional content which is either background material for the tweet or advertised by the tweet itself (Sankaranarayanan et al. 2009). Tweets from organizations such as NGOs are particularly likely to have this property, as they typically broadcast information about a report or something they have published on their website (details on linking are given below). By retrieving the linked content, rather than the tweet itself, and using it for further processing, classical IE techniques can be applied. In this context, Twitter can be seen as a focused online channel that directly provides links to articles without the need to crawl and analyze full websites.[168] Using linked content instead of the tweet itself has the additional advantage of increasing the available context to perform EE. Context is important for surveillance, in order to be able to specify events (Denecke 2012, 22). Other researchers have already made use of linked content in their text processing software (Sankaranarayanan et al. 2009).

Therefore, we built a list of NGOs that potentially post on Twitter, allowing us to retrieve the related content. In order to cope with the amount of data, we restricted ourselves to two countries (where possible): India, given the importance of English (Crystal 2004) and prevalence of child labor (Understanding Children's Work 2010), and Indonesia, due to its high Twitter use (Bennett 2012) and the presence of child labor (Understanding Children's Work 2009). As a consequence, different NGO databases were used to create the list. Table 91 shows the websites analyzed to compile a list of 12,339 NGOs that deal with India or Indonesia.[169] If possible, the search for NGOs was constrained to child labor, child(ren), human rights, and development NGOs. All of these sources indicated a website for a NGO, if available. Consequently, a total of 6,170 website addresses were collected. Given that the different databases overlap, web addresses were checked for similarity, resulting in a final list of 5,138 (almost) unique website addresses.[170] Table 92 provides an overview of the distribution of these NGO websites according to different categories roughly provided by the databases. Most NGO

[168] Here it is assumed that texts references in tweets are written in a news-like manner, making the algorithms developed in the last chapter applicable. As NGOs publishing longer texts which aim to present a reliable image, they will likely formulate these in a news-like manner. Moreover, although originally intended for longer texts, the text extraction approaches may also be sufficient for relatively short texts as they rely on the distance relations discussed above.

[169] Only 9 NGOs on this list were placed in categories other than India or Indonesia. These have been designated as "Asia" or "worldwide".

[170] For this correction the websites have not been fully automatically adjusted – roughly 3% of addresses have a slash at the end. Moreover, no detailed subfolders or subdomains have been cleaned. This might result in some duplicates. However, given that the twitter feeds have been used, a second uniqueness correction has been applied to the dataset later.

websites are categorized to India with an unknown domain. This might be partly because India uses English as an official language. Only about a quarter of the websites are categorized as a special domain. As a result, most of the NGOs do not deal with child labor specifically or post content on the matter. Nevertheless, specialized NGOs may exist.

			NGOs retrieved	NGOs with website link provided	
ID	**Source**	**Description**	**# Total**	**# Total**	**# Unique Websites**
1	ILO	NGO on Child Labor in India	13	12	12
2	Forum: Karmayog	NGOs in India	163	70	65
3	DevDir	NGOs in India	2,747	2,086	1,825
4	Choike	NGOs on Human Rights	10	10	10
5	IdeaList Search	Organizations in different countries	470	349	266
6	Tigweb	NGOs worldwide per country	108	63	58
7	UN	NGOs worldwide per country	671	454	393
8	NGOs of India	NGOs in India	6,423	1,771	1,468
9	US Aid	US NGOs worldwide per country	124	124	71
10	DevDir	NGOs in Indonesia	628	506	374
11	Semru	NGOs in Indonesia	117	117	90
12	ProPoor	NGOs in India	856	599	506
Sum			**12,330**	**6,161**	**5,139**

Sources (referenced by ID): 1: http://www.ilo.org/legacy/english/regions/asro/newdelhi/ipec/responses/india/other.htm; 2: http://www.karmayog.org/childlabour/childlabour_18034.htm; 3: http://www.devdir.org/files/India.pdf; 4: http://www.choike.org/2009/eng/5/1/links.html; 5: http://www.idealist.org/search/v2/; 6: http://www.tigweb.org/resources/orgs/; 7: http://esango.un.org/civilsociety/displayAdvancedSearch.do; 8: http://www.ngosindia.com/ngos.php; 9: http://www.pvo.net/usaid/; 10: http://www.devdir.org/files/Indonesia.PDF; 11: http://www.smeru.or.id/ngolist.php; 12: http://www.propoor.org/search.php; Dates of collection: 9.11.-18.11.2013 and 16.2.2014 (ProPoor)

Note: An ascending order of IDs was used to determine unique websites – "unique website" is referenced with the first source mentioned in this order

Table 91: Overview of number of NGOs collected with and without websites, including sources (author's representation)

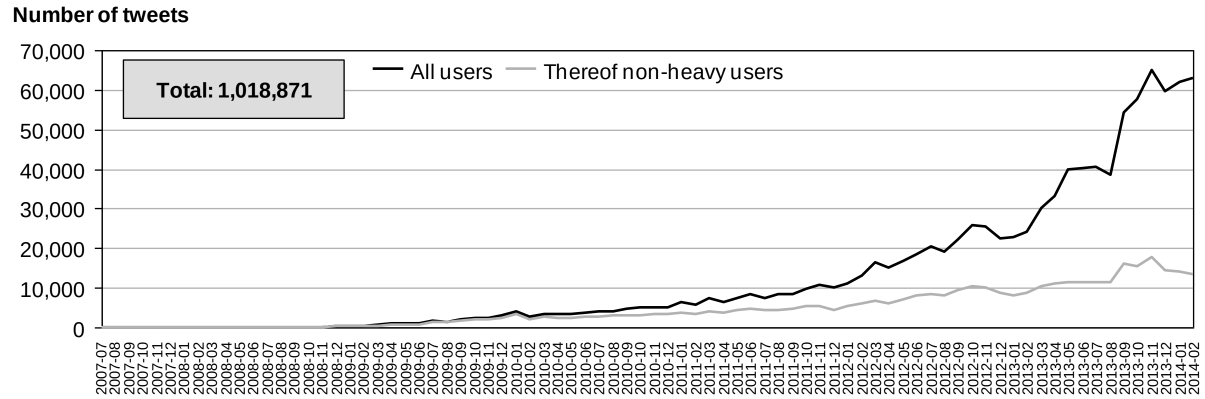
Geo. area	Child(ren)/child labor	Human rights/development	Unknown	Total
India	177	1,139	3,294	**4,610**
Indonesia	5	59	464	**528**
Total	**182**	**1,198**	**3,758**	**5,138**

Table 92: Overview of categorized geographic and subject area for collected NGO websites (author's representation)

The 5,138 unique NGO websites (predominantly in India) were automatically parsed to determine whether a link to Twitter was provided on the first page that opened by following the website link.[171] Altogether, this resulted in a set of 778 unique twitter accounts. Using the Twitter search API iteratively, we gathered the maximum amount of tweets from each user, constrained by the theoretical maximum of roughly 3,200 status messages that can be downloaded per user (Twitter 2015a)[172]. This produced a set of roughly one million tweets, published between 11th July 2007 and 10th March 2014.[173] Figure 76 shows an overview of the

[171] Date of retrieval: 17.2.2014.
[172] Application authorization without following Twitter users downloading the public statuses via GET statuses/user_timeline has been used.
[173] This date has been used for the corresponding article or articles.

tweets within the dataset. When downloading each tweet, external links (included in the tweet) were followed. This website data is stored together with the tweet.[174]



Note: 6,245 tweets from March 2014 are excluded to have a full month as the final month; "Heavy users" are those with more than 3,000 tweets in the dataset

Figure 76: Number of tweets per year and month in the dataset by type of user (author's representation)

More than two thirds of the tweets in the dataset are written in English (76.4%), followed by Bahasa Indonesian (11.1%) and 48 other languages with shares no higher than 2.3% (12.4% total). More than half of these tweets (54.2%) contain links to referenced content. Around 27% of the links provided in the tweets could not be retrieved and have therefore been excluded.[175] In total, 204 users within the dataset have more than 3,000 tweets. Nearly two-thirds of the tweets come from these "heavy users" (63.9%). However, excluding these heavy users still shows an increase in the overall traffic over the years (Figure 76). Consequently, Figure 76 may only be understood as a lower bound for the number of tweets posted in earlier years, given that only a constrained dataset could be retrieved for heavy tweeters. In fact, for 71 users the earliest tweet that could be retrieved was from during or after 2013, and for three users the earliest retrievable tweet was from 2014.

The tweets have been reduced to a set where each linked text contains "child" at any place in the text, similar to the assumption used in the text mining methodology. Consequently, 85,020 texts were then stored as a new set for further processing. Particularly large files, and by extension particularly long documents, were removed from the analysis due to problems with external APIs.[176]

[174] Boilerpipe verison 1.2.0 was used to scrape markup texts in the web pages (Boilerpipe 2011).
[175] For six percent of the total links an exception occurred during the retrieval process (e.g., the website link was not reachable), and for another nine percent Boilerplate did not return any content.
[176] Altogether, 422 documents were removed due to size constraints (larger than 1 MB text when stored as a GATE document or containing a very large number of special characters). In total this was around 0.5% of documents considered.

In summary, the dataset gathered for more detailed analysis is based on a broad set of different NGOs that can also be monitored on a regular basis. Twitter offers different ways to access their datastream.[177]

7.2 Method

The two sources were input into the prototype described in the previous chapters. Given the constraints on dimensional processing, the analysis focused on the classification task of applying the best ML approach in terms of F1 score. Using the prototype with preprocessing and the ML classifier, the two datasets were reduced to produce the analysis sets.[178]

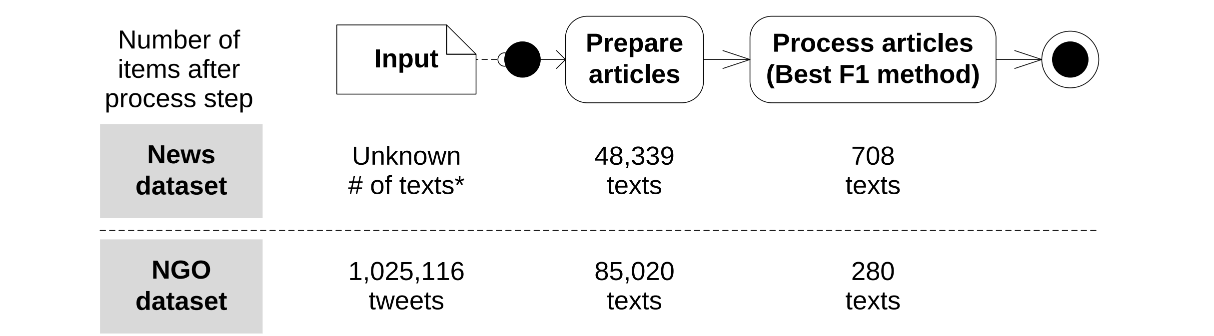
In order to produce a more content-driven analysis of the resulting datasets, a manual analysis was conducted based on a random sample of 100 texts for each of the final, classified sets. Data were classified based on whether they contained a child labor incident, were positive, mentioned a dimension specified in text mining, mentioned a dimension other than a geographic reference on the country level, and whether they detailed the geographic reference at least to the city level. Although the classification was performed solely by the author of this thesis, it should provide an indication of the quality of the datasets produced. Text excerpts from different illustrative texts provide additional insights.

7.3 Results

The average number of news articles found using the news aggregators was 3,743 per month in 2014. Assuming roughly 22 working days per month, 170 articles on child labor would need to be read each working day. Given that child labor is only one of many social sustainability topics, this appears to present difficulties. Automatic processing using the classifiers significantly reduces the number of articles to below two percent for news articles and below one percent for the NGO texts. The number of texts retrieved after processing the articles with all classifiers is shown in Figure 77, highlighting the results after applying the classifier with the highest F1 score.

[177] Another service offered by Twitter is its Streaming API which allows ongoing access to the Twitter stream, including filtering options (Twitter 2015b).

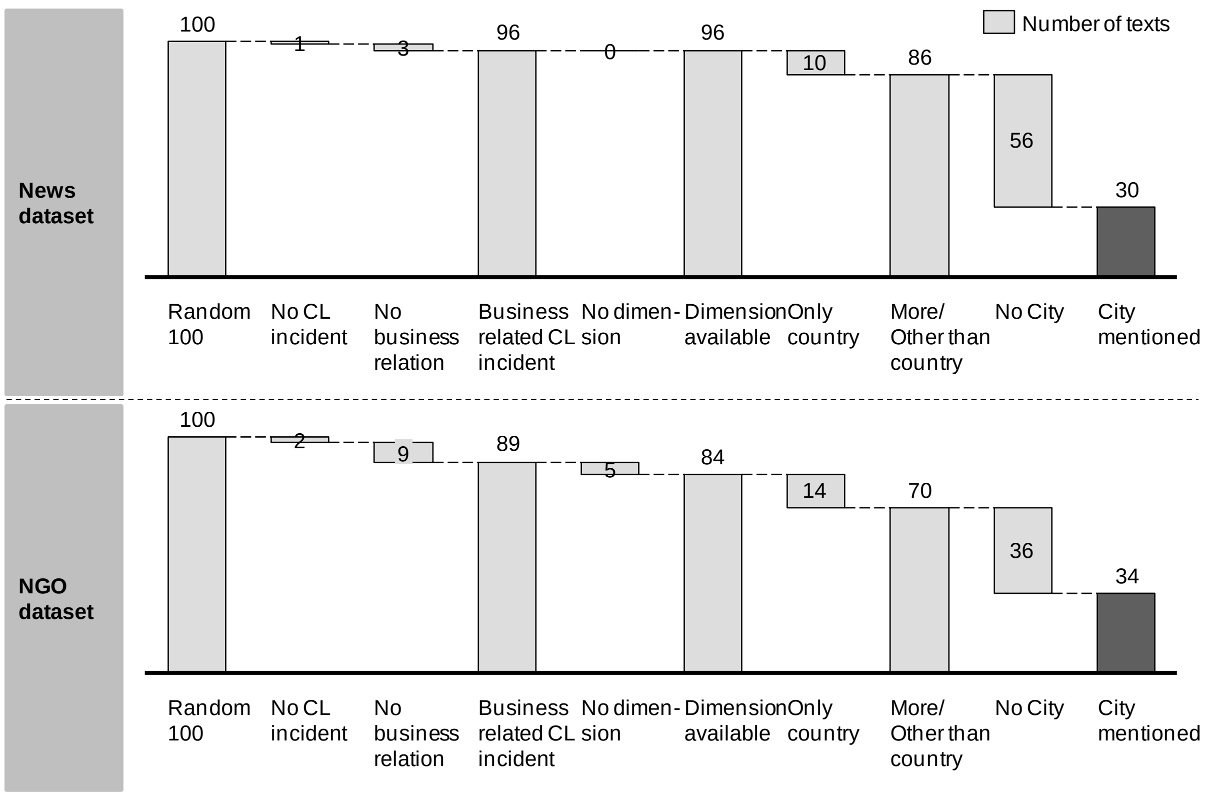
[178] For more details, see the chapter on text mining methodology and the evaluation in the final chapters. Here, the classifier has been trained on all available input data.



* Unknown due to IR approach through news aggregators

Figure 77: Number of articles in different steps of input data analysis for news and NGO datasets (author's representation)

The results of the analysis of the random selection of 100 articles from the news and NGO datasets are presented in Figure 78. Manual inspection shows that the large majority of articles in the sets (96% and 89% respectively) do in fact include business-related child labor incidents. Incidents without a link to business referred to domestic child labor. However, the data quality of individual incident descriptions varies significantly.



Note: "Other than country" does not mean that a geographic reference is available – sometimes only a sector is referenced

Figure 78: Analysis of news and NGO dataset based on a random sample of 100 items (author's representation)

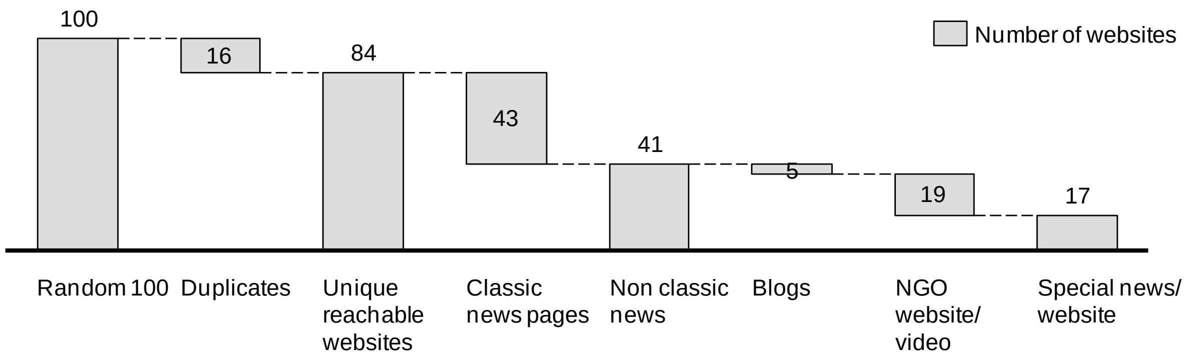
Only five cases had no dimension[179] and only 24 cases mention only the country, but in many cases the additional detail does not go significantly beyond this. In fact, most articles also

[179] i.e., the text contains a reference which can be classified as a child labor incident under the definition used in this thesis but which is too broad to be considered a dimension for TM.

provide the sector (not shown above) without giving any details as granular as a geographic reference to the city level. However, the sample from the NGO dataset has more geographically detailed cases. It should be noted that the lowest geographic detail may also come from a "light" reference in a caption, e.g., "Child labourers at a rope factory in Keraniganj, Dhaka."[180] Moreover, using the strict definition of a positive report, only 5% of the news dataset and 12% of the NGO dataset purely describe initiatives focused on eliminating or condemning child labor.

Although the assembly of the NGO dataset was strongly focused on India and Indonesia, its references also link to articles describing child labor cases in different parts of the world, highlighting the internationality of NGO work. In particular, mentions of child labor in the supply chain for Samsung (see Table 93 #8) appear in several articles in the news and NGO datasets, with only limited or no additional information in each new report.

Analyzing the types of links in the NGO dataset random sample reveals that a large share of NGO posts redirect to classic news pages such as The Guardian[181] when the tiny URLs[182] in the posts are expanded (see Figure 79). Nevertheless, many unique references (41 in total) still link to non-classical news pages such as blogs, NGO websites or videos (with descriptions), and special news pages or special websites. This thesis defines special news pages as pages that are not focused on a special domain and particularly not on the sustainability domain (e.g., news posted by international organizations such as the ILO). Special websites refer to comparable non-NGO and non-news websites.



* Website not reachable at the time of writing – text was retrievable at the time of collection

Figure 79: Types of websites in the NGO manual analysis sample (author's representation)

[180] http://www.theguardian.com/global-development/2013/sep/23/child-labour-falls-third-168-million, 23.9.2013.

[181] http://www.theguardian.com/.

[182] Tiny URLs provide an ID-based redirect to longer URLs in order to be able to better post URLs in social media settings with character size constraints.

The most NGO posts referencing child labor were found in 2013, followed by 2012 (see Figure 80). This may be due to the upper limit when downloading tweets, but it may also represent an increase in volume by year.

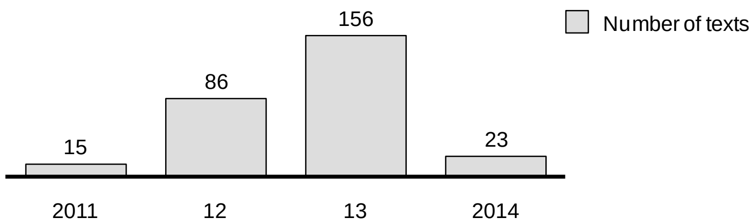


Figure 80: Time distribution of child labor referencing texts in NGO posts (n=280; author's representation)

ID	Data-set	Date	Type	Text
1	News	Oct 09 2013	General report	Title: "ILO will miss 2016 target to end worst forms of child labor"
2	News	Jul 28 2014	General report	Title: "Explainer: Child Labor Legislation in South America"
3	NGO	Jan 16 2013	General report	"More than 1.3 million children in Yemen are involved in child labour, including 469,000 in the 5-11 age group, according to a new study. This means 17 per cent of Yemen's 7.7 million children in the 5-17 age group and 11 per cent of those aged 5-11 are involved in child labour, according to the study which is Yemen's first-ever national child labour survey carried out with support from the International Labour Organization , the Social Development Fund and UNICEF. […]"
4	NGO	Aug 28 2013	Multiple incidents	"As many as 36 cases from Koderma and 22 from Khunti were brought to Dube's notice when he visited these districts. Such incidents include a girl from Khunti missing since 2009 when she went to Delhi for work. […] The survivors also discussed their fate and the torture they were subjected to while they worked in Delhi and other big cities"
5	NGO	Aug 26 2013	Multiple incidents	"He and 11 other children from his village were taken by bus on a 2,100km (1,300 miles) journey and put to work in Ho Chi Minh City (formerly Saigon), south Vietnam. […] The charity helps children forced into a variety of jobs from prostitution to begging, but in the past year just over a quarter of that number have been rescued from garment factories in Ho Chi Minh City, the country's largest metropolis and industrial centre."
6	NGO	Dec 22 2012	Incident report, narratives	"Aslam, twelve, is a native of Sipur village, Azam Nagar in the Katihar district. Despite being promised a good education he too ended up in the same dingy sweatshop in Delhi. Like the other rescued children he worked very long days, sleeping in the same room where he worked."
7	NGO	Jun 13 2012	Incident report	"A young Indian bonded child laborer cries as he is walked away after being rescued during a raid by workers from Bachpan Bachao Andolan, or Save the Childhood Movement, at a garment factory in New Delhi, India, Tuesday, June 12, 2012."
8	News	Aug 29 2014	Incident report (Samsung)	"New York-based China Labor Watch (CLW) claimed in a report Thursday that Samsung supplier HEG Technology had hired people under the age of 16 at its facility in Huizhou, China. The report said the factory, where components for Lenovo are also built, had hired new workers, […]"
9	News	Jun 19 2014	Incident report	"Twelve minors, working as labourers in various garages, hotels, bars and dhabas in and around Chhatrapur in Ganjam district, were rescued on Wednes-day."
10	News	Jul 14 2014	Incident report	"Samsung's audit was conducted after the campaign group China Labor Watch and found evidences of illegal hiring process that took place on June 29. Ac-cording to the campaign group: "During the busy season and in urgent need of labor, Shinyang hires child labor and underage student workers."
11	NGO	Jun 13 2012	Incident report	"Social activists, with the support of the police, have raided a New Delhi factory employing children aged between six and 13. The children, who had been working there for more than a year, came from other regions of India like West Bengal, Bihar and Uttar Pradesh, according to the activists."

Sources (referened by ID):
1: http://www.trust.org/item/20131009102946-p4t5j/;
2: http://www.as-coa.org/articles/explainer-child-labor-legislation-south-america;
3: Tweet ID: 291591675367797000, http://www.unmultimedia.org/radio/english/2013/01/more-than-1-3-million-child-labourers-in-yemen-ilo/;
4: Tweet ID: 3726461851162378e17, http://timesofindia.indiatimes.com/city/ranchi/NCPCR-pledges-action-against-child-trafficking/articleshow/22105876.cms;
5: Tweet ID: 371912419372986000, http://www.bbc.co.uk/news/world-asia-23631923;
6: Tweet ID: 282409517118001000, http://www.huffingtonpost.com/gordon-brown/india-child-labor-laws_b_2345756.html;
7: Tweet ID: 212881625070051000, http://www.utsandiego.com/news/2012/jun/12/factory-raids-reveal-child-labor-persists-in-india/;
8: http://www.bangkokpost.com/tech/computer/429443/samsung-denies-child-labour-at-chinese-supplier
9: http://timesofindia.indiatimes.com/city/bhubaneswar/12-child-labourers-rescued-in-Ganjam/articleshow/36808587.cms
10: http://www.digitaljournal.com/news/world/samsung-has-used-child-labor-in-china/article/389632;
11: Tweet ID: 212825335249641000, http://www.aljazeera.com/video/asia/2012/06/2012612164116177802.html

Table 93: Excerpts from texts from the 100 text sample for each dataset (author's representation)

The text excerpts in Table 93 give an impression of the diverse way in which child labor is depicted in various texts. The following discusses several details while referencing the lines in Table 93 using their IDs (indicated with hashes). First, the datasets also include broader re-ports. They often reference a broader array of different child labor incidents (#1, #2), together with contextual references (e.g. #3). Beyond these general reports, other texts also give broader references to a combined set of multiple child labor cases (#4, #5). Incidents may be depicted in narrative fashion building on a single individual case (#6) or at least referencing it directly (#7). However, incidents are also reported directly, as can be seen in the later text excerpts (#8 to #11). Thereby, the reports on child labor can also include child labor categories such as prostitution, begging, or domestic work that are less relevant from a company perspective (#4).

7.4 Summary and discussion

In summary, applying the classifiers led to a significant reduction in the number of articles and texts that need to be processed by the system (or rechecked by a human operator). In particular, given the significant input sets, the approach using links in NGO tweets only led to a limited number of child labor incidents. However, these contain details that might not be found using news aggregators. This is especially likely if they are published on blogs or other NGO-related websites.

Given the large geographic distances in a country like India, details at a city level can be particularly helpful when an incident needs to be referenced to supplier locations. Consequently, although they were present in roughly a quarter of the instances, the overall lack of these details significantly reduces the usefulness of the data. Moreover, not filtering out non-relevant sector references may lead to the inclusion of texts referencing non-company relevant forms of child labor. Nevertheless, they may provide indications for companies situated in a specific area.

The wide variety of different ways to describe child labor incidents, along with the density of describing different references to child labor in texts, might require a more context-focused definition of a child labor incident that allows for more than one child labor occurrence per document (while potentially still keeping the observations as a sub level). This would help to more clearly split the references within a single article without exaggerating individual dimensional values through a (potentially erroneous) cross-product to produce observations or the need to skip multiple relevant child labor references by limiting the analysis to the minimum distance relationship. In fact, the child labor distance relations might provide the basis for the definition of such incidents.

By focusing on child labor as a specific case of social sustainability, a condensed set of particularly relevant reports containing a detailed reference to a child labor incident could be established. This set is particularly relevant for monitoring child labor in supply chains. Given that the classification approaches used have no perfect recall rates (73.7% using cross validation on the closed Reuters corpus), even more undetected texts might be found in the original data sources. In addition, the techniques used in this chapter are further constrained by several decisions. First, they focus on India and Indonesia during the selection of relevant NGOs. Second, only two news aggregation services were used. Third, Twitter was the only microblogging service considered. Finally, English was the only language used. Using local languages could provide additional insights. One other way to extend the analysis would be to include user-generated content or citizen journalism (e.g. Chua, Razikin, and Goh 2011).

8 Discussion

The main aim of this thesis is to analyze how to better support ongoing social sustainability monitoring of supply chains using an automated risk management system that integrates private and public data sources, including news. The system's output is a prioritized list of supplier locations based on their risk of breaching social sustainability standards. Hence, a model determining this breach risk was developed and evaluated. In addition, a context-specific text mining approach was tested that could also be used for other application domains. The availability of input sources was also investigated. Child labor, being a major part of social sustainability, was chosen as a use case. While it has been proposed in the literature (Grimm, Stölzle, and Hofstetter 2013; Foerstl et al. 2010; C. Reuter et al. 2010), this specific type of system has not been elaborated in the scientific literature before now. Hence, this thesis tries to provide an estimate of the risk of rare social sustainability events that are also hard for experts to assess (Brooker 2011).

The focus on social sustainability is particularly relevant as it has been rarely covered in earlier literature given that it is hard to quantify and has an ambiguous nature (W. Norman and MacDonald 2004). Specific assessment approaches such as tailored auditing approaches and particular assessment schemes have been proposed (Keating et al. 2008; Miemczyk, Johnsen, and Macquet 2012). Social sustainability issues can induce severe economic effects through reputational damage (M. B. Taylor, Zandvliet, and Forouhar 2009), even if they only occur once in a company's supply chain (Lemke and Petersen 2013; Vermeulen and Seuring 2009).

By focusing on child labor incidents, this thesis covers a type of social sustainability issue that is widely recognized in different sustainability standards. Although ILO give a definition of child labor, there is still much debate over an exact definition and the actual level of negative impact caused by child labor (Bourdillon et al. 2010). Nevertheless, it is certain that companies affected by a child labor scandal face severe reputational damage (Lemke and Petersen 2013). Thus this is an important issue even before any moral questions are considered. Other areas of social sustainability such as corruption or bonded labor are even less clearly defined.

8.1 Results and interpretation

The results of this thesis can be split into three areas: those for the risk model, the TM approach, and the analysis of real data availability.

8.1.1 Risk model

The first research question is how an integrative risk model can be used to ongoingly monitor supplier locations while including news. A BN-based risk management model was proposed that estimates the likelihood of a breach in child labor standards at a given supplier location and allows the integration of evidence using relevance and credibility scores. This has not been used in earlier studies employing BNs, although they have been applied to supply chain risk management (Badurdeen et al. 2014), environmental risk analysis (Joseph Amundson et al. 2014; Yen and Zeng 2010), and supplier performance (Maleki, Bashkite, and Machado 2012). Hence, this thesis introduces this aspect with a focus on social sustainability. The probabilistic model underlying a BN allows the mean, median, or quartiles of a node within the BN to be used to rank supplier locations. As well as a relative ranking of supplier locations at time t, suppliers can also be monitored over time using these numbers.

Ongoing model updates through new evidence lead to an increase in the likelihood that a supplier will breach a code of conduct. The evidence has to conflict with the basic assumption that suppliers comply with a company's standards. Additionally, the update process requires that the items entered as evidence are independent of each other and, thus, each update carries additional information that leads to an increase in the breach likelihood. This behavior is corroborated by sensitivity tests and has also been shown mathematically. Practically, even slight evidence from a source with low credibility should also lead to an update, increasing the overall breach likelihood for a supplier location.

Relevance and credibility are two particularly relevant factors when differentiating the quality of news inputs. Parameters were proposed to attribute a relevance to a news text based on its level of detail. The idea that influence on the overall breach likelihood increases as articles become more detailed was supported by the responses from an expert questionnaire. This also underlines the fact that cases which do not occur directly at a supplier location should still affect the breach likelihood if they can be related to a supplier location through an article's content. The influence of credibility on the importance of a source of evidence was only touched briefly in the questionnaire. However, the literature also makes this connection, along with suggestions for deriving it automatically (Schwarz and Morris 2011; Nagura et al. 2006; R. Tang et al. 2003).

Besides ongoing input from news sources, the network is also updated using new audit results. The more recent audit score together with the time since the last audit are integrated. Concentrating on the last known audit result means the best in-depth data on a supplier is used. This is supported by the expert questionnaire; the experts view audits as the most important source of information when needing to estimate the risk level of supplier locations with regard to child labor. Nonetheless, the experts still assign a residual child labor risk, even if a supplier achieves the best possible audit score. Instead of audit data, this input can also come from certification processes, partner companies, or through platform-based exchange (for example through SEDEX; Teuscher, Grüninger, and Ferdinand 2006). Focusing on the most recent audit result ignores other possible input from earlier data, which could provide additional conclusions such as diminishing performance.

The prior integrates data from statistical sources to mathematically determine the probability of child labor per country, sector, and ruralness. The cases tested together with the experts show comparable rankings for supplier locations. However, the experts provided a wide variety of different answers and rankings, which decreases the interpretability of the results. This variation may be due to the difficulty in estimating the probability value based on the numbers provided without an additional calculation framework, or it may be due to the different heuristics experts use to determine risk probabilities given limited information. Nevertheless, the prior value provides a mathematically derived quantified number, and its components have been agreed on by the experts. In a different context than child labor, another structure for the prior might be necessary given the data availability and underlying driving forces of a different social sustainability factor.

BNs have the advantage that they are more easily understandable than other probabilistic frameworks (Duespohl, Frank, and Doell 2012; Koks and Challa 2005; Wooldridge 2003). The old suggestion that “[N]o MIS [management information system, A.T.] should ever be installed unless the managers for whom it is intended are trained to evaluate and hence control it rather than be controlled by it” (Ackoff 1967, B–153), makes this property even more important. The nodes of the BN proposed in this thesis can be directly explained to sustainability managers. Moreover, as the expert questionnaire highlights, the requirements incorporated into the BN (together with the surrounding system) are strongly supported by experts in supply chain, sustainability, and risk management. Only one requirement did not see significant agreement – experts want to be able to manually change the final input into the risk model. Comparable editing functions have been suggested before (Atkinson and Piskorski 2011). But as the current system design (apart from the initial configuration) allows unbiased input into further processing steps, the amount of user input, if allowed, has to be discussed in detail. Experts do not necessarily make the correct judgments, and biases may affect the manual input, leading to a questionable ranking of supplier locations. There are also numerous factors that can unconsciously influence decisions (Bazerman 2006). Raiffa highlights this based on the example of a lawyer asked why he persistently chose an option with lower mathematical probability (white instead of black when selecting the color indicating the outcome of a game): “Yes, I understand, but in my experience at the bar, *life is just plain perverse, and I would still bet on predominantly white*! But I really am not a betting man.” (Raiffa 1968, 20) Consequently, also the configuration needs to be cross-checked. Nevertheless, experts nearly significantly attribute job relevance to a model and system incorporating the chosen requirements (at the 6% level, but true for large companies), supporting the overall assumption that the system has a high likelihood of being adopted by an organization.

8.1.2 Text mining approach

The TM approach was discussed in two steps: the detection of child labor incidents in texts and the extraction of the related event information. The system has been specifically tailored to child labor risks, given the domain-relatedness of TM applications (Arendarenko and Kakkonen 2012; Coppola et al. 2009; IJntema et al. 2012; Van Landeghem et al. 2011; M. Naughton, Stokes, and Carthy 2010). Specifically, both ED and EE use child labor distance relations (i.e., a word indicating “child” or “labor” and the nearest item indicating the other element) as a feature introduced to concentrate the analysis on relevant text zones (Manning, Raghavan, and Schütze 2008, 339). While this may restrict the usefulness to other applications, most domains of social sustainability can be detailed using two-worded key-terms. The detection system achieves F1 measures above 80% for the texts containing a child labor incident of the gold standard and, consequently, allows drastic filtering of large streams for child labor incidents. Given that the interviewed experts did not uniformly prefer either precision or recall, an F1 measure optimization was performed.

The extraction of incident events and independent incident observations for the domain was based on different heuristics which strongly utilize child labor distance relations. EE typically shows precision and recall rates of around 60% (Tanev, Piskorski, and Atkinson 2008, 208). The results show high recall values analyzed separately for the three dimensions used in this thesis: geography, company, and sector. The F1 measures only meet this benchmark for the geography and roughly for the company dimensions depending on the calculation type used. The new approaches suggested for sector tagging using a standard industry classification generate a broad variety of different possible sector tags with relatively high recall, but disambiguation still seems insufficient for practical use. Therefore, as also requested by the experts, a manual check of the extracted evidence before its integration into the risk model seems necessary. Stricter pattern matching-based approaches, as referenced in the related work, could potentially improve performance.

When tagging the gold standard used for EE, the data model with one child labor incident and several observations per news text storing multiple values per dimension appeared sufficient. There were only an isolated number of cases where a broader association between entities and child labor had to be assumed. For example, a sentence stating "[…] children and mentally disabled were forced or lured to work in kilns and mines in Shanxi and neighboring Henan […]"[183] was judged to relate Shanxi and Henan provinces to both kilns and mines. This "cross-product assumption" to derive observations from incidents appears less correct after examining the data from the crawled articles. Some of the texts describe several very different child labor incidents, and hence the cross-product leads to partly incorrect results. One solution could be to automatically split a text into multiple parts. Similar text partitioning systems have been presented in the past (e.g. Hearst and Plaunt 1993), but this is left for future work.

8.1.3 Empirical data availability

Finally, the third research question aimed to find online public data sources that can be used to detect child labor incidents. In the top-down part, the analysis focused on India and Indonesia and screened texts identified through links posted by NGOs concerned with human rights or children on social media (Twitter). The bottom-up analysis used two different news aggregation services to collect news articles published in various online news outlets, not restricted to a specific country.

The analysis revealed that relevant data is indeed published on the web, particularly by news organizations. Nonetheless, additional content provided by NGO through links in posts may include additional incidents and dimensional details that can improve risk analysis. This NGO content may not be covered by standard news reports. Given the small field of child labor and the restriction of NGOs to those operating in India or Indonesia (although the incidents col-

[183] Document in Reuters TRC2; Timestamp: 2008-03-20 07:51:15; BEIJING.

lected cover a broader range of regions), the top-down technique also appears to be a relevant supplement to classical news aggregators for other domains.

8.2 Recommendations and ethical reflection

Altogether, the findings lead to several recommendations for practical application. Using a quantified risk model that is ongoingly updated based on observations can help to focus a company's auditing resources and other activities such as supplier development where they are most effective. Designing a model and system that fulfills the key requirements and is understandable is feasible. Consequently, its implementation may also allow a more unbiased and objective discussion of social sustainability activities across the supply chain. Additionally, leveraging text mining may free up resources currently focused on screening activities and detect more issues. Given that companies frequently already leverage text sources, e.g., for environmental analysis (Chih-Ping and Yen-Hsien 2004; Wei and Lee 2004), these sources could also be employed to monitor child labor. However, they might require some extensions, e.g., through social media monitoring as discussed for the top-down NGO-centric approach above.

Consequently, to be effective, companies would need to have the resources (internal or external) and processes to be able to work with the outcome of the risk model. They need to analyze potential issues more deeply, potentially also performing on-site visits. At the same time, the observations and risks detected would still need to be checked manually until a TM approach reaches a sufficiently high F1 rate for event extraction (probably comparable to classification). While the focus here was on child labor, this process can also be adapted for other social sustainability risk sources such as forced labor. This could be integrated into a broader system that includes the different elements, with components as suggested in Figure 81.

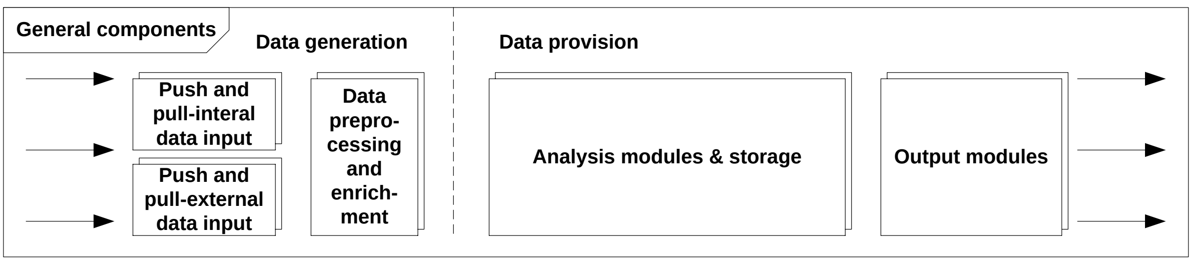


Figure 81: General component overview of a potential overarching sustainability risk information system (Thöni, Madlberger, and Schatten 2013b, 272)

A key question when performing a risk-based analysis of suppliers or when a social sustainability issue is detected at a supplier location is how to deal with the particular supplier. The appropriate reaction strongly depends on the type of issue. For child labor, discontinuing a supplier due to bad performance may not improve the child or their family's situation (Kolk and van Tulder 2002b). Besides financial aid for the child or children (as suggested for example by Social Accountability International 2008b), supplier development can offer further opportunities (Harms, Hansen, and Schaltegger 2013). Increased commitment, collaboration, and supportive treatment of suppliers can be steps that may already be taken in order to mitigate issues, particularly for high-risk suppliers (Locke, Amengual, and Mangla 2009).

One question that needs to be addressed when introducing a comparable system in a practical context is the degree to which organizations trust the news reports incorporated into the model. Texts from stakeholders with self-interest, for example NGOs that also want to stress the importance of their existence, might over-exaggerate or even misinform. Consequently, before reaching out to suppliers directly for evaluation, different escalation levels for each case appear useful to prevent damage from premature or unwarranted accusations.

8.3 Limitations

A key prerequisite for the analysis proposed in this thesis is the availability of a mapping of supplier locations as far as possible across the supply chain. This includes, if possible, transparency regarding sub- and sub-sub suppliers that could still trigger a reputational effect in case of issues. This level of transparency is often very hard to achieve and may significantly constrain the approach presented in this thesis. Some options for overcoming this include using social media to approximate supplier locations (Madlberger et al. 2014), leveraging industry networks, or adapting the model to a regional level if supplier regions instead of exact locations are known.

As in any model, the main limitation is the abstraction of reality to a limited set of variables and the impossibility of fully modeling all relevant aspects of an expert's behavior. Furthermore, at a general level, it is constrained by the availability of input data. In particular, the statistics needed to calculate the prior may be difficult to find, ambiguous, not comparable, or even not available. This has been addressed by leveraging the broad numbers published by ILO, Worldbank, and UNICEF (e.g. for Indonesia Understanding Children's Work 2009).

The model has been evaluated using an expert questionnaire, requirements, and sensitivity tests. In a practical context even further problems may arise that can only be identified when implementing the proposed system into an organization. Additional requirements such as processing speed or volume restrictions will need to be fulfilled (Aggarwal 2014, 9). Nevertheless, the expert-based review of the suggested system and model approach make the usefulness of a practical implementation clear.

Mathematically, the nodes of a BN are assumed to be conditionally independent. However, in this model this might not be completely true for all nodes affecting the breach likelihood; for example, an audit score or news report might be partly related to statistical data depending on the sector or region. However, this effect should be very limited for audits as they mainly leverage company internal information. For news reports, statistical information is typically significantly older than the evidence that can be gathered from news articles, which is often affected by recent events. In addition, as discussed above, news should only influence the breach likelihood strongly if it embodies sufficiently detailed information (potentially more detailed than the dimensions of the model's prior). Another technical problem for BNs is the requirement for discrete variable input and, thus, continuous variables need to be transformed (i.e., "discretized") before inclusion. This may result in decreased precision of the overall value (Duespohl, Frank, and Doell 2012; Pai et al. 2003; Uusitalo 2007).

Analyzing news on child labor involves a further constraint: basing an analysis solely on news articles automatically includes a "sensory gap between real-world happenings and what are captured in media." (Xie, Sundaram, and Campbell 2008). This has been addressed through the suggestion of extending input sources to also include posts covered in social media. However, these may also spare important events or depict them differently from reality. Consequently, the credibility parameter attached to the input source is particularly important.

Linguistic processing in general poses several challenges (Denecke 2012, 22): linguistic variation, polysemy, negations, contextual information, and coreference. The methodologies suggested for TM try to overcome these to a certain extent. The TM approach used here focuses on identifying child labor incidents and particularly is tailored to the specific "two-wordedness" of the term "child labor". Consequently, reusing it for other applications is limited to contexts with similar constraints. Moreover, the gold standards for the evaluation of the ED and EE approaches assume that articles on child labor include the word "child" at some point in the text. Furthermore, using these gold standards for model training implicitly shows a bias towards the type of news texts published there (as can be seen in the incident event construct discussed in Chapter 8.1). Also, the gold standards are constrained by the limited number of true positive articles on child labor in them. However, applying the trained classifiers to crawled datasets suggested a precision comparable to the closed datasets; however, this was only analyzed by a single researcher on a random sample of 200 articles.

Given the low frequency of child labor events, this thesis does not differentiate whether several articles complement or contradict one another. Based on the definition of a "child labor incident" provided in the introduction, no further details such as negation or other constraints are used from the text. This potential for improvement is left for future work and has already been discussed in more detail in other domains with higher frequency events such as activist activities (Ploeger et al. 2013). Moreover, folding relies on the correctness of the underlying facts. This rigid assumption may be relaxed in later work by embodying techniques from information fusion which can handle errors and noise (H. Ji 2010). Finally, in an ideal case merging would also require including the input quality of the events being merged (Appelt 1999).

The EE approach for the domain relies on different methods and heuristics. These have been evaluated using a set of tagged dimensions built on the assumption that the lowest detail has the highest importance, leaving out other entities that may have been included in the articles. Consequently, recall and precision values might differ if these were included. Likewise, the limited set of true positives used to test EE could also diminish the quality of the result. This is also strengthened by "double" articles.

Finally, the analysis of the availability of input data focused on India and Indonesia for NGOs and was limited to two news aggregators. Given that the aim of the third research question can be interpreted as falsifying the hypothesis that insufficient data is available in public news sources, the results presented can be seen as sufficient proof for the falsification. Nevertheless, also these results are limited by the performance of the classifier. Only English was used, and using additional local languages could provide additional insights. While not further detailed here, this seems strongly supported by the fact that many more references to child labor were found for India than Indonesia given. This is probably because India has English as one of its official languages.

9 Conclusion

Issues with child labor and other social sustainability themes can cause severe reputational damage to companies, even if found in very remote areas of the supply chain. Moreover, the societal impact of global supply chains has come under particular scrutiny in recent years. This thesis suggests using a Bayesian network (BN) ongoingly fed by reports on child labor observations to estimate the risk of a breach of corporate sustainability standards at a particular supplier location. BNs have already been implemented for risk management and are noted for their clear and explicit treatment of information uncertainty. BNs can also be configured using disparate inputs, including quantitative, statistical, and subjective expert information. Consequently, a risk model was proposed that fulfills requirements agreed upon by experts and which allows expert strategies to be incorporated. Tests were partly performed using special or limited data from India and Indonesia.

The BN risk model estimates the probability of a breach of a company's "code of conduct" with regard to child labor for a specific supplier location. It builds on a statistically derived prior and is updated using the most recent audit results for the location and news items containing child labor issues that can be related to the location. The latter makes use of Bayesian updating to incorporate credibility and relevance of news items as well as their number, yielding an observational probability. Through its probabilistic nature, the BN provides a quantified ranking of supplier locations based on their level of risk which may be used for further mitigating actions in the supply chain. Using an expert questionnaire, the risk model received a positive evaluation in terms of its usefulness based on the technology acceptance model. Domain experts have also been asked to provide their feedback on different requirements and on the calibration of the model.

This thesis also suggests a text mining (TM) approach that relies on a domain-tailored technique for child labor incident event detection and extraction which can also be used for different areas of social sustainability or other domains. One key feature of the approach is the focus on document zones using "child labor distance relations." The TM method partly reuses existing system parts and produces items called "independent incident observations" (IIOs) that can be related to supplier locations based on their attributes and used to update the respective BN risk models. Results tested broadly on a closed English news corpus and in a more limited fashion on a German corpus suggested the results were useful for event detection but less practical one for event extraction, still requiring manual correction steps. Nevertheless, the latter result strongly depends on the event frame (dimension) extracted, with better results for geography and company and worse for sectors based on the standard industry classification (ISIC).

Finally, using the TM event detection approach developed for this thesis, two public online text sources were analyzed for the availability of relevant input messages. Two approaches were taken: A bottom-up approach relying on two news aggregators and a top-down approach analyzing texts linked by NGO social media posts. The analysis of a random sample from each dataset supported the usefulness of news data as a risk input and also highlighted the usefulness of additional input beyond classical news.

In summary, this thesis provides several scientific novelties and adds the following to the current stack of knowledge:

- A BN-based risk model for ongoing determination of the breach risk at supplier locations with regard to child labor sustainability issues. In particularly, it allows for the incorporation of news information and can be generically adapted to other sustainability issues apart from child labor as well as risk contexts which depend on the environment (e.g., country, sector, and ruralness), ongoing data from news, and evaluation results (e.g., from audits).
- A text mining approach tailored to identifying and extracting child labor incidents with better results for event detection than event extraction. This can also be transferred to other sustainability issues. It can also be potentially reused in any scenario where the risk can be described by a two-word expression (such as “child labor”).
- An analysis of online public data sources for their usefulness with regard to child labor risk monitoring in supply chains. Particularly, the top-down approach used to gather texts referred to by NGOs may be employed in company settings.
- A gazetteer- and rule-based approach for tagging sector keywords preserving hierarchical information and based on a standard industry classification. This approach has been compared to a machine learning-based classifier that has been trained on classification labels.
- A suggestion for a technical approach combining the data into a system that can be used for risk monitoring.
- An RDF-based data model directly incorporating dimensional information in triple format together with extending linked open data.

Still this thesis indicates some very interesting and prospectively fruitful ideas for further research in multiple scientific domains. The questionnaire used to evaluate the model can be extended in further research to cover even more experts or to retest parts such as the prior, the news articles, or parameters in more depth. Additionally a conjoint study could be conducted to derive more empirical parameter values for article relevance and source credibility. Furthermore, the model ongoingly adds evidence to the model. To exclude articles, they need to be deleted from the database. A future approach could handle this more elegantly and include fading to reduce the effect of older evidence. For example, Allan et al. (1999, 27) have experimented with an exponential decay function.

In addition, the system only allows reviewing observations in the database. Further work could analyze the opportunities that rely on active learning (Manning, Raghavan, and Schütze 2008, 336). They could be employed to show uncertain cases of child labor incident observations. Experts could then decide whether the text should be further processed and included. The information gathered could also be used to retrain the original classifier and improve performance for future iterations (Torii et al. 2011). While the two-word relationship between "child" and "labor" has only been explored within the context of this thesis, this approach and its usefulness can be tested in other social sustainability contexts. Additionally, the distance relationships can be explored to implement document zones and extract child labor incidents for each zone, extending the model employed in this thesis. One further extension could be the alignment of the domain model with known, more limited models for events such as the grounded annotation framework (Fokkens et al. 2013).

From a mathematical standpoint, Bayesian theory can be seen as a special case of the Dempster-Shafer theory (Shafer and Srivastava 1990, 129). Consequently, one idea for generalization could be to employ Dempster-Shafer Theory to the problem which might allow a more direct representation of conflicting information[184] and avoid reliance on the assumption that suppliers comply with not employing children. Fei and Shilei (2010) have already suggested a supply chain risk management approach utilizing evidence theory without a focus on sustainability or a discussion of combination methods inherently important to Dempster-Shafer Theory.

The domain-tailored TM approach discussed here could be improved at several levels. Additional feature selection techniques might further enhance the classifier (Dasgupta et al. 2007;

[184] One of the core specifics of belief for the Dempster-Shaver Theory is that the sum of belief in a certain outcome might not add up to 1 (called additivity; Shafer and Srivastava 1990).

Manning, Raghavan, and Schütze 2008, 272–279; Rogati and Yang 2002). This thesis used a $tf - idf$ based weighting of features together with a document zone based approach to feature selection. Nevertheless, this could be improved by employing the techniques suggested by the referenced authors. Likewise, switching to different named entity extractors could further improve performance (Gangemi 2013). Additionally, the TM approach followed here focuses on the methodology, excluding processing speed. Although sufficient hardware resources may compensate for slow processing speed, assessing and improving processing speed has been left for future work. The use of online web services during tagging might eventually require a switch to different named entity annotators. As a special case, the tagging of sectors based on a broad, standard industry classification could also be a focus of future work. For example, hierarchical classification might be employed to improve the process (Manning, Raghavan, and Schütze 2008, 337).

Multilingualism is only slightly discussed in this thesis and could be seen as a further extension to cover even more input sources. As URIs are used to reference named entities, the same structures might be used in different languages. This follows the abstraction of surface forms and conceptual representation introduced earlier (e.g. Ferber 1997). Multi-language support might help to gather better and more granular data if local NGOs publish in regional languages rather than English.

Another data gathering technique could be used for including statistical data. Linked open data could be an option for reducing the amount of manual work required. However, this is currently strongly constrained by data availability and limited trust, as it is often third parties rather than the original data owners who publish linked data on the web (see Thöni 2013 for an example on corruption). Additional opportunities and risks have been outlined by Weichselbraun et al. (2014). In terms of architecture, a multi-agent system approach could be employed, instantiating individual agents for each supplier location that assess sustainability implications for each supplier individually (LaComb, Interrante, and Aggour 2007).

Overall, this thesis provides a first quantitative risk model for social sustainability monitoring in supply chains based on a Bayesian network and text mining. Analyzing with a focus on child labor suggests that sufficient relevant input data is available for risk management purposes. In organizations, a system building on the techniques suggested would still need some level of manual interaction and adaptation beyond IT, and organizational processes would need to be established to trigger appropriate responses to changes in risk measures. Still, it may be seen as a step towards greater supply chain responsibility and the improvement of working conditions in developing countries.

Literature

Abukhader, Sajed M. 2008. “Eco-Efficiency in the Era of Electronic Commerce – Should ‘Eco-Effectiveness’ Approach Be Adopted?.” *Journal of Cleaner Production* 16 (7): 801–8.

Abukhader, Sajed M., and Gunilla Jönson. 2004. “E-Commerce and the Environment: A Gateway to the Renewal of Greening Supply Chains.” *International Journal of Technology Management* 28 (2): 274–88.

Ackoff, Russell L. 1967. “Management Misinformation Systems.” *Management Science* 14 (4): B – 147.

Aggarwal, Charu C. 2014. “Mining Text and Social Streams: A Review.” *ACM Special Interest Group on Knowledge Discovery & Data Mining (SIGKDD) Explorations Newsletter* 15 (2): 9–19.

Aggour, Kareem S., John Interrante, and Ibrahim Gokcen. 2006. “Integrating Techniques for Event-Based Business Intelligence Gathering.” In *Papers from the 2006 Association for the Advancement of Artificial Intelligence (AAAI) Workshop*, 30–35.

Ahi, Payman, and Cory Searcy. 2013. “A Comparative Literature Analysis of Definitions for Green and Sustainable Supply Chain Management.” *Journal of Cleaner Production* 52 (August): 329–41.

———. 2014. “Assessing Sustainability in the Supply Chain: A Triple Bottom Line Approach.” *Applied Mathematical Modelling*, in press.

———. 2015. “An Analysis of Metrics Used to Measure Performance in Green and Sustainable Supply Chains.” *Journal of Cleaner Production* 86 (January): 360–77.

Ahn, David. 2006. “The Stages of Event Extraction.” In *Proceedings of the Workshop on Annotating and Reasoning About Time and Events*, edited by Branimir Boguraev, Rafael Munoz, and James Pustejovsky, 1–8. ARTE ’06. Stroudsburg: Association for Computational Linguistics.

Aktiv gegen Kinderarbeit. 2014. “OTTO Group.” http://www.aktiv-gegen-kinderarbeit.de/firma/otto-group/#footnote_8_3871 (accessed 13.01.15).

Ali, Jabir, and Sushil Kumar. 2011. “Information and Communication Technologies (ICTs) and Farmers’ Decision-Making across the Agricultural Supply Chain.” *International Journal of Information Management* 31 (2): 149–59.

Allais, Federico Blanco, Frank Hagemann, and International Programme on the Elimination of Child Labour. 2008. *Child Labour and Education: Evidence from SIMPOC Surveys*. Geneva: International Labour Organization.

Allan, James. 2002. “Introduction to Topic Detection and Tracking.” In *Topic Detection and Tracking*, edited by James Allan, 1–16. The Information Retrieval Series 12. Springer US.

Allan, James, Jaime G. Carbonell, George Doddington, Jonathan Yamron, and Yiming Yang. 1998. *Topic Detection and Tracking Pilot Study Final Report*. http://repository.cmu.edu/compsci/341/?utm_source=repository.cmu.edu%2Fcompsci%2F341&utm_medium=PDF&utm_campaign=PDFCoverPages (accessed 05.11.14).

Allan, James, Hubert Jin, Martin Rajman, Charles L. Wayne, Daniel Gildea, Victor Lavrenko, Rose Hoberman, and David Caputo. 1999. *Topic-Based Novelty Detection: 1999 Summer Workshop at CLSP, Final Report*. The Center for Language and Speech Processing (CLSP): John Hopkins University. https://www.academia.edu/2788199/Topic-based_novelty_detection_1999_summer_workshop_at_CLSP_final_report (accessed 05.11.14).

Allen, Elane, and Christopher Seaman. 2007. “Statistics Roundtable: Likert Scales and Data Analyses.” *Quality Progress*. http://asq.org/quality-progress/2007/07/statistics/likert-scales-and-data-analyses.html (accessed 10.12.14).

Amerding, Taylor. 2013. “Big Data Without Good Analytics Can Lead to Bad Decisions.” *CIO.com*. August 26. http://www.cio.com/article/738767/Big_Data_Without_Good_Analytics_Can_Lead_to_Bad_Decisions (accessed 26.09.13).

Amindoust, Atefeh, Shamsuddin Ahmed, Ali Saghafinia, and Ardeshir Bahreininejad. 2012. “Sustainable Supplier Selection: A Ranking Model Based on Fuzzy Inference System.” *Applied Soft Computing* 12 (6): 1668–77.

Amundson, J., W. Faulkner, S. Sukumara, J. Seay, and F. Badurdeen. 2012. “A Bayesian Network Based Approach for Risk Modeling to Aid in Development of Sustainable Biomass Supply Chains.” In *Computer Aided Chemical Engineering*, edited by Ian David Lockhart Bogle and Michael Fairweather, 30:152–56. 22nd European Symposium on Computer Aided Process Engineering. Elsevier.

Amundson, Joseph, Adam Brown, Matthias Grabowski, and Fazleena Badurdeen. 2014. “Life-Cycle Risk Modeling: Alternate Methods Using Bayesian Belief Networks.” *Procedia CIRP*, Variety Management in Manufacturing Proceedings of the 47th CIRP Conference on Manufacturing Systems, 17: 320–25.

Anastasi, G., M. Antonelli, A. Bechini, S. Brienza, E. D’Andrea, D. De Guglielmo, P. Ducange, B. Lazzerini, F. Marcelloni, and A. Segatori. 2013. “Urban and Social Sensing for Sustainable Mobility in Smart Cities.” In *Proceedings of Sustainable Internet and ICT for Sustainability*, 1–4. SustainIT ’13. Laxenburg: International Federation for Information Processing.

Anderson, Dan R., and Kenneth E. Anderson. 2009. “Sustainability Risk Management.” *Risk Management and Insurance Review* 12 (1): 25–38.

Anjewierden, Anjo, and Suzanne Kabel. 2001. “Automatic Indexing of PDF Documents with Ontologies.” In *Proceedings of the 13th Belgium–Netherlands Conference on Artificial Intelligence*, 23–30. BNAIC ’01. Amsterdam: Benelux Association for Artificial Intelligence.

Ansoff, Igor. 1975. “Managing Strategic Surprise by Response to Weak Signals.” *California Management Review* 18 (2): 21–33.

Anthony Alexander, Helen Walker, and Mohamed Naim. 2014. “Decision Theory in Sustainable Supply Chain Management: A Literature Review.” *Supply Chain Management: An International Journal* 19 (5/6): 504–22.

