Understanding and exploiting social media data to support decision making in fashion and apparel supply chains

Samaneh Beheshti-Kashi

Understanding and exploiting social media data to support decision making in fashion and apparel supply chains

von Samaneh Beheshti-Kashi

Dissertation

vorgelegt am Fachbereich 3 der Universität Bremen
zur Erlangung des Grades
eines Doktors der Ingenieurwissenschaften
– Dr.-Ing. –

Erstgutachter: Prof. Dr.-Ing. Klaus Dieter Thoben
Zweitgutachter: Prof. Dr. Michael Lawo

Tag der mündlichen Prüfung: 30.11.2020

Acknowledgement

I completed this dissertation as a doctoral candidate and a member of the International Graduate School for Dynamics in Logistics (IGS). During this journey different people have contributed to the completion of my thesis. I would like to thank my supervisor Prof. Dr.-Ing. Klaus-Dieter Thoben for his supervision, feedback and discussions. Also I would like to thank Prof. Dr. Michael Lawo for his supervision. I wish to thank the IGS for the funding. But most importantly, I wish to express my deepest gratitude to Dr.-Ing. Ingrid Rügge for her invaluable support. The IGS enables research and friendships beyond genders, cultures and borders. Also I wish to express my sincere appreciation to Prof. Dr. Hamid Reza Karimi for all the fruitful discussions. A very special thanks goes to Prof. Dr. Jürgen Pannek for his guidance. His valuable input played an essential role in the completion of this work. I would like to thank the BIBA - Bremer Institut für Produktion und Logistik GmbH, in particular, my colleagues at the department of Data Analytics and Process Optimization.

Special thanks also to Prof. Dr. Aseem Kinra for giving me the opportunity of researching at the Copenhagen Business School and diving deeper into the field of Text Mining. But also for all the discussions, feedback and a wonderful experience of supervising two extraordinary master students. I would also like to thank Prof. Dr. Tsan-Ming Choi for providing me the chance of a research stay at The Hong Kong Polytechnic University. That stay enabled me a deeper understanding of the fashion and apparel supply chain.

I wish to express my deepest gratitude to my colleagues and friends Darja Wagner, Moritz Quandt and Thies Beinke for all the discussions and valuable input on the one hand, but most importantly for their support and invaluable assistance through the different stages of my doctoral study. Similarly, sincere appreciation to my dearest friends: Marcella, Sylvia, Molin, Fang, Ping, Haniyeh, Kishwer, Elaheh and Juan. And last, but not least, I wish to acknowledge the love, support, patience and understanding of my husband, my parents, my family and my dearest friends.

ZUSAMMENFASSUNG

Wertschöpfungsketten in der Bekleidungsindustrie zeichnen sich durch eine hohe Komplexität und Dynamik aus. Über die gesamte Lieferkette hinweg sind Entscheidungen der jeweilig Beteiligten mit hohen Unsicherheiten behaftet, da in der Regel nicht ausreichend Daten zur Verfügung stehen, die als Entscheidungsgrundlage für zukünftige Bedarfsprognosen dienen könnten. In Ergänzung dieser Datenbasis werden zusätzliche Informationsquellen, wie beispielsweise historische Verkaufszahlen, genutzt. Diese stehen für modische Bekleidungsartikel jedoch nur selten zur Verfügung. Mit dem Aufkommen von sozialen Medien bietet sich für die Beteiligten der Bekleidungswertschöpfungskette eine weitere potentielle Datenquelle, die zur Erweiterung ihrer Entscheidungsbasis führen kann.

Das Extrahieren und Integrieren von Daten und Informationen aus sozialen Medien in Entscheidungsprozesse soll hierbei nicht nur aus rein technischer Perspektive betrachtet werden, da soziale Medien bestimmte Merkmale besitzen, die bei der Verwertbarkeit von Daten eine entscheidende Rolle spielen. Neben des hohen Volumens, der Geschwindigkeit, sowie der Vielzahl an Datenarten, nimmt die Verlässlichkeit von Daten aus sozialen Medien eine entscheidende Rolle ein, wenn es um die Unterstützung bei Entscheidungen in der Lieferkette geht. Diese werden allerdings bei der Extraktion und Nutzung von Daten aus sozialen Medien kaum in Betracht gezogen. Besonders gilt das für die meisten bestehenden Ansätze, die sich mit der Nutzung von Informationen aus sozialen Medien im Kontext von Wertschöpfungsketten der Bekleidungsindustrie befassen.

An diesem Punkt setzt die vorliegende Dissertation an. Sie sieht den Handlungsbedarf in der Gestaltung einer Methodik zur Verwertbarkeit von Daten aus sozialen Medien zur Unterstützung von Entscheidungsprozessen von Bekleidungswertschöpfungsketten. Hierbei wird der Fragestellung nachgegangen, ob und inwieweit Daten aus den sozialen Medien, als ergänzende Informationsquelle für Entscheidungen in Bekleidungswertschöpfungsketten genutzt werden können. Dabei orientiert sich die Arbeit als übergeordnete Forschungsmethodik am Ansatz des ”Design Science Research”. Das dabei entwickelte Artefakt ist als Prozessmodell gestaltet, und beruht auf Anforderungen, die auf Merkmalen von Bekleidungswertschöpfungsketten, den Informationsbedürfnissen von Entscheidungsträgern der Lieferkette, sozialen Medien, Text Mining sowie der Prozessmodellierung basieren. Entsprechend beinhaltet das Prozessmodell die folgenden vier Layer: Process Layer, Informationsource

Layer, Social Media Layer sowie das Text Mining Layer. Die Realisierung des Prozessmodells erfolgt in der Modellierungssprache Business Process Model and Notation 2.0.

Die Anwendbarkeit des Prozessmodells wird im Rahmen eines Anwendungsfalls an der Produkteigenschaft Farbe untersucht. Als Datenbasis dienen deutschsprachige Blog-Daten. Um das Potenzial von sozialen Medien als ergänzende Informationsquelle aufzeigen zu können, muss ein ökonomischer Zeitversatz zwischen Daten aus sozialen Medien sowie realer Verkaufsdaten gezeigt werden. Durch experimentelle Ergebnisse wird dieser Zeitversatz zwischen Blog- und realen Verkaufsdaten über verschiedene Lieferkettenstufen hinweg nachgewiesen. In Anlehnung an den ”Design Science Research Process Model”- Ansatz wird der Nutzen des Prozessmodells anhand der Verwendung des Artefakts evaluiert, welches die Einbeziehung realer Nutzer impliziert. Die Befragung von Experten aus der Lieferkette hat gezeigt, dass das Potential der extrahierten Informationen zur Unterstützung bei Entscheidungen besonders in frühen Stufen der Wertschöpfungskette gesehen wird.

Abstract

Fashion and apparel supply chains are highly dynamic, complex and challenged by high uncertainties. Stakeholders are typically confronted with incomplete information at the time when they do require it. For complementing this lack of information, additional sources are explored involving historical sales data. However, for fashion items historical sales data are often sparse. Ever since the emergence of social media, its relevance for fashion and apparel supply chains keeps rising. The exploitation of social media should not only be considered as a purely technical task, as social media comes with a range of characteristics. Being classified as a big data source, the 5 V's apply also to social media. In addition to volume, velocity, variety of data, the veracity feature is fundamental in terms of using the extracted information for supporting decision processes. Nevertheless, most existing approaches which use social media data for fashion and apparel problems do not consider neither these characteristics nor fashion and apparel supply chain characteristics. In addition, stakeholders' needs and requirements are neglected when exploiting social media data. This said, the present research claims the necessity of a methodology for exploiting social media data for fashion and apparel supply chain decisions considering the stakeholders' perspective. In this manner, this thesis poses the question if and how social media can be used as an addition information source to support fashion and apparel supply chain decisions.

For targeting this research question, as an overall framework a "Design Science Research" approach is selected. Following the "Design Science Research Methodology Process Model", this thesis first designs and develops an artifact, designed as a process model enabling an understanding of social media data and a systematical exploitation of textual social media data to support fashion and apparel SCs decisions, secondly, shows its use in a demonstration case and finally evaluates its utility by the judgment of experts working in the field of fashion and apparel supply chains. For the development of the process model, characteristics of fashion and apparel supply chains and social media, supply chain stakeholders, as well as text mining and process model design features are considered. Accordingly, the process model consists of four layers: the Process Layer, the Information Source Layer, the Social Media Layer and the Text Mining Layer. The realization of the process model is conducted in Business Process Model and Notation 2.0.

Following the "Design Science Research Methodology Process Model", the utility of the

artifact designed is shown in a demonstration by implementing a case study around the product feature colour. The process model is applied on blog data. In order to show the feasibility of social media data as an additional information source, the existence of an economically advantageous time offset between sales and blog data is examined. The findings show that it is possible to identify an advantageous economic value for two colour groups, even over different supply chain stages. Having demonstrated the utility of the process model by the case study, the evaluation is performed based on its results. This involves comparing the objectives of the artifact with the oberserved results generated from its use. An ex post naturalistic approach is applied and manufacturers and retailers from fashion and apparel supply chains are surveyed. The evaluation has demonstrated that supply chain stakeholders see an added value from the extracted information in particular when available two or four months in advance of the selling season.

Contents

List of Tables

List of Figures

Abbreviations

The following abbreviations are used in this dissertation:

AHP	Analytic Hierarchy Process
ANN	Artificial neural network
API	Application Programming Interface
BPMN 2.0	Business Process Model and Notation
CA	Content Analysis
DM	Data Mining
DSR	Design Science Research
DSRM	Design Science Research Methodology
DSS	Decision Support System
GfK	Growth from Knowledge
IQ	Information Quality
IQA	Information Quality Assessment
KNIME	Konstanz Information Miner
NLP	Natural Language Processing
OECD	Organisation for Economic Co-operation and Development
POS	Part-of-Speech
RSS	Rich Site Summary
SC	Supply Chain
SCM	Supply Chain Management
SCA	Supply Chain Analytics
SMA	Social Media Analytics
SNA	Social Network Analysis
TM	Text Mining
UGC	User Generated Content
XML	Extended Markup Language

1 INTRODUCTION

Big data and big data sources are increasingly discussed as a formidable opportunity for decision makers in logistics and Supply Chain Management (SCM) (Waller and Fawcett, 2013). Hence, the usage of big data analytics in supply chain (SC) has been studied by researchers (e.g. (Awwad et al., 2018; Fosso Wamba et al., 2018; Tiwari et al., 2018)) and practitioners. Big data analytics applies advanced analytics techniques such as Data Mining (DM) (Russom et al., 2011) or Text Mining (TM) and involves the analysis of large and diverse data sets for uncovering hidden values applied to increase business benefits (Hofmann, 2017). The usage of analytics in SCs per se is not new for decision makers (Souza, 2014). The utilization of big data analytics, with its variety of sources and techniques such as the Internet of Things, cloud computing, wireless sensor networks and social media, is considered to have a huge impact on SCM and logistics (Tiwari et al., 2018; Waller and Fawcett, 2013). The term Supply Chain Analytics (SCA) has emerged in the literature referring to the usage of advanced big data analytics in SCM (Souza, 2014; Wang et al., 2016). Souza (2014) states that SCA focusses on the use of information and analytical tools to make better decisions regarding material flows in the SC. The use of big data analytics is explored in a range of SCM areas such as SC network design (Gereffi and Memedovic, 2003; Piller and Müller, 2004), product design and development (Barnes and Lea-Greenwood, 2006; Hines and Bruce, 2007), demand planning (Bhardwaj and Fairhurst, 2010; Bruce and Daly, 2006; Cheng and Choi, 2010), procurement (Birtwistle et al., 2003; Wang et al., 2016), production (Ghemawat et al., 2003; Zülch et al., 2011), inventory (Christopher, 2000; Wang et al., 2016), logistics and distribution (Lee and Chen, 1999), transportation (Da Silveira et al., 2001), and SC agility and sustainability (Waller and Fawcett, 2013; Yeung et al., 2010). For obtaining a detailed review of big data analytics in SCM consult Awwad et al. (2018); Fosso Wamba et al. (2018); Russom et al. (2011); Tiwari et al. (2018). The large and diverse data sets used in big data analytics have also been generated by new technologies such as social networks, smart devices or RFID tags (Hofmann, 2017; Souza, 2014). In addition to typical structured data, social media, and other online activities generate a large volume of unstructured data. A SC manager cannot only have access to this data (Sanders and Ganeshan, 2015) but also

benefits from it in various processes, for instance in forecasting (Tiwari et al., 2018) or other processes lacking historical data, e.g., new product development. A effective exploitation of big data sources might lead to multitude improvements of SC processes, for example, the establishment of responsive and agile SCs and the proactive response to SC risks. Moreover, the triangulation of different data sources is highly valuable. In this context, data from social media, news, events, weather sources or other static and dynamic data should be combined and applied in SC decisions (Fosso Wamba et al., 2018). Different authors agree that social media based sources are typical examples of big data sources (Choi, 2018; Guellil and Boukhalfa, 2015; Lin, 2015; Sebei et al., 2018).

While the application of big data sources in logistics and SCs is still in its early phase (Hofmann, 2017), its potential has already been shown is some industries such as consumer electronics (Akundi et al., 2018) or apparel (Cui et al., 2018). Though, in general, social media research in the context of SCM needs still further investigations (Swain and Cao, 2019).

1.1. Motivation and Problem Statement

The apparel industry is considered to be one of the largest industries in the global economy (Seram et al., 2016). Fashion and apparel SCs are characterized by short life-cycles, high-volatility, low-predictability and high-impulse purchasing (Christopher et al., 2004). Production plans have to be placed several months before the start of the season because most production plants are located in Asian countries while the target countries are mostly in Europe (Mostard et al., 2011). Therefore, the time-to-market for fashion products is longer compared to their short selling period. Moreover, the demand for fashion and apparel products is often unstable and influenced by external factors such as sales promotion or consumer purchasing power (Thomassey, 2014). Fashion markets are further characterized by a large product variety and dynamic changes with respect to style or colour (Hui and Choi, 2016). As the purchasing of items is highly impulse-driven, the availability of the products is crucial (Christopher et al., 2004). However, the particularities of fashion markets coupled with the high seasonality of apparel products, makes accurate forecasting of fashion products difficult. Inaccurate and unreliable sales forecasts prevent a proactive response to market changes and often cause poor production planning or lost orders (Wong, 2013). In contrast, more accurate sales forecasts will have fundamental benefits for different SC operations such as reduced safety inventory costs or more consistent delivery and production planning (Cui et al., 2018; Stevenson and Spring, 2007).

Decision-making in fashion and apparel SCs is thus different in comparison to other industries. Although product availability is fundamental (Christopher et al., 2004), demand information is often hardly available (Tokar et al., 2014). Decision makers are, therefore, confronted with incomplete information.

These peculiarities regarding fashion and apparel SCs lead to ubiquitous uncertainties (Flynn and Lu, 2016) and can have different sources. Still, essential sources of uncertainty are related to quantities, timings and specifications of customer demand. This is the primary cause of the bullwhip effect (Stevenson and Spring, 2007). To deal with this uncertainty, stakeholders typically use additional information sources for enriching their decision base. In this regard, the literature demonstrates that upstream order information profits from integrating downstream information such as point of sale data (Cui et al., 2015; Gaur et al., 2005). The more uncertainty exists, the more information needs to be collected to reach a given objective (Bode et al., 2011). Acquiring suitable information is then highly crucial for the SC to function effectively (Flynn and Lu, 2016). This information should involve customer demand, sales forecasts, order status, inventory levels, capacity availability, lead times and quality (Stevenson and Spring, 2007). Since access to historical data provides information about the past and may enable information about the future, it constitutes the basis for quantitative forecasting (Sanders and Manrodt, 2003). However, in the case of fashion and apparel, a lack of historical data exist as most fashion items are ephemeral (Thomassey, 2010) and, if available, challenging to use for predicting future designs (Singh et al., 2019).

The literature discusses different approaches to deal adequately with the difficulties of the fashion industry concerning accurate forecasting, and there are mainly two research streams. The first research stream focuses on the adequate integration of expert judgement and its combination with statistical forecasts (Belvedere and Goodwin, 2017; Davydenko and Fildes, 2014). The use of experts incorporates market intelligence which is otherwise not included in the statistical algorithms (Belvedere and Goodwin, 2017) and is followed by practitioners in different industries, including the fashion industry (Davydenko and Fildes, 2014; Fildes et al., 2009). The second stream examines the use of statistical methods, the combination of advanced methods, as well as the comparisons of their performances (Au et al., 2008; Sun et al., 2008; Thomassey and Happiette, 2007; Wong and Guo, 2010). However, most forecasting literature focuses on the development and integration of algorithms (Hofmann and Rutschmann, 2018). As these advanced methods are hard to use for practitioners (Thomassey, 2010), most fashion and apparel companies apply first basic statistical methods for calculating so-called baseline forecasts using different tools (Thomassey, 2014) and then adjust the baseline forecast according to individual decision makers (Thomassey, 2014; Tokar et al., 2014; Wong, 2013). Explanatory factors such as weather data, competition or calendar data are typically included in these adjustments (Thomassey, 2010). While a large number of apparel companies follow this two-step procedure, it has several drawbacks. Firstly, the adjustment of the baseline forecasts is highly subjective. Secondly, relying on only a few individuals makes the companies dependent and inflexible. Should employees leave the organization, their expertise is also gone (Wong, 2013), as it cannot be stored in a system(Hofmann and Rutschmann, 2018). Thirdly, not all relevant variables can be included, making analysis highly complicated and imprecise (Thomassey, 2014). This goes along with

the limited ability of decision makers to process such large amounts of information (Sanders and Manrodt, 2003).

This is where the use of additional information sources and their adequate handling comes into play. As mentioned above, social media is considered to be one promising source in the SCA context. This thesis is motivated by the increased relevance of social media and advocates its value for fashion and apparel SCs for several reasons. Fashion is specially discussed on social media, and many fashion blogs publish different fashion-related topics. User numbers have increased on different social media, and customers share information on products and services and their product preferences on these platforms. Besides, the user's purchasing decisions are increasingly impacted by social media friends products references (Cui et al., 2018) or influencers (Lee and Watkins, 2016). Considering the relationship between customer sentiment and demand, it is stated that their development will have a similar form over time (Curtin, 1982). In other words, a potential measurement of customer sentiment enables a better prediction of the development of future demand. While capturing these sentiments used to be equated with high costs and resources, accessing social media might allow this measurement beforehand (Hofmann, 2017).

Social media data is characterized as follows. Vast volumes of social media data is generated (volume) and published in real-time (velocity) in different type of data (variety) often by the person who is not identified. The latter results in the fact that social media data is often considered to have believability issues (veracity) (Stieglitz et al., 2018). The fifth characteristic of social media value refers to the ability of the decision makers to have adequate knowledge to harness the data. Thus, social media can be classified as a big data source (Guellil and Boukhalfa, 2015; Lin, 2015) since it complies with the 5 V's, i.e., volume, velocity, variety, veracity and value (Stieglitz et al., 2018). For harnessing social media data, it is crucial to consider all of these characteristics, as each of them leads to different challenges (Agarwal and Yiliyasi, 2010). The veracity attribute especially challenges the exploitation of social media data. The quality assessment of social media data plays a fundamental requirement while harnessing data from social media-based sources. Huge impacts on data quality were brought already by the internet, as collecting and sharing information had been fundamentally changed. This impact is increased through the User-Generated-Content (UGC) published on social media (Borchers, 2009). A quality-based assessment of social media data turns more crucial when the extracted data is used for decision-making since poor data and information quality (IQ) have an impact on it (Redman, 1998). Data quality research (Wang and Strong, 1996) suggests a range of dimensions for assessing the quality of data. However, these dimensions are developed to be applicable to transactional data and not on social media data.

The utilisation of social media data, and in particular, unveiling its potential value, is linked to additional challenges for researchers and practitioners. The main challenge is in the nature of most data generated through social media as social media data is typically in the

form of unstructured data (Chan et al., 2017), either in the form of textual or multimedia data. It is a challenge because companies are typically confronted with structured data Cui et al. (2018). Like other big data sources, social media-based sources can only be of economic value in SCM if they are exploited adequately, generate useful insights for practitioners, and are reliable and trustworthy. In particular, the predictive validity and associated forecasting risk of the derived information need to be analysed further. Researchers need to feel more confident in using such data sources and respective methods (Wenzel and Van Quaquebeke, 2018). How to generate such a value out of the information remains a challenge (Chan et al., 2017).

Given the peculiarities of apparel products, fashion sales forecasting is highly complex and specific. For implementing an apparel forecast system, in addition to expertise in the field of forecasting, it is also necessary to have a deep understanding of the challenges and operations of the fashion industry and its SC processes (Thomassey, 2014). To tackle the entire SC, it can therefore be stated that there is a need for a systematic methodological approach to the integration of social media as an additional information source into fashion and apparel SC decisions which considers the characteristics of social media, the peculiarities of fashion and apparel SCs as well as the stakeholders' information needs and requirements.

1.2. Research Objective, Research Questions and Research Approach

The main objective of this research is the development of a methodology for the exploitation of social media data to support fashion and apparel SC decisions. The stakeholder perspective represented by its information needs and the harnessing of social media data for a given information need are specifically in the focus of this dissertation. This is done by introducing a process model designed as a nominal process for the exploitation of social media data as a support for decision makers dealing with social media data, incorporating the decision-makers' perspective and social media characteristics into the process model. Another objective is to demonstrate the utility of the process model by experiments illustrating the potential value of social media data for the participants of fashion and apparel SC stakeholders

Based on these objectives, the dissertation answers the following main research question:

Considering stakeholder added-value, how can social media data be exploited to support decision processes in fashion and apparel SCs?

For answering this research question, three sub-questions are formulated:

- *Research Question 1 (RQ1):* Which properties of real-world commercial data and social media data are sufficient for the existence of an economically advantageous setting in SCs?

- *Research Question 2(RQ2):* Considering a SC stakeholder perspective, how can a method be designed to exploit social media and obtain information relevant to the decision process?
- *Research Question 3 (RQ3):* Given these properties and information for fashion and apparel SCs, which stakeholders, processes and decisions are potentially impacted?

RQ1 is stated for elaborating the sufficient properties which have a real added-value for the stakeholder in the form of an economic advantage. Defining the sufficient properties of commercial and social media data related to the SC will enable setting out the added value of fashion and apparel SC participants and decision makers. Conducting a literature review on fashion and apparel SC characteristics gives the answer to RQ 1. The answer to RQ1 is only the first step in fulfilling the main research question. While RQ 1 targets to establish the sufficient conditions required for responding to the main question, RQ 2 addresses the development of a methodology to exploit social media for obtaining information relevant to fashion and apparel decision processes. For laying the foundation for this methodology, literature reviews are conducted for elaborating both characteristics of fashion and apparel SCs and social media first. Second, current approaches to assessing and exploiting social media data are analysed, and the use of social media in fashion and apparel SCs is examined. These synthesised characteristics and approaches lay the base for deriving requirements for the proposed methodology, which enables an understanding of social media data and a systematical exploitation of textual social media data for supporting fashion and apparel SC decisions. For targeting RQ 3, and thus illustrating the utility of the proposed process model, a case study is framed by establishing a demonstration on a fashion and apparel decision process. In this regard and as an overall frame, it is defined to follow a posteriori analysis performed after the point of sale. It is, however, a crucial step as it is fundamental to explore a potential connection between real-world sales data and social media data. To be of economic value, e.g., in forecasting, ordering or manufacturing, the similarity must be obtainable before the point of sale and should be accompanied by a respective reliability factor. The process model is thus applied to the framed case study. Its utility is illustrated by conducting an evaluation by involving both real-world data and decision makers from the fashion and apparel SC.

Figure 1.1 shows the steps followed to answer the research questions and achieve the goal of this dissertation. This section only serves as a broad overview of the research approach. Chapter 2 provides an in-depth description of the research methodology used in this dissertation.

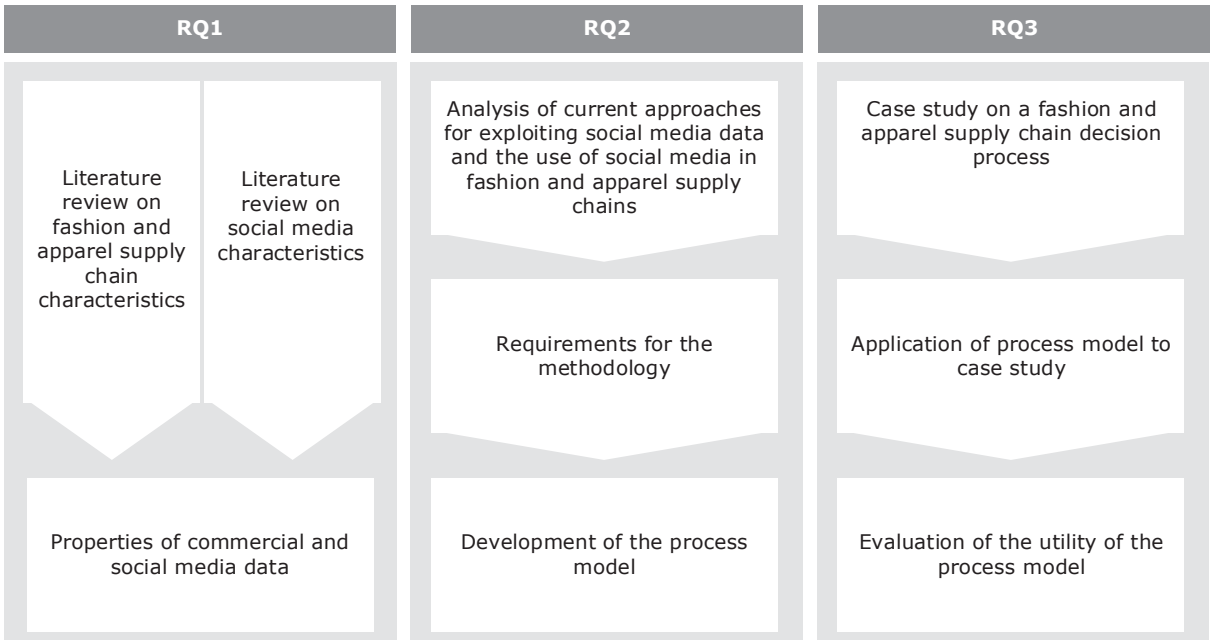


Figure 1.1.: Overview research approach

1.3. Structure of Dissertation

In this section, the structure of the thesis is elaborated, as illustrated by Figure 1.2. Chapter 1 sets the research framework of the present thesis by motivating this research and elaborating the problem statement in Section 1.1, outlining the research objective and deriving the research questions in Section 1.2. In the last section, the structure of the thesis is described.

Chapters 2 and 3 lay the foundation of this research. To this end, Chapter 2 outlines the research methodology, starting with an overview on design science research (DSR) and the design science research methodology (DSRM) in 2.1, illustrating the application of the DSRM on the objective of the presented thesis, showing the approaches for the design and development in 2.2, for the demonstration in 2.3 and for the evaluation in 2.4. The structure of the thesis is based on the DSRM process presented in Chapter 2.

Chapter 3 gives an overview of the foundation of the research by pointing out the relevant fields tackled in the research. Thus, Chapter 3 consists of the following sections. An overview on fashion and apparel SCs focussing on the different SC participants, their decision processes, and their information needs is provided by Section 3.1. This is followed by Section 3.2 introducing social media characteristics, challenges and quality assessment mechanisms. The last section of Chapter 3, outlines the most relevant research dealing with the use of social media for fashion and apparel SCs (see Section 3.3).

In Section 4.1, requirements for the process model are derived based on fashion and apparel SCs and stakeholders, social media characteristics, TM and process model design elements. Based on these requirements, the process model is designed and developed in Chapter 4. The process model consists of four layers, namely, the Process Layer, Information Source

Layer, Social Media Layer, and the Text Mining Layer. These layers are described successively in the Sections 4.2 to 4.5.

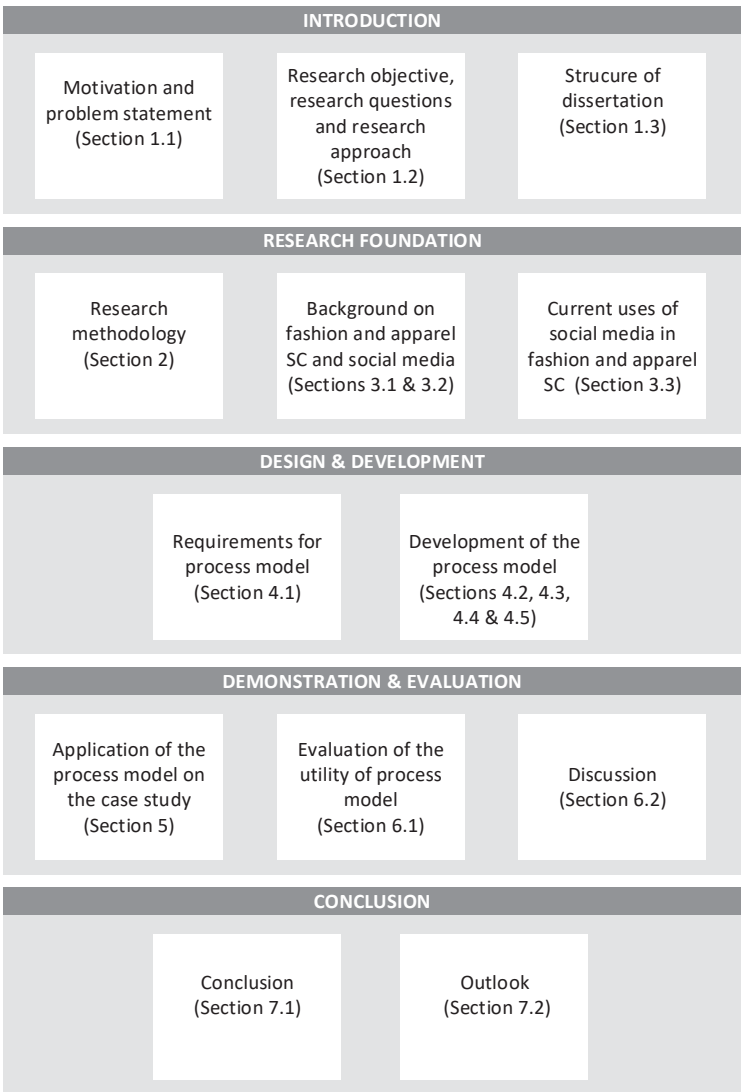


Figure 1.2.: Structure of the dissertation

For demonstrating the use of the developed process model, in Chapter 5 the process model is applied to a real-world case. To this end, the case study is framed in Section 5.1. The remaining sections of Chapter 5 are structured in accordance with Chapter 4. Thus, Section 5.2 illustrates the applied Process Layer, Section 5.3 focusses on the information sources, the applied Social Media Layer is presented in Section 5.4 and lastly, the Text Mining Layer in Section 5.5. Based on the generated results by the use of the process model in the case study, its utility is evaluated by involving potential real-world users. The findings of the evaluation are presented in Section 6.1. A discussion of the findings is provided by Section 6.2. The thesis will close with a summary and conclusion and a further research section in Chapter 7.

2 Research Methodology

This chapter introduces the research methodology used in this thesis. Chapter 2 is divided into four sections. In Section 2.1 DSR principles and guidelines are described. Furthermore, the DSRM process model by Peffers et al. (2007) is presented. The DSRM process model constitutes the overall framework for this thesis. As the last point in Section 2.1, the application of the DSRM in the context of this research is illustrated. Section 2.2 demonstrates the applied procedure for the design and development of the artifact. Furthermore, Section 2.3 describes the demonstration approach. In Section 2.4 the evaluation approach is elaborated, which is mainly based on the four-step method for DSR evaluation research design introduced by Venable et al. (2012). [1]

2.1. Design Science Research Methodology

Having "its roots in engineering and the sciences of the artificial (Simon, 1996)" (Hevner et al. 2004, p. 76), DSR advocates the creation and evaluation of IT artifacts for solving identified organizational problems (Hevner et al., 2004). This major principle of DSR demonstrates the problem-solving paradigm targeted by it (Hevner et al., 2004). Hevner et al. (2004) propose seven guidelines for the conduction of DSR. They point out that "the creation of an innovative, purposeful artifact" (Hevner et al. 2004, p. 82) is required. The artifact needs to be designed for a specific problem, and its utility needs to be demonstrated by means of evaluation. In addition, the novelty of the artifact plays a vital role in DSR. As a further guideline, it is recommended that rigorous methods are applied in the development and evaluation of the artifact. As the next guideline, the search process of an effective design is focussed, which involves an iterative approach and advocates knowledge of the application and solution domain. The final guideline claims an effective communication of the results to "technology-oriented as well as management-oriented audiences" (Hevner et al. 2004, p. 90).

[1] Some content of this chapter is published in Beheshti-Kashi and Kinra (2020)

Peffers et al. (2007) argue that previous research had defined the principles of DSR as in Hevner et al. (2004), its targeted objectives as well as guidelines for conduction as in Hevner et al. (2004) or in Fulcher and Hills (1996), and justification as in Nunamaker Jr et al. (1990) or in Walls and Sawy (1992). However, considering typical characteristics of a methodology, Peffers et al. (2007) state that so far, research missed to introduce a procedure proposing a process for conducting DSR. Only the two characteristics principle and practise rules were focussed on in the DSR context. A crucial remark is that the provided process requires to be generally accepted (Peffers et al., 2007). Based on existing principles and guidelines of DSR, Peffers et al. (2007) introduce the DSRM process model as demonstrated in Figure 2.1.

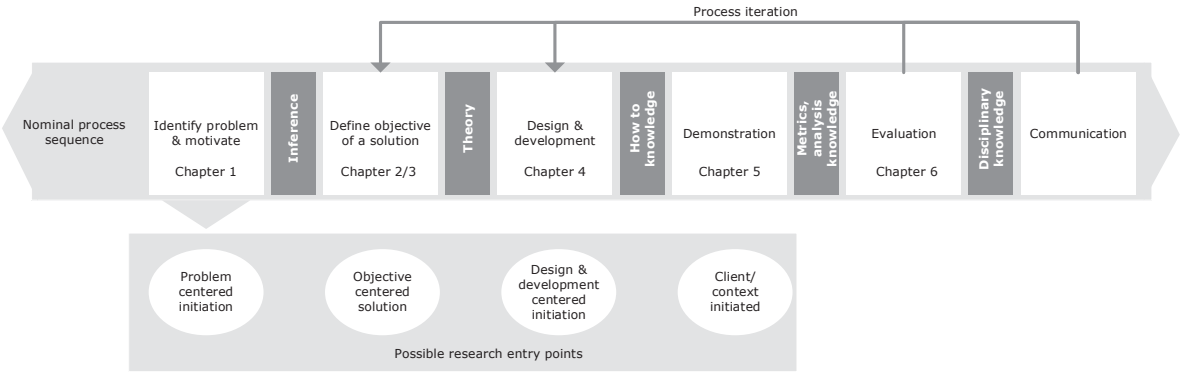


Figure 2.1.: Design Science Research Methodology (DSRM) Process Model based on Peffers et al. (2007)

The DSRM is introduced by means of a process model and defines six activities: identify problem & motivate, define objectives of a solution, design & development, demonstration, evaluation and communication. Once the problem has been defined, and the research has been motivated, the objective of the solution is formulated. Based on this, the articfact is designed and developed. Within the demonstration, a suitable context should be derived, and the artifact used to solve the problem. After that, the utility of the artifact is evaluated. The last step of a DSR is its communication, which is realized through scientific or professional publications. Although the process is structured in a nominal sequence, Peffers et al. (2007) advocate a context-dependent choice for starting the research based on the initial idea. That is, the DSRM process defines four different entry points to the research that are problem-centered initiation, objective-centered solution, design & development-centered initiation as well as client/context initiated.

Applying the DSRM process model to the current research (see, Figure 2.1), the entry point of this thesis is the problem-centred initiation. In this way, the DSRM is followed in its nominally sequential order starting with identifying the specific problem and illustrating the relevance of the importance of the research. This is conducted in Chapter 1. Based on the problem definition, the objective of a potential artifact is elaborated in Chapter 3. According to Peffers et al. (2007) the objective of the artifact may be formulated quantitatively or qual-

itatively. The former involves defining terms for demonstrating the superiority of the artifact compared to existing solutions. The latter implies a "description of how the new artifact is expected to support solutions to problems not hitherto addressed" (Peffers et al. 2007, p. 12). The design & development involved in the creation of the artifact is conducted in Chapter 4. This is followed by the demonstration of utility of the artifact in Chapter 5, and its evaluation in Chapter 6. Finally, communication activity is conducted through the completion and publishing of this thesis.

2.2. Design and Development

Design and development is the third activity in the DSRM process, and it includes the creation of the artifact (see Figure 2.1). An artifact can have different forms such as constructs, models, methods or instantiations (Hevner et al., 2004). Besides the actual development of the artifact, this activity also includes defining the functionality and architectures of the artifact.

Requirements definition

Figure 2.2 illustrates the approach for defining the requirements to be considered for the design and development of the targeted methodology.

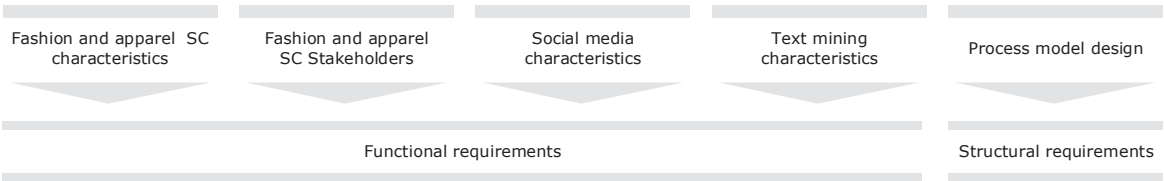


Figure 2.2.: Approach for deriving requirements for developing the process model

For deriving requirements, literature provides different approaches and taxonomies on the different types of requirements (Laplante, 2013). In this research, we differentiate between functional and structural requirements. By formulating functional requirements, services of a system are documented, and the reaction of input is dealt with (Laplante, 2013). Applying this to the development of the targeted methodology, the formulation of functional requirements will mainly focus on how to extract data, how to deal with the extracted information and how to exploit information from social media data which provides an added value for SC stakeholders. For targeting this, the foundation for deriving and elaborating the requirements and functionality of the methodology to support fashion and apparel SC decisions with social media data is led in Chapter 3. Firstly, this is done by sketching a schematically and exemplified fashion and apparel SC, stakeholders, and the main processes. Secondly, by elaborating on the characteristics of social media and social media data and the challenges

involved in dealing with social media data, which also addresses the harnessing of textual data.

The main concern while designing the process model was to ensure that SC stakeholders will have an added-value from its use. For this purpose, it was necessary to map the requirements from SC stakeholders. To this end, we focussed on the manufacturer and retailer and mapped requirements for the buyer and designer (retailer) and product developer and sales representative (manufacturer). For representing their perspective and needs into the process model, user stories are formulated. These are built based on the role descriptions (see Section 3.1.2) and besides based on the domain knowledge of the author and discussions with experts working in the field of fashion and apparel SC.

Similarly to Schieber and Hilbert (2014a), we develop structural requirements for considering organising and considering the process model design aspects. An addition to these requirements, characteristics of process design and TM frameworks lay the foundation for drafting the structural requirements. Based on characteristics of fashion and apparel SCs, SC stakeholder, social media, TM and functional requirements are derived. In addition to functional requirements, structural requirements are derived based on characteristics of process model design. Based on the described characteristics, features/categories are derived for each field. From these categories, the specific requirements are derived. In addition, it is also necessary to look at prior research on the handling of social media data in general as well as in the SC context. This is done in Chapter 3.3.

Similarly to TM (Schieber and Hilbert, 2014a), Social Media Analytics (SMA) is considered to be an analytical process. Therefore, the selected methodology is a process model. The development of the methodology requires the design of a social media process model, which is the artifact according to DSR (Hevner et al., 2004). As the main artifact is designed as a process model, the following section gives an overview of process models.

PROCESS MODELS

A process model illustrates the sequence of required activities that have to be conducted in order to carry out (IT-) projects, applications or processes (Abts and Muelder, 2013; Fischer et al., 1998; Grob et al., 2004; Stahlknecht and Hasenkamp, 2005). Process models belong to the group of reference models. The term reference model is used in a variety of ways. It describes any exemplified and abstracting process model, recommendation, process, or guidelines defined for a defined/limited problem area, and thus are applicable in as many individual cases as possible (Stahlknecht and Hasenkamp, 2005).

Phases in a process model are ordered sequentially. In each phase, a number of activities are conducted. These activities can also be conducted in parallel within a phase. By ordering the different activities, a complex process is typically defined in manageable units (Grob et al., 2004). Defining activities and sequences reduce the complexity of processes (Breitner, 2012). Feedback loops to prior phases are allowed if defined objectives are not achievable

(Krcmar, 2015). In addition to the sequentially ordered process models, iterative ordered process models also exist, which advocate taking iterative steps to deal with the complexity of the processes (Krcmar, 2015). Process models specify which procedures, methods and techniques are to be used (Stahlknecht and Hasenkamp, 2005). A typical process model consists of different components (Abts and Muelder, 2013; Fischer et al., 1998; Stahlknecht and Hasenkamp, 2005).

- **Activities:** Workflows producing certain results by using procedures, methods and techniques
- **Procedures, methods and techniques:** Obtain precise instructions in order to produce the desired results
- **Phases:** In order to structure the processes in terms of content or time, coherent activities may be aggregated into phases. Phases serve the grouping of activities in order to conduct them in planned and controlled units

To realize the development, it is necessary to look into existing research to determine the suitable components of such a process model and build upon prior research in the field. Approaching literature also enables the delimitation of the process model, i.e. to clearly define which elements are not be included in the social media process model. The most important method for developing the process model is thus a literature review.

For the development of the process model, two approaches, deductive and inductive, are used. While a deductive approach is pursued by integrating and building up requirements from the literature review, an inductive approach is established by considering specific application-based frameworks (Fettke, 2014). In this way, the overall methodology for developing the social media process model constitutes a hybrid approach.

Given the multidisciplinary nature of this research, it is required to review literature in different disciplines. Based on the objective of the thesis, prior research in the following fields is considered: Fashion and apparel SCs, SMA and TM.

Process models have different features, one of which is the level of formalisation (Filß et al., 2005). Process models can have a broad range when it comes to formalisation/representation (Krcmar, 2015). Verlage (1998) advocates the formalisation of process models and highlights the advantages in terms of making the process model accessible to tools or increasing the quality of the process model itself. Formalisation enables the representation of complex processes with all relevant elements such as data streams, events, branches or roles (Allweyer, 2015). Notation languages are typically used in formalisation. One modelling language for visualising business processes is the Business Process Model and Notation (BPMN). In 2006, BPMN was officially picked up as an Object Management Group -Standard. Within the last years, BPMN has developed into a standard language in the field of business processes, and it is used in different process models. For instance, Schieber

and Hilbert (2014a) use BPMN for the creation of a generic TM process model. Similarly, Schieber and Hilbert (2014b) introduce a process model for content extraction from weblogs documented in BPMN 2.0. The Text Mining Layer of the proposed process model is mainly based on the generic TM model presented by Schieber and Hilbert (2014a). For fulfilling the formalisation requirements, the process model is designed in BPMN 2.0. Figure 2.3 summarises the elements which are used for developing the process model.

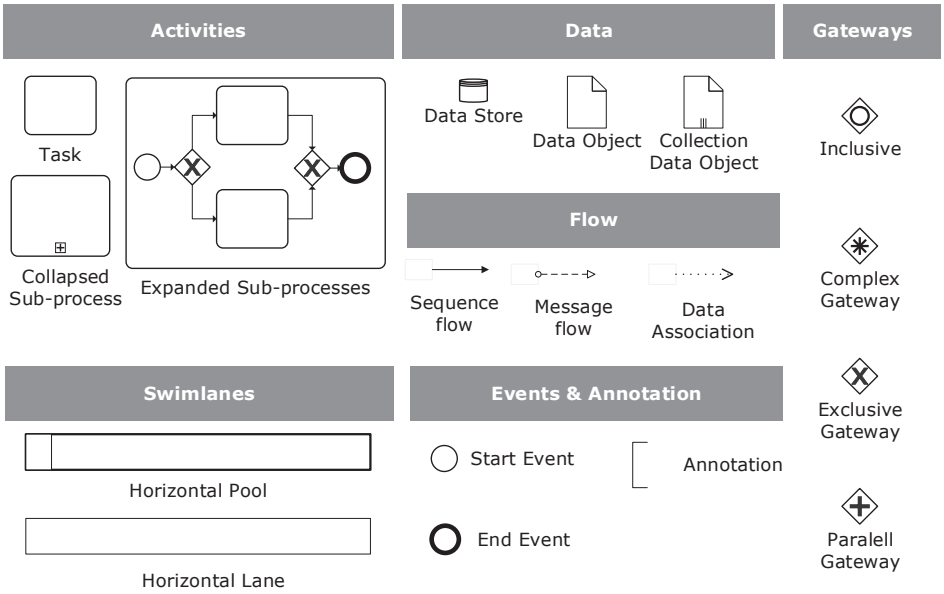


Figure 2.3.: Utilized BPMN 2.0 elements based on (Allweyer, 2015)

The process model is structured in one pool with four different swimlanes. The lanes typically represent responsibilities for activities in a process and can stand for an organization, role or system. In this thesis, the lanes are used to differentiate the layers targeted (Allweyer, 2015). The activities are modelled on different levels with sub-processes that are visualized, collapsed and expanded. In order to document the information flow along with the processes, data objects, collection data objects as well as data stores are utilized.

