Towards automatically generating supply chain maps from natural language text

Pascal Wichmann

Department of Engineering
University of Cambridge

This dissertation is submitted for the degree of
Doctor of Philosophy

Wolfson College May 2021

Declaration

I hereby declare that except where specific reference is made to the work of others, the contents of this dissertation are original and have not been submitted in whole or in part for consideration for any other degree or qualification in this, or any other university. This dissertation is my own work and contains nothing which is the outcome of work done in collaboration with others, except as specified in the text and Acknowledgements. This dissertation contains fewer than 65,000 words including appendices, bibliography, footnotes, tables and equations and has fewer than 150 figures.

Pascal Wichmann
May 2021

Towards automatically generating supply chain maps from natural language text

Pascal Wichmann

Supply chains have become increasingly global, complex, and multi-tiered. Consequently, companies have been gradually losing visibility of their supply network topology, that is, the structure formed by supply chain participants and their inter-dependencies across multiple tiers. This is a problem for firms as information about their extended supply chains is a valuable input when making supply chain decisions, such as how to improve their efficiency, resilience, or sustainability. Supply chain mapping, whereby information about the supply network structure is collected and visualised, is often cited in the literature to be key to addressing the problem. Yet the challenge of acquiring the necessary information remains largely unaddressed.

This thesis aims to tackle this challenge by presenting an approach for automatically generating basic supply chain maps from unstructured, natural language text by using machine learning methods. The focus of this research is on automatically extracting individual buyer-supplier ("who supplies whom") relations, a pre-requisite for automating the creation of supply chain maps from text. Such text might be sourced from openly available documents, such as news articles obtained from the Web, or alternatively from privately acquired documents. This thesis focusses mainly on the former although the results provided apply equally to the latter. A classifier for buyer-supplier relations was obtained in two steps: Firstly, a reference dataset ("corpus") was created by having human annotators assign a "label" to each pair of organisational named entities in each given sentence of the dataset. Labels indicate the type of relation between these two organisations expressed in this sentence. Secondly, a classifier was designed and trained on the dataset. A selection of different classifier architectures were tested and compared against each other. A further part of this research extends the scope from extracting individual buyer-supplier relations to the end-to-end processing pipeline in which a collection of text documents is converted into a basic supply chain map. The end-to-end approach, including a pre-trained classifier, was validated on a large, unlabelled and previously unseen real-world dataset.

The approach proposed in this study shall be understood as a first step towards the vision of automating supply chain mapping from text. It is not yet an equivalent substitute for manual research. It could, however, provide an initial (partial) supply chain map or help with checking existing supply chain maps for completeness. The effectiveness of any automated approach to supply chain mapping is clearly dependent on obtaining sufficiently rich data to work with in the first place. For example, in an experiment using openly available Web data

only, it was possible to extract 229 distinct Boeing suppliers. Assuming a total number of 13,000 Boeing suppliers, this would correspond to 1.8% of Boeing's suppliers, albeit some of the key ones.

A central contribution of this work is a method to automatically extract individual buyer-supplier relations from text. Using the proposed method, questions regarding the achievable level of agreement among the human annotators ("inter-annotator agreement"), the classification performance on the reference dataset, as well as the achieved performance on large unlabelled datasets could be addressed. A further contribution is the identification of challenges in developing an end-to-end approach for automating the complete supply chain mapping process as well as a conceptual framework for such an approach.

Acknowledgements

My special thanks go to my supervisor Professor Duncan McFarlane, head of the Distributed Information & Automation Laboratory (DIAL) within the Institute for Manufacturing (IfM) at the Department of Engineering, for his kind support and constructive feedback. Similarly, I would like to thank Dr. Ajith Parlikad, Reader in Asset Management at the IfM, for his helpful advice. Furthermore, I would like to express my deep gratitude to Simon Baker, postdoctoral researcher at the Cambridge University Language Technology Lab (LTL), who spent his free time providing expert advice on the NLP parts of this work. Finally, I would like to thank all my colleagues on my research project VIPr, at DIAL and the IfM as well as the industrial sponsor and the EPSRC.

I dedicate this thesis to my grandfather for inspiring my curiosity as well as to my parents for supporting my brother and myself in giving us the freedom to go where our hearts take us.

Table of contents

List of figures

List of tables

Chapter 1

Introduction

The purpose of this chapter is to provide the motivation behind this research into automated supply chain mapping from text and outline the problem that this thesis will address. This chapter will also present the research aim, research questions, and overall structure of this thesis.

1.1 Problem background and motivation

Modern products, such as cars or aircraft, can be made of millions of parts[1]. These parts are often sourced from many suppliers which are spread across the world, as shown in the case of a BMW 3 Series in Figure 1.1.

As substantial parts of a company's value creation are outsourced to suppliers who in turn outsource to their suppliers, supply chains are becoming increasingly multi-tiered, complex, and geographically distributed (Christopher and Peck, 2004). As a result, companies gradually lose visibility of the topology of their supply network. Knowledge that relates to the supply chain structure, including all supply chain participants and their inter-relations across different tiers, shall henceforth be referred to as *structural supply chain visibility*. This does not include dynamic or behavioural aspects of supply chain visibility, such as the current condition of goods in transit or current inventory levels. Companies know their direct suppliers and customers, but knowledge of sub-tier suppliers tends to be limited. Indeed, a study found that "40% of companies who sourced only in the UK, and almost 20% who sourced globally, had no supply chain information beyond their direct suppliers" (Achilles Group, 2013)[2], that is the identity of sub-tier suppliers remained unknown in those

[1] A Boeing 747-8, as an extreme example, consists of about 6 million parts (Lufthansa, 2018).

[2] The study was commissioned by Achilles, a provider of supply chain management solutions, and conducted by IFF Research

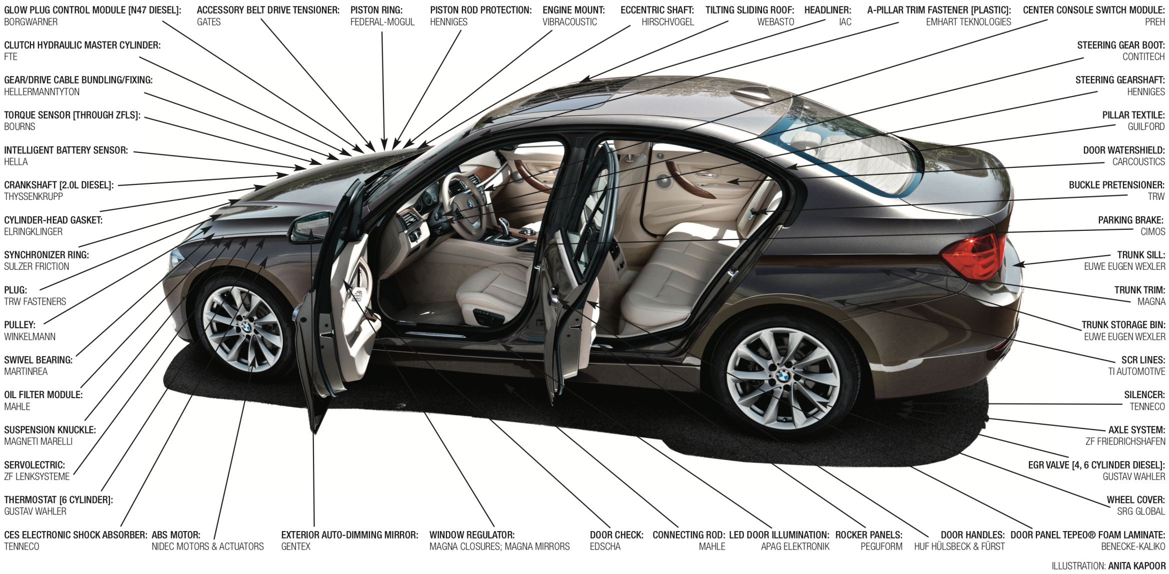


Fig. 1.1 Small selection of known suppliers of the BMW 3 Series; reprinted with permission from Automotive News Europe (Clark, 2013)

cases. Figure 1.2 illustrates this conceptually for the case of a fictitious Original Equipment Manufacturer (OEM).

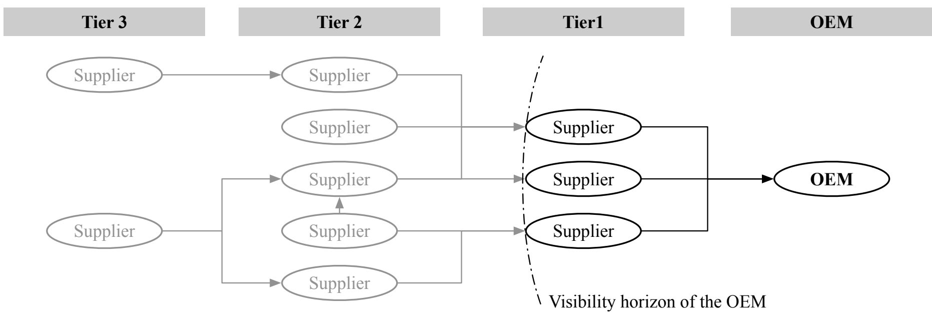


Fig. 1.2 Conceptual illustration of limited structural supply chain visibility

There are solutions for collaborating companies to exchange real-time information about the location and condition of goods in transit, inventory levels, or other dynamic aspects of supply chain performance. For instance, companies can agree to use track-and-trace technology based on RFID or other IoT solutions. However, these technologies are not designed to discover unknown suppliers and reveal their inter-relations. As such, they cannot address the lack of *structural* supply chain visibility.

The reason why structural supply chain visibility cannot easily be achieved is a combination of multiple factors. One major reason is the "proprietary nature of each supplier's relationships with its partners" (Sheffi, 2005). Indeed, suppliers have an incentive not to disclose their own supply network to their customers, especially if they run the risk of being cut out as the middleman or losing bargaining power. That said, suppliers can be contractually obliged by the OEM to disclose their own suppliers. However, it is difficult for the OEM to verify that this information is correct or complete[3]. The difficulty of obtaining the required data is exacerbated by the fact that supply chains are dynamic (Lambert and Cooper, 2000), that is suppliers and inter-relations can change over time. To understand why poor structural supply chain visibility is a problem, one can think of a supply network as a "network of dependencies on suppliers". As such, the performance of a company's supply chain is crucial to its operations. More generally, information about its extended supply chain structure is a valuable input to various decision-making processes of a firm, such as managing the efficiency, resilience, and sustainability of its supply chain. Furthermore, a loss of visibility also corresponds to a loss of control over the extended supply chain.

In particular, management of supply chain risk without visibility of the supply network poses a problem to a company. As another consequence of the emergence of longer, geographically distributed supply chains, companies are now also exposed to more risks and a wider range of risks. For instance, a German car manufacturer can now experience a parts shortage because of a tsunami in Japan[4]. Disruptions at a sub-tier supplier can propagate through a supply network. This ripple effect can halt the production lines of downstream companies which never knew that the disrupted sub-tier supplier was a part of their network. In a special report on global supply chains, The Economist notes in July 2019: "In terms of the risks, most MNCs [multinational companies] do not know who supplies the supplier to their supplier, but they might be held hostage if that distant vendor cannot fulfil its obligations. The dangers are occasionally brought to light by external shocks. Sometimes these are delivered by natural disasters. In the wake of the Japanese tsunami in 2011, a global semiconductor giant tried to map its vulnerabilities to third- and fourth-tier vendors; it took a team of 100 executives more than a year to work out which firms were in its extended supplier networks." (TheEconomist, 2019).

Because a company's performance increasingly depends on the performance of its supply network and "individual businesses no longer compete as solely autonomous entities, but rather as supply chains" (Lambert and Cooper, 2000), ignorance about one's extended supply chain can easily lead to ignorance about one's competitiveness. Studies show that the share of supply chain disruptions that originate at suppliers further upstream than direct suppliers can be as high as 50% (Business Continuity Institute, 2014; KPMG International & The

[3]A situation known as *information asymmetry* in management and organisation theory, e.g. see Vining and Weimer (1988)

[4]See the case of the Xirallic shortage in 2011, described in Section 4.2.3

Economist Intelligence Unit, 2013). Furthermore, they also show that suppliers critical to continued operations can be located *anywhere* in a multi-tiered network and do *not* have to correspond to large sales volumes (Yan et al., 2015) or a large value of the individual component[5]. Knowledge of the supply chain structure would enable a company to better detect and mitigate risks in advance as well as to more quickly and appropriately react to risk events. A simple example scenario would be a situation in which all of a company's direct suppliers depend on the same sub-tier supplier. In this case, the risk associated with the potential single point of failure at the sub-tier supplier is obscured. Should a company instead have this knowledge, it could mitigate the risk by increasing inventory levels, demanding suppliers to diversify their supplier base, or sourcing substitute parts from different suppliers.

Supply chain mapping, whereby information about supply chain participants and their inter-dependencies across multiple tiers is collected and visualised, is often cited in literature as a means of addressing the problem of limited supply chain visibility. Various tools exist for *visualising* buyer-supplier relations. Yet the actual issue of *acquiring* the required information in the first place remains largely unaddressed (Farris, 2010).

Even though data that can readily be used for supply chain mapping is still scarce, we live in an age of an ever-increasing availability of massive amounts of data. Vast amounts of data have become abundantly available at low cost via the Web. E.g. the Common Crawl[6] project maintains a freely accessible repository of archived websites. Despite the efforts to provide humanity's collective knowledge in structured format[7], publicly available ontologies[8] have hardly been populated with information about buyer-supplier relations. Because relevant information is not generally available in a structured format yet, natural language text appears to be the next most promising source data format to address the problem of limited supply chain visibility. Natural language text is promising for a variety of reasons: A large proportion of published human communication is in text form and text documents. For example, news articles and blog posts contain valuable information about buyer-supplier relations. Furthermore, text in natural language generally adheres to grammatical rules and contains words with known meanings for a given context. But since the text documents are typically unstructured, that is running text in natural language instead of in tables with a known data schema, extracting information from it is a challenging problem. Thanks to advances in Machine Learning, in particular Deep Learning, and Natural

[5]"[...] some of the lowest-cost items, from minor suppliers, can cause the biggest and costliest disruptions. [...] In many cases [...] companies have alternative suppliers and ample inventories for big-ticket parts but no such backstops for bits and pieces whose absence nevertheless could halt production." (Hagerty, 2013)

[6]`http://commoncrawl.org/`; last accessed: 2019-05-21

[7]E.g. data in a fixed schema, such as database tables, or ontologies

[8]E.g. dbpedia (`https://wiki.dbpedia.org/`) or wikidata (`https://www.wikidata.org`)

Language Processing, the extraction of information can now be automated (e.g. Wu and Weld (2010), Zeng et al. (2014)).

A solution that would automate the generation of supply chain maps from text documents could have many beneficiaries and use cases. Companies could be enabled to better protect themselves from parts shortages, to avoid damages to their reputation or even legal risks from depending on undesirable sub-tier suppliers, or to make their supply chains more sustainable or more efficient. Government agencies could profit from having a better overview of how whole industries depend on each other or on other countries. Insurance companies, too, may have an interest in better understanding their customers' exposure to supply chain risk.

The overall aim of this study is to examine the extent to which supply chain maps can be automatically generated from unstructured[9], openly available[10], natural language[11] text, such as news articles, and how this can be achieved. The main focus of this study is the extraction of individual buyer-supplier relations, which are a prerequisite for the creation of supply chain maps. The findings of this study will help companies to discover and better monitor their extended supply chain structure.

Figure 1.3 illustrates the conceptual model of the research problem. Variations of this model will be used throughout this thesis to illustrate different aspects of this study. As depicted, the research *scope* is the automated *end-to-end generation of supply chain maps* from unstructured text. In the context of this research study, "automated" shall refer to the fact that the designated tasks do not have to be carried out by a human but are carried out by a computer program. The term shall not imply that the approach can fully replace human research skills in supply chain mapping. The research *focus* is the automated extraction of *individual buyer-supplier relations*. A generated supply chain map, in turn, may assist companies with decision-making processes, such as ones concerning the management of supply chain risks, supply chain efficiency or supply chain sustainability. Despite being shown in separation in Figure 1.3, the management domains are evidently intertwined: for instance, an unsustainable supplier may quickly turn into a supply risk. The ellipsis in the illustration shall indicate that further use cases of supply chain maps exist[12].

[9]The term unstructured is used here to clarify that information shall be extracted from the text itself without the need for additional metadata or specific arrangements of the text, such as tables.

[10]Openly available text includes text extracted from publicly accessible Web pages or news archives. This does not represent a a technical limitation. The same approach could also be applied to privately held text repositories.

[11]A natural language is a language that naturally evolved among humans, such as English – as opposed to constructed languages, such as programming code.

[12]See Section 4.3 for an overview of use cases for supply chain maps.

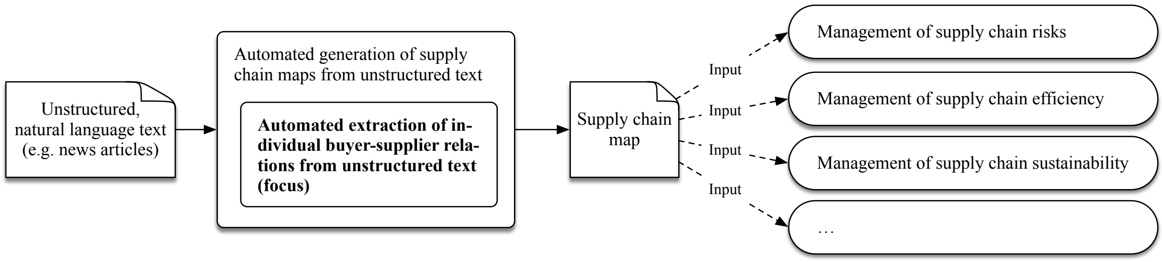


Fig. 1.3 Conceptual model of the research problem

Because supply chain maps are a crucial input to a variety of management processes, instead of focussing on a single use case, this study aims to address the basic information needs that typical supply chain mapping use cases have in common: the identification of supply chain participants and their inter-relations.

This research builds upon a large body of existing academic work in the NLP domain regarding the extraction of entities and their relations from text. Researchers have also used macro-economic data to map how different industries depend on each other (input-output analysis). However, to the author's knowledge, no tools or methods have been proposed to address the issue of automatically generating supply chain maps from text in a comprehensive manner.

The approach proposed and tested in this study shall be understood as a first attempt at systematising and automating supply chain mapping from text. At least at the stage presented in this thesis, it is *not* an equivalent substitute for manual research. But it could, for example, provide an initial supply chain map or help with checking existing supply chain maps for completeness. Furthermore, the effectiveness of any automated approach to supply chain mapping is dependent on obtaining sufficiently rich data to work with in the first place. That said, it is possible to improve the performance further, e.g. by increasing the size of the training data or by inferring relations from multiple statements.

The next section presents the research questions that this study will aim to answer to address this gap.

1.2 Research questions

This section presents an overview of the research questions that this study will aim to answer. A more detailed break-down and discussion of the research questions can be found in Chapter 3. The above discussion has demonstrated a need for an approach to automatically generate supply chain maps from unstructured text. A prerequisite for automated supply chain mapping from text is the ability to automatically extract individual buyer-supplier relations from text. The extraction of individual buyer-supplier relations shall be considered first and in separation, before the broader problem is addressed.

Following the structure presented above, the research questions are as follows:

Research Question 1 (RQ1):
To what extent and how can the extraction of buyer-supplier relations from unstructured text be automated?

Research Question 1 is quantitative and technical in nature and deals with the extraction of individual buyer-supplier relations from unstructured text.

Research Question 2 then addresses the overall problem by widening the scope from the extraction of buyer-supplier relations to the end-to-end process of generating useful supply chain maps from text.

Research Question 2 (RQ2):
What are the challenges in developing a suitable end-to-end approach for automating the generation of supply chain maps from unstructured text?

Research Question 2 is qualitative in nature and aims to identify and discuss the challenges that are likely to arise during the development of an end-to-end approach to automating supply chain mapping from text. These challenges may be a result of the characteristics of supply chains or the limited information quality of the input text document. An example of such a challenge is the extraction of indirect buyer-supplier relations, such as the extraction of relations to sub-tier suppliers.

Figure 1.4 locates the research questions within the conceptual model of the research problem.

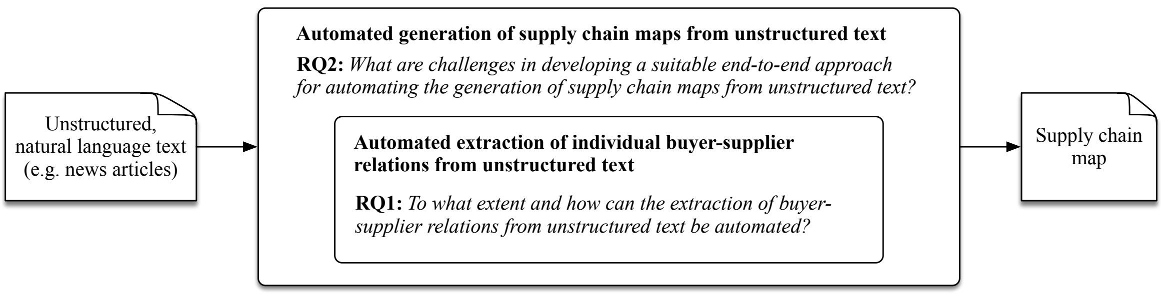


Fig. 1.4 Research questions located within the conceptual model of the research problem

Regarding the scope of this research study, the following caveats should be carefully considered.

Important caveats

1. *Automated*: The term "automated" shall refer to the fact that the designated repeated tasks do not have to be carried out by a human but are intended to be carried out by a computer program with little or no human intervention. The term shall *not* imply that the proposed approach can fully replace human research skills in supply chain mapping, is complete or that its implementations are error-free and could not benefit from manual corrections or complementary manual work. It shall also *not* imply that, for example, the initial acquisition of training data is not the result of manual work. Since large quantities of text have to be processed, it is important to design a scalable end-to-end approach that does not rely on repeated manual processing steps. The term "partial automation" is avoided as it could be confused with an approach that contained designated and repeated manual processing steps, such as having to manually detect company names.

2. *End-to-end*: In this context, the term "end-to-end" shall refer to a process that consumes natural language text (one end) and outputs a basic supply chain map (the other end). The process described in this research cannot possibly address the information needs of all supply chain mapping use cases but will focus on the basic common ones.

3. *Limited data availability*: Even if this process was fully automated and worked perfectly well, this would *not* mean that 100% of any company's supply chain could be reconstructed. Among other factors, this heavily depends on the availability of text documents containing the required information. For smaller companies, suitable text data may not be openly available if at all. Even for larger companies, only a small fraction of the supply chain may be described in publicly or privately available text documents. To understand how much of a company's supply chain can be reconstructed, one first needs to design and apply an automated approach that can process vast amounts of data.

A detailed discussion of the research problem and the methodology can be found in Chapter 3. In the next section, the organisation of the thesis is presented.

1.3 Organisation of the thesis

This thesis consists of eight chapters as shown in Figure 1.5. A summary of each of the following chapters describing their main aims is provided below.

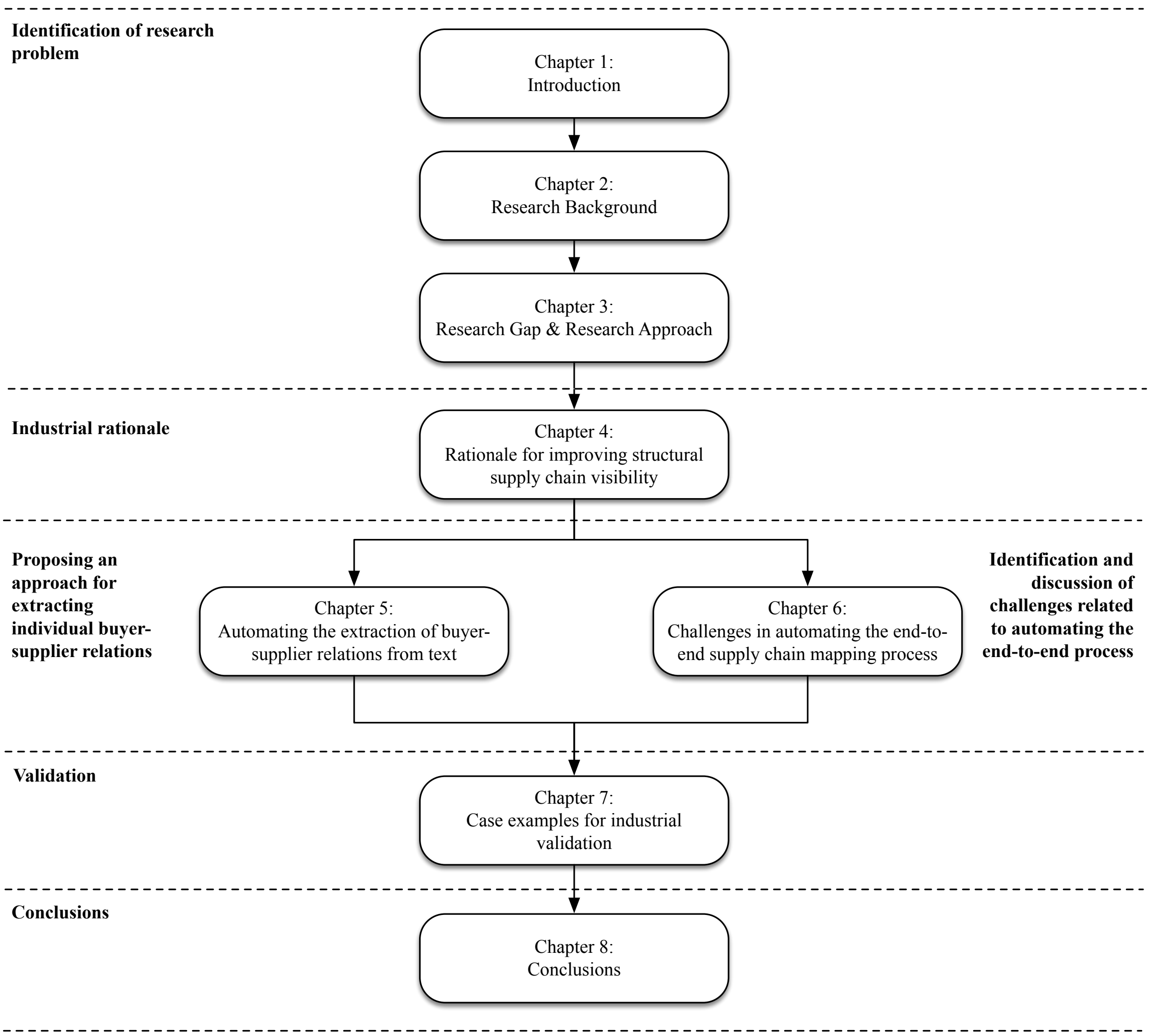


Fig. 1.5 Organisation of the thesis

Chapter 2: Research Background This chapter provides a review of the literature in the field of this study. The aim of this chapter is to provide the academic background for (i) the problem domain, that is the concepts of supply chain and structural supply chain visibility, and (ii) the solution domain for the research problem, that is Machine Learning and Natural Language Processing. In a final section of this chapter (iii), academic work at the intersection of both is discussed.

Chapter 3: Research Gap This chapter summarises the research gap from the literature review and the key issues of the industrial practice. This provides the justification for the research scope, research focus and the research questions that are defined subsequently. Finally, the research approach is discussed.

Chapter 4: Rationale for improving structural supply chain visibility This chapter provides an industrial perspective on the problem of limited visibility of the supply chain structure. The aim is to establish that structural supply chain visibility is limited in practice and to provide a rationale for improving it.

Chapter 5: Automating the extraction of buyer-supplier relations from text This chapter focusses on the core problem of improving structural supply chain visibility: the need for a means of extracting individual buyer-supplier relations from a given piece of text. The aim of this chapter is to analyse this problem, develop a suitable approach for addressing it and to design options of how to evaluate the performance. To establish a ground truth, a text corpus containing buyer-supplier information has to be created and labelled by humans. Only then the classification performance of different algorithms can be determined and Machine Learning classifiers be trained.

Chapter 6: Challenges in automating the end-to-end supply chain mapping process This chapter broadens the scope to examine what is required for the automation of the entire process of generating supply chain maps from text. The aim of this chapter is to address the challenges in automating the end-to-end process, such as the challenge to correctly identify a company's tier in a supply chain. A challenge may be discussed even if a solution cannot be designed or implemented within the scope of this research.

Chapter 7: Case examples for industrial validation This chapter examines industrial supply chain mapping cases. In particular, it provides performance metrics for the corpus creation, relation classification and their use in the application of a trained model to previously unseen data. Furthermore, to validate the applicability and usefulness of the proposed approach, the approach is applied to a large, previously unseen dataset of industry news scraped from the internet. The approach is also specifically analysed for an automotive manufacturer and an aerospace manufacturer. By applying the approach to a large, previously unseen dataset, some preliminary conclusions about data availability and data sparsity can be drawn. Some of the challenges identified in the previous chapter are addressed and remaining ones are illustrated using concrete examples from the data.

Chapter 8: Conclusions and future research Finally, the last chapter concludes this research by summarising the key results and findings. Furthermore, the limitation of the proposed approaches and a recommendation for future work are discussed.

The role of chapters 4, 5, and 6 The chapters 4, 5, and 6 of this thesis have been designed to follow a specific logic, as shown in Figure 1.6.

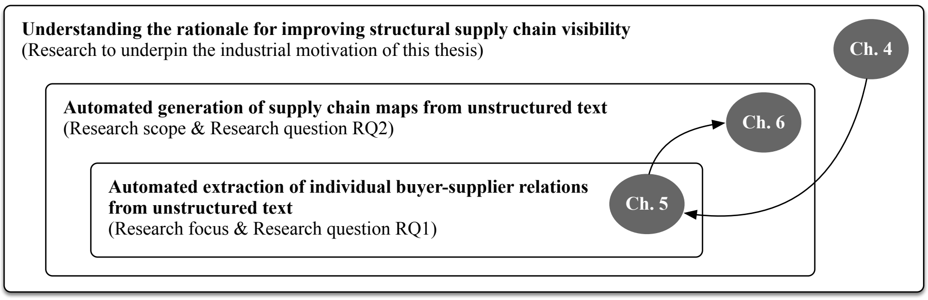


Fig. 1.6 Role of chapters 4, 5, and 6

Chapter 4 provides the rationale for improving structural supply chain visibility from the point of view of the industrial practice. The chapter aims to provide a better understanding of the underlying reasons for the need of understanding supply chain structures, summarises findings and provides arguments that underpin the motivation of this thesis. It also looks at the problem from a wider angle by providing insights into existing options of increasing structural supply chain visibility and by hypothesising about potential beneficiaries. Because this chapter goes beyond merely providing the industrial background and offers new insights, it has been placed after the research gap and approach.

Chapter 5 represents the core chapter of this thesis and focusses on a single aspect and necessary building block of an approach that aims to automatically extract supply chains maps from text: the extraction of individual buyer-supplier relations. This chapter corresponds to the research focus and proposes an appropriate methodology.

Chapter 6 then widens the scope again and examines which other tasks and processes are relevant for an end-to-end automation of supply chain mapping from text. This chapter can be regarded as a detailed gap analysis between the achieved extraction of individual relations and the larger vision of automating the overall process of supply chain mapping from text. Findings from Chapter 5, the shortcomings and achievements of the proposed approach, are required to produce a more comprehensive picture of automated supply chain mapping from text.

Appendix Additional content can be found in the Appendix of this thesis. Further supplementary material, such as code examples and technical details, has been made available in an online repository[13].

[13] https://github.com/pwichmann/phd_thesis

Chapter 2

Research background

2.1 Introduction

The previous chapter introduced the research problem and overall aim of this study: to examine to what extent and how supply chain maps can be automatically generated from unstructured, natural language text. The purpose of this chapter is to provide the relevant academic background required to understand this research, to identify existing methods that have been used to solve similar problems, and to identify the research gap. This chapter focusses on the academic literature, whereas aspects of the industrial practice can be found in Chapter 4. The literature review is split into three parts, as depicted in Figure 2.1. The first part contains the review related to the problem space, such as supply chain concepts. The second part refers to prior academic work related to the solution space, such as Machine Learning and Natural Language Processing. The third part summarises known prior work at the intersection of problem space and solution space, such as academic work on automatically generating supply chain maps from unstructured text[1]. The domain of network science has been omitted in this review of the relevant academic background since the focus of this research at this early stage is on correctly extracting networks of buyer-supplier relations – and not yet on analysing the resulting graphs or using them to infer statements.

[1]Two papers have been produced on the research presented in this thesis. In Wichmann et al. (2018), the idea of automatically generating supply chain maps from natural language text was first introduced and its challenges discussed. For the experiments in this paper, manually pre-defined lexico-syntactic patterns were used to extract relations. In Wichmann et al. (2020), the focus is on automated classification of buyer-supplier relations by creating a text corpus and using it to train and test a Deep Learning classifier.

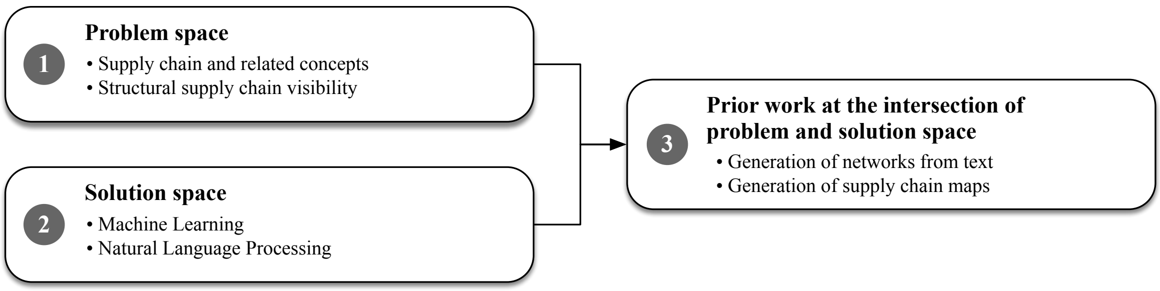


Fig. 2.1 The literature review is split into three sections: (1) problem space, (2) solution space, and (3) prior work at the intersection of problem and solution space

The conclusion of the academic background can be found at the beginning of Chapter 4 where the research gap is derived in combination with the key issues from the review of the industrial practice.

2.2 Supply chain and related concepts

2.2.1 Supply chain

The term *supply chain* can be misleading since it is rarely used in the strict sense of a "chain of businesses with one-to-one, business-to-business relationships, but a network of multiple businesses and relationships" (Lambert and Cooper, 2000, p.65). Within the context of this thesis, the terms *supply chain* and *supply networks* will be used synonymously and shall always imply a network structure.

Since supply chains are networks, they consist of nodes – the members of the supply chains, typically organisations – and directed links – often conceptualised as "process links" (Lambert and Cooper, 2000) of various kinds[2] or as "flows of products, services, finances, and/or information from a source to a customer" (Mentzer et al., 2001). The notion of flows is interesting since it provokes the question to what extent a supply chain map should also be capable of representing monetary or informational flows.

The combination of nodes and links give the network its structural dimensions. The *horizontal* structure refers to the number of tiers across the supply chain. A company's horizontal position refers to the position within the supply chain which can be near the initial source of supply or closer to the ultimate customer. The *vertical* structure refers to the number of suppliers or customers represented within each tier (Lambert and Cooper, 2000).

[2] Lambert and Cooper (2000) distinguish four different types of process links: managed process links, monitored process links, not-managed process links, and non-member process links.

The term “upstream” is used to denote the direction towards to original supplier whereas “downstream” refers to the direction towards the ultimate customer. A supply chain for a particular company typically does not include the company’s competitors (unless they also play the role of suppliers or customers). Another characteristic is the supply chain’s scope

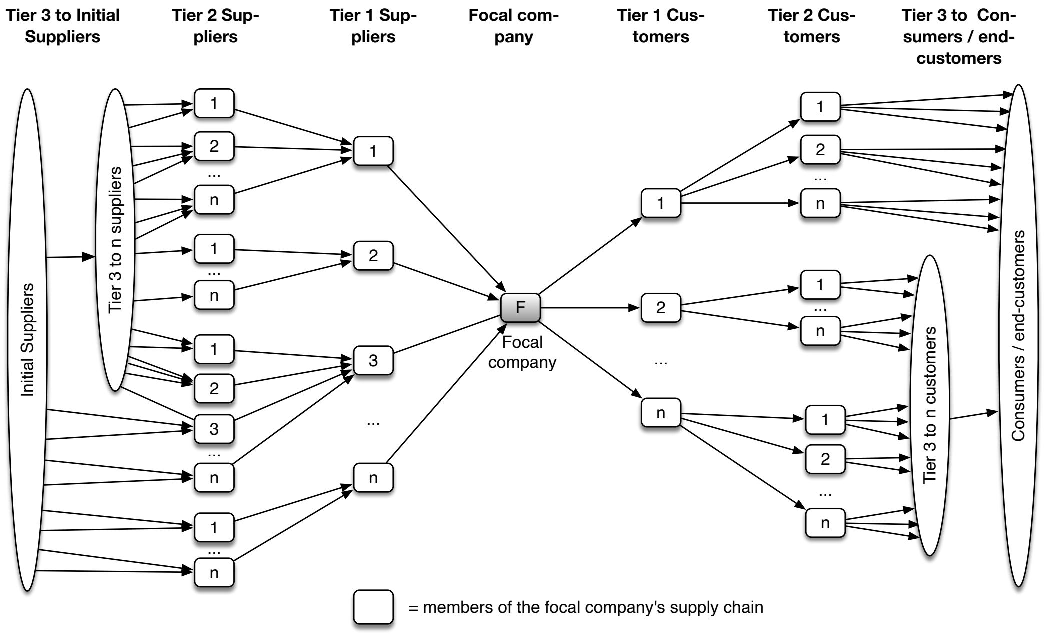


Fig. 2.2 Illustration of a typical supply chain structure for a focal company; source: author, adapted from Lambert and Cooper (2000)

and level of detail one would like to consider. For example, the highest level of complexity that Mentzer et al. (2001) considered is the “ultimate supply chain” which incorporates all the organisations involved from the ultimate supplier to the ultimate customer including supporting organisations such as third party logistics providers or market research firms.

Supply chain networks are not static. Depending on existing regulatory restrictions, the availability of alternative suppliers, the standardisation of products and other company-specific and industry-specific factors, supply chains can be highly dynamic, with companies frequently switching between multiple alternative suppliers. This generally makes it difficult to know the supply chain structure at any given point in time.

It is noteworthy that supply chain networks represent flows between companies and, hence, *dependencies*. This is substantially different from the structure of, for example, communication networks where the links may indicate *possible* alternative pathways (but not dependencies). This difference easily leads to confusion when (e.g. resilience-related) findings from one domain are to be applied to the other. If supply networks are depicted with directed flows, these flows tend to indicate the direction for the provision of goods and services to the buyer. Yet, financial and informational flows often have the opposite direction: from the buyer to the supplier.

Distinction between supply chain and related concepts The terms *supply chain*, *supply network*, *logistics network*, *value chain* and *value stream* sound similar and can easily be confused. The term "value stream" is only of relevance in so far as the concept of value stream mapping can have some similarity to supply chain mapping. A discussion of these concepts and their differences can be found in the Supplementary Material[3].

2.2.2 Supply chain risk

Supply chain risk and vulnerability

Even though there is no generally accepted definition of risk, a prominent one from classical decision theory is "variation in the distribution of possible outcomes, their likelihoods, and their subjective values" (March and Shapira, 1987). In other words, the uncertainty about future events gives rise to risk. Risk can be quantified as the multiplicative product of the probability of the event occurring and the resulting, undesirable impact (e.g. a monetary loss) should the event occur (Kenett and Raanan, 2011): $\text{Risk} = \text{Probability} \cdot \text{Impact}$.

Broadly speaking, *supply chain risks* are related to mismatches of demand and supply, and hence to disruptions of the flows between organisations. Jüttner (2005) stresses the relatedness of supply chain risks and flows between organisations: "Risk in the supply chain centres around the disruption of 'flows' between organisations. These flows relate to information, materials, products and money. They are not independent of each other but are clearly connected". Supply chain risks, by definition, extend "beyond the boundaries of the single firm and, moreover, the boundary-spanning flows can become a source of supply chain risks" (Jüttner, 2005).

Supply chain risks arise from fluctuations in the performance of the supply chain either in the upstream or the downstream operations. The general term *supply chain risk* is often imprecisely used as a synonym to *supply risk*, the latter of which explicitly refers to risks that may delay or disrupt the *supply* of goods and services in the desired quantity and quality. But in fact supply chain risks also include demand-side risks, such as disruptions in the outbound logistics or changes in the preference of customers. Some types of risks can be considered supply chain risks but do not immediately relate to disrupted flows: Reputation risks and legal risks, e.g. resulting from using suppliers with unethical practices, may arise with only

[3] `https://github.com/pwichmann/phd_thesis`

an indirect impact on supply or demand. Since supply chains are networks, supply chain risks are transmitted, and risk events can propagate through the network (Waters, 2011). Given that there is rarely a single cause for a risk event, but a complex causal chain from initial cause to final impact, a classification of risks by their source or effect is therefore likely to be tenuous and easy to criticise. Tang and Tomlin (2008) distinguish and define six main types of supply chain risks: supply risks, process risks, demand risks, intellectual property risk, behavioural risk, and political/social risks. An alternative, more comprehensive categorisation of supply chain risks and risk drivers can be found in Chopra and Sodhi (2004). Yet a another different categorisation of risks is based on the degree to which a risk is within the control of the firm: Christopher and Peck (2004) distinguish risks that are *within the firm* (either process- or control-related), that are *external to the firm but internal to the supply chain network* (either demand- or supply-related), and those that are *external to the network* (environmental risks).

It is noteworthy that supply chain risks can arise on very different abstraction levels, as illustrated by Figure 2.3. A single machine breakdown or warehouse fire may interrupt the local flow of parts between production sites at specific geo-locations. Power outages, port strikes, or a damaged submarine communications cable may disrupt an entire infrastructure, like power, IT, or transportation networks. Lawsuits or bankruptcies are risks to the network of legal entities an organisation consists of. The market environment and political risks, such as regulations or embargoes, impact companies on a regional, national or even global level.

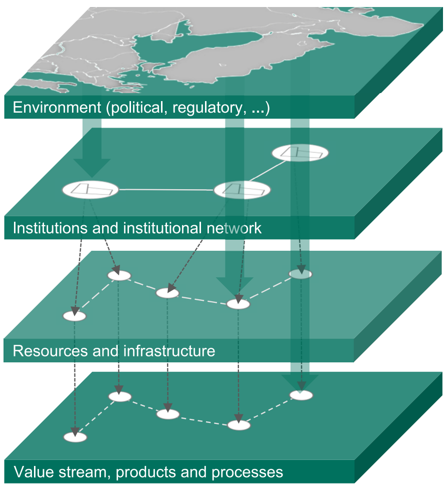


Fig. 2.3 Layers of supply chain risk; adapted from Berbner (2017, p.66)

A supply chain's risk exposure determines its vulnerability. *Supply chain vulnerability* can be defined as "an exposure to serious disturbance arising from supply chain risks and affecting the supply chain's ability to effectively serve the end customer market" (Jüttner, 2005). A discussion of the term *supply chain resilience* in this context can be found in the Supplementary Material.

2.2.3 Supply chain risk management

The purpose of this section is to provide a brief summary of supply chain risk management and to establish that there are different stages of the risk management process with different objectives. This background may help with the subsequent discussion of the usefulness of structural supply chain visibility at different stages of the supply chain risk management process.

Generic risk management process

The question about how to deal with risk has been addressed in risk management research and generic risk management processes have been proposed. The generic risk management process typically includes steps for the identification, assessment, treatment, and monitoring of risks. Figure 2.4 shows a more clearly defined risk management process according to the ISO 31000 family of standards for risk management. This standard also proposes generic risk treatment strategies such as risk sharing.

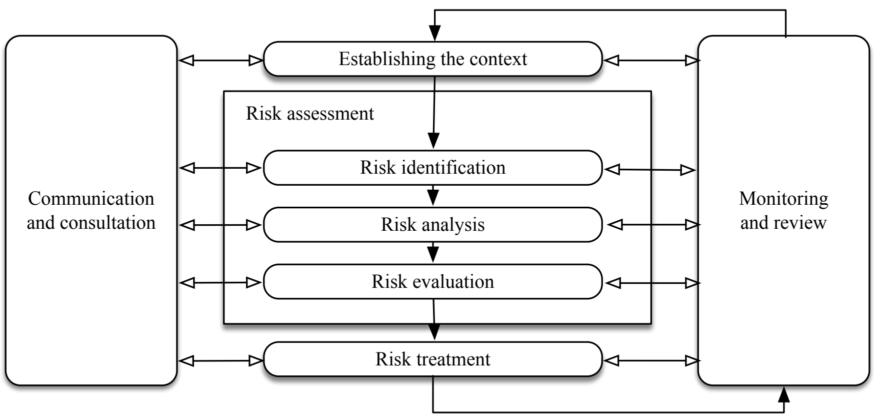


Fig. 2.4 Risk management framework according to ISO 31000:2009; adapted from the corresponding figure in the ISO standard

General tactics to reduce risk

One can distinguish two main tactics to reduce risk, such as the risk of supply chain disruptions (Tomlin, 2006):

- A *mitigation* action is what is done *prior* to the occurrence of the risk event. And, hence, the costs of the action are incurred irrespective of whether the risk event occurs. The purpose of mitigation is to reduce the probability of the risk event and/or its potential impact.

- A *contingency* action is taken *after* the risk event has occurred (and hence the costs are incurred only if the event actually has happened). Since the incident has already occurred, the purpose of contingency is only to reduce its impact.

Based on the above definition of risk, actions to manage risk can either attempt to reduce the probability of the event or to reduce the impact in case the event happens. Arguably, a further characteristic of risk is lead time – that is the time between the occurrence of the incident and its detection or prediction (Sheffi, 2005). Hence, risk management actions could also attempt to increase the lead time to enlarge the set of available mitigation or contingency options.

Examples of measures to address supply chain risk

A wide range of mitigation and contingency measures have been identified and discussed in the literature. It is essential to understand that these mechanisms are generally associated with a cost; otherwise, they would be in general use already.

Tang (2006b) provides a comprehensive and systematic overview of general tactical and strategic plans to address supply chain risks. Tang identifies four different main approaches (supply management, product management, demand management, and information management) and presents a range of specific mechanisms for each approach. Chopra and Sodhi (2004) also present a range of approaches to mitigate supply chain risk. Examples of specific *mitigation measures* are, for example, having multiple (redundant) suppliers (Tang, 2006b) to reduce the dependency on any single supplier or spreading multiple suppliers across multiple countries to reduce exposure to risk on a country-level (Tang, 2006b). Examples of specific *contingency measures* are, for example, demand-side management to shift demand from a disrupted product to one that is available, e.g. by adjusting prices, so that no demand-supply mismatch occurs. Tang (2006a) provides the example of Dell offering special price incentives to shift demand to computers that utilised components from countries that were not disrupted after an earthquake hit Taiwan in 1999.

2.3 Structural supply chain visibility

2.3.1 Supply chain visibility in general

In academic literature, the term *supply chain visibility* (or sometimes also called *supply chain transparency*) has been defined inconsistently. Table 2.1 provides an overview of selected definitions; for a more comprehensive overview, the reader may be referred to Goh et al. (2009), Francis (2008), Caridi et al. (2010) and Barratt and Oke (2007).

Table 2.1 Overview of selected definitions of *supply chain visibility*

Reference	**Definition**
Fawcett and Magnan (2002)	No explicit definition provided; aspects that were examined as part of supply chain visibility: existence of a supply chain map and visibility of sub-tier suppliers and/or customers; existence of process maps
Christopher and Peck (2004)	"Put very simply, supply chain visibility is the ability to see from one end of the pipeline to the other. Visibility implies a clear view of upstream and downstream inventories, demand and supply conditions, and production and purchasing schedules for example. It also implies internal visibility with clear lines of communications and agreement on 'one set of numbers"'; the authors distinguish "supply chain understanding" and define it as "an understanding of the network that connects the business to its suppliers and their suppliers and to its downstream customers"
Francis (2008)	"the identity, location and status of entities transiting the supply chain, captured in timely messages about events, along with the planned and actual dates/times for these events." In this context, an entity moving though the supply chain can be an item, a shipment, a vehicle etc.
Jüttner and Maklan (2011)	"the extent to which actors within the supply chain have access to or share timely information about supply chain operations, other actors and management which they consider as being key or useful to their operations"

The term supply chain visibility is often used to refer to the identity, location or current status of ordered goods in transit – especially in the context of RFID track-and-trace technology. Other definitions also include the sharing of operational information, such as inventory levels. Again other definitions also explicitly include knowledge about sub-tier suppliers, their inter-dependencies and the process flows across company boundaries in a multi-tiered supply chain, such as Fawcett and Magnan (2002).

Within the scope of this study, the broad definition of supply chain visibility by Jüttner and Maklan (2011) is adopted. The notion of supply chain visibility shall also include what Christopher and Peck (2004) call "supply chain understanding" and define as "an understanding of the network that connects the business to its suppliers and their suppliers and to its downstream customers". It would appear counter-intuitive to assign the label "perfect visibility" to a supply chain in which all suppliers (and their inter-dependencies) on the various tiers remained unknown.

2.3.2 Definition of structural supply chain visibility

For the context of this study, *structural* supply chain visibility shall be defined as those aspects of supply chain visibility that correspond to knowing the supply chain participants and their inter-relations across multiple tiers, that is the supply network *topology*. The term is introduced to distinguish structural properties from more dynamic or operational ones that are commonly associated with the term supply chain visibility and relate to knowledge of the real-time performance of the supply network. Figure 2.5 aims to depict aspects of *structural* supply chain visibility as a subset of the overall concept of supply chain visibility.

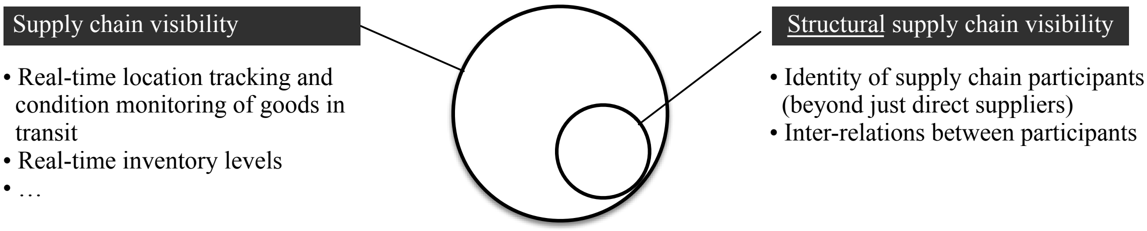


Fig. 2.5 Structural supply chain visibility as part of the broader term supply chain visibility

2.3.3 Importance of structural supply chain visibility

In academia, the importance of knowing the extended supply chain structure has been discussed and established from various different perspectives, such as supply risk or corporate social responsibility (CSR). On an abstract level, the value of knowing the supply chain structure can be viewed from a value of information (cf. Howard (1966)) perspective. This section shall summarise relevant academic work that provides evidence regarding the importance of structural supply chain visibility specifically with respect to *supply risk*.
Beyond this section, additional evidence for the importance of structural supply chain visibility can be found in other parts of this thesis:

- Section 2.3.4 of this thesis where the work by Farris (2010) is discussed in the context of *strategic supply chain mapping*.
- Chapter 4 which offers an *industrial perspective* on the rationale for improving structural supply chain visibility. It provides further evidence drawn from case studies and interviews and also summarises potential beneficiaries and use cases.

Since a supply chain structure is the network of dependencies that exposes a company to supply chain risks in the first place and through which risks propagate, knowledge of that structure can help with managing supply chain risks. The existence of risk is the reason why supply chains or operations in general need to be resilient and why inventory can be regarded as a surrogate for information ("replacing inventory with information" (Mangan et al., 2008)).

In the literature, there is ample evidence for the importance of structural supply chain visibility for supply chain risk management. Basole and Bellamy (2014) examine the link between structural supply chain visibility and risk management and find that "structural visibility into the lower tiers of the supply network has a significant mitigating impact on cascading risks" and that "enhanced visibility is an important and perhaps essential capability for effective supply chain risk identification and mitigation. Supply chain managers must therefore move beyond a simplified dyadic or triadic view to a more holistic approach when developing risk identification and mitigation strategies". Examples of obscured risk include suppliers depending on the same sub-tier supplier or high-risk sub-tier suppliers.

Yan et al. (2015) introduce the idea of a "nexus supplier". Contrary to the intuition that strategic, direct suppliers are the critical ones due to their direct and large impact on a buying firm's profit and risk position, a nexus supplier could be located in any (sub-)tier of the supply chain, does not have to relate to a large sales volume, but has a potentially large impact on the buying firm if it was disrupted. The existence and identity of such a nexus supplier on a sub-tier could only be revealed with better visibility into the supply chain structure. The network structure also determines how risk events propagate through the network and if they get absorbed or even amplified (Jüttner et al., 2003). An early detection of and response to risk events would require knowledge about which events are relevant to a company's supply chain. For this, too, knowledge of the supply chain structure is necessary. Kim et al. (2014) use graph theory to investigate the importance of supply chain structure for its resilience. The authors ask the question how the supply network structure influences disruptions, and how one can assess the resilience of supply network structures. Their results "indicate that [the] network structure significantly determines the likelihood of disruption", and that "different structural relationships among network entities have different levels of resilience" (Kim et al., 2014). Gardner and Cooper (2003) state that "[t]here are examples of companies that found a critical component or raw material supplier that was in danger of going out of business [...], was the only supplier of a critical component [...], or dumped toxic waste improperly [..]. Quickly identifying sole or critical suppliers more than one tier up can suggest further investigation and monitoring for supply chain bottlenecks." Christopher and Peck (2004) state that a "fundamental pre-requisite for improved supply chain resilience is an understanding of the network that connects the business to its suppliers and their suppliers and to its downstream customers. Mapping tools can help in the identification of 'pinch points' and 'critical paths"'.

2.3.4 Supply chain mapping

Definition Supply chain mapping aims to address the problem of limited supply chain visibility – in particular the visibility of the supply chain structure. Gardner and Cooper (2003) define supply chain maps as "a representation of the linkages and members of a supply chain along with some information about the overall nature of the entire map". This basic definition is adopted in this thesis and shall be further refined by the following observations.

Since supply chain maps serve a mostly tactical or strategic purpose, the scope and level of detail can vary from one to another. They range from a geographic vulnerability map which "simply depicts which supplier of what parts are located in each area of the world" (Sheffi, 2005) to maps that incorporate the "connectedness of the supply chain to help understand interdependencies" (Sheffi, 2005). "Such a supply chain map highlights the flow of parts out of given regions, depicting who is involved and the plants in other parts of the world that are dependent on them. Such a map can become a tool for understanding the extent to which a flood in Brazil will affect production in Singapore or sales in Germany." (Sheffi, 2005, p.32-33). Gardner and Cooper distinguish a strategic supply chain map "by its direct tie-in to corporate strategy. This type of map can be either an integral part of the strategic planning process or a tool for implementing the supply chain strategy. Supply chain maps come in a number of shapes and styles. The focus could be on a particular use or user, on a theme such as a type of value added, or generic, covering all aspects of supply chain structure. The maps can depict organizations, flows, facilities, and/or processes" (Gardner and Cooper, 2003).

The term supply chain *map* has the following two connotations: Most importantly, it suggests a *visual* representation of the supplier dependencies. It may also allude to an explicitly spatial, *geographical* depiction of a supply chain. From a risk management perspective, the geographical aspect can be important since disruptions are often linked to geolocations (natural disasters, country risk etc.). Gardner and Cooper (2003) note that "[w]hile cartographers might insist that a true map has spatial relationships depicted, for a supply chain map this information could be present or absent".

The minimal set of elements of a supply chain map is composed of the companies (nodes of the network) and their inter-relations (arcs of the network). Further basic elements of supply chain maps include additional information about the companies (e.g. company key information, including its geographical location) and about the relations (e.g. the type of goods or services provided, or the end-product to which the provided goods or services contribute, or the time period during which this relation existed).

Gardner and Cooper (2003) show numerous examples of supply chain maps found in literature, such as the one shown in Figure 2.6 from Fine (1999). By choosing possible combinations from their table of supply chain map attributes, Gardner and Cooper (2003) also provide sample maps for hypothetical supply chains, such as the one shown in Figure 2.7.

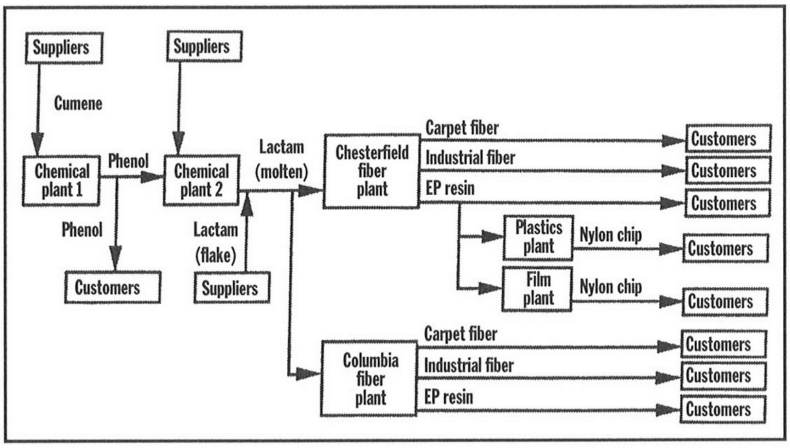


Fig. 2.6 Example of a supply chain map, from Fine (1999), showing individual plants

Even though supply chain mapping addresses the problem of supply chain transparency, the technique itself (if solely seen as a visualisation technique) does not help with acquiring the necessary information (cf. Gardner and Cooper (2003), Farris (2010)). The main challenge for any supply chain mapping project is an obvious one: unavailability of data beyond the first tier. Indeed, obtaining the required information is challenging. In one of the few academic papers addressing how the necessary data can be acquired, Farris (2010) noticed that "the first obstacle in developing a strategic supply chain mapping procedure was the ability to obtain actual data spanning the supply chain".

Benefits and applications of supply chain maps Gardner and Cooper (2003) identify a wide range of potential *benefits* of supply chain mapping for a firm. These go beyond just

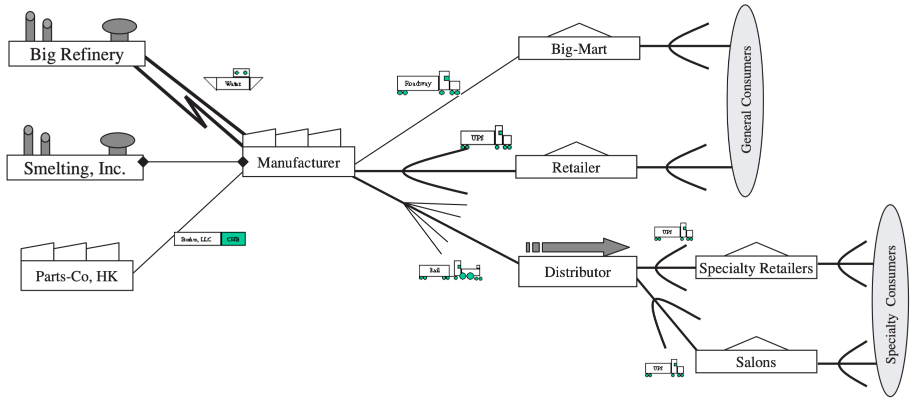


Fig. 2.7 Sample map for a hypothetical supply chain, from Gardner and Cooper (2003)

improving structural supply chain visibility for a firm. Gardner and Cooper (2003) consider a supply chain map “a communication tool to reach across firms, functions, and corporate units”. Farris (2010) build upon the work by Gardner and Cooper (2003) and provide an overview of “strategic areas of interest” for Tier 1 to Tier 2 relationships that could be revealed by a supply chain map that goes beyond just direct suppliers. As depicted by Figure 2.8, use cases of strategic supply chain maps go far beyond supply risk management or even just an upstream perspective. For instance, by understanding “your customer’s customer”, one may be able to design better or alternative sales channels. Moreover, by structuring the use cases in a 2-by-2 matrix, two perspectives are highlighted that are otherwise less obvious: There can also be value in understanding “your customer’s other suppliers” and ”your suppliers’ other customers”. These important perspectives may be easily missed in a classical supply chain mindset. Numerous papers report on the *application* of supply chain mapping to specific scenarios. E.g. Choi and Hong (2002) provide supply chain mapping case studies in the automotive industry. The supply chain maps were limited to the centre console assembly of three different product lines. The data was collected *manually* through interviews, from documents provided by the automotive companies, and via observations during a plant tour. Choi and Hong (2002) compare the three resulting supply network structures from the points of view of formalisation, centralisation, and complexity. Another example of a manual supply chain mapping exercise is a report by the US Geological Survey. This report “uses the supply chain of tantalum (Ta) to investigate the complexity of mineral and metal supply chains in general and show how they can be mapped” (Soto-Viruet et al., 2013).

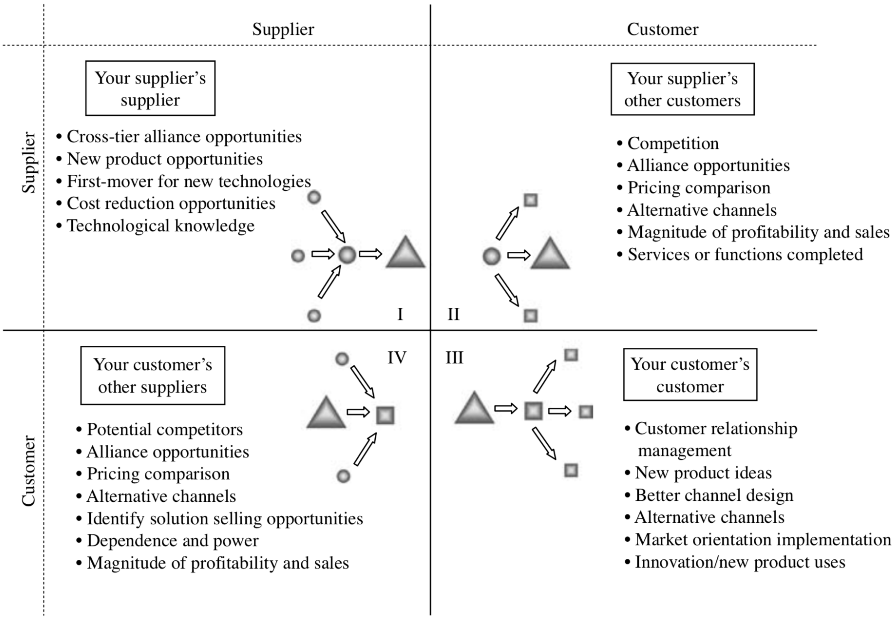


Fig. 2.8 Strategic areas of interest for Tier 1 to Tier 2 relationships, from Farris (2010)

2.4 Machine Learning

The concept of Machine Learning is relevant to this research for several reasons. The domain of Natural Language Processing (NLP) has been increasingly revolutionised by the advances in Machine Learning, and in particular Deep Learning. What had been painstakingly handcrafted rules before is now increasingly replaced by learning from the data itself. Existing methods of Machine Learning and NLP inform the methodology of this research and the chosen research approach.

2.4.1 Definition

In 1959, Arthur Samuel defined Machine Learning as the "field of study that gives computers the ability to learn without being explicitly programmed". Tom Mitchell's definition from 1997 of what it means for a computer program to learn is more technical: It is "said to learn to perform a task T from experience E, if its performance at task T, as measured by a performance metric P, improves with experience E over time" (Mitchell, 1997).

2.4.2 Types of machine learning

Today, one can arguably distinguish three main types of machine learning: *supervised learning*, *unsupervised learning*, and *reinforcement learning* (Murphy, 2012). For the context of this study, supervised learning is of particular importance.

Supervised learning In *supervised learning*, the objective during a training phase is to learn a mapping from provided inputs (called *features*) to provided outputs (called *response variables* or *labels*). Labels often have to be obtained from humans that annotate data. Labelled data (i.e. data comprising features *and* correct labels) is sometimes also referred to as *gold-standard* data. After the training phase, this learned mapping can be applied to predict the response variable for previously unseen inputs, as shown in Figure 2.9. If the response variable is a discrete value (like the brand of a car), the problem is called *classification*. On the other hand, if the response variable is continuous (like the price of a share), the problem is called *regression* (cf. Chollet (2017)).

The trade-off between too little and too much training as well as between a too simplistic and an overly complex model is called the *"bias-variance trade-off"*. A model can be *"overfit"* to the data such that the model works very well on the training data but does not generalise to unseen data. In this case, the model has been fit to the noise in the training data. In order to examine how well a model generalises, the model is not trained on all the labelled data. Instead, the data is split into partitions: a training dataset and the test dataset. Sometimes, a third partition (called validation dataset) is used to fine-tune hyper-parameters of the model.

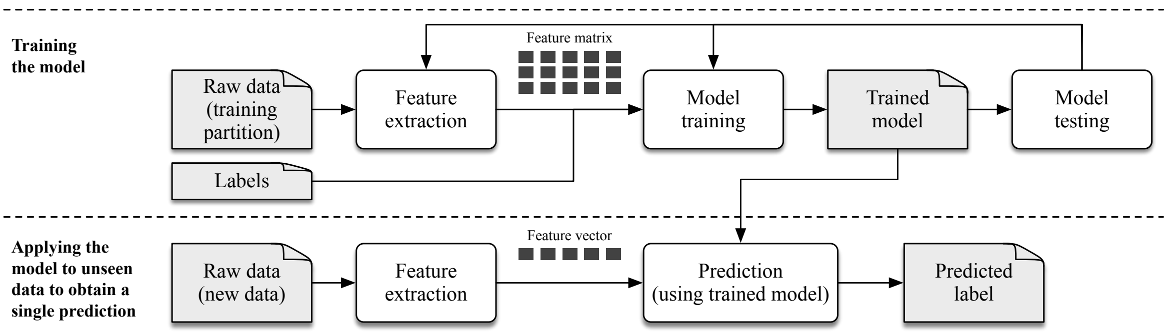


Fig. 2.9 Generic pipeline for a supervised learning task

Unsupervised learning As labels often have to be provided by human annotators, the label acquisition can be costly, and may introduce undesirable biases. *Unsupervised learning* works without the need for labels but addresses different types of problems. In unsupervised learning, only the inputs are given and the objective is to find interesting patterns (like groups of similar items or anomalies in a list of financial transactions) in the data or otherwise describe or summarise the data, e.g. in form of dimensionality reduction or clustering (cf. Chollet (2017, p. 94)). The spectrum between fully unsupervised and fully supervised learning is called *semi-supervised learning*, where the learner is trained on a combination of labelled and unlabelled data in the attempt to leverage the advantages of both.

Reinforcement learning In *reinforcement learning*, an agent can observe and interact with the environment. Without being provided with an explicit description of how to achieve an objective, the agent is taught by receiving a delayed reward signal. The agent then attempts to maximise the expected reward (cf. Chollet (2017, p. 95)).

2.4.3 Deep Learning and selected algorithms

Neural networks

Neural networks, or artificial neural networks (ANN), are a family of machine learning algorithms that are loosely inspired by the biological brain. Each unit in a neural network "resembles a neuron in the sense that it receives input from many other units and computes its own activation value" (Goodfellow et al., 2016). Neural networks are composed of stacked layers, and those layers between the input layer and the output layer are called *hidden* layers.