Aone, Chinatsu, and Mila Ramos-Santacruz. 2000. “REES: A Large-Scale Relation and Event Extraction System.” In *Proceedings of the Sixth Conference on Applied Natural Language Processing*, edited by Sergei Nirenburg, 76–83. ANLC ’00. Stroudsburg: Association for Computational Linguistics.

APA – Austria Presse Agentur. 2013. “APA Basisdienst Corpus (directly Provided).” http://www.apa.at/Site/Nachrichtenagentur/Die_Redaktion.de.html (accessed 23.12.13).

Appelt, Douglas E. 1999. “Introduction to Information Extraction.” *AI Communications* 12 (3): 161–72.

Arendarenko, Ernest, and Tuomo Kakkonen. 2012. “Ontology-Based Information and Event Extraction for Business Intelligence.” In *Artificial Intelligence: Methodology, Systems, and Applications*, edited by Allan Ramsay and Gennady Agre, 89–102. Lecture Notes in Computer Science 7557. Berlin Heidelberg: Springer.

Ashby, Alison, Mike Leat, and Melanie Hudson-Smith. 2012. “Making Connections: A Review of Supply Chain Management and Sustainability Literature.” *Supply Chain Management-an International Journal* 17 (5): 497–516.

Asia Monitor Resource Centre. 2008. "Child Labour in China's Informalized Urban Industrial Sector." October 15. http://www.amrc.org.hk/alu_article/informalization_of_labour_in _asia/child_labour_in_chinas_informalized_urban_industrial_s (accessed 02.04.14).

Association for Computational Linguistics. 1995. *MUC6 '95: Proceedings of the 6th Conference on Message Understanding*. MUC6 '95. Stroudsburg: Association for Computational Linguistics.

Atkinson, Martin, and Jakub Piskorski. 2011. "Frontex Real-Time News Event Extraction Framework." In *Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 749–52. KDD '11. New York: ACM.

Atkinson, Martin, Jakub Piskorski, Bruno Pouliquen, Ralf Steinberger, Hristo Tanev, and Vanni Zavarella. 2008. "Online-Monitoring of Security-Related Events." In *Proceedings of the 22nd International Conference on Computational Linguistics*, 18–22. COLING '08. Manchester: Coling 2008 Organizing Committee.

Australian Government. 2010. "Risk Based Approach to Audit Frequency." *Department of Health and Ageing*. http://www.tga.gov.au/industry/manuf-audit-frequency.htm (accessed 11.01.13).

Avornicului, Mihai-Constantin, Silviu Claudiu Popa, and Constantin Avornicului. 2010. "The Development and the Evaluation of a System for Extracting Events from Web Pages." *Informatica Economica* 14 (4): 114–23.

Ayoub, N., K. Wang, T. Kagiyama, H. Seki, and Y. Naka. 2006. "A Planning Support System for Biomass-Based Power Generation." Edited by W. Marquardt and C. Pantelides. *Computer Aided Chemical Engineering*, 16th European Symposium on Computer Aided Process Engineering and 9th International Symposium on Process Systems Engineering, 21 (2006): 1899–1904.

Azadi, Majid, Mostafa Jafarian, Reza Farzipoor Saen, and Seyed Mostafa Mirhedayatian. 2015. "A New Fuzzy DEA Model for Evaluation of Efficiency and Effectiveness of Suppliers in Sustainable Supply Chain Management Context." *Computers & Operations Research* 54 (February): 274–85.

Azadnia, Amir Hossein, Muhamad Zameri Mat Saman, and Kuan Yew Wong. 2015. "Sustainable Supplier Selection and Order Lot-Sizing: An Integrated Multi-Objective Decision-Making Process." *International Journal of Production Research* 53 (2): 383–408.

Azadnia, Amir Hossein, Muhamad Zameri Mat Saman, Kuan Yew Wong, Pezhman Ghadimi, and Norhayati Zakuan. 2012. "Sustainable Supplier Selection Based on Self-Organizing Map Neural Network and Multi Criteria Decision Making Approaches." *Procedia - Social and Behavioral Sciences* 65: 879–84.

Baccianella, Stefano, Andrea Esuli, and Fabrizio Sebastiani. 2010. "SentiWordNet 3.0: An Enhanced Lexical Resource for Sentiment Analysis and Opinion Mining." In *Proceedings of the International Conference on Language Resources and Evaluation*, edited by Nicoletta Calzolari, Khalid Choukri, Bente Maegaard, Joseph Mariani, Jan Odijk, Stelios Piperidis, Mike Rosner, and Daniel Tapias, 2200–2204. LREC '10. Paris: European Language Resources Association (ELRA).

Badan Pusat Statistik, and International Labour Organisation. 2010. *Working Children in Indonesia, 2009*. Jakarta: Statistics Indonesia, International Labour Organization.

Badurdeen, Fazleena, Mohannad Shuaib, Ken Wijekoon, Adam Brown, William Faulkner, Joseph Amundson, I. S. Jawahir, Thomas J. Goldsby, Deepak Iyengar, and Brench Boden. 2014. "Quantitative Modeling and Analysis of Supply Chain Risks Using Bayesian Theory." *Journal of Manufacturing Technology Management* 25 (5): 631–54.

Bai, Chunguang, and Joseph Sarkis. 2010. "Integrating Sustainability into Supplier Selection with Grey System and Rough Set Methodologies." *International Journal of Production Economics* 124 (1): 252–64.

———. 2012. "Performance Measurement and Evaluation for Sustainable Supply Chains Using Rough Set and Data Envelopment Analysis." In *Sustainable Supply Chains*, edited by Tonya Boone, Vaidyanathan Jayaraman, and Ram Ganeshan, 174:223–41. New York: Springer.

———. 2014. "Determining and Applying Sustainable Supplier Key Performance Indicators." *Supply Chain Management: An International Journal* 19 (3): 275–91.

Bai, Chunguang, Joseph Sarkis, and Xiaopeng Wei. 2010. "Addressing Key Sustainable Supply Chain Management Issues Using Rough Set Methodology." *Management Research Review* 33 (12): 1113–27.

Barboza, David. 2014. "Samsung Suspends China Supplier Over Child Labor Case." *The New York Times - Sinosphere Blog*, July 14. http://sinosphere.blogs.nytimes.com/2014/07/14/samsung-contractor-suspended-over-child-labor-allegations/ (accessed 18.07.14).

Barkemeyer, Ralf, Frank Figge, and Diane Holt. 2013. "Sustainability-Related Media Coverage and Socioeconomic Development: A Regional and North–South Perspective." *Environment and Planning C: Government and Policy* 31 (4): 716–40.

Barkemeyer, Ralf, Frank Figge, Diane Holt, and Barbara Wettstein. 2009. "What the Papers Say: Trends in Sustainability. A Comparative Analysis of 115 Leading National Newspapers Worldwide." *Journal of Corporate Citizenship* 2009 (33): 68–86.

Barrett, William M., Svetlana Strunjaš-Yoshikawa, and Jonathan H Bell. 2007. "Extension of Computer-Aided Process Engineering Applications to Environmental Life Cycle Assessment and Supply Chain Management." Edited by Valentin Pleşu and Paul Şerban Agachi. *Computer Aided Chemical Engineering*, 17th European Symposium on Computer Aided Process Engineering, 24: 1187–92.

Bartlett, Paul A., Denyse M. Julien, and Tim S. Baines. 2007. "Improving Supply Chain Performance through Improved Visibility." *The International Journal of Logistics Management* 18 (2): 294–313.

Baskaran, Venkatesan, Subramanian Nachiappan, and Shams Rahman. 2012. "Indian Textile Suppliers' Sustainability Evaluation Using the Grey Approach." *International Journal of Production Economics*, Green Manufacturing and Distribution in the Fashion and Apparel Industries, 135 (2): 647–58.

Basu, Gautam, Karen Butner, Eric Cope, Léa Deleris, Jin Dong, Mary Helander, Kann Katircioglu, and Bonnie Ray. 2007. "Supply Chain Risk Management: A Delicate Balancing Act." *Management Science* 51 (5): 695–711.

Bazerman, Max H. 2006. *Judgment in Managerial Decision Making*. 6th edition. J. Wiley.

Beach, Lee Roy, Terence R. Mitchell, Marcia D. Deaton, and Joyce Prothero. 1978. "Information Relevance, Content and Source Credibility in the Revision of Opinions." *Organizational Behavior and Human Performance* 21 (1): 1–16.

Bearzotti, Lorena, Erica Fernandez, Armando Guarnaschelli, Enrique Salomone, and Omar Chiotti. 2011. "Supply Chain Event Management System." In *Supply Chain Management - Applications and Simulations*, edited by Mamun Habib, 59–82. InTech.

Bebbington, Jan, Carlos Larrinaga, and Jose M. Moneva. 2008. "Corporate Social Reporting and Reputation Risk Management." *Accounting, Auditing & Accountability Journal* 21 (3): 337–61.

Becker, Joerg, Christian Brelage, Alexander Dreiling, and Michael Ribbert. 2004. "Business Process-Driven Information Requirements Engineering." In *Innovations Through Information Technology*, 2:352–61. Proceedings of the 2004 International Information Resource Management Association Conference. Hershey: Idea Group Inc (IGI).

Bennett, Shea. 2012. "REVEALED: The Top 20 Countries And Cities On Twitter [STATS] | SocialTimes." *SocialTimes*. August 13. http://www.adweek.com/socialtimes/twitter-top-countries/468210 (accessed 11.02.15).

Benoît, Catherine, and Gina Vickery-Niederman. 2011. *Social Sustainability Assessment Literature Review*. White Paper #102. The Sustainability Consortium, Arizona State

University, and University of Arkansas. http://www.sustainabilityconsortium.org/wp-content/themes/sustainability/assets/pdf/whitepapers/Social_Sustainability_Assessment.pdf (accessed 04.11.12).

Benz, Thomas, Norddin El Ghouti, Florian Krietsch, Paul Mathias, Florian Peyeron, and Philipp Themann. 2012. "The eCoMove Simulation Test Bed." *Transport Research Arena 2012* 48 (0): 1983–92.

Berners-Lee, Tim, James Hendler, and Ora Lassila. 2001. "The Semantic Web." *Scientific American* 284 (5): 28–37.

Berns, Maurice, Andrew Townend, Zayna Khayat, Balu Balagopal, Martin Reeves, Michael S. Hopkins, and Nina Kruschwitz. 2009. "The Business of Sustainability: What It Means to Managers Now." *MIT Sloan Management Review* 51 (1): 20–26.

Bhagwat, Rajat, and Milind Kumar Sharma. 2007. "Performance Measurement of Supply Chain Management: A Balanced Scorecard Approach." *Computers & Industrial Engineering* 53 (1): 43–62.

Bhutta, Khurrum S., and Faizul Huq. 2002. "Supplier Selection Problem: A Comparison of the Total Cost of Ownership and Analytic Hierarchy Process Approaches." *Supply Chain Management: An International Journal* 7 (3): 126–35.

Bibliographisches Institut GmbH. 2013. "Duden | Kind | Rechtschreibung, Bedeutung, Definition, Synonyme, Herkunft." http://www.duden.de/rechtschreibung/Kind (accessed 21.01.14).

Bieker, Thomas, and Bernhard Waxenberger. 2002. "Sustainability Balanced Scorecard and Business Ethics - Developing a Balanced Scorecard for Integrity Management." In *10th International Conference of the Greening of Industry Network*. Göteborg.

Bilder, Geoffrey. 2010. "Data and Text Mining: The Search for Unknown Knowns." *UCL Department of Information Studies*. http://www.youtube.com/watch?v=F9oOhXVVPLk (accessed 14.12.12).

Bizer, Christian, Tom Heath, and Tim Berners-Lee. 2009. "Linked Data-the Story so Far." *International Journal on Semantic Web and Information Systems (IJSWIS)* 5 (3): 1–22.

Björk, Anders, Martin Erlandsson, Janne Häkli, Kaarle Jaakkola, Åsa Nilsson, Kaj Nummila, Ville Puntanen, and Antti Sirkka. 2011. "Monitoring Environmental Performance of the Forestry Supply Chain Using RFID." *Computers in Industry* 62 (8–9): 830–41.

Blackhurst, Jennifer V., Kevin P. Scheibe, and Danny J. Johnson. 2008. "Supplier Risk Assessment and Monitoring for the Automotive Industry." *International Journal of Physical Distribution & Logistics Management* 38 (2): 143–65.

Black, William J., John McNaught, Argyris Vasilakopoulos, Kalliopi Zervanou, Babis Theodoulidis, and Fabio Rinaldi. 2005. *CAFETIERE Conceptual Annotations for Facts, Events, Terms, Individual Entities, and Relations*. Parmenides Technical Report TR-U4.3.1. http://www.nactem.ac.uk/files/phatfile/cafetiere-report.pdf (accessed 29.01.15).

Blasch, Erik, Kathryn B. Laskey, Anne-Laure Jousselme, Valentina Dragos, Paulo CG Costa, and Jean Dezert. 2013. "URREF Reliability versus Credibility in Information Fusion (STANAG 2511)." In *Proceedings of 16th International Conference on Information Fusion*, 1600–1607. FUSION '13. IEEE.

Boilerpipe. 2011. "Boilerpipe - Boilerplate Removal and Fulltext Extraction from HTML Pages - Version 1.2.0." https://code.google.com/p/boilerpipe/ (accessed 13.02.15).

Borsato, Milton. 2014. "Bridging the Gap between Product Lifecycle Management and Sustainability in Manufacturing through Ontology Building." *Computers in Industry* 65 (2): 258–69.

Bortz, Jürgen, and Nicola Döring. 2006. *Forschungsmethoden und Evaluation: für Human- und Sozialwissenschaftler: Fur Human- Und Sozialwissenschaftler*. 4th ed. Heidelberg: Springer.

Bosona, Techane, Ingrid Nordmark, Girma Gebresenbet, and David Ljungberg. 2013. “GIS-Based Analysis of Integrated Food Distribution Network in Local Food Supply Chain.” *International Journal of Business and Management* 8 (17): 13–34.

Botsis, Taxiarchis, Thomas Buttolph, Michael D. Nguyen, Scott Winiecki, Emily Jane Woo, and Robert Ball. 2012. “Vaccine Adverse Event Text Mining System for Extracting Features from Vaccine Safety Reports.” *Journal of the American Medical Informatics Association* 19 (6): 1011–18.

Botsis, Taxiarchis, Michael D. Nguyen, Emily Jane Woo, Marianthi Markatou, and Robert Ball. 2011. “Text Mining for the Vaccine Adverse Event Reporting System: Medical Text Classification Using Informative Feature Selection.” *Journal of the American Medical Informatics Association* 18 (5): 631–38.

Bourdillon, Michael, Deborah Levison, William Myers, and Ben White. 2010. *Rights and Wrongs of Children’s Work*. New Brunswick: Rutgers University Press.

Boyd, D. Eric, Robert E. Spekman, John W. Kamauff, and Patricia Werhane. 2007. “Corporate Social Responsibility in Global Supply Chains: A Procedural Justice Perspective.” *Long Range Planning* 40 (3): 341–56.

BPS Statistic Indonesia. 2008. *THE INDONESIAN 2006 ECONOMIC CENSUS*. The 12th East Asian Statistical Conference. Tokyo. http://www.stat.go.jp/english/info/meetings/eastasia/pdf/t2indpa.pdf (accessed 20.08.14).

Brandenburg, Marcus, Kannan Govindan, Joseph Sarkis, and Stefan Seuring. 2014. “Quantitative Models for Sustainable Supply Chain Management: Developments and Directions.” *European Journal of Operational Research*, Eco-Efficient Green Supply Chain Management, 233 (2): 299–312.

Brank, Janez, Dunja Mladenić, and Marko Grobelnik. 2006. “Gold Standard Based Ontology Evaluation Using Instance Assignment.” In *Proceedings of the 4th International EON Workshop 2006*. EON ’06. Edinburgh: CEUR Workshop Proceedings.

Bremer, Jenifer, and John Udovich. 2001. “Alternative Approaches to Supply Chain Compliance Monitoring.” *Journal of Fashion Marketing and Management* 5 (4): 333–52.

British Telecommunications. 2012. “Assessing Environmental and Social Risks.” http://www.btplc.com/Responsiblebusiness/Ourstory/Sustainabilityreport/section/index.aspx?sectionid=68d95d30-36ed-4f1f-a138-6344031a2bbe (accessed 04.12.12).

Brooker, Peter. 2011. “Experts, Bayesian Belief Networks, Rare Events and Aviation Risk Estimates.” *Safety Science* 49 (8–9): 1142–55.

Brooks, Stoney, Xuequn Wang, and Saonee Sarker. 2012. “Unpacking Green IS: A Review of the Existing Literature and Directions for the Future.” In *Green Business Process Management*, edited by Jan vom Brocke, Stefan Seidel, and Jan Recker, 15–37. Berlin Heidelberg: Springer.

Bühl, Achim, and Peter Zöfel. 2002. *SPSS 11: Eine Einführung in Die Moderne Datenanalyse Unter Windows*. 1st ed. München: Addison-Wesley.

Bullinger, Hans-Jörg, J.von Steinaecker, and A Weller. 1999. “Concepts and Methods for a Production Integrated Environmental Protection.” *International Journal of Production Economics* 60–61 (April): 35–42.

Business Social Compliance Initiative. 2009. *BSCI Code of Conduct English*. http://www.bsci-intl.org/system/files/2_bsci_codeofconduct_english_pdf_1.pdf (accessed 02.01.13).

Büyüközkan, Gülçin, and Gizem Çifçi. 2011. “A Novel Fuzzy Multi-Criteria Decision Framework for Sustainable Supplier Selection with Incomplete Information.” *Computers in Industry*, Fuzziness in Industry and Applications, 62 (2): 164–74.

Caceres, Cesar, Alberto Fernandez, Sascha Ossowski, and Matteo Vasirani. 2006. “Agent-Based Semantic Service Discovery for Healthcare: An Organizational Approach.” *IEEE Intelligent Systems* 21 (6): 11–20.

Cachon, Gerard, and Christian Terwiesch. 2013. *Matching Supply with Demand: An Introduction to Operations Management*. 3rd ed. Mcgraw-Hill Higher Education.

Cahill, Lawrence. 2011. “Using Risk Factors to Determine EHS Audit Frequency - EHS Journal.” *EHS Journal*. http://ehsjournal.org/http:/ehsjournal.org/lawrence-b-cahill/using-risk-factors-to-determine-ehs-audit-frequency/2011/ (accessed 11.01.13).

Cao, Ya-nan, Peng Zhang, Jing Guo, and Li Guo. 2014. “Mining Large-Scale Event Knowledge from Web Text.” *Procedia Computer Science*, 2014 International Conference on Computational Science, 29: 478–87.

Capet, Philippe, Thomas Delavallade, Takuya Nakamura, Agnes Sandor, Cedric Tarsitano, and Stavroula Voyatzi. 2008. “A Risk Assessment System with Automatic Extraction of Event Types.” In *Intelligent Information Processing IV*, edited by Zhongzhi Shi, E. Mercier-Laurent, and D. Leake, 220–29. IFIP – The International Federation for Information Processing 288. Springer US.

Carlson, R, M Erixon, P Forsberg, and A.-C Pålsson. 2001. “System for Integrated Business Environmental Information Management.” *Advances in Environmental Research* 5 (4): 369–75.

Carpineto, Claudio, and Giovanni Romano. 2012. “A Survey of Automatic Query Expansion in Information Retrieval.” *ACM Computing Surveys* 44 (1): 1:1–1:50.

Carter, Craig R., and P. Liane Easton. 2011. “Sustainable Supply Chain Management: Evolution and Future Directions.” *International Journal of Physical Distribution & Logistics Management* 41 (1): 46–62.

Carter, Craig R., and Marianne M. Jennings. 2002. “Social Responsibility and Supply Chain Relationships.” *Transportation Research Part E: Logistics and Transportation Review* 38 (1): 37–52.

———. 2004. “The Role of Purchasing in Corporate Social Responsibility: A Structural Equation Analysis.” *Journal of Business Logistics* 25 (1): 145–86.

Carter, Craig R., and Dale S. Rogers. 2008. “A Framework of Sustainable Supply Chain Management: Moving toward New Theory.” *International Journal of Physical Distribution & Logistics Management* 38 (5): 360–87.

Castellanos, Malu, Chetan Gupta, Song Wang, Umeshwar Dayal, and Miguel Durazo. 2012. “A Platform for Situational Awareness in Operational BI.” *Decision Support Systems* 52 (4): 869–83.

Ceres, and Sustainalytics. 2014. *Gaining Ground: Corporate Progress on the Ceres Roadmap for Sustainability*. http://www.ceres.org/resources/reports/gaining-ground-corporate-progress-on-the-ceres-roadmap-for-sustainability/view (accessed 15.01.15).

Chang, Chia-Hua. 2013. “An Intelligent Supplier Selection System Based on Self Organizing Map, Rough Set Theory, and Bayesian Belief Network.” *International Journal of Electronic Business Management* 11 (2): 100–112.

Chang-Shing Lee, Zhi-Wei Jian, and Lin-Kai Huang. 2005. “A Fuzzy Ontology and Its Application to News Summarization.” *IEEE Transactions on Systems, Man & Cybernetics: Part B* 35 (5): 859–80.

Charniak, Eugene. 1991. “Bayesian Networks without Tears.” *Artificial Intelligence Magazine* 12 (4): 50–63.

Chen, Ming-Kuen, Teng-Wang Tai, and Tsu-Yi Hung. 2012. “Component Selection System for Green Supply Chain.” *Expert Systems with Applications* 39 (5): 5687–5701.

Chen, Wei, Chun Chen, Li-jun Zhang, Can Wang, and Jia-jun Bu. 2010. “Online Detection of Bursty Events and Their Evolution in News Streams.” *Journal of Zhejiang University-Science C-Computers & Electronics* 11 (5): 340–55.

Chen, Ye, and Yun Peng. 2003. “An Extended Bayesian Belief Network Model of Multi-Agent Systems for Supply Chain Managements.” In *Innovative Concepts for Agent-Based Systems*, edited by Walt Truszkowski, Mike Hinchey, and Chris Rouff, 335–46. Lecture Notes in Computer Science 2564. Berlin Heidelberg: Springer.

Chen, Yong, Yu Wang, Xiaolian Duan, Huatang Zhang, and Yuyue Jia. 2014. “Text Mining for Information Screen in Risk Assessment of Environmental Endocrine Disruptive Chemicals.” In *Advances in Computer Science and Its Applications*, edited by Hwa

Young Jeong, Mohammad S. Obaidat, Neil Y. Yen, and James J. Park, 967–72. Lecture Notes in Electrical Engineering 279. Berlin Heidelberg: Springer.

Chih-Ping, Wei, and Lee Yen-Hsien. 2004. "Event Detection from Online News Documents for Supporting Environmental Scanning." *Decision Support Systems* 36 (4): 385–401.

Chiou, Cherng-Ying, Shih-Hung Chou, and Chun-Yuan Yeh. 2011. "Using Fuzzy AHP in Selecting and Prioritizing Sustainable Supplier on CSR for Taiwan's Electronics Industry." *Journal of Information & Optimization Sciences* 32 (5): 1135–53.

Christopher, Martin, and Helen Peck. 2004. "Building the Resilient Supply Chain." *The International Journal of Logistics Management* 15 (2): 1–14.

Chua, Alton Y. K., Khasfariyati Razikin, and Dion H. Goh. 2011. "Social Tags as News Event Detectors." *Journal of Information Science* 37 (1): 3–18.

Ciliberti, Francesco, Pierpaolo Pontrandolfo, and Barbara Scozzi. 2008a. "Logistics Social Responsibility: Standard Adoption and Practices in Italian Companies." *International Journal of Production Economics* 113 (1): 88–106.

———. 2008b. "Investigating Corporate Social Responsibility in Supply Chains: A SME Perspective." *Journal of Cleaner Production* 16 (15): 1579–88.

Cimiano, Philipp, Steffen Staab, and Julien Tane. 2003. "Automatic Acquisition of Taxonomies from Text: FCA Meets NLP." In *Proceedings of the International Workshop & Tutorial on Adaptive Text Extraction and Mining Held in Conjunction with the 14th European Conference on Machine Learning and the 7th European Conference on Principles and Practice of Knowledge Discovery in Databases*, 10–17. ATEM '03. Cavtat–Dubrovnik.

Clausen, Uwe, Christiane Geiger, and Carl Behmer. 2012. "Green Corridors by Means of ICT Applications." *Procedia - Social and Behavioral Sciences* 48 (January): 1877–86.

Clift, Roland. 2004. "Metrics for Supply Chain Sustainability." In *Technological Choices for Sustainability*, edited by Dr Subhas K. Sikdar, Prof Dr Peter Glavič, and Prof Dr Ravi Jain, 239–53. Berlin Heidelberg: Springer.

Closs, David J., Cheri Speier, and Nathan Meacham. 2011. "Sustainability to Support End-to-End Value Chains: The Role of Supply Chain Management." *Journal of the Academy of Marketing Science* 39 (1): 101–16.

Clough, Paul, and Mark Sanderson. 2013. "Evaluating the Performance of Information Retrieval Systems Using Test Collections." *Information Research* 18 (2): paper 582.

Colas, Fabrice, and Pavel Brazdil. 2006. "Comparison of SVM and Some Older Classification Algorithms in Text Classification Tasks." In *Artificial Intelligence in Theory and Practice*, edited by Max Bramer, 169–78. IFIP International Federation for Information Processing 217. Springer US.

Collingsworth, Terry. 2003. "Alien Tort Claims Act: Holding Corporations Accountable for Human Rights Violations in the Global Economy | International Labor Rights Forum." *Human Rights Dialogue Spring 2003*. May 1. http://www.laborrights.org/in-the-news/alien-tort-claims-act-holding-corporations-accountable-human-rights-violations-global (accessed 01.02.15).

Committee of Sponsoring Organizations of the Treadway Commission. 2004. *Enterprise Risk Management — Integrated Framework - Executive Summary*. http://www.coso.org/documents/coso_erm_executivesummary.pdf (accessed 14.08.14).

Conway, Mike, Son Doan, Ai Kawazoe, and Nigel Collier. 2009. "Classifying Disease Outbreak Reports Using N-Grams and Semantic Features." *International Journal of Medical Informatics*, Mining of Clinical and Biomedical Text and Data Special Issue, 78 (12): e47–58.

Coppola, Bonaventura, Aldo Gangemi, Alfio Gliozzo, Davide Picca, and Valentina Presutti. 2009. "Frame Detection over the Semantic Web." In *The Semantic Web: Research and Applications*, edited by Lora Aroyo, Paolo Traverso, Fabio Ciravegna, Philipp Cimiano, Tom Heath, Eero Hyvönen, Riichiro Mizoguchi, Eyal Oren, Marta Sabou, and Elena Simperl, 126–42. Lecture Notes in Computer Science 5554. Berlin Heidelberg: Springer.

Correa, M., C. Bielza, and J. Pamies-Teixeira. 2009. “Comparison of Bayesian Networks and Artificial Neural Networks for Quality Detection in a Machining Process.” *Expert Systems with Applications* 36 (3, Part 2): 7270–79.

Cortes, Corinna, and Vladimir Vapnik. 1995. “Support-Vector Networks.” *Machine Learning* 20 (3): 273–97.

Cotton Connect. 2012. “Introduction.” http://www.cottonconnect.org/what-we-do/introduction.aspx (accessed 28.12.12).

Cousins, Paul D., Richard C. Lamming, and Frances Bowen. 2004. “The Role of Risk in Environment-Related Supplier Initiatives.” *International Journal of Operations & Production Management* 24 (6): 554–65.

Crystal, David. 2004. “Subcontinent Raises Its Voice.” *The Guardian*. November 19. http://www.theguardian.com/education/2004/nov/19/tefl (accessed 11.02.15).

Cunningham, Hamish, Kalina Bončeva, and Diana Maynard. 2011. *Text Processing with GATE*. Sheffield: University of Sheffield Dept. of Computer Science.

Cunningham, Hamish, Allan Hanbury, and Stefan Rüger. 2010. “Scaling up High-Value Retrieval to Medium-Volume Data.” In *Advances in Multidisciplinary Retrieval (the 1st Information Retrieval Facility Conference)*, edited by Hamish Cunningham, Allan Hanbury, and Stefan Rüger. Lecture Notes in Computer Science 6107. Vienna: Springer.

Cunningham, Hamish, Diana Maynard, Kalina Bontcheva, Valentin Tablan, Niraj Aswani, Ian Roberts, Genevieve Gorrell, et al. 2011. *Text Processing with GATE (Version 6)*.

Cunningham, Hamish, Diana Maynard, Kalina Bontcheva, Valentin Tablan, and Pierre Isabelle. 2002. “GATE: A Framework and Graphical Development Environment for Robust NLP Tools and Applications.” In *Proceedings of the 40th Anniversary Meeting of the Association for Computational Linguistics*, 168–75. ACL ’02. Stroudsburg: Association for Computational Linguistics.

Cunningham, Hamish, Diana Maynard, and Valentin Tablan. 2000. *JAPE: A Java Annotation Patterns Engine (Second Edition)*. Research Memorandum CS–00–10. Department of Computer Science, University of Sheffield. http://www.dcs.shef.ac.uk/ diana/Papers/jape.ps (accessed 16.02.15).

Cunningham, Hamish, Valentin Tablan, Angus Roberts, and Kalina Bontcheva. 2013. “Getting More Out of Biomedical Documents with GATE’s Full Lifecycle Open Source Text Analytics.” *PLOS Computational Biology* 9 (2): e1002854, 1–16.

Daiber, Joachim, Max Jakob, Chris Hokamp, and Pablo N. Mendes. 2013. “Improving Efficiency and Accuracy in Multilingual Entity Extraction.” In *Proceedings of the 9th International Conference on Semantic Systems*. I-Semantics ’13. New York: ACM.

Dai, Jing, and Jennifer Blackhurst. 2012. “A Four-Phase AHP–QFD Approach for Supplier Assessment: A Sustainability Perspective.” *International Journal of Production Research* 50 (19): 5474–90.

Danilova, Vera, Mikhail Alexandrov, and Xavier Blanco. 2014. “A Survey of Multilingual Event Extraction from Text.” In *Natural Language Processing and Information Systems*, edited by Elisabeth Métais, Mathieu Roche, and Maguelonne Teisseire, 85–88. Lecture Notes in Computer Science 8455. Springer International Publishing.

Dao, Viet, Ian Langella, and Jerry Carbo. 2011. “From Green to Sustainability: Information Technology and an Integrated Sustainability Framework.” *The Journal of Strategic Information Systems* 20 (1): 63–79.

Dasgupta, Anirban, Petros Drineas, Boulos Harb, Vanja Josifovski, and Michael W. Mahoney. 2007. “Feature Selection Methods for Text Classification.” In *Proceedings of the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 230–39. KDD ’07. New York: ACM.

Das, Sanjiv. 2010. *News Analytics: Framework, Techniques and Metrics*. 11-08. SCU Leavey School of Business Research Paper. http://papers.ssrn.com/sol3/papers.cfm? abstract_id=1814258 (accessed 22.11.13).

Davies, Ian, Robert Mason, and Chandra Lalwani. 2007. "Assessing the Impact of ICT on UK General Haulage Companies." *International Journal of Production Economics* 106 (1): 12–27.

Davis, Fred D. 1989. "Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology." *MIS Quarterly* 13 (3): 319–40.

Davis, Fred D., Richard P. Bagozzi, and Paul R. Warshaw. 1989. "User Acceptance of Computer Technology: A Comparison of Two Theoretical Models." *Management Science* 35 (8): 982–1003.

De Boer, Luitzen, Eva Labro, and Pierangela Morlacchi. 2001. "A Review of Methods Supporting Supplier Selection." *European Journal of Purchasing & Supply Management* 7 (2): 75–89.

Dellschaft, Klaas, and Steffen Staab. 2006. "On How to Perform a Gold Standard Based Evaluation of Ontology Learning." In *The Semantic Web - ISWC 2006*, edited by Isabel Cruz, Stefan Decker, Dean Allemang, Chris Preist, Daniel Schwabe, Peter Mika, Mike Uschold, and Lora M. Aroyo, 228–41. Lecture Notes in Computer Science 4273. Berlin Heidelberg: Springer.

Denecke, Kerstin. 2012. *Event-Driven Surveillance: Possibilities and Challenges*. Heidelberg New York Dordrecht London: Springer Science & Business Media.

Dey, Asoke, Paul LaGuardia, and Mahesh Srinivasan. 2011. "Building Sustainability in Logistics Operations: A Research Agenda." *Management Research Review* 34 (11): 1237–59.

Dhanda, Kanwalroop Kathy, and Ronald Paul Hill. 2005. "The Role of Information Technology and Systems in Reverse Logistics: A Case Study." *International Journal of Technology Management* 31 (1-2): 140–51.

Diallo, Yacouba, Alex Etienne, Farhad Mehran, International Labour Office, and International Programme on the Elimination of Child Labour. 2013. *Global Child Labour Trends 2008 to 2012*. Geneva: International Labour Office.

Diallo, Yacouba, International Labour Office, International Programme on the Elimination of Child Labour, and Statistical Information and Monitoring Programme on Child Labour. 2010. *Global Child Labour Developments: Measuring Trends from 2004 to 2008*. Geneva: ILO.

Ding, Grace K. C. 2005. "Developing a Multicriteria Approach for the Measurement of Sustainable Performance." *Building Research & Information* 33 (1): 3–16.

Dinopoulos, Elias, and Laixun Zhao. 2007. "Child Labor and Globalization." *Journal of Labor Economics* 25 (3): 553–79.

Dobson, Ian, and Jim Hietala. 2011. *Risk Management: The Open Group Guide*. Berkshire: Van Haren Publishing.

Dogan, Ibrahim. 2012. "Analysis of Facility Location Model Using Bayesian Networks." *Expert Systems with Applications* 39 (1): 1092–1104.

Dogan, Ibrahim, and Nezir Aydin. 2011. "Combining Bayesian Networks and Total Cost of Ownership Method for Supplier Selection Analysis." *Computers & Industrial Engineering* 61 (4): 1072–85.

Donaldson Soberanis, Ivy Elizabeth. 2010. "An Extended Bayesian Network Approach for Analyzing Supply Chain Disruptions." Dissertation, University of Iowa.

Dorman, Peter. 2008. *Child Labour, Education and Health: A Review of the Literature*. Geneva: International Labour Organization.

Dotoli, Mariagrazia, Maria Pia Fanti, Carlo Meloni, and MengChu Zhou. 2006. "Design and Optimization of Integrated E-Supply Chain for Agile and Environmentally Conscious Manufacturing." *IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans* 36 (1): 62–75.

Dou, Yijie, and Joseph Sarkis. 2008. *A Joint Location and Outsourcing Sustainability Analysis for a Strategic Offshoring Decision*. SSRN Scholarly Paper ID 1125496. Rochester: Social Science Research Network. http://papers.ssrn.com/abstract=1125496 (accessed 20.01.15).

Draicchio, Francesco, Aldo Gangemi, Valentina Presutti, and Andrea Giovanni Nuzzolese. 2013. “FRED: From Natural Language Text to RDF and OWL in One Click.” In *The Semantic Web: ESWC 2013 Satellite Events*, edited by Philipp Cimiano, Miriam Fernández, Vanessa Lopez, Stefan Schlobach, and Johanna Völker, 263–67. Lecture Notes in Computer Science 7955. Berlin Heidelberg: Springer.

Dreyer, Louise, and Michael Hauschild. 2006. “Scoping Must Be Done in Accordance with the Goal Definition, Also in Social LCA.” *The International Journal of Life Cycle Assessment* 11 (2): 87–87.

Dreyer, Louise, Michael Hauschild, and Jens Schierbeck. 2005. “A Framework for Social Life Cycle Impact Assessment.” *The International Journal of Life Cycle Assessment* 11 (2): 88–97.

Dreyer, Louise, Michael Z. Hauschild, and Jens Schierbeck. 2010a. “Characterisation of Social Impacts in LCA.” *The International Journal of Life Cycle Assessment* 15 (3): 247–59.

———. 2010b. “Characterisation of Social Impacts in LCA. Part 2: Implementation in Six Company Case Studies.” *The International Journal of Life Cycle Assessment* 15 (4): 385–402.

Duespohl, Meike, Sina Frank, and Petra Doell. 2012. “A Review of Bayesian Networks as a Participatory Modeling Approach in Support of Sustainable Environmental Management.” *Journal of Sustainable Development* 5 (12): 1–18.

Du, Mian, Jussi Kangasharju, Ossi Karkulahti, Lidia Pivovarova, and Roman Yangarber. 2013. “Combined Analysis of News and Twitter Messages.” In *Joint Workshop on NLP&LOD and SWAIE: Semantic Web, Linked Open Data and Information Extraction*, edited by Diana Maynard, Marieke van Erp, Brian Davis, Petya Osenova, Kiril Simov, Georgi Georgiev, and Preslav Nakov, 41–48. Hissar: INCOMA Ltd. Shoumen.

Dutkiewicz, Jakub, Maciej Nowak, and Czeslaw Jedrzejek. 2014. “R2E: Rule-Based Event Extractor.” In *Proceedings of the RuleML 2014 Challenge and the RuleML 2014 Doctoral Consortium*, edited by Theodore Patkos, Adam Wyner, and Adrian Giurca. Prague: CEUR Workshop Proceedings.

Dyer, Jeffrey H., and Harbir Singh. 1998. “The Relational View: Cooperative Strategy and Sources of Interorganizational Competitive Advantage.” *The Academy of Management Review* 23 (4): 660–79.

Dyllick, Thomas, and Kai Hockerts. 2002. “Beyond the Business Case for Corporate Sustainability.” *Business Strategy and the Environment* 11 (2): 130–41.

ECLT Foundation. 2013. “History & Background « ECLT Foundation.” http://www.eclt.org/site/about-child-labour/introducing-child-labour/#anchor3 (accessed 30.10.13).

Edmonds, Eric V. 2008. *Defining Child Labour: A Review of the Definitions of Child Labour in Policy Research*. Geneva: ILO.

Edmonds, Eric V., and Nina Pavcnik. 2005. “Child Labor in the Global Economy.” *The Journal of Economic Perspectives* 19 (1): 199–220.

Edwards, James W., Andrew Lyons, and Dennis Kehoe. 2004. “A Web-Enabled Demand Planning System (DPS) to Enable Coordinated Re-Manufactur of Automotive Components.” *International Journal of Internet and Enterprise Management* 2 (4): 366–82.

Edwards, Ward. 1954. “The Theory of Decision Making.” *Psychological Bulletin* 51 (4): 380–417.

Egels-Zandén, Niklas. 2007. “Suppliers’ Compliance with MNCs’ Codes of Conduct: Behind the Scenes at Chinese Toy Suppliers.” *Journal of Business Ethics* 75 (1): 45–62.

———. 2014. “Revisiting Supplier Compliance with MNC Codes of Conduct: Recoupling Policy and Practice at Chinese Toy Suppliers.” *Journal of Business Ethics* 119 (1): 59–75.

Ehrig, Marc, and Jérôme Euzenat. 2005. “Relaxed Precision and Recall for Ontology Matching.” In *Proceedings of K-Cap 2005 Workshop on Integrating Ontology*, edited by Benjamin Ashpole, Marc Ehrig, Jérôme Euzenat, and Heiner Stuckenschmidt, 25–32. K-Cap ’05. Banff: CEUR Workshop Proceedings.

Elkington, John. 1998. *Cannibals with Forks: The Triple Bottom Line of 21st Century Business*. New Society Publishers.

Elliot, Steve, and Derek Binney. 2008. “Environmentally Sustainable ICT: Developing Corporate Capabilities and an Industry-Relevant IS Research Agenda.” In *Proceedings of Pacific Asia Conference on Information Systems*, Paper 209. PACIS ’08. Atlanta: Association for Information Systems (AIS).

Ellis, A. F., International Labour Organisation, and International Programme on the Elimination of Child Labour. 2005. *Child Labour – The Role of Labour Inspection: A Resource Booklet for Labour Inspectors, Other Enforcement Agencies and Key ILO Partners*. Geneva: International Labour Organisation. http://www.ilo.org/ipecinfo/product/download.do?type=document&id=10010 (accessed 30.10.13).

Elloumi, Samir, Ali Jaoua, Fethi Ferjani, Nasredine Semmar, Romaric Besançon, Jihad Al-Jaam, and Helmi Hammami. 2013. “General Learning Approach for Event Extraction: Case of Management Change Event.” *Journal of Information Science* 39 (2).

Emmelhainz, Margaret A., and Ronald J. Adams. 1999. “The Apparel Industry Response to ‘Sweatshop’ Concerns: A Review and Analysis of Codes of Conduct.” *Journal of Supply Chain Management* 35 (2): 51–57.

Erol, Ismail, Safiye Sencer, and Ramazan Sari. 2011. “A New Fuzzy Multi-Criteria Framework for Measuring Sustainability Performance of a Supply Chain.” *Ecological Economics* 70 (6): 1088–1100.

Erwin, Patrick M. 2011. “Corporate Codes of Conduct: The Effects of Code Content and Quality on Ethical Performance.” *Journal of Business Ethics* 99 (4): 535–48.

Esuli, Andrea, and Fabrizio Sebastiani. 2006. “Sentiwordnet: A Publicly Available Lexical Resource for Opinion Mining.” In *Proceedings of the International Conference on Language Resources and Evaluation*, edited by Nicoletta Calzolari, Aldo Gangemi, Bente Maegaard, Joseph Mariani, Jan Odijk, and Daniel Tapias, 6:417–22. LREC ’06. European Language Resources Association, Evaluation and Language resources Distribution Agency, Istituto di Linguistica Computazionale.

Ethical Trading Initiative. 2012. *Child Labour Briefing*. http://www.ethicaltrade.org/sites/default/files/resources/Child%20labour%20briefing.pdf (accessed 30.10.13).

European Commission. 2012. “The 5 Targets for the EU in 2020.” http://ec.europa.eu/ europe2020/europe-2020-in-a-nutshell/targets/index_en.htm (accessed 20.10.12).

European Statistical System. 2012. *Quality Assurance Framework of the European Statistical System - Version 1.1*. http://epp.eurostat.ec.europa.eu/cache/ITY_PUBLIC/QAF_2012/EN/QAF_2012-EN.PDF (accessed 02.10.14).

Eurostat. 2008. *NACE Rev. 2: Statistical Classification of Economic Activities*. Luxembourg: Office for Official Publications of the European Communities.

Euzenat, Jérôme. 2007. “Semantic Precision and Recall for Ontology Alignment Evaluation.” In *Proceedings of International Joint Conferences on Artificial Intelligence*, 348–53. IJCAI ’07. California: International Joint Conferences on Artificial Intelligence Organization.

Euzenat, Jérôme, and Pavel Shvaiko. 2013. *Ontology Matching*. Berlin, Heidelberg: Springer Berlin Heidelberg.

Ezawa, Kazuo J., and Til Schuermann. 1995. “Fraud/Uncollectible Debt Detection Using a Bayesian Network Based Learning System: A Rare Binary Outcome with Mixed Data

Structures." In *Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence*, 157–66. UAI'95. San Francisco: Morgan Kaufmann Publishers Inc.

Fabbe-Costes, Nathalie, Christine Roussat, and Jacques Colin. 2011. "Future Sustainable Supply Chains: What Should Companies Scan?" *International Journal of Physical Distribution & Logistics Management* 41 (3): 228–52.

Fabbe-Costes, Nathalie, Christine Roussat, Margaret Taylor, and Andrew Taylor. 2014. "Sustainable Supply Chains: A Framework for Environmental Scanning Practices." *International Journal of Operations & Production Management* 34 (5): 664–94.

Fair Labor Association. 2011. *FLA Workplace Code of Conduct and Compliance Benchmarks*. http://www.fairlabor.org/sites/default/files/fla_complete_code_and_benchmarks.pdf (accessed 25.10.13).

Fares, Jean, and Dhushyanth Raju. 2007. *Child Labor across the Developing World: Patterns and Correlations*. WPS4119. World Bank Policy Research Working Paper. http://papers.ssrn.com/sol3/papers.cfm?abstract_id=959765 (accessed 05.02.14).

Faris, Craig, Brian Gilbert, Brendan LeBlanc, Brian Ballou, and Dan L. Heitger. 2013. *Demystifying Sustainability Risk - Integrating the Triple Bottom Line into an Enterprise Risk Management Program*. Committee of Sponsoring Organizations of the Treadway Commission. http://www.coso.org/documents/COSO-ERM%20Demystifying%20Sustainability%20Risk_Full%20WEB.pdf (accessed 14.08.14).

Fei, Wang, and Wang Shilei. 2010. "Risk Early Warning of Supply Chain Based on D-S Evidence Theory." In *2010 International Conference on Artificial Intelligence and Education (ICAIE)*, 418–20.

Fellbaum, Christiane. 1998. *WordNet: An Electronic Lexical Database*. MIT Press.

Ferber, Reginald. 1997. "Automated Indexing with Thesaurus Descriptors: A Co-Occurrence Based Approach to Multilingual Retrieval." In *Research and Advanced Technology for Digital Libraries*, edited by Carol Peters and Costantino Thanos, 233–52. Lecture Notes in Computer Science 1324. Berlin Heidelberg: Springer.

Fernández, Erica, Enrique Salomone, and Omar Chiotti. 2010. "Model Based on Bayesian Networks for Monitoring Events in a Supply Chain." In *Advances in Production Management Systems. New Challenges, New Approaches*, edited by Bruno Vallespir and Thècle Alix, 358–65. IFIP Advances in Information and Communication Technology 338. Berlin Heidelberg: Springer.

Fernando, Samuel, and Mark Stevenson. 2012. "Mapping WordNet Synsets to Wikipedia Articles." In *Proceedings of the International Conference on Language Resources and Evaluation*, edited by Nicoletta Calzolari, Khalid Choukri, Thierry Declerck, Mehmet Uğur Doğan, Bente Maegaard, Joseph Mariani, Asuncion Moreno, Jan Odijk, and Stelios Piperidis, 590–96. LREC '12. Paris: European Language Resources Association (ELRA).

Ferreira, Luciano, and Denis Borenstein. 2012. "A Fuzzy-Bayesian Model for Supplier Selection." *Expert Systems with Applications* 39 (9): 7834–44.

Fiscus, Jonathan G., and George R. Doddington. 2002. "Topic Detection and Tracking." In , edited by James Allan, 17–31. Norwell: Kluwer Academic Publishers.

Foerstl, Kai, Carsten Reuter, Evi Hartmann, and Constantin Blome. 2010. "Managing Supplier Sustainability Risks in a Dynamically Changing environment—Sustainable Supplier Management in the Chemical Industry." *Journal of Purchasing and Supply Management* 16 (2): 118–30.

Fokkens, Antske, Marieke Van Erp, Piek Vossen, Sara Tonelli, Willem Robert Van Hage, Luciano Serafini, Rachele Sprugnoli, and Jesper Hoeksema. 2013. "GAF: A Grounded Annotation Framework for Events." In *Proceedings of the 1st Workshop on EVENTS: Definition, Detection, Coreference, and Representation*, edited by Eduard Hovy, Teruko Mitamura, and Martha Palmer, 11–20. Stroudsburg: Association for Computational Linguistics.

Fombrun, Charles J., Naomi A. Gardberg, and Michael L. Barnett. 2000. "Opportunity Platforms and Safety Nets: Corporate Citizenship and Reputational Risk." *Business and Society Review* 105 (1): 85–106.

Foran, Barney, Manfred Lenzen, Christopher Dey, and Marcela Bilek. 2005. "Integrating Sustainable Chain Management with Triple Bottom Line Accounting." *Ecological Economics* 52 (2): 143–57.

Forest Stewardship Council. 2012. *FSC Principles and Criteria for Forest Stewardship*. http://igi.fsc.org/download.fsc-pc-v5-with-explanatory-notes.4.pdf (accessed 02.01.13).

Forman, George. 2003. "An Extensive Empirical Study of Feature Selection Metrics for Text Classification." *Journal of Machine Learning Research* 3 (March): 1289–1305.

Fors, Heather Congdon. 2012. "Child Labour: A Review of Recent Theory and Evidence with Policy Implications." *Journal of Economic Surveys* 26 (4): 570–93.

Forstmoser, Peter P.F., and N.H. Herger. 2006. "Managing Reputational Risk: A Reinsurer's View." *Geneva Papers on Risk & Insurance - Issues & Practice* 31 (3): 409–24.

Fozza, Sara, and Valerio Recagno. 2012. "Sustainable Technologies and Innovation for Green Corridors: Survey and Application." *Transport Research Arena 2012* 48 (0): 1753–63.

Frasincar, Flavius, Jethro Borsje, and Leonard Levering. 2009. "A Semantic Web-Based Approach for Building Personalized News Services." *International Journal of E-Business Research* 5 (3): 35–53.

Frehe, Volker, and Frank Teuteberg. 2014. "The Role of ICT in Green Logistics: A Systematic Literature Review." In *Information Technology in Environmental Engineering*, edited by Burkhardt Funk, Peter Niemeyer, and Jorge Marx Gómez, 53–65. Berlin Heidelberg: Springer.

Freifeld, Clark C., Kenneth D. Mandl, Ben Y. Reis, and John S. Brownstein. 2008. "HealthMap: Global Infectious Disease Monitoring through Automated Classification and Visualization of Internet Media Reports." *Journal of the American Medical Informatics Association* 15 (2): 150–57.

Fuge, Mark, Katherine McKinstry, and Kevin Ninomiya. 2013. "Impactmap: Designing Sustainable Supply Chains by Incorporating Data Uncertainty." In *Proceedings of the International Symposium on Sustainable Systems and Technologies*, edited by Melissa Bilec and Jun-ki Choi. ISSST '13. Oakland: Sustainable Conoscente Network.

Gaizauskas, Robert, and Yorick Wilks. 1998. "Information Extraction: Beyond Document Retrieval." *Journal of Documentation* 54 (1): 70–105.

Galibert, Olivier, Sophie Rosset, Cyril Grouin, Pierre Zweigenbaum, and Ludovic Quintard. 2011. "Structured and Extended Named Entity Evaluation in Automatic Speech Transcriptions." In *Proceedings of the 5th International Joint Conference on Natural Language Processing*, 518–26. IJCNLP '11. Asian Federation of Natural Language Proceesing.

Gambelli, D., and V. Bruschi. 2010. "A Bayesian Network to Predict the Probability of Organic Farms' Exit from the Sector: A Case Study from Marche, Italy." *Computers and Electronics in Agriculture* 71 (1): 22–31.

Gangemi, Aldo. 2013. "A Comparison of Knowledge Extraction Tools for the Semantic Web." In *The Semantic Web: Semantics and Big Data*, edited by Philipp Cimiano, Oscar Corcho, Valentina Presutti, Laura Hollink, and Sebastian Rudolph, 351–66. Lecture Notes in Computer Science 7882. Berlin Heidelberg: Springer.

García-Rodríguez, Francisco J., Carlos Castilla-Gutiérrez, and Carlos Bustos-Flores. 2013. "Implementation of Reverse Logistics as a Sustainable Tool for Raw Material Purchasing in Developing Countries: The Case of Venezuela." *International Journal of Production Economics*, Special Issue on Service Science, 141 (2): 582–92.

Gardner, John T., and Martha C. Cooper. 2003. "Strategic Supply Chain Mapping Approaches." *Journal of Business Logistics* 24 (2): 37–64.

Garvey, Myles D., Steven Carnovale, and Sengun Yeniyurt. 2015. “An Analytical Framework for Supply Network Risk Propagation: A Bayesian Network Approach.” *European Journal of Operational Research* 243 (2): 618–27.

Gauthier, Caroline. 2005. “Measuring Corporate Social and Environmental Performance: The Extended Life-Cycle Assessment.” *Journal of Business Ethics* 59 (1-2): 199–206.

Gayo, Jose Emilio Labra, and Jose María Alvarez Rodríguez. 2011. “Product Scheme Classifications - Datasets.” http://www.josemalvarez.es/web/2011/11/16/product-scheme-classifications/ (accessed 22.07.14).

GeoNames. 2014. “GeoNames Feature Codes.” http://www.geonames.org/export/codes.html (accessed 20.11.14).

Gerner, Martin, Farzaneh Sarafraz, Casey M. Bergman, and Goran Nenadic. 2012. “BioContext: An Integrated Text Mining System for Large-Scale Extraction and Contextualization of Biomolecular Events.” *Bioinformatics* 28 (16): 2154–61.

Ghadge, Abhijeet, Dani Samir, and Roy Kalawsky. 2012. “Supply Chain Risk Management: Present and Future Scope.” *The International Journal of Logistics Management* 23 (3): 313–39.

Giannakis, Mihalis, and Michalis Louis. 2011. “A Multi-Agent Based Framework for Supply Chain Risk Management.” *Journal of Purchasing and Supply Management* 17 (1): 23–31.

Giblin, Carmel. 2012. “Eliminating Child Labour from the Supply Chain.” *The Guardian*, June 12. http://www.theguardian.com/sustainable-business/eliminating-child-labour-supply-chain (accessed 07.10.13).

Gildea, Daniel, and Daniel Jurafsky. 2002. “Automatic Labeling of Semantic Roles.” *Computational Linguistics* 28 (3): 245–88.

Gimenez, Cristina, and Elcio M. Tachizawa. 2012. “Extending Sustainability to Suppliers: A Systematic Literature Review.” *Supply Chain Management* 17 (5): 531–43.

Giunipero, Larry C., Robert E. Hooker, and Diane Denslow. 2012. “Purchasing and Supply Management Sustainability: Drivers and Barriers.” *Journal of Purchasing and Supply Management* 18 (4): 258–69.

Glavaš, Goran, and Jan Šnajder. 2014. “Event Graphs for Information Retrieval and Multi-Document Summarization.” *Expert Systems with Applications* 41 (15): 6904–16.

Global Reporting Initiative. 2011. *G3.1 Sustainability Reporting Guidelines*. Amsterdam: Global Reporting Initiative. https://www.globalreporting.org/reporting/latest-guidelines/g3-1-guidelines/Pages/default.aspx (accessed 28.12.12).

Global Social Compliance Programme. 2010. *GSCP Reference Code Version 2*. http://gscpnet.com/gscpfiles/GSCP_Reference_Code_V2_April_2010.pdf (accessed 02.01.13).

Gluchowski, Peter, and Roland Gabriel. 2008. *Management Support Systeme und Business Intelligence. Computergestützte Informationssysteme für Fach- und Führungskräfte*. 2nd ed. Berlin Heidelberg: Springer.

Gluchowski, Peter, Roland Gabriel, and Peter Chamoni. 2005. *Management Support Systeme Und Business Intelligence: Computergestützte Informationssysteme Für Fach-Und Führungskräfte*. New York: Springer.

Godfrey, Michael, and Andrew Manikas. 2012. “Integrating Triple Bottom Line Sustainability Concepts into a Supplier Selection Exercise.” *Business Education & Accreditation* 4 (1): 1–12.

Gopalakrishnan, Kavitha, Yahaya Y. Yusuf, Ahmed Musa, Tijjani Abubakar, and Hafsat M. Ambursa. 2012. “Sustainable Supply Chain Management: A Case Study of British Aerospace (BAe) Systems.” *International Journal of Production Economics*, Sustainable Development of Manufacturing and Services, 140 (1): 193–203.

Gopal, P.R.C., and Jitesh Thakkar. 2014. “Development of Composite Sustainable Supply Chain Performance Index for the Automobile Industry.” *International Journal of Sustainable Engineering*, in press.

Govindan, Kannan, Roohollah Khodaverdi, and Ahmad Jafarian. 2012. "A Fuzzy Multi Criteria Approach for Measuring Sustainability Performance of a Supplier Based on Triple Bottom Line Approach." *Journal of Cleaner Production* 47: 345–54.

Gray, Philip C.R., and Peter M. Wiedemann. 1999. "Risk Management and Sustainable Development: Mutual Lessons from Approaches to the Use of Indicators." *Journal of Risk Research* 2 (3): 201–18.

Greiner, Lynn. 2007. "IT Infrastructure Library (ITIL) Definition and Solutions." *CIO*. March 7. http://www.cio.com/article/2439501/infrastructure/it-infrastructure-library--itil--definition-and-solutions.html (accessed 17.02.15).

Grimmer, Justin, and Brandon M. Stewart. 2013. "Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts." *Political Analysis* 21 (3): 267–97.

Grimm, Jörg H., Joerg S. Hofstetter, and Joseph Sarkis. 2014. "Exploring Sub-Suppliers' Compliance with Corporate Sustainability Standards." *Journal of Cleaner Production*, November, in press.

Grimm, Jörg H., Professor Dr Wolfgang Stölzle, and Assistenz-Professor Dr Joerg S. Hofstetter. 2013. "Conceptualizing the Capability of Supplier Sustainability Risk Management." In *Kompetenz, Interdisziplinarität Und Komplexität in Der Betriebswirtschaftslehre*, edited by Wolfgang Kersten and Jochen Wittmann, 39–56. Wiesbaden: Springer Fachmedien.

Grishman, Ralph. 1998. *TIPSTER Text Architecture Design*. Version 3.1. New York: New York University. http://www.itl.nist.gov/iaui/894.02/related_projects/tipster/docs/arch31.doc.

Grishman, Ralph, Silja Huttunen, and Roman Yangarber. 2002a. "Real-Time Event Extraction for Infectious Disease Outbreaks." In *Proceedings of the Second International Conference on Human Language Technology Research*, edited by Mitchell Marcus, 366–69. HLT '02. San Francisco: Morgan Kaufmann Publishers.

———. 2002b. "Information Extraction for Enhanced Access to Disease Outbreak Reports." *Journal of Biomedical Informatics*, Sublanguage - Zellig Harris Memorial, 35 (4): 236–46.

Grober, Ulrich. 2009. "Hans Carl von Carlowitz: Der Erfinder Der Nachhaltigkeit." *Die Zeit*, November 9. http://www.zeit.de/1999/48/Der_Erfinder_der_Nachhaltigkeit (accessed 17.02.15).

Groth, Sven S., and Jan Muntermann. 2011. "An Intraday Market Risk Management Approach Based on Textual Analysis." *Decision Support Systems*, Enterprise Risk and Security Management: Data, Text and Web Mining, 50 (4): 680–91.

Guarcello, L., I. Kovrova, S. Lyon, M. Manacorda, and F. C. Rosati. 2010. *Towards Consistency in Child Labour Measurement: Assessing the Comparability of Estimates Generated by Different Survey Instruments*. Understanding Children's Work Programme Working Paper Series. http://www.ucw-project.org/attachment/st_Towards_consistency_in_child_labour_measurement20110517_123409.pdf (accessed 20.08.14).

Guenther, Edeltraud, and Vera Greschner Farkavcová. 2010. "Decision Making for Transportation Systems as a Support for Sustainable Stewardship: Freight Transport Process Evaluation Using the ETIENNE-Tool." *Management Research Review* 33 (4): 317–39.