Text Mining

Tapping into these valuable data sources is mainly possible through advancements in the field of TM. "Text Mining and text analytics are broad umbrella terms describing a range of technologies for analysing and processing semi-structured and unstructured text data. The unifying theme behind each of these technologies is the need to "turn text into numbers" so powerful algorithms can be applied to large document databases. Converting text into a structured, numerical format and applying analytical algorithms require knowing how to both use and combine techniques for handling text, ranging from individual words to documents to entire document databases" (Miner et al. 2012, p. 30). The basic unit of automatic textual analysis is a document (Feldman and Sanger, 2007) which can adopt different forms

such as a book, a page of a book or a paragraph. In the context of social media data, it may be the so-called post, the message in different variations such as a tweet, blog post, a Facebook post or a review on a product. Therefore, a document may vary in terms of length and number of words (tokens). Since TM algorithms are not able to deal with whole documents, it is required to define document features that enable feature-based representations of documents for the algorithms. Although these features are usually characters, words, terms and concepts, the features are however not limited to these (Feldman and Sanger, 2007). In order to process these documents automatically, a range of methods and approaches are required.

Typical approaches in TM are Social Network Analysis (SNA), multilingual TM, spam classification, use of k-means clustering to group documents, anomaly detection, trend detection and analysis of streaming text data as described in Berry and Kogan (2010). Miner et al. (2012) define seven practice areas of Text Analytics (TA). Search and information retrieval (IR), document clustering, document classification, web mining, information extraction (IE), Natural Language Processing (NLP) and concept extraction. The seven practice areas overlap considerably since many practical TM tasks sit at the intersection of multiple practice areas. Moreover, a typical TM project tackles techniques from all areas. These practice areas originate from the following related fields: DM, AI and machine learning, statistics, computational linguistics, library and information science and databases (Miner et al., 2012). Table 4.7 summarizes the TM practise areas with their corresponding topics as well as commonly used algorithms. For running TM projects, it is necessary to follow specific activities in order to exploit textual data. Figure 2.4 depicts one general TM process suggested by Hippner and Rentzmann (2006). This framework involves task definition, document selection, document processing, TM methods, interpretation/evaluation and application as the main TM steps. While preprocessing activities as data preparation or data cleansing may be applied in DM approaches for improving the analysis results, these and further preprocessing tasks essentially deal with textual data. Hence, without the preprocessing activity, textual data cannot be further processed.

Figure 2.4.: A general TM process based on (Hippner and Rentzmann, 2006)

A range of further frameworks for exploiting textual data is presented and discussed in literature. Table 2.1 distinguishes the frameworks in case-specific and application neutral frameworks. While the case-specific frameworks are typically tailored to the corresponding application, the application-neutral frameworks remain at a high-level consideration of the TM process, in a similar way as it is demonstrated in Figure 2.4.

Table 2.1.: Selected TM Frameworks

Case specific TM Frameworks	**Application-neutral TM Frameworks**
Archak, Ghose and Ipeirotis, 2011 (Archak et al., 2011)	Abdous and He, 2011 (Abdous and He, 2011)
Atkinson-Abutridy, Mellish and Aitken, 2004 (Atkinson-Abutridy et al., 2004)	Chou,Sinha and Zhao, 2010 (Chou et al., 2010)
Chau and Yeh, 2004 (Chau and Yeh, 2004)	Choudhary et al. 2009 (Choudhary et al., 2009)
Dasgupta and Sengupta, 2016 (Dasgupta and Sengupta, 2016)	Fan et al. 2006 (Fan et al., 2006)
Chou, P. Sinha and Zhao, 2008 (Chou et al., 2008)	Hippner & Rentzmann 2006 (Hippner and Rentzmann, 2006)
Ur-Rahman and Harding, 2012 (Ur-Rahman and Harding, 2012)	Hotho et al. 2005 (Hotho et al., 2005)
Duan et al., 2011 (Duan et al., 2011)	Natarajan, 2005 (Natarajan, 2005)
Lee et al., 2011 (Lee et al., 2011)	Schieber & Hilbert, 2014 (Schieber and Hilbert, 2014a)
Li and Liu, 2012 (Li and Liu, 2012)	Kobayashi et al., 2017 (Kobayashi et al., 2018)

This is underlaid by the circumstance that the specific activities and methods used in a TM process are not specified within the majority of the frameworks. Schieber and Hilbert (2014a) argue that a consolidation of the individual models in an integrated, generic process model contributes to an improved structure of TM activities. Moreover, this approach would reduce the complexity of such projects by linking them to concrete processes. Accordingly, the authors propose a generic TM process model with a set of activities, providing concrete methods for these activities. The generic TM process model proposed by Schieber and Hilbert (2014a) contains six phases. These are task definition, document selection and analysis, document processing, application of TM methods, evaluation of results and derivation of actions and application. These phases incorporate a range of activities and procedures to be conducted. The authors consider feedback loops as functional requirements of TM process models, which ensure that TM results can be improved during the entire process. Another relevant work in the context of this research is Kobayashi et al. (2018). They examine TM processes from an organizational perspective and cluster the TM process in the three main process steps of preprocessing, TM operations and postprocessing. Taking the generic TM process models as a structural base, the following section gives a detailed overview of generic TM processes' different activities and procedures.

2.3. Demonstration

According to Peffers et al. (2007), a range of methods is applicable for demonstrating the use of the developed artifact. These may involve simulations, experimentations, proof or case studies. For this research, a case study method is considered as the method for demonstrating the utility of the artifact, since a case study approach is especially suited for explanatory examinations in which the phenomenon has not been sufficiently explored (Meredith, 1998). The demonstration involves the framing of the case study as a first step and the application of the developed process model as a second step.

For framing the case study, three main steps are pursued: determining the overall frame, formulating specific questions based on the defined frame and lastly, elaborating case study features. As an overall frame, it is defined to follow a posteriori analysis, which is performed after the point of sale and thus contrasts an a priori approach. Although it is a crucial step as it is fundamental to explore a potential connection between real-world sales data and social media data. This approach will not allow verifying whether proper forecasting is possible. Furthermore, any insight into the connection between textual social media data and sales data only represents added value for a SC decision maker if it is readily available before an upcoming decision has to be taken by the decision maker, i.e., trend information needs to be known before it hits the market. Thus, the utility of the artifact can be demonstrated by exploring this potential connection. After that, specific questions are formulated to target this overall objective and frame in the second step. Finally, for the framing of the case, it is required to define features to delimitate the real-world complexity, which is done in the third step. The detailed elaboration of the three steps is set out in Section 5.1. As the second main element of the demonstration, the process model is applied to the framed case. The phases and activities proposed in the process model are executed. The process model proposes methods for conducting the different activities, and suitable methods are used accordingly where the examining relationships between the sales and social media data, the spearman's rho is used. For processing the social media data, the KNIME Analytics Platform [2] (KNIME) is used. KNIME is an open-source software for data science. KNIME was selected as a suitable working environment for several reasons. First, KNIME generally provides strong text processing extensions. For the German language, in particular, a range of nodes are already implemented. Second, KNIME is freely accessible with all functionalities. Third and last, KNIME has a large and active community, which allows access to a range of existing functionalities.

2.4. Evaluation

A major element of the evaluation is to "observe and measure how well the artifact supports a solution to the problem" (Peffers et al. 2007, p. 13). It addresses the comparison of the objectives of a solution to the observed results obtained by applying the artifact to the demonstration. The evaluation may involve different forms (e.g. quantitative performance measures, satisfaction surveys, client feedback or simulations). Though any appropriate empirical evidence is applicable as evaluation in the context of DSRM (Peffers et al., 2007). In this light, Venable et al. (2012) propose a framework for the evaluation of DSR based on Hevner et al. (2004) and Peffers et al. (2007). The framework includes four main steps, which are depicted in Figure 2.5.

[2] https://www.knime.com/

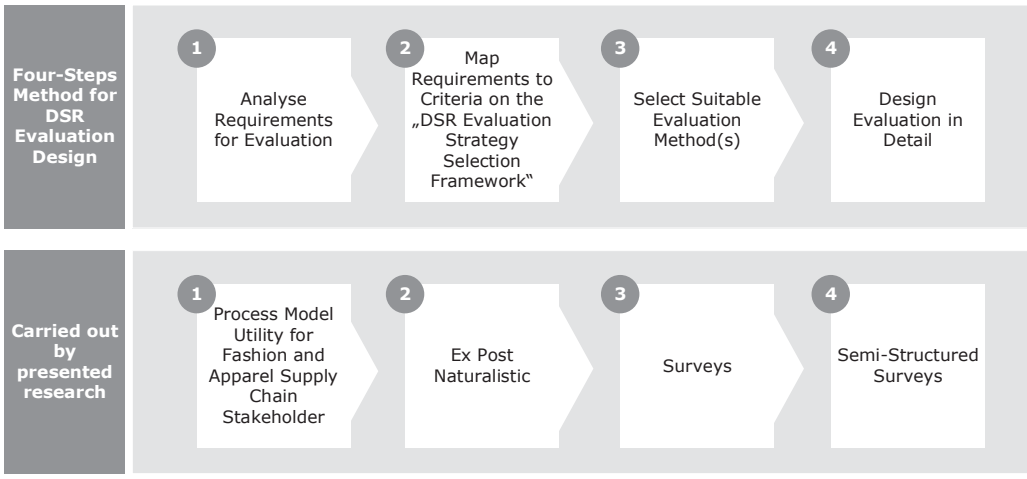


Figure 2.5.: Four-steps method for DSR Evaluation Research Design by Venable et al. (2012) carried out for the present research

The objective of the first step is to analyse the requirements for the evaluation. The definition of the requirements is the base for the entire evaluation. The first step includes the determination of the evaluands, the nature of the artifact, the properties to be evaluated or the purpose of the evaluation. These requirements will be mapped to the criteria of the DSR Evaluation Strategy Selection Framework. Venable et al. (2012) proposes an ex-ante vs ex-post and artificial vs naturalistic evaluation scheme. Within this classification, a range of criteria are formulated, which needs to be analysed against the requirements defined in the first step. The mapping serves the definition of the evaluation strategy on a high level. It does not imply the selection of specific evaluation methods. After defining the evaluation strategy on this high level, the next step is to select specific evaluation methods (Venable et al., 2012). Based on the selected method(s), the final step of this method is setting the evaluation's specific design, for instance, a questionnaire. The bottom part of Figure 2.5 depicts the main outputs when carrying out the four-step model on the present research. These four steps are described in the following.

- Analyse requirements for evaluation: The main requirement for the evaluation is to demonstrate the utility of the process model for fashion and apparel SCs stakeholders. This implies that the real world user has to be included in the evaluation. This is the main priority.

- Map requirements to criteria on the DSR evaluation strategy selection framework: Accordingly, following the DSR Evaluation Strategy Selection Framework, the evaluation is aligned with a naturalistic evaluation. Secondly, a summative evaluation is sought, which represents an ex-post evaluation strategy. This means that a naturalistic and ex-post evaluation is selected as appropriate for the evaluation.

- Select Suitable Evaluation Methods: The combined evaluation strategy offers different specific evaluation methods such as focus groups, case studies, participant observation

or surveys (qualitative or quantitative) (Venable et al., 2012). For this research, the use of surveys is selected. In the last step, a self-completion questionnaire/survey is designed.

- Design evaluation in detail: Survey Design. The designed survey consists of three parts: demographic information, Description of research, Description of scenario. In the first section, demographic data is queried. The second part gives a short description of the research. The third part provides the selected scenario. A figure of the SC with the main results will be presented. And potential impacts on the different processes of the SC are queried, together with the time of availability. The question will be designed in the form of an open-ended question. While closed-ended questions are restricted to the predefined responses, the main advantages of open or open-ended questions are that the respondents have the possibility to formulate replies in their own words and are not restricted to the defined replies (Bell et al., 2018). As the stakeholders' perspective is focussed in this research, open-ended has the advantage of giving the respondents the possibility to formulate their unfiltered opinion. Moreover, open questions allow unusual replies that the survey designer did not consider while designing the questions and responses. And finally, open-ended questions are recommended when new areas are explored (Bell et al., 2018). The last point is a highly valuable advantage for the objective of this evaluation. For testing purposes, the survey was piloted by providing it to two researchers and making slight adaptations after receiving their feedback. The survey is distributed via e-mail and is inserted in the appendix (see Figure A.15).

2.5. Communication

Elements of this research were published. Besides, to fulfil the last step of the DSR-approach, the finalized research in the form of this document will be published.

3 Research Foundation

The main objective of this chapter is to elaborate on the characteristics of fashion and apparel SCs as well as social media in order to lay the foundations for RQ1, RQ2 and RQ 3. For this, the chapter provides first background information on the involved fields starting with fashion and apparel SCs in Section 3.1 and 3.2. Secondly, it gives an overview of the current handling of social media in fashion and apparel SC processes. In the last section, a summary and conclusion of the research gap are presented.

3.1. Fashion and Apparel Supply Chains

Fashion and apparel markets and SCs are highly dynamic, complex and fragmented. Section 3.1.1 gives a short overview of these peculiarities. SC stakeholders and processes are sketched out in Section 3.1.2. As a key decision process within fashion and apparel SC, fashion sales forecasting processes and current approaches are presented in Section 3.1.3. The information needs matching with the individual SC processes are illustrated in Section 3.1.4. Section 3.1 ends with a summary and an interim conclusion on the research object.

3.1.1. Basic Concepts and Characteristics

[3]In the last decades, the fashion and apparel industry has undergone severe challenges and constant changes. The main drivers of this development are the dynamic changes on end-consumers preferences (Webb, 2007) and purchasing behaviour together with technological advancements (Seram et al., 2016). These changes are characterised by a low affinity for fashionable products, the desire to quick changes of fashion items (Christopher et al., 2004), and the desire of individualised items (Piller and Müller, 2004; Webb, 2007). In order to be responsive to these changes, the industry was forced to implement suitable SC strategies (Seram et al., 2016) addressing mainly the adaptation of internal processes and operations. Until the 1980s, mostly mass production of basic apparel articles was common in

[3]Parts already published in Beheshti-Kashi (2017)

the industry. First changes arose in the mid of 1980s with the demand for more fashionable products. As a response to this demand, producer-driven value chains were transformed into buyer-driven chains (Gereffi and Memedovic, 2003). The desire for quickly changeable fashion items resulted in the launch of 3-5 mid-season collections per year (Barnes and Lea-Greenwood, 2006). This was possible through the implementation of Quick Response (Hines and Bruce, 2007) which addresses the responsiveness of the apparel SC. The implementation of Quick Response came along with the integration of new technology such as Electronic Point of Sale, barcodes and Electronic Data Interchange (EDI) (Cheng and Choi, 2010).

Fast fashion retailers are highly oriented towards customers' demand, and they target responding to market changes in a timely manner (Bruce and Daly, 2006). The main features of fast fashion retailing are small and medium volumes, reduced lead times, postponement, a backward vertical of the SC (Birtwistle et al., 2003) as well as a high variety of customer preferences (Zülch et al., 2011). In particular, the reduced lead time contrasts with traditional retailing (Ghemawat et al., 2003). As a response to the increased need for individualised and customised products (Zülch et al., 2011) apparel companies put the implementation of Mass Customisation strategies forward. The huge challenges within the last decade heavily grounded on the strong growth of e-commerce in its traditional form and new innovative concepts such as mobile commerce or curated shopping approaches. Launched within the last years, these concepts focus on an increased engagement and participation of customers. Taking the technological advancements as a main driver for the emerged changes within apparel and SC processes, big data analytics is put forward.

Christopher et al. (2004) refer to fashion as a "broad term that typically encompasses any product or market where there is an element of style that is likely to be short-lived" (Christopher et al. 2004, p.367). The nature of fashion markets is summarized through short life-cycles, high-volatility, low-predictability and high-impulse purchasing. The short-life cycles refer to the short time in which the product is saleable. This time frame is typically very short and seasonal, measurable even in weeks. The demand for these products is often unstable and influenced by external factors making the accurate forecasting of products difficult. As the purchasing of items is highly impulse-driven, the availability of the products is crucial (Christopher et al., 2004). Moreover, fashion markets are characterized by a large product variety, and the dynamic changes with respect to style or colour (Hui and Choi, 2016) as well as complex SC structure (Fisher, 1997).

Apparel types are broadly differentiated into two categories: outer and inner clothing, each of which comes with a high variety of styles and colours (Nayak and Padhye, 2015). Furthermore, apparel products are differentiated into different levels. Different ways of classifying the products are introduced by literature. Wojaczek (1996) uses high fashion, fashion and basics articles. Hammond (1999) suggests a fashion triangle in which the relationship between fashion grade and demand uncertainty is illustrated in Figure 3.1.

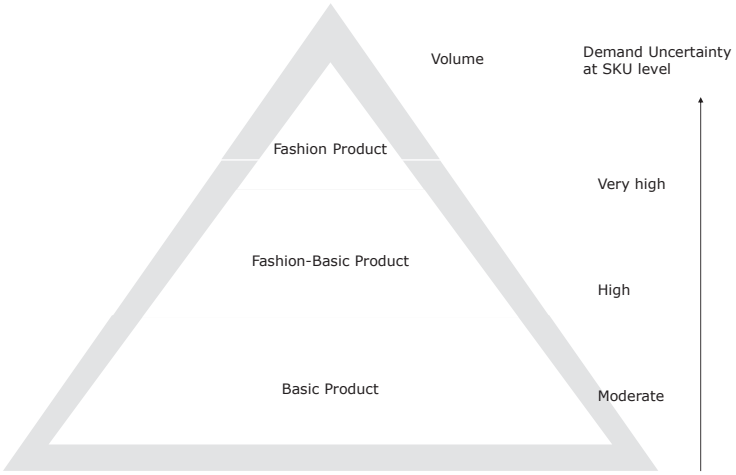


Figure 3.1.: Fashion Triangle based on Hammond (Hammond, 1999)

He distinguished between basic products, fashion-basic products and fashion products and argued that the demand uncertainty increases with the fashion grade. Lowson (2003) makes a difference between basic products, seasonal (or fashion basic) and short-season (fast fashion) products. In all three classifications, demand uncertainty plays a crucial role with regard to the product type. This thesis mainly considers fashionbasic and fashion products as their trading is more challenging due to their high demand uncertainty. It is relevant to mention such differentiation since the type of product has a large impact on the SC strategies (Fisher, 1997). Lam and Postle (2006) illustrate the relevance of differentiating between product types for selecting an appropriate SC strategy for textile and apparel companies in Hong Kong. They differentiate between functional and innovative products and suggest a responsive SC for innovative products and an efficient SC for functional products. Similarly, in the context of fashionable products, fast fashion strategies are often applied (Bhardwaj and Fairhurst, 2010). Despite the different configurations of the SC, in a schematic SC the same stakeholders are involved and similar processes have to be conducted to produce apparels. These are introduced in the following.

3.1.2. Stakeholders and processes

In fashion and apparel SCs, two extreme operation modes exist, namely “ready-to-order” and “make –to-order”. While in the make-to-order mode, the process is initiated by an individual user request, the ready-to-order mode starts with the design of an apparel item based on the prediction of customer preferences (Hui and Choi, 2016). This work considers the ready-to-order operating mode since it is mostly applied by fashion companies (Hui and Choi, 2016).

The objective of a typical apparel SC is to satisfy the market needs with the lowest possible costs, the fastest speed and with a maximized profit all at the same time (Hui and Choi, 2016). Figure 3.2 illustrates the fashion and apparel SC as an input-output system in which the customer’s preferences and demographical information are considered input factors and

"the" fashion product is the output. The major steps are product design & development, demand forecasting, ordering & replenishment, price negotiation, quality control as well as information sharing. The customers' preferences and demographical information may come in diverse forms and from a range of different sources. Traditionally, customers' preferences are captured through typical market research instruments such as surveys, interviews or focus groups, which are high resources and time-intensive approaches (Hui and Choi, 2016). Nowadays, these preferences are often publically articulated by customers themselves through their social media engagement.

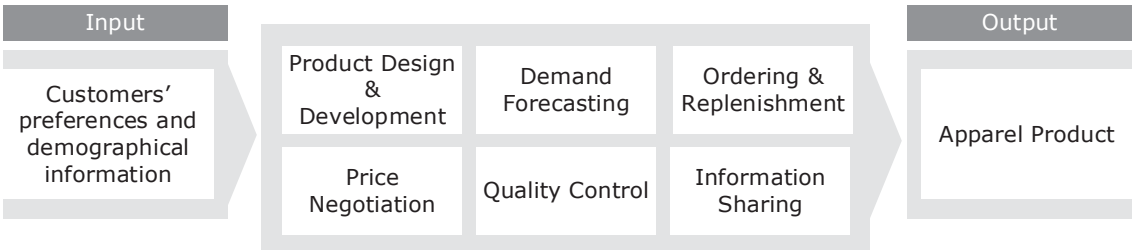


Figure 3.2.: Fashion and apparel SC as an Input-Output-System (Hui and Choi, 2016)

Product design & development aim at determining whether the SC can provide the right product to the market. Demand forecasting is crucial for driving the sourcing and inventory planning of the SC. Ordering & replenishment are related to the inventory quantity decision and ordering time. It also affects the inventory costs and the flexibility of the SC. Price negotiation is related to the transactions taking place between the SC partners. Quality control affects price and product quality. Information sharing enables the enhancement of the SC performance and lowers the risk for all SC members from challenges such as the bullwhip effect. The major processes within the apparel and SC system can be assigned to different stakeholders. Figure 3.3 illustrates an exemplified fashion and apparel SC system with its generalized processes. Note that the time spanned on the figure displays 18 months prior to the start of the selling season. Apparel companies follow different SC strategies and have implemented different SC configurations that imply a difference in lead and time-to-market times. Nowadays, in particular, so-called fast fashion retailers such as Zara [4] have reduced lead times of fashion and apparel products (Bhardwaj and Fairhurst, 2010). With their backward vertical integration and centralized distribution facilities, Zara has managed to react quickly to market changes and consumer demands (Hines, 2007). Traditional long lead times are thus compressed, and in some cases, slightly modified products are delivered to their stores only within two weeks (Ghemawat et al., 2003). Though the different strategies and configurations of SCs are not the subject of this research, accordingly, we will not expand further on this and will not dive into comparing these different configurations along with their advantages or disadvantages. At this point, it is crucial to understand that the illus-

[4]Zara is owned by the Spanish textile company Inditex and is often considered to be the innovator of fast fashion (Hines and Bruce, 2007)

tration of the SC in Figure 3.3 and the sketched lead times are a schematical and exemplified configuration of the SC. This setting is not based on a particular company, and the lead times may vary depending on the actual configuration and internal processes of the corresponding SCs.

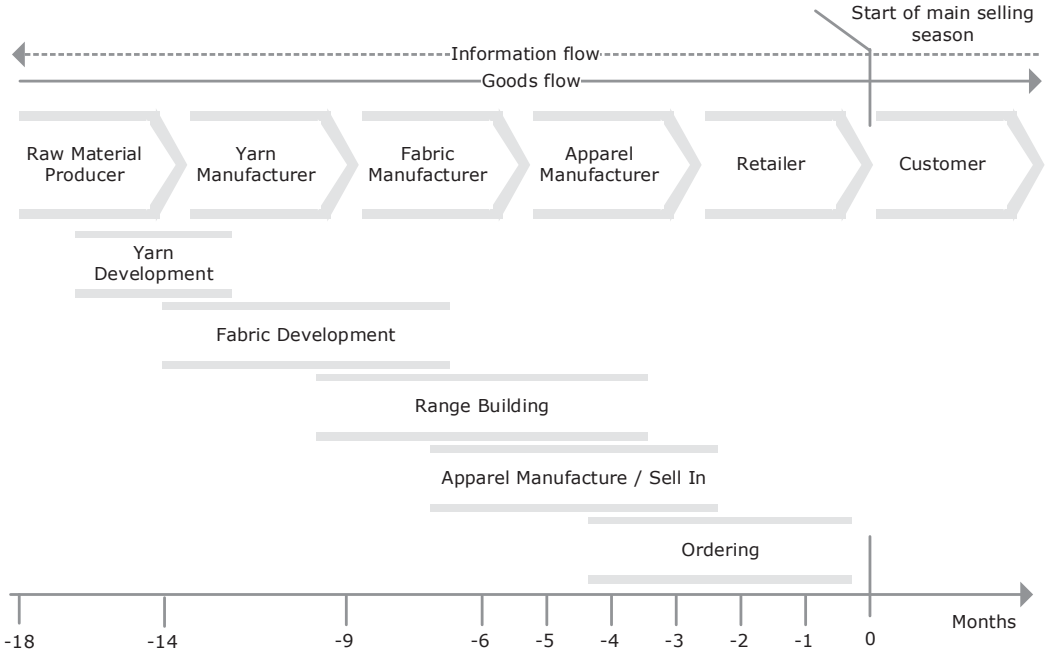


Figure 3.3.: Schematic illustration of a Fashion and Apparel SC (based on Birtwistle et al. (2003); Hines (2007); Hines and Bruce (2007); Şen (2008)

The figure illustrates the main stakeholders of a fashion and apparel SC as described inside the arrows. The processes are presented below to the stakeholders. The main stakeholders of a fashion and apparel SC are raw material producer, yarn manufacturer, fabric manufacturer, garment manufacturer and retailer. The major fashion and apparel SC operations are fibre and yarn production, fabric production, apparel manufacture, range building and ordering (Şen, 2008). Although each stakeholder has its key process, the processes also overlap with other stakeholders along with the SC. The goods and information flow are sketched accordingly. The stakeholders and processes are matched on a time axis in a backward approach to illustrate the dilemma of fashion and apparel SCs. All products need to be available on the point-of-sales and at the start of a selling season. This is indicated by 0 months on the time axis. However, for finishing the goods and making them available at time 0, more than one year prior to this the first processes need to be conducted by different stakeholders. This circumstance contrasts with the fact that information on customer preferences is hardly available when the different stakeholders would require this information for running their processes.

The raw material producer handles fibre production as the first step of each fashion and apparel SC. The yarn development involves the conversion of natural and man-made fibre into yarn (Giri and Rai, 2013). Fibres are categorized into natural (for instance cotton, linen,

jute, bamboo or wool, fur, silk) and man-made fibres, also known as synthetic fibres (for instance polyester, nylon, rayon). The fibres produced need to be transformed into fabric. The process of yarn development usually starts up to 18 months prior to the start of the season. The fabric development involves weaving, knitting and non-woven processes such as looping, fixing, knotting, plaiting or twisting yarn for manufacturing the fabric (Şen, 2008). Textile mills play an essential role within the fabric production stage (Hines and Bruce, 2007). The fabric development starts nearly 12 months before the selling season and lasts several months. The range building process is one of the most crucial processes of the fashion and apparel SC. It includes three major tasks which are range building planning concept, planning and the actual building. These include a set of sub-processes conducted mainly by the manufacturer and retailer. Table 3.1 provides an overview of the main tasks and sub-processes of the range building process.

Table 3.1.: Range building steps

Main task	**Sub-processes**	**Involved stake-holders**
Range building concept	- Collecting ideas from sources processing - Sales forecasting - Adapting to retailers own requirements - Creating documentation for manufacturer	Retailer
Range building planning	- Discuss queries with buyers and designers - Place fabrics and trims according to theme (to fabric manufacturer)	Manufacturer
Range building	- Prepare samples - Calculate costs - Present to retailer for sample selection (manufacturer) - Sample selection from the presented range (retailer)	Manufacturer and retailer

The apparel production itself consists of a range of processes (Giri and Rai, 2013) (Gereffi and Memedovic, 2003). Typically, the apparel production involves the processes of cut order planning, marker planning, spreading, cutting and sewing. These main processes require many different steps. The manufacturing process is made more complex by the wide range of apparel items that have to be produced, i.e. a range of different designs, styles and colours (Nayak and Padhye, 2015). At the managerial level, the garment manufacturer is further involved in constant fabric quality inspections (Ngai et al., 2014) and shipment related activities, including ticketing, labelling and packing. The actual apparel production can start up to 6 months before the season and often lasts until 2-3 months prior to the selling season. A range of processes is conducted within the retail cycle. The retailer has interfaces to both manufacturer and customer, which may be considered as the link between supply and

demand (Hofmann and Rutschmann, 2018). While all fashion and apparel SC stakeholders are challenged due to the described peculiarities, it is particularly the retailer who operates under high volatility, uncertainty, complexity and ambiguity (Hofmann and Rutschmann, 2018). The retailer cycle is the process closest to the selling season and the customer. The garment manufacturer and retailer have a crucial role within the SC and traditionally have shared an interface between the account manager on the manufacturer side and the buyer on the retailer side (Fiddis, 1997), changing to multi-functional collaborating teams meeting the complex requirements of the markets (Fernie, 2014). A crucial process is sales forecasting, which is a fundamental process. Poor forecasts will result in poor production planning and customer service and lost orders (Wong et al., 2013).

3.1.3. Fashion Sales Forecasting

[5] Wong (2013) defines the following decision points for fashion and apparel SCs: selection of the plant's location, production scheduling, assembly line balancing control, cutting room related processes including cut order planning, marker planning, spreading and cutting scheduling, and fashion sales forecasting. The author claims that all these key decision points are dependent on the experience and subjective assessment of management and decision makers. When considering these decision points, sales forecasting is declared as the foundation for different operation planning phases (Wong, 2013). Sales forecasting plays a key role in SCs for several reasons. Due to high dynamic market demands, sales forecasting has a huge impact on fashion retailers. Inaccurate forecasts and unreliable forecasts result in insufficient production planning, lost orders, inadequate customer services, and poorly utilized resources resulting from the retroactive response. Recent research has demonstrated an increase in SC performance generated by effective sales forecasting (Wong, 2013) since accurate forecasts support order management, product placement, scheduling and production. This leads to the question of how forecasts are incorporated in fashion and apparel SCs.

[5] Some content of this section is published in Beheshti-Kashi et al. (2016, 2015a)

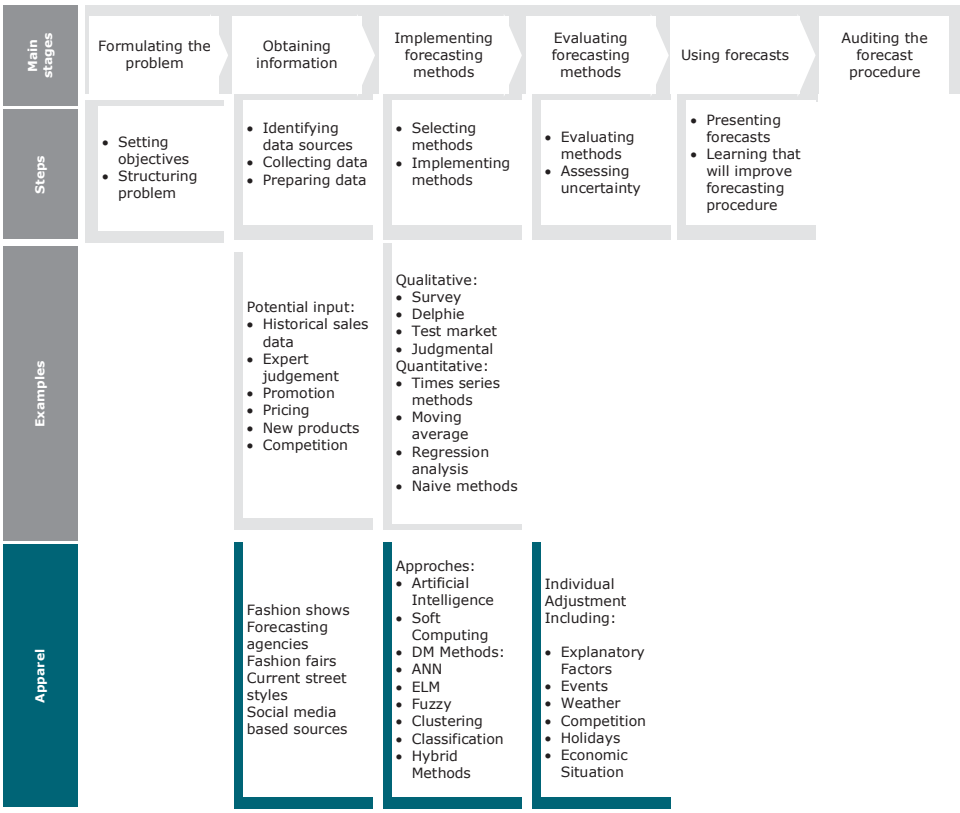


Figure 3.4.: A Simplified Sales Forecasting Process Adapted to Fashion Products Requirements based on (Armstrong, 2001; Ivanov et al., 2017; Thomassey, 2010, 2014)

As illustrated in Figure 3.4, a schematic sales forecasting process includes six stages: formulating the problem, obtaining information, implementing forecasting methods, evaluating forecasting methods, using forecasts and auditing forecasting procedure. These stages typically involve a range of steps which are displayed in the second row of the figure. The problem formulation stage, for instance, starts with setting the objective of the forecasting procedure. This involves, for instance, describing the decisions which might be affected by the forecasts or obtaining the decision makers' agreement on methods. The structuring problem step includes, amongst others, the identification of possible outcomes. The stage obtaining information summarises the identification of data sources, its collection and preparation. Examples include using diverse data sources, obtaining the most recent data and cleaning data activities. Typical data sources are historical sales data or expert judgements. Based on the selected data, the stage of implementing forecasting methods is put forward. At this point, it is defined which methods, e.g. quantitative or qualitative methods to follow and how to integrate and combine them. Typical quantitative methods include time-series or moving average, qualitative methods constitute surveys or Delphi methods (Armstrong, 2001; Ivanov et al., 2017). After the actual implementation of the forecast methods, the results are evaluated, and the uncertainties are assessed. The results are presented, and lessons learned documented for improving the forecasting procedure. The last stage of a typical forecasting process should include auditing the entire procedure. This could be done by pro-

viding forecast checklists prior to conducting the procedure in order to ensure the reliability of the results. The upper part of Figure 3.4 enhanced the process by including apparel specific data sources, more advanced methods, and individual adjustment factors typically used in fashion sales forecasting.

Given the peculiarities of fashion sales forecasting, an increased number of research works has focused on other methods than simple statistical approaches. These include the use of artificial neural networks (ANNs), the so-called extreme learning machine (ELM) (Sun et al., 2008), evolutionary neural network (ENN) (Au et al., 2008), machine learning approaches such as random forecast (Singh et al., 2019) or hybrid approaches (Wong and Guo, 2010). The use of ANNs for sales forecasting and reports improved performances compared to conventional approaches. Similarly, in the case of noisy data, ENN shows promising results (Au et al., 2008). In all these works, historical sales data is applied. Apart from the application of advanced methods, additional input data are required as well in sales forecasting.

Moreover, the evaluation and adaptation of base forecasts require further attention. For adjusting the base forecasts, individual adjustments of decision-maker are used. Apart from the individual expertise and experiences, the adjustments are based on explanatory factors such as holidays, competitions or future events.

For dealing with the lack of historical data, Thomassey and Happiette (2007) and Thomassey (2010) propose soft computing methods, e.g. fuzzy inference systems and neural networks. Apart from the lack of historical sales data, the availability of descriptive data in quantitative (average sales period) or nominal (style) and a huge number of historical items belonging to the previous collection are preconditions for using their proposed model. Meeting these conditions, the proposed approach shows promising results (Thomassey, 2010). However, the author reports that these methods can hardly be adopted by apparel companies (Thomassey, 2014). Nevertheless, many commercial software often applies these techniques for their predictions (Jain, 2007). However, these forecasts often only serve as a baseline for decision makers for individual adjustments.

A different approach is followed by Mostard et al. (2011). The authors use pre-order demand information, apply three approaches to enhance demand forecasting and report promising forecasting results. Similarly, Teucke et al. (2016) focus on pre-orders of seasonal clothing items and propose a hybrid two-step prediction model which is able to estimate the additional post-orders before the actual production is started. Applying this hybrid approach, more accurate results have been achieved. In contrast to most sales forecasting methods, both approaches Mostard et al. (2011) and Teucke et al. (2016) base their calculations on pre-order demand information instead of sales data from previous collection items. This approach seems promising, in particular in the light that a large part of the season's collection consists of new products or modified products. Kahn (2014) sees a difference in forecasting new products compared to existing products. He considers quality assumptions, judgements and processes as essential factors and proposes that new products will be more successful

after conducting these steps. Different authors such as Choi et al. (2014) motivate their research from a situation in which a forecasting model faces both data and time constraints. They introduce the Fast Fashion Forecasting algorithm (3F) and apply it to three years of real sales data as well as on an artificial database. The algorithm shows reasonable forecast results, even with the lack of efficient time and data.

An apparel product is composed of different product features (see Figure 3.5). However, colour is a main characteristic factor (King, 2012). Both the complex and time-consuming nature of SCs plus the interdependencies established through global sourcing in fashion and apparel SCs make colour forecasting an economic necessity (Scully and Cobb, 2012). Choi et al. (2012) focus on colour forecasting for fashionable products with limited data. For this purpose, they examine several different forecasting methods and compare them with regard to how they deal with the constraint of little data. The authors report that hybrid models produce the most accurate forecasts compared to the other models since they can produce reasonable forecasts, even with very little data. Gu and Liu (2010) propose a computer-assisted colour database for colour forecasting. The main parts of their model are colour information selecting and arranging, colour information transfer and processing, and colour data optimization. They conclude that human intervention is necessary, in particular, for image recognition. The use of user-generated-content in the forecasting of new products was not considered by the authors presented before.

3.1.4. Information Needs

The main factor in the development of apparel items is trends. These trends can come from the apparel industry or related industries such as textiles or shoes, from other industries such as entertainment or sports industries, or even have their origin in a wider context such as in cultural, social or technological developments (Le Pechoux et al., 2007). A fashion trend is defined as follows: “The term fashion trend refers to aspects of the appearances and construction of fashion products that relate to a particular season. Such trends are manifest in the appearance of fashion products, which are designed and manufactured prior to being delivered in a season” (Jackson 2007, p.170). Fashion trends may be further distinguished in long and short- term trends. While short-term trends refer to a certain season, long-term usually “underpin future designs” (Jackson 2007, p.170). Furthermore, the attributes colour, fabric, print, silhouette, styling detail, and trim “can be manipulated to reflect changing fashion” (Jackson 2007, p.170). Although these features may manifest a trend, fashion trend is often a combination of different features. It is suggested that style and brand be included as additional features which might manipulate a fashion trend (Jackson, 2007). All these features may constitute a fashion trend (see Figure 3.5).

Figure 3.5.: Trend Manipulating Features (based on Jackson (2007))

Fashion trends impact decisions along with the SC as they are often consulted by the different stakeholders within the processes. Le Pechoux et al. (2007) conclude that the creative process of developing new apparel items and collections are built as problem-solving systems. To solve the problem of apparel creation, information is required. The information needs vary according to their position and functions along the chain. In Figure 3.6 the Fashion and Apparel SC with its stakeholders and processes is sketched as in Figure 3.3 and the most important apparel features are mapped to the corresponding stakeholders. These features constitute the information needs of the stakeholders. Information about these features should ideally be available to the stakeholders for running their processes smoothly. The customers' preferences are often available only during the main selling period. However, this information should be available sooner, as displayed in figure 3.6.

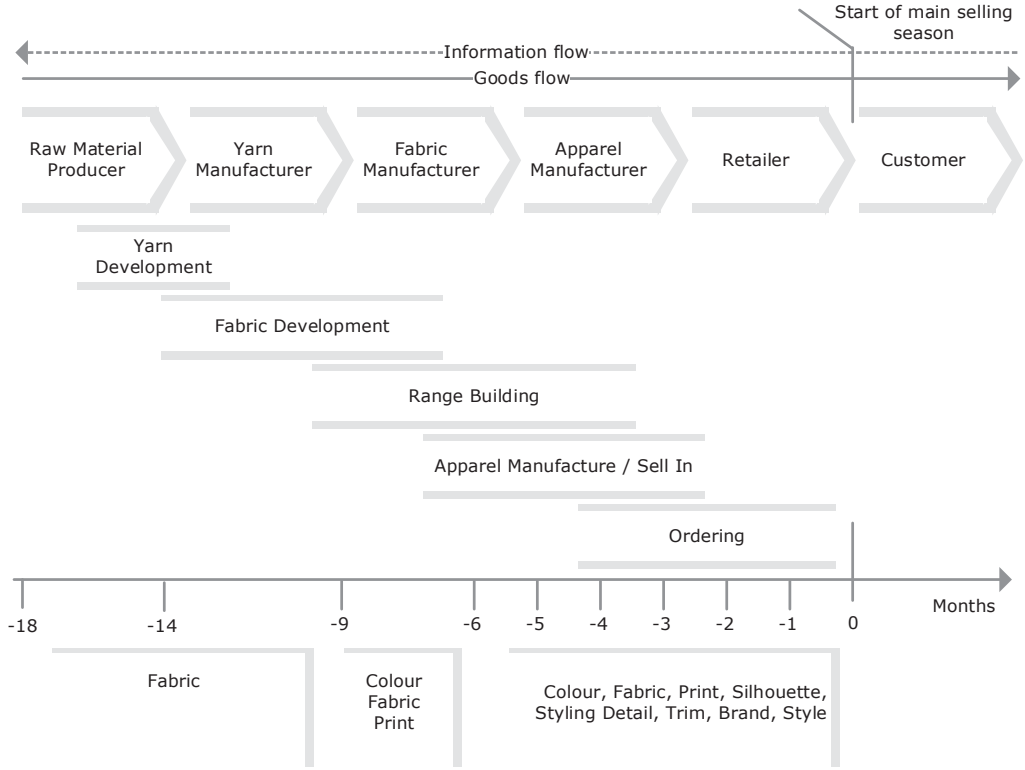


Figure 3.6.: Information Needs matched on SC

This results in significant challenges and uncertainties for stakeholders and decision-makers within the fashion and apparel SC. Furthermore, the time perspective is added to the figure to display the high time shift between the information needs for the respective processes and the beginning of the main retail-selling period. To obtain information to satisfy each stakeholder's need, a range of sources is available for consultation. Figure 3.7 summarizes the sources traditionally accessed by the different stakeholders for gaining information on the features colour, fabric, print, silhouette and style.

Colour	Fabric	Print	Silhouette and Style
Colour services Historical colour data Colour testing results New 'emerging colour stars' Purchased garments Trends Yarn samples Colour swatches Colour shows	Structural fabric characteristics (quality, weight, performance) Seasonal theme or timing of the line Aesthetics Marketplace trends Past sales history Fabric price Perceived customer benefits Textile mill availability	Original designs by the company Pre-developed fabrics from fabric companies Books and magazines Fabric services (fabric libraries/design services/forecasting services) Fabric samples Fabric mills Market trends Textile studios	Original designs Branded merchandise Market place (domestic and international) Current trends Past successful style blocks

Figure 3.7.: Sources Traditionally Accessed for Obtaining Information on Colour Ranges, Fabrics, Print, Silhouette and Styles Based on Wickett et al. (1999)

In addition to these sources, social media has gained increased interest from fashion and apparel stakeholders and customers. In order to utilize social media for supporting decisions alongside the fashion and apparel SC, social media and its different facets are described in the following section.