Deep learning

The number of layers in a neural network determines the *depth* of the model. And it is this terminology that gave rise to the name *"deep learning"* which refers to neural networks with at least one, but commonly dozens if not hundreds of hidden layers. The *width* of the model refers to the number of units in a particular layer, whereas the *size* of the model refers to the overall number of units in the model. Neural networks form weighted linear combinations of the non-linear activation function of each unit. The "universal approximation theorem" states that a feed-forward neural network with a single hidden layer containing a finite (but potentially very large) number of neurons can approximate any continuous function arbitrarily well (e.g. see Hornik (1991)). Weights of the neural network are updated using the backpropagation algorithm which successively minimises the error between predicted output and provided label. Comprehensive introductions to deep learning can be found in Chollet (2017) and Goodfellow et al. (2016).

Selected deep learning network architectures

A wide variety of Deep Learning model architectures have been developed in recent years. In the following, a selection of neural network designs shall be summarised that are relevant for the problem at hand. Figure 2.10 shall illustrate conceptually how the following architectures inter-relate.

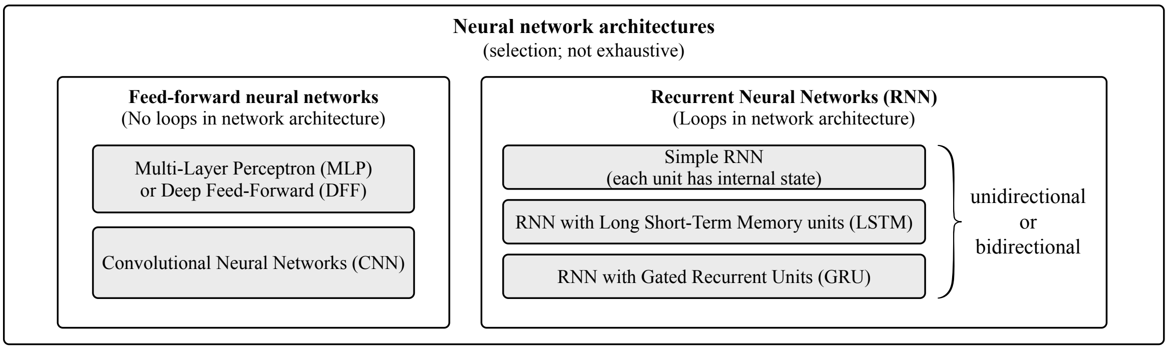


Fig. 2.10 Selection of relevant neural network architectures

Feed-forward neural network A *feed-forward neural network* is the simplest form of a neural network. The name refers to the fact that the connections between the nodes in the network do not form a cycle. A feed-forward neural network with an input layer, an output layer and at least one hidden layer is often called a *multi-layered* perceptron (MLP); networks with multiple hidden layers are commonly referred to as "deep", such as in "*deep* feed-forward" (DFF) network (e.g. Goodfellow et al. (2016, p. 163)).

Convolutional neural network (CNN) A specific type of feed-forward multi-layer neural network are *convolutional neural networks* (commonly abbreviated as *CNN* or *convnets*). A CNN typically consists of stacks of convolution layers and pooling layers. Convolution layers look at local spatial patterns in an input tensor, whereas the pooling layers spatially downsample the data. A fully-connected layer at the end of the CNN then allows for a classification or regression. CNNs are not suited for all types of tasks but work well for data with spatial relations, such as the "grid-like topology" (Goodfellow et al., 2016, p. 163) provided by the grid of pixels in an image. CNNs can also be used for text classification: A 1-dimensional convolving filter (1D CNN) can be applied to a text sequence to detect specific word sequences ('ngrams'). Jacovi et al. (2018) provide a summary of the use of CNNs for text classification.

Recurrent neural network (RNN) *Recurrent neural networks* (commonly abbreviated as *RNN*) (cf. Goodfellow et al. (2016, p. 368)), as opposed to feed-forward neural network, contain loops. This way, they allow the behaviour of neurons not just to be determined by activations in previous hidden layers but also by activations at earlier times or even a neuron's own activation at an earlier time. RNNs are particularly suited to sequential data, such as time series data, speech or text sequences, since they can consider the order of the sequence for a prediction task.

Long short-term memory (LSTM) networks *Long short-term memory* (LSTM) networks are a type of RNN that contains LSTM units. LSTM units were introduced in 1997 by Hochreiter and Schmidhuber (1997) and address the problem of the vanishing gradient: "Learning to store information over extended time intervals by recurrent backpropagation takes a very long time, mostly because of insufficient, decaying error backflow" (Hochreiter and Schmidhuber, 1997). Long short-term memory units are added to the network structure and allow the network to learn much faster by storing information. LSTM units are composed of input gate, output gate and forget gate. Each of these gates can be thought of as a "neuron" in a feed-forward network, producing an activation at a time step t.

Gated Recurrent Units (GRU) (Cho et al. (2014) and Chung et al. (2014)) have a similar function as LSTM units. They tend to be computationally less expensive but also with less representational power compared to LSTM (Chollet, 2017, p.215).

Bidirectional RNNs The idea of RNNs in general is that with respect to sequential data and specific tasks at hand, the *order* of the data can matter and would need to be considered by the network. Interestingly, for some types of sequential inputs, models perform similarly well if the data is read "anti-chronologically". However, because these RNNs trained on the reversed sequence learn a different representation, it is useful to combine the outputs of RNNs trained on the normal *and* the reversed sequence (Chollet, 2017). Such network architectures are called *bidirectional* RNNs, and bidirectional versions also exist for RNN sub-types (e.g. B(i)LSTM (Graves and Schmidhuber, 2005)).

Attention and transformers Recent developments, such as Google's attention-based transformers (Vaswani et al., 2017), were not considered within the scope of this thesis. Large-scale unsupervised language models, such as GPT-2 (Radford et al., 2019) and its successors, could at least help with identifying relations.

2.4.4 Metrics of labelling consistency

Whenever labels are acquired from human annotators in the domain of Machine Learning, the question arises to what extent these labels are correct. Since the true label is not known in advance, strategies to assess the correctness are based on redundant labelling and metrics of consistency. Labelling consistency is then used as a proxy for the achieved labelling quality. The labelling consistency can be measured across annotators (*inter*-annotator agreement) or for a specific annotator across redundant labelling tasks (*intra*-annotator agreement). A wide range of metrics for inter- or intra-annotator agreement have been designed, such as Cohen's Kappa (κ) (Cohen, 1960) or Fleiss' Kappa (Fleiss, 1971). A more detailed discussion including the relevant formula as well as worked examples can be found in the Supplementary Material[4].

2.5 Natural language processing

The domain of natural language processing (NLP) provides the tools used within this research to extract organisations and their relations. Thus, this section shall summarise relevant key concepts and prior work. This section is structured along the following four main objectives:

- Define NLP
- Introduce the concept of *information retrieval* and the key metrics of *precision, recall and F-score* which will later be used to measure how well an algorithm is able to detect buyer-supplier relations
- Introduce the concept of *information extraction* and, in particular, the sub-tasks *named entity recognition* and *relation extraction*
- Introduce *vectorisation* as a way to convert text into numeric vectors that can be fed into machine learning algorithms

2.5.1 Definition

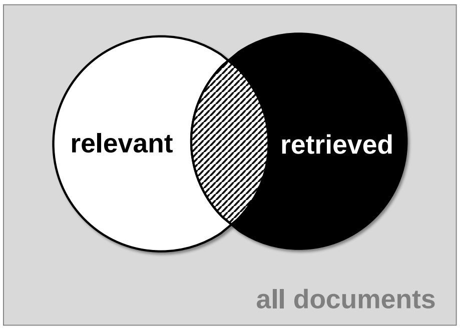
Natural Language Processing (NLP) is a sub-field of Artificial Intelligence and an "area of research and application that explores how computers can be used to understand and manipulate natural language text or speech to do useful things" (Chowdhury, 2005). Because of the complexity of natural language resulting in an intractable number of possibilities to

[4] `https://github.com/pwichmann/phd_thesis`

be realistically captured by rule-based programming techniques, modern NLP methods rely on Machine Learning. There are several common tasks in NLP, such as automatic machine translation of natural language, natural language understanding, and question answering.

2.5.2 Information retrieval

Information retrieval (IR) can be defined as "finding material (usually documents) of an unstructured nature (usually text) that satisfies an information need from within large collections (usually stored on computers)" (Manning et al., 2009). A typical example use case of information retrieval is a Web search but the activity of information retrieval predates the Web and was already relevant for library catalogues. The effectiveness of an IR system (that is the quality of its search results) is typically measured in *recall* and *precision*. Precision refers to the *fraction of the returned results that are relevant to the information need*, while recall refers to the *fraction of the relevant documents in the collection that were returned by the system* (Manning et al., 2009). Figure 2.11 illustrates their inter-relation. The formulae for precision and recall use a set notation with vertical bars on each side indicating the cardinality of a set.



$$\text{precision} = \frac{|\{\text{relevant documents}\} \cap \{\text{retrieved documents}\}|}{|\{\text{retrieved documents}\}|}$$

$$\text{recall} = \frac{|\{\text{relevant documents}\} \cap \{\text{retrieved documents}\}|}{|\{\text{relevant documents}\}|}$$

Fig. 2.11 Conceptual illustration of recall and precision; based on Manning et al. (2009)

Given the trade-off between recall and precision, the so-called F measure, the harmonic mean of both metrics, is commonly used to combine both in a single metric:

$$F_\beta = (1+\beta^2) \cdot \frac{\text{precision} \cdot \text{recall}}{(\beta^2 \cdot \text{precision}) + \text{recall}} \tag{2.1}$$

The parameter β in above formula specifies the weight assigned to recall and precision. More specifically, the formula places β times as much importance to recall as precision. In the case of an equal weighting of recall and precision ($\beta = 1$), the F measure is referred to as F_1 score:

$$F_1 = 2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}} \tag{2.2}$$

2.5.3 Information extraction

The task of information extraction (IE) can be defined as identifying "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. In other words, IE is about deriving structured factual information from unstructured text." (Piskorski and Yangarber, 2013). Information extraction is *different* from information retrieval: Whereas the goal of information retrieval typically is to return a list of documents selected and ranked by their relevance, the objective of information extraction is to "extract from the documents salient facts [...] in order to build more meaningful, rich representations of their semantic content" (Piskorski and Yangarber, 2013). The term *Open* (domain) Information Extraction, or *Open IE* for short and introduced by Banko et al. (2007), refers to an "extraction paradigm where the system makes a single data-driven pass over its corpus and extracts a large set of relational tuples without requiring any human input" (Banko et al., 2007). The idea is to extract all types of relations instead a small set of explicitly pre-specified relations. Generally, Open IE systems perform well when the objective is to extract general knowledge of an a-priori unknown domain from text. If there is a closed domain of interest, specialised systems will still tend to outperform Open IE systems. Typical tasks within the domain of information extraction are (Piskorski and Yangarber, 2013):

- **Named Entity Recognition (NER):** The detection and classification of predefined types of named entities, such as organisations or persons
- **Co-reference Resolution (CO):** The detection and resolution of multiple (co-referring) mentions of the same entity in the text, e.g. a pronoun used in a following sentence
- **Relation(ship) Extraction (RE):** The detection and classification of predefined relationships between entities in a given text, such as "LocatedIn"
- **Event Extraction (EE):** The detection of events in free text and the extraction of detailed and structured information about them (which might involve the recognition of multiple entities and possibly various relations between them)

Data mining and Web mining

Data mining is inconsistently defined and definitions range from activities of exploratory data analysis to a comprehensive process of knowledge discovery. Van Wel and Royakkers

(2004) define data mining as "the process of extracting previously unknown information from (usually large quantities of) data, which can, in the right context, lead to knowledge" (Van Wel and Royakkers, 2004). Based on above definition, *Web mining* is "the whole of data mining and related techniques that are used to automatically discover and extract information from web documents and services" (Van Wel and Royakkers, 2004).

2.5.4 Named entity recognition

Named Entity Recognition (NER) is the task of detecting and classifying named entities. Named entities are phrases that contain the names of persons, organisations, locations, times and quantities. An important feature for the detection of named entities is the capitalisation of the word. To improve the performance of NER systems, a pre-defined list of known named entities, called gazetteer, can be used. On common datasets, such as the Conll-03 (4 classes[5]), modern NER taggers[6] achieve an F_1 score of more than 90%. On datasets with an extended set of classes, such as Ontonotes 5.0[7] with 11 named entity classes and further 7 classes for times and quantities, current NER systems achieve F_1 scores of more than 85%[8].

There are a few important caveats with respect to the performance of NER systems: (1) The performance of NER systems is dependent on the class (e.g. person, location, organisation). Averaged F_1 scores may obscure poor performance for the class of organisations which is most relevant for this research. (2) NER systems often heavily rely on the capitalisation of the named entity. If the whole text sequence is written in lower or upper case (e.g. in headlines) and the NER system has been trained on text with different formatting, the detection becomes more difficult. (3) Current NER systems still do not appear to make full use of the wider context of a text sequence to disambiguate concepts. The context is important because the mention "Woodward" may refer to a company (Woodward, Inc.), a city (Woodward, Oklahoma) or a person (Bob Woodward). Especially family names are often also used as company names. The context, such as the verb used in combination with the mention, may indicate which class of named entity is present.

[5]Persons, locations, organisations and names of miscellaneous entities

[6]E.g. Flair (`https://github.com/zalandoresearch/flair`; last accessed: 2019-03-30), Spacy (`https://spacy.io/usage/facts-figures#benchmarks`; last accessed: 2019-03-30)

[7]`https://catalog.ldc.upenn.edu/docs/LDC2013T19/OntoNotes-Release-5.0.pdf`; last accessed: 2019-03-30

[8]`https://github.com/sebastianruder/NLP-progress/blob/master/named_entity_recognition.md`; last accessed: 2019-03-30

2.5.5 Named entity disambiguation

The same name can refer to different entities in different contexts. For instance, the mention "Delta" may refer to at least the following companies, Delta Airlines or Delta Faucet, depending on the context. Moreover, the same entity can be referred to by different names, such as "Toyota", "Toyota Motor Corporation", "Toyota Jidosha KK". Thus, there is a need to map mentions to the corresponding unique entity. *Named entity disambiguation* (NED) is the "problem of establishing these mappings between the mentions and the actual entities" (Hoffart et al., 2011). Ironically, the same problem (or slight variations thereof) is also known under different names, i.a. entity name disambiguation, named entity resolution, named entity linking, named entity normalisation or named entity grounding. An example for a system that performs NED is AIDA (Yosef et al., 2011), which leverages the YAGO knowledge base to disambiguate mentions based on the textual context they appear in.

The overall process of recognising mentions in unstructured texts and disambiguating them by mapping them to unique entities stored in a knowledge base is sometimes also referred to as *Named Entity Recognition and Disambiguation* (NERD). While the two tasks have often been performed sequentially, there have been more recent attempts to perform both tasks jointly (e.g. see Luo et al. (2015)).

2.5.6 Relation extraction

Definition

Relation Extraction is the task of detecting and classifying relations between entities from unstructured text. Commonly, the process consists of named entity recognition, followed by a relation classification. For binary (dyadic) relations, the relation between each pair of named entities is classified. The output of the relation extraction would then be a set of triples, each consisting of first entity, second entity, and the identified relation type.

Relation extraction methods

Relation extraction methods are not necessarily mutually exclusive but can be used in combination (in parallel or in sequence). The following general approaches for relation extraction can be distinguished: (a) Pattern-based, (b) Bootstrapping, (c) Supervised (learning), (d) Distant supervision, (e) Unsupervised.

(a) Pattern-based Before the rise of Machine Learning, the common way to extract relations was using pre-defined lexico-syntactic patterns, also referred to as Hearst (Hearst,

1992) patterns named after the author who first popularised the method. Using stemming or lemmatisation, the patterns can be generalised so that patterns do not need to be created for each possible verb inflection. The advantage of this method are the high precision (that is, results returned by well-defined patterns tend to be correct) and the ease at which the method can be understood. The obvious downside is the effort required to manually define the patterns, and the generally low recall. Furthermore, it is difficult to come up with any quantitative rationale why some patterns are used and others are not since inclusion/exclusion criteria are not learned from the data.

(b) Bootstrapping Bootstrapping (Mintz et al., 2009) is the iterative process of starting a relation extraction process with a few seeds, such as confirmed tuples of entities in a specific relation. These seeds can then be used to extract a new set of patterns from a large text corpus. These new patterns can then be used to extract more instances, which can in turn be used to identify more patterns, and so on. Despite being more efficient than pattern-based approaches, and being able to learn patterns from the data, the downside of this approach is the *semantic drift*. Without supervision, the corrective mechanism is missing to prevent the algorithm from slowly starting to return false positives resulting in low precision.

(c) Supervised Supervised relation extraction consists of first hand-labelling sentences in a corpus with respect to the presence of entities and relations between them (cf. Mintz et al. (2009)). A supervised classifier can then be trained given a pair of entities, the provided label, and a wide range of features extracted from each sentence. The advantage of a supervised approach over a pattern-based approach is that it can learn the patterns from the data itself and without any manual effort. This way supervised learning is much more capable at dealing with the large number of possible ways a relation can be expressed in a language. The classification performance can always be improved by adding further annotated data to the training set. Features for the classification process can be extracted automatically and feature weights can be learned. Supervised classifiers also output a probability score that can be taken into consideration. Supervised relation extraction suffers from the obvious disadvantage that labelling is costly and labelled training data is therefore often limited in quantity. To learn a pattern from the data itself, a large number of positive and negative examples are required. Furthermore, a general problem of supervised methods is the bias that can be introduced by the chosen text corpus or the labelling process.

(d) Distant supervision Distant supervision was first introduced by Mintz et al. (2009). Distant supervision attempts to address the problem that labelled data is difficult to obtain.

The idea of distant supervision is to leverage a database of known relations between entities and a database of documents containing those entities. This way, distant supervision combines a supervised approach (database of known relations between entities) with an unsupervised one (database of unstructured text). The basic assumption of distant supervision is that "if two entities participate in a relation, any sentence that contain those two entities might express that relation" (Mintz et al., 2009).

(e) Unsupervised Unsupervised relation extraction methods do not require labels and use named entities and the text itself as input. Examples of unsupervised methods are based on statistical co-occurrence (i.e. entities that are frequently named in the same context have some form of relation) or attempt to extract and simplify the text between entities (e.g. Shinyama and Sekine (2006), Banko et al. (2007)). The advantage of unsupervised methods is that they do not require costly labels and no biases are introduced by a labelling process. When using only unsupervised methods, however, relations cannot be classified, just clustered. In particular, simple co-occurrence would not be sufficient to derive the directionality of a relation.

2.5.7 Vectorisation and word embeddings

Machine Learning algorithms generally expect numeric tensors as input. In order to use a sequence of text as input to a Machine Learning algorithm, it first needs to be converted into numeric tensors. This is typically done in two steps:

1. *Tokenisation*: the text is broken down into tokens, e.g. individual words
2. *Vectorisation*: each unique text token is converted into a numeric vector

The overall process is depicted in Figure 2.12 and explained in the following sections.

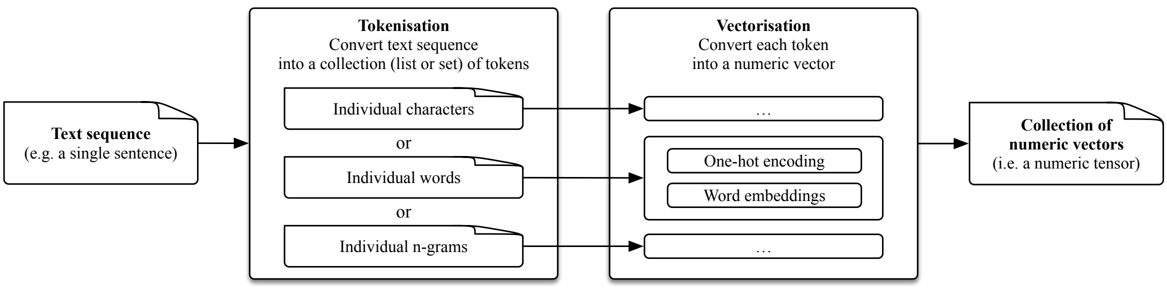


Fig. 2.12 Converting text into numerical tensors through tokenisation and vectorisation

Tokenisation

Tokenisation is the process of segmenting a text sequence into smaller text chunks. Commonly used tokens for Machine Learning are individual characters, individual words, or individual n-grams[9] of words or characters.

Vectorisation

Vectorisation is the process of converting a text token into a numeric vector. Commonly used vectorisation methods are *one-hot encoding* or *(word) embeddings* (Chollet, 2017, p. 180).

One-hot encoding One-hot encoding refers to the process of assigning every unique word to a unique integer index. The vector for a word i then becomes a binary vector of size N where N is the size of the vocabulary. The vector only has the value 1 in entry i but is zero for all other entries (Chollet, 2017, p. 181). The obvious problem with one-hot encoding is that the resulting vectors are very high-dimensional (same dimension as the number of unique words in the vocabulary) and sparse (mostly consisting of zeros). Another downside is that words lose their meaning once they are encoded. Distances between vectors also do not correspond to semantic distances between the encoded words.

(Word) embeddings Similar to one-hot encoding, word embeddings are vectors that represent unique words. However, word embeddings are real-valued, low-dimensional, and dense vectors. Furthermore, the remarkable property of word embeddings is that these vectors are able to encapsulate semantic relationships between different words. For example, geometric distances correspond to semantic distances (e.g. synonyms are encoded by similar vectors) and directions in the geometric space are meaningful (e.g. from the 'male' to the 'female' version of a concept, or from the singular to the plural form of a concept).

Word embeddings are learned from the data: They are either learned anew as part of the task at hand using an *embedding layer* of a neural network, or they are re-used as pre-trained word embeddings that were obtained in another task. The basic idea for obtaining word embeddings is to train a Deep Learning classifier on a large number of text sequences where a single word has been removed and needs to be guessed based on the context. The assumption underlying the techniques that convert text into word embeddings is the *distributional hypothesis* (Harris, 1954) which suggests that words occurring in the same context tend to have similar meanings. The context for each word is its nearby words.

[9] N-grams are all overlapping groups of n or less *consecutive* words or characters that can be generated from an input text; for the case $n = 2$, the n-gram is called a bigram.

In 2013, a team at Google led by Tomas Mikolov published an approach and toolkit for the efficient learning of word embeddings, called *word2vec* (Mikolov et al., 2013), that popularised the concept of using word embeddings. Another subsequently developed method is *GloVe*[10] (Pennington et al., 2014).

These methods have been applied to large datasets, such as news archives and website archives, and the results have been made available as pre-trained word embeddings. The advantage of these pre-trained word embeddings is that they have been trained on large-scale datasets, and the large computational effort has already been invested. One example are the Word2Vec Google News[11] word embeddings that were trained on parts of the Google News dataset (about 100 billion words) and consist of 300-dimensional vectors for 3 million words and phrases. One of the available GloVe word embeddings dataset was trained on 840 billion tokens obtained from a Common Crawl dataset, and resulted in a vocabulary size of 2.2 million where each word was mapped onto a 300-dimensional word embedding vector. Subsequently, word embeddings have also been created for languages other than English, such as Facebook's fastText[12] that was trained on Wikipedia articles.

Young et al. (2017) summarise the recent development as follows: "For decades, machine learning approaches targeting NLP problems have been based on shallow models [...] trained on very high dimensional and sparse features. In the last few years, neural networks based on dense vector representations have been producing superior results on various NLP tasks. This trend is sparked by the success of word embeddings and deep learning methods. Deep learning enables multi-level automatic feature representation learning. In contrast, traditional machine learning based NLP systems liaise heavily on hand-crafted features. Such hand-crafted features are time-consuming and often incomplete".

2.6 Automated generation of networks or supply chain maps from text

This section summarises work at the intersection of both the problem domain and the solution domain. The focus is on the automated generation of networks or graphs, such as supply networks, from text. Figure 2.13 provides a indicative and subjective positioning of selected prior academic work with respect to the research presented in this thesis. This landscape shows how prior work compares to the problem domain of the thesis (supply chain mapping)

[10]Global Vectors for Word Representation; `https://nlp.stanford.edu/projects/glove/`; last accessed: 2019-03-30

[11]`https://code.google.com/archive/p/word2vec/`; last accessed: 2019-03-30

[12]`https://github.com/facebookresearch/fastText`; last accessed: 2019-03-30

and the proposed solution space (automated relation extraction from text). The literature is discussed in the context of key themes arising in this field.

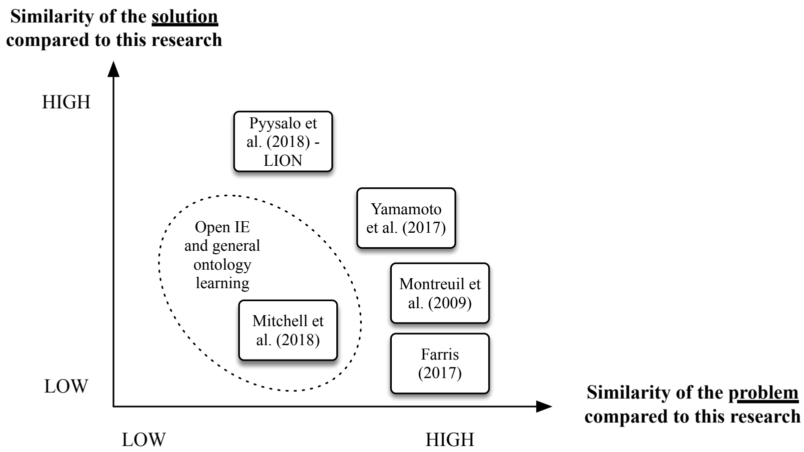
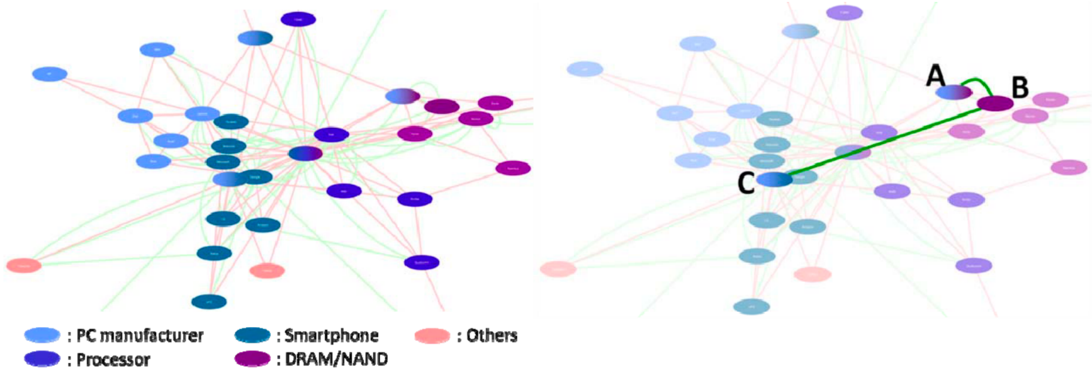


Fig. 2.13 Indicative positioning of this research in the landscape of a selection of most closely related prior work.

Company relation extraction from unstructured text A recent paper by Yamamoto et al. (2017) attempts to extract company relations from text but only focusses on extracting cooperative[13] versus competitive relations using distant supervision and manual labels. A resulting map showing the network of collaborative versus competitive company relations is depicted by Figure 2.14. Company roles were labelled manually; only the cooperative or competitive relations were extracted automatically. Within cooperative relations, the length of the edge indicates different sub-types: joint ventures were depicted with a shorter edge, indicating close cooperation. Buyer-supplier relations ("supply-demand" relations) were depicted with a longer edge. The paper does not clarify if this sub-categorisation was performed manually or automatically; the methodology section of the paper appears to suggest this was not done automatically.

While the classification into cooperative versus competitive relations may seem to also be helpful with the extraction buyer-supplier relationships, it actually is not: Competitive clues include terms like "sues", "lawsuit" or "loses". Because buyers and suppliers not uncommonly happen to have legal disputes, these clues can be misleading for the purpose of supply chain mapping. Furthermore, the directionality of relations was not considered.

[13]The paper uses both terms "cooperative" and "collaborative" interchangeably.



(a) Network colored for each product **(b) Example of two types of cooperative relation**

Fig. 2.14 Network of collaborative vs. competitive company relations; source: Yamamoto et al. (2017))

Concept of a supply chain mapping tool Montreuil et al. (2009) present the concept of a tool, dubbed "Supply Web Mapper", which "is a visualisation, mining, and assessment application allowing analysts or decision-makers to explore in a summarised and efficient way the Supply Web created through the interactions of multiple organizations". The tool would allow "a visual mining of a Supply Web by exploiting available partner's databases while protecting their confidentiality". The work is conceptual in nature and assumes data being "supplied by inter organizational information systems" rather than being extracted from text documents.

Mapping supply chains on an industry-level using macro-economic data (input-output analysis) Input-output models (see Christ (1955) for a review) are an established[14] technique in economics and are quantitative models that represent the interdependencies between different industries as well as regional or national economies. An example of a network visualisation resulting from an input-output analysis is shown in Figure 2.15. This map shows interdependencies between national motor vehicle clusters. Each box contains the name of a country. Arrows between these boxes point in the direction of payments, whereas goods flow in the opposite direction. Numbers indicate the percentage of imports that is imported from a particular source country.

[14] The economist Leontief (1936) is generally credited for the idea; the original idea can be attributed to Francois Quesnay and his Tableau Economique (1759)

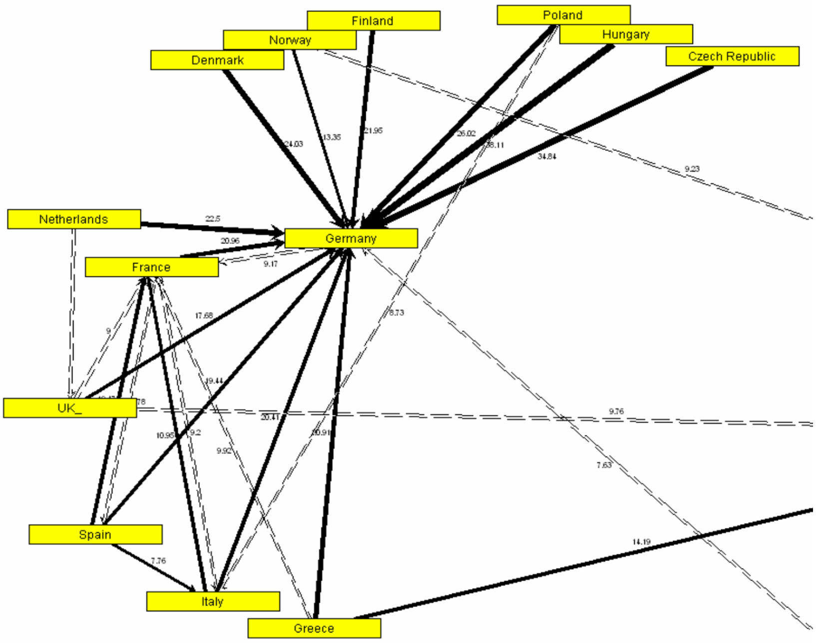


Fig. 2.15 Example of an input-output analysis: Interdependencies between national motor vehicle clusters 1995; cutout of the map by Wixted (2006)

Farris (2010) attempts to address the problem of finding actual data for use specifically in strategic supply chain mapping by using economic input-output data (input-output analysis). This data was then converted into macro industry supply chain maps. Figure 2.16 shows a plastics and rubber products industry supply chain macro map by Farris (2010). The nodes are represented by different geometric shapes depending on whether industries are suppliers or customers (or both) with respect to the focal industry (triangle). The nodes stand for industries (not companies) and their names correspond to bureau of economic analysis identifier codes. Arcs indicate the *financial flows* related to the buying (dashed arcs) and selling (solid arcs) of goods and services. Percentages indicate the share of total purchase spend (dashed arcs) or sales revenue (solid arcs).

The process of generating the maps was manual and the input-output analysis based on macro-economic data generally does not allow for any maps on a company-level. However, these maps could provide a useful additional top-down perspective for supply chain maps on company-level.

Ontologies and ontology learning An ontology is “an explicit specification of conceptualization” (Gruber, 1995) for the purpose of sharing and reusing knowledge among software entities. “Ontologies serve as metadata schemas, providing a controlled vocabulary of con-

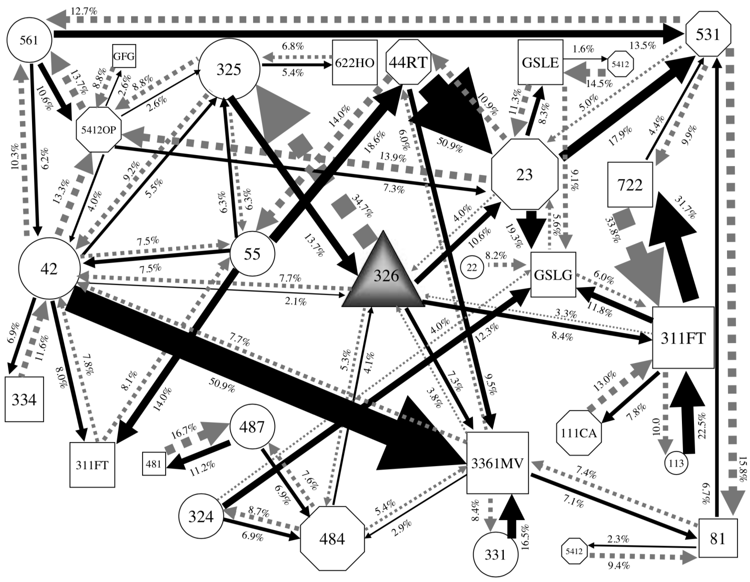


Fig. 2.16 Industry supply chain macro map; source: Farris (2010)

cepts, each with explicitly defined and machine-processable semantics." (Maedche and Staab, 2001). Ontologies are used in the context of the Semantic Web and Linked (Open) Data. The technical standard maintained by the W3C is the Resource Description Framework (RDF[15]) which relies on triples to form large graphs. These triples commonly take the form of Subject-Predicate-Object or Subject-Property-Value. An simple example of such a triple would be "ex:cat rdfs:subClassOf ex:animal". These structures can be defined in an RDF Schema (RDFS) or using the Web Ontology Language (OWL[16]) as an extensible knowledge representation data model. Populated data graphs can then be queried using languages such as SPARQL[17]. Such queries can follow long chains of predicates or properties (predicate paths / property paths[18]). Once a data graph has been built for buyer-supplier relations, it could be queried this way to check if a company is an Nth-tier supplier to another company.

Ontologies have already been designed to store information about organisations and various Linked Open Vocabularies (LOV) offer specifications for organisations and even their inter-relations. E.g. the W3C[19] offers an ontology for the concept of an organisation. To store which companies stand in some relation to another company, the relation "linkedTo" is offered. The specification even mentions supply chain relations as an example of which

[15] `https://www.w3.org/RDF/`; last accessed: 2019-07-31
[16] `https://www.w3.org/TR/owl2-overview/`; last accessed: 2019-07-31
[17] `https://www.w3.org/TR/rdf-sparql-query/`; last accessed: 2019-07-31
[18] `https://www.w3.org/TR/sparql11-query/#propertypaths`; last accessed: 2019-07-31
[19] `https://www.w3.org`; last accessed: 2019-07-31

relations may be describes this way: “Indicates an arbitrary relationship between two organizations. Specializations of this can be used to, for example, denote funding or supply chain relationships.”[20]. Schema.org[21] provides a wide-ranging vocabulary covering entities and relations. The vocabulary covers specifications for the concept “Organization”[22] as well as “Corporation”[23]. The definition of a corporation also includes the Legal Entity Identifier (LEI[24]), which can be used to uniquely identify a company and to determine ownership relations between legal entities. Other ontologies have been designed to store and manage information about contracts and contractual relationships[25]. *Web Data Commons*[26] is a project that aims to extract “extract structured data from the Common Crawl, the largest web corpus available to the public”. The project does not analyse sentences in natural language but instead collects embedded Linked Data, which news organisations sometimes include in their articles. In summary, ontologies, the concept of Linked Data and related technologies could be relevant for various reasons:

- Machine-understandable knowledge representation data model
- Existing querying options for data graphs
- Potential to integrate data from other data sources (either buyer-supplier relations if they have been collected or other relevant information, such as company’s alternative names or associated brand names and products)
- Potential to reuse outputs of the supply chain data graph in other applications

However, existing ontologies have not been populated with sufficient data at the right level of granularity to be useful for supply chain mapping at this point. But they already provide a basic data structure to store and manage such information that can be further extended. The idea of *ontology learning* proposed by Maedche and Staab (2001) involves the use of Machine Learning techniques to (semi-)automatically construct ontologies and, thus, to acquire knowledge – typically from text sources. Buitelaar et al. (2005) organise the aspects of ontology development in form of a layer cake of increasingly complex subtasks, as illustrated in Figure 2.17.

[20] `https://www.w3.org/ns/org`; last accessed: 2019-07-31
[21] `https://schema.org`; last accessed: 2019-07-31
[22] `https://schema.org/Organization`; last accessed: 2019-07-31
[23] `https://schema.org/Corporation`; last accessed: 2019-07-31
[24] `https://www.gleif.org/en/about-lei/introducing-the-legal-entity-identifier-lei`; last accessed: 2019-07-31
[25] E.g. Open Contracting Data Standard (OCDS): `https://standard.open-contracting.org/latest/en/schema/reference/`; last accessed: 2019-07-31
[26] `http://webdatacommons.org/`; last accessed: 2019-07-31

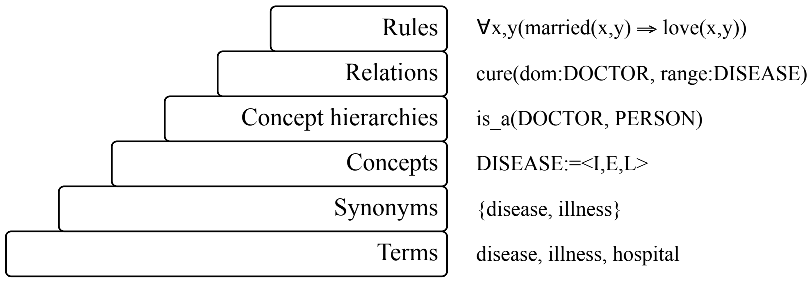


Fig. 2.17 Ontology layer cake (illustration adapted from Buitelaar et al. (2005))

Dong et al. (2014) cluster the literature on *automatic knowledge base construction*[27] into 4 groups:

1. Approaches which are built on Wikipedia infoboxes and other (already) structured data sources (e.g. YAGO[28] , DBPedia[29], and Freebase[30])
2. Approaches which use open information (schema-less) extraction techniques applied to the entire web (e.g. OLLIE[31] by the University of Washington)
3. Approaches which extract information from the entire web, but use a fixed ontology / schema (e.g. NELL[32] of the "Read The Web" project by Carnegie Mellon University)
4. Approaches which construct taxonomies (is-a hierarchies) as opposed to general knowledge bases with multiple types of predicates (e.g. Probase[33])

The "Knowledge Vault" proposed by Dong et al. (2014) extracts information from the entire Web but fuses together facts extracted from text with prior knowledge derived from the Freebase graph. Biemann (2005) provides a comprehensive overview of the different methods of ontology learning, in particular from unstructured text. Not surprisingly, they are similar to the methods of relation extraction and include Hearst patterns, supervised (classification) and unsupervised Machine Learning techniques. Mitchell et al. (2018) define a never-ending learning paradigm for machine learning, and present the Never-Ending Language Learner

[27] It appears that the popularity of the term ontology has decreased over the last years, and similar work is now referred to as the automated creation of *knowledge bases* or *knowledge graphs* while still using the identical underlying concepts and technologies known from ontologies.
[28] `https://www.mpi-inf.mpg.de/departments/databases-and-information-systems/research/yago-naga/yago/`; last accessed: 2019-04-25
[29] `https://wiki.dbpedia.org/`; last accessed: 2019-04-25
[30] Now deprecated; `https://developers.google.com/freebase/`; last accessed: 2019-04-25
[31] `https://knowitall.github.io/ollie/`; last accessed: 2019-04-25
[32] `http://rtw.ml.cmu.edu/rtw/`; last accessed: 2019-04-25
[33] `http://haixun.olidu.com/probase/index.htm`; last accessed: 2019-04-25

(NELL). The system has been learning to read the Web 24 hours per day since January 2010, and acquired a vast knowledge base with 120mn diverse, confidence-weighted beliefs (e.g. servedWith(tea,biscuits)). The authors state that NELL has also learned to reason over its knowledge base to infer new beliefs – based on those beliefs it already has. And NELL has "a primitive but already-useful ability to extend its representation by suggesting new relational predicates (e.g., RiverFlowsThroughCity(x,y)) between existing categories (e.g., river, city).". The system was, however, given an initial ontology, some initial labelled examples, and some feedback through occasional interactions with humans.

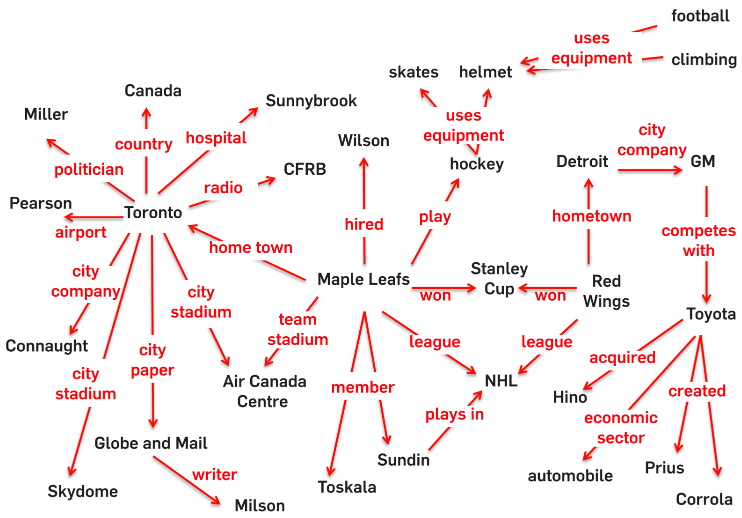


Fig. 2.18 NELL knowledge fragment; from Mitchell et al. (2018)

The proposed method and system is related to automated supply chain mapping in the sense that a knowledge graph can be generated by the collection of semantic triples[34] (two concepts linked by a relational predicate). As opposed to a domain-specific ontology, like one for supply chain mapping, the ontology created by NELL holds common, general knowledge. Ontologies, such as DBpedia[35], contain relations that are potentially relevant from a supply chain mapping perspective. The class "Organisation"[36] provides properties such as the location, the product or the parent organisation. While the data structure of ontologies like DBpedia could in principle store buyer-supplier relations, such information was not available in tests[37]. Fu et al. (2009) propose a technical concept for an ontology-based supply chain

[34]E.g. see the atomic data entity in the Resource Description Framework (RDF), a common data model for ontologies; `https://www.w3.org/TR/rdf11-concepts/`; last accessed: 2019-03-30
[35]`http://wiki.dbpedia.org/services-resources/ontology`; last accessed: 2019-03-30
[36]`http://dbpedia.org/ontology/Organisation`; last accessed: 2019-03-30
[37]E.g. see the entry for a large company, such as BMW (`http://dbpedia.org/page/BMW`; last accessed: 2019-04-12)

information management system. However, the system is not designed to discover the supply chain structure or extract this information from natural language text.

Mining graphs of co-occurrences from biomedical texts The LION Literature-Based Discovery (LBD) project[38] (Pyysalo et al., 2018) uses statistical co-occurrence to extract relations from bio-medical texts and visualise them as an interactive network. Even though the problem domain is unrelated to the research problem at hand, the solution is similar: A network of relations is automatically generated from natural language text. The purpose of the proposed method and tool is to facilitate the discovery of new knowledge from scientific papers in the bio-medical domain. Because the system uses co-occurrence only, relations are not classified (or directional).

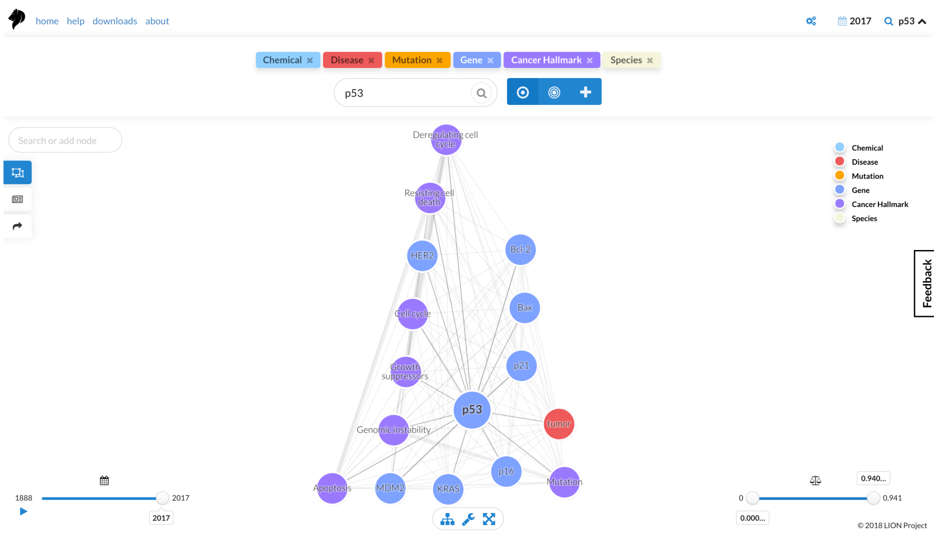


Fig. 2.19 Project LION: Extracting relations from biomedical text (Screenshot taken from `http://lbd.lionproject.net`)

The identified research gap as well as the research approach will be summarised in the following chapter.

[38]`http://lbd.lionproject.net`; last accessed: 2019-02-24

Chapter 3

Research gap & research approach

3.1 Introduction

The previous chapters outlined the industrial motivation (Chapter 1) and provided the necessary academic background (Chapter 2). The purpose of this chapter is to present the research gap, to derive a research scope, and to develop an appropriate approach to address the research questions.

This chapter is structured as follows: The first section summarises the research gap from the Research Background chapter. The research gap provides the academic motivation for this work as it documents the novelty of the research. Based on these findings, the research scope is derived in the next section and further narrowed down to the research focus in a subsequent section. Next, the research questions are presented as well as the arguments underpinning the decisions. Lastly, the research approach is outlined.

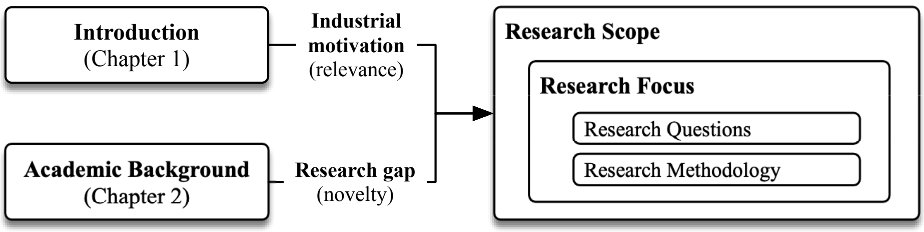


Fig. 3.1 Structure of Chapter 3: The research scope is derived from the research gap and industrial motivation and then narrowed down to the research focus.

3.2 Research gap from literature review

The literature review in Chapter 2 summarised prior academic work in the problem space (supply chain management and structural supply chain visibility), the solution space (machine learning and natural language processing), and prior academic work at the intersection of both.

Section 2.3 distinguished structural aspects of supply chain visibility from more dynamic, operational ones. Previous research not only indicates that a company is impacted by its extended supply chain structure, it also suggests that knowledge of the structure ("structural supply chain visibility") is a valuable input to a wide range of management decisions. The criticality of suppliers does not always correlate with large sales volume, and critical suppliers can be located in any (sub-)tier of the supply network. Supply chain mapping is frequently cited in literature to be the solution for establishing visibility into the supply chain structure. However, obtaining the necessary information required to map the structure is a problem. Existing manual approaches, e.g. using supplier surveys, have drawbacks and could at least be complemented by automated approaches.

The application of machine learning techniques to text made it possible to automatically extract (structured) information from (unstructured) text. Achieving F_1 scores of above 85% at least on shared datasets, automated systems for named entity recognition now perform reasonably well, as shown in Section 2.5.4. Relation extraction, the detection and classification of relations between entities in a given text, can also be automated using Machine Learning classification methods applied to text inputs. The concept of word embeddings allows to convert words into real-valued, dense vectors that encapsulate the semantic relationships between different words. A number of different Relation Extraction methods have been developed (Section 2.5.6). Analogous to Machine Learning in general, they can be distinguished according to the extent of supervision required. Simple statistical co-occurrence of entities does not require supervision but cannot distinguish different types of relationships. Supervised methods can distinguish different types of relationships but require labelled training datasets.

Open IE solutions and general ontology learning methods exist that aim at extracting basic relations, such as "partOf" or "locatedIN", from a text. However, these solutions are typically only able to extract basic relations and cannot classify buyer-supplier relations. The extraction of more specific relations from text in a closed domain, such as biomedicine, is a common topic of applied NLP research. A recent paper by Yamamoto et al. (2017) attempts to extract company relations from text but only focusses on cooperative versus competitive relations. This is insufficient for extracting buyer-supplier relations in the context of supply chain mapping.

A literature review did not reveal prior academic work where specifically the process of generating supply chain maps from text was automated using Natural Language Processing methods.

Summarising the research gap In summary, the research gap with respect to automatically generating supply chain maps from unstructured, natural language text can be characterised by the following aspects:

- **Lacking an overall approach for automating supply chain mapping:** What challenges need to be addressed to fully automate the complete supply chain mapping process? A comprehensive approach for automating the mapping process does not appear to have been published.
- **Unknown availability and sparsity of information about supply chain structure:** Because an automated approach does not exists yet, it has not been established how much information about buyer-supplier relations can be expected to be contained in any large, openly available text archive (e.g. general news). This refers to both information availability in absolute terms (how many relations?) as well as relative terms – often referred to as the "density" or "sparsity" of a dataset (ratio of relevant relations to overall text quantity).
- **Unknown ambiguity of the buyer-supplier relation labelling task:** It is unclear how ambiguous the task of labelling is for a human annotator who is provided with a sentence and general labelling instructions to identify buyer-supplier relations. A labelled dataset is required to measure the performance of any automated classification and to train a classifier. If human annotators cannot agree on the correct label, this would indicate a training dataset of poor quality and suggest too much ambiguity for an attempt to automate the process (be it because of the task itself or the provided instructions). What is the achievable inter-annotator agreement and what is a sensible approach for the label collection?
- **Unclear buyer-supplier relation class labels and relation sub-types:** In advance of this research, it is not immediately obvious what useful class definitions for buyer-supplier relations could be and which sub-types would require distinction. Which classes should be distinguished to ensure high-quality labels?
- **Unknown performance of the buyer-supplier relation classification:** It is currently unknown how well an (automatic) classifier will perform on a labelled dataset. The usefulness of the approach will depend on the misclassification costs for any particular

use case. It is also not known if common features like word embeddings are sufficient for this task and which relation classes are most likely to be confused. Additionally, it is unknown how large a labelled dataset needs to be to achieve reliable results during testing or when the classifier is applied to different datasets. A large training dataset may be required to capture the large number of linguistic variations.

In addition, although not a research gap in the strict sense, there is also a *lack of labelled datasets* to be used as a reference for this problem. Labelled datasets for buyer-supplier relations are datasets that contain both the sentence as well as labels indicating the type of relation that is expressed between organisational named entities mentioned in that sentence. Those datasets could be used for training a classifier but do not appear to be available. Even if confirmed buyer-supplier tuples could be obtained from a database, some form of human supervision is required since these companies can be mentioned together in different contexts.

The research presented in this thesis aims to address the gaps mentioned above. However, given the complexity of the problem, this study can only provide first conclusions rather than the ultimate answers. For example, any attempt to answer questions about information availability and sparsity is severely constrained by imperfect information processing, limited corpus sizes and other factors. Nevertheless, these early indications appear useful in guiding future attempts at automating the generation of supply chain maps from unstructured text.

3.3 Research scope: Automatically generating supply chain maps from text

This section defines the chosen research scope given the research problem outlined before. It also provides the rationale for the scoping decisions as well as makes explicit some of the underlying assumptions for this research to clarify expectations. The research scope shall be defined as follows:

Research scope:
Automatically generating supply chain maps from unstructured, natural language text

As stated before, this research does *not* claim to fully replace the human research currently necessary to map supply chains – neither regarding the breadth of the research tasks nor regarding the achievable performance level. The objective, however, is to design and test an end-to-end approach for converting text into basic supply chain maps that does not rely on repeated manual processing steps. Such an approach makes an initial contribution to automated supply chain mapping development.

In the following, the rationale for various scoping decisions shall be provided. Table 3.1 summarises the decisions; the decisions are discussed subsequently.

Table 3.1 Scoping decisions

Scoping aspect	Chosen scope
(A) Data type	Unstructured, natural language text
(B) Degree of automation	Automated supply chain mapping approach
(C) Industry	Industry-agnostic where possible
(D) Accessibility: Public versus privately held data	From technical standpoint no difference once data has been obtained. Assumption that publicly available text sources will at least also be used.
(E) Language	Language-agnostic where possible; English language where focus is necessary (e.g. practical experiments)
(F) Use case	Use-case-agnostic where possible; focus on the common information requirements shared by use cases of supply chain maps: extraction of supply chain participants and their inter-relations

(A) Decision: Selecting unstructured, natural language text (vs. other data types)

To obtain supply chain participants and their inter-relations, a variety of data types can be considered. Data types are an important criterion because the required technical solution may differ substantially across different types. And so the required effort needs to be weighed against the expected benefit of being able to process this data type, which in turn is determined by the availability of data of that type and the expected usefulness. An overview of common data types is provided by Figure 3.2.

It needs to be justified why the focus of this research is on unstructured natural language text as opposed to other data types. Table 3.2 provides a summary of that justification.

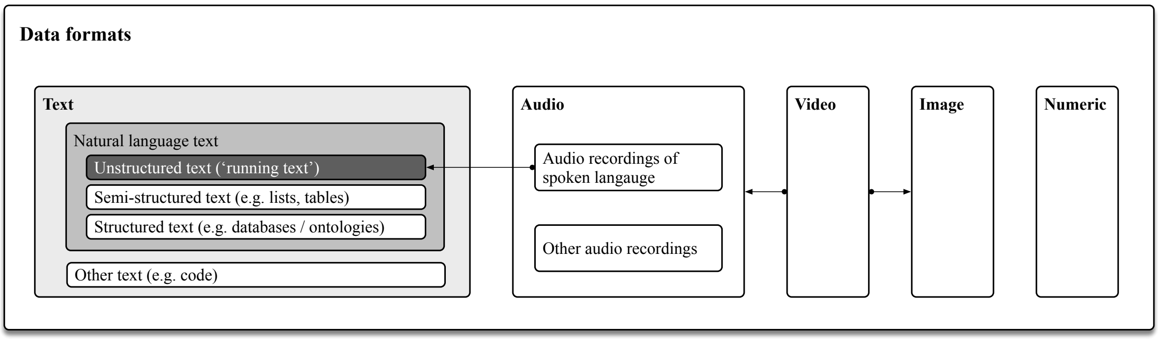


Fig. 3.2 Common data types: Focus on unstructured natural language text

Table 3.2 Data types

Data type	Discussion	Focus?
Numeric	Purely numeric data does not appear to be helpful when it comes to *identifying* supply chain participants and their relations. However, numeric delivery data may be useful to detect if parts shortages from different direct suppliers are correlated and may indicate shared sub-tier suppliers. Economic data also may help with quantifying the value of goods and services exchanged between industries or economies.	NO
Image	There are images on the Web that visually describe from which suppliers parts of a product are sourced. However, the automated analysis of these images is difficult for at least two reasons: (1) The relations are often described using a combination of text and visual clues, and (2), there is no standard for the visual clues so that any automated information extraction becomes difficult.	NO
Audio	Audio of *spoken language* can nowadays be automatically transcribed into text. The processing of recorded language is therefore treated identical to text. Other types of audio data do not appear to be useful for solving the problem at hand.	NO
Video	Video data can be split into image and audio data; see arguments there	NO
Text	A large proportion of published human communication is in text form and text documents. For example, news articles and blog posts contain valuable information about buyer-supplier relations. Furthermore, text in natural language follows a standardised structure (grammar) and uses words with known meanings for a given context.	YES

The data type “text” can be further broken down allowing for a more refined scope:

Natural language Text can be categorised into *natural language* text (e.g. text in in any natural language like English) and other text forms (e.g. text written in a programming language like JavaScript). Supply chain information is assumed to be contained in human communication, and hence *natural language* text.

Unstructured text Natural language can further be characterised by the extent to which it is structured. Normal running text is considered *unstructured* text. Text with some structure but no explicit schema, e.g. text in lists or tables or in JSON or XML format, is called *semi-structured*[1]. Text stored in databases or ontologies / knowledge graphs with an explicit formal schema that describes the data structure is called *structured* text. Since information relevant for supply chain mapping is not widely available in (semi-)structured format, the focus is on *unstructured text*. Common ontologies, like dbpedia, provide the structure that would allow to store such information. However, these structures have hardly been populated with supply chain related data – most likely because that information is difficult to obtain and the extraction of buyer-supplier relations cannot be automated yet. The fact that the research scope is on unstructured, natural language text does *not* mean that extracted information cannot or should not be combined with information from other databases or knowledge bases if they exist. In fact, information extraction from unstructured text aims at creating and extending such knowledge bases so that the contained information can be easily (re-)used.

Processing natural language text is challenging due to the large number of variations. But intuitively, considering news reports, market reports, company websites, it appears that most of the stored information about supply chain structures is likely stored in text form. The text corpus that is the Web is continuously updated, and contains both official accounts (press releases, SEC filings, legal documents, ...) and unofficial accounts (blog posts, social media entries, ...). The content of websites can be accessed automatically, and large archives of websites have been made available for free[2].

(B) Decision: Automation of supply chain mapping (vs. manual approaches)

Manual attempts of acquiring information for supply chain mapping are time-consuming and costly, e.g. see Farris (2010). The dynamic nature of supply chains would also require continuous updates to the supply chain maps. Thus, it appears desirable to design an

[1] A comprehensive definition of semi-structured text can be found in Abiteboul (1996).

[2] The Common Crawl project stores scraped website contents on a daily basis and makes the archives publicly available.

automated approach that could at least complement any manual processes. If the data sources are large archives of scraped Web documents, then manually analysing these would not be feasible anyways.

(C) Decision: Determining industry focus

The aim of this research is to be industry-agnostic where it is possible. If the scope has to be reduced, the automotive and aerospace industry shall be prioritised. These industries are known to have large global supply networks, require the shipping of physical material and components, and tend to be well-covered by general news articles. By attempting to map supply chains from text documents, there may be an implicit bias towards companies whose supply chains revolve around physical goods rather than information or software. Language can also be industry-specific, such that a classifier trained on text about one industry may not perform as well as on text about another industry.

(D) Decision: Accessibility – public data versus privately held data

From a technical standpoint, there is no difference in processing publicly available or privately held text once it has been obtained. Differences, however, may occur in the data acquisition and possibly in the pre- and post-processing. Text obtained from the Web comes from diverse sources, in diverse formats, with unknown credibility. Companies could potentially leverage their own internal data, such as the text content of internal emails or presentations. It is generally assumed that publicly available text sources will at least also be used since data is available in vast quantities and at low cost (if not for free). The question of copyright and usage rights shall not be considered as a scoping criterion. Recent jurisdiction in the US appears to suggest that scraping of public websites is legal, such as in the case of HiQ Labs. vs. LinkedIn[3]. Public domain text sources are also available in large quantities, such as Wikipedia dumps[4].

(E) Decision: Language focus

The aim of this research is to be language-agnostic where it is possible. However, for the experimental parts of this research the English language shall be preferred since most research focusses on English due to the availability of English texts, and thus, models for the English language tend to perform best.

[3] `http://cdn.ca9.uscourts.gov/datastore/opinions/2019/09/09/17-16783.pdf`; last accessed: 2019-09-17

[4] `https://dumps.wikimedia.org/`; last accessed: 2019-09-17

(F) Decision: Single specific use case vs. commonly shared information needs

An important consideration is the decision between the option of focussing on a single specific use case and the option of focussing on commonly shared information needs across many use cases. Because supply chain maps can serve very different purposes and, hence, can have very different forms and informational requirements, there are two options how the research problem can be approached.

One option is to investigate a particular decision problem a company is confronted with in depth, and then work backwards what the information needs are and how these can be addressed. Following this approach ensures that the particular needs of the problems are considered. On the other hand, this approach is likely to result in a large list of diverse information requirements that have to be met by using data of different data types from a wide range of sources and processing it with various different techniques.

Another option is to acknowledge that different kinds of supply chain maps are a valuable input for a variety of decisions a company has to make and to identify the informational needs that all these use cases have in common. This way the information needs shared by all of the typical use cases are prioritised and the research results are relevant for all of these use cases. Figure 3.3 illustrates one aspect of this decision conceptually.

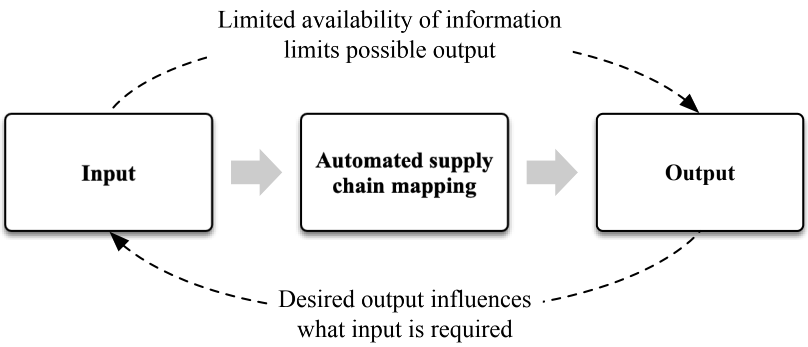


Fig. 3.3 Research scoping decisions are inter-dependent

To ensure general usefulness, the scope of this research shall be on the common elements of supply chain maps rather than specific information needs of a particular use case. This way the scope is set to be the minimal set of informational elements required for supply chain mapping in general rather than to consider all the particular informational requirements of a specific use case. By definition, the absolute minimum of a supply chain network are the extraction of supply chain participants and their inter-relations.

Decisions (A) to (F) shape the specific focus of the research reported here.

3.4 Research focus: Automatically extracting buyer-supplier relations

This section narrows down the research scope and defines a research focus. It also provides the rationale for the scoping decisions. The research focus shall be defined as follows:

Research focus:
Automatically extracting buyer-supplier relations from unstructured, natural language text

Please note that this focus is narrower (extraction of buyer-supplier relations) than the research scope (generation of supply chain maps). In the following, it shall be justified why the focus is on extracting relations and in particular on extracting binary buyer-supplier relations.

Rationale for focussing on extracting relations Whereas the overall research scope is the automatic generation of *supply chain maps* from unstructured, natural language text, the research focus of this study is on automatically extracting *individual buyer-supplier relations* from unstructured, natural language text, as depicted in Figure 3.4. This problem shall be considered first and in separation from the overall problem, before the broader problem is addressed conceptually in a subsequent chapter.

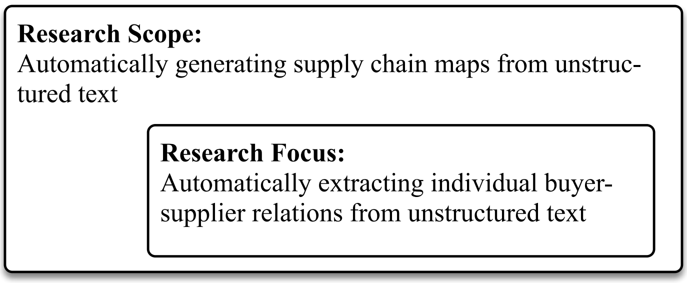


Fig. 3.4 Research focus: Automatically extracting buyer-supplier relations from unstructured, natural language text

There are a number of reasons that led to this scoping decision of focussing on extracting individual buyer-supplier relations (beyond the scoping decisions already made in Section 3.3):

- By definition, the ability to automatically extract individual buyer-supplier relations is a *prerequisite and fundamental building block* for automated supply chain mapping.

- Because any supply chain map will at least require companies and their inter-relations, extracting buyer-supplier relations is necessary *independent of the use case of the map*. In contrast, other steps in the mapping process, such as the visualisation, tend to be more specific to the use case.

- Extracting buyer-supplier relations is a confined problem where the desired performance can be expressed quantitatively and measured with a high degree of objectivity.

- Natural language processing is generally performed in “pipelines”, that is chains of tasks where one uses the output of the other as input. The implication is that errors will propagate and may result in additional errors further downstream. Hence, it makes sense to start with a shorter processing pipeline and optimise its performance before adding any further processing steps.

Rationale for focussing on explicitly expressed binary buyer-supplier relations A further scoping decision has to be made with respect to the *type* of relations that shall be extracted. One could aim to extract relations with more than the two elements. This would allow to capture more information. An example of such a (here: ternary) relation could be “[company] supplies [company] with ... for end-customer [company]”. To reduce complexity of the problem, the scope is reduced to binary relations (relations with two elements) and more complex relations have to be broken down into binary ones, as it is common practice in ontologies. Furthermore, one could aim to extract relations between elements that are not (only) organisations, such as the provided part, the end-product, the location where a part is produced, or the point in time when a buyer-supplier relation was reported to be active. Examples of such relations could be “[company] produces [product]” or “[product] is produced in [location]”. By collecting these relations as well, buyer-supplier relations could be inferred even though it is not explicitly stated. Examples are:

- (1) Company A is the only producer of material M in the world. (2) Company B uses material M. (1) and (2) would suggest that Company A supplies Company B at least indirectly.

- (1) In the aluminium market, Company A and Company B have a market share of 60% and 40%, respectively. (2) Company C uses aluminium. (1) and (2) would suggest there is a prior probability of 60% that a piece of aluminium used by Company A was (directly or indirectly) sourced from Company A.

For the sake of simplicity, the scope shall be limited to *explicitly expressed* relations between organisations only. Upon success, the scope can then be extended in future research.

3.5 Research questions

This section presents the research questions this thesis aims to answer. Research Question 1 corresponds to the research focus of automatically extracting individual buyer-supplier relations, whereas Research Question 2 addresses the overall problem of automating the end-to-end process of generating supply chain maps from text. Figure 3.5 illustrates the conceptual model of this research problem. This figure now provides a higher level of detail than the corresponding figure previously shown in Chapter 1, and shall be explained in the following.