Gupta, Vishal, and Darvinder Kaur. 2010. "A Survey of Named Entity Recognition in English and Other Indian Languages." *International Journal of Computer Science Issues* 7 (6): 239–45.

Guyon, Olivier, Nabil Absi, Dominique Feillet, and Thierry Garaix. 2012. "A Modeling Approach for Locating Logistics Platforms for Fast Parcels Delivery in Urban Areas." *Procedia - Social and Behavioral Sciences*, Seventh International Conference on City Logistics, 39: 360–68.

Hack, Stefan, and Christian Berg. 2014. “The Potential of IT for Corporate Sustainability.” *Sustainability* 6 (7): 4163–80.

Hadiguna, Rika Ampuh. 2012. “Decision Support Framework for Risk Assessment of Sustainable Supply Chain.” *International Journal of Logistics Economics and Globalisation* 4 (1/2): 35.

Hagemann, Frank, Yacouba Diallo, Alex Etienne, Farhad Mehran, and International Labour Office. 2006. *Global Child Labour Trends 2000 to 2004*. Geneva: International Labour Organization.

Hagemann, Frank, Hakki Ozel, Lorenzo Guarcello, International Labour Office, and International Programme on the Elimination of Child Labour. 2007. *Children’s Non-Market Activities and Child Labour Measurement: A Discussion Based on Household Survey Data*. Geneva: International Labour Organization.

Hagenau, Michael, Michael Liebmann, and Dirk Neumann. 2013. “Automated News Reading: Stock Price Prediction Based on Financial News Using Context-Capturing Features.” *Decision Support Systems* 55 (3): 685–97.

Hahn, Udo, and Klemens Schnattinger. 1998. “Towards Text Knowledge Engineering.” In *Proceedings of the Fifteenth National/Tenth Conference on Artificial Intelligence/Innovative Applications of Artificial Intelligence*, edited by Bruce G. Buchanan and Ramasamy Uthurusamy, 524–31. AAAI ’98/IAAI ’98. Palo Alto: American Association for Artificial Intelligence.

Hamp, Birgit, and Helmut Feldweg. 1997. “Germanet-a Lexical-Semantic Net for German.” In *Proceedings of ACL Workshop Automatic Information Extraction and Building of Lexical Semantic Resources for NLP Applications*, edited by Piek Vossen, Geert Adriaens, Nicoletta Calzolari, Antonio Sanfilippo, and Yorick Wilks, 9–15. Stroudsburg: Association for Computational Linguistics.

Hansen, Art, and Pablo Diego Rosell. 2012a. *Children Working in the Carpet Industry of Pakistan: Prevalence and Conditions*. United States Department of Labor, ICF International, Inc. http://www.dol.gov/ilab/iclre/Downloads/Research/Report/PC_Pakistan_Report.pdf (accessed 02.04.14).

———. 2012b. *Children Working in the Carpet Industry of Nepal: Prevalence and Conditions*. United States Department of Labor, ICF International, Inc. http://www.dol.gov/ilab/iclre/Downloads/Research/Report/PC_Nepal_Report.pdf (accessed 02.04.14).

Hansen, Erik G., and Stefan Schaltegger. 2014. “The Sustainability Balanced Scorecard: A Systematic Review of Architectures.” *Journal of Business Ethics*, September, 1–29.

Hansen, Ursula, and Ulf Schrader. 2005. “Corporate Social Responsibility Als Aktuelles Thema Der Betriebswirtschaftslehre.” *Die Betriebswirtschaft* 65 (4): 373–95.

Harland, Christine, Richard Brenchley, and Helen Walker. 2003. “Risk in Supply Networks.” *Journal of Purchasing and Supply Management*, Supply Chain Management: Selected Papers from the European Operations Management Association (EurOMA) 8th International Annual Conference, 9 (2): 51–62.

Harmon, Robert R., and Nora Moolenkamp. 2012. “Sustainable It Services: Developing a Strategy Framework.” *International Journal of Innovation & Technology Management* 9 (2): 1250014–1.

Harms, Dorli, Erik G. Hansen, and Stefan Schaltegger. 2013. “Strategies in Sustainable Supply Chain Management: An Empirical Investigation of Large German Companies.” *Corporate Social Responsibility and Environmental Management* 20 (4): 205–18.

Harrington, Kevin, and John O’Connor. 2009. “How Cisco Succeeds at Global Risk Management.” *Supply Chain Management Review*, no. July/August: 10–17.

Harwood, Ian, and Stuart Humby. 2008. “Embedding Corporate Responsibility into Supply: A Snapshot of Progress.” *European Management Journal* 26 (3): 166–74.

Hasle, Geir. 1999. “Transportation Management in Distributed Enterprises.” *Human Systems Management* 18 (3,4): 203–12.

Hassini, Elkafi, Chirag Surti, and Cory Searcy. 2012. “A Literature Review and a Case Study of Sustainable Supply Chains with a Focus on Metrics.” *International Journal of Production Economics* 140 (1): 69–82.

Hayes, Matthew, and Peter F. Nardulli. 2011. *SPEED’s Societal Stability Protocol and the Study of Civil Unrest: An Overview and Comparison with Other Event Data Projects*. Cline Center for Democracy, University of Illinois at Urbana-Champaign. http://www.clinecenter.illinois.edu/publications/SPEED-Comparison_With_Other_Projects.pdf (accessed 11.11.14).

Hearst, Marti A., and Christian Plaunt. 1993. “Subtopic Structuring for Full-Length Document Access.” In *Proceedings of the 16th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, 59–68. SIGIR ’93. New York: ACM.

Heath, Tom. 2008. “How Will We Interact with the Web of Data?” *IEEE Internet Computing* 12 (5): 88–91.

Hecking, Matthias, and Tatjana Sarmina-Baneviciene. 2010. “A Tajik Extension of the Multilingual Information Extraction System ZENON.” In *Proceedings of 15th International Command and Control Research and Technology Symposium*, 1–14. ICCRTS ’10. Santa Monica: Command and Control Research Program.

Hendrik Reefke, and Mattia Trocchi. 2013. “Balanced Scorecard for Sustainable Supply Chains: Design and Development Guidelines.” *International Journal of Productivity and Performance Management* 62 (8): 805–26.

Hepp, Martin, and Jos de Bruijn. 2007. “GenTax: A Generic Methodology for Deriving OWL and RDF-S Ontologies from Hierarchical Classifications, Thesauri, and Inconsistent Taxonomies.” In *The Semantic Web: Research and Applications*, 4519:129–44. Lecture Notes in Computer Science. Berlin Heidelberg: Springer.

Hiltunen, Elina. 2006. “Was It a Wild Card or Just Our Blindness to Gradual Change.” *Journal of Futures Studies* 11 (2): 61–74.

Hobbs, Jerry R. 1993. “The Generic Information Extraction System.” In *Proceedings of the 5th Conference on Message Understanding*, 87–91. MUC5 ’93. Stroudsburg: Association for Computational Linguistics.

Hobbs, Jerry R., Douglas Appelt, John Bear, David Israel, Megumi Kameyama, Mark Stickel, and Mabry Tyson. 1997. “FASTUS: A Cascaded Finite-State Transducer for Extracting Information from Natural-Language Text.” In *Finite-State Language Processing*, edited by Emmanuel Roche and Yves Schabes, 383–406. The Massachusetts Institute of Technology.

Hofmann, Hannes, Christian Busse, Christoph Bode, and Michael Henke. 2014. “Sustainability-Related Supply Chain Risks: Conceptualization and Management.” *Business Strategy and the Environment* 23 (3): 160–72.

Hofman, Wout. 2011. “Supply Chain Visibility with Linked Open Data for Supply Chain Risk Analysis.” In *Workshop on IT Innovations Enabling Seamless and Secure Supply Chains*, 20–31.

———. 2013. “Compliance Management by Business Event Mining in Supply Chain Networks.” In *Proceedings of 7th International Workshop on Value Modeling and Business Ontology*. VMBO ’13. Delft.

Hogenboom, Alexander, Frederik Hogenboom, Flavius Frasincar, Uzay Kaymak, Otto van der Meer, and Kim Schouten. 2011. “Detecting Economic Events Using a Semantics-Based Pipeline.” In *Database and Expert Systems Applications*, edited by Abdelkader Hameurlain, Stephen W. Liddle, Klaus-Dieter Schewe, and Xiaofang Zhou, 440–47. Lecture Notes in Computer Science 6860. Berlin Heidelberg: Springer.

Hogenboom, Alexander, Frederik Hogenboom, Flavius Frasincar, Kim Schouten, and Otto van der Meer. 2013. “Semantics-Based Information Extraction for Detecting Economic Events.” *Multimedia Tools and Applications* 64 (1): 27–52.

Hogenboom, Frederik, Flavius Frasincar, Uzay Kaymak, and Franciska de Jong. 2011. “An Overview of Event Extraction from Text.” In *Workshop on Detection, Representation, and Exploitation of Events in the Semantic Web*, edited by Marieke van Erp, Willem Robert van Hage, Laura Hollink, Anthony Jameson, and Raphaël Troncy, 779:48–57. DeRiVE ’11. Bonn: CEUR Workshop Proceedings.

Holt, Diane, and Ralf Barkemeyer. 2012. “Media Coverage of Sustainable Development Issues – Attention Cycles or Punctuated Equilibrium?” *Sustainable Development* 20 (1): 1–17.

Holton, Carolyn. 2009. “Identifying Disgruntled Employee Systems Fraud Risk through Text Mining: A Simple Solution for a Multi-Billion Dollar Problem.” *Decision Support Systems* 46 (4): 853–64.

Holzmann, Robert, and Steen Jørgensen. 1999. *Social Protection as Social Risk Management*. 9901. The World Bank: Social Protection Discussion Paper. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.200.9746&rep=rep1&type=pdf (accessed 25.04.13).

———. 2001. “Social Risk Management: A New Conceptual Framework for Social Protection, and Beyond.” *International Tax and Public Finance* 8 (4): 529–56.

Hopkins, Kathryn. 2008. “Child Labour: Primark Caught out.” *The Guardian*, June 23. http://www.theguardian.com/business/2008/jun/23/primark.children (accessed 01.02.15).

Ho, William, Xiaowei Xu, and Prasanta K. Dey. 2010. “Multi-Criteria Decision Making Approaches for Supplier Evaluation and Selection: A Literature Review.” *European Journal of Operational Research* 202 (1): 16–24.

Hsu, Li-Chang, Shang-Ling Ou, and Yih-Chang Ou. 2015. “A Comprehensive Performance Evaluation and Ranking Methodology under a Sustainable Development Perspective.” *Journal of Business Economics and Management* 16 (1): 74–92.

Huang, LiGuo, Daniel Port, Liang Wang, Tao Xie, and Tim Menzies. 2010. “Text Mining in Supporting Software Systems Risk Assurance.” In *Proceedings of the IEEE/ACM International Conference on Automated Software Engineering*, 163–66. ASE ’10. New York: ACM.

Hubbard, Graham. 2009. “Measuring Organizational Performance: Beyond the Triple Bottom Line.” *Business Strategy and the Environment* 18 (3): 177–91.

Humpherys, Sean L., Kevin C. Moffitt, Mary B. Burns, Judee K. Burgoon, and William F. Felix. 2011. “Identification of Fraudulent Financial Statements Using Linguistic Credibility Analysis.” *Decision Support Systems*, On quantitative methods for detection of financial fraud, 50 (3): 585–94.

Hung, Sheng-Hao, Chia-Hung Lin, and Jen-Shin Hong. 2010. “Web Mining for Event-Based Commonsense Knowledge Using Lexico-Syntactic Pattern Matching and Semantic Role Labeling.” *Expert Systems with Applications* 37 (1): 341–47.

Hunkeler, David. 2006. “Societal LCA Methodology and Case Study.” *The International Journal of Life Cycle Assessment* 11 (6): 371–82.

Hunkeler, D., and G. Rebitzer. 2005. “The Future of Life Cycle Assessment.” *The International Journal of Life Cycle Assessment* 10 (5): 305–8.

Hunter, Anthony, and Rupert Summerton. 2006. “Merging News Reports That Describe Events.” *Data & Knowledge Engineering* 59 (1): 1–24.

Hutchins, Margot J., and John W. Sutherland. 2008. “An Exploration of Measures of Social Sustainability and Their Application to Supply Chain Decisions.” *Journal of Cleaner Production* 16 (15): 1688–98.

Huttunen, Silja, Roman Yangarber, and Ralph Grishman. 2002a. “Complexity of Event Structure in IE Scenarios.” In *Proceedings of the 19th International Conference on Computational Linguistics - Volume 1*, 1–7. COLING ’02. Stroudsburg: Association for Computational Linguistics.

———. 2002b. "Diversity of Scenarios in Information Extraction." In *Proceedings of the International Conference on Language Resources and Evaluation*, 1443–50. LREC '02. Paris: European Language Resources Association (ELRA).

Iacob, Maria-Eugenia, Marten J. van Sinderen, M. Steenwijk, and P. Verkroost. 2013. "Towards a Reference Architecture for Fuel-Based Carbon Management Systems in the Logistics Industry." *Information Systems Frontiers* 15 (5): 725–45.

Iakovou, Eleftherios T. 2001. "An Interactive Multiobjective Model for the Strategic Maritime Transportation of Petroleum Products: Risk Analysis and Routing." *Safety Science* 39 (1-2): 19–29.

IBM Corp. 2013. *IBM SPSS Statistics for Windows* (version Version 22.0). Armonk: IBM Corp.

III Global Conference on Child Labour. 2013. *The Brasilia Declaration on Child Labour*. Brasilia. http://www.ilo.org/ipecinfo/product/download.do?type=document&id=23480 (accessed 09.12.13).

IJntema, Wouter, Jordy Sangers, Frederik Hogenboom, and Flavius Frasincar. 2012. "A Lexico-Semantic Pattern Language for Learning Ontology Instances from Text." *Web Semantics: Science, Services and Agents on the World Wide Web* 15 (September): 37–50.

Ikonomakis, M., S. Kotsiantis, and V. Tampakas. 2005. "Text Classification Using Machine Learning Techniques." *WSEAS Transactions on Computers* 8 (4): 966–74.

Institute for Human Development. 2009. *Documenting Child Labour Monitoring System (CLIMS) of INDUS Child Labour Project*. New Delhi: International Labour Organization. http://www.ilo.org/ipecinfo/product/download.do?type=document&id=14140 (accessed 30.10.13).

Intel. 2015. "Moore's Law Inspires Intel Innovation." http://www.intel.com/content/www/us/en/silicon-innovations/moores-law-technology.html (accessed 05.02.15).

International Labour Office. 2004. *Child Labour: A Textbook for University Students*. Geneva: International Labour Organization.

———. 2013. *World Report on Child Labour Economic Vulnerability, Social Protection and the Fight against Child Labour.* Geneva: International Labour Organization.

International Labour Organization. 1973a. "Convention C138 - Minimum Age Convention, 1973 (No. 138)." *NORMLEX*. http://www.ilo.org/dyn/normlex/en/f?p=NORMLEXPUB:12100:0::NO:12100:P12100_ILO_CODE:C138 (accessed 02.12.13).

———. 1973b. "Recommendation R146 - Minimum Age Recommendation, 1973 (No. 146)." *NORMLEX*. http://www.ilo.org/dyn/normlex/en/f?p=NORMLEXPUB:12100:0::NO:12100:P12100_INSTRUMENT_ID:312484:NO (accessed 02.02.15).

———. 1982. *Resolution Concerning Statistics of the Economically Active Population, Employment, Unemployment and Underemployment, Adopted by the Thirteenth International Conference of Labour Statisticians*. Thirteenth International Conference of Labour Statisticians (October 1982). http://www.ilo.org/wcmsp5/groups/public/---dgreports/---stat/documents/normativeinstrument/wcms_087481.pdf (accessed 21.08.14).

———. 1999a. "Convention C182 - Worst Forms of Child Labour Convention, 1999 (No. 182)." *NORMLEX*. http://www.ilo.org/dyn/normlex/en/f?p=NORMLEXPUB:12100:0::NO::P12100_ILO_CODE:C182 (accessed 02.12.13).

———. 1999b. "Recommendation R190 - Worst Forms of Child Labour Recommendation, 1999 (No. 190)." *NORMLEX*. http://www.ilo.org/dyn/normlex/en/f?p=NORMLEXPUB:12100:0::NO:12100:P12100_INSTRUMENT_ID:312528:NO (accessed 02.02.15).

———. 2008a. "Child Labour: Subregional Office for Eastern Europe and Central Asia (SRO-Moscow)."

http://www.ilo.org/public/english/region/eurpro/moscow/areas/ipec/causes.htm (accessed 25.10.13).
———. 2008b. *Resolution II - Resolution Concerning Statistics of Child Labour*. The 18th International Conference of Labour Statisticians. http://ilo.org/wcmsp5/groups/public/---dgreports/---stat/documents/normativeinstrument/wcms_112458.pdf (accessed 02.02.15).
International Organization for Standardization. 2009. "ISO/Guide 73:2009(en) Risk Management — Vocabulary." https://www.iso.org/obp/ui/#iso:std:iso:guide:73:ed-1:v1:en:term:1.1 (accessed 15.08.14).
———. 2011. "ISO/IEC 20000-1:2011(en) - Terms and Definitions." https://www.iso.org/obp/ui/#iso:std:iso-iec:20000:-1:ed-2:v1:en (accessed 09.03.15).
International Programme on the Elimination of Child Labor. 2012. *Towards the Elimination of Hazardous Child Labour... Practices with Good Potential*. Geneva: International Labour Organization.
International Programme on the Elimination of Child Labour. 2005. *Overview of Child Labour Monitoring*. Geneva: International Labour Organization.
———. 2010. *Child Labour Monitoring System (Tanzania)*. Geneva: International Labour Organization.
———. 2013. *Marking Progress against Child Labour - Global Estimates and Trends 2000-2012*. Geneva: International Labour Organization.
International Programme on the Elimination of Child Labour, and International Labour Office. 2011. *Children in Hazardous Work: What We Know, What We Need to Do*. Geneva: International Labour Organization.
International Programme on the Elimination of Child Labour, Statistical Information and Monitoring Programme on Child Labour, and International Labour Office. 2004. *Child Labour Statistics: Manual on Methodologies for Data Collection through Surveys*. Geneva: International Labour Organization.
Inyaem, Uraiwan, Phayung Meesad, Choochart Haruechaiyasak, and Dat Tran. 2009. "Ontology-Based Terrorism Event Extraction." In *2009 1st International Conference on Information Science and Engineering (ICISE)*, 912–15.
ITILFoundations. 2014. "ITIL Foundations Incident Management, a Definition." http://www.itilfoundations.com/processes/incident-management/definition/ (accessed 10.04.14).
Jakhar, Suresh Kumar. 2015. "Performance Evaluation and a Flow Allocation Decision Model for a Sustainable Supply Chain of an Apparel Industry." *Journal of Cleaner Production* 87 (January): 391–413.
Jamieson, Susan. 2004. "Likert Scales: How to (ab)use Them." *Medical Education* 38 (12): 1217–18.
Jenkin, Tracy A., Jane Webster, and Lindsay McShane. 2011. "An Agenda for 'Green' Information Technology and Systems Research." *Information and Organization* 21 (1): 17–40.
Ji, Heng. 2010. "Challenges from Information Extraction to Information Fusion." In *Proceedings of the 23rd International Conference on Computational Linguistics: Posters*, edited by Chu-Ren Huang and Dan Jurafsky, 507–15. COLING '10. Stroudsburg: Association for Computational Linguistics.
Ji, Heng, and Ralph Grishman. 2008. "Refining Event Extraction through Cross-Document Inference." In *ACL*, 254–62. Citeseer.
Ji, Heng, Xiang Li, Angelo Lucia, and Jianting Zhang. 2010. "Annotating Event Chains for Carbon Sequestration Literature." In *Proceedings of the International Conference on Language Resources and Evaluation*, edited by Nicoletta Calzolari, Khalid Choukri, Bente Maegaard, Joseph Mariani, Jan Odijk, Stelios Piperidis, Mike Rosner, and Daniel Tapias, 2632–36. LREC '10. Paris: European Language Resources Association (ELRA).

Jin, Yun, Sung Hyon Myaeng, and Yuchul Jung. 2007. "Use of Place Information for Improved Event Tracking." *Information Processing & Management* 43 (2): 365–78.

Jones, Karen Spärck. 1972. "A Statistical Interpretation of Term Specificity and Its Application in Retrieval." *Journal of Documentation* 28: 11–21.

Jørgensen, Andreas, Lufanna C. H. Lai, and Michael Z. Hauschild. 2010. "Assessing the Validity of Impact Pathways for Child Labour and Well-Being in Social Life Cycle Assessment." *International Journal of Life Cycle Assessment* 15 (1): 5–16.

Jørgensen, Andreas, A. Le Bocq, L. Nazarkina, and M. Hauschild. 2008. "Methodologies for Social Life Cycle Assessment." *The International Journal of Life Cycle Assessment* 13 (2): 96–103.

Julius Raab Stiftung. 2014. *Unternehmerische Verantwortung in Der Sozialen Marktwirtschaft*. Wien: noir. http://juliusraabstiftung.at/resources/files/2014/6/6/972/studie-unternehmerische-verantwortung-2014-ds-final.pdf (accessed 13.05.15).

Jungermann, Felix, and Katharina Morik. 2008. "Enhanced Services for Targeted Information Retrieval by Event Extraction and Data Mining." In *Natural Language and Information Systems*, edited by Epaminondas Kapetanios, Vijayan Sugumaran, and Myra Spiliopoulou, 335–36. Lecture Notes in Computer Science 5039. Berlin Heidelberg: Springer.

Jusoh, Shaidah, and Hejab M. Alfawareh. 2012. "Techniques, Applications and Challenging Issue in Text Mining Ues, Applications and Challenging Issue in Text Mining Ues, Applications and Challenging Issue in Text Mining." *9* 6 (2): 431–36.

Kabir, Golam, Solomon Tesfamariam, Alex Francisque, and Rehan Sadiq. 2015. "Evaluating Risk of Water Mains Failure Using a Bayesian Belief Network Model." *European Journal of Operational Research* 240 (1): 220–34.

Kaiser, Katharina, and Silvia Miksch. 2005. *Information Extraction*. Asgaard-TR-2005-6. A Survey. Technical report, Vienna University of Technology, Institute of Software Technology and Interactive Systems. http://ieg.ifs.tuwien.ac.at/techreports/Asgaard-TR-2005-6.pdf (accessed 06.11.14).

Kang, Ning, Bharat Singh, Chinh Bui, Zubair Afzal, Erik M. van Mulligen, and Jan A. Kors. 2014. "Knowledge-Based Extraction of Adverse Drug Events from Biomedical Text." *Bmc Bioinformatics* 15 (March): 64.

Kao, Han-Ying, Chia-Hui Huang, Chu-Ling Hsu, and Chiao-Ling Huang. 2010. "Diagnosis from Bayesian Networks with Fuzzy Parameters – A Case in Supply Chains." In *Advances in Grid and Pervasive Computing*, edited by Paolo Bellavista, Ruay-Shiung Chang, Han-Chieh Chao, Shin-Feng Lin, and Peter M. A. Sloot, 353–62. Lecture Notes in Computer Science 6104. Berlin Heidelberg: Springer.

Kao, Han-Ying, Chia-Hui Huang, and Han-Lin Li. 2005. "Supply Chain Diagnostics with Dynamic Bayesian Networks." *Computers & Industrial Engineering* 49 (2): 339–47.

Keating, Burion, Ali Quazi, Anton Kriz, and Tim Coltman. 2008. "In Pursuit of a Sustainable Supply Chain: Insights from Westpac Banking Corporation." *Supply Chain Management: An International Journal* 13 (3): 175–79.

Keller, Ann C. 2009. "Credibility and Relevance in Environmental Policy: Measuring Strategies and Performance among Science Assessment Organizations." *Journal of Public Administration Research and Theory*, January, mup001.

Kemper, Hans-Georg, Henning Baars, and Walid Mehanna. 2010. *Business Intelligence - Grundlagen und praktische Anwendungen: Eine Einführung in die IT-basierte Managementunterstützung*. Wiesbaden: Vieweg+Teubner.

Kenyon, Paul, Fiona Campbell, and Esella Hawkey. 2000. "Gap and Nike: No Sweat? October 15 2000." *BBC*, October 15. http://news.bbc.co.uk/2/hi/programmes/panorama/970385.stm (accessed 24.07.13).

Khan, Aurangzeb, Baharum Baharudin, Lam Hong Lee, and Khairullah Khan. 2010. "A Review of Machine Learning Algorithms for Text-Documents Classification." *Journal of Advances in Information Technology* 1 (1): 4–20.

Khan, Faisal I., Rehan Sadiq, and B. Veitch. 2004. “Life Cycle iNdeX (LInX): A New Indexing Procedure for Process and Product Design and Decision-Making.” *Journal of Cleaner Production* 12 (1): 59–76.

King, Jack L. 1996. “Providing Assurance on the Absence of Unknown Activities.” In *Probabilistic Safety Assessment and Management ’96*, edited by P. Carlo Cacciabue and loannis A. PapazogIou, 1:217–22. Thrid International Conference on Probability and Safety Assessment Conference (PSAM III). London: Springer.

Klassen, Robert D., and Ann Vereecke. 2012. “Social Issues in Supply Chains: Capabilities Link Responsibility, Risk (opportunity), and Performance.” *International Journal of Production Economics* 140 (1): 103–15.

Klöpffer, Walter. 2008. “Life Cycle Sustainability Assessment of Products.” *The International Journal of Life Cycle Assessment* 13 (2): 89–95.

Knott, Graham, and Andrew Fox. 2010. “A Model of Sustainable Risk Management.” *The Magazine of the Emergency Planning Society*, Emergency Managment, , no. August: 38–42.

Kogg, Beatrice. 2009. “Responsibility in the Supply Chain: Interorganisational Management of Environmental and Social Aspects in the Supply Chain : Case Studies Form the Textile Sector.” Dissertation, Lund: Lund University.

Kogg, Beatrice, and Oksana Mont. 2012. “Environmental and Social Responsibility in Supply Chains: The Practise of Choice and Inter-Organisational Management.” *Ecological Economics* 83: 154–63.

Koh, S.C. Lenny, Andrea Genovese, Adolf A. Acquaye, Paul Barratt, Nasir Rana, Johan Kuylenstierna, and David Gibbs. 2013. “Decarbonising Product Supply Chains: Design and Development of an Integrated Evidence-Based Decision Support System - the Supply Chain Environmental Analysis Tool (SCEnAT).” *International Journal of Production Research* 51 (7): 2092–2109.

Koks, Don, and Subhash Challa. 2005. *An Introduction to Bayesian and Dempster-Shafer Data Fusion*. Edinburgh: DSTO Systems Sciences Laboratory. http://oai.dtic.mil/oai/oai?verb=getRecord&metadataPrefix=html&identifier=ADA417895 (accessed 13.02.13).

Kolcz, Aleksander, Vidya Prabakarmurthi, and Jugal Kalita. 2001. “Summarization As Feature Selection for Text Categorization.” In *Proceedings of the Tenth International Conference on Information and Knowledge Management*, edited by Henrique Paques, Ling Liu, and David Grossman, 365–70. CIKM ’01. New York: ACM.

Kolk, Ans, and Rob van Tulder. 2002a. “Child Labor and Multinational Conduct: A Comparison of International Business andStakeholder Codes.” *Journal of Business Ethics* 36 (3): 291–301.

———. 2002b. “The Effectiveness of Self-Regulation: Corporate Codes of Conduct and Child Labour.” *European Management Journal* 20 (3): 260–71.

Koplin, Julia, Stefan Seuring, and Michael Mesterharm. 2007. “Incorporating Sustainability into Supply Management in the Automotive Industry – the Case of the Volkswagen AG.” *Journal of Cleaner Production*, The Automobile Industry & Sustainability, 15 (11–12): 1053–62.

Krajnc, Damjan, and Peter Glavič. 2005. “How to Compare Companies on Relevant Dimensions of Sustainability.” *Ecological Economics* 55 (4): 551–63.

Kreng, Victor B., and Chia-Hua Chang. 2003. “Bayesian Network Based Multiagent System—application in E-Marketplace.” *Computers & Mathematics with Applications* 46 (2–3): 429–44.

Krikorian, Raffi. 2013. “New Tweets per Second Record, and How!” *Twitter Blogs*. August 16. https://blog.twitter.com/2013/new-tweets-per-second-record-and-how (accessed 11.02.15).

Krstajic, Milos, Christian Rohrdantz, Michael Hund, and Andreas Weiler. 2012. *Getting There First: Real-Time Detection of Real-World Incidents on Twitter*.

http://kops.ub.uni-konstanz.de/bitstream/handle/urn:nbn:de:bsz:352-229499/Getting-There-First%20edit.pdf?sequence=3 (accessed 18.04.14).

Krysiak, Frank C. 2009. “Risk Management as a Tool for Sustainability.” *Journal of Business Ethics* 85 (3): 483–92.

Kumaran, Giridhar, and James Allan. 2004. “Text Classification and Named Entities for New Event Detection.” In *Proceedings of the 27th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, 297–304. SIGIR ’04. New York: ACM.

———. 2005. “Using Names and Topics for New Event Detection.” In *Proceedings of the Conference on Human Language Technology and Empirical Methods in Natural Language Processing*, edited by Raymond Mooney, Chris Brew, Lee-Feng Chien, and Katrin Kirchhoff, 121–28. HLT/EMNLP ’05. Stroudsburg: Association for Computational Linguistics.

Kurnia, Sherah, Rahim MD Mahbubur, and Marianne Gloet. 2012. “Understanding The Roles Of Is/It In Sustainable Supply Chain Management.” In *Proceedings of Pacific Asia Conference on Information Systems*, Paper 50. PACIS ’12. Atlanta: Association for Information Systems (AIS).

Labuschagne, Carin, Alan C. Brent, and Ron P.G. van Erck. 2005. “Assessing the Sustainability Performances of Industries.” *Journal of Cleaner Production* 13 (4): 373–85.

LaComb, Christina, John Interrante, and Kareem S. Aggour. 2007. “Monitoring Key Company Events through Deliberative Learning.” *Information Systems and eBusiness Management* 5 (4): 295.

Lai, Robert S.Q., Li-Ling Hsu, and Jason C.H. Chen. 2012. “Green Supply Chain Management Systems: A Case Study in the Textile Industry.” *Human Systems Management* 31 (2): 111–21.

Lam, Wai, and Kei Shiu Ho. 2001. “FIDS: An Intelligent Financial Web News Articles Digest System.” *IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans* 31 (6): 753–62.

Langlais, Philippe, Michel Simard, and Jean Véronis. 1998. “Methods and Practical Issues in Evaluating Alignment Techniques.” In *Proceedings of the 17th International Conference on Computational Linguistics-Volume 1*, 711–17. COLING ’98. Montreal: University of Montreal.

Lauritzen, Steffen L., and David J. Spiegelhalter. 1988. “Local Computations with Probabilities on Graphical Structures and Their Application to Expert Systems.” *Journal of the Royal Statistical Society. Series B (Methodological)* 50 (2): 157–224.

Lee, Eunchang, Yongtae Park, and Jong Gye Shin. 2009. “Large Engineering Project Risk Management Using a Bayesian Belief Network.” *Expert Systems with Applications* 36 (3, Part 2): 5880–87.

Lee, Gunwoong, and Uday Kulkarni. 2011. “Business Intelligence in Corporate Risk Management.” In *Proceedings of the Americas Conference on Information Systems*, Paper 420. AMCIS ’11. Atlanta: Association for Information Systems (AIS).

Lee, J. 2003. “E-Manufacturing-Fundamental, Tools, and Transformation.” *Robotics and Computer-Integrated Manufacturing* 19 (6): 501–7.

Leetaru, Kalev. 2011. “Culturomics 2.0: Forecasting Large-Scale Human Behavior Using Global News Media Tone in Time and Space.” *First Monday* 16 (9): online.

Leetaru, Kalev, and Philip A. Schrodt. 2013. *GDELT: Global Data on Events, Location and Tone, 1979-2012.* San Francisco. http://gdelt.utdallas.edu/data/documentation/ISA.2013.GDELT.pdf (accessed 26.12.13).

Legris, Paul, John Ingham, and Pierre Collerette. 2003. “Why Do People Use Information Technology? A Critical Review of the Technology Acceptance Model.” *Information & Management* 40 (3): 191–204.

Lehmann, Richard J., Robert Reiche, and Gerhard Schiefer. 2012. “Future Internet and the Agri-Food Sector: State-of-the-Art in Literature and Research.” *Computers and Electronics in Agriculture* 89 (0): 158–74.

Leidner, Jochen L., and Frank Schilder. 2010. “Hunting for the Black Swan: Risk Mining from Text.” In *Proceedings of the ACL 2010 System Demonstrations*, 54–59.

Leire, Charlotte, and Oksana Mont. 2010. “The Implementation of Socially Responsible Purchasing.” *Corporate Social Responsibility and Environmental Management* 17 (1): 27–39.

Lei, Zhen, Ling-da Wu, Ying Zhang, and Yu-chi Liu. 2005. “A System for Detecting and Tracking Internet News Event.” In *Advances in Multimedia Information Processing - PCM 2005*, edited by Yo-Sung Ho and Hyoung Joong Kim, 754–64. Lecture Notes in Computer Science 3767. Berlin Heidelberg: Springer.

Lemke, Fred, and Henry L. Petersen. 2013. “Teaching Reputational Risk Management in the Supply Chain.” *Supply Chain Management* 18 (4): 413–29.

Levenshtein, VI. 1966. “Binary Codes Capable of Correcting Deletions, Insertions and Reversals.” *Soviet Physics Doklady* 10 (8): 707–10.

Lianghui, Yu, Qiang Wang, and Suling Jia. 2012. *Application of Text Mining in the Country Risk Management Based on Ann*. Edited by H. Wang and K. Takahashi. Beijing: Beihang Univ Press.

Liao, Shasha, and Ralph Grishman. 2010. “Using Document Level Cross-Event Inference to Improve Event Extraction.” In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, edited by Jan Hajič, Sandra Carberry, Stephen Clark, and Joakim Nivre, 789–97. ACL ’10. Stroudsburg: Association for Computational Linguistics.

———. 2011. “Acquiring Topic Features to Improve Event Extraction: In Pre-Selected and Balanced Collections.” In *Proceedings of the International Conference Recent Advances in Natural Language Processing 2011*, edited by RANLP 2011 Organising Committee, 9–16. RANLP ’11. Hissar: RANLP 2011 Organising Committee.

Li, Jian, and Yang Gao. 2010. “Base on Bayesian Network of Supply Chain Enterprises Collaborative Sensitivity Analysis.” In *2010 International Conference on Logistics Systems and Intelligent Management*, 1:160–63.

Li-li Chen, Gui Yun-miao, and Li Zhen. 2011. “Supply Disturbances Risk Situation Analysis Based on Bayesian Network.” *Journal of Anhui Polytechnic University* 4.

Li, Maozhen, Bin Yu, O. Rana, and Zidong Wang. 2008. “Grid Service Discovery with Rough Sets.” *IEEE Transactions on Knowledge and Data Engineering* 20 (6): 851–62.

Lin, Fu-ren, and Chia-Hao Liang. 2008. “Storyline-Based Summarization for News Topic Retrospection.” *Decision Support Systems*, Special Issue Clusters, 45 (3): 473–90.

Liu, Bing, and Lei Zhang. 2012. “A Survey of Opinion Mining and Sentiment Analysis.” In *Mining Text Data*, edited by Charu C. Aggarwal and ChengXiang Zhai, 415–63. Springer US.

Liu, Duen-Ren, Meng-Jung Shih, Churn-Jung Liau, and Chin-Hui Lai. 2009. “Mining the Change of Event Trends for Decision Support in Environmental Scanning.” *Expert Systems with Applications* 36 (2, Part 1): 972–84.

Liu, Lingzhe, Hennie Daniels, and Wout Hofman. 2014. “Business Intelligence for Improving Supply Chain Risk Management.” In *Enterprise Information Systems*, edited by Slimane Hammoudi, José Cordeiro, Leszek A. Maciaszek, and Joaquim Filipe, 190–205. Lecture Notes in Business Information Processing 190. Springer International Publishing.

Liu, Mingrong, Yicen Liu, Liang Xiang, Xing Chen, and Qing Yang. 2008. “Extracting Key Entities and Significant Events from Online Daily News.” In *Intelligent Data Engineering and Automated Learning – IDEAL 2008*, edited by Colin Fyfe, Dongsup Kim, Soo-Young Lee, and Hujun Yin, 201–9. Lecture Notes in Computer Science 5326. Berlin Heidelberg: Springer.

Liu, Shaofeng, Zhihong Wang, and Li Liu. 2012. "An Integrated Sustainability Analysis Approach to Support Holistic Decision Making in Sustainable Supply Chain Management." In *Fusing Decision Support Systems into the Fabric of the Context*, 391–402. Frontiers in Artificial Intelligence and Applications. Amsterdam: IOS Press.

Li, Xiangyang, and Charu Chandra. 2007. "A Knowledge Integration Framework for Complex Network Management." *Industrial Management & Data Systems* 107 (8): 1089–1109.

Li, Yaoyong, and Kalina Bontcheva. 2008. "Adapting Support Vector Machines for F-Term-Based Classification of Patents" 7 (2): 7:1–7:19.

Li, Yaoyong, Kalina Bontcheva, and Hamish Cunningham. 2005a. "SVM Based Learning System for Information Extraction." In *Deterministic and Statistical Methods in Machine Learning*, edited by Joab Winkler, Mahesan Niranjan, and Neil Lawrence, 319–39. Lecture Notes in Computer Science 3635. Berlin Heidelberg: Springer.

———. 2005b. "Using Uneven Margins SVM and Perceptron for Information Extraction." In *Proceedings of the Ninth Conference on Computational Natural Language Learning*, edited by Ido Dagan and Daniel Gildea, 72–79. CoNLL '05. Stroudsburg: Association for Computational Linguistics.

———. 2007. "Experiments of Opinion Analysis on the Corpora MPQA and NTCIR-6." In *Proceedings of the Sixth NTCIR Workshop Meeting on Evaluation of Information Access Technologies*, 323–29. NTCIR-6. Tokyo: NII Test Collection for IR Systems Project.

———. 2009. "Adapting SVM for Data Sparseness and Imbalance: A Case Study on Information Extraction." *Natural Language Engineering* 15 (2): 241–71.

Li, Yaoyong, and John Shawe-Taylor. 2003. "The SVM with Uneven Margins and Chinese Document Categorization." In *Proceedings of The 17th Pacific Asia Conference on Language, Information and Computation*, edited by Dong Hong Ji and Kim Teng Lua, 216–27. PACLIC17. Singapore: COLIPS Publications.

Li, Yaoyong, Hugo Zaragoza, Ralf Herbrich, John Shawe-Taylor, and Jaz S. Kandola. 2002. "The Perceptron Algorithm with Uneven Margins." In *Proceedings of the Nineteenth International Conference on Machine Learning*, edited by Claude Sammut and Achim G. Hoffmann, 379–86. ICML '02. San Francisco: Morgan Kaufmann Publishers.

Li, Zhiwei, Bin Wang, Mingjing Li, and Wei-Ying Ma. 2005. "A Probabilistic Model for Retrospective News Event Detection." In *Proceedings of the 28th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, 106–13. SIGIR '05. New York: ACM.

Li, Zuofeng, Feifan Liu, Lamont Antieau, Yonggang Cao, and Hong Yu. 2010. "Lancet: A High Precision Medication Event Extraction System for Clinical Text." *Journal of the American Medical Informatics Association* 17 (5): 563–67.

Lockamy III, Archie. 2011. "Benchmarking Supplier Risks Using Bayesian Networks." *Benchmarking: An International Journal* 18 (3): 409–27.

———. 2014. "Assessing Disaster Risks in Supply Chains." *Industrial Management & Data Systems* 114 (5): 755–77.

Lockamy III, Archie, and Kevin McCormack. 2009. "Examining Operational Risks in Supply Chains." *Supply Chain Forum* 10 (1): 2–14.

———. 2010. "Analysing Risks in Supply Networks to Facilitate Outsourcing Decisions." *International Journal of Production Research* 48 (2): 593–611.

———. 2012. "Modeling Supplier Risks Using Bayesian Networks." *Industrial Management & Data Systems* 112 (2): 313–33.

Locke, Richard, Matthew Amengual, and Akshay Mangla. 2009. "Virtue out of Necessity? Compliance, Commitment, and the Improvement of Labor Conditions in Global Supply Chains." *Politics & Society* 37 (3): 319–51.

Locke, Richard, Thomas Kochan, Monica Romis, and Fei Qin. 2007. "Beyond Corporate Codes of Conduct: Work Organization and Labour Standards at Nike's Suppliers." *International Labour Review* 146 (1-2): 21–40.

Locke, Richard, Fei Qin, and Alberto Brause. 2007. "Does Monitoring Improve Labor Standards? Lessons from Nike." *ILRReview* 61 (1): 3–31.

Loeser, Fabian. 2013. "Green IT and Green IS: Definition of Constructs and Overview of Current Practices." In *Proceedings of the Americas Conference on Information Systems*, Presentation 4. AMCIS '13. Atlanta: Association for Information Systems (AIS).

Lomas, Natasha. 2014. "Digital News Finally As Popular As Newspapers In The UK." *TechCrunch*, June 25. http://social.techcrunch.com/2014/06/25/print-vs-digital/ (accessed 11.02.15).

Lubber, Mindy. 2012. "U.S. Companies Must Raise The Bar On Supply Chain Conditions." *Forbes*. http://www.forbes.com/sites/mindylubber/2012/05/17/raising-the-bar-on-supply-chains/ (accessed 28.12.12).

Luhn, H.P. 1957. "A Statistical Approach to Mechanized Encoding and Searching of Literary Information." *IBM Journal of Research and Development* 1 (4): 309–17.

Lu, Zhiyong. 2011. "PubMed and beyond: A Survey of Web Tools for Searching Biomedical Literature." *Database: The Journal of Biological Databases and Curation* 2011 (January): baq036.

Madlberger, Lisa, Heidelinde Hobel, Andreas Thöni, and A Min Tjoa. 2014. "Analysing Supplier Locations Using Social and Semantic Data: A Case Study Based on Indonesian Factories." In *Proceedings of the 14th International Conference on Knowledge Technologies and Data-Driven Business*, Article No. 21. iKnow '14. New York: ACM.

Madlberger, Lisa, Andreas Thöni, and Peter Wetz. 2013. "Ontology-Based Data Integration for Corporate Sustainability Information Systems." In *Proceedings of the 15th International Conference on Information Integration and Web-Based Applications & Services*, edited by Edgar Weippl, Maria Indrawan-Santiago, Matthias Steinbauer, Gabriele Kotsis, and Ismail Khalil, 353–57. iiWAS '13. New York: ACM.

Maedche, Alexander, and Steffen Staab. 2002. "Measuring Similarity between Ontologies." In *Knowledge Engineering and Knowledge Management: Ontologies and the Semantic Web*, edited by Asunción Gómez-Pérez and V. Richard Benjamins, 251–63. Lecture Notes in Computer Science 2473. Berlin Heidelberg: Springer.

Maes, Sam, Karl Tuyls, Bram Vanschoenwinkel, and Bernard Manderick. 1993. "Credit Card Fraud Detection Using Bayesian and Neural Networks." In *Interactive Image-Guided Neurosurgery. American Association Neurological Surgeons*, edited by R.J. Maciunas, 261–70.

Mair, Judith. 2011. "Exploring Air Travellers' Voluntary Carbon-Offsetting Behaviour." *Journal of Sustainable Tourism* 19 (2): 215–30.

Makris, Sotiris, P. Zoupas, and George Chryssolouris. 2011. "Supply Chain Control Logic for Enabling Adaptability under Uncertainty." *International Journal of Production Research* 49 (1): 121–37.

Maleki, Meysam, Viktoria Bashkite, and Virgilio Cruz Machado. 2012. "Integration of Supply Chain Performance with Customer Values through Combining Analytical Network Process and Bayesian Network." In *Annals of DAAAM for 2012 & Proceedings of the 23rd International DAAAM Symposium*, edited by Branko Katalinić, 297–300. Vienna: DAAAM International.

Maleki, Meysam, and V. Cruz-Machado. 2013. "Supply Chain Performance Monitoring Using Bayesian Network." *International Journal of Business Performance and Supply Chain Modelling* 5 (2): 177–97.

Mani, V., Rajat Agrawal, and Vinay Sharma. 2014. "Supplier Selection Using Social Sustainability: AHP Based Approach in India." *International Strategic Management Review* 2 (2): 98–112.

Manning, Christopher D., Prabhakar Raghavan, and Hinrich Schütze. 2008. *Introduction to Information Retrieval*. Cambridge: Cambridge University Press.

Manning, Christopher D., and Hinrich Schütze. 1999. *Foundations of Statistical Natural Language Processing*. Cambridge: MIT Press.

Manning, Christopher D., Mihai Surdeanu, John Bauer, Jenny Finkel, Steven J. Bethard, and David McClosky. 2014. "The Stanford CoreNLP Natural Language Processing Toolkit." In *Proceedings of 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, edited by Kristina Toutanova and Hua Wu, 55–60. ACL '14. Stroudsburg: Association for Computational Linguistics.

March, James G. 1978. "Bounded Rationality, Ambiguity, and the Engineering of Choice." *The Bell Journal of Economics* 9 (2): 587–608.

———. 1994. *Primer on Decision Making: How Decisions Happen*. New York: Simon and Schuster.

Mares, Radu. 2010. "The Limits of Supply Chain Responsibility: A Critical Analysis of Corporate Responsibility Instruments." *Nordic Journal of International Law* 79 (2): 193–244.

Marett, Kent, Robert F. Otondo, and G. Stephen Taylor. 2013. "Assessing the Effects of Benefits and Institutional Influences on the Continued Use of Environmentally Munificent Bypass Systems in Long-Haul Trucking." *MIS Quarterly* 37 (4): 1301–12.

Maron, M. E. 1961. "Automatic Indexing: An Experimental Inquiry." *Journal of the ACM* 8 (3): 404–17.

Martinez-Sala, Alejandro S., Esteban Egea-Lopez, Felipe Garcia-Sanchez, and Joan Garcia-Haro. 2009. "Tracking of Returnable Packaging and Transport Units with Active RFID in the Grocery Supply Chain." *Computers in Industry* 60 (3): 161–71.

Martin, Mervyn. 2013. "Child Labour: Parameters, Developmental Implications, Causes and Consequences." *Contemporary Social Science* 8 (2): 156–65.

Maybury, Mark T. 1995. "Generating Summaries from Event Data." *Information Processing & Management*, Summarizing Text, 31 (5): 735–51.

Maynard, D. G., and Sophia Ananiadou. 1999. *Incorporating Linguistic Information for Term Extraction*. http://cswww.essex.ac.uk/cluk/maynard.PDF (accessed 17.07.14).

Maynard, Diana, and Sophia Ananiadou. 2000. "Identifying Terms by Their Family and Friends." In *Proceedings of the 18th Conference on Computational Linguistics-Volume 1*, COLING '00:530–36. Saarbrücken: DFKI.

Maynard, Diana, Wim Peters, and Yaoyong Li. 2006. "Metrics for Evaluation of Ontology-Based Information Extraction." In *Proceedings of 4th International Evaluating Ontologies for the Web Workshop*. EON '06. Edinburgh.

———. 2008. "Evaluating Evaluation Metrics for Ontology-Based Applications: Infinite Reflection." In *Proceedings of the International Conference on Language Resources and Evaluation*, edited by Nicoletta Calzolari, Khalid Choukri, Bente Maegaard, Joseph Mariani, Jan Odijk, Stelios Piperidis, and Daniel Tapias, Paper 273. LREC '08. Paris: European Language Resources Association (ELRA).

Maynard, Diana, Milena Yankova, Alexandros Kourakis, and Antonis Kokossis. 2005. "Ontology-Based Information Extraction for Market Monitoring and Technology Watch." In *Proceedings of the Workshop on End User Aspects of the Semantic Web*, edited by Martin Dzbor, Hideaki Takeda, and Maria Vargas-Vera, 33–42. ESWC '05. Heraklion: CEUR Workshop Proceedings.

Mayring, Philipp. 2010. *Qualitative Inhaltsanalyse: Grundlagen Und Techniken*. 11th ed. Beltz.

McDougall, Dan, and Stefan Schmitz. 2007. "Otto-Konzern: Kinderarbeit Für Den Heine-Versand." *Stern*. February 11. http://www.stern.de/wirtschaft/news/otto-konzern-kinderarbeit-fuer-den-heine-versand-581923.html?p=2&nv=ct_cb (accessed 13.01.15).

Mcintyre, Kristie, Hugh A. Smith, Alex Henham, and John Pretlove. 1998. “Logistics Performance Measurement and Greening Supply Chains: Diverging Mindsets.” *International Journal of Logistics Management* 9 (1): 57–68.

McKeown, Kathleen. 1992. *Text Generation: Using Discourse Strategies and Focus Constraints to Generate Natural Language Text*. Cambridge New York Oakleigh: Cambridge University Press.

McKinnon, Alan C. 2010. “Product-Level Carbon Auditing of Supply Chains: Environmental Imperative or Wasteful Distraction?” *International Journal of Physical Distribution & Logistics Management* 40 (1/2): 42–60.

Meacham, Jeramy, Lisa Toms, Kenneth W. Green, and Vikram S. Bhadauria. 2013. “Impact of Information Sharing and Green Information Systems.” *Management Research Review* 36 (5): 478–94.

Meadows, Donella H., Dennis L. Meadows, Jorgen Randers, and William W. Behrens III. 1972. *The Limits to Growth; a Report for the Club of Rome’s Project on the Predicament of Mankind*. New York: Universe Books.

Melton, Genevieve B., and George Hripcsak. 2005. “Automated Detection of Adverse Events Using Natural Language Processing of Discharge Summaries.” *Journal of the American Medical Informatics Association : JAMIA* 12 (4): 448–57.

Melville, Nigel P. 2010. “Information Systems Innovation for Environmental Sustainability.” *MIS Quarterly* 34 (1): 1–21.

Mendes, Pablo, and Max Jakob. 2011. “Simple Web Service-Based Annotation Client.” https://github.com/dbpedia-spotlight/dbpedia-spotlight/blob/master/eval/src/main/java/org/dbpedia/spotlight/evaluation/external/DBpediaSpotlightClient.java (accessed 16.10.14).

Mendes, Pablo N., Max Jakob, Andrés García-Silva, and Christian Bizer. 2011. “DBpedia Spotlight: Shedding Light on the Web of Documents.” In *Proceedings of the 7th International Conference on Semantic Systems*, edited by Chiara Ghidini, Axel-Cyrille Ngonga Ngomo, Stefanie Lindstaedt, and Tassilo Pellegrini, 1–8. I-Semantics ’11. New York: ACM.

Meng, Li-yuan, and Rui Zhao. 2010. “Reputation Risk Management: A Corporate Social Responsibility Reporting Perspective.” In *2010 International Conference on Management and Service Science (MASS)*, 1–4.

Mentzer, John T., William DeWitt, James S. Keebler, Soonhong Min, Nancy W. Nix, Carlo D. Smith, and Zach G. Zacharia. 2001. “Defining Supply Chain Management.” *Journal of Business Logistics* 22 (2): 1–25.

Miemczyk, Joe, Thomas E. Johnsen, and Monica Macquet. 2012. “Sustainable Purchasing and Supply Management: A Structured Literature Review of Definitions and Measures at the Dyad, Chain and Network Levels.” *Supply Chain Management* 17 (5): 478–96.

Miller, George A. 1995. “WordNet: A Lexical Database for English.” *Communications of the ACM* 38 (11): 39–41.

Ministry of Statistics and Programme, and Implementation of India. 2006. *Provisional Results of Economic Census 2005 - All India Report*. New Delhi. http://mospi.nic.in/mospi_new/upload/economic_census_prov_results_2005.pdf (accessed 12.04.14).

Mittermayer, MA, and G Knolmayer. 2007. *Text Mining Systems for Market Response to News: A Survey*. No 184. Working Papers of the Institute of Information Systems. Bern: University of Bern. hhttp://www2.ie.iwi.unibe.ch/publikationen/berichte/resource/WP-184.pdf (accessed 11.11.14).

Mladenić, Dunja, Janez Brank, Marko Grobelnik, and Natasa Milic-Frayling. 2004. “Feature Selection Using Linear Classifier Weights: Interaction with Classification Models.” In *Proceedings of the 27th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, 234–41. SIGIR ’04. New York: ACM.

Modapothala, J.R., and B. Issac. 2009. “Study of Economic, Environmental and Social Factors in Sustainability Reports Using Text Mining and Bayesian Analysis.” In *Proceeedings of IEEE Symposium on Industrial Electronics Applications*, 1:209–14. ISIEA ’09. IEEE.

Modapothala, J. R., B. Issac, and E. Jayamani. 2010. “Appraising the Corporate Sustainability Reports – Text Mining and Multi-Discriminatory Analysis.” In *Innovations in Computing Sciences and Software Engineering*, edited by Tarek Sobh and Khaled Elleithy, 489–94. Springer Netherlands.

Moore, Gary C., and Izak Benbasat. 1991. “Development of an Instrument to Measure the Perceptions of Adopting an Information Technology Innovation.” *Information Systems Research* 2 (3): 192–222.

Moore, Gordon E. 1965. “Cramming More Components onto Integrated Circuits.” *Electronics* 38 (8): 56.

Moriizumi, Sara, Bongsung Chu, Haiyan Cao, and Hiroaki Matsukawa. 2011. “Supply Chain Risk Driver Extraction using Text Mining Technique.” *Information-an International Interdisciplinary Journal* 14 (6): 1935–45.

MSCI. 2014. *Global Industry Classification Standard Structure - GICS*. http://www.msci.com/products/indexes/sector/gics/gics_structure.html (accessed 09.03.15).

Muñoz, Edrisi, Elisabet Capón-García, José M. Laínez, Antonio Espuña, and Luis Puigjaner. 2013. “Considering Environmental Assessment in an Ontological Framework for Enterprise Sustainability.” *Journal of Cleaner Production* 47 (0): 149–64.

MySQL. 2009. “MySQL 5.5 Reference Manual: 12.9.4 Full-Text Stopwords.” http://dev.mysql.com/doc/refman/5.5/en/fulltext-stopwords.html (accessed 09.03.15).

Nagura, Ryosuke, Yohei Seki, Noriko Kando, and Masaki Aono. 2006. “A Method of Rating the Credibility of News Documents on the Web.” In *Proceedings of the 29th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, 683–84. SIGIR ’06. New York: ACM.

Naik, Ajaya Kumar. 2009. “Informal Sector and Informal Workers in India.” In *Proceedings of Special IARIW-SAIM Conference on “Measuring the Informal Economy in Developing Countries.”* Kathmandu.

Nakamura, Takuya, Philippe Capet, Thomas Delavallade, Julien Jacquelinet, Claude Martineau, Aude Rebotier, Ágnes Sándor, and Stavroula Voyatzi. 2007. “Intelligent Awareness: Event Extraction, Information Evaluation & Risk Assessment.” In *Proceedings of the 3rd Language and Technology Conference*, edited by Zygmunt Vetulani, 539–43. Poznan: Widawnictwo Poznanskie.

National Institute of Standards and Technology. 2014a. “Automatic Content Extraction (ACE) Evaluation.” http://www.itl.nist.gov/iad/mig/tests/ace/ (accessed 03.02.15).

———. 2014b. “Text Analysis Conference (TAC).” http://www.nist.gov/tac/ (accessed 03.02.15).

———. 2014c. “Topic Detection and Tracking Evaluation.” http://www.itl.nist.gov/iad/mig/tests/tdt/ (accessed 03.02.15).

National Sample Survey Office. 2011. *Employment and Unemployment Situation in India 2009-10*. Ministry of Statistics & Programme Implementation of India. http://www.indiaenvironmentportal.org.in/files/file/NSS_Report_employment%20and%20unemployment.pdf (accessed 12.04.14).

Naughton, Martina, Nicholas Kushmerick, and Joseph Carthy. 2006. “Event Extraction from Heterogeneous News Sources.” In *Proceedings of the AAAI Workshop Event Extraction and Synthesis*, edited by Naveen Ashish, Doug Appelt, Dayne Freitag, and Dmitry Zelenko, 1–6. Palo Alto: Association for the Advancement of Artificial Intelligence.

Naughton, M., N. Stokes, and J. Carthy. 2010. “Sentence-Level Event Classification in Unstructured Texts.” *Information Retrieval* 13 (2): 132–56.

Neapolitan, Richard E. 2003. *Learning Bayesian Networks*. Chicago: Prentice Hall.

Nikolaou, Ioannis E., Konstantinos I. Evangelinos, and S. Allan. 2013. "A Reverse Logistics Social Responsibility Evaluation Framework Based on the Triple Bottom Line Approach." *Journal of Cleaner Production* 56 (October): 173–84.

Nishihara, Yoko, Keita Sato, and Wataru Sunayama. 2009. "Event Extraction and Visualization for Obtaining Personal Experiences from Blogs." In *Proceedings of the Symposium on Human Interface 2009 on Human Interface and the Management of Information. Information and Interaction. Part II: Held As Part of HCI International 2009,* edited by Gavriel Salvendy and Michael J. Smith, 315–24. Berlin Heidelberg: Springer.

Nobelprize.org. 2014a. "Kailash Satyarthi - Nobel Lecture: Let Us Globalise Compassion, and Set Our Children Free." *Nobel Media AB 2014*. http://www.nobelprize.org/nobel_prizes/peace/laureates/2014/satyarthi-lecture_en.html (accessed 14.04.15).

———. 2014b. "The Nobel Peace Prize 2014 - Press Release." *Nobel Media AB*. http://www.nobelprize.org/nobel_prizes/peace/laureates/2014/press.html (accessed 13.01.15).

Norman, Geoff. 2010. "Likert Scales, Levels of Measurement and the 'laws' of Statistics." *Advances in Health Sciences Education* 15 (5): 625–32.

Norman, Wayne, and Chris MacDonald. 2004. "Getting to the Bottom of 'Triple Bottom Line.'" *Business Ethics Quarterly* 14 (2): 243–62.

Norrman, Andreas, and Ulf Jansson. 2004. "Ericsson's Proactive Supply Chain Risk Management Approach after a Serious Sub-Supplier Accident." *International Journal of Physical Distribution & Logistics Management* 34 (5): 434–56.

Norsys. 2013a. "Netica Application." http://www.norsys.com/netica.html (accessed 29.10.13).

———. 2013b. "Netica-J 5.04, for Programming Bayesian Networks in JAVA." http://www.norsys.com/netica-j.html (accessed 13.02.15).