3.2. Social Media

A definition of social media is illustrated in Section 3.2.1. This is followed by an illustration of different social media channels. In the third step, social media content is defined. Subsequently, characteristics assigned to social media are outlined. This is followed by presenting approaches to access but more importantly to exploit social media data in Section 3.2.2. Based on this, the resulting challenges for exploiting social media content are derived in Section 3.2.3. The relevance of considering information quality assessment (IQA) of social media content is elaborated in Section 3.2.4. In each section, the corresponding points are put in the perspective of the research scope. [6]

3.2.1. Basic Concepts and Characteristics

Within the literature discourse on social media, the terms and concepts discussed such as social media, social networking sites, social network sites, social software, User Generated Content (UGC), Web 2.0 as well as participative web and participatory internet are often used interchangeably. Kaplan and Haenlein (2010) "define social media as a group of Internet-based applications that build on the ideological and technological foundations of Web 2.0, and allow the creation and exchange of User Generated Content" (Kaplan and Haenlein 2010, p. 61). The two key concepts of this definition are UGC and Web 2.0. Both concepts have emerged in different contexts. The Organisation for Economic Co-operation and Development (OECD) suggested three characteristics of UGC in order to identify a possible spectrum. Firstly, the content should be published on a publicly available website. Secondly, the content is generated by putting some creative effort to construct a new one. Lastly, the content should be created outside of professional routines and practise. The report stated that these criteria might change over time (OECD, 2007). With the increased market penetration of social media, users have been engaged in commercial activities. In this regard, the third characteristic may be blurred and shall be reviewed. With the rise of Web 2.0 and emerging technologies, ordinary users have been given a new role. Social media empowered the formerly passive and purely consuming user to an active and producing entity. For this role, literature introduced the term produser (Bruns, 2006). The term Web 2.0 was introduced by O'Reilly (2004). Social media include different tools such as social networks, blogs,

[6] Some content of this chapter is published in Beheshti-Kashi (2020); Beheshti-Kashi et al. (2016); Beheshti-Kashi and Kinra (2020); Beheshti-Kashi et al. (2018)

microblogging services, wikis, video and photo sharing platforms (Boyd and Ellison, 2007; Kaplan and Haenlein, 2010). One approach of classifying social media channels is displayed in Figure 3.8.

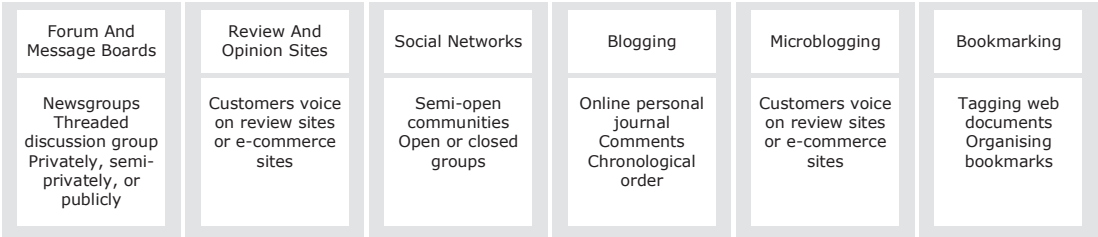


Figure 3.8.: Social media categories based on Sterne (2010)

The classified categories are the following: forum and message boards, review and opinion sites, social networks, blogging, microblogging and bookmarking. However, it is notable that the boundaries between the categories are not strict, and in some cases, hybrid forms can exist. This results from the highly dynamic nature of social media. In literature, a range of characteristics is assigned to social media. Agarwal and Yiliyasi (2010) compared social media to industrial media and identified five relevant features, namely, accessibility, permanence, reach, recency, and usability. In terms of accessibility, the main characteristics of social media are its public availability and that it is mostly freely accessible. In contrast, industrial media is often privately owned and not freely available. Social media provides its users with the possibility to edit content anytime, while in industrial media, this is often not possible. In terms of reach, both media have a global reach. A further relevant property of social media is recency, which enables instantaneous communication. In terms of usability, the authors argue that to use social media, no skills are required to generate content, and for industrial media, special skills and training are required. It is important to mention that with a stronger penetration and emergence of social media over the last years, the boundaries between social media and industrial media have been blurred. In this research, it is claimed that the defined characteristics are still applicable to social media and that industrial media have been adapting by using social media as an additional media channel for reaching the audience. In a further approach, Peters et al. (2013) elaborate characteristics of social media content based on five studies (Berger and Milkman, 2012; Kozinets et al., 2010; Leeflang et al., 2012; Liu-Thompkins and Rogerson, 2012; van Noort et al., 2012). They summarise these studies and identify three main characteristics; content quality, content valence, and content volume. Content quality considers content characteristics such as interactivity or vividness and the content domain and narrative styles. Content valence targets emoticons standing for a range of different feelings such as anger or joy. The content valence also addresses the tonality (positive, negative) of social media content. Finally, the content volume summarizes counts and volumes of social media content.

Moreover, social media is considered as one big data source referring to social media big

data (Lin, 2015) or social big data (Guellil and Boukhalfa, 2015). Similarly to Stieglitz et al. (2018) and Sebei et al. (2018), this research assumes that the 5 V's characterizing big data may also apply to social media content. The 5V's are applied in the context of social media as follows.

Volume and Velocity

A massive amount of data is generated through social media on a range of different channels. Content on social media is published in real-time. The emergence of smartphones in combination with high mobile volumes, as well as the availability of mobile application versions of the social media tools, have accelerated the speed of publishing content. In Germany, for example, mobile access to social media from smartphones has increased since 2014 by 23 per cent in 2017. Figure 3.9 [7] provides some examples of the volume and velocity of social media activities per minute on selected social media channels.

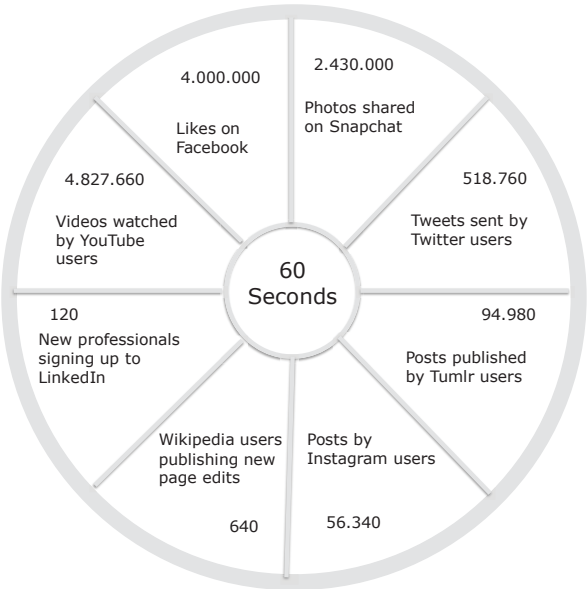


Figure 3.9.: Social media in 60 seconds

Fashion and apparel SC stakeholder will benefit from this massive amount of data generated. Though, it is necessary to implement appropriate technologies and techniques for exploiting this data.

Variety

Social media content is abundant since it involves different types, as illustrated in Figure 3.10. Examples for the richness of social media data are textual data, tags, metadata (posting time/ time stamp/geotag/author's personal information) as well as network data (Hu and Liu, 2012). A post on social media is considered as a basic unit, thus named social media unit.

[7] exracted from https://www.internetlivestats.com/ in September 2019

A social media unit usually consists of the actual content, the message and metadata. Like traditional web data, a social media unit is considered semi-structured data, as it typically comes in XML format.

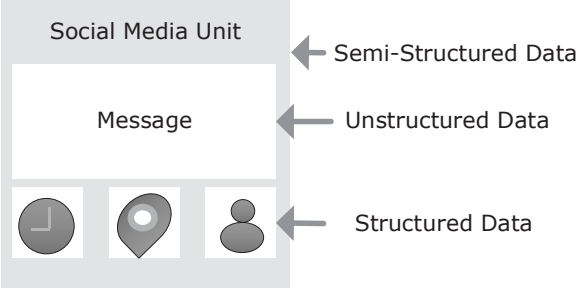


Figure 3.10.: Data Types of a Social Media Unit based on Hu and Liu (2012); Page et al. (2014); Sebei et al. (2018)

Regarding the message, most of the data published on social media are in the form of unstructured data, which include textual data, pictures and videos. Metadata, which is considered information about information (Page et al., 2014), such as the timestamp, location-based data or additional user profile data, is often created automatically and is regarded as structured data. In addition to the message, metadata created on social media may be valuable in decision support since the data can be mined from third parties and applied as additional information to enhance their decision base. In some cases, it is the combination of metadata and message which makes a social media unit useable for a decision maker. For instance, information on the colour or fabric of an apparel product is only relevant for the decision maker with a matched timestamp. Otherwise, the decision maker cannot assign this information to the corresponding season and, in this case, there is no benefit for the decision maker to use this information extracted from social media. In addition to the automatically created metadata, users can assign labels to their posts. Bloggers, for instance, often create labels for each blog post. Hashtags used in Twitter can also be considered as a metadata since they classify the respective Tweet. Page et al. (2014) define this type of metadata as social media metadata. This definition is expanded by referring to Smith-Yoshimura and Shein (2011), including information such as tagging, comments or recommendations. Decision makers can also apply this type of community-based metadata to widen their decision base.

Veracity

In terms of veracity, different issues have to be considered. First, as stated in Agarwal and Yiliyasi (2010) barriers to using social media are low as no skills are required. Second, in contrast to traditional media, there is no filter such as an editor who has to release content before it can be published. These two facts together with automated communication through chatbots, highlight the importance of veracity issues in social media. In big data literature, veracity is mentioned in the context of information security (Demchenko et al., 2013). In

the work of Lukoianova and Rubin (2014), veracity is defined by three categories; credibility, objectivity, and truthfulness. In the social media literature, a crucial factor is that the source of published content on social media is often unknown or unidentifiable, which leads to the challenge of measuring the believability of content published on social media (Shankaranarayanan and Blake, 2017). Believability is a multidimensional construct and consists of different dimensions. Source credibility is one major factor that determines the believability of data (Wang and Strong, 1996). Metadata can be one approach for assessing the credibility of a source (Shankar and Watts, 2003). Thus, social media metadata can also be used for assessing the credibility of a social media unit. Other authors suggest applying the three constructs of identity, expertise and reputation for evaluating the credibility of a source (Shankaranarayanan and Blake, 2017). The consideration of the veracity feature is in particular crucial in the context of decision-making as unverified information might have negative impacts on decision-making processes (Ashwin Kumar et al., 2016).

Value

Similarly to other Big Data Sources, social media content's value becomes evident unless it is processed with adequate methods and approaches. That is, it requires proper processing and exploitation to be useful in terms of decision support (Marr, 2015). Moreover, the value represents the financial profit of an organization generated by big data (Yin and Kaynak, 2015).

Predictive Value of Social Media

Apart from the described characteristics, the predictive value of social media data is addressed in a range of research works. Different application fields, methods and social media channels are hereby focussed. Examples for application fields are political elections (Bermingham and Smeaton, 2011; Chung and Mustafaraj, 2011), flu outbreaks (Broniatowski et al., 2013; Lamb et al., 2013) the prediction of stock markets (Bollen et al., 2011; Pagolu et al., 2016; Rao and Srivastava, 2012), box office revenues (Liu et al., 2016), and product sales (Dijkman et al., 2015; Lassen et al., 2014). Although different social media based sources are used to show the predictive value, Twitter is the most frequently used in the earlier research works as in Bermingham and Smeaton (2011); Bollen et al. (2011); Dhar and Chang (2009); Tumasjan et al. (2010); Zhang et al. (2011). A range of review papers has been published on forecasting using social media data (Kalampokis et al., 2013; Phillips et al., 2017; Schaer et al., 2019). In Kalampokis et al. (2013)the authors classify the existing literature based on different criteria: application area, sources and evaluation approach. The authors identify seven application areas: disease outbreaks, elections, macroeconomics, movies, natural phenomena, product, sales and the stock market. As sources, the authors classify blogs, web search, message boards, reviews, Twitter, Facebook updates and social

multimedia (YouTube, Flickr). Phillips et al. (2017) identify elections, stocks, marketing and sales, public health, threat detection and user characteristics as main application fields. Schaer et al. (2019) classify the following application areas economic indicators, financial markets, public health and environment, services and consumer goods. As data sources, they identify forum and blogs, reviews, search traffic and social networks.

3.2.2. Accessing and Exploiting Social Media Data

The utilization of social media data assumes its availability. Different approaches exist for accessing social media. It is essential to highlight that the process of accessing social media only implies the collection and generation of a data set, and it is not to be used synonymously with its exploitation, which addresses the whole processing phase. This research differentiates between the pure accessing and exploiting process. A significant advantage of social media data is that it is publicly available. For accessing social media data, different approaches exist. The most frequently used approaches for accessing social media data are outlined in the following.

Beheshti-Kashi et al. (2016) propose a tool for collecting and processing blog data. For this purpose, the use of RSS (Rich Site Summary) feeds is suggested. Schieber and Hilbert (2014b) introduce a BPMN based process model for content extraction from blogs. Chaniotakis et al. (2016) suggest a generic framework on data collection from social media. The first decision is on the necessity of building an application for the data collection. This does not include channels in which users publish publicly such as Twitter. The authors distinguish between application-based data collections and non-application-based data collections. A range of technical approaches can be considered for implementing these approaches such as Web Scraping, Application Programming Interfaces (API) provided by social media channel, RSS feeds, search filter and real-time streaming. These approaches are described as follows.

Accessing Social Media data through Web Scraping

Web Scraping is a commonly used approach for accessing web data in general. It describes the collection of data through any means other than a program interacting with an API (Mitchell, 2018). A program queries a web server, requests data which mostly is an HTML format, and lastly parses this data for obtaining the required information (Mitchell, 2018).

Although Web Crawling and Web Scraping are utilized interchangeably, there is a distinguishing feature. While Web Crawling indexes information, Web Scraping is an automated approach for the extraction of content. A Web Crawler is a system for the bulk downloading of web pages (Olston et al., 2010) and is applied for different purposes such as web search engines, Web archiving, web DM and web monitoring services (Olston et al., 2010). An informative overview of Web Crawling in science and practice is presented by Olston et al. (2010). Most of the well-established web technologies offer packages for building an

individual Web Scraper. Some examples are Scrapy in Python or rvest in R.

In the context of social media, the term social media scraping has evolved. Since most social media services restrict the amount and frequency of data sent out for/by bots, scraping information from these services is challenging. Most social media services only allow the scraping of the data through their APIs to keep control of the data extracted. While some social media services such as Facebook and Linkedin even incorporate a mechanism to block potential bots, Twitter, Pinterest, or Instagram do not have such strict regulations.

Accessing Social Media Data directly through official APIs

Most social media companies provide the possibility to access their data through their application programming interfaces (APIs). APIs are usually used by third parties to develop individual software clients to incorporate, for instance, Facebook in other social media services. However, the APIs can also be used to collect (retrieve), store or manipulate data from social media services (Lomborg and Bechmann, 2014).

The individual APIs have structural differences and limitations in accessing the data (Lomborg and Bechmann, 2014). Twitter, for example, used to have quite open regulations in terms of accessing their data. However, the regulations have been restricted within the last years (Lomborg and Bechmann, 2014). In contrast, private profiles on Facebook are mostly not accessible, as privacy settings are often deployed by its users (Lomborg and Bechmann, 2014). Nonetheless, public pages such as company pages are completely accessible (Cui et al., 2018). These examples demonstrate how the functionalities of the individual APIs differ between different channels. In this case, the degree of access to the data type which is usable is different. The terms of use of the APIs may also change over time, without prior notice (Chaniotakis et al., 2016). This may result in disturbances within the data extraction phase of a given project. This also demonstrates a dependency on the respective service. For this reason, it is crucial to carefully consider the different APIs in order not only to be aware of their terms of uses and their limitations but also to monitor their development for any change in the terms of use to be noticed.

When using APIs, it is also required to have alternative extraction strategies. It is not efficient to focus only on one extraction strategy. The main advantage in applying an official API is the access to environments that require authentication as the API runs through the back-end of the respective social media service (Lomborg and Bechmann, 2014). Facebook does not allow developers to have access to the timestamps of likes and shares (Cui et al., 2018). Another advantage of the APIs is that they often come with additional libraries (Twitter4J [http://twitter4j.org] and Facebook4J) and in different programming languages, e.g. in R, Java or Python (Chaniotakis et al., 2016), which facilitate the integration in the individual infrastructure of an organization. There are some challenges in working with APIs in terms of collecting, structuring and analysing data, although the tool used as an interface and the researcher, for instance, may influence the type of questions that are to be answered

using the collected data (Lomborg and Bechmann, 2014). To obtain data from APIs, there are free open-source tools, individual tools developed by research groups/institutions, and commercial SMA tools.

Mitchell (2018) argues that if an API exists, then it is preferable to use it rather than build a bot to access the same data. In some cases, an API may not exist. One possible reason for the absence of an API is small and finite data, which meant no need for the provider to implement an API. Another reason may be a lack of infrastructure and technical ability to provide an API. Even in the presence of an API, its functionalities have to be considered in order to examine if the API meets the requirements of the defined objective, for example, in terms of requested volume, type or format of data (Mitchell, 2018). If the requirements cannot be met, it is recommended to create a scraper.

Accessing social media data through RSS feeds

RSS is a data type for web feeds that provide users with the possibility to access updates from previously subscribed web services. An RSS document usually has a standardized format, for example, in the form of an XML file format. A typical RSS document also includes metadata such as information about the author or published date in addition to the published text. RSS feed has several advantages, one of which is that the standardized structured format allows the integration of a feed in different programmes and tools. Another advantage is that using RSS feeds, the user accesses only the main information/textual data of a website. The navigation bar, for instance, or further social feeds, are not included within the RSS feeds. Some social media services provide the possibility of subscribing to RSS feeds. The feeds give the reader notice about updates on the blog. Different services can be used to aggregate RSS from different websites or social media services and visualize them in one feed in terms of storing the RSS feeds. Moreover, the feeds may be stored in individual databases.

Accessing Social Media Through Search Filter

Since users can publish on every topic on social media, a large amount of data is available, and the variety of topics is large. Nevertheless, the objective is not to collect a massive amount of data from social media, but to collect data by a clearly defined analysis objective. It is, therefore, necessary to define search filters based on both the information needs and analysis objective. The data collection conducted by search filters requires the prior definition of keywords(Kim et al., 2016). A consideration of search filters is highly crucial for the success of the analysis to prohibit the collection of irrelevant information and ensure that all required information is collected. In both stages, the conclusions may turn biased (Kim et al., 2016). The literature describes different approaches for using the right search filter and evaluating the retrieved data. One approach for developing, applying and validating search filters is described in Kim et al. (2016). They introduce the retrieval precision and retrieval

recall for accessing the quality of the retrieved data (see Section 4.5.5).

Accessing Social Media through Real Time Streaming

The velocity in which content is published on social media services is high (see Section 3.2.1). In this way, accessing social media through real-time streaming has certain advantages. One of them is that it also enables real-time analysis. One potential application of this form of accessing social media is sentiment analysis. A real-time sentiment analysis permits the monitoring of customer opinions towards a newly launched product, service, policy or regulation. This monitoring may prevent an emerging shit storm and give the chance to react adequately and develop counteractions. Social media real-time streaming is used for sentiment analysis. Different tools are available for social media real-time processing, such as Azure stream analytics. For detailed information on real-time data processing in Facebook consult Chen et al. (2016). Table 3.2 summarizes the approaches presented for accessing social media data by showing their advantages and disadvantages and examples of social media channels in which such approaches could be used. Based on these, decision-makers can decide which approach is the most suitable one.

Table 3.2.: Overview on extraction methods for social media data

Extraction approach	Advantages	Disadvantages	Examples
Scraping	- If not blocked, no limitations of volume and time - Individual configuration according to requirements	- Mechanism of blocking scrapers - Non-public data not accessible	- Twitter - YouTube - Blogs
API	- Access to data which requires authentication	- Technical imitations in terms of accessible time and volume - Independency to social media service	- Twitter - Facebook - YouTube
RSS feeds	- Structured way of accessing the content	- Availability of RSS feeds	- Twitter - Facebook - Blogs
Search filter	- Targeted extraction	- Missing relevant data	- Twitter
Real time streaming	- Short reaction time		- Twitter - Facebook

It is notable that due to the highly dynamic social media environment, the elaboration of the extraction methods is based on the time of the thesis. Accordingly, there is no claim for

completeness.

Technical Limitations in Accessing Social Media Data

Although there are different approaches for accessing social media data, technical limitations appear in different channels. The limitations are mostly related to the volumes of data, access rate, timely limitations, e.g. how far back in time data can be accessed, type of information, i.e. public/private, and which metadata can be gathered. Due to the fast-paced character within the social media landscape, these technical extraction barriers and regulations may vary without notice. These limitations and regulations have to be considered before selecting the final extraction strategies to allow for a reaction to changes on time and the development of equivalent alternatives. Using social media data involves more than the pure extraction of the data. For supporting fashion and apparel SCs decisions, the extracted data needs adequate exploitation.

Exploiting Social Media Data

As mentioned above, working with social media data is not only limited to its access. Social media data can also be exploited. To exploit social media data, the discipline of SMA has emerged (Stieglitz et al., 2014a). It is defined as "handling with developing and evaluating informatics tools and frameworks to collect, monitor, analyze, summarize, and visualize social media data, usually driven by specific requirements from a target application" (Zeng et al. 2010, p.14). Nevertheless, as this research intends to target SC decision stakeholders and their decision processes, a business perspective is added which is defined as follows: "Business Social Media Analytics refers to all activities related to gathering relevant social media data, analyzing the gathered data, and disseminating findings as appropriate to support business activities such as intelligence gathering, insight generation, sense making, problem recognition/opportunity detection, problem solution/opportunity exploitation, and/or decision making undertaken in response to sensed business needs" (Holsapple et al. 2014, p.4).

The works of Stieglitz and Dang-Xuan (2013) and Stieglitz et al. (2014a) push forward the SMA research by developing a SMA framework. This framework is presented in Figure 3.11. It introduces SMA as "a new research field combining knowledge from multiple disciplines to provide IS research with methodological foundations for gathering, modeling, analyzing, and mining large-scale social media data in order to make business, economic, social, and technical claims from both research and practical perspectives. Furthermore, we argue that SMA can help IS research develop decision-making or decision-aiding frameworks for enterprises that employ social media" (Stieglitz et al. 2014a, p.90).

In their framework, approaches including structural attributes, opinion and sentiment related and topic, and trend related are introduced as the most frequently applied approaches. As methods, they introduce statistical analysis, Social Network Analysis (SNA), sentiment

analysis, content analysis (CA) and trend analysis. In the approaches of structural attributes, SNA is one of the most frequently applied methods that describes the relationships between entities such as persons, organizations or states based on the structure of their connections (Stieglitz et al., 2014a). Possible fields of application are identifying influential or opinion leaders or in general relevant communities in social media. Trend and topic-related analysis are applied to predict emerging topics. Hidden Markov models are often the basis for these trend detecting algorithms. For the detection of emerging topics, dictionary learning may be used, (Kasiviswanathan et al., 2011) where a dictionary learning-based framework is introduced for the detection of emerging topics in social media.

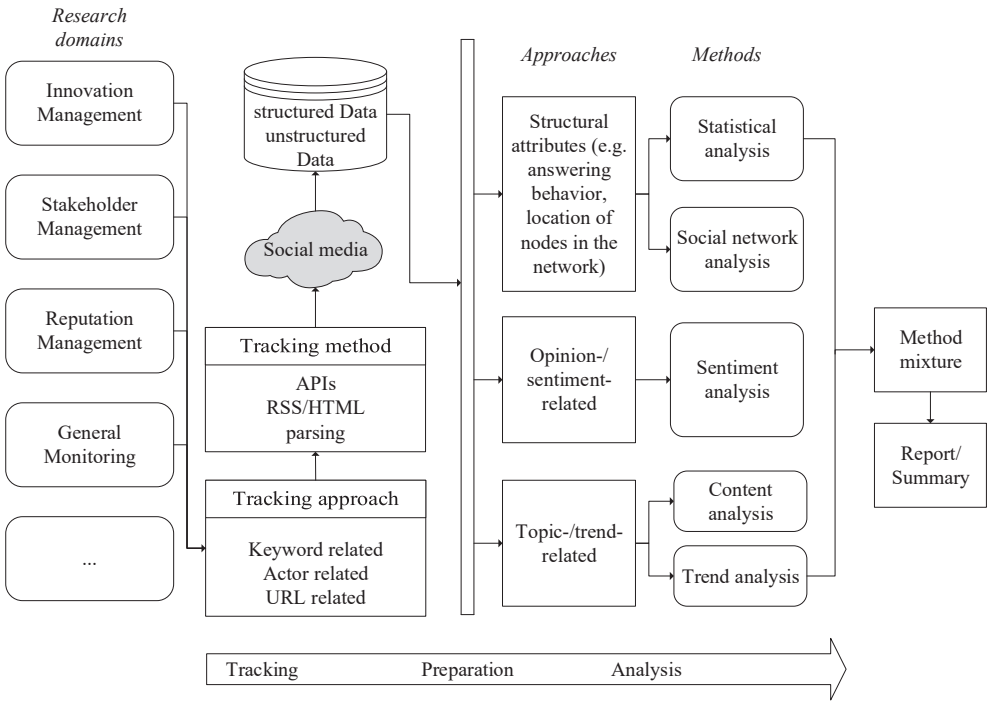


Figure 3.11.: Social Media Analytics Framework (Stieglitz et al., 2014a)

Sentiment analysis examines the value of opinions discussed in textual data, often classified in positive, neutral and negative opinions (Liu et al., 2010). The stream of sentiment analysis has been pushed forward mainly through the development in machine learning methods of NLP and IR and the availability of aggregated review website on the web (Liu et al., 2010). In the literature, different approaches to sentiment analysis are discussed. Two main basic approaches are distinguished: machine learning and lexicon-based. The machine learning approaches incorporate supervised and unsupervised techniques. Supervised techniques are mainly applied where a pre-labelled data set is available. Examples of such techniques are decision tree classifier, probabilistic classifiers (e.g., Naïve Bayes, Bayesian network), linear classifiers (support vector machines, neural networks) and rule-based classifiers. Unsupervised techniques are used in the context of unlabelled data sets. Lexicon-based approaches mainly consist of dictionary and corpus-based approaches (statistical, semantic methods) (Miner et al., 2012).

In their framework Stieglitz et al. (2014a) do not consider the characteristics and challenges of SMA, although in Stieglitz et al. (2018) these challenges are considered within the context of the developed SMA framework for the different stages of topic discovery, data collection and data preparation. At the topic discovery stage, they summarize the challenges of interdisciplinary nature and event and topic detection. At the data collection stage, software architecture challenges are identified. Lastly, in the data preparation stage data quality and data visualization, related changes are brought up. The characteristics of social media result in challenges for exploitation. These are elaborated in the following section. Sebei et al. (2018) conduct a review on SMA process and big data pipeline and introduce a social big data analytics pipeline including data collection, data storage, data preprocessing, data processing, data analysis and data interpretation. The following challenges are assigned to the data fidelity, privacy, security; scalability, availability, integrity; data quality; data streaming, real-time response, heterogeneity, and data visualization.

3.2.3. Exploitation Challenges

According to Marr (2015), the most relevant feature is value since without proper handling, it is not possible to gain the value of the social media content it may reveal. However, before selecting suitable exploitation methods, it is necessary to examine the challenges resulting from the exploitation of social media content.

The main challenges for exploiting social media are spam, contextual relevance, the freshness of information, colloquial usage and intentional misspelling as well as information overload. Figure 3.12 presents demonstrates an overview of these challenges.

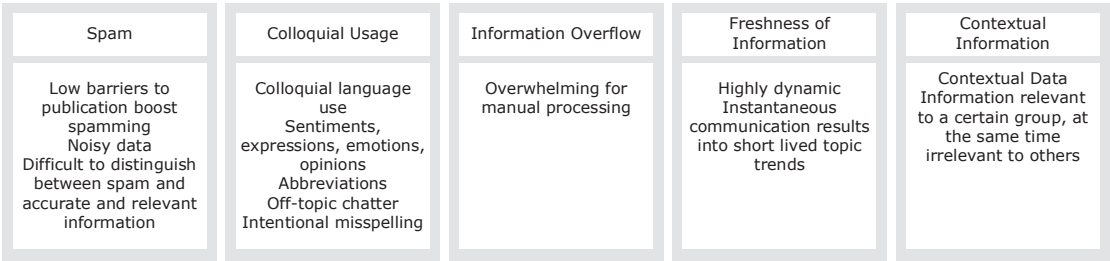


Figure 3.12.: Challenges in social media based on Agarwal and Yiliyasi (2010)

Due to the low barriers in accessing social media, spamming is a widespread problem. Spammers create fake accounts on Twitter or Facebook, create blogs or publish irrelevant comments on YouTube videos. The main objective for spamming is to host content-based advertisements, which may generate revenues if users click on it. Therefore, a significant amount of noisy data exists, making it difficult to distinguish between spam and relevant information. Varol et al. (2017) propose an approach for identifying the so-called social bots, which are (fake) accounts not owned and controlled by real users but by software producing content and interacting with real users. This algorithmic-based content is typically used

to disseminate news (Lokot and Diakopoulos, 2016) and is thus co-responsible for a large amount of spam in social media. The massive amount of spamming gives the feature veracity more importance in terms of decision support.

Another challenge in exploiting social media content is casual and colloquial language usage. Examples are the use of abbreviations, slang or spelling mistakes. Therefore, this data requires more intensive preprocessing. Social media users often intentionally modify spellings such as in ”sooooo nice”. This type of spelling indicates emphasis and should not be disregard as a simple spelling mistake. It should not be treated as an unintentional spelling mistake and not be corrected within the preprocessing phase since a corrective adaption may change the content valence and may result in incorrect deductions for the decision-maker. Within the challenge of colloquial usage, there is a high emergence of off-topic communication, underpinning the necessity of extensive preprocessing measures.

A further significant challenge is dealing with information overflow, which results from the volume, velocity and variety of social media. The velocity additionally leads to a highly dynamic environment, in which topics are short-lived and which fosters the freshness of information (Hayes, 2007).

The contextual relevance of social media data also results in exploitation challenges, as some information may be relevant to a specific group and at the same time irrelevant to others. More advanced exploitation methods are required to understand this relevance.

The described characteristics of social media-based sources and the published content illustrate the relevance of assessing the IQ when working with social media content, particularly when utilizing social media content for decision support. However, IQA is not a trivial task since quality is a multidimensional construct (Borchers, 2009) involving different aspects. For this reason, IQ is not directly measurable.

3.2.4. Information Quality Assessment in Social Media

The objective of this section is to give an overview of quality assessment approaches of social media data. For this, IQ dimensions are introduced, and the assessment of IQ is put forward.

Information Quality Dimensions

In literature, the terms information quality and data quality are often used synonymously. The definitions follow the concept of fitness of use as elaborated by Tayi and Ballou (1998). This concept was also used by Wang and Strong (1996) who adopted the Total Quality Management approach to data quality and introduced the term data consumers. They follow an empirical approach to identify data quality attributes relevant to data consumers and introduce a hierarchical framework. The authors identify 20 dimensions generated by four data quality categories. The four categories are Intrinsic Data Quality, Contextual Data Quality, Representational Data Quality, and Accessibility Data Quality (see Figure 3.13). A data

quality dimension is defined as a set of data quality attributes that represent a single aspect or construct of it (Wang and Strong, 1996).

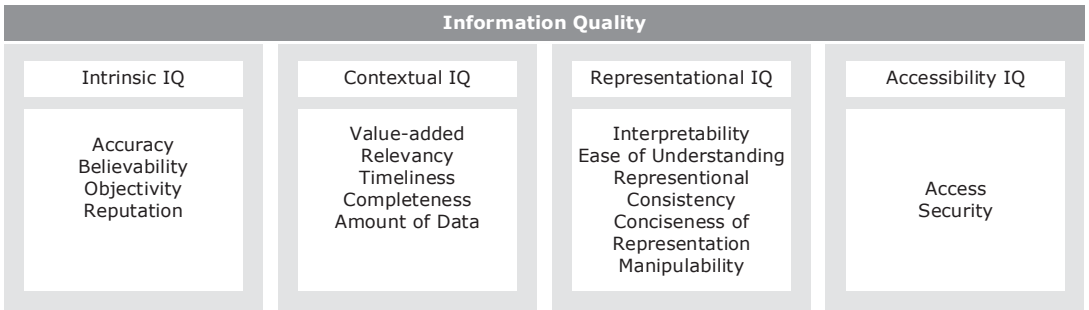


Figure 3.13.: Data Quality categories and dimensions (Wang and Strong, 1996)

A high correlation between the factors accuracy, believability, objectivity, and reputation of data was identified in the intrinsic category. Moreover, the authors state that companies often focus too much on accuracy as the only dimension of data quality. In contrast, professionals working in the field of Information Systems concentrate more on the accuracy and objectivity of data. Fisher et al. (2011) suggest that believability might be much more important than accuracy since people are driven by their beliefs. Furthermore, they reveal that several factors might determine the believability dimension, such as knowledge, experience and the degree of uncertainty in the data. In the second category, contextual IQ, the dimensions relevancy, completeness, value-added, timeliness and amount of data are summarized. The third category is representational IQ consisting of the dimensions interpretability, ease of understanding, representational consistency, conciseness of representation and manipulability. The fourth category is accessibility IQ and consists of the dimensions of access and security. These dimensions are intended to assess the quality of transactional data. While transactional data gives a complete picture of a transaction and is collected and run by organizations, the original purpose of data published on social media is not its collection (Shankaranarayanan and Blake, 2017). Moreover, the source is often not identifiable or not known (Shankaranarayanan and Blake, 2017). This leads to the question of how to understand the believability or relevance of social media content. The unknown and unidentifiable source is considered a key challenge in measuring the believability of social media content. Hence, it is necessary to elaborate on the IQA literature on social media-based sources.

Information Quality Assessment of Social Media data

As one of the earliest researchers targeting the topic of quality assessment of social media content, Agarwal and Yiliyasi (2010) mapped the challenges relating to spam, contextual relevance, colloquial usage and intentional misspelling, information overload and freshness of information to the IQ dimensions suggested by Wang and Strong (1996). They propose that the dimensions accuracy, believability, reputation, value-added, relevancy, timeliness,

amount of data, ease of understanding, manipulability and conciseness cover the challenges of social media content.

Shankaranarayanan et al. (2012) examine the applicability of the dimensions accessibility, accuracy, believability, completeness, consistency, relevance and timeliness on social media content. The authors select those dimensions that have an impact on practice and organizations. As this research examines the utilization of social media data in an organizational context, the work discussed in Shankaranarayanan et al. (2012) is taken as a baseline for selecting adequate dimensions. This work concludes that the dimensions accuracy, believability, timeliness and relevance are essential when assessing the quality of social media data. As the measurement of completeness is based on a given data structure and social media data lacks such predefined structure, the assessment of completeness for social media data is not possible (Shankaranarayanan et al., 2012). Similarly, measuring the consistency of social media content is hardly possible due to the lack of predefined structures (Shankaranarayanan et al., 2012). The dimension accessibility assesses the ease of the attainability of data (Shankaranarayanan et al., 2012). While accessing transactional data may be restricted for different reasons such as security, sensitivity, or privacy, social media data is meant to be easily accessible through the API's offered by social media providers. For this reason, the authors do not evaluate the importance of the accessibility dimension in the context of social media as relevant as for transactional data.

This research agrees on the difficulties in measuring the completeness and consistency of social media content and does not consider both dimensions for assessment. Nevertheless, the accessibility dimension is regarded as relevant in the context of social media for two reasons. First, the provided API's often have restrictions, for example, in terms of volume or time. Second, considering the perspective of data consumers, accessibility can also be defined in terms of language skills. If information is published in a different language than a data consumer is skilled in, then the information is not easily accessible for the data consumer. The need to consider the believability dimension when working with social media content emerges through both the low entry barriers in social media usage and the lack of filter and editing before the publishing process. In particular, for transactional data, the accuracy dimension has often been considered as the most relevant one while assessing quality. In literature, it refers to the correctness of given data compared to a given baseline As this is often not available, its measurement is difficult. This is especially true in the social media environment. However, similarly, for transactional data, the accuracy dimension is considered an important dimension for assessing social media content (Shankaranarayanan et al., 2012). The importance of agility and speed in decision making in the context of SCM is undisputed. The currentness of data has always been a crucial factor gaining even higher importance in the SC context (Hofmann, 2017), which gives a high relevance to the timeliness of data. In summary, this research considers the following dimensions of IQ: accessibility, accuracy, believability, relevance, timeliness.

Among these dimensions, the literature focuses on believability. However, believability is transported by different terms such as truthfulness (Lukoianova and Rubin, 2014) or credibility (Cai and Zhu, 2015; Metzger and Flanagin, 2013). In particular, the terms credibility and believability are considered to be the same in most contexts (Fogg and Tseng, 1999). Source credibility is a crucial factor in defining believability (Shankaranarayanan et al., 2012). According to Shankaranarayanan et al. (2012) source credibility in social media can be defined by the constructs of identity, expertise and reputation. Furthermore, the data consumer's domain expertise plays a crucial role in assessing the believability of social media data as demonstrated in Figure 3.14.

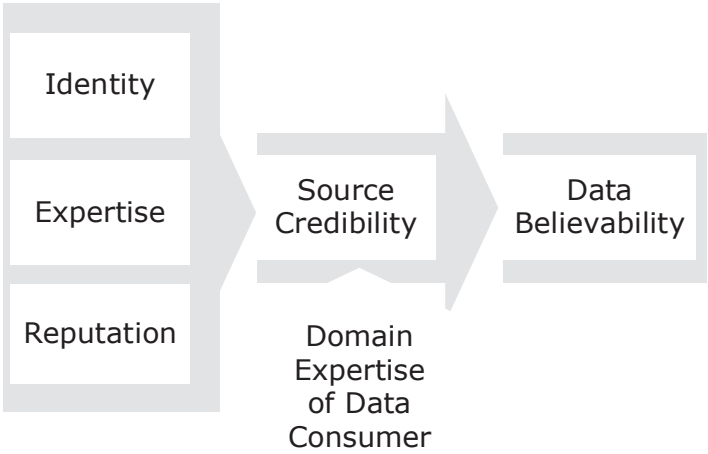


Figure 3.14.: A model for assessing data believability (Shankaranarayanan et al., 2012)

To gauge identity, it is crucial to be identifiable. In terms of assessing expertise, it is important to see how much expertise the data provider has. Reputation considers the influence of the data provider within its community. Notably, in the context of fashion and apparel, the term influencer has emerged. Using the influencer concept, reputation may be assessed. For this purpose, a range of tools exists for the identification of key influencers in social media. Network structure metrics, such as the number of followers or page rank as well as activity-centric metrics often based on the frequency of posts, or the number of likes, are typical approaches for identifying key influencers (Peng and Sun, 2012). Nilsson and Alserud (2017) analyse in their master thesis the different dimensions of believability and refine the model of Shankaranarayanan et al. (2012). They report a larger importance of the dimensions identity and reputation compared to expertise. At the same time, they confirm the significance of the source credibility in the context of social media. This thesis is aligned with both works in the relevance of the source credibility, though, the model as displayed in Figure 3.14 is taken as reference point.

3.3. Exploiting Social Media Data for Fashion and Apparel Supply Chain Decisions

Following the DSRM to develop a methodology for exploiting social media data to support fashion and apparel SC decisions, it is necessary to consider the current handling of information in fashion and apparel SC decisions.[8]

3.3.1. Social Media use in Fashion and Apparel Supply Chains

This section aims to give an overview of research using social media for fashion and apparel SC processes. Several researchers use Facebook as one source of applying for fashion and apparel SC processes. One of the main works conducted on using social media information in fashion sales forecasting is the work of Cui et al. (2018). They examine the operational value of social media information by incorporating social media information in sales forecasting. The main objective of their research is to quantify value. They use operational information from an online apparel retailer including sales, advertising and promotions, and publicly available social media information from Facebook (the company's official Facebook page). The Facebook data includes the number of comments, the number of posts representing the volume feature, the average length of comments, and the average sentiment of comments representing the valence feature. Sentiment analysis approaches are applied to capture the sentiment of the comments, and each comment is classified into positive, neutral and negative. As forecasting models, a range of statistical models such as linear regression, support vector machine or random forest models is adapted. The authors report a statistically significant improvement in the company's internal forecasts when incorporating the proposed approaches.

Similarly, Boldt et al. (2016) incorporate social media data for the sales forecasting of sports apparel using Nike's ten most active Facebook pages based on total likes. Social media features such as the number of likes, total posts, total comments, total shares, total unique actors, and unique commenters are selected from their Facebook pages. Quarterly global sales figures are taken from Nike's financial reports. Using these two datasets, correlation analysis was conducted in order to explain global sales. The authors report a high forecasting accuracy for some of the simple regression models. Additionally, an event study was conducted for examining the impact of Nike's real-world activity on the activity of the Facebook page(s). Overall, the authors contribute to assessing the informational value of social media data for companies in marketing, sales operations and SCs. Both Boldt et al. (2016) and Cui et al. (2018) investigate the potential of companies using their Facebook pages to improve

[8]Some content of this chapter is published in Beheshti-Kashi and Kinra (2020)

the forecasting accuracy of apparel companies. af Rosenborg et al. (2017) examine the relation between the activity of HM's Facebook page and the company's financial performance by considering total posts, comments and number of likes, and a relationship between the Facebook and corresponding business data.

A recently published work by Iftikhar and Khan (2019) presents a framework for demand forecasting in SCs utilising social media data. In this framework, Twitter and Facebook serve as data sources, and a case study is conducted around a retail apparel company. Their framework is structured into data collection and preprocessing, sentiment extraction and forecast model. In this way, sentiment, trend and word analysis are included in an extended Bass emotion model to predict sales. Brand related, product type related and fashion trend related data are extracted and further processed. Following the framework developed, the authors report a positive effect on the obtained accuracy.

Choi (2018) focuses on Quick Response and shows an approach where retailers use social media comments from different platforms to adjust their ordering decisions towards future demand. The author reports that having good social media comments for coming products has a crucial impact on the retailer's decision-making. Good social media comment in this context is the consistency with other market demand observations. It is shown that the retailer is more confident in its forecast accuracy if social media comments and other market demand observations are consistent.

A different focus on social media considerations in SCs is presented by Shen et al. (2017). They focus on the fashion retailing industry and develop an analytical model for examining the effects of demand changes on a luxury fashion SC with social influences. The consumers of luxury fashion are usually categorised into fashion-leaders and fashion followers. The authors assign fashion leaders great social power and refer to the followers as normal users who make their purchases based on their information via social media. Demand changes are defined as the consumer response to the information obtained through social media on purchasing behaviour. Based on their results, the authors propose managerial implications. In terms of the potential impacts of demand changes on decision making in the luxury SC, it is suggested that in the case of small demand changes, for instance, it is enough to adjust the price on the product web page.

Yet another approach is followed by See-To and Ngai (2018). The authors focus on short-run models for demand distributions and sales forecasting using the term nowcasting. They develop a method for visualising demand distributional characteristics and apply it to a Chinese apparel e-commerce platform using customer reviews. The author report that analysed customer reviews contain useful information and can result in improved forecasting accuracy even by applying simple models. Based on their results, the authors claim to incorporate customer reviews for nowcasting sales of online shops. In this way, optimised inventory management can be implemented.

3.3.2. Use of Blogs for Fashion Processes

Apart from Facebook, fashion blogs are discussed as an additional potential source for SC processes. Fashion blogs are the biggest group amongst the blogosphere (Halvorsen et al., 2013) and are defined as "blogs that focus on fashion brands, trends, products, e-commerce, street style and personal style" (Halvorsen et al. 2013, p.213). Fashion bloggers have gained increased relevance from the fashion industry and are considered opinion leaders or influencers (Uzunoğlu and Misci Kip, 2014). Usually, they have broad outreach to their readers and followers. Their impact on them is often explained by the phenomenon that the influence of interpersonal communication on an individual's behaviour is larger than the mass medias' impact (Weimann, 1994). The industry has realized the actual impact from the blogger on their reader, which might be a potential customer of fashion companies. The first fashion bloggers were invited to report from the New York Fashion Week in 2006. Accordingly, the companies have developed different strategies for cooperating with fashion bloggers in order to be able to use their reach out to the readers of the bloggers. Bloggers are invited to events, products are provided to them, and presented on their blogs (Kulmala et al., 2013). It is reported that blogs are fundamental in the decision-making process of consumers (Esteban-Santos et al., 2018).