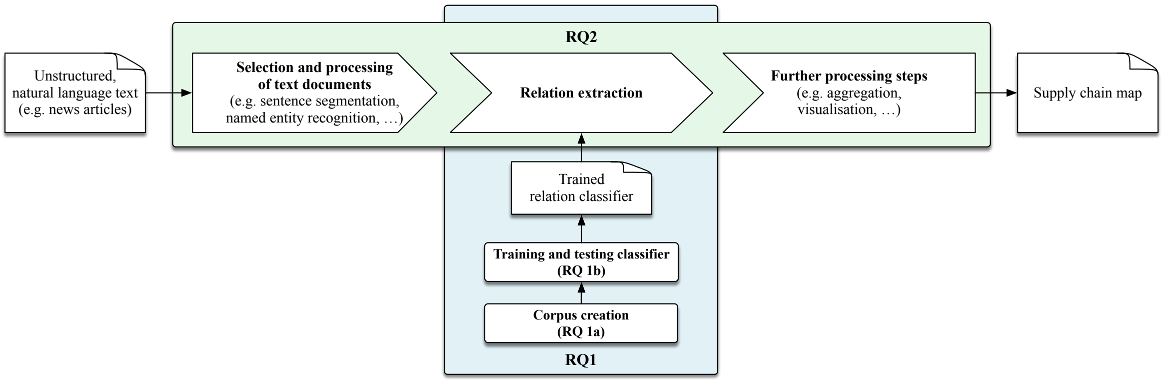


Fig. 3.5 Detailed conceptual model of the research problem

Following the structure presented above, the research questions are as follows:

Research question 1 (RQ1):
To what extent and how can the extraction of buyer-supplier relations from unstructured text be automated?

The research question relates to the need for a classifier that is able to extract individual buyer-supplier relations from text. The question is further broken down into the following sub-questions:

(1a) *What is the achievable inter-annotator agreement of humans attempting to classify buyer-supplier relations?*

(1b) *What is a suitable approach for the classification of buyer-supplier relations and what classification performance can be achieved?*

The sub-questions follow the logical sequence of how the overall question can be addressed. Firstly (1a), a reference dataset (“gold standard” or “ground truth”) needs to be established to test the performance of any automated classification attempt. This reference dataset contains both text as well as labels provided by human annotators. Each label indicates the type of relation expressed between two organisations in a sentence, such as the first organisation being a supplier to the second. By letting annotators provide labels redundantly, the degree of agreement among them can be measured. This *inter-annotator agreement* provides indications for the difficulty of the task and the quality of the obtained labels. Then (1b) a relation classification approach can be designed, trained on and tested against the reference dataset. Once a classifier has been trained, it can be used repeatedly without further human input.

It is important to clearly separate these two stages as they answer different questions and require different performance metrics. In stage (1a), one can measure *inter-* and *intra-*annotator agreement to quantify the degree of consensus and consistency in the labelling responses. A ground truth is not available yet, thus the need for the human labelling. In stage (1b), one can now measure the performance of a classifier against the training dataset generated in the previous stage which is now considered the ground truth. Having a ground truth makes it possible to use metrics such as recall and precision to capture aspects of completeness and correctness. Given how frequently new algorithms and network architectures are proposed in Machine Learning and NLP, the aim is *not* to identify the best possible algorithm but to test the general feasibility of the idea of automatically extracting buyer-supplier relations.

Research Question 2 then addresses the overall problem by widening the scope from extracting individual buyer-supplier relations to the end-to-end generation of useful supply chain maps.

Research question 2 (RQ2):

What are the challenges in developing a suitable end-to-end approach for automating the generation of supply chain maps from unstructured text?

This question is qualitative in nature and aims to identify and discuss the challenges that are likely to arise during the development of an end-to-end approach to automating supply chain mapping from text. These challenges may be a result of the characteristics of supply chains or the limited information quality of the input text document. Research Question 2 aims to bridge the gap between the extraction of individual buyer-supplier relations and the generation of a basic supply chain map from text. This gap is analysed in detail in Chapter 6. It should be noted that it is out of the scope of this thesis to produce a complete, automated supply chain mapping approach that solves all identified challenges.

One example to illustrate the gap is the need to disambiguate mentions of organisations and link them to a unique real-world entity. Another example is the need to learn how much confidence one wants to place on some extracted information given what else has been extracted or given what the source of that information is. A further challenge is the unknown *data availability* and *sparsity* with respect to buyer-supplier relations in a real-world news dataset. In this context, availability shall be measured in absolute terms, such as the absolute number of buyer-supplier relations that can be found for a company. Sparsity, on the other hand, shall be measured in relation to the size of the input data, such as a ratio of the number of buyer-supplier relations that can be extracted from a dataset and the overall size of that dataset. By applying a trained classifier to a large dataset of previously unseen data, one can attempt to answer questions about data sparsity. However, because the dataset is unlabelled and vast, recall can only be measured on a small sample.

3.6 Research approach

This section will present the way that the research questions of this study will be addressed. After reflecting on the underlying research paradigm, the research methodology is presented.

3.6.1 Research paradigm

Research paradigms or inquiry paradigms "define for inquirers what it is they are about, and what falls within and outside the limits of legitimate inquiry" (Guba and Lincoln, 1994). This research paradigm can be summarised by the answers to three fundamental questions (Guba and Lincoln, 1994):

1. The *ontological* question: "What is the form and nature of reality, and therefore, what is there that can be known about it?"

2. The *epistemological* question: "What is nature of the relationship between the knower or would-be knower and what can be known?"

3. The *methodological* question: "How can the inquirer (or would-be knower) go about finding out whatever he or she believes can be known?"

Table 3.3 provides an overview of alternative research paradigms.

Ontology The ontological spectrum ranges from so-called realism to relativism. Realism can briefly be described as the belief that facts are real and objective independent of the subjective consciousness of the human mind and can be apprehended via direct observations. Proponents of relativism on the other hand would proclaim that realities are apprehendable only "in the form of multiple, intangible mental constructions" (Guba and Lincoln, 1994) that are subjective and "dependent for their form and content on the individual persons or groups holding the constructions. Constructions are not more or less 'true', in any absolute sense, but simply more or less informed and/or sophisticated" (Guba and Lincoln, 1994).

Epistemology The epistemological spectrum corresponds to the ontological one in the sense that any answer to the question how knowledge can be acquired depends on the ontological position. Realists will assume that they can study an object without influencing it (or being influenced by it). The investigator and investigated object are assumed to be independent entities. Relativists on the other hand will take an interpretivist or subjectivist view: the investigator and the investigated object are "assumed to be interactively linked so that the 'findings' are literally created" rather than objectively discovered.

Methodology Again, since the methodological position builds upon ontology and epistemology, the methodological spectrum ranges from a position of empirical testing of hypotheses and quantitative methods (in the case of positivism / realism) to hermeneutical or dialectical methods (in the case of constructivism / relativism). The latter implies that the constructed knowledge can only be refined via the interaction between investigator and respondent. This last perspective is particularly common in social sciences.

Table 3.3 Overview of alternative research paradigms according to Guba and Lincoln (1994)

	Positivism	**Postpositivism**	**Critical Theory**	**Constructivism**
Ontology	naive realism – "real" reality but apprehendable	critical realism – "real" reality but only imperfectly and probabilistically apprehendable	historical realism – virtual reality shaped by social, political, cultural, economic, ethnic and gender values; crystallised over time	relativism – local and specific constructed realities
Epistemology	dualist / objectivist; findings true	modified dualist / objectivist; critical tradition/community; findings probably true	transactional / subjectivist; value-mediated findings	transactional / subjectivist; created findings
Methodology	experimental / manipulative; verification of hypotheses; chiefly quantitative methods	modified experimental / manipulative; critical multiplism; falsification of hypotheses; may include qualitative methods	dialogic / dialectical	hermeneutical / dialectical

Based on the above categorisation, this study will take a (post-)positivistic perspective. Generally, this also appears to be in line with the typical (post-)positivist paradigms seen in the relevant domains of research.

3.6.2 Research methodology

General choice of methods

Following the (post-)positivist research paradigm, the study will be based on quantitative methods. This, however, does not imply that qualitative methods cannot or should not also be applied. In a paper from 2005, Golicic et al. criticise the lack of qualitative research in the domain of supply chain management: "Logistics scholars agree that logistics and supply chain management are steeped in the positivist paradigm and that past research is primarily normative and quantitative." and they call for a "balanced approach" that also incorporates qualitative methods as "[r]esearchers who exclusively choose one approach or the other seriously delimit the scope of their inquiry and, thereby, their ability to contribute to the body of knowledge" (Golicic et al., 2005). Their balanced approach consists of using a quantitative deductive approach as well as using a qualitative inductive approach. The *quantitative deductive* approach starts with reviewing the "appropriate literature in order to develop a conceptual framework that specifies relevant variables and expected relationships among them" (Golicic et al., 2005), followed by building a formal theory to explain the phenomenon and capable of generating predictive statements that can be tested. Before any data is collected, hypotheses are formulated that have been derived from the theory via deductive reasoning. The last step consists of collecting data and verifying the developed theory by testing the hypotheses which express the proposed relationships between variables. The *qualitative inductive* approach, on the other hand, starts with the data collection to understand the phenomenon in depth (e.g. via field visits). The second step, according to Golicic et al., is to "to describe the phenomenon from the point of view of the informants"; information is obtained using techniques such as open-ended questions. The next step is to build a substantive theory from the descriptive data, which typically takes the form of process models.

Especially during the exploratory first phase, this study makes use of semi-structured interviews with industry partners. These interviews were important for context and background: to provide a more in-depth understanding of the problem, to ensure relevance of the problem, to reveal requirements for modelling and method development and to ensure the validity of the findings as well as applicability of proposed methods. That said, this study is of *quantitative* rather than qualitative nature and the interviews are not a formal part of the methodology.

Research phases and detailed map of research methodology

Research phases A high-level overview of the research phases is depicted in Figure 3.6. The research problem arose from a collaboration with an industrial research partner and was further refined in semi-structured interviews with various stakeholders within the firm.

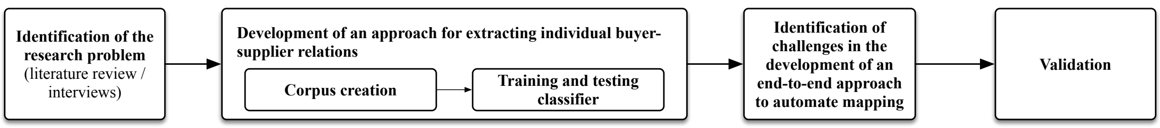


Fig. 3.6 Research phases

Detailed map of the research methodology A more detailed plan of the research methodology, including the sub-elements of each research phase, is given in Figure 3.7.

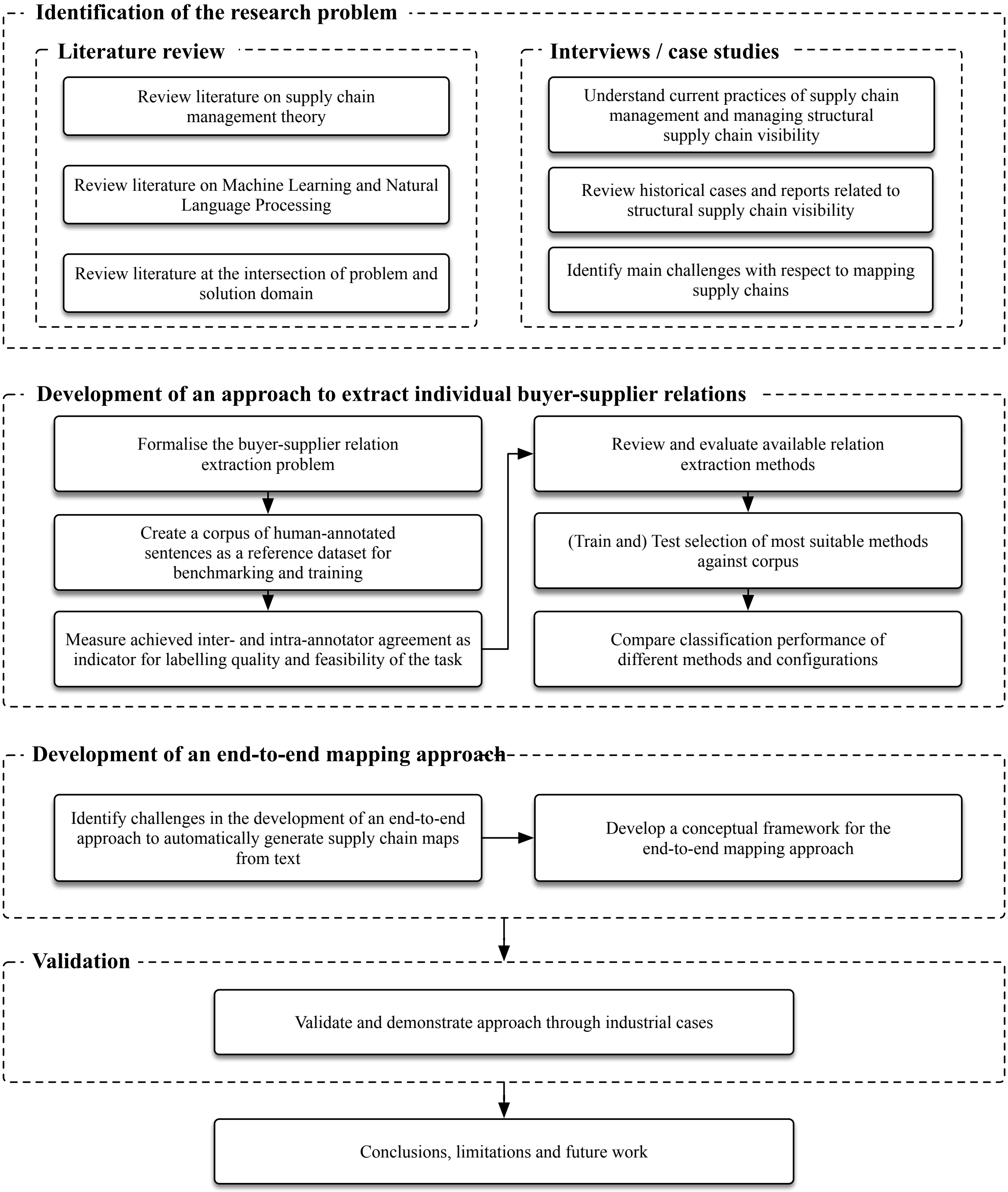


Fig. 3.7 Research methodology

To identify and refine the research problem, input from both academic literature and industrial practice was considered. The literature review covered the problem domain of supply chain management and the importance of structural supply chain visibility. This also included academic case studies related to structural supply chain visibility. By definition, the automatic extraction of information from text falls into the domain of Natural Language Processing which was also covered by the literature review. Since the field of Natural Language Processing increasingly consists of applying Machine Learning methods to text, this topic was also reviewed in depth. Lastly, prior academic work at the intersection of problem and solution domain was identified.

The industrial perspective was considered in two ways: Semi-structured interviews were conducted with a range of supply chain managers to understand current practices of the management of supply chains and their structural visibility. Furthermore, in order to consider a wider range of companies, additional industrial case studies from the literature were included. One important purpose of the case studies was to ensure that the identified problem is a relevant problem that is worth solving.

The next phase focusses on the prerequisite for any comprehensive mapping approach, that is the extraction of individual buyer-supplier relations from text. First, the problem is formally defined as a classification problem. A corpus of human-annotated sentences is then created as a benchmark but also to serve as a training dataset for any learning algorithm. Relevant extraction methods are then reviewed and the most suitable ones are trained and tested on the corpus.

Based on the findings from the extraction of individual buyer-supplier relations, a gap analysis can be performed to identify further challenges that need to be addressed for a complete automation of the supply chain mapping process.

The approach is validated and demonstrated by applying a prototypical implementation to a real-world dataset. By applying the proposed approach to a large unlabelled dataset first indications regarding the availability and sparsity of information about buyer-supplier relations can be obtained.

3.7 Summary

This chapter has presented the identified research gap as well as the research approach. The overall research scope is the automatic generation of supply chain maps from unstructured text with a specific focus on the extraction of individual buyer-supplier relations.

The next chapter provides an industrial perspective on the research problem and aims to present the rationale for improving structural supply chain visibility.

Chapter 4

Rationale for improving structural supply chain visibility

4.1 Introduction

In the previous chapters, the research background was presented, a research gap was identified and an approach was suggested to address this gap. The previous chapters focussed on the academic novelty and suggested that a comprehensive study of automating the generation of supply chain maps from text has not been conducted yet.

In contrast, the purpose of this chapter is to provide an *industrial perspective* on the problem of limited structural supply chain visibility. It thereby focusses on the industrial relevance of this research and aims to provide a rationale for why efforts to improve structural supply chain visibility can be worthwhile. This chapter also provides an overview of potential beneficiaries and use cases. There are a number of assumptions that need to hold for this research to become industrially relevant, shown in Figure 4.1. The first assumption is that companies in practice actually experience limited structural supply chain visibility. If that is the case, the next assumption is that this limited visibility actually represents a problem for those companies and increased visibility would improve a management decision. To ensure relevance, a further assumption is that there are no effective alternative solutions already available that could fully address the problem.

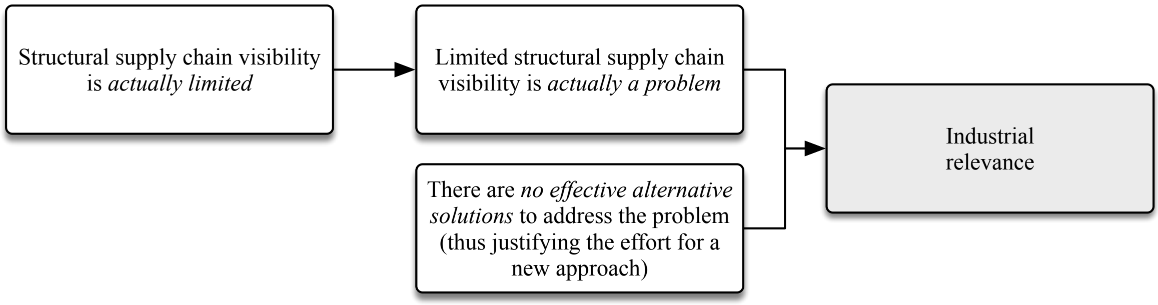


Fig. 4.1 Assumptions for relevance

Overall, the aims of this chapter are:

1. To provide an insight into how limited visibility of the supply chain structure impacts organisations as captured by case studies in the literature as well as conducted within the scope of this research

2. To summarise the challenges that companies face with regards to obtaining the data required for mapping their supply chain structure

3. To provide a structured overview of potential beneficiaries and use cases

4. To demonstrate the industrial rationale behind this research and the practical relevance of a solution

To achieve these aims, exploratory case studies are conducted using the literature as well as interviews conducted within the scope of this research. Finally, some of the existing tools are reviewed. The conclusion of the industrial background can be found at the beginning of Chapter 4 where the research gap is derived in combination with the key issues from this chapter's review of the industrial practice.

4.2 Exploratory case studies and interviews

This section provides a selection of case studies for the purpose of exploring the topic of structural supply chain visibility from an industrial point of view.

4.2.1 Objectives

The purpose of these exploratory case studies is to better understand the research problem from an industrial perspective and ensure its practical relevance – as opposed to validating any research findings. The following questions were phrased to guide the exploratory process. They are structured according to the previously stated assumptions and aim to help identify (counter-)evidence.

- What is the evidence that structural supply chain visibility is actually (not) limited in practice?
- What is the evidence that limited structural supply chain visibility is actually (not) a problem in practice?
- What is the evidence that existing tools and approaches are (in)sufficient for addressing the problem of limited structural supply chain visibility?

4.2.2 Methodology

To better understand the research problem, a series of semi-structured interviews were conducted with practitioners, such as supply chain managers at manufacturing firms. In order to add further perspectives, additional case studies were drawn from the academic literature as well as business news where limited visibility of the supply chain structure had an impact on a company's ability to mitigate risk or appropriately respond to a risk event. Inclusion criteria for those were the extent to which the cases have been researched or documented, the potential or realised impact of the risk as well as the degree to which structural supply chain visibility reportedly was a key aspect. The overall methodology for the exploratory case studies is illustrated by Figure 4.2.

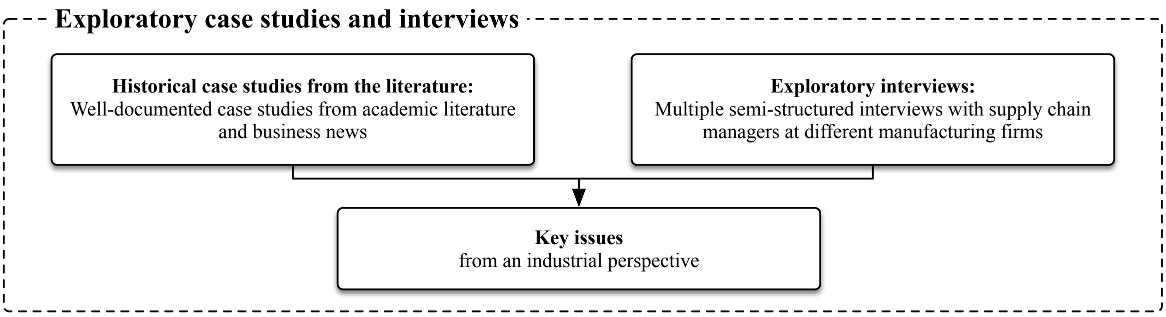


Fig. 4.2 Exploratory case studies and interviews

4.2.3 Historical case studies from the literature

In the following, a selection of exploratory, historical case studies from the literature is presented:

- 2011 Tohoku earthquake and tsunami in Japan
- 2011 Thailand floods
- 2012 PA12 shortage due to factory explosion

2011 Tohoku earthquake and tsunami in Japan

In March 2011, an earthquake and a subsequent tsunami hit Japan resulting in thousands of deaths and major destruction. The disaster also hit Japanese car manufacturers hard – not because their own factories had been destroyed but because supply with critical parts had been disrupted. The production at all of Toyota's 12 assembly plants in Japan had to be temporarily suspended, "domestic production at Toyota, the world's largest automaker, plummeted 63 percent in March" (Topham, 2011) and Toyota's quarterly profits dropped by 99% (BBC, 2011). In the aftermath, The Economist reported on 19.05.2011: "The disaster barely dented factories belonging to the big Japanese carmakers. [...] But it ravaged the suppliers of critical parts and raw materials in north-eastern Japan. Toyota faces a shortage of 30 components. That is much better than the 500 it lacked shortly after the quake, but it only takes one missing part to bring an assembly line stuttering to a halt. [...] The quake cost Toyota ¥100 billion ($1.2 billion), the most of any Japanese carmaker so far." (The Economist, 2011).

What makes this case study interesting in the context of this thesis are the findings in the aftermath of the disaster. About a year after the disaster, Automotive News, considered the newspaper of record for the automotive industry, provided the following analysis: "Automakers thought they had hedged risk by diversifying Tier 1 and Tier 2 suppliers. They had assumed the supply base looked like a tree's roots, spreading out further down the line. But the supply chain was actually more diamond shaped, with rival suppliers turning to the same subsuppliers for parts. These Tier 3 and Tier 4 suppliers were all but ignored by the carmakers themselves." (Greimel, 2011). Figure 4.3 shows a symbolic illustration of Toyota's supply chain structure based on the verbal descriptions by Shinichi Sasaki, Toyota's global procurement chief, as reported in the March 2012 article by Automotive News (Greimel, 2011). Before the disaster, Toyota had falsely believed they had a good understanding of their supply chain structure: "Since the quake, Toyota Motor Corp. has cataloged nearly 1,000 at-risk parts that could be trouble in the next disaster, says Shinichi Sasaki, Toyota's global procurement chief. It has identified another 300 supplier factories in precarious locations – straddling fault lines or lying in the path of a potential tsunami. 'It was quite delusional of us to have thought before the quake we had a good grasp of the supply chain,' Sasaki said." (Greimel, 2011). Even after the disaster, 50% of the suppliers refused to disclose their own

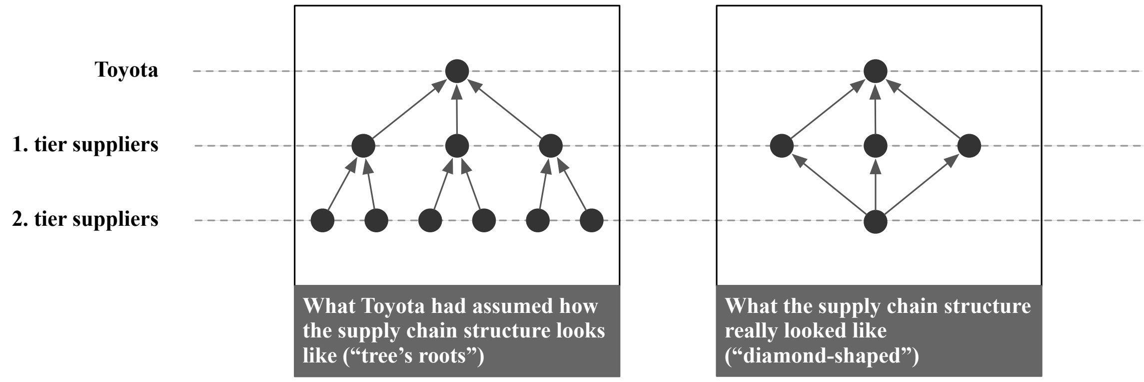


Fig. 4.3 Symbolic illustration of Toyota's supply chain structure: assumed vs. real

suppliers: "Toyota has more than 500 Tier 1 suppliers in Japan, including global companies operating here. Since the quake, about half have cooperated by detailing their supply chains to Toyota. But the other half refused to share, saying it was proprietary information, Sasaki said. 'Half of that is black-boxed to us,' he said." (Greimel, 2011). Risk mitigation options are generally available – some at a cost. Given a better understanding of the supply chain structure, these options could have mitigated at least parts of the risks: "By the end of March, Toyota plans to decide how to shore up the weak links it has pinpointed. The company wants to fix them by autumn, Sasaki said. Steps will include asking suppliers to make the same part at multiple locations or getting the same part from different suppliers. Toyota may even carry larger inventories of sensitive parts." (Greimel, 2011)

At least one further vulnerability in the automotive supply chain was revealed by the earthquake: Xirallic, a pigment that is used in automobile paints and provides a glittering shine, had *only* been produced in a *single plant* that happened to be based in Japan (cf. Seetharaman (2011) and Boudette and Bennett (2011)). The production at this plant, operated by German chemical company Merck KGaA, had to be suspended and the resulting shortage of the pigment hindered the production of cars of certain colours at various automakers worldwide.

2011 Thailand floods

Overview The Chao Phraya River basin in Thailand includes densely populated areas, along with large areas of manufacturing industry in and around Bangkok. In summer 2011, a severe flooding of unprecedented nature occurred in this area, resulting in more than 800 deaths. The flood was not the largest in terms of submerged area, but it lasted 158 days from July to December and, thus, longer than any other flood event in that region (Promchote et al., 2016). As damage and economic losses of USD 45.7 bn (approx. 13% of Thailand's GDP) (IAJ, 2013) were unprecedented (Haraguchi and Lall, 2014), this flood has been considered the worst in the last 50 years (Promchote et al., 2016). Besides the amount of rainfall, the size of the flooded area, and the duration of the floods, the massive economic losses were also due to the geographical concentration of major (and now flooded) factories and the subsequent disruption of world wide supply chains (IAJ, 2013). Companies, such as Nikon, Canon, Sony, Toyota, Honda, Nissan, Isuzu and Western Digital, all had large factories in the flood-affected areas north of the capital Bangkok (Carey, 2011).

Disruption of HDD production Besides being a major automobile manufacturer, Thailand also was the world's largest manufacturer of hard disk drives (HDD), devices used to store data on computers. In 2010, Thailand's share in the global HDD production had been 43% (Haraguchi and Lall, 2014). During the floods, Western Digital Corporation, which produced one third of the world's hard disks, lost 45% of its shipments because one of their factories was inundated (Haraguchi and Lall, 2014). And not only were factories of the HDD manufacturers, such as Western Digital, closed but also some of their suppliers were affected. Japan's Nidec Corp had to close some plants in Thailand, as had disk parts maker Minebea. Nidec controlled about 80% of the world's output of a key HDD component – the motor – and now became a major bottleneck for the HDD production (Arthur, 2011). Due to the floods, prices of HDDs on the global market increased by 80–190% for desktop HDD and by 80–150% for mobile HDD (Haraguchi and Lall, 2014). The New York Times quotes John Monroe, an expert on storage devices at the technology research firm: "Surely one of the inevitable impacts of this is that never again will so much be concentrated in so few places" (Fuller, 2011). The newspaper also quoted Hajime Yamamoto, the Thailand director of IHS Automotive, who also predicted that car manufacturers and their suppliers would seek to diversify their operations to other countries: "They will try to balance their expansion so they don't have concentration of risk in Thailand" (Fuller, 2011). The flooding in 2011 revealed the extent to which the global economy had relied on components made in this region of Thailand, and also the extent to which various manufacturers and component producers had been geographically concentrated in industrial clusters in flood-prone areas. Thailand had incentivised high-tech manufacturers to settle in industrial parks, and, hence, to settle in clusters of multiple facilities in geographical proximity to each other. And few facilities produced large shares of the overall Thai or even global output. E.g. a single facility in Bang Pa-In owned by Western Digital had produced one quarter of the global supply of "sliders", an integral part of HDD, before it got flooded (Fuller, 2011). The recovery from the floods would also take longer than in other industries since HDD manufacturing requires "clean room" environments. A further interesting aspect was revealed by the flooding, illustrated conceptually by Figure 4.4: Intel had to warn investors that its fourth-quarter revenue in 2011 would fall about $ 1 billion below its own targets (Tibken, 2011). The reason for this was *not* that Intel itself or its own supply chain was majorly affected by the floods in Thailand. But because of the shortage of HDD, PC makers had to slow down production, and, hence, demanded fewer processor chips[1].

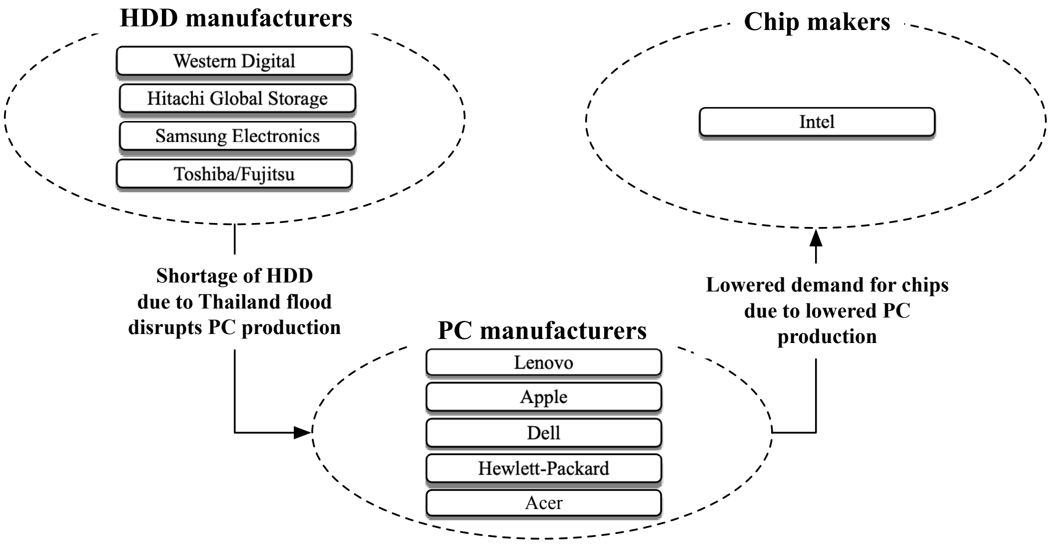


Fig. 4.4 Conceptual illustration how Intel got affected by 2011 Thailand floods

2012 PA12 shortage due to factory explosion

On March 31, 2012, a fire and explosion occurred at Evonik's production site "Marl Chemical Park" in Marl, Germany, and killed two workers. The plant produced CDT (cyclododecatriene) which serves as a starting material in plastics manufacturing. CDT is used to make laurolactam which, in turn, is used as a monomer in polyamide 12 (PA12). Plastic parts made of PA12 are "key components in automobile production, in the photovoltaic industry, and in offshore pipelines" (Evonik, 2012). This tragic event led to to a severe supply shortage. Evonik was the world's largest PA12 supplier and the plant shutdown took out "around 40% of the 100,000 tpa of global capacity" (Platt, 2012). PA12 is the "sole-specified material for automotive fuel lines due to its unique combination of thermal, physical, chemical and mechanical properties" (Platt, 2012). One of the automotive suppliers, TI Automotive, an automotive fluid systems manufacturer, warned its customers, global automotive manufacturers, about an impending shortage with the potential to stop car production lines in factories

[1]Cf. the aspect of "your customer's other suppliers" mentioned later in Section 4.3.2.

around the world. Other suppliers of PA12 were unlikely to cover the shortfall as they were already at full capacity and fully committed to existing customers (Platt, 2012), and, thus, automakers and suppliers held a summit meeting in Detroit to discuss PA12 supply and alternative materials. In this case, Evonik had acted largely as a sub-tier supplier to the global carmakers. And nearly half of the global production of a material required for the production of cars had been produced in a single facility.

4.2.4 Exploratory interviews

A series of exploratory interviews were conducted to better understand the problem of limited structural supply chain visibility. Table 4.1 provides an overview of those interviews whose results shall be summarised in the following. More detailed interview summaries can be found in the Appendix A.1.

Table 4.1 Exploratory interviews conducted as part of this research

	Interviewee
Interview series with aerospace company	Series of interviews with supply chain managers of a major aerospace manufacturer in March/April 2015
Interview with supply chain manager / consultant	Interview in June 2018 with former supply chain manager of an electronics manufacturer and now consultant

Interview series with supply chain managers of an aerospace manufacturer

Profile The findings summarised below stem from a series of 7 semi-structured on-site interviews with supply chain managers of a major aerospace manufacturer in March/April 2015. The supply chain managers worked in different departments of the manufacturer and covered relevant functions, such as business continuity or supplier management.

Summary of selected findings

- The company does not possess a map of the extended supply chain and there is no clear responsibility for systematically increasing structural supply chain visibility. The lack of supply chain maps prevents the company from carrying out further tasks, such as predicting supply disruptions. The true risk is often obscured: Direct suppliers may share the identical sub-tier supplier that is the true risk factor. It is known to supplier managers that the company's suppliers tend to keep some of their own suppliers to themselves even though they are contractually obliged to disclose them. Main reasons include the lack of positive incentives (monetary or otherwise) for revealing all suppliers and the company's inability to check the supplier-provided information for completeness.

- There is a need to not only see the company's own supply chain but the interdependencies of the complete aerospace industry. A supplier may be struggling because another one of its customers increased the load.
- The sourcing in the aerospace manufacturer is generally carried out as a programme-specific[2] function and not as a cross-programme function. Only basic materials are sourced enterprise-wide.
- Logistics companies can be the cause of supply disruptions and should also be considered in supply chain maps.
- Knowing the extended supply chain could help with the compliance with more recent environmental regulations, e.g *"REACH"* (short for: *Registration, Evaluation, Authorisation and Restriction of Chemicals*).
- Especially smaller and less experienced suppliers are critical from a risk management perspective. Significant effort is spent on mentoring and capability building with these.
- A number of monitoring services are in place to detect anomalies or warn suppliers about impending dangers, such as hurricanes. However, these can only be applied to known suppliers. Sub-tier suppliers could easily be added to these services if their identities were known.

Interview with supply chain consultant

Profile The interviewee is an accomplished manufacturing professional and supply chain consultant, with more than 15 years' experience in electronics manufacturing companies where she held a variety of senior management positions in supply chain, manufacturing engineering and operations management. The interviewee answered the questions by drawing on her experience as a consultant with the aerospace industry (a) and as a previous supply chain management professional at an electronics company (b) herself.

[2] Programme in this context is to be understood as a particular aircraft model, such as the Boeing 737 or the Airbus A380.

(a) Summary of selected findings with respect to the aerospace and defence industry in general

- In the aerospace domain, supply chains for a particular programme do not change much. Even over time spans of 10 years, the companies will largely remain the same. However, once a new programme is started, the supply chain may look completely different.

- Some aerospace companies can be oblivious that second-tier suppliers in turn depend on very small and often highly specialised companies. Some of these second-tier companies are single-sourced and struggling. The second-tier suppliers do not only depend on components manufacturers. Because they are relatively small, these suppliers do not perform all the processing steps in-house but send out to other specialised companies. The dependencies on *processing* specialists are often neglected in risk assessments compared to the dependencies on parts or material suppliers.

- In the aerospace domain, environmental EU regulations, such as *"RoSH"* (short for: *Restriction of Hazardous Substances*) and *"REACH"* (short for: *Registration, Evaluation, Authorisation and Restriction of Chemicals*) restrict the type of chemicals and processes that a company is allowed to use. By better understanding supply chains, such as discovering that dubious or known-to-be-incompliant sub-tier suppliers participate in the supply chain, it may become easier to understand where your company is not compliant with regulations. A supply chain map could at least tell how complicated it would be to make sure that every participant is compliant.

- Extended supply chain maps could also help to understand where a company's products are being used further downstream.

(b) Summary of selected findings with respect to the supply chain management at an electronics company

- Supply risk management is not only relevant once a product is being manufactured. When a new product is designed, knowledge about supply risks can be used to inform the product design process, such as avoiding bespoke parts or avoiding process steps that only few suppliers can carry out. Supply chain maps could therefore not only help with risks in the current supply chain but help with planning future ones.

- The electronics company prioritised the supplier risk management based on the financial value a supplier corresponded to. This, however, was "somewhat of a red herring" given that sales volume does not always correspond to criticality.

- To ensure a healthy supplier base, the electronics company would monitor its own share in the revenue of their suppliers. The rule was to not exceed 20% of any supplier's portfolio to ensure the suppliers remained robust.
- The structural visibility of the electronics company was limited to the distributors (first tier) and the manufacturers (second tier). The manufacturers were known because they directly collaborated with them, and just supplied via a distributor. Beyond the manufacturer, there was no visibility. The manufacturer was expected to know and to inform them about problems.

4.3 Potential beneficiaries of supply chain maps and possible use cases

Structural supply chain visibility may be important for one or more different beneficiaries in a variety of use cases. This section aims to hypothesise on potential beneficiaries and use cases as well as to provide a tentative structured overview.

The following limitations of the overview should be kept in mind: (a) The provided overview is not meant to be exhaustive. The aim of this overview is to demonstrate the potential value of discovering and monitoring supply chain structures as well as to demonstrate the variety of beneficiaries and use cases. (b) Some of the use cases are overlapping. E.g. there may be legal, supply risk and altruistic reasons why a company aims to avoid forced labour in its supply chain. (c) The overview shall also not suggest that supply chain mapping or even any automated supply chain mapping solution will *fully* solve the listed problems but that it *may* be useful in addressing them. (d) It is far beyond the scope of this section to prove or even quantify the value of supply chain maps for the different use cases. This exercise should merely be understood as the author hypothesising on potential beneficiaries and use cases.

The following section focusses on main potential beneficiaries and use cases. Further potential beneficiaries and use cases are listed towards the end of this section.

4.3.1 Main beneficiaries and use cases

An overview of main potential beneficiaries and use cases is provided by Table 4.2. Use cases can distinguished along different dimensions, e.g.:

- the motivation that drives the information need (e.g. managing supply risk, supply chain efficiency or sustainability)

- the actor or beneficiary (e.g. a manufacturing company, an insurance company or governmental agency)
- the scope of the required supply chain map (e.g. a company's own extended supply chain or a supply chain map spanning whole industries or economies)

Table 4.2 Importance of structural supply chain visibility: Hypothesising on potential beneficiaries and use cases

Motivation	Beneficiary (scope of the map)	Use case
Supply risk management (supply disruptions)	OEM (own supply chain)	Early & effective mitigation: Understanding risk exposure & detecting vulnerabilities in the extended supply chain, e.g. a single shared supplier on a sub-tier or geographical concentration of suppliers.
	OEM (own supply chain)	Fast & effective contingency: Draw links between world events and resulting potential impacts on supply and then react swiftly and appropriately, e.g. by ordering parts from an alternative supplier.
	Governmental agencies and policymakers (any scope)	Understand risk exposure of companies, industries or a country as a whole, e.g. in case of embargoes, tariffs or other trade barriers like "Brexit"
	(Re-)Insurance companies (any scope)	Understand risk exposure of companies, industries or across industries; use of supply chain and risk mapping as a tool in the sales process for supply chain risk insurances
Legal or regulatory reasons	OEM (own supply chain)	Ensuring compliance with, e.g. general embargoes or regulations that prohibit the use of materials sourced from specific countries for specific purposes (e.g. EU regulations like *RoSH* and *REACH*) as well as so-called *KYC* regulations for financial institutions
Responsible sourcing & Corporate Social Responsibility (CSR)	OEM (own supply chain)	Ensuring that only sustainable and ethically responsible companies participate in extended supply chain (e.g. modern slavery or conflict minerals)

In the following, the main beneficiaries and use cases of Table 4.2 shall be discussed in more detail.

Supply risk management

As defined in Section 2.2.2, supply risks are those that may delay or disrupt the *supply* of goods and services in the desired quantity and quality. Supply chain maps can help with managing these risks. In the following, different supply risk phases shall be distinguished. Each phase is then discussed with respect to the impact of knowing the extended supply chain structure on the ability to manage supply risks.

Shared mental model A dimension along which supply risk use cases can be distinguished are the general *tactics* to reduce risk: mitigation and contingency (see Section 2.2.3). In the case of supply chain risk, it may even be useful to subdivide contingency and distinguish *three* phases overall: (1) before a risk event has occurred (mitigation), (2) after a specific risk event has occurred and while its impacts are propagating through the network (contingency in a broader sense), and (3) once the impacts of that risk event have reached the focal company (contingency in a narrow sense), as illustrated in Figure 4.5. Due to the latency in the supply network, companies may have sufficient information and time to react to a specific risk event before its impacts reach them. Closely related to the tactics is the trigger for a supply chain

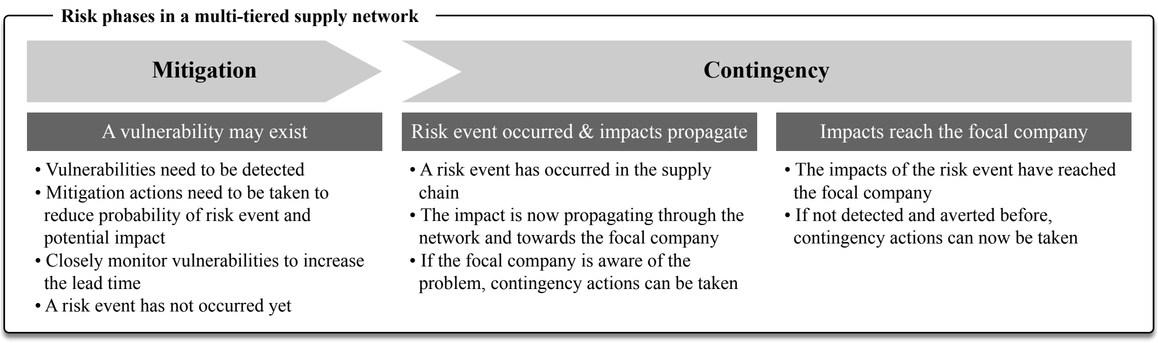


Fig. 4.5 Risk phases in a multi-tiered supply network

mapping need, which may be a singular event or the desire to continuously monitor the supply network.

Focal company managing its own supply chain (mitigation phase) In the *mitigation phase*, a risk event has not occurred yet but supply chain vulnerabilities may exist. The motivation in this phase is to assess the risk exposure in the current or planned extended supply chain, and to identify and choose appropriate mitigation actions. Assessing the risk exposure includes both the detection of specific vulnerabilities (“pinch points”; see Christopher and Peck (2004)) as well as the creation of an overall risk profile. For example, a direct supplier like General Electric (GE) may account for 20% of a company's material. But, in fact, GE may also act as a supplier to other suppliers. And, thus, the real risk exposure actually corresponds to 50% of the company's material. Table 4.3 attempts to summarise example situations where knowledge of the supply chain structure is required for or can at least help with identifying vulnerabilities.

Table 4.3 Mitigation situations

	Vulnerability of an existing or planned future supply chain
Various forms of supplier concentration	Dependencies on a single shared supplier on a sub-tier level
	Geographical concentration of sub-tier suppliers
	Undesirable levels of suppliers depending on specific countries or other political unions (relevant in cases of tariffs etc.)
	Limited number of available alternative suppliers or trends towards such a market consolidation
Various forms of undesirable supply chain participants	Dependencies on previously unknown, undesirable sub-tier suppliers (legal risks, quality risks, availability risks, blacklisted by authorities, unsustainable, unethical etc.)
	Dependencies on individual sub-tier suppliers located in high-risk areas (e.g. prone to natural disasters)
Various forms of overlapping, competing or otherwise interfering supply chains	Seemingly unrelated industries competing for the same resources as the OEM and potentially negatively impacting supply
	Dependencies on the health of a supply chain that produces complementary parts (see the case study in Section 4.2.3 where demand for Intel's chips was impacted by an HDD shortage)
	Competitors and otherwise risky (e.g. financially unstable) businesses using (and potentially negatively impacting) identical supply chain participants
Dependencies on problematic parts or materials	E.g. dependencies on parts or materials on a sub-tier level with uncertain supply (e.g. rare earth elements that are mostly mined in China)

Additional benefits may arise indirectly: Discovering and continuously monitoring their extended supply chain structures allows companies to apply their existing supply chain risk management tools to more than just direct suppliers. For instance, tools that provide information about the financial health of a company can then also be applied to sub-tier suppliers. Lastly, it may be possible to reduce companies' risk premium on supply chain insurances by demonstrating existing and successful efforts to achieve supply chain visibility.

The discovery and monitoring process may be triggered by a singular need but is more likely to be part of a continuous risk monitoring process. This includes detecting mere changes in the supply network topology, such as a supplier switching the sub-tier supplier, which could be indicative of changes in the risk profile – not only because the supplier changed but because supply chain further upstream this supplier is now likely to be different as well.

Focal company managing its own supply chain (contingency phase) Even once a specific risk event has occurred somewhere in the supply chain and its impacts have started to propagate through the network and towards the focal company, knowledge of the supply chain structure may still be beneficial, as Table 4.4 attempts to summarise.

Table 4.4 Potential benefits in contingency situations

	Potential benefits in contingency situations
After risk event has occurred but before the impacts have reached the focal company	Realise the relevance of a world event to a company's own operations. For example, knowing that an earthquake happened in Japan is not actionable until it is known if and which sub-tier suppliers or parts are affected. Provided with knowledge of the extended supply chain, a company can react swiftly and appropriately (e.g. by ordering parts from an alternative supplier faster than the similarly affected competition).
	Predict how the impacts will propagate through the network, i.e. if the impact of the risk event will be amplified or may get absorbed and how long it may take to reach the focal company.
Even after impacts have already reached focal company	Help identify contingency options and choosing the most appropriate one. E.g. knowledge of the supply chain structure may help a company to identify alternative suppliers that are not affected by the risk event.
	Help with identifying and addressing the root cause. Knowledge of the supply chain structure may help identify the root causes, for instance of a parts shortage, and the location of the problem within the multi-tiered supply chain.

Organisations trying to understand other companies' supply risks Apart from companies trying to manage their own supply risks, there are various organisations with an interest in understanding other companies' supply risks. *Governmental agencies and policymakers* may have an interest to understand and reduce the risk exposure of companies, industries or a country as a whole (e.g. in case of embargoes, tariffs or other trade barriers like "Brexit"). In this case, the scope of the supply chain map could go beyond a single company's supply network. *B2B insurance companies and reinsurance companies* have a similar interest. In addition, the ability to reveal previously unknown vulnerabilities could help convince customers to buy a supply chain insurance. Another benefit may be a better understanding of supply chain risk by the insurer: If multiple supply chain insurance customers depend on the same (sub-tier) supplier, the risk for the insurance company may be much higher than expected in a portfolio of random (and thus less strongly correlated) insurance customers. *Strategy consultancies* may benefit from knowledge of the supply chain structure in the due diligence phases of M&A[3] projects or to quickly gain an overview of how an industry works. *Private equity firms and other investment firms* may be interested in supply chain maps to understand which types of supply disruptions may harm or benefit which companies. Short-term investors may be interested how supply chain are temporarily disrupted and which companies may be positively or negatively affected.

Legal or regulatory reasons

Organisations may have an interest in discovering the extended supply chain of their own or other companies to ensure or verify compliance with legal or other regulatory requirements. For instance, there are regulations that prohibit the use of materials sourced from specific countries, e.g. the use of titanium sourced from Russia in US military aircraft, embargoes prohibiting the trade with specific countries or rules regarding food safety. Further examples are environmental EU regulations, such as *"RoSH"*[4] and *"REACH"*[5] that restrict the type of chemicals and processes that a company is allowed to use. If it was known that specific companies are not compliant, then knowledge of the extended supply chain would reveal if these regulations may have been violated. Or vice versa: knowing supply chain participants on a sub-tier would enable a company to check for their compliance with regulations.

Financial institutions could have an interest in understanding a potential credit receiver's exposure to risks. They could also potentially benefit from the perspective of compliance with *"Know Your Customer"* (KYC) or "Customer Identification Program" (CIP)[6] regulations which require banking institutions, credit companies, and insurance agencies to verify the identify of their customers and to ensure that customers are not involved in illegal activities. An extensions of this process is known as *"Know Your Customer's Customer"* (KYCC).

[3] M&A stands for mergers and acquisitions

[4] "RoSH" is short for the "directive on the restriction of the use of certain hazardous substances in electrical and electronic equipment"; `https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32011L0065`; last accessed: 2019-08-01

[5] "REACH" is short for: Registration, Evaluation, Authorisation and Restriction of Chemicals; `https://echa.europa.eu/regulations/reach/understanding-reach`; last accessed: 2019-08-01

[6] E.g. United States CIP regulation in 31 CFR § 1020.220: `https://www.law.cornell.edu/cfr/text/31/1020.220`; last accessed: 2019-08-01

Responsible sourcing & Corporate Social Responsibility (CSR)

A variety of use cases can be subsumed under *"sustainable sourcing"*. As suggested by van den Brink et al. (2019), "responsible sourcing" in this context shall be understood as an umbrella term "encompassing all sourcing designed to be 'socially responsible', 'green' or 'sustainable"'. This shall also include all facets of what has been coined *Corporate Social Responsibility* (CSR), such as avoiding inhumane production conditions or other human rights violations on a sub-tier level (human rights due diligence). These use cases for structural supply chain visibility are often not clearly separable: Besides altruistic motives, companies managing their supply chains can have an interest in ethical sourcing to avoid legal or reputation risks. Furthermore, in cases of unethical sourcing, existing inhumane working conditions often coincide with actions that also harm the environment. The general management approaches to ensure responsible sourcing include a "supply chain due diligence" (van den Brink et al., 2019), the adoption of "sustainability schemes" (van den Brink et al., 2019), such as the Initiative for Responsible Mining Assurance (IRMA), or establishing a "chain of custody" (van den Brink et al., 2019). Of particular importance in the context of responsible sourcing are the topics of *modern forms of slavery*, *conflict minerals*, and *greenhouse gas emissions*. A discussion of why these topics matter and how structural supply chain visibility may be able to help is provided in the Appendix A.2.

4.3.2 Further beneficiaries and use cases

Further beneficiaries and use cases are summarised in Table 4.5.

Table 4.5 Importance of structural supply chain visibility: Hypothesising on further potential beneficiaries and use cases

Motivation	Potential beneficiary (scope of the map)	Use case
Food safety and other contamination scenarios	OEM & Governmental agencies (any scope)	Preventing and tracing back the spread of diseases or other types of "contamination", such as the spread of dioxin in chicken feed or low-quality steel
Supply chain efficiency & network design	OEM (own supply chain)	Adjusting the supply chain architecture to make it more efficient (e.g. faster, less costly, or involving fewer participants)
Supplier discovery	OEM (industry supply chain)	Identifying alternative suppliers
Detecting supplier consolidation	OEM (own and industry supply chain)	Detecting a consolidation trend in the extended supply chain – potentially indicating a shift of negotiation power away from the OEM
Competitive intelligence	OEM (industry supply chain)	Monitoring of competitors, such as benchmarking supply chain architecture against competitors or other companies
Supply chain research	Academics (any scope)	Enabling academics to work with larger datasets for research on supply chains

A more detailed explanation of these can be found in the Appendix A.2. Farris (2010) provides a framework for use cases of structural supply chain visibility from the perspective of a focal company, as discussed and shown before in Figure 2.8 in Section 2.3.4. Especially use cases related to knowledge of "your customer's *other suppliers*" (e.g. new product opportunities) and "your suppliers' *other customers*" (e.g. redesigned or new sales channels) could be easily missed as they correspond to a less common point of view.

4.3.3 Key information to be provided by supply chain maps

As mentioned in Section 2.3.4 of Chapter 2, the minimal set of supply chain map elements are the companies (nodes of the network) and their inter-relations (arcs of the network). A rudimentary map may use the company name or some other unique identifier as the nodes' labels. And a possible inter-relation could be a "supplies" relation that may be visually represented by a directed arc from the supplier to the buyer, such as shown in Figure 2.7.

With the knowledge of potential beneficiaries and use cases discussed in this Section 4.3, one can conclude that such a rudimentary map can already represent actionable and, thus, valuable information:

- For instance, it may reveal the (previously unknown) suppliers of another company and, thus, information that could be used for supplier discovery, competitive intelligence or to learn that another company shares the same supplier.
- If such a map spans multiple tiers, it may already be sufficient to detect (previously hidden) supplier concentration on a sub-tier – either as confirmed or merely potential sub-tier suppliers.

The types of further key information provided by a supply chain map differ by use case:

- As mentioned in Section 2.3.4, for the purpose of supply risk management, *geographical information* can be of special importance since some types of risks have a geographical component (e.g. natural disasters, political risks). The location of the headquarters or the region a company is active in could provide a first approximation. Even more precise and useful for detecting a geographical concentration of sub-tier suppliers would be a geographical mapping of material flows between production sites and warehouses rather than legal entities. Figure 2.3 in Chapter 2 showed how supply chain risks can arise on different abstraction levels – and supply chain maps would have to mirror the relevant abstraction levels.
- To ensure that decisions are not made based on outdated information, a supply chain map may also provide *timestamps* or state a *time period* during which a relation was found to exist.
- A further common key information is the *type of product or service* that is provided by one company to another.
- If obtainable, the information for which *end-product* or programme (in the case of the aerospace industry) a product or service is used can help distinguish different supply chains and help with the identification of sub-tier suppliers.

In Section 6.6 of Chapter 6, some of these additional information requirements will be discussed again.

4.4 Review of existing supply chain mapping options

An exhaustive review of existing tools for supplier management, supply chain monitoring or supply chain mapping is not within the scope of this study. Nevertheless, even a rough attempt at mapping out the landscape of existing tools can provide a useful perspective. This section shall provide such overview but the examples given are in no way intended to represent a comprehensive list nor should the structure be interpreted as final.

4.4.1 Structuring the landscape of supply chain mapping solutions

To structure the landscape of supply chain mapping solutions, it appears useful to distinguish the following domains:

- Solutions for supply chain mapping versus supply chain management
- Solutions for gathering supply chain information versus visualising supply chain information

Supply chain mapping as part of supply chain management Supply chain management comprises a wide range of functions, such as order management or supplier relationship management. Often as part of supplier risk management, supply chain mapping is just one of the tasks that fall within the category of supply chain management, as illustrated in Figure 4.6. And, thus, a plethora of supply chain management tools exist – from often highly integrated software suites (e.g. SAP / Ariba Supply Chain Management, Oracle's E-Business Suite or Epicor's SCM) covering a wide range of supply chain management tasks to specialised tools just for specific functions. Limited visibility into the extended supply chain structure is not at the core of general SCM software if it is addressed at all.

In particular, there are a number of service providers and technical solutions for the problem of real-time supply chain visibility (e.g. Clearmetal[7]). This, however, refers to the tracking and tracing of goods in transit and the real-time monitoring of inventory levels. These solutions do not address the problem of discovering and monitoring the extended supply chain structure.

Information gathering vs. visualisation A key differentiator for supply chain mapping options is whether raw information is assumed to be provided by the user and can merely be visualised or whether new information about buyer-supplier relations is gathered and provided by the system to the user.

[7] `www.clearmetal.com`; last accessed: 2019-04-25

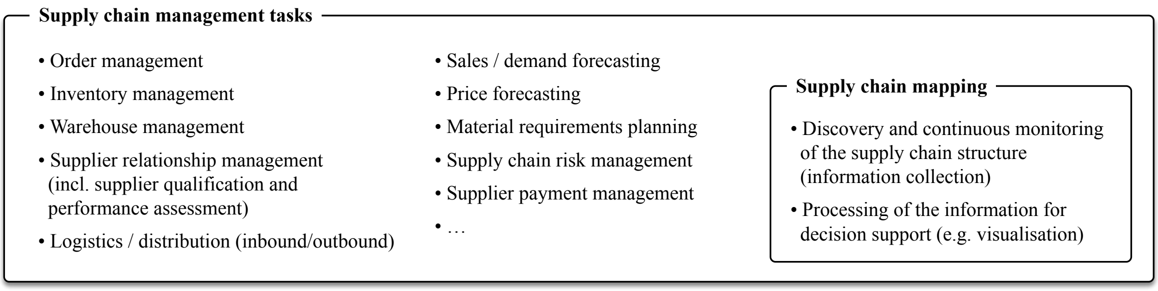


Fig. 4.6 Supply chain mapping is just one of the tasks that fall within the category of supply chain management

4.4.2 Existing options and limitations

Traditional options of obtaining buyer-supplier relations include supplier interviews, supplier surveys or interviews with domain experts. These options are limited in their scalability, and suppliers may withhold some of the information.

Modern options for acquiring information include third-party data providers that have often specialised in a particular industry, such as Markline[8] (automotive industry), IHS SupplierInsight[9] (i.a. automotive industry) or AirFramer[10] (aerospace industry) that collect and maintain information about direct suppliers (“who supplies whom”) and provided parts or services. The information is sold to customers via subscriptions or one-off payments. General business data providers, such as Reuters or Bloomberg, commonly also provide supplier information. In a corporate brochure, Bloomberg showcases a case study on the Volkswagen “Dieselgate” scandal where they are able to identify “all of Volkswagen’s American and European suppliers that derived at least 10% of their revenue from Volkswagen at the time of the diesel emissions revelations” (Bloomberg, 2018). The underlying information is likely to be sourced from manual analyst research as well as annual reports of publicly traded companies. The International Financial Reporting Standards for Operating Segments (IFRS[11]) require disclosures when an entity receives more than 10% (the percentage that Bloomberg stated in their corporate brochure) of its revenue from a single customer. If so, the entity must disclose this fact, the total amount of revenue earned from each such customer, and the name of the operating segment that reports the revenue. Although it appears that the name of the customer does not have to be disclosed. The Bloomberg

[8] `https://www.marklines.com/en/market_report/`; last accessed: 2019-04-25
[9] `https://ihsmarkit.com/products/supplierinsight.html`; last accessed: 2019-04-25
[10] `http://www.airframer.com`; last accessed: 2019-04-25
[11] `https://www.ifrs.org/issued-standards/list-of-standards/ifrs-8-operating-segments/`; last accessed: 2019-04-25

Terminal also offers some information about the direct suppliers and direct customers of companies, likely to be manually sourced from financial statements and news reports. Figure 4.7 provides a screenshot of such an overview for the aerospace company Safran. In this case, 96 (presumably direct) suppliers and 93 (presumably direct) customers were identified. Unexpectedly, an investment firm, Berkshire Hathaway, is listed as a supplier. Sourcemap[12] is a tool to visualise and analyse supply chains. Their limitation, however,

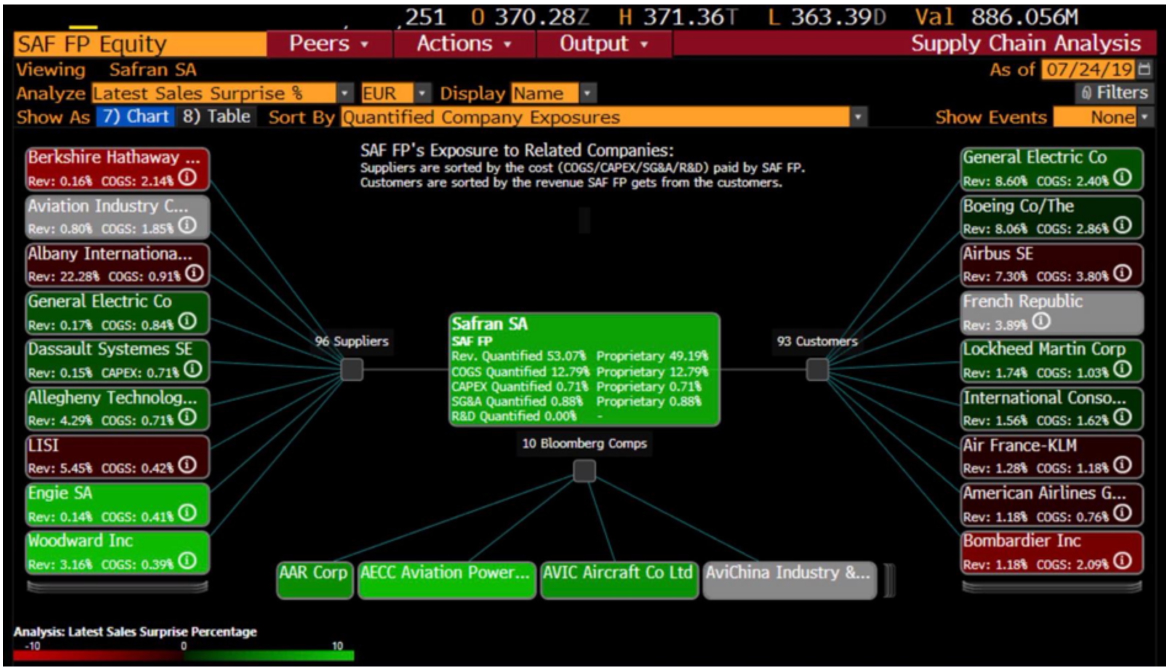


Fig. 4.7 Bloomberg Terminal: direct suppliers and direct customers of aerospace company Safran

appears to be that they do not acquire the required input to map and analyse supply chains. The discovery process appears to be limited to automatically dispatch "cascading sub-supplier mapping surveys to hundreds of suppliers at once". Figure 4.8 depicts the supply chain map visualisation provided by Sourcemap[13]. Recently, companies have emerged that aim to extract structured information from Web pages, such as Diffbot[14]. It is not known to what extent services like Diffbot have been designed to specifically capture and reason about company inter-relations.

[12] `http://www.sourcemap.com/`; last accessed: 30.08.2016

[13] The screenshot was taken from `https://www.sourcemap.com/blog/2019/6/11/evolving-tech-for-supply-chain-transparency-sustainable-brands-2019-panel`; last accessed: 20190-07-08

[14] `https://www.diffbot.com/`; last accessed on 2020-01-20

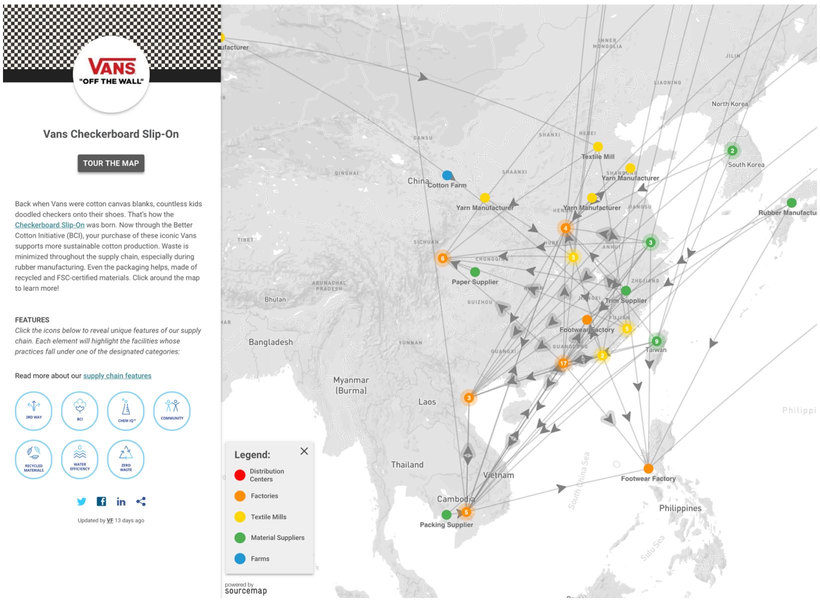


Fig. 4.8 Supply chain visualisation as provided by Sourcemap.com

A further high-quality data source for buyer-supplier relations is *customs duties information* which some countries make available. In the US, this information can be accessed thanks to the Freedom of Information Act. Using shipping addresses as well as information about the goods and shipper, inferences about direct buyer-supplier relations could be made.

Supply chain operating networks[15] (or supply chain collaborative networks) are typically cloud-based analytic applications that aim to improve supply chain visibility by sharing information (e.g. supplier KPIs) along the supply chain in real-time. Examples of such networks are SAP's Ariba/Supplier InfoNet[16], GT Nexus[17] or Elemica[18] (for the chemical industry). There are a number of either company- or industry-*specific* or industry-*agnostic* supply chain operating networks. These are B2B networks that integrate features known from Facebook, Amazon and Ebay of the B2C world. One of their limitations is that a company can typically only discover its sub-tier suppliers after a request has been sent to and permission has been granted by the direct supplier. Another limitation is the unwillingness of larger companies to share information about their supply chain with third party providers, such as supply chain operating networks. Consequently, larger companies often have established their

[15] E.g. see Banker (2016) for a use of the term
[16] `http://go.sap.com/uk/product/srm/supplier-performance-infonet.html`; last accessed: 30.08.2016
[17] `http://www.gtnexus.com/`; last accessed: 30.08.2016
[18] `http://www.elemica.com/`; last accessed: 30.08.2016

own networks for information sharing, such as Exostar[19] (Boeing) or SupplyOn AirSupply[20] (EADS, Airbus and others).

Risk management service providers Not dissimilar to the supply chain operating networks, there are a number of highly specialised service providers for supply chain risk management that also systematically collect data about supplier relations and other risk-relevant data points. One example of these providers is Resilinc[21] which considers as its core functionality to also provide customers with their supplier network as well as information about each supplier. The information about the supply network structure allows Resilinc to also offer event-based services on top – such as alerts and incident response planning. Another player in this field is Achilles which similarly claims to improve supply chain resilience by providing an overview of a company's total supply chain: "Understand your total supply chain, including suppliers who indirectly contribute components or services. Without this information, you have no way to rapidly detect risks and mitigate the failures that might impact customers, brand value, revenue and margins."[22]. Achilles, however, appears to rely on information collected from supplier questionnaires.

4.5 Summary

Companies are under increasing pressure to understand and manage their extended supply chains. On one hand, supply chains have become more global and complex so that the competitiveness of a company is also determined by the supply chain architecture. On the other hand, the exposure to supply chain risks has increased while the visibility into the supply chain structure has remained very limited.