Object Management Group. 2011. *OMG Unified Modeling LanguageTM (OMG UML), Superstructure*. Version 2.4.1. http://www.omg.org/spec/UML/2.4.1/ (accessed 09.03.15).

OECD Glossary. 2015. "Informal Sector – ILO Definition." http://stats.oecd.org/glossary/detail.asp?ID=1350 (accessed 09.03.15).

Ofcom. 2014. *News Consumption in the UK: 2014 Report*. http://stakeholders.ofcom.org.uk/binaries/research/tv-research/news/2014/News_Report_2014.pdf (accessed 02.11.15).

Olugu, Ezutah Udoncy, and Kuan Yew Wong. 2012. "An Expert Fuzzy Rule-Based System for Closed-Loop Supply Chain Performance Assessment in the Automotive Industry." *Expert Systems with Applications* 39 (1): 375–84.

Openlink. 2014. "Virtuoso Open-Source Edition 7.0." https://github.com/openlink/virtuoso-opensource (accessed 13.02.15).

Oxford Dictionary. 2012. "Information Technology: Definition of Information Technology." http://oxforddictionaries.com/definition/english/information%2Btechnology?q=information+technology (accessed 10.12.12).

———. 2013. "Ranking: Definition of Ranking." http://oxforddictionaries.com/definition/english/ranking?q=ranking (accessed 02.10.13).

———. 2014a. "Factory: Definition of Factory." http://www.oxforddictionaries.com/de/ definition/english/factory (accessed 20.11.14).

———. 2014b. "Independent: Definition of Independent." http://www.oxforddictionaries.com/definition/english/independent (accessed 26.06.14).

———. 2014c. "Coherence: Definition of Coherence." http://www.oxforddictionaries.com/definition/english/coherence (accessed 03.12.14).

———. 2014d. "Logical: Definition of Logical." http://www.oxforddictionaries.com/ definition/english/logical (accessed 03.12.14).

———. 2014e. “Consistency: Definition of Consistency.” http://www.oxforddictionaries.com/definition/english/consistency?searchDictCode=all (accessed 03.12.14).

———. 2014f. “Comparable: Definition of Comparable.” http://www.oxforddictionaries.com/definition/english/comparable?q=comparability#comparable__6 (accessed 03.12.14).

Öztürk, Burcu Avcı, and Funda Özçelik. 2014. “Sustainable Supplier Selection with A Fuzzy Multi-Criteria Decision Making Method Based on Triple Bottom Line.” *Business & Economics Research Journal* 5 (3): 129–47.

Padró, Lluís, Zeljko Agic, Xavier Carreras, Blaz Fortuna, Esteban García-Cuesta, Zhixing Li, Tadej Stajner, and Marko Tadić. 2014. “Language Processing Infrastructure in the XLike Project.” In *Proceedings of the International Conference on Language Resources and Evaluation*, edited by Nicoletta Calzolari (Conference Chair, Khalid Choukri, Thierry Declerck, Hrafn Loftsson, Bente Maegaard, Joseph Mariani, Asuncion Moreno, Jan Odijk, and Stelios Piperidis, 3811–16. LREC ’14. Paris: European Language Resources Association (ELRA).

Pagell, Mark, and Anton Shevchenko. 2014. “Why Research in Sustainable Supply Chain Management Should Have No Future.” *Journal of Supply Chain Management* 50 (1): 44–55.

Pai, Roshan R., Venkata R. Kallepalli, Reggie J. Caudill, and MengChu Zhou. 2003. “Methods toward Supply Chain Risk Analysis.” In *IEEE International Conference on Systems, Man and Cybernetics, 2003*, 5:4560–65 vol.5.

Paolucci, Massimo, Takahiro Kawamura, Terry R. Payne, and Katia Sycara. 2002. “Semantic Matching of Web Services Capabilities.” In *The Semantic Web — ISWC 2002*, edited by Ian Horrocks and James Hendler, 333–47. Lecture Notes in Computer Science 2342. Berlin Heidelberg: Springer.

Papka, Ron. 1999. “On-Line New Event Detection, Clustering, and Tracking.” Dissertation, University of Massachusetts Amherst.

Paradis, François, Jian-Yun Nie, and Arman Tajarobi. 2005. “Discovery of Business Opportunities on the Internet with Information Extraction.” In *Proceedings of Workshop on Multi-Agent Information Retrieval and Recommender Systems Held at Nineteen International Joint Conference on Artificial Intelligence*, edited by Esma Aïmeur, Robin Burke, and Abdel-Illah Mouaddib, 47–54. IJCAI ’05. Edinburgh.

Patcha, Animesh, and Jung-Min Park. 2007. “An Overview of Anomaly Detection Techniques: Existing Solutions and Latest Technological Trends.” *Computer Networks* 51 (12): 3448–70.

Patwardhan, Siddharth, and Ellen Riloff. 2009. “A Unified Model of Phrasal and Sentential Evidence for Information Extraction.” In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing: Volume 1-Volume 1*, edited by Philipp Koehn and Rada Mihalcea, 151–60. EMNLP ’09. Stroudsburg: Association for Computational Linguistics.

Pearl, Judea. 1988. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. San Francisco: Morgan Kaufmann Publishers Inc.

———. 1997. *Bayesian Networks*. http://www.stat.ucla.edu/papers/preprints/223/223.pdf (accessed 04.03.14).

Pedersen, Esben Rahbek, and Mette Andersen. 2006. “Safeguarding Corporate Social Responsibility (CSR) in Global Supply Chains: How Codes of Conduct Are Managed in Buyer-Supplier Relationships.” *Journal of Public Affairs* 6 (3-4): 228–40.

Peng, Qing Song. 2011. “Combine Influence Diagram with SCM Network.” *Advanced Materials Research* 271-273 (July): 98–101.

Petrak, Johann. 2013. “GATE Plugin String Annotation.” *Google Project Hosting*. https://code.google.com/p/gateplugin-stringannotation/ (accessed 16.02.15).

Petrini, Maira, and Marlei Pozzebon. 2009a. “Managing Sustainability with the Support of Business Intelligence: Integrating Socio-Environmental Indicators and Organisational Context.” *The Journal of Strategic Information Systems* 18 (4): 178–91.

———. 2009b. “Managing Sustainability with the Support of Business Intelligence: Integrating Socio-Environmental Indicators and Organisational Context.” *The Journal of Strategic Information Systems* 18 (4): 178–91.

Petrović, Sasa. 2012. “Real-Time Event Detection in Massive Streams.” Dissertation, Edinburgh: University of Edinburgh.

Pimentel, Marco A. F., David A. Clifton, Lei Clifton, and Lionel Tarassenko. 2014. “A Review of Novelty Detection.” *Signal Processing* 99 (June): 215–49.

Piotrowicz, Wojciech, and Richard Cuthbertson. 2009. “Sustainability - a New Dimension in Information Systems Evaluation.” *Journal of Enterprise Information Management* 22 (5): 492–503.

Piskorski, Jakub, Martin Atkinson, Jenya Belyaeva, Vanni Zavarella, Silja Huttunen, and Roman Yangarber. 2010. “Real-Time Text Mining in Multilingual News for the Creation of a Pre-Frontier Intelligence Picture.” In *Proceedings of ACM SIGKDD Workshop on Intelligence and Security Informatics*, 4:1–4:9. ISI-KDD ’10. New York: ACM.

Piskorski, Jakub, Hristo Tanev, Martin Atkinson, Eric van der Goot, and Vanni Zavarella. 2011. “Online News Event Extraction for Global Crisis Surveillance.” In *Transactions on Computational Collective Intelligence V*, edited by Ngoc Thanh Nguyen, 182–212. Lecture Notes in Computer Science 6910. Berlin Heidelberg: Springer.

Piskorski, Jakub, and Roman Yangarber. 2013. “Information Extraction: Past, Present and Future.” In *Multi-Source, Multilingual Information Extraction and Summarization*, edited by Thierry Poibeau, Horacio Saggion, Jakub Piskorski, and Roman Yangarber, 23–49. Theory and Applications of Natural Language Processing. Berlin Heidelberg: Springer.

Ploeger, Thomas, Maxine Kruijt, Lora Aroyo, Frank de Bakker, Iina Hellsten, Antske Fokkens, Jesper Hoeksema, and Serge ter Braake. 2013. “Extracting Activist Events from News Articles Using Existing NLP Tools and Services.” In *Proceedings of the Third Workshop on Detection, Representation and Exploitation of Events in the Semantic Web*, edited by Marieke van Erp, Laura Hollink, Raphaël Troncy, Willem Robert van Hage, Pierre van de Laar, David A. Shamma, and Lianli Gao, 30–41. DeRiVE ’13. Sydney: CEUR Workshop Proceedings.

Pouliquen, Bruno, Ralf Steinberger, and Camelia Ignat. 2003. “Automatic Annotation of Multilingual Text Collections with a Conceptual Thesaurus.” In *Proceedings of the Workshop Ontologies and Information Extraction at EUROLAN’2003*, 9–28. Bucharest.

Pouliquen, Bruno, Ralf Steinberger, Camelia Ignat, and Tom De Groeve. 2004. “Geographical Information Recognition and Visualization in Texts Written in Various Languages.” In *Proceedings of the 2004 ACM Symposium on Applied Computing*, 1051–58. New York: ACM.

Princeton University. 2010. “About WordNet.” *WordNet*. http://wordnet.princeton.edu (accessed 19.11.14).

Qazi, Abroon, John Quigley, Alex Dickson, Pedro Crespo Del Granado, Martim Joyce-Moniz, and Stefan Ravizza. 2014. “A Novel Framework for Quantification of Supply Chain Risks.” In *Proceedings of 4th Student Conference on Operational Research*, edited by Pedro Crespo Del Granado, Martim Joyce-Moniz, and Stefan Ravizza, 1–14. SCOR ’14. Saarbrücken Wadern: Dagstuhl Publishing.

Quinlan, J. Ross. 1993. *C4.5: Programs for Machine Learning*. San Francisco: Morgan Kaufmann Publishers Inc.

Rabenasolo, Besoa, and Xianyi Zeng. 2012. “A Risk-Based Multi-Criteria Decision Support System for Sustainable Development in the Textile Supply Chain.” In *Handbook on*

Decision Making, edited by Jie Lu, Lakhmi C. Jain, and Guangquan Zhang, 151–70. Intelligent Systems Reference Library 33. Berlin Heidelberg: Springer.

Radinsky, Kira, and Eric Horvitz. 2013. "Mining the Web to Predict Future Events." In *Proceedings of the Sixth ACM International Conference on Web Search and Data Mining*, 255–64. WSDM '11. New York: ACM.

Raiffa, Howard. 1968. *Decision Analysis: Introductory Lectures on Choices under Uncertainty*. New York: Random House.

Ramani, Karthik, Devarajan Ramanujan, William Z. Bernstein, Fu Zhao, John Sutherland, Carol Handwerker, Jun-Ki Choi, Harrison Kim, and Deborah Thurston. 2010. "Integrated Sustainable Life Cycle Design: A Review." *Journal of Mechanical Design* 132 (9): 1–15.

Rao, N. H. 2007. "A Framework for Implementing Information and Communication Technologies in Agricultural Development in India." *Technological Forecasting and Social Change* 74 (4): 491–518.

Raskin, Victor, Brian Buck, Arthur Keen, Christian F. Hempelmann, and Katrina E. Triezenberg. 2008. "Accessing and Manipulating Meaning of Textual and Data Information for Information Assurance and Security and Intelligence Information." In *Proceedings of the 4th Annual Workshop on Cyber Security and Information Intelligence Research*, edited by Frederick Sheldon, Axel Krings, Robert Abercrombie, and Ali Mili, 25:1–25:3. CSIIRW '08. New York: ACM.

Reckhow, Kenneth H. 1999. "Water Quality Prediction and Probability Network Models." *Canadian Journal of Fisheries and Aquatic Sciences* 56 (7): 1150–58.

Reefke, Hendrik, David Sundaram, and M. Daud Ahmed. 2010. "Maturity Progression Model for Sustainable Supply Chains." In *Advanced Manufacturing and Sustainable Logistics*, edited by Wilhelm Dangelmaier, Alexander Blecken, Robin Delius, and Stefan Klöpfer, 308–19. Lecture Notes in Business Information Processing 46. Berlin Heidelberg: Springer.

Refaeilzadeh, Payam, Lei Tang, and Huan Liu. 2009. "Cross-Validation." In *Encyclopedia of Database Systems*, edited by Ling Liu and M. Tamer Özsu, 532–38. Springer US.

Reips, Ulf-Dietrich. 2002. "Standards for Internet-Based Experimenting." *Experimental Psychology (formerly Zeitschrift Für Experimentelle Psychologie)* 49 (4): 243–56.

Reuter, Carsten, Kai Foerstl, Evi Hartmann, and Constantin Blome. 2010. "Sustainable Global Supplier Management: The Role of Dynamic Capabilities in Achieving Competitive Advantage." *Journal of Supply Chain Management* 46 (2): 45–63.

Reuter, Nadine, Svitlana Vakulenko, Jan vom Brocke, Stefan Debortoli, and Oliver Müller. 2014. "Identifying the Role of Information Systems in Achieving Energy-Related Environmental Sustainability Using Text Mining." In *Proceedings of the European Conference on Information Systems*. ECIS '14. Atlanta: Association for Information Systems (AIS).

Reuters, and National Institute of Standards and Technology. 2009. "Retuers TRC2 Corpus." http://trec.nist.gov/data/reuters/reuters.html (accessed 23.12.13).

Rivera, Samuel J., Barbara S. Minsker, Daniel B. Work, and Dan Roth. 2014. "A Text Mining Framework for Advancing Sustainability Indicators." *Environmental Modelling & Software* 62 (December): 128–38.

Rizzi, Stefano. 2012. "Collaborative Business Intelligence." In *Business Intelligence*, edited by Marie-Aude Aufaure and Esteban Zimányi, 186–205. Lecture Notes in Business Information Processing 96. Berlin Heidelberg: Springer.

Rizzo, Giuseppe, and Raphaël Troncy. 2012. "NERD: Named Entity Recognition and Disambiguation, Ontology." http://nerd.eurecom.fr/ontology#Location (accessed 01.06.14).

Rizzo, Giuseppe, Raphaël Troncy, Sebastian Hellmann, and Martin Bruemmer. 2012a. "NERD Meets NIF: Lifting NLP Extraction Results to the Linked Data Cloud." *LDOW* 937.

———. 2012b. “NERD Meets NIF: Lifting NLP Extraction Results to the Linked Data Cloud.” In *Proceedings of Workshop on Linked Data on the Web*, edited by Christian Bizer, Tom Heath, Tim Berners-Lee, and Michael Hausenblas. Vol. 937. LDOW ’12. Lyon: CEUR Workshop Proceedings.

Rizzo, Giuseppe, Marieke van Erp, and Raphaël Troncy. 2014. “Benchmarking the Extraction and Disambiguation of Named Entities on the Semantic Web.” In *Proceedings of the International Conference on Language Resources and Evaluation*, edited by Nicoletta Calzolari (Conference Chair, Khalid Choukri, Thierry Declerck, Hrafn Loftsson, Bente Maegaard, Joseph Mariani, Asuncion Moreno, Jan Odijk, and Stelios Piperidis. LREC ’14. Paris: European Language Resources Association (ELRA).

Rockström, Johan, Will Steffen, Kevin Noone, Asa Persson, F. Stuart Chapin, Eric F. Lambin, Timothy M. Lenton, Marten Scheffer, Carl Folke, and Hans Joachim Schellnhuber. 2009. “A Safe Operating Space for Humanity.” *Nature* 461 (7263): 472–75.

Rogati, Monica, and Yiming Yang. 2002. “High-Performing Feature Selection for Text Classification.” In *Proceedings of the Eleventh International Conference on Information and Knowledge Management*, 659–61. CIKM ’02. New York: ACM.

Rogers, Emma. 2011. “Exercising Responsibility in the Seafood Supply Chain.” Masterthesis, Lund: Lund University.

Rönninger, Stephan, and Malcolm Holmes. 2009. “A Risk-Based Approach to Scheduling Audits.” *PDA Journal of Pharmaceutical Science and Technology* 63 (6): 575–88.

Rosell, Pablo Diego, and Art Hansen. 2012. *Children Working in the Carpet Industry of India: Prevalence and Conditions*. Washington, Calverton: United States Department of Labor, ICF International. http://www.dol.gov/ilab/iclre/Downloads/Research/Report/PC_India_Report.pdf (accessed 02.04.14).

Ruggie, John. 2008. *Promotion and Protection of All Human Rights, Civil, Political, Economic, Social and Cultural Rights, Including the Right to Development*. Human Rights Council. http://www.reports-and-materials.org/Ruggie-companion-report-15-May-2008.pdf (accessed 20.01.14).

Sabou, Marta, Chris Wroe, Carole Goble, and Gilad Mishne. 2005. “Learning Domain Ontologies for Web Service Descriptions: An Experiment in Bioinformatics.” In *Proceedings of the 14th International Conference on World Wide Web*, 190–98. WWW ’05. New York: ACM.

Sakaki, Takeshi, Makoto Okazaki, and Yutaka Matsuo. 2010. “Earthquake Shakes Twitter Users: Real-Time Event Detection by Social Sensors.” In *Proceedings of the 19th International Conference on World Wide Web*, 851–60. New York: ACM.

Sándor, Ágnes. 2009. “Automatic Detection of Discourse Indicating Emerging Risk.” *Critical Approaches to Discourse Analysis across Disciplines* 4 (2): 171–79.

Sankaranarayanan, Jagan, Hanan Samet, Benjamin E. Teitler, Michael D. Lieberman, and Jon Sperling. 2009. “TwitterStand: News in Tweets.” In *Proceedings of the 17th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, 42–51. GIS ’09. New York: ACM.

Santos, António Paulo, Paulo Bengala, and Nuno C. Marques. 2013. “Economic Activity Recognition.” In *Proceedings of the 16th Portuguese Conference on Artificial Intelligence*, edited by Luís Paulo Reis, José Cascalho, Luís Gomes Mendes, Hélia Guerra, and Pedro Cardoso, 332–43. EPIA ’13. Angra do Heroısmo: University of Azores.

Sarawagi, Sunita. 2007. “Information Extraction.” *Foundations and Trends in Databases* 1 (3): 261–377.

Sarkis, Joseph, and Dileep G. Dhavale. 2014. “Supplier Selection for Sustainable Operations: A Triple-Bottom-Line Approach Using a Bayesian Framework.” *International Journal of Production Economics*, in proof.

Sarkis, Joseph, Marilyn Michelle Helms, and Aref A. Hervani. 2010. "Reverse Logistics and Social Sustainability." *Corporate Social Responsibility and Environmental Management* 17 (6): 337–54.

Sarkis, Joseph, Qinghua Zhu, and Kee-hung Lai. 2011. "An Organizational Theoretic Review of Green Supply Chain Management Literature." *International Journal of Production Economics* 130 (1): 1–15.

Sarkki, Simo, Jari Niemelä, Rob Tinch, Sybille van den Hove, Allan Watt, and Juliette Young. 2013. "Balancing Credibility, Relevance and Legitimacy: A Critical Assessment of Trade-Offs in Science–policy Interfaces." *Science and Public Policy*, August, 1–13.

Schaltegger, Stefan, and Roger Burritt. 2014. "Measuring and Managing Sustainability Performance of Supply Chains." *Supply Chain Management: An International Journal* 19 (3): 232–41.

Schaltegger, Stefan, Sarah E. Windolph, and Dorli Harms. 2010. *Corporate Sustainability Barometer*. http://www.leuphana.de/institute/csm/forschung-projekte/corporate-sustainability-barometer.html (accessed 20.05.13).

Scharl, Arno, Alexander Hubmann-Haidvogel, Albert Weichselbraun, Heinz-Peter Lang, and Marta Sabou. 2013. "Media Watch on Climate Change -- Visual Analytics for Aggregating and Managing Environmental Knowledge from Online Sources." In *Proceedings of 46th Hawaii International Conference on System Sciences*, edited by Ralph H. Sprague, 955–64. HICSS '13. Los Alamitos Washington Tokyo: IEEE.

Schatten, Alexander. 2009. "'Green' Supply Chains: Using Information Integration for Sustainable Development." In *Proceedings of Conference on Intelligent-Systems for Engineering Environmental Knowledge Held on International Conference on Complex, Intelligent and Software Intensive Systems*, 1124–27. CISIS/i-SEEK '09. IEEE.

Schiebel, Walter, and Siegi Pöchtrager. 2003. "Corporate Ethics as a Factor for Success–the Measurement Instrument of the University of Agricultural Sciences (BOKU), Vienna." *Supply Chain Management: An International Journal* 8 (2): 116–21.

Schiller, Ben. 2012. "Transport's Cleaner Future. (cover Story)." *Financial Management (14719185)*, February, 26–30.

Schmid, Helmut. 1994. "Probabilistic Part-of-Speech Tagging Using Decision Trees." In *Proceedings of International Conference on New Methods in Language Processing*, 44–49. Manchester.

———. 1995. "Improvements In Part-of-Speech Tagging With an Application To German." In *Proceedings of the ACL SIGDAT Workshop*, 47–50.

Schmidt, Isabell, Manfred Meurer, Peter Saling, Andreas Kicherer, Wolfgang Reuter, and Carl-Otto Gensch. 2004. "SEEbalance: Managing Sustainability of Products and Processes with the Socio-Eco-Efficiency Analysis by BASF." *Greener Management International*, no. 45 (Spring): 79–94.

Schwarz, Julia, and Meredith Morris. 2011. "Augmenting Web Pages and Search Results to Support Credibility Assessment." In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 1245–54. ACM SIGCHI '11. New York: ACM.

Sebyala, Abdallah Abbey, Temitope Olukemi, and Lionel Sacks. 2002. "Active Platform Security through Intrusion Detection Using Naive Bayesian Network for Anomaly Detection." In *London Communications Symposium*. London: Communications Engineering Doctorate Centre.

Serrano, Laurie, Maroua Bouzid, Thierry Charnois, Stephan Brunessaux, and Bruno Grilheres. 2013. "Events Extraction and Aggregation for Open Source Intelligence: From Text to Knowledge." In *Proceedings of the 25th International Conference onTools with Artificial Intelligence*, edited by Randall Bilof, 518–23. ICTAI '13. Los Alamitos Washington Tokyo: IEEE.

Seuring, Stefan. 2013. "A Review of Modeling Approaches for Sustainable Supply Chain Management." *Decision Support Systems* 54 (4): 1513–20.

Seuring, Stefan, and Stefan Gold. 2012. “Conducting Content-Analysis Based Literature Reviews in Supply Chain Management.” *Supply Chain Management: An International Journal* 17 (5): 544–55.

Seuring, Stefan, and Martin Müller. 2008a. “Core Issues in Sustainable Supply Chain Management – a Delphi Study.” *Business Strategy and the Environment* 17 (8): 455–66.

———. 2008b. “From a Literature Review to a Conceptual Framework for Sustainable Supply Chain Management.” *Journal of Cleaner Production* 16 (15): 1699–1710.

Shafer, Glenn, and Rajendra P. Srivastava. 1990. “The Bayesian and Belief-Function Formalisms: A General Perspective for Auditing.” *Auditing: A Journal of Practice & Theory* 9: 110–37.

Shanteau, James. 1988. “Psychological Characteristics and Strategies of Expert Decision Makers.” *Acta Psychologica* 68 (1–3): 203–15.

Shevtshenko, Eduard, and Yan Wang. 2009. “Decision Support under Uncertainties Based on Robust Bayesian Networks in Reverse Logistics Management.” *International Journal of Computer Applications in Technology*, 36 (3/4): 247–58.

Shi, Xianliang, Ling Xia Li, Lili Yang, Zhihua Li, and Jung Y. Choi. 2012. “Information Flow in Reverse Logistics: An Industrial Information Integration Study.” *Information Technology & Management* 13 (4): 217–32.

Shuaib, M., H. Metta, T. Lu, F. Badurdeen, I. S. Jawahir, and T. Goldsby. 2011. “Design and Performance Evaluation of Sustainable Supply Chains: Approach and Methodologies.” In *Advances in Sustainable Manufacturing*, edited by Günther Seliger, Marwan M. K. Khraisheh, and I. S. Jawahir, 347–52. Springer.

Siddiqi, Ahmed F. 2013. “Important Determinants of Child Labor: A Case Study for Lahore.” *American Journal of Economics and Sociology* 72 (1): 199–221.

Sigala, Marianna. 2014. “Customer Involvement in Sustainable Supply Chain Management: A Research Framework and Implications in Tourism.” *Cornell Hospitality Quarterly* 55 (1): 76–88.

Simchi-Levi, David, Philip Kaminsky, and Edith Simchi-Levi. 2003. *Designing & Manageing the Supply Chain - Concepts, Strategies & Case Studies*. 2nd ed. New York: McGraw-Hill/Irwin.

Smith, David A. 2002. “Detecting and Browsing Events in Unstructured Text.” In *Proceedings of the 25th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, 73–80. SIGIR ’02. New York: ACM.

Social Accountability International. 2008a. “Social Accountability 8000:2008.”

———. 2008b. *Social Accountability 8000:2008*. http://www.sa-intl.org/_data/n_0001/ resources/live/2008StdEnglishFinal.pdf (accessed 11.01.13).

Sodhi, ManMohan S., and Christopher S. Tang. 2009. “Managing Supply Chain Disruptions via Time-Based Risk Management.” In *Managing Supply Chain Risk and Vulnerability*, edited by Teresa Wu and Jennifer Blackhurst, 29–40. London: Springer.

Soosay, Claudine, Andrew Fearne, and Mohsen Varsei. 2014. “Extending Sustainable Practices Beyond Organizations to Supply Chains.” In *Linking Local and Global Sustainability*, edited by Sukhbir Sandhu, Stephen McKenzie, and Howard Harris, 71–90. The International Society of Business, Economics, and Ethics Book Series 4. Springer Netherlands.

Sourirajan, K., P. Centonze, M. E. Helander, K. Katircioglu, M. Ben-Hamida, and C. Boucher. 2009. “Carbon Management in Assembly Manufacturing Logistics.” *IBM Journal of Research and Development* 53 (3): 433–48.

Spence, Laura, and Michael Bourlakis. 2009. “The Evolution from Corporate Social Responsibility to Supply Chain Responsibility: The Case of Waitrose.” *Supply Chain Management: An International Journal* 14 (4): 291–302.

Spiegelhalter, David J., A. Philip Dawid, Steffen L. Lauritzen, and Robert G. Cowell. 1993. “Bayesian Analysis in Expert Systems.” *Statistical Science* 8 (3): 219–47.

Spiegelhalter, David J., and Steffen L. Lauritzen. 1990. “Sequential Updating of Conditional Probabilities on Directed Graphical Structures.” *Networks* 20 (5): 579–605.

Srivastava, Samir K. 2007. “Green Supply-Chain Management: A State-of-the-Art Literature Review.” *International Journal of Management Reviews* 9 (1): 53–80.

Statistik Austria. 2010. “Integrated System of International Classifications.” http://www.statistik.at/web_en/classifications/classification_database/integrated_system_of_international_classifications/ (accessed 10.01.15).

Steinberger, Ralf, Bruno Pouliquen, and Erik van der Goot. 2009. “An Introduction to the Europe Media Monitor Family of Applications.” In *Proceedings of the SIGIR 2009 Workshop - Information Access in a Multilingual World*, edited by Fredric Gey, Noriko Kando, and Jussi Karlgren, 1–8. SIGIR-CLIR ’09. New York: ACM.

Stoll, Richard J., and Devika Subramanian. 2006. *Hubs, Authorities, and Networks: Predicting Conflict Using Events Data*. Annual Meeting of International Studies Association 2006. San Diego. http://www.cs.rice.edu/~devika/conflict/papers/stollsubramanian.pdf (accessed 10.11.14).

Stolz, Alex, Bene Rodriguez-Castro, Andreas Radinger, and Martin Hepp. 2014. “PCS2OWL: A Generic Approach for Deriving Web Ontologies from Product Classification Systems.” In *The Semantic Web: Trends and Challenges*, edited by Valentina Presutti, Claudia d’ Amato, Fabien Gandon, Mathieu d’ Aquin, Steffen Staab, and Anna Tordai, 644–58. Lecture Notes in Computer Science 8465. Springer International Publishing.

Suh, Kyo, Timothy Smith, and Michelle Linhoff. 2012. “Leveraging Socially Networked Mobile ICT Platforms for the Last-Mile Delivery Problem.” *Environmental Science & Technology* 46 (17): 9481−9490.

Sun, Aixin, and Ee-Peng Lim. 2001. “Hierarchical Text Classification and Evaluation.” In *Proceedings of the 2001 IEEE International Conference on Data Mining*, edited by Nick Cercone, T.Y. Lin, and Xindong Wu, 521–28. ICDM ’01. Los Alamitos Washington Brussels Tokyo: IEEE.

Sustainable Trade Initiative. 2012. *Child Labour Platform: Report 2010 – 2011*. http://www.hivos.nl/dut/content/download/79767/691352/file/Child_Labour_Platform_Report_2010-11.pdf (accessed 18.11.13).

Svensson, Göran. 2009. “The Transparency of SCM Ethics: Conceptual Framework and Empirical Illustrations.” *Supply Chain Management: An International Journal* 14 (4): 259–69.

Szyperski, Norbert. 1980. “Informationsbedarf.” *Handwörterbuch Der Organisation* 2: 904–13.

Tablan, Valentin, and Ian Roberts. 2013. *Mímir User Guide: 5.0 Snapshot*. University of Sheffield Dept. of Computer Science. https://gate.ac.uk/mimir/doc/mimir-guide.pdf (accessed 22.01.14).

Tablan, V., K. Bontcheva, I. Roberts, and H. Cunningham. 2015. “Mímir: An Open-Source Semantic Search Framework for Interactive Information Seeking and Discovery.” *Journal of Web Semantics* 30 (January): 52–68.

Taghaboni-Dutta, Fataneh, Amy J.C. Trappey, and Charles V. Trappey. 2010. “An XML Based Supply Chain Integration Hub for Green Product Lifecycle Management.” *Advances in Aligning Knowledge Systems, Improving Business Logistics, Driving Innovation and Adapting Customer Centric Services* 37 (11): 7319–28.

Tajbakhsh, Alireza, and Elkafi Hassini. 2014. “A Data Envelopment Analysis Approach to Evaluate Sustainability in Supply Chain Networks.” *Journal of Cleaner Production*, in press.

Tanaka, Ryuichi. 2003. “Inequality as a Determinant of Child Labor.” *Economics Letters* 80 (1): 93–97.

Tanev, Hristo, Jakub Piskorski, and Martin Atkinson. 2008. “Real-Time News Event Extraction for Global Crisis Monitoring.” In *Natural Language and Information Systems*, edited by Epaminondas Kapetanios, Vijayan Sugumaran, and Myra Spiliopoulou, 207–18. Lecture Notes in Computer Science 5039. Berlin Heidelberg: Springer.

Tang, Christopher S. 2006. "Perspectives in Supply Chain Risk Management." *International Journal of Production Economics* 103 (2): 451–88.

Tang, Rong, Kwong Bor Ng, Tomek Strzalkowski, and Paul B. Kantor. 2003. "Automatically Predicting Information Quality in News Documents." In *Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology: Companion Volume of the Proceedings of HLT-NAACL 2003–short Papers - Volume 2*, 97–99. NAACL '03. Stroudsburg: Association for Computational Linguistics.

Taticchi, Paolo, Patrizia Garengo, Sai S. Nudurupati, Flavio Tonelli, and Roberto Pasqualino. 2014. "A Review of Decision-Support Tools and Performance Measurement and Sustainable Supply Chain Management." *International Journal of Production Research*, 1–22.

Taticchi, Paolo, Flavio Tonelli, and Roberto Pasqualino. 2013. "Performance Measurement of Sustainable Supply Chains: A Literature Review and a Research Agenda." *International Journal of Productivity and Performance Management* 62 (8): 782–804.

Taylor, Mark B., Luc Zandvliet, and Mitra Forouhar. 2009. *Due Diligence for Human Rights: A Risk-Based Approach*. Working Paper No. 53. Corporate Social Responsibility Initiative. http://www.hks.harvard.edu/m-rcbg/CSRI/publications/workingpaper_53_taylor_etal.pdf (accessed 25.10.13).

Taylor, Shirley, and Peter A. Todd. 1995. "Understanding Information Technology Usage: A Test of Competing Models." *Information Systems Research* 6 (2): 144–76.

Techopedia. 2014. "What Is an Instance? - Definition from Techopedia." *Techopedia.com*. http://www.techopedia.com/definition/16325/instance (accessed 04.08.14).

Tesnière, Lucien. 1959. *Eléments de Syntaxe Structurale*. Librairie C. Klincksieck.

Teuscher, Peter, Beat Grüninger, and Niels Ferdinand. 2006. "Risk Management in Sustainable Supply Chain Management (SSCM): Lessons Learnt from the Case of GMO-Free Soybeans." *Corporate Social Responsibility and Environmental Management* 13 (1): 1–10.

Teuteberg, Frank, and David Wittstruck. 2010. "A Systematic Review of Sustainable Supply Chain Management Research." In *Proceedings Der Multikonferenz Wirtschaftsinformatik 2010*, edited by Kolbe Schumann and Breitner, 1001–15. Universitätsverlag Göttingen.

The Association for Computing Machinery. 2015. "ACM Computing Surveys: Information for Authors." April 10. http://surveys.acm.org/author_info.html (accessed 10.04.15).

The Association of Insurance and Risk Managers, The Public Risk Management Association, and The Institute of Risk Management. 2010. *A Structured Approach to Enterprise Risk Management (ERM) and the Requirements of ISO 31000*. http://www.theirm.org/media/886062/ISO3100_doc.pdf (accessed 14.08.14).

The Danish Institute for Human Rights. 2006. *Human Righs Compliance Assessment - Quick Qheck Version 1.1.4*. https://hrca2.humanrightsbusiness.org/docs/file/HRCA%20Quick%20Check_English.pdf (accessed 16.10.13).

The International Business Leaders Forum, and The International Finance Corporation. 2010. *Human Rights Impact Assessment*. http://www.ifc.org/wps/wcm/connect/8ecd35004c0cb230884bc9ec6f601fe4/hriam-guide-092011.pdf?MOD=AJPERES (accessed 25.10.13).

The Telegraph. 2013. "Warren Buffett: His Best Quotes," February 14. http://www.telegraph.co.uk/finance/newsbysector/banksandfinance/8381363/Warren-Buffett-his-best-quotes.html (accessed 04.06.15).

The World Bank Group. 2013. "World Development Indicators | The World Bank." http://wdi.worldbank.org/table/2.6 (accessed 05.03.14).

Thielsch, Meinald T., and Simone Weltzin. 2009. “Online-Befragungen in Der Praxis.” In *Praxis Der Wirtschaftspsychologie: Themen Und Fallbeispiele Für Studium Und Praxis*, edited by Torsten Brandenburg and Meinald T. Thielsch, 69–85. Münster: MV Wissenschaft.

Thöni, Andreas. 2013. “Integrating Linked Open Data for Improved Social Sustainability Risk Management in Supply Chains.” In *Lecture Notes in Informatics*, edited by Matthias Horbach, 916–27. Bonn: Gesellschaft für Informatik.

Thöni, Andreas, Jack King, and A Min Tjoa. 2014. “Ongoing Social Sustainability Monitoring in Supply Chains.” In *Proceedings of the European Conference on Information Systems (ECIS) 2014*. Tel Aviv.

Thöni, Andreas, Lisa Madlberger, and Alexander Schatten. 2013a. “Companies as Drivers of Sustainability – Towards Requirements for an Integrative Sustainability Risk Management System.” *International SERIES on Information Systems and Management in Creative eMedia*, SAME ’13, Proceedings of the 6th International Workshop on Semantic Ambient Media Experience (2): 1–7.

———. 2013b. “Towards a Data-Integration Approach for Enterprise Sustainability Risk Information Systems.” In *CONFENIS-2013: 7th International Conference on Research and Practical Issues of Enterprise Information Systems*, edited by Josef Basl, Pavel Jasek, Ota Novotny, and A Min Tjoa, 269–77. Linz: Trauner.

Thorleuchter, Dirk, Tobias Scheja, and Dirk Van den Poel. 2014. “Semantic Weak Signal Tracing.” *Expert Systems with Applications* 41 (11): 5009–16.

Thorleuchter, Dirk, and Dirk Van den Poel. 2013. “Weak Signal Identification with Semantic Web Mining.” *Expert Systems with Applications* 40 (12): 4978–85.

Thoroe, Lars, Adam Melski, and Matthias Schumann. 2009. “RFID in Reverse Logistics Research Framework and Roadmap.” *Wirtschaftinformatik Proceedings 2009*, January, Paper 77.

Thurston, Matthew, and Matthew J. Eckelman. 2011. “Assessing Greenhouse Gas Emissions from University Purchases.” *International Journal of Sustainability in Higher Education* 12 (3): 225–35.

Toh, Keith T. K., Pieter Nagel, and Roger Oakden. 2009. “A Business and ICT Architecture for a Logistics City.” *International Journal of Production Economics* 122 (1): 216–28.

Torii, Manabu, Lanlan Yin, Thang Nguyen, Chand T. Mazumdar, Hongfang Liu, David M. Hartley, and Noele P. Nelson. 2011. “An Exploratory Study of a Text Classification Framework for Internet-Based Surveillance of Emerging Epidemics.” *International Journal of Medical Informatics* 80 (1): 56–66.

Toutanova, Kristina, Dan Klein, Christopher D. Manning, and Yoram Singer. 2003. “Feature-Rich Part-of-Speech Tagging with a Cyclic Dependency Network.” In *Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology-Volume 1*, 173–80. NAACL ’03. Stroudsburg: Association for Computational Linguistics.

Toutanova, Kristina, and Christopher D. Manning. 2000. “Enriching the Knowledge Sources Used in a Maximum Entropy Part-of-Speech Tagger.” In *Proceedings of the 2000 Joint SIGDAT Conference on Empirical Methods in Natural Language Processing and Very Large Corpora-Volume 13*, edited by Hinrich Schfitze and Keh-Yih Su, 63–70. EMNLP ’00. Stroudsburg: Association for Computational Linguistics.

Trappey, Amy J. C., Charles V. Trappey, and Chang-Ru Wu. 2010. “Genetic Algorithm Dynamic Performance Evaluation for RFID Reverse Logistic Management.” *Expert Systems with Applications* 37 (11): 7329–35.

Travis, Alan. 2014. “UK Firms to Face New Rules Aimed at Ending Slavery in Supply Chains.” *The Guardian*, October 13, sec. World news. http://www.theguardian.com/world/2014/oct/13/uk-firms-new-rules-ending-slavery-supply-chains (accessed 15.10.14).

Trienekens, Jacques H., P. M. (Nel) Wognum, Adrie J. M. Beulens, and Jack G. A. J. van der Vorst. 2012. "Transparency in Complex Dynamic Food Supply Chains." *Advanced Engineering Informatics* 26 (1): 55–65.

Tsukayama, Hayley. 2012. "Foxconn and Apple Supply: What Are the Effects?" *The Washington Post*. September 24. http://www.washingtonpost.com/business/economy/foxconn-and-apple-supply-what-are-the-effects/2012/09/24/126b9966-0670-11e2-afff-d6c7f20a83bf_story.html (accessed 04.12.12).

Tushi, Bonny, Darshana Sedera, and Jan Recker. 2014. "Green IT Segment Analysis: An Academic Literature Review." In *Proceedings of Twentieth Americas Conference on Information Systems*, 2566–80. AMCIS '14. Atlanta: Association for Information Systems (AIS).

Twitter. 2014. "Character Counting." *Twitter Developers*. https://dev.twitter.com/overview/api/counting-characters (accessed 11.02.15).

———. 2015a. "Things Every Developer Should Know." *Twitter Developers*. https://dev.twitter.com/overview/general/things-every-developer-should-know (accessed 10.02.15).

———. 2015b. "The Streaming APIs." *Twitter Developers*. https://dev.twitter.com/ streaming/overview (accessed 11.02.15).

Uhart, Mathieu, Laurent Bourguignon, Pascal Maire, and Michel Ducher. 2012. "Bayesian Networks as Decision-Making Tools to Help Pharmacists Evaluate and Optimise Hospital Drug Supply Chain." *European Journal of Hospital Pharmacy: Science and Practice* 19 (6): 519–24.

Understanding Children's Work. 2007. *Measuring Child Labour: Discussion Note for Country Consultation in Bangladesh*. http://www.ucw-project.org/attachment/ICLS_ discussion_note_Bangladesh_11sept2007.pdf (accessed 20.08.14).

———. 2009. "Indonesia ,Child Labour Survey 2009." http://www.ucw-project.org/Pages/Tables.aspx?id=1372 (accessed 11.02.15).

———. 2010. "India, National Sample Survey Round 66 (NSS-R66)." http://www.ucw-project.org/Pages/Tables.aspx?id=1540 (accessed 10.04.14).

———. 2012. *Understanding Children's Work and Youth Employment Outcomes in Indonesia*. Rome: Understanding Children's Work Programme. http://www.ucw-project.org/attachment/Child_labour_Youth_employment_Indonesia20120625_ 162630.pdf (accessed 08.08.14).

———. 2014. "UCW Project - Youth Employment Outcomes, Youth Employment Statistics, Child Labour Legal Standards, Understanding Child Labour, Understanding Employment Outcomes, Domestic Chores." http://www.ucw-project.org/Pages/ChildLab Indicator.aspx (accessed 12.04.14).

Understanding Children's Work, and International Labour Office. 2010. *Child Labour Trends, Challenges and Policy Responses: Joining Forces against Child Labour*. Geneva: International Labour Organization.

UNEP. 2009. *Guidelines for Social Life Cycle Assessment of Products*. United Nations Environment Programme. http://socialhotspot.org/userfiles/guidelines-sLCA.pdf (accessed 04.11.12).

UN Global Compact Nordic Network. 2013. "Global Compact Self Assessment." http://www.globalcompactselfassessment.org/ (accessed 15.10.13).

United Nations Statistics Division. 2008a. *International Standard Industrial Classification of All Economic Activities (ISIC), Rev.4*. New York: United Nations.

———. 2008b. *Central Product Classification (CPC) Ver.2*. http://unstats.un.org/unsd/cr/registry/cpc-2.asp (accessed 22.07.14).

———. 2014. "ISIC Rev.4." http://unstats.un.org/unsd/cr/registry/isic-4.asp (accessed 22.07.14).

United States Department of Labor. 2013. “Reducing Child Labor and Forced Labor - Addressing Root Causes.” http://www.dol.gov/ilab/child-forced-labor/Addressing-Root-Causes.htm (accessed 25.10.13).

University of Illinois. 2012. “Technology: About the Carbon Capture Report.” http://www.carboncapturereport.org/about-technology.html (accessed 06.11.14).

Uusitalo, Laura. 2007. “Advantages and Challenges of Bayesian Networks in Environmental Modelling.” *Ecological Modelling* 203 (3–4): 312–18.

Vandenberg, Paul, Anne-Brit Nippierd, and Sandy Gros-Louis. 2007a. *Eliminating Child Labour: How Employers Can Eliminate Child Labour*. Geneva: Bureau for Employers’ Activities, International Labour Office, and the International Organization of Employers (IOE).

———. 2007b. *Eliminating Child Labour: The Role of Employers’ Organizations in Combating Child Labour*. Geneva: Bureau for employers’ activities, International Labour Office, and the International Organization of Employers (IOE).

Van Landeghem, Sofie, Bernard De Baets, Yves Van de Peer, and Yvan Saeys. 2011. “High-Precision Bio-Molecular Event Extraction from Text Using Parallel Binary Classifiers.” *Computational Intelligence* 27 (4): 645–64.

Vannieuwenhuyse, B., L. Gelders, and L. Pintelon. 2003. “An Online Decision Support System for Transportation Mode Choice.” *Logistics Information Management* 16 (2): 125–33.

Van Rijsbergen, C.J. 1975. *Information Retrieval*. London: Butterworths.

Vargas-Vera, M., and D. Celjuska. 2004. “Event Recognition on News Stories and Semi-Automatic Population of an Ontology.” In *Proceedings of IEEE/WIC/ACM International Conference on Web Intelligence*, edited by Ning Zhong, Henry Tirri, Yiyu Yao, Lizhu Zhou, Jiming Liu, and Nick Cercone, 615–18. WI ’04. Los Alamitos Washington Brussels Tokyo: IEEE.

Venkatesh, Viswanath, and Hillol Bala. 2008. “Technology Acceptance Model 3 and a Research Agenda on Interventions.” *Decision Sciences* 39 (2): 273–315.

Venkatesh, Viswanath, and Fred D. Davis. 2000. “A Theoretical Extension of the Technology Acceptance Model: Four Longitudinal Field Studies.” *Management Science* 46 (2): 186–204.

Verdecho, M. J., J. J. Alfaro-Saiz, and R. Rodriguez–Rodriguez. 2014. “A Performance Measurement Framework for Monitoring Supply Chain Sustainability.” In *Annals of Industrial Engineering 2012*, 331–38. London: Springer.

Verma, Sanjeeve, and Ranjan Chaudhuri. 2009. “Creating Customer Satisfaction And Profitable Value Chain With E-Commerce In Rural India: A Case Based Approach.” *International Journal of Business Insights & Transformation* 2 (1): 51–63.

Vermeulen, Walter J. V., and Stefan Seuring. 2009. “Sustainability through the Market – the Impacts of Sustainable Supply Chain Management: Introduction.” *Sustainable Development* 17 (5): 269–73.

Von Carlowitz, Hans Carl. 1713. *Sylvicultura Oeconomica, Oder Haußwirthliche Nachricht Und Naturmäßige Anweisung Zur Wilden Baum Zucht*. Leipzig.

W3C. 2014. “RDF Schema 1.1.” February 25. http://www.w3.org/TR/2014/REC-rdf-schema-20140225/ (accessed 09.03.15).

Walker, Helen, and Stephen Brammer. 2012. “The Relationship between Sustainable Procurement and E-Procurement in the Public Sector.” *International Journal of Production Economics* 140 (1): 256–68.

Waltinger, Ulli. 2010a. “GermanPolarityClues: A Lexical Resource for German Sentiment Analysis.” In *Proceedings of the International Conference on Language Resources and Evaluation*, edited by Nicoletta Calzolari, Khalid Choukri, Bente Maegaard, Joseph Mariani, Jan Odijk, Stelios Piperidis, Mike Rosner, and Daniel Tapias, 1638–42. LREC ’10. Paris: European Language Resources Association (ELRA).

———. 2010b. “Sentiment Analysis Reloaded: A Comparative Study On Sentiment Polarity Identification Combining Machine Learning And Subjectivity Features.” In *Proceedings of the 6th International Conference on Web Information Systems and Technologies,* edited by Joaquim Filipe and José Cordeiro, 1:203–10. WEBIST ’10. INSTICC Press.

Wang, Hsiao-Fan, and Surendra M. Gupta. 2011. *Green Supply Chain Management: Product Life Cycle Approach*. 1st ed. McGraw-Hill Professional.

Watthayu, W., and Y. Peng. 2004. “A Bayesian Network Based Framework for Multi-Criteria Decision Making.” In *Proceedings of the 17th International Conference on Multiple Criteria Decision Analysis*. MCDM ’04. Whistler.

Webbink, Ellen, Jeroen Smits, and Eelke de Jong. 2013. “Household and Context Determinants of Child Labor in 221 Districts of 18 Developing Countries.” *Social Indicators Research* 110 (2): 819–36.

Webby, Robyn. 2006. “Insulating the Supply Network from Corporate Social Responsibility Risk.” In *Proceedings of 91st Annual International Supply Management Conference*. Tempe: Institute for Supply Management.

Weber, Olaf, Roland W. Scholz, and Georg Michalik. 2010. “Incorporating Sustainability Criteria into Credit Risk Management.” *Business Strategy and the Environment* 19 (1): 39–50.

Wei, Chih-Ping, and Yen-Hsien Lee. 2004. “Event Detection from Online News Documents for Supporting Environmental Scanning.” *Decision Support Systems,* Knowledge Management Technique, 36 (4): 385–401.

Weichselbraun, Albert, Daniel Streiff, and Arno Scharl. 2014. “Linked Enterprise Data for Fine Grained Named Entity Linking and Web Intelligence.” In , 1–11. ACM Press.

Weiler, Andreas, Svetlana Mansmann, and Marc H. Scholl. 2012. “Towards an Advanced System for Real-Time Event Detection in High-Volume Data Streams.” In *Proceedings of the 5th Ph.D. Workshop on Information and Knowledge,* 87–90. PIKM ’12. New York: ACM.

Weisbrot, Mark. 2014. “What Standards of Child Labor Should Apply Overseas?” *The New York Times,* July 17. http://www.nytimes.com/roomfordebate/2014/07/16/what-standards-of-child-labor-should-apply-in-developing-countries/to-end-child-labor-washington-must-press-companies-to-act (accessed 18.07.14).

Wen, Lei, Longxiu Xu, and Rui Wang. 2013. “Sustainable Supplier Evaluation Based on Intuitionistic Fuzzy Sets Group Decision Methods.” *Journal of Information and Computational Science* 10 (10): 3209–20.

Wigas, Wade M. 2006. *Likert-Type Scale Response Anchors*. Clemson International Institute for Tourism & Research Development. http://www.clemson.edu/centers-institutes/tourism/documents/sample-scales.pdf (accessed 27.10.14).

Williams, Andrew. 2003. “Linking the Environmental and Social Dimensions of Corporate Social Responsibility.” In *Proceedings of the the 10th International Conference of the Greening of Industry Network*. GIN 2002. Göteborg.

Willis, Craig, and Robert M. Losee. 2013. “A Random Walk on an Ontology: Using Thesaurus Structure for Automatic Subject Indexing.” *Journal of the American Society for Information Science and Technology* 64 (7): 1330–44.

Willoughby, William F. 1890. “Child Labor.” *Publications of the American Economic Association* 5 (2): 5–70.

Winter, Marc, and A. Michael Knemeyer. 2013. “Exploring the Integration of Sustainability and Supply Chain Management: Current State and Opportunities for Future Inquiry.” *International Journal of Physical Distribution & Logistics Management* 43 (1): 18–38.

Wittstruck, David, and Frank Teuteberg. 2012. “Integrating the Concept of Sustainability into the Partner Selection Process: A Fuzzy-AHP-TOPSIS Approach.” *International Journal of Logistics Systems and Management* 12 (2): 195–226.

Wognum, P. M. (Nel), Harry Bremmers, Jacques H. Trienekens, Jack G.A.J. van der Vorst, and Jacqueline M. Bloemhof. 2011. “Systems for Sustainability and Transparency of Food Supply Chains – Current Status and Challenges.” *RFID and Sustainable Value Chains* 25 (1): 65–76.

Wood, Donna J. 1991. “Corporate Social Performance Revisited.” *The Academy of Management Review* 16 (4): 691–718.

Wooldridge, S. 2003. *Bayesian Belief Networks*. CSIRO Centre for Complex Systems Science. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.112.6230&rep=rep1&type=pdf (accessed 13.12.12).

WordNet. 2014. “WNGLOSS - Glossary of Terms Used in WordNet System.” Accessed November 19. http://wordnet.princeton.edu/man/wngloss.7WN.html (accessed 19.11.14).

World Commission on Environment and Development. 1987. *Our Common Future*. Oxford: Oxford Paperbacks.

World Economic Forum. 2009. *Supply Chain Decarbonization - The Role of Logisitcs and Transport in Reducing Supply Chain Carbon Emissions*. https://members.weforum.org/pdf/ip/SupplyChainDecarbonization.pdf (accessed 29.10.12).

———. 2012. *New Models for Addressing Supply Chain and Transport Risk*. Industry Agenda. Cologny/Geneva: World Economic Forum. http://www3.weforum.org/docs/WEF_SCT_RRN_NewModelsAddressingSupplyChainTransportRisk_IndustryAgenda_2012.pdf (accessed 04.12.12).

Wu, Chung-Min, Ching-Lin Hsieh, and Kuei-Lun Chang. 2013. “A Hybrid Multiple Criteria Decision Making Model for Supplier Selection.” *Mathematical Problems in Engineering* 2013 (September): Article ID 324283.

Wu, Desheng Dash, Shu-Heng Chen, and David L. Olson. 2014. “Business Intelligence in Risk Management: Some Recent Progresses.” *Information Sciences* 256 (January): 1–7.

Wunderwald, Martin. 2011. “NewsX - Event Extraction from News Articles.” Dissertation, Dresden: Dresden University of Technology.

Xie, Lexing, Hari Sundaram, and Murray Campbell. 2008. “Event Mining in Multimedia Streams.” *Proceedings of the IEEE* 96 (4): 623–47.

Xu, Lei, D. Thresh Kumar, K. Madan Shankar, Devika Kannan, and Gang Chen. 2013. “Analyzing Criteria and Sub-Criteria for the Corporate Social Responsibility-Based Supplier Selection Process Using AHP.” *International Journal of Advanced Manufacturing Technology* 68 (1-4): 907–16.

Yakovleva, Natalia, Joseph Sarkis, and Thomas Sloan. 2012. “Sustainable Benchmarking of Supply Chains: The Case of the Food Industry.” *International Journal of Production Research* 50 (5): 1297–1317.

Yangarber, Roman, Ralph Grishman, and Pasi Tapanainen. 2000. “Automatic Acquisition of Domain Knowledge for Information Extraction.” In *Proceedings of the 18th International Conference on Computational Linguistics*, 940–46.

Yang, Yiming, and Xin Liu. 1999. “A Re-Examination of Text Categorization Methods.” In *Proceedings of the 22nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, 42–49. SIGIR ’99. New York: ACM.

Yang, Yiming, Tom Pierce, and Jaime Carbonell. 1998. “A Study of Retrospective and On-Line Event Detection.” In *Proceedings of the 21st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, 28–36. SIGIR ’98. New York: ACM.

Yen, B.P.-C., and Bingcong Zeng. 2010. “A Hierarchical Assessment Method Using Bayesian Network for Material Risk Detection on Green Supply Chain.” In *Proceedings of IEEE International Conference on Industrial Engineering and Engineering Management*, 1184–88. IEEM ’10. IEEE.

Yilmaz, Ayse Kucuk, and Triant Flouris. 2010. “Managing Corporate Sustainability: Risk Management Process Based Perspective.” *African Journal of Business Management* 4 (2): 162–71.

Yoon, Janghyeok. 2012. “Detecting Weak Signals for Long-Term Business Opportunities Using Text Mining of Web News.” *Expert Systems with Applications* 39 (16): 12543–50.

Yuan, Changhe, Feng Cheng, Henry Dao, Markus Ettl, Grace Lin, and Karthik Sourirajan. 2012. “A Bayesian Framework for Supply Chain Risk Management Using Business Process Standards.” In *The Handbook of Integrated Risk Management in Global Supply Chains*, edited by Panos Kouvelis, Lingxiu Dong, Onur Boyabatli, and Rong Li, 537–64. John Wiley & Sons, Inc.

Yuan, Wen-Feng, Si-feng Liu, Zhi-Geng Fang, and Hong-Zhuan Chen. 2009. “Comprehensive Evaluation on Suppliers of Aircraft Based on Grey Bayesian Network Model.” In *Proceedings of the IEEE International Conference on Grey Systems and Intelligent Services (GSIS) 2009*, 552–58.

Zeit Online. 2014. “Armut: Bolivien Erlaubt Kinderarbeit Ab Zehn Jahren.” *Die Zeit*, July 4, sec. Ausland. http://www.zeit.de/politik/ausland/2014-07/bolivien-erlaubt-kinderarbeit (accessed 01.02.15).

Zeydan, Mithat, Cüneyt Çolpan, and Cemal Çobanoğlu. 2011. “A Combined Methodology for Supplier Selection and Performance Evaluation.” *Expert Systems with Applications* 38 (3): 2741–51.

Zha, Hongyuan. 2002. “Generic Summarization and Keyphrase Extraction Using Mutual Reinforcement Principle and Sentence Clustering.” In *Proceedings of the 25th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, 113–20. SIGIR ’02. New York: ACM.

Zhang, Qiannan, Tian Huang, Yongxin Zhu, and Meikang Qiu. 2013. “A Case Study of Sensor Data Collection and Analysis in Smart City: Provenance in Smart Food Supply Chain.” *International Journal of Distributed Sensor Networks*, article ID 382132.

Zhang, Yi, and Bing Liu. 2007. “Semantic Text Classification of Emergent Disease Reports.” In *Knowledge Discovery in Databases: PKDD 2007*, edited by Joost N. Kok, Jacek Koronacki, Ramon Lopez de Mantaras, Stan Matwin, Dunja Mladenič, and Andrzej Skowron, 629–37. Lecture Notes in Computer Science 4702. Berlin Heidelberg: Springer.

Zhao, Yan, and Yuguang Wang. 2012. “Text Categorization Based on Emergency Domain Words: A System Engineering View.” *Systems Engineering Procedia*, Safety and Emergency Systems Engineering, 5: 8–14.

Zheng, Ying, and H. Zhou. 2012. “An Intelligent Text Mining System Applied to SEC Documents.” In *Proceedings of IEEE/ACIS 11th International Conference on Computer and Information Science*, edited by Juan E. Guerrero, 155–60. ICIS ’12. Los Alamitos Washington Tokyo: IEEE.

Zografos, Konstantinos G., Norbert Sedlacek, and Jeroen Bozuwa. 2012. “A Comparative Assessment of Freight Transport and Logistics Policies in Europe.” *Procedia - Social and Behavioral Sciences* 48 (0): 2523–32.

Zsidisin, George A., Lisa M. Ellram, Joseph R. Carter, and Joseph L. Cavinato. 2004. “An Analysis of Supply Risk Assessment Techniques.” *International Journal of Physical Distribution & Logistics Management* 34 (5): 397–413.

Zutshi, Ambika, Andrew Creed, and Amrik Sohal. 2009. “Child Labour and Supply Chain: Profitability or (mis)management.” *European Business Review* 21 (1): 42–63.

Appendix prototype implementation

This chapter of the appendix will detail technical features. Particularly, three parts may be differentiated. First, an overall prototype was implemented that integrates the methodologies described in the main text. It initializes a risk framework based on pre-defined supplier locations, accepts texts from an input source, and integrates all available evidence to produce a risk-based ranking of suppliers. The second and a third parts of the prototype are the technical details of the risk model and the TM application. All three parts of the system are outlined below. Although the focus of the evaluation is on TM and the risk model, the proposed prototype fulfils three important roles: In the design stage it served as a system to verify the necessity of different parameters, it may be seen as a design suggestion for future implementations of comparable systems, and it functions as a viability check of the integration potential of the different technologies applied.