Research on blogs and fashion processes is conducted on their use in influencing consumer behaviour, such as in Halvorsen et al. (2013), or in comparing brand associations from blogs and textual data published by fashion companies themselves in Crawford Camiciottoli et al. (2014). Similarly, Esteban-Santos et al. (2018) conduct several analyses based on a survey for exploring the influence of fashion bloggers in the Millennials' buying behaviour in the Spanish context. While these studies focus on the consumer perspective, the following research works consider a potential value for decision-making in SC processes.

A first approach to the use of fashion blogs in fashion processes was suggested by Rickman and Cosenza (2007). They proposed a weblog-text trending approach and focus more on catching the actual *buzz* from the fashion posts. They conclude that for the use of fashion blogs for trend forecasting purposes, different changes and developments are needed, including rich accumulation of fashion communication or the acceptance of fashion bloggers by the marketers. In this research, it is claimed that both developments have occurred within the last years.

Beheshti-Kashi et al. (2015b) introduce the TrendFashion tool for the collection and analysis of fashion blog posts. The focus is on the identification of fashion trends for supporting forecasting decisions. They report the identification of features such as colour or style information to a specific product. In a further study, Beheshti-Kashi et al. (2016) examine Italian fashion blogs and real-world data from an Italian retailer to find similarities between both data sets. For this study, the focus is on the product feature colour. The authors find out that for some colours, the insights of both data sets correspond. Beheshti-Kashi et al.

(2018) illustrate a decision-support case in the SMA framework using fashion blogs. Experimental analyses are conducted. The potential of fashion blogs in fashion and apparel SCs are showcased, and the potential benefits of SC stakeholders are discussed in particular fashion buyers. The illustration focuses on topic detection (applied on colour detection) and the tracking of fashion topics. For the tracking task, a five-year time period of fashion blogs posts was used. The authors report that the development of some fashion topics was observable. Following these analyses in this context, before finalizing the production plans, the buyer might have the possibility to extract information on colours, co-occurred colours or track topics presented on the fashion shows from the time between the shows and their order deadlines for the next collections. This procedure would allow him to catch the actual coverage on the different fashion topics from potential customers and, consequently, reduce the buyers' own decisions in selecting the collection.

3.4. Summary

Chapter 3 presented and elaborated the basic concepts and characteristics of fashion and apparel SCs, social media and TM. The problem setting formulated in Chapter 1.1 is specified by sketching a schematic of a fashion and apparel SC and demonstrating the dilemma of the stakeholders.

Within fashion and apparel SCs, decisions have to be made at specific time instances relative to the point of sales. Due to the peculiarities of fashion and apparel SCs, such as high volatility, these decisions are typically made with uncertainty as necessary information are hardly fully available.

For supporting these decisions, company knowledge regarding, e.g., historical sales data is used to predict and anticipate properties such as volume or colour of products (Hofmann, 2017). As a consequence, different stakeholders are dependent on the availability of colour information for their decisions. As elaborated in Section 3.1.3, sales forecasting is a highly crucial decision process in fashion and apparel SCs. Different stakeholders along a fashion and apparel SC need to conduct forecasting. For sales forecasting processes, various data and information sources are used. Due to the seasonal nature of fashion products, decision-makers are often confronted with limited historical data. Accordingly, the literature also proposes methods which are able to handle the sparse data or approaches which require access to different data and information sources. For enhancing the limited data, additional information sources are thus consulted by decision-maker.

In accordance with the research scope, Chapter 3.2 further elaborates on social media. Classified as a big data source, this chapter applies the big data's 5 V's features for characterising social media, namely volume, velocity, variety, veracity and value. These characteristics pose challenges. In particular, the veracity characteristics is considered crucial in this dissertation. This research argues that the veracity characteristic can be approached by

incorporating IQ. Therefore, IQA approaches discussed in the literature are presented. The analysis outlines that for social media quality assessment, the dimensions of accessibility, accuracy, believability, relevance and timeliness are of high significance. The IQ dimensions are incorporated into the elaboration of the requirements in Section 4.1.

The public availability of social media data stands out as the main advantage of social media. However, the question of how to access and exploit social media data remains. This is brought by Section 3.2.2 providing an overview of access and exploitation of social media data. This research argues that pure access to social media data is insufficient for utilising social media data as additional information sources for decision processes. The utilisation of social media data has to be integrated into a systematic process model. The literature introduces SMA as a framework to be used when handling social media data. Although existing SMA frameworks are designed for managing social media data, social media is not elaborated in detail. For example, this means that the characteristics of social media and social media data are not considered. This thesis mainly targets bridging the gap between the social media characteristics and a methodology developed correspondingly for the exploitation of social media data.

Besides structured data, semistructured data, a significant part of data being generated in social media is unstructured data. Of this unstructured data, a large fragment is textual data. While structured data can be directly analysed, textual data requires adequate preprocessing in order to be exploited. For this reason, a methodology for exploiting social media data for SC decisions requires the consideration of TM methods. The handling with textual data is incorporated in the requirements in Section 4.1.

Having elaborated on the characteristics of fashion and apparel SCs, social media and TM, the foundation for developing a systematical methodology incorporating social media as an additional source for fashion and apparel SCs is laid.

Apart from laying this foundation, the overview of existing approaches of social media use in Fashion and Apparel SCs is given by Section 3.3. When looking at these works, it becomes evident that social media data is used as an additional data source for different operational processes without considering social media further.

For instance, one social media channel is selected without considering others and without following a predefined approach. Furthermore, the peculiarities of fashion and apparel SCs are hardly considered while selecting social media data. Involved stakeholders in particular need to be considered and their potential added-value. Ideally, a decision-maker would seek further information sources in order to reduce the given uncertainty. The work of Iftikhar and Khan (2019) try to enable a little understanding of the data by including word analysis and apply different preprocessing steps of the extracted social media data. Still, there is a lack of considering the characteristics of social media data in their framework proposed. Therefore, a proposed methodology for exploiting social media data should provide usable information for the decision-maker. In order to meet both elaborated challenges and prevealed lacks,

Section 4.1 will derive the requirements for the process model. Based on these requirements, the process model is developed.

4 Development of the process model

The objective of Chapter 4 is to illustrate the design & development of the artifact, the process model. The development of the process model is aligned with the research scope as stated in the research objective presented in Section 1.2: developing a methodology for the exploitation of social media data for supporting fashion and apparel SCs decisions. The methodological foundation for the process model development is laid in Chapter 2. In this chapter, firstly, the requirements for the process model are presented (see Section 4.1). Based on the requirements, four main components are developed. These are the **Process Layer** (see Section 4.2), **Information Source Layer** (see Section 4.3), **Social Media Layer** (see Section 4.4) and **Text Mining Layer** (see Section 4.5). While the Process Layer focusses on the processes, objective and needs of the decision-maker, the Information Source Layer considers the potential sources for the defined processes and needs. The Social Media Layer is designed for the exploitation of social media and the Text Mining Layer introduces steps for processing textual social media data.[9]

4.1. Requirements

This section elaborates on the requirements and functionality of the process model. The requirements are distinguished in structural and functional requirements. The foundation for deriving and elaborating these requirements of a process model for exploiting social media data to support fashion and apparel SC decisions with social media data is laid in chapter 3. The requirements are derived based on the characteristics of fashion SCs and social media described in the literature. Moreover, the requirements are prioritised according to their importance for the design & development of the process model. The overall framework of the requirement derivation is illustrated in Figure 2.2. Based on the requirement identification approach, the requirement section is organised accordingly. The requirements derived

[9]Some content of this chapter is published in Beheshti-Kashi (2020); Beheshti-Kashi and Kinra (2020); Beheshti-Kashi et al. (2018)

from the fashion and apparel SC characteristics are presented in Section 4.1.1. Section 4.1.2 is dedicated to the requirements of manufacturers and retailers. This is followed by the requirements derived from the social media characteristics in Section 4.1.3 and the requirements based on the characteristics of textual data and TM frameworks in Section 4.1.4. Apart from the functional requirements, structural requirements are derived based on the characteristics of process models (see Section 4.1.5). The summary of the requirements is presented in Section 4.1.6.

4.1.1. Fashion and Apparel Supply Chains

Perspective of Supply Chain stakeholder

It is important to ensure the perspective of the data consumer, e.g., an SC manager, while dealing with social media data to support fashion and apparel SC decisions. The relevance of involving the data consumer's perspective arises from two observations. Firstly, following the concept of *fitness for use* (Tayi and Ballou, 1998), data should be usable for the data consumer, i.e., fit for the purpose (see Section 3.2.4). In order to ensure this fitness, it is proposed to capture the data consumer's information needs and requirements in an early stage of the data processing phase. Secondly, the data consumer's domain expertise should be considered while assessing the credibility of a source (see Figure 3.2). A typical sales forecasting process often incorporates individual knowledge from SC managers or buyers, in this case, the data consumer. In practice, the buyer uses its expertise and assesses baseline forecasts based on historical sales data. The assessment is incorporated mainly in decreasing or increasing the predicted sales data (see Section 3.1.3). Similarly, it is suggested that the domain expertise of the data consumer is be applied for assessing the source credibility of potential sources.

- Req_FASC 1.1: The process model should include the perspective of the decision-maker.
- Req_FASC 1.2: The process model should include the objective of the decision-maker.
- Req_FASC 1.3: The process model should provide methods for extracting information relevant to the stakeholders decision-process.

Different needs of Supply Chain stakeholder

An apparel product constitutes different features such as colour,fabric, print, or style (see Section 3.1.4). A decision-maker may have different information needs for fulfilling the analysis task. The process model should accordingly provide the option for the decision-maker to customize the analysis according to his current needs. This also implies the diversity of product features.

- Req_FASC 1.4: The process model should include the different needs of the decision-maker.

INVOLVEMENT OF DIFFERENT STAKEHOLDER

Fashion and apparel SC processes are complex. Decision-making processes in fashion and apparel SCs are specific. A range of processes and decisions with different stakeholders has to be coordinated. Different stakeholders are involved in decision-making processes (see Section 3.1.2). For this reason, it is important to identify the involved stakeholders conducting a decision process.

- Req_FASC 1.5: The process model should include the different stakeholders involved in a decision process.

4.1.2. FASHION AND APPAREL SUPPLY CHAIN STAKEHOLDER

In order to ensure that the process model addresses the needs of fashion and apparel SC stakeholder, this section recalls the different stakeholders' roles (based on Section 3.1), and derives requirements based on their main decision processes. The focus is on the manufacturer and retailer, as they play a crucial role in fashion and apparel SCs.

MANUFACTURER

The **sales representative** or **the sales team** on the manufacturer side is responsible for managing the day-to-day relationship and providing services to designated B2B customer. The sales representative work closely with the buying and product design team of the customer (retailer) and requires to understand the needs of the customers. Based on this role description, the following user story and requirement meeting the sales representatives' needs are formulated.

1. *User story: As a sales representative I need to understand the preferences of the target group of the retailer so that I can provide the retailer the right samples.*
    - Req_M1: The sales representative shall obtain information relevant to product features.

The **product developer** or the **product development team** design and engineer products that are serviceable for their retailers. The design process begins with a product line idea, which explains the mood, theme, and other key elements that contribute to the identity of the line. Product developers require material specifications, analyse cost, and finally develop the product at a quality, and level, pleasing both producer and consumer of the designed products.

1. *User story: As a product developer I need to understand market trends so that I can develop a range compatible to production.*
    - Req_M2: The product developer shall obtain information relevant to product features.
    - Req_M3: The product developer shall assess believability of extracted information.

RETAILER

The **fashion buyer** or the **buying team** have a crucial function, and their decisions impact the success of the company (Wong, 2013). Typically, buyers have various responsibilities in a company. They manage the range and are responsible for negotiations with suppliers, negotiating, among other factors, about prices and deliveries. Furthermore, buyers identify target groups and examine the buying behaviour of the customers. They predict the sales of the products. Moreover, they continuously monitor upcoming trends to react appropriately to market changes (Jackson and Shaw, 2000). Buyers should be aware of past seasons' sales, upcoming trends, and in particular, of consumer needs. Based on the tasks and responsibilities of the buyer, the following user stories can be derived.

1. *User story: As a buyer, I need to collect customer preferences so that the range can be selected according to customer needs.*
    - Req_R1: The buyer shall obtain information relevant to product features.
2. *As a buyer, I need to collect reliable customer preferences so that the accuracy of sales forecasts increases.*
    - Req_R2: The buyer shall be able to assess the believability of extracted information.

The **designer** or the **designer team** handle all the design processes from the beginning, such as concept, sketching, working with the pattern makers or fitting. A key responsibility is to constantly track current and future fashion market trends to design a garment that meets consumer demand. Designers draw inspiration from a wide range of sources. Form these responsibilities, the following user story and based on this, requirements are derived.

1. *User story: As a designer, I need to collect information related to garment product features so that I can design garments according to customer preferences.*
    - Req_R3: The designer shall obtain information relevant to product features.
    - Req_R4: The designer shall obtain information on product features with respect to the time.

4.1.3. Social Media

As described in Section 3.2 social media is characterized by several features. Based on these characteristics, requirements are defined in the following. Overall, 18 requirements addressing social media characteristics are derived.

Variety of data types in Social Media

The richness of social media data constitutes the availability of unstructured data such as textual data or multimedia data, structured data, which is mostly metadata as such posting time, timestamp, geotag and also network data. Following the variety features (see Section 3.2.1) describing the diversity in data that is published in social media, it is proposed that the process model should consider different types of data. To make the various data usable for the decision-maker, the process model must be able to manipulate the different types of data that exist on social media.

- Req_SM 1.1: The process model should process textual data.
- Req_SM 1.2: The process model should deal with the processing of multimedia data.
- Req_SM 1.3: The process model should deal with the processing of network data.
- Req_SM 1.4: The process model should deal with the processing of metadata.

Variety of channels in Social Media ecosystem

Within the social media landscape, different social media tools exist (see Figure 3.8). The literature presents different categorization schemes for classifying social media channels. Though, due to the dynamic and fast-paced nature of social media, new channels arise. Apart from different channels, each channel has its individual characteristics, a different structure, different features or different target groups. From this, it is derived, that the process model should give the decision-maker the opportunity to include different channels in the analysis. Moreover, it can be that for different decision processes, specific channels are more relevant than others.

- Req_SM 2.1: The process model should deal with the diversity of social media channels.
- Req_SM 2.2: The process model should deal with the differences in the structure of the channels.
- Req_SM 2.3: The process model should provide the decision-maker activities for identifying channels specific to the needs.

Social Media Analytics

As elaborated on in Section 3.2.2 SMA targets the development of tools or frameworks for collecting, monitoring, analyzing, summarizing, and visualizing social media data. The development of such frameworks is driven typically by specific target applications. From a business decision-making perspective, these activities have to be seen in the context of data relevant to the decision processes of the SC stakeholder (Holsapple et al., 2014). To this end, a deployment of suitable methods that assure that the exploitation is in accordance with the defined analysis task and the stakeholder's information needs is necessary. This also includes that, for instance, if network data is collected, then the decision-maker should have the possibility to apply network analysis. This goes along with the fifth feature of social media, i.e. the value. Unless social media data is not processed correctly, its value can not be exploited (see Section 3.2).

- Req_SM 3.1: The process model should contain phases dedicated to SMA relevant activities.
- Req_SM 3.2: The process model should provide methods for ensuring the extracted data are relevant to the SC stakeholders decision process.
- Req_SM 3.3: The process model should contain different methods for extraction of social media data.

Accessing Social Media data

The availability of social media data is the prerequisite for its use in fashion and apparel SC decision processes. There exists a range of options for accessing it (see Section 3.2.2). Commonly used approaches are web scraping, the use of the official API's, or RSS. Different social media channels provide different accessing approaches. Due to the various channels and different structures that these channels have, the decision-maker needs to decide on a suitable extraction method.

- Req_SM 4.1: The process model should provide different methods for extraction of social media data.

Colloquial Usage in Social Media

Exploiting social media data comes with a range of challenges (see Section 3.2.3). One noticeable challenge is the use of casual and colloquial use of language. Examples are the use of abbreviations, slangs or spelling mistakes or intentional misspellings. Social media users often utilize these spelling mistakes for underlying their expressions. For this reason, a simple correction of these mistakes would change the tonality of the statement. This might lead to wrong conclusions made by an SC stakeholder.

Apart from the language aspect of colloquial usage, the availability of sentiments, expressions or opinions is also considered under this challenge (see Figure 3.12). This goes along with the feature of content valence, addressing both the availability of emoticons representing different feelings and the existence of tonality (positive, negative) on social media content (see Section 3.2.1). From the SC stakeholder perspective, a correct catching of these expressions can unlock considerable value. Typically, the SC stakeholder relies on its individual expertise when adjusting baseline forecasts based on historical sales data. Thus, being able to adequately exploit these available sentiments, the SC decision-maker can base these kinds of adjustments on the basis of a range of sentiments, not only on its individual expertise.

- Req_SM 5.1: The process model should deal with colloquial language use.
- Req_SM 5.2: The process model should deal with intentional misspellings.
- Req_SM 5.3: The process model should provide methods for exploiting the tonality of social media content.
- Req_SM 5.4: The process model should deal with methods for dealing the existence of emoticons.

Veracity of Social Media

Veracity is one noticeable feature of social media (see Section 3.2.1). Given that the barriers to using social media are low and there is a lack of filter before the publishing process, social media faces spamming and noisy data (see Section 3.2.3). The importance of the veracity characteristic makes the assessment of social media data necessary when applied to fashion and apparel decisions. From an SC stakeholder perspective, potential data that are likely to be used for supporting decision processes need to meet some quality aspects because poor data quality impacts decisions (see Section 1.1). As it is pointed out in Section 3.2.4, quality assessment constitutes the consideration of several dimensions. The dimensions accessibility, accuracy, believability, relevance, and timeliness, are considered suitable dimensions for assessing the quality of social media data. Accordingly, it is proposed to apply them within the process model.

- Req_SM 6.1: The process model should include IQ mechanisms.

Data provider

Similarly to incorporating the data consumer's perspective, the data provider should be reviewed. The consideration of the data provider is particularly crucial in the context of the veracity feature. As elaborated and pointed out in Section 3.2.4, the believability dimension is amongst other terms transported through the credibility. Source credibility is a significant

factor in defining the believability (Shankaranarayanan et al., 2012). As illustrated in Figure 3.14, the source credibility is defined by the constructs of identity, expertise, and reputation. By assessing the source credibility, the data provider is taken into account. For extracting believable content, it is necessary to assess the believability of the data provider. In this regard, the following two requirements are stated. Conducting a believability assessment, the SC stakeholder can weight exploited data in a more structured manner. For instance, content extracted from specific fashion bloggers may have a higher impact on its decisions than the information extracted from the ”mass”. Or the other way around, depending on the information needs and requirements of the decision-maker.

- Req_SM 7.1: The process model should include the perspective of the data provider.
- Req_SM 7.2: The process model should contain a believability assessment mechanism.

4.1.4. Text Mining

A large portion of social media data is in the form of unstructured data. This type of data needs proper handling and can not be covered by methods for structured data. In this way, the process model must be able to manipulate textual data. The general requirement for handling textual data is already considered in Req_SM 1.1. However, looking into existing TM frameworks, it turns evident that the Text Mining process itself underlies a range of requirements (see Section 2.2). In this regard, Schieber and Hilbert (2014a) introduce a generic TM process model based on existing TM frameworks.

- Req_TM 1.1: The process model should include requirements of a TM framework.
- Req_TM 1.2: The process model should include methods required for processing textual data.
- Req_TM 1.3: The process model should include the SC stakeholders perspective in evaluating generated results.

4.1.5. Process model design

In addition to the derived functional requirements based on characteristics of fashion and apparel SCs, social media and TM, it is necessary to consult characteristics assigned to the design of process models from a methodological perspective. To this end, the specifications of process models are analyzed and presented in Section 2.2. A sequential order, cluster of similar activities or formalization are the key features of process models. In accordance with process model characteristics, three structural requirements are stated in the following.

- Req_SR 1.1: The process model should be designed in compliance with typical components of a process model.
- Req_SR 1.2: The process model should be designed based on a formalization.
- Req_SR 1.3: The process model should involve feedback loops enabling a steady improvement of the generated results.

4.1.6. Summary

Figure 4.1 summarizes the derived requirements classifying into functional and structural requirements. Overall, the requirements definition resulted into 33 functional requirements and 3 structural requirements. The functional requirements consists of 12 requirements based on fashion and apparel SCs, 18 based on social media characteristics, and 3 derived from the field of TM. The 3 requirements derived as structural requirements are based on the specifics of process models. Based on these derived requirements the process model is designed in the following sections. The process model is designed around four main phases, namely, Process Layer, Information Source Layer, Social Media Layer and Text Mining Layer.

	Section 4.1.1	Section 4.1.2	Section 4.1.3	Section 4.1.4
	Fashion and Apparel Supply Chain Characteristics	Fashion and Apparel Supply Chain Stakeholders	Social Media Characteristics	Text Mining Characteristics
Functional Requirements	• Req_FASC 1.1: The process model should include the perspective of the decision-maker. • Req_FASC 1.2: The process model should include the objective of the decision-maker. • Req_FASC 1.3: The process model should provide methods for extracting information relevant to the stakeholders decision-process. • Req_FASC 1.4: The process model should include the different needs of the decision-maker. • Req_FASC 1.5: The process model should include the different stakeholders involved in a decision process.	• Req_M1: The sales representative shall obtain information relevant to product features. • Req_M2: The product developer shall obtain information relevant to product features. • Req_M3: The product developer shall assess believability of extracted information. • Req_R1: The buyer shall obtain information relevant to product features. • Req_R2: The buyer shall be able to assess the believability of extracted information. • Req_R3: The designer shall obtain information relevant to product features. • Req_R4: The designer shall obtain information on product features with respect to the time.	• Req_SM 1.1: The process model should process textual data. • Req_SM 1.2: The process model should deal with the processing of multimedia data. • Req_SM 1.3: The process model should deal with the processing of network data. • Req_SM 1.4: The process model should deal with the processing of metadata. • Req_SM 2.1: The process model should deal with the diversity of social media channels. • Req_SM 2.2: The process model should deal with the differences in the structure of the channels. • Req_SM 2.3: The process model should provide the decision-maker activities for identifying channels specific to the needs. • Req_SM 3.1: The process model should contain phases dedicated to social media analytics relevant activities. • Req_SM 3.2: The process model should provide methods for ensuring the extracted data are relevant to the SC stakeholders decision process. • Req_SM 3.3: The process model should contain different methods for extraction of social media data. • Req_SM 4.1: The process model should provide different methods for extraction of social media data. • Req_SM 5.1: The process model should deal with colloquial language use. • Req_SM 5.2: The process model should deal with intentional misspellings. • Req_SM 5.3: The process model should provide methods for exploiting the tonality of social media content. • Req_SM 5.4: The process model should deal with methods for dealing the existence of emoticons. • Req_SM 6.1: The process model should include information quality mechanisms. • Req_SM 7.1: The process model should include the perspective of the data provider. • Req_SM 7.2: The process model should contain a believability assessment mechanism.	• Req_TM 1.1: The process model should include requirements of a TM framework. • Req_TM 1.2: The process model should include methods required for processing textual data. • Req_TM 1.3: The process model should include the SC stakeholders perspective in evaluating generated results.
Structural Requirements	• Req_SR 1.1: The process model should be designed in compliance with typical components of a process model. • Req_SR 1.2: The process model should be designed based on a formalization. • Req_SR 1.3: The process model should involve feedback loops enabling a steady improvement of the generated results.			

Figure 4.1.: Summary of functional and structural requirements

4.2. Process Layer

With establishing the Process Layer, the perspective of the SC stakeholder is included in the social media exploitation process in an early stage. According to Marr (2015), companies incorporating big data initiatives require setting up a particular case to understand the business value that these big data strategies will bring. There may be a risk of falling into the buzz trap. Before the collection of large volumes of data, it is then crucial to define and develop a case by incorporating objectives, information needs and information sources. Consequently, it is claimed that the Process Layer holds a crucial relevance in the success of big data initiatives, regardless of the data that will be exploited. The Process Layer is led by three sub-questions: What do we need to know? What are the most important unanswered questions? Who needs to know what, when and why? With designing the Process Layer, the perspective of the SC stakeholder is considered and highlighted. These questions are elaborated into six steps that constitute the Process Layer. These steps are displayed in Figure 4.2. Please note, that this figure focuses on the Process Layer only, the complete process model is displayed in Figure 4.19.

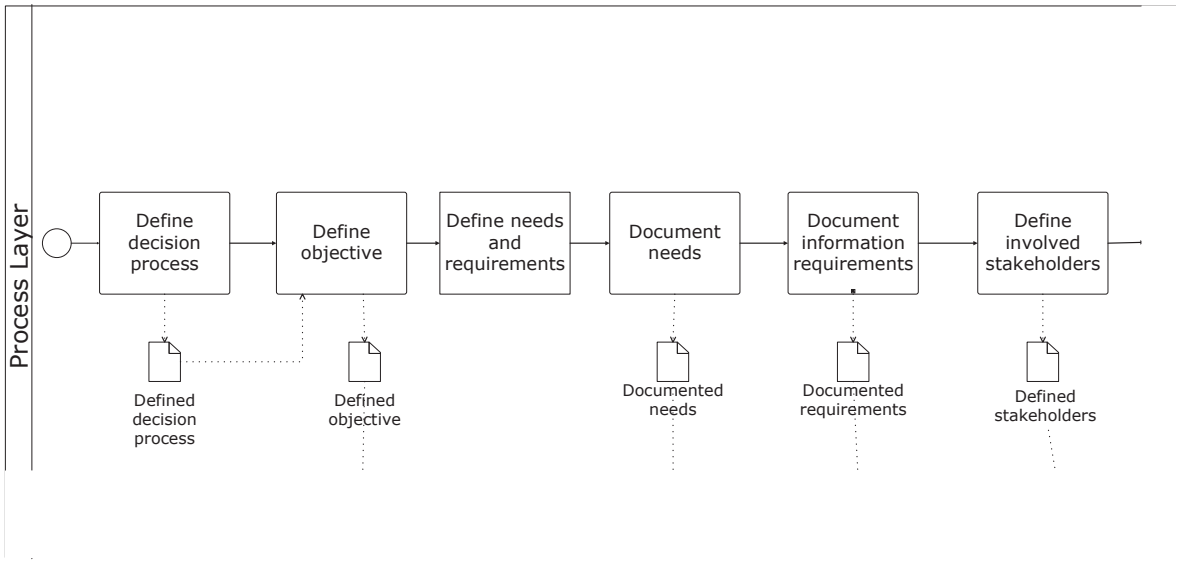


Figure 4.2.: Tasks -Process Layer-

The first step is highly dependent on the decision-maker and stakeholders engaged in the decision process and should not set out through BI applications. Marr (2010) pointed out that often organizations rely firstly on the available data and proceed with analysis without having identified the information needs. However, the focus should be on the identification of the information needs, and accordingly, the time invested in determining the required information might be carefully set out as a fundamental step before conducting further analyses (Marr, 2010).

The first task to be conducted is to define the decision process for which information is

required to be exploited. With an accurate definition of the decision process, the stakeholder is involved from the first step into the exploitation process of social media data. This is a first step for ensuring that the extracted information from social media is relevant to the decision process of the respective stakeholder.

Based on the defined decision process, the objective for the exploitation process is defined. Different approaches, such as the Balanced Scorecard scheme or strategic map for the identification of goals and according information needs, can be used for this definition process. This approach ensures that the data-driven approach is aligned with the overall strategic objectives of an organization. Once the decision process and the objective are defined, in the third step, the information needs related to these are defined.

In the fourth step, the information needs and requirements of the stakeholder are made accurate. Different tasks and roles prompt particular information needs. These needs refer to a state when an individual realizes that it is uncomfortable with its current state of knowledge (Case and Spink, 2012). According to Taylor, needs evolve from an unexpressed need into a conscious, a formalized and a compromised one (Case and Given, 2016; Taylor, 1968). The formalized need is the seeker's expression of the needed information, while the compromised need incorporates the constraints of the available sources (e.g. formats and query languages). We define a conscious need as an unexpressed understanding of the desired content, and a formalized need is an expressed description of the desired content. Concerning problem-solving, we consider general information needs about the problem situation, the domain of the problem and methods to come to solutions (Byström and Järvelin, 1995). In this way, conscious needs are turned into formalized needs, by, for instance, writing them down. The formalized needs support the decision-maker to distinguish relevant from irrelevant content. Besides, the formalized needs can be easily shared with other people involved in the business task or specific problem situation. Apart from the formalized needs, information requirements might be formalized. Their purpose is to assess the usefulness of relevant information. Information meeting the formalized need and the requirements can be assessed as useful. Table 4.1 lists the quality dimension accessibility, accuracy, believability, relevance and timeliness. Based on these, information requirements are derived. These requirements are used in the following steps by the stakeholder to assess if a piece of information meets the defined requirements or not.

Table 4.1.: Using IQ Dimensions for deriving requirements

IQ Dimensions	Derived Requirements
Accessibility	Information should be ideally accessible without large effort.
Accuracy	Information should be ideally free of error.
Believability	Information should ideally come from a credible source
Relevance	The relevance of information constitutes of the information needs.
Timeliness	Information are ideally timely available.

In the subsequent step, the SC decision-maker can choose between different data types. As elaborated in Section 4.1.3, social media owns a variety of data types. This task allows the decision-maker to choose from the richness of social media data. The data types should be selected based on the stakeholder's defined objective, information needs, and requirements. The decision-maker can choose between textual data, multimedia data, network data or metadata. Based on this selection in further steps, adequate processing methods can be applied.

In the last step of the Process Layer, the decision process's involved stakeholders have to be defined. As described in Section 3.1 fashion and apparel SCs are highly complex. In order to ensure the running of the chain, stakeholders are engaged in constant coordination processes. To this end, it is necessary to know what information is needed by which stakeholder. Thus, it is required to define the involved stakeholders together within their responsibilities and needs. The output of this step can be a list of engaged stakeholders with corresponding tasks (responsibilities). These steps define stakeholder, the responsibilities of the stakeholder and the decision-maker within the stakeholder's organization.

Having conducted all the steps of the Process Layer, the decision process, objective, needs and requirements, and involved stakeholders are defined. The process continues with the Information Source Layer that is introduced in the following section.

4.3. Information Source Layer

The main objective of the Information Source Layer is to outline, classify, and select relevant information sources. A prerequisite of the exploitation of data for supporting decisions is the availability of relevant information sources and relevant content. Relevant content refers to the predefined objectives and information needs within the Process Layer. In the context of a holistic approach to support fashion and apparel SC decisions by social media data, it is required to consider different types of data and information sources. For this reason, the output of the Information Source Layer is designed in modules, incorporating unstructured and structured data sources. Nevertheless, since in this work, we focus only on social media-

based sources, i.e. structured data sources, are not used in this research. The exploitation of traditional textual data sources is not targeted in this thesis either. Figure 4.3 illustrates the steps followed to construct the Information Source Layer. Please note, that this figure focuses on the Information Source Layer only, the complete process model is displayed in Figure 4.19. The steps of the Information Source Layer are described in the following sections.

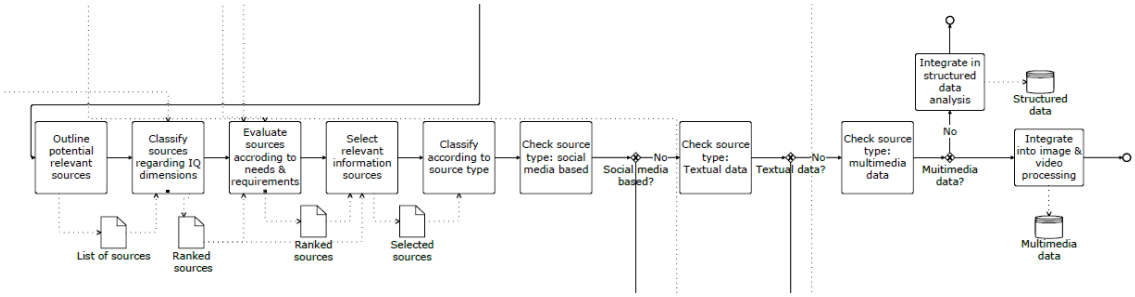


Figure 4.3.: Tasks -Information Source Layer-

4.3.1. Outline potential relevant sources

As a first step, it is required to outline the potential data sources which are usually addressed for the defined task. In this step, all potential sources relevant to the defined task should be outlined, without any filtering. Concerning the information needs or requirements, the main question answered here is what sources are usually considered for the defined decision process. The outlining of the potential sources is the starting point for the selection of the information sources, which is done in the last step.

4.3.2. Classify sources regarding information quality dimensions

After outlining the potential relevant sources, it is required to evaluate them against further criteria. This step is heavily based on existing information needs, requirements for existing problems. In this research, the factor perceived IQ is focused. The quality assessment is more critical supporting decisions by exploiting social media content, as for instance, for traditional information sources, such as manuals, newspapers or material provided by companies, the consideration of IQ turns evident. For this reason, we suggest applying IQ dimensions (see Section 3.2.4) for evaluating the outlined information sources. The Analytic Hierarchy Process (AHP) or utility analyses are potential methods for evaluating the sources with respect to the IQ dimensions. The input for this step is the outlined sources from the previous step. After conducting the second step, we should ideally have classified all the sources with respect to the IQ dimensions. It is worth mentioning that the assessment of the sources for each IQ dimension is done following the defined goal.

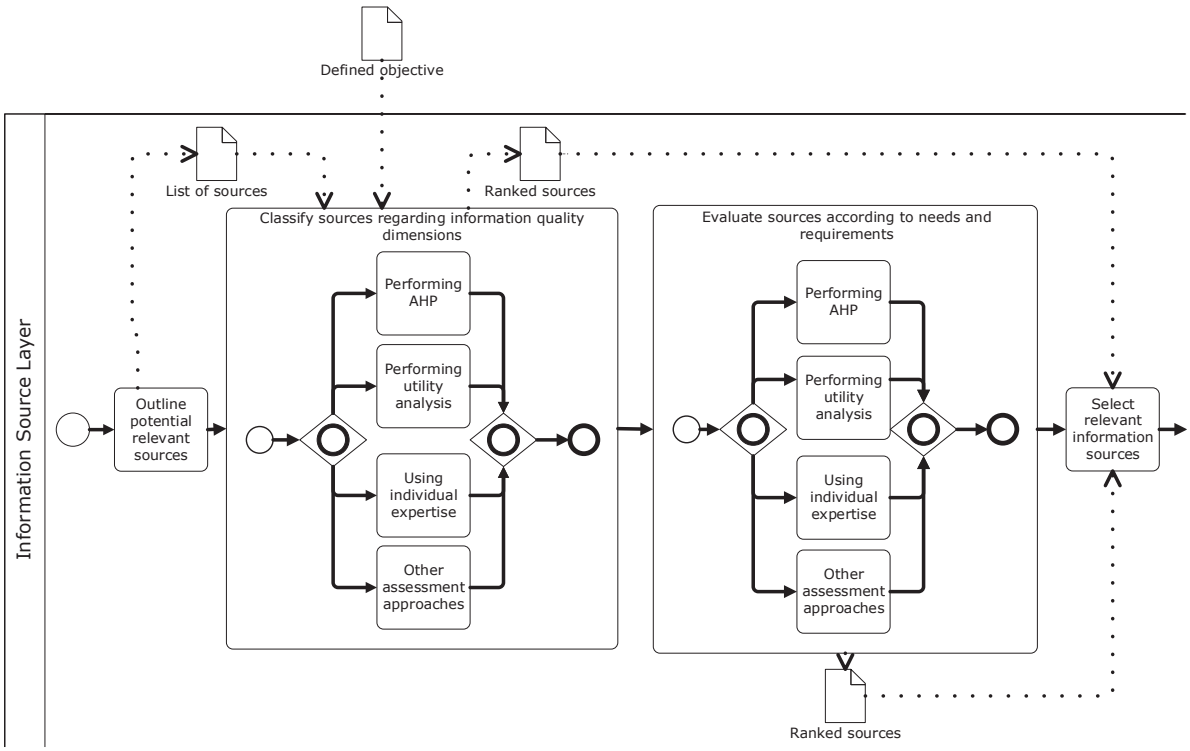


Figure 4.4.: Sub-processes- expanded- Classify sources regarding IQ dimensions and Evaluate Sources according to needs and requirements-

4.3.3. Evaluate Sources according to needs and requirements

After classifying sources based on the quality dimensions, information needs and requirements are acknowledged. The question to be replied to is what information needs are more likely to occur in what information sources. This question is derived from the assumption that if the needs and requirements are evaluated regarding the information sources in an early stage (before the actual collection of data), data can be collected in a more directed manner. Different methods can be applied for the evaluation of the sources based on the information needs and requirements. Besides, this evaluation can be done by not only using structured methods such as AHP methods but also by domain experts based on their individual experiences and knowledge in a more unshaped way.

4.3.4. Select relevant information sources and classify according to source type

The aim of this step is first to select the potentially relevant information sources and then classify them. Data needs proper processing depending on its type. To identify suitable processing, it is required to classify the selected information types according to their type. The information sources can be classified as social media-based sources, traditional text

sources or structured data sources. Based upon step 2 and step 3, i.e. the classification according to the IQ dimensions and the evaluation of the sources focussing on the needs and requirements, the relevant information sources are selected. The classification of the selected sources according to their types enables adequate exploitation of the data in proceeding steps. In this research, we will focus on the exploitation of social media-based sources, which is presented through the proceeding Social Media Layer.

4.4. Social Media Layer

In accordance with the research scope of this thesis, the Social Media Layer becomes highly relevant in the entire process model. The objective of the Social Media Layer is to enable suitable exploitation of social media data for supporting fashion and apparel decisions. As described in Section 3.2.1, social media data has certain characteristics which require appropriate handling. The overall frame and structure of such a Social Media Layer are derived from the SMA framework proposed by Stieglitz et al. (2014a). This framework includes tracking, preparation, and analysis. In the context of the Social Media Layer, these elements are translated into channel selection, data extraction, and availability check. Due to the fast-paced and dynamic nature of social media, current channels can dissolve, or further channels may rise. For this reason, in order to accommodate upcoming social media types, the Social Media Layer is designed more generically. Table 4.2 displays the main tasks constituting the Social Media Layer together with their main components, leading questions and methods. The steps are described in the following sections.

Table 4.2.: Social Media Layer- main components, leading questions and methods

Main tasks	Main components	Leading questions	Methods
Channel Selection	- Upper level channel selection - Medium level channel selection	Which social media channels are relevant in the context of the defined objective and information needs?	- Application of IQ dimensions
Data Extraction	- Examination of channel structure - Selection of information components - Setting of extraction strategy - Generation of sample corpus	Which data to extract and how to extract data from potential relevant social media sources?	- RSS feeds - Official API - Commercial Web Crawler - Individual Web Crawler - Individual applications
Availability Check	- Content Analysis (CA)	Does the extracted data contain relevant content in terms of the defined information needs?	- Web Content Analysis (WebCA)

The components and questions targeted by the Social Media Layer illustrated by Table 4.2 are converted into tasks and activities and formalised in a subsequent step (see Figure 4.5).

Please note, that this figure focuses on the Social Media Layer only, the complete process model is displayed in Figure 4.19.

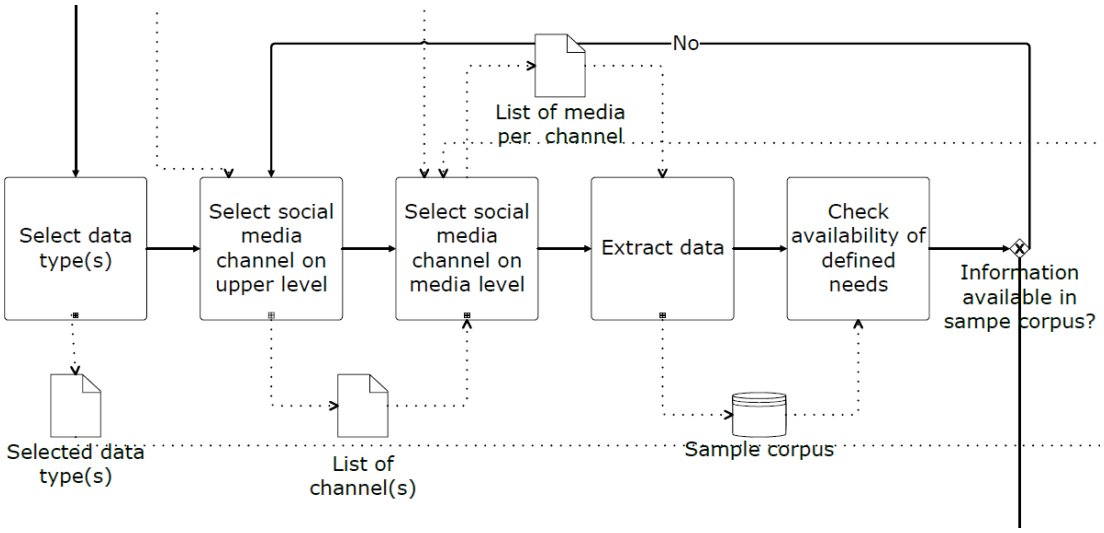


Figure 4.5.: Tasks collapsed -Social Media Layer-

4.4.1. Select data type(s)

In accordance with the requirements addressing the *Variety of data types in social media* (see Section 4.1.3) the first task to be conducted within the Social Media Layer is to select one or more data types for further processing. Figure 4.6 shows the expanded sub-process of the select data type(s) process.

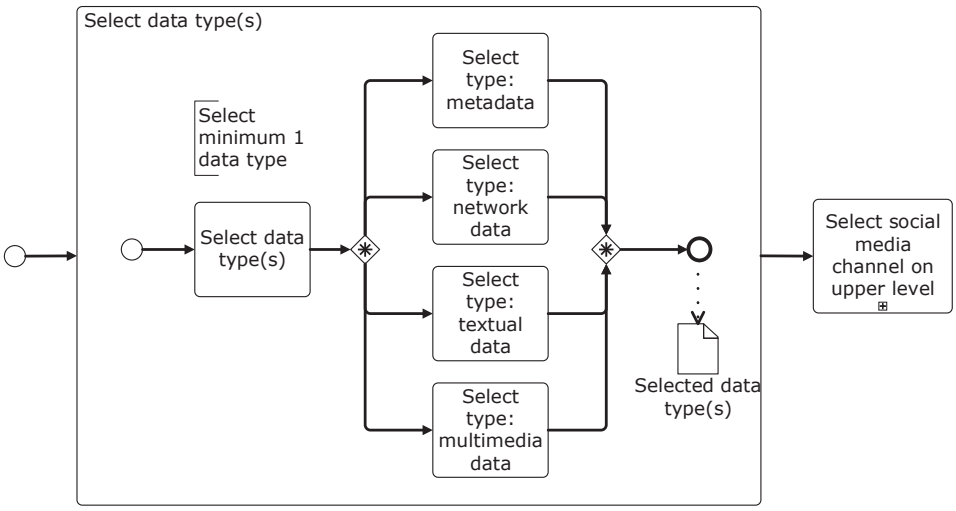


Figure 4.6.: Sub-processes expanded - select data type(s)-

4.4.2. Channel Selection

Channel selection includes the two processes of select social media channel on upper and on media level. The main question to be answered in this step is which social media channels

are relevant in terms of the predefined information needs and requirements. The needs and requirements defined in the Process Layer serve as input for the channel selection. The channel selection consists of two selection processes. In the first process, the upper level, an appropriate social media channel is selected. In the second phase, suitable media are selected within the defined channel. For instance, given that fashion blogs are considered as the main channel resulted from the upper level analysis, the selection on medium level defines which individual fashion blogs to include.

In order to be usable for fashion and apparel SC decisions, the social media data extracted requires careful assessment of the quality of published content. This is mainly true due to the illustrated characteristics of social media data. Therefore, the baseline for the channel selection is IQ dimensions (see Req_SM 6.1). Accordingly, the selection of channels at both the upper and media level is conducted by the application of IQ dimensions. The same IQ dimensions used in the Process Layer (see Table 4.1) are used for channel selection at both levels.

SELECT SOCIAL MEDIA CHANNEL ON UPPER LEVEL

The expaned sub-process of *select social media channel on upper level* is demonstrated by Figure 4.7.