Supply chain visibility is limited for multiple reasons Besides the increased complexity of supply chains which gave rise to the problem in the first place, the main reason is the "proprietary nature of each supplier's relationships with its partners" (Sheffi, 2005, p.13). Consequently, suppliers might not even fully know their own sub-tier supply network. Given the information asymmetry, a manufacturer would also not be in a good position to check the provided data for completeness and accuracy. Hence, contractually forcing the suppliers to disclose all their suppliers tends to be unsuccessful. This has been confirmed by interviews

[19] `https://www.exostar.com/`; last accessed: 30.08.2016
[20] `http://www.supplyon.com/en/airsupply.html`; last accessed: 30.08.2016
[21] `https://www.resilinc.com/`; last accessed: 30.08.2016
[22] `http://www.achilles.co.uk/en/for-buyers/supply-chain-solutions-portfolio/supply-chain-mapping`; last accessed: 30.08.2016

with an industry partner: Despite contractual obligations, suppliers would regularly not disclose some of their own suppliers to keep them as a "secret sauce"[23] of their recipe. Traditionally, companies also fail to provide *positive* incentives for sharing supply chain data (e.g. one could think of providing risk management services to suppliers in exchange for information about the sub-tier network). Finally, secondary sources of information such as third party data providers are also problematic for various reasons, e.g. questionable data quality, costs, lacking up-to-dateness. The following overall conclusions from this industrial study are made:

a) There is a need for visibility of the supply chain structure Knowledge of the supply chain structure is a valuable input to a wide range of decision problems in the corporate world, such as supply risk management, food security, legal requirements (e.g. prevention of modern forms of slavery), competitive intelligence, supply chain efficiency, and supply chain sustainability. Beneficiaries are not only the companies themselves that manage their supply chains but also governmental agencies, insurance companies, consultancies, or research institutes.

b) Existing options for increasing structural supply chain visibility are problematic Direct suppliers have an incentive to *not* disclose all of their suppliers, and the adherence to contractual agreements demanding their disclosure cannot be easily checked. Supply chain operating networks tend to require the permission of suppliers before they can be disclosed to companies downstream the supply chain. Datasets about supply chains can be purchased but are expensive, and quality and origin of the information cannot be easily checked. Technologies like RFID tracking & tracing help with exchanging real-time information of goods in transit but generally do not address the problem of discovering supply chain participants in the first place.

These conclusions drive the main focus for the remainder of this thesis: namely exploring methods which contribute towards automated supply chain mapping.

[23]Quote from one of the initial interviews with supply chain risk managers

Chapter 5

Automating the extraction of buyer-supplier relations from text

5.1 Introduction

In Chapter 3, the research scope was defined to be the overall problem of automatically generating supply chain maps from unstructured, natural language text which was supported by the industrial findings in Chapter 4. This scope was further narrowed down to the research focus: the core problem of extracting *individual buyer-supplier relations*. It was argued that a supply chain map in its most basic form consists of nodes and arcs which represent individual organisation and their buyer-supplier relations. Thus, the automated extraction of individual buyer-supplier relations is a fundamental and necessary building block for any, more comprehensive approach to automate supply chain mapping from text. A collection of extracted individual relations could already be directly visualised as a (non-transitive[1]) network.

This chapter represents the core research and main contribution of this thesis. The aim of this chapter is to address the problem of automatically extracting individual buyer-supplier relations from unstructured, natural language text, as illustrated by Figure 5.1 for a single sentence and relation. The aim is to extract simple buyer-supplier relations between two organisational entities ("who supplies whom" as indicated by directed arcs) from unstructured, natural language text, such as news reports.

[1]See Chapter 6 for a discussion of the transitivity property.

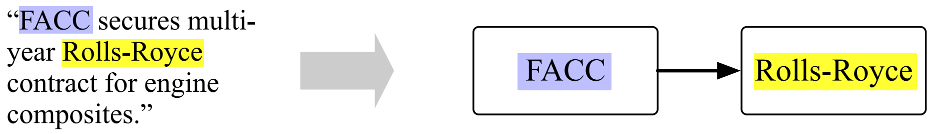


Fig. 5.1 Buyer-supplier extraction

Across a collection of documents, the problem can be illustrated as shown in Figure 5.2: A collection of documents is converted into a list of triples. Each triple consists of:

a) two organisational named entities which are identified first

b) the relationship class which is predicted to exist between them

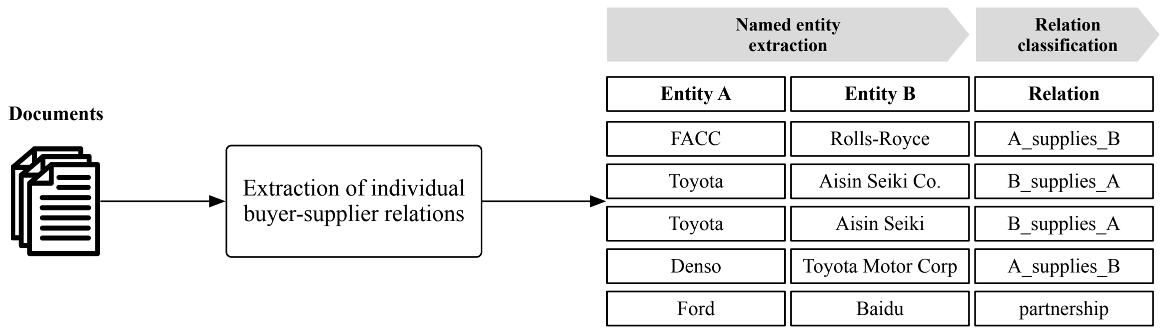


Fig. 5.2 Buyer-supplier extraction across multiple documents: The aim is to extract triples of Entity A, Entity B, and an appropriate relationship class.

Borrowing from the language used in the domains of ontologies and Linked (Open) Data[2], the problem can also be represented using a subject-predicate-object logic, with 'isSuppliedBy' and 'suppliesTo' as mutually inverse predicates. The resulting list of triples can then be converted into a directed (or mixed[3]) graph.

By addressing the above problem, this chapter proposes a methodology to answer Research Question 1 and its sub-questions. After the problem has been further defined and formalised, the problem is broken down into two stages in a natural sequence: (1) the creation of a labelled text corpus and (2) the design, training and testing of a classifier. Each of these stages correspond to one of the sub-questions of Research Question 1.

[2] Cf. Chapter 2 on ontologies and Linked (Open) Data

[3] Directed arcs for 'suppliesTo' and 'isSuppliedBy' relations and undirected arcs for partnerships or other undirected trading relations.

In an industrial context, the classification model would, in principle, only need to be trained once. A pre-trained model can then be applied arbitrarily often to unlabelled and previously unseen datasets. That said, in practical applications, the sequence may be less linear and more *cyclical* (e.g. in active learning scenarios with "humans in the loop" who may frequently provide new training examples and trigger a re-training of the classifier). The process is shown in Figure 5.3. As long as the confidence score for relations is not reliable, sentences predicted to have a relevant relation can be manually labelled and added to the training data. This way, the share of false positives can be successively reduced. Once confidence scores are more reliable, an "active learning" approach[4] can be pursued where specifically those data points the classifier was most uncertain about are labelled by a human.

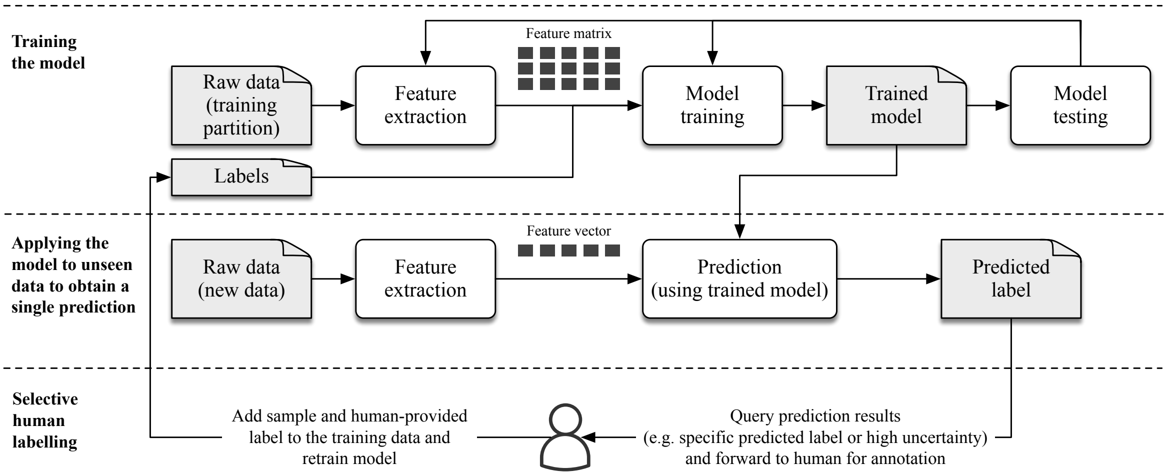


Fig. 5.3 Common procedure for iterative improvement of the training data with a human-in-the-loop, such as known in active learning.

5.2 Buyer-supplier relation extraction as classification problem

5.2.1 Problem scope and requirements

This section shall provide the rationale for focussing on buyer-supplier relation extraction and additionally discusses further scoping decisions.

[4] Active learning was outside the scope of this thesis but shall be mentioned as a common option for continuously improving any classification model.

Rationale for focussing on relation classification

Figure 5.4 illustrates the focus of this chapter conceptually. The focus lies solely on the extraction of individual buyer-supplier relations expressed between two organisational named entities. The chapter thereby deliberately ignores research problems and processing steps before and after the relation extraction phase, which will be returned to again in Chapter 6.

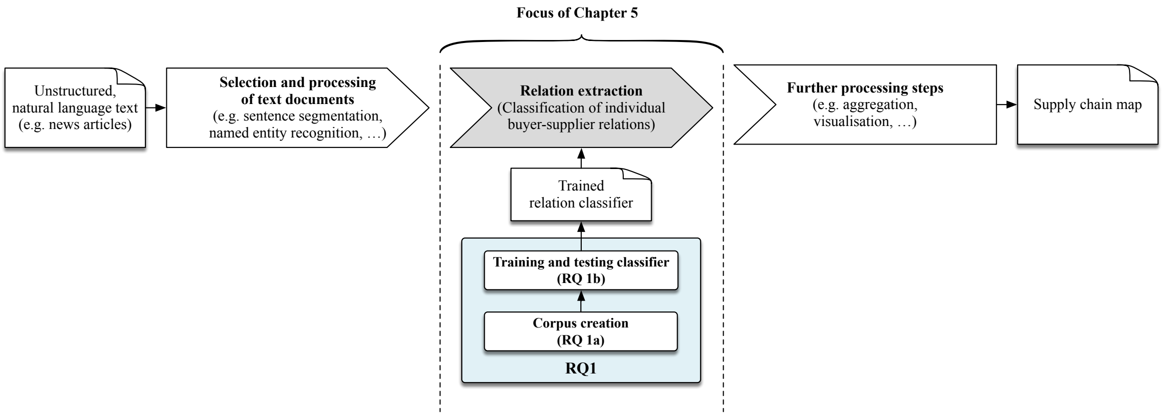


Fig. 5.4 Chapter focus on the classification of buyer-supplier relations

Further scoping decisions and requirements

The scope is further refined along various dimensions and shall be made explicit at this point. A short summary of these scoping decisions and requirements is provided by Table 5.1. For better readability, a discussion of these aspects and the rationale for the scoping decision are provided in the Appendix.

Table 5.1 Scoping decisions for relation extraction

Scoping aspect	Requirement / chosen scope
Language	English language
Ability to cope with large number of variations	Approach needs to be able to cope with thousands of variations of how buyer-supplier relations can be expressed.
Relation types	Focus on explicitly stated relations between two organisations.
Relation classes	Ability to distinguish at least four classes: “company A supplies company B”, “company B supplies company A”, “undirected buyer-supplier relationship or partnership”, “no buyer-supplier relationship stated”.
Mention-wise classification	Each possible pairing of two organisational named entity mentions in a given text needs to be classified. Even if the same named entity is mentioned multiple times, these mentions shall be considered in separation.
Company names cannot be features	To achieve a generalisable classifier, names of specific companies cannot be features and need to be hidden from the classifier.
Contextual scope	The contextual scope is limited to single sentence and no co-reference resolution shall be performed.
Inference	Instead of also inferring buyer-supplier relations from context, relations shall be extracted from explicit statements.
Text source	The classifier shall be able to work with general news.
Task scope	The scope of the task is relation extraction only. Further tasks, such as named entity disambiguation, shall not be considered at this stage. Similarly, the veracity of the statements found in the input text is *not* checked.
Information quality (incl. availability)	When the classification performance is measured, the quality of the input data, such as availability or correctness, shall be ignored.
Time-dependence of the structural information	Time-dependence of structural information shall be ignored at this stage.

5.2.2 Problem formalisation

In conclusion of the requirements above, the problem can be formalised as follows:

Problem formalisation:

For a given single, self-contained sentence in the English language, all pairs of detected organisational entity mentions shall be classified with respect to the existence of a buyer-supplier relation explicitly stated between them.

The term “mentions” is used in above statement to clarify that repeated mentions of the *same* organisational entity have to be considered separately. Relations that would require the resolution of pronouns to company names (co-reference resolution) will be ignored. Furthermore, more multi-faceted relations, such as extracting supplied products or further companies involved, shall be considered out of scope at this stage. Each relation shall only be assigned one single class, so that the overall problem can be characterised as a *multi-class classification problem* as opposed to a multi-label classification problem. Each class may then represent a relation type, such as “company A supplies company B” , “company A is supplied by company B” (inverted direction), “company A and B engage in a partnership” or “no buyer-supplier relation stated”. The class definitions used can be found in the corresponding section below.

5.2.3 Structuring the problem

The problem of developing an approach for extracting individual buyer-supplier relations can be split into three stages, as illustrated by Figure 5.5. The first two stages correspond to the two research sub-questions of this chapter. The third stage is shown for the sake of completeness: Once a relation classifier has been trained, the pre-trained model can be applied repeatedly. Stage 1 and 2 do not have to be executed every time the classifier is applied.

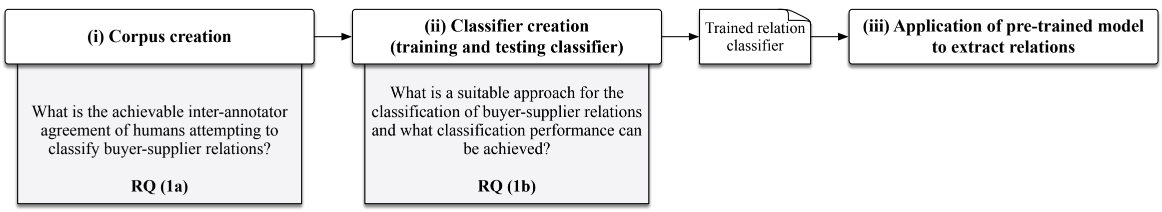


Fig. 5.5 Three stages for relation extraction

These stages have a natural sequence:

(i) Corpus creation

This stage consists of both the collection and preparation of the text input to be used for the corpus as well as the process of labelling sentences by human annotators. The importance of the corpus is two-fold:

- *"Gold standard" text corpus:* In order to evaluate classification performance, a human-labelled text corpus is required that will be considered the ground truth against which the classifiers' predictions can be compared. The gold standard data will allow us to measure both recall and precision. If a Machine Learning classifier is used (as opposed to one based on pre-defined patterns), the dataset also serves as a training dataset. The dataset should be representative of the data that the classifier is expected to encounter when applied in practice.
- *Suitability for automation:* Furthermore, higher inter-annotator agreement suggests a more manageable, formalisable task that is more likely to be suitable for automation. If it is impossible to establish a ground truth among human annotators, a classifier cannot be expected to perform well on the problem. In advance of this research, it is not obvious at all that the task of classifying buyer-supplier relations is simple or formalisable enough for annotators to agree.

(ii) Classifier creation

Only once a gold standard dataset has been established, classifiers can be designed, trained and tested. Recall, precision and their harmonic mean, F_1 score, can be used to identify the best-performing one. The achievable performance is dependent on the size and quality of the corpus. A high recall would *not* yet suggest that supply chains can be fully reconstructed; this would depend on the information availability which is not tested in this stage. Similarly, errors in any processing steps prior to the relation classifications are *not* captured, such as errors in the detection and disambiguation of organisational named entities.

(iii) Application of pre-trained model

Once a classification model has been developed and trained, it can be repeatedly applied to previously unseen, unannotated data.

5.3 Creation of a human-labelled text corpus

In this section, the creation of an initial human-labelled reference text corpus will be discussed. A number of datasets have been published to allow users to train classifiers of semantic relations[5]. A dataset for the problem of buyer-supplier relations has not been created yet. Hence, the need to create such a text corpus.

5.3.1 Objective and requirements

The objective is to define useful relationship classes and to create a text corpus of several thousand human-annotated sentences. The purpose of the corpus is to enable a quantitative assessment of how well any relation extraction method performs, and to serve as a training dataset for a Machine Learning classifier. The annotation process also aims at answering questions regarding the difficulty (including the ambiguity) of the task.

The text corpus shall be representative of a general news set, such that a classifier's performance measured on the dataset is a good predictor of its performance on a previously unseen general news dataset. This is a challenging requirement. With limited funds and time, the expected size of the text corpus achievable within the scope of this research is relatively small. In addition, the share of sentences describing buyer-supplier relations in a general news dataset is low.

5.3.2 Methodology

The following methodology for the corpus creation is structured as illustrated by Figure 5.6.

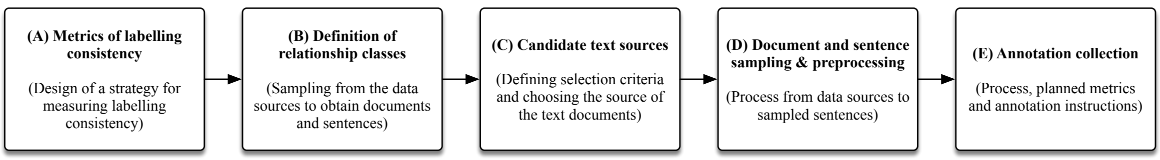


Fig. 5.6 Stages of corpus creation

The general aspects of the methodology shall be discussed in the following subsections, using the structure provided by Figure 5.6. The specific choices, experimental settings and results can be found in Chapter 7 where the methodology is applied.

[5]A collection of these datasets can be found at `https://github.com/davidsbatista/Annotated-Semantic-Relationships-Datasets/tree/master/datasets`

(A) Metrics of labelling consistency

The metrics of inter- and annotator agreement have already been discussed. Sentences will be annotated redundantly both across annotators (*inter*-annotator agreement) as well as for the same annotator (*intra*-annotator agreement). For the sake of simplicity, all collected annotations (the class for each pair of organisational entities in a sentence) are assumed to be independent. In reality, this is not strictly the case: Within a sentence, the class selection for one pair of organisational entities may affect the class selection for the next pair. For the calculation of inter-annotator agreement and the training, all annotations are treated as independent.

(B) Definition of relationship classes

Classes for the distinction of relevant relations need to be defined *in advance* of the human labelling process. Due to the importance of directionality in supply relations, two classes are designed for explicitly expressed, directed buyer-supplier relations: (a) “company A supplies company B”, and (b) “company A is supplied by company B” (inverted direction). The actual sentence in the news article may use different words to express this relation, such as “purchasing from” or “using parts from”. These relations may be expressed in any tense (past, present or future) since the tense could later automatically be identified if necessary. Furthermore, these relations should be expressed as certain, factual statements rather than a possibility. These relations are not limited to the purchase of physical goods, parts and material but also include the use of services, such as logistics services, or intellectual property.

This leaves a set of other ways a buyer-supplier relation may be expressed: Collaborations, joint ventures, and other forms of partnerships do not have an obvious directionality but may still result in dependencies that are relevant for a supply chain map. An example for this type of sentences could be “company A and company B sign a contract for the supply of”. An initial analysis also showed that actual buyer-supplier relations are often euphemistically described as partnerships or collaborations. Furthermore, buyer-supplier relations can sometimes only be implied, may be ambiguous or can be explicitly stated as uncertain. Examples of such cases are: “company A is in talks with company B over the purchase of” or “company A plans to buy from company B”. To avoid too many different classes and to ensure that only one class is ever applicable, these cases are grouped into a single third class of buyer-supplier relations (c) that are undirected, or implied or stated as uncertain. This class also aims to avoid “contamination” of the first two classes with examples that lack directionality, certainty, or explicitness.

A further class of relations (d) is distinguished for statements in which one organisation owns another (fully or partially) or is part of another organisation. Without this class, such relations could be misinterpreted as normal buyer-supplier relations. Examples of this class are instances where a company buys a stake in another company or where a company's department provides something to the parent company. The purpose of this class is less to obtain ownership relations which could be obtained from publicly available reports or databases but to facilitate the annotation decisions and ensure the purity of the other buyer-supplier relationship classes. Investment relations can sometimes also indicate existing or future buyer-supplier relations[6]. Because the named entity recognition is performed automatically as part of the pre-processing of a sentence, errors may occur where a labelled text sequence is not an organisation or was incorrectly segmented. To keep the task complexity manageable and to ensure the identical NER tagging results as a starting point for all annotators, annotators were not asked to rectify incorrect NER tags. Instead, a relation can be classified as 'reject' (e) in that case or other circumstances where an annotator felt incapable of assigning a class. This way an annotator is not forced to provide an incorrect or otherwise inappropriate response. Finally, the case that none of the above classes are appropriate is captured by a last class (f). This is the most common case for sentences randomly obtained from news articles. The final class definitions are shown in Figure 5.7.

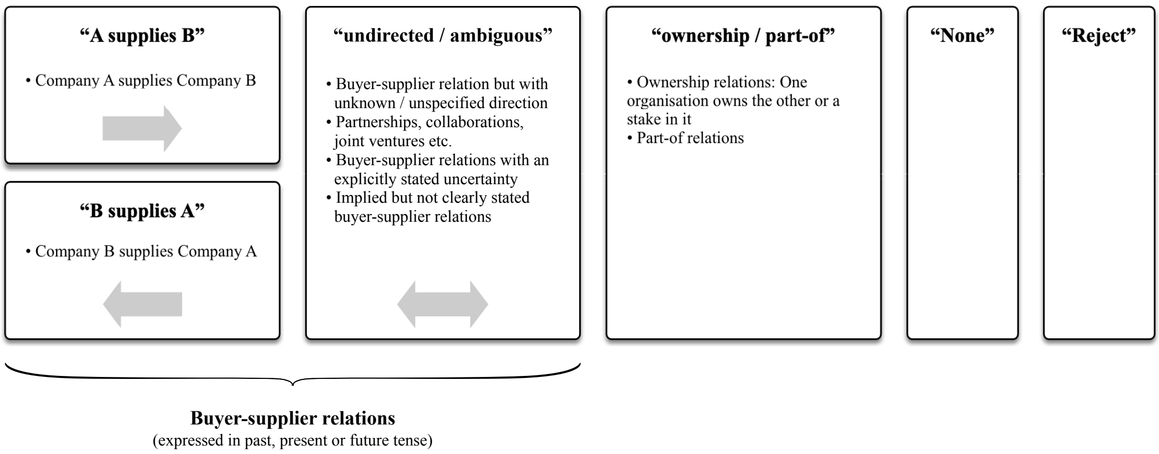


Fig. 5.7 The created corpus uses the depicted relationship class definitions.

The class definitions presented above represent a compromise between labelling simplicity and and the expressive power of the chosen set of labels. Additional or other class defini-

[6]E.g. "Boeing reports that it has invested in Gamma Alloys, a US aluminum alloy specialist which is developing metal-matrix composites for use in aerospace, automotive and other industries."

tions and even a multi-labelling scenario were considered but have been excluded here. A discussion of these aspects can be found in the Appendix.

(C) Candidate text sources

A wide range of general news datasets are available; Table 5.2 shows a small selection of datasets that were identified as candidate text sources for this research. All listed datasets contain business news and, thus, articles that mention buyer-supplier relations.

Table 5.2 Example selection of datasets identified as candidate text sources for this research

Name	Description
Reuters TRC2	Thomson Reuters Text Research Collection (TRC) consisting of 1.8 million general news articles from 2008 to 2009 (`http://trec.nist.gov/data/reuters/reuters.html`)
Reuters RCV1	Thomson Reuters dataset consisting of 800k articles in the English language, from 1996 to 1997 (`http://trec.nist.gov/data/reuters/reuters.html`)
Reuters RCV2	Thomson Reuters dataset consisting of 400k articles in 13 languages, from 1996 to 1997 (`http://trec.nist.gov/data/reuters/reuters.html`)
Custom data from webhose.io	Webhose.io (`www.webhose.io`) is a commercial data provider that grants users to download a certain number of Web documents per month for free. Via filter criteria, the type of the scraped Web content can be specified.
Signal Media	1 million news and blog Web documents mostly from 2015 provided within the context of the First International Workshop on Recent Trends in News Information Retrieval (`http://research.signalmedia.co/newsir16/signal-dataset.html`)
Common Crawl News	The Common Crawl News dataset (`http://commoncrawl.org/2016/10/news-dataset-available`) consists of terabytes of news articles that have been scraped off the Web on a monthly basis since August 2016. The data is available via AWS S3 in WARC format and consists of raw HTML documents.
BBC news	The BBC news dataset (`http://mlg.ucd.ie/datasets/bbc.html`) consists of 2225 documents from the BBC news website corresponding to stories in five topical areas (incl. business, politics, and tech) from 2004-2005.
Kaggle Million Headlines	This dataset published on Kaggle (`https://www.kaggle.com/therohk/million-headlines`) consists of 1.1 million news headlines (only) published by the Australian Broadcasting Corp. over a period of 15 years.

For this research, the datasets are chosen for the corpus creation based on the following criteria:

- Large number of documents (at least 50,000 documents) in the English language

- Documents should contain both headlines and the text body of the article because either may contain buyer-supplier relations but tend to use a different language style
- Ideally, a standard corpus used in NLP research comprised of documents that have already been cleansed and brought into a standard format
- Unbiased collection (general news, at least covering multiple months)

Since the text corpus is not used to learn new buyer-supplier relations and it can be assumed that the linguistic patterns use to describe buyer-supplier relation did not change significantly over the last decades, the age of the datasets is considered less important for training a classifier.

(D) Methodology for document and sentence sampling

The aim of this phase is to create a pool of sentences from which sentences can be randomly selected and presented to the annotators. To obtain an unbiased dataset representative of general news, sentences should be *randomly* sampled from such general news sources. However, because sentences need to be manually annotated, the dataset cannot be too sparse so that annotators spend most of their time annotating sentences without any buyer-supplier relation. These sentences shall henceforth be referred to as negative examples, whereas positive sentence express at least one relevant relation. Because most sentences in any news dataset are negative, annotated negative sentences are not that valuable. Any sentence that was randomly sampled from a general news dataset would likely be negative, and hence the availability of negative samples to the training dataset is not a bottleneck, but the availability of positive examples. This data imbalance results in a trade-off between a perfectly unbiased dataset and the time or money required for labelling the dataset.

Overall sampling process The overall sampling and pre-processing methodology is shown by Figure 5.8 and combined two sampling approaches to equal parts.

One approach samples documents in three partitions: one partition for random documents drawn from a general news dataset, a second partition for documents that are retrieved using keywords related to selected focus industries (aerospace and automotive), and a third partition for documents that are retrieved based on a search for company names in these focus industries. The underlying rationale for this approach was that the trade-off between the expected relevance of a sentence and its bias could in principle be steered by adjusting the proportion of each partition in the final sample. However, initial tests quickly revealed that the proportion of positive sentences was still too small for an efficient annotation by humans. Human annotators would spend most of their valuable time annotating negative samples. To address this problem, candidates for positive sentences need to be manually collected by multiple researchers and stored in a further data partition. Only sentences with two or more detected organisational named entities are admitted to the annotation process. Sentences can be manually or automatically NER-tagged. The automatic NER tagging has benefits and drawbacks that are discussed in the Appendix. The details for each of the two approaches can also be found in the Appendix.

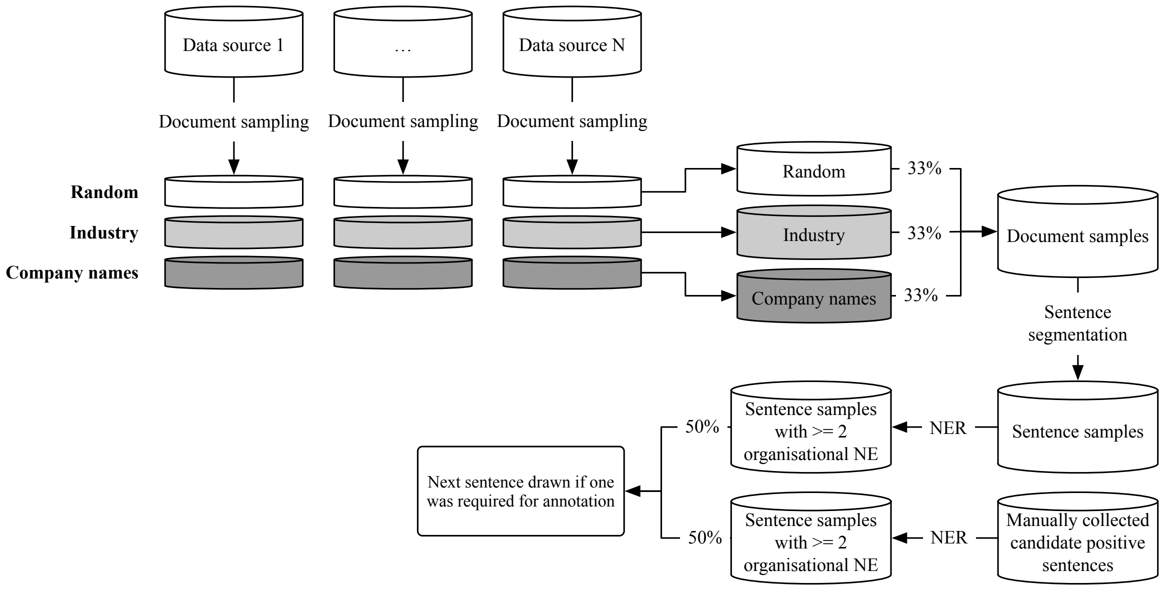


Fig. 5.8 Document and sentence sampling methodology

(E) Annotation collection

This section shall outline what annotations need to be collected and how this process can be conducted.

Pair-wise labelling Annotators have to label *each possible pair* of named organisational entities. The order within the organisation pairs can be ignored since the relation class captured the directionality already.

Redundant labelling for inter- and intra-annotator agreement To measure *inter*-annotator agreement, a subset of all sentences has to be labelled by all annotators. The first, e.g. 40, sentences of the first annotation session are presented to all annotators in identical order. To measure inter-annotator agreement beyond this inter-annotator set, sentences are also labelled redundantly. That is, identical sentences can get labelled by more than just one annotator. Inter-annotator agreement is measured on the basis of two types of annotation sets:

1. A fixed set of sentences (the *"inter-annotator set"*) that were randomly sampled. These sentences are the first sentences that each annotator had to annotate. They have to be identical for *all* annotators and also have to be annotated in identical order.

2. Any set of sentences that any two annotators both annotated. This includes but may far exceed the inter-annotator sentences.

In addition, within each labelling session of some fixed number of sentences from the beginning of a session are randomly re-injected towards the end to measure *intra*-annotator agreement. These sentences do *not* have to be identical for different annotators and are just repeated within a specific session of a specific annotator. A schematic overview of this redundant labelling is provided by Figure 5.9.

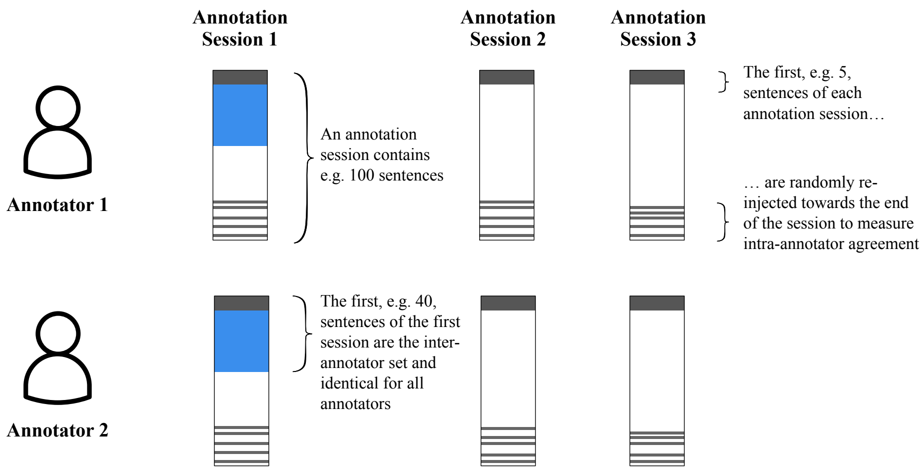


Fig. 5.9 Schematic overview of redundant labelling for inter- and intra-annotator agreement.

Despite the redundant labelling, the classification algorithm needs to be presented with a *single*, assumed to be correct answer per pair of two organisational entities in a sentence. To obtain this single answer, a *simple majority vote* is performed over all responses for the same pair of organisational entities in a specific sentence[7]. For the sake of simplicity, no details of the voting results, such as the level of agreement, are forwarded to the classification algorithm.

[7]The simple majority vote is performed across all annotation responses. If an annotator annotated a sentence multiple times, each annotation is a separate vote. A future extension could consider the aggregation votes for each annotator first, or even weights based on past performance.

Entity masking If the sentences were to be presented to the classifier as the original text, then the classifier may start to consider the mentioned organisations as features. It would start to memorise company names and inter-relations between specific companies. This, however, would lead to a model that is not able to generalise to other organisational mentions. The context is supposed to be the feature, but not the entity name itself.

Thus, specific names of organisational entities need to be masked in the text corpus. Three types of masks can be used[8]. One mask each can be used for the two organisational named entities in question ("__NE_FROM__" and "__NE_TO__"). The names of these masks do *not* refer to the origin or destination of a material flow. They merely indicate the origin and destination of the relationship arc. A further mask ("__NE_OTHER__") can be used for all other organisational named entities also mentioned in the sentence. Instead of just ignoring other organisational named entities, this mask is used to indicate their existence. This can help the classifier, e.g. with recognising instances where a list of suppliers is mentioned. Each of the three masks is subsequently assigned a separate embedding vector. The specific text strings used for the masks do not matter; however, the text strings should be sufficiently uncommon so that no tokens in a news article accidentally get confused as masks. The way the masking is conducted is illustrated by Figure 5.10. For each combination of two organisational named entities in a sentence, a new sentence version with corresponding masks is generated. Each of these masked sentences represents a *single* potential relation between two named organisational entities and has been assigned a relationship label during the annotation process.

Fig. 5.10 Entities in an NER-tagged sentences are masked.

[8] A similar approach of masking in the medical domain is described by Bar (2016) to identify the existence of a relation between microRNA and genes.

By only choosing two masks for the two named entity mentions in focus, the problem is simplified. In some sentences, a mention may be repeated and this information gets lost in the masking process. See the following authentic sentence as an example:

Example sentence
"In 2013, Boeing and Mubadala announced a new Framework Strategic Agreement to increase the long-term role of Mubadala as a direct supplier, including support as Mubadala developed prepreg manufacturing in the UAE."

The company name "Mubadala" is repeated multiple times. From reading the first words, it can only be concluded that Boeing and Mubadala have announced an agreement, which makes the relation a "partnership / undirected / ..." (class 3). However, after considering the first repeated mention of "Mubadala", it becomes clear that "Mubadala" is the direct supplier to Boeing. These repeated mentions will not be considered as such and will be masked the same as any other organisational named entity mention.

Terminology When a sentence is annotated, an "annotation item" is generated. Because a sentence can get annotated multiple times (by the same annotator for the purpose of measuring intra-annotator agreement or other annotators to measure inter-annotator agreement), each sentence may have multiple associated annotation items. Each annotation item contains human-generated labels about the sentence itself but also labels for each pair of organisational entities. Labels of pairs of organisational entities are also referred to as "arcs" since they describe the relation between the two organisations.

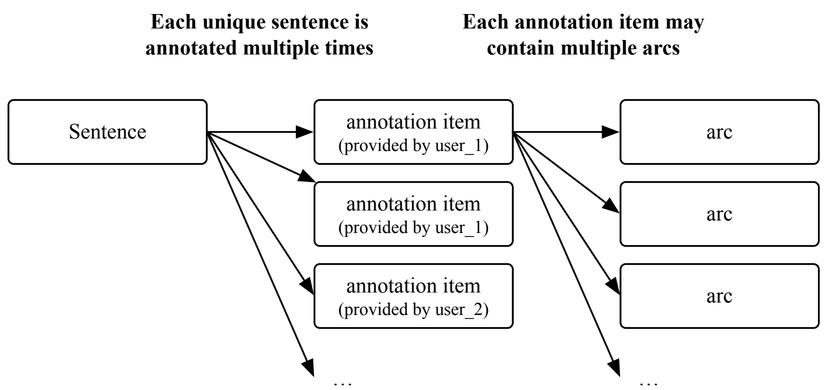


Fig. 5.11 Each sentence can be associated with multiple annotation items, and each annotation item can contain multiple arcs.

Overall labelling process The overall process for an example sentence is illustrated by Figure 5.12. Each sentence may contain multiple pairs of organisational named entities. Any such pair – sometimes visualised as an "arc" between two organisations – requires a label. To prevent the classifier from "memorising" relations between specific organisations, their names need to be *masked*. As a proxy for the feasibility of the labelling task and the labelling quality, *inter-annotator agreement* and *intra-annotator agreement* are measured. For this, at least some random examples need to be labelled *redundantly* by different annotators and even the same annotator. A simple majority vote then determines the single label that is assumed to be the ground truth.

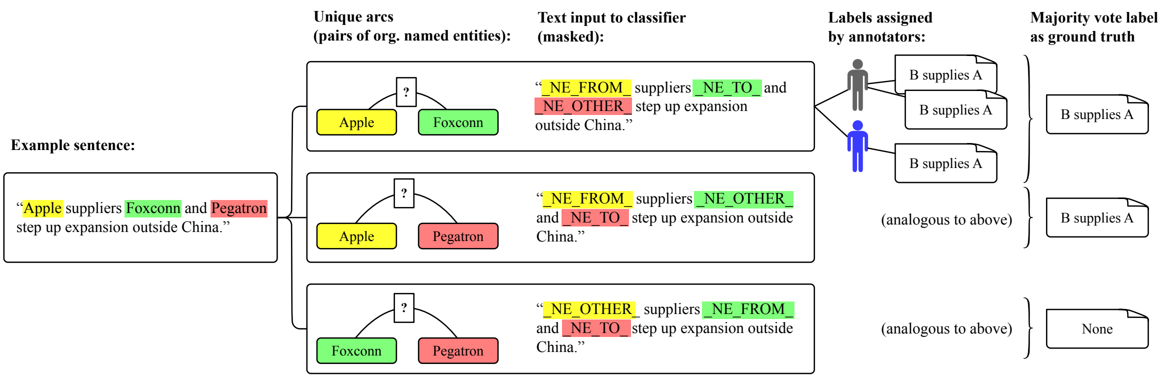


Fig. 5.12 Overall labelling process

Summary of the methodology

This section presented the methodology for creating a corpus that can be used to assess the performance of a relation classifier and that can serve as a training dataset. It was decided to measure labelling consistency using inter- and intra-annotator agreement. This agreement will then serve as an indicator for the labelling quality and the feasibility of the task. Five basic relation types, including two directed and one undirected buyer-supplier relations, were defined as well as one additional class in case an annotator feels the need to abstain from a labelling decision. Selection criteria for the text sources were defined mainly to ensure that a classifier trained on the data will generalise. An approach to sample documents from the general text sources needed to be designed. A fully random approach would result in too few positive examples. Thus, an approach with 3 partitions (random, focus industries, focus companies) was designed. Since even this approach would result in a low share of positive examples, a decision was made to manually sample and add candidate positive sentences. Labels had to be collected for each pair of two organisational named entity mentions in a sentence and, to measure inter- and intra-annotator agreement, the labelling had to be performed redundantly for the same sentence and even for the same annotator. Simple majority voting was chosen as a mechanism for defining which label shall count as the ground truth. The names of companies needed to be masked to ensure that any classifier trained on the data will be able to generalise beyond specific company names.

Figure 5.13 shows how a labelled sample can be represented. The masked sentence is stored as “x”, the label is stored as “y” and was determined via a simple majority vote. In this example, the sentence was only labelled once. A collection of these “x” and corresponding “y” can be used to train a classifier.


```
{
  "id": "e4c77037c3567b45566492d308cf386debb6d51f",
  "originalText": "Teijin to Supply Carbon Fiber Composite to Airbus",
  "relations": {
    "f454addcc81c242379ab82e6edcf3da5028170ec": {
      "id": "f454addcc81c242379ab82e6edcf3da5028170ec",
      "votes": {
        "2": [
          "5ac41236acb33b3ef8e0d678"
        ]
      },
      "x": "__NE_FROM__ to Supply Carbon Fiber Composite to __NE_TO__",
      "y": 2
    }
  }
},
```


Fig. 5.13 Structure of a sample input

5.4 Creation of a buyer-supplier relationship classifier

5.4.1 Objective and requirements

As stated in Section 5.2.3, the objective of this stage is to identify a suitable relation classification method and test its performance against the gold standard dataset, i.e. the previously created text corpus. Once a classifier has been trained, it can be applied to large quantities of previously unseen data to extract buyer-supplier relations.

A key requirement for the classification method is the ability to distinguish the directionality of a buyer-supplier relation. The classification method should also allow for an easy, time-efficient addition of new “patterns” to look for. Ideally, the general approach should also be extendable to additional languages apart from English.

5.4.2 Methodology

The methodology for the relation classification shall discuss the following aspects:

(i) Classes

(ii) Relation extraction methods

(iii) Feature set

(iv) Metrics for evaluating classification performance

(v) Process for training and testing

(i) Classes

For the training and testing of the classifier, the classes 'no buyer-supplier relation' and 'rejected by annotator' are merged as any application based on the classifier would likely treat the 'reject' class the same as a non-existing relation, especially in case of incorrectly identified organisations. This results in *5 classes* that need to be distinguished by the classifier.

(ii) Relation extraction method

Options for relation extraction methods The general options for relation extraction have been mentioned in Chapter 2, and shall be revisited briefly from the perspective of the problem at hand.

A fully *unsupervised approach* based on the co-occurrence of organisational named entities is not sufficient for the problem at hand. The reason for this is the required directionality of the buyer-supplier relations. Unsupervised approaches (especially word embeddings or deep language models), however, can be used in addition to the chosen method. Simple *lexico-syntactic patterns* were used in the early stages of this research. As expected, they proved to achieve high precision but poor recall. In addition, the definition of these patterns is time-consuming. As opposed to using word embeddings, these patterns also do not capture the underlying semantics, such as the similarity in meaning of "to buy" and "to purchase". *Semi-supervised approaches*, like distant supervision or bootstrapping, run the risk of *semantic drift*. This requires relabelling to identify false positives. Figure 5.14 illustrates how a Distant Supervision process could look like for the problem at hand. The figure also shows an example of an incorrectly derived label: "Apple stock rises after Foxconn reports earnings." is not a positive example. There can be other reasons why a stock rises after another company reports earnings. It does not have to be a buyer-supplier relation.

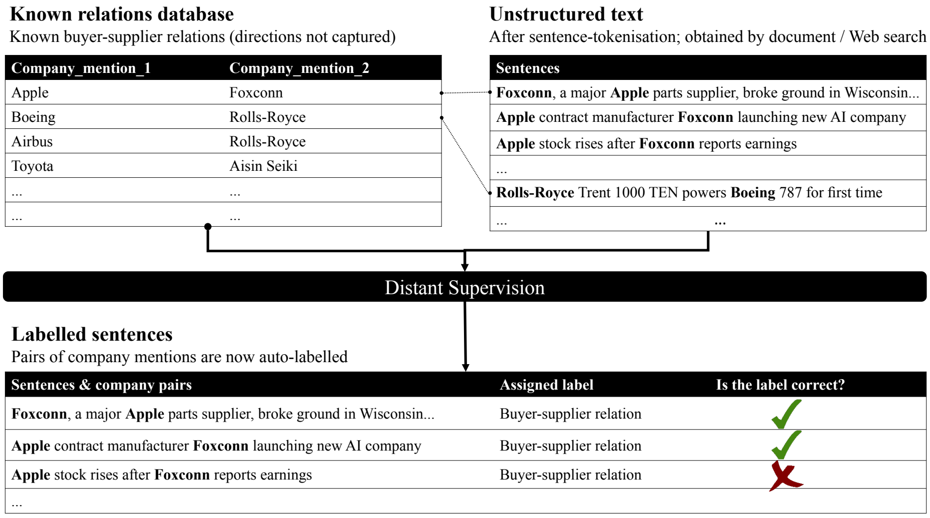


Fig. 5.14 Distant supervision leverages a database of entities with a known relation and a database of unstructured text.

A number of reasons appear to suggest that distant supervision alone may *not* be sufficient for the problem at hand: Firstly, companies are often mentioned together even if no buyer-supplier relation is expressed. E.g. “Today, shares of Apple and Foxconn rose by 10%.”. If the sentence was just labelled based on the co-occurrence of both entity mentions, the wrong label would be assigned. Secondly, the existence of two entity mentions does not allow to distinguish between different directionalities. Just knowing that two entities are related is not enough, we require the direction. Thirdly, if a company is mentioned multiple times in a sentence, it is unclear which mention is having a relation with the other company. Lastly, there is a limited availability of buyer-supplier databases. Some are available, such as so-called approved supplier lists of aerospace companies. Distant supervision could, however, be used to efficiently filter text to collect candidate positive examples that could be injected into the annotation process.

The remaining option is a *supervised learning approach*. The major drawback of the supervised approach is that it requires a labelled and representative training dataset of sufficient size.

Selection of suitable relation extraction method Following the rationale outlined above, a supervised learning approach is chosen. Recent research in NLP has shown that current Deep Learning based methods using pre-trained word embeddings tend to outperform simpler classifiers. Nevertheless, before immediately relying on more complicated methods, it appears sensible to first try simpler supervised classification methods, such as a linear support vector machine (SVM) classifier (Boser et al., 1992). In the domain of Machine Learning and NLP, new network architectures are published frequently and the state-of-the-art is a fast-moving target. As summarised in Section 2.4.3, current state-of-the-art classifiers include simple multi-layer perceptrons, convolutional neural networks and recurrent neural networks, the latter group includes LSTM (Hochreiter and Schmidhuber, 1997), BiLSTM (Graves and Schmidhuber, 2005), and GRU (Cho et al., 2014; Chung et al., 2014). The algorithms selected to be used for this research are as follows:

- Linear SVM (with bag-of-words as features)
- Linear SVM (with bag-of-words and positional features)
- Simple neural network (multi-layer perceptron)
- BiLSTM

A simple linear SVM is chosen as a baseline Machine Learning classifier. The SVM is tested in two configurations explained below. In addition, two types of neural network architectures are selected: A simple feed-forward neural network (MLP) and a BiLSTM. Even though the MLP is not a typical sequence model, like RNNs, it is expected to "memorise" n-grams. The BiLSTM is chosen as representative of the state-of-the-art at the time of conducting this research and is supposed to be able to deal with longer dependencies in a given text sequence. Even more recent developments, such as transformers[9], were not considered.

Baseline performance To establish a baseline performance, two dummy classifiers are used: a random one and a stratified one. As the name suggests, the random dummy classifier votes fully randomly, resulting in a uniform distribution of assigned labels. The stratified baseline classifier[10] votes randomly but respects the training set's class distribution. That is if the class "None" represents 70% of all assigned class labels, then the baseline classifier would vote "None" randomly but 70% of the time. The stratified classifier is expected to outperform the fully random classifier.

Another naive dummy classifier would be one that always votes "none". Due to the class imbalance, this classifier would obviously achieve a very high recall and precision on the "none" class. It would even achieve a relatively high averaged F_1 score. But this classifier

[9] Recent examples are BERT (Bidirectional Embedding Representations from Transformers) or ULMFiT.

[10] Stratified dummy classifier provided by the scitkit-learn library (`https://scikit-learn.org/stable/modules/generated/sklearn.dummy.DummyClassifier.html`; last accessed: 2019-06-13)

would never be able to identify any buyer-supplier relations and completely miss the aim of this exercise. This dummy classifier is, thus, not further considered as a sensible baseline classifier.

(iii) Feature set

Feature set for the Deep Learning classifiers (MLP and BiLSTM) The *feature set* is identical for both Deep Learning classifiers and is based on the word embeddings obtained from the GloVe dataset. Instead of just using the complete set of pretrained word embedding vectors, two changes have to be performed:

1. *Filtering*: The dataset, initially consisting of millions or even billions of tokens, needs to be filtered down to only those tokens actually present in the training data. Embedding vectors beyond those of the training data do not add any value. The used GloVe embeddings, originally obtained by training on a corpus of 840B tokens of text from Common Crawl, are 300-dimensional vectors for a vocabulary of 2.2 million tokens. These embeddings were *filtered* by those tokens actually present in the training data.

2. *Vectors for masks*: Names of organisational entities need to be masked to allow a learning algorithm to generalise. Three types of masks are used: one mask each for the two organisational named entities in question, and one "other entity" mask for all other organisational named entities not in question but mentioned in the sentence. Each mask is assigned a separate embedding vector, e.g. the mask "__NE_FROM__" is assigned the 300-dimensional vector $\begin{pmatrix} 1 & \dots & 1 & 1 & 1 & 0 \end{pmatrix}$. The other masks are assigned the vectors $\begin{pmatrix} 1 & \dots & 1 & 1 & 0 & 1 \end{pmatrix}$ and $\begin{pmatrix} 1 & \dots & 1 & 0 & 1 & 1 \end{pmatrix}$, respectively.

Feature engineering for the linear SVM classifier For the simpler linear SVM model, a different data representation had to be chosen which requires some feature engineering. Two alternative data representations are tested:

1. In one configuration (A), the SVM does not consider word order nor does it consider positional information about the organisational named entities. The feature vector is a one-hot-vector based on a simple bag-of-words-approach.

2. In another configuration (B), the one-hot-vector is extended by two positional arguments for each of the two organisations in question.

Configuration (A) uses a simple bag-of-words approach (Joachims, 1998), where the order of words is disregarded, and a one-hot-vector is used to represent a sentence. A one-hot-vector is a vector with a the length equal to the vocabulary size of the training dataset, a value of ‘1’ is assigned to to the index of the vector if a word appears in the given sentence, ‘0’ otherwise. The SVM does not consider word order nor does it consider positional information about the organisational named entities. Similar to the other algorithms, the SVM is *not* provided with the company names. The SVM classifier is trained on this representation using a *one-vs-rest* setup.

The configuration (B) extends the previous configuration by basic positional information and is illustrated by Figure 5.15. The one-hot-vector is extended by four additional positional values (see bottom right of the figure). For each of the two organisational named entities in question for any potential relation, the number of tokens before and the number of tokens after the organisational mention are added as components of the feature vector.

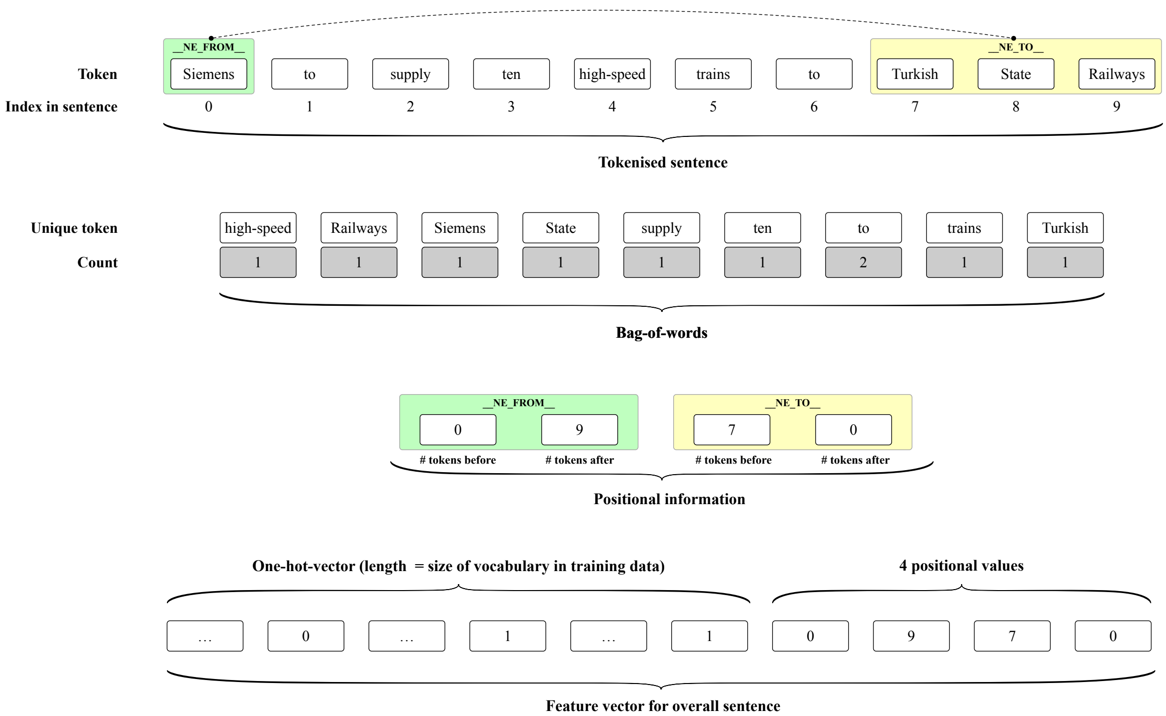


Fig. 5.15 Feature engineering for the SVM model

(iv) Metrics for evaluating classification performance

The commonly used performance metrics for binary classification are *recall* and *precision*. Generally, there is a trade-off between recall and precision. It is trivial to achieve perfect recall or perfect precision by either classifying all relations as positive or none at all. In order to optimise for a single variable, these metrics are combined in the F_1 *score* which is the harmonic mean of recall and precision. For multi-class classification problems, such as the problem at hand, the metrics for binary classification can still be used. The classification performance for each class is computed against all other classes ("one-vs-all"). The achieved classification performance for each class can then be averaged across all classes. For the averaging, there are two options: The “macro” average provides the mean performance across all classes and assumes equal weights for each class. Class imbalances can distort this macro average. The “micro” average is based on a count of all false positives, true positives etc. and generally represents a less biased performance metric.

(v) Process for training and testing

As is standard practice in Machine Learning, the text corpus is split into three partitions: for training, validation (such as tuning hyperparameters) and testing. A model is trained and tested multiple times and the results are averaged over all runs. As common for single-label multiclass classification problems, categorical cross-entropy is used as a loss function for the deep neural networks. Annotated data allows to assess classification performance automatically. Recall, precision, F_1 score and other metrics can be computed by comparing the classification results with the gold standard annotations.

5.5 Summary

This chapter addressed the problem of automatically extracting individual buyer-supplier relations between two organisational named entities. The problem was subdivided into corpus creation as well as classifier creation (i.e. the design, training and testing of one or more classifiers). Once trained, the classifier can be applied an arbitrary number of times to extract relations from previously unseen text.

Section 5.3 described the methodology for obtaining a labelled text corpus. The labelling consistency can be measured using inter- and intra-annotator agreement and serves as a proxy for the quality of the obtained labels and the feasibility of the labelling task. To measure the agreement, pairs of organisational mentions in a sentence need to be labelled redundantly. The definition of relationship classes represents a trade-off between the simplicity of the labelling task and the expressive power of the chosen set of labels. Three types of buyer-supplier relations are distinguished: “A supplies B”, “B supplies A” (inverse direction), and an additional class for all other direct buyer-supplier relations. The latter includes undirected or ambiguous ones as well as partnerships. Further three classes are defined for ownership/part-of relations, rejections, and “none” of the other classes. This set of classes is a substantial simplification and some types of relations or various nuances are not captured. Examples are stated *sub-tier* relations or direct buyer-supplier relations which *just ended* but existed before. For the text corpus, documents are sampled from English general news datasets. Using English as a language ensures a large availability of news sources as well as a variety of NLP tools and models to build upon. However, this also implies that a trained classifier would only work on English text – unless retrained on labelled data in other languages. The documents and sentences sampled for annotation have to be enriched with manually collected sentences that appear to describe a buyer-supplier relation. Otherwise, annotators would spend most of their time annotating negative examples. Only sentences with at least two auto-detected named entity mentions are presented to the annotators. Entity mentions in the text corpus are masked, i.e. substituted by a specific placeholder token, to force any classifier to generalise beyond specific organisation names. Once the corpus has been created, different classifiers can be trained and tested on that data, as described in Section 5.4. A supervised learning approach is chosen to avoid the problem of semantic drift (inherent to semi-supervised approaches) and to be able to extract directed relations (unlike unsupervised approaches). The downside of a supervised approach, however, is that it requires a labelled and representative training dataset of sufficient size. The classification models chosen to be used for this research are two Linear SVMs using different features, a simple MLP, and a BiLSTM deep neural network. Two dummy classifiers, uniform random and stratified random, are also selected whose performance will serve as a baseline to benchmark the other classifiers. Pre-trained word embeddings are used as features. The three types of masks used to hide specific organisational mentions from the classifier are also assigned an embedding vector each.

Following the methodology outlined in this chapter, buyer-supplier relations between two identified organisational entities can now be automatically classified. This approach is validated in Chapter 7.

The next chapter shall first bridge the gap between this chapter and an end-to-end approach for automating supply chain mapping from text. In particular, it needs to be discussed which processing steps have to occur before the relation classification can be applied and which steps have to occur to convert a set of extracted relation triples into useful supply chain maps.

Chapter 6

Challenges in automating the end-to-end supply chain mapping process

6.1 Introduction

The extraction of buyer-supplier relations examined in the previous Chapter 5 is a fundamental building block of any approach that aims to automatically generate supply chain maps from natural language text. By applying the methodology outlined in the previous chapter, buyer-supplier relations between two detected organisational named entities can be classified. Figure 6.1 illustrates the focus so far in the context of the conceptual model that has been used throughout this thesis.

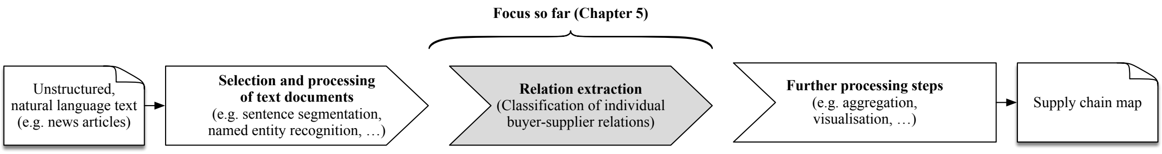


Fig. 6.1 Conceptual model: Focus so far

This chapter now widens the scope again to address the overall supply chain mapping problem. It aims to analyse the gaps between extracting individual buyer-supplier relations and an end-to-end approach to automating supply chain mapping from text. Before an end-to-end approach can be discussed, two major gaps have to be addressed:

- The "pre-processing" of documents prior to the relation extraction phase. In the Corpus Creation stage, these steps were largely performed manually and still need to be automated.

- The "post-processing" of relation triples so that they can be presented in form of a supply chain map.

Figure 6.2 illustrates the focus and structure of this chapter. The terms "pre-processing" and "post-processing" are written in quotes since they are just used in the context of this thesis for the steps before and after relation extraction.

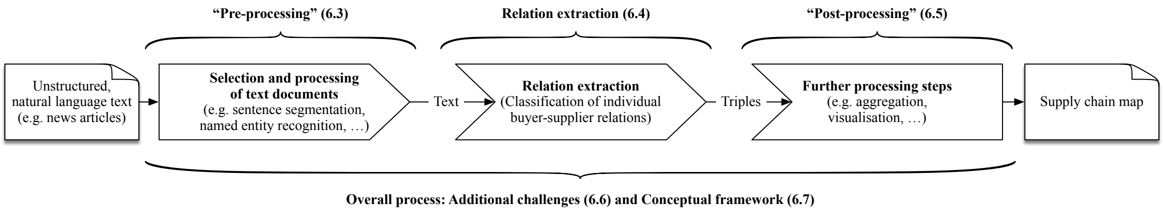


Fig. 6.2 Conceptual model: Focus and structure of this chapter

The structure of this chapter is as follows: In Section 6.2, the methodology for this chapter is discussed. The subsequent section outlines the tasks that are required to convert a set of relation triples – the output of applying the methodology of Chapter 5 – into a visual map. Section 6.3 outlines the necessary "pre-processing" steps. In Section 6.4, the process to convert text into relation triples is summarised. Section 6.5 outlines the "post-processing" steps to convert a set of triples into a supply chain map. In Section 6.6, additional challenges and requirements for an end-to-end approach of automating supply chain mapping from text are identified. These requirements are derived from the characteristics of supply chain maps or are a result of limited information quality. Selected challenges are discussed in more detail in a further section. For these selected challenges, possible solutions are outlined. Section 6.7 attempts to summarise the requirements and challenges in an overall framework.

6.2 Methodology

This chapter aims to identify the most immediate challenges one encounters when trying to build upon the work presented in the last chapter to automate the end-to-end process of supply chain mapping from text. These challenges shall be addressed by a conceptual framework. The necessary steps for "pre-processing" text documents as preparation for the relation extraction step are derived based on the findings of Chapter 5. Similarly, required "post-processing" steps are derived to convert relation triples into a simple visual map. Additional challenges for the automated supply chain mapping from text are identified. For each challenge, an explanation, rationale and additional evidence is provided, such as references to literature or concrete examples from the data demonstrating the discussed need. Lastly, a conceptual framework for an end-to-end process to automate the generation of supply chain is derived. Ultimately, automating supply chain mapping from text could be as complex and challenging as solving the problem of language understanding and human research. Thus, not all identified challenges can be addressed within the scope of this research. The key propositions from this chapter are validated in Chapter 7 where the end-to-end approach is practically applied.

6.3 “Pre-processing” of text documents

A number of steps need to be performed to provide the relation classifier with the expected input. To automate the end-to-end process, these steps, too, have to be performed automatically. Some of the pre-processing steps are required, such as the detection and masking of organisational named entities in a sentence. Other steps improve expected performance (e.g. execution speed or the efficiency of the annotation process). This includes discarding text that can safely be assumed to be irrelevant, such as documents in other languages or script elements in a Web document. Automatically removing as much irrelevant text as possible is especially crucial as a step prior to the labelling by human annotators. This way their valuable time is spent most usefully. Co-reference resolution and named entity resolution could be further beneficial pre-processing steps but were not conducted as part of this research.

6.3.1 Basic “pre-processing” steps

The classifier expects a single sentence where all organisational named entities have already been identified and appropriately masked. Figure 6.3 shows a minimal processing pipeline for preparing text documents.

If the document collection contains Web documents, such as HTML files, these may have to be cleansed to remove HTML tags and other script elements and extract the pure text content from heading and text body. Some of the documents’ meta-data may be relevant for subsequent tasks and it may be beneficial to extract and store this information separately before stripping the Web document of all HTML tags. It may be beneficial to remove (or mask) stock ticker symbols so that these cannot be accidentally be “memorised” by the classification algorithm. Because NLP tasks are language-specific, it is beneficial to detect the language of the document. This way documents in irrelevant languages can be filtered out as early as possible. The pure text content can then be split into individual sentences in a step called sentence segmentation or sentence tokenisation. Each sentence is then split into individual tokens. Depending on the quality of the documents, it may be useful to remove

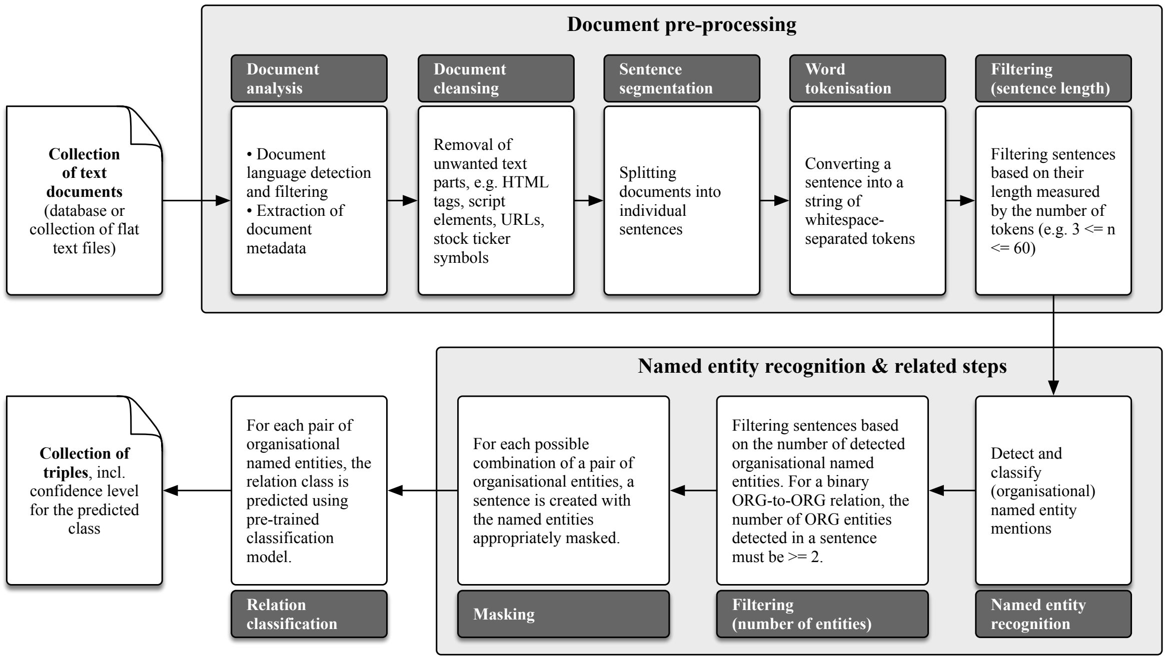


Fig. 6.3 A minimal "pre-processing" pipeline for text documents

sentences above or below a certain length (measured in number of tokens). In a subsequent step, named entities can be automatically detected and classified using an off-the-shelf named entity recognition system. Only sentences with at least two named organisational entities need to be processed. Sentences with fewer can be discarded named organisational entities because no co-reference resolution is applied. During the masking step, for each possible pair of two organisational entities, a masked sentence version is created. This can then be fed into the classification algorithm to detect the relation between these two entities. The output of the classifier is a collection of relation triples.

6.3.2 Beneficial additional steps and potential for improvement

The described process is limited in various ways to maintain a manageable scope of the research project. These limitations represent potential for future improvements of the process. Potential extensions include:

Co-reference resolution Adding co-reference resolution for pronouns but also nouns, such as "the carmaker", would increase the number of relations that can be detected. The addition of co-reference resolution would come at the cost of introducing a further source of errors.

The co-reference resolution would be applied *before* sentences are filtered based on the number of named organisational entities.

Named entity resolution Named entity resolution at an early stage of the processing pipeline may prevent errors at later stages, such as the misclassification of relations. A combined and context-aware named entity recognition and resolution step could ensure that classifier is only presented with masked organisations rather than, for instance, accidentally masked product names. Ideally, the named entity resolution system considers the whole context of the document and is able to make use of stock ticker symbols.

6.4 Buyer-supplier relation extraction

This step refers to the classification of the relationship for each pair of organisational entity mentions in a sentence. The necessary steps to obtain a pre-trained classifier capable of performing this tasks have already been described in Chapter 5. Following that methodology, text is first annotated by humans to obtain a labelled text corpus. This text corpus is then used to train and test the classifier. Once trained, it can be applied to single, previously unseen sentences where all organisational entities have been masked. Given the sentence, the classifier is able to return a confidence score for each type of relationship.

Whereas obtaining the text corpus, designing and training the classifiers (as described in Chapter 5) is comparatively complex and laborious, the mere application of a pre-trained classifier to pre-processed text is not.