Overarching system design and implementation

Architecture

The system is partitioned into three parts according to a general logic for Business Intelligence (BI) systems that differentiates three layers (Kemper, Baars, and Mehanna 2010, 11): one for providing the data, one for generating information and distributing it, and finally one for accessing the generated information. Given that the system does not provide a graphical user interface (apart from an administrative system state viewer), the focus is on the former two components. Several additional architecture principles were also followed:

- Add all observations to database independent of supplier locations
- Allow manual access of observations by user to control risk ratings
- Decouple database access, BNs, and observations

The data model is based on Semantic Web RDF triples, and Virtuoso (Openlink 2014) is used as the underlying triple store.[185] This store is loaded with the ontologies used in TM (alignment ontology, CPC ontology, ISIC ontology, domain model in RDF). A domain-specific access-layer is used throughout the application to access the data. Generating and distributing information is performed using a controller thread started by another class which subsequently triggers different system components which tag the text, extract and fold observations, and integrate the data into the risk model. Each component is provided to the controller through an interface (see Figure 82).

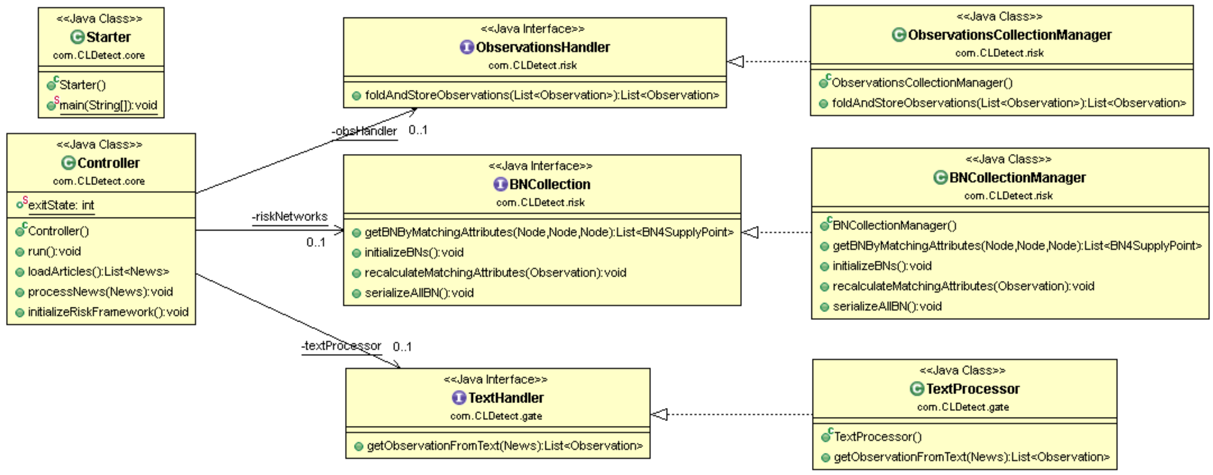


Figure 82: Classes and interfaces representing the main components of the system (author's representation; created with ObjectAid)[186]

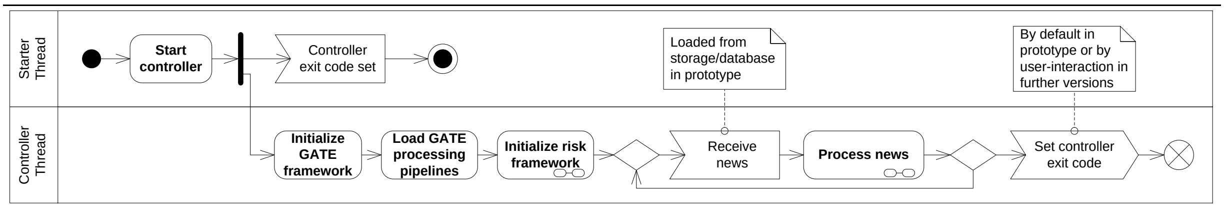
This structural, static view of the system will be supplemented with the dynamic view in the next section.

Workflow

The basic system workflow is dominated by a starter and a controller running in different threads. Figure 83 depicts this overarching system workflow. A starter thread starts the controller thread until an exit code is set.[187] The triggered controller thread begins by loading the text processor, which analyzes incoming news using the General Architecture for Text Engineering (GATE), the GATE ANNIE framework (Cunningham et al. 2013; Cunningham et al. 2002; Cunningham et al. 2011), and the observation handler that processes the observations retrieved from the text (e.g. folding; see below). The next step initializes the risk framework,

[185] The Virtuoso database is provided by Openlink in both commercial and open source versions (http://virtuoso.openlinksw.com/). It has previously been used, among others, to create DBpedia, a Semantic Web version of Wikipedia. Triple stores allow triples to be directly stored and retrieved using SPARQL queries.
[186] http://www.objectaid.com/.
[187] The exit code is monitored at second-based intervals.

which maintains a Netica risk model (Norsys 2013a) for each supplier location. The program is then ready to receive and process news.



Note: "Initialize risk framework is detailed in Figure 84 and "process news" in Figure 85

Figure 83: Overarching system workflow (author's representation)

Initializing the risk framework involves several steps, outlined in Figure 84. After loading the supply points (i.e., the supplier locations) a distinct BN is created for each location. At first the more static values are retrieved from the database, then the initially available observations are integrated into the BN network. Updates are directly performed by the BN class. After these processing steps the BN is ready to integrate further observations.

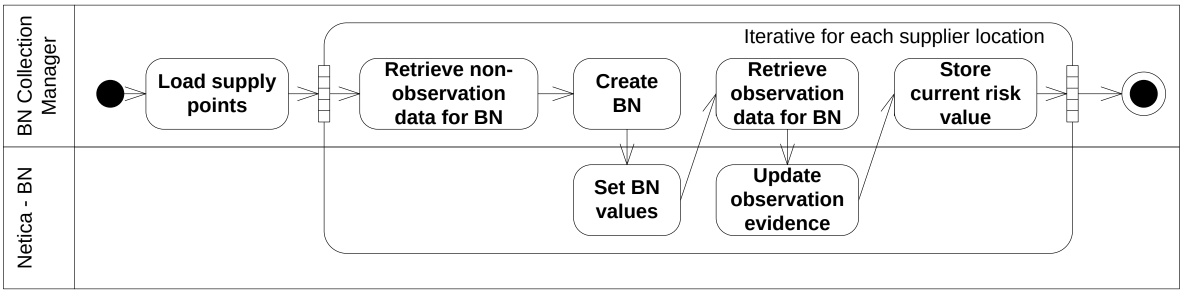


Figure 84: Initialize risk framework workflow (author's representation)

When a text is received, it is processed in multiple steps. First, the text is handled by the GATE processing pipeline outlined in the right half of Figure 85. This pipeline will be depicted in more detail below. After filtering the dimensions, the child labor incident observations are loaded from the text. These observations are subsequently folded, meaning that new observations (i.e., those which do not overlap or conflict with existing dimensions) are stored in the database and observations that relate to existing observations (i.e., those which are similar or overlap with existing observations) are updated in the database.

The observations are then used to update the risk values for different supplier locations. The affected BNs for the supplier locations are determined for each observation with changes (i.e. new or updated). For each of these BNs, the observations are loaded and the observational risk value is recalculated. Finally, the overall risk prioritization is updated.[188]

[188] Details on additional packages used are given later in the appendix.

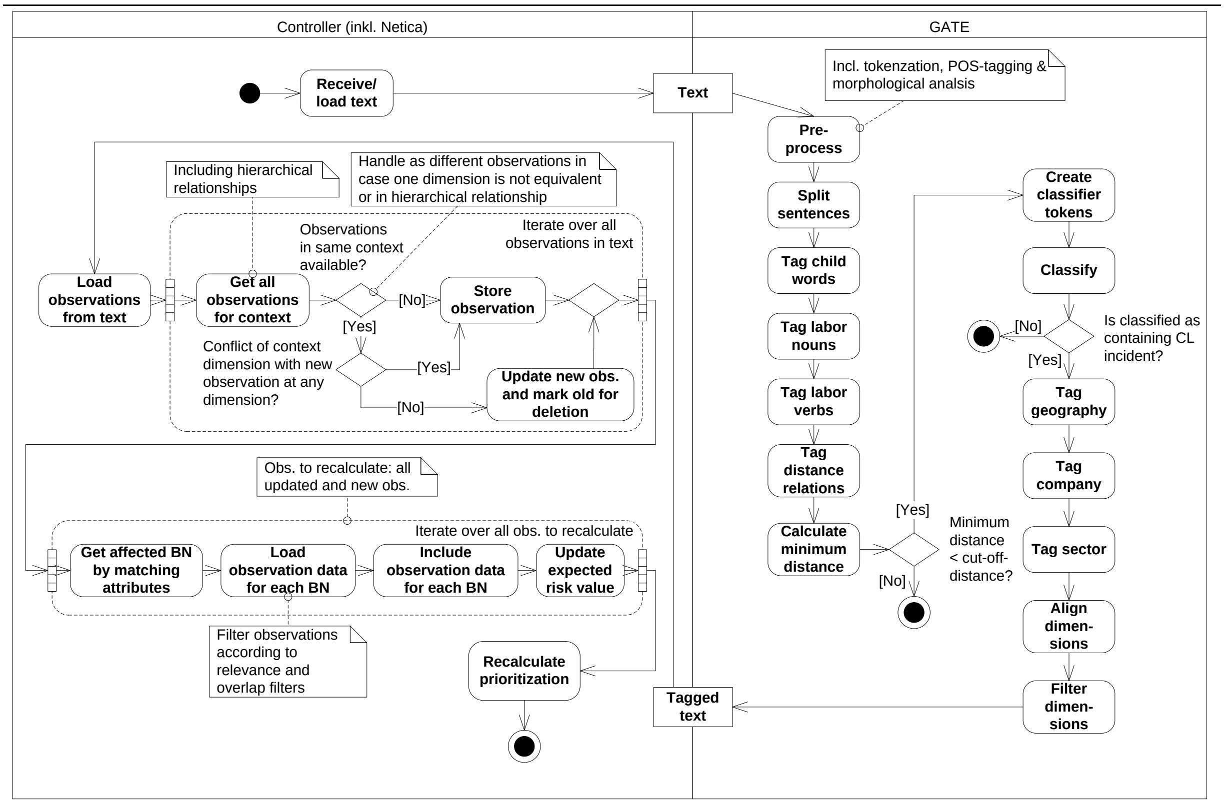


Figure 85: Process news workflow for one text document (author's representation)

Technical risk model implementation (Bayesian network)

The following paragraphs will outline the technical details of the risk model implementation. The risk model is based on the Bayesian network implementation Netica, developed by Norsys (2013a). The model was designed using the graphical Netica interface. Within the overall prototype, the NeticaJ application programming interface (API) is used to access the data via an API instead of the graphical interface (Norsys 2013b).[189]

Risk model structure

The risk model can be decomposed into four components: the prior, the audits, the observations, and the combined overall breach likelihood. The following paragraphs will outline these structural parts.

Contextual prior

As described in the main text, the prior depends on the country, the sector, and the ruralness. To reflect this, three discrete Netica nodes are used to differentiate the different values that these variables may take. The country variable needs to include all supplier countries, the sector node all highest-level ISIC sectors, and the ruralness node the states "urban" and "rural".

[189] The final Netica model can be seen in "Appendix L: Bayesian network implementation in Netica".

The combining node reflecting the prior then needs to differentiate several normal distributions where the mean μ_{prior} is selected dependent on the input states (Script 1 shows the equation of the prior node in Netica syntax).[190] The sigma is set to be constant for all combinations. To include the sigma of the prior σ_{prior} a constant node has been used (Figure 86).

```
P (P_prior | Ruralness, Sector, Country) =
 (Country == India && Sector == Manufacturing && Ruralness == Urban) ?
NormalDist(P_prior,3.9,s) :
 (Country == India && Sector == Manufacturing && Ruralness == Rural) ?
NormalDist(P_prior,5.9,s) :
 (Country == India && Sector == Construction && Ruralness == Urban) ?
NormalDist(P_prior,23.6,s) :
 (Country == India && Sector == Construction && Ruralness == Rural) ?
NormalDist(P_prior,44.7,s) :
 (Country == Indonesia && Sector == Manufacturing && Ruralness == Urban) ?
NormalDist(P_prior,4.9,s) :
 (Country == Indonesia && Sector == Manufacturing && Ruralness == Rural) ?
NormalDist(P_prior,2.6,s) :
 (Country == Indonesia && Sector == Construction && Ruralness == Urban) ?
NormalDist(P_prior,0.0,s) :
 (Country == Indonesia && Sector == Construction && Ruralness == Rural) ?
NormalDist(P_prior,1.5,s) :
0
```

Script 1: Netica equation for prior probability P_{prior} (author's representation)

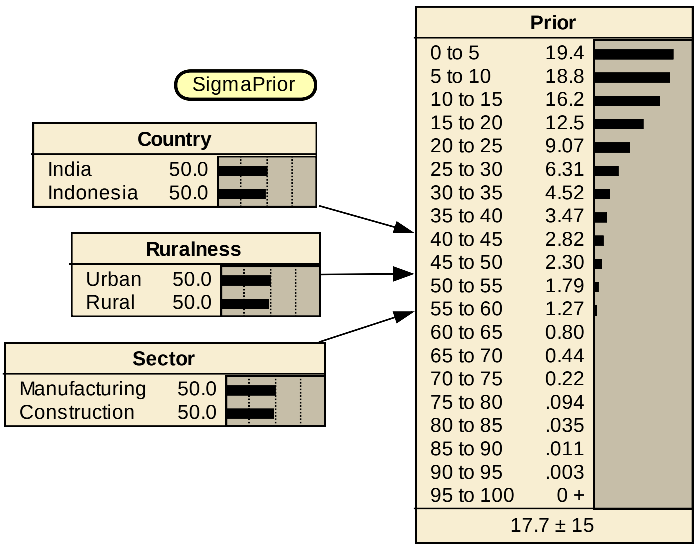


Figure 86: BN nodes for prior likelihood P_{prior} (author's representation)

Audits

The audit likelihood is based on the result of the last audit as well as the time since the last audit. To represent the audit score in the BN, it is discretized into five intervals. For the prototype the minimum audit score a_{min} is set to 0 and the maximum a_{max} to 5. For the prior a normal distribution is assumed with the mean at 4 and a standard deviation of 1. If an introductory audit for a supplier is assumed and a finding is entered (i.e., marking one discretiza-

[190] A later version could implement this using variables.

tion with 100% probability), the prior for the node has no influence on the testing outcome. For the time since the last audit, we assume a continuous stochastic variable T with a continuous value greater or equal to 0. The time is measured in months since the last audit. For representation in the model, this value is again discretized into three phases of "short" (< 6 months), "medium" (6 < 24 months), and "long" (≥ 24 months). For the prototype, we assume a prior value of a mean of 9 months with a deviation of 3 months. Later this will need to be set to company specific data, e.g., based on historic audit results over a year. Again, entering a finding eliminates the influence of the prior. Consequently, the values in Table 94 are prototypically assumed depending on the time horizon since the last audit.

Time since last audit	Time since last audit [months]	Assumed standard deviation [percent points]
Short	>= 0 && <6	5
Medium	>= 6 && <24	10
Long	>=24	15

Table 94: Audit probability standard deviation values depending on time since last audit (author's representation)

The Netica equation used for P_{audit} is defined in Script 2, resulting in the representation in Figure 87. The formula used for the combination was explained in the main text. The values are also based on the input from the expert questionnaire presented in the evaluation section of this thesis.

```
P (P_audit | taudit, ascore) =
(taudit==short) ? NormalDist (P_audit, 57-ascore*8.8, 5) :
(taudit==medium) ? NormalDist (P_audit, 57-ascore*8.8, 10) :
                   NormalDist (P_audit, 57-ascore*8.8, 15)
```

Script 2: Netica equation for audit likelihood P_{audit} (author's representation)

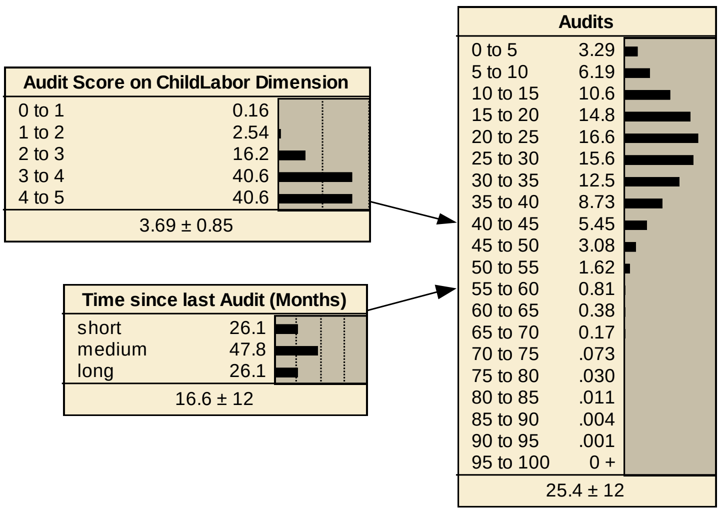


Figure 87: BN nodes for audit likelihood P_{audit} (author's representation)

Observations

The nodes preceding the observational likelihood are used to reflect the current state of observational evidence for a supplier location. Each new child labor incident observation triggers cases for relevance and credibility as well as an increase in the count for the frequency. The first section below explains the Netica variables, while the second explains the update process. Similar to the prior and audit nodes, the observational likelihood P_{obs} is modeled as a continuous stochastic variable. A normal distribution is assumed where the probability varies around a mean depending on the frequency f (the number of relevant observations for a location), the expected value of the credibility $E(c)$, and the average relevance $E(r)$. These three factors are combined in a function x that is then transformed with a scaling function s with an upper bound of 100 as required by the observational likelihood. Ultimately, the equation from Script 3 and described in the model description is used for the observational probability, resulting in the overall network outlined in Figure 88.

```
P (P_obs | f, r, c) =
NormalDist(P_obs, ((1-(1/(1+(f*(c+r))/5)))*100), (10.204-(0.765*f*r*c)))
```

Script 3: Netica equation for observational probability P_{obs} (author's representation)

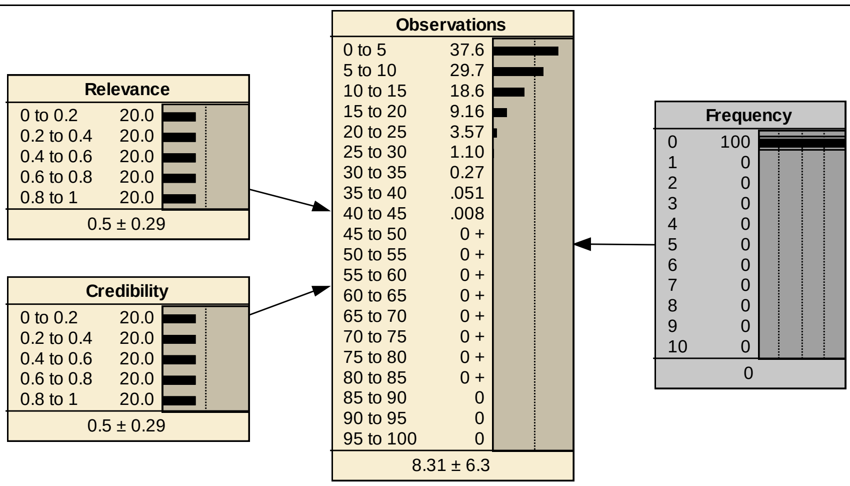


Figure 88: BN nodes for observational likelihood P_{obs} (author's representation)

Variables

Credibility and relevance are modeled using random variables reflecting the probability distributions of the credibility and relevance of child labor incident observations related to a supplier location. A uniform prior distribution is assumed for both nodes. The empirical distribution is determined via continuous updates. For each child labor incident observation a specific credibility and relevance value is calculated and used as an input to the network. Both values are deduced by the program based on the features of the observation. These are the configuration parameters of the risk model.[191] The values used in the prototypical implementation are configured as outlined in Table 95 and Table 96 and align with the values gathered by the expert survey.

Attributes of child labor incident observation		Credibility c
Publishing channel	**Examples**	
Social media	Twitter, Facebook	0.2
Web 2.0 Website	Blogspot	0.5
News channel	CNN, BBC, Reuters	0.8

Table 95: Credibility values based on publishing channels (own representation)

Attributes of child labor incident observation			Relevance r
Geographic features	**Company features**	**Sector features**	
None	Company name	None	0.5
None	None	ISIC sector	0.0
None	Company name	ISIC sector	0.8
Country	None	None	0.1
Region	None	None	0.3
City	None	None	0.5
Country	None	ISIC sector	0.2
Region	None	ISIC sector	0.5
City	None	ISIC sector	0.7
Country	Company name	None	0.7
Region	Company name	None	0.8
City	Company name	None	0.9
Country	Company name	ISIC sector	0.7
Region	Company name	ISIC sector	0.8
City	Company name	ISIC sector	0.9

Table 96: Relevance values based on dimensional attribute availability (author's representation)

The frequency f is also included as a "count node". It is defined as the number of child labor incident observations related to supplier location l. For the prototype it is assumed that no more than 10 observations are available for one supplier location. In practical cases, this node can be extended to also allow for more cases. By including the frequency as a discrete node instead of using a constant, changes in the frequency can have a direct effect on probabilities without needing to resample conditional probability tables.

Updating

Netica allows users to include findings (i.e., observations in the case of this thesis) via case files. Each line in a case file represents a new finding where the columns (i.e., credibility and relevance) indicate the values for the different variables. Consequently, for each independent observation a new line in a case file is created and then incorporated into the model. This updates the credibility and relevance variables. Additionally, the frequency is increased by one for each line in the case file.

[191] In the implementation these parameters are configured through Java property files. They may also be stored in the database system.

Breach likelihood

The final node combines the prior, audit, and observational probabilities. It is modeled via a continuous node representing the possible states between 0 and 100 (Figure 89). Moreover, it is assumed to be a combination of the distributions of the parent nodes. Its distribution is modeled via a sampling process.[192] Additionally, the final node incorporates the distinction between supplier locations that have signed the “code of conduct” and (potential) supplier location that have not.[193] For locations that have signed “codes of conducts”, the prior probability’s mean is shifted by a user-defined factor. This factor is modeled via a discrete node with two states representing whether the supplier has signed a code of conduct or not.

Several parameters influence the final probability of a breach. The “signed code of conduct parameter” φ only has an effect if the respective node is set to true, otherwise it is set to 1. The other three parameters influencing the final node’s probability represent a linear combination of the prior likelihood P_{prior}, the observational likelihood P_{obs}, and the audit likelihood P_{audit}. Consequently, the following formula with the three additional parameters (ϕ, χ, ψ) is used:

$$P_{breach} = P_{prior} * \phi * \varphi + P_{obs} * \chi + P_{audit} * \psi;\ \phi + \chi + \psi = 1; 0 < \varphi \leq 1$$

This results in the Netica implementation from Script 4. Values for the weights were extracted from the expert questionnaire depicted in the chapter on evaluation.

```
BP (P_prior, SignedCodeOfConduct, P_obs, P_audit) =
SignedCodeOfConduct == true ?
   (P_prior*0.25*0.238+P_obs*0.333+P_audit*0.429) :
   (P_prior*0.238+P_obs*0.333+P_audit*0.429)
```

Script 4: Netica equation for P_{breach} if $\varphi = 0.25, \phi = 0.238, \chi = 0.333, \psi = 0.429$ (author’s representation)

[192] Netica creates this using a Monte Carlo sampling based on the model equations.
[193] It is assumed that suppliers need to comply with the code of conduct irrespective of a signature. However, not signing it increases the breach risk significantly.

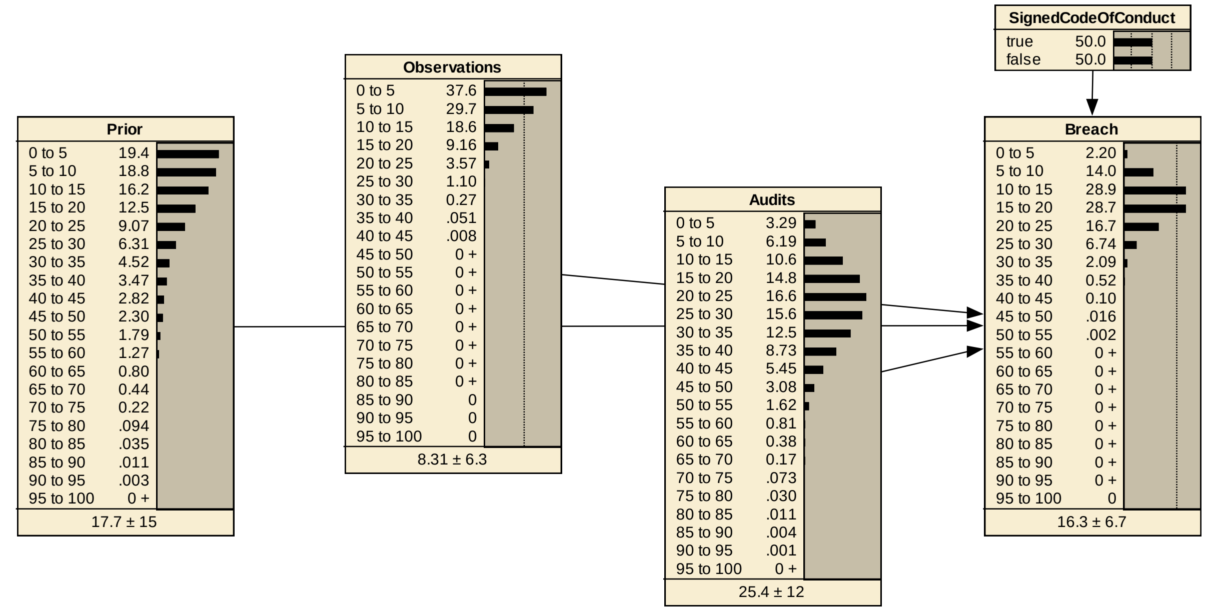


Figure 89: BN nodes directly influencing breach likelihood P_{breach} (author's representation)

Technical text mining implementation

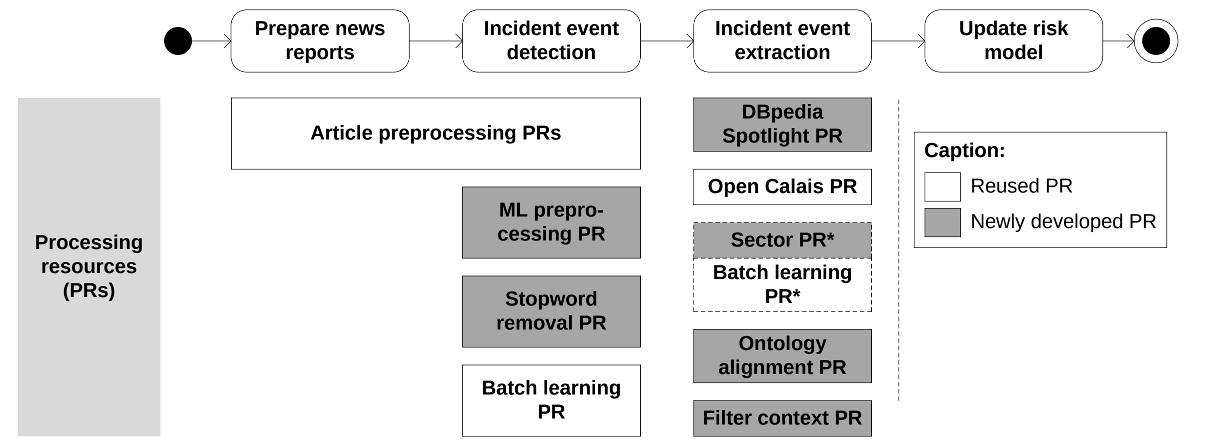
The technical TM application takes a text as input and returns a child labor incident event. For this purpose the text is processed in a number of steps, including annotating parts and applying a pre-trained ML classifier. In the following the words "tag" and "annotation" are used synonymously.

Underlying technical architecture

The TM implementation uses GATE (Cunningham et al. 2013; Cunningham, Bončeva, and Maynard 2011)[194], a tool that has been used in a number of publications and is widely known (e.g. A. Hogenboom et al. 2011). GATE is developed by the University of Sheffield and builds on the earlier TIPSTER architecture introduced by a DARPA research project (Gaizauskas and Wilks 1998; Grishman 1998). GATE allows the user to specify processing pipelines that are executed on a per-document basis. For each document the processing resources (PRs) included in a pipeline of documents are executed one after the other. A PR is a software component with a standardized interface which may be configured through initialization and runtime parameters. A large collection of different PRs is already available for GATE, and they can also be programmed using Java[195]. The following mapping will intro-

[194] GATE 7.1 was used.
[195] https://www.java.com/.

duce the different PRs used in the application, highlighting the ones specifically developed for this thesis (Figure 90). The structure refers to the process depicted in the main text.[196]



* The use of sector PR or batch learning PR during event extraction depends on whether a rule- or ML-based approach is chosen

Figure 90: Overview of processing resources used for TM according to process steps (author's representation)

Most PRs were used for incident event detection and extraction. The subsequent paragraphs explain each block in Figure 9 after describing the data model used for the TM.

GATE uses a data model for TM that centers on documents. Each document may contain several annotation sets which can contain multiple annotations. An annotation covers part of a text (defined by a start and end offset, i.e., the i^{th} character in a document) and has both a name and a feature set containing features. For example, the surface form (i.e., the text characters in a document) might be "Vienna University of Technology," with the corresponding tag "University" with a start offset of 7 and an end offset of 16 if it is the only text in the document. Additional features such as "city=Vienna" or "country=Austria" can provide complementary meta-information in the feature set. Moreover, each document can have a document feature set. Figure 91 provides an overview of these connections as a class diagram. All the information can be serialized together with the document in a file.

[196] To test the different process steps to evaluate the approaches, different combinations of the processing resources were used. The combinations tested are explained in a later chapter. Mímír, an indexing service, was used to retrieve information from the datasets (V. Tablan et al. 2015). Version 4.1.3 was used, building on the Grails 1.3.7 library (https://grails.org/).

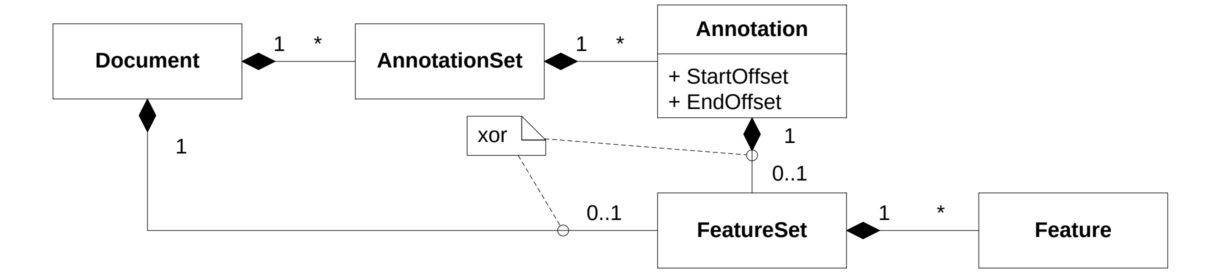
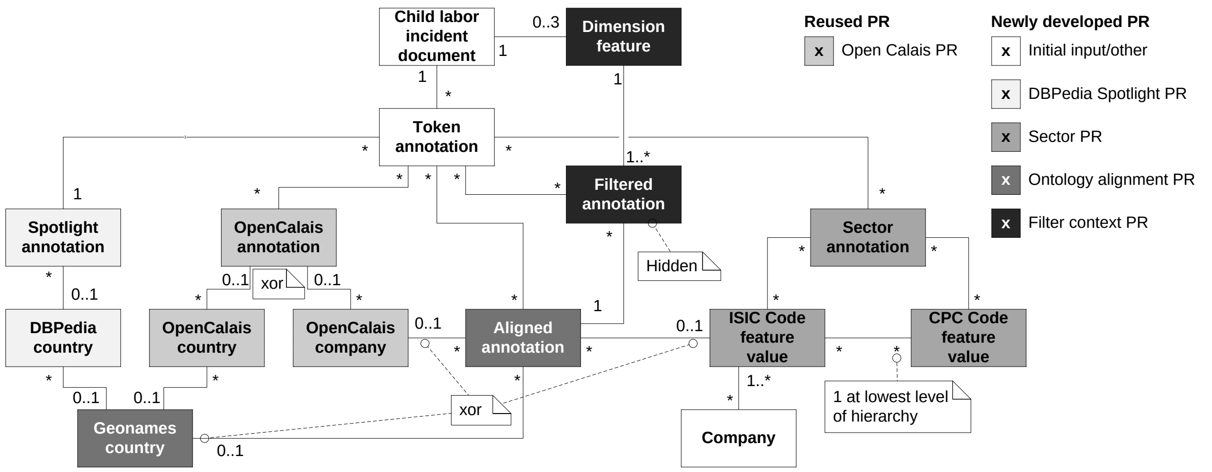


Figure 91: General tagging structure used by GATE (author's representation)

Building on the GATE data model, the different PRs used in this thesis create a set of annotations in multiple layers for each child labor incident document. This results in the domain relationships outlined in Figure 92 when focusing on event extraction.[197] For each domain item the principle corresponding PR is shown in the figure's caption. This overview will be explained in the following subsections.



Note: Intermediary annotation objects such as features or feature sets are partly not shown for better visibility

Figure 92: Conceptual domain model for TM information (author's representation)

Text mining processing resources

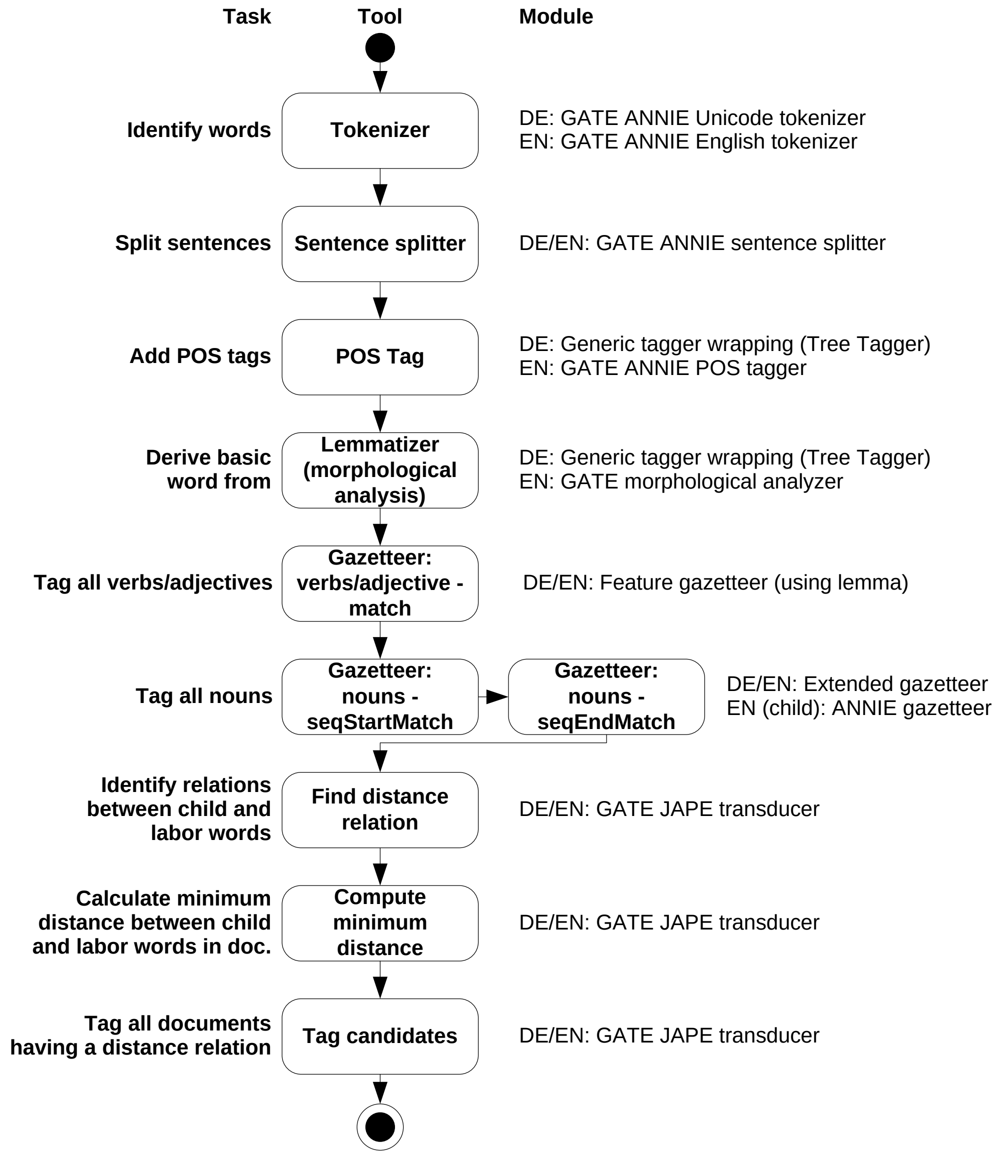
The sections below will detail each processing resource block described in Figure 90. The default parameter configuration is detailed for each PR. If not otherwise stated, these parameters are used by the software or evaluation steps.

Article preprocessing PRs

Article preprocessing involves all steps for processing a natural text before all necessary tokens for the first character distance-based decision have been created. This process is performed via a GATE pipeline using different modules, as shown in Figure 93. While all modules reuse existing GATE modules, new scripts are suggested for the JAPE transducers, a

[197] The relationships for child labor distance relationships along with child and labor words are not detailed here graphically.

configurable pattern-matching component for GATE (Cunningham, Maynard, and Tablan 2000). The components used are the GATE ANNIE pipeline (Cunningham, Hanbury, and Rüger 2010; Cunningham et al. 2002), a feature and an extended gazetteer (Petrak 2013), and Tree Tagger for German POS-tagging/lemmatization (Schmid 1995; Schmid 1994).[198]



Note: DE = German; EN = English

Figure 93: Preprocessing pipeline to derive distance relations (author's representation)

The first four steps in the preprocessing pipeline establish a tagset with separated words and sentences. Words are tagged with token annotations. This includes adding the lemma (i.e., basic form) for each word as a feature. In the following steps, different types of gazetteers are used to tag the words relating to “child” and “labor” in English or German. For verbs and adjectives a feature gazetteer is applied that tags a word if the words inside the gazetteer list are found within the lemma of a token (in any position). For nouns this is used more restrictively:

[198] The TreeTagger software can be found under http://www.cis.uni-muenchen.de/~schmid/tools/TreeTagger/.

the match must be at the beginning or end of a word (the surface form of the word is used[199]). Words are marked with separate tags for “child” and “labor”. These are used in the next step to find the distance relations. Distance relations are found between two adjacent “child” and “labor” annotations. No overlaps are created. Script 11 in “Appendix M: Child labor distance relations and minimum distance” depicts the JAPE script applied if the “child” word comes before the “labor” word.[200] Each distance relation has a feature with the character distance between the tags. The minimum distance JAPE script (Script 12 in “Appendix M: Child labor distance relations and minimum distance”) is then applied to find the distance for the child labor relation with the lowest character distance. It then adds this feature to the document as a separate attribute. Finally, each document receives another document feature with the value “true” if the document contains at least one child labor relation. If stopword removal (in the specific sense of this thesis – see main text) is applied, then this processing step is included before calculating the minimum distance for the document.

Stopword removal PR

Name	Type	Required	Standard	Description
Initialization parameters				
pathToStopwordList	String	x	<Path>	Path to the file already containing or due to contain the stopwords
Runtime parameters				
childLaborFeature-containingAnnoName*	String		DOC	Defines the name of the annotation containing the child labor feature indicating child labor incident documents in the training set
childLaborFeatureName*	String		childLabor	Defines the name of the feature indicating child labor incident documents in the training set
creationMode	Boolean	x	False	Enables or disables the creation mode in which a stopword list is defined
distanceRelationName	String	x	CL_Distance Relation	Name of the child labor distance relation
inputAnnotationSetName	String	x	NE	Name of the annotation set containing all tokens to handle

* Only needed for creating a list of stopwords

Table 97: Parameters for stopword removal PR (author’s representation)

The stopword removal processing resource deletes one word distance relations. These are child labor distance relations that contain tags referring to both “child” and “labor” within the same word (i.e., token). Although this can be performed within one processing step, the resource has been divided into two modes – one for creating a list of stopwords representing words that lead to such situations and one for deleting child labor distance relations formed

[199] Further work could try the lemma. However, this was avoided at this point due to the risk that lemmas are partly not available/found.
[200] The script uses German words for child (Kind) and labor (Arbeit). A second, analogous script is used for the case that the word for “labor” comes before the word for “child”.

over such words. This allows the stopword list to be trained on a training corpus for application elsewhere, and it also allows the user to review the list.[201]

During the creation phase, stopwords are only added to the list if they are found within a document that has not been tagged as a document referring to a child labor incident in the training set. This should reduce the likelihood of this word being a keyword for the set of incidents.[202] While being applied, the stopword removal PR removes distance relations from the annotation set specified in the input parameters. Each surface form is compared with the pre-created list of stopwords.

Machine learning preprocessing PR

Name	Type	Requ.	Standard	Description
Runtime parameters				
allowMultipleAnnotations	Boolean	x	false	Defines whether multiple annotations for the same token are allowed
annotationSetName	String		ML	Name of the annotation set to be used
checkForExportWords*	Boolean	x	false	Defines whether only annotations are entered for words in export list
creationMode	Creation Mode	x	MIN_RELA-TIONLEVEL	Sets creation mode for features
distanceRelationName	String	x	CL_Distance Relation	Name of the distance relation if used
exportWords*	Boolean	x	false	Switches between export word creation and usage mode (true=creation)
outputAnnotationName	String	x	ML_Token	Name of the output annotation (i.e. feature produced for ML)
pathToSentimentFile	URL		-	Path to sentiment resource used
pathToWordsToExportFile*	URL		-	Path to export file with words to be used for checkForExport-Words-mode
pathToWordsToRemoveFile*	URL		-	Path to “words to remove” file to be used for removeWords-Mode
removeWords*	Boolean	x	false	Defines whether words in the remove file should be excluded from created features
sentimentResource	Sentiment Resource	x	NONE	Mode used for sentiments – whether they are created for machine learning or not (dependent on resource used)

* Only relevant for preprocessing steps preceding sector tagging ML

Table 98: Parameters for ML preprocessing PR (author’s representation)[203]

The ML preprocessing PR allows the creation of features used for the ML algorithm based on the approaches discussed above (here “features” refers to the input for the ML and not the attributes of a GATE annotation). This is the implementation of the feature selection step.

[201] On some occasions words with this “stopword feature” are relevant. For example, “Kinderarbeit” in German would match the condition without being an actual stopword. These words can be removed via a manual review of the automatically created stopword list.

[202] The German version had a set for both true and false words and the list was then combined manually. This manual step was excluded for English, given that only one stopword was found. Consequently, no training data bias is associated with the calculation for the English set.

[203] Additional in-class parameters were used for WordNet features and token feature names.

Consequently, the different strategies can be selected using the “creationMode” parameter. The parameter distinguishes seven methods of producing ML features: (1) using all tokens; (2) considering only tokens within child labor relations or (3) in sentences touched by child labor relations; (4,5) the previous two techniques restricted to a random minimum distance relation; and, finally (6,7), the previous two techniques using all minimum distance relations.

The main role of the PR is to select and copy tokens. In an extended step, it can also enrich tokens with additional features used by the ML step. Therefore, this PR is able to add sentiment tags to token features. For this purpose SentiWordNet[204] is implemented which is a language resource that provides positive and negative sentiment clues for words and also includes a Java API to access its resources (Baccianella, Esuli, and Sebastiani 2010; Esuli and Sebastiani 2006). The sentiments are determined based on string-matches.

The PR includes all the tags needed for ML in the annotation set defined in its parameters. For ML, the same annotation set also needs to include an annotation for each document covering all tokens created by this PR, which also includes the class that will be learned during the ML’s learning phase.

The ML preprocessing PR is also used to create the feature set used to learn the classifiers for the ML-based sector tagging approach. For this purpose the PR offers two additional modes. The first method is to define a list of words (via the parameters “removeWords” and “pathToWordsToRemoveFile”) that is removed from the ML feature set during creation. The second method is to export a set of words (through the parameters “exportWords”, “pathToWordsToExportFile”, and “checkForExportWords”) that defines the words to be recognized during application. Both modes were used to train different versions for the sector ML classifier tests.

[204] http://sentiwordnet.isti.cnr.it/.

Batch learning PR

Name	Type	Required	Standard	Description
Initialization parameters				
configFileURL	URL	X	config.xml	Configuration file in the folder containing learned models or models to learn
Runtime parameters				
inputASName	String		ML	Name of the input annotation set
learningMode	RunMode	X	APPLICATION	Mode for the batch learning PR – whether application or learning (with sub types)
outputASName	String		ML	Name of the output annotation set for application mode
runProtocolDir	URL		-	Location of protocol file if required

Table 99: Static initialization parameters for batch learning PR (author's representation)

The ML reuses the GATE batch learning PR (Y. Li, Bontcheva, and Cunningham 2009) which implements several machine learning approaches, including SVM, PAUM, KNN, Naïve Bayes, and C4.5. The batch learning PR has been used in earlier classification tasks (e.g. Y. Li, Bontcheva, and Cunningham 2007; Y. Li, Bontcheva, and Cunningham 2005a).[205] The PR is configured via an XML-file that specifies the parameters used during ML. Multiple configurations were tested, as detailed in the results section. Script 5 shows one example configuration file.

The batch learning PR is also used for the ML sector tagging approach described above. In this case, the learned model is applied as a classifier during information extraction. Nevertheless, general settings are kept similar.

[205] The LibSVM (http://www.csie.ntu.edu.tw/~cjlin/libsvm/) implementation of SVM provided by WEKA (http://www.cs.waikato.ac.nz/ml/weka/) is an important part of the batch learning PR incorporated in GATE. It has previously been used to successfully deal with skewed data (Van Landeghem et al. 2011).

```xml
<?xml version="1.0"?>
<ML-CONFIG>
  <SURROUND value="false"/>
  <FILTERING ratio="0.2" dis="near" />
  <IS-LABEL-UPDATABLE value="true"/>
  <IS-NLPFEATURELIST-UPDATABLE value="true"/>
  <EVALUATION method="kfold" runs="8"/>
  <PARAMETER name="thresholdProbabilityClassification" value="0.1"/>
  <VERBOSITY level="1" />
  <ENGINE nickname="SVM" implementationName="SVMLibSvmJava"
        options=" -s 0 -t 1 -d 3 -c 10 -tau 1 -m 1000 "/>
  <DATASET>
     <INSTANCE-TYPE>DOC</INSTANCE-TYPE>
     <NGRAM>
        <NAME>DOCML1gram</NAME>
        <NUMBER>1</NUMBER>
        <CONSNUM>1</CONSNUM>
        <CONS-1>
            <TYPE>ML_Token</TYPE>
            <FEATURE>root</FEATURE>
        </CONS-1>
     </NGRAM>
     <ATTRIBUTE>
        <NAME>Class</NAME>
        <SEMTYPE>NOMINAL</SEMTYPE>
        <TYPE>DOC</TYPE>
        <FEATURE>childLabor</FEATURE>
        <POSITION>0</POSITION>
        <CLASS/>
     </ATTRIBUTE>
   </DATASET>
</ML-CONFIG>
```

Script 5: Example XML-configuration file for batch learning PR (author's representation)

DBpedia Spotlight PR

Name	Type	Required	Standard	Description
Initialization parameters				
servicePath	URL	x	localhost[206]	URL of DBpedia Spotlight service
Runtime parameters				
annotationSetName	String	x	NE	Name of the annotation set used
Confidence	Double	x	0.1	Confidence parameter for the DBpedia Spotlight service
creationMode	String	x	SENTENCE-LEVEL	Parameter specifying whether text is sent in full or in chunks of sentences to the service
keepExisting	Boolean	x	true	Flag to condition the execution on the existence of tags with the same annotation name
outputAnnotationName	String	x	NELookup_Spotlight	Name of the output annotation generated
Spotter	String	x	Default	Type of spotter used by the DBpedia Spotlight service
Support	Integer	x	10	Support parameter for the DBpedia Spotlight service

Table 100: Static initialization parameters for DBpedia Spotlight PR (author's representation)

A dedicated local server was used to set up DBpedia Spotlight as a service. Two different versions of DBpedia Spotlight have been set up. The first version of Spotlight is based on Apache Lucene (P. N. Mendes et al. 2011), while the second version uses an enhanced statistical core (Daiber et al. 2013). Both approaches use the same processing resource. At the time of initialization, the client is started using the servicePath specified by the user. The client is based on the "Simple web service-based annotation client" published by Mendes and Jakob

[206] Differentiated depending on the type of service used (Lucene-based or statistical).

(2011). Using HTTP GET, it transports the text to the Spotlight service and receives a JSON object containing the result, which is then returned reusing a data object called DBpediaResource. The parameters “confidence”, “spotter,” and “support” are used as input for this process and handed to the service as HTTP-encoded parameters. In order to improve processing speed, the PR can either send the text to the Spotlight service in full or in chunks determined by the sentences identified in the text.

Open Calais PR

Name	Type	Required	Standard	Description
Initialization parameters				
licenseID	String	x	<UserID>	License key to use the Open Calais service
openCalaisURL	URL	x	<Open Calais API>	URL of the Open Calais web service
Runtime parameters				
allowDistribution	Boolean	x	false	Open Calais parameter allowing submission results to be shared
allowSearch	Boolean	x	false	Open Calais parameter making the content submitted searchable
calculateRelevanceScore	Boolean	x	false	Open Calais parameter leading to the inclusion of a relevance score in the tags
docRDFaccessible	Boolean	x	false	Open Calais parameter that allows results to be saved as linked data
enableMetadataType	Metadata Type			Allows the inclusion of social tags or relations
externalID	String			Specifies an ID for the submission for reference
outputASName	String		NE	Name of the output annotation set
submitter	String			Allows a name to be included into the submission

Table 101: Parameters for Open Calais PR (author’s representation)

To tag text using Open Calais, the Tagger_OpenCalais PR provided by GATE was used. No changes were made to the tagging process. The ontology alignment PR was then used to create an aligned tag set.

Rule-based sector PR

Name	Type	Required	Standard	Description
Initialization parameters				
cpcOntologyUrl	URL	X	cpc.rdf	Path to CPC ontology in RDF
isicOntologyUrl	URL	x	isic.rdf	Path to ISIC ontology in RDF
javaWordNetPropertiesUrl	URL	x	prop.xml	Properties for WordNet access
topOntologyNameSpace	String	x	http://at.ifs. tuwien.ac.at	Namespace of the ontology used to establish the aligned tagset (URI name space of tags used as input)
Runtime parameters				
annotationSetName	String	x	NE	Name of the annotation set used
nrOfWordsForMatches	Integer	x	4	Number of words used for comparison in gazetteer
outputAnnotationName	String	x	NELookup_ Sector	Name of the output annotation produced
stopwordPath	URL	x	<path>	Location of file containing stopwords to delete from label keywords
useSpecification	Boolean	x	false	Flag defining whether the context of a keyword is checked for containing words within the specification
wordPartMatches	Boolean	x	false	Flag specifying if word part matches are tested for the keywords

Table 102: Parameters for sector PR (author's representation)[207]

The PR described here is only used if the rule-based sector tagging is applied. The creation of the underlying ontologies, the metadata enrichment, and the tag sets and tagging are all explained below.

Transform to OWL ontology

In order to create OWL ontologies for ISIC and CPC, the Python scripts developed by Stolz et al. (2014) were applied to tabular representations of the ISIC V4 (United Nations Statistics Division 2014) and the CPC V2 (United Nations Statistics Division 2008b; United Nations Statistics Division 2008b) classifications.[208] The scripts create a semantically and logically sound ontology including canonical URIs while retaining the original hierarchical relationships. These are then transformed to RDF and serialized (Stolz et al. 2014). The script has been slightly adapted for ISIC.[209] As the taxonomic concepts are transformed using owl:subClassOf relationships, these were used for further processing steps. Script 10 shows an example section of the CPC ontology.

[207] Additional in-class parameters have been used for classes and constants of more constant ontology elements as well as the path to the stopwords.
[208] The transformation builds on the GenTax approach, initially described in Hepp and de Bruijn (2007). Within this approach, each category is transformed into two ontology classes – one taxonomic reflecting the category and one context-specific (Stolz et al. 2014, 6)
[209] To include cross-references between CPC and ISIC in the final ontologies, the internal format was extended to represent them. A generic top-node was also added to each of the ontologies.

```
<!-- http://at.ifs.tuwien.ac.at/ontologies/pcs2owl/cpc/C_0-tax -->
    <owl:Class rdf:about="&cpc;C_0-tax">
        <rdfs:label xml:lang="en">Agriculture, forestry and fishery products [Taxonomy
Concept: Anything that may be an instance of this category in any context]</rdfs:label>
        <hierarchyCode>0</hierarchyCode>
        <rdfs:comment xml:lang="en">This class subsumes everything that is a member of the
following category of products or services in any relevant context: Agriculture, forestry and
fishery products. It includes both related types of goods (e.g. accessories, supplies, ...)
and items that are no actual goods of this kind but related to the respective category (e.g.
expenses reflecting such goods).</rdfs:comment>
        <rdfs:isDefinedBy rdf:resource="http://at.ifs.tuwien.ac.at/ontologies/pcs2owl/cpc/"/>
    </owl:Class>
```

Script 6: Outline of RDF serialization of CPC ontology (author's representation)

Derive tagset based on metadata

To process the ontologies within the GATE environment, the OWLIM plugin was used to load the RDF files to an internal GATE ontology representation. This allows the ontology to be accessed via an API interface and, hence, enables traversing the ontology's elements. By traversing the elements, all labels are processed in the next step.

During the initialization of the processing resource, keywords, specifying words, and contextual data are loaded into two hashmaps that allow fast data access to the information encapsulated in the ontologies. This is helpful as ontology access methods provided by GATE, in particular OWLIM, rely on broader data structures that encapsulate a wider array of data not needed for the implementation of a gazetteer like the tagging mechanism for sectors. Figure 13 outlines the basic data structure used.

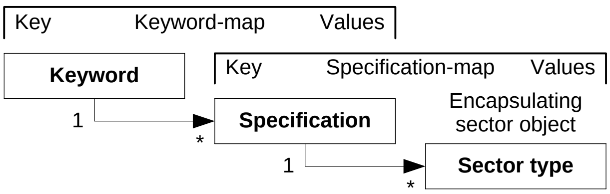


Figure 94: Data structure used for fast access to sector-related data (author's representation)

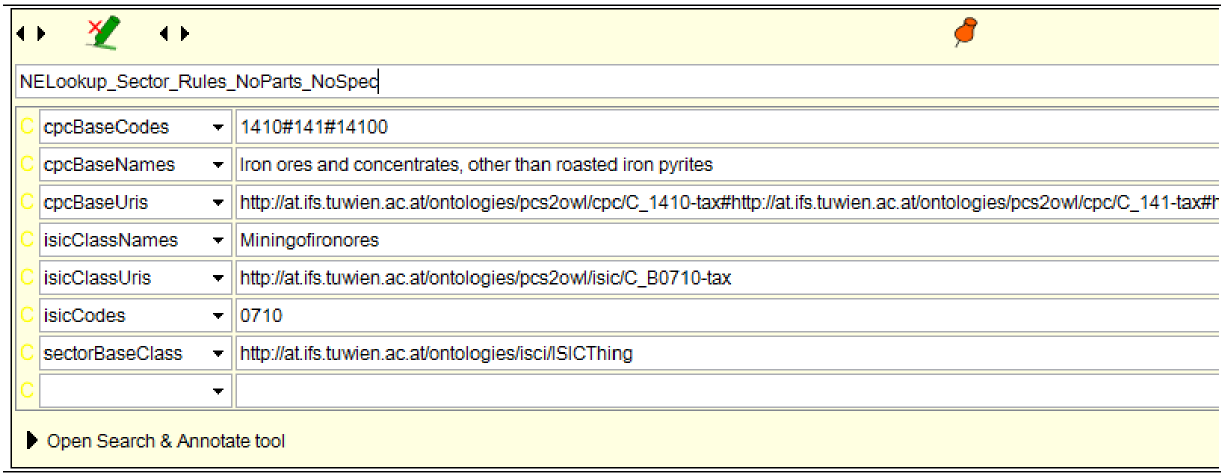
If a key (i.e., a surface form or lemma) is available more than once, then the initial value of the hash key is subsequently increased. Therefore, if a word or lemma corresponds to more than one ISIC code, then all ISIC codes are added to the respective tag. This is also true for the CPC codes.[210] The most important step in loading the ISIC and CPC ontologies is the rule-based processing of sector tags with the function "getGazeteerValuesForLabel". Calling this function leads to the transformation of a single label into the specifications and keywords in the structure described above.[211]

[210] In the implementation the lookup of the ISIC code is sped up via another hashmap.
[211] Lemmas are derived using Wordnet within the Stanford Core NLP framework (Manning et al. 2014; Toutanova and Manning 2000; Toutanova et al. 2003).

Perform tagging

An annotation is added to a text if a match is found by a lookup in the respective hashmap. Direct matches of the surface form are given priority over matches of lemmas. The match strings are checked in ascending order of the number of words in the string.

The tagging is performed in a process of multiple steps and conditionally allows the tagging to be restricted to the occurrence of specifying words in the context of a keyword. If the specification condition is set to true, then the sentences touched by a keyword match are checked for any matching surface form. If the check is positive, a tag is added. If part matches are enabled, match strings for the different possible word parts are created, starting with length one up to the full length of the match string. All matches are added to a combined set (see Figure 95 for an example in GATE).