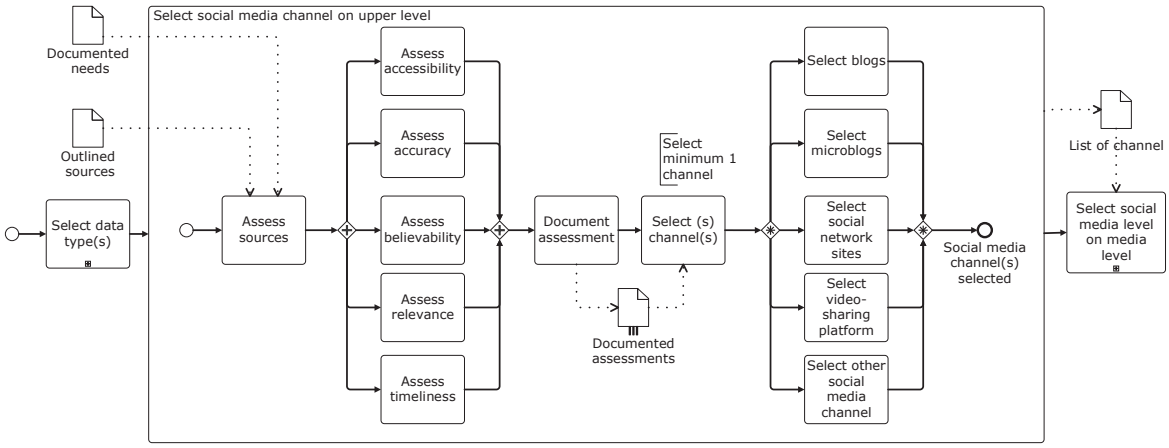


Figure 4.7.: Sub-process expanded-select social media channel on upper level-(Beheshti-Kashi and Kinra, 2020)

The following list states relevant questions and gives some guidance for using these dimensions for the channel selection on the upper level.

- Accessibility: How can the data be accessed? Which data can be accessed in terms of time, volume, and frequency? Does the channel provide an official API? Is the data accessible for the individual data consumer?
- Accuracy: Assessing the accuracy of a social media channel at the upper level is hardly possible since the content is not considered. For this reason, it is proposed to handle

the accuracy on this level more softly.

- Believability: As it is suggested that the believability of information be measured by assessing the source quality, it is difficult to assess whether a whole social media channel is believable or not believable. Accordingly, at the upper level, it is necessary to examine which options a social media channel generally provides; for example, to derive hints to then assess the identity, expertise and reputation of a particular social medium.

- Relevance: The relevance is a contextual dimension. One channel may be relevant for data consumer A with a certain need, while the same channel may not be relevant to data consumer B. In terms of the upper channel selection, the relevance of a social media channel can be elaborated by examining the general thematic availability of targeted information.

- Timeliness: Availability of a timestamp is crucial for assessing the timeliness and the relevance of the information in the context of fashion and apparel SC decisions. This is particularly true due to the high seasonality of fashion products. In the upper level channel selection, the given social media channels are accordingly scrutinized in terms of the availability of a timestamp.

SELECT SOCIAL MEDIA CHANNEL ON MEDIA LEVEL

After the selection of suitable social media channels at the upper level, the media has to be chosen on each selected channel. As each social media channel has its own functionalities and structures, this selection can be related to accounts, websites, hashtags and influences. The media selection approach is illustrated in Figure 4.8. It includes two main stages. While in the first stage, sub-categories for the quality dimensions are defined, in the second, measurable features are determined. These two stages are required since the quality dimensions are not directly measurable.

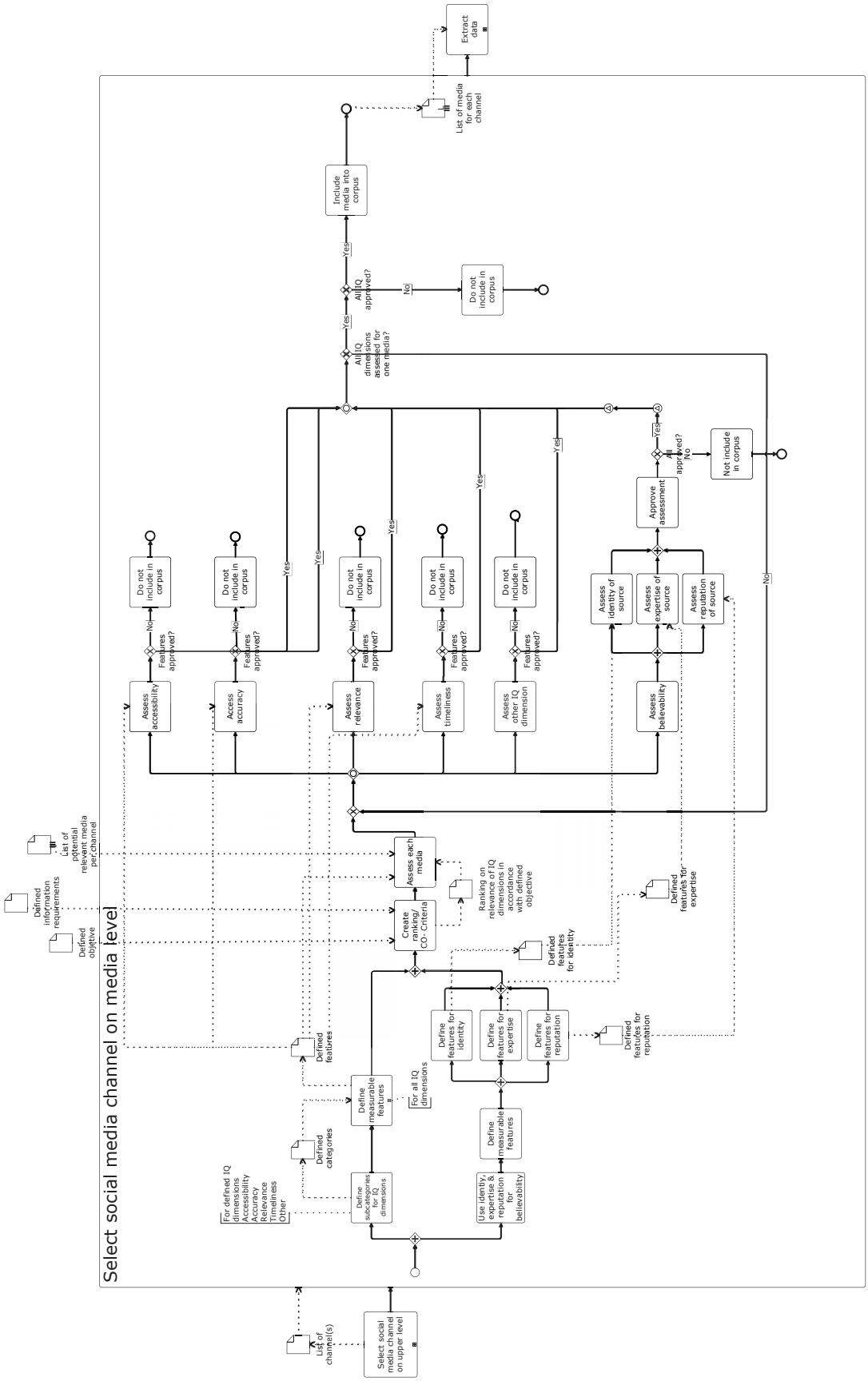


Figure 4.8.: Sub-process expanded -select social media channel on media level- (Beheshti-Kashi and Kinra, 2020)

Assessing the accessibility, accuracy, relevance and timeliness at the mediua level can be handled in a more straightforward way compared to the believability assessment as believability is itself a multidimensional construct. As described in Section 3.2.4 source credibility is the key factor in believability assessment. Shankaranarayanan et al. (2012) propose a model in which the three constructs of identity, expertise and reputation are used for assessing believability (see Figure 3.14). We have adopted this model in the context of this research (see Figure 4.9). In contrast to Shankaranarayanan et al. (2012), the data consumer's domain expertise in this work is already incorporated into the assessment of the identity, expertise, and reputation. One approach for assessing the identity of the data provider is to search for available information on the data provider or check other social media profiles. For assessing expertise, the data consumer can examine other posts or comments on specific topics or search for references showing the expertise of the data provider. As reputation construct considers the potential influence of the data provider within their community, the main question to follow is if the data provider has influence on its community. This can be examined quantitatively, e.g. number of followers or qualitatively, e.g. quality of posts. Following these steps, the data consumer can assess the data provider in terms of credibility. If the data provider (source) is considered credible, then the believability of data is also met.

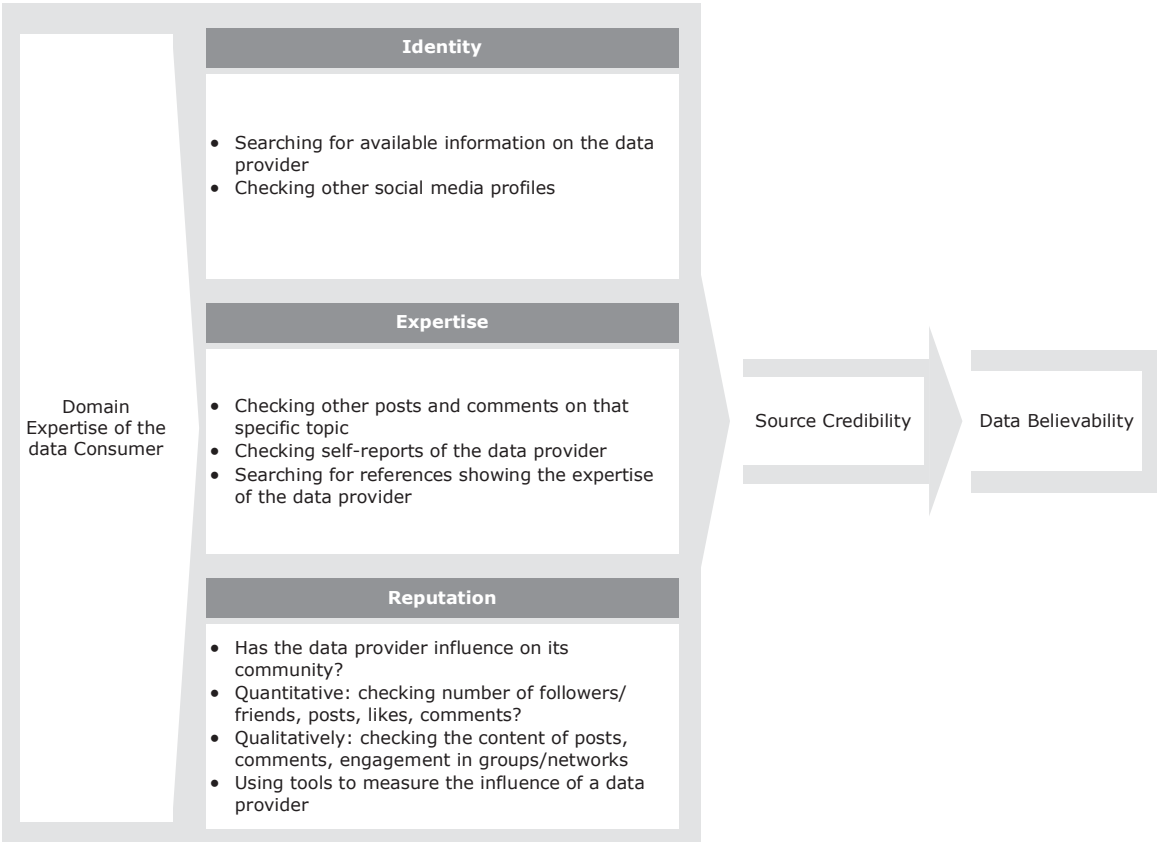


Figure 4.9.: Believability assessment based on (Shankaranarayanan et al., 2012)

The outcome of the channel selection is thus twofold. The channel selection at the upper level gives the decision-maker a list of relevant social media channels, whereas, on the media

level, a list of media inside each of the selected social media channels is provided.

4.4.3. Extract Data

There are two objectives for the data extraction task. The first is to enable a precise examination of both the selected channels in terms of structural elements and information components. The second is to generate a sample corpus, which will be further examined based on the content within the Availability Check. The extract data processes are based on the proposition of targeted exploitation of social media content and compromise four main processes. The main tasks, components and questions are summarized in Table 4.3.

Table 4.3.: Data Extraction- main tasks, components and leading questions

Main tasks	Components	Questions
Examining Structure	Structural examination of the - channel - media - unit	How is the structure of the selected - channel(s) - media - unit(s)?
Selecting Relevant Information Components	Matching identified information components to IQ dimensions and derive requirements	Which information components are relevant?
Setting Extraction Strategy	Elaborating different extraction strategies Selection most adequate extraction strategy	Which extraction strategy is most suitable?
Generating Sample Corpus	Using selected information components and extraction strategy for generating a sample corpus	Does the identified channel(s) include the required information?

The data extract sub-process is displayed in Figure 4.10. While the input for the data extraction, which is a list of selected channels and selected media, is delivered through the channel selection, the output is a sample corpus. It serves as input for the Availability Check. The four phases of the data extraction step are explained in the following.

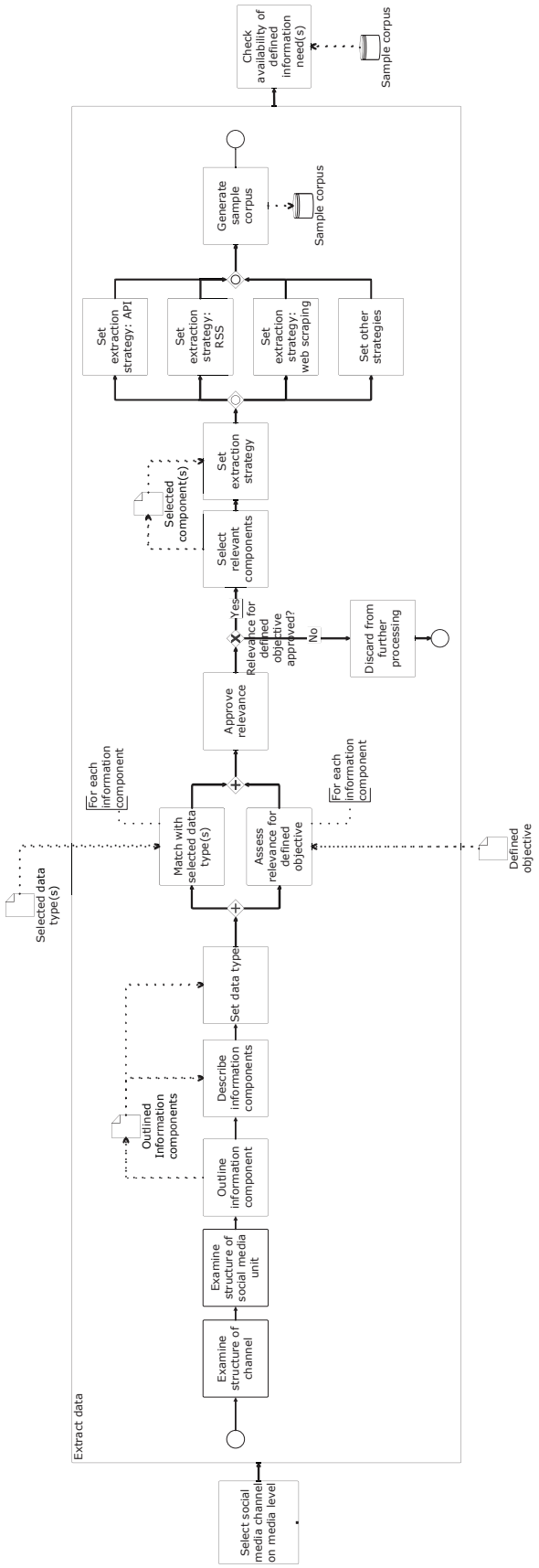


Figure 4.10.: Sub-process expanded-extract data-

Examine structure of the channel

The objective of the structural examination of the selected channels is to identify structural elements of the respective information sources. This task works as the preparation phase for the selection of relevant information components within the selected sources. In accordance with the overall objective of the Information Source Layer that aims at efficient exploitation of the selected sources, the structural examination enables the targeted exploitation of the relevant information components. The basis for the structural examination of the selected channels is the output of the channel selection, in which all potential relevant channels are listed. A detailed examination of the structure of the channels enables a targeted and efficient collection and generation of the corpus. As different social media channels are organised differently, it is required to examine each of the channels individually. Since these channels have different ways of displaying the information published by the user, the structural examination is not sufficient only on the upper channel level.

The structural examination includes three main tasks. First, the general structure of the selected channel has to be examined. For this purpose, the concept of wireframing can be used reversely. Wireframes are often used for layouting websites. Accordingly, it can be used to bring out the structure of the targeted object (channel, media, social media unit). In the second stage, the media is targeted. However, the structural examination at the media level may be omitted in some cases because the corresponding channel already provides the structure. For instance, Facebook pages (one Facebook page is media of the social media channel Facebook) have similar structural elements. For blogs, the structure of the different media may be different. In the third and last task, the social media unit is considered. A document may vary in different social media services. For instance, a document on Twitter is a Tweet, a document on Facebook is a post, and in blogs, a document is a complete post from the author. The structural examination includes, for instance, the consideration of the potential length of the messages. This is relevant in terms of storage, processing time, and the working environment. While Twitter, for instance, has limited the number of characters for a long time to 140, it changed to 280 characters in 2017. Facebook messages are potentially possibly 20,000 for private messages and 63,206 characters. At the end of the structural examination, for each selected channel, media type and social media unit, the structures are defined.

Select the relevant information components

The objective of this task is to identify and select the relevant information components. The main question to answer in this phase, is what pieces of information should be collected with regard to the information needs. Selecting the relevant information components based on the information needs and requirements enables targeted extraction of the social media content. In order to do so, it is necessary to refer to the previously defined information needs and

requirements in Table 4.1. The basis for selecting the relevant information components is the structural examination of the channel types and social media units. From the IQ dimensions, requirements can be derived. The defined objectives may also be included in this phase. For instance, if in the requirements, it is stated that the information need is critical in terms of time, the timestamp which is published along with the message has to be selected to be collected and included in the generated corpus. In the case that the message does not provide a time, alternative dates can be used, which matches the timestamp as closest as possible. For example, in the case of the extraction of blog posts, Beheshti-Kashi et al. (2015b) propose to extract the date of the latest user comment along with the blog post in order to have at least an approximate time reference. Having stored the time stamp along with the message, further analysis options will be open. One of them is Trend Analysis which enables the tracking of specific topics in a set of documents (Chakraborty et al., 2014). If one requirement refers to location-based data, then it is needed to select location base data for the collection. Figure 4.11 serves as a guideline and template for selecting the relevant information components. First, the components are listed and described. Then the relevance of the components for the objective and the corresponding data types set are defined. Elaborating on the guideline points gives a better understanding of the selection of the relevant information components. The completed template is the output of this phase and is a structured preparation for setting the extraction step.

Information Component	Description	Relevance for Defined Objective	Data Type	Selection
Component 1 Component 2 Component 3 ...				x x

Figure 4.11.: Template for the Selection Process of the Relevant Information Components

Set Extraction Strategy

After selecting the required information components, the extraction strategy has to be specified. The objective of this task is to define and set the most suitable extraction alternative for the given information needs and requirements and defined channels. The main question in this step is which extraction strategy is most suitable. The input for setting the extraction strategy is the selected relevant information components. Gathering different extraction strategies and setting the most suitable can save time and resources. Several options exist for acquiring data from social media channels, which depend on the channel(s) themselves. As described in Section 3.2.2, the following methods are applicable to accessing and extracting data from social media: Web scraping, API provided by social media channel, RSS feeds, Search filter, Real-time streaming. The advantages, disadvantages and further details of the accessing methods are summarized in Table 3.2. It is important to highlight that the listed

approaches are the most common during the time of this thesis. Due to the dynamic environment of social media, other methods may have emerged or will be appeared in future. In any case, the following recommendations should be taken into account when setting the extraction strategy and for securing a smooth corpus generation. First, it is vital to outline and consider several different strategies so as to operate independently. Second, if the utilized infrastructure for accessing the data is not developed internally, it is crucial to examine precisely the terms and conditions of the infrastructure used. This may apply to used RSS feeds, APIs provided by social media channels or commercial scraping tools. Apart from the initial examination, regular monitoring of the terms and conditions or for commercial tools, the pricing models is recommended since these may change over time. Social media companies often change their terms and conditions without prior notice and a high media echo. For this reason, it is possible that in half of the project time, not all of the information which was previously accessible and included as a relevant information component will no longer be available. Apart from the terms and conditions, the provided functionalities for accessing the data may vary over time. Moreover, it is recommendable to use one extraction strategy for all relevant information components. At this stage, all required stages for preparing the corpus creation are completed. In this way, the next step focusses on the corpus generation.

GENERATE SAMPLE CORPUS

After deciding how to collect the data and which parts of the respective sources (documents) are relevant and should be included in the corpus, the data collection needs to be conducted. A further aspect that has to be considered is the time relevant for the respective problem to be solved. Questions as to the reliability and validity of a social media corpus are addressed in this phase. This last phase of the data extraction step ends with loading in the data an adequate working environment.

4.4.4. CHECK AVAILABILITY OF DEFINED INFORMATION NEEDS

The objective of the availability check step is to examine the availability of relevant content. Relevant content in this context is related to the defined information needs. Does the identified information source contain relevant content which will meet the information needs? The input for the data availability is the generated data sample.

Once the extracted data is stored in a suitable working environment, it is proposed to conduct CA on the corpus. Although CA is a method originating from communication sciences, it is used in a range of social science research. Human coders could be incorporated for this task. However, manual CA is highly time and resource-intensive. Krippendorff (1980) suggest a five-step procedure for conducting CA. His work and follow-up works are usually addressed with traditional CA, but they might find their limitations when working with internet data (Herring, 2010). Herring (2010) introduces the approach of WebCA, in which she

outlines the different disciplines where CA and similar concepts are used (social Network analysis and discourse analysis) and how they might be applied to the web. Figure 4.12 displays the WebCA paradigm, including the different analysis methods which may be followed in reliance on the objective and source of analysis.

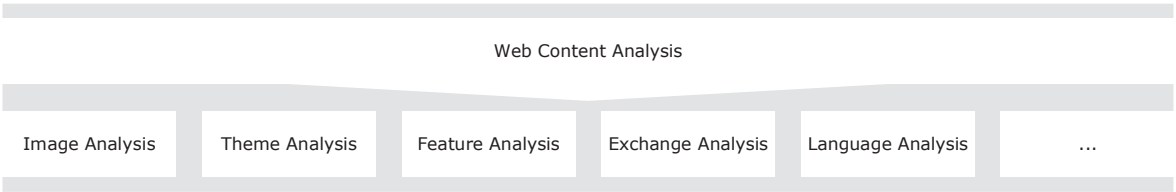


Figure 4.12.: WebCA: An expanded paradigm Herring (2010))

In the last step of the Social Media Layer, the suggestion is to follow theme analysis for examining whether the information needs relevant for the decision-maker and the targeted task are available in the generated corpus. The theme analysis is understood as a first step in approaching the content. Therefore, this step is considered a more qualitative step in the Social Media Layer. The quantitative analysis of the extracted data is suggested to be followed in the Text Mining Layer.

4.5. Text Mining Layer

The objective of the Text Mining Layer is the processing of textual social media data. As textual data requires proper processing for the purpose of applying advanced mining methods as stated in the requirements (see 4.1), the process model includes a Text Mining Layer as the last step for exploiting social media data for fashion and SC decisions.

As presented in Section 2.2 literature describes various process models or frameworks of Text Mining. Most of these models are either custom-designed for the elaborated case or highly generic, illustrating the general function/process of Text Mining (Schieber and Hilbert, 2014a). In contrast, Schieber and Hilbert (2014a) propose a generic process model for Text Mining purposes. The model consists of six phases, namely, task definition, document selection and analysis, document preparation or document processing, application of TM methods, evaluation of the results, derivation of actions, and application of results. In this research, there are two main reasons to follow this model for the elaboration of the Text Mining Layer. First, the model is application-neutral and includes details and the main subprocesses and methods. Second, it is developed by considering an extensive number of text mining process models from the literature. However, since this research focusses on the exploitation of social media data for fashion and apparel SC decisions, it is particularly relevant for the last step of the TM process to consider the perspective of TM in organizational environments. The organizational perspective is not focussed on the generic process model of Schieber and Hilbert (2014a). At this point, the text mining process suggested by Kobayashi

et al. (2018) comes into play because it is designed from an organizational perspective. The authors consider the TM process as a three step process. Starting with the preprocessing as a first step, including text collection, text cleaning and text transformation. The preprocessing is followed by the operation of TM techniques (such as classification, clustering topic modelling etc.). In Figure 4.13 both process models are contrasted in order to illustrate the different phases and steps.

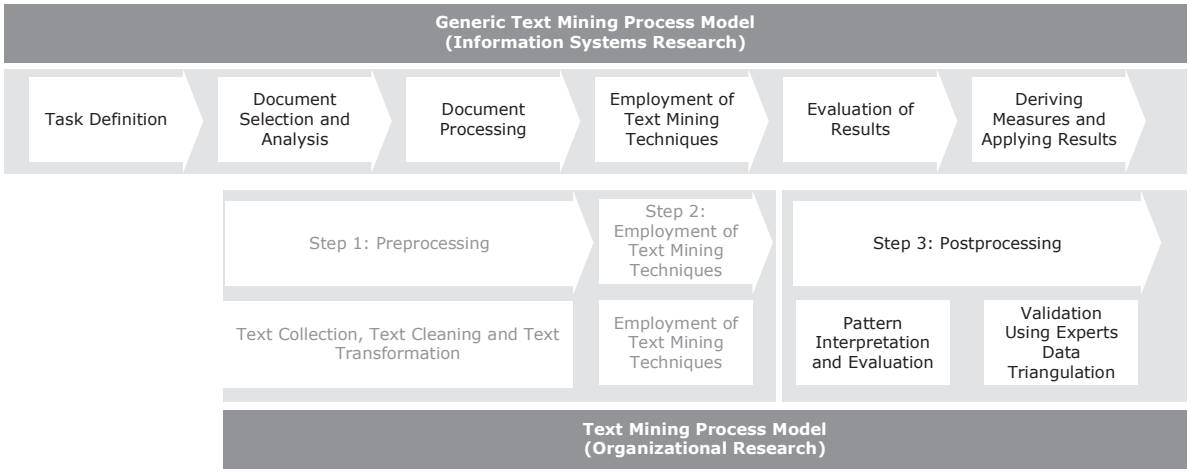


Figure 4.13.: Comparison of the Text Mining Process Models of Schieber and Hilbert (2014a) and Kobayashi et al. (2018)

A simplified version of the process model of Schieber and Hilbert (2014a), showing the main phases without the internal connections and feedback loops, is displayed in the upper part of the figure. The model proposed by Kobayashi et al. (2018) is shown in the lower part. While in (Schieber and Hilbert, 2014a) each Text Mining process starts with a Task definition, Kobayashi et al. (2018) do not include the task definition process as a part of the Text Mining process. Their first step is the *Preprocessing* phase consisting of *Text Collection, Text Cleaning and Text Transformation*. Also, Schieber and Hilbert (2014a) include collection, cleaning and transformation steps; however, with a different naming and aggregation level. For example, Kobayashi et al's step *Text Collection* is included in the second phase of the process model of Schieber and Hilbert (2014a) that is *Document Selection and Analysis of Source Documents*. Both models propose the operation of text mining techniques. As the last phase, Kobayashi et al. (2018) suggest postprocessing including pattern interpretation and evaluation, and validation. In contrast, Schieber and Hilbert (2014a) consider the evaluation of the results and the derivation of the measures and the application of the results as the two last phases.

Considering the different models and comparing them, it is decided to use both for designing the Text Mining Layer. The generic process model proposed in Schieber and Hilbert (2014a) is selected as the backbone for this process model due to its generic approach and detail level regarding the activities and procedures. However, the organizational perspective is lacking. For this reason, the model by Kobayashi et al. (2018) is used to supplement it.

It is the postprocessing phase from Kobayashi et al. (2018) that is included as the last phase of our Text Mining Layer because it includes the use of experts and additional data for validation purposes. We assume that combining phases and steps of both models establishes the validity of the TM output. This fact is crucial to the context of supporting fashion and apparel SC decisions. Based on these two frameworks, the processes of the Text Mining Layer are presented in Figure 4.14. Please note, that this figure focuses on the Text Mining Layer only, the complete process model is displayed in Figure 4.19.

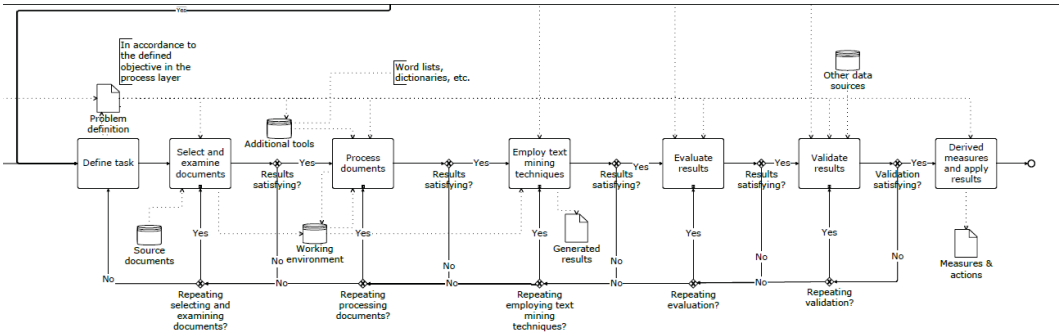


Figure 4.14.: Tasks collapsed -Text Mining Layer-

Figure 4.15 illustrates the expanded Text Mining Layer, and the expanded sub-processes for *perform linguistic processing, perform technical processing* and *perform social media processing*. The following sections give a detailed overview of each task and sub-processes.

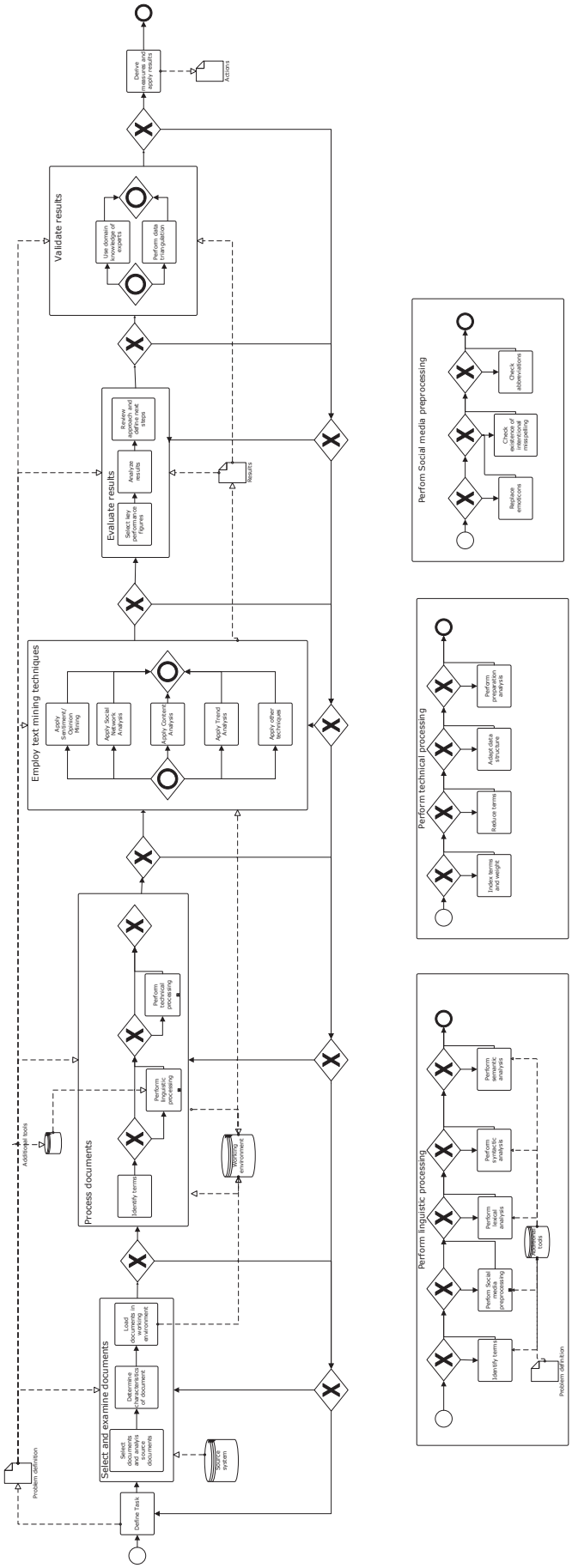


Figure 4.15.: Tasks expanded -Text Mining Layer-based on Kobayashi et al. (2018); Schieber and Hilbert (2014a)

4.5.1. Define Task

According to Hotho et al. (2005), as a first step, it is crucial to understand the problem to be solved in the first phase. In the CRISP - DM, the first step is termed business understanding. Hotho et al. (2005) define the grouping of similar documents in a document selection as one typical business understanding goal in the context of text mining. The first step is crucial since it also impacts all the following phases. Since different aspects are focused on the analysis, we have to formulate different goals and corresponding pipelines for achieving them. In the following, the goals are formulated in the first step. After they have been defined, the actual pipelines/procedures are displayed in order to illustrate how these goals are going to be achieved. It is noted that the preprocessing of the corpus is similar for all of the goals.

Within a TM project, different aspects may be focused on. Figure 4.16 illustrates a decision tree which structures the selection of appropriate TM approaches. This selection is lead by the level of the research target: is the interest in the results about words or on a higher level such as sentence, paragraph or document? According to the selected level, a range of further questions needs to be answered to find the adequate practice area of Text Mining.

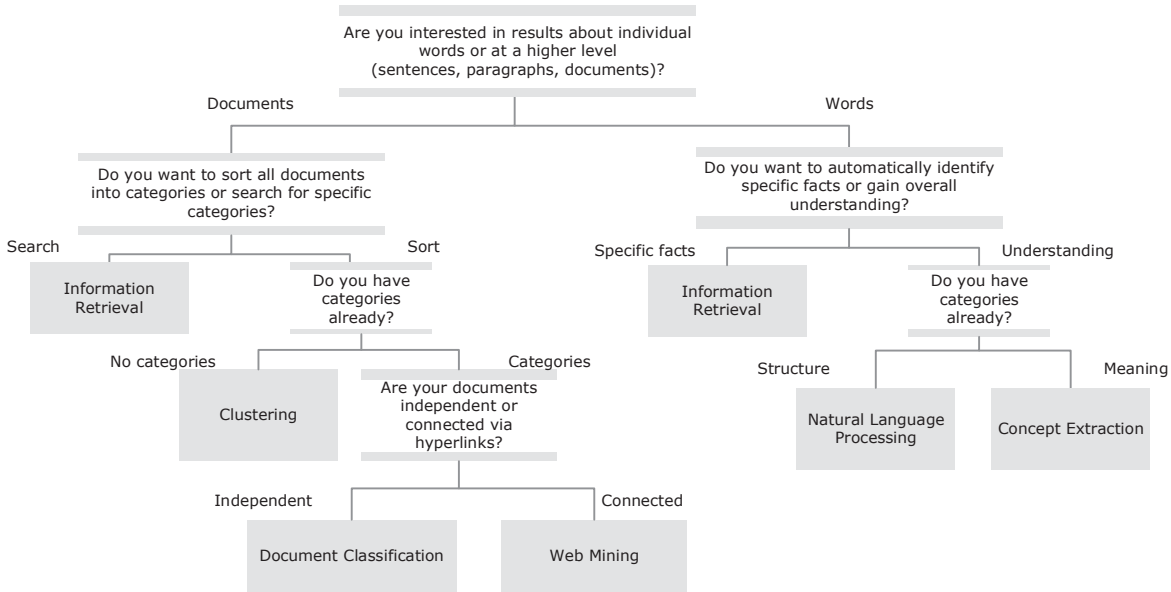


Figure 4.16.: Decision Tree for selecting appropriate TM methods (Miner et al., 2012)

4.5.2. Select documents and examine source documents

The second main process of the Text Mining Layer is the *select and examine documents* process. This process is performed in accordance with the defined task. It consists of three tasks. The first task is to select and examine the source documents in their structure. This task goes along with the select social media channel on media level in the Social Media Layer. The second task is referred to the actual collection of the content. This is based on

the extraction strategy set in the data extraction process of the Social Media Layer. The last task of this sub-process is loading in the working environment.

4.5.3. Process documents

Due to the unstructured nature of the textual documents, the document processing phase is a relevant phase in order to be able to apply classical DM methods. The sub-processes of the document processing phase are depicted in Figure 4.17.

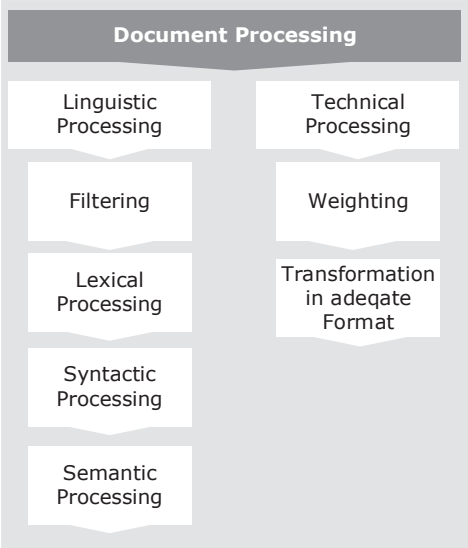


Figure 4.17.: Document Processing Steps based on Schieber and Hilbert (2014a)

Linguistic processing includes filtering, lexical, syntactic and semantic processing. They provide a repertoire of methods that are followed to achieve specific objectives. Table 4.4 summarizes the objective of the sub-processes of the linguistic processing with the main tools and methods for achieving these objectives.

Table 4.4.: Sub-processes of the linguistic processing

Sub-process	Objective of sub-process	Main methods	Description
Word Filtering	- Filtering irrelevant words	- Stop word lists	- Filtering so-called stopwords
Lexical Processing	- Reduction of terms to original forms	- Lemmatization - Stemming	- Reduction to base form - Reduction to word stem
Syntactic Processing	- Targeted Extraction	- Part-of-Speech (POS) Tagging - Parsing	- Assignment of POS to all terms - Identification of the syntax and the word functions of a sentence
Semantic Processing	- Word meaning and relationships within the sentences	- Calculation of collocations - Thesauri and ontologies	- Collocations are calculated identifying word combinations that occur often together - Usage of word databases such as WordNet (English), GermaNet (German)

For example, in the case of word filtering, so-called stop word lists that include articles or prepositions are used in order to filter irrelevant words and reduce the number of terms in a given text corpus. Within the process of stop word filtering, terms such as articles, prepositions or conjunctions are filtered out since they are mostly irrelevant for further analysis. The term identification process, which is often named tokenisation (Webster and Kit, 1992), aims at splitting the free text into terms. For this purpose, all punctuation is erased; paragraphs are segmented into sentences and sentences into words.

In the lexical analysis, the remaining terms have to be analysed regarding their morphology. In order to do so, it is crucial to reduce all the terms to their original forms (Schieber and Hilbert, 2014a). This process can be conducted via two different methods: Lemmatisation and Stemming. In the process of lemmatisation, the terms are processed to their simple forms, namely for nouns into the first case and for verbs into the infinitive (Bird et al., 2009). Stemming is not the process of returning to the word origin, but rather through the reduction of the term itself (Bird et al., 2009). For the English language, the Porter Stemmer is often applied (Porter, 1980). Both methods can be used, though, depending on the analysis objective. The advantage of applying either lemmatisation or stemming lies in the reduction of terms, which will often result in performance advantages. Language-dependent methods are summarized in Table 4.5.

Table 4.5.: Language dependent processes

Process	Purpose	Tool kit
Removal of Stop words	Removal of irrelevant words	Stop word lists with individual words
Stemming	Reducing words	Rules
Lemmatization	Reducing words	Rules
POS-Tagging	Assigning POS for a targeted extraction of tokens	Dictionary, Morphological Rules
Sentiment Analysis	Catching tonality of a given document	Word lists

The syntactic analysis follows the lexical analysis. During this process, two main sub-processes are conducted: Part-of-Speech (POS) tagging and Parsing. Within the POS-tagging all terms have to be assigned to their corresponding POS (Bird et al., 2009). This is performed with the help of lexicons and probabilistic models (Bahl and Mercer, 1976). The advantage of this process is that depending on the problem, the terms may be directly targeted. For the German language, the TreeTagger is used, which is based on a decision tree (Schieber and Hilbert, 2014a). For the English language, the Brill algorithm is usually applied (Bird et al., 2009). The Parsing is focussed on the identification of the syntax and the word functions of a sentence. Therefore, at the end of this process, information on the subject or object of a sentence become evident.

The objective of the semantic process is to find the word meaning and relationships within the sentences. For this purpose, two different approaches are commonly applied: calculation of collocations and the use of thesauri and ontologies. In the case of collocations, word combinations that often occur together are calculated. The lexical database WordNet is often used for English. GermaNet is the equivalent of German. In addition to these databases, case-specific word databases may also be required.

Technical Processing

After the linguistic processing, technical processing is conducted which follows the objectives of weighting the terms and transforming the data into a structure supported by the TM methods/classical DM methods applied in a further step (Schieber and Hilbert, 2014a). For the term weighting procedure, measurements metrics are required. These metrics can be classified in simple such as absolute, relative or document frequency, and weighted metrics. Table 4.6 summarizes the metrics.

Table 4.6.: Metrics for weighting the relevance of the terms/tokens based on Schieber and Hilbert (2014a)

Weighting Scheme	**Type (weighted/simple)**	**Description**
Binary	Simple	Occurrence of word - occurring vs not occurring
Term Frequency	Simple	Measures the occurrence of one term in one document
Document Frequency	Simple	Measures in how many documents a term occurs
TF*IDF	Weighted	Gives a high value for terms that occur often in some documents and low values for terms that occur in almost all documents
TF*ICF	Weighted	It does not consider the whole corpus but do the weighting for a certain document category

According to Schieber and Hilbert (2014a), in addition to the described metrics, it is also useful to introduce domain-specific weighting measures such as in Li and Liu (2012) or Beheshti-Kashi et al. (2015b) in particular if the existing metrics do not deliver satisfactory accuracy. The term weighting process provides additional advantages. One advantage is that the number of terms will be reduced by filtering the irrelevant words from the relevant words. Furthermore, the term weighting increases the accuracy of the results while conducting TM methods (Schieber and Hilbert, 2014a).

The second objective of the technical processing is to transform the data into a format suitable for the application of TM/DM operations. The common data structure is the vector space model introduced by Salton et al. (1975). The different documents will be transformed into a vector and joined into a matrix. The organization of the matrix is as follows: the terms will be organized in columns, and the documents will be put into rows (Figure 4.18, Output 3).

Without the technical processing, the application of TM methods or classical DM methods is not possible. After transforming the text into a numerical format, the actual application of classical DM algorithms such as classification or clustering follows. Once these algorithms have been applied, the results will be evaluated, and the phase actions are derived. In order to obtain a better understanding of the actual output of the different sub-processes, Figure 4.18 shows a typical output of the tokenization, linguistic (POS tagging) and technical processes.

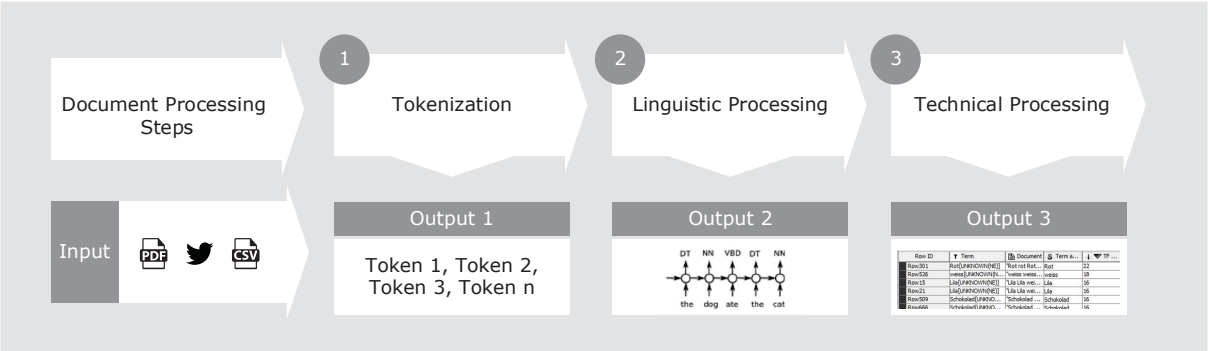


Figure 4.18.: Document Processing Steps Output

4.5.4. Employ text mining techniques

After the technical processing step, the unstructured data is transformed into a suitable format for applying Text Mining techniques. Table 4.7 gives an overview of the seven practise areas of Text Mining with their corresponding approaches and commonly used algorithms. As stated in the task definition, selecting the equivalent practise area depends on the main research objective and interest.