6.5 "Post-processing" of predicted buyer-supplier relations

Additional steps are required to convert the set of relation triples into a visual map. The resulting supply chain map is still simple but should already provide some limited utility. The post-processing steps include the *aggregation* of relation triples and the subsequent *visualisation*. These steps presented below are the bare minimum. Sections 6.6 and 6.7 show that further post-processing steps may be necessary based on the supply chain maps' intended use. E.g. it may be a useful addition to compute confidence scores for the extracted information, e.g. by leveraging how often a relation was reported by different but reputable sources and using different sentences structures.

6.5.1 Basic aggregation

This step refers to the aggregation of a collection of individual relations into a data structure that can be visualised in form of a map. Even a simple visualisation of a supply chain will typically aggregate the data. The following steps achieve a minimal level of aggregation by the deduplication of entity mentions and relation occurrences.

Collapsing identical entity mentions By feeding in the set of triples into visualisation tools, such as D3.js or Cytoscape, entity mentions with identical names will be collapsed into one, as shown in Figure 6.4. This step is an implicit, naive form of entity disambiguation where entity mentions with identical names are assumed to refer to the identical entity, and entity mentions with different names are assumed to refer to different entities. For instance, "Toyota Motor Corporation" and "Toyota" would be considered two separate entities.

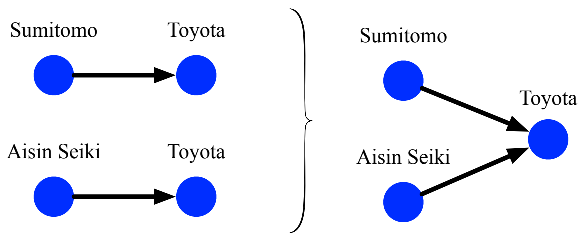


Fig. 6.4 Basic aggregation: Collapsing nodes

Collapsing repeated relation occurrences A further aggregation step is to treat multiple occurrences of the same relation in different sentences or documents as an attribute of the relation. This attribute is visualised not as separate links but, for instance, as the width or colour of a link. This is illustrated in Figure 6.5.

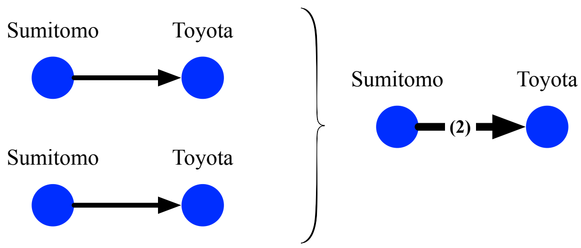


Fig. 6.5 Basic aggregation: Collapsing relation occurrences

6.5.2 Basic visualisation

At the very least, a supply network visualisation consists of nodes, representing the organisations, and links between them, representing the buyer-supplier relations. Since multiple classes have been defined, it appears sensible to distinguish these classes visually, e.g. with different colours or line types for different relations. Directional relations may also be visualised using arrows. Furthermore, the confidence score of the relation classification as well as repeated mentions of the same relation can be visually indicated. Further visualisation options require the extraction of further information, e.g. the spatial distances between nodes could represent geographical distances, if the organisations' locations are known, or other features, such as the intensity of the collaboration.

Example

Figure 6.6 shows a simple, automated visualisation of automatically extracted buyer-supplier relations in form of a basic supply chain map. This particular visualisation was implemented in d3.js. Different relation classes are represented by different line types. Directional relations may be visualised using arrows pointing in the direction of the material flow. To demonstrate that a basic map can be automatically generated, the following hand-picked but authentic sentences were processed:

> “ASCO manufactures and supplies Toyota with these water pumps.” “ASCO, manufacturer of high lift device mechanisms, complex mechanical assemblies and major functional components, signed a long term contract with Airbus for the production of hybrid complex frames.” “Denso supplies Toyota with approximately half of its components.” “GKN Aerospace has been awarded a contract by Airbus.” “Velocity Composites has signed a new contract that will see it supply aerospace manufacturer GKN Aerospace with structural plies for the next five years.” “Japanese car brands Toyota and Suzuki have announced wide-ranging global collaboration plans.”

Relations that were identified as directed ones are indicated as such in the map. The arrow head indicates the detected material flow. The classifier interpreted the relation between Toyota and Suzuki as non-directional / partnership, as indicated by the lacking arrow head for this link. The size of this knowledge graph can become arbitrarily large by processing additional news data. It is obvious that supply chain maps generated solely based on simple company-to-company relations are limited. For instance, the visualisation appears to suggest that Velocity Composites is a sub-tier supplier of Airbus. However, the provided text example alone does not provide sufficient evidence for this inference. One way to address this “transitivity problem” is to also extract the end-product for which a part is intended if this fact is mentioned in the context. The visualisation can also be further enriched. E.g. the confidence in each relation classification could be indicated.

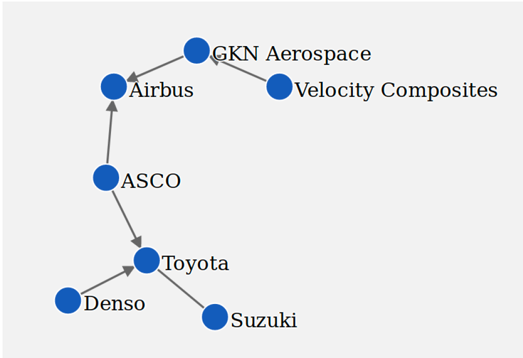


Fig. 6.6 Basic supply chain map automatically generated from text

6.6 Additional challenges for automated supply chain mapping from text

The focus so far has only been on classifying relations between exactly two organisational named entities. Some information that is relevant for supply chain mapping cannot be obtained using this approach. The gap analysis below is structured in correspondence to the following questions: How could the scope of the relation extraction be increased beyond what was considered in Chapter 5? Which additional challenges need to be considered that may arise (a) from characteristics of supply chains or (b) from limited information quality?

6.6.1 Challenges of relation extraction scope

So far, the focus was on answering the question "who supplies whom?" and directed buyer-supplier relationships between organisations are by definition the minimal information required for describing a supply network (nodes and links). The question "who supplies whom *with what*?" already requires a *ternary* relation between two organisational named entities and, for instance, a branded product (named entity) or a description of a product or service. Regarding just the extraction of buyer-supplier relations, at least the following dimensions can be used to increase the scope further: the *entity types* considered in a relation (e.g. relations between an organisation and a product), the *number of entities considered in a single relation* (e.g. ternary relations, such as "who supplies whom with what?"), the *number and type of attributes* captured for a relation (e.g. indicated uncertainty or time), the *relation types* (e.g. "competes-with" or "has-suppliers-located-in"). In the attempt to capture all relevant information, the scope of the relation extraction can become infinitely complex. A discussion of options to increase the scope can be found in the Appendix in Section B.2.

6.6.2 Challenges of process automation

Table 6.1 provides an overview of challenges. These challenges were arrived at based on insights from the literature review, the industry review and/or the research process. Subsequently, each challenge is discussed in a separate paragraph which also provides sources, collected evidence or examples.

Table 6.1 Additional challenges for an end-to-end approach of extracting supply chain maps from text

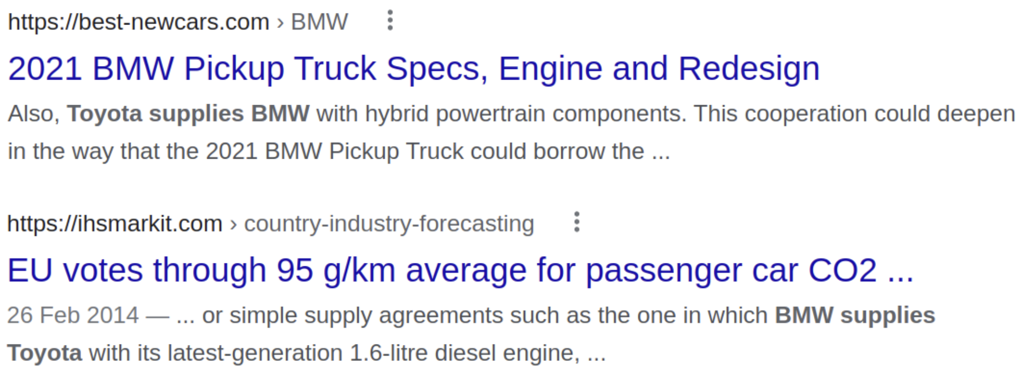
(a) Challenges resulting from supply chain characteristics	• Extracting geographical information (e.g. location of production facilities) • Extracting supplied part, material or service • Disambiguating different roles of the same company to solve seeming contradictions • Extracting aspects of time to capture supply chain dynamics • Extracting information from text in various languages • Dealing with countries as actors within supply chains • Separating suppliers of primary and support functions, such as financial services • Identifying logistics providers as special type of supplier
(b) Challenges resulting from limited information quality	• Coping with incomplete information / limited data availability • Coping with imperfect (e.g. inaccurate or contradictory) information • Coping with an abundance of positive information and lack of negative information (a company does not or no longer supply another company) • Incorporating that the frequency of occurrence does not necessarily indicate the criticality of a buyer-supplier relation • Learning company- or industry-specific models, e.g. due to specific language • Extracting information on higher aggregation levels than individual buyer-supplier relations • Indirectly inferring buyer-supplier relations from other types of relations • Establishing the correct tier of a company in a supply chain • Disambiguating entity mentions using the context • Coping with ambiguous language by inferring buyer-supplier dependencies using evidence from multiple sources • Coping with named entities within named entities (product name containing company name)

(a) Challenges resulting from supply chain characteristics

An ideal solution for the overall process of automating supply chain mapping from text would need to address the following challenges resulting from the characteristics of supply chains.

Extracting geographical information: Risk sources often have a geographical context. Extracting where a company is located or where a part is produced is important for a number of supply chain mapping use cases. The geolocation of a company's headquarters as a first approximation allows for a rough geographical vulnerability assessment. The importance of geolocations is mentioned in the literature (e.g. see Figure 2.3 adapted from Berbner (2017, p.66) or Sheffi's definition of supply chain maps in Section 2.3.4). The importance of geolocations was also mentioned in the exploratory interviews, e.g. in the context of warning suppliers of impending dangers, such as hurricanes. The *geographical* supplier concentration also played a crucial role in the 2011 Thailand floods in Section 4.2.3.

Extracting supplied part, material or service: It is possible that two companies supply each other with different parts or services. To separate these relations, the supplied parts or services need to be extracted as well. Consider the example shown in Figure 6.7. One article states that Toyota supplies BMW with hybrid powertrain components. The other article states a buyer-supplier relation in the opposite direction: BMW supplies Toyota with diesel engines. Thus, the directionality of the supplier relation can be different for different parts or services.


https://best-newcars.com › BMW

2021 BMW Pickup Truck Specs, Engine and Redesign

Also, **Toyota supplies BMW** with hybrid powertrain components. This cooperation could deepen in the way that the 2021 BMW Pickup Truck could borrow the ...

https://ihsmarkit.com › country-industry-forecasting

EU votes through 95 g/km average for passenger car CO2 ...

26 Feb 2014 — ... or simple supply agreements such as the one in which **BMW supplies Toyota** with its latest-generation 1.6-litre diesel engine, ...

Fig. 6.7 Example of two companies supplying each other

Disambiguating different roles of the same company: The same company can assume multiple different roles in the same supply network (Brintrup, 2010), making the learning task difficult because the information seems to be contradictory. A company can, for example, deliver one component as a first tier supplier and another part as a second tier supplier (Figure 6.8).

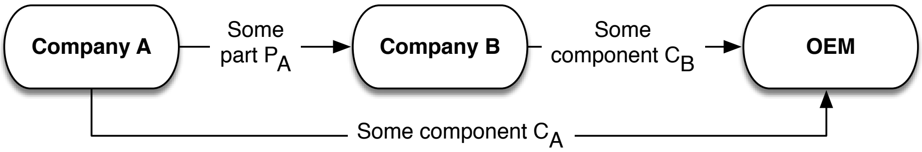


Fig. 6.8 Company A with multiple roles in the same supply network.

Extracting aspects of time to capture supply chain dynamics: Supply chain structures change over time, e.g. in- or outsourcing decisions may change length and width of the supply chain (Lambert and Cooper, 2000). As extracted information can be outdated, it ideally needs to be assigned with a time stamp. Supply chain structures, or parts thereof, may be more static (e.g. in the case of parts suppliers in aerospace programmes for the duration of the programme) or highly dynamic (e.g. in the case of switching to a different logistics supplier to expedite shipping). Furthermore, if a time period of validity is explicitly stated, such as the duration of a contract or relationship, this information should also be captured. Similarly, statements of changes in a relationship should be captured, such as a company dropping a supplier or being added as a supplier. An example of such an event is provided by Figure 6.9. Capturing aspects of time as well as changes in the relationship does not only help with visualising an up-to-date representation of a supply chain, it would also allow to detect and visualise trends over time, such as an increasing consolidation of the supplier base.

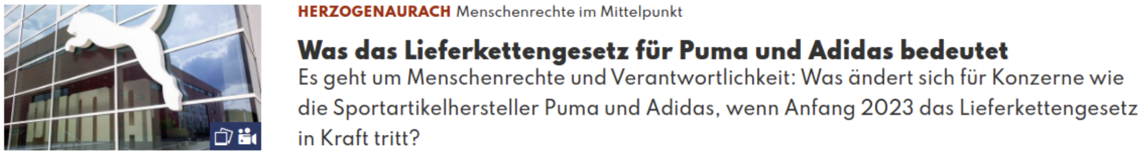
VW cancels contracts with Prevent Group after supply problems

By Marcus Williams | 10 April 2018

Volkswagen has cancelled the majority of its contracts with parts and material supplier Prevent Group following a series of supply problems affecting production in Brazil between 2015 and 2016. The carmaker said the problems had cost it a total of €250m ($307m).

Fig. 6.9 Example of a company ending a buyer-supplier relation; screenshot from a 2018 article in Automotive Logistics

Extracting information from text in various languages: Supply networks tend to be global (Christopher and Peck, 2004). Hence, they span many countries and language areas and, to capture as much of the supply network as possible, any approach should ideally consider multiple languages. Local newspapers located near the headquarters or main facilities of companies can have a particularly good access to information to address the particularly high level of interest of the local population. E.g. The Seattle Times was found to frequently report about Boeing and the aerospace industry. Similarly, some news about Puma and Adidas, both located in the small city of Herzogenaurach in Bavaria, may be best covered by a regional German newspaper, as shown by Figure 6.10.

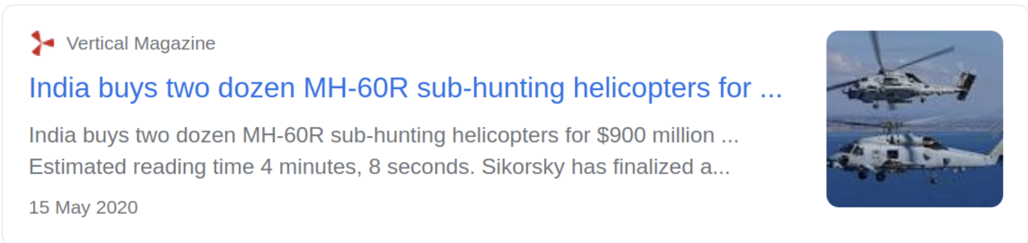

HERZOGENAURACH Menschenrechte im Mittelpunkt

Was das Lieferkettengesetz für Puma und Adidas bedeutet

Es geht um Menschenrechte und Verantwortlichkeit: Was ändert sich für Konzerne wie die Sportartikelhersteller Puma und Adidas, wenn Anfang 2023 das Lieferkettengesetz in Kraft tritt?

Fig. 6.10 Regional German newspaper discussing the impact of supply chain regulation on Puma and Adidas; source: nordbayern.de

Dealing with countries as actors within supply chains: Especially in the defense industry, sellers and buyers can be countries, as shown by Figure 6.11. NER systems may not detect these as organisations but as geographical entities. In experiments conducted during this research project, some information did not get extracted due to this reason.


Vertical Magazine

India buys two dozen MH-60R sub-hunting helicopters for ...

India buys two dozen MH-60R sub-hunting helicopters for $900 million ... Estimated reading time 4 minutes, 8 seconds. Sikorsky has finalized a...

15 May 2020

Fig. 6.11 Example for a country acting as a buyer; screenshot from Google News

Separating suppliers of primary and support functions A supply chain map is typically expected to show companies that directly contribute to a product or service. However, some purchased products or services represent support functions. These can be hard to distinguish from primary functions, especially for an automatic classifier. Support functions include legal services, consulting services, HR services, financial services, or, as shown by Figure 6.12, the fire detection system for a warehouse. Without background knowledge, support functions and primary functions can be hard to distinguish.

> Fike Safety Technology provides fire detection system to protect Unilever's new warehouse facility in Nigeria.

Fig. 6.12 Example of a support function

Identifying logistics providers as special type of supplier Within the context of this research, logistics providers are considered part of the supply chain. However, logistics providers may change more abruptly than, for example, a supplier of parts. If the delivery process needs to be expedited, a quick decision may be made to transport goods by plane rather than by ship. In addition, information about the logistics provider may be less abundantly available than information about suppliers of goods or processing services. The exploratory interview with the supply chain managers of an aerospace manufacturer presented in Section 4.2.4 revealed that logistics providers had been the cause for supply chain disruptions and should also be considered in supply chain maps. The 2021 Suez Canal obstruction by the container ship “Ever Given” is just one of the more drastic examples.

(b) Challenges resulting from limited information quality

Further characteristics can be derived from limitations of the information quality. This includes the challenges inherent in natural language. The following aspects are not meant to be exhaustive and are not always clearly separable.

Coping with incomplete information or limited data availability: The main reason for supply chain data unavailability is the “proprietary nature of each supplier’s relationships with its partners” (Sheffi, 2005). It is, therefore, evident that not all desired information can be expected to be reported on the Web. Among other consequences, an overall approach to automated supply chain mapping from text needs to consider that not all suppliers on all tiers for every company and every industry will be identifiable. Different levels of data availability are to be expected for different types of companies and industries.

Coping with imperfect (e.g. inaccurate or contradictory) information: Sources of information, especially on the Web, can be unreliable and information may be contradictory. So far, the approach proposed in Chapter 5 assumes that sentences contain true statements. To address the issue of veracity, additional processing steps are required. A confidence score for the information source could be maintained so that reputable sources, such news agencies or the official press release websites of established companies, are generally considered more trustworthy than social media posts. If the information source is stored together with the extracted relations, the confidence score for the relation can then be combined with the confidence score for the information source in general. Furthermore, information can be cross-checked via inter-source agreement. If different sources agree, ideally in text that is worded differently and not just a direct copy, then a higher confidence can be placed in the correctness of the statement. However, official information can be expected to contain information that is already generally available. Unofficial, by default less credible, accounts should not be fully neglected. But it is crucial to highlight these accounts as such.

Coping with an abundance of positive information and lack of negative information : News reports typically contain a 'positive' information bias in the sense that they state that one company supplies the other. News reports are less likely to state that a company does *not* or *no* longer supply the other. Especially press releases will report on contract wins rather than losses, as became clear during the early stages of this research. Extracted information might have to be assigned an industry- or part-specific half-life to reduce confidence over time.

Incorporating that the frequency of occurrence does not necessarily indicate the criticality: In some NLP tasks, the frequency of a relation may be used to derive how important that relation is. Similarly, if it is frequently reported that company A supplies company B, we might assume that company A is a major supplier. However, from a risk management perspective, a buyer-supplier relation can be critical even though it is not mentioned frequently and does not correspond to a large sales volume (Simchi-Levi et al., 2014; Yan et al., 2015).

Learning company- or industry-specific models: Some of the aforementioned characteristics, such as supply chain dynamics and data availability, are company- or industry-specific. This makes it difficult to predict how well a supply chain can be mapped since the information extraction performance and information quality are affected by the type of company or industry. Furthermore, the language used to describe buyer-supplier relations can be industry-specific. E.g. a chemical may be "synthesized" as opposed to "produced". Different industries also have very different lengths of supply chains: a fresh produce grocery supply chain tends to be shorter (in terms of the number of tiers) than the aerospace supply chain.

Extracting information on higher aggregation levels: Relevant information about supply chains might be provided on different aggregation levels, as illustrated by Figure 6.13. For instance, an individual buyer-supplier relation may be referenced as "company A supplies company B with part C". But important information might also be provided on a higher aggregation level, such as “50% of company A’s suppliers are located in the Los Angeles area” or “Product P consists of 100,000 parts sourced from over 1,500 suppliers in Japan, Germany, and China”. Other aggregation levels include flows between production sites, legal entities, or company groups.

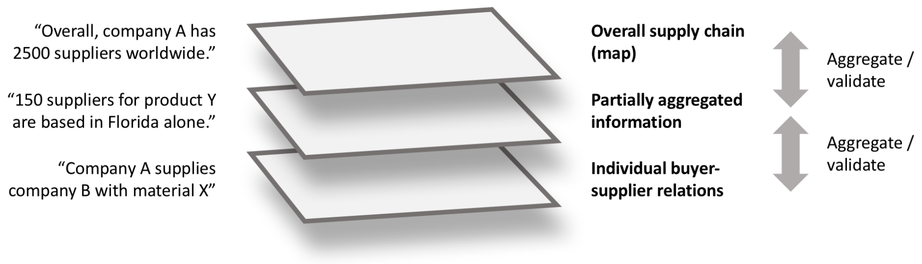


Fig. 6.13 Information on different aggregation levels.

Indirectly inferring buyer-supplier relations from other types of relations: Buyer-supplier relations can be expressed indirectly which makes any automated extraction difficult. Consider the two statements that may be taken from two different documents: (1) “Company A is the only producer of material M in the world.’ (2) “Company B uses material M.”. Both statements together would imply that Company A supplies Company B at least indirectly. Seemingly irrelevant individual statements can thus become important to collectively infer relevant relations.

Establishing the correct tier of a company in a supply chain (incl. transitivity problem) Information may be insufficient or insufficiently unambiguous to correctly establish the tier a company is operating on for a particular supply chain. There are at least the following aspects to this problem.

1. The information that allows for identifying the (sub-)tier is provided but cannot be captured by the proposed approach without adjustment

2. The information required to determine the presence of a direct or indirect relation is lacking or ambiguous

3. Inferring sub-tier relations from binary ORG-to-ORG relations is problematic (transitivity problem)

The latter aspect (transitivity problem) shall be discussed in more detail: The approach proposed in Chapter 5 is based on extracting binary buyer-supplier relations. This introduces the problem of establishing the mathematical property of *transitivity* when reconstructing a supply network: If one can establish that company A supplies company B and also that company B supplies company C, does that imply that company A is a sub-tier supplier of company C? Unfortunately, this inference cannot always be drawn (Figure 6.14). This issue shall be referred to as *transitivity problem*.

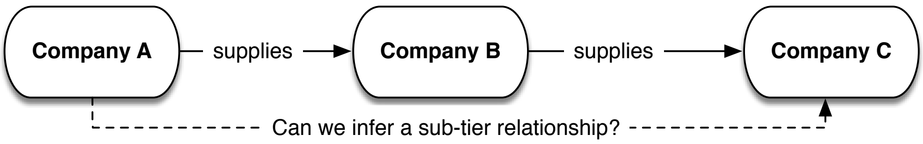


Fig. 6.14 Illustration of the "transitivity problem"

More formally, the approach proposed in Chapter 5 will result in a set of relationship triples aRb for companies a,b and a buyer-supplier relation R. This set of triples can be visualised in a graph, a basic supply chain map. However, this map will not necessarily have the desirable property of *transitivity*, i.e. $\forall a,b,c \in X : (aRb \wedge bRc) \Rightarrow aRc$ (for companies a,b,c and buyer-supplier relation R). Instead, the resulting supply chain map will only show a *myopic* (one-tier-up or one-tier-down) view of any company's (assumed to be) direct suppliers and customers. Thus, it will only show *likely sub-tier suppliers* but not confirmed ones. This becomes most obvious and unsatisfactory in the case of highly diversified conglomerates where only a smaller subset of all direct suppliers are relevant for any given product. Still, the extracted information may already be useful since it would allow to identify a set of likely or at least potential sub-tier suppliers.

To extract sub-tier relations, the approach proposed in Chapter 5 can be extended to capture end-products, to extract ternary relations, or to extract information about tiers if explicitly mentioned. However, the more tiers need to be described, the less likely is it to find a complete description in the text sources.

Disambiguating entity mentions using the context: Mentions of organisational named entities need to be disambiguated and linked to the unique real-world entity. This is a challenging task for a number of reasons.

Brand names and names of the legal entity are often used interchangeably. For example, it is not immediately clear if the sentence "*Interstate supplies Mercedes with batteries in the US*" really implies a relation between the legal entities "*Interstate Batteries*" and "*Mercedes-Benz USA, LLC*", or "*Daimler AG*". Similarly, brand name "Blackberry" and its manufacturer – then called "Research In Motion" – are used synonymously in news articles.

In fact, "Research In Motion" is often just referred to as "RIM". Moreover, NER tagging systems will often confuse brand names for products and companies with the names of actual organisations, especially if these brands *are* operated as separate units within an organisation. A similar problem exists for last names and company names.

A further challenge are industry-specific abbreviations that can easily be confused with company names. For instance, "LPG" could be a company name but could also stand for liquefied petroleum gas. Especially, abbreviations used in the context of contracts can be important, such as "LTA" (long-term agreement) or "MOU" (memorandum of understanding).

Companies, like Toyota or Boeing, tend to consist of a network of hundreds of legal entities. Legal entities can be identified using their full name or their Legal Entity Identifier (LEI). However, in news articles, companies are referred to by their common name. It is not immediately clear which exact legal entity entered into a contract with which other exact legal entity. Articles written by professional business information providers, such as Bloomberg or Reuters, often mention the stock ticker symbol or other custom company identifiers after the company name. These identifiers can be used for entity linking.

Coping with ambiguous language by inferring buyer-supplier dependencies using evidence from multiple sources: A "deal" can be a supply contract but can also refer to an acquisition or other kinds of agreements. A different example of ambiguous language is depicted by Figure 6.15. What looks like active voice in the heading is actually passive voice. The words "has been" were omitted for brevity in the title compared to the first sentence of the text body. And, thus, the title can easily be misinterpreted as stating the opposite directionality. The linguistic clues that help distinguish these cases are often subtle and tend to require context or even background knowledge not provided by the context.

Tetra Tech awarded $50M U.S. Navy biological resources contract

Tetra Tech has been awarded a $50 million contract to perform biological resources management services for the U.S. Navy. The work will be performed primarily at locations within the Naval Facilities Engineering Command (NAVFAC) Atlantic's area of responsibility, which includes the continental United States, Europe, southwest Asia, and the Caribbean. This single-award, indefinite delivery/indefinite quantity contract has one base year and four one-year options...

Fig. 6.15 Language used in headings

In the aerospace industry, language can be particularly ambiguous due to the concepts of seller-furnished equipment (SFE) and buyer-furnished equipment (BFE). BFE is equipment purchased by the buyer of the aircraft and given to the aircraft manufacturer to be installed before delivery by the manufacturer. Given a news report, this makes it difficult to understand which party is paying for the parts.

Coping with named entities within named entities: It can occur that a named entity, such as a product, contains another entity, such as the name of the organisation (e.g. "Boeing 737" or "Toyota Corolla"). In those cases, NER tagging systems may not be able to provide overlapping segmentation boundaries. As a consequence, the whole text sequence may be classified as product and the entity is not considered for classifying relations between organisations.

6.7 A conceptual framework for an end-to-end approach

In this section, a conceptual framework is proposed that provides a structure for thinking about an end-to-end approach for automating the generation of supply chain maps from unstructured text. This framework shall *not* be understood as a rigorous academic model and not all tasks could be implemented within the scope of this research (e.g. extraction of company-to-product relationships or the inference of sub-tier relations). Due to scenario-specific circumstances, for example, additional tasks may also need to be added.

Figure 6.16 depicts the conceptual framework. The framework is structured using stacked abstraction layers – starting with document retrieval / collection and ending with the visualisation & decision support. For each abstraction layer, example tasks are listed. These address challenges previously identified in this chapter. The framework also indicates the artefacts that are generated on one abstraction layer and passed onto the next one. Figure 6.16 corresponds to Figure 6.2 which showed the same process on a higher abstraction level earlier.

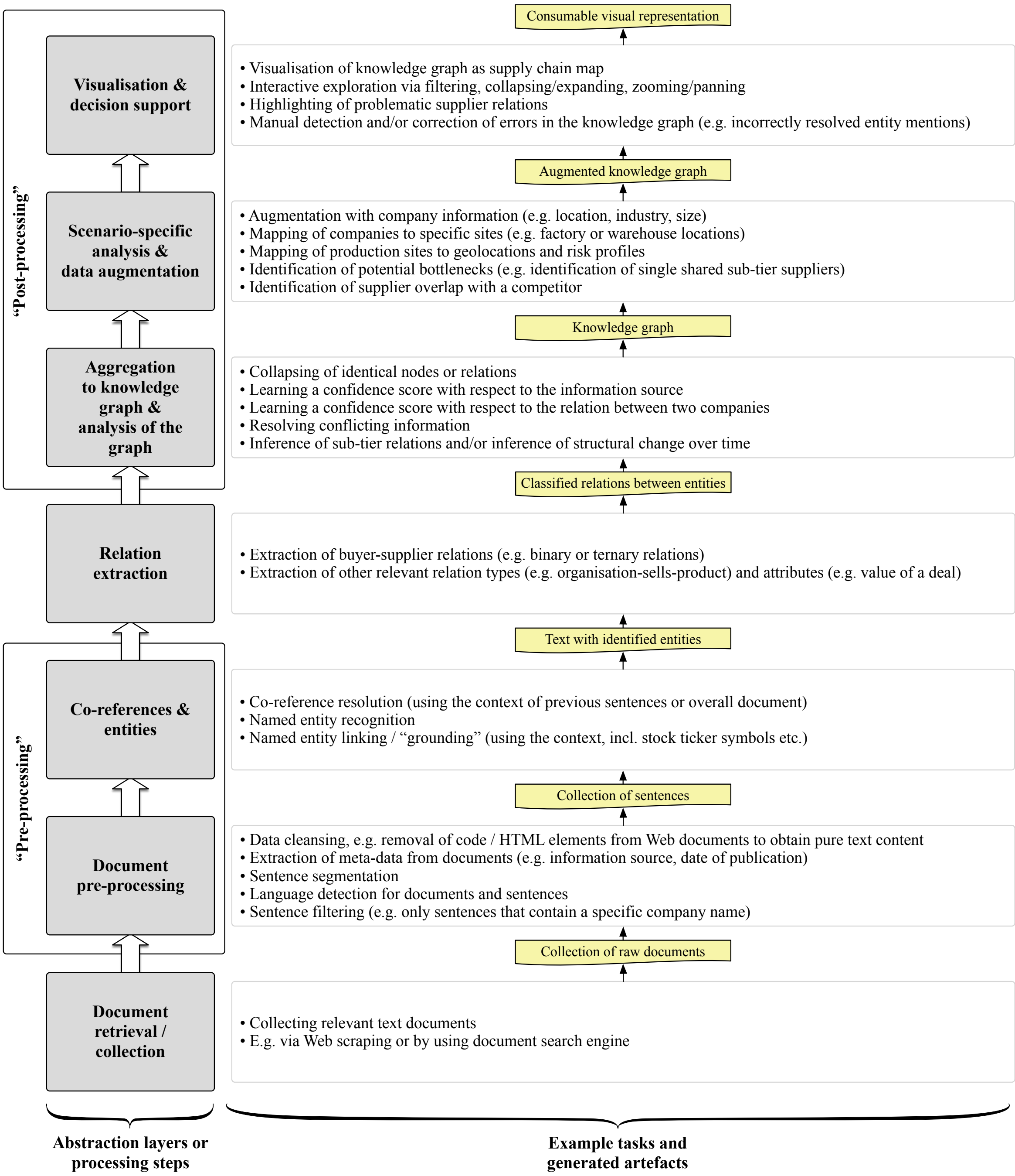


Fig. 6.16 Conceptual model for extracting supply chain maps from text

Document retrieval / collection Prerequisite for the overall process is a collection of documents. These may already be existent or have to be collected via Web scraping or by using a document search engine. Due to the complex reasoning in subsequent layers, it is likely that the overall process will not be run on-demand but that a knowledge graph will be created based on a large collection of documents. Only the querying of that knowledge graph would then occur on-demand to satisfy a specific information need. If applicable, documents can be filtered for the existence of company names.

Document pre-processing Documents are pre-processed as described in Section 6.3.1.

Co-references & entities The purpose of this layer is to detect, classify (NER) and disambiguate (entity linking) named entities as well as to resolve co-references. This does not only include organisational named entities but may also include names of products or locations. Once entities have been fully identified, the relations between them can be classified. The linking of entities can be done using tools that are based on Wikipedia or Yago data, such as Ambiverse or Aida Yago. A further option is to use the recently introduced international legal entity identifiers (LEI)[1] which are freely accessible online. The datasets also provide information about the "ultimate parent, defined as the highest level legal entity preparing consolidated financial statements, as well as the direct accounting consolidating parent".

Relation extraction The purpose of this layer is to classify the relations between all the entities detected in the previous stage. This includes binary buyer-supplier relations between organisational named entities. But this may also encompass other types of relations relevant to the problem at hand, such as "ORGANISATION-sells-PRODUCT" relations.

Aggregation to knowledge graph & analysis of the graph The set of individual relations can then be aggregated, e.g. by appropriately aggregating multiple mentions of the same relation. By cross-referencing statements across different information sources, a confidence score for each information source may be generated and continuously updated. The same can be done for the confidence in a particular relation, such as "Tesla is a Toyota supplier", by asking questions about the agreement across different pieces of evidence as well as the recency of such statements. A confidence decay function could ensure that older information is considered to be less reliable than information that was published more recently. By cross-referencing different statements, one may want to attempt to further clarify the tier on which a company is providing goods or services to another company. Statements that seemed to suggest a direct relationship (e.g. "company B used material from company A") can then be overruled by more specific statements (e.g. "company A is a second-tier supplier

[1]GLEIF; `https://www.gleif.org/en/`; last accessed: 2019-07-10

to company B”). This way conflicting information can be reconciled to a single view on the world.

Scenario-specific analysis & data augmentation So far, the layers have been largely scenario-agnostic and simply aimed at creating a knowledge graph of buyer-supplier and other relevant relations. Depending on the use case, the knowledge graph can be augmented with external (structured) data. Once organisational named entities have been grounded, additional information about the company, such as size, location, revenue or industry, can be imported, e.g. from dbpedia. If the use case is a supply risk scenario, the knowledge graph can be analysed to identify potential bottlenecks, such as single shared suppliers on a sub-tier level. By mapping companies to production sites and production sites to geolocations, one could integrate risk profiles for a particular geographical region as well as collect the data required for a geographical representation.

Visualisation & decision support The last layer encompasses the tasks that are user-facing, such as the visualisation of the supply chain map. The users of such a system may need to interactively explore the map by filtering of companies of interest, by collapsing or expanding nodes or by zooming and panning in the map. Users may also require to flag or correct errors, such as an incorrectly grounded organisation name. Problems, such as shared suppliers on a sub-tier, could be highlighted for users interested in such information, e.g. those managing supply risk.

6.8 Summary

The idea of this chapter was to build upon the work of extracting individual buyer-supplier relations and widen the scope to an end-to-end approach to automating supply chain mapping from text. This chapter can be regarded as a detailed gap analysis between the achieved extraction of individual relations and the higher goal of automating the overall process of supply chain mapping from text (the research scope). The main purpose was to identify and discuss the challenges that one will likely encounter.

Using some basic aggregation steps, such as collapsing identical entity mentions, the “<entity1><entity2><relation >” triples obtained applying the approach described in Chapter 5 can already be converted in a visual map. However, for automating an overall approach to map supply chains from text, additional challenges need to be considered. Different types of relations, e.g. ternary ones, could be considered in addition to the ones considered in the previous chapter. Furthermore, there are additional requirements arising from the characteristics of supply chains as well as from limited information quality. In particular, establishing the correct tier of a company in a supply chain is difficult. This includes the problem of identifying sub-tier relations (“transitivity problem”) in text documents. It may help to structure the tasks supposed to address the aforementioned challenges in a conceptual framework. The proposed conceptual framework consists of different abstraction layers, from the collecting the documents to more scenario-specific functions, such as providing a visual representation and decision support.

In the next chapter, the approach for extracting individual buyer-supplier relations proposed in Chapter 5 will be applied and validated in case examples. Furthermore, the results will be aggregated and visualised as supply chain maps. During the process of applying the approach to real-world data, some of the challenges discussed in this chapter will reappear in the context of the validation cases.

Chapter 7

Industrial validation

7.1 Introduction

The previous two chapters address the research problem by (1) examining if and how individual buyer-supplier relations can be automatically extracted, and by (2) widening the scope and providing a framework for the end-to-end process of automating supply chain mapping. The purpose of this chapter is to apply the proposed approach and validate it in industrial case studies. Two case examples were chosen, Boeing for the aerospace industry and Toyota for the automotive industry, and presented in a separate section.

The structure of this chapter mirrors the structure of the previous chapters and follows the methodology presented therein, as illustrated by Figure 7.1. First, a labelled corpus is created in Section 7.2. In the subsequent Section 7.3, a number of classifiers are designed as well as trained and tested on the corpus. In Section 7.4, a trained classifier is then used to run the overall processing pipeline on previously unseen, unlabelled data.

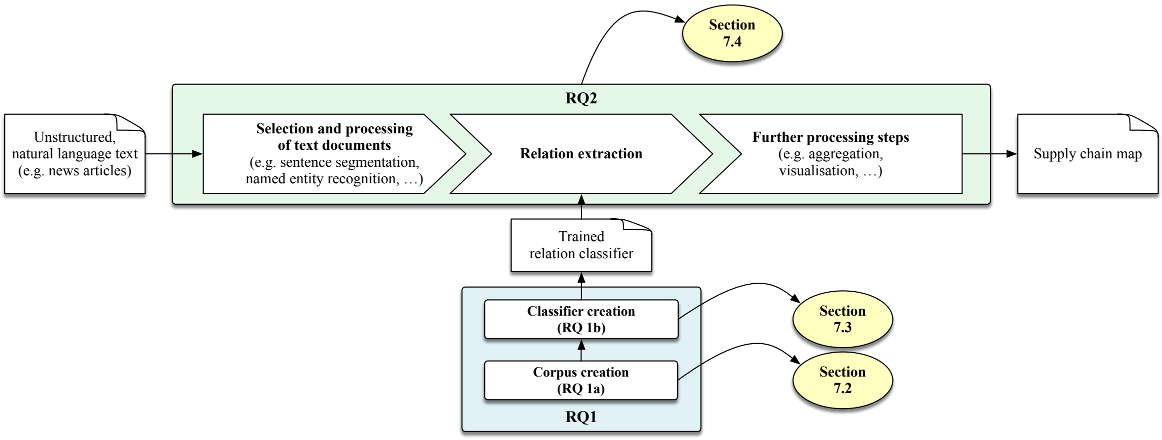


Fig. 7.1 Structure of Chapter 7

7.2 Assessing the corpus creation

This section describes the creation of the labelled text corpus. This includes selecting data sources, sampling from them, defining a process for the annotation and collecting these from human annotators. The reason why this step is described in detail is the significance of the corpus for the performance of the classifier later.

7.2.1 Objectives

The objectives of this experiment include:

- obtain a labelled text corpus that can serve as training and testing dataset for a relation classifier
- assess the labelling quality and validate the overall process

Inter- and intra-annotator agreement of the redundantly labelled initial corpus will serve as a quantitative metric and proxy for the quality of the obtained annotations. This, in turn, also provides an indication regarding the feasibility of the task and the quality of the overall annotation process.

7.2.2 Overview

The creation was done in two stages: an initial corpus was created by having multiple annotators label sentences redundantly to obtain inter- and intra-annotator agreement. To enlarge the corpus and incorporate some of the findings from the first stage, the corpus was then further extended. Figure 7.2 illustrates these stages and already summarises relevant aspects of each stage. Explanations, such as descriptions of the underlying dataset and of the labelling process, will be provided in the corresponding subsections.

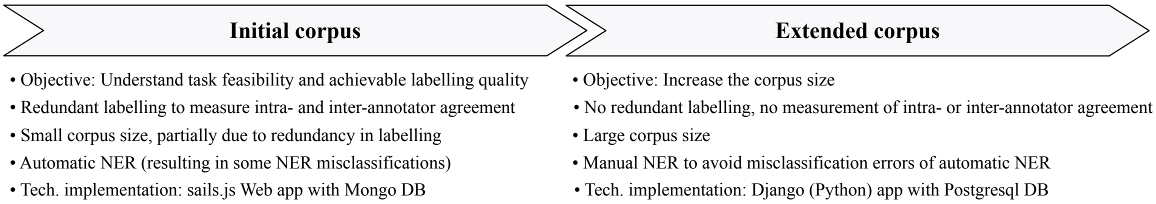


Fig. 7.2 Structure of the section on Corpus Creation

The following sections describe the creation of the initial corpus as well as the extended corpus and summarise the findings.

7.2.3 Formation of the initial corpus

Selected data sources To avoid any potential bias from just a single dataset, multiple datasets were chosen using the selection criteria defined in Section 5.3.2. The documents for the text corpus were obtained from the following data sources: the Reuters corpora TRC2 and RCV1[1], the NewsIR'16 dataset[2], and a customised dataset obtained from webhose[3].

Document and sentence sampling The documents and sentences were sampled following the approach discussed in Section 5.3.2. Documents were sampled from four different sources and assigned to three different partitions each. Manually collected candidate positive sentences were added to the corpus to increase the share of positive examples. The process as well as the four selected data sources are depicted by Figure 7.3.

Documents were segmented into sentences using *spacy*[4] as off-the-shelf solution. To facilitate the subsequent annotation, sentences were automatically NER-tagged[5], again using spacy. The organisation names were *not* provided to the NER tagger in advance. To reduce

[1]Reuters corpora TRC2 and RCV1 (`https://trec.nist.gov/data/reuters/reuters.html`)

[2]NewsIR'16 dataset (`https://research.signal-ai.com/newsir16/signal-dataset.html`); last accessed: 2019-01-10

[3]Webhose.io (`https://webhose.io/`); last accessed: 2019-01-10

[4]Spacy.io (`https://spacy.io/`; last accessed: 2019-01-09) is an open-source general purpose NLP software library. It provides various NLP functions, such as NER, POS tagging, dependency parsing, and word vectors.

[5]NER-tagging is the process of detecting and classifying named entities in a text and tagging these text sequences accordingly. NER stands for Named Entity Recognition.

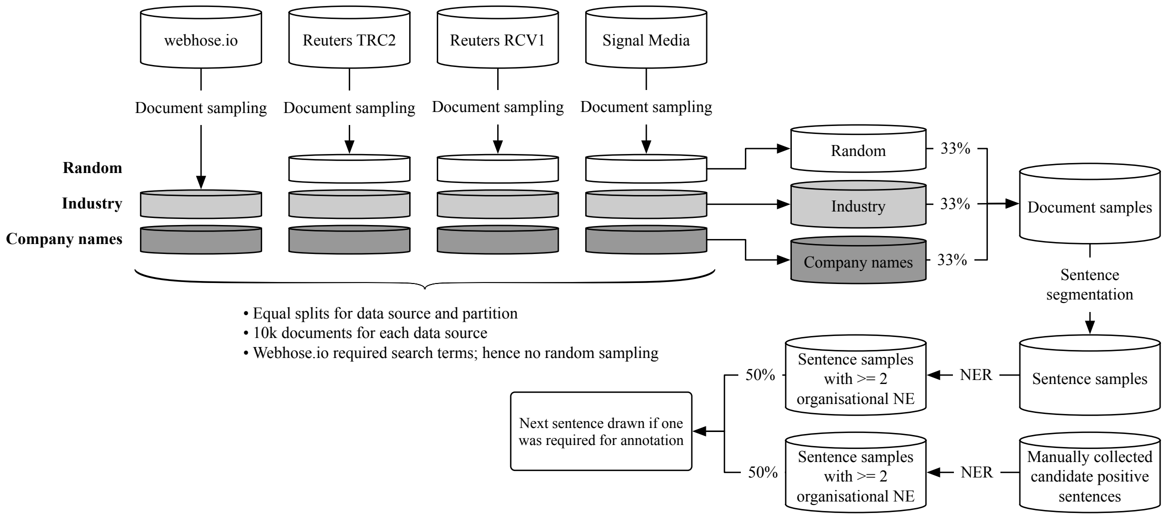


Fig. 7.3 Sampling methodology

false positives, *Flair*[6] and the *Stanford CoreNLP*[7] NER taggers were used in combination with spacy in a simple ensemble. The results of all three libraries had to match for an organisational named entity to be considered in the subsequent steps. Since automatic NER tagging performs well but is still imperfect, organisations may not have been detected, erroneously detected, or detected with incorrect segmentation boundaries. Initially, about 4,100 candidate positive sentences were collected and stored without a label as it was up to the annotators to classify the sentences. Not all of these sentences ended up getting labelled for two main reasons: The automatic named entity recognition system did not identify all organisational named entities and sentences with less than two were not admitted to the annotation process.

Class definitions The class definitions used were the ones proposed in Section 5.3.2. The classes 'no buyer-supplier relation' and 'rejected by annotator' were merged, as discussed in Section 5.4.2.

Annotation collection The annotations were collected following the methodology outlined in Section 5.3.2.

Options for obtaining human annotations are highly scalable platforms like Amazon's Mechanical Turk where micro-tasks can be outsourced for a few US cents per task or directly

[6] Flair (`https://github.com/zalandoresearch/flair`; MIT-licensed NLP software library; last accessed: 2019-06-11)

[7] Stanford CoreNLP (`https://stanfordnlp.github.io/CoreNLP/index.html`; last accessed: 2019-06-11)

recruiting people to perform the task, ideally via some Web-based form or interface. Given the complexity of the task, Mechanical Turk was ruled out as a platform since this would have required substantial efforts to automate the quality management. Instead, annotators were recruited that were known to the author and could be personally instructed to perform the task. This way problems could be quickly flagged and the annotation process be adjusted if required. The small number of annotators was expected to provide a large number of annotation each.

The labelling was conducted independently by seven annotators. To avoid inconsistent labels, annotators were provided with identical, detailed written instructions, including a few example sentences with solutions[8]. They also went through an introductory labelling session. The annotators were the author of this thesis, four academics from the University of Cambridge as well as two friends with an excellent command of the English language. To facilitate the labelling, a Web application was developed to provide an interactive user interface. This Web app also provided the following crucial functionalities, such as:

- Providing labelling options specific to the sentence, i.e. not more selection fields than required to annotate a specific sentence.
- Ensuring that a small partition of the sentences get labelled by all annotators redundantly (inter-annotator agreement)
- Randomly re-introducing the first x sentences of a session towards the end of the same session (intra-annotator agreement)
- Ensuring a pre-set redundancy level, e.g. each sentence gets redundantly labelled by n annotators[9]
- Capturing user feedback about any errors in the previous processing steps (sentence segmentation and named entity recognition)

For the initial corpus, a Web application was developed based on Sails.js and Mongo DB and only aimed to collect relation labels and measure intra- and inter-annotator agreement. The user interface of this application is shown in Figure 7.4.

For each pair of two organisational named entities that had been automatically detected, the most appropriate relation needed to be chosen from a drop-down menu. For the eventual case that a sentence as a whole was not intelligible to label any relation, annotators could reject the whole sentence and flag it as “unclassifiable”.

[8]The used annotation instructions can be found in the Supplementary Material (`https://github.com/pwichmann/phd_thesis`).

[9]This redundancy level was initially set to 3 and then later relaxed to obtain more examples.

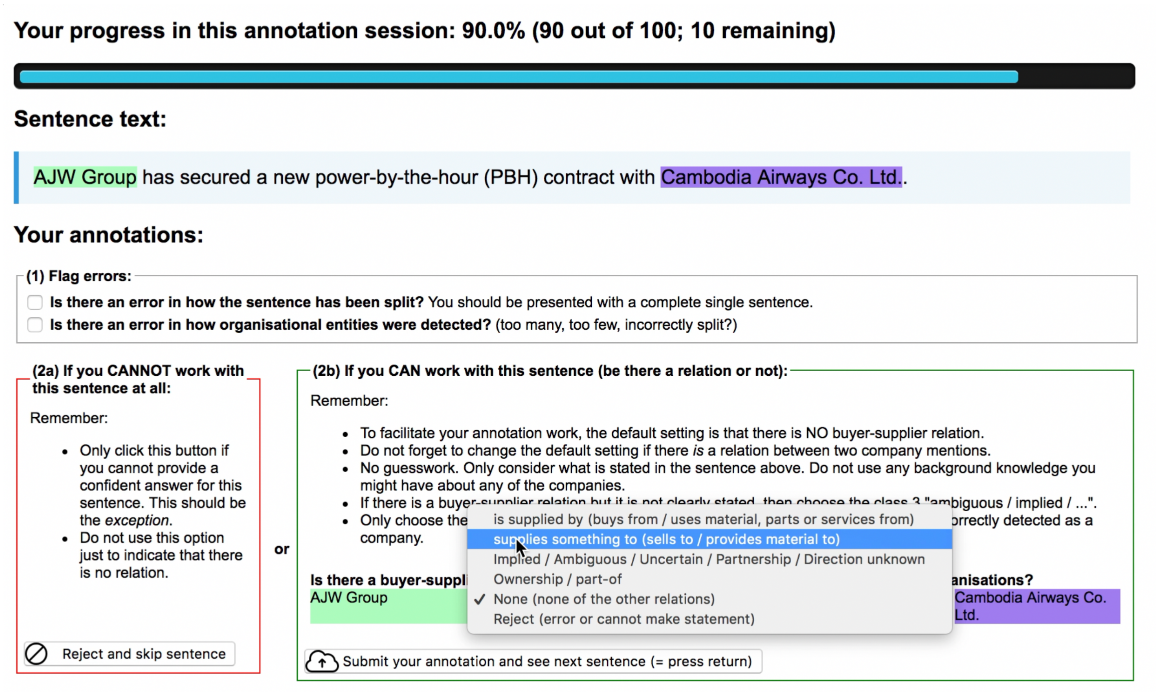


Fig. 7.4 Interface of annotation Web app: Each pair of already detected and highlighted organisations needs to be classified.

An important design decision is whether to not only mask company names for the classification algorithm but also for the *annotators*. Masking can help to prevent annotators from using background knowledge in the labelling process. However, masking would make the labelling task more difficult by adding a layer of abstraction. Hence, the organisational named entities are not masked but revealed to the annotators. To avoid incorrect labels, annotators are specifically instructed not to use the company names as a clue for their labelling decision, to only consider the information provided by the sentence at hand, and to not use any personal background knowledge about the relationship between two organisations.

Corpus characteristics and achieved labelling quality

Overview of corpus characteristics

Table 7.1 below provides key summarising statistics for the created text corpus. For this research, a dataset of 3,887 annotated unique sentences was used, resulting in 8,231 labelled unique arcs. Roughly half of the sentences came from the randomly sampled pool of sentences and another half from the pool of sentences that had been manually collected. Each unique arc could be labelled redundantly by multiple annotators (inter-annotator agreement) and even by the same annotator (intra-annotator agreement). Thus, the number of assigned class labels (14,632) was higher than the number of unique arcs.

Table 7.1 Overview of (initial) corpus characteristics

Metric	**Value for initial corpus**
Number of unique annotators	7
Number of unique sentences that were labelled	3,887
Number of finished annotation items (an annotation provided for a sentence by an annotator)	14,632
Number of unique arcs (after majority vote)	8,231
Number of relation classes	6
Achieved *inter*-annotator agreement (inter-annotator set)	Cohen's $\kappa = 0.90$; Fleiss' $\kappa = 0.90$
Achieved *inter*-annotator agreement (for all sentences annotated by a pair)	Cohen's $\kappa = 0.80$
Achieved *intra*-annotator agreement	Cohen's $\kappa = 0.86$

The contribution of labels across annotators varied and is shown in Figure 7.5.

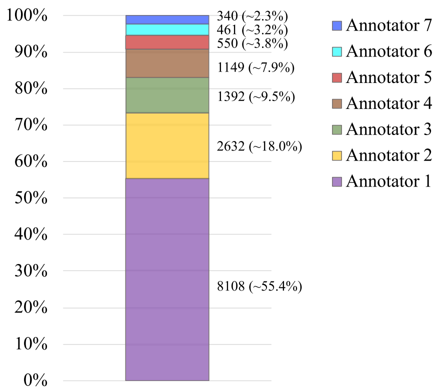


Fig. 7.5 Label contribution by annotator before majority vote (N=14,632 assigned labels)

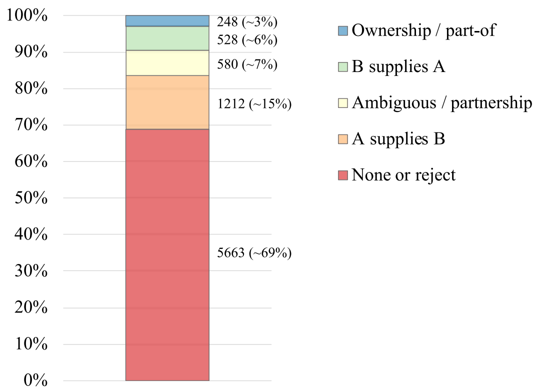


Fig. 7.6 Label distribution for 8,231 arcs (after majority vote)

As expected, the resulting dataset was imbalanced: Nearly 70% of assigned labels were "none" (~60%) or "reject" (~10%). The label distribution after majority vote are shown in Figure 7.6. Because of the imbalance, F_1 score (as opposed to accuracy) is considered the metric to optimise for.

Inter-annotator set (40 sentences) Based on the special inter-annotator set of 40 sentences that all annotators had to process, the achieved inter-annotator agreement is shown in

Table 7.2. Fleiss' (Fleiss, 1971) and Cohen's (Cohen, 1960) κ statistics were chosen to measure annotator agreement. To adapt Cohen's κ from a pair-wise comparison to more than two annotators, the arithmetic average of Cohen's κ across all pairs of annotators was computed. The achieved average *inter*-annotator agreement was $\kappa = 0.90$ for both[10] Fleiss' and Cohen's κ. This suggests annotations of good quality.

Table 7.2 Inter-annotator agreement on the inter-annotator set

Metric of inter-annotator agreement	**Description**	**Value**
Cohen's Kappa	Calculated as simple arithmetic average of Cohen's kappa across all pairwise measurements of inter-annotator agreement	$\kappa = 0.90$
Fleiss' (Multi-)Kappa	Standard calculation	$\kappa = 0.90$

The inter-annotator agreement can also be illustrated in a heatmap as shown in Figure 7.7.

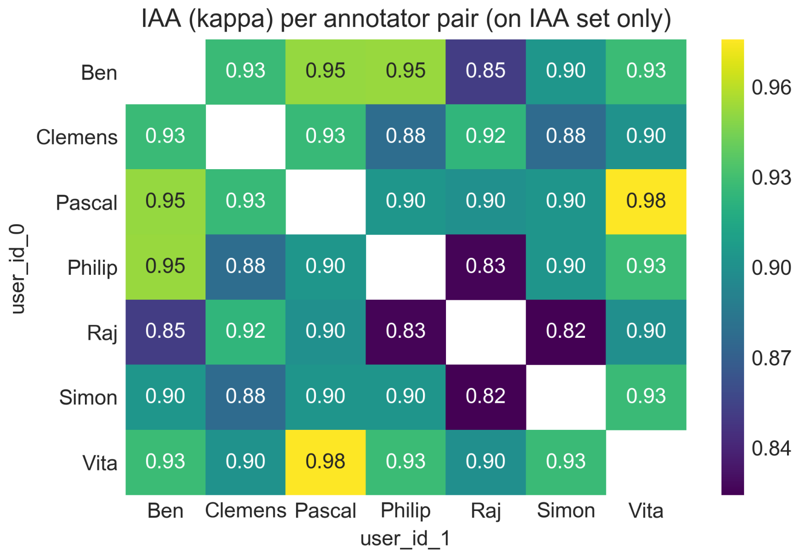


Fig. 7.7 Pair-wise inter-annotator agreement (Cohen's Kappa) assessed on only the inter-annotator set.

Pair-wise inter-annotator agreement (all shared sentences) A better assessment of inter-annotator agreement may be obtained by measuring the agreement that any two annotators show for sentences they both annotated which may potentially far exceed the 40 inter-annotator sentences. Please note that for this evaluation, the sentences as well as the number

[10] The first 3 decimal places were identical for Fleiss' and Cohen's κ.

of sentences was different for each pair of annotators. The result of the pair-wise assessment is shown in Figure 7.8 using Cohen's κ.

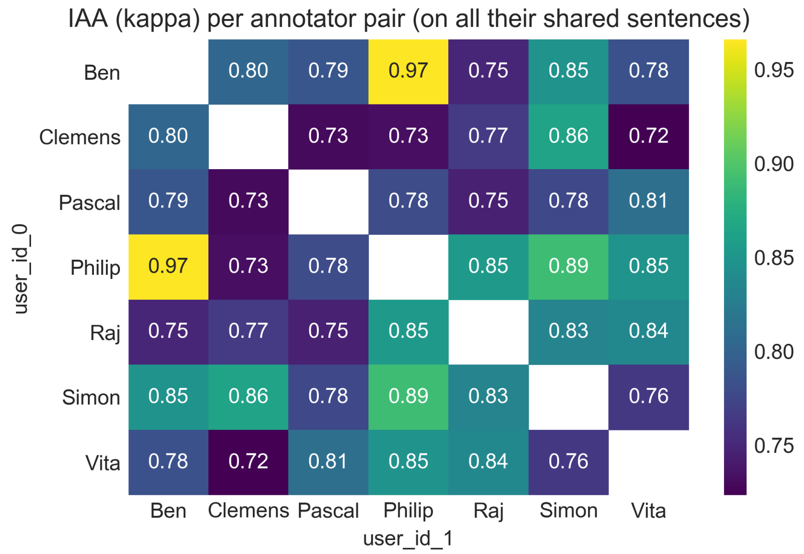


Fig. 7.8 Pair-wise inter-annotator agreement (kappa) assessed on all sentences annotated by both annotators of any pair.

The simple arithmetic mean of all these pairwise agreements was $\kappa = 0.80$ and, thus, lower than the one measured for the inter-annotator set that all annotators had to label.

Intra-annotator agreement Intra-annotator agreement was assessed by re-injecting the first 5 sentences of each annotation session (in random order and interrupted by new sentences) towards the end. The achieved average *intra*-annotator agreement was $\kappa = 0.86$. This suggests annotations of good quality.

7.2.4 Formation of the extended corpus

In early tests, classifiers trained on the initial corpus alone already performed moderately well. However, as expected in Section 5.3.2, the trained classifiers appeared biased by – what could be called – 'trigger words'. These were words that are over-represented in the dataset in those sentences that express a buyer-supplier relationship. Consequently, the classifiers showed a tendency to return false positives when those 'trigger words' were present. The initial corpus was subsequently extended by further annotated sentences. For this extension, the objective differs slightly compared to the initial corpus.

Objective

Since the initial corpus already allowed to draw conclusions about task complexity and labelling quality, measuring inter-annotator agreement was no longer in focus. Instead, it was more important to increase the size of the corpus rather than to label redundantly. To improve labelling quality and to ensure that no data is lost, this time the NER tagging was performed manually for sentences to be added to the text corpus. Performing this task manually addressed the drawbacks of automated NER mentioned in 5.3.2. In case a company name was mentioned as part of a product, such as “Boeing 737”, the organisational named entity mention “Boeing” was still supposed to be tagged as such. Because the NER task was performed manually, the ‘reject’ relation class was no longer required. Instead, a sentence could be directly sent back into the NER process. In subsequent sections, it will be stated which of the two corpora have been used to achieve the results. Where possible, both corpora will be used in separation. The term “Extended corpus” shall refer to the *combination* of initial corpus and additional annotated sentences.

Process

In contrast to the initial text corpus, the sentences were not randomly sampled. Instead, to improve precision, sentences for which a classifier had predicted a buyer-supplier relation of some kind were collected and inserted into the labelling process. Sentences that were identical to sentences already present in the corpus were automatically rejected and not labelled again. Sources were identical to the ones used for the initial corpus. Manually collected candidate positive examples were also added again, in particular those where NER errors had previously resulted in a sentence not being used for training.

Building on the learnings from the first labelling Web application, a second one was developed that allowed for manual NER tagging, for custom named entity classes and relation classes. This second version was developed in the Django Web framework using Python and a Postgresql database. Figure 7.9 provides a screenshot of the NER interface. Named entities could be first selected by double-clicking on a single word or by dragging the cursor from the beginning to the end of the entity mention. Once highlighted, the entity could be classified below using a drop-down menu. The subsequent labelling of relations between two organisational named entities is shown in Figure 7.10 and was similar to the process of the initial corpus.

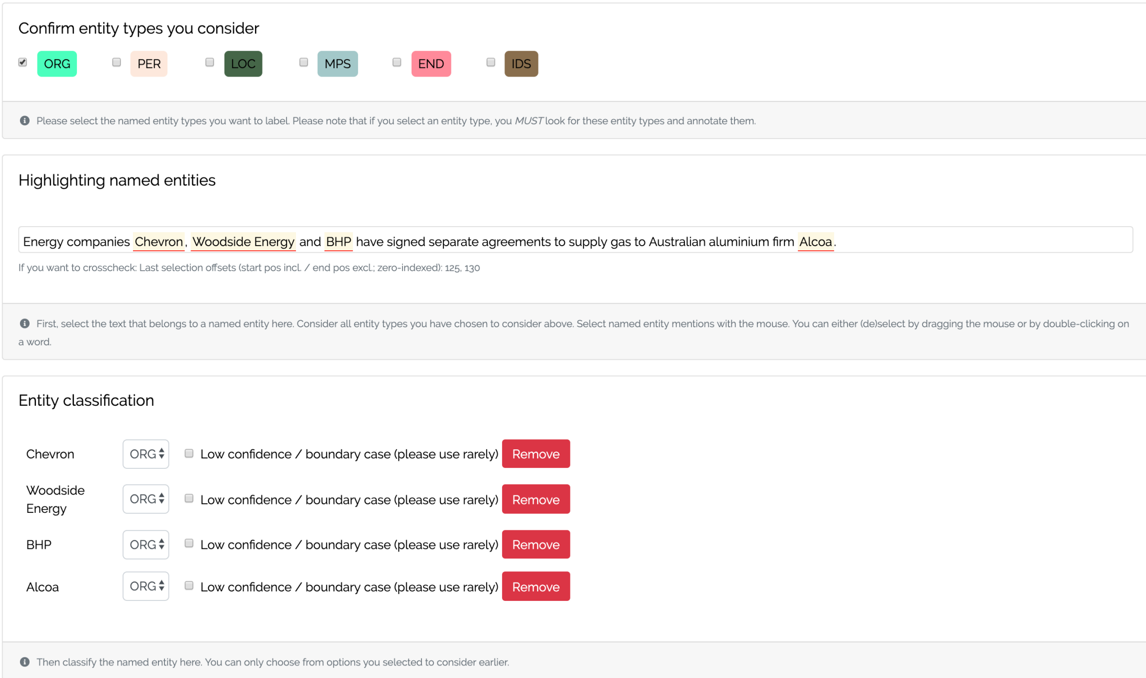


Fig. 7.9 NER tagging interface of rewritten annotation Web app

Corpus characteristics

The corpus characteristics are summarised by Table 7.3. Since the additional sentences were not labelled redundantly this time, no inter- or intra-annotator agreement could be measured.

Table 7.3 Overview of corpus characteristics

Metric	**Initial corpus**	**Extended corpus (incl. initial corpus)**	**Difference**
Number of unique sentences that were labelled	3,887	11,161	$+7,274$
Number of unique arcs	8,231	24,809	$+16,578$

Table 7.4 provides an overview of the label distribution. Generally, “A supplies B” relations occur more than twice as often as the relation of opposite directionality “B supplies A”. A possible reason for this could be that phrases in active voice, such as “company A provides company B with”, are common. “B supplies A” does not necessarily require a sentence in passive voice though. The sentence “Company A purchases material from company B.” would need to be labelled “B supplies A”.

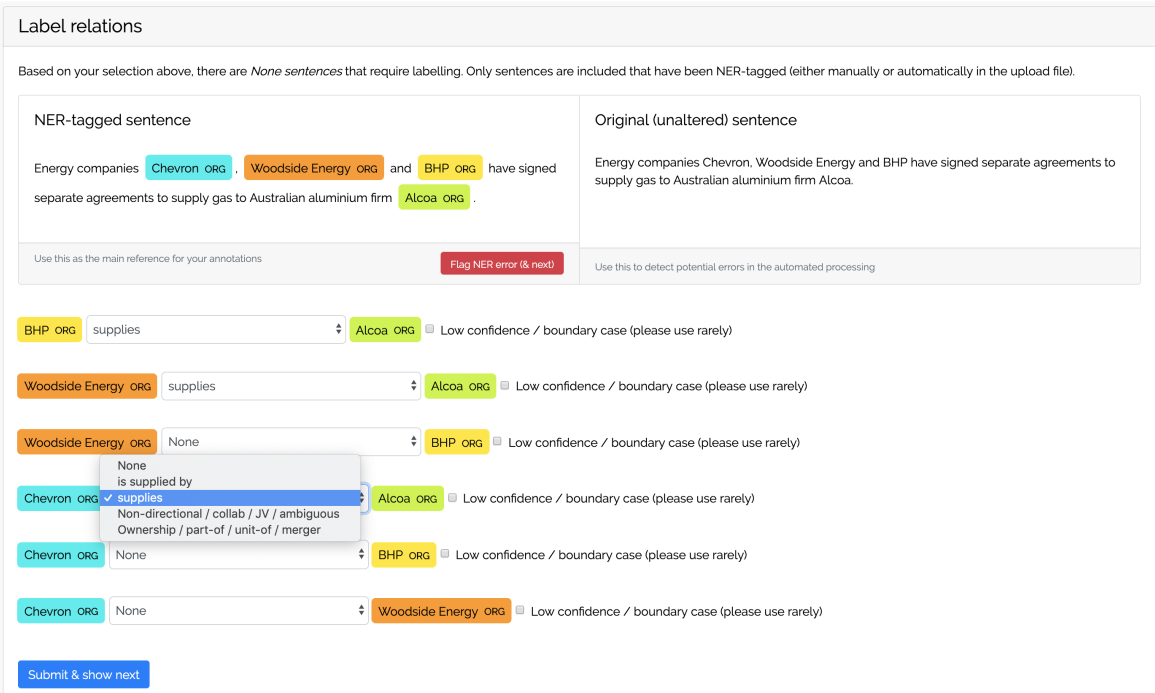


Fig. 7.10 Relation classification interface of rewritten annotation Web app

Table 7.4 Label distribution in corpus

Label	Share in initial corpus	Share in extended corpus (excl. initial corpus)	Share in combined dataset
none or reject	69%	66%	67%
A supplies B	15%	14%	14%
B supplies A	6%	5%	6%
ambiguous / partnership	7%	9%	8%
ownership / part of	3%	6%	5%

7.2.5 Discussion

As stated earlier, the inter- and intra-annotator agreement of the redundantly labelled *initial* corpus will serve as a quantitative metrics for validating the corpus creation process. Strictly speaking, these metrics only measure the consistency of the provided labels. But these metrics also serve as a proxy for the quality of the obtained annotations, provide an indication for the feasibility of the task (given the provided text and instructions) and for the quality of the overall annotation process. The achieved *inter*-annotator agreement of $\kappa = 0.90$ on the inter-annotator set of 40 sentences across all annotators suggests annotations of good quality. It appears that the labelling task was feasible and class definitions were sufficiently understandable. A perfect inter-annotator agreement was not expected due to the generally existing ambiguity in natural language text. The *intra*-annotator agreement of $\kappa = 0.86$ suggests that annotators were also consistent with themselves and did not, for example, guess answers but had followed rules for their voting behaviour.