Note: The image has been cropped on the right for better visibility

Figure 95: Example annotation created via sector PR (author's representation)

Ontology alignment PR

Name	Type	Requ.	Standard	Description
Initialization parameters				
baseOntologyNamespace	String	x	http://at.ifs. tuwien.ac.at/ MappingOwl/	URI of mapping ontology base namespace
importMappingTextFilePath	URL	x	importmappint.txt	Path to file with path mappings for imported ontologies (ISIC and CPC)
mappingOntology	URL	x	mappingOnto.owl	Path to alignment (i.e. mapping) ontology file
topCLIncidentDimClassUri	String	x	http://at.ifs.tuwien.a c.at/Mapping Owl/CLIncident Dimension	URI of root node of dimensions hierarchy in alignment ontology
Runtime parameters				
annotationSetName	String	x	NE	Name of the annotation set used
conflictResoluationPriority	String	x	OpenCalais, Spot-light, SectorTaggingPR	Ordered list of input PRs defining their priority for conflict resolution
geoNamesUser	String		<Geonames Username>	User name for the Open Geonames API service[212]
mapOpenCalais, mapSectorTagging, mapSpotlight	Boolean	x	True	Flag indicating whether to include Open Calais PR, sector PR, and DBpedia Spotlight PR
maxTypesConsidered	Integer	x	2	Limits the number of original class URIs (e.g., provided by Spotlight) used during the alignment pro-cess, i.e., to identify relevant tags.
openCalaisAnnoName	String	x	OpenCalais	Name of the annotation created by the Open Calais PR used as input
outputAnnotationName	String	x	NELookup_Aligned	Name of the output annotation
sectorTaggingAnnoName	String	x	NELookup_ Sector	Name of the annotation created by the Sector Tagging PR used as input
spotlightAnnoName	String	x	NELookup_ Spotlight	Name of the annotation created by the DBpedia Spotlight PR used as input

Table 103: Static initialization parameters for ontology alignment PR (author's representation)[213]

The ontology alignment PR is designed to combine the tags produced by the DBpedia Spotlight PR, the Open Calais PR, and the sector PR to create a combined set of tags without duplicates for each dimension.

When the PR is initialized, the alignment ontology is loaded.[214] The execution creates a cleaned list of tags for each document, only containing relevant tags based on their type. Tags are added to an offset-sorted queue based on the predefined priority specified using the parameter "conflictResoluationPriority". A parser script is used to convert the annotations to an aligned object covering all attributes needed for later processing. Using the alignment ontolo-

[212] http://www.geonames.org/.

[213] Additional in-class parameters have been used for dimension names (i.e., class names in the alignment ontology), namespaces, and the root of the ISIC ontology.

[214] After loading the ontology into a standard GATE OWLIM ontology class, a hashmap is created for faster processing in subsequent steps. This map relates each URI contained in the alignment ontology to the corresponding dimension associated with this URI. This allows for fast lookups without needing to traverse the entire ontology in order to determine the dimension linked to a specific URI.

gy, class-URIs in the input tag are mapped to the corresponding dimension. The maximum number of classes is specified by the parameter “maxTypesConsidered”. Figure 96 shows the class diagram of the aligned tags created for each input tag. Table 104 at the end of this section provides an overview of the mappings performed at the field level.

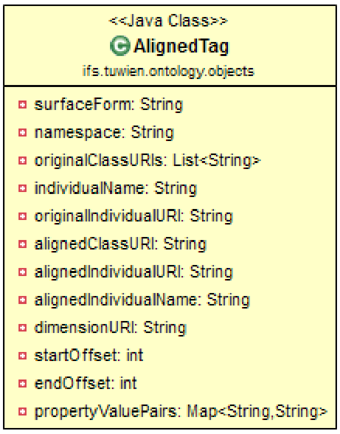


Figure 96: Class diagram for aligned tag (author’s representation; created with ObjectAid)[215]

In a subsequent step, conflicts are resolved via the established ordering of tokens. For each dimension a bit set is used where each bit represents an offset and set bits mark offset areas already covered. If an overlap is found, the tag is not added to the cleaned queue of tokens.

Based on the cleaned, ordered list of tags, the final tags are created by updating the aligned tag classes with additional information. The necessary conversion is shown in Table 104. Three attributes have special importance: the URIs identifying the class and the individual (i.e., the instance of the class) and the name of the individual. Particularly for the geographic dimension this includes the use of LOD and the OpenGeonames Webservice[216]. The end result is the implementation of a multi-step process as depicted in Script 7. The processes for the company and sector dimensions are less complex as the relevant resources to create the aligned tag are either directly provided by the input tag or retrievable in an ontology used. Finally, the annotations are created.

[215] http://www.objectaid.com/.

[216] For this purpose the “Java Client for GeoNames Webservices” is reused which has been downloaded from http://www.geonames.org/source-code/; additionally, for Open Calais the “RESTful Java Client With Apache HttpClient” has been included, available at http://www.mkyong.com/webservices/jax-rs/restful-java-client-with-apache-httpclient/.

```
// Get geonames URI from linked open data
if (tag.getOriginalIndividualURI() != null && tag.getOriginalIndividualURI() != "") {
   // Original individual URI available -> use linked open data
   if (tag.getNamespace().contains("opencalais")) {
      alignedIndivUri = queryLinkedOpenData.getGeoNamesUriForOpenCalaisLocation(
                              tag.getOriginalIndividualURI());
   } else if (tag.getNamespace().contains("dbpedia")) {
      alignedIndivUri = queryLinkedOpenData.getGeoNamesUriForDBPediaLocation(tag.getOriginalIndividualURI());
   }
}
if (tag.getOriginalIndividualURI() == null
      || tag.getOriginalIndividualURI() == ""
      || alignedIndivUri == null) {
   // Original individual URI not available/resolved -> use OpenGeonames API
   if (this.geoNamesUser != null && this.geoNamesUser != "") {
      alignedIndivUri = AlignmentPR.NS_GEONAMES +
            client.getGeoNamesIdForSurfaceForm(tag.getSurfaceForm(), this.geoNamesUser);
   }
}
if (alignedIndivUri != null && alignedIndivUri != "") {
   // Use geonames feature code as class URI
   alignedClassUri = queryLinkedOpenData.getGeoNamesFeatureCodeForGeoNamesUri(alignedIndivUri);
}
if (tag.getIndividualName() != null && tag.getIndividualName() != "") {
   alignedIndivName = tag.getIndividualName();
}
if (alignedIndivUri != null) {
   tag.setAlignedIndividualURI(alignedIndivUri);
}
if (alignedClassUri != null) {
   tag.setAlignedClassURI(alignedClassUri);
}
if (alignedIndivName != null) {
   tag.setAlignedIndividualName(alignedIndivName);
}
```

Script 7: Resolving attributes for the geographic dimension (author's representation)

From…		…To	Aligned tag					
Source	**Feature**	**Cardinality***	**alignedClassURI**	**alignedIndividualURI**	**individualName**	**namespace**	**originalIndividualURI**	**originalClassURI**
		Cardinality	*1*	*1*	*1*	*1*	*1*	*N*
Open Calais Geo-graphy	_type	1	-	-	-	-	-	-
	_typeGroup	1	-	-	-	-	-	-
	_typeReference	1	-	-	-	excl. last element	-	copy
	name	1	-	-	copy	-	-	-
	nationality	1	-	-	-	-	-	-
	resolutions	1	get feature code**	get OpenGeonames URI from original URI**^	-	-	-	-
	surface form	1	-	get OpenGeonames URI from original web service**^	-	-	String transformation (result used in query)	-
Spot-light Geo-graphy	URI	1	get feature code**	get OpenGeonames URI from original URI**^	-	excl. last element	copy (result used in query)	-
	percOfSecond	1	-	-	-	-	-	-
	similarityScore	1	-	-	-	-	-	-
	support	1	-	-	-	-	-	-
	types	N	-	-	-	-	-	expand schema namespace URIs
	surface form	1	-	get OpenGeonames URI from original web ser-vice**^-	-	-	-	-
Open Calais Company	_type	1	-	-	-	-	-	-
	_typeGroup	1	-	-	-	-	-	-
	_typeReference	1	copy	-	-	excl. last element	-	copy
	name	1	-	-	copy	-	-	-
	nationality	1	-	-	-	-	-	-
	resolutions	1	-	-	-	-	-	-
	surface form	1	-	copy or create from name	-	-	String transformation	-
Sector Tagging PR	cpcBaseCodes	N	-	-	-	-	-	-
	cpcBaseNames	N	-	-	-	-	-	-
	cpcBaseURIs	N	-	-	-	-	-	-
	isicClassURIs	N	copy from original class URI per tag	-	-	use original class URI excl. last URI part	-	split from N ele-ments to 1 tag***
	isicClassNames	N	-	-	split from N ele-ments to 1 tag***°	-	-	-
	sectorBaseClass	1	-	-	-	-	-	-
	isicCodes**	N	-	-	-	-	-	-

* Indicates that a single field may contain multiple values; ** Use SPARQL to query LOD; *** Similar number of elements for split applies (number of tokens created = number of ISIC codes in isicCodes field); ^ Do either or for same source (“xor”); ° For the sector, the class name is used as the individual name (could also be implemented with additional field class name if needed)

Table 104: Mapping between tags from different sources in order to create one aligned tag (cardinalities refer to items per feature attribute; author’s representation)

Filter context PR

Name	Type	Requ.	Standard	Description
Runtime parameters				
annotationSetName	String	x	NE	Name of the annotation used to retrieve aligned tokens
closestLimit	Integer	x	0	Character distance limit for “closest” filtering approach (0 indicates unlimited distance)
filterAllCapitalSurfaceForms	Boolean	x	true	Indicates whether aligned tags with only upper case letters are removed
filterMode	Mode	x	<no standard>	Specifies the filtering approach applied
filterToLowestGeography	Boolean	x	true	Shows only geographical attributes at the lowest hierarchical level found
filterToLowestSector	Boolean	x	true	Shows only sectoral attributes at the lowest hierarchical level found
inputAnnotationName	String	x	NELookup_ Aligned	Name of aligned tag used as input
inputChildAnnotationName	String	x	ChildLabor_ Kind	Name of the child annotation
inputDistanceRelationName	String	x	CL_Distance_ Relation	Name of the child labor distance annotation
inputLaborAnnotationName	String	x	ChildLabor_ Arbeit	Name of the labor annotation
maximumDistance	Integer	x	50	Character distance limit for distance-based filtering approaches
minRelationOnly	Boolean	x	true	Indicates whether only minimum distance relation should be taken as context
nameInsteadOfOrgUri	Boolean	x	false	Defines whether the name instead of the URI should be used for the organizational dimension

Table 105: Parameters for filter context PR (author’s representation)[217]

The filter context PR extracts child labor incident events from a set of aligned tags. It begins by loading the aligned tags and any additional tags depending on the mode (e.g., distance relations) while respecting constraints such as minimum distances. The filtering is then applied separately for each dimension. If this is set to filter to the lowest hierarchical level for sector or geography, then all values found, apart from the lowest level, are excluded from the incident event.

Depending on the mode, different functions are used. One function determines which tags are close to other tags within a predefined set of tags. For example, all the closest aligned tags to all child labor distance relations can be determined. This can be done with and without limits (see Script 8). A second function allows finding those annotations that fall within the bounds defined by other annotations (see Script 9).[218] The last function returns the URI with the highest count. The final list of URIs resulting from the application of these functions for each dimension is included as a character-separated list as a document attribute.

[217] Additional in-class parameters have been used for alignment ontology class names and sentence annotation.
[218] The bounds are implemented using bit sets that use 1 to define annotation offsets within the bounds and 0 for annotation offsets outside the bounds.

```
List<String> filteredURIs = new ArrayList<String>();
Map<Integer, List<String>> forwardMap = getDistanceHashMap(alignedAnnos, false, nameInsteadOfUri);
Map<Integer, List<String>> reverseMap = getDistanceHashMap(alignedAnnos, true, nameInsteadOfUri);
int annoMinOffset = -1; int annoMaxOffset = -1; int currentDist = -1;

int limit = document.getContent().toString().length();
if (this.closestLimit > 0) {
   if (this.closestLimit < limit) {
      limit = this.closestLimit;
   }
}

// Increase distance until first element is found
for (currentDist = 1; (currentDist < document.getContent().toString().length() && currentDist <
      limit); currentDist++) {
   // Check for every annotation in set
   for (Annotation annotation : baseAnnotations) {
      annoMinOffset = Integer.valueOf(annotation.getStartNode().getOffset().toString());
      annoMaxOffset = Integer.valueOf(annotation.getEndNode().getOffset().toString());
      // After annotation
      if ((annoMaxOffset + currentDist) < document.getContent().toString().length()) {
         if (forwardMap.get(annoMaxOffset + currentDist) != null) {
            filteredURIs.addAll(forwardMap.get(annoMaxOffset + currentDist));
         }
      }
      // Before annotation
      if ((annoMinOffset - currentDist) > 0) {
         if (reverseMap.get(annoMinOffset - currentDist) != null) {
            filteredURIs.addAll(reverseMap.get(annoMinOffset - currentDist));
         }
      }
   }
   // Stop if at least one element found
   if (filteredURIs.size() > 0) {
      return filteredURIs;
   } else {
      Return null;
   }
}
```

Note: The function getDistanceHashMap() returns a hashmap that with annotations at respective offset positions (end offset if reversed, otherwise start offset)

Script 8: Finding closest annotations for a set of annotations (author's representation)[219]

```
List<String> filteredURIs = new ArrayList<String>();
int nextSetBit = -1;
for (AlignedTag tag : alignedAnnos) {
   // tag.print();
   nextSetBit = bounds.nextSetBit(tag.getStartOffset());
   if (nextSetBit > -1 && nextSetBit < tag.getEndOffset()) {
      if (!nameInsteadOfUri) {
         if (!filteredURIs.contains(tag.getAlignedRelevantURI())) {
            filteredURIs.add(tag.getAlignedRelevantURI());
         }
      } else {
         if (!filteredURIs.contains(tag.getAlignedRelevantName())) {
            filteredURIs.add(tag.getAlignedRelevantName());
         }
      }
   }
}
if (filteredURIs.size() > 0) {
   return filteredURIs;
} else {
   return null;
}
```

Note: Set bits indicate offset positions relevant to analysis

Script 9: Finding annotations within bounds defined by a bit set on annotation offsets (author's representation)

[219] The getDistance hashmap function used in the script returns a hashmap with the distance as the key and a list of tags with the start offset as this value (for the reversed map the distance is calculated starting at the document's end).

Summary

Altogether, this chapter provided details on the implementation of the risk and TM methodology. Several further details can be found in the rest of the appendix.[220]

[220] Particularly “Appendix O: Class diagrams of prototype” gives technical overviews with UML diagrams.

General appendices

Appendix A: IT for sustainable supply chain management[221]

IT has been applied to a large amount of domains. Business intelligence IT-based tools have been applied in a wide array of fields in general risk management (D. D. Wu, Chen, and Olson 2014; G. Lee and Kulkarni 2011). Their connection has also been underlined for general sustainability management (Petrini and Pozzebon 2009b; Petrini and Pozzebon 2009b). In supply chains, IT may help in risk management in a variety of areas, for example in risk analysis (Harrington and O'Connor 2009; Lingzhe Liu, Daniels, and Hofman 2014; D. D. Wu, Chen, and Olson 2014).[222] Nevertheless, the amount of related work that focuses on IT for sustainability in supply chains is limited while IT's usefulness has been stressed (Hack and Berg 2014). Consequently, this section will focus on a survey on IT for sustainable supply chains.

Sustainability and IT are interlinked in two ways: Either IT itself can become more sustainable or sustainability can be improved by applying IT to increase efficiency (Schatten 2009). This duality has been observed and supported by various researchers. For example, Piotrowicz and Cuthbertson (2009) stress the necessity to also include the effects on sustainability into the evaluation of IT systems. In addition, Melville (2010) acknowledges the potential of IT in improving environmental sustainability and Srivastava (2007) highlights that IT will play an important role in the future of "green" (i.e. environmentally friendly) SCM. Nevertheless, the current state of research in the domain of IT for the broad area of SSCM is still quite diffuse. Up to now a consolidated survey of academic approaches of how IT is seen to drive sustainability in supply chains appears to be missing (see related work section below).

Previous works have focused on the intersections of any two of the domains of sustainability, IT and SCM. Three review attempts tackle all three areas. One of them addresses green logistics specifically (Frehe and Teuteberg 2014), while another attempt focuses on radio frequency identification (Thoroe, Melski, and Schumann 2009). Kurnia, Mahbubur, and Gloet (2012) present the closest related article to the review presented in this section. However, they aim at

[221] This section reflects parts of a to-be-published paper written together with A Min Tjoa.
[222] Cisco implemented an improved supply chain risk management system that enables the company to detect risks upfront (Harrington and O'Connor 2009).

developing the roles that IT (or information systems) may play in SSCM and lesser at studies where IT is presented and evaluated for this purpose (i.e. also more strongly drawing from the IT/information systems literature base). Moreover, they only cover papers until 2011.[223]

All in all, given the limitations of the current work, this survey addresses the intersection of IT and SSCM with a structured literature survey including both, environmental and social sustainability, and presents a more strongly technology-driven perspective of the topic. In light of this thesis, this allows for a more thorough analysis of potentially existing work in the domain of IT for SCRM or related ideas. Additionally, this survey also captures the related work addressing the environmental domain to allow for a comparison with this, more advanced field.

Context of review and keywords used

In general, this paper focuses on peer-reviewed literature that covers how IT may be used within SCM to improve either environmental or social[224] sustainability in supply chains. The basic questions are: (1) How can IT be used to positively affect supply chain sustainability? (2) How does IT currently affect the sustainability of supply chains in practice? Articles including these discussions are the basis to then develop suggestions for further research. This particularly excludes papers where IT is only used as a tool and the use of IT itself is not addressed in the research (e.g. papers that only use software packages to enable a calculation such as MATLAB). The sustainability dimensions are treated similarly: As stated, only academic papers in which either environmental or social sustainability are affected by IT are taken into account. Moreover, papers that focus solely on sustainability aspects of the IT domain are not included (e.g. the life-cycle assessment of a laptop).

[223] Several further reviews have been published either combining IT with sustainability or with SCM have been published (see “Appendix B: Overview of related literature reviews on IT for SSCM” for details – for reference the articles are indexed by their categories (a) or (b) and a consecutive number). At the intersection of IT and SCM (a), three studies review the wider domain by addressing the use and effects of IT for SCM and logistics (1,3,5) while other papers address narrower domains such as secure collaboration or RFID. However, no in-depth connection to sustainability issues can be identified in these studies. Eight review papers, published between 2008 and 2014, deal with the interaction between sustainability and IT (b). Elliot and Binney (2008) present a first article giving a brief overview. Melville (2010) stresses information system innovation to develop a conceptual framework. While Brooks, Wang, and Sarker (2012) and Tushi, Sedera, and Recker (2014) address “green” IT in general, limit Harmon and Moolenkamp (2012) their research to IT services. Loeser (2013) in his review combines Green IT with Green IS (information systems). A broader view is presented by Dao, Langella, and Carbo (2011) who take a resource-based view and discuss the integration of IT with human resources and SCM to develop sustainability capabilities. Nevertheless, their focus is on resources and SCM is a subcategory of the analysis and treated equally as human resources management. Finally, Jenking, Webster, and McShain (2011) present a multi-level research framework with their review.

[224] Social sustainability may consist of a large set of different parameters including factors such as health, wages, justice for all, or the creation of social capital (Mani, Agrawal, and Sharma 2014).

To identify SCM papers, the high-level keywords of *supply chain* and *SCM* are used. Further, the keyword *logistic** is included, as logistical activities are particularly key in SCM. Asterisks ("*") represent wildcards that allow for the automatic search of different but similar words.

The IT perspective in SCM is closely linked to the information flow in supply chains. Hence, IT is defined as "the study or use of systems (especially computers and telecommunications) for storing, retrieving and sending information" (Oxford Dictionary 2012). Therefore, *information technology* is used as another keyword. For better coverage, additional "information" keywords are added: *information system*, *ICT* (information and communication technology), *information management system* and *software*. To include "common" technologies relevant in the context of SCM and expected to be used in abstracts, the set of keywords is extended by *Internet*, *Web service*, *e-technology*, *e-communication*, *e-service*, *data exchange* and *data hub* – these are all keywords that are very much linked to the online world. Specific technologies have been omitted assuming that relevant papers also include one of the more general terms in the metadata fields analyzed.

While research in the past has focused on economic sustainability, this thesis addresses the two other dimensions of environmental and social sustainability. Environmental impacts can include carbon emissions, noise or waste, whereas social impacts cover improvements for the individual or society. The general term *sustainab** and dimension-specific terms such as *green**, *carbon*, *environment**, *ecolog**, *social* and *ethic** are also used as keywords for the literature research.[225]

Research methodology

According to the Association for Computing Machinery (ACM) a survey paper is "[A] paper that summarizes and organizes recent research results in a novel way that integrates and adds understanding to work in the field. A survey article assumes a general knowledge of the area; it emphasizes the classification of the existing literature, developing a perspective on the area, and evaluating trends." (The Association for Computing Machinery 2015). Hence, this paper first retrieves current work in a structured way before classifying and discussing it. For the primary discussion, a content analysis as introduced by Mayring (2010) has been applied. Content analysis builds on two layers: "The first level analyses the manifest content of texts and documents by statistical methods. On a second level, latent content of the text and docu-

[225] To generate the search strings, the keywords for SCM, IT and sustainability are combined separately using "OR". In a second step, the three resulting strings are connected using "AND".

ments is excavated requiring interpretation of the underlying meaning of terms and arguments." (Seuring and Gold 2012, 546). Seuring and Müller (2008b) as well as Seuring and Gold (2012), originally defined by Mayring (2010), outline a four-step process to follow when conducting a content analysis. Its general, structured approach allows deriving reliable and, hence, reproducible results for literature-based analyses, particularly also in the SCM domain (Seuring and Gold 2012).

Adapted for this survey, first research articles were collected in a structured way. Second, this was followed by a descriptive, formal analysis of the material and its general metadata. The third step delivered a (theory-led) selection and definition of structural dimensions and categories for categorisation. Initial categories for a structured discussion were defined deductively and later adapted based on inductive ideas (Mayring 2010). The last step uses these categories to analyse the papers and document the results. Thus, steps 3 and 4 were performed iteratively multiple times to ensure a balanced approach and to reduce the risk of miscategorisation.

Delimitations and the search for literature

The scientific publications of this survey originate from a variety of different origins limited to peer-reviewed journals in English. Thompson Reuters SCI and SSCI indexes[226] served as the main sources for the survey and extended with other established academic sources: EBSCO Business Sources Premier,[227] T&I ProQuest[228] and SciVerse ScienceDirect.[229] Restricting to several domains of interest (e.g. computer or decision science) allowed to minimise the number of false positives. The literature search used the data fields of title, abstract and keywords (if available; including non-author keywords). Relevant papers can also cover the domain of interest only in parts.

The authors analysed all retrieved papers in a first cross-reading of abstracts to identify mismatches. These can occur, for example, due to popular constructions in other contexts such as "logistic regression". In total, 55 papers belong to the final selection of the literature survey.[230]

[226] http://apps.webofknowledge.com/.
[227] http://web.ebscohost.com/.
[228] http://search.proquest.com/.
[229] http://www.sciencedirect.com/.
[230] E-commerce is one topic linked to IT and SCM that stands out in the discussion. E-commerce relates to the end user and to many intermediate steps in the supply chain. Nevertheless, this investigation addresses how IT affects sustainability in a given supply chain and not how IT requires a fundamentally different supply chain due to changes in the consumption behaviour of end customers. Consequently, the authors excluded papers without a focus on e-commerce as a tool in the interorganisational B2B parts of the supply chain.

Content analysis

Figure 97 presents the perspectives structuring this paper's analysis of literature on IT for SSCM.

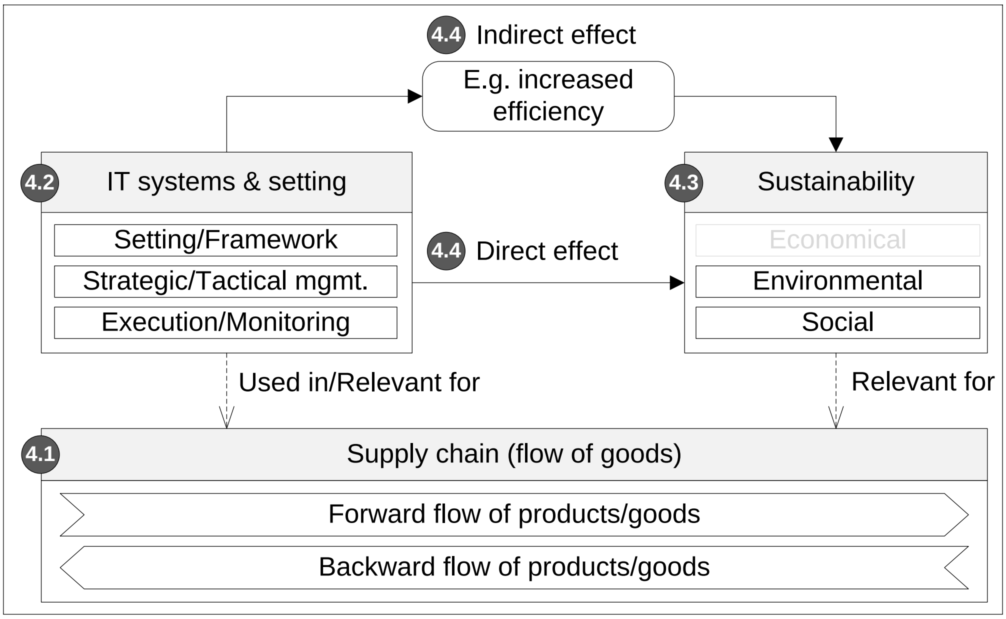


Figure 97: Overview of the classification dimensions (letters represent the references to the respective chapters; author's representation)

The first dimension of categorization (A) follows classical supply chain discussions. A supply chain can typically be split into a forward and a backward (often also called "reverse") flow of products and goods that are separated by consumption of customers (García-Rodríguez, Castilla-Gutiérrez, and Bustos-Flores 2013). Occasionally neither the forward nor the backward network is the focus of interest within SCM. Thus, some papers take a product-centric view, whereas others present an overarching viewpoint. The second dimension (B) distinguishes three different IT perspectives: Systems for strategic or tactical management, systems for supply chain execution and monitoring and, at an overarching level, the supporting frameworks or general setting (e.g. policies) for IT within SCM. As a third perspective (C), we differentiate the impact of IT along the environmental and social dimensions of sustainability. Finally (D), we distinguish two types of effects on the relationships of IT with sustainability. This relationship can be either direct (when the use of IT aims directly at a change in at least one dimension of sustainability explicitly) or indirect (when the use of IT influences an intermediary variable such as efficiency that causes improvements in sustainability and sustainability is not the primary focus). A detailed discussion of all dimensions based on the key papers in each section is provided below.

Descriptive analysis

The analysis includes 55 available papers in total. The first of these papers were published in 1999 with few new papers being published until 2008. However, there was a significant increase in the number of publications in the sample in 2009 that lasted until 2012 (according to the final publication date). A spike may be noticed in 2012 with 15 papers. This number declined to seven papers in 2013. In 2014, until May, two papers were available.

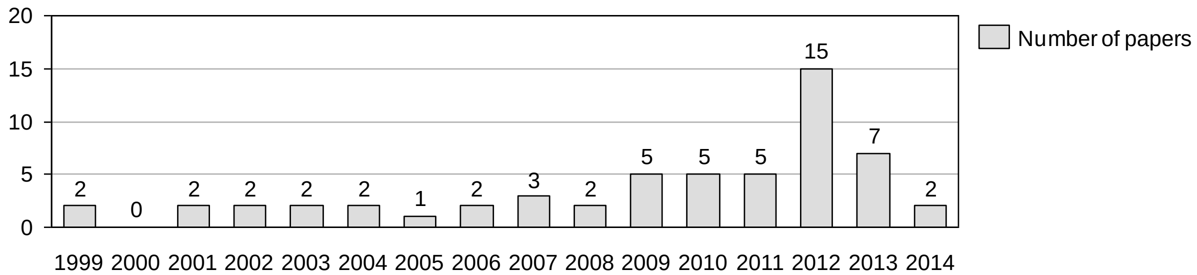


Figure 98: Time-distribution of papers in scope (n=55; author's representation)

As depicted in Figure 99, most identified papers (33 out of 55) have a theoretical or conceptual primary focus. All other categories, except for case studies, show similar low publication numbers with 4 to 6 peer-reviewed articles each. Some of the theoretical and conceptual papers also include some type of software implementation.

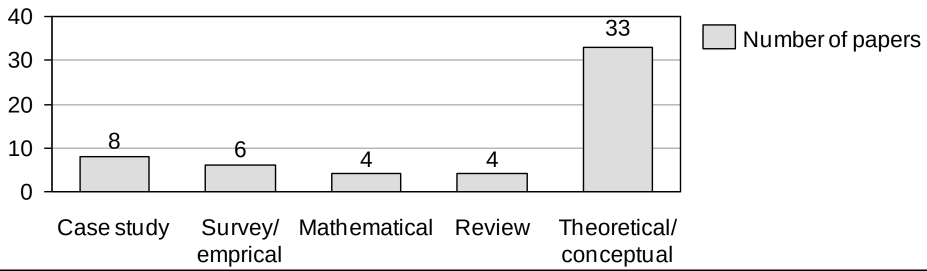


Figure 99: Methodological focus of papers in scope (n=55; author's representation)

Conceptual analysis

The following sections discuss the significant research ideas and streams. A full classification can be found in "Appendix C: Overview of classification of papers".

Focus on the supply chain

Most papers discussing IT in SSCM deal with the forward flow, take a product perspective or have an overarching viewpoint. Only some papers address IT in the context of the backward/reverse flow. Here, the reverse flow specifically addresses the questions of reuse, recycling and end-of-life, while transportation activities are discussed in the forward flow. In any of these contexts IT especially touches the flow of information in a supply chain next to the physical way of goods.

Forward flow

Literature in the context of the forward flow mainly focuses on transportation and the calculation of its environmental impacts. IT planning tools that allow calculation of carbon emissions (see Sourirajan et al. 2009 as an example) have already led multiple companies to enlarge their warehouses and shorten their travel distances (Schiller 2012). Other suggested software tools take an even broader range of environmental effects of transportation into consideration (Guenther and Farkavcová 2010). In an effort to improve precision, Iacob et al. (2013) propose a reference-architecture for a transport carbon calculation and management system based on true fuel consumption. At least until 2008 no attention was given to leverage IT for transport mode choice (Abukhader 2008).

Another aspect arising from the surveyed literature is that IT can improve sustainability at a vehicle-routing level. In 1999 Hasle presented a tool to improve vehicle-routing efficiency that explicitly considers the environmental impact (Hasle 1999). In a different approach, Suh, Smith, and Linhoff (2012) focus on the last-mile delivery problem and how the current standard system of door-to-door delivery can be improved using a social network. A comparable influence on routing is taken by intelligent transportation systems that enable truck drivers to bypass potential bottlenecks (Marett, Otondo, and Taylor 2013).

Some papers address how IT improves the sustainability of nodes in a supply chain network and where nodes such as production facilities are located. Verma and Chaudhuri (2009) show how the complete Indian agricultural value chain has been changed by IT. They highlight that the largest effects on sustainability are at the production and consolidation levels due to improved and directed information, while transportation is least affected. Other authors discuss comparable issues in India's agricultural sector (Rao 2007; Ali and Kumar 2011). Still, significant sustainability impact may be achieved by choosing optimal facility locations. This theme is supported by several authors (Ayoub et al. (2006) for production or storage facilities, Guyon et al. (2012) for city logistics platform and by Bosona et al. (2013) for warehouses).

Reverse flow

Supply chains show close links to sustainability if they cover reuse and recycling activities. Some IT approaches specifically aim at these functions driving their efficiency. These can start with light-weight web-based systems for information exchange (J. W. Edwards, Lyons, and Kehoe 2004) and be particularly important in more complex reverse logistics scenarios (Dhanda and Hill 2005). García-Rodríguez, Castilla-Gutiérrez, and Bustos-Flores (2013) show that in less developed countries (management) information systems can have a significant impact on the implementation of a reverse logistics system for raw materials (García-Rodríguez, Castilla-Gutiérrez, and Bustos-Flores 2013). Performance improvement of the reverse flow may unfold in measuring and assessing reverse logistics chains using IT. For this purpose Trappey, Trappey, and Wu (2010) build on radio frequency identification technology (RFID) linked with fuzzy cognitive maps and Olugu and Wong (2012) present a more general IT system.

Product

Current product-related work can be divided into two streams. On the one hand, suggestions for IT-driven sustainability ensure that products (M.-K. Chen, Tai, and Hung 2012) or suppliers meet environmental/ecological standards (Lai, Hsu, and Chen 2012). For this purpose Taghaboni-Dutta, Trappey, and Trappey (2010) propose a green-parts information platform that covers relevant sustainability information for the product design process.

On the other hand, IT systems can help improve the shortcomings in calculating product-specific sustainability impacts. McKinnon (2010) accused current product-specific carbon labeling of being too inaccurate and, by using existing life-cycle assessment or RFID data, highlights the possibly enabling role of IT. This is similar to what Björk et al. (2011) developed for the specific case of a wood supply chain. As the level of detailed life-cycle information might be too low internally in a company, Barrett, Strunjaš-Yoshikawa, and Bell (2007) present a suitable object-oriented data model to exchange product-specific life-cycle assessment data between organizations.

Overarching supply chain perspective

IT systems that tackle sustainability on an overarching supply chain level have been purposed in multiple areas, either topic or system driven. Bullinger, Steinaecker, and Weller (1999) describe an early IT system for environmental production management that includes life-cycle information, while building on overarching data models. Current papers suggest ontologies for the exchange of product-specific sustainability data (Muñoz et al. 2013; Borsato 2014) or aim for covering the whole supply chain or life-cycle in multiple industries (Koh et al. 2013).

Industry-specific papers deal with multiple areas and do not show specific spikes. Trienekens et al. (2012), for example, focus on the food supply chain. They describe how an information system focused on information exchange along the value chain could satisfy a broad array of information needs from different stakeholders. Lehmann, Reiche, and Schiefer (2012) highlight that an increased use of IT at all stages of the agricultural supply chain can improve efficiency, as well as environmental and social performance. They postulate that the sustainability information (e.g. use of pesticides) needed by consumers is very similar to that which is needed by enterprises along the value chain. In addition, Sigala (2014) discusses the potential use and impact of social media on sustainability in tourism supply chains and Toh, Nagel, and Oakden (2009) deal with cities. Overarching viewpoints may also be driven by international institutions such as the European Union (EU). Research within the EU-funded project "SuperGreen" specifically discusses IT technologies that enable environmental sustainability in freight transport corridors in Europe (Clausen, Geiger, and Behmer 2012; Fozza and Recagno 2012).

IT focus topic

IT systems may be split into systems for strategic and tactical management purposes and systems for execution and monitoring in a supply chain. Furthermore, the environment or framework that supports the use of IT systems on a meta level presents a third perspective.

Systems for strategic and tactical management

Sustainability factors usually affect route and network planning directly. Sourirajan et al. (2009) give details on a carbon-analyzer tool that facilitates the calculation of transport carbon emissions in a network. Besides, planning systems often make use of geographical information systems (GIS) that could help improve sustainability of production activities (e.g. in agriculture see Rao 2007 or for warehouse locations see Bosona et al. 2013) and of transportation links. For example, Ayoub et al. (2006) utilize a GIS as the primary tool for mapping locations along the value chain, which allows expert users to consider environmental and social factors when optimizing the overall network. Additionally, several expert systems have been suggested to support in environmental sustainability related situations. Within the papers under consideration, Vannieuwenhuyse, Gelders, and Pintelon (2003) developed an interactive online tool to decide on modes in a transportation network. Iakovou (2001) proposed a decision support system for risks in crude oil maritime transport operations. Chen, Tai, and Hung (2012) promote an expert system that allows optimum component selection in "green" supply chain settings, especially given hazardous material regulations. The system of Koh et al. (2013) allows collaborative identification of carbon emission hot spots, as well as options for intervention (e.g. estimated impact).

Some authors specifically focus on e-commerce. E-commerce typically enables SCM activities before the product traverses the supply chain (Davies, Mason, and Lalwani 2007). In the B2B context within supply chains e-commerce can have three sustainability effects (Abukhader and Jönson 2004): It can (1) enable smoother operations, thus reducing, for example, energy and material consumption, (2) increase in energy-consuming hardware equipment, and (3) change in consumption behavior. The total effects are yet unclear (Abukhader 2008). Even the sometimes attributed effects of increased efficiency in processes are still under discussion (Abukhader and Jönson 2004; Dotoli et al. 2006). Nevertheless, Walker and Brammer (2012) investigate e-procurement as an option to improve sustainability, while Lee (2003) mentions e-manufacturing in connection with the greening of processes.

In contrast, life-cycle management systems specifically help manage and minimize the environmental effects of goods and services sourced. Thurston and Eckelman (2011) postulate that feasible software tools are already publicly available and supporting proposals such as a data model for the exchange of life-cycle assessment data between organisations have been made (Barrett, Strunjaš-Yoshikawa, and H Bell 2007). Data exchange in supply chains may also support designing sustainable products. With their integrated data platform, Taghaboni-Dutta, Trappey and Trappey (2010) try to show how part-specific environmental information can improve sustainability through facilitating the selection of “green” parts for product development. In contrast, through the use of social media Sigala (2014) considers the potential of involving customers in sustainable supply chain decision making.

Systems for execution and monitoring

IT systems can also be used for the operational execution of supply chains and for monitoring activities, especially those activities that enable decisions. An example can be drawn from the agricultural sector in India where farmers’ input is used to monitor production activities (Rao 2007). In Europe information exchange is regarded as one of the major drivers for more sustainable European transport systems, which is based on IT-supported single transport documents and (intermodal) route planning with tracking and tracing as options (Zografos, Sedlacek, and Bozuwa 2012). Carlson et al. (2001) propose, together with an integrative data perspective, the exchange of environmentally relevant data between supply chain partners. Meacham et al. (2013) even underline that real-time exchange of data in a supply chain is a prerequisite for an environmental management system.

Information needs with regard to sustainability can also come from the public, as Trienekens et al. (2012) highlight. This has led them to promote a common software infrastructure based on defined standards. Edwards, Lyons, and Kehoe (2004) propose using Web technology to implement cost effective planning systems. Moreover, multiple RFID systems have been discussed in the context of SSCM. Björk et al. (2011) use RFID tags to show how tracking and tracing along the whole wood supply chain can be a step towards improved environmental performance (see also McKinnon 2010). Furthermore, Trappey, Trappey, and Wu (2010) show how an IT system that integrates RFID can be used to improve the performance of reverse logistics and thus increase sustainability.

Moreover, multiple RFID systems have been discussed in the context of SSCM enabling further improvements through their tracking and tracing options (see also McKinnon 2010). These systems allow the monitoring of the sustainability performance (Björk et al. 2011 for the whole wood supply chain) or allow sustainability-relevant performance improvements as for example in reverse logistics (Trappey, Trappey, and Wu 2010) or waste management (Martinez-Sala et al. 2009). Even customer-health may be improved when tracking livestock in food supply chains (Wognum et al. 2011). Similarly, sensor networks embedded in an "Internet of Things" may secure food supply chains in cities (Q. Zhang et al. 2013). Nevertheless, the latter has not been stressed in the other papers.

A further category encompasses intelligent transport systems (ITS) that use technology to improve transportation and, thus, potentially sustainability (Zografos, Sedlacek, and Bozuwa 2012). This is not only true for the systems commonly known in road transport (Iacob et al. 2013; Marett, Otondo, and Taylor 2013), but also for new systems; for example in rail infrastructure, like the European Rail Transport Management System (ERTMS). Benz et al. (2012) present several applications and a test bed to optimize network-level greenhouse gas emissions in a general logistics network. Finally, expert systems can be used to analyse the operational performance of supply chains (Olugu and Wong 2012) or for ad-hoc improvements. Suh, Smith, and Linhoff (2012) show how a social network in combination with location-based information of participants can be used to improve last-mile package delivery in real time.

General framework and setting

Not only IT systems themselves but also the general framework and setting (e.g. policies) contribute to how IT affects sustainability in supply chains. Zografos, Sedlacek, and Bozuwa (2012) analyse European policies and compare them with the current status of policy implementation and direction in the member states. They highlight that, next to other factors, improvement in IT can enhance Europe's position on sustainability. Additionally, within the

EU-funded project “SuperGreen”, Clausen, Geiger, and Behmer (2012) develop general recommendations (application and infrastructure) concerning IT technologies which can help to enable ecological sustainability along multiple freight corridors within the EU. Moreover, Toh, Nagel, and Oakden (2009) underline the importance of IT as an infrastructure that fosters (sustainable) logistics by enabling shifts in freight movement within cities and, therefore, increases environmental sustainability.

Focus-dimension of sustainable development

As discussed earlier, this thesis focuses on the environmental and social dimensions of SSCM. Figure 100 presents the distribution of the papers in the survey over the dimensions. The results are comparable to Seuring and Müller (2008b) who also identified a strong focus on environmental sustainability. Moreover, not only greenhouse gases (GHGs), but broader environmental effects are also touched on by 26 out of the 55 papers. Work on the social dimension is very limited with only one paper addressing the dimension specifically and 16 papers addressing it next to environmental concerns.

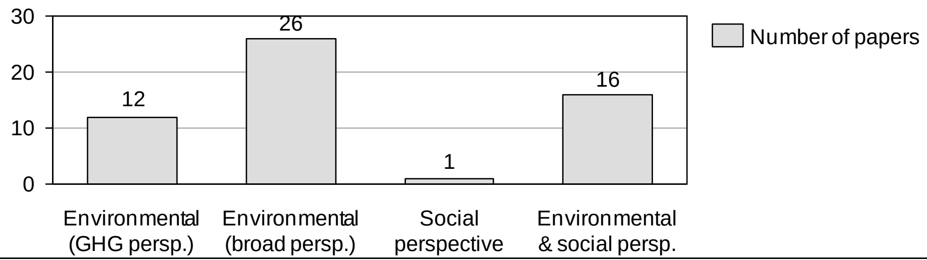


Figure 100: Environmental and social sustainability focus in the papers analyzed (n=55; author’s representation)

Environmental focus

Within research on IT for environmental sustainability in supply chains two major substreams can be identified. One stream focuses on the elimination and reduction of GHG emissions and is strongly driven by recent regulations (e.g. European Commission 2012). The other stream also takes a multitude of other environmental concerns into account, such as noise or particle pollution. Here, different regulations such as restriction of the use of certain hazardous substances (RoHS) in electrical and electronic equipment or registration, evaluation, authorization and restriction of chemical substances (REACH) are the drivers. IT systems supporting the optimization of environmental impacts strongly focus on GHG emissions (Sourirajan et al. 2009; Schiller 2012; Guenther and Farkavcová 2010). In particular, some authors take a look at carbon emissions of road freight traffic (Benz et al. 2012; Suh, Smith, and Linhoff 2012; Iacob et al. 2013; Bosona et al. 2013). Also, Shi et al. (2012) attribute their efforts of improving reverse logistics information flow to establishing a low carbon environment.

Nevertheless, several authors also considered a wider perspective by including the analysis of congestion (Vannieuwenhuyse, Gelders, and Pintelon 2003), particle emissions (Clausen, Geiger, and Behmer 2012) or a multitude of environmental factors (Carlson et al. 2001 as well as the domain ontologies of Muñoz et al. 2013 and Borsato 2014). Reverse SCM also covers broader effects when recycling and end-of-life management of products is considered and promoted; thus, reducing waste, for example (Olugu and Wong 2012; García-Rodríguez, Castilla-Gutiérrez, and Bustos-Flores 2013; Martinez-Sala et al. 2009; Trappey, Trappey, and Wu 2010). As early as 1999, Bullinger, Steinaecker, and Weller already considered material reuse and other environmental factors in their product life cycle model (Bullinger, Steinaecker, and Weller 1999). IT systems supporting the verification of standards have an environmental scope as wide as the standards they verify. As standards usually take an approach beyond GHG, a wider environmental perspective is ensured (see, for example, the systems described by Lai, Hsu, and Chen 2012 or Chen, Tai, and Hung 2012).

Social focus

Only one paper specifically addresses social issues in sustainability even though social sustainability is a broad concept. Ali and Kumar (2011) focus on agricultural efficiency in developing countries and draw a connection to social issues. They underline that an information system in an agricultural supply chain can improve the efficiency of production. Hence, it can also increase the income level of farmers through increased production in the long run, while also reducing the risk of social unrest. However, an advanced educational level of participants is required to ensure efficiency effects through the use of IT.

Environmental and social focus

A significant number of papers simultaneously address environmental and social sustainability together. Particularly, four areas are discussed: IT systems to fulfill multidimensional information needs, IT systems for sustainable agriculture in developing countries, IT systems to improve the sustainability of public procurement, and verification of whether applied IT systems drive sustainability in supply chains. Customer information usually needs to touch on all dimensions of sustainability as government and consumers, both want to be broadly informed. Therefore, companies in the food or agricultural supply chain, for example, need a high level of transparency in order to fulfill these demands, which is only possible using IT (Trienekens et al. 2012; Lehmann, Reiche, and Schiefer 2012; Q. Zhang et al. 2013). Additionally, although Koh et al. (2013) primarily consider carbon emissions, they also consider the social implications of evaluating carbon reduction measures in their system. In a different sense, including customers and their understanding of sustainability in the discussion can broaden the discussion even more (Sigala 2014).

Some authors discuss the use of IT for sustainability as a developmental aspects based on the Indian agricultural domain. Rao (2007) describes the role of ICT in improving the growth of agriculture and connects the environment, population growth and income levels with it. Verma and Chaudhuri (2009) instead suggest that e-commerce reduces farmer exploitation by reducing information asymmetries. Walker and Brammer (2012) show empirically that in developed countries e-commerce can also lead to improvements in environmental supply practices and of some (labor, safety and health), but not all (community, philanthropic activities and buying from minority-owned SMEs) social aspects in public procurement. In another food supply chain setting, Wognum et al. (2011) define social sustainability mainly as safeguarding the health of customers. Further, the right approach can ensure that IT drives supply chain sustainability. Piotrowicz and Cuthbertson (2009) postulate that (especially supply chain) information systems have to be evaluated based on a framework that includes all dimensions of sustainability with specific sub-categories. Zografos, Sedlacek, and Bozuwa (2012) ask for policies that improve ICT standardization and usage to ensure increased environmental and social sustainability. Only one paper explicitly combines environmental and social factors with an IT system optimizing them: Ayoub et al. (2006) address environmental and social factors in dealing with optimal locations of facilities along the supply chain. Implicitly, this is also what the system of Guyon et al. (2012) should enable. Therefore, within the scope of this survey, no IT system enables combined environmental and social optimisation in supply chains.

Immediacy of IT's influence on sustainability

Our literature survey distinguishes between direct and indirect effects of IT on sustainability. Usually direct effects mean environmental or social effects that occur through providing an IT system. These primary effects result from the use of IT devices and, for example, the material and energy needed for production. In contrast, secondary or indirect effects are driven by the use of IT (Piotrowicz and Cuthbertson 2009). However, in this paper we use a different definition of direct effects: Direct effects are created by IT that explicitly aims to reduce negative environmental and social effects, while indirect effects result from the use of IT systems that improve intermediary variables, for example efficiency. They are, therefore, said to improve sustainability, but sustainability is not the primary focus. This is an "egg or chicken problem" (i.e., does the increase in efficiency lead to emissions reduction or does emissions reduction lead to efficiency increase). Therefore, the primary aim of the IT-based improvement is considered.

Direct effects

Most direct effects named in the papers are related to environmental sustainability and not social sustainability. One paper directly relates to social effects by describing IT as an enabler for communicating knowledge about diseases, thereby improving social issues in the respective community (Rao 2007). Transmitting sensor data for food safety could be understood in this sense (Q. Zhang et al. 2013). Direct environmental effects of IT can be attributed to the optimization of carbon emissions as implemented in the IBM carbon tool (Sourirajan et al. 2009) or in the SCEnAT tool for supply chain decarbonization (Koh et al. 2013). Other tools aim to directly improve energy consumption in transport (Guenther and Farkavcová 2010) or road freight networks (Benz et al. 2012), for example, on the basis of life-cycle inventory data. Moreover, IT systems can help assess environmental impacts. The RFID system of Björk et al. (2011) allows a direct optimization and analysis of the environmental effects. Olugu and Wong's approach (2012) describes how an environmental assessment can be conducted using fuzzy values in the context of closed-loop SCM. Another theme arises when companies directly exchange data on sustainability factors to improve sustainability along the value chain (Carlson et al. 2001; Trienekens et al. 2012; Meacham et al. 2013). Benefits are already given if the exchange within a company is better supported (Muñoz et al. 2013; Borsato 2014). Direct optimization of environmental effects also occurs when the product's complete life cycle, until disassembly, is considered (Bullinger, Steinaecker, and Weller 1999). Enhanced models like multi-criteria optimization integrated into IT systems take the inclusion of "soft" parameters into consideration if the environmental impact cannot be quantified (Vannieuwenhuyse, Gelders, and Pintelon 2003). Another form of "direct impact" results from avoiding sustainability effects with the help of an IT system; for example, by complying with international regulations (M.-K. Chen, Tai, and Hung 2012) or by improving e-procurement (Walker and Brammer 2012).

Indirect effects

Many papers only focus on the indirect effects through increased efficiency in supply chain processes. Indirect effects on sustainability can vary in the "level of indirectness". "Very indirect" relations can be seen when increased IT usage improves the educational level of supply chain participants or other capabilities, which in the end (for example through better usage of resources) leads to improved sustainability (Rao 2007; Dao, Langella, and Carbo 2011). For example, e-commerce in agricultural supply chains can be a catalyst for growth and development (Verma and Chaudhuri 2009). Another extended link is described by Lehmann, Reiche, and Schiefer (2012), who underline consumer awareness as a mediating factor. Nevertheless, most papers that point to indirect effects address performance improvements, primarily environmental improvements, through an optimized operation (Hasle 1999; Melville 2010). The already mentioned EU research project "SuperGreen" (Clausen, Geiger, and Behmer 2012; Fozza and Recagno 2012) also uses ICT as an enabler for increased efficiency in the use of transport resources.

In a similar sense the effects of e-commerce on the environment are typically indirect, as IT improves the performance of processes (Abukhader and Jönson 2004). Davies, Mason, and Lalwani (2007) show how e-commerce and IT affect general haulage in the United Kingdom. The impact of IT is driven by improved visibility and planning accuracy along the supply chain and reductions in transaction, administrative and raw material/service costs. Furthermore, Dotoli et al. (2006) show that e-links in supply chains can increase agility and reconfigurability, thereby improving environmental sustainability. In reverse logistics increased information exchange can lead to indirect positive sustainability effects through efficiency and reduced waste (García-Rodríguez, Castilla-Gutiérrez, and Bustos-Flores 2013). Finally, risk can be seen as an indirect effect. Iakovou (2001) presents an interactive solution methodology to minimize the risks of environmental pollution due to maritime shipment of petroleum. In a comparable sense, Wognum et al. (2011) highlight the importance of transparency in achieving sustainability and Lai, Hsu, and Chen (2012) suggest an information system that allows verification of environmental certificates.

Summary

Several suggestions can be derived from the survey with relevance for this thesis. First, research on the output and effects of IT in SSCM can be improved with more empirical studies that try to develop a perspective on the current status of IT support, its relevance, and the possibilities of further practical support. This would be advisable also in view of the currently limited amount of empirical research in the domain. Second, only a small number of papers on IT for SSCM currently use models or has a quantitative approach. The analysis of decision-making approaches (e.g. the application of fuzzy sets, multi-criteria approaches, etc.) has been present only to a limited extent. This could also include the topic of risk analysis. A further application area for quantitative models could lie in supporting quantitative sustainability modeling.

The structured discussion of IT for sustainable SCM reveals several additional insights. In categorizing along the supply chain, little focus is placed on nodes and reverse activities, which could possibly provide opportunities for future research. Moreover, in previous studies in the general field of sustainability in SCM (Seuring and Müller 2008b), the research on social issues is underrepresented in the context of IT usage. Only one paper focused specifically on social issues, although some papers at least include social arguments in their discussions with an environmental focus.

When focusing on risk management only three papers of the survey are of particular interest: Iakovou’s decision support system (2001) to improve maritime crude transport and the related environmental risks, Ali and Kumar’s system (2011) supporting social factors and, thus, reducing the risk of social unrest through an indirect connection, and the approaches of Lai Hsu, and Chen (2012) as well as Wognum et al. (2011) who draw a connection between the possibility to achieve supply chain transparency through IT and reduced environmental risks. Nevertheless, none of them directly relates to the domain of this thesis.

Some entrepreneurial approaches without academic discussion exist that are related to SCRM. They have, up to the author’s knowledge, not been academically published nor evaluated. To mention few these are for example Sedex (2012) which allows suppliers to input sustainability data and sourcing deputies to see supplier inputs as well as to determine the basic risks associated with a supplier or Maplecroft (2012), allowing country and sector specific risk assessments. Other offerings such as Dun & Bradstreet (D&B, 2012), Thomson Reuters (2013) or Spotter (Spotter Europe, 2012) integrate data retrieved through text mining (e.g. news feeds or social media), but their intention is not a sustainability-specific analysis of risk. Generally, they are limited in the extent of how they cover local and supplier location-related risk aspects.

Appendix B: Overview of related literature reviews on IT for SSCM

Authors	Year	Sust. focus	# of papers*	Period	Research focus	Data gathering
a) Intersection of IT and SCM						
1. Jain, Wadhwa, and Deshmukh	2009	N/A	34R	N/A	IT systems for dynamic supply chains	NFM
2. Arzu Akyuz and Erman	2010	N/A	24	N/A	Performance management for supply chains incl. some IT aspects	LRJ
3. Perego, Perotti, and Mangiaracina	2011	N/A	44	1994-2009	IT and communication technology for logistics and freight transportation	KWS, LRJ, ASE
4. Sarac, Absi and Dauzere-Peres	2010	N/A	143R	N/A	Current state of applying RFID in SCM	NFM
5. Zhang, van Donk, and Dirk	2011	N/A	40	1995-2010	Survey studies on the influence of IT on SCM and supply-chain performance	KWS, LRJ, ASE
6. Costa et al.	2012	N/A	101R	N/A	Focus on RFID traceability in agri-food supply chains	NFM
7. Zeng et al.	2012	N/A	136R	N/A	Secure collaboration in supply chains	Environment-based design
8. Măzăreanu	2013	N/A	52R	N/A	Risk during implementation of information systems	NFM
9. Ngai et al.	2014	N/A	77	1994-2009	Decision support/intelligent systems; focus on textile and apparel supply chains	KWS, LRJ, ASE
10. Esposito and Evangelista	2014	N/A	41	1990-2012	Specific discussion of virtual enterprises	KWS, ASE, AEA
11. Phahl and Moxham	2014	N/A	103	N/A	Focus on integration of ECR, RFID and visibility	KWS, ASE
b) Intersection of Sustainability and IT						
1. Elliot and Binney	2008	ENV	39	N/A	General review with a split according to corporate capabilities	NFM
2. Melville	2010	ENV	35	2000-2007	Focus on information systems innovation	LRJ, ASE
3. Dao, Langella, and Carbo	2011	E&S	98R	N/A	Integration of IT with human resources management and supply chain management based on a resource-based view of the firm	NFM
4. Jenkin, Webster, McShain	2011	ENV	38	N/A-2009	General review with multi-level research framework	KWS, ASE
5. Brooks, Wang, and Sarker	2012	ENV	22**	N/A	Green IT, including practitioner magazines	KWS, LRJ, ASE
6. Harmon and Moolenkamp	2012	E&S	64R	N/A	Direct and indirect sustainability effects of IT services incl. a strategy framework	NFM
7. Loeser	2013	ENV	78	2008-2012	Construct definition and practice overview for Green IT measures and Green IS initiatives	KWS, LRJ, ASE
8. Tushi, Sedera, and Recker	2014	ENV	98	2007-2013	Green IT including a taxonomy of area	KWS, ASE
c) Intersection of Sustainability, SCM, and IT						
1. Thoroe, Melski, and Schumann	2009	N/A	38R	N/A	Focus on RFID and reverse logistics	NFM
2. Kurnia, Mahbubur, and Glohet	2012	E&S	43	N/A	Discussion of role of IT in sustainable SCM	KWS, ASE
3. Frehe and Teuteberg	2014	ENV	51	2000-N/A	Focus on green logistics with slight elements of SCM	KWS, LRJ, ASE

* "R"s indicate numbers of references instead of numbers of papers included in literature review; ** Academic papers only. Topics: SSCM: Intersection of Sustainability and SCM; IT&SCM: Intersection of IT and SCM; IT&S: Intersection of IT and Sustainability. Sustainability focuses: ENV: Environmental; E&S: Environmental and social; N/A: No focus on either environmental or social sustainability. Data gathering methods: KWS: Keyword search; LRJ: Research limited to certain journals; NFM: No stated formal material collection; ASE: Author-based subject & paper exclusion; AEA: Additional empirical analyses; ARD: Automatic relevance detection (by search engine).

Table 106: Existing, relevant literature reviews on intersecting fields of IT, SCM, and sustainability (own representation)[231]

[231] Data gathering methods partly based on Seuring and Gold (2012).