Table 4.7.: A summary of potential TM techniques based on Miner et al. (2012)

Practise Area	Description	Topics	Common Algorithms
- Search and information retrieval	- Storage and retrieval of text documents	- Keyword search - Inverted index	- Van Rijsbergen algorithm (Van, Rijsbergen, 1971)
- Document clustering	- Grouping terms, snippets, paragraphs, or documents	- Document clustering - Document similarity	- k-means Singular value decomposition
- Document classification	- Grouping/categorizing snippets, paragraphs, or documents based on models trained on labelled examples	- Feature selection - Sentiment analysis - Dimensionality reduction - eDiscover	- Naïve Bayes, Logistic regression, Decision trees - Neural network, Support vector machines, MARSplines - Regression, k-nearest neighbors
- Web mining	- Data and TMon the Internet	- Sentiment analysis - Web crawling - Link analytics	- Page Rank Algorithm (Brin and Page, 1998) - Hyperlink-Induced Topic Search (HITS) (Kleinberg, 1999)
- Information extraction	- Identification/extraction of relevant facts and relationships from unstructured text	- Entity extraction - Link extraction	- Conditional random fields - Hidden Markov models
- Natural language processing (NLP)	- Low-level language processing and understanding tasks	- POS tagging - Tokenization - Question answering	- Naïve Bayes - Hidden Markov models
- Concept extraction	- Grouping of words and phrases into semantically similar groups	- Topic modeling - Synonym identification	- Link Analysis - Word clustering - Latent Dirichlet Allocation (LDA) (Blei, Ng and Jordan, 2003)

4.5.5. Evaluate results

The purpose of this step is to evaluate the results regarding the defined tasks and the quality of the results. For this, key performance figures matching the tasks and the selected Text Mining methods are selected, for example, for text classification accuracy, precision, recall or the f-measure are used (Schieber and Hilbert, 2014a). For Text Similarity, Cosine similarity or the Jaccard Coefficient may be applied (Subhashini and Kumar, 2010). The results can be evaluated based on these figures.

Kobayashi et al. (2018) summarize the steps required after the application of the TM operations postprocessing and include pattern interpretation and evaluation as well as validation of the TM results by applying data triangulation or using experts. The integration of domain experts is crucial since it has to make sure that the extracted patterns convey real patterns and not only random occurrences which may be generated due to the volume of the data. For this purpose, the Bonferroni's principle may be applied (Kobayashi et al., 2018). Furthermore, it is necessary to create reliability, validity (e.g., content, predictive, and discriminant validity) and credibility of the output of TM models. It is not the TM procedures that are validated but the generated output. Kobayashi et al. (2018) outline that before using the output generated using TM, the validity of the findings needs to be established. For this purpose, the authors propose using comparative evaluation to compare the TM output with other data sources such as survey or expert data. Furthermore, they emphasize the importance of involving domain experts.

4.5.6. Validate results

In the sub-process validate results, the stakeholders are asked to validate the results generated by the Text Mining Layer. For this purpose, it is suggested that SC stakeholders make use of their individual expertise and domain knowledge. Furthermore, following Kobayashi et al. (2018), data triangulation processes should be applied. The validation process considers the defined task at the beginning of the Text Mining Layer. Moreover, the involved stakeholders defined in the Process Layer, are asked to conduct the validation as they have been determined to be involved in the targeted decision process (see Figure 4.2).

4.5.7. Derive measures and apply results

In the last step, the generated TM results need to be translated into specific actions. This may include the formulation of hypothesis, trends or recommendation for actions. Some authors also refer to the storage of the generated knowledge for future decisions (Schieber and Hilbert, 2014a).

4.6. SUMMARY

Chapter 4 constitutes the main contribution of this research. Following the DSRM process model, the design & development phase is performed. Before the actual development, it is necessary to derive requirements upon the process model that can be developed. The foundations for the requirements are laid by elaborating on characteristics of fashion and apparel SCs, social media and text mining process models. Based on the derived requirements, the process model consists of four layers: process, information, social media, and TM. These are displayed on Figure 4.19. For a better comprehensibility and visibility, the layers are marked with the corresponding figure numbers in the grey underlined boxes.

The Process Layer addresses the objective, needs, and requirements of the respective stakeholders. Based on these needs and requirements, potential information sources can be relevant. The steps within the Information Source Layer are conducted for outlining, classifying and selecting relevant information sources based on the information needs and requirements as defined in the Process Layer. The major part of the process model is the Social Media Layer. It consists of three components, channel selection, data extraction and availability check. The Social Media Layer considers the characteristics of social media data. Given the fact that the veracity feature is a crucial feature when exploiting social media data to support decisions, the proposed activities inside the steps of the Social Media Layer provide adequate tools for handling the veracity feature. In particular, the use of IQ dimensions in channel selection where the source credibility is assessed provides a structured methodology for assessing the believability of data published in social media. The fourth layer of the process model is TM which is based both on a generic TM process model and an organizational perspective-based TM framework. The proposed process model provides a structured methodology for exploiting social media data to support fashion and apparel SC decisions.

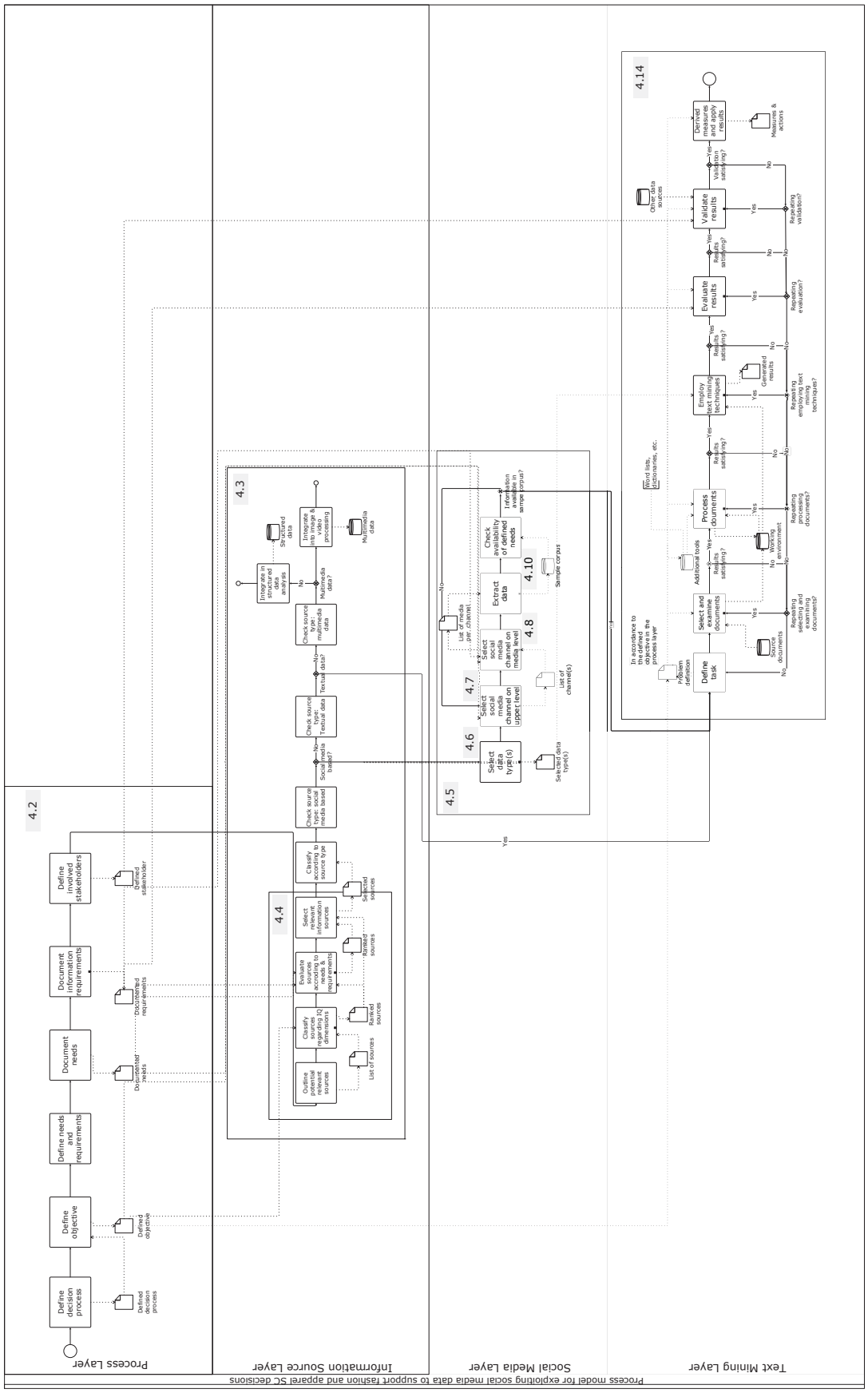


Figure 4.19.: Overall process model for exploiting social media data for supporting fashion and apparel SC decisions

5 Application of the Process Model on Fashion and Apparel Supply Chains

Following the DSRM (see Section 2.3), the use of the developed artifact, the process model, is demonstrated in Chapter 5. For this purpose, a case study approach is selected. As the application of the process model is conducted within this demonstration, this chapter is structured according to Chapter 4. The case study is framed in Section 5.1. Based on the case study setting, the different steps of the Process Layer are applied in Section 5.2. This is followed by the application of the Information Source Layer in Section 5.3. Section 5.4 shows the different tasks realized in the Social Media Layer. The processing of the textual data and the generated results are shown in Section 5.5. The chapter is concluded in Section 5.6. [10]

5.1. Case Study Setting

For exploring the methodological potential of textual social media data for SCs, we consider textual social media data and sales data to identify correlations. In this case study, a posteriori analysis is conducted. This contrasts with a priori analysis performed after the point of sale. Hence, with our approach, we cannot verify whether proper forecasting is possible. To approach the long-term aim outlined, we need to clarify whether or not a connection between blog and sales data exists, i.e. whether trends may be anticipated. Following this approach, two questions are targeted:

1. Are the dynamics with respect to time of sales data and product colour information extracted from fashion social media data positively correlated?

2. Does an economically advantageous time offset exist between the two sources?

The answer to the first question enables us to conclude whether or not textual social media data-based information may improve a decision-making process. Since we cannot expect to have complete clarity regarding the time-dependent probability distribution of the sales

[10] Some content of this chapter is published in Beheshti-Kashi (2020); Beheshti-Kashi and Kinra (2020); Beheshti-Kashi et al. (2019)

data, we focus on first principle model dynamics, i.e., the time development of the expected value. Moreover, it is notable that the first question targets correlations only. In this way, the answer to the first question will not allow us to draw conclusions on the economic advantage of the information. Only by identifying the economically advantageous time offset between the two data sources can show what decision-making processes in an SC may be targeted and impacted.

Following the methodological approach in Section 2.3 the case study is elaborated. For building the case study, six features are chosen: target group, product type, product feature, language, location and time. These features, together with their selected values, are presented in Table 5.1.

Table 5.1.: Case study features

Case study feature	Value
Target group	Women
Product type	Women's outer apparel
Product feature	Colour
Language	German
Location	Germany
Time	2015-2016

As target group and product type 'women' and 'women outer apparel' are selected. This decision is based on two main factors. In order to be reliable, the law of large numbers needs to be applied, which requires a large dataset. Firstly, women's apparel products have more variety compared to menswear. Secondly, companies have higher sales with womenswear compared to menswear. Figure 5.1 illustrates the sales figures for womenswear and menswear in Germany from 2010 to 2016.

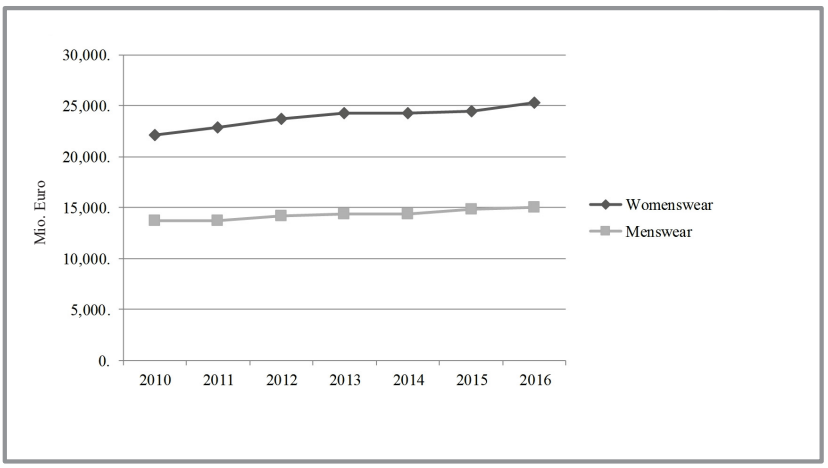


Figure 5.1.: Sales womenswear and menswear 2010-2016 in Germany

Comparing these figures, it becomes evident that companies have higher sales with wom-

enswear. Accordingly, the women's case will have a higher relevance. Due to the availability of sales data in Germany, the German market is considered and blogs targeting women only. The target language is German. In order to keep the case simple but yet realistic, the most basic feature in fashion and apparel is focussed, which is the colour. Colour is a main feature of fashion items (Jackson, 2007) and impacts decisions on the entire SC. As a consequence, different stakeholders are dependent on the availability of colour information for their decisions. This information is extracted from blogs. As sales data, data from GfK [11] is used. The GfK data is from 2016. To cover for time offsets as targeted by our second research question, we needed a larger time frame for the textual data. In order to be economically advantageous to an SC, a decision based on this information must be made before the point of sale. To cover for a wide variety of possible time shifts, we consequently fixed the range for textual data to the years 2015-2016.

For applying the process model to the framed case study, the author took the perspective of an SC member and conducted the tasks and sub-processes of the four layers consecutively. However, in order to exemplify the potential use of the process designed, the interaction of SC members with it is mapped in Figure 5.2 designed as a collaboration diagram in BPMN 2.0. The upper pool depicts the fashion and apparel SC, divided into four swimlanes which represent the retailer, apparel, fabric and yarn manufacturer together with their main tasks. The SC pool faces the process model designed as a black box in the lower part of the diagram. The expanded sub-process *apparel manufacturing* is depicted in the appendix by Figure A.1. The interaction between the pools is characterized by two messages represented as message flows. These are "Query formulated" and "Extracted information received based on defined needs & requirements". Based on the exemplified process and role descriptions in Section 3.1.2, and discussions with experts working in fashion and apparel SCs, it is suggested that the developed process interacts in particular within the range building concept phase in which the sub-processes *collecting ideas from different sources*, as well as *sales forecasting* are mainly targeted. In addition, the process of *discuss query with buyer & designers* and *sample selection from the presented range* seems most applicable. The two message flows represent an interaction on a high level. In this case, the query is conducted by the different activities and sub-processes within the four layers, and a final output is presented. However, applying the process involves high engagement and interaction for the SC members.

[11] GfK Growth for Knowledge - GfK Consumer Panel Fashion, http://www.gfk.de

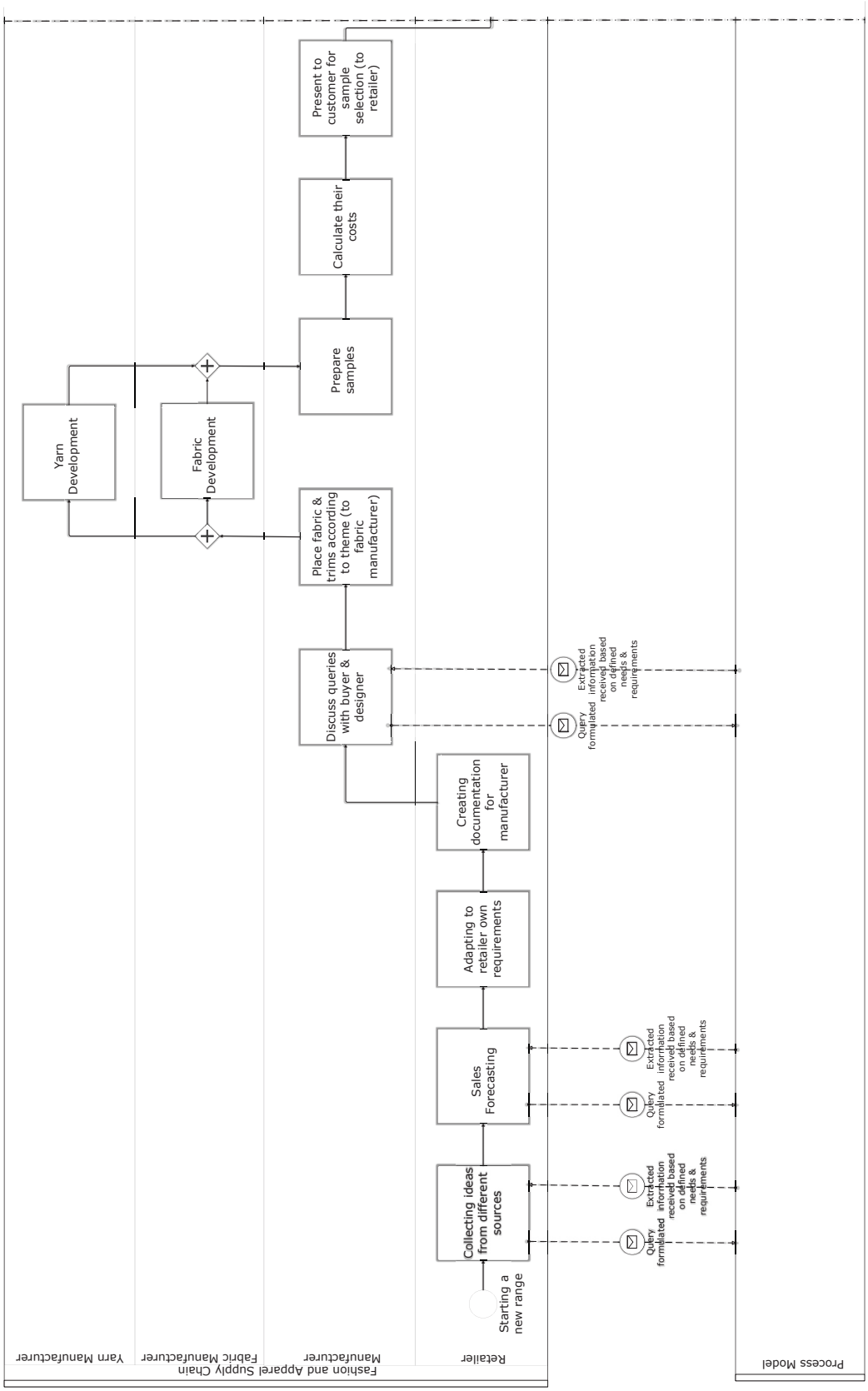


Figure 5.2.: Collaboration diagramm Process Model with Fashion and Apparel SC (1/2)

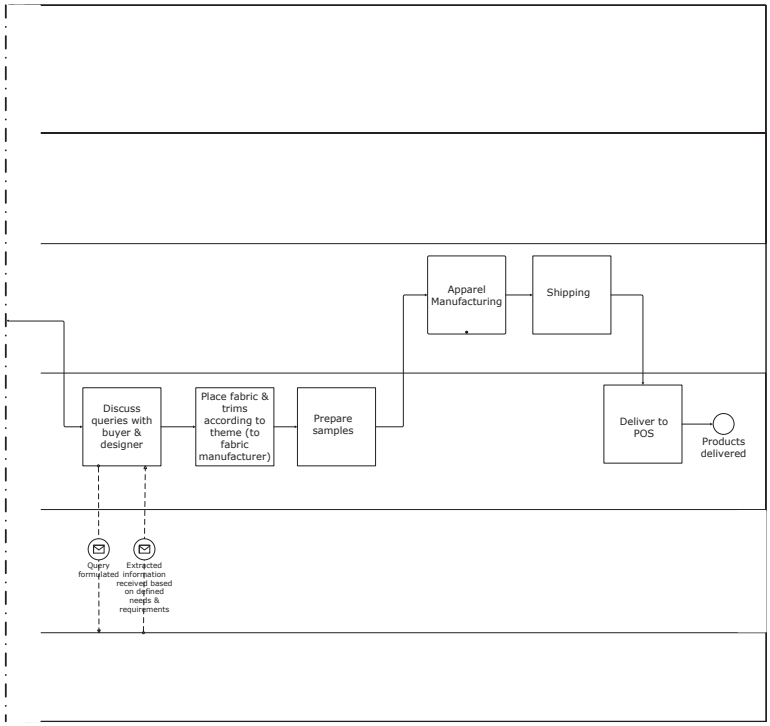


Figure 5.3.: Collaboration diagramm Process Model with Fashion and Apparel SC (2/2)

In order to follow the application of the process model on the case study, Table 5.2 is included to link the process model with the case study. The link is established by the generated data objects created at different processes. The generated data objects would also represent the messages received, processed and give back to the process model by the SC stakeholders. For this purpose, the corresponding layer, sub-process, the name of the generated data object in the process model on the one hand, and figures and tables created through conducting the process model are included on the other hand. For instance, the data object ”Outlined information components” is produced inside the ”Extract data” sub-process which is an activity in the Social Media Layer. Running the case study, this outcome is visualized by the Table ”Relevant information components” (5.8). In some cases, the result of the generated data objects is not visualized by a table or figure in the thesis but only described within the text, which is referred to in the column ”Note”.

Table 5.2.: Generated documents fo the conducted case study

Layer	Subprocess	Generated Data Documents	Table/Figure Name	Table No.	Figure No.	Note
Process Layer	- Define decision process	- Defined decision process	–	–	–	Described in Section 5.2
Process Layer	- Define objective	- Defined objective	–	–	–	Described in Section 5.2
Process Layer	- Document needs	- Defined needs	–	–	–	-Described in Section 5.2
Process Layer	- Document information requirements	- Documented requirements	-Information Requirements for Trend Predictions	5.3	–	–
Process Layer	- Define involved stakeholders	- Defined stakeholders	-Stakeholders impacted by the product feature colour	–	5.4	–
Information Source Layer	- Outline potential relevant sources	Outline potential relevant sources	-Potentially relevant information sources used for trend predictions	–	–	–
Information Source Layer	–	-Ranked Sources	-Potentially relevant information sources used for trend predictions	–	–	–
Information Source Layer	-Select relevant information sources and classify according type	-Selected sources	-Potential relevant channels for fashion related content	5.6	–	–
Social Media Layer	-Select data type	-Selected data type(s)	-Potential relevant channels for fashion related content	–	–	-Described in Section 5.4.1
Social Media Layer	-Select social media channel on upper level	-Documented Assessments	-Application of IQ dimensions on selected social media based sources	5.7	–	–

Social Media Layer	-Select social media channel on upper level	-List of channels	–	–	–	-Described in Section 5.4.2
Social Media Layer	-Select social media channel on upper level	- Defined categories - Defined features	-Application of IQ Dimensions of Fashion Blog	–	5.5	–
Social Media Layer	-Select social media channel on upper level	- Defined features for identity - Defined features for expertise - Defined features for reputation	-Believability Assessment of Fashion Blogs	–	5.6	–
Social Media Layer	-Select social media channel on upper level	-Ranking on relevance of IQ dimensions in accordance with defined objective	-Application of IQ Dimensions of Fashion Blog	–	5.5	–
Social Media Layer	-Select social media channel on upper level	–	-Selected blogs	–	A.2	–
Social Media Layer	-Extract data	-Outlined information components	-Relevant information components	5.8	–	–
Social Media Layer	-Extract data	-Outlined information components	-Selected component(s)	5.8	–	–
Text Mining Layer	-Define Task	-Problem definition	–	–	–	-Described in Section 5.5.1

5.2. Applied Process Layer

According to Figure 4.2 showing the necessary steps to follow so that the process is conducted, first, the objective and information needs are defined, then the information needs and requirements are made explicit, and lastly, the involved stakeholders are defined. The objective is to support decisions for trend prediction processes. Trend monitoring or predicting activities are highly important for the range building and sales forecasting processes. Trend monitoring involves the collection of information. Since the selected feature is colour, the information needs refer to colour-related information. In order to make the information needs and requirements explicit, both of them need to be documented. For the purpose of selecting the information requirements, the IQ dimensions are examined. All IQ dimensions, apart from accuracy, are necessary as an information requirement for trend monitoring in this case study (see Table 5.3).

Table 5.3.: Information Requirements for Trend Monitoring

Information Quality Dimensions	**Applied for Trend Predictions**	**Information Requirement**
Accessibility	Information should ideally be accessible without large effort	x
Accuracy	Information should ideally be free from errors	
Believability	The source should ideally provide additional information	x
Relevance	The relevance of information constitutes the information needs	x
Timeliness	Information available together with corresponding time stamp	x

For defining the stakeholders involved in the defined decision-processes, two questions are answered.

- Which stakeholder is impacted by information on the product feature colour?
- When is the information on the colour required?

Based on these questions, Figure 5.4 illustrates the stakeholders who are usually mostly impacted by the product feature colour. Mainly, fabric manufacturer, garment manufacturer and retailer.

Figure 5.4.: Stakeholders impacted by the product feature colour

5.3. Applied Information Source Layer

As defined in Section 4.3, the objective of the Information Source Layer is to select appropriate information sources which will adequately be processed in further steps. All four steps that must be taken so that the Information Source Layer is conducted (see Figure 4.3) are applied to the case study. Their implementation is described in the following sections.

5.3.1. Outline potential relevant sources

In order to obtain information on the defined information needs, which in the case of trend monitoring are typically trend manipulating features (see Figure 3.5), the buyer needs to outline a list of potentially relevant sources. Table 5.4 indicates potential relevant information sources in fashion and apparel SC for trend predictions.

Table 5.4.: Potentially relevant information sources used for trend predictions

Information sources	Descriptions
Historical Sales Data	Companies sales data on past collections
Real-time point of sales data	Real-time sales data obtained by point of sales
Fashion Shows	Designer uses fashion shows to present their upcoming collections
Fashion Fairs	Fashion companies showing and demonstrating clothing/cuts/-colours/fabrics
Fashion Magazines	Online or print fashion magazines publishing on fashion related topics
Current Street Styles	Trendscouts visiting different cities, campuses with the goal to identify street styles
Media (TV/Movies)	TV shows, movies often impact the customers' purchasing decisions
Forecasting Agencies	Forecasting agencies generate reports on upcoming trends, operating globally and locally
Consumer Data	Consumer data often provided by market research institutes
Social Media	Individual user can publish experiences on products on different social media

In order to obtain information about the product feature colour, fashion companies access different information sources. One primary source is the companies own historical sales database. One disadvantage of this historical sales data is that it is based on past sales. A large part of fashion items in fashion collections is often replaced by new items. Therefore, usually a little amount of historical sales data exist. Besides, since fashion products are often highly seasonal and designed for a specific time, complete historical sales data are hardly available for newly launched products. Therefore, companies integrate real-time sales data

coming directly from the point of sale. The inditex company Zara, for instance, transfers the sales data daily to their design teams. In this way, these insights can be integrated into future product development activities. Fashion shows, fashion fairs or fashion magazines also are information channels used for trend prediction. Companies often employ so-called trend scouts who look for current street styles, e.g. on university campuses. Media (TV/movies) is in general a relevant information channel that is monitored. Moreover, companies often work with forecasting agencies generating prediction reports on upcoming trends. Another possible way of obtaining consumer data is purchasing this type of data from market research institutions. This method of capturing consumer preferences is related to high costs. With the emergence of social media services, fashion and apparel companies have realized their importance and have started profiting from this valuable source in terms of obtaining direct information from potential customers as well as reaching out to potential customer in a more direct manner.

5.3.2. Classify Sources regarding information quality dimensions

For classifying the sources regarding the IQ dimensions, there are several tools available for conducting the AHP method, for instance, excel. We used the AHP priority calculator and performed pairwise comparisons (Goepel, 2018). In doing so, the importance of each dimension for the respective sources is assessed. As in this case study, the goal was to only include social media-based sources, the pairwise comparison was performed only for social media-based sources. The targeted criteria are listed in the following:

- *Criteria:* Accessibility, Accuracy, Believability, Objectivity, Reputation, Value-added, Relevancy, Timeliness, Completeness, Amount of data, Interpretability, Ease of Understanding, Consistency, Manipulability, Conciseness, Security

Based on the assessment, a ranking of the quality dimensions was generated, and only the five highest-ranked quality dimensions are to proceed with in the following steps of the process model. For social media channels, dimensions are ranked the highest and are therefore used for the proceeding assessment steps (see Table 5.5): accessibility, accuracy, believability, relevancy and timeliness.

Table 5.5.: Pairwise comparison of IQ dimensions for social media based sources

Information Quality Dimensions	Value (%)	Ranks
Accessibility	18,5	1
Accuracy	9,1	5
Believability	17,7	2
Relevance	10,3	4
Timeliness	10,5	3

5.3.3. Evaluate Sources according to needs and requirements

As it was mentioned in Section 4.3.3, based on the existing experience, it is not always necessary to perform a pairwise comparison to evaluate the sources according to the needs. For this case study, different discussions with experts have resulted that it is expected that colour related information would be available in social media-based sources. This contrasts, for instance, the availability of information on the silhouette on social media.

5.3.4. Select relevant information sources and classify according to source type

Social media-based sources are selected as one relevant type of information sources. As described in Section 3.2.1, many social media channels exist. Nevertheless, it is necessary to select only the channels relevant to the defined objective. Table 5.6 exhibits the social media channels [12] which are relevant in the context of fashion and apparel.

[12] In accordance with the scope of the present thesis and the exploitation of textual social media content, media sharing platforms are not considered in the following.

Table 5.6.: Potential relevant channels for fashion related content

Channel	Description
Fashion blogs	A fashion blog is a weblog genre dealing with fashion, lifestyle and related topics.
Social networking services (Facebook)	Fashion and Apparel related topics exist on Facebook in form of official Facebook pages of apparel companies. In addition, fashion blogger having their own blogs, and are in addition engaged in Facebook. "Normal" user publishing fashion related pictures, purchases. Events such as fashion weeks have their own fan pages.
Microblogging services (Twitter)	Similarly, to Facebook, fashion and apparel companies as well as fashion blogger are engaged in Twitter with their own accounts. In contrast to Facebook, fashion related content is often tagged with corresponding # (hashtags).
Media sharing platform	In media sharing platforms fashion related content is available in form of videos and pictures. YouTube and Instagram are two examples of relevant channels in the context of fashion related content. Fashion and Apparel companies, fashion blogger are involved in both channels. Textual data exists in the two channels in form of user comments and feedback.

5.4. Applied Social Media Layer

The social media-based sources selected in the previous section, the Information Source Layer, are adopted in the Social Media Layer. Based on the objective of the Social Media Layer, the steps (see Figure 4.5) followed to conduct this layer in the case study are reported as follows.

5.4.1. Select data type(s)

In accordance with the sub-process of *select data type(s)*, in this case study, textual data and metadata are selected as the relevant data types (see Figure 4.6).

5.4.2. Channel Selection

The channel selection process is conducted for both the upper level and medium level.

Select social media channel on upper level

The IQA at the upper level (see Figure 4.7) is performed based on two outputs of the Information Source Layer, namely, selection IQ dimensions and selected relevant channels. Table 5.7 describes every selected IQ dimension for each chosen relevant channel. On the upper level, a direct measurement of the quality dimensions is not targeted. The assessment

is conducted on a structural level to examine the potential of the channels for measurement of media level.

Table 5.7.: Application of IQ dimensions on selected social media based sources

IQ Dimensions	Fashion Blogs	Social Networking Sites -Facebook	Microblogging Sites- Twitter
Accessibility	- Content freely accessible for user - Archives enable extraction of older content - Multi-Language Content	- Content freely accessible through official API - Limitations in terms of time and volume	- Content freely accessible through official API - Limitations in terms of time and volume
Believability	- Identity: Access to further social media engagement, section about me-personal information available - Expertise: Large data base for assessing the expertise, availability of archive - Reputation: Cooperation with companies visible and mostly transparent	- Identity: Not easy to identify, higher percentage of fake accounts - Expertise: Difficult to access domain specific language only from one post difficult to monitor company collaboration - Reputation: Number of friends, number of likes and shares	- Identity: Not easy to identify, higher percentage of fake accounts - Expertise: Difficult to assess expertise from short message, reduced information value due to short message - Reputation: Number of followers, number of retweets
Relevance	- To be defined contextually - Fashion related topics available	- To be defined contextually - Fashion related topics available, but not the main topic	- To be defined contextually - Fashion related topics available but not the main topic
Timeliness	- Mostly time stamp available on post level - If not available, on user comment level	- Time stamp available	- Time stamp available

In terms of accessibility, the content published on fashion blogs is freely accessible. The access barriers are low as there is no need for registration. Furthermore, fashion blogs often provide an archive in which older posts can be accessed. However, this functionality is not available on all blogs. Lastly, posts are often published in more than one language, which broadens the accessibility. In contrast to fashion blogs, the accessibility of social networking sites is often limited. Content is freely accessible through the official developers API[13] . Moreover, registration is required. In terms of legal barriers, only public accounts can be accessed, and only the level of privacy which they have granted can be accessed. It should be noted that everything but the official API is illegal and accordingly violates the copyrights. [14] In terms of technical barriers, extraction is limited to one URL at a time. Keyword search is not possible. Similarly to Facebook, the microblogging service Twitter provides an official API for accessing the data. However, there are some limitations that have to be noted. In terms of legal barriers, everything but the official API [15] is illegal and accordingly violates the copyrights. In terms of technical barriers, it is notable that for "normal users" there is no access to historical data older than seven days over the official API [16]. Moreover, the official API returns only a sample of tweets (ca. 1 pct.). One further access option would be purchasing tweets.

In terms of believability, the three categories, identity, expertise and reputation are considered and elaborated on with regard to how these categories may be approached on the different channels. While in terms of reputation, all three channels provide structural elements to measure the reputation and potential influence on the community, expertise, but in particular, the identity assessment is more transparent on fashion blogs. Fashion blogs often provide a section on the author(s) where personal information on the blogger is published. Furthermore, the engagement in other social media channels is easier to track as most social media icons of corresponding channels and links are transparently available. This enables easier tracking of general consistency in terms of identity. Moreover, Twitter and Facebook are more vulnerable to fake accounts. Although the relevance assessment is highly contextual, at a higher level, the availability of fashion-related content can be elaborated on for assessing the relevance for the upper level channel selection. As a central topic in fashion blogs is fashion, the relevance of this channel is considered compared with Facebook or Twitter. On which of course fashion-related content is available.

In terms of timeliness for fashion blogs often a timestamp is available on post level. Moreover, timestamps are available at the user comment level. For Facebook and Twitter, timestamps are available as metadata. On Facebook, timestamps of likes and shares cannot be accessed (Cui et al., 2018). The output of the channel selection on upper level: Following

[13]https://developers.facebook.com/

[14]https://developer.twitter.com/en/developer-terms/agreement-and-policy.html

[15]https://developer.twitter.com/

[16]https://developer.twitter.com/en/developer-terms/agreement-and-policy.html

Table 5.7 and the elaborations on the different quality dimensions, the defined information needs, fashion blogs are considered to be the most suitable channels for conducting further analysis.

SELECT SOCIAL MEDIA CHANNEL ON MEDIA LEVEL

After the selection of fashion blogs as the social media channel, the selection at the medium level is followed as a second step of the channel selection (see Figure 4.8). This includes the selection of individual blogs. It is necessary to set a starting point for the selection process. The following two alternatives were sketched: either generating an individual fashion blog ranking or using an existing fashion blog ranking. Since a ranking is based on certain criteria, for the first alternative it would be necessary to specify criteria and apply them to the blogs. As the generation of an individual ranking is not in the scope of the present research, the second alternative was selected. The ranking provided by Teads (2015), which shows the 100 most influential fashion blogs in 2015, was selected. The ranking is based on a range of criteria such as backlinks or shares on Twitter or Facebook and has been used in a range of scientific works, e.g. Kopetzky (2013). In this manner, the starting point for the individual blog selection process were these 100 blogs upon which the IQA was conducted. For the scope of the present thesis, the dimensions relevance, believability, timeliness, and accessibility are the dimensions which have to be proven in order to include the blogs in the fashion blog corpus. While for the evaluation of relevance and believability, it is required to look at the blog as a whole, for others such as timeliness and accessibility, it is necessary to consider the post level. Figure 5.5 illustrates the application of the IQ dimensions relevance, believability, timeliness, and accessibility for the selection of individual fashion blogs. As the accuracy dimension is not measurable at the content level, it was therefore not considered in the selection process of the individual blogs. The process starts with assessing the relevance because it is defined as required. If a blog does not meet the relevance criteria, it is already sorted out. Timeliness, believability and accessibility assessment follow. For a better overview, the believability assessment is condensed in Figure 5.5 and the detailed assessment process is illustrated in Figure 5.6

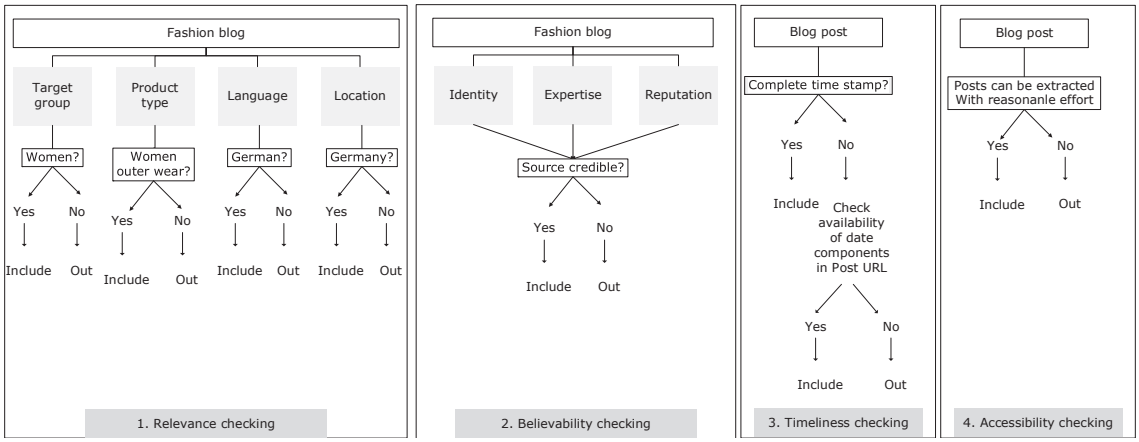


Figure 5.5.: Application of IQ Dimensions of Fashion Blogs

With regard to ”fitness of use” as described in Section 3.2.4 the definition of the relevance is defined in accordance with the defined case study scope and the case study features (see Table 5.1). For each of the case study features, it was checked whether the respective feature of the blog met the defined value. e.g. target group equal to women. The relevance is only proven if the four features of the blog are correct, that is, that they match their respective value. In this case, the blog is included in the fashion blog corpus. Otherwise, the blog is dismissed. This information dimension is defined as hard. Figure 5.6 illustrates the believability assessment according to the defined in Section 4.4.2. For operationalizing the believability dimension, source credibility is used as the main factor constituting this dimension. The credibility of a source is proven when the source shows all identity, expertise and reputation. In order to prove if these three components are confirmed, each one of the three components must be broken down into measurable features. For this breakdown, the features are designed in a manner that a binary assessment is possible. Figure 5.6 shows the process of proving the identity, expertise and reputation for each blog. This process starts by checking the identity of the blogger based on available author information. Other social media platforms were also checked to monitor the consistency of the different social media platforms on which the author was present. Since most of the blogger provide links to their social media engagement, these links were followed. Subsequently, for assessing the expertise of the source, we inspected the involvement with companies, as it was assumed that these cooperate with bloggers to whom they assign expertise. For this purpose, the availability of sponsored posts or affiliated links were examined. Moreover, domain-specific language was checked as its use considered to show expertise in the field of fashion. For assessing the reputation, four features were checked. First, the engagement of the blogger in other social media was checked. Bloggers usually link their other social media engagements on their blogs. These links were used for the assessing engagement in other social media. Other indicators for the reputation are the number of followers and friends. Moreover, the availability of comments on published posts was assessed. Similar to the expertise assessment, the cooperation with companies was considered as an indicator for the reputation of the blog. It is

assumed that companies preferably cooperate with bloggers who have a broad reach and enable coverage to potential customers for the companies. The overall believability assessment was conducted through the perspective of the data consumer who is the domain expert, and who is thus able to assess the information regarding the described fashion related topics by her/his knowledge and expertise. When identity, expertise and reputation had been proved, the blog was considered to be a credible source. Thus, data believability is achieved and the blog included in the fashion blog corpus. The result of the believability assessment was that all of the blogs had passed the assessment.

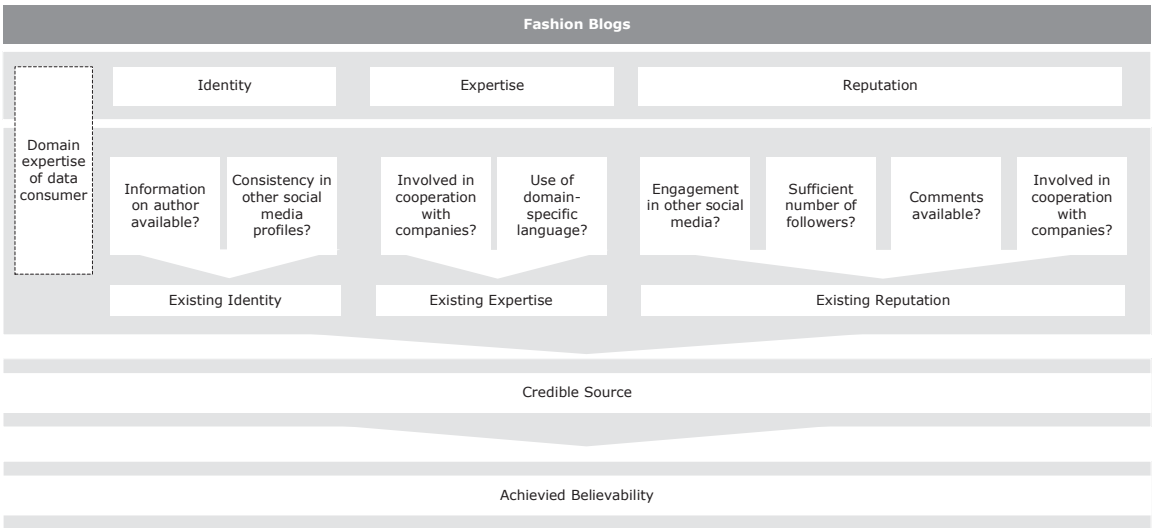


Figure 5.6.: Believability Assessment (Beheshti-Kashi and Kinra, 2020)

For checking the timeliness criteria, a two-step workflow was defined. Firstly, it was checked whether the blog provides dates and complete timestamps displayed on the posts. If so, the blog is included in the fashion blog corpus; otherwise, the blog is checked as to whether the post URL includes date information, at least the month and year. If both were provided, the blog was included in the corpus. Thereafter, the given date information was extracted from the URL and converted into a dd.mm.yyyy-format. In order to obtain a complete URL, a default value was generated by adding the first day of the month to month and year information from the URL. For example, July 2015, is transformed into 01.07.2015. This is done based on the fact that the time reference to a specific month is sufficient.

As the third and final requirement, the accessibility of a blog is related to the potential extraction of the posts. Therefore, the main question followed while examining the accessibility is whether the extraction of the targeted posts (time frame) is possible by the given resources. For the operationalization of this question, the accessibility of each blog was tested by applying test runs with KNIME. A further dimension of accessibility is defined through the language. Blogs publishing content in a language other than German were omitted from the blog corpus. Figure A.2 lists all blogs included in the fashion blogs corpus. By selecting the individual fashion blogs, the last step in the Social Media Layer is finished.

5.4.3. Extract Data

Once the fashion blogs had been selected, it was necessary to conduct the data extraction. Following the four steps displayed in Figure 4.10, the structure of a fashion blog is first examined. Based on this, relevant information components are selected, an extraction strategy set and, a sample corpus generated.

Examine the structure of the selected fashion blogs

Although there is no standardized template for designing a fashion blog and structuring the posts, similar structures have emerged within the fashion blogosphere. Figure 5.7 illustrates a typical post published on a fashion blog. The length of the text differs from post to post. The main focus is often its visual illustration (pictures/videos). Consequently, the proportion of textual information to images is in favour of the images. A characteristic of a fashion blog post is a list with each outfit item combined with the brand/store and a link to where the item was or might be purchased. By displaying this information and combining it with a potential point-of-sale, the blogger might directly influence the success of an item since the reader can easily purchase the item with some clicks.