However, during the process of creating the labelled corpus, a number of aspects could be observed that were not immediately obvious in advance.

Corpus size The size of the corpus was still rather small – even for a confined task like classifying buyer-supplier relations. Using the same methodology as described above, the corpus can easily be extended. The use of "active learning" could make the process of labelling further examples more efficient by only requesting labels for examples where the uncertainty of the classifier is highest.

NER errors A significant share of the manually collected positive examples were also lost due to false negatives in the NER stage. To address this, positive samples can be manually NER tagged in the revised process for the extended corpus.

Data imbalance and biases with consequences As it was unclear in advance what the share of positive sentences in the news datasets would be, the proposed methodology may have placed too much emphasis on the document and sentence sampling. As explained above, due to the low share of positive sentences in the news datasets, candidate positive sentences had to be manually collected and added to the corpus, thereby potentially introducing biases that the document and sentence sampling had aimed to avoid.

Labelling complexity and ambiguity The class definition and annotation instruction may give the impression that the annotation task is a relatively simple one. However, some fraction of the sentences were extremely complex to mentally parse and convert into relations between entities. Furthermore, there were edge cases where correctly distinguishing between classes is practically impossible.

Some relations described *complex buyer-supplier relationships* involving subcontractors. A prime contractor may have been assigned work by a customer. The prime contractor outsources the work to a sub-contractor which delivers the goods directly to the customer. This creates ambiguity as to what extent the material flow is representative of the (inverse) financial flow. Understanding which organisation is buyer and which is supplier becomes difficult, as illustrated by Figure 7.11.

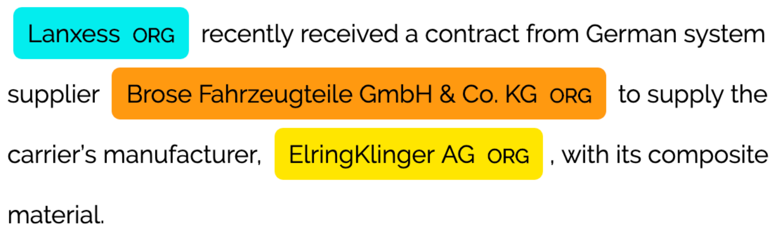


Fig. 7.11 Example of a complex relation involving an intermediary or subcontractor

Abbreviations or other company identifiers, like stock symbols, were often mentioned after the company itself had been named. The NER, however, was often not able to distinguish normal company mentions from other company identifiers. Consequently, annotators were also requested by the labelling app to annotate relations involving these identifiers. Even though the instructions covered this aspect, the treatment of such identifiers was problematic. They represented useful information for entity linking but rather disturbed the process of relation classification. It appears sensible to try to remove these identifiers from the text but store the information for later stages of the processing pipeline. A naive approach could be to remove all short text sequences directly following a detected organisational named entity, do not contain whitespace characters and are enclosed in brackets.

7.3 Assessing the classifier creation

In this section, the annotated text corpus from the previous step can now be used to train and test a classifier.

7.3.1 Objectives

The objectives of this experiment include:

- measure the achievable classification performance to validate the idea of automatically extracting individual buyer-supplier relations from text
- obtain a pre-trained classifier that can be used to extract buyer-supplier relations from a large, unlabelled dataset

To assess the classification performance, the annotated text corpus is considered the gold standard against which any classifier can be tested. As common in machine learning, by training on one partition of the text corpus and testing on another one, the classifiers' achieved performance can be measured in terms of recall, precision and F_1 score. Since an annotated corpus is available, the achieved performance can be measured automatically.

A more meaningful validation result is achieved by comparing the achieved performance levels to the ones produced by *two naive implementations of a dummy baseline classifier*: A random dummy classifier uses a *uniform* probability distribution to vote for any class. A *stratified* dummy classifier is slightly more sophisticated and randomly votes for any class but respects the class distribution in the training set. If the trained classifiers clearly outperform these baseline classifiers, these classifiers can be considered useful.

7.3.2 Details on the technical implementation of classifiers

Word embeddings as features

The feature set consisted of word embeddings obtained from the GloVe dataset. The dataset, originally consisting of 840B tokens and 300-dimensional vectors trained on Common Crawl, was filtered by those tokens actually present in the training data.

Deep Learning classifiers (BiLSTM and MLP)

The BiLSTM and the MLP architecture were designed to expect 380 features as input. This means that, for each classification task, a text sequence of up to 380 "words" could be fed into the network. Each "word" was represented by its corresponding 300-dimensional embedding. In the case of both neural network classifiers, a Softmax layer was used to normalise the outputs so that they could be interpreted as probabilities. The BiLSTM had an embedding layer and considered 16 features in the hidden state of the LSTM layer. It also used a dropout with a probability of 0.5. The last layer was a dense layer with 5 output units, one for each class. The BiLSTM model used the standard hyperbolic tangent function ("tanh") activation function. The MLP architecture was identical in terms of input and output. An embedding layer was followed by a single hidden layer of size 128. This then connected to the output layer of size 5. In initial tests, increasing the network depth did not lead to noticeable performance improvements. ReLU was used as an activation function for the MLP. As common for single-label multiclass classification problems, categorical cross-entropy was used as a loss function for the deep neural networks. "Adam" (Kingma and Ba, 2014) was used as optimisation algorithm. Each network was trained and tested multiple times and the results were averaged over all runs.

Linear SVM classifier

Generally, an SVM is a discriminative classifier that estimates a separating hyperplane in a high-dimensional feature space given labelled training data. The algorithm outputs an optimal hyperplane which can be used to categorise new examples. A grid search is used to tune the hyperparameter of the SVM classifier that is commonly referred to as C. Simply speaking, SVMs aim to fit a hyperplane to separate data points such that (1) the largest minimum margin between different classes is achieved and (2) as many instances as possible are correctly separated. As it is not always possible to optimise both, the C parameter determines the importance of (1).

7.3.3 Achieved classification performance

Given the class imbalances, the micro-averaged metrics are reported in this study, instead of macro-averaged ones[11]. Furthermore, in multi-class single-label scenarios, the micro-averaged recall equals the micro-averaged precision, and hence the F_1-score. Oversampling the minority classes was conducted but did not visibly improve classification performance. The dataset was partitioned into training set (70%), validation set (10%), and test set (20%). The neural networks were implemented in Python 3.6 using PyTorch and trained on a single Linux desktop machine using an NVidia GeForce GTX 1080 Ti GPU. As is common practice, the loss and the F_1 score on the validation data were observed while increasing the number of epochs to avoid over- or underfitting. The model with the best score was automatically saved to avoid under- or overfitting with respect to the number of epochs. The training was conducted up to 100 epochs, and in an initial trial up to 1000 epochs. With a batch size of 32, the best BiLSTM model in our tests was obtained between epoch 17 and 37. Using above system and the initial corpus, a single training and testing run (e.g. BiLSTM trained over 22 epochs and using a batch size of 32) could be completed in approximately one minute.

Overview of results

As expected, the stratified dummy classifier outperformed the random one and achieved a micro-averaged F_1 score of 0.36 compared to 0.20 for the fully random one on the initial corpus. The actual classifiers performed well-above this baseline. Smaller differences in the performance levels were not statistically significant and may be due to chance even though results have already been averaged over 10 separate runs. In particular, the Deep Learning algorithms included stochastic processes. Table 7.5 provides an overview of the results for the

[11] A macro-averaged metric would initially be computed independently for each class and then averaged over all classes. This would treat all classes equally despite their different sizes.

different models. The results in this table were achieved by training and testing on the initial corpus or by training and testing on the extended corpus. The purpose of this table is to show the achievable performance levels and compare them to baseline classifiers to demonstrate the general feasibility of the task. The purpose is *not* to identify a single algorithm guaranteed to always perform best for this type of problem. A conclusion, such as that the SVM with positional features generally outperforms all other algorithms for classifying buyer-supplier relations, would *not* be valid.

Table 7.5 Classification results (micro-averaged F_1 score)

Method	**Configuration**	**Initial corpus**	**Extended corpus**
Random dummy classifier	Fully random dummy baseline classifier (uniform assignment of class labels)	$F_1 = 0.20$	$F_1 = 0.20$
Stratified dummy classifier	Stratified dummy baseline classifier (random voting respecting the training set's class distribution)	$F_1 = 0.36$	$F_1 = 0.36$
SVM	Bag-of-words converted into one-hot-vector (word order and position of organisational named entities are not considered)	$F_1 = 0.68$	$F_1 = 0.66$
SVM	Bag-of-words converted into one-hot-vector; feature vector also contains *positional* information for organisational named entities	$F_1 = 0.75$	$F_1 = 0.77$
MLP	GloVe embeddings; input sequence length of 380; batch size of 32	$F_1 = 0.74$	$F_1 = 0.76$
BiLSTM	GloVe embeddings; input sequence length of 380; batch size of 64	$F_1 = 0.73$	$F_1 = 0.73$

Selected detailed results

For better readability, only the results for the stratified dummy classifier and the BiLSTM on the extended corpus are presented in this section. The detailed results for each classifier and each of the two corpora can be found in the Appendix.

Stratified dummy baseline classifier The performance of the stratified *dummy* classifier can be found in Table 7.6. The results only differed slightly compared to the results on just the initial corpus.

Table 7.6 Classification results per class (averaged over 10 runs) – Stratified dummy classifier; extended corpus

	Accuracy	Precision	Recall	F_1 score
Class 0: None or reject	0.49	0.69	0.47	0.55
Class 1: B supplies A	0.91	0.06	0.06	0.06
Class 2: A supplies B	0.78	0.18	0.14	0.16
Class 3: ambiguous/undirected	0.85	0.07	0.09	0.09
Class 4: ownership/part-of	0.69	0.05	0.30	0.09
Micro-averaged				0.36

BiLSTM The class-wise classification performance for the BiLSTM is shown in Table 7.7. The results are shown for a batch size of 64. Different batch sizes (32, 128) were tested and resulted in slightly lower performance levels (approx. 0.68 to 0.72). Since the process was stochastic, achieved performance levels could differ slightly with each new run. The table shows an average over 10 separate runs.

Table 7.7 Classification results per class (averaged over 10 runs) – BiLSTM; extended corpus

	Accuracy	Precision	Recall	F_1 score
Class 0: None or reject	0.79	0.83	0.86	0.84
Class 1: B supplies A	0.93	0.29	0.29	0.29
Class 2: A supplies B	0.88	0.60	0.55	0.57
Class 3: ambiguous/undirected	0.92	0.47	0.55	0.51
Class 4: ownership/part-of	0.94	0.42	0.26	0.32
Micro-averaged				0.73

Discussion

Overall, it was possible to train classifiers that clearly outperform the dummy baseline classifiers and achieved F_1 scores above 0.75. The trained networks appeared to be able to distinguish well between Class 0 and all others, as demonstrated by the F_1 score of above 0.85 for Class 0 achieved by most of the (non-dummy) classifiers. These results can be considered a validation for the proposed approach of automatically extracting individual buyer-supplier relations from natural language text.

However, a number of limitations need to be considered: First of all, only one aspect of the envisioned end-to-end processing pipeline was examined in this experiment: the prediction of relationship classes for already NER-tagged sentences. The errors introduced by the NER itself and other potential pre-processing steps were not considered in the stated classification performance. A further limitation is that the achieved precision for classes 1 to 3 is substantially lower than 1. If such a classifier is applied to large quantities of text, this will result in a large absolute number of false positives. Recall of the Classes 1 to 4 still remained relatively low which was likely due to the small size of the obtained annotated dataset. More concretely, the classifier may have encountered completely new linguistic expressions in the test phase that it had not encountered during the training phase. This is true in particular for the classes with small sample size, such as "ownership / part-of" and "B supplies A". As was shown in Table 7.4, the class "A supplies B" was more than twice as common the dataset as the equivalent relation with inverted directionality "B supplies A". This may be one reason for the different F_1 score. The obtained dataset focussed on two manufacturing industries, automotive and aerospace, which may limit its usefulness for other, industries with different supply network structures. While some generic expressions, such as "supplies with", are used across industries and the classifier should perform well for these, other expressions may be more industry-specific and including examples of these in the dataset could lead to the discovery of further relations. With regards to defining relationship classes, there appears to be a trade-off between the number of relationship classes and the simplicity of the classification task. More classes may lead to a longer annotation time or lower labelling quality. On the other hand, the defined classes are still limited in their ability to distinguish more subtle semantic differences in relationships. For example, it may be useful to distinguish buyer-supplier relations that have just ended from those who explicitly never existed, and a relationship that is explicitly said to have never existed may have to be distinguished from one where information is just lacking.

7.4 Assessing the end-to-end process

In this section, the previously proposed, end-to-end approach is applied to large, unlabelled, and previously unseen datasets, as illustrated by Figure 7.12. This implies that – as opposed to the testing of the classifier in the previous section – all steps from the pre-processing of documents to a basic visualisation of the detected supply chains were performed automatically this time.

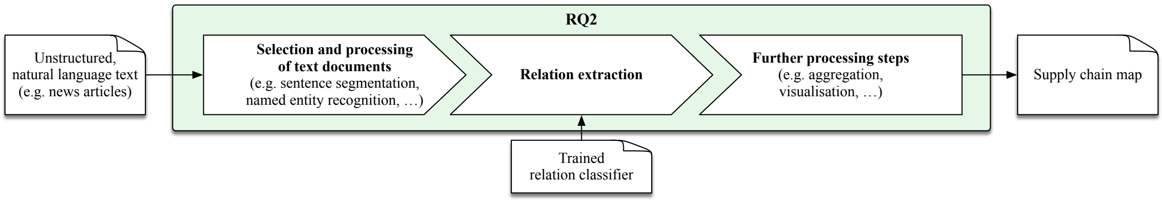


Fig. 7.12 Application to unseen data

7.4.1 Objectives

Validating the end-to-end processing pipeline that includes the extraction of buyer-supplier relations from a large, previously unseen dataset is less straightforward than for the corpus creation or classifier creation stage. There are more facets to be examined and the performance level, such as the recall of buyer-supplier relations, cannot be automatically assessed anymore. The validation also aims to provide indicative answers to questions of data sparsity (how much relevant information is in a given dataset) and information availability (how much information about a supply chain can be found).

The objectives of this experiment include:

- test the end-to-end processing pipeline from text to the visualisation of basic supply chain map, including automatic sentence segmentation and NER tagging
- test the classifier on previously unseen, large real-world datasets covering a wide range of topics and, hence, predominantly out-of-domain data
- gain insights regarding the representativeness of the training data and the ability of the trained classifier to generalise beyond the corpus
- draw basic[12] conclusions about the information availability and data sparsity in a typical news dataset, and, thus, the usefulness of the overall approach

7.4.2 Methodology

Subsequently, the conceptual framework proposed in Section 6.7 will be followed. However, not all proposed tasks, such as the inference of sub-tier relations, could be implemented within the scope of this research. Thus, the extracted individual buyer-supplier relations will

[12]Only basic conclusions are possible since only a few datasets can be tested, co-reference resolution is not being used and all conclusions are conditional on the latent error rates along the processing pipeline, such as NER tagging.

be aggregated and visualised but the resulting supply chain maps will not be transitive (cf. the transitivity problem discussed in the previous chapter). An NER system will be used to automatically detect mentions of organisations but no named entity resolution will be performed. Hence, duplicate references to the same organisation using different names are to be expected in the results.

The validation was performed on three different datasets: a random sample of sentences, a dataset with sentences that contained the word “Boeing”, and a dataset with sentences that contained the word “Toyota”. After applying the proposed approach to these datasets, it is possible to quantitatively analyse the results. For instance, one can determine which share of sentences was found to contain a buyer-supplier relation. A manual inspection of a random sample of the results allows to produce a confusion matrix and sources of errors along the processing pipeline can be identified.

The nature of this experiment differs from previous ones: The proposed approach was supposed to be tested on a *large* dataset that has *previously not been seen* by the classifier and is *unlabelled*. Furthermore, all processing steps were supposed to be executed *automatically* – without human annotations at any stage. Compared to previous experiments, this led to a number of fundamental changes:

Large, previously unseen dataset The classifier was applied to a large, previously unseen dataset. This was supposed to reveal if the relation classifier is able to generalise beyond just the (potentially biased) training corpus. The size of the dataset had a further consequence: Buyer-supplier relations are relatively rare in a general news dataset. In combination with a vast dataset of millions of sentences, a less-than-perfect precision of the processing pipeline means that a large share of returned buyer-supplier relations was now expected to be false positives.

Introduction of new error sources Previously, during the Corpus Creation phase, some steps may have been done manually. This time, these needed to be executed fully automatically, such as language detection, sentence segmentation, and named entity recognition. In particular, the automated named entity recognition could be expected to introduce additional misclassification errors. For instance, products may get incorrectly labelled as organisations. In previous experiments, these errors were not considered. These errors are not directly observable unless samples are manually inspected.

Unlabelled dataset leads to changing performance measurement Due to the lack of labels and the large size of the dataset, the available evaluation options were limited: The performance of the approach (especially recall, and this F_1 score) could no longer be automatically measured. The lack of labels means that results had to be inspected manually. The large size of the dataset means that it may only be possible to check small random samples rather than all results. Especially recall was difficult to measure since the metric would require knowledge of how many relations of a specific type existed in the complete dataset or at least in a random sample. Precision was easier to measure since the metric only required knowledge of the returned results which are considerably fewer than the size of the complete dataset. Furthermore, it was no longer possible to separate out different error sources. Errors could occur in the stages prior to the classification, such as NER tagging. Without manual inspection, errors that arose at different stages of the processing pipeline were not separable.

Selection of a text corpus The requirements and options for a text corpus were identical to Section 5.3.2 where a text corpus had to be created for the training data. A further requirement was that the dataset the classifier was applied to should be large so that the classifier was exposed to a wide variety of text examples. This is different from the training data where only a small sample has to be used. The dataset should also have little overlap with the previously created text corpus to test if the classifier can generalise.

7.4.3 Experiment and results

Applied model Out of the set of tested algorithms, the BiLSTM model was chosen. In the previous tests, the model had performed clearly better than any dummy classifier and the bag-of-words SVM classifier. Even though the BiLSTM did *not* clearly outperform the MLP or the SVM classifier that uses positional information, the BiLSTM was also used for these tests in the hopes that knowledge of the full token sequence in a sentence may in the end turn out to be beneficial. In the following experiments, the word tokeniser used to preprocess sentences was not set to split off contractions, e.g. the "'s" in "Boeing's". However, the NER system Flair was pre-trained on data where this is the case. Thus, the actually achievable NER performance and relation extraction performance may be slightly better than the performance stated in this thesis.

Dataset To apply a trained classifier to a previously unseen and different dataset, a dataset was collected by scraping a range of blogs and news websites for the automotive and aerospace industry, including a few selected supplier industries (such as plastics and composite materials). The dataset contained only the text content of the articles and had a size of approximately 1.4 GB (more than 40 million sentences).

Overview of subsequent analyses Figure 7.13 provides an overview of the subsequent three sets of analyses. All analyses are based on the dataset of automotive and aerospace news articles. In a first set of analyses, a random sample of sentences was processed. This way, a large number of sentences was processed to test the general effectiveness of the approach. In the subsequent two sets of analyses, the dataset was filtered for sentences that contain a specific company name. Thus, the complete dataset was considered but only the fraction of the data was used that was relevant for the focal companies. Because sentences may have contained the name of the focal company but also described other buyer-supplier relations, the detected relations were then filtered again: either the target or the source entity needed to contain the name of the focal company. The first case used Boeing as a representative of the aerospace industry, the second case used Toyota as a representative of the automotive industry.

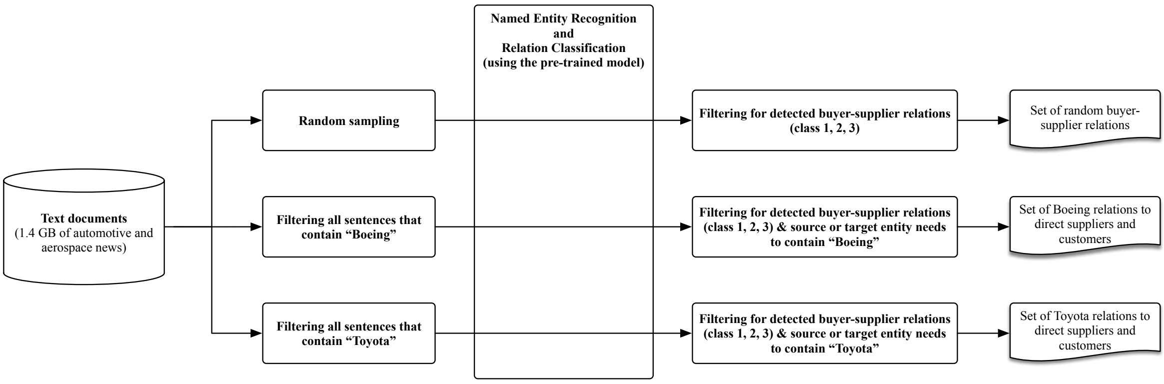


Fig. 7.13 Process overview of subsequent analyses

7.4.4 Random sample

A random sample of approximately 582k sentences sentences was processed for the following analyses. Duplicate sentences were ignored and not processed. It is possible that, by chance[13], some examples from the training dataset were also included in the large, unannotated corpus. However, the unlabelled corpus was orders of magnitude larger than the training dataset.

Analysis on sentence-level

The first set of analyses examined the results on a sentence-level. Figure 7.14 provides an overview of how many sentences actually contained a specific number of organisational named entity mentions. This analysis was important because only sentences with at least two

[13]E.g. due to copied press releases reported by multiple sources

explicit organisational mentions would be considered for further analysis. In this case, the results were based on the prediction of the Flair NER system (which may have included false positives and false negatives).

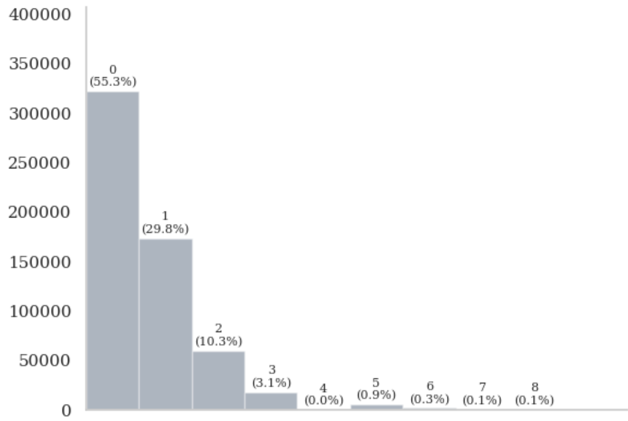


Fig. 7.14 Distribution of ORG mentions per sentence: Number of sentences containing x organisational named entities, for $x = 0, 1, 2, \ldots, 8$

Figure 7.14 shows that $55.3\% + 29.8\% = 85.1\%$ of the sentences did not contain two or more organisational named entities. Thus, only about 15% of sentences "qualified" for a relation classification.

On a sentence-level, the funnel looked as illustrated by Figure 7.15. Of about 582k sentences, approximately 15% contained two or more organisational named entities. About 12 percentage points of those related to sentences without any detected buyer-supplier relation. About three percentage points related to sentences for which at least one buyer-supplier relation (class 1, 2 or 3) could be detected. Ignoring any misclassification errors in the NER or relation extraction phase, this provided a first rough indication for the information sparsity that can be expected in an industry-specific news dataset (in this case: automotive and aerospace).

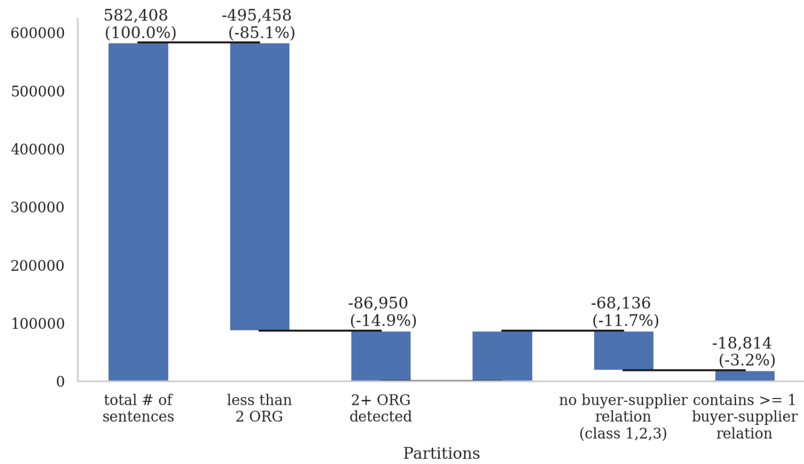


Fig. 7.15 Waterfall on sentence-level

Analysis on arc-level

As mentioned before, if a sentence contained more than two organisational named entities, then each possible (unordered) pair of such entities represented an arc that required a class label. As illustrated above, approximately 87k sentences contained two or more organisational named entities. This corresponded to approximately 208k arcs that required a label prediction.

Analysis of label distribution Figure 7.16 provides an analysis of the results on an arc-level. Out of the approximately 208k arcs, 86.7% were classified as “none” (class 0). About 7.1% were classified as “B supplies A” (class 1) and about 2.4% were classified as “A supplies B” (class 2). 1.5% were classified as “partnership / ambiguous / ...” (class 3). A remaining 2.3% were classified as stating an ownership relation (class 4). It needs to be kept in mind that these percentages are skewed by false positives and false negatives for each class and can only provide an approximate indication of the true shares.

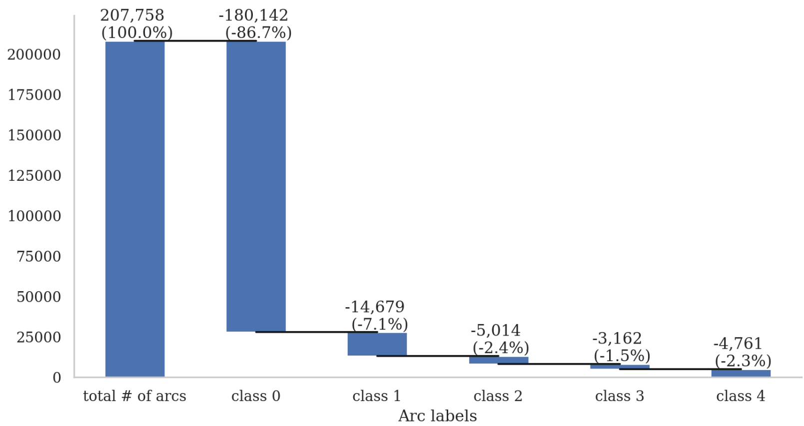


Fig. 7.16 Waterfall on arc-level: Share of class labels in the overall number of arcs

The large share of class 1 predictions was not expected since the labelled text corpus contained more class 2 than class 1 examples. This may be indicative of a larger share of false positives in class 1 – potentially caused by a smaller number of examples for that class in the corpus. This is supported by the fact that – when tested on the labelled corpus – the classifiers showed a lower precision (and recall) for class 1 that for class 2.

Analysis of confidence score distribution In addition to providing the most likely label, the classifier also provided the corresponding confidence score. Since there were 5 classes, a "winning" label could not have a confidence score of less than 0.2.

Figure 7.17 shows the distribution of confidence scores for "winning" labels for each class. The distribution for class 0 was negatively skewed: the mode of the distribution was at the right end where the confidence was highest. For all other classes, the mode was much more centred around a confidence of 0.5. Overall, the classifier was much more confident in identifying negative examples (class 0) than in assigning any of the other class labels. A large number of buyer-supplier relations were still predicted with a confidence close to minimum required for a "winning" label (0.2). With more training data, these distributions are expected to shift further to the right. The y axes of the diagrams represent the number of times an arc was assigned a particular class label with a particular confidence score. The layout reflects the fact that the relations of interest are predominantly the classes 1, 2 and 3.

Manual inspection of a sample

To assess the results, a sample of 500 randomly sampled arcs and their assigned class labels was inspected manually. This was done by reading the original as well as masked sentence

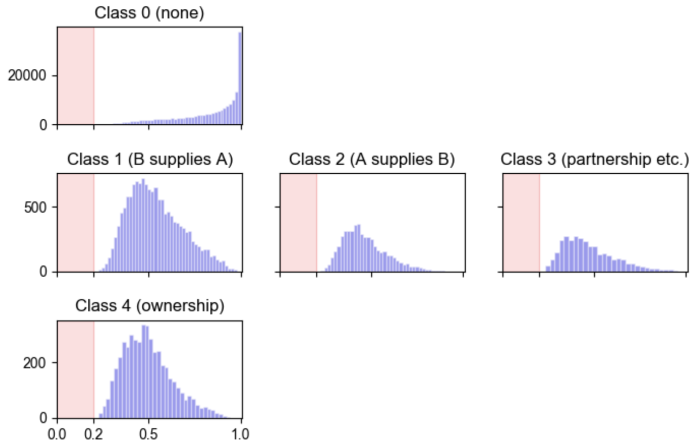


Fig. 7.17 Distribution of confidence scores for "winning" labels

and recording the true label next to the automatically predicted one. In addition, it was noted if any of the two organisational named entities in question for a specific arc had been detected erroneously.

Confusion matrix By comparing the vector of true labels with the vector of predicted labels, a confusion matrix could be built, which is provided by Figure 7.18 (without normalisation) and Figure 7.19 (with normalisation) below. The informative value of this confusion matrix is still limited by the fact that only a small fraction of the results could be inspected manually. The confusion matrices showed that the classifier was able to detect examples of the "none" class reasonably well. 82% of the "none" examples were classified as such. In 11% of the cases, these were still misclassified as "B supplies A" (class 1). However, especially examples of the classes 1 to 4 were still frequently confused. A perfect classifier would have had values of 1.0 along the diagonal of the normalised confusion matrix.

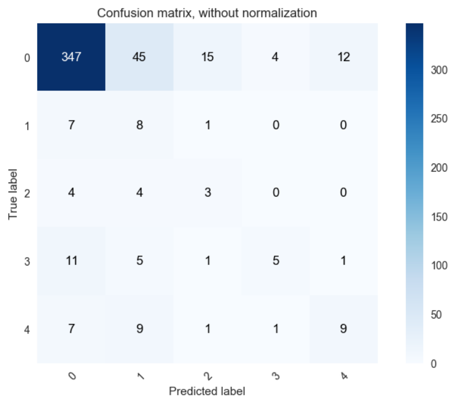


Fig. 7.18 Confusion matrix without normalisation

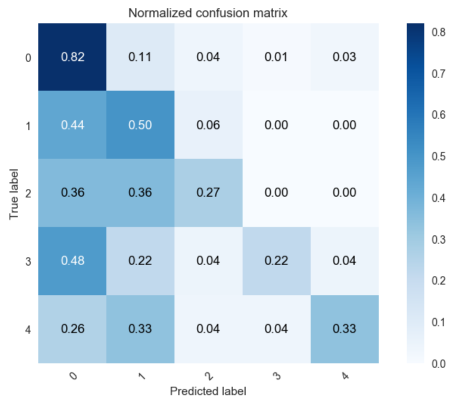


Fig. 7.19 Confusion matrix with normalisation

By restricting the predictions to those above a confidence score threshold, the number of misclassifications could be reduced but not fully eliminated, as shown in Figure 7.20 and Figure 7.21. In this case, predictions with a confidence score < 0.7 were ignored.

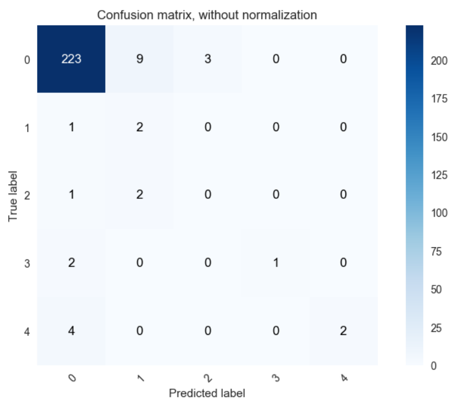


Fig. 7.20 Confusion matrix without normalisation (min score 0.7)

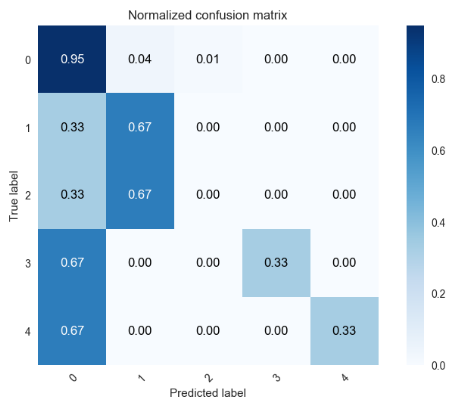


Fig. 7.21 Confusion matrix with normalisation (min score 0.7)

NER errors Incorrectly detected organisational named entities did not automatically lead to an incorrectly detected relation. However, in some instances this may have been the case, even for a human who misunderstood a named entity as a company rather than, for instance, a product name. NER errors were common in the inspected sample. In 50 out of 500 sentences (10%), at least one of the two organisational named entities in question was incorrectly detected as such (false positives or major segmentation errors). The most common error was the misclassification of a product name or brand as a company (e.g. the car models "Zafira Life" or "Pontiac"). In addition to collecting more training data, an NER system that is able to better distinguish mentions of organisations from mentions of products would improve the relation extraction performance.

Visualisation of the resulting supply chain map

Figure 7.22 illustrates some of the extracted buyer-supplier relations as a simple supply chain map. Before the visualisation, basic aggregation steps were carried out – as described in Section 6.5.1. No manual corrections were performed for this map. Thus, the network visualisation is based on *predicted* organisational named entities and their *predicted* relations. The complete network was too large to be shown; only a small segment is presented here. The network was also filtered to only include arcs between two entities which were both detected with a confidence above 0.9 and a confidence for the relation class label of above 0.8. In this particular visualisation example, classes 1 to 4 were not optically distinguished and are shown as undirected arcs. Arcs labelled as "none" were not included in the visualisation.

In this example, the nodes for GM and General Motors Co. were manually highlighted to show that named entity linking was not conducted. Using named entity linking, these two nodes could be collapsed into a single one. The common NER error type of mistaking a brand name for a company name could also be observed in this network segment (here: Ford Ranger). In some cases, the relation between an organisation and a product misclassified as an organisation could be misclassified as a buyer-supplier relation.

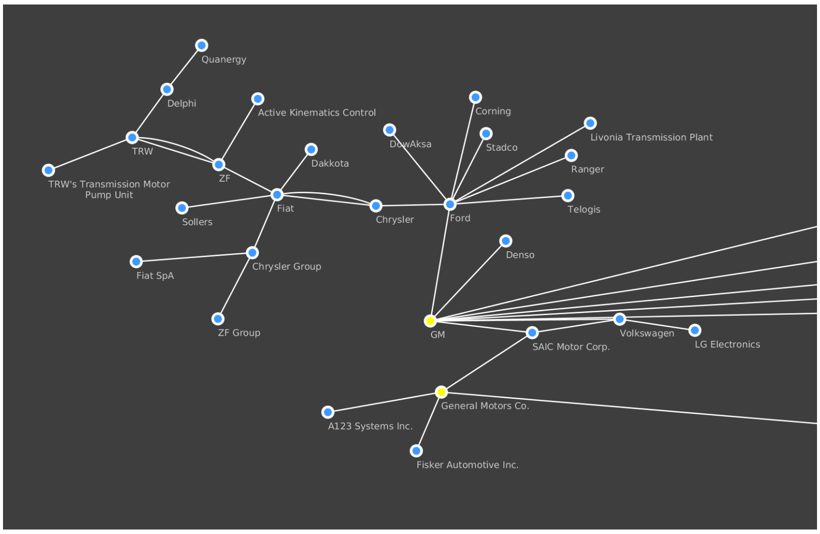


Fig. 7.22 Small segment of the resulting network; visualised using the open source network visualisation software Cytoscape

Figure 7.23 shows a small cut-out of a network visualisation as provided by the open source software Gephi. Unlike the typical representation of a focal company's supply chain map in clearly separated tiers, this automated layout shows the dense network of interrelations. The data was visualised as produced by the processing pipeline without any manual correction of incorrect nodes or arcs. As in the example visualisation before, the graph demonstrates the need for further post-processing. For instance, Airbus and Airbus Group (surrounded by ellipses in Figure 7.23) could be collapsed into a single node after entity linking. Due to a word tokenizer anomaly[14] in this experiment which was only discovered later, possessive forms were not correctly identified and considered as part of the organisational named entity, such as "Liebherr's", as opposed to just extracting "Liebherr". This may have resulted in some misclassification errors. Some names of products or product features were incorrectly identified as organisations, such as "PW1100G" (the name of a Pratt & Whitney engine type).

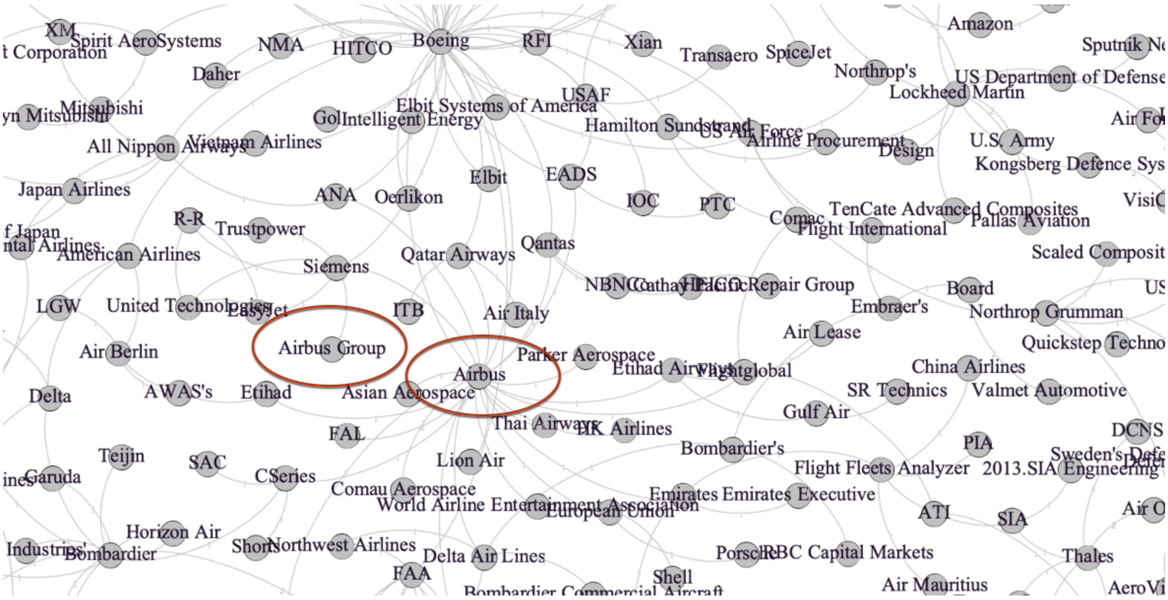


Fig. 7.23 Network visualisation with Gephi; small cut-out of the overall network and only based on a small fraction of the corpus

Summary

Overall, it was possible to generate a rudimentary supply chain map from text without manual processing steps. Since only direct buyer-supplier relations got extracted and aggregated,

[14]Unexpectedly, the segtok word tokenizer did not split off apostrophe and apostrophe "s" at the end of named entity mentions.

the resulting supply chain maps could only show potential sub-tier suppliers as opposed confirmed sub-tier supplier relations (cf. transitivity problem discussed in Chapter 6). Since all processing steps in the end-to-end pipeline could be executed automatically, it was possible to process large amounts of data and filter it down to a subset of sentences that were likely to be relevant for supply chain mapping. In this experiment, ca. 582k sentences could get filtered down to ca. 19k sentences that were classified to contain at least one buyer-supplier relation of some kind. Ignoring any misclassification and other processing errors along the way, roughly 3% of sentences of the industry news dataset described a buyer-supplier relation. The data sparsity in a general news or even general Web dataset will be even more pronounced. Unfortunately, entity mentions, such as product names, were frequently incorrectly classified as organisations by the NER system. In subsequent processing steps, these false positives get masked in a sentence and the sentence is presented to the relation classifier. This can result in subsequent errors in the classification of the relation.

7.4.5 Aerospace manufacturer

For this case problem, the aerospace manufacturer Boeing was selected.

Case problem and motivation

Boeing was an interesting case since approximately "70% of the development and production activities" of the 787 Dreamliner had been outsourced to suppliers (Tang and Zimmerman, 2009). After a number of supply chain issues and production delays, the degree of outsourcing was later reduced for subsequent programmes (e.g. see Gates and Grunbaum (2011)). Understanding the structure of its extended supply chain has thus been of great interest to Boeing.

Estimating the number of direct Boeing suppliers Estimating the number of Boeing's direct suppliers was difficult for a number of reasons. The context for published numbers is often not sufficiently specific:

- Were only *direct* suppliers considered?
- Were also suppliers considered that relate to secondary activities (e.g. HR services, accounting, catering, ...)?
- Were different sub-organisations counted separately, such as country organisations or sales organisations?

- Which of Boeing's divisions were considered (BCA[15], BDS[16] and/or BGS[17])? If only a subset was considered, what is the supplier overlap?
- Were all suppliers actually active or just certified/approved?

In 2005, the Boeing magazine Frontiers reported: "79 Percentage reduction in the number of suppliers that Boeing business units work with today. This number has dropped from more than 30,000 in 1998 to 6,450 now. These 6,450 suppliers are based in more than 100 countries. [...] 3,000 [is the] [a]pproximate number of suppliers that can trade through Exostar, an aerospace Internet "virtual marketplace" founded by Boeing, Lockheed Martin, Raytheon, BAE Systems and Rolls-Royce." (Arkell, 2005). It is not immediately obvious if this number includes only material, parts and component providers or also providers of software and services. And it is not clear if the number only comprises first-tier suppliers. In October 2018 (13 years after the Frontiers article), the online magazine SupplyChainDive reports that "[t]hree million parts arrive at Boeing facilities every day from 5,400 suppliers" (Cosgrove, 2018). If the wording was deliberate, this would suggest 5,400 *first-tier* suppliers for both the commercial and military Boeing units. It is not immediately clear if different legal entities of the same company (e.g. country organisations or similar) would count as different suppliers. A Boeing press release from April 2018 states a very different number of suppliers: "In 2017, Boeing spent almost $60 billion with nearly 13,000 suppliers from all 50 U.S. states and 57 countries. Supplier-provided components, services and engineering support make up approximately 65 percent of the cost of Boeing products." (Boeing, 2018). A conversation of the author with a BDS manager in 2019 revealed that BDS reported approximately 4,680 active direct suppliers for its division. It was not possible to obtain the exact definition of these suppliers.

Experimental design

The unlabelled text corpus was first filtered for sentences containing the word "Boeing". The NLP pipeline, including NER and relation classifier, was then applied to these pre-filtered sentences to identify Boeing suppliers and customers. Once these had been identified by name, the corpus could again be filtered by those names and the pipeline could be applied to them.

[15] Boeing Commercial Airplanes
[16] Boeing Defense, Space & Security
[17] Boeing Global Services

Results

Named entity recognition Using Flair as NER system on the text corpus resulted in 76,120 detected named entities that contained the word “Boeing”. As expected, not always was the term identified as an *organisational* named entity. When mentioned as part of a product name, such as “Boeing 737”, the whole sequence would be classified as MISC (miscellaneous named entity). Moreover, depending on the context, sometimes even just the word “Boeing” was (often correctly) classified as a product. However, a mention that was not classified as organisational was lost at this stage of the processing pipeline – dramatically reducing the recall that was achievable on an unlabelled dataset. Future extensions of this work should address this limitation. Figure 7.24 provides an analysis of the detected named entity mentions that contained the word “Boeing”. Around 60% (approx. 46,000 mentions) of these entities were classified as ORG. Approximately 40% (approx. 30,000 mentions) of these entities were classified as MISC. The class labels PER (for person) and LOC (for location) are not necessarily incorrect. E.g. “William Boeing” was indeed a person and “Boeing Field” in “[...] flight to Boeing Field in Seattle , Washington” was identified as a location.

Fig. 7.24 Distribution of class labels for named entities containing the word "Boeing"

Figure 7.25 provides an analysis of the detected named entity mentions that contained the word “Boeing” and that were classified as organisational. About 90% of mentions were covered by the six text sequences, only about 10% were other text sequences. The commonly extracted organisational mentions were generally correct.

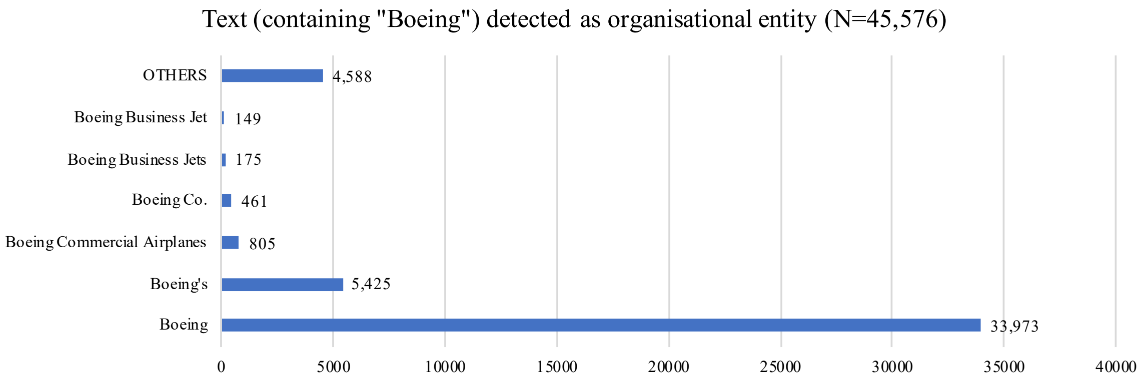


Fig. 7.25 Top unique text sequences containing "Boeing" and classified as organisational named entity

Figure 7.26 provides the same analysis as before but for instances that were classified as miscellaneous (MISC), i.e. as a product. The commonly extracted organisational mentions were generally correct, and it can be argued that in some contexts the single word "Boeing" indeed referred to the product rather than the company. Apart from some errors, most mentions were correctly classified as MISC, given that they were Boeing products. But, as stated before, all these mentions classified as MISC were lost for the further analysis even though they may have been useful in identifying buyer-supplier relations.

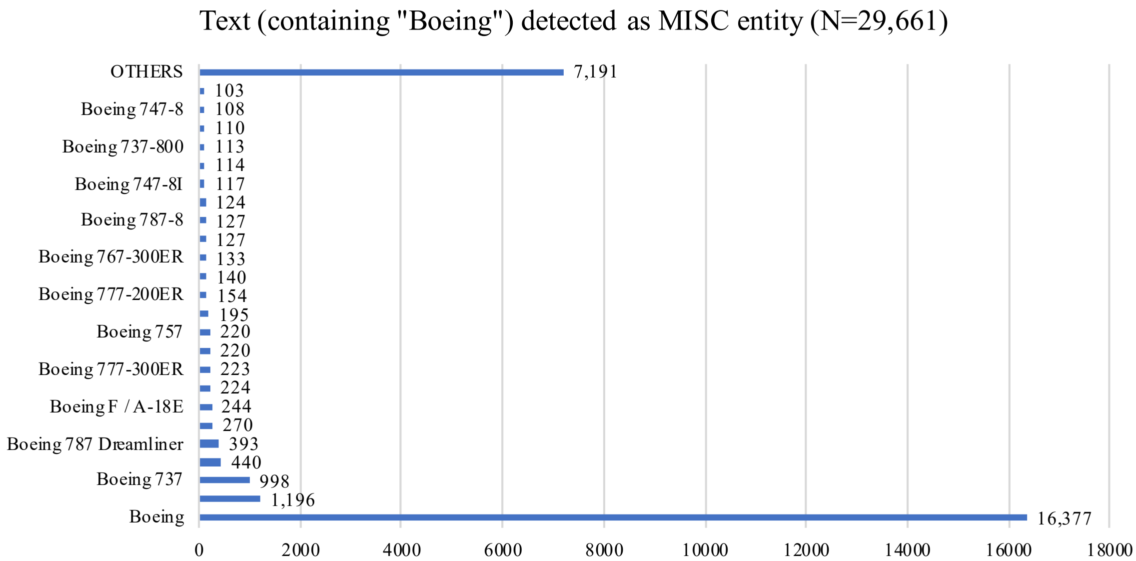


Fig. 7.26 Top unique text sequences containing "Boeing" and classified as MISC named entity

Analysis on sentence-level On a sentence-level, the filtering process is illustrated by Figure 7.27. Because sentences had been filtered to contain the word “Boeing”, the share of sentences with less than two organisational named entity mentions was lower (approx. 63%) than on a random sample of the corpus (approx. 85% as shown in Figure 7.15). Text sequences containing “Boeing” but being classified as products still kept this share artificially high. About 12% of sentences contained at least one predicted buyer-supplier relation.

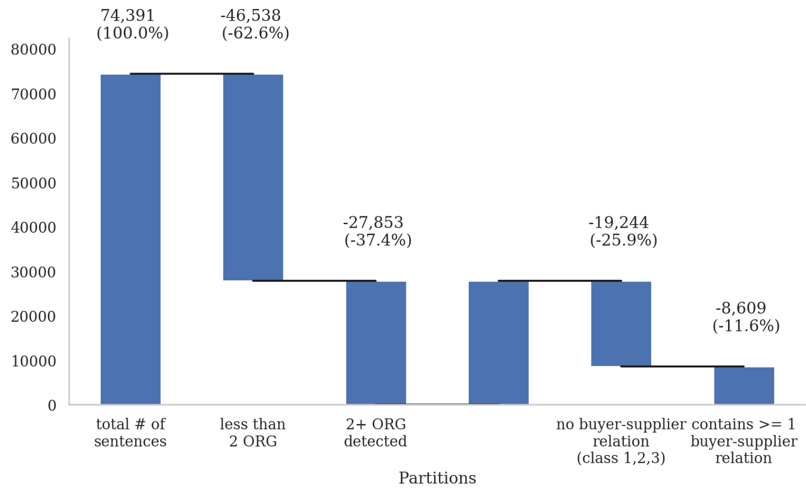


Fig. 7.27 Waterfall on sentence-level (only sentences containing “Boeing”)

Analysis on arc-level On an arc-level, the distribution of relation class labels was similar to the previously analysed random sample. Figure 7.28 provides the corresponding waterfall illustration. Approximately 85% of all arcs were predicted to have no relevant relation. Class 1 (“B supplies A”) is the largest class with approximately 11% (compared to approximately 7% in the random sample). Based on previous inspections, a large share of these class 1 instances were expected to be false positives due to overfitting of the classifier on the rather small training data.

Not all of these arcs included Boeing in the pair of considered entities. A sentence containing the word “Boeing” may have contained other organisational named entities and each possible pair would be considered an arc that required a label. Overall, 6,555 arcs were predicted to describe a buyer-supplier relationship (class 1, 2, 3) in which one of the two organisational named entities contained the word “Boeing”.

Figure 7.29 provides an overview of the distribution of confidence scores for the different relation classes. This analysis was conducted for arcs extracted from sentences that contained the word “Boeing”. Similar to the random sample, the classifier tended to be highly confident

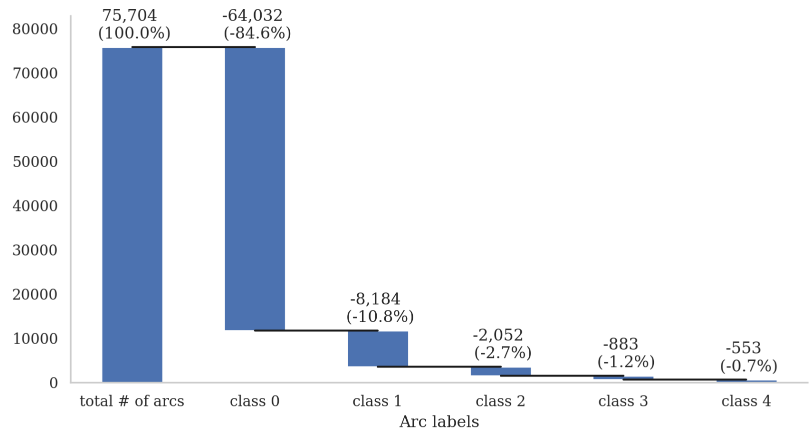


Fig. 7.28 Waterfall on arcs-level (extracted from sentences containing "Boeing")

for a large fraction of the examples assigned the "None" class. For all other classes, the confidence was still relatively low with a mode around or below 0.5. With more training data, the distributions are expected to shift further to the right. See Figure 7.17 for a more detailed explanation of the type of analysis.

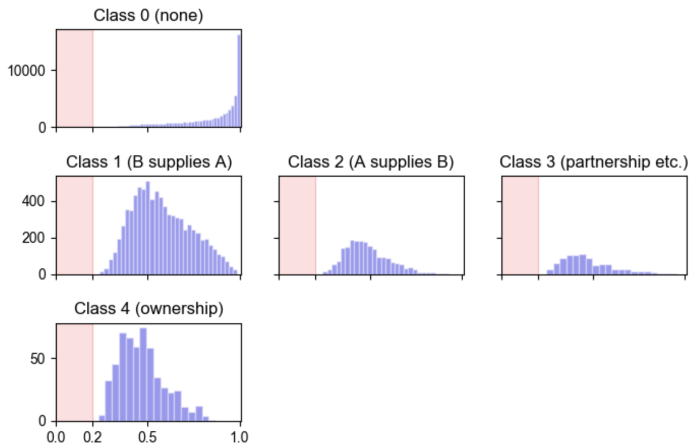


Fig. 7.29 Distribution of confidence scores for "winning" labels (for arcs extracted from sentences containing "Boeing")

Visualisation of first-tier relations For the visualisation, no manual corrections were carried out; the result represents what was produced automatically by the processing pipeline. Repeated mentions of identical relations (same named entity mentions, same relation class) were aggregated. Out of 6,555 arcs 3,743 unique ones remained. The result had not yet been filtered by the confidence score for the named entity or the relation. Filtering was likely to improve precision but would reduce recall. Error types included wrong directionality, incorrectly recognised named entity, incorrectly classified relation.

Figure 7.30 shows the entities predicted to be organisations that were in a (direct) relationship with Boeing, first-tier suppliers *and* customers. Arcs were filtered for class 1, 2 and 3, and one of the named entities needed to contain the word "Boeing".

The network had been zoomed out to illustrate the following aspects: (1) The disconnected graphs consisting of just two or three nodes were likely to contain entities that were incorrectly detected as organisational named entities and contained the word "Boeing". Examples of such cases were "FlightSafetyBoeing" and "BoeingAmid" (the latter was likely to have been caused by a failed sentence segmentation). Some "chains" were longer than just a single tier due to multiple organisations in this chain containing the word "Boeing". (2) In the centre of the large cluster, the common names for Boeing could be found, such as "Boeing" or "Boeing Co.", which were connected to a large number of other organisations.

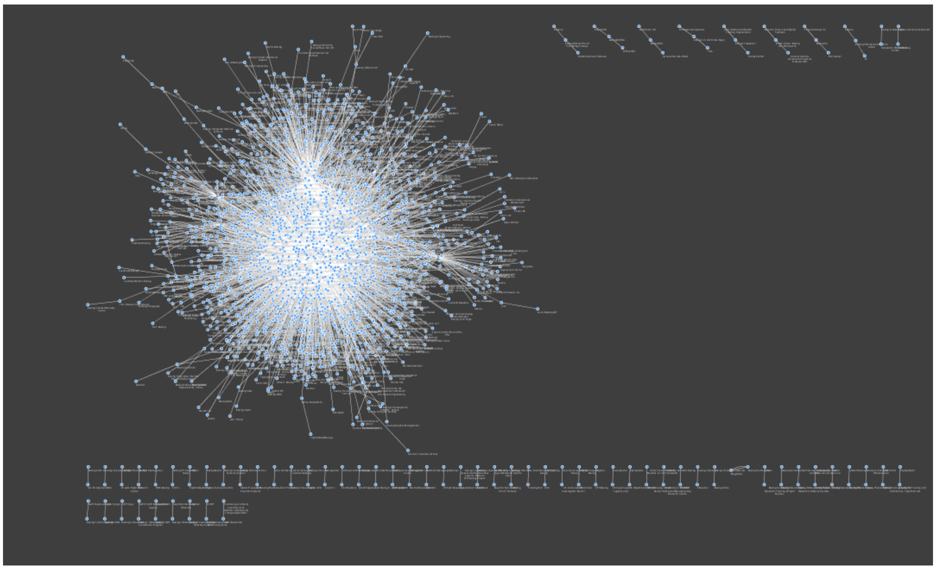


Fig. 7.30 Overall network of supposed Boeing direct customers and suppliers; visualised in Cytoscape

Figure 7.31 shows the network in a different visualisation layout. This visualisation shows how – without entity linking – different name variations for the same entity created different clusters, e.g. "Boeing", "Boeing's" or "Boeing Co.".

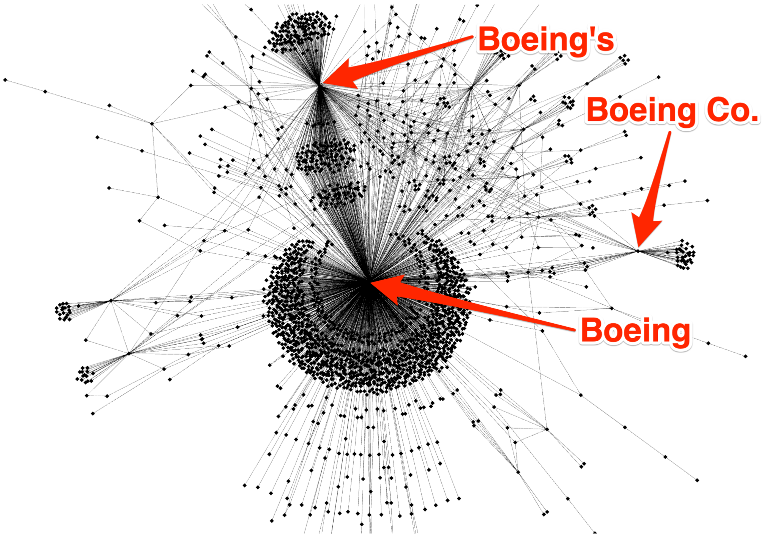


Fig. 7.31 Overall network of supposed Boeing direct customers and suppliers; visualised in Gephi

Identified customers and suppliers 2,089 distinct entities were predicted to be Boeing suppliers, customers or partners. These distinct entities have not been grounded, and thus include duplicates, and also may have contained erroneously detected named entities as well as misclassified relations. E.g. due to the large number of instances where Boeing and Airbus are mentioned together in one sentence, it is not unexpected that Airbus was misclassified as customer and supplier of Boeing. The number of instances was so large that a classifier with less than perfect precision would assign incorrect class labels numerous times. A further extension of this work could consider a mechanism similar to tf-idf[18] to normalise for co-occurrences. Another option could be to identify rivals via a separate relation class to avoid misclassifications.

To better aggregate the extracted information, all nodes containing "Boeing" were collapsed into a single node. Additionally, organisations predicted to stand in a buyer-relationship had now been classified as customers, suppliers, or "partners / ambiguous". Table 7.8 shows, in descending order, the top 30 organisations based on the number of times it was predicted as "supplier" to Boeing. In addition, it is also shown the number of times this organisation was classified to be Boeing's "customer" or "partner / ambiguous / ...". Seemingly contradictory labels did not always signal an extraction error since companies can take different roles over time or even simultaneously. The following sentence is such an example:

[18] tf-idf stands for term frequency–inverse document frequency

Example sentence
Boeing also had a major presence at the show, bringing a brand new AH-64E Apache attack helicopter that it leased from the US Army.

Within the context of this sentence, the US Army in fact supplied Boeing with a leased helicopter – inverting the commonly expected directionality of the relation between the two organisations. However, in most cases shown in the table, labels for both classes "supplier" and "customer" were indicative of remaining classification errors somewhere along the processing pipeline. Confidence scores for named entities or the relation labels were not considered for this table.

Examples of *correctly* extracted customers or suppliers The following examples were correctly processed. The considered pair of organisation named entities is shown in brackets ([...]), *other* detected named entities are shown in curly braces ({...}). The organisational named entity mentioned first is referred to as "A", the second one is referred to as "B".

Example sentence
[Boeing] partners with [Thermwood] on 3D printed tool for 777X.

In the example above, the named entities were correctly detected and classified as organisations. The relation was correctly predicted to be class 3 (partnership / ambiguous / ...) – albeit only with a rather low confidence of 0.54.

Example sentence
The [USN] awarded [Boeing] a $60.8 million contract as part of the Increment 3 Block 2 improvements for the Poseidon on 5 August.

The named entities were correctly detected and classified as organisations. In this context, "USN" stands for US Navy. The correctly predicted class label for the relation was "B supplies A", that is Boeing supplies the USN. The confidence of the model for this prediction was 0.83.

Example sentence
He does elaborate, but that deal, announced in December 2018, set conditions related to [Spirit's] long-term supply of commercial aircraft components to [Boeing].

The named entities were correctly detected. The contraction (apostrophe s) should have been split off, which was a word tokenisation error caused by an incorrect configuration of the tokeniser. The correctly predicted class label for the relation was "A supplies B", that is Spirit supplies Boeing. The confidence of the model for this prediction was rather low at 0.38.

Table 7.8 Top 30 extracted organisations predicted to be in direct relation with Boeing, sorted by the descending number of times the label "customer" was assigned

Detected named entity mention	cust. ↓	supplier	ambiguous / partner	Ratio cust. / supplier	Comment on plausibility using background knowledge
Airbus	329	189	26	1.74	Imperfect precision and large number of arcs containing Boeing and Airbus
USAF	51	30	1	1.70	USAF stands for US Air Force and is likely to be customer
FAA	45	33	2	1.36	FAA is the US Federal Aviation Administration; neither supplier nor customer
US Air Force	44	29	5	1.52	see USAF
ANA	34	25	2	1.36	All Nippon Airways (ANA) is a Japanese airline and likely a customer
Spirit AeroSystems	27	20	5	1.35	Spirit is an aerostructures manufacturer and likely a supplier
Spirit Aerosystems	27	20	5	1.35	see above
All Nippon Airways	22	35	0	0.63	see ANA
Spirit	22	14	7	1.57	see Spirit AeroSystems
United	21	6	0	3.5	In this context, United stands for United Airlines and is likely a customer
MAX	19	36	2	0.53	Max is a result of an NER error; "737 MAX" (aircraft type)
Max	19	36	2	0.53	see MAX
US Navy	19	25	3	0.76	The US Navy is likely to be customer rather than supplier to Boeing
Embraer	18	32	37	0.56	Embraer is a Brazilian aerospace conglomerate and aircraft manufacturer; Boeing and Embraer maintain a partnership
Qatar Airways	16	8	0	2.00	Qatar Airways is an airline and likely customer
NMA	15	21	5	0.71	NMA is result of an NER error; NMA stands for New Midsize Airplane and is a Boeing aircraft concept, not an organisation
Lockheed Martin	15	17	5	0.88	Lockheed Martin is an aerospace and defense company and is both rival as well as partner to Boeing
American Airlines	15	17	1	0.88	American Airlines is an airline and as such likely to be a customer
Southwest Airlines	15	14	1	1.07	Southwest Airlines is an airline and as such likely to be a customer
SOUTHWEST Airlines	15	14	1	1.07	see Southwest Airlines
Air New Zealand	15	14	0	1.07	Air New Zealand is an airline and as such likely to be a customer
US Federal Aviation Administration	14	15	0	0.93	see FAA
Lufthansa	14	13	0	1.08	Lufthansa is an airline and as such likely to be a customer
Air Force	14	9	1	1.56	The US Air Force is likely to be a Boeing customer
United Airlines	14	9	0	1.56	see United
US Army	13	20	2	0.65	The US Army is likely to be a Boeing customer
Ryanair	12	12	2	1.00	Ryanair is an airline and as such likely to be a customer
Sky Interior	12	8	0	1.50	Sky Interior is result of an NER error; this is not an organisation but a new interior aircraft design branded by Boeing
Singapore Airlines	12	6	1	2.00	Singapore Airlines is an airline and as such likely to be a customer
Lockheed	12	4	5	3.00	see Lockheed Martin

Examples of *incorrectly* extracted customers or suppliers The following examples were *incorrectly* processed and selected to illustrate specific errors that can occur along the processing pipeline. With respect to understanding the capabilities and limitations of the approach, the incorrectly processed examples tended to be much more informative than the correctly processed ones.

Example sentence
[Boeing] has been the prime contractor for [GMD] since 2001 and works with an industry team to incorporate improvements.

Without any background knowledge, it looks as if the sentence had been correctly processed. "Boeing" and "GMD" were classified as organisations and the predicted relation label was class 2 ("A supplies B"), which appeared to be correct. However, "GMD" is not actually an organisation and stands for "Ground-based Midcourse Defense system". This easily made error in the classification of the named entity led to subsequent errors in the classification of the relation. Without background knowledge, a human is likely to make the same mistake. If, in some sentence, a named entity is misclassified as an organisation, the classifier (or a human) can only misclassify the relationship.

Example sentence
[Air Lease Corp] has ordered two new Boeing 787-9s and finalised a previously announced order for 12 737 Max aircraft, [Boeing] says on 8 August.

The organisational named entities have been correctly detected. The predicted relation label was "B supplies A". At first glance, this appears to be correct. However, strictly speaking, the relation exists between "Air Lease Corp" and the first mention of "Boeing", not the second one. If the second mention got replaced by a news agency, such as Reuters, there would be no buyer-supplier relation. These cases are ambiguous since a human would be able to understand which role the second organisation is likely to play: If it is a news agency, it cannot be a supplier but just a neutral reporter. However, if in this context, the organisation is an aircraft manufacturer, then it is at least likely to be that supplier of the ordered aircraft.

Example sentence
{Tata} has signed deals with several Western companies, including one to manufacture components for [Boeing] and another to produce helicopter cabins for [Sikorsky].

The named entities were correctly detected, and this sentence clearly expresses some buyer-supplier relationships. However, the arc between "Boeing" and "Sikorsky" needed to be labelled as "None" – and not as "B supplies A" which the classifier did with a confidence of 0.35. Instead, Boeing and Sikorsky are said to be supplied *by Tata* in separate deals.

Example sentence
[Boeing] formed a joint venture with {India's Tata Sons} in 2008, called [Tata Boeing Aerospace].

The named entities were correctly extracted[19]. The assigned label is class 3 (ambiguous / partnership). This is plausible at first sight since a joint venture between two companies is supposed to be assigned this label. However, the arc that is being labelled is between the co-owner of the joint venture (Boeing) and the joint venture itself. Thus, it would require the ownership class label.