Appendix C: Overview of classification of papers

Paper	Metho-dology	Supply chain focus	IT focus	Sustain-ability	Imme-diacy of IT
Abukhader 2008	T	O	STM	E	I
Abukhader and Jönson 2004	T	O	STM	E	I
Ali and Kumar 2011	S	FF	STM, EM	S	I
Ayoub et al. 2006	T	FF	STM	E&S	D
Barrett, Strunjaš-Yoshikawa, and Bell 2007	T	P	STM	E	D
Benz et al. 2012	T	FF	STM, EM	E	D
Björk et al. 2011	T	P	EM	E	D
Borsato 2014	CS	O	EM	E	D
Bosona et al. 2013	T	FF	STM	E	I
Bullinger, Steinaecker, and Weller 1999	T	O	STM, EM	E	D
Carlson et al. 2001	T	O	STM, EM	E	D
Chen, Tai, and Hung 2012	T	P	STM	E	D
Clausen, Geiger, and Behmer 2012	CS	O	STM, EG	E	D, I
Dao, Langella, and Carbo 2011	T	O	STM	E&S	I
Davies, Mason, and Lalwani 2007	CS	FF	STM	E	I
Dhanda and Hill 2005	CS	RF	STM	E	I
Dotoli et al. 2006	M	O	STM	E	I
Edwards, Lyons, and Kehoe 2004	T	RF	STM, EM	E	I
Fozza and Recagno 2012	S	O	STM, EG	E&S	D, I
García-Rodríguez, Castilla-Gutiérrez, and Bustos-Flores 2012	S	RF	STM	E	I
Guenther and Farkavcová 2010	T	FF	STM	E	D
Guyon et al. 2012	T	FF	STM	E&S	D
Hasle 1999	T	FF	STM	E	I
Iacob et al. 2013	T	FF	EM	E	D
Iakovou 2001	M	FF	STM	E	I
Koh et al. 2013	T	O	STM	E&S	D
Lai, Hsu, and Chen 2012	CS	P	STM	E	I
Lee 2003	T	O	STM, EM	E	I
Lehmann, Reiche, and Schiefer 2012	R	O	STM, EM	E&S	I
Marett, Otondo, and Taylor 2013	S	FF	EM, EG	E	D, I
Martinez-Sala et al. 2009	T	P	STM, EM	E	I
McKinnon 2010	T	P	STM, EM	E	D
Meacham et al. 2013	S	O	STM, EM	E	D
Meixell and Norbis 2008	R	FF	STM	E	D, I
Melville 2010	R	O	STM, EM	E	D, I
Muñoz et al. 2013	T	O	STM, EM	E	D
Olugu and Wong 2012	T	RF	EM	E	D
Piotrowicz and Cuthbertson 2009	T	O	STM	E&S	D
Rao 2007	CS	FF	STM, EM, EG	E&S	D, I
Schiller 2012	T	FF	STM	E	I
Shi et al. 2012	T	RF	STM, EM	E	I
Sigala 2014	T	O	STM	E&S	D
Sourirajan et al. 2009	M	FF	STM	E	D
Suh, Smith, and Linhoff 2012	M	FF	EM	E	I
Taghaboni-Dutta, Trappey, and Trappey 2010	T	P	STM	E	I
Thurston and Eckelman 2011	CS	O	STM	E	D
Toh, Nagel, and Oakden 2009	T	O	EG	E&S	D, I
Trappey, Trappey, and Wu 2010	T	RF	EM	E	I
Trienekens et al. 2012	T	O	EM	E&S	D
Vannieuwenhuyse, Gelders, and Pintelon 2003	T	FF	STM	E	D
Verma and Chaudhuri 2009	CS	FF	STM, EM	E&S	I
Walker and Brammer 2012	S	O	STM	E&S	D
Wognum et al. 2011	T	O	EM	E&S	I
Zhang et al. 2013	T	O	EM	E&S	D
Zografos, Sedlacek, and Bozuwa 2012	R	FF	EM, EG	E&S	I

CS…Case study; S...Survey/Empirical model; M...Mathematical model; R...(Literature) review; T...Theoretical/Conceptual model
FF…Forward flow; RF…Reverse flow; P…Product; O…Overarching
STM...Systems for strategic and tactical management; EM...Systems for execution and monitoring;

EG...Supporting (policy) environment and general framework E…Environmenal; S…Social; E&S…Environmental&Social D…Direct effects; I...Indirect effects

Table 107: Overview of classification of papers in literature survey on IT for SSCM (author's representation)

Appendix D: Overview of questionnaire to derive information needs

I	**Sustainability Monitoring**
A	**Existing Processes**
1	How do you monitor environmental and social sustainability? (e.g. material, pollution, energy consumption, corruption, child labor, …)
1.1	SUPPLEMENTARY: Who does it? External?
1.2	SUPPLEMENTARY: Manually or technically supported?
2	Which IT systems do you use for this purpose?
3	Which specific data are you collecting for this purpose?
3.1	SUPPLEMENTARY: Which role does auditing play in this context, particularly in-person on-site auditing?
4	For which purpose are you using the data?
B	**Evaluation**
1	How would you assess the current processes?
1.1	SUPPLEMENTARY: Is there an overhead with regard to existing IT systems? Are these too complex or too complicated to use?
1.2	SUPPLEMENTARY: Is in-person on-site auditing necessary? Is it (too) expensive?
C	**Improvement Potential**
1	What would you like to change and how would you do it?
1.1	SUPPLEMENTARY: Are there any other internal or external data that would improve monitoring?
II	**Supplier Rankings with regard to Social Sustainability**
	Focus on Suppliers (w/a Human Rights & Corruption).
A	**Existing Processes**
1	In which business processes do you currently rank suppliers (e.g. supplier selection, supplier performance management, supply chain risk management, …)?
2	Up to which extent is sustainability (environmental and social) included in these processes?
3	How do you calculate supplier ratings?
3.1	SUPPLEMENTARY: Is the result product, supplier or location specific?
3.2	SUPPLEMENTARY: Is uncertainty included in the calculation process?
4	Which IT systems do you use for the ranking of suppliers?
5	Which data do you use for the ranking of suppliers?
5.1	SUPPLEMENTARY: Is real-time data included in the ranking processes?
B	**Evaluation**
1	How would you assess the current processes?
C	**Improvement Potential**
1	What would you like to change and how would you do it?
1.1	SUPPLEMENTARY: What do you think of automatic data integration?
1.2	SUPPLEMENTARY: Do you think, that ranking should be related to geographic locations?
III	**Stakeholder Integration & Sustainability Reporting**
A	**Existing Processes**
1	How do you disclose information about environmental and social sustainability of your company?
2	Who are your target groups?
3	How do you collect feedback from stakeholders concerning environmental and social topics (e.g. complaints, ideas, or comments)?
4	Do you include stakeholders into tasks in the context of sustainability management (e.g. planning tasks)?
5	Which IT systems support you to disclose sustainability information and to communicate with stakeholders?
B	**Evaluation**
1	How satisfied are you with the existing processes?
C	**Improvement Potential**
1	With which stakeholder groups would you like to intensify personal or virtual contact in the future?
2	Which trends do you currently see, which will require you to increasingly integrate stakeholders?

Table 108: Overview of questionnaire used to derive information needs (author's representation)

Appendix E: Bayesian networks for supply chain risk management

This section provides an overview of literature dealing with BN applied in the context of supply chain risk management (such as supplier selection and performance management). The literature search resulted in a list of 33 papers (Table 109; up to December 2014). Two additional variables have been ascribed to each paper. The first indicates whether sustainability is included in the model and the second indicates whether input from news (or comparable) sources is explicitly included.

Source	Domain	Focus of BN	Sustain-ability	News sources
Amundson et al. (2012)	Input availability risk	Model for the availability of feedstock	No	No
Amundson et al. (2014)	Life-cycle risk	Comparison of numeric and Boolean modeling strategies for risk analysis	Env.^	No
Badurdeen et al. (2014)	General risk	Model of general risks in supply chains incl. interdependencies	Yes	No
Basu et al. (2007)	Operational risk	Model of general risks that can impact supply chain operations	No	No
Bearzotti et al. (2011)	Disruption risk	Estimating delay probability	No	No
Chang (2013)	Supplier selection	Modeling negotiation between buyers and suppliers in a multi-agent framework; inclusion of factors influencing buy decision	No	No
Chen and Peng (2003)	Disruption risk	Extended BN for probability of uncertain events	No	No
Dogan and Aydin (2011)	Supplier selection	Combination of BN with total cost of ownership (TCO) for selecting suppliers	No	No
Fernández, Salomone, and Chiotti (2010)	Disruption risk	Proactive monitoring of disruption risks modeling supply chain processes	No	No
Ferreira and Born-stein (2012)	Supplier selection	Combination of BN (influence diagrams) with fuzzy logic to rank suppliers	No*	No
Garvey, Carnovale, and Yeniyurt (2014)	Risk propagation	Risk propagation in supply networks constructing a dedicated risk network graph	No	No
Kao et al. (2005)	Supply chain diagnostics	SC diagnostics with a dynamic BN covering variables from a quick scan focused on schedule adherence	No	No
Kao et al. (2010)	Supply chain diagnostics	Combination of BN with fuzzy parameters	No	No
Kreng and Chang (2003)	Supplier selection	Multi-agent approach with BN to model supplier selection decision	No	No
Li and Chandra (2007)	Knowledge integration	Modeling of "classical" SC risks across the supply chain ("big picture of system state")	No	No
Li and Gao (2010)	Supplier collaboration	Development of collaborative index	No	No
Lockamy III (2011)	Supplier benchmarking	Determination of supplier risk for benchmarking; combining evidence for different risk categories	No**	No
Lockamy III (2014)	Disruption risk	Modeling disaster risk in supply chains incl. value at risk (VAR)	No**	No
Lockamy III and McCormack (2009)	Operational risk	Modeling operational risks in supply chains incl. value at risk (VAR)	No**	No
Lockamy III and McCormack (2010)	Outsourcing risk	Modeling outsourcing risks in supply chains incl. value at risk (VAR)	No**	No
Lockamy III and McCormack (2012)	General risk	Modeling general risks in supply chains incl. value at risk (VAR)	No**	No
Makris, Zoupas, and Chryssolouris (2011)	Buyer behavior	Purchase probability for a highly customized product	No	No
Maleki and Cruz-Machado (2013)	Supplier performance	Supply chain performance incl. interdependencies between variables	Env.^	No
Maleki et al. (2012)	Supplier performance	Combination of analytical network process and BN with BN focusing on customer preferences	Env.^	No
Pai et al. (2003)	Disruption risk	Using thread-asset pairs to infer disruption probabilities for a production or logistics node	No	No
Peng (2011)	Overall SC forecasting	Suggestion of BN (influence diagrams) for time and cost forecasting	No	No
Qazi, Quigley, and Dickson (2014)	Project risk	Combination of BN with game theory risk assessment to model impact on cost and time of a project	No	No
Shevtshenko and Wang (2009)	Recycling/Reverse logistics	Reverse logistics together with probabilities split into uncertainty and variability (i.e. inherent randomness) - use of interval-valued imprecise probabilities	No	No
Soberanis (2010)	Disruption risk	Modeling supply chain disruption risk defining two different impact factors	No	No

Uhart et al. (2012)	General risk	Reduction in medicinal errors in drug supply chain for hospital management	No	No
Yen and Zeng (2010)	Material risk	Hierarchical assessment for material risk in a green supply chain setting	Env.^	No
Yuan et al. (2009)	Supplier evaluation	Supplier assessment under the inclusion of grey theory	No	No
Yuan et al. (2012)	Operational risk	Modeling operational risks in business processes	No	No

* Only the "planted area" indicator can be considered a triple-bottom line criterion**; **** Inclusion of "Supplier HR problem" and "EPA and FDA report sharing" from a disruption point of view and not as dedicated sustainability items; ^ Environmental

Table 109: Overview of literature on BNs for supply chain risk management and related areas (author's representation)[232]

Only five cases include sustainability in their modeling approach. Four cases integrate environmental sustainability specifically (Joseph Amundson et al. 2014; Maleki, Bashkite, and Machado 2012; Yen and Zeng 2010; Maleki and Cruz-Machado 2013). Only one paper particularly includes social sustainability elements in the BN. Badurdeen et al. (2014) use BNs as a way to combine different risk types.[233] However, they do not go beyond taking a single variable per risk type as input, and they are more concerned with quality, cost, and time factors than with testing the compliance of a social sustainability or other aspect. The other papers do not incorporate sustainability.[234] Nevertheless, Ferreira and Borenstein (2012) suggest a more update-oriented, dynamic approach incorporating new evidence over time; however, this is neither news related nor focused on social sustainability. Amundson et al. (2014) also discuss a generic inclusion of evidence. In comparison, Ho, Xu, and Dey (2010) performed a broad literature review of supplier selection models based on multi-objective programming. They do not find a single paper referencing BNs for multi-criteria decision making. Overall, none of the papers considers the explicit integration of news or other comparable sources into the model to adjust the respective risk factors.

In summary, to the best of the author's knowledge BNs have only been applied to a limited extent in order to rank suppliers with regard to risk. No paper considers ongoing social sustainability compliance analysis as well as the explicit integration of news evidence (i.e., allowing the inclusion of multiple news reports). The application of BNs to social sustainability in supply chains can therefore yield new insights.

[232] An additional paper on supply disturbances is presented by Li-li, Yun-miao, and Zhen (2011). However, the paper is not published in English and according to the abstract does not include sustainability.

[233] Factors included by Lockamy III with and without McCormack (2011; 2012; 2009; 2014; 2010) such as "Supplier HR problem" and "EPA and FDA report sharing" could be considered to refer to environmental and social concerns. However, sustainability is not highlighted in the papers and the factor is only considered as input with single variables.

[234] 28 papers fall into this category (J. Amundson et al. 2012; Basu et al. 2007; Bearzotti et al. 2011; Chang 2013; Ye Chen and Peng 2003; Dogan and Aydin 2011; Ferreira and Borenstein 2012; Garvey, Carnovale, and Yeniyurt 2015; Kao, Huang, and Li 2005; Kao et al. 2010; Kreng and Chang 2003; X. Li and Chandra 2007; J. Li and Gao 2010; Lockamy III 2011; Lockamy III and McCormack 2012; Lockamy III and McCormack 2009; Lockamy III 2014; Makris, Zoupas, and Chryssolouris 2011; Maleki and Cruz-Machado 2013; Pai et al. 2003; Peng 2011; Qazi et al. 2014; Shevtshenko and Wang 2009; W.-F. Yuan et al. 2009; Lockamy III and McCormack 2010; Uhart et al. 2012; C. Yuan et al. 2012; Fernández, Salomone, and Chiotti 2010; Donaldson Soberanis 2010).

Appendix F: Sources for child labor indicators

Indicator	Sources
Accessibility of education	(International Labour Office 2013, 13); (International Labour Organization 2008a); (International Labour Office 2004, 85); (International Programme on the Elimination of Child Labour and International Labour Office 2011, 65); (International Programme on the Elimination of Child Labour, Statistical Information and Monitoring Programme on Child Labour, and International Labour Office 2004, 29); (Martin 2013, 159); (United States Department of Labor 2013)
Age checks with schools	HRCA A.2.1.5
Belonging to minority population suffering social exclusion	(International Labour Office 2004, 87); (International Labour Office 2013, 21); (United States Department of Labor 2013)
Business exchange with stakeholder groups (NGO, schools, etc.)	HRCA Quick Check 19/20
Control of risk incl. suppliers	GRI 3.1 Human rights
Costs of education	(Fors 2012, 576); (Vandenberg, Nippierd, and Gros-Louis 2007b, 18–19)
Difficult family situations	(Dorman 2008, 6); (ECLT Foundation 2013); (Fors 2012, 579); (International Labour Office 2004, 87); (International Programme on the Elimination of Child Labour, Statistical Information and Monitoring Programme on Child Labour, and International Labour Office 2004, 29)
Effects of HIV/AIDS	(ECLT Foundation 2013); (International Labour Office 2004, 87)
Existence of certifications	GRI 3.1 Human rights
Existence of child labor policy/institutions	HRCA check (International Programme on the Elimination of Child Labour and International Labour Office 2011, 41, 44, 48, 65); (International Programme on the Elimination of Child Labour, Statistical Information and Monitoring Programme on Child Labour, and International Labour Office 2004, 29, 30); (United States Department of Labor 2013); (Vandenberg, Nippierd, and Gros-Louis 2007b, 11, 14, 21, 23)
Existence of management system for human rights	GRI 3.1 Human rights; HRCA Quick Check A.2.3.8
Existence of remediation/education programs	HRCA check; SA 8000 IV.1.2; BSCI 7; GSCP 2.2; HRCA A.2.2.1; GC self assessment LA.4.f
Felt need for education	(International Labour Office 2004, 90)
Geographic area	GRI 3.1 HR6 2.2; (International Programme on the Elimination of Child Labour and International Labour Office 2011, 8); (International Programme on the Elimination of Child Labour 2013, 17); (United States Department of Labor 2013); (Vandenberg, Nippierd, and Gros-Louis 2007b, 10)
Household size	(Dorman 2008, 9); (International Labour Office 2004, 83); (International Labour Organization 2008a); (Siddiqi 2013, 211); (Webbink, Smits, and Jong 2013, 832)
Household assets/debt	(ECLT Foundation 2013); (International Labour Office 2004, 81)
Household income fluctuations	(International Labour Office 2004, 95); (International Programme on the Elimination of Child Labor 2012, 13)
Income inequality	(Fors 2012, 581); (Tanaka 2003, 96)
Insurance/credit programs/ financial support/social spending	(Dorman 2008, 7); (ECLT Foundation 2013); (Institute for Human Development 2009, 29); (International Labour Office 2004, 96, 115); (International Labour Office 2013, 9, 39); (International Programme on the Elimination of Child Labour and International Labour Office 2011, 65); (United States Department of Labor 2013); (Vandenberg, Nippierd, and Gros-Louis 2007b, 24)
Level of education (child / parents)	(Dorman 2008, 8); (International Labour Office 2004, 87); (Martin 2013, 159); (Webbink, Smits, and Jong 2013, 829)

Low household income	(Dorman 2008, 8); (Ellis, International Labour Organisation, and International Programme on the Elimination of Child Labour 2005, 33); (Fors 2012, 575); (Giblin 2012); (International Labour Organization 2008a); (International Programme on the Elimination of Child Labour, Statistical Information and Monitoring Programme on Child Labour, and International Labour Office 2004, 29); (Martin 2013, 158); (Siddiqi 2013, 211–212); (United States Department of Labor 2013)
Minimum wage	(International Labour Office 2013, 13); (International Programme on the Elimination of Child Labour, Statistical Information and Monitoring Programme on Child Labour, and International Labour Office 2004, 30); (Vandenberg, Nippierd, and Gros-Louis 2007b, 20)
National income (partly) / GDP per capita (country level poverty)	(Diallo et al. 2010, 39); (International Labour Office 2004, 84); (International Labour Office 2013, 12); (International Programme on the Elimination of Child Labour 2013, 7); (International Programme on the Elimination of Child Labour, Statistical Information and Monitoring Programme on Child Labour, and International Labour Office 2004, 29); (Fors 2012, 572)
Oversight through audits/inspections	(Closs, Speier, and Meacham 2011, 110); HRCA A.2.1.8, A.2.1.9, A.2.2.3, A.2.3.9 (International Programme on the Elimination of Child Labour and International Labour Office 2011, 45, 66); (Vandenberg, Nippierd, and Gros-Louis 2007b, 29)
Parental skill level	(International Labour Office 2004, 87)
Poor family values	(International Labour Office 2004, 87); (International Programme on the Elimination of Child Labour and International Labour Office 2011, 65)
Power vs. customers/stakeholders	(Williams 2003, 20–21)
Product characteristics	(ECLT Foundation 2013); (International Labour Office 2004, 100–101); (International Programme on the Elimination of Child Labour, Statistical Information and Monitoring Programme on Child Labour, and International Labour Office 2004, 30); (Vandenberg, Nippierd, and Gros-Louis 2007b, 22)
Production peaks/labor shortages	(International Labour Office 2013, 13); (International Programme on the Elimination of Child Labour, Statistical Information and Monitoring Programme on Child Labour, and International Labour Office 2004, 30); (Vandenberg, Nippierd, and Gros-Louis 2007b, 21–22)
Quality of education	(International Labour Office 2004, 85); (International Labour Office 2013, 13); (International Labour Organization 2008a); (International Programme on the Elimination of Child Labour 2005, 3); (International Programme on the Elimination of Child Labour and International Labour Office 2011, 65); (Vandenberg, Nippierd, and Gros-Louis 2007b, 18)
Ratified international conventions	(Vandenberg, Nippierd, and Gros-Louis 2007b, 11)
Ruralness of geography/neighborhood	(Allais, Hagemann, and International Programme on the Elimination of Child Labour 2008, v); (Martin 2013, 159); (Siddiqi 2013, 212); (Vandenberg, Nippierd, and Gros-Louis 2007b, 10); (Webbink, Smits, and Jong 2013, 832)
School enrolment rates	(Allais, Hagemann, and International Programme on the Elimination of Child Labour 2008, v); (Diallo et al. 2010, 25); (Dorman 2008, 3); (Edmonds 2008, 12); (Institute for Human Development 2009, 11) (International Programme on the Elimination of Child Labor 2012, 60); (International Programme on the Elimination of Child Labour and International Labour Office 2011, 65)
Sector/Operation	(Diallo et al. 2010, 13); GRI 3.1 – HR6_2.2; (International Programme on the Elimination of Child Labour and International Labour Office 2011, 9); (International Programme on the Elimination of Child Labour 2013, 7); (United States Department of Labor 2013); (Vandenberg, Nippierd, and Gros-Louis 2007b, 10); (Williams 2003, 20)
Societal acceptance of child labor	(Giblin 2012); (International Labour Office 2004, 81, 83); (International Labour Office 2013, 13); (International Labour Organization 2008a); (International Programme on the Elimination of Child Labour 2005, 3) (International Programme on the Elimination of Child Labour and International Labour Office 2011, 65);

	(International Programme on the Elimination of Child Labour, Statistical Information and Monitoring Programme on Child Labour, and International Labour Office 2004, 30); (Martin 2013, 160); (United States Department of Labor 2013); (Vandenberg, Nippierd, and Gros-Louis 2007b, 14, 20); (Williams 2003, 20–21)
Socio-economic dislocation (economic crisis, political and social transition)	(Dorman 2008, 6); (International Labour Office 2004, 87); (International Labour Office 2013, 13, 22, 25); (International Labour Organization 2008a); (International Programme on the Elimination of Child Labour, Statistical Information and Monitoring Programme on Child Labour, and International Labour Office 2004, 29); (United States Department of Labor 2013); (Vandenberg, Nippierd, and Gros-Louis 2007b, 15)
Strength of unions, NGOs, etc.	(International Programme on the Elimination of Child Labor 2012, 3); (International Programme on the Elimination of Child Labour 2010, 1); (International Programme on the Elimination of Child Labour and International Labour Office 2011, 52, 65); (United States Department of Labor 2013); (Vandenberg, Nippierd, and Gros-Louis 2007b, 11–12)
Strong peer group and external influences, with material values	(International Labour Office 2004, 87); (International Programme on the Elimination of Child Labour, Statistical Information and Monitoring Programme on Child Labour, and International Labour Office 2004, 29)
Supervision of apprentice systems	HRCA A.2.1.7, A.2.2.2; GC self assessment LA.4.e; FLA CL.8
Work organization & access to capital	(International Labour Office 2004, 101)
Works system productivity	(International Labour Office 2004, 85); (International Programme on the Elimination of Child Labour and International Labour Office 2011, 46); (International Programme on the Elimination of Child Labour, Statistical Information and Monitoring Programme on Child Labour, and International Labour Office 2004, 30); (Vandenberg, Nippierd, and Gros-Louis 2007b, 14, 23, 30); (Webbink, Smits, and Jong 2013, 832)
Written age-check/identification procedures	GC self assessment – LA.4.d; (Giblin 2012); HRCA A.2.1.2, A.2.1.3, A.2.1.4; (International Programme on the Elimination of Child Labour and International Labour Office 2011, 48, 65); (Vandenberg, Nippierd, and Gros-Louis 2007b, 11)

Note: GRI…Global Reporting Initiative (2011); HRCA…Human Rights Compliance Assessment (The Danish Institute for Human Rights 2006); SA…Social Accountability (Social Accountability International 2008b); BSCI…Business Social Compliance Initiative (2009); GSCP…Global Social Compliance Programme (2010); GC…Global Compact (UN Global Compact Nordic Network 2013); FLA…Fair Labor Association (2011)

Table 110: Overview of sources for indicators for child labor (author's representation)

Appendix G: Child labor per ruralness

The following calculation derives the number of children in rural/urban areas based on the variable definition used in the risk model methodology. Valid for the case that $PCL_{C,R_1} \neq PCL_{C,R_2}$.

$NC_C = NC_{C,R_1} + NC_{C,R_2}$
$NCL_C = NCL_{C,R_1} + NCL_{C,R_2}$
$NCL_{C,R_1} = NC_{C,R_1} * PCL_{C,R_1}$
$NCL_{C,R_2} = NC_{C,R_2} * PCL_{C,R_2}$

$NCL_C = NC_{C,R_1} * PCL_{C,R_1} + NC_{C,R_2} * PCL_{C,R_2}$
$= NC_{C,R_1} * PCL_{C,R_1} + \left(NC_C - NC_{C,R_1}\right) * PCL_{C,R_2}$
$= NC_{C,R_1} * PCL_{C,R_1} - NC_{C,R_1} * PCL_{C,R_2} + NC_C * PCL_{C,R_2}$
$= NC_{C,R_1} * \left(PCL_{C,R_1} - PCL_{C,R_2}\right) + NC_C * PCL_{C,R_2}$

$$NC_{C,R_1} = \frac{NCL_C - NC_C * PCL_{C,R_2}}{PCL_{C,R_1} - PCL_{C,R_2}}$$

$$NC_{C,R_1} = \frac{NC_C * PCL_C - NC_C * PCL_{C,R_2}}{PCL_{C,R_1} - PCL_{C,R_2}} > 0 \ (see\ below)$$

Greater than zero is shown in the following:

$$\frac{NC_C * PCL_C - NC_C * PCL_{C,R_2}}{PCL_{C,R_1} - PCL_{C,R_2}} > 0$$

Two resulting cases (not defined for equality) – case (1) $PCL_{C,R_1} - PCL_{C,R_2} < 0$:

$NC_C * PCL_C - NC_C * PCL_{C,R_2} < 0$
$PCL_C - PCL_{C,R_2} < 0$
$\frac{NCL_{C,R_1} + NCL_{C,R_2}}{NC_C} - PCL_{C,R_2} < 0$
$NCL_{C,R_1} + NCL_{C,R_2} < NC_C * PCL_{C,R_2}$
$NC_{C,R_1} * PCL_{C,R_1} + NC_{C,R_2} * PCL_{C,R_2} < NC_C * PCL_{C,R_2}$
$NC_{C,R_1} * PCL_{C,R_1} + \left(NC_C - NC_{C,R_1}\right) * PCL_{C,R_2} < NC_C * PCL_{C,R_2}$
$NC_{C,R_1} * PCL_{C,R_1} + NC_C * PCL_{C,R_2} - NC_{C,R_1} * PCL_{C,R_2} < NC_C * PCL_{C,R_2}$
$NC_{C,R_1} * PCL_{C,R_1} - NC_{C,R_1} * PCL_{C,R_2} < 0$
$NC_{C,R_1} * \left(PCL_{C,R_1} - PCL_{C,R_2}\right) < 0$

As the expression in the bracket is negative in this case and $NC_{C,R_1} > 0$ the whole term is negative. Q.e.d. The same holds true for the second case with $PCL_{C,R_1} - PCL_{C,R_2} > 0$.

Appendix H: Proof: Monotonically increasing expected value of observational likelihood

This section aims at showing that the expected value of the observational likelihood is monotonically strictly increasing in the number of IIOs (i.e. $(k_t f_t)_t$ given the frequency of (i.e. the number of) IIOs entered f and the expected value of a node k is monotonically strictly increasing). It is an applied version of learning CPTs through counting-learning introduced above (main text). The "node" is a leaf node (i.e. without parents) and can for example be the credibility c which is updated by incident reports. The same is then also true for the relevance r. The statement will be proved for the case of a discretized (continuous) credibility node with uninformed prior as used in the risk model. Learning from a finding results in an update of a node's probability table (see Figure 101).

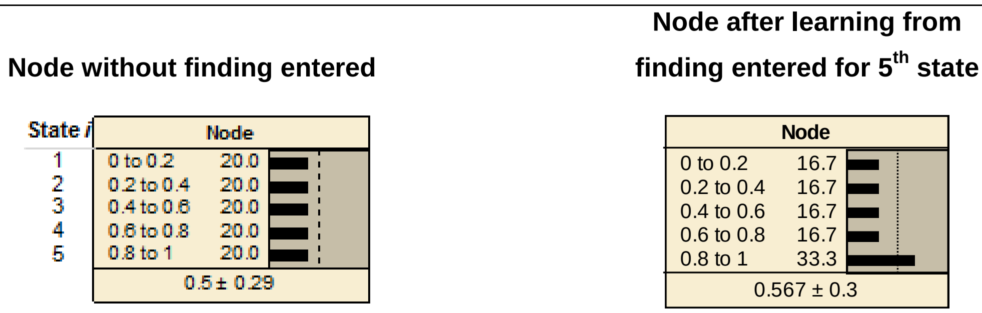


Figure 101: Exemplary nodes of BN with finding included at 5th state for right node (on left hand side with variables; author's representation)

Assumptions and definitions:

$f; f \in \mathbb{N}; \mathbb{N} > 0$ … Number of IIOs (news reports) related to a specific supplier location l

$i, i \in \{1, \dots, 5\}$ … Indices of discretized node states from lowest to highest value (roughly: 1 for 0-0.2; 2 for 0.2-0.4; 3 for 0.4-0.6; 4 for 0.6-0.8; 5 for 0.8-1)

$X, |X| = 4$ … Set of node states without an included finding

t … Time index of IIOs (indicates the arriving of new items over time

$s_i, s_i \in (0,1), s_{i,t} = s_{i,t+1} \forall t, s_i < s_{i+1}\ \forall i$ … Avg. of discretized state i (e.g. 0.1 for 0 to 0.2 discretization – these are the values of the discretized states in Netica)

$p(s_{i,t}); p(s_{i,t}) \in (0,1)$ … Probability of discretized state i in time t based on state averages (each state represented with discretized average)

We further differentiate states if a finding has been entered in time t

$x_{j,t}, x_{j,t} \in (0,1), j \in \{1, \dots, 4\}, x_{j,t} < x_{j+1,t}\ \forall j$ … Avg. of discretized state j without an included finding with index i in time t

$y_t, y_t \in (0,1) \ldots$ Average of discretized state with an included finding in time t –this is the avg. value of the state in each period a finding is entered

$\forall x_j, j \in (1, \ldots, 4)\ \exists i, s_i = x_j, x_j \neq s_k, k \neq i, i \in (1, \ldots, 5), k \in (1, \ldots, 5)$

Given the assumed discretization of the node into intervals of similar sizes and the usage of the discretized states for the update of findings, the overall mean value of the node in time t will be calculated based on the states with:

$$E(node) = E(c) = \sum_i s_i * p(s_i)$$

Here, we use BN concepts implemented by Netica[235], a standard software for BN. If a finding has been included in time t then the following formula applies in t given that the state averages for x and y change over time depending on which state of the node "receives" a finding (IIOs) f.

$k_t, k_t \in (0,1) = \sum_{j=1}^{4} x_{j,t} * p(x_{j,t})_t + y_t * p(y_t)_t \ldots$ Expected value of a node with a finding entered in time t

The probability of a node state after including a finding is calculated with the concept of counting-learning using experience as described above in the main text. Here it can be applied in the following form if one finding is entered:

$exp_t = exp_{t-1} + 1, exp_0 \geq 0, exp_t = exp_0 + f_t \ldots$ Experience for the node at time t based on the number of findings entered for the node

$$\begin{cases} Finding\ included\ for\ state\ s_i: p(s_i)_t = p(y_t) = \dfrac{p(s_i)_{t-1} * exp_{t-1} + 1}{exp_t} \\ No\ finding\ included\ for\ state\ s_i: p(s_i)_t = p\left(x_{j,t}\right) = \dfrac{p(s_i)_{t-1} * exp_{t-1}}{exp_t} \end{cases}$$

Altogether it will be shown that the following statement of monotonicity holds in the case concerned within the Netica framework. In each time t a finding is entered per assumption.

$k_t * f_t < k_{t+1} * f_{t+1};\ f_{t+1} = f_t + 1$

Proof:

Expanding k leads to:

---

[235] https://www.norsys.com/netica.html

$$f_t * \left(\sum_{j=1}^{4} x_{j,t} * p(x_{j,t})_t + y_t * p(y_t)_t\right) < f_{t+1} * \left(\sum_{j=1}^{4} x_{j,t+1} * p(x_{j,t+1})_{t+1} + y_{t+1} * p(y_{t+1})_{t+1}\right)$$

For better visibility of changes the right side is fully expanded:

$$f_t * \left(\sum_{j=1}^{4} x_{j,t} * p(x_{j,t})_t + y_t * p(y_t)_t\right) < f_{t+1} * (x_{1,t+1} * p(x_{1,t+1})_{t+1} + x_{2,t+1} * p(x_{2,t+1})_{t+1} + x_{3,t+1} * p(x_{3,t+1})_{t+1} + x_{4,t+1} * p(x_{4,t+1})_{t+1} + y_{t+1} * p(y_{t+1})_{t+1})$$

Substituting the calculation of p through update formulas used by Netica:

$$f_t * \left(\sum_{j=1}^{4} x_{j,t} * p(x_{j,t})_t + y_t * p(y_t)_t\right) < f_{t+1} * \left(x_{1,t+1} * \frac{p\left(x_{1,t+1}\right)_t * exp_t}{exp_{t+1}} + x_{2,t+1} * \frac{p\left(x_{2,t+1}\right)_t * exp_t}{exp_{t+1}} + x_{3,t+1} * \frac{p\left(x_{3,t+1}\right)_t * exp_t}{exp_{t+1}} + x_{4,t+1} * \frac{p\left(x_{4,t+1}\right)_t * exp_t}{exp_{t+1}} + y_{t+1} * \frac{p(y_{t+1})_t * exp_t + 1}{exp_{t+1}}\right)$$

At this point, we have to make assumptions without loss of generality about the relationships of the different x. The statement will be proved for the case that a message with the highest credibility case possible is followed by a message with the lowest credibility case possible. This sequence represents the case which is the most likely to break the monotonicity assumption. In other words it represents the largest possible reduction of the expected credibility by an update through an IIO. The proof will show that the strict monotonicity condition holds in this most extreme case.

Knowing this leads to the following consequences:

$y_t = s_5, x_{1,t} = s_1, x_{2,t} = s_2, x_{3,t} = s_3, x_{4,t} = s_4$

$y_{t+1} = s_1, x_{1,t+1} = s_2, x_{2,t+1} = s_3, x_{3,t+1} = s_4, x_{4,t+1} = s_5$

Therefore, we can substitute in the formula above:

$$f_t * \left(\sum_{i=1}^{4} s_i * p(s_i)_t + s_5 * p(s_5)_t\right) < f_{t+1} * \left(s_2 * \frac{p(s_2)_t * exp_t}{exp_{t+1}} + s_3 * \frac{p(s_3)_t * exp_t}{exp_{t+1}} + s_4 * \frac{p(s_4)_t * exp_t}{exp_{t+1}} + s_5 * \frac{p(s_5)_t * exp_t}{exp_{t+1}} + s_1 * \frac{p(s_1)_t * exp_t + 1}{exp_{t+1}}\right)$$

Given that $s_1 * \frac{p(s_1)_t * exp_t + 1}{exp_{t+1}} = s_1 * \frac{p(s_1)_t * exp_t}{exp_{t+1}} + \frac{s_1}{exp_{t+1}}$ the bracket can be recombined. Moreover the first term is reformulated:

$$f_t * \left(\sum_{i=1}^{5} s_i * p(s_i)_t\right) < f_{t+1} * \left(\frac{1}{exp_{t+1}} * \left(exp_t * \sum_{i=1}^{5} s_i * p(s_i)_t + s_1\right)\right)$$

Utilizing the definition of $E(node)$:

$$f_t * E(node) \ < f_{t+1} * \left(\frac{exp_t * E(node)}{exp_{t+1}} + \frac{s_1}{exp_{t+1}}\right)$$

Including the definition of f_{t+1} the right side can be reformulated:

$$f_t * E(node) \ < f_t * \frac{exp_t * E(node)}{exp_{t+1}} + \frac{exp_t * E(node)}{exp_{t+1}} + f_{t+1} \frac{s_1}{exp_{t+1}}$$

Subtracting by $f_t * \frac{exp_t * E(node)}{exp_{t+1}} > 0$:

$$f_t * E(node) - f_t * \frac{exp_t * E(node)}{exp_{t+1}} < \frac{exp_t * E(node)}{exp_{t+1}} + f_{t+1} \frac{s_1}{exp_{t+1}}$$

Expanding the first term we get:

$$f_t * \frac{exp_{t+1} * E(node)}{exp_{t+1}} - f_t * \frac{exp_t * E(node)}{exp_{t+1}} < \frac{exp_t * E(node)}{exp_{t+1}} + f_{t+1} \frac{s_1}{exp_{t+1}}$$

Including that $exp_{t+1} = exp_t + 1$:

$$\frac{f_t * exp_t * E(node) + f_t * E(node) - f_t * exp_t * E(node)}{exp_{t+1}} < \frac{exp_t * E(node)}{exp_{t+1}} + f_{t+1} \frac{s_1}{exp_{t+1}}$$

Therefore reducing the left term and subtracting $\frac{exp_t * E(node)}{exp_{t+1}} > 0$ we get:

$$\frac{f_t * E(node)}{exp_{t+1}} - \frac{exp_t * E(node)}{exp_{t+1}} < f_{t+1} \frac{s_1}{exp_{t+1}}$$

Reformulating the left side and multiplying with $exp_{t+1} > 0$ we get:

$$E(node) * (f_t - exp_t) < f_{t+1} * s_1$$

Including that $exp_t = exp_0 + f_t$:

$$E(node) * (f_t - exp_0 - f_t) < f_{t+1} * s_1$$

Reformulating the left side we get:

$$-E(node) * exp_0 < f_{t+1} * s_1$$

$$-E(node) * exp_0 < f_t * s_1 + s_1$$

Given that we know that $E(node) \in (0,1)$, $exp_0 \geq 0$, $f_t > 0$ (as findings have been entered), and $s_1 > 0$, then the inequality must hold for the case discussed here. Q.e.d.

Appendix I: Synonym sets used for document classification

English

Seed	Words
Labor	**Nouns**: labor; labour; proletariat; toil; working class **Verbs**: dig; drudge; fag; grind; labor; labour; moil; toil; travail
Work	**Nouns**: employment; piece of work; work; workplace **Verbs**: crop; cultivate; do work ; forge; form; knead; mold; mould; shape; work

Table 111: Words used in small set for gazetteers of program (author's representation)

Seed	Words
Labor	**Nouns**: corvee; donkeywork; drudgery; effort; elbow grease; exertion; grind; hackwork; haymaking; hunt; hunting; labor; labor force; labor pool; labour; lumpenproletariat; manual labor; manual labour; organized labor; overwork; overworking; plodding; proletariat; roping; slavery; sweat; toil; travail; working class; **Verbs**: dig; do work; drudge; fag; grind; labor; labour; moil; toil; travail; work
Work	**Nouns**: action; aid; attention; bakehouse; bakery; bakeshop; beehive; booking; brokerage; brokerage house; busywork; caning; care; central; chef-d'oeuvre; coaching; coaching job; colliery; coursework; creamery; drill site; duty; employment; engagement; exchange; farm; fish farm; fishery; followup; follow-up; forge; gasworks; glassworks; handcraft; handicraft; handiwork; handwork; heavy lifting; housekeeping; housewifery; housework; investigating; investigation; ironing; ironwork; ironworks; job; lab; labor; laboratory; labour; lacework; lacquerware; laundry; lavation; leatherwork; location; logging; loose end; lumberyard; make-work; masterpiece; metalwork; ministry; mission; missionary work; navigation; nightwork; openwork; operation; oyster bank; oyster bed; oyster park; paper route; paperwork; piece of work; piecework; piscary; pit; polishing; polychrome; procedure; project; proving ground; public service; publication; research lab; research laboratory; rope yard; ropewalk; roundhouse; sailing; science lab; science laboratory; seafaring; service; services; shining; shipyard; shop; shop floor; silverwork; smithy; social service; spadework; studio; subbing; substituting; tannery; task; telecommuting; telephone exchange; teleworking; tending; test bed; timework; toil; undertaking; unfinished business; wash; washing; waterworks; welfare work; wicker; wickerwork; woodwork; work; work in progress; work load; workload; workpiece; workplace; workshop; **Verbs**: assist; bank; beat; beaver; beaver away; blackleg; boondoggle; break one's back; buckle down; busy; carpenter; carve; cast; chip; clerk; coil; collaborate; cooperate; crop; cultivate; cut out; dig; do work ; drive; drudge; electioneer; fag; farm; fill; fink; forge; form; freelance; get together; go through; grind; handbuild; hand-build; hill; intern; job; join forces; keep one's nose to the grindstone; keep one's shoulder to the wheel; knead; knuckle down; labor; labour; layer; machine; man; masticate; mess around; minister; model; moil; mold; monkey; monkey around; moonlight; mould; mound; muck about; muck around; occupy; overcrop; overcultivate; page; peg away; plug away; potter; preform; proof; puddle; pull one's weight; putter; rat; remold; reshape; roughcast; run through; scab; scant; sculpt; sculpture; serve; shape; sinter; skimp; skipper; slave; slog; specialise; specialize; stamp; subcontract; swage; take; throw; tinker; toil; travail; turn a trick; upset; volunteer; wait; waitress; whore; work; work at; work on; work through

Table 112: Words used in extended set for gazetteers of program (author's representation)

German

Seed	Synonyms
Kind	**Nouns**: Mädchen, Junge, Tochter, Sohn, Bursche, Bubi, Bürschchen, Kleiner, Halbwüchsiger, Teen, Jugendliche, Heranwachsende, Halbwüchsige, Heranwachsender, Kids, Jugendlicher, Kerlchen, Kleine, Nachkomme, Enkel, Nachfahre, Enkelkind, Enkelsohn, Enkeltochter, Enkelin, Minderjährige, Unmündige **Adjectives**: jung, minderjährig, nicht volljährig, unmündig, unreif, kindsköpfig, kindlich, kindhaft, klein, jünger, jugendlich, adoleszent, halbwüchsig, heranwachsend
Arbeit	**Nouns**: Beschäftigung, Anstellung, Festanstellung, Arbeitsplatz, Arbeitsstelle, Arbeitsverhältnis, Dienst, Beruf, Arbeitsfeld, Handwerk, Metier, Arbeitsbereich, Aufgabenbereich, Aufgabengebiet, Betätigungsfeld, Fachgebiet, Broterwerb, Erwerb, Gelderwerb, Engagement, Erwerbstätigkeit, Gewerbe, Geschäft, Handel, Transaktion, Verkauf, Position, Posten, Beschäftigungsverhältnis, Stelle, Stellung, Tätigkeit, Betätigung, Bedienung, Handhabung, Steuerung, Hantierung, Werk, Produktion, Verfertigung, Fertigung, Fabrikation, Fertigstellung, Bau, Errichtung, Verrichtung, Herstellung, Erzeugung, Anfertigung, Tätigkeit, Tun, Berufsausübung, Berufstätigkeit, Job **Verbs**: arbeiten, hantieren, tätig, werken, betätigen, wirken, beschäftigen, wirtschaften, bedienen, führen, dirigieren, vorantreiben, handhaben, betätigen, erwerbstätig, beschäftigen, betreiben, ausführen, verfertigen, anfertigen, bereiten, bauen, erstellen, erzeugen, produzieren, herstellen, fertigen, verrichten, abwickeln, nachgehen, praktizieren, ausüben, versehen, betreiben, vollführen, handeln, anbieten, andienen, antragen, hausieren, verkaufen, vermarkten, vertreiben, fertigstellen
Zwangsarbeit	Knechtschaft, Abhängigkeit, Leibeigenschaft, Hörigkeit, Unfreiheit, Unterdrückung, Knechtung, Unterjochung, Versklavung, Sklaverei
Ausbeutung,	**Nouns**: n/a + Missbrauch included additionally **Verbs**: ausbeuten, missbrauchen, knechten, beherrschen, gefügig machen, unterdrücken, unterjochen, versklaven, tyrannisieren, schinden, zusetzen, schlauchen, strapazieren, traktieren
Verdienst	**Nouns**: Verdienst, Bezahlung, Bezüge, Einkommen, Einkünfte, Einnahmen, Entgelt, Entlohnung, Gehalt, Lohn, Vergütung **Verbs**: verdienen

Table 113: Synonyms used in gazetteers of program (author's representation)

Appendix J: Word lists used for rule-based sector tagging

Remove unnecessary words		**Filter parts**	**Get specification segments**
Cut part after phrase including phrase	**Delete phrase from labels**	**Delete segment if phrase included in segment**	**Move to specification segments from phrase onwards**
including	women's or girls'	%	made up of
without	babies'	related	of
of subclass	men's or boys'	the like	incidental to
not	in shell	its	to
except	similar	any	-
other than	provided	its compounds	doped for
concerning	parts thereof	its derivatives	designed for
but not	and the like	processed	related to
and the like	parts of	fitted with	used for
and their	of a kind	resulting	suitable for
suitable for	furnished by	thereof	for
preserved by	therefor	their	based on
whether	consisting solely	parts	manufactured from
therefor	general	related	obtained from
in other	substitutes	other articles	derived from
and similar	activities of	class	made from
made by	activities		from
otherwise			primarily with
comprising			used with
presented			with
			through
			which are
			in the
			containing
			incorporating
			involving
			on a
			by
			via

Table 114: Overview of terms defined for tagset extraction rule-set (author's representation; spaces included intentionally)

```java
private static List<String>[] getGazeteerValuesForLabel(String label) {
    @SuppressWarnings("unchecked")
    List<String>[] result = new List[2];
    // Keywords
    List<String> allKeywords = new ArrayList<String>();
    List<String> keywords = new ArrayList<String>();
    List<String> cleanedKeyords = new ArrayList<String>();
    String prefix = null;
    String suffix = null;
    // Specifications
    List<String> allSpecifications = new ArrayList<String>();
    List<String> specifications = new ArrayList<String>();
    List<String> cleanedSpecifications = new ArrayList<String>();

    // CLEAN LABEL
    if (label.contains("[")) {
        label = label.substring(0, label.indexOf("["));
    }
    while (label.contains("(") && label.contains(")")) {
        label = label.substring(0, label.indexOf("(") - 1) + label.substring(label.indexOf(")") + 1, label.length());
    }
    label = label.trim();
    label = label.toLowerCase();

    // Split along ";" to form segments
    String[] labels = label.split(Pattern.quote(";"));
    for (int i = 0; i < labels.length; i++) {
        prefix = null;
        suffix = null;
        label = labels[i];
        label = label.trim();
        keywords.clear();

        // DELETE TOO SPECIFIC OR TOO GENERAL SEGMENTS
        if ((label.contains("used solely") || label.contains("n.e.c.")) ||
                (label.startsWith("other") || label.endsWith("other"))) {
            label = "";
        }
```

```java
// CHECK SPECIFICATIONS
// Check if trailing, detailing word after 1 comma and delete if
if (label.contains(",") && label.indexOf(",") == label.lastIndexOf(",")) {
    if (label.split(Pattern.quote(",")).length > 1) {
        String secondTerm = label.split(Pattern.quote(","))[1];
        if (!secondTerm.contains("and") && !secondTerm.contains("other") && (!secondTerm.contains("or") ||
                secondTerm.contains("or in"))) {
            specifications.add(secondTerm);
            label = TextUtils.removeStartingWithStringInclTrailingSpace(label, ",", 0);
        }
    }
}
// Get other specifications
for (String phrase : specificationPhrases) {
    label = TextUtils.addToListStartingWithExclString(label, phrase, specifications);
}
// Get specification exemptions
if (label.startsWith("parts for") || label.startsWith("parts of")) {
    specifications.add(label.substring("parts for".length()));
    specifications.add(label.substring("parts of".length()));
    label = "parts";
}
if (label.contains("used") && label.contains("in") && label.indexOf("used") < label.indexOf("in") && !label.contains("not
    in")) {
    label = TextUtils.addToListStartingWithExclString(label, " in", specifications);
} else {
    label = TextUtils.removeStartingWithStringInclTrailingSpace(label, " in", 0);
}

// REMOVE UNNECESSARY WORDS FROM SEGMENTS
// Cut parts from keyword segments according to cut phrases
for (String phrase : cutPhrases) {
    label = TextUtils.removeStartingWithStringInclTrailingSpace(label, " " + phrase, 0);
}
// Remove strings according to delete phrases
for (String phrase : deletePhrases) {
    label = TextUtils.removeSubString(label, phrase);
}
// Special handling of manufacturing and services
if (label.contains("manufacturing services")) {
    label = TextUtils.removeSubString(label, "manufacturing services");
    specifications.add("services");
} else if (!label.startsWith("service") && !label.startsWith("Public administrative service")) {
    label = TextUtils.removeSubString(label, "services");
}

// PREPARE PARTS
// Process on part strings -> initialize
keywords.add(label.trim());
// Find keywords
keywords = TextUtils.splitAtPattern(keywords, " / ");
keywords = TextUtils.splitAtPattern(keywords, " or ");
keywords = TextUtils.splitAtPattern(keywords, ", ");
keywords = TextUtils.splitAtPattern(keywords, " and ");

// FILTER PARTS BY FILTER PHRASES
keywords = TextUtils.deleteStringsIncludingWords(keywords, filterPhrases);

// GET PRE- AND SUFFIXES
if (keywords.size() >= 2) {

    // Get prefixes
    String firstPart = keywords.get(0);
    String[] parts = firstPart.split(Pattern.quote(" "));
    String[] nextElems = keywords.get(1).split(Pattern.quote(" "));
    if (nextElems.length <= 1) {
        if (parts.length == 2) {
            // If two before "," then use first as prefix
            keywords.set(0, parts[1].trim());
            prefix = TextUtils.addToStringWithSpace(prefix, parts[0].trim());
        } else if (parts.length == 3) {
            keywords.set(0, parts[2].trim());
            prefix = TextUtils.addToStringWithSpace(prefix, parts[0].trim() + " " + parts[1].trim());
        }
    }

    // Get suffixes
    String last = keywords.get(keywords.size() - 1);
    // Check other
    boolean containedOther = false;
    for (int j = 0; j < keywords.size(); j++) {
        String keyword = keywords.get(j);
        if (keyword.contains("other")) {
            containedOther = true;
            keywords.set(j, TextUtils.removeSubString(keyword, "other"));
        }
    }
    if (!containedOther) {
        if (last != null
                && !last.equals("") && NlpUtils.containsPosTag(last, "NN") && (!NlpUtils.containsPosTag(keywords.get(0),
                "NN")
                || (NlpUtils.containsPosTag(keywords.get(0), "NN") && !keywords.get(0).endsWith("s")))) {

            String[] secondToLastElems = keywords.get(keywords.size() - 2).split(Pattern.quote(" "));
            if (secondToLastElems.length <= 1) {

                parts = last.split(Pattern.quote(" "));
                if (parts.length == 2 && parts[0].length() > 2) {
                  keywords.set(keywords.size() - 1, parts[0].trim());
                  suffix = TextUtils.addToStringWithSpace(suffix, parts[1].trim());
                } else if (parts.length == 3 && parts[1].length() > 2) {
                  keywords.set(keywords.size() - 1, parts[0].trim() + " " + parts[1].trim());
                  suffix = TextUtils.addToStringWithSpace(suffix, parts[2].trim());
                } else if (parts.length == 4 && parts[2].length() > 2) {
                  keywords.set(keywords.size() - 1, parts[0].trim() + " " + parts[1].trim() + " " + parts[2].trim());
```

```
                            suffix = TextUtils.addToStringWithSpace(suffix, parts[3].trim());
                        }
                    }
                }
            }
        }

        // REMOVE STOPWORDS
        keywords = TextUtils.deleteStopwords(keywords, stopwordList);

        // Remove all parts containing specific pos tags
        if (prefix == null && suffix == null && !keywords.contains("wholesale")) {
            keywords = NlpUtils.deleteStringsWithAllMatchingPosTag(keywords, "JJ"); // adjectives
            keywords = NlpUtils.deleteStringsWithAllMatchingPosTag(keywords, "RB"); // adverbs
            keywords = NlpUtils.deleteStringsWithAllMatchingPosTag(keywords, "VBN"); // verbs
        } else {
            // ADD PRE- AND SUFFIXES
            for (String removeword : filterPhrases) {
                prefix = TextUtils.removeSubString(prefix, removeword);
                suffix = TextUtils.removeSubString(suffix, removeword);
            }
            keywords = TextUtils.addPrefixes(keywords, prefix);
            keywords = TextUtils.addSuffixes(keywords, suffix);
        }

        // FINAL CLEANING
        // Delete values shorter or equal two characters
        keywords = TextUtils.deleteStringsShortThanIndicated(keywords, 2);
        // Remove all double blanks and trim; remove all empty
        for (String value : keywords) {
            if (value.endsWith("-"))
                value = value.substring(0, value.length() - 1);
            if (value.startsWith("-"))
                value = value.substring(1);
            value = value.replace("  ", " ");
            value = value.replace(",", "");
            value = value.trim();
            if (value != null && !value.equals("") && !value.equals(" ") && !allKeywords.contains(value)) {
                cleanedKeyords.add(value);
            }
        }

        // Add to final list
        allKeywords.addAll(cleanedKeyords);
        cleanedKeyords.clear();
    }

    // HANDLE SPECIFICATIONS
    // Remove unnecessary words in specifications
    for (int i = 0; i < specifications.size(); i++) {
        for (String phrase : cutPhrases) {
            specifications.set(i, TextUtils.removeStartingWithStringInclTrailingSpace(specifications.get(i), phrase, 1));
        }
        for (String phrase : deletePhrases) {
            specifications.set(i, TextUtils.removeSubString(specifications.get(i), phrase));
        }
    }
    // Split into parts
    specifications = TextUtils.splitAtPattern(specifications, " or ");
    specifications = TextUtils.splitAtPattern(specifications, ", ");
    specifications = TextUtils.splitAtPattern(specifications, " and ");
    specifications = TextUtils.splitAtPattern(specifications, " of ");
    // Filter parts and remove stopwords
    specifications = TextUtils.deleteStringsIncludingWords(specifications, filterPhrases);
    specifications = TextUtils.deleteStopwords(specifications, stopwordList);
    // Final cleaning
    for (String specification : specifications) {
        specification = specification.replaceAll("\\d", ""); // rem. numbers
        while (specification.contains("  ")) {
            specification = specification.replace("  ", " ");
        }
        specification = specification.replace(",", "");
        specification = specification.trim();
        if (specification != null && !specification.equals("") && !specification.equals(" ")
                && !allSpecifications.contains(specification)
                && !cleanedSpecifications.contains(specification)
                && !allKeywords.contains(specification)) {
            cleanedSpecifications.add(specification);
        }
    }
    allSpecifications.addAll(cleanedSpecifications);

    // Swap if keywords empty
    if (allKeywords.size() == 0) {
        allKeywords.addAll(allSpecifications);
        allSpecifications.clear();
    }

    result[0] = allSpecifications;
    result[1] = allKeywords;
    return result;
}
```

Note: Algorithm references word lists as depicted in Table 114.