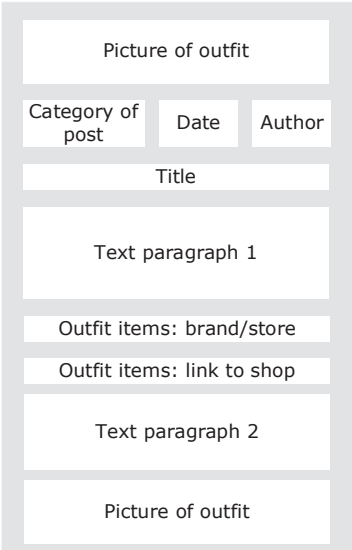


Figure 5.7.: Schematic structure of a Fashion Posts (Social Media Unit Fashion Blog Post)

Select the relevant information components of fashion blog post

The baseline for the selection of the relevant information components is the structural examination of the fashion blog post displayed in Figure 5.7. Following Figure 4.11, which is based on these defined structural elements, the different information components are described and defined along with their relevance on the overall defined objective. Matching the information component with their relevance to the defined objective enables a targeted extraction of the date and prevents the “blind” streaming of data. Table 5.8 outlines the

identified information components with a short description, their relevance for the defined objective and indicates its selection state. The definition of the data type is based on Figure 3.10.

Table 5.8.: Relevant information components

Information component	Description	Relevance for the defined objective	Data type	Selection
URL of post	- Reference to the post	- Relevant in terms of traceability of the data	semi-structured	x
Title of post	- Mostly catching title - First impression for readers	- Relevant - Title may be used for a getting a quick overview of the data without deep processing	semi-structured	x
Date of post	- Data is handed differently, in the blogs in terms of displaying the date or not. Also, the format of the date is differently organized throughout the fashion blogs	- Relevant - Crucial for assigning post to a particular time period/season - Season-dependent	structured	x
Name of author	- Not a direct reference to the post	- Not relevant	structured	
Category of post	- Fashion blogs write in different categories	- Relevant for filtering out non-relevant categories such as food/traveling	structured	
Pictures	- Pictures of the outfit in different categories - Differences in number	- Not relevant since textual data is focused	unstructured	
Textual information/ Text paragraphs	- Description of the outfit or location or event - Differences in length	- Relevant - Main focus	unstructured	x
Comments	- Blogs allow comments - Often blogger comments other blogs	- Not relevant - Direct user feedback	unstructured	
Social media icons	- Show the number of follows/shares in the different channels	- Interesting for monitoring purposes and measuring degree of influence	structured	
Number of comments of the articles	- Shows the number of comments of the particular post	- Not relevant	structured	
Listing of the outfit items with brands name/shops name	- All outfit items are listed separately together with the brands	- Not relevant	semi-structured	
Links to brands/shops	- In a second box often included with more pictures and links to the shops - In case the item worn in the pictures is sold out, an older item or handmade, similar articles are selected and linked to	- Not relevant - Not relevant	semi-structured	

In accordance with the defined objective, the title of a post, the date, the textual paragraphs and the URL are selected. Due to the fast-paced nature of the fashion industry, the date is considered as one of the most important information components available on a fashion blog. Without the indication of a date, the assignment of a described garment or garment feature to a season is hardly possible. Furthermore, the date enables the tracking of the development of garments and their features over a certain period of time. Therefore, the dates have to be converted into a standard format. Besides the date, the textual paragraphs are considered as the main information component in accordance with the defined objective. In addition, the title and URL are selected. The title may enable a quick overview of the corpus, without deep processing of the whole textual data. For tracing the posts, it is useful to collect the URL along with the corresponding posts.

It is notable that this selection does not imply that the remaining information components are not relevant in general. Further components such as pictures/videos or user comments

may be relevant in a different context, but as thus are not in the focus of this research, they are not selected for the further examination.

SET EXTRACTION STRATEGY FOR THE FASHION BLOGS

After defining the relevant information components to be extracted, the extraction strategy was set. Different options to collect data from blogs are available. One option is to use RSS feeds. The collection through RSS feeds have one great advantage concerning the structure: in RSS feeds, the posts are already structured, and the data can be stored more accurately through the analysis of the blogs. When setting RSS as an extraction strategy, we encountered two challenges. First, not all of the blogs provide the opportunity to subscribe to an RSS feed. Second, depending on the defined setting of the RSS feed, not all posts, and particular old ones, are not contained. Often a period of two weeks was covered by the feeds. In this case, either the feeds need to be collected over a longer period, or they can be used for short-term exploitation. In addition, the extraction strategy should be based on the best available option in order to reduce costs and efforts which will arise when setting a range of extraction strategies instead of one consistent strategy. Since in the case of RSS feeds it was not possible to set a consistent strategy, RSS feeds were not set as extraction strategy, and since no API was available for the blogs, web scraping was set as the extraction strategy.

GENERATE A FASHION BLOG SAMPLE CORPUS

Once the extraction strategy had been set for the blogs, a sample corpus was generated. For this purpose, we used KNIME Analytics Platform[17] and Import.io[18]. From each selected blog, a set of blog post has been collected and a sample corpus generated. For the sample corpus, all the selected information components are respectively collected. For the sample corpus it is not necessary to collect "all" the blog posts. The extensive corpus generation is conducted in Section 5.5.2.

5.4.4. CHECK AVAILABILITY OF COLOUR RELATED CONTENT

Following Section 4.4.4, the availability of the defined information needs were to be examined in the last step of the Social Media Layer. For this purpose, the generated sample corpus was analysed regarding the availability of colour occurrences. Selected fashion blog posts were examined by CA. The availability check revealed that colour occurrences were available. Based on this finding, the Text Mining Layer can be conducted in the following.

Summarising the Social Media Layer, the channel selection resulted in the selection of fashion blogs. The data extraction generated the URL, title, data and message of each fashion

[17] https://www.knime.com/
[18] https://www.import.io/

blogs post as relevant information components and web scraping as a suitable extraction strategy. The availability check showed the availability of colour occurrences.

5.5. Applied Text Mining Layer

Following the Text Mining Layer described in Section 4.5, the task was defined, the documents were selected, their structure analysed and loaded in the working environment. After that, the processing of the documents was conducted, and text mining operations applied, and the results evaluated. In the last step, a validation of the results is conducted. The operationalisation of the Text Mining Layer is conducted in KNIME and illustrated by Figures A.3 and A.4.

5.5.1. Define Task

In accordance with the overall objective defined in Section 5.2, we define the following tasks: detection and extraction of individual colour occurrences, classification of single colour occurrences in colour groups, and trend analysis: tracking of colour groups over time. The first task defined requires different steps. In order to detect colour occurrences in the fashion blog corpus, it was required to generate a dictionary containing colour information. This enables the tagging of the colour occurrences within the fashion blog corpus, which is the first step in reaching the overall objective. After detecting and extracting the individual colour occurrences, it was necessary to classify them. This classification into colour groups is essential for two reasons. Firstly, it would otherwise hardly be possible to make claims about their distributions due to the high variety of individual colour tones. Secondly, this classification helps to create a baseline and generate comparability to other data sets, in our case, to the sales data. As already pointed out, the time aspect is particularly crucial in apparel and fashion SCs. For fulfilling the overall objective of the case study, it is essential to track the colour group over time. In this way, and as a final task, we target the tracking of the colour groups over time.

5.5.2. Select documents and examine source documents

As illustrated in Figure 5.8, the document selection and analysis step consists of mainly three sub-steps: namely, identifying the source systems and documents, determining document properties and loading in working environment.

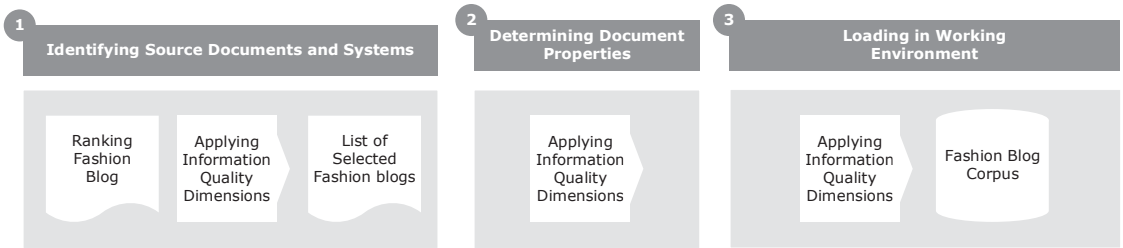


Figure 5.8.: Document Selection and Analysis for fashion blogs

In identifying the source system and documents, the question to be answered was which blogs could be included in the data set. For this purpose, the IQ dimensions accessibility, accuracy, believability and timeliness are applied to the fashion blog ranking. The output of this process is a list of selected fashion blogs. The detailed selection of the individual blogs is described in Section 5.4.2. For the determination of the document properties, the examination of the document structure is a task within the data extraction (see Section 5.4.3) of the Social Media Layer. In this stage, the document properties are determined in terms of the working environment. For loading documents in a working environment in the last sub-step, we generated the blog corpus by applying the methodology displayed in Figure 5.9. The steps are described in the following.

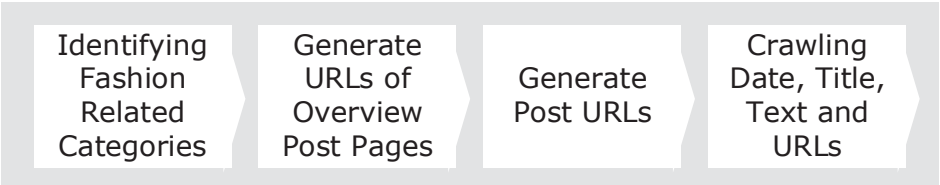


Figure 5.9.: Methodology for crawling blog posts

As we are interested in fashion-related content only, it starts by identifying the fashion-related categories of each blog. For the selected blogs, the following categories are selected: ”outfits”, ”outfit”, ”looks”, ”fashion”, ”Mode” (German word for fashion) or ”trend”. For the purpose of accessing the individual posts, we needed to have access to the overview sites. A typical overview page, lists a certain number of posts, linking to the actual post articles. In this manner, in the second phase, we generated lists of overview post pages for each blog. In the third step, we use these lists for crawling the post URLs. In the last phase, the generated post links served as input for accessing the required content, date, title and text paragraphs. The post generation and crawling of date, title and text were operationalized through the KNIME workflow as displayed in Figure A.6 is described in the following. The post overview links are used as input for obtaining the generating the single post links. We used a web reader developed within the KNIME community and provided by Palladian. The HtmlParser node downloads the HTML content and includes it as an XML column in the output port (Wiswedel et al., 2013). Since we are not interested in the entire XML content, we need to

apply XML parsing in order to select the relevant information components set up in Section 5.4.3. KNIME offers a range of XML operations. For this purpose, the XPath node is used in order to perform XPath queries on the XML document, which provides the functionality of "performing an XPath query on a column of XML cells and can produce another XML cell (a Node), a collection of XML cells (a Node-Set), or just a single value (number, string, Boolean, etc...) depending on the result of XPath query" (Wiswedel et al. 2013, p. 8). Path expressions are used by XPath to select nodes in the corresponding XML document. As fashion blogs are different in terms of structure, it was required to configure the XPath node for each blog individually in order to crawl the title, date and textual paragraph. Finally, the posts are saved as CSV and Excel files and then stored in a Microsoft Access database. The Access database is used to conduct some preprocessing, applying some SQL queries to organize the database, for instance, to sort in years and months.

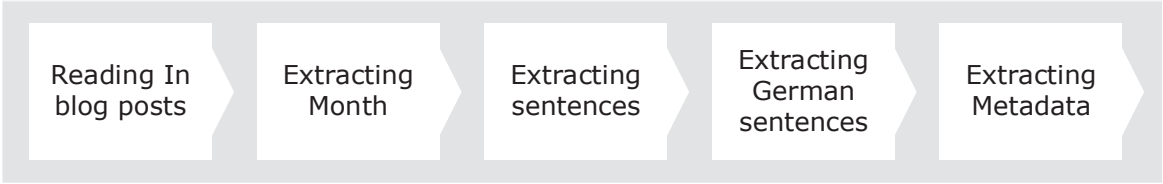


Figure 5.10.: Reading In and first pre-processing steps

After loading in the data in KNIME, the following sub-processing are applied: extracting month, extracting sentences and filtering German sentences. The operationalized workflows in KNIME are illustrated in the appendix (see Figures A.5, A.7, A.8, A.9). Filtering only German sentences was particularly relevant at this stage, as through the scraping of the blogs, in addition to the German parts, also the English chunks were scraped. Since the English parts of the posts are translations of the German posts, it was required to filter them out. Otherwise, they could affect the results. For the filtering of the posts, German and English stop words were used, setting up corresponding tagging rules for German and English sentences. As the last step, additional metadata (website and time stamp) are extracted. In particular, the extraction of the timestamp is crucial for the tracking of colour groups over time.

5.5.3. Process documents

Figure 5.11 provides a schematic overview of the document processing phases conducted. The operationalization of the conducted steps is displayed in the workflow by Figure A.11.

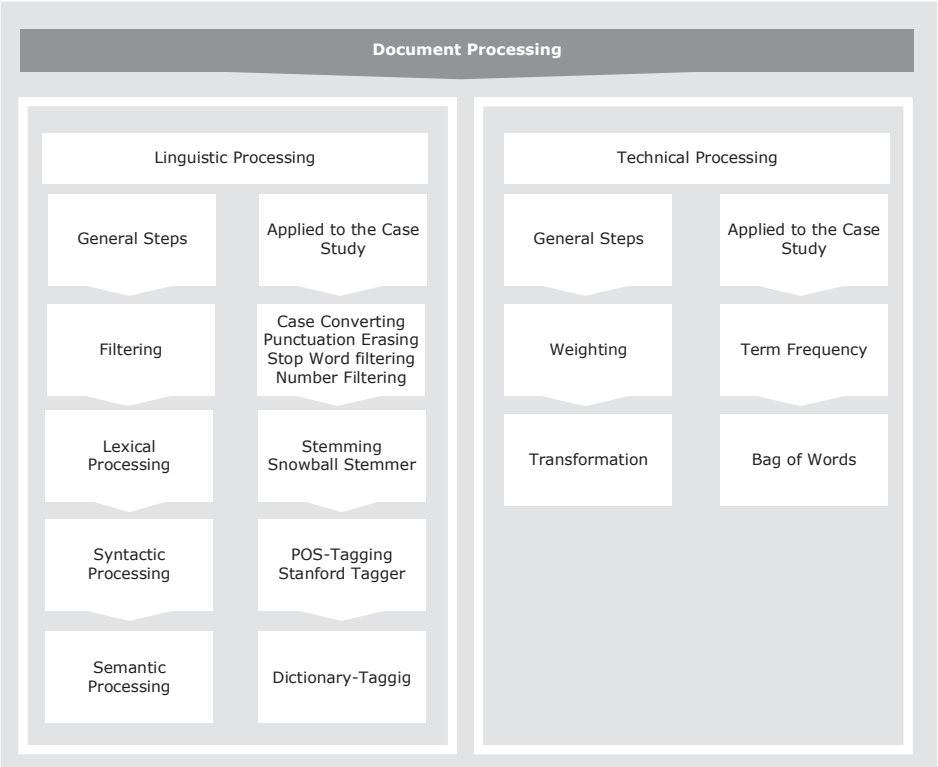


Figure 5.11.: Document Processing steps applied to the case study

The document processing phase includes the linguistic and technical processing. The linguistic processing is initiated by performing basic pre-processing such as case converting, stop word filtering, punctuation erasing, and number filtering. After that, lexical processing was conducted by performing stemming operations. The snowball stemmer [19] was used, as it provides the functionality for German. In the syntactic processing, POS tagging is applied. Conducting POS tagging before the detection of individual colour occurrences enables a more efficient extraction as only adjectives were extracted. In KNIME, the Stanford Tagger was used which is based on models underlying the Stanford NLP group [20]. A POS tagger requires an underlying tag set. For German, the Stuttgart-Tübingen Tag set [21] underlines the Stanford Tagger.

A Dictionary Tagger filled with a colour dictionary is used to semantically enrich the data and is used for tagging the adjective list accordingly. The semantic enrichment is performed as illustrated by the KNIME workflow in Figure A.10. In the second phase, the technical processing term frequency was applied as the weighting measure. Bag of words is applied in order to transform the data into a suitable format for further analysis. In both phases of linguistic and technical processing, we applied a range of evaluation steps, tested different workflows until the final workflow was created. As these pre-processing steps set the foundations for the subsequent steps, it was necessary to examine the quality of the results

[19] https://snowballstem.org/

[20] http://nlp.stanford.edu/software/tagger.shtml

[21] https://www.ims.uni-stuttgart.de/forschung/ressourcen/lexika/GermanTagsets.html

conducted by different workflows. For this purpose, two measures were taken. Firstly, the position of the nodes and secondly, their configuration in the workflow were adapted. For instance, the workflow was tested with and without POS-Tagging. Without POS-Tagging, the occurrence of the word ”weiss” is higher compared to the results running the workflow with a POS-Tagger because ”weiss” also represents the first person singular of the verb ”wissen”. If we run the workflow without including a POS tagger, all occurrences of the word ”weiss” are considered. However, when including the POS-Tagger, only the adjective ”weiss” tokens are correctly filtered. Another example is the testing of the positions of case converting and stemming before and after the semantic enrichment (POS Tagging/Dictionary Tagging). The tests showed that applying both operations before improves the results of the semantic enrichment. In summary, testing played an essential role in the reliability and quality of the generated text mining results.

5.5.4. Employ text mining techniques

After we had extracted the required data and processed it adequately, we applied different operations to the processed data. The TM operations were selected in accordance with the defined objectives of the Text Mining Layer stated in Section 5.5.1. Overall, we generated an approach which consists mainly of the three analyses streams: Dictionary Based Keyword Extraction, Rule-Based Classification and Automatic Frequency Analyses. For these streams, workflows were created in KNIME.

For generating a dictionary for tagging and extracting the keywords, which in this case are colour names, we used the GfK colour names used by the households in the first analysis stream. One reason for that was to include colour naming used by ordinary people. We repeated this process of all the colour groups. After conducting frequency analyses for the single colour occurrences, the necessity of grouping the colours arose in order to obtain a baseline for comparing the blog and the GfK data. For this purpose, we established a range of rules for classifying the single colour occurrences into colour groups. In the third and final analyses stream, we conducted automatic frequency analyses at the level of the single colour occurrences and for the classified colour groups.

The sunburst chart displayed in Figure 5.12 illustrates the individual colour occurrences together with the corresponding colour groups. Classifying the colour occurrences became evident through discussions with experts working in different positions along with the fashion and apparel SC. These colour occurrences were classified according to the colour classification scheme used by the GfK 2016.

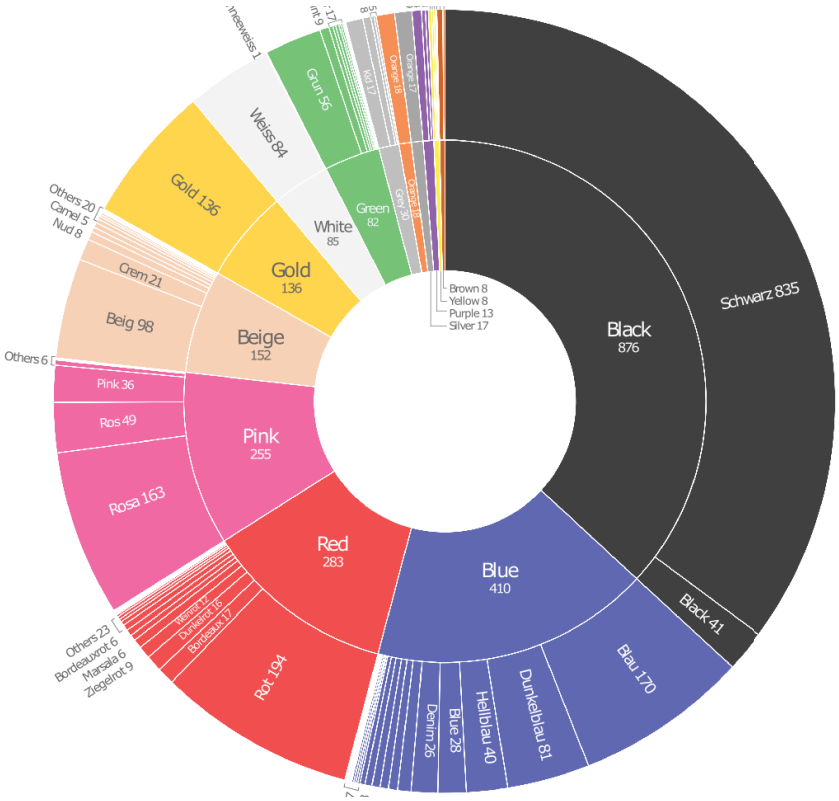


Figure 5.12.: Extracted colours 2016 – individual colour occurrences - classified

For visualization purposes, the individual colours are displayed in their respective classification group colours. The different font sizes and colour shapes are based on the frequencies of the colour occurrences[22]. From the sunburst chart, three observations can be made. Firstly, it shows the dominance of the colour ”schwarz” (black). Secondly, the different variations of the tones in the group ”blau”, ”rot” and ”gruen” can be observed. Thirdly, it shows the necessity of the classification step. When comparing the colour occurrences at the individual colour tone level, it is hardly possible to draw conclusions from the data set due to the high variety of the individual colour tones. The colours black and white are neglected in the following analyses since they are considered basic colours. For a more fluent readability, the group names of the classifications are translated into English (see Table 5.9). They are used in the following.

[22]The colour”weiss” is displayed in black in terms of readability

Table 5.9.: Translated colour groups

Colour group	Translation
Beige	Beige
Blau	Blue
Braun	Brown
Gelb	Yellow
Gold	Gold
Grau	Grey
Grün	Green
Lila	Purple
Orange	Orange
Rosa	Pink
Rot	Red
Silber	Silver

Table 5.10 gives an overview of the blog statistics for 2015 and 2016. In total, 2318 blog posts for 2015 and 1741 blogs from 2016 are extracted. As described previously, the blog posts are extracted into sentences in order to remove duplicates and to filter out German sentences only.

After cleaning the posts from the duplicates and English, 31193 sentences remain for 2015 and 25905 for 2016. Table 5.10 further shows statistics of the extracted colours on different levels. Overall, we identified 1302 colour entries in 2015 and 1204 in 2016. These contain 96 (2015) and 88 (2016) different colours. Based on the GfK classification scheme, we have classified these colours into 14 colour groups.

Table 5.10.: Descriptive Blog Statistics

Year	Number of documents	Number of sentences	Identified colour entries	Different colours	Colour groups
2015	2314	31193	1302	96	14
2016	1741	25905	1204	88	14

For targeting correlations between the blog data and the GfK data in a later step, months are used as units. Therefore, it is worth examining the blog data on a monthly basis. From the monthly distribution of the blog posts it can be concluded that posts are more or less equally distributed over the 12 month period.

In order to track the colour occurrences over time, it is necessary to add the respective time information to the posts. Figure A.13 illustrates the classified colours over a time period of 12 months for the year 2016.

In order to track the colour occurrences over time, it is necessary to add the respective time information to the posts. Figure A.13 illustrates the classified colours over a time period of

12 months for the year 2016.

Figure 5.13 illustrates the colour distributions for 2015 and 2016 on a yearly basis. The top-ranked colour groups in both years are blue, red, pink, beige, gold and green. It is noticeable that the order in the ranking is the same in both years. Nevertheless, there are differences in their developments. While for the groups beige, blue, red and green a decrease from 2015 to 2016 can be noted, gold and pink have an increase. In particular, the group pink was boosted in 2016. Furthermore, it can be observed that the percentile distributions of the individual colour groups are similar over both years. The greatest difference is a 5.58% increase for the group pink. The least ranked colour groups have a percentile distribution of less than 5 %. These are the groups silver, yellow, brown, grey, purple and orange. In this case, there is a slightly different order in the ranking in both years compared to the top-ranked colour groups.

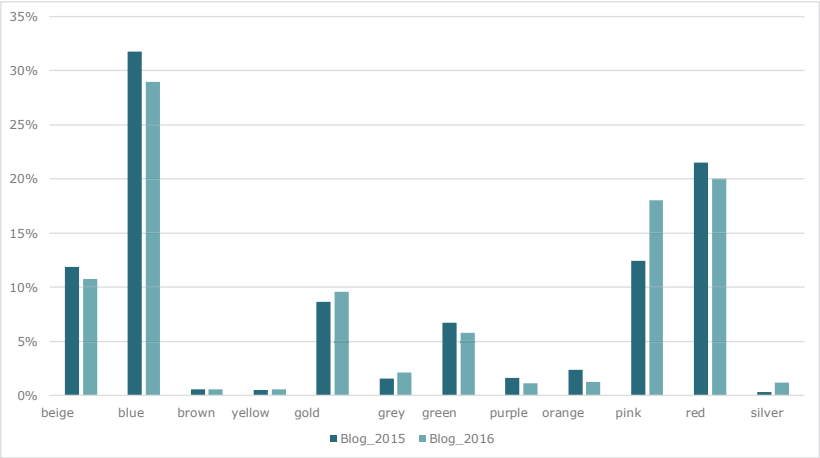


Figure 5.13.: Colour distributions comparison 2015 and 2016 yearly

Due to their low precentral distributions, the least ranked colour groups are neglected. The colours beige, blue, green and pink were focussed on the following analyses. Their monthly distribution is displayed in Figure A.13 for 2015 and Figure A.14 for 2016.

5.5.5. Evaluate Results

The results have been evaluated iteratively. In order to find the most suitable processing workflow according to the defined task (see Section 5.5.1), it was necessary to make a range of adaptions to the different configuration of some specific nodes. Table 5.11 lists the processes and their adaptions. The processes of filtering German documents, creating a dictionary, stemming and classifying colour were adapted several times and were given as feedback loops into the process. A range of iterations was required to obtain the final workflow considering all requirements for the defined task.

Table 5.11.: Feedback loops

Process	Adaptation measures
Filtering German documents	Justification of the rules for identifying German and English written parts
Creating dictionary	Revisions of the dictionary as input for the dictionary tagger
Stemming	Usage and location of the stemming functionality
Classifying colour	Classification rules for classifying colour tones into colour classifications/groups

5.5.6. Validate results

Following Figure 4.13, which illustrates the main steps in the Text Mining Layer, the last step involves validating the TM results, which can be conducted using data triangulation or involving experts. For the purposes of the research, a second data set is applied for triangulation purposes. In the last step, experts are involved in elaborating the impacts of the results for the stakeholders of a fashion and apparel SC.

Data Triangulation

For examining the potential impact of the extracted blog colour occurrences, real-world data was required. We first considered two alternatives: sales data from individual fashion and apparel companies or sales data obtained from consumer data. We selected the latter alternative since the sales data is not limited to a specific company. Therefore, we did not have to consider other variables such as the focus on a certain colour or brand. Table 5.12 exhibits the structure of the GfK data. In this structure, the input of the customer was classified according to the GfK classification scheme, in which the input of the customer can be categorized into 15 classes. From these 15 classes, the categories ”mehrfarbig, bunt”, ”schwarz” and ”weiss” were excluded from further analysis for the same reasons as for the blog data. The proportional distribution without these three categories is as displayed in Figure 5.14. The colour distributions are displayed for the whole year 2016.

Table 5.12.: Structure of GfK data set

Timestamp	Original input customer	GfK colour code	GfK colour name
01.01.2016	nude	2	beige
...	...	...	...
31.12.2016	marin	8	blue

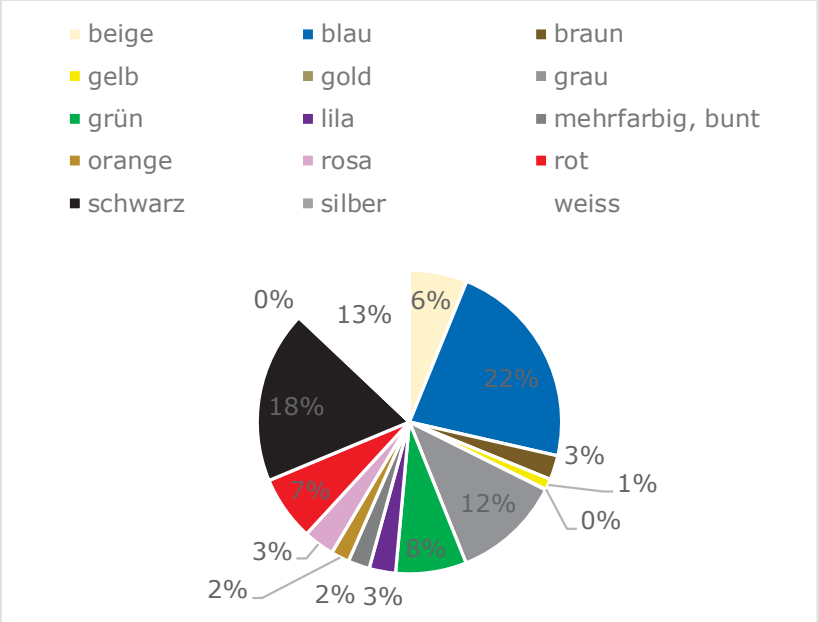


Figure 5.14.: GfK Colour Distribution 2016

We used the GfK data and contrasted it with the data extracted from the fashion blogs. Table 5.13 summarizes the statistics of both blog and GfK data. From these documents, we identified 1302 colour entries for the textual data set 2015 and a total of 1204 entries for 2016. As we also observed a large number of different colour types in the GfK data, we unified colour groups such as different types of “red" based on the GfK sales data classification scheme. This unification was then applied to both the textual and the sales data set.

	GfK data	**Blog data**	
	2016	**2015**	**2016**
Number of documents	56198	2318	1741
Number of sentences	-	31193	25905
Identified colour entries	56198	1302	1204
Different colours	14	96	88
Filtered colour types	14	14	14

Table 5.13.: GfK and Blog data

Figure 5.15 compares the colour distributions obtained from the blog data and purchase data in 2016. Few observations can be made from this figure. Firstly, while the distributions of some colour groups in the GfK data set, namely blue, brown, yellow, grey green, purple and orange, are higher than the respective distributions in the blog corpus, the distribution of the colour groups beige, gold, pink, red and silver is lower in the GfK. Secondly, for some groups, the differences within them are slightly lower (beige or blue), and in some cases, there is a higher discrepancy, for instance in the case of pink, red or grey. Thirdly, there are similar distributions in both data sets such as for beige, blue, brown, yellow, silver, orange and purple. Another noticeable observation is that the purchase of grey items are much

higher than its occurrence in the blog corpus. This might be due to the fact that greyish colours may be purchased as basic products. A similar observation can be made with the colour group gold, only the other way around, meaning that its occurrence is much higher in the blog corpus than in the purchases. Besides grey and gold, the least occurred groups in both data sets which were brown, yellow, purple, orange, and silver were removed from analysis.

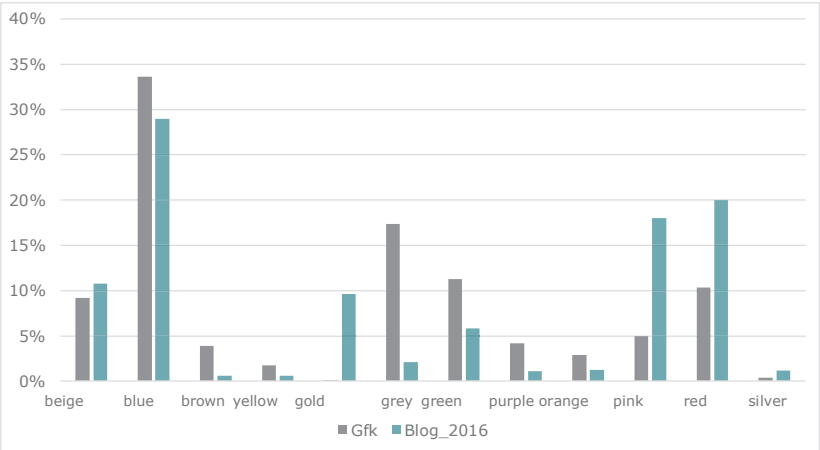


Figure 5.15.: Matching GfK and Blog datasets yearly distribution

After this analysis, the time horizon was reduced to 6 months for both datasets and the year 2015 was included. Figure 5.16 illustrates this new comparison.

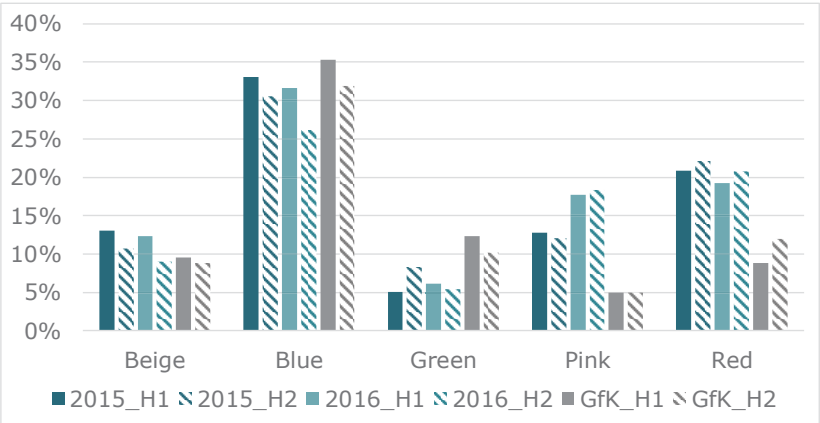


Figure 5.16.: Biannual comparison between Blogs and GfK colour distributions

From Figure 5.16, it can be observed that for some groups, the trend is similar. For instance, for blue, it can be observed that in 2016 the first six months of purchases were higher than in the second half of the year. The same trend can be stated for the blog corpus in both years, 2015 and 2016. This trend can also be observed for the colour beige, where even the decrease in the second half-year is small. The opposite direction is observable for red. The proportions for reddish colour occurrences tend to increase in the second half of the year. This trend can be observed in the purchases as well as in the blog corpora for 2015 and 2016.

For pink, the difference between the first and second half year are slightly minor in both blog corpora. In the GfK data set is exactly the same distributions. For green, a similar trend is observable between the purchases and the blog corpus of 2016.

After this descriptive analysis, we performed a correlation analysis between extracted blog colours and GfK data since the descriptive analysis of the two data sets does not illustrate respective relations. In order to examine potential relations between the extracted colours from the blogs and the purchases, we require calculating correlations for the blog corpora and the GfK data set. In order to target the time offset concern, we divided the blog data into 12-month frames, with an offset to the starting of the sales data set. Figure 5.17 portrays this approach. It is important to mention that the blog data stays unchanged; it is only the time frame which was shifted.

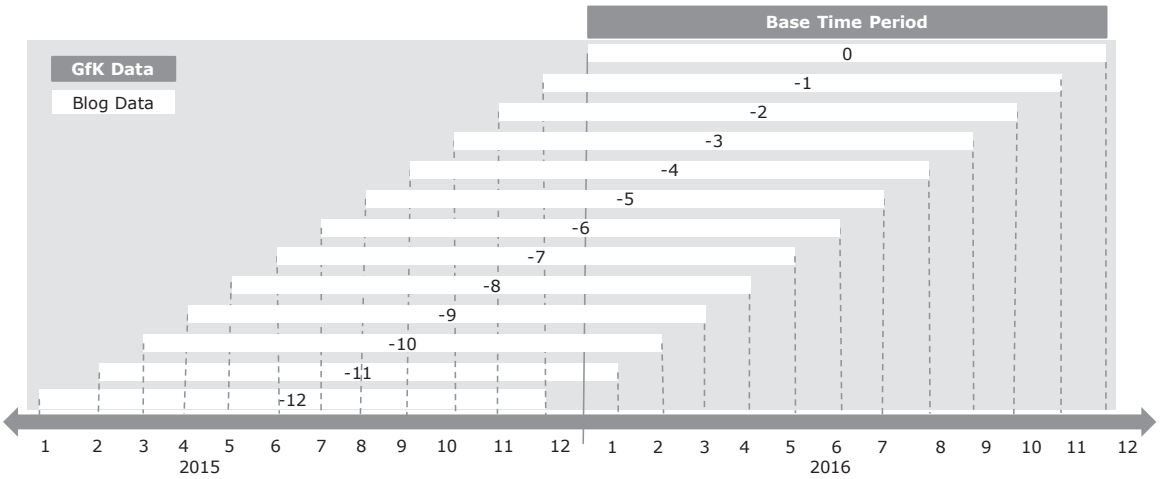


Figure 5.17.: Approach for matching GfK and blog data on a time line

For this approach, the GfK data set for 2016 and the blog corpus for 2015 and 2016 are the inputs. The base is a 12 month period ranging from January 2016 to December 2016. Therefore, the first match was conducted between the GfK data (January 2016 to December 2016) and blog data (January 2016 to December 2016), one bar displaying 0 time shift. From this point, the blog data was shifted on a monthly manner backwards. For the current analysis, the GfK data was not adapted to the equivalent time horizon and remained static. According to this setup, the examination is conducted individually for the colour groups blue, beige, green, pink and red. This approach targets the offset on a static level, and trends are therefore dismissed by this approach. It does not deal with the dynamics and volatility of fashion and apparel SCs as well as markets. In order to consider these characteristics, we generated an approach of the sales data which is given by the dynamics of the expected value and the standard deviation. In addition, we divided the sales data period into sub-periods, following the idea of the moving horizon estimator (Allgöwer et al., 1999). Depending on the trend cycle length, Shannon's Theorem (Shannon, 1948) states that the maximal length of a sampling period has to be less than half of the minimal cycle length in order to be able

to completely reconstruct the trend signal. As mentioned in Section 3.1, the change rate of fashion collections is high and in contrast, the repetition of fashion trends is slow and in the range of years.

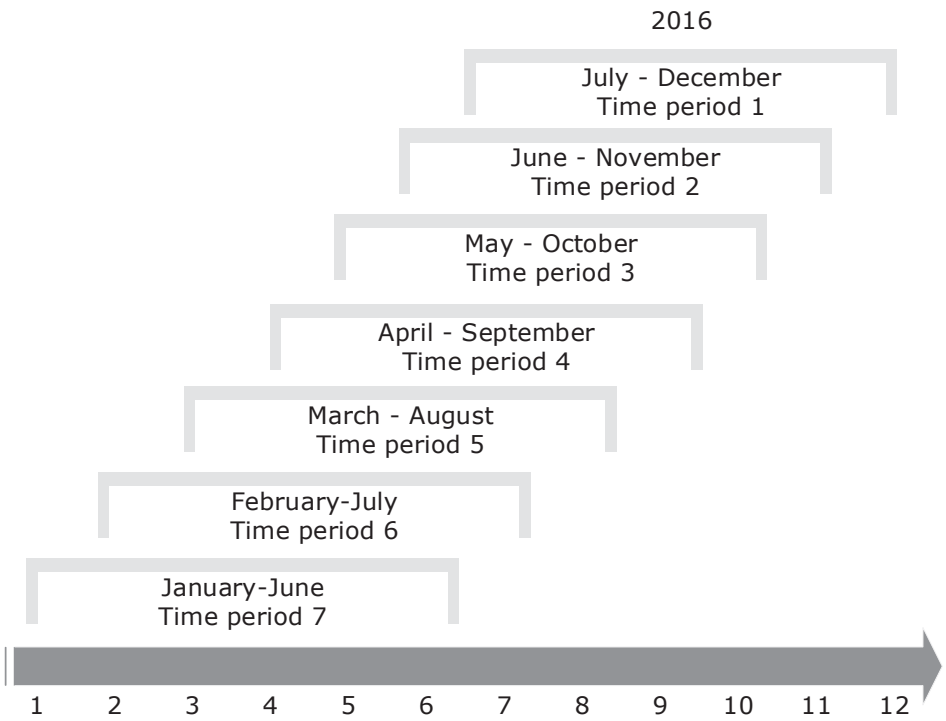


Figure 5.18.: Regarded Time Periods

Accordingly, we consider a six-month period as suitable with respect to Shannon's Theorem. Figure 5.18 displays this setting. Setting a six-month period firstly enables us to observe seasonal changes which typically happen twice a year, and secondly, to apply the law of large numbers. Alternatively, we could have used disjunct periods, but in this case, the trend information would have been lost. Furthermore, it was not clear where to set the separation points for those periods. For example, if a seasonal change occurs in March, considering the January-June period would blur the switch. The setting in Figure 5.18 was executed for the five colours by following the approach displayed in Figure 5.19. It can be seen from this figure that seven periods were applied to the Gfk data. The blog data was matched to these periods, starting at the same time, inserting one-month time shifts, and going backwards for 5 months. Within one period, only the blog data was shifted backwards. Calculating the correlations for different periods and monthly shifting is necessary due to the fast-paced and dynamic characteristics of fashion and apparel SCs. This analysis enables us to draw conclusions on how much in advance observations can be made on colour occurrences prior to purchase. As a correlation measure, Spearman's rho has been selected. Using this setup, all short, dynamic and volatile characteristic of fashion products and markets were taken.

Figure 5.19.: Framework for correlation analyses

Having defined the data and the model, our matching criterion is based on Spearman's rank correlation coefficient, which is given by

$$\rho = \frac{\mathrm{Cov}(\mathrm{rg}_X, \mathrm{rg}_Y)}{\sigma_{\mathrm{rg}_X} \sigma_{\mathrm{rg}_Y}} \tag{5.1}$$

where rg_X and rg_Y represent the rank for the textual data X and the sales data Y, $\mathrm{Cov}(\cdot, \cdot)$ denotes the covariance of the ranks and σ_{rg_X}, σ_{rg_Y} represent the standard deviations. A Spearman rank correlation coefficient close to 1 indicates that two random variables show similar trends. In contrast to that, a coefficient close to -1 indicates that no similarities exist. Evaluating this for each colour, we obtained a table of rank correlation coefficients, which is spanned by the number of subdivisions of the sales data and the number of considered time-delayed frames. To simplify the latter and reduce the impact of outliers on the result,

we introduced a quantization using two threshold values 0 and r. We then assigned -1 for all rank correlation coefficients lower than 0. For coefficients between 0 and r we assign 0, and for coefficients larger than r we assigned 1. Based on the quantization, we define the performance criterion for each colour and each time offset by the sum of the quantized rank correlation coefficients over all subdivisions k

$$\rho_{\text{total}} = \sum_{k=1}^{n} \chi(\rho_k > r) + \chi(\rho_k < 0) \tag{5.2}$$

where we use the characteristic function

$$\chi(\rho_k > r) = \begin{cases} 1 & \text{if } \rho_k > r \\ 0 & \text{else} \end{cases} \tag{5.3}$$

and indicate the time-shifted Spearman rank correlation coefficients via

$$\rho = \frac{Cov(rg_{X_k}, rg_Y)}{\sigma_{rg_{X_k}} \sigma_{rg_Y}} \tag{5.4}$$

To showcase the results and the outlined procedure, we considered the colour beige. From the performance criterion, we first obtained the Spearman rank correlation coefficients given in Table 5.14. Respective tables for the colour groups blue, green, pink and red are inserted in the appendix (see Table A.3)

Table 5.14.: Rank correlation coefficients for the colour beige

Offset	Period 1	Period 2	Period 3	Period 4	Period 5	Period 6	Period 7
0	-0.2	-0.2	-0.257	-0.485	0.028	-0.2	-0.314
-1	0.028	0.028	0.371	0.942	0.714	0.485	0.485
-2	0.314	0.314	0.542	-0.085	0.028	0.314	0.085
-3	-0.428	-0.428	-0.485	-0.428	-0.028	-0.428	-0.6
-4	0.485	0.6	0.542	0.771	0.428	0.2	0.714
-5	0.142	0.257	0.257	0.2	-0.085	0.428	0.371

Based on Table 5.14, we observed a wide variety of strong positive and strong negative correlations but no clear trend. For example, the maximal value occurs in period 4 for offset -1, yet the remaining row entries are comparably small. Utilizing the threshold $r = 0.5$, we can distinguish between strong positive and strong negative correlations and obtain a clearer picture of the quantized correlation coefficients over time, see Figure 5.20.