Example sentence
[Boeing] has brought its new BBJ Convertible to [NBAA] for the first time to highlight the multi-mission business jet.

This example was chosen to illustrate the inherent ambiguity in natural language. The named entities were correctly extracted. At first glance, it may even seem to a human reader that Boeing supplies a product to another organisation. However, NBAA is the National Business Aviation Association, which is in fact an organisation but, in this context, probably stands for their yearly convention and exhibition. Thus, there is no buyer-supplier relationship stated in the sentence.

Manual inspection of a sample of 150 arcs For this analysis, all arcs were collected where one entity contained the word "Boeing" and the arc was predicted to be a buyer-supplier relation (class 1, 2, or 3). The model used to predict buyer-supplier relations was identical to the one used throughout this chapter[20]. Out of the set of all returned arcs, 150 arcs were randomly sampled and manually inspected. The results of the inspection are as follows:

- 46 of the 150 arcs (approximately 31%) had been correctly classified, that is the assigned class label 1, 2, or 3 had indeed been the right one.
- In 63 out of 150 cases (approximately 42%), the correct class label would have been class 0. These errors could only partially (17 arcs) be explained by NER errors in any of the two relevant entities.

[19] Apart from including "India's" in the named entity which is a preventable error introduced by the word tokenisation step.

[20] The prediction model used was identical. However, the word tokenizer was set to split off contractions, such as apostrophes. This time, 6,350 arcs were identified as opposed to 6,555 arcs before (with contractions not being split off). Splitting off contractions is expected to result in a better NER performance since Flair was trained on data where contractions had been split off. Unfortunately, the training data for the model contained sentences where contractions had been treated differently.

- In all other 41 cases (approximately 27%), a misclassification occurred by confusing the class labels 1, 2, or 3.

Common errors were the misclassification of products as organisations which then resulted in the wrong relation class being assigned to the corresponding arcs. Approximately 71% of the sentences (corresponding to the 150 sampled arcs) did, in fact, describe at least one buyer-supplier relation (class 1, 2, 3) between entities mentioned in the sentence.

Iteratively adding potential sub-tier suppliers By extending the scope to sub-tier suppliers of Boeing, at least two aspects can be further investigated: (1) Is it possible to identify sub-tier suppliers and potentially even identify risks on a sub-tier level? (2) Companies, like Boeing, are widely and frequently covered in the news. This will be different for smaller suppliers. To what extent is there still some information in general industry news about those suppliers and their relations?

Honeywell was one of the organisations identified as a direct supplier to Boeing. The exercise that was previously carried out for Boeing could now be carried out for Honeywell.

Figure 7.32 provides a network visualisation of the result in Gephi. Again, without entity linking, multiple clusters form around distinct name variations, such as "Honeywell" and "Honeywell Aerospace". Honeywell International Inc. is the official name of the international conglomerate. Honeywell Aerospace is the aerospace division of the conglomerate giant. Overall, 747 arcs were classified as buyer-supplier (class 1, 2, 3) relations. Figure 7.32 shows the resulting network after a basic aggregation step. As expected, more arcs related to Honeywell's customers rather than Honeywell's suppliers.

An example of an identified supplier to Honeywell is provided by the following example.

Example sentence
[Elbit] had purchased the [Honeywell] helmet product line for the Boeing AH-64 Apache, which is called the integrated helmet display and sighting system.

The organisations Elbit and Honeywell were correctly identified. The Boeing helicopter was identified as a product (MISC), and Boeing was not separately detected as an organisation. The predicted relationship class was "B supplies A", which is correct given the provided sentence. This sentence appears to describe a sub-tier relation: Honeywell provides the helmet product line to Elbit which is a supplier to Boeing's Apache helicopter. The confidence for the relation label was 0.78.

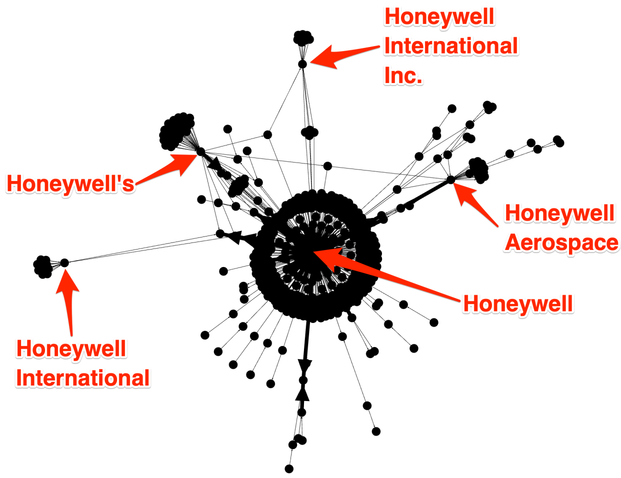


Fig. 7.32 Overall network of supposed Honeywell direct customers and suppliers; visualised in Gephi

Example sentence
Additive manufacturing company [Sintavia] has received internal approval to make production parts for [Honeywell Aerospace] using the powder bed fusion AM process.

The organisations were correctly detected in the sentence above. The predicted label "B supplies A" has the incorrect direction, the correct label would have been "A supplies B". This confusion is likely due to the phrase "has received". The relation label was assigned with the confidence 0.45. However, the sentence reveals a potentially interesting sub-tier supplier for aerospace customers of Honeywell.

Estimating the number of identified Boeing suppliers A further interesting question is how many distinct Boeing *suppliers* could be identified in the data. Answering this question is difficult for at least the following reasons:

- It is required to perform entity linking to remove different mentions of the same entity (or same company group), e.g. "Tata" and "Tata Advanced Systems". This step has to be performed manually for best accuracy.
- Not all arcs classified as positive state a Boeing supplier, some of them state a customer. Some relations do not involve Boeing and some are false positives. Again, this means arcs have to be manually inspected.

- Companies may have changed their names or may have been bought since the time of reporting. Without substantial research efforts, this leads to double-counting if the underlying data covers a large time period (e.g. “Subaru, formerly known as Fuji Heavy Industries, is a major supplier for Boeing.”).

- Distinguishing direct from sub-tier suppliers is not always possible given the provided text.

- It is often impossible to distinguish which Boeing department is supplied by an organisation, e.g. Boeing BCA or Boeing BDS.

- The number of suppliers heavily depends on the size and quality of the input data.

- The overall number of Boeing suppliers is highly speculative. Uncertainty exists with respect to the aspects first-tier vs. sub-tier suppliers, legal entities vs. company groups, types of suppliers (e.g. suppliers providing goods and services for the production of aircraft versus suppliers providing HR services, catering) etc.

Figure 7.33 shows a diagram of the process and the results. Please note that this manual exercise was only conducted for the purpose of analysing the results. This manual process is not required to produce supply chain maps. This analysis also deliberately only considers aspects of completeness and ignores false positives. 6,350 arcs were classified as positive (class 1, 2, or 3) and had a source or target node containing the word “Boeing”. All of these arcs were *manually inspected* by the author. During this process, 1,428 of these arcs were found to state a buyer-supplier relation with Boeing being supplied by another organisation. This includes joint ventures and partnerships that appeared to be on the supply side (upstream). Overall, 301 *distinct* organisations were stated to supply Boeing based on the manual inspection. 229 of which appeared to be correctly extracted and classified by the system. If the directionality was incorrectly extracted, a buyer-supplier relation was *not* counted as correctly extracted. The difference between the two numbers is not larger because multiple arcs may state the same supplier. Thus, missing a supplier once can be corrected by extracting it another time.

If it is assumed that Boeing has 13,000 suppliers, then about 1.8% of the suppliers would have been extracted. However, this number has to be interpreted with caution due to the problems stated above. Given the low recall of the classifier in the tests, the underlying data is likely to have mentioned more suppliers. Increasing the recall of the classifier will further increase the number of positive predictions. Simply extrapolating from this number would be misleading. It does not appear likely that the number of extracted unique suppliers will increase linearly with an increase in the size of the data. Instead, some form of saturation curve appears more likely. Additional documents may increasingly tend to restate already known buyer-supplier relations. It appears unknowable what the achievable maximum level of completeness is and what the bottleneck will be once data size or classifier performance are increased or once co-reference resolution has been added.

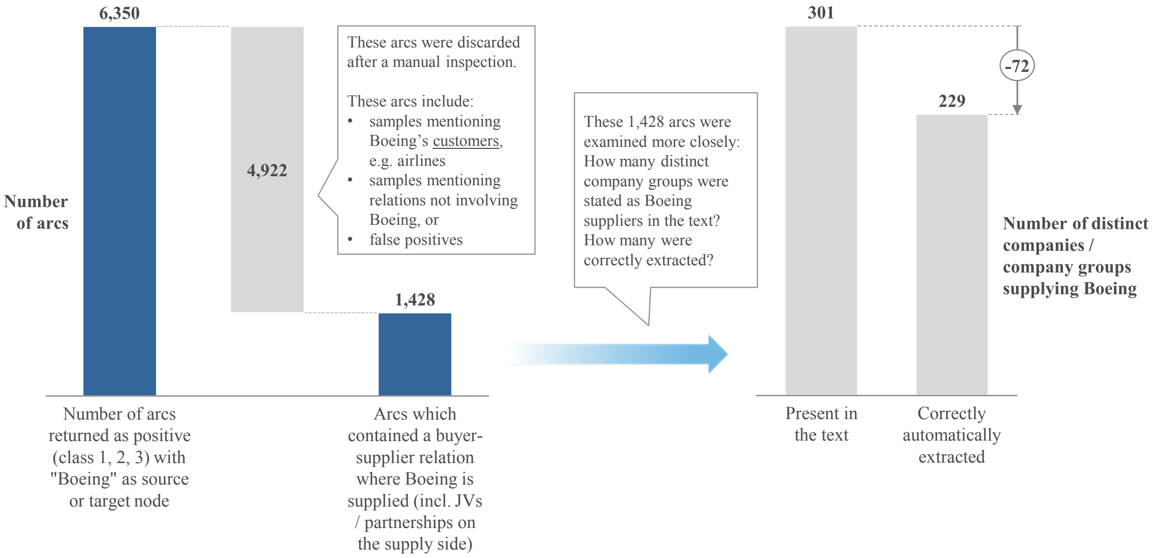


Fig. 7.33 Extracted Boeing suppliers

Discussion of results

Overall, it was possible to correctly extract a large number of direct Boeing suppliers and customers. The relation classification was still imperfect, as the manual inspection of a sample revealed. Approximately 31% of all inspected 150 arcs were fully correctly classified, in a further 27% of the cases buyer-supplier relation class labels (class 1, 2, 3) had been confused.

A major limitation of the extraction performance was the classification of products containing the string “Boeing” as MISC entities. This generally occurred when the context described a Boeing product, such as “Boeing 747-8”. Approximately 40% of all Boeing mentions were classified as such and were consequently lost for subsequent steps. This illustrates the need to either design an NER system that is able to detect organisational named entities within a product name or to extend the relation extraction to also encompass products.

It was possible to correctly extract 229 distinct organisations that appear to be Boeing suppliers. Assuming 13,000 suppliers, then this number corresponds to about 1.8% of Boeing suppliers. This result is still far away from a complete list of first-tier suppliers.

It was possible to follow an iterative approach to identify potential sub-tier suppliers: Honeywell was first identified as a direct supplier to Boeing. Honeywell itself then became the starting point for the next supplier search.

Whereas there is reason to believe that the relation classification performance can further be increased by increasing the size of the training data, a fundamental limitation remains the myopic view of the resulting supply chain maps: The mere chaining of individual buyer-supplier relations does not result in proven sub-tier relations. Instead, only a list of *potential* sub-tier suppliers can be identified at this stage. From Boeing's perspective, it can be insightful to investigate the suppliers of their (direct) suppliers, as done in the case of Honeywell.

The collected examples often contained more valuable information than was extracted, such as product names or the value of contracts. This provides an interesting opportunity to extend the approach to also capture this information.

7.4.6 Automotive manufacturer

For this case problem, the automotive manufacturer Toyota was selected.

Case problem and motivation

Toyota is a Japanese multinational automotive manufacturer with a global supply chain. Past problems with lacking visibility into the supply chain structure have been discussed in detail in Section 4.2.3.

Experimental design

The experimental design was identical to the one used for the case study on Boeing. The unlabelled text corpus was first filtered for sentences containing the word "Toyota". The NLP pipeline, i.a. consisting of NER and relation classifier, was then applied to these pre-filtered sentences to identify Toyota suppliers and customers. Once these have been identified by name, the corpus could again be filtered by those names and the pipeline be applied to them.

Results

Named entity recognition Similar to Boeing, it was expected that not all mentions of Toyota would be detected as organisational named entities. A large fraction of the mentions would be part of product names, such as "Toyota Prius", which were likely to get detected as MISC by the Flair NER system. Figure 7.34 provides an overview of the distribution of class labels for detected named entities containing the word “Toyota”. Around 85% (ca. 50,000) of these entity mentions were classified as ORG. Approximately 14% (ca. 8,000) of these entity mentions were classified as MISC, e.g. mentions of “Toyota Prius” or “Toyota Camry”. Most PER labels (for persons) were incorrect and resulted from a word tokenisation error. Contractions (e.g. apostrophe s) that were not split off correctly misled the NER system into considering “Toyota’s” followed by a person’s name as a person, e.g. “Toyota’s Didier Auriol” was classified as PER.

Fig. 7.34 Distribution of class labels for named entities containing the word "Toyota"

Figure 7.35 provides an analysis of the detected named entity mentions that contained the word “Toyota” and that were classified as organisational. About 88% of mentions are covered by the 12 text sequences. The commonly extracted organisational mentions were generally correct.

Figure 7.36 provides provides the same analysis as before but for instances that were classified as miscellaneous (MISC). The mentions commonly detected as such were products or branded features. Due to the large number of products, the group “OTHERS” accounted for approximately 45% of unique text sequences. As stated before, all these mentions classified as MISC were lost for the further analysis even though they may be useful in identifying buyer-supplier relations. “Toyota Division” should have probably been classified as ORG in all cases.

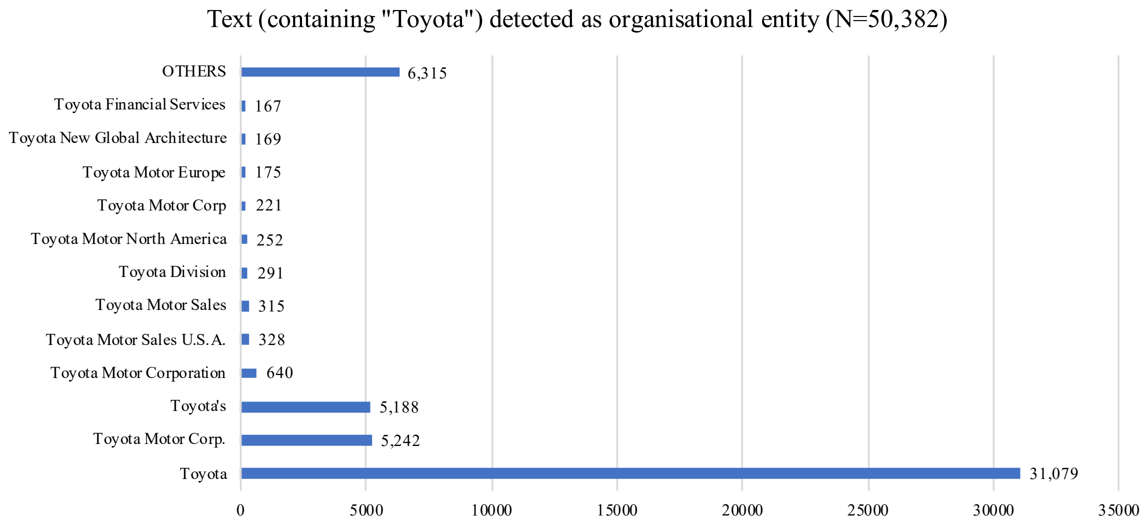


Fig. 7.35 Top unique text sequences containing “Toyota” and classified as organisational named entity

Analysis on sentence-level On a sentence-level, the filtering process is illustrated by Figure 7.37. About 8% of sentences containing the word “Toyota” were predicted to also contain at least one buyer-supplier relation (class 1, 2 or 3).

Analysis on arc-level On an arc-level, the distribution of relation class labels is provided by Figure 7.38. Compared to the Boeing case study, a larger percentage of arcs was predicted to have no relevant relationship (approx. 93% compared to approx. 85% for Boeing). One explanation could be a potential bias towards Boeing or aerospace in the training data.

Figure 7.39 provides an overview of the distribution of confidence scores for the different relation classes. This analysis was conducted for arcs extracted from sentences that contained the word “Toyota”. See Figure 7.17 for a more detailed explanation of the type of analysis.

Visualisation of first-tier relations For the visualisation, no manual corrections were carried out; the result represents what was produced automatically by the processing pipeline. Repeated mentions of identical relations (same named entity mentions, same relation class) were aggregated. The dataset contained 4,829 arcs predicted to be buyer-supplier relationships (class 1, 2, or 3). The result had not been filtered by the confidence score for the named entity or the relation yet. Figure 7.40 shows the resulting network.

Examples of *correctly* extracted customers or suppliers The following examples were correctly processed. The considered pair of organisation named entities is shown in brackets,

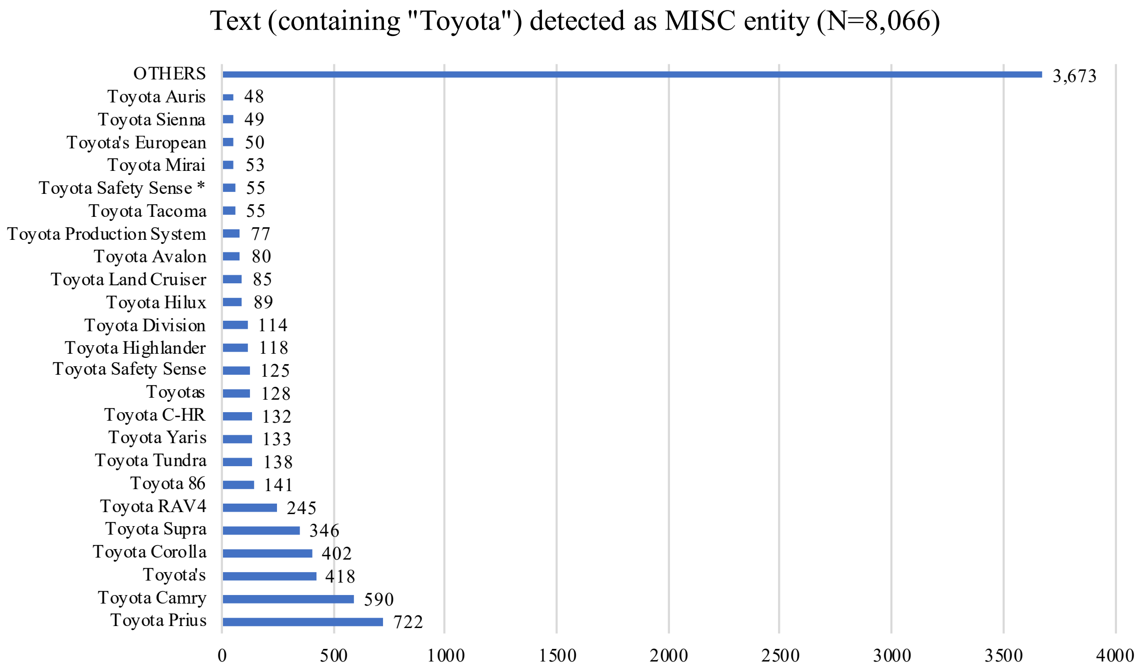


Fig. 7.36 Top unique text sequences containing "Toyota" and classified as MISC named entity

other named entities are shown in curly braces. The organisational named entity mentioned first is referred to as "A", the second one is referred to as "B".

Example sentence
The new affiliate will leverage [Toyota's] existing alliance with [Microsoft Corp.] by using {Microsoft's Azure} cloud technology to expand the carmaker's data management and services capabilities.

The organisational named entities in question were correctly detected. The relation was correctly predicted to be class 3 (ambiguous / partnership / ...) with a confidence of 0.59.

Example sentence
[Toyota Motor Corp.] is adding 1.6 million vehicles to the 3.1 million cars it is already recalling in the U.S. to replace [Takata Corp.] airbag inflators that could explode.

The organisational named entities were correctly extracted. The relation was correctly classified as "B supplies A" with a confidence of 0.53. Product recalls due to faulty parts proved to be occasions where buyer-supplier relations were revealed and could be extracted.

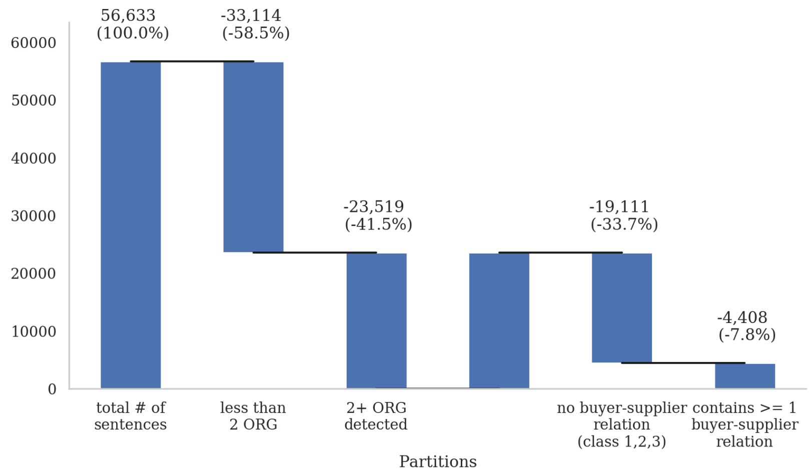


Fig. 7.37 Waterfall on sentence-level (only sentences containing “Toyota”)

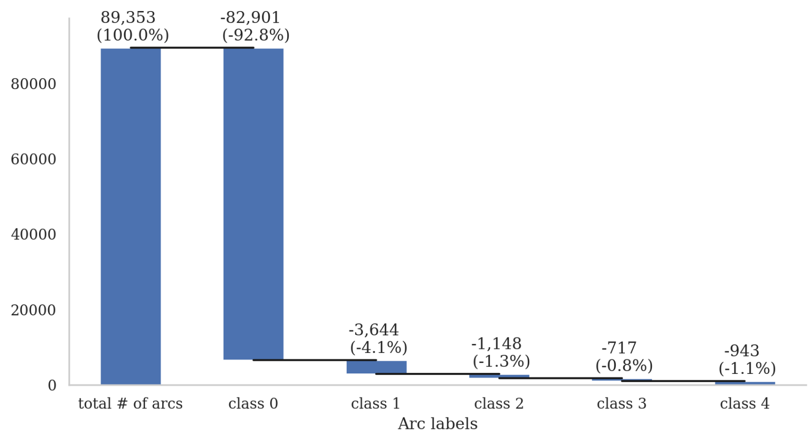


Fig. 7.38 Waterfall on arc-level (extracted from sentences containing “Toyota”)

Examples of *incorrectly* extracted customers or suppliers The following examples were *incorrectly* processed and selected to illustrate specific errors that can occur along the processing pipeline.

Example sentence
[Toyota] and [BMW AG] said earlier this year they would co-develop a midsize sports car to bow several years from now.

The entities were correctly detected. Unfortunately, the relation was incorrectly classified as “B supplies A” with a confidence of 0.33. The term “co-developed” would only justify class 3 (ambiguous / partnership).

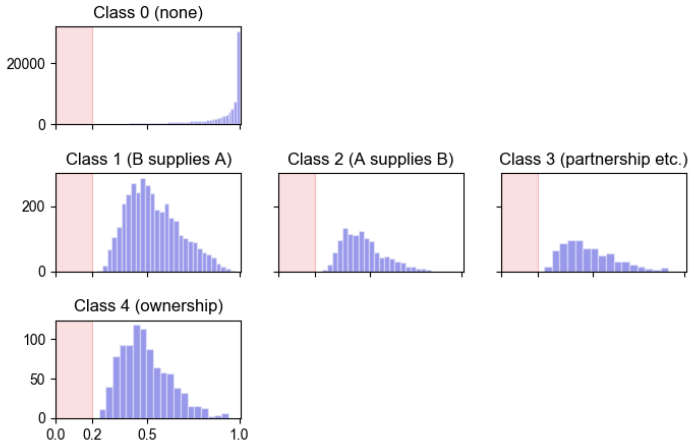


Fig. 7.39 Distribution of confidence scores for "winning" labels (for arcs extracted from sentences containing "Toyota")

Example sentence
[That's] largely due to [Toyota's] recall of 8 million vehicles early this year because of unintended acceleration.

This sentence demonstrates an unfortunate error made during this research. The word tokeniser was configured in such a way that contractions (such as the apostrophe s in "Toyota's") were not split off. However, the NER system Flair was trained on data were this had been done. Consequently, the NER system produced avoidable false positives. Words with a contraction were detected as organisations. Without this processing inconsistency, some of the relation classification errors could have been avoided. This error could not be easily rectified since already the training data at the beginning of the processing pipeline would have required some re-processing. This error also caused the additional name variations due to apostrophe and apostrophe s.

Example sentence
[NAPO] was established to improve local parts sourcing, and serves as a parts distribution network to supply all North American [Toyota] distributors, as well as the export of North American-produced parts to foreign distributors worldwide.

This sentence illustrates how complex and ambiguous sentences can be. This example asks for the relation between NAPO and Toyota. Even though the sentence contained numerous indicators of a buyer-supplier relation, the relation between NAPO and Toyota should have probably been labelled as "none" (class 0).

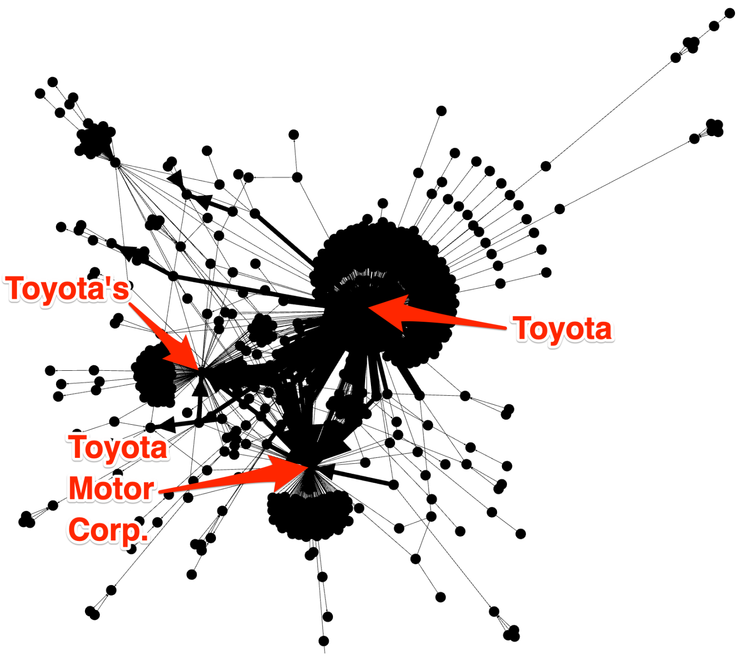


Fig. 7.40 Overall network of supposed Toyota direct customers and suppliers; visualised in Gephi

Manual inspection of a sample of 150 arcs A random sample of 150 arcs was manually inspected. Only about 16% of the assigned class labels were fully correct, compared to the 31% measured in the case of Boeing. One major reason for this lower performance were frequent NER errors. Car models mentioned in the sentences (e.g. Prius, Camry, Tacoma, Fortuner, Corolla etc.) were commonly misclassified as organisations. Once masked as an organisation, the relation classifier was then likely to make subsequent mistakes. The distinction was not always obvious: Lexus, for example, can be considered the luxury vehicle division of Toyota and, thus, correctly be classified as an organisation. This is probably not the case for simple car models, like the Toyota Camry or Toyota Tacoma. Another source of NER errors were named production plants, e.g. "TMMGT" (Toyota Motor Manufacturing de Guanajuato, the name of a Mexican production site), the "Tahara plant" or the "Deeside factory". These or the locations in their names were misclassified as organisations.

Discussion of results

The analyses on the dataset for Toyota provided results in line with the ones generated for Boeing. The dataset contained 4,829 arcs predicted to be buyer-supplier relationships (class 1, 2, or 3). The problem of products containing the name "Toyota" and being classified as

MISC was less prominent. This occurred in only 14% of the cases (compared to 40% of the cases for Boeing). However, the NER system frequently misclassified mentions of car models as organisations, which led to subsequently misclassified relations. This illustrates the need for a robust NER system, which may have to be trained on domain-specific data.

7.4.7 Summary of the end-to-end process validation

This section aims to revisit the objectives (cf. Section 7.4.1) for the validation of the end-to-end processing pipeline. As laid out in the objectives, the end-to-end processing pipeline was applied to convert text into a basic supply chain map, including automatic sentence segmentation and NER tagging. The text dataset consisted of a large number of articles covering a wide range of topics and was predominantly sourced from automotive and aerospace news websites and blogs. The overall approach was successful in the sense that it was possible to automatically generate a basic supply chain map in an end-to-end process from text to visual map.

However, the validation attempts showed that, despite significant progress in supply chain mapping, serious limitations remain especially with respect to misclassification errors, lacking inference of sub-tier relations, data sparsity and general information availability.

Since the classifier had been designed and trained to only extract direct buyer-supplier relations, the resulting map could not show confirmed but only potential sub-tier suppliers – as opposed to a transitive network of relations (cf. the discussion of transitivity in Chapter 6).

Initially, the processing pipeline, including the pre-trained relation classifier, was applied to a random sample of the data. Using the model's predictions, it was possible to derive rough indications regarding the data sparsity. Approximately 3% of sentences were classified as describing one or more buyer-supplier relations.

The complete unlabelled corpus was also filtered for occurrences of the company names "Boeing" and "Toyota". Only these sentences were then processed using an NER system as well as the trained model for extracting buyer-supplier relations.

In the case of Boeing, 6,555 arcs were predicted to describe a buyer-supplier relationship (class 1, 2, 3) with an entity containing the word "Boeing". A manual inspection of 150 arcs for which the classifier predicted a buyer-supplier relation revealed the following: 46 of the 150 arcs (approximately 31%) had been correctly classified, that is the assigned class label 1, 2, or 3 had indeed been the right one. In 63 out of 150 cases (approximately 42%), the correct class label would have been class 0. In all other 41 cases (approximately 27%), a misclassification occurred by confusing the class labels 1, 2, or 3. Common errors were the misclassification of products as organisations which then resulted in the wrong relation class being assigned to the corresponding arcs. Approximately 71% of the sentences (corresponding to the 150 sampled arcs for which a buyer-supplier relation was predicted) did, in fact, describe at least one buyer-supplier relation (class 1, 2, 3) between entities mentioned in the sentence.

In the case of Toyota, only about 16% of the assigned class labels were fully correct as determined during a manual inspection of 150 randomly sampled arcs that were predicted to be buyer-supplier relations. This compares to 31% measured in the case of Boeing. Frequent NER errors contribute to this lower performance. For example, car models (e.g. Prius, Camry, Tacoma, Fortuner, Corolla etc.) mentioned in the sentences were commonly misclassified as organisations.

Information availability, i.e. how much of a company's supply chain is publicly described, cannot be directly measured since it is impossible to manually inspect large text datasets and extrapolating from a sample could be misleading. A manual inspection of the results returned for Boeing resulted in 229 distinct organisations that were apparently correctly identified as Boeing suppliers. This is a small fraction (less than 2% if it is assumed that Boeing has 13,000 suppliers) of Boeing's supplier base. Even though one could increase this fraction by processing more data and improving the performance of NER and relation classifier, the generated supply chain maps are unlikely to be complete.

Nevertheless, it is conceivable that the approach could be usefully applied in practice since it is faster and cheaper than human research and can in principle be run for many different languages. And lastly, even human research may not be able to produce complete and fully accurate supply chain maps and, thus, the practically achievable benchmark is unlikely to be a 100% complete and accurate supply chain map.

7.5 Summary

The aim of this chapter was to apply and validate the approach proposed in Chapter 5 as well as demonstrate some of the challenges identified in Chapter 6.

An initial corpus was created and the inter- and intra-annotator agreement was measured. The achieved inter-annotator agreement of $\kappa = 0.90$ suggested annotations of good quality and this result was considered as validating the overall labelling process. This corpus was subsequently extended by sentences where the organisational named entities, too, had been manually identified. However, these additional sentences were no longer redundantly annotated to measure annotator agreement.

Given the annotated corpus, it was then possible to train and test classifiers on this data. To measure the achievable relation classification performance a number of classification models were compared against baseline classifiers. A linear SVM classifier using positional features, an MLP, and a BiLSTM achieved a micro-averaged F_1 score in the range of 0.73 to 0.77 – well-above the F_1 score of 0.36 achieved by the best dummy baseline classifier in the test. This showed that the newly trained models generally work (as they are better than the baseline classifiers). But it also demonstrated that the F_1 score is near a level where the idea of usefully deploying such a model no longer appears to be outside the realm of possibility. With more training data, the performance of the models is likely to increase further.

The pre-trained BiLSTM model was chosen for the assessment of the end-to-end pipeline that converts large quantities of text into basic supply chain maps. It was possible to generate supply chain maps without the need for manual processing steps. This in itself could be considered a partial validation of the proposed approach. However, especially the addition of the NER processing step introduced further errors, such as the misclassification of products as organisations and the subsequent misclassification of relations. Data sparsity and general information availability remain a problem. If large quantities of data have to be processed to offset data sparsity, then imperfect precision will result in a large number of false positives. In the case of Boeing, only 229, and thus a small fraction of the assumed to be 13,000 direct suppliers, could be extracted from the used dataset.

The next chapter concludes this research by summarising the key results and findings. Furthermore, the limitation of the proposed approaches and a recommendation for future work are discussed.

Chapter 8

Conclusions

8.1 Overview

In the previous chapters, the research conducted as part of this study has been presented and applied to case examples for industrial validation. The purpose of this chapter is to provide the synthesis and conclude this research. In the following sections, the research results are summarised and critically examined from the perspective of the previously identified research gap and stated aim of this research. The research questions are revisited and answered by the key findings of this research. In addition, the academic contribution and industrial value of this research are summarised. This chapter also presents limitations of this research as well as provides recommendations for future research.

8.2 Summary of research

This research investigated the automatic generation of supply chain maps from unstructured, natural language text. The focus of this research was on the extraction of buyer-supplier relations as a fundamental building block and the contribution of the results in this area to the overall problem of automating supply chain mapping. The research was motivated by the problem of limited visibility of the multi-tiered supply chain structure. Knowing supply chain participants and their inter-dependencies across tiers could be beneficial in various use cases, such as detecting supply chain vulnerabilities or avoiding unethically sourced materials. Supply chain mapping is commonly named as an approach to increase visibility into the supply chain structure. However, acquiring the necessary information remains challenging. This research aimed to address this problem by investigating the use of Natural Language Processing and Machine Learning to automatically extract buyer-supplier relations. After reviewing the relevant academic literature in Chapter 2, the research gap and research approach were defined in Chapter 3. This chapter was also used to refine the scope of the research problem. In Chapter 4, the rationale for improving structural supply chain visibility was discussed from an industrial point of view. To gather evidence, a series of semi-structured interviews were conducted and additional case studies from the literature were analysed. Furthermore, potential use cases and beneficiaries were identified and discussed. The analysis confirmed that structural supply chain visibility is, in fact, limited in practice and that there is a need for improving it. The subsequent Chapter 5 focussed on the core problem of classifying individual buyer-supplier relations between two organisational named entities. This problem was split into two stages. (1) A labelled text corpus had to be created using human annotations. (2) A classifier needed to be designed as well as then trained and tested on the corpus. Chapter 6 widened the scope again to consider the end-to-end processing pipeline from the pre-processing of documents over the extraction of individual buyer-supplier relations to the visualisation of a supply chain map. Additional requirements for a complete supply chain mapping process were derived from the characteristics of supply chains as well as from the limited information quality. In Chapter 7, the proposed approach was executed and applied to validation case studies. This provided results for the achievable inter-annotator agreement and classification performance.

8.3 Revisiting research questions

This section aims to revisit the research questions and provides a summary of the corresponding research results. To help with readability, a high-level plain English summary in bullet point form is provided first, followed by a more precise summary of the details that assumes more background knowledge.

Research question 1 (RQ1):
To what extent and how can the extraction of buyer-supplier relations from unstructured text be automated?

The research question was further broken down into the following sub-questions:

(1a) *What is the achievable inter-annotator agreement of humans attempting to classify buyer-supplier relations?*

(1b) *What is a suitable approach for the classification of buyer-supplier relations and what classification performance can be achieved?*

Summary (RQ1):

- Human annotators were asked to manually and independently label relations between any two organisations mentioned in an English sentence. The classes to choose from were: (1) “A supplies B”, and (2) “A is supplied by B” (inverted direction), (3) undirected or ambiguous buyer-supplier relation, (4) part-of or ownership relation, (5) no relevant relation.

- On this task, the human annotators achieved a high level of agreement ($\kappa = 0.9$). To remind the reader, an inter-annotator agreement of $\kappa = 1$ indicates perfect agreement, whereas a value of $\kappa = 0$ indicates no agreement beyond what would be expected by chance.

- The high level of agreement suggests that the labelling was feasible given the chosen task design and that the obtained annotations were of good quality. Low agreement levels, closer to random chance, would have indicated a problem with the task and it would not have been sensible to continue with using the obtained dataset for benchmarking or training a classifier.

- A suitable approach for the classification of buyer-supplier relations was found to be supervised learning. A selection of such classification algorithms was tested and benchmarked against two dummy classifiers: one that chooses classes randomly and a stratified one that chooses classes randomly but in proportion to the class sizes.

- The trained classifiers achieved F_1 scores in the range of 0.66 to 0.77. They all thereby clearly outperformed the dummy classifiers, which only achieved F_1 scores of 0.2 (random classifier) and 0.36 (stratified). To remind the reader, classification performance is commonly measured using the F_1 score. This score combines recall and precision to equal parts into a single metric. Recall is the ability of a classifier to find all the positive samples, whereas precision is its ability to avoid false positives. The best possible F_1 score is 1, the worst is 0.

- On the one hand, the achieved F_1 score suggests that the approach works. On the other hand, it is questionable if the classification performance achieved in this research would already suffice for use in an industrial setting. The performance levels reported in this thesis should not be understood as the theoretically achievable maximum but can be further increased by a larger training dataset and by making use of more recent developments in NLP, such as transformers.

A fixed set of 40 sentences had to be labelled by all annotators at the beginning of their first labelling session. On this inter-annotator set, the achieved agreement across all annotators was $\kappa = 0.9$ (Cohen's Kappa) and $\kappa = 0.9$ (Fleiss' Kappa). When extended to all sentences a pair of annotators had processed, the average of these pairwise agreements was slightly lower, at $\kappa = 0.8$ (Cohen's Kappa). In addition to the *inter*-annotator agreement, the *intra*-annotator agreement was also measured at $\kappa = 0.86$ (Cohen's Kappa). Even though agreement measures are not standardised, this appears to suggest annotations of good quality. Furthermore, if seen as a proxy, the high level of inter-annotator agreement also validates the overall labelling process including the class definitions, instructions, selected annotators and the developed labelling tools. Even though the inter-annotator agreement was not only tested on the inter-annotator set of 40 sentences (across all annotators) but also on all sentences that any two annotators had both labelled, the achievable inter-annotator agreement may be different in other circumstances, such as with different annotators or text samples. Nevertheless, this appears to be evidence that the labelling task itself is feasible given the class definitions, instructions, and provided labelling tools.

To identify suitable approaches for the classification of buyer-supplier relations, multiple options were considered. A fully unsupervised approach was deemed insufficient for the problem at hand due to the required directionality of buyer-supplier relations. Given the intractably large number of ways how buyer-supplier relations can be expressed in natural language, an approach using manually pre-defined lexico-syntactic patterns was also ruled out. Generally, a suitable approach for the classification of buyer-supplier relations is likely to be found among those classification models that can learn from the data[1]. Semi-supervised approaches were considered problematic as well, i.a. due to semantic drift. Thus, it was concluded to use a supervised learning approach with the major drawback of requiring a large dataset of annotated sentences. Different algorithms and data representations were tested, including an SVM, an MLP and a BiLSTM. In all cases, the actual organisational named entities needed to be masked. Three types of masks were used: one mask each for the two organisational named entities in question, and one "other entity" mask for all other organisational named entities not in question but mentioned in the sentence. The SVM was tested in two configurations. The first configuration was a simple bag-of-words approach that ignores word order. The second configuration also used a bag-of-words approach but at least also considered the position of the two organisational named entities in question. By definition, the BiLSTM model considers the sequence of text.

[1] Early experiments with hard-coded lexico-syntactic patterns that were conducted as part of this research proved to be inappropriate for the task given the wide variety of ways a buyer-supplier relation can be expressed in natural language.

When tested on the labelled corpus, all algorithms clearly outperformed the base line classifiers. A random dummy classifier achieved a micro-averaged F_1 score of 0.20, compared to 0.36 achieved by a stratified dummy classifier that considers class sizes. The simple bag-of-words SVM achieved a micro-averaged F_1 score of 0.68 on the initial corpus and 0.66 on the extended corpus. The other algorithms (SVM considering the named entity positions, MLP, and BiLSTM) achieved micro-averaged F_1 scores of 0.73 to 0.77. The SVM classifier that also considered the positions of named entity mentions performed best in the test. However, the differences between this model and the MLP or BiLSTM were too small to be considered significant.

For some individual classes, the achieved F_1 score was still only around 0.3. Better classification performance is achievable, i.a. by increasing the size of the training corpus. Since the training data had already been NER-tagged, NER errors were not included in these performance measurements. If applied to a previously unseen dataset, additional error sources had to be considered that will reduce the achievable performance, such as errors that occurred during sentence segmentation or named entity recognition. The examined models were relatively confident with regards to identifying that no buyer-supplier relation was present (as is evident from the distribution of confidence scores for the class 0). However, all other classes were still frequently confused due to biases in the training dataset and its still rather small size. In summary, research question RQ1 can be answered as follows: Since human annotators were able to consistently classify buyer-supplier relations (inter-annotator agreement of 0.9), creating a high-quality labelled text corpus and automating this classification process is not immediately beyond the realms of possibility. Using a supervised learning approach, it is, in fact, possible to train a classifier to extract individual buyer-supplier relations. The achieved classification performance of 0.73 to 0.77 (micro-averaged F_1 scores; averaged across 10 runs) is significantly better than a stratified dummy classifier (0.36). False positives, in particular, still limit the usefulness of the trained classifiers.

Research Question 2 (RQ2):

What are the challenges in developing a suitable end-to-end approach for automating the generation of supply chain maps from unstructured text?

Summary (RQ2):

- It was possible to design and test an end-to-end approach to convert natural language text into basic supply chain maps. The approach is scalable, fast and cost-efficient. In one of the experiments, 229 apparent Boeing suppliers could be correctly extracted.
- However, the validation case studies also demonstrated a number of serious limitations. Some of these limitations may be overcome by more and better training data or better classification algorithms. Other limitations are more systemic and would also apply to human researchers attempting to build supply chain maps from reading news articles. The latter limitations are, thus, much harder to overcome.
- Key *short-term* limitations of the approach:
  - The detection of organisations and the classification of relations between them still suffer from relative high error rates. Imperfect precision combined with large quantities of text that are being processed results in a large number of false positives. This can be addressed by more training data, a database of companies and products to help with disambiguation, and improved classification algorithms.
  - Limited to English language. This can be addressed by creating additional corpora for other languages of interest.
  - Limited to relations between two organisation. Provided products or services currently do not get extracted. This can be addressed by obtaining the corresponding annotations and retraining the classifiers.
- Key *systemic* limitations of the approach:
  - Limited information availability: It is unlikely that all existing buyer-supplier relations of all organisations in the world have been expressed in publicly accessible text documents.
  - It is hard to establish the correct tier of a company in a supply chain from processing text documents. Is a company a direct supplier or a sub-tier supplier? If it is a sub-tier supplier, via which other companies is the product passed along the supply chain? Only in rare cases is this explicitly stated. Chaining together individual buyer-supplier relations can only reveal *potential* sub-tier suppliers ("transitivity problem").
  - Absence of evidence is not evidence of absence. The approach cannot be used to conclude which organisation is *not* supplying another company (unless explicitly stated in a text).

Challenges in developing a suitable overall approach arise from the characteristics of supply chains as well as the limited information quality (including quantity).

This research project focussed on the extraction of buyer-supplier relations between two organisations. However, it may be crucial to understand the geographical context as well as to identify which part, material or service was provided. The same company can take multiple roles within the same supply chain. In those cases, extracted information may seem to be contradictory. Supply chains are dynamic and information about buyer-supplier relations may get outdated. Due to the global nature of supply chains, information about supply chains is distributed across the globe and in various languages. An overall approach would need to be able to extract information in these different languages. Another challenge is the separation of primary from secondary functions within the supply chain: the supply with parts required for production may be more relevant for a supply risk management use case than accounting services provided to a company. Linguistically, these buyer-supplier relations may look similar. Logistics providers play a special role in supply chains. They need to be considered as they may introduce disruptions but they may be quickly switched for alternative logistics options.

Challenges introduced by the limited information quality include the fact that supply chain maps will likely remain incomplete or inaccurate. Not all the information will be available or correct and, thus, conclusions have to consider that some information may just be lacking or incorrect. A further challenge is that news reports are more frequently stating new or existing buyer-supplier relations than non-existing or recently ended buyer-supplier relations. An industry-specific information half-life model could address this problem. If a buyer-supplier relation is mentioned more frequently, this does not imply that this relation is particularly important, e.g. from a risk management perspective. This means that an extracted relation with low confidence due to only few occurrences in the data cannot simply be discarded. A factor that adds further complexity is that many characteristics, such as supply chain dynamics, are company- or industry-specific. Moreover, information may be provided on different aggregation levels. The statement that a company has 100 suppliers in a particular region could be useful information. This study focussed on individual, self-contained sentences. Some information, however, can only be extracted by inferring from a larger context. The individual pieces of information may not even explicitly talk about supply relations but, in combination, an inference may be possible. A major challenge is to establish the correct tier of a company in a supply chain. This includes the "transitivity problem": Chaining together individual buyer-supplier relations can only reveal *potential* sub-tier suppliers. Ambiguity in natural language is a general limitation: This includes ambiguous entity names as well as ambiguous language used to describe buyer-supplier relations.

A useful structure for addressing these challenges may be the proposed conceptual framework which consists of seven stacked abstraction layers along the end-to-end process from text collection to supply chain map visualisation and decision support.

The attempt to validate the proposed end-to-end approach on a large dataset of real-world news showed that, in principle, it is possible to automatically generate rudimentary supply chain maps from text. This is possible on a large scale, fast and cost-efficiently.

However, the validation attempt also showed a number of serious limitations. Some of these limitations might be overcome by more and better training data or better classification algorithms, whereas other limitations are more systemic.

The processing pipeline still suffered from relatively high error rates, mainly in the NER tagging and relation extraction steps. This resulted in a large number of false positives. In the Boeing case, for example, a manual inspection of 150 arcs for which the classifier predicted a buyer-supplier relation revealed the following: 46 of the 150 arcs (approximately 31%) had been correctly classified, that is the assigned class label 1, 2, or 3 had indeed been the right one. This can be interpreted as the precision in this inspection. In 63 out of 150 cases (approximately 42%), the correct class label would have been class 0. In all other 41 cases (approximately 27%), a misclassification occurred by confusing the class labels 1, 2, or 3. Common errors were the misclassification of products as organisations which then resulted in the wrong relation class being assigned to the corresponding arcs.

The combination of imperfect precision and data sparsity is another challenge. Without having applied co-reference resolution, around 85% of sentences got discarded since the NER system alone did not detect at least two organisational named entities. And around 3% of sentences were classified to contain at least one predicted buyer-supplier relation.

Another remaining limitation is the fraction of a supply chain that is actually described in publicly available text datasets, such as news reports. As a proxy, the returned results for Boeing were manually inspected to at least obtain a rough estimate of the share of Boeing suppliers that were correctly extracted from the data. 6,350 arcs were manually inspected and 1,428 of these were found to state a buyer-supplier relation in which Boeing is supplied by another organisation. This included joint ventures and partnerships on the supply side. In a further step, these arcs were analysed again to extract distinct Boeing suppliers. 301 were found to be present in the text, 229 of these got correctly extracted by the system. Assuming a total number of 13,000 Boeing suppliers, this would correspond to about 1.8% of Boeing suppliers. Even though performance improvements appear possible, it is likely that supply chain maps generated by the automated approach based on publicly available datasets will not be complete. However, as a counterargument one could bring forward that supply chain maps generated by humans based on publicly available data will also likely be incomplete. An automated approach would at least provide the benefit of being fast, cost-efficient and, thus, scalable.

Conclusions regarding the data sparsity and information availability have to be made with caution since both can only be observed through the “eyes” of an imperfect processing pipeline applied to a large but still limited amount of data.

In summary, research question RQ2 can be answered as follows: Even though it was possible to design and test a simple end-to-end approach to convert natural language text into basic supply chain maps, remaining challenges are manifold. For example, establishing the correct tier of a company in a supply chain is difficult given the limited availability of information and the ambiguity in natural language. A further challenge is that the combination of the classifiers’ imperfect precision and data sparsity leads to a large number of false positives. However, an automated approach can provide the benefit of being fast, cost-efficient, multi-lingual in principle and, thus, highly scalable.

8.4 Contribution of the thesis

8.4.1 Academic contribution

This study has made the following key contributions to academic research:

A methodology for automatically extracting individual buyer-supplier relations The methodology covered corpus creation, relation extraction as well as applying a trained model to an unlabelled corpus. The idea of using supervised learning to automatically extract relations between two named entities is not novel. However, the application to the problem of supply chain mapping, specifically the extraction of buyer-supplier relations, required the careful design of appropriate class definitions – weighing up the expressive power of the chosen set of classes and the labelling simplicity. Using the proposed relation classes, directed and undirected buyer-supplier relations can be captured. To avoid confusion with buyer-supplier relations, a separate ownership relation was defined. Following this methodology, buyer-supplier relations can be extracted from large quantities of unlabelled text.

Estimating achievable inter- and intra-annotator agreement Before this research study, achievable levels of inter- and intra-annotator agreement for the extraction of buyer-supplier relations had not been established. Given the proposed class definitions and the provided labelling instructions, an inter-annotator agreement of $\kappa = 0.90$ across all annotators could be achieved[2]. Despite remaining ambiguity in the task instructions as well as ambiguity in natural language, the achieved annotator agreement is high and suggests that the task could potentially be automated.

Estimating achievable classification performance By training and testing various classification algorithms on the labelled text corpus, a lower bound on the achievable classification performance could be established. When tested on the labelled corpus, all algorithms clearly outperformed the base line classifiers. A random dummy classifier achieved a micro-averaged F_1 score of 0.20, compared to 0.36 achieved by a stratified dummy classifier that considers class sizes. The simple bag-of-words SVM achieved a micro-averaged F_1 score of 0.68 on the initial corpus and 0.66 on the extended corpus. The other algorithms (SVM considering the named entity positions, MLP, and BiLSTM) achieved micro-averaged F_1 scores of 0.73 to 0.77. By increasing the size of the labelled text corpus, the performance is expected to increase further.

Identified challenges of automating the end-to-end process of supply chain mapping from text In advance of this study, it was unclear which challenges needed to be addressed to automate the end-to-end process of supply chain mapping from text. Future research could be designed to explicitly address these challenges, such as the inference of sub-tier relations.

8.4.2 Industrial contribution and implications

The proposed approach for classifying buyer-supplier relations can be applied to process large amounts of previously unseen text data to extract buyer-supplier relations or at least identify sentences that are likely to describe buyer-supplier relations. This knowledge, in turn, can be a valuable input in decision-making processes, such as risk mitigation and contingency scenarios. There are multiple ways in which the proposed automation can prove to be useful to a company or other actors. The automated approach may be cheaper or faster than collecting the information by other means, and it may discover previously unknown parts of the supply chain structure. The approach could be used to quickly generate a first draft of a supply chain map or may be used to augment or validate an existing map.

[2]On the inter-annotator set across all annotators, the achieved agreement was 0.9 (Cohen's Kappa) and 0.9 (Fleiss' Kappa). The averaged pairwise inter-annotator agreement on all sentences that two annotators had both labelled was 0.8 (Cohen's Kappa)

8.5 Limitations

The following main limitations should be noted regarding the methodology, proposed approach and the results of this study:

Small training dataset Since text had to be manually annotated, the resulting training dataset is still relatively small for a problem in the domain of NLP.

Difficulty of measuring what has been missed Given the large quantities of text that are being processed and the sparsity of relevant information in that text, it is difficult to measure what supply chain information did exist but has been missed by the approach. A manual inspection can only be performed on small text samples but never exhaustively.

Veracity of statements not scrutinised Testing the veracity of statements was deliberately excluded from the scope. If a sentence stated that a company supplies another company, this had to be taken at face value. This is less of a problem if the news source is reputable but can become a problem if rumours or intentional misinformation is picked up.

Information availability Even if there was a perfect relation extraction model, its success would still remain limited by the availability of written information about a particular company and its supplier relations. In the case of less (media-)prominent companies, this is likely to be the limiting factor.

Limited relation classification performance The relation classification performance achieved during this research study is still limited. This also means that any conclusions based on these predictions can only be regarded as indicative. For instance, conclusions drawn about the availability of relevant information in a large set of text documents can only be indicative since the classifiers used to extract the relevant information are still far from perfect.

Difficulty of establishing the correct tier and the "transitivity problem" Establishing the correct tier of a company within a supply chain is difficult. In particular, it is difficult to establish sub-tier relations that would make a supply chain map a graph of transitive buyer-supplier relations. This limitation is not only a result of using triples (entity1, entity2, relation). It is a more general limitation due to the fact that humans rarely discuss supply chains across more than three tiers.

English language only Even though the proposed approach could be extended to other languages, the generated classification models were trained and tested on English text only.

Limited scope of relation extraction Not all expressed buyer-supplier relations were captured by the proposed approach. For the sake of simplicity, only single sentences and relations between explicitly mentioned organisations were considered. No co-reference resolution was performed and not all relevant types of relations were considered, e.g. company-to-product relations were not extracted even though they would allow for relevant inferences. Furthermore, the material flow was not mapped between physical locations but rather between organisations only.

8.6 Conclusions and recommendations for future research

In summary, to address the problem of limited visibility of extended supply chain structures, it was proposed to automate the extraction of supply chain maps from natural language text. A fundamental building block for this approach is the extraction of individual buyer-supplier relations between two organisations from natural language text. The results have shown that it is, in principle, possible to automatically extract a large number of buyer-supplier relations and visualise these in form of basic supply chain maps. The performance of the proposed approach is still limited by both the Named Entity Recognition as well as the relation classification. The achieved performance of the classifier and the end-to-end approach was not sufficient to fully replace manual efforts. And the availability of large and rich datasets will ultimately remain a limiting factor for the success of automated mapping approaches. That said, the performance of the classifier can be improved by increasing the size of the training dataset and named entity recognition and entity linking solutions will continue to get better. The proposed approach may be useful in situations where a first rudimentary supply chain map needs to be created or an existing supply chain map needs to be checked for completeness.

Taking a step back from the proposed approach, the following conclusions can be drawn from the overall work:

- The problem is real: Initially, the work was motivated by the specific problem of a particular aerospace manufacturer attempting to better manage supplier risk. The work has shown that the problem of limited visibility into the multi-tiered structure of a supply chain is experienced by *many* companies and has relevance *far beyond the realm of supplier risk management*. Companies and other entities, such as governmental agencies, desire to better understand extended supply chain structures and spend significant amounts of resources towards that end. However, a satisfying solution for obtaining the required information still does not appear to exist.

- Relevant data is publicly available: During the labelling exercises, thousands of sentences had to be carefully examined. This work revealed that there is a wealth of information relevant for understanding supply chain structures contained in openly available text documents. This wealth will largely remain untapped unless information can be extracted automatically, on a massive scale and ideally in all languages. It is unlikely that all relevant data is publicly available – but extracting the part that is can already provide a benefit.
- A first stab at the problem is easy: Automatically extracting *some* buyer-supplier relations is not difficult. However, if recall and precision are to be improved, the nuances and the ambiguity in natural languages require a much larger number of carefully labelled example sentences and potentially databases of companies and products to provide the necessary background information.

Recommendations for future research can be derived from the challenges identified in Chapter 6. With respect to challenges resulting from supply chain characteristics, the following recommendations can be made:

- Augmenting the knowledge graph with geolocations could be beneficial. Geolocations can be extracted from the input text (if mentioned). Or they can be obtained by linking entity mentions to public ontologies, e.g. to obtain at least the headquarters' geolocation. Once this is done, existing risk profiles of geolocations or companies can be used to augment the supply chain maps.
- A further direction for future extension is to capture more complex relations or more attributes of a buyer-supplier relation. An obvious next step would be to capture provided parts, material or services. This can be done with ternary relations ("who supplies whom with what") but it requires a classifier specifically trained to detect provided parts, material or services. Using the typical bill of materials (BoM) of a product, one could attempt to systematically probe for suppliers of specific parts or check a generated supply chain map for completeness. Other useful attributes of a buyer-supplier relation are the value or the duration of the contract or any quantification of the provided goods or services. Instead of learning the relationship and the directionality in one process, it could be beneficial to separate these: to first detect if

there is a buyer-supplier relation and then to detect the directionality as an attribute of that relation.

- The labelling system could be extended to capture changes in the supply chain structure, such as when a supplier is dropped or replaced by another one.

- This research only used text in the English language. By incorporating the ability to process documents in other languages, one should be able to generate more complete supply chain maps. A first simple approach would be to use one of the publicly available translation models to translate the already existing training dataset into another widely spoken language and train a new model on this translated corpus. The automatic translation may introduce some errors but it would not require any manual labelling. The labels from the English dataset could be re-used.

- The restriction to only consider organisational entities could be relaxed to also allow for countries that act as buyers, e.g. in the case of military products.

With respect to challenges resulting from limited information quality, there are also some promising directions to extend the research:

- Future research could extend the proposed approach to capture more relations by employing co-reference resolution. Instead of requiring the organisation to be mentioned by name, it could then be referred to by a pronoun or a more abstract description like "the carmaker".

- The research could also be extended to capture information on higher aggregation levels, such as "Company A has 150 suppliers in Florida". This information cannot be easily integrated into the knowledge graph of company-to-company relations but it could help with checking the collected information for completeness and correctness.

- An obvious extension is to capture explicitly stated sub-tier relations as ternary relations if they are mentioned. This would address (but not fully solve) the "transitivity problem".

- An in-depth analysis of the resulting knowledge graph was not within the scope of this study. Once the performance of the NER and the relation classifier has been further improved, analysing these large global supplier networks using the tools of network science could be insightful. In addition, it appears useful to further understand which company- or industry-specific factors influence the availability of buyer-supplier information. This could, for example, take the form of a supervised prediction model.

> Given a rudimentary supply chain map, it may also be possible to infer if one company is a sub-tier supplier to another company. It may require additional market information, such as the market share of a supplier, to estimate the probability of a sub-tier relation.

In addition to the recommendations related to the identified challenges, exciting research opportunities have opened up thanks to the advances in Machine Learning and NLP in recent years. Future research should consider the more recent transformers and contextual embeddings as promising alternatives to the classification models and word embeddings used in this research. Even more recent NLP developments, such as large pre-trained language models provided by Google, Facebook or OpenAI, were not considered for this research. In early tests by the author, these language models were able to complete sentences, such as "GKN Aerospace is a supplier to" with credible text one would expect in a news report. The predicted completion may state realistic customers or suppliers that were learned from the training data. However, it may not be possible to assess the truthfulness of the predicted completion or to trace back where information was derived from. Completions may look credible (which is the whole idea of language models) but could, in fact, be incorrect. Figure 8.1 shows an example[3] of an auto-completed sentence. In the predicted completion, customers are explicitly named. It is possible that these self-supervised models could support the supervised relation extraction process. The GPT-3 language model published by OpenAI shortly before the submission of this thesis could be another promising path to explore in the attempt to discover supply chain structures.

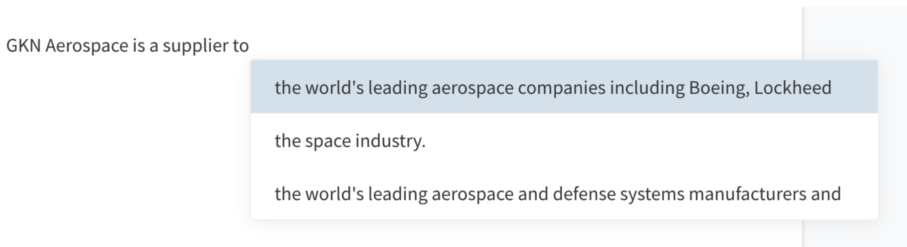


Fig. 8.1 Auto-completion using transformers reveals the names of customers; source: Huggingface

[3] Used online demo: `https://transformer.huggingface.co/`; last accessed: 2019-09-25

References

S Abiteboul. Querying Semi-Structured Data. Technical Report 1996-19, Stanford InfoLab, 1996. URL http://ilpubs.stanford.edu:8090/144/.

Achilles Group. Supply Chain Mapping, 2013. URL http://www.achilles.co.uk/images/pdf/communitypdfs/Automotive/Achilles-Supply-Chain-Mapping.pdf.

Debby Arkell. The evolution of creation, 2005. URL https://www.boeing.com/news/frontiers/archive/2005/march/mainfeature1.html.

Charles Arthur. Thailand's devastating floods are hitting PC hard drive supplies, warn analysts, 2011. URL https://www.theguardian.com/technology/2011/oct/25/thailand-floods-hard-drive-shortage.

Myo Min Aung and Yoon Seok Chang. Traceability in a food supply chain: Safety and quality perspectives. *Food Control*, 39(1):172–184, 2014. ISSN 09567135. doi: 10.1016/j.foodcont.2013.11.007. URL http://dx.doi.org/10.1016/j.foodcont.2013.11.007.

Steve Banker. Supply Chain Operating Networks are Finally Beginning to Emerge, 2016. URL https://logisticsviewpoints.com/2016/05/24/supply-chain-operating-networks-are-finally-beginning-to-emerge/.

Michele Banko, Michael J. Cafarella, Stephen Soderland, Matt Broadhead, and Oren Etzioni. Open information extraction from the web. *IJCAI International Joint Conference on Artificial Intelligence*, pages 2670–2676, 2007. ISSN 10450823.

Nadav Bar. Training a Machine Learning Classifier for Relation Extraction from Medical Literature, 2016. URL https://www.microsoft.com/developerblog/2016/09/13/training-a-classifier-for-relation-extraction-from-medical-literature/.

Mark Barratt and Adegoke Oke. Antecedents of supply chain visibility in retail supply chains: A resource-based theory perspective. *Journal of Operations Management*, 25(6): 1217–1233, 2007. ISSN 02726963. doi: 10.1016/j.jom.2007.01.003.

Rahul C. Basole and Marcus A. Bellamy. Supply Network Structure, Visibility, and Risk Diffusion: A Computational Approach. *Decision Sciences*, 45(4):753–789, aug 2014. ISSN 00117315. doi: 10.1111/deci.12099.

BBC. Toyota profits plunge after earthquake disruption (2011-08-02), 2011. URL https://www.bbc.co.uk/news/business-14369137.

Ulrich Berbner. *Situationsadäquate Gestaltung des Supply-Chain-Risikomanagments unter Berücksichtigung des Informationstechnologieeinsatzes: Ein konfigurationstheoretischer Ansatz zur Auswahl adäquater Strukukturtypen*. PhD thesis, Darmstadt Technical University, Department of Business Administration, Economics and Law, Institute for Business Studies (BWL), 2017.

Chris Biemann. Ontology Learning from Text: A Survey of Methods. *GLDV-Journal for Computational Linguistics and Language Technology*, 20(2):75–93, 2005.

Bloomberg. Predictive modeling with Bloomberg's Supply Chain data., 2018. URL `https://data.bloomberglp.com/professional/sites/10/233552_CDS_REF_SupplyChain_CASE_DIG-2.pdf`.

Boeing. Boeing Honors Suppliers for Outstanding Performance, 2018. URL `https://boeing.mediaroom.com/2018-04-12-Boeing-Honors-Suppliers-for-Outstanding-Performance`.

Bernhard E Boser, Isabelle M Guyon, and Vladimir N Vapnik. A training algorithm for optimal margin classifiers. In *Proceedings of the fifth annual workshop on Computational learning theory*, pages 144–152. ACM, 1992.

Neal E. Boudette and Jeff Bennett. Pigment Shortage Hits Auto Makers (2011-03-26), 2011. URL `https://www.wsj.com/articles/SB10001424052748703696704576222990521120106`.

Alexandra Brintrup. Behaviour adaptation in the multi-agent, multi-objective and multi-role supply chain. *Computers in Industry*, 61(7):636–645, 2010.

Paul Buitelaar, Philipp Cimiano, and Bernardo Magnini. *Ontology Learning from Text : An Overview*. IOS Press, Amsterdam, 2005. ISBN 978-1-58603-523-5. doi: 10.1.1.70.3041.

Business Continuity Institute. Supply Chain Resilience 2014 Annual Survey, 2014. URL `http://www.bcifiles.com/supply-chain.pdf`.

Adam Carey. Thailand floods to put the grinch in Christmas (2011-10-29), 2011. URL `https://www.theage.com.au/national/thailand-floods-to-put-the-grinch-in-christmas-20111028-1mo4a.html`.

Maria Caridi, Luca Crippa, Alessandro Perego, Andrea Sianesi, and Angela Tumino. Do virtuality and complexity affect supply chain visibility? *International Journal of Production Economics*, 127(2):372–383, 2010.

Kyunghyun Cho, Bart Van Merriënboer, Dzmitry Bahdanau, and Yoshua Bengio. On the properties of neural machine translation: Encoder-decoder approaches. *arXiv preprint arXiv:1409.1259*, 2014.

Thomas Y. Choi and Yunsook Hong. Unveiling the structure of supply networks: Case studies in Honda, Acura, and DaimlerChrysler. *Journal of Operations Management*, 20 (5):469–493, 2002. ISSN 02726963. doi: 10.1016/S0272-6963(02)00025-6.

Francois Chollet. *Deep Learning with Python*. Manning Publications Co., 1 edition, 2017.

Sunil Chopra and ManMohan S. Sodhi. Managing risk to avoid supply-chain breakdown. *MIT Sloan management review*, 46(53-61), 2004. ISSN 0957-4093. doi: 10.1108/IJOPM-10-2012-0449.

Gobinda G. Chowdhury. Natural language processing. *Annual Review of Information Science and Technology*, 37:51–89, 2005. ISSN 00664200. doi: 10.1002/aris.1440370103.

Carl F Christ. A review of input-output analysis. In *Input-output analysis: An appraisal*, pages 137–182. Princeton University Press, 1955.

Martin Christopher and Helen Peck. Building the resilient supply chain. *International Journal of Logistics Management*, 15(2):1–13, 2004. ISSN 0957-4093. doi: 10.1108/09574090410700275. URL `http://dx.doi.org/10.1108/09574090410700275`.

Junyoung Chung, Caglar Gulcehre, KyungHyun Cho, and Yoshua Bengio. Empirical evaluation of gated recurrent neural networks on sequence modeling. *arXiv preprint arXiv:1412.3555*, 2014.