Script 10: Java code for rule-based sector tagging approach (author's representation)

Appendix K: Turtle code of alignment ontology

```
@prefix foaf: <http://xmlns.com/foaf/0.1/> .
@prefix pext: <http://www.ontotext.com/proton/protonext#> .
@prefix geospecies: <http://rdf.geospecies.org/ont/geospecies#> .
@prefix er: <http://s.opencalais.com/1/type/er/> .
@prefix geoFeatures20040307: <http://www.mindswap.org/2003/owl/geo/geoFeatures20040307.owl#> .
@prefix s: <http://s.opencalais.com/1/> .
@prefix adms: <http://www.w3.org/ns/adms#> .
@prefix resource: <http://dbpedia.org/resource/> .
@prefix ptop: <http://www.ontotext.com/proton/protontop#> .
@prefix schema: <http://schema.org/> .
@prefix ns: <http://rdf.freebase.com/ns/> .
@prefix relators: <http://id.loc.gov/vocabulary/relators/> .
@prefix dc-term: <http://purl.org/dc/terms/> .
@prefix vann: <http://purl.org/vocab/vann/> .
@prefix cpc: <http://at.ifs.tuwien.ac.at/ontologies/pcs2owl/cpc/> .
@prefix dbpedia: <http://dbpedia.org/ontology> .
@prefix xsd: <http://www.w3.org/2001/XMLSchema#> .
@prefix xml: <http://www.w3.org/XML/1998/namespace> .
@prefix isic: <http://at.ifs.tuwien.ac.at/ontologies/pcs2owl/isic/> .
@prefix ontology: <http://dbpedia.org/ontology/> .
@prefix MappingOwl2: <http://at.ifs.tuwien.ac.at/MappingOwl/> .
@prefix error: <http://org.semanticweb.owlapi/error#> .
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix MappingOwl: <http://at.ifs.tuwien.ac.at/MappingOwl#> .
@prefix mondeca: <http://labs.mondeca.com/foaf/mondeca.rdf#> .
@prefix ontology3: <http://linkedgeodata.org/ontology/> .
@prefix : <http://www.w3.org/2002/07/owl#> .
@prefix wgs84_pos: <http://www.w3.org/2003/01/geo/wgs84_pos#> .
@prefix person: <http://data.semanticweb.org/person/> .
@prefix spatial: <http://geovocab.org/spatial#> .
@prefix Geo: <http://s.opencalais.com/1/type/er/Geo/> .
@prefix geo: <http://rdf.insee.fr/geo/> .
@prefix ontology2: <http://www.geonames.org/ontology#> .
@prefix business: <http://rdf.freebase.com/ns/business/> .
@prefix ns2: <http://creativecommons.org/ns#> .
@prefix dc: <http://purl.org/dc/elements/1.1/> .
@prefix skos: <http://www.w3.org/2004/02/skos/core#> .
@prefix schema2: <http://at.ifs.tuwien.ac.at/schema/> .
@prefix e: <http://s.opencalais.com/1/type/em/e/> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix owl: <http://www.w3.org/2002/07/owl#> .
@base <http://at.ifs.tuwien.ac.at/MappingOwl/> .
<http://at.ifs.tuwien.ac.at/MappingOwl/> rdf:type :Ontology ;
                                         :imports cpc: ,
                                                  isic: ,
                                                  <http://www.geonames.org/ontology> .
###  http://at.ifs.tuwien.ac.at/MappingOwl#DBpediaThing
MappingOwl:DBpediaThing rdf:type :Class .
###  http://at.ifs.tuwien.ac.at/MappingOwl#OpenCalaisThing
MappingOwl:OpenCalaisThing rdf:type :Class .
###  http://at.ifs.tuwien.ac.at/MappingOwl/CLIncidentDimension
MappingOwl2:CLIncidentDimension rdf:type :Class .
###  http://at.ifs.tuwien.ac.at/MappingOwl/CityDimension
MappingOwl2:CityDimension rdf:type :Class ;
                          rdfs:subClassOf MappingOwl2:Geographical .
###  http://at.ifs.tuwien.ac.at/MappingOwl/CountryDimension
MappingOwl2:CountryDimension rdf:type :Class ;
                             rdfs:subClassOf MappingOwl2:Geographical .
###  http://at.ifs.tuwien.ac.at/MappingOwl/Geographical
MappingOwl2:Geographical rdf:type :Class ;
                         rdfs:subClassOf MappingOwl2:CLIncidentDimension .
###  http://at.ifs.tuwien.ac.at/MappingOwl/OrganizationalDimension
MappingOwl2:OrganizationalDimension rdf:type :Class ;
                                    rdfs:subClassOf MappingOwl2:CLIncidentDimension .
###  http://at.ifs.tuwien.ac.at/MappingOwl/RegionalDimension
MappingOwl2:RegionalDimension rdf:type :Class ;
                              rdfs:subClassOf MappingOwl2:Geographical .
###  http://at.ifs.tuwien.ac.at/MappingOwl/SectoralDimension
MappingOwl2:SectoralDimension rdf:type :Class ;
                              rdfs:subClassOf MappingOwl2:CLIncidentDimension .
###  http://at.ifs.tuwien.ac.at/ontologies/pcs2owl/cpc/C_CPCThing-tax
cpc:C_CPCThing-tax rdfs:subClassOf MappingOwl2:SectoralDimension .
###  http://at.ifs.tuwien.ac.at/ontologies/pcs2owl/isic/C_IsicThing-tax
isic:C_IsicThing-tax rdfs:subClassOf MappingOwl2:SectoralDimension .
###  http://at.ifs.tuwien.ac.at/schema/OpenGeoNamesThing
schema2:OpenGeoNamesThing rdf:type :Class .
###  http://dbpedia.org/ontology/AdministrativeRegion
dbpedia:/AdministrativeRegion rdf:type :Class ;
                              rdfs:subClassOf dbpedia:/Region .
```

```
###  http://dbpedia.org/ontology/Agent
dbpedia:/Agent rdf:type :Class ;
               rdfs:subClassOf MappingOwl:DBpediaThing .
###  http://dbpedia.org/ontology/Airline
dbpedia:/Airline rdf:type :Class ;
                 rdfs:subClassOf dbpedia:/Company .
###  http://dbpedia.org/ontology/City
dbpedia:/City rdf:type :Class ;
              rdfs:subClassOf dbpedia:/Settlement .
###  http://dbpedia.org/ontology/CityDistrict
dbpedia:/CityDistrict rdf:type :Class ;
                      rdfs:subClassOf dbpedia:/Settlement .
###  http://dbpedia.org/ontology/ClericalAdministrativeRegion
dbpedia:/ClericalAdministrativeRegion rdf:type :Class ;

                                      rdfs:subClassOf dbpedia:/AdministrativeRegion .
###  http://dbpedia.org/ontology/Company
dbpedia:/Company rdf:type :Class ;
                 rdfs:subClassOf dbpedia:/Organisation .
###  http://dbpedia.org/ontology/Country
dbpedia:/Country rdf:type :Class ;
                 rdfs:label "Country"@en ;
                 rdfs:subClassOf MappingOwl2:CountryDimension ,
                                 dbpedia:/PopulatedPlace .
###  http://dbpedia.org/ontology/Deanery
dbpedia:/Deanery rdf:type :Class ;
                 rdfs:subClassOf dbpedia:/ClericalAdministrativeRegion .
###  http://dbpedia.org/ontology/Department
dbpedia:/Department rdf:type :Class ;
                    rdfs:subClassOf MappingOwl2:RegionalDimension ,
                                    dbpedia:/PopulatedPlace .
###  http://dbpedia.org/ontology/Diocese
dbpedia:/Diocese rdf:type :Class ;
                 rdfs:subClassOf dbpedia:/ClericalAdministrativeRegion .
###  http://dbpedia.org/ontology/District
dbpedia:/District rdf:type :Class ;
                  rdfs:subClassOf dbpedia:/GovernmentalAdministrativeRegion .
###  http://dbpedia.org/ontology/DistrictWaterBoard
dbpedia:/DistrictWaterBoard rdf:type :Class ;
                            rdfs:subClassOf dbpedia:/GovernmentalAdministrativeRegion .
###  http://dbpedia.org/ontology/GovernmentalAdministrativeRegion
dbpedia:/GovernmentalAdministrativeRegion rdf:type :Class ;
                                          rdfs:subClassOf dbpedia:/AdministrativeRegion .
###  http://dbpedia.org/ontology/IranSettlement
dbpedia:/IranSettlement rdf:type :Class ;
                        rdfs:subClassOf dbpedia:/Settlement .
###  http://dbpedia.org/ontology/LawFirm
dbpedia:/LawFirm rdf:type :Class ;
                 rdfs:subClassOf dbpedia:/Company .
###  http://dbpedia.org/ontology/Municipality
dbpedia:/Municipality rdf:type :Class ;
                      rdfs:subClassOf dbpedia:/GovernmentalAdministrativeRegion .
###  http://dbpedia.org/ontology/NaturalRegion
dbpedia:/NaturalRegion rdf:type :Class ;
                       rdfs:subClassOf dbpedia:/Region .
###  http://dbpedia.org/ontology/OldTerritory
dbpedia:/OldTerritory rdf:type :Class ;
                      rdfs:subClassOf dbpedia:/Territory .
###  http://dbpedia.org/ontology/Organisation
dbpedia:/Organisation rdf:type :Class ;
                      rdfs:subClassOf MappingOwl2:OrganizationalDimension ,
                                      dbpedia:/Agent .
###  http://dbpedia.org/ontology/OverseasDepartment
dbpedia:/OverseasDepartment rdf:type :Class ;
                            rdfs:subClassOf dbpedia:/Department .
###  http://dbpedia.org/ontology/Parish
dbpedia:/Parish rdf:type :Class ;
                rdfs:subClassOf dbpedia:/ClericalAdministrativeRegion .
###  http://dbpedia.org/ontology/Place
dbpedia:/Place rdf:type :Class ;
               rdfs:subClassOf MappingOwl:DBpediaThing .
###  http://dbpedia.org/ontology/PopulatedPlace
dbpedia:/PopulatedPlace rdf:type :Class ;
                        rdfs:subClassOf dbpedia:/Place .
###  http://dbpedia.org/ontology/Prefecture
dbpedia:/Prefecture rdf:type :Class ;
                    rdfs:subClassOf MappingOwl2:RegionalDimension ,
                                    dbpedia:/PopulatedPlace .
###  http://dbpedia.org/ontology/Province
dbpedia:/Province rdf:type :Class ;
                  rdfs:subClassOf dbpedia:/GovernmentalAdministrativeRegion .
###  http://dbpedia.org/ontology/Publisher
```

```
dbpedia:/Publisher rdf:type :Class ;
                   rdfs:subClassOf dbpedia:/Company .
###  http://dbpedia.org/ontology/RecordLabel
dbpedia:/RecordLabel rdf:type :Class ;

                     rdfs:subClassOf dbpedia:/Company .
###  http://dbpedia.org/ontology/Regency
dbpedia:/Regency rdf:type :Class ;
                 rdfs:subClassOf dbpedia:/GovernmentalAdministrativeRegion .
###  http://dbpedia.org/ontology/Region
dbpedia:/Region rdf:type :Class ;
                rdfs:subClassOf MappingOwl2:RegionalDimension ,
                                dbpedia:/PopulatedPlace .
###  http://dbpedia.org/ontology/Settlement
dbpedia:/Settlement rdf:type :Class ;
                    rdfs:subClassOf MappingOwl2:CityDimension ,
                                    dbpedia:/PopulatedPlace .
###  http://dbpedia.org/ontology/State
dbpedia:/State rdf:type :Class ;
               rdfs:subClassOf MappingOwl2:RegionalDimension ,
                               dbpedia:/PopulatedPlace .
###  http://dbpedia.org/ontology/SubMunicipality
dbpedia:/SubMunicipality rdf:type :Class ;
                         rdfs:subClassOf dbpedia:/GovernmentalAdministrativeRegion .
###  http://dbpedia.org/ontology/Territory
dbpedia:/Territory rdf:type :Class ;
                   rdfs:subClassOf MappingOwl2:RegionalDimension ,
                                   dbpedia:/PopulatedPlace .
###  http://dbpedia.org/ontology/Town
dbpedia:/Town rdf:type :Class ;
              rdfs:subClassOf dbpedia:/Settlement .
###  http://dbpedia.org/ontology/Village
dbpedia:/Village rdf:type :Class ;
                 rdfs:subClassOf dbpedia:/Settlement .
###  http://geovocab.org/spatial#Feature
spatial:Feature rdfs:subClassOf schema2:OpenGeoNamesThing .
###  http://org.semanticweb.owlapi/error#Error1
error:Error1 rdfs:subClassOf schema2:OpenGeoNamesThing .
###  http://org.semanticweb.owlapi/error#Error2
error:Error2 rdfs:subClassOf schema2:OpenGeoNamesThing .
###  http://s.opencalais.com/1/IndustryTerm
s:IndustryTerm rdf:type :Class ;
               rdfs:label "Industry Term" ;
               rdfs:subClassOf s:type/em/e/MarkupEntity ;
               rdfs:comment """Extracts a description of an industry, segment, product family,
             or business.""" .
###  http://s.opencalais.com/1/type/em/e/City
s:type/em/e/City rdf:type :Class ;
                 rdfs:subClassOf MappingOwl2:CityDimension ,
                                 s:type/em/e/MarkupEntity ,
                                 s:type/em/e/Place .
###  http://s.opencalais.com/1/type/em/e/Company
s:type/em/e/Company rdf:type :Class ;
                    rdfs:subClassOf MappingOwl2:OrganizationalDimension ,
                                    s:type/em/e/MarkupEntity .
###  http://s.opencalais.com/1/type/em/e/Country
s:type/em/e/Country rdf:type :Class ;
                    rdfs:subClassOf MappingOwl2:CountryDimension ,
                                    s:type/em/e/MarkupEntity ,
                                    s:type/em/e/Place ;
                    rdfs:comment """Extracts a name of a country, normalized to the
             fullest name occurring within the document.""" .
###  http://s.opencalais.com/1/type/em/e/MarkupEntity
s:type/em/e/MarkupEntity rdf:type :Class ;
                         rdfs:subClassOf MappingOwl:OpenCalaisThing .
###  http://s.opencalais.com/1/type/em/e/Organization
s:type/em/e/Organization rdf:type :Class ;
                         rdfs:subClassOf MappingOwl2:OrganizationalDimension ,
                                         s:type/em/e/MarkupEntity .
###  http://s.opencalais.com/1/type/em/e/Place
s:type/em/e/Place rdf:type :Class ;
                  rdfs:subClassOf s:type/em/e/MarkupEntity .
###  http://s.opencalais.com/1/type/er/Company
er:Company rdf:type :Class ;

           rdfs:subClassOf MappingOwl2:OrganizationalDimension ,
                           er:ResolvedEntity .
###  http://s.opencalais.com/1/type/er/Geo
er:Geo rdf:type :Class ;

       rdfs:subClassOf er:ResolvedEntity .
###  http://s.opencalais.com/1/type/er/ResolvedEntity
```

```
er:ResolvedEntity rdf:type :Class ;
                  rdfs:subClassOf MappingOwl:OpenCalaisThing .
###  http://s.opencalais.com/1/type/er/Geo/City
er:Geo/City rdf:type :Class ;
            rdfs:subClassOf MappingOwl2:CityDimension ,
                            er:Geo .
###  http://s.opencalais.com/1/type/er/Geo/Country
er:Geo/Country rdf:type :Class ;
               rdfs:subClassOf MappingOwl2:CountryDimension ,
                               er:Geo .
###  http://s.opencalais.com/1/type/er/Geo/ProvinceOrState
er:Geo/ProvinceOrState rdf:type :Class ;
                       rdfs:subClassOf MappingOwl2:RegionalDimension ,
                                       er:Geo .
###  http://schema.org/AdministrativeArea
schema:AdministrativeArea rdf:type :Class ;
                          rdfs:label "AdministrativeArea"@en ;
                          rdfs:subClassOf MappingOwl:DBpediaThing ,
                                          MappingOwl2:RegionalDimension .
###  http://schema.org/City
schema:City rdf:type :Class ;
            rdfs:label "City"@en ;
            rdfs:subClassOf MappingOwl:DBpediaThing ,
                            MappingOwl2:CityDimension .
###  http://schema.org/Organization
schema:Organization rdf:type :Class ;
                    rdfs:subClassOf MappingOwl:DBpediaThing ,
                                    MappingOwl2:OrganizationalDimension .
###  http://schema.org/Place
schema:Place rdf:type :Class ;
             rdfs:subClassOf MappingOwl:DBpediaThing .
###  http://www.geonames.org/ontology#Feature
ontology2:Feature rdf:type :Class ;
                  rdfs:label "Feature"@en ;
                  rdfs:subClassOf schema:Place ,
                                  wgs84_pos:SpatialThing ;
                  rdfs:comment "A geographical feature"@en .
###  http://www.mindswap.org/2003/owl/geo/geoFeatures20040307.owl#GeographicFeature
geoFeatures20040307:GeographicFeature rdfs:subClassOf schema2:OpenGeoNamesThing .
###  http://www.w3.org/2003/01/geo/wgs84_pos#SpatialThing
wgs84_pos:SpatialThing rdf:type :Class ;
                       rdfs:subClassOf MappingOwl:DBpediaThing .
###  http://www.w3.org/2004/02/skos/core#Concept
skos:Concept rdfs:subClassOf schema2:OpenGeoNamesThing .
###  http://www.w3.org/2004/02/skos/core#ConceptScheme
skos:ConceptScheme rdfs:subClassOf schema2:OpenGeoNamesThing .
###  http://xmlns.com/foaf/0.1/Document
foaf:Document rdfs:subClassOf schema2:OpenGeoNamesThing .
###  http://xmlns.com/foaf/0.1/Organization
foaf:Organization rdfs:subClassOf schema2:OpenGeoNamesThing .
###  http://xmlns.com/foaf/0.1/Person
foaf:Person rdfs:subClassOf schema2:OpenGeoNamesThing .
```

Appendix L: Bayesian network implementation in Netica

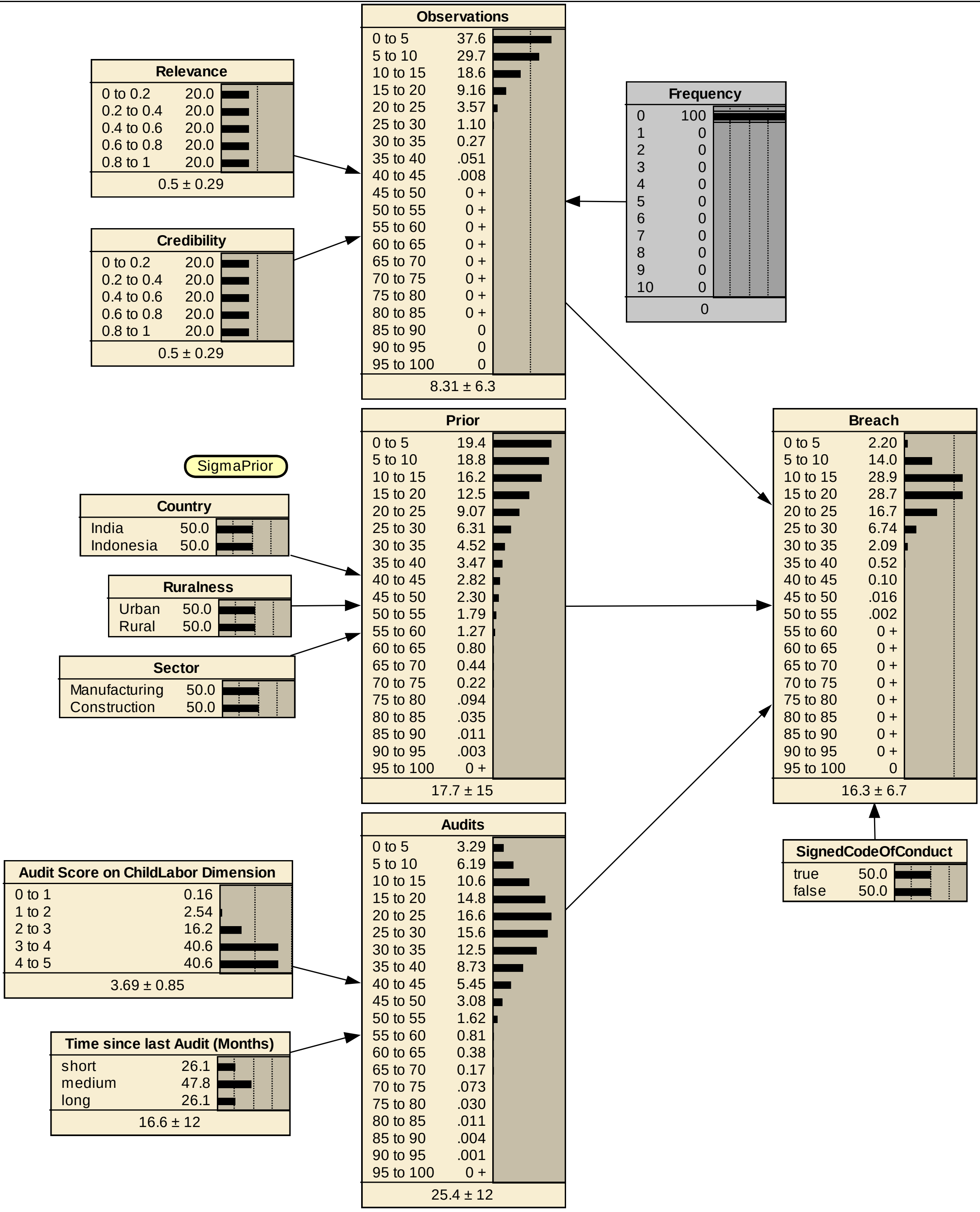


Figure 102: Full view of Bayesian risk network (author's representation)

Appendix M: Child labor distance relations and minimum distance

```
Phase: EvidenceAnnotation
Input: ChildLabor_Kind ChildLabor_Arbeit
Options: control = appelt

Rule: CL_KindBeforeArbeit

({ChildLabor_Kind}):kind
({ChildLabor_Arbeit}):arbeit
-->
{
   // Get annotations in lists
   List arbeitList = new ArrayList((AnnotationSet)bindings.get("arbeit"));
   List kindList = new ArrayList((AnnotationSet)bindings.get("kind"));

   //Get current arbeit
   Annotation currentArbeit = (Annotation)arbeitList.get(0);
   //Get current kind
   Annotation currentKind = (Annotation)kindList.get(0);

   // Get token
   AnnotationSet tokenAS = inputAS.get("Token");
   AnnotationSet tokensArbeit = tokenAS.get(currentArbeit.getStartNode().getOffset(),
      currentArbeit.getEndNode().getOffset());
   AnnotationSet tokensKind = tokenAS.get(currentKind.getStartNode().getOffset(),
      currentKind.getEndNode().getOffset());

   int distance = -1;
   if (tokensKind.firstNode().getOffset()==tokensArbeit.firstNode().getOffset() &&
      tokensArbeit.lastNode().getOffset()==tokensKind.lastNode().getOffset()) {
      distance = 0;
   } else {
      //Calculate distance
      int begOffset = tokensKind.lastNode().getOffset().intValue();
      int endOffset = tokensArbeit.firstNode().getOffset().intValue();
      distance = endOffset - begOffset;
   }

   // Add annotation
   FeatureMap features = Factory.newFeatureMap();
   features.put("distance", distance);
   features.put("rule", "CL_KindBeforeArbeit");
   try {
      outputAS.add(tokensKind.firstNode().getOffset(), tokensArbeit.lastNode().getOffset(),
      "CL_DistanceRelation", features);
   } catch (InvalidOffsetException e) {
      // Do nothing
   }

}
```

Script 11: JAPE rule for creating distance relations (author’s representation)

```
Phase: EvidenceAnnotation
Input: CL_DistanceRelation
Options: control = all

Rule: CL_MinimumDistance

({CL_DistanceRelation}):distance
-->
{
   //Get current document features and current minimum distance
   FeatureMap features = doc.getFeatures();
   Object o = features.get("CL_distance");
   //System.out.println("CurrentDistance: "+o);

   //Get distance feature for annotation found
   List foundList = new ArrayList((AnnotationSet)bindings.get("distance"));
   Annotation currentDistance = (Annotation)foundList.get(0);
   FeatureMap currentFeatures = currentDistance.getFeatures();
   Object cO = currentFeatures.get("distance");
   String currentDist = cO.toString();
   int currentCompareDistance = Integer.parseInt(currentDist);

   //Replace only if smaller or not available
   boolean replace = false;
   int replaceDistance = -1;

   if (o != null) {
      String dist = o.toString();
      int currentMinDistance = Integer.parseInt(dist);

      //Check if smaller
      if(currentCompareDistance<currentMinDistance) {
         replace = true;
         replaceDistance = currentCompareDistance;
      }

   } else {
      //Not existing
      replace = true;
      replaceDistance = currentCompareDistance;
   }

   if(replace) {
      doc.getFeatures().put("CL_distance", replaceDistance);
   }

}
```

Script 12: JAPE rule for determining the minimum distance relation’s distance (author’s representation)

Appendix N: Stopword lists

Language	Stopwords
English	childcare
German	UNO-Kinderhilfswerk; Kinderhilfswerk; UN-Kinderhilfswerk; Weltkinderhilfswerk; Betriebskindergärten; Kindergartenarbeit; Kinderhandel; Betriebskindergartens; Betriebskindergarten; Kindergärtnerberuf; Kinderbetreuungsarbeit; Arbeiterkinder; Kinderausstellung; UNO-Kinderhilfwerk; Kindervorstellung; Kindermodengeschäft; Kinderschutzarbeit; Arbeiterkind; Kinderstrumpffabrik; SOS-Kinderdorf-Arbeit

Table 115: Stopwords defined for stopword removal based on evaluation corpuses (author's representation)

Appendix O: Class diagrams of prototype

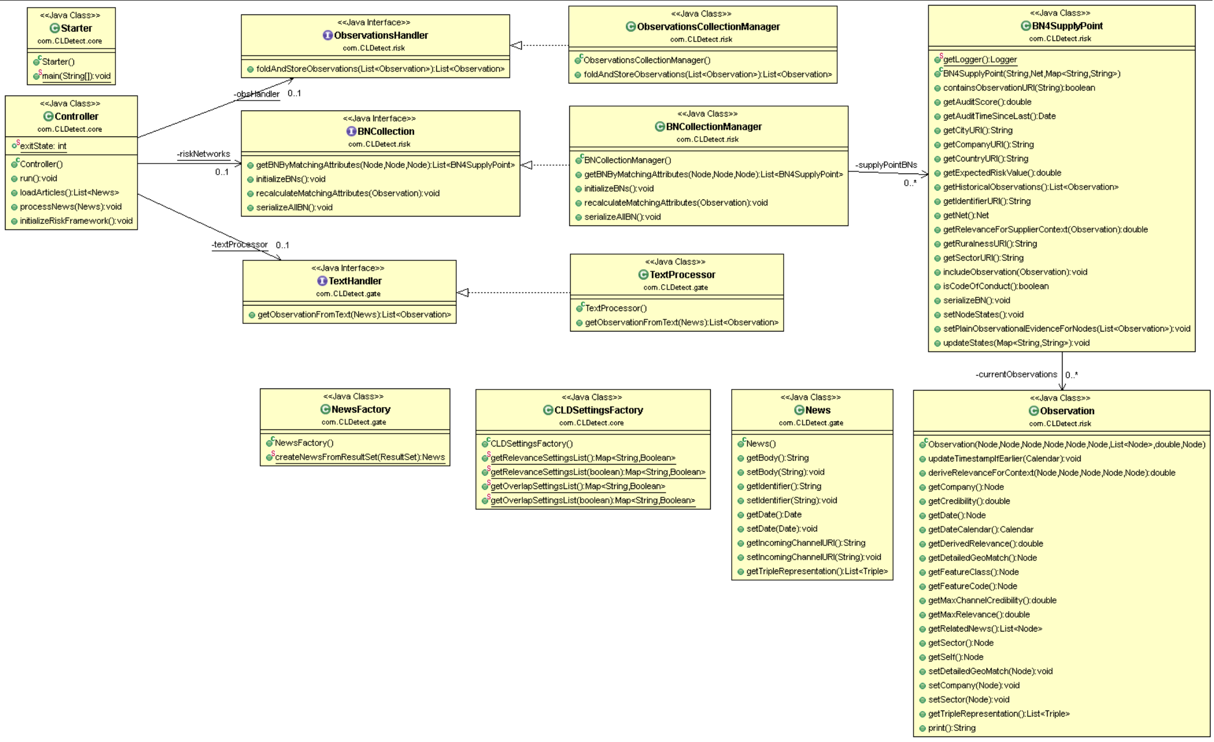


Figure 103: Classes for data interaction (author's representation; created with ObjectAid)[236]

[236] http://www.objectaid.com/.

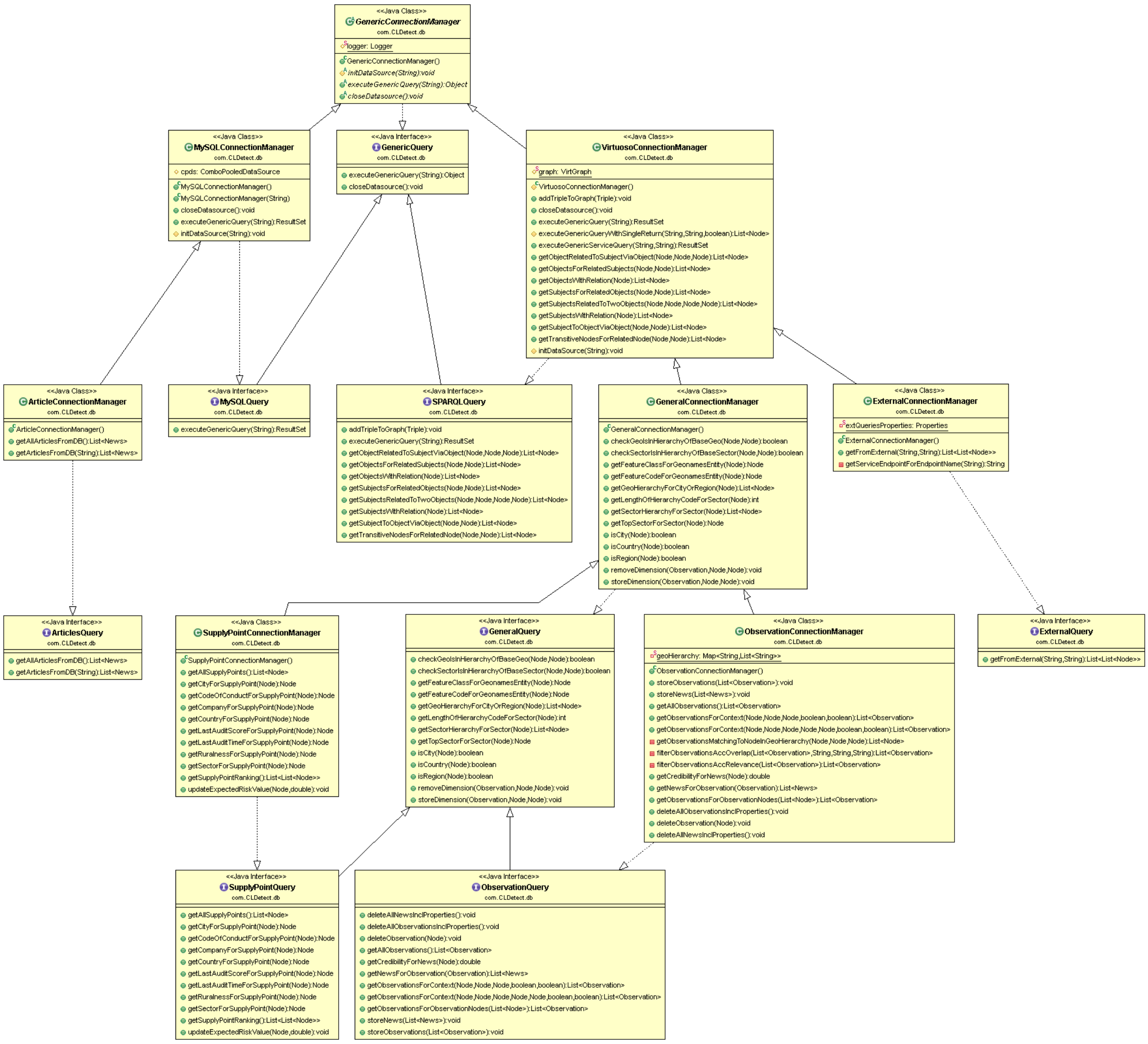


Figure 104: Classes for database access (author's representation; created with ObjectAid)[237]

[237] http://www.objectaid.com/.

Appendix P: Gold standard items including dimension tags

English – Reuters TRC2

The following articles have been marked as containing child labor – referenced by publication timestamps as provided in the Reuters TRC2 set:[238]

ID	Timestamp	Geography	Sector	Company
1	04.01.2008 20:14		Growing of beverage crops (ISIC: A0127)	
2	08.01.2008 18:34	Ivory Coast	Growing of beverage crops (ISIC: A0127)	
3	10.01.2008 17:47	China		
4	11.01.2008 04:23	Shanxi provice	Manufacturing (ISIC: C)	
5	11.01.2008 04:30	Shanxi provice	Manufacturing (ISIC: C)	
6	30.01.2008 13:46	Mozambique		
7	04.02.2008 11:05	Bangladesh, Uzbekistan	Growing of fibre crops (ISIC: A0116), Weaving of textiles (ISIC: C1312)	Tesco, Hennes & Mauritz, Marks and Spencer
8	11.02.2008 20:59	Ivory Coast	Growing of beverage crops (ISIC: A0127)	
9	13.02.2008 02:18	Australia	Raising of cattle and buffaloes (ISIC: A0141)	
10	13.02.2008 03:02	Australia	Raising of cattle and buffaloes (ISIC: A0141)	
11	13.02.2008 11:58	Ivory Coast	Growing of beverage crops (ISIC: A0127)	
12	17.02.2008 09:44	Lahore	Manufacture of plastics and synthetic rubber in primary forms (ISIC: C2013), Manufacture of paper and paper products (ISIC: C170)	
13	18.02.2008 02:14	Lahore	Manufacture of plastics and synthetic rubber in primary forms (ISIC: C2013), Manufacture of paper and paper products (ISIC: C170)	
14	27.02.2008 08:04	Yemen, Saudi Arabia	Agriculture, forestry and fishing (ISIC: A), Activities of households as employers of domestic personnel (T9700), Sports activities (R931)	
15	01.03.2008 11:27	Sierra Leone		
16	06.03.2008 16:55	Likasi	Mining and quarrying (ISIC: B)	
17	10.03.2008 00:03	Ivory Coast	Growing of beverage crops (ISIC: A0127)	
18	10.03.2008 08:03	Ivory Coast	Growing of beverage crops (ISIC: A0127)	
19	11.03.2008 06:59	Turkey		
20	14.03.2008 10:06	China	Manufacture of luggage, handbags and the like, saddlery and harness (ISIC: C1512)	Nike
21	20.03.2008 07:51	Shanxi provice, Henan province	Mining and quarrying (ISIC: B), Manufacturing (ISIC: C)	
22	07.04.2008 11:58	Niger	Agriculture, forestry and fishing (ISIC: A), Activities of households as employers of domestic personnel (T9700)	
23	07.04.2008 16:33	Niger	Agriculture, forestry and fishing (ISIC: A), Activities of households as employers of domestic personnel (T9700)	
24	08.04.2008 14:47	Germany		
25	09.04.2008 07:09	Dhaka	Retail trade, except of motor vehicles and motorcycles (ISIC: G47), Other personal service activities n.e.c. (S9609), Other credit granting (K6492)	
26	11.04.2008 02:00	Tibet		
27	11.04.2008 08:24	Thailand		
28	11.04.2008 12:27	Thailand		
29	11.04.2008 14:21	Thailand		
30	21.04.2008 12:20	Afghanistan		
31	22.04.2008 13:22	Pakistan	Manufacture of sports goods (ISIC: C3230)	Credit Suisse
32	22.04.2008 14:12		Manufacture of sports goods (ISIC: C3230)	
33	22.04.2008 15:37		Manufacture of sports goods (ISIC: C3230)	
34	22.04.2008 15:45		Manufacture of sports goods (ISIC: C3230)	
35	22.04.2008 18:31		Manufacture of sports goods (ISIC: C3230)	
36	23.04.2008 13:30	Thailand, Bangladesh	Processing and preserving of fish, crustaceans and molluscs (ISIC: C1020)	Costco, Cub Foods, Giant, Giant Eagle, Harris Teeter, IGA, Tops Markets, Trader Joe's, Wal-Mart
37	23.04.2008 18:11	Thailand, Bangladesh	Processing and preserving of fish, crustaceans and molluscs (ISIC: C1020)	
38	29.04.2008 04:25	Shanxi Province, Henan Province, Sichuan Province, Guangdong Province, Liangshan	Manufacturing (ISIC: C), Mining and quarrying (ISIC: B)	

[238] It might be possible that some IDs are not unique, however, such cases should be easily detectable.

39	30.04.2008 02:08	Sichuan Province, Guangdong Province	Manufacturing (ISIC: C)	
40	30.04.2008 04:58	Dongguan	Manufacturing (ISIC: C), Mining and quarrying (ISIC: B)	
41	30.04.2008 06:08	China		
42	30.04.2008 08:47	China		
43	30.04.2008 15:07	China		
44	01.05.2008 06:16	China		
45	01.05.2008 06:22	Sichuan Province, Guangdong Province, Shanxi Province, Henan Province	Manufacturing (ISIC: C), Mining and quarrying (ISIC: B)	
46	01.05.2008 08:09	Shipai, Dongguan	Manufacturing (ISIC: C)	
47	01.05.2008 08:55	China		
48	01.05.2008 10:19	China		
49	02.05.2008 01:16			
50	02.05.2008 01:17			
51	02.05.2008 01:52	China		
52	02.05.2008 02:18			
53	02.05.2008 03:05	China		
54	02.05.2008 06:40			
55	02.05.2008 09:14			
56	14.05.2008 01:05		Manufacture of sports goods (ISIC: C3230)	Credit Suisse
57	14.05.2008 06:05		Manufacture of sports goods (ISIC: C3230)	Credit Suisse
58	15.05.2008 06:20	Bangladesh		Ericsson, Telenor
59	20.05.2008 07:01	Bangladesh		Telenor
60	20.05.2008 10:46	Bangladesh	Manufacture of communication equipment (ISIC: C2630)	Telenor, Grameenphone
61	20.05.2008 14:33	Bangladesh	Manufacture of communication equipment (ISIC: C2630)	Telenor, Grameenphone
62	22.05.2008 15:02	Bangladesh	Manufacture of communication equipment (ISIC: C2630)	Telenor, Grameenphone
63	30.05.2008 16:05	Mayanmar	construction (ISIC: F)	
64	02.06.2008 06:46	India		
65	04.06.2008 19:16	-		
66	09.06.2008 16:39	United States	Agriculture, forestry and fishing (ISIC: A)	
67	10.06.2008 20:31		Growing of beverage crops (ISIC: A0127)	
68	17.06.2008 02:16	Tirupur	Manufacture of textiles (ISIC: C13)	Primark, Associated British Foods
69	24.06.2008 04:38	Sri Lanka	Manufacture of textiles (ISIC: C13)	
70	26.08.2008 04:35	Sri Lanka	Retail trade, except of motor vehicles and motorcycles (ISIC: G47)	
71	05.09.2008 09:12	Bangladesh	Telecommunications (ISIC: J61)	Telenor, Grameenphone
72	05.09.2008 16:17			Grameenphone, Telenor
73	05.09.2008 21:00	Bangladesh		Grameenphone, Telenor
74	09.09.2008 10:26	India	Growing of fibre crops (ISIC: A0116)	Monsanto
75	09.09.2008 11:32	India	Growing of fibre crops (ISIC: A0116)	Monsanto
76	09.09.2008 17:15	India	Growing of fibre crops (ISIC: A0116)	Monsanto
77	09.09.2008 20:38	Postville	Processing and preserving of meat (ISIC: C1010)	Agriprocessors Inc
78	09.09.2008 21:40	India	Growing of fibre crops (ISIC: A0116)	Monsanto
79	09.09.2008 21:41	India	Growing of fibre crops (ISIC: A0116)	Monsanto
80	11.09.2008 16:42	Ivory Coast	Growing of beverage crops (ISIC: A0127)	
81	12.09.2008 13:12	Ivory Coast	Growing of beverage crops (ISIC: A0127)	
82	30.09.2008 13:08	Uzbekistan	Growing of fibre crops (ISIC: A0116)	Wal-Mart Stores
83	30.09.2008 14:49	Uzbekistan	Growing of fibre crops (ISIC: A0116)	Wal-Mart Stores, Hennes & Mauritz
84	13.10.2008 11:13	Bangladesh		Telenor
85	13.10.2008 11:59	Bangladesh		Telenor
86	16.10.2008 06:09	Belitung	Manufacturing (ISIC: C)	
87	18.10.2008 07:55	Bangladesh, Uzbekistan	Growing of fibre crops (ISIC: A0116), Weaving of textiles (ISIC: C1312)	Wal-Mart, Tesco, Hennes & Mauritz, JC Penney, Marks and Spencer
88	22.10.2008 08:27	China	Manufacturing (ISIC: C)	Wal-Mart Stores
89	22.10.2008 08:42	China	Manufacturing (ISIC: C)	Wal-Mart Stores
90	22.10.2008 09:42	China	Manufacturing (ISIC: C)	Wal-Mart Stores
91	23.10.2008 04:07	India	Other manufacturing n.e.c. (ISIC: C3290)	
92	23.10.2008 04:27	Deeg	Other manufacturing n.e.c. (ISIC: C3290)	
93	23.10.2008 06:02	Deeg	Other manufacturing n.e.c. (ISIC: C3290)	
94	23.10.2008 07:24	Deeg	Other manufacturing n.e.c. (ISIC: C3290)	
95	23.10.2008 09:15	New Delhi	Other manufacturing n.e.c. (ISIC: C3290)	
96	23.10.2008 10:09	India	Other manufacturing n.e.c. (ISIC: C3290)	
97	23.10.2008 10:13	New Delhi	Other manufacturing n.e.c. (ISIC: C3290)	
98	23.10.2008 10:16	New Delhi	Other manufacturing n.e.c. (ISIC: C3290)	
99	23.10.2008 12:12	Deeg	Other manufacturing n.e.c. (ISIC: C3290)	
100	30.10.2008 11:11	Bangladesh		

101	15.11.2008 10:01	Bangladesh		
102	19.11.2008 09:14	Nepal		
103	08.01.2009 21:59	Costa Rica	Other retail sale not in stores, stalls or markets (ISIC: G4799)	
104	08.01.2009 22:39	Costa Rica	Other retail sale not in stores, stalls or markets (ISIC: G4799)	
105	09.01.2009 18:43	Costa Rica	Other retail sale not in stores, stalls or markets (ISIC: G4799)	
106	09.01.2009 21:33	Costa Rica	Other retail sale not in stores, stalls or markets (ISIC: G4799)	
107	10.01.2009 02:08	Costa Rica	Other retail sale not in stores, stalls or markets (ISIC: G4799)	
108	14.01.2009 02:54	Shanxi provice, Henan province, Guangdong province	Mining and quarrying (ISIC: B), Manufacturing (ISIC: C)	
109	02.02.2009 06:35	North Korea		
110	09.02.2009 17:44	China		
111	09.02.2009 18:08	China		
112	10.02.2009 02:49	Indonesia		
113	12.02.2009 17:00		Fishing (ISIC: A031), Finishing of textiles (ISIC: C1313), Growing of beverage crops (ISIC: A0127)	
114	17.02.2009 17:19	Angola		
115	20.02.2009 13:30			
116	23.02.2009 17:41	India		Ikea
117	25.02.2009 18:38	South Africa		

Table 116: Overview of TRC2 child labor gold standard including dimensions (author's representation)

German – APA

The following articles have been marked as containing child labor – referenced by APA IDs:

APA_20080121_APA0429, APA_20080802_APA0140, APA_20080806_APA0675, APA_20080819_APA0121, APA_20080822_APA0533, APA_20080828_APA0223, APA_20080919_APA0667, APA_20080930_APA0357, APA_20081017_APA0159, APA_20081017_APA0160, APA_20081023_APA0041, APA_20081023_APA0376, APA_20081028_APA0013, APA_20081030_APA0063, APA_20081111_APA0304, APA_20081125_APA0017, APA_20081202_APA0532, APA_20081229_APA0400, APA_20080312_APA0374, APA_20080917_APA0296, APA_20080929_APA0461, APA_20080401_APA0248, APA_20080711_APA0083, APA_20080418_APA0449, APA_20080422_APA0660, APA_20080422_APA0710, APA_20080501_APA0295, APA_20080507_APA0720, APA_20080514_APA0503, APA_20080613_APA0515, APA_20080615_APA0209, APA_20080718_APA0182

Appendix Q: Augmented precision and recall

In the following the augmented precision and recall metrics based on lexical precision will be detailed as it has been introduced by Dellschaft and Staab (2006). The approach is split into a local and a global measure. While the former handles the comparison of two concepts, the latter allows comparing two hierarchies. For the local level a formula will be used that includes the semantic context in the hierarchy:

Building on the work of Maedche and Staab (2002, 255), Dellschaft and Staab reuse the concept of semantic cotopy sc to derive a measure for local taxonomic precision tp_{sc}. Basically the semantic cotopy includes all super- and subtypes of a concept including the concept itself. It is used to calculate the alternative measure introduced in the main text (Dellschaft and Staab 2006, 233).

$O_1, O_2 \ldots$	Ontologies 1 and 2
$c_1, c_2, c_1 \in O_1, c_2 \in O_2 \ldots$	Concepts 1 and 2 as parts of the respective ontologies

$$sc(c, O) \coloneqq \{c_i | c_i \in C \wedge (c_i \leq c \vee c \leq c_i)\}$$

$$tp_{sc}(c_1, c_2, O_1, O_2) \coloneqq \frac{|sc(c_1, O_1) \cap sc(c_2, O_2)|}{|sc(c_1, O_1)|}$$

The global construct for taxonomic precision is then defined as the following if full lexical details are included:

$O_G, O_R \ldots$	Ontologies ($G \cdots$Gold standard, $R \ldots$Retrieved)
$c_1, c_2, c_1 \in O_G, c_2 \in O_R \ldots$	Concepts 1 and 2 as parts of the respective ontologies

$$TP_{sc}(O_R, O_G) \coloneqq \frac{1}{|O_R|} \sum_{c \in C_R} \begin{cases} tp_{sc}(c, c, O_R, O_G), & if\ c \in C_G \\ 0, & if\ c \notin C_G \end{cases}$$

$$TR_{sc}(O_R, O_G) \coloneqq TP(O_G, O_R)$$

The resulting F-measure is then calculated with the following formula:

$$TF_{sc}(O_R, O_G) \coloneqq \frac{2 * TP_{sc}(O_R, O_G) * TR_{sc}(O_R, O_G)}{TP_{sc}(O_R, O_G) + TR_{sc}(O_R, O_G)}$$

Appendix R: Structure of expert survey

Nr	Group	Question	Scale
A	Demo-graphics	Please respond to the following questions to allow a demographic classification of your response.	-
A.1		Do you believe that child labor should be eliminated in a supply chain?	Strongly agree; Agree; Neither agree nor disagree; Disagree; Strongly disagree
A.2		What is the size of your company?	<10 employees; <50 employees; <250 employees; <1000 employees; >=1000 employees
A.3		Which industry is the main focus of your company?	<ISIC industry codes level 2>
A.4		Does your company have customers within the European Union?	Yes; No
A.5		Is your company sourcing from international locations outside the European Union?	Yes; No
A.6		Where is the main location of your personal workplace (i.e. your personal work place/"home office")?	<country list>
A.7		Do you have professional experience with supply chain, risk, or sustainability management?	Yes; No
A.8		How long do you have experience in the area of your position?	<1 year; <5 years; >= 5 years
B	Effects	In the following a theoretical case study is presented, partly based on real, practical data. Names have been disguised in order to focus on the data provided. The first case presents a Europe-based company which has four suppliers. You will be asked to rank the suppliers according to the probability of having a child labor incident at a supplier location based on your professional experience, but out of your company's context. In the second case only one of the suppliers will be considered for simplification and you will receive additional information from external news sources. You will then be asked regarding the influence of each report revealed. Finally, independent of the former questions, some questions on auditing will be asked.	-
B.1'		1. Initial supplier ranking Your case company (size undisclosed) is based in Europe and imports several hundred products from around the world. Main sourcing locations include Blueland and Pinkland. Suppliers to your company do not only produce goods that are used by your company, but are also used for services in an international context. No audit has been performed yet. All four suppliers have signed a code of conduct with your company that prohibits them from employing children. Each supplier has only one production facility at the headquarter's (HQ) location for simplification. The suppliers show different characteristics. Background-Table: Child labor and company numbers[239] - see Table 118	-
B.1''		Please rank the suppliers according to their risk of having a child labor incident based on the introductory data provided: All your answers must be different. Please number each box in order of preference from 1 to 4	<Ranking list>
B.2'		2. External data from news and social media[240] In the following only Supplier B will be considered. You will be presented several news reports that reveal information that may relate to the supplier. You will be asked how the new report would affect your rating of Supplier B with regard to the probability of a child labor incident (Pinkland, Pinkregion, Pinkcity, urban, Manufacturing):	
B.2.1		2.1 First article received online through classic news channel Greencompany's supplier in CHILD STAFF SHAME: Factory used 'underage workers' - report Greencompany will do 30 per cent less business with a Pinkcountry's parts maker reportedly caught with underage workers at its factory"	Extremely influential; Very influential; Somewhat influential; Slightly influential; Not at all influential

[239] Here and in the following, additional color codes have been used for easier visual differentiation of suppliers, geographies, and sectors.

[240] The actual survey data has been presented with visual emphasis and color coding.

		despite the supplier's promises to comply with labour laws in future. Greencompany suspended business with Supplier B when US charity Yelloworg claimed it had found at least five child workers without contracts at the supplier. Pinkcountry's authorities discovered that Supplier B was not directly responsible for employing the children, finding that a subcontractor had hired them through a labour agency. As a result, Greencompany said it would resume business with Supplier B, albeit about a third less than before. [...] How much does the news report influence your perceived probability of a child labor incident at Supplier B:	
B.2.2		2.2 Second article received online through classic news channel Bluecountry, ILO boost co-op The government of Pinkcountry and the International Labor Organization (ILO, the UN specialized agency) have signed a program of technical cooperation for 2014-2016. The document entitled 'Decent Work Agenda' was signed by the Minister of Labour and Social Protection, the chairman of the Chamber of Commerce and the chairman of the Federation of Trade Unions. [...] Under the program, the ILO will support the implementation of the national action plan on elimination of child labor. The program also aims to improve working conditions and employment in agriculture. [...] How much does the news report influence your perceived probability of a child labor incident at Supplier B:	Extremely influential; Very influential; Somewhat influential; Slightly influential; Not at all influential
B.2.3		2.3 Third article received online through classic news channel Over 10,000 kids working in Pinkregion As per the statistics, of the 10,009 child workers in Pinkregion, 4,738 children are available in the job market "seeking work". Pinkregion may pride itself for its high human development indicators, but the latest statistics of Census 2011 released on Wednesday do not bear good news on this front. A total of 10,009 children in Pinkregion, aged between 5 and 14 are employed and shockingly 6,920 of these children are working as 'main workers' or have been employed constantly over a prolonged period of time. Despite implementation of the Right of Child to Free and Compulsory Education (RTE) Act, which makes schooling mandatory for children up to the age of 14, there are 5,351 male children and 4,658 female children who have seemingly become victims of child labour in Pinkregion. [...] Of these, around 1,300 have worked as manual workers. [...] Experts in the field of economics, urban development and statistics present at the session analyzing the latest Census data, said that though statistics are unable to reveal the nature of employment, generally a majority of these children appear to be employed as domestic help in households, as waiters in small eateries and as help along with parents in manual labour jobs such as manufacturing. How much does the news report influence your perceived probability of a child labor incident at Supplier B:	Extremely influential; Very influential; Somewhat influential; Slightly influential; Not at all influential
B.2.4		2.4 Fourth article received online through social media posting Six minors rescued NGO Greyorg's staff have rescued six minors from a State Transport bus in Greencity. They were being trafficked to Pinkcity to work as child labourers, an official said Tuesday. A man who was taking the children along was arrested and is being interrogated. The children, in the age group of 12 to 16 years, were sent to a shelter Monday night after the Child Welfare Committee (CWC) chairperson visited them and requested the police to take appropriate action. The coordinator of Greyorg's special anti-trafficking project in Greencity said he and his colleague saw a middle-aged man with six children boarding a bus at Greencity's district. [...] Children from the state are trafficked to Pinkcity to work in construction. Though the situation has improved over the years, cases of trafficking still surface. How much does the news report influence your perceived probability of a child labor incident at Supplier B:	Extremely influential; Very influential; Somewhat influential; Slightly influential; Not at all influential
B.3		3. Internal data from audits Which probability of having a child labor incident (if only audit data is taken into consideration) would you associate with a random supplier reaching either a mininimum (worst), a medium, or a maximum (best) audit score? [Percent]	Exemplary minimum score of 1 (worst) leading to ... Exemplary medium score of 3 leading to ... Exemplary maximum score of 5 (best) lead-

			ing to ...
B.4		Which weight would you give the following three probabilities if they are combined in order to calclate an overall probability of a child labor incident at a supplier location? The sum should equal to 100%:	Probability based on statistical data (e.g. child labor statistics) Probability based on audit data (e.g. last audit result/time) Probability based on evidence from external reports (e.g. news)
C	Usefulness	-	-
C.1		Do you agree that the following requirements should be fulfilled by a supply chain child labor risk management system that ranks suppliers with regard to their risk level of breaching child labor standards (e.g. for dispatching audits)? The system may be used as an input for further risk management process steps. <see chapter 6 for requirements listed>	strongly disagree; moderately disagree; somewhat disagree; neutral; somewhat agree; moderately agree; strongly agree
C.2		Please answer the questions below. "Such a system" refers to a supply chain child labor risk management system fulfilling all requirements listed above: - In my job, usage of such a system is important - In my job, usage of such a system is relevant - In my job, usage of such a system would reduce the risk of - supply chain surprises - In my job, usage of such a system would reduce the risk of - biased supply chain decisions - I have no difficulty telling others about the results of using such - a system - I believe I could communicate to others the consequences of - using such a system - The results of using such a system are apparent to me - I would have difficulty explaining why using such a system may - or may not be beneficial	strongly disagree; moderately disagree; somewhat disagree; neutral; somewhat agree; moderately agree; strongly agree
C.3		Which of the following two statements is more important according to your experience: An IT system for supply chain sustainability risk management that automatically includes evidence from news should… - Maximize recall, thus, not miss any potential news about a possibly relevant sustainability incident with the risk of including irrelevant data - Maximize precision, thus, focus on reducing the risk of including irrelevant data with the risk of missing news about an incident Maximize precision, thus, focus on reducing the risk of including irrelevant data with the risk of missing news about an incident	(1) or (2)

Table 117: Complete survey structure of expert questionnaire for risk model evaluation (author's representation)

Supplier	Country	Rural/ Urban	Sector	Est. child labor rate per country and rural/urban [%]*	Sector's share of overall child labor per country and rural/urban** [%]	Est. # of children working per country, rural/urban and sector [absolute]*	Est. # of companies in sector per country and rural/urban**
Supplier A	Blueland	rural	Construction	4.9	0.1	1,475	95,226
Supplier B	Pinkland	urban	Manufacturing	0.5	34.2	71,435	1,808,369
Supplier C	Blueland	urban	Construction	1.9	0	0	85,093
Supplier D	Blueland	urban	Manufacturing	1.9	17.6	51,103	1,027,252

Table 118: Table with fictional supplier related data used in questionnaire (author's representation)

Appendix S: Invitation email for expert questionnaire

Dear Ladies and Gentlemen,

We are reaching out to you **to understand and better support how experts handle child labor risks in supply chains**. This is part of a dissertation's scientific effort at the **Vienna University of Technology.** The dissertation tries to **develop an information system to facilitate the ongoing monitoring of social sustainability issues in supply chains** and is focused on **child labor** specifically.

We would be grateful if you could fill in the survey attached, **allowing us to evaluate the system proposed**. The survey's target audience is **supply chain managers, risk managers, and sustainability managers (or quality managers if concerned with comparable topics)**.

Your data will be processed anonymously, only including basic demographic information of your company and yourself.

Completing the questionnaire takes roughly 20 minutes in total. **As a service in return** we would like to offer **an executive summary** of the results. If you wish to obtain this summary, could you please write a brief email to the email address stated below.
It would be great if you could answer the questions **until November 21st, 2014**.

We highly appreciate your effort and would like to thank you very much for your help in our scientific work! And, we hope that you could also benefit from answering the questions.

Click here to do the survey: {SURVEYURL}

With kindest regards,
Andreas Thöni

Supervised by:
o.Univ.Prof. Dr. A Min Tjoa (Vienna University of Technology)
o.Univ.Prof. Dr. Alfred Taudes (Vienna University of Economics and Business)

Mag. Dipl.-Ing. Andreas Thöni
PhD Candidate
Information & Software Engineering Group
Institute of Software Technology and Interactive Systems
Vienna University of Technology
Favoritenstrasse 9-11/188
Email: <left out for publication>
Phone: <left out for publication>

Appendix T: Raw data of expert questionnaire (usefulness)

	strongly disagree	moderat-ely disagr.	somewhat disagree	neutral	somewhat agree	moderate-ly agree	strongly agree
Use of public statistical data sources (e.g. child labor statistics)in order to calculate an initial incident probability for a supplier location	0	2	1	1	8	8	8
Differentiate initial probability by country, sector and rural/urban	0	1	0	1	8	11	7
Automated calculation of the initial probabil-ity (i.e. prior probability) of having at least one child working at a random company	0	1	2	4	12	6	3
Automated use of publicly available news reports on child labor incidents (e.g. via internet) in order to update the initial proba-bility	0	0	2	2	9	10	5
Automated use of publicly available non-news reports on child labor incidents (e.g. from social media sites as Twitter)in order to update the initial probability	0	1	5	5	9	4	4
Adjustment of initial probability depending on the specifity in which an incident report refers to a specific supplier location (i.e. relevance)	0	2	3	3	8	7	5
Adjustment of initial probability depending on the credibility of the channel through which an incident report has been received (e.g. news or social media)	0	1	3	0	9	6	9
Automated update of initially calculated probability without user-involvement (i.e. unbiased)	0	4	5	4	6	6	3
Use of last auditing result to update the initial probability	0	1	1	2	5	6	13
Use of whether Code of Conduct has been signed to adjust the initial probability values	2	1	2	7	5	4	7
Risk model calculates the probability for the risk of a breach of child labor standards for each supplier locations	0	3	3	1	11	6	4
System output (risk rank/score) should be integrateable into further supplier evaluation processes	1	2	1	0	5	9	10
System output (risk rank/score) should be unbiased with regard to different suppliers	0	2	2	2	5	11	6

Table 119: Frequencies for question: "Do you agree that the following requirements should be fulfilled by a supply chain child labor risk management system that ranks suppliers with regard to their risk level of breaching child labor standards (e.g. for dispatching audits)- The system may be used as an input for further risk management process steps." (author's representation)

	strongly disagree	moderat-ely disagr.	somewhat disagree	neutral	somewhat agree	moderate-ly agree	strongly agree
In my job, usage of such a system is im-portant	2	3	3	3	7	3	7
In my job, usage of such a system is rele-vant	1	4	1	3	7	6	6
In my job, usage of such a system would reduce the risk of supply chain surprises	0	2	1	4	10	4	7
In my job, usage of such a system would reduce the risk of biased supply chain decisions	0	1	0	6	7	10	4

Table 120: Frequencies for question: "Please answer the questions below. "Such a system" refers to a supply chain child labor risk management system fulfilling all requirements listed above:"

	strongly disagree	moderat-ely disagr.	somewhat disagree	neutral	somewhat agree	moderate-ly agree	strongly agree
I have no difficulty telling others about the results of using such a system	0	1	0	3	7	7	10
I believe I could communicate to others the consequences of using such a system	0	0	1	3	6	10	8
The results of using such a system are apparent to me	0	1	1	3	8	9	6
I would have difficulty explaining why using such a system may or may not be beneficial	5	11	3	4	3	1	1

Table 121: Frequencies for question: “Please answer the questions below. "Such a system" refers to a supply chain child labor risk management system fulfilling all requirements listed above:”

Curriculum vitae

Mag.rer.soc.oec. Dipl.-Ing. Andreas Thöni, BSc

Date and Place of Birth 05.07.1985, Rum
Military Service High Mountains Division 6, StbB 6, 6. JgBrig, Absam

School Education

1991-1995 Elementary School Seefeld/Tyrol

1995-2003 Bundesrealgymnasium Adolf-Pichler-Platz (6020 Innsbruck)
A Levels: with distinction

University Education

2012-... Doctorate in Business Informatics at the Vienna University of Technology

2007-2010 Master of Science in Business Informatics
at the Vienna University of Technology
Graduation: Dipl.-Ing. (aquivalent to MSc; with distinction)
Thesis: „Die Beteiligung der Unternehmensleitung an der IT-Governance - Eine empirische Untersuchung von Unterschieden in österreichischen Unternehmen“ (Grade: excellent)

2005-2010 Diploma Study of International Business
at the Vienna University of Economics and Business Administration
Specialisation No. 1: International Transportation and Logistics
Specialisation No. 2: Strategic Management and Controlling
Graduation: Mag.rer.soc.oec. (with distinction)
Thesis: „Betriebswirtschaftliche Optimierung in mehrfachverbundenen Multi-Hub Netzwerken am Beispiel europäischer Airlines“
(Grade: excellent)

2009 Participation at the ATHENS-Programm at the TELECOM Paris Tech

2008 Exchange Semester at the Hong Kong University of Science and Technology (Business School)

2004-2007 Bachelor of Science in Business Informatics
at the Vienna University of Technology
Specialisation: Software Quality Assurance
Graduation: BSc (with distinction)

Internships/Jobs

- ▪▫ Constultant (Fellow Senior Associate) at
  McKinsey & Company, Inc. (on educational leave since 10/2012)
  Business Technology Office
  Since October 2010
- ▪▫ Fellow Intern at **McKinsey & Company, Inc.**
  Business Technology Office
  July/August 2009
- ▪▫ Junior Consultant/Internship **Deloitte Consulting** GmbH,
  Deloitte Consulting/CFO Services
  July/August 2008
- ▪▫ Internship **Austrian Airlines AG**,
  Corporate Project & IT-Services/ITP
  July/August 2007
- ▪▫ Internship **Siemens Austria AG**,
  Program- und System Development (PSE)/SMC
  July/August 2006
- ▪▫ Internship **Siemens Austria AG**,
  Programm- und System Development (PSE)/SMC
  July 2005
- ▪▫ Internship **Tyrolean Airways** Tiroler Luftfahrt
  GmbH; Dispatch, IT
  3.5.2004-31.8.2004
- ▪▫ Ski Instructor **Ski School Seefeld in Tirol**
  Saisons 2001/02, 2002/03, 2004/05, 2005/06
  (Christmas and Mid-Semester Breaks)

Member of

- ▪▫ Austrian Computer Society (OCG)
- ▪▫ Uni Management Club Wien Alumni (UNIMC Alumni)
- ▪▫ Global Advancement Program Alumni (GAP Alumni)

Andreas Thöni m.p.
(01.06.2015)