James Clark. Automotive News Europe, 2013. URL `https://europe.autonews.com/article/20130927/CUTAWAY01/309279999/trw-helps-bmw-3-series-earn-top-safety-rating`.

Jacob Cohen. A coefficient of agreement for nominal scales. *Educational and psychological measurement*, 20(1):37–46, 1960.

Emma Cosgrove. Inside Boeing's digital supply chain turnaround, 2018. URL `https://www.supplychaindive.com/news/Boeing-digital-supply-chain-transformation/538630/`.

Xin Dong, Evgeniy Gabrilovich, Geremy Heitz, Wilko Horn, Ni Lao, Kevin Murphy, Thomas Strohmann, Shaohua Sun, and Wei Zhang. Knowledge vault: A web-scale approach to probabilistic knowledge fusion. In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 601–610. ACM, 2014.

Evonik. Fire at the CDT plant on the grounds of the Marl Chemicals Park, 2012. URL `https://corporate.evonik.com/en/Pages/article.aspx?articleId=103929`.

M. Theodore Farris. Solutions to strategic supply chain mapping issues. *International Journal of Physical Distribution & Logistics Management*, 40(3):164–180, 2010. ISSN 0960-0035. doi: 10.1108/09600031011035074.

Stanley E Fawcett and Gregory M Magnan. The rhetoric and reality of supply chain integration. *International Journal of Physical Distribution & Logistics Management*, 32(5):339, 2002. ISSN 09600035. doi: 10.1108/09600030210436222.

Charles H Fine. *Clockspeed: Winning industry control in the age of temporary advantage*. Basic Books, 1999.

Joseph Fleiss. Measuring Nominal Scale Agreement among many raters. *Psychological Bulletin*, 76(5):378–382, 1971.

Vernon Francis. Supply chain visibility: lost in translation? *Supply Chain Management: An International Journal*, 13(3):180–184, 2008. ISSN 1359-8546. doi: 10.1108/13598540810871226. URL http://dx.doi.org/10.1108/13598540810871226.

Zetian Fu, Jun Yue, and Zhenbo Li. Ontology and Its Application in Supply Chain Information Management. In *Supply Chain the Way to Flat Organisation*. IntechOpen, 2009. ISBN 9789537619350. doi: 10.5772/6669.

Thomas Fuller. Thailand Flooding Cripples Hard-Drive Suppliers, 2011. URL https://www.nytimes.com/2011/11/07/business/global/07iht-floods07.html.

John T Gardner and Martha C Cooper. Strategic Supply Chain Mapping Approaches. *Journal of Business Logistics*, 24(2):37–64, 2003. ISSN 07353766. doi: 10.1002/j.2158-1592.2003.tb00045.x.

Dominic Gates and Rami Grunbaum. A 'prescient' warning to Boeing on 787 trouble, 2011. URL http://old.seattletimes.com/html/sundaybuzz/2014125414_sundaybuzz06.html.

Mark Goh, Robert De Souza, Allan N. Zhang, Wei He, and P. S. Tan. Supply chain visibility: A decision making perspective. *2009 4th IEEE Conference on Industrial Electronics and Applications, ICIEA 2009*, pages 2546–2551, 2009. doi: 10.1109/ICIEA.2009.5138666.

Stefan Gold, Alexander Trautrims, and Zoe Trodd. Modern slavery challenges to supply chain management. *Supply Chain Management*, 20(5):485–494, 2015. ISSN 13598546. doi: 10.1108/SCM-02-2015-0046.

S.L. Golicic, D.F. Davis, and T.M. McCarthy. A Balanced Approach to Research in Supply Chain Management. In H. Kotzab, S Seurig, M Mueller, and G. Reiner, editors, *Research Methodologies in supply chain management*, pages 15–30. Physica-Verlag, Heidelberg, 2005. ISBN 3790815837.

Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep learning*. The MIT Press, Cambridge, MA, USA, 1 edition, 2016.

Alex Graves and Jürgen Schmidhuber. Framewise phoneme classification with bidirectional lstm and other neural network architectures. *Neural Networks*, 18(5-6):602–610, 2005.

Hans Greimel. Tsunami: The aftermath (2012-03-12), 2011. URL https://www.autonews.com/article/20120312/OEM01/303129960/tsunami-the-aftermath.

Thomas R. Gruber. Toward principles for the design of ontologies used for knowledge sharing, 1995.

Egon G. Guba and Yvonna S. Lincoln. Competing Paradigms in Qualitative Research. In N.K. Denzin and Y.S. Lincoln, editors, *Handbook of qualitative research*, chapter Competing, pages 105–117. Sage, 1994.

James R. Hagerty. Factories Look for Supply-Chain Risks in All the Wrong Places, 2013. URL https://blogs.wsj.com/corporate-intelligence/2013/12/18/factories-look-for-supply-chain-risks-in-all-the-wrong-places/.

Masahiko Haraguchi and Upmanu Lall. Flood risks and impacts: A case study of Thailand's floods in 2011 and research questions for supply chain decision making. *International Journal of Disaster Risk Reduction*, 14:256–272, 2014. ISSN 22124209. doi: 10.1016/j.ijdrr.2014.09.005. URL `http://dx.doi.org/10.1016/j.ijdrr.2014.09.005`.

Zellig S. Harris. Distributional structure. *WORD*, 10(2-3):146–162, 1954. doi: 10.1080/00437956.1954.11659520. URL `https://doi.org/10.1080/00437956.1954.11659520`.

Marti a. Hearst. Automatic Acquisition of Hyponyms from Large Text Corpora. *Proceedings of the 14th conference on Computational Linguistics*, 2:23–28, 1992.

Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural computation*, 9 (8):1735–1780, 1997.

Johannes Hoffart, Mohamed Amir Yosef, Ilaria Bordino, Hagen Fürstenau, Manfred Pinkal, Marc Spaniol, Bilyana Taneva, Stefan Thater, and Gerhard Weikum. Robust disambiguation of named entities in text. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pages 782–792. Association for Computational Linguistics, 2011.

Kurt Hornik. Approximation capabilities of multilayer feedforward networks. *Neural Networks*, 4(2):251–257, 1991. ISSN 0893-6080. doi: https://doi.org/10.1016/0893-6080(91)90009-T. URL `http://www.sciencedirect.com/science/article/pii/089360809190009T`.

R. A. Howard. Information Value Theory. *Systems Science and Cybernetics, IEEE Transactions on*, 2(1):22–26, 1966.

IAJ. 2011 Thai Floods : Impact and Lessons Learned for Japanese Non-Life Insurers, 2013. URL `https://www.casact.org/education/annual/2013/handouts/Paper_2566_handout_1507_0.pdf`.

Alon Jacovi, Oren Sar Shalom, and Yoav Goldberg. Understanding Convolutional Neural Networks for Text Classification. In *Proceedings of the 2018 EMNLP Workshop BlackboxNLP: Analyzing and Interpreting Neural Networks for NLP (Association for Computational Linguistics)*, pages 56–65, 2018. URL `http://arxiv.org/abs/1809.08037`.

Thorsten Joachims. Text categorization with support vector machines: Learning with many relevant features. In *European conference on machine learning*, pages 137–142. Springer, 1998.

Uta Jüttner. Supply chain risk management: Understanding the business requirements from a practitioner perspective. *The International Journal of Logistics Management*, 16(1): 120–141, 2005. ISSN 13598546. doi: 10.1108/13598540410527079.

Uta Jüttner and S Maklan. Supply chain resilience in the global financial crisis: An empirical study. *Supply Chain Management*, 16(4):246–259, 2011. ISSN 1359-8546. doi: 10.1108/13598541111139062.

Uta Jüttner, Helen Peck, and Martin Christopher. Supply chain risk management: outlining an agenda for future research. *IJ of Logistics: Research and Applications*, 6(4):197–210, 2003.

Ron Kenett and Yossi Raanan, editors. *Operational Risk Management: A Practical Approach to Intelligent Data Analysis*. John Wiley & Sons, 2011.

Yong H Kim and Gerald F Davis. Challenges for Global Supply Chain Sustainability: Evidence From Conflict Minerals Reports. *Academy of Management Journal*, 59(6): 1896–1916, 2016. ISSN 0001-4273. doi: 10.5465/amj.2015.0770.

Yusoon Kim, Yi-Su Chen, and Kevin Linderman. Supply network disruption and resilience: A network structural perspective. *Journal of Operations Management*, 33-34(1):43–59, 2014. doi: 10.1016/j.jom.2014.10.006. URL `https://onlinelibrary.wiley.com/doi/abs/10.1016/j.jom.2014.10.006`.

Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

KPMG International & The Economist Intelligence Unit. The Global Manufacturing Outlook 2013 - Competitive advantage: enhancing supply chain networks for efficiency and innovation, 2013. URL `http://www.kpmg.com/DE/de/Documents/global-manufacturing-outlook-2013-kpmg.pdf`.

Douglas M Lambert and Martha C Cooper. Issues in Supply Chain Management. *Industrial Marketing Management*, 29(1):65–83, 2000.

Lufthansa. BOEING 747-8 - One plane, six million parts, 2018. URL `https://magazin.lufthansa.com/xx/en/fleet/boeing-747-8-en/one-plane-six-million-parts/`.

Gang Luo, Xiaojiang Huang, Chin-Yew Lin, and Zaiqing Nie. Joint Named Entity Recognition and Disambiguation. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing (ACL)*, pages 879–888, 2015.

Alexander Maedche and Steffen Staab. Ontology Learning for the Semantic Web. *IEEE Intelligent Systems*, 16(2):72–79, 2001.

John Mangan, Chandra Lalwani, and Tim Butcher. *Global logistics and supply chain management*. John Wiley & Sons, 2008.

Christopher D. Manning, Prabhakar Raghavan, and Hinrich Schutze. *An Introduction to Information Retrieval*. Cambridge University Press, 2009. ISBN 0521865719. doi: 10.1109/LPT.2009.2020494. URL `https://nlp.stanford.edu/IR-book/information-retrieval-book.html`.

James G March and Zur Shapira. Managerial Perspectives on Risk and Risk Taking. *Management Science*, 33(11):1404–1418, 1987. ISSN 0025-1909. doi: 10.1287/mnsc.33.11.1404.

JT John T. Mentzer, William Dewitt, JS James S Keebler, Soonhong Min, Nancy W Nix, Carlo D Smith, and Zach G Zacharia. Defining supply chain management. *Journal of Business Logistics*, 22(2):1–25, 2001. ISSN 07353766. doi: 10.1002/j.2158-1592.2001.tb00001.x. URL `http://onlinelibrary.wiley.com/doi/10.1002/j.2158-1592.2001.tb00001.x/abstract`.

Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. Efficient Estimation of Word Representations in Vector Space. *Proceedings of the International Conference on Learning Representations (ICLR 2013)*, pages 1–12, 2013. ISSN 15324435. doi: 10.1162/153244303322533223. URL `http://arxiv.org/abs/1301.3781`.

Mike Mintz, Steven Bills, Rion Snow, and Dan Jurafsky. Distant supervision for relation extraction without labeled data. *Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP: Volume 2 - ACL-IJCNLP '09*, 2(2005):1003, 2009. ISSN 1932432469. doi: 10.3115/1690219.1690287. URL `http://portal.acm.org/citation.cfm?doid=1690219.1690287`.

T M Mitchell. *Machine Learning*. McGraw-Hill Higher Education, New York, USA, 1997.

Tom Mitchell, William W. Cohen, E Hruschka, Partha Talukdar, B Yang, Justin Betteridge, Andrew Carlson, B Dalvi, Matt Gardner, Bryan Kisiel, J Krishnamurthy, Ni Lao, K Mazaitis, T Mohamed, N Nakashole, E Platanios, A Ritter, M Samadi, B Settles, R Wang, D Wijaya, A Gupta, X Chen, A Saparov, M Greaves, and J Welling. Never-Ending Learning. *Communications of the Acm*, 61(5):2302–2310, 2018.

Benoit Montreuil, Olivier Labarthe, Driss Hakimi, Anne Larcher, and Mathieu Audet. Supply Web Mapper. In *International Conference on Industrial Engineering and Systems Management, IESM' 2009 May 13 - 15, MONTREAL - CANADA*, 2009.

Kevin P Murphy. *Machine Learning: A Probabilistic Perspective*. The MIT Press, 2012. ISBN 9780262018029.

Jeffrey Pennington, Richard Socher, and Christopher Manning. Glove: Global Vectors for Word Representation. *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1532–1543, 2014. ISSN 10495258. doi: 10.3115/v1/D14-1162. URL `http://aclweb.org/anthology/D14-1162`.

Jakub Piskorski and Roman Yangarber. Information Extraction: Past, Present and Future. In Thierry Poibeau, Horacio Saggion, Jakub Piskorski, and Roman Yangarber, editors, *Multi-source, Multilingual Information Extraction and Summarization*, pages 23–50. Springer-Verlag Berlin Heidelberg, 1 edition, 2013. ISBN 978-3-642-28568-4. doi: 10.1007/978-3-642-28569-1.

David K Platt. PA 12 shortage sparks industry action, 2012. URL `https://www.plasticsnewseurope.com/article/20120529/PNE/305299966/pa-12-shortage-sparks-industry-action`.

Parichart Promchote, S-Y Simon Wang, and Paul G Johnson. The 2011 great flood in Thailand: Climate diagnostics and Implications from climate change. *Journal of Climate*, 29(1):367 – 379, 2016.

Sampo Pyysalo, Simon Baker, Imran Ali, Stefan Haselwimmer, Tejas Shah, Andrew Young, Yufan Guo, Johan Högberg, Ulla Stenius, Masashi Narita, et al. Lion lbd: a literature-based discovery system for cancer biology. *Bioinformatics*, 2018.

Alec Radford, Jeffrey Wu, Rewon Child, David Luan, Dario Amodei, and Ilya Sutskever. Language models are unsupervised multitask learners. *OpenAI Blog*, 1(8), 2019.

Deepa Seetharaman. Automakers face paint shortage after Japan quake (2011-03-26), 2011. URL https://www.reuters.com/article/us-japan-pigment/automakers-face-paint-shortage-after-japan-quake-idUSTRE72P04B20110326.

Yossi Sheffi. *The Resilient Enterprise: Overcoming Vulnerability for Competitive Advantage*, volume 1. The MIT Press, 2005. ISBN 0262195372.

Yusuke Shinyama and Satoshi Sekine. Preemptive information extraction using unrestricted relation discovery. In *Proceedings of the Human Language Technology Conference of the North American Chapter of the ACL*, pages 304–311, 2006. doi: 10.3115/1220835.1220874.

David Simchi-Levi, William Schmidt, and Yeshua Wei. From Superstorms to Factory Fires. *Harvard Business Review*, Jan-Feb(February):96–101, 2014.

Yadira Soto-Viruet, W. David Menzie, John F. Papp, and Thomas R. Yager. USGS Open-File Report 2013–1239 - An Exploration in Mineral Supply Chain Mapping Using Tantalum as an Example. Technical report, United States Geological Survey (USGS), 2013. URL http://pubs.usgs.gov/of/2013/1239/.

Christopher Tang. Robust strategies for mitigating supply chain disruptions. *International Journal of Logistics*, 9(1):33–45, 2006a. ISSN 1367-5567. doi: 10.1080/13675560500405584.

Christopher Tang and Brian Tomlin. The power of flexibility for mitigating supply chain risks. *International Journal of Production Economics*, 116:12–27, 2008. ISSN 09255273. doi: 10.1016/j.ijpe.2008.07.008.

Christopher S. Tang. Perspectives in supply chain risk management. *International Journal of Production Economics*, 103:451–488, 2006b. ISSN 09255273. doi: 10.1016/j.ijpe.2005.12.006.

Christopher S Tang and Joshua D Zimmerman. Managing New Product Development and Supply Chain Risks: The Boeing 787 Case. *Supply Chain Forum*, 10:74–87, 2009.

The Economist. The global car industry - After the quake (2011-05-19), 2011. URL https://www.economist.com/business/2011/05/19/after-the-quake.

TheEconomist. Multinational companies are adjusting to shorter supply chains, 2019. URL https://www.economist.com/special-report/2019/07/11/multinational-companies-are-adjusting-to-shorter-supply-chains.

Shara Tibken. Intel Cuts Its Outlook (2019-12-13), 2011. URL https://www.wsj.com/articles/SB10001424052970203518404577094200200314384.

Brian Tomlin. On the Value of Mitigation and Contingency Strategies for Managing Supply Chain Disruption Risks. *Management Science*, 52(5):639–657, 2006. ISSN 0025-1909. doi: 10.1287/mnsc.1060.0515.

James Topham. Auto production faces bigger hit after Japan quake (2011-04-26), 2011. URL https://www.reuters.com/article/us-autos-japan-production/auto-production-faces-bigger-hit-after-japan-quake-idUSTRE73O0U020110426.

Susan van den Brink, René Kleijn, Arnold Tukker, and Jaco Huisman. Approaches to responsible sourcing in mineral supply chains. *Resources, Conservation and Recycling*, 145(November 2018):389–398, 2019. ISSN 09213449. doi: 10.1016/j.resconrec.2019.02.040. URL https://linkinghub.elsevier.com/retrieve/pii/S092134491930103X.

Lita Van Wel and Lambèr Royakkers. Ethical issues in web data mining. *Ethics and Information Technology*, 6:129–140, 2004. ISSN 1388-1957. doi: 10.1023/B:ETIN.0000047476.05912.3d.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *Advances in neural information processing systems*, pages 5998–6008, 2017.

Aidan R Vining and David L Weimer. Information asymmetry favoring sellers: a policy framework. *Policy Sciences*, 21(4):281–303, 1988. URL http://www.springerlink.com/content/q5j5367430n16821/.

Donald Waters. *Supply Chain Risk Management: Vulnerability and Resilience in Logistics*. Kogan Page Publishers, 2 edition, 2011. ISBN 0092-2102. doi: 10.1007/978-0-387-72776-9.

Pascal Wichmann, Alexandra Brintrup, Simon Baker, Philip Woodall, and Duncan McFarlane. Towards automatically generating supply chain maps from natural language text. *IFAC-PapersOnLine*, 51(11):1726 – 1731, 2018. ISSN 2405-8963. doi: https://doi.org/10.1016/j.ifacol.2018.08.207. 16th IFAC Symposium on Information Control Problems in Manufacturing INCOM 2018.

Pascal Wichmann, Alexandra Brintrup, Simon Baker, Philip Woodall, and Duncan McFarlane. Extracting supply chain maps from news articles using deep neural networks. *International Journal of Production Research*, 58(17):5320–5336, 2020. doi: 10.1080/00207543.2020.1720925.

Brian Wixted. KNOWLEDGE, INNOVATION AND COMPETITIVENESS: DYNAMICS OF FIRMS, NETWORKS, REGIONS AND INSTITUTIONS. In *DRUID Summer Conference*, pages 1–27, 2006.

Fei Wu and Daniel S Weld. Open Information Extraction using Wikipedia. In *Proceedings ofthe 48th Annual Meeting ofthe Association for Computational Linguistics*, pages 118–127, 2010. ISBN 1581138741. doi: 10.1145/1519103.1519113.

Ayana Yamamoto, Yuichi Miyamura, Kouta Nakata, and Masayuki Okamoto. Company relation extraction from web news articles for analyzing industry structure. In *Semantic Computing (ICSC), 2017 IEEE 11th International Conference on*, pages 89–92. IEEE, 2017.

Tingting Yan, Thomas Y Choi, Yusoon Kim, and Yang Yang. A theory of the nexus supplier: a critical supplier from a network perspective. *J of Supply Chain Management*, 51(1): 52–66, 2015.

Mohamed Amir Yosef, Johannes Hoffart, Ilaria Bordino, Marc Spaniol, and Gerhard Weikum. Aida: An online tool for accurate disambiguation of named entities in text and tables. *Proceedings of the VLDB Endowment*, 4(12):1450–1453, 2011.

Tom Young, Devamanyu Hazarika, Soujanya Poria, and Erik Cambria. Recent Trends in Deep Learning Based Natural Language Processing. *IEEE Computational Intelligence Magazine*, 13(July):55–75, 2017. ISSN 1556-603X VO - 13. doi: arXiv:1708.02709v6. URL `http://arxiv.org/abs/1708.02709`.

Daojian Zeng, Kang Liu, Siwei Lai, Guangyou Zhou, and Jun Zhao. Relation Classification via Convolutional Deep Neural Network. In *Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers*, pages 2335–2344, 2014. ISBN 978-1-941643-26-6. doi: http://aclweb.org/anthology/C/C14/C14-1220.pdf.

Appendix A

Details related to Chapter 4

This part of the appendix contains more detailed content related to Chapter 4 which provided the rationale for improving structural supply chain visibility. It includes detailed summaries of the exploratory interviews as well as an in-depth discussion of use cases for structural supply chain visibility.

A.1 Exploratory interviews

A.1.1 Interview series with aerospace manufacturer

Profile The findings summarised below stem from a series of 7 semi-structured on-site interviews with supply chain managers of a major aerospace manufacturer in March/April 2015. The supply chain managers worked in different departments of the manufacturer and covered relevant functions, such as business continuity or supplier management.

Upon request by the aerospace manufacturer and for reasons of confidentiality, the interview summaries have been redacted in this version of the thesis.

Upon request by the aerospace manufacturer and for reasons of confidentiality, the interview summaries have been redacted in this version of the thesis.

Upon request by the aerospace manufacturer and for reasons of confidentiality, the interview summaries have been redacted in this version of the thesis.

Upon request by the aerospace manufacturer and for reasons of confidentiality, the interview summaries have been redacted in this version of the thesis.

A.1.2 Interview with supply chain consultant

Profile The interviewee is an accomplished manufacturing professional, with more than 15 years' experience in electronics manufacturing companies where she held a variety of senior management positions in supply chain, manufacturing engineering and operations management. The interviewee is also an Industrial Fellow in the Education and Consultancy Services (ECS) group of the Institute for Manufacturing (IfM) at the University of Cambridge and works with a wide range of businesses of various sizes, tailoring and applying the IfM's toolset to enable companies to develop their strategies and capabilities and make tactical and operational changes to deliver sustainable, profitable growth. The interviewee is currently working as part of the ECS team in the ongoing Sharing in Growth programme, a Government-funded programme of intensive supplier development in the UK aerospace and nuclear industries. The interviewee answered the questions by drawing on her experience as a consultant with the aerospace industry (a) and as a previous supply chain management professional at an electronics company (b) herself.

(a) Findings with respect to the aerospace and defence industry in general Second-tier suppliers to Boeing or Rolls-Royce have long contracts for a particular programme, the production of a particular aircraft type. Once these suppliers enter a programme, they tend to stay on that programme. Companies like Boeing are oblivious that these second-tier suppliers in turn depend on very small, and often highly specialised companies. Some of these second-tier companies were single-sourced and struggling. The more parts are sourced from areas around the world, the more the buyers are exposed to risks, such as natural disasters. The second-tier suppliers do not only depend on components manufacturers. Because they are relatively small, these suppliers do not perform all the processing steps in-house but send out to other specialised companies. The dependencies on *processing* specialists are often neglected in risk assessments compared to the dependencies on parts or material suppliers. In the aerospace domain, environmental EU regulations, such as *"RoSH"* (short for: *Restriction of Hazardous Substances*) and *"REACH"* (short for: *Registration, Evaluation, Authorisation and Restriction of Chemicals*) restrict the type of chemicals and processes that a company is allowed to use. Because of required approvals and because for some processes alternative methods have not been found yet, companies are reluctant to change their processes and chemicals are still being used that are not “REACH”-compliant. By better understanding supply chains, such as discovering that dubious or known-to-be-incompliant sub-tier suppliers participate in the supply chain, it may become easier to understand where your company is not compliant with regulations. A supply chain map could at least tell how complicated it would be to make sure that every participant is compliant. One aerospace company discovered that their products were being used by a major film studio and tried to understand what the use case for their product was. From a business development perspective it may be valuable to understand where a product is being used further downstream to develop and extend the product portfolio and identify new customer groups. A supply chain map could help with discovering where a product is being used. In the aerospace domain, supply chains for a particular programme do not change much. Even over time spans of 10 years, the companies will largely remain the same. In fact, once processes have been approved, even switching sites might not be possible due to site-based approvals. Even poorly performing companies tend to stay on the programme. However, once a new programme is started, the supply chain may look completely different.

(b) Findings with respect to the supply chain management at an electronics company
The electronics company suffered from a shortage of a specific part after the distributor’s warehouse in the UK burned down. The part was a cheap metal contact procured from Malaysia. A little kink in the contact made the part a bespoke one. Thousands of these contacts would be used every day and, hence, the warehouse would hold large quantities of that contact and a distributor would just regularly fill up the inventory. After the warehouse fire, a replacement order of parts via ship took six weeks to arrive and the electronics company had to spend significant funds on air freight. Before that incident, the cheap metal part had not been considered critical; a warehouse fire had not been considered a risk. The warehouse fire revealed that the manufacturer also had not held sufficient stock either – because of an undisclosed tooling problem.

Supply risk management is not only relevant once a product is being manufactured. When a new product is designed, knowledge about supply risks can be used to inform the product design process. When the designers had a choice, the supply chain risk managers of the electronics company could influence which parts or technologies were chosen. This could take the form of favouring commodity parts over bespoke ones or avoiding specific processes if only few companies could perform these tasks, such as sand casting aerospace parts in the UK. Supply chain maps could therefore not only help with risks in the current supply chain but help with planning future ones.

To understand their supply chain, the electronics companies used magazines and obtained a dataset every year that was used to understand the supply chain for the electronic parts. But because the supply chain was relatively short, they were able to ask their suppliers to find out information from their suppliers. In case of bespoke parts, they also spoke directly with the manufacturer, e.g. to understand tooling. To understand risk, it was key to know the geo-location of the suppliers. The electronics company prioritised based on the financial value a supplier corresponded to despite it being “somewhat of a red herring” given that sales volume does not always correspond to criticality. The health of suppliers was monitored using financial analyses, such as the development of investments, revenue and profit over time.

To ensure a healthy supplier base, the electronics company would monitor its own share in the revenue of their suppliers. The rule was to not exceed 20% of any supplier’s portfolio to ensure the suppliers remained robust (“You want to be big enough that they really care about you, but you want to be part of a package, so when there is ebb and flow in demand, they are sustainable.”). One of the tasks in supply chain management was to see where the company was too dominant for a particular supplier. This is something that can be seen in the aerospace industry in particular: Some of the smaller companies get so excited that they got on board of a big programme and that they “put all their eggs in one basket” and this programme accounts for 40 or 50% of their revenue. In those cases and if the programme is paused for 6 weeks, these smaller companies are in danger.

The structural visibility of the electronics company was limited to the distributors (first tier) and the manufacturers (second tier). The manufacturers were known because they directly collaborated with them, and then let them supply via a distributor. Beyond the manufacturer, there was no visibility. The manufacturer would know and was expected to inform them about problems.

Selection of prior assumptions that were confirmed during the interview For a supply chain map, it is important to have directed relations to understand how risk propagates through the network. Two companies can have more than just a single directed buyer-supplier relation. A company may be a direct supplier on one stream, and supply indirectly via another supplier to the same end-customer. Supply chains change over time but in the aerospace industry, supply chains tend to be stable on a programme-basis.

A.2 Use cases of structural supply chain visibility

A.2.1 Selected corporate social responsibility topics

Modern slavery Shockingly, modern forms of slavery and forced labour are still common around the world. The Global Slavery Index[1] by the Walk Free Foundation estimated that 40.3 million people lived in modern slavery in 2016. A heatmap showing the prevalence of modern slavery is depicted by Figure A.1.

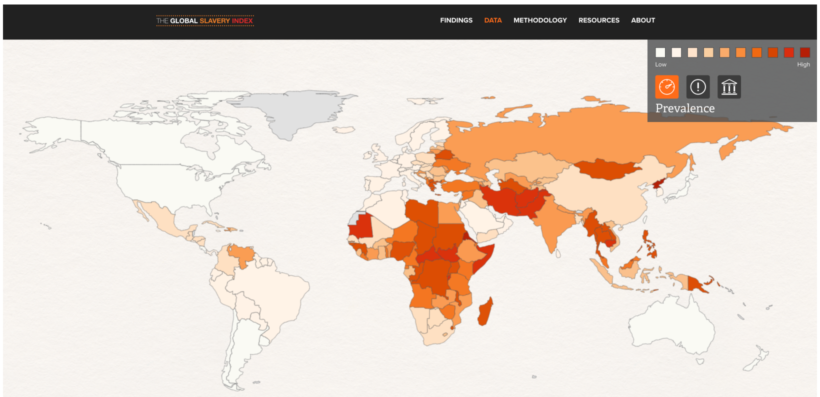


Fig. A.1 Word map – Global Slavery Index; source: Global Slavery Index[2]

Companies with complex supply chains may unknowingly rely on forced labour. In 2016, the CORE coalition states[3]: “"[T]he risk of involvement in modern slavery through the supply chain increases as supplier and / or contracting and sub-contracting chains grow. Increased risks associated with difficulties in supply chain monitoring and management arise in complex global supply chains. Actions by suppliers, sub-contractors and business partners can place the entire supply chain at risk."

Recently, new regulations against modern forms of slavery, human trafficking, and other human rights violations have come into force. Table A.1 provides an overview of selected such regulations.

[1] `https://www.globalslaveryindex.org/2018/data/maps/#prevalence`; last accessed: 2019-03-07

[3] `https://corporate-responsibility.org/wp-content/uploads/2016/03/CSO_TISC_guidance_final_digitalversion_16.03.16.pdf`; last accessed: 2019-06-23

Table A.1 Overview of Modern Slavery regulations (selection)

Legislative region	Signed into law	In effect since	Name
California (US)	2010	01.01.2012	California Transparency in Supply Chains Act of 2010
UK	2015	29.10.2015	UK Modern Slavery Act 2015
Australia	2018	01.01.2019	Modern Slavery Bill 2018
Netherlands	2019	(2020)	Child Labour Due Diligence Law

Typically, these regulations require larger companies to regularly and publicly disclose which efforts they undertake to understand their extended supply chain and prevent human rights violations in their supply chains. The idea is to increase transparency for consumers and investors, and to let them assert pressure on those companies that fail to address these issues. The "California Transparency in Supply Chains Act" spearheaded this movement, followed by the UK Modern Slavery Act in 2015[4], and the Australian "Modern Slavery Bill 2018". Other countries have announced plans for similar regulations (e.g. the Dutch "Child Labour Due Diligence Law" which is expected to come into effect in 2020).

It is noteworthy that this type of legislation increases pressure for companies to understand their supply chains *beyond just their direct suppliers*. The Australian regulation[5] clarifies the reporting scope and states that the term supply chain is "intended to refer to the products and services that contribute to the entity's own products and services and is not restricted to 'tier one' or direct suppliers.". The CORE coalition states the following: "Mapping supply chains is essential to understanding and managing a business's exposure to risk. In many cases, companies are aware of the first-tier factories and production facilities in their supply chains, but know little about the practices of the businesses supplying these vendors or producers. Companies should know as much as possible about their supply chains, about who is involved in the provision of their goods and services, and about the working conditions at all levels of the supply chain."[6]

In addition to the harm done to the people working under such conditions, the reputation risk, and the supply risk, it appears that the legal risk has also increased in recent years. Companies are increasingly held accountable for the misconduct of their suppliers – even

[4] `http://www.legislation.gov.uk/ukpga/2015/30/section/54`

[5] Bills Digest 12, 2018-19 - Modern Slavery Bill 2018

[6] `https://corporate-responsibility.org/wp-content/uploads/2016/03/CSO_TISC_guidance_final_digitalversion_16.03.16.pdf`; last accessed: last accessed: 2019-06-23

multiple tiers downstream. A number of lawsuits have been filed not against the company that was directly responsible for the crimes but against against companies further down the supply chain claiming a joint responsibility. Norton Rose Fulbright provide an in-depth overview of the legal background, list various law suits and discuss the potential liability of companies for misconduct of their sub-tier suppliers[7]. An excellent overview of the challenges modern slavery poses to supply chain management is provided by Gold et al. (2015). The authors also discuss management tools for detecting slavery in a supply chain.

Conflict minerals Materials considered *"conflict minerals"* are, for instance, tin, tungsten, tantalum and gold, often referred to as *"3TG"*, or *cobalt* which is commonly sourced from the Democratic Republic of Congo (DRC) and surrounding areas under inhumane conditions and via poorly regulated *"artisanal and small-scale mining"* (ASM).

In 2010, the Dodd–Frank Wall Street Reform and Consumer Protection Act ("Dodd-Frank Act") was signed into US federal law. Section 1502 of this act gave companies over three years to determine and report on whether their products contained conflict minerals from the Democratic Republic of Congo. In a paper published in 2016, Kim and Davis (2016) analysed every conflict minerals report submitted to the US Securities and Exchange Commission (SEC). They find that out of 1,300 corporations which had submitted these reports "almost 80% admitted they were unable to determine the country of origin of such materials, and only 1% could certify themselves conflict-free with certainty beyond reasonable doubt". The authors conclude that "[i]nternationally diversified firms and those with large and more dispersed supply chains were less likely to declare their products conflict-free: complexity reduces the visibility of a firm's supply chain". Kim and Davis (2016) confirm the problem of limited structural supply chain visibility from a Corporate Social Responsibility perspective: "[...] as production becomes increasingly disaggregated, corporations are called on to be more accountable for the practices of their suppliers and even the actions of the states in which they operate. This reflects a 'responsibility paradox': demands for corporate social responsibility (CSR) increase even as companies' ability to deliver shrinks".

Greenhouse gas emissions In some sense similar to the issue of modern slavery, companies are also increasingly held responsible for greenhouse gas (GHG) emissions in their extended supply chain as opposed to just those they are directly responsible for. The Greenhouse

[7] `https://human-rights-due-diligence.nortonrosefulbright.online/publications/2018/q2/components-of-effective-supply-chain-management/`; last accessed: 2019-06-23

Gas Protocol (GHG Protocol)[8] defines so-called “Scope 3”[9] emissions that are not directly caused by the focal company (Scope 1) and not caused by the purchased electricity, heat and steam (Scope 2) but represent the footprint of the upstream and downstream value chain. Estimating and managing this footprint is made even more difficult when the supply chain participants are unknown. Even though Scope 3 emissions are more difficult to control, they “can represent the largest source of emissions for companies and present the most significant opportunities to influence GHG reductions”[10].

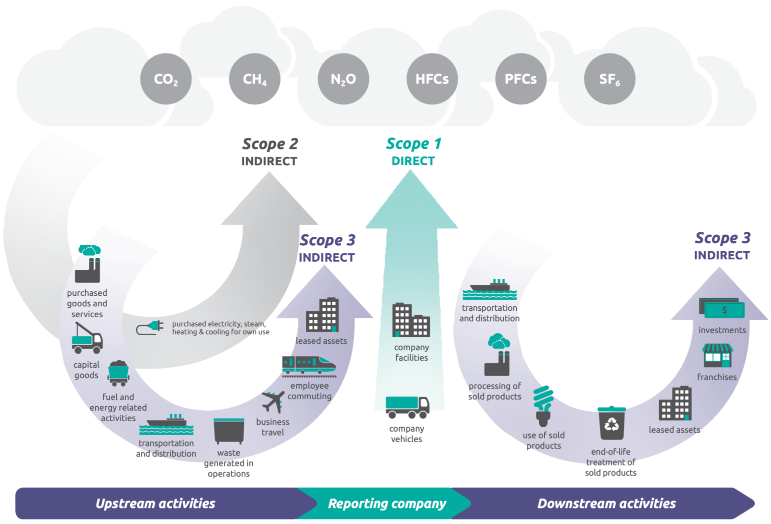


Fig. A.2 Scopes of the GHG Protocol: Scope 3 considers the footprint of the upstream and downstream value chain (source: ghgprotocol.org)

A.2.2 Further use cases of structural supply chain visibility

This chapter shall provide a summary for further use cases of structural supply chain visibility.

Food safety and other “contamination” scenarios The topology of a supply network can be valuable information in the context of food safety and other “contamination” scenarios. Unlike the supply risk scenarios, then primary motivation in these cases is *not* to avoid supply disruptions but to prevent or trace back the spread of a potentially harmful “contamination” through the supply chain. In the case of food safety, the motivation is to prevent or trace

[8] `http://ghgprotocol.org/`; last accessed: 2019-07-23
[9] `http://ghgprotocol.org/standards/scope-3-standard`; last accessed: 2019-07-23
[10] `http://ghgprotocol.org/sites/default/files/standards/Corporate-Value-Chain-Accounting-Reporing-Standard_041613_2.pdf`; last accessed: 2019-07-23

back the spread of diseases through the global food processing chain. Notable cases in the literature are, among many others, the BSE food crisis and Dioxin in chicken feed (e.g. Aung and Chang (2014)). But there are also other “contamination” scenarios that require companies or governmental agencies to understand where materials originated from or where they ended up being used. A prominent example is the 2017 Kobe Steel case[11], where inspection data had been falsified and steel of questionable quality had been delivered to and used by various customers in the automotive and aerospace industry, potentially across multiple tiers.

Supply chain efficiency & network design Apart from ensuring a steady supply with materials, parts, and services, adjustments of the supply chain architecture can pursue further objectives, such as making the supply chain faster, less costly, or less complex by involving fewer participants. For these purposes, information about the current or planned supply chain structure may be valuable.

Supplier discovery Knowledge about where own suppliers or competitors obtain their materials, parts, and services from may help discover new, alternative suppliers.

Detecting supplier consolidation A consolidation trend in the extended supply chain may lead to a shift of negotiation power away from the focal company. This may not immediately impact supply but it could lead to price increases and a limited ability to change to alternative suppliers.

Competitive intelligence In addition to discovering and monitoring their own extended supply networks, companies may have in interest to do the same for their direct competitors, e.g. for benchmarking their supply chain architecture against the ones of their competitors or other companies.

Supply chain research Due to the limited availability of information about supply chain structures, researchers are often limited to simplified (dyadic or triadic) supply chain models. Academics may find it valuable to be able to work on larger supply network graphs that potentially span multiple industries.

[11] E.g. see `https://www.ft.com/content/a1e494c2-ad9f-11e7-beba-5521c713abf4`; last accessed: 2019-05-30

Discovering new customer groups or use cases One may also hypothesise about the possibility to *identify new customer groups* by increasing the visibility downstream. Companies could ask the question who ends up using their products further downstream beyond their direct customer and for what purpose.

Discovering investment opportunities Financial investors could *identify potential investment targets* by discovering the (sub-tier) suppliers of a company that is expected to grow. Instead of investing directly in that (already no longer undervalued) company, they may want to invest in their (still undervalued) supply chain participants instead.

Appendix B

Details related to Chapter 5

This part of the appendix contains more detailed content related to Chapter 5 which addressed the problem of automating the extraction of individual buyer-supplier relations from text. The content provided below covers an in-depth reflection on the scoping decisions as well as on selected aspects of the methodology presented in Chapter 5.

B.1 Scoping decisions

In Section 5.2.1 scoping decisions were summarised. A discussion of these aspects and a rationale for the decision shall be provided here.

English language Various processing steps, such as the NER and relation classification, are language-dependent. For the sake of simplicity, the scope is limited to extracting information from text documents in the English language. The support of NLP libraries tends to be best for the English language and news datasets are most easily available in English. However, ideally the classification approach should work for any common natural language by switching language-specific modules.

Ability to cope with a large number of variations Any approach needs to be able to cope with thousands of variations of how buyer-supplier relations can be expressed. A few hard-coded patterns would be insufficient.

Classification of explicitly stated relations between two organisations The aim is to extract explicitly stated buyer-supplier relations between two named organisations. For the sake of simplicity especially in the labelling process, extracting other relation types is *not* considered within the scope – even if these would collectively allow to infer buyer-supplier relations between two organisations. Examples of relation types beyond the scope of this research would be company-to-product relationships. Furthermore, the focus is on relations that are binary with respect to the number of elements in them and these entities shall be organisations. Lastly, the organisations need to be mentioned by name in the sentence. This is a limitation as some information may then not be captured. For instance, consider the authentic sentence: “Safran’s joint venture with General Electric provides Boeing with engines for the 737.”. It can be extracted that Safran and General Electric have a joint venture. However, since the joint venture is unnamed, no relation can be extracted between it and its customer Boeing. In the domain of Linked Data, and specifically the Resource Description Framework (RDF[1]), there is the concept of a *blank node*. A blank node is an unnamed or unidentified entity that nevertheless participates in a data graph, so that one can navigate through it to other relationships. Within the scope of this research, blank nodes will not be created. However, future extensions of this work could profit from including these, successively collecting more information about these blank nodes and potentially inferring the identity of these nodes.

Requirements regarding the relation classes The simplest class definition would be a binary one: either there is some form of buyer-supplier relationship or there is none. However, this would ignore the directionality of buyer-supplier relations which is important for common cases of supply chain maps as it indicates how risks and disruptions can propagate through the network. A step up from the binary classification scheme would be a multi-class classification with directionality considered. The classes would then be:

- company A supplies company B (A ‘suppliesTo’ B)
- company B supplies company A (A ‘isSuppliedBy’ B)
- no buyer-supplier relationship stated

By considering directionality, one is able to build a *directed graph*, as is commonly expected for a supply chain map. However, some sentences clearly express a buyer-supplier relationship but do not specify the direction, e.g. in the sentence "BMW and Continental sign supply agreement". The directionality could only be derived using background knowledge about the companies’ roles in the industry. Furthermore, joint ventures, partnerships and similar arrangements between two organisations will have to be considered special cases of buyer-supplier relations where the direction is commonly not knowable. Using the predicate logic from the domain of Linked Data, this relation could be phrased as ‘tradesWith’. This provides the rationale for at least 4 classes that need to be distinguished:

[1] `https://www.w3.org/RDF/`; last accessed: 2019-07-31

- company A supplies company B (A 'suppliesTo' B)
- company B supplies company A (A 'isSuppliedBy' B)
- undirected buyer-supplier relationship or some form of partnership or joint venture with unspecified directionality (A 'tradesWith' B)
- no buyer-supplier relationship stated

Additional relations classes could be considered. Section 5.3.2 below will elaborate on the definition of relationship classes.

A mention-wise classification is required As the same company may be mentioned multiple times in the same sentence and not all of these mentions are in a buyer-supplier relation, the classification task has to be performed for each pair of two mentions. Otherwise, if the trained classifier is applied to new data, it would produce a large number of false positives. Pairs of *identical* mentions, potentially resulting in self-referential relations, can be ignored. Even if the same named entity is mentioned multiple times, these mentions shall be considered in separation because the context may be different for each mention. Figure B.1 shows a sentence to illustrate this need. Organisational named entities have already been labelled and are highlighted ("ORG"). The first mention of "NTT" does not stand in a buyer-supplier relationship with "Softbank Corp". The second mention, however, does.

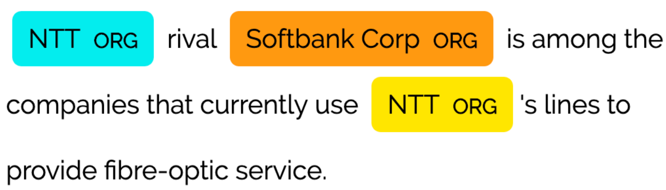


Fig. B.1 Example sentence illustrating the need for mention-wise classification

The classification shall not consider the specific company names as features In order to be able to generalise, the specific company names shall not considered be features to the classification process. An appropriate form of masking needs to be chosen to prevent any algorithm from "memorising" relations between specific companies and generalise beyond the mentions of specific companies.

Contextual scope limited to a single sentence and, even within a single sentence, co-reference resolution shall not be performed The contextual scope for this problem shall deliberately be restricted to a single, self-contained sentence where the relevant organisations are explicitly mentioned by name. This is done to keep the NLP pipeline short and the number of inter-dependent error sources low. This is a deliberate simplification that will lead to the loss of useful information but it also does not impact the nature of the classification task. For the problem at hand, co-reference resolution for pronouns (“it”, “they” etc.) or nouns (“the company”, “the carmaker” etc.) shall *not* be performed. It shall also not be aimed to extract any other information from the sentences before or after the sentence in question or any other sources of information. A text input, such as “Sumitek International LLC is a supplier of construction and mining machinery. The company supplies Toyota forklift loaders.” can, thus, not result in a detected buyer-supplier relation since “The company” will not get detected as an organisational named entity.

Extraction from explicit, complete statements instead of inference Buyer-supplier relations could be inferred from multiple statements, such as “Company A is the only producer of product X; X is used by Company B.” or “Company A is the only company that produces X in region Y. Company B procures its product X from a supplier in region Y.”. These types of inferences would require additional classification steps, such as the extraction of company-to-product relations, and the number of relations that would allow for such inferences appears to be virtually unlimited. For the sake of simplicity, this type of inference shall be considered out of scope. The focus is on explicitly stated buyer-supplier relations.

The distinction between what is explicitly stated and what is inferred may not always be clear. For instance, if an airline operates an aircraft from a specific manufacturer, one can reasonable assume that the aircraft was purchases from that manufacturer. But that is not always true. Aircraft can be leased or bought after they have been used.

General news The classifier that is trained and tested on the annotated text corpus is supposed to be able to work with general news. This also includes likely irrelevant topics, such as sports, entertainment and politics. This way, general news do not need to be pre-filtered before they can be run through the classifier. General news also appears to be an appropriate scope for the data input because general news datasets are available and NLP research has been conducted on general news datasets, so that word embeddings trained on general news are available.

Non-core and meta tasks are ignored The aim is solely to extract individual buyer-supplier relations between two organisational entities. This is *required* for supply chain mapping but *not sufficient*. All challenges that have to be addressed by prior, subsequent or parallel tasks but do not directly impact the extraction of buyer-supplier relations are ignored. These tasks are discussed in Chapter 6.

Information availability is ignored for the relation classification When the classification performance is measured, information availability is ignored. Extracted relations are compared against the ones that were stated in the input text, not against the actual supply chain.

Time-dependence of the structural information Even though the data may provide timestamps, the time-dependence of the structural information shall be ignored at this stage of the research. Once buyer-supplier relations can be reliably extracted, one may want to extend the research to show structural changes over time or attempt to identify the supply chain structure at a particular point in time.

B.2 Discussion of alternative labelling options

The class definitions used to distinguish relations between two organisational entities used in this thesis represent a compromise between labelling simplicity and and the expressive power of the chosen set of labels. Different classes were considered as were a multi-labelling system that could have captured additional facets of buyer-supplier relations.

The scope of the relation extraction can be further increased along at least the following dimensions:

- entity types considered in a relation (e.g. relations between an organisation and a product)
- number of entities considered in a single relation (e.g. ternary relations, such as "who supplies whom with what?")
- number and type of attributes captured for a relation (e.g. indicated uncertainty of a statement or time)
- relation types (e.g. "competes-with" or "has-suppliers-located-in")

The more relation types are captured, the more options are there to infer buyer-supplier relations from different contexts. E.g. the statement “Company A is the only supplier of part P.” and the statement “Company B uses part P in their products.” taken together suggest a dependency of Company B on Company A.

B.2.1 Multi-labelling instead of multi-class classification

Instead of using a multi-class classification with mutually exclusive classes, the problem could have been formulated as a multi-labelling scenario. This would have increased complexity but would have also increased the expressive power of the labelling system. For instance, one label could have indicated if a buyer-supplier relation is present. A further label could have then indicated the direction of that relation. And even further labels could have signalled aspects, such as the stated or suggested confidence of the information.

While such a system increases the labelling complexity, it could in fact facilitate some annotation decisions since the annotation decision is split into multiple smaller and often independent ones. For instance, annotators may disagree on the direction or certainty of a buyer-supplier relation but not the fact that a buyer-supplier relation is present. The currently proposed system of class labels treats an “A supplies B” relation that is stated as uncertain the same as a certain “partnership” relation.

B.2.2 Additional aspects that could have been considered

A number of additional aspects can be considered for additional classes or different class definitions. In the following, these aspects are explained, discussed and reasons for *not* considering them in separate classes are provided.

The *grammatical tense* of a statement could represent important information. For the use of the supply chain map, it may make a difference if a company supplied, supplies, has been supplying or will supply another company. The tense, however, can generally be detected automatically using existing NLP libraries and this step could be performed at a later stage. Hence, human annotations for tense are not necessarily required and would add to the complexity of the task.

The *grammatical mood*, such as irrealis (“would have if the deal had taken place”) are not considered separately.

Information about the *time period* during which a statement is valid is not captured. This is more crucial than just detecting the grammatical tense. The following sentence is from a news article: “Michelin has been supplying tires to Gulf Air’s Airbus fleet for the last decade". Statements stretching across periods of time are especially valuable for verifying information from different sources.

Statements of *negated buyer-supplier relations* are not explicitly considered in the class definitions. A sentence may be explicitly negated, such as "company A does not supply company B". Following the proposed class definition, this case would fall into the category "None". The downside of the class definition is that information is lost. The fact that one company does not supply another may be relevant, e.g. to assess the veracity of other statements.

Language that implies a *change in the relations* and, hence, the supply chain structure is not used yet. Examples are sentence structures, such as "company A has been dropped as a supplier to company B", "company A no longer provides company B with..." or "company A has been won as a new customer for company B". To keep the annotation task at a manageable level, these cases are just treated based on what is the new state. The information that there was a change and information about the previous state is lost.

"Directness" is a further aspect cannot be captured by the proposed class definitions. In rare cases, the "directness" of a buyer-supplier relation is explicitly specified in a sentence. That is the sentence states if a supplier acts as a tier-1 or tier-2 supplier or potentially both. Figure B.2 shows such a sentence.

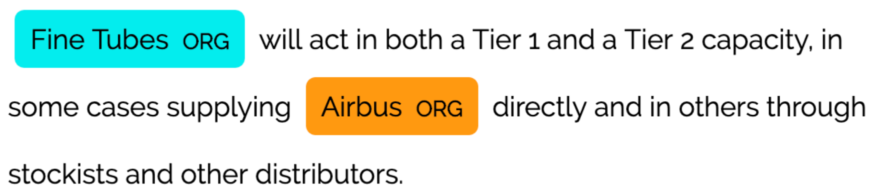


Fig. B.2 Example sentence found in the news dataset that specifies the supplier's tier

In some cases, the *number of alternative suppliers* is explicitly mentioned in a sentence. From a risk management perspective, so-called "sole" or "single-source" suppliers are especially critical. The sentence shown in Figure B.3 indicates such a relationship.

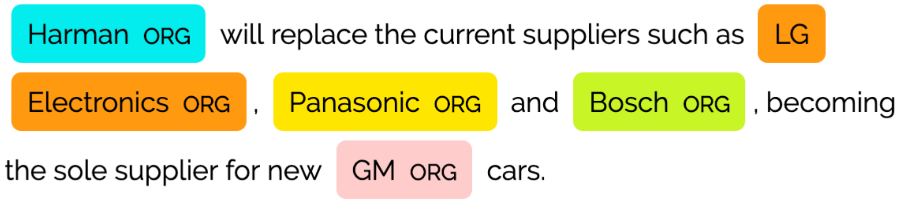


Fig. B.3 Example sentence found in the news dataset that specifies that a company acts as a sole supplier.

Questions and indirect speech are not captured separately at this point. Future extensions of this research could attempt to store this information.

The *continuum of (un)certainty* is not captured beyond just the separation certain and uncertain statements. However, the language may provide more fine-grained information, such as “highly likely”, “may”, “aims to become supplier”, etc.

Similarly, *negotiation stages* are not distinguished beyond the classes defined above. However, some industries have formal processes for contract negotiations that indicate an upcoming buyer-supplier relation, such as signing a Memorandum of Understanding (“MoU”) before actually signing an order, but are not yet buyer-supplier relations.

The *type of sale* is not further distinguished by the proposed class definition. That means, for instance, that primary services cannot be distinguished from secondary services, such as the outsourcing of HR functions. Similarly, it is impossible to distinguish the sale of an asset, such as a factory, from one company to another even though it is not a typical buyer-supplier relation. Also hard to distinguish are suppliers of financial funds, such as banks, when no transfer of ownership is involved. However, if a whole business is sold, this can be classified as an ownership relation between the sold business and the new owner.

B.2.3 Summary of options to increase the scope of relation extraction

Table B.1 provides specific examples how the scope could be increased along different dimensions.

Table B.1 Overview option for increasing scope of the relation extraction

Aspect	Options for increasing the scope
Entity types in a relation	• Named entities, e.g. names of organisations (ORG), products (PROD), locations (LOC) • Other elements, such as the provided part or material (e.g. steel), which are regular words and not capitalised unlike named entities
Number of entities considered in a single relation	• 2 (binary relations), e.g. ORG-ORG or ORG-PROD • 3 (ternary relations), e.g. explicit sub-tier relations (ORG-ORG-ORG) or “company produces part in location” (ORG-PROD-LOC) • Relations with more than 3 elements
Number and type of attributes of a relation	• Indicated (un)certainty or stages in the contract negotiation process, e.g. “may”, “in advanced talks to purchase” • Monetary value of the contract (if mentioned) • Sole (single-source) supplier or sole buyer (if mentioned) • Changes in the buyer-supplier relation • ...
Type of relations	• Buyer-supplier relations • Relations that describe the supply chain on a higher level (e.g. the number of a company’s suppliers located in a specific region) • Relations that do not describe buyer-supplier relations but collectively allow for the inference of buyer-supplier relations by reasoning using this background knowledge

B.3 Detailed methodology for sentence sampling

Approach A: Sampling of documents into 3 partitions One approach to address this trade-off is to sample documents into 3 separate partitions: one partition for random documents drawn from a general news dataset, a second partition for documents that are retrieved using keywords related to selected focus industries (aerospace and automotive), and a third partition for documents that are retrieved based on a search for company names in these focus industries. This way, the trade-off between the expected relevance of a sentence and its bias could in principle be steered by adjusting the proportion of each partition in the final sample.

Aerospace and automotive are chosen as focus industries. The rationale behind this decision is (a) that these industries are known for having complex and global supply chains, and (b) that their supply chains are assumed to be well-covered in general news. For the aerospace industry, for example, documents can be filtered for the existence of keywords, such as “aerospace”, “aircraft” or “planemaker”. If datasets also contain industry codes, these datasets can more accurately be filtered using these codes as opposed to keywords. The RCV1 Reuters dataset contains industry codes (“Thomson Reuters Business Classification”) and can more accurately be filtered using 23 of these codes instead. 50% of documents were sampled from the aerospace industry and 50% from the automotive industry. To filter documents by company names, a list of the top 100 global automotive company names and brands can be used as well as a list of the top 100 global aerospace and defence companies. Because headlines follow different linguistic rules than common running text, both headlines as well as the news content are considered. Relevant documents in each partition are subsequently segmented into individual sentences that would then be drawn randomly. Initial tests quickly revealed that, even in the more relevant data partitions, the proportion of positive sentences was still too small for an efficient annotation by humans.

Approach B: Manual collection of candidate positive sentences To address this problem but without compromising the overall human annotation, *in addition* to the already created dataset, candidates for positive sentences can be manually collected by multiple researchers and stored in a further data partition. To prevent biases, these positive sentences can not just be obtained via a Web search that used potential features, such as "supplies Toyota with". Otherwise, a classifier trained on the data would be biased towards the patterns the positive sentences were found with in the first place. Instead, the following strategies are deemed acceptable and can be used to collect the sentences:

- Using a Web search engine by using as a search term (a) a single company name or (b) the names of two companies of which one is known or merely suspected to supply the other.
- Manually analyse websites that tend to publish industry news, such as recent deals and partnerships.

In all of these cases, sentences are manually identified in the search result summaries, headings or the original articles that could describe a buyer-supplier relation, partnership or collaboration. Ambiguous sentences are not ignored but are also collected so that the overall dataset is rather too inclusive than too exclusive. Similar to the previous approach, the focus is on aerospace and automotive companies but any by-catch from other industries is also added to the collection.

The drawback of adding manually collected candidate positive sentences is the introduction of additional bias. This is unavoidable as it is a direct consequence of the objective of manually collecting sentences: The label distribution is no longer representative of a general news dataset. For instance, specific words, such as “award”, may become – what could be called – trigger words. If the word “award” is over-represented in the dataset in sentences that express a buyer-supplier relationship (supplier awards, awarding a contract), then a classifier may be prone to errors on previously unseen data where the word “award” is much less of an indicator of a buyer-supplier relation. Words that may result in a similar effect are good indicators of specific relation classes in such a limited training dataset but are not *always* reliable indicators, such as “deal”, “agreement”, “contract”, or “order”. To correct for these overfitting errors, sentences with a predicted buyer-supplier relation may later have to be fed back into the labelling process to improve precision. Once confidence scores have become sufficiently reliable, an active learning approach can be employed.

Benefits and drawbacks of automatic NER tagging To facilitate the subsequent annotation, sentences can be manually or automatically NER-tagged. Only sentences with two or more detected organisational named entities are admitted to the annotation process. In order to express a relationship, a sentence must at least contain two organisational mentions (if no co-reference resolution is used).

The benefits of using automated NER tagging are the following:

- Sentences with less than two organisational named entities can be automatically discarded before entering any manual annotation process.
- Annotators only have to classify relations instead of also identifying organisational entities.
- If annotators had to first detect the named entities, then the ambiguity of the task would have led inconsistent sets of entities for each sentence and annotator, and subsequent classification labels would no longer be comparable.

Drawbacks of automated NER tagging The automated NER tagging also has drawbacks:

- Automatic NER tagging performs well but is still imperfect and organisations were not detected, were erroneously detected, or were detected with incorrect segmentation boundaries. Manually tagging named entities is likely to be more correct.

- Especially for the dataset of candidate positive sentences, the automated NER tagging resulted in a loss of data: Undetected named entities lead to the discarding of perfectly valid positive examples.

- The common NER tag is "Organisation" (as opposed to "Company") which can also include non-company entities like governmental agencies or football clubs.

- Company names are often also used as a product specifier, such as "Boeing 737". In those cases, the whole string is likely to be detected as a product (PRODUCT) or miscellaneous (MISC) named entity. If not accounted for by additional processing steps, the information that a company name was stated would then get lost. If tagged manually, these instances could be recorded as organisational named entity mentions.

Appendix C

Details related to Chapter 7

This part of the appendix contains more detailed content related to Chapter 7 which addressed the industrial validation of the research presented in this thesis. The content provided below provides the detailed results of the achieved performance for each classification model.

C.1 Detailed results of the achieved classification performance

This section provides the detailed results for each classifier and corpus that have been summarised in Section 7.3.3.

Detailed results (initial corpus)

The following results have been achieved on the initial corpus.

Random dummy baseline classifier The performance of the random *dummy* baseline classifier can be found in Table C.1.

Table C.1 Classification results per class (averaged over 10 runs) – Random dummy classifier; initial corpus

	Accuracy	Precision	Recall	F_1 score
Class 0: None or reject	0.38	0.69	0.20	0.31
Class 1: B supplies A	0.76	0.07	0.20	0.10
Class 2: A supplies B	0.71	0.15	0.20	0.17
Class 3: ambiguous/undirected	0.76	0.06	0.20	0.09
Class 4: ownership/part-of	0.78	0.03	0.20	0.05
Micro-averaged				0.20

Stratified dummy baseline classifier The performance of the stratified *dummy* classifier can be found in Table C.2.

Table C.2 Classification results per class (averaged over 10 runs) – Stratified dummy classifier; initial corpus

	Accuracy	Precision	Recall	F_1 score
Class 0: None or reject	0.48	0.69	0.46	0.56
Class 1: B supplies A	0.89	0.07	0.05	0.06
Class 2: A supplies B	0.74	0.14	0.16	0.15
Class 3: ambiguous/undirected	0.88	0.06	0.07	0.06
Class 4: ownership/part-of	0.73	0.04	0.29	0.06
Micro-averaged				0.36

Support vector machine (SVM) without positional features The performance of the support vector machine (SVM) classifier without positional features can be found in Table C.3. The size of the vocabulary was 10,803 (initial corpus).

Table C.3 Classification results per class (averaged over 10 runs) – SVM without positional features; initial corpus

	Accuracy	Precision	Recall	F_1 score
Class 0: None or reject	0.73	0.80	0.82	0.81
Class 1: B supplies A	0.92	0.39	0.26	0.31
Class 2: A supplies B	0.82	0.38	0.42	0.40
Class 3: ambiguous/undirected	0.92	0.35	0.35	0.35
Class 4: ownership/part-of	0.97	0.47	0.35	0.40
Micro-averaged				0.68

Support vector machine (SVM) *with* positional features The performance of the support vector machine (SVM) classifier *with* positional features can be found in Table C.4. The size of the vocabulary was 10,803 (initial corpus).

Table C.4 Classification results per class (averaged over 10 runs) – SVM with positional features; initial corpus

	Accuracy	Precision	Recall	F_1 score
Class 0: None or reject	0.78	0.78	0.94	0.86
Class 1: B supplies A	0.94	0.61	0.19	0.29
Class 2: A supplies B	0.87	0.59	0.44	0.51
Class 3: ambiguous/undirected	0.94	0.49	0.21	0.29
Class 4: ownership/part-of	0.97	0.73	0.22	0.33
Micro-averaged				0.75

Multi-layer perceptron The class-wise classification performance for the MLP is shown in Table C.5. The results are shown for a batch size of 32. Different batch sizes (64, 128) were tested and resulted in similar performance levels (approx. 0.73). Since the process was stochastic, achieved performance levels could differ slightly with each new run. The table shows an average over 10 separate runs.

Table C.5 Classification results per class (averaged over 10 runs) – MLP; initial corpus

	Accuracy	Precision	Recall	F_1 score
Class 0: None or reject	0.79	0.80	0.92	0.86
Class 1: B supplies A	0.93	0.48	0.25	0.33
Class 2: A supplies B	0.85	0.50	0.42	0.46
Class 3: ambiguous/undirected	0.94	0.44	0.27	0.33
Class 4: ownership/part-of	0.97	0.92	0.19	0.32
Micro-averaged				0.74

BiLSTM The class-wise classification performance for the BiLSTM is shown in Table C.6. The results are shown for a batch size of 64. Different batch sizes (32, 128) were tested and resulted in similar performance levels (approx. 0.71 to 0.72). Since the process was stochastic, achieved performance levels could differ slightly with each new run. The table shows an average over 10 separate runs.

Table C.6 Classification results per class (averaged over 10 runs) – BiLSTM; initial corpus

	Accuracy	Precision	Recall	F_1 score
Class 0: None or reject	0.79	0.82	0.89	0.85
Class 1: B supplies A	0.92	0.35	0.14	0.20
Class 2: A supplies B	0.84	0.46	0.52	0.49
Class 3: ambiguous/undirected	0.93	0.36	0.23	0.28
Class 4: ownership/part-of	0.97	0.79	0.16	0.27
Micro-averaged				0.73

Detailed results (extended corpus)

The following results have been achieved on the *extended* corpus that includes *both* the initial corpus as well as the subsequently added annotated sentences.

Random dummy baseline classifier The performance of the random *dummy* baseline classifier can be found in Table C.7. The results only differ slightly compared to the results on just the initial corpus.

Table C.7 Classification results per class (averaged over 10 runs) – Random dummy classifier; extended corpus

	Accuracy	Precision	Recall	F_1 score
Class 0: None or reject	0.40	0.67	0.20	0.31
Class 1: B supplies A	0.77	0.05	0.20	0.08
Class 2: A supplies B	0.71	0.15	0.20	0.17
Class 3: ambiguous/undirected	0.75	0.08	0.20	0.11
Class 4: ownership/part-of	0.77	0.05	0.20	0.08
Micro-averaged				0.20

Stratified dummy baseline classifier The performance of the stratified *dummy* classifier can be found in Table C.8. The results only differed slightly compared to the results on just the initial corpus.

Table C.8 Classification results per class (averaged over 10 runs) – Stratified dummy classifier; extended corpus

	Accuracy	Precision	Recall	F_1 score
Class 0: None or reject	0.49	0.69	0.47	0.55
Class 1: B supplies A	0.91	0.06	0.06	0.06
Class 2: A supplies B	0.78	0.18	0.14	0.16
Class 3: ambiguous/undirected	0.85	0.07	0.09	0.09
Class 4: ownership/part-of	0.69	0.05	0.30	0.09
Micro-averaged				0.36

Support vector machine (SVM) without positional features The performance of the support vector machine (SVM) classifier without positional features can be found in Table C.9. The size of the vocabulary was 17,625 (extended corpus). The overall micro-averaged F_1 score was slightly lower than the same model trained and tested on the initial corpus.

Table C.9 Classification results per class (averaged over 10 runs) – SVM without positional features; extended corpus

	Accuracy	Precision	Recall	F_1 score
Class 0: None or reject	0.72	0.81	0.76	0.78
Class 1: B supplies A	0.92	0.29	0.39	0.33
Class 2: A supplies B	0.84	0.47	0.52	0.49
Class 3: ambiguous/undirected	0.91	0.44	0.45	0.44
Class 4: ownership/part-of	0.93	0.34	0.39	0.36
Micro-averaged				0.66

Support vector machine (SVM) *with* positional features The performance of the support vector machine (SVM) classifier *with* positional features can be found in Table C.10. The size of the vocabulary was 17,625 (extended corpus). The overall micro-averaged F_1 score was slightly higher than the same model trained and tested on the initial corpus.

Table C.10 Classification results per class (averaged over 10 runs) – SVM with positional features; extended corpus

	Accuracy	Precision	Recall	F_1 score
Class 0: None or reject	0.80	0.80	0.94	0.86
Class 1: B supplies A	0.96	0.65	0.26	0.37
Class 2: A supplies B	0.89	0.71	0.50	0.59
Class 3: ambiguous/undirected	0.93	0.58	0.54	0.56
Class 4: ownership/part-of	0.95	0.60	0.16	0.25
Micro-averaged				0.77

Multi-layer perceptron The class-wise classification performance for the MLP is shown in Table C.11. The results are shown for a batch size of 32. Different batch sizes (64, 128) were tested and resulted in slightly lower performance levels (approx. 0.74 to 0.76). Since the process was stochastic, achieved performance levels could differ slightly with each new run. The table shows an average over 10 separate runs.

Table C.11 Classification results per class (averaged over 10 runs) – MLP; extended corpus

	Accuracy	Precision	Recall	F_1 score
Class 0: None or reject	0.81	0.82	0.91	0.86
Class 1: B supplies A	0.95	0.48	0.35	0.40
Class 2: A supplies B	0.88	0.63	0.57	0.60
Class 3: ambiguous/undirected	0.93	0.60	0.46	0.52
Class 4: ownership/part-of	0.95	0.52	0.26	0.35
Micro-averaged				0.76

BiLSTM The class-wise classification performance for the BiLSTM is shown in Table C.12. The results are shown for a batch size of 64. Different batch sizes (32, 128) were tested and resulted in slightly lower performance levels (approx. 0.68 to 0.72). Since the process was stochastic, achieved performance levels could differ slightly with each new run. The table shows an average over 10 separate runs.

Table C.12 Classification results per class (averaged over 10 runs) – BiLSTM; extended corpus

	Accuracy	Precision	Recall	F_1 score
Class 0: None or reject	0.79	0.83	0.86	0.84
Class 1: B supplies A	0.93	0.29	0.29	0.29
Class 2: A supplies B	0.88	0.60	0.55	0.57
Class 3: ambiguous/undirected	0.92	0.47	0.55	0.51
Class 4: ownership/part-of	0.94	0.42	0.26	0.32
Micro-averaged				0.73