Parameter Optimization

Finally, we gained the best correlation between sales data and information extracted from textual data by the maximal values of the performance criterion. Since the result of the latter

is given by a table spanning the number of considered colours and the number of considered time-delayed frames, the best fit can be identified by finding the maximal entry of the table. It displays the seven periods with the corresponding time shifts. Blue highlighted cells display correlations, whereas red marked cells show no correlation.

In the next step, based on the calculated coefficients, r was classified into $-1, 0$ and 1 to identify patterns in the correlations over time period and time shift. For $r < 0$ the r's are classified as -1. This means that there is no correlation between the GfK data and the blog data. For $r > 0$, in a first step, the level is set at $r > 0,5$, tend to have a correlation, thus classified as 1. And lastly, $0 < r < 0{,}5$ classified as 0. This classification has been conducted for the five colour groups. Figure 5.20 illustrates the transferred correlations for $r > 0$ for the seven periods and corresponding time shifts for the colour group beige. In case of an existing correlation, the equivalent time shift bar is coloured in blue, otherwise in red. A positive correlation can be seen between GfK and blog data -4 months, which occurs in 5 out of 7 subperiods of the sales data. From the SC manager's perspective, this shows a positive example of both correlations of blog and sales data, which is additionally economically useful.

Figure 5.20.: Transferred correlation classification on periods and time shifts for the colour group Beige for r=0.5

This approach was also conducted for the other colour groups, as displayed in Figure 5.21. For each period, seven values per time shift can be existing. In Figure 5.21 we indicated the intensity of the positive quantized correlation by more intensely coloured boxes. Note that the intensity shows the reliability of the information. We observed that ”beige” shows a strong correlation for a time offset of -4, while for other colours no clear conclusion can be drawn.

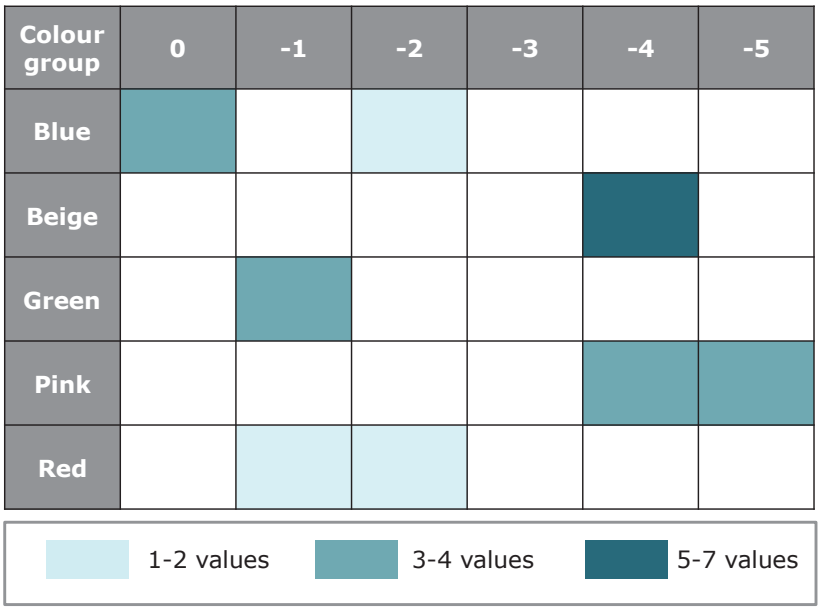


Figure 5.21.: Sum of quantized correlations for r = 0:5 over time for all colours

We set the threshold $r = 0.5$ high, but arbitrarily in order to support correlations between the sales and blog data. However, it provides useful insights from a decision support perspective. In particular, in the context of risk, the threshold provides a simple setscrew to in-/decrease the accuracy of the performance criterion as prewarn indicator. To test the dependency of the outcome on the choice of the threshold value r, we conducted a series of tests with different choices ranging from $r = 0.25$to $r = 0.5$ (see Figure 5.22). In doing so, it was possible to identify the most suitable level in the GfK dataset and the corresponding blog data in respective time shifts.

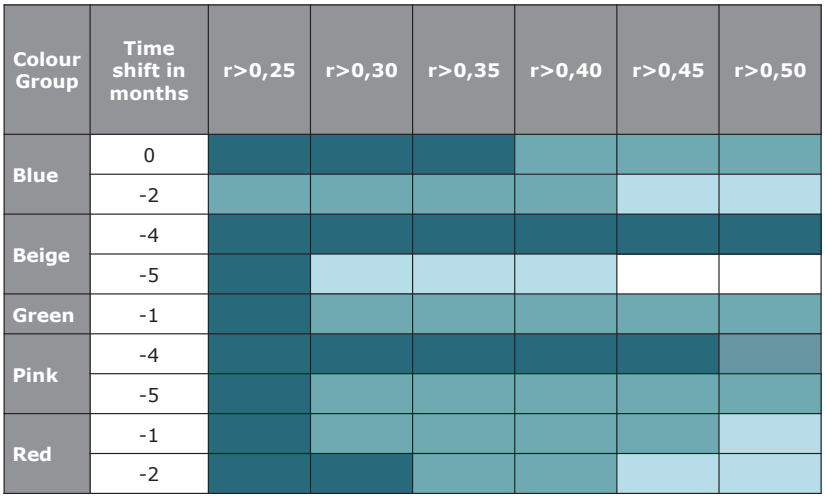


Figure 5.22.: Variations in r in the different time shifts for each colour

We observed that by reducing the threshold value r, other colour/time-shift combinations also showed high performance criteria values. For instance, in the case of beige, for $r > 0,25$ the time shifts -4 and -5 obtained a higher relevance. Transferring the results of Figure 5.22

to a fashion and apparel SC, we generated Figure 5.23. This allows us to identify time shifts for which a high-performance criterion value indicates a similar trend between the sales and the blog data.

With the notable exception of the colour ”blue”, we observed an economically advantageous time offset for the remaining colour groups. For "beige" and "pink" in particular, the offset is large enough to cover the next stage within an SC and in our case, may allow not only for operational but also tactical decision support. As displayed in Figure 5.23, the four-time shifts, 0,-1,-2 and -4 months, affect different stakeholders and processes along with the SC. Yet, lowering the threshold results in an imperfect trend correlation between the sales data and the blog data. Hence, for a decision-maker, lower thresholds represent higher risks for decisions to be taken as the accepted variability between these variables rises. Still, plotting these values over time may allow for the detection of upcoming trends at the cost of reliability.

The main question to ask as a next step is what this result implies and how this information can be used by these stakeholders in the corresponding processes in order to support their decisions. This is examined in Section 6.1.

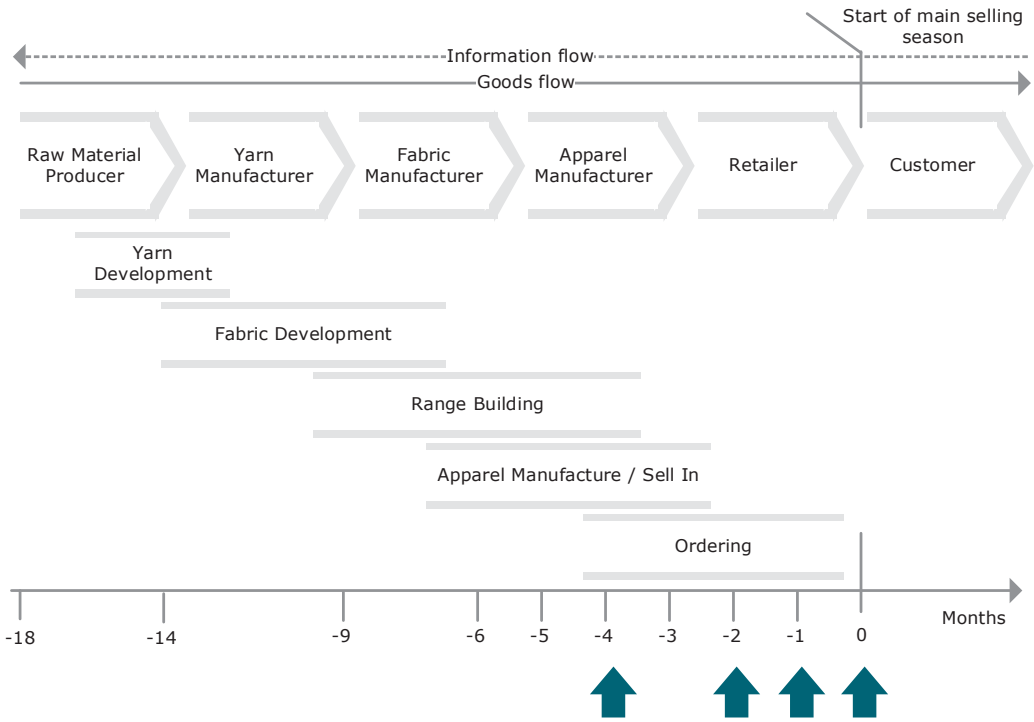


Figure 5.23.: Correlation results transferred to Fashion and Apparel SC

5.6. Summary

Following the DSRM process model, the use of the artifact was demonstrated in the preceding chapter. For this purpose, a case study was framed. For exploring the methodological potential of social media data as an additional source for SC decisions, it was necessary to

address two questions. First, it was examined if real-world sales data and data extracted from social media are positively correlated. This question only targets correlations and does not allow to conclude any economic value in terms of decision support. For this reason, it is also necessary to examine if both sources have an economically advantageous time offset. Stating these two questions, the process model was framed around women's apparel, the product feature colour, Germany, German and the year 2015-2016. Although colour is a basic feature, the availability of colour information is crucial for different fashion and apparel SC stakeholders as it impacts their decision-making processes.

The four layers of the process model were applied to the case study. First, in the Process Layer, the trend prediction was set as the decision process to focus upon which the decision makers' needs and requirements were defined, and the involved stakeholders were determined. Based on the needs and requirements, potential sources were outlined in the Information Source Layer. Typically used sources such as historical sales data, fashion shows or consumer data are listed along with social media-based sourced. The sources were classified according to IQ dimensions and to the formulated needs. Having defined social media-based sources as one relevant source type in trend predictions, fashion blogs, social networking sites, and microblogging services were considered relevant sources to include in the Social Media Layer.

In the first phase of the Social Media Layer, the channel selection was conducted on the upper level. Applying the assessment in terms of accessibility, believability, relevance and timeliness, fashion blogs were defined to be a relevant source to continue exploiting. The media level channel selection was conducted based on a fashion blog ranking. For this purpose, a relevance, believability, timeliness and accessibility assessment on the individual blogs was conducted. In the data extraction phase, the structure of a blog post was examined, and the relevant information components selected. These are URL, title, date and the message. Web Scraping was defined as the extraction strategy, and a sample corpus was generated. In the last phase of the Social Media Layer, the availability of the colour occurrences was determined.

In the first step of the Text Mining Layer, the objective of the analysis was defined. Based on the stated questions for showing the process model's use, three analysis objectives were defined: the detection of individual colour occurrences, classification of colour groups and tracking of colour occurrences. In a second step, the individual blogs are defined as source system documents, their properties are defined, the posts are crawled and loaded into the working environment KNIME. As document processing step, linguistic and technical processing is applied to the corpus. In a following step, dictionary-based keyword extraction, rule-based classification and automatic frequency analyses are applied. In the last step, data triangulation is used by applying purchasing data.

The application of the process model enables us to demonstrate its use. In doing so, we illustrated the potential of the process model for a decision-maker as we showed that blog

and sales data not only correlate but also showed an economically advantageous time offset. The observed time offset is large enough to cover the next stage in an SC. Given this, a decision-maker can use the extracted data from fashion blogs for enhancing its decision base on an operational and tactical level.

6 Evaluation and Discussion

After demonstrating the use of the process model, Chapter 6 addresses the utility of it for fashion and apparel SC stakeholders. The overall methodology of the evaluation is presented in Section 2.4. The four-steps method for DSR Evaluation Research Design by Venable et al. (2012) is followed (see Figure 2.5). In order to elaborate on this, experts have been consulted by means of qualitative surveys. The results of the evaluation are presented in Section 6.1. The presented discussion in Section 6.2 is based on the responses to the research questions.

6.1. Evaluation

Following the DSRM process model, the evaluation of an artifact aims to observe and measure the performance of the developed artifact in support of a solution to the defined problem. Recalling the entry point of this research, a problem-centered initiation addressing the uncertainty of fashion and apparel SC stakeholders in meeting customers' preferences due to the complexity of decision processes and lack of required information at the required time in the fashion and apparel SC, is followed. The objective of the process model is to provide the SC stakeholder with information relevant to its decision processes while at the same time generating an added-value. This said, the evaluation of the process model should involve the comparison of the objectives of the solution to the observed results from using the artifact in the demonstration (Peffers et al., 2007). To this end, the results obtained through the case study, and in particular the observed economically advantageous time offsets, are used as the baseline for the conduction of this evaluation. For this reason, the results are transferred back to the SC as shown in Figure 5.23.

Considering the time differentiating factor, it is possible to classify the stakeholder, according to the degree of impact through the results, into least impacted, indirectly impacted, and most impacted stakeholders (see Figure 6.1). It is notable to mention that the stakeholder and processes are schematically outlined.

In Figure 6.1 it is observed how the raw material producer and the yarn manufacturers are

classified as the least impacted stakeholders. The reason for this is that the yarn production has to be conducted more than 12 months in advance and 8-12 months prior to the selling season. This means that having the information generated from the blog up to 4 months in advance will not have a noticeable impact on their decision processes.

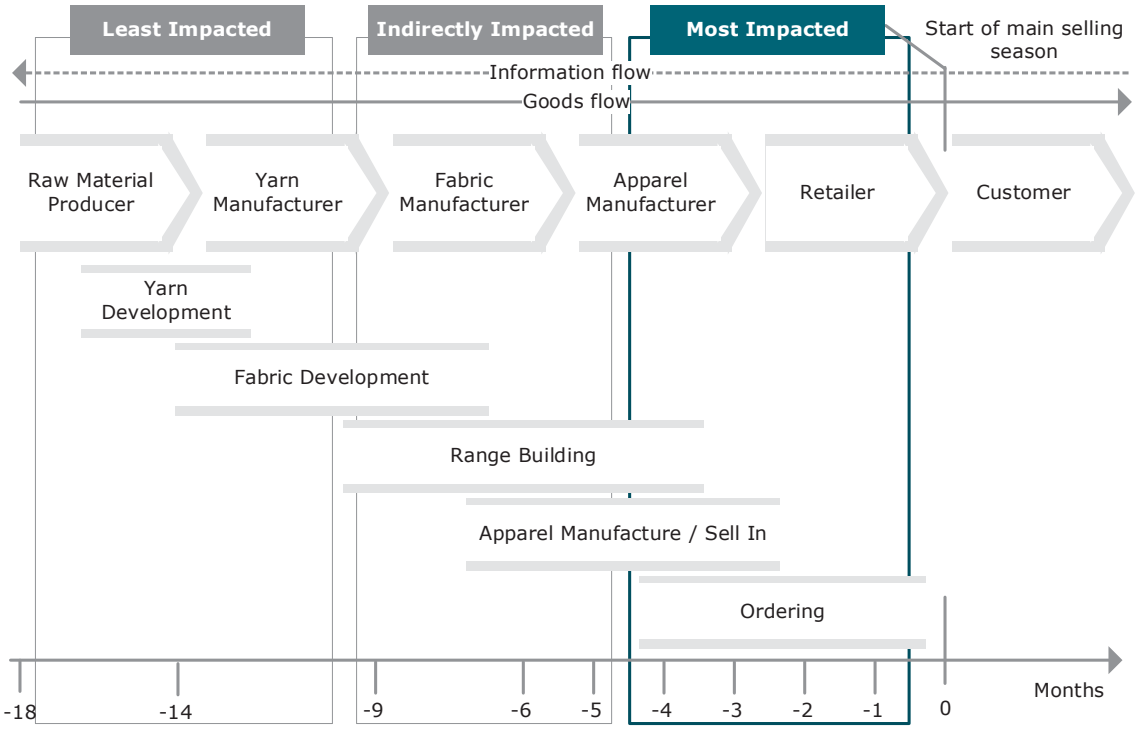


Figure 6.1.: Impacted Stakeholder

Decision processes for the product feature colour for the fabric manufacturer may be indirectly impacted by information extracted from social media through the manufacturer. However, this depends on the type of fabric and fabric manufacturer. For instance, the denim fabric manufacturer would not be impacted since denim is mostly blue. For other fabrics such as cotton, there is a larger range of colours which can be produced. Therefore, denim fabric manufacturers, for example, are less impacted compared to non-denim fabrics.

The most impacted stakeholders are the retailer and the garment manufacturers. From the economic perspective, it is the retailer who would be able to utilize the extracted information from social media for respective tasks. As the retailer and manufacturer are the most impacted, the following evaluation of the results was focused on these two stakeholders. Based on the evaluation methodology following the four-steps method for DSR Evaluation Research Design by Venable et al. (2012) established in Section 2.4, real-world users, that are decision-makers/experts working in fashion and apparel SCs, are included in the evaluation. The main requirement for the evaluation is to demonstrate the utility of the process model for fashion and apparel SC stakeholders. This is done by involving experts from manufacturers to retailers.

In the following paragraphs, we illustrate how the retailer and the manufacturer might

utilize the data from the blogs. This is done by considering the four time shifts as four different cases, and examining the processes which are meant to be conducted by the respective stakeholder.

Based on these results and Figure 6.1, it is concluded that the following three processes might be relevant to look at in more depth, namely, range building (sample collection), confirmation of orders between retailer and manufacturer.

In order to examine how the blog information can be utilized in the three processes from the manufacturer and retailer side, we conducted expert consultation. For this purpose, experts from Pakistani companies were consulted since Pakistan belongs to the largest exporter of textile products. These companies operate on an international level, working with international brands and retailers. Table 6.1 summarizes the sample.

Table 6.1.: Overview of sample

Organization	**Stakeholder role**	**Product focus**	**Number of employees**	**Position**	**Responsibilities**
Company 1 (respondent M1)	Manufacturer	Women and men	251-500	General Manager	Responsible for complete business unit's operational and financial performance, reporting to 4 shareholders of the company
Company 2 (respondent M2)	Manufacturer	Women and men	Over 500	Marketing Executive	Merchandising and Marketing tasks
Company 1 (respondent M3)	Manufacturer	Women and men	251-500	———	———
Company 3 (respondent M4)	Manufactuer	Women and men	Over 500	Account Manager	Create, manage, oversee, and execute one of the biggest UK retailer
Company 4 (respondent M5)	Manufactuer	Women and men	Over 500	Sr. Product Manager	R&D Denim Products
Company 5 (respondent R1)	Retailer	Women	Over 500	Business Head	Managing Budget & Overall Profitability of Unstitched Business
Company 6 (respondent R2)	Retailer	Women and men	Over 500	Business Head	Managing Budget & Overall Profitability of Unstitched Business
Company 7 (respondent R3)	Retailer	Women and men	Over 500	———	Marketing & Sales Manager

The following sections focus on the manufacturers and their perspectives on the potential impact on the three processes of range building, confirmation with the retailer and garment manufacturing which are presented one after another considering the four cases (0 month, 1 months, 2 months, and 4 months prior the selling season). The replies are documented by the Tables A.16, A.17, A.18, A.19, A.20, A.21 in the appendix.

Impact on decisions on Range Building (sample collection)

Sample selection or range building is the first step towards production orders placement for the upcoming season. That is why sales forecasting plays a vital role in building the base for the upcoming season sale. Within the range building process, the garment manufacturer will prepare a collection of samples for the retailer in order to have maximum chances of selection by the retailer. The processes involved in the range building are listed in Table 3.1.

Case 1: 0 Month

All the respondents agreed that in case 1, the extracted information regarding colours would have no impact on decisions for the range building process. At this time, important decisions have already been finalized (respondent M1, M2), and therefore there is no opportunity for the manufacturer to make new decisions (respondent M2). With the start of the selling season, the goods have already placed in the warehouses of the retailer (respondent M4). One respondent pointed out that the extracted information can be integrated into upcoming range building processes and future orders (respondent M2). Similarly, another respondent states that this information can be included for the next season; however, due to the pace of fashion processes, including that information will ”not be fruitful” as other trends would already have been placed (respondent M5).

Case 2: 1 Month

Similar to case 1, one month in advance will have either no (respondent M2) or only little impact (respondent M4 and M5) on decisions related to the range building process. At this time, the goods have left the manufacturer and are either in transit or have already been delivered to the retailer and stored at their warehouses (respondent M4). In this case, the information may also be integrated for upcoming orders (respondent M2).

Case 3: 2 Months

In the case of two months left prior to the start of the selling season, the experts pointed out that the extracted information will have a considerable or medium impact on decisions on range building. Respondent M1 stated that the information gathering process had been completed at this time. The information gathering process includes accessing a range of different information sources as listed in Table 5.4. Therefore, at this point, there would be some opportunity for combining the different information accessed and using it for the range building. At this time, some adjustments can take place. For instance, swatches can be compared to the extracted colour information. If applicable, styles can be replicated according to the compared swatches (respondent M2). At the same time, these adjustments may bear risk in terms of production, and these steps cannot always be dealt with (respondent M2). One respondent pointed out that at this time, the goals regarding colours are set (respondent M3). One interesting procedure was stated by another respondent. He mentioned that the extracted blog information would be reviewed against the existing samples in order to check the similarity between the extracted blog information and the existing collection. If discrepancies

exist, the retailer and buyers will be consulted to access their opinions on those differences. If the retailer is interested in adjusting their collection to the blog information, quick sampling can be performed, otherwise, the blog information will not be considered (respondent M4). Another opportunity for using this information is to share the ”samples with clients for quick delivery orders” (respondent M5).

Case 4: 4 Months

Given the case that the blog information is available four months prior to the selling season, it will have the highest impact on sample collection and range building, according to the surveyed experts. In this case, preliminary decisions on sample collection and range building are made. Decisions on colour are conducted during this phase (respondent M1). Since manufacturers are constantly seeking ”new inspirations” which can be included, a four month time period gives them enough time for discussions regarding the selection process. There would be sufficient time for the retailer to review the samples, and in general, the whole process of generating a collection will go ”in the right direction with all the flavours” (respondent M4). Particular styles from the range can be replicated or collected with the trend colour for marketing purposes (respondent M2). The extracted information may also be included in the selling plan decided on during this phase. Similarly, marketing efforts can be conducted around this period (respondent M1). Having clear information on a trend at that time is of interest to both for manufacturer and retailer. On the one hand, the manufacturer can pitch styles more easily in front of the retailer. On the other hand, retailers are more comfortable in their decision making (respondent M5).

IMPACT ON DECISION ON CONFIRMATION OF ORDERS WITH THE RETAILER

In this part, we present the results of the experts' evaluation on the impact of the extracted blog data on the process of confirming orders between manufacturer and retailer.

Case 1: 0 months

In case 1, the extracted information is available right at the start of the selling season, and all respondents agreed that in this situation, the manufacturer's decisions are not going to be impacted by the extracted information, because as stated by respondent 1, at this time contracts are already ”signed and stamped”, and ideally the ordered goods have been delivered to the retailer (respondent M4). Respondent M5 assesses, however, low to medium impact on decisions.

Case 2: 1 month

Case 2, in which the extracted information is available 1 month prior to the selling season, has little if any impact on decisions in confirming orders according to the experts (respondents M1-4). One month in advance is still too subtle time for the manufacturer to use the extracted blog information for making decisions. Similarly, as in case 1, goods are either in transit or have already been delivered to the retailer (respondent M4). At this time, the more significant part of the required processes, such as the raw materials and contracts, have already been

determined (respondent M1). Only one respondent sees a medium impact (respondent M5).

Case 3: 2 months

The level of impact on case 2 is seen differently by the experts. While for two respondents, the level of impact is high, since typical decisions to be taken during this time, are order placements, contracts drafting or deadlines agreement (respondent M1 and M5), for other experts, the level of impact is low. It is pointed out that if the extracted colour information already matches with the set production colour, an increase of the order quantity can be expected (respondent M2). Another respondent sees no general impact for the manufacturer as it is stated that orders would already be in production. However, at the same time, this respondent outlined that if the retailer asks to include the extracted blog information into the current range, then the manufacturer will be flexible at this time and provide more options, for example, the case of denim production, the manufacturer may provide different wash options according to the blog information to the retailer and renegotiate on the range. Moreover, if the fabric is available, the manufacturer may book a top-up quantity for new shades to meet the retailer's needs (respondent M4).

Case 4: 4 months

Case 4 assumes that the extracted information is available four months prior to the selling season. All experts surveyed agreed that the time assessed in this case provides an adequate period for blog information to be included in the required decisions. This means that the impact of the extracted information on decision processes related to confirming orders with the retailer is high. In this context, respondent M2 saw an impact on the potential increase of order quantities from the retailer because incorporating the extracted blog information in samples would leave a good impression on buyers, and the quantity may thus increase. Another respondent similarly stated that the extracted information should be included in the sample collection for retailer selection for ordering. If these samples are selected by the retailer, then the manufacturer not only needs to check lead times but also negotiate and agree on terms with the retailer and will proceed with fabrics and trim buying (respondent M4).

IMPACT ON DECISIONS IN GARMENT PRODUCTION

Case 1: 0 months

In terms of garment production, the experts agree that having the blog information at the start of the selling season will have no impact on decisions to be made since at this time, the goods have already been dispatched by the retailer. One respondent argues that the manufacturing facility can take steps both to become acquainted with the new trend for synchronization as well as develop the cheapest methods to materialize products. However, in terms of running orders, there would not be any impact.

Case 2: 1 month

In case 2, in which the extracted information is available one month before the selling season,

no impact will be made on decisions in garment production processes, according to the experts. Respondent M1 ,however, argues that chronic issues may have to be dealt with during this period. Another respondent sees a very low impact (respondent M5). For the other respondents, there is no impact, as ideally, the goods have already been delivered to the retailer.

Case 3: 2 months

The judgment of the experts according to the impact of extracted blog information for this period is diverse. One respondent did not see any impact. Others saw the potential of minor changes as production can be updated with insights received through the blogs (respondent M1). At this stage, consultation with the retailer may be helpful. If the retailer agrees on the adaption, the manufacturer can adjust and cooperate with the retailer. Another respondent sees a medium impact (respondent M5).

Case 4: 4 months

The highest potential impact that the extracted blog information will have on the garment production is when it is available 4 months before the selling season. As stated by respondent M1, this happens because a garment production process is preplanned several months before the delivery. If trends target embellishment and embroideries in particular, then the manufacturer will profit from the blog insights. Samples are made based on production feasibility and efficiencies. The samples will be made production friendly, and in case of foreseen problems with these samples, manufacturers will have time to make backup plans (respondent M4).

RETAILER

The following sections focus on the retailers perspective. For this purpose, responses of three experts are included. The potential impact on the three processes of range building, confirmation with the retailer and garment manufacturing which are presented one after another considering the four cases (0 month, 1 months, 2 months, and 4 months prior the selling season).

IMPACT ON DECISIONS ON RANGE BUILDING (SAMPLE COLLECTION)

Case 1: 0 months

In case 1, in which the extracted information is available by the start of the selling season, both respondents agree that no impact on decisions regarding range building processs is assigned by all respondents.

Case 2: 1 month

In the case of having the extracting information one month before the start of the selling season, two respondents do not consider any impact for this decision making processes (respondents R1 and R3).

Case 3: 2 months

Obtaining the information from the blogs 2 months prior to the selling period is seen valuable for both respondents. Respondent R1 pointed out that based on the available information, new samples can be collected. Similarly, another respondent state that the information can be used for some specific range of samples.

Case 4: 4 months

incorporating for upcoming collections (respondent R1). This contrasts the statement of respondent R2 who argues that this time frame would be too long for having an impact on decisions related to the range building process. Respondent R3 considers 4 months in advance as designers would have sufficient to build the range according to the customer requirement extracted from the blogs.

IMPACT ON DECISION ON CONFIRMATION OF ORDERS WITH THE MANUFACTURER

Case 1: 0 months

Respondent R2 does not see any impact on order-related decisions for this case. An interesting value of using this information at that stage is pointed out by respondent R3. With no time in advance to the selling season, the information can be used for store merchandising or visual outlook arrangement according to the extracted colours.

Case 2: 1 months

Respondent R2 does not see any impact on order-related decisions for this case. Having the information 1 month before the start of the selling season would allow to use for order confirmation purposes (respondent R1). Respondent R3 does not consider any impact, as orders will already be at the production stage.

Case 3: 2 months

Case 3 assumes that information extracted from social media data is available 2 months prior to the selling season. In this case, the retailer can use this information for the ordering process of placing some urgent orders (respondent R3).

Case 4: 4 months

With a 4 months time in advance on the retailer side is mainly the pre-order process that can be tackled, that is decisions which require to be made in the pre-order phase can be supported by the extracted information from the blogs. In particular, this extracted information can be used to customize samples provided by the manufacturer according to the own requirements. If the analysis, for example, shows a trend towards pinkish colour, the retailer will have the chance to adjust the suggested samples, accordingly, if they think that the selected samples do not push pinkish colours.

IMPACT ON DECISIONS IN GARMENT PRODUCTION

Case 1: 0 months

In the case 1, none of the respondents see any impact on decisions related to the garment production.

Case 2: 1 month

Respondent R1 points out that 1 month prior to the start of the selling season, the availability of the information enables to analyse data from competitive brands. The other two respondents do not assign any impact on this information for the case of 1 month availability for garment production decisions.

Case 3: 2 months

In case 3, respondent R1 argues that the production might be speeded up in order to be able to include the new obtained information in production processes of the garments.

Case 4: 4 months

The four-month time frame in advance of the start of the selling season enables the integration of the extracted information in garment production processes (respondent R1). At this stage, the garments manufacture's product development team can make a collection as the data in order to support their sales team, and the production team can align their production lines as required product (respondent R3).

6.2. DISCUSSION OF RESULTS

The discussion of the results is structured around the three sub-questions stated in Section 1.2.

PROPERTIES OF SOCIAL MEDIA AND REAL WORLD SALES DATA

The overall target of this question is to enable an understanding of social media and real-world sales data in the context of fashion and apparel SCs. Thus, RQ 1 addresses determining sufficient properties of sales data and social media for the existence of an economically advantageous setting in SCs. For this purpose, fashion and apparel SCs processes need to be analyzed, and decision processes illustrated considering the time perspective. Along with fashion and apparel SCs, different processes need to be conducted for the production of an apparel item. SC stakeholders require information at a time when the information is not available. While information on customer preferences is typically needed by the stakeholders in an early stage of the processes, this information is mostly available in a later SC stage. This situation indicates the most fundamental problem of fashion and apparel SCs.

Establishing social media as an additional information source for these decisions is only possible if information extracted from social media correlates with real-world data. The conditions necessary for using social media data as an additional information source for

fashion and apparel SCs is that real-world commercial data and social media data needs to be positively correlated. This means that the information would be available in the future. However, information that is only available in the future does not represent an economic advantage. Thus, a further condition is the existence of an economic advantageous time offset.

Focussing only the problem structure and neglecting improvement for estimators or utilization of different methods, in a second step, a parametric model for forecasting may be utilized. The model will allow the estimation/information extraction phase from the forecasting phase to be separated and can additionally be (recursively) adapted using real-time information extracted from textual data via, e.g., a Kalman filter. The identified model may then be used for forecasting. Note that the proposed quantization of the correlation may also be used for the result of the Kalman filter, which computes both expected value and covariances, yet before applying the forecasting technique. Upholding this sequence, the superposition principle between estimation and forecasting is conserved. Still, the quantization threshold may serve as a risk appetite setscrew for the decision support tool as outlined in our application example. Using this setscrew, while weak trends may be obtained, the reliability of these trends to be reflected in sales data decreases.

The latter property of a meaningful yet simple setscrew is of particular importance for practitioners. The complexity of the problem can be concealed and, yet it provides valuable insights regarding likely or unlikely trends. Consequently, textual blog data may be suitable to add on existing information extracting techniques for decision support tools.

METHOD FOR EXTRACTING INFORMATION RELEVANT TO DECISION PROCESSES FROM SOCIAL MEDIA

RQ 2 addresses the design of a method that processes relevant information from social media by the perspective of SC stakeholder. It was necessary to consider methodological implications, current decision processes of fashion and apparel SCs as well as current dealing with social media data.

METHODOLOGICAL CONSIDERATIONS

Following the *design & development* activity of the DSRM process, the focus lay on the development of an artifact. From a methodological perspective, this research puts forward the design of a process model as elaborated in Section 2.2. From the method design aspect, we target a nominal structured methodology involving characteristics of fashion and apparel SCs, of social media, and text mining process models. Based on these characteristics, requirements for the design of the process model are derived (see Section 4.1). The process model is designed in a manner that upcoming social media can be included without great effort. One characteristic feature of process models is the level of formalization. In order to increase the quality of the process model and make it accessible to tools, standardized

formalization of the process model was conducted using BPMN 2.0. For the designing of the process model at a general and at a detailed level, characteristics of fashion and apparel SCs, social media, and text mining are examined. Based on this information, requirements are derived which have to be met by the process model (see Section 4.1).

Addressing fashion and apparel supply chains

The analysis of fashion and apparel SC enables us to gain knowledge on SC decision processes and the challenges that fashion and apparel SC stakeholders are confronted with. In addition, the information needs and requirements of involved SC stakeholder, i.e. SC managers or buyers are examined (see Section 3.1). These information needs are confronted with the lack of timely available information and result in an ubiquitous uncertainty in the stakeholder's decision processes. Further information sources are consulted and decisions are often made based on the expertise and experience of the stakeholder. The involvement of the SC stakeholder's perspective in the exploitation process, is one crucial measure taken towards exploiting information relevant to decision processes. This is particularly essential because extracted information will be used for decision processes. Following the concept of fitness for use, data should be usable for the data consumer purpose (see Section 3.2.4). For ensuring that the SC stakeholder can use the extracted data for the intended decision processes, the need to consider the perspective of the stakeholder at an early stage of the exploitation process arises. Based on this, requirements are derived in Section 4.1.1.

Targeting fashion and apparel supply chain stakeholders' perspectives

These requirements address the involvement of the SC stakeholder's perspective in the exploitation processes. This implies the consideration and mapping of the targeted decision process and the associated information needs and requirements. As most fashion and apparel decision processes are complex and involve different stakeholders in the decision process, it is essential to define other stakeholders in the exploitation process. Considering information relevant to decision processes indicates that the "fitness of use" approach has to be made feasible in its application. To this end, the requirements which the data has to fulfil, need to be defined and should be made in accordance with the defined objective. For this purpose, it is suggested to use IQ dimensions for deriving requirements for a specific decision process. Having defined the required information needs, the decision-maker can be sure that the extracted information is relevant to his/her information needs, thus increasing the reliability of the extracted information. For addressing these issues, the first layer of the Process Layer is designed with the focus on SC stakeholders (see Section 4.2).

Apart from social media, a range of other information sources exists which can be used by SC stakeholders. Targeting the long-term objective of establishing social media as an additional information source for fashion and apparel SC processes, it is necessary to consider

social media not as an isolated information source. To this end, to ensure the extraction of information relevant to the defined decision processes, the sources can be further classified according to the information needs and requirements, and IQ dimensions. Having mapped the sources to the needs, one or more suitable sources can be selected for further processing. In order to conduct adequate processing measures, it is suggested to classify the selected information sources based on their type, meaning social media-based sources, structured data sources, traditional text sources and multimedia sources (see Section 4.3). This classification defines the successive processing steps. Social media-based sources continue to be processed separately. These activities are involved in the second layer of the process model, the Information Source Layer (see Section 4.3).

Including social media characteristics

This brings us to the social media aspect addressed by RQ 2. In order to extract information from social media, characteristics that come with social media should be involved. A first step towards utilizing social media data for SC stakeholders is access to it. There is not a standard approach for accessing social media data. The different approaches are outlined in Section 3.2.2. Dealing with social media data is more than just the access to it. It includes not only the collection of the data, but in particular, monitoring, analyzing, or visualizing the data. SMA has emerged as an interdisciplinary research field in the last years. It fosters these activities and deals with the development and evaluation of tools and frameworks for these tasks (see Section 3.2.2). Examining SMA frameworks such as in Stieglitz et al. (2014b), social media is not sufficiently scrutinized regarding its characteristics. Considering these, is crucial as the characteristics result in exploitation challenges (see Sections 3.2.1 and 3.2.3). To this end, as the variety of data types and channels and the veracity of social media data and exploitation challenges are required for extracting information relevant for decision processes and making it usable for the decision-maker, their consideration is inevitable.

In particular, the consideration of the veracity attribute which tackles data quality aspects, is needed due to the existence of spamming and fake content on social media. This means that a quality assessment mechanism needs to be considered at the Process Layer, in particular, because the extracted data from social media should serve as an additional information source for fashion and apparel SC decision processes. In the social media landscape, a range of social media tools exist (see Section 3.2.1). For extracting data relevant to the SC stakeholder's decision process, the defined information requirements will support the selection of suitable channels. For instance, fashion blogs may be more relevant to some decision processes while microblogs are more relevant in the case of another decision process. For this purpose, one major component of the Social Media Layer is the channel selection step (see Section 4.4.2). The design of the channel selection ensures a structural selection of relevant social media channels. In dealing with the increased challenge of fake content, the selection of the media, is designed also in a structured manner in the channel selection stage of the

Social Media Layer. For increasing the validity of the extracted information, it is suggested to apply the IQ dimensions of accessibility, accuracy, believability, relevance and timeliness in the selection process. The approach for this process is presented in Figure 4.8). A large focus is put on assessing the belivability of the data. In accordance with Shankaranarayanan et al. (2012),it is assumed that the credibility of a source has a noticeable impact when assessing the believability of data. The believability assessment is based on the source credibility; thus means, that if a given source is assessed as credible, believability is established for the published content. The measurement of believability is not directly possible because in itself it is a multidimensional construct. Based on Shankaranarayanan et al. (2012), the source credibility in the context of social media is assessed by considering the identity, reputation and expertise of the data provider, e.g. a fashion blogger (see Figure 4.9). A further measure to ensure the usefulness of the extracted information for the decision-maker is the selection of relevant information components (components which are likely to be collected) based on the relevance for the defined objective. Following this approach, only information that is likely to be relevant for the decision process is collected. In dealing with the reliability and the validity of a social media corpus, the step of generating a sample corpus is included. To ensure that the relevant information is available in the content published by the selected media, the last phase of the Social Media Layer is dedicated to checking the availability of the information needs of the SC stakeholder. Following an analytical approach, the use of WebCA approaches seems promising. According to the defined objective and data type, an adequate approach (e.g., features analysis, theme analysis) should be selected. In the case of textual data, the use of theme analysis is proposed for examining the availability of the defined information needs and requirements (see Section 4.4.4). If the availability check reveals that current sample cannot provide information mapped with the information needs and requirements, feedback loops to the selection of suitable channels is included.

Addressing the processing of textual data

A large portion of data available in social media is textual data. In contrast to structured data, textual data requires adequate processing methods to turn the data into information for decision-makers. While companies typically have strong knowledge and the necessary infrastructure to process structured data, both are often lacking when dealing with textual data. However, advances in the field of text mining have moved the processing of textual data forward. In order to meet the requirements for the processing of textual data, the fourth layer of the process model is designed.

An existing text mining process model is put forward, focussing on structured processing of textual data. Often text mining models are designed from the perspective of a specific application, such as in Ur-Rahman and Harding (2012) or Li and Liu (2012). When matching the requirements for the process model, we followed two requirements which seemed to oppose one another. First, the text mining should be designed in an application neutral way

in order to ensure that already by the design of the Text Mining Layer no manipulation in the data is provoked. Secondly, we claim that the Text Mining Layer also has to be designed in a manner which generates results relevant for the SC decision-maker. To this end, two existing process models were merged. Furthermore, to ensure systematic processing of the textual data, the Text Mining Layer should include a process model providing steps, activities for the decision-maker to follow. The generic text mining process model by Schieber and Hilbert (2014a) is included in the process model to ensure a generic approach that provides at the same time a range of detail steps. The framework of Kobayashi et al. (2018) is included to assure that the organization perspective is included in the exploitation process. In particular, the postprocessing steps of the Text Mining Layer are designed in such a way, that the stakeholders'perspective is considered. The SC decision-maker is involved in assessing the results generated by the text process model (see Section 4.5).

Mapping the SC stakeholders' perspective in each of the four layers increases the reliability of the relevance of the extracted and exploited information from social media for stakeholders' decision processes.

UTILITY OF PROCESS MODEL

While RQ 2 addresses the design and development of the process model, RQ3 aims to evaluate its utility. In this regard, the evaluation of the process model involves comparing the objectives of the solution to the actual observed results after using the process model in the demonstration. This demonstration was established by framing a case study and applying the process model to it. The case study setting is defined in Section 5.1.

Starting with the **Process Layer**, the decision process of trend prediction is put forward. For this decision process, objectives, information needs and requirements, as well as involved stakeholders are defined. The product feature colour is selected as the main information need of the decision-maker for several reasons. The colour is a basic feature of an apparel product. At the same time, colour is a main feature of fashion items (Jackson, 2007) and impacts decisions on the entire SC. As a consequence, different stakeholders are dependent on the availability of colour information for their decisions. Scully and Cobb (2012) even claim that an early definition of the right colours will "eliminate waste, save time and encourages profitable results for all along the SC" (Scully and Cobb 2012, p.30). To this end, in showing the utility of the process model for decision-makers for the product feature colour, SC decision makers are likely to have an economic added value by applying it.

Having decided on the colour, as the relevant information need for the decision process of trend predictions, different information sources which are typically consulted for obtaining information on colours are outlined. Subsequent to the **Information Source Layer** (see Section 4.3), the last step is the classification of the selected sources according to their type. Fashion blogs, social networking sites, microblogging services and multimedia platform are

outlined as potential relevant social media-based sources. However, as the focus of this thesis is the exploitation of textual social media data, multimedia platforms are not included for further processing.

As the first step in the **Social Media Layer**, preferred data types that should be collected in later stages should be defined. This is followed by the channel selection involving the application of IQ dimensions. Fashion blogs were selected as one output of the channel selection. When choosing the individual blogs, the suggested approach, which involves the IQ dimensions of relevance, believability, accessibility and timeliness, are applied. This research claims that the application of these dimensions will not only help to extract information relevant to the decision process but will also increase the reliability of the extracted information within the selection of the individual sources. Conducting the suggested approach, relevance and believability of data is established. In addition, accessibility is assured, and it is ensured that only information is included which is related to the targeted seasons, by including the timeliness dimension (see Section 5.4.2).

Following the Social Media Layer steps, after the selection of the individual blogs, based on the defined objective, relevant URL, title, date and message of a blog were selected as relevant information components.

Using KNIME and Import.io a sample corpus was generated. In the availability check phase of the Social Media Layer, theme analysis was conducted to examine if colour-related information was available in the corpus. Creating a sample corpus and examining the availability of defined information needs were measures to ensure that the extraction of information is relevant to the decision process and increases the reliability of results in an early stage of the exploitation process. Conducting both steps proved that colour the existence of colour related information.

As a next step, the **Text Mining Layer** was conducted. Running all the steps required for extracting colour occurrences, 96 different single colour occurrences for 2015 were identified for 2015 and 88 from the corpus of 2016. For examining the economic added value for SC stakeholders when using this extracted information, it aims to answer whether the time dynamics of sales data and information extracted from blogs data regarding a specific product property is positively correlated. If yes, then secondly, targeting the existence of an economically advantageous time offset between the two sources and the assessment of the respective forecast is addressed. Based on the approach outlined in Section 5.5.6, it can be observed that for the colour ”beige” a positive correlation between sales data and information extracted from the fashion blog corpus for -4 months, which occurs in 5 out of 7 subperiods of the sales data, exists. From the SC manager side, this shows a positive example of both correlations of blog and sales data, which is additionally economically beneficial. As we observed in the presented application example, it is possible to identify trends using textual data before they hit the market, i.e., information extracted from textual data shows a positive correlation with regards to sales data. This connection may be exploited economically

by considering observable trends social media as a prewarn indicator to improve prediction-based decision making within an SC. Based on the results displayed by Figure 5.21 which is based on a threshold r=0.5, a strong correlation is observed for the colour group ”beige” whereas, for the other colour groups, it can not be concluded on the correlations. Yet, the threshold r = 0.5 was chosen high but arbitrary to support a good correlation between the two data sets. Still, it provided a useful insight from a decision support point of view. In the context of forecasting risk, the threshold represents a simple setscrew to increase or decrease the accuracy of the performance criterion as a prewarn indicator. To test the dependency of the outcome on the choice of the threshold value r, we conducted a series of tests with different choices ranging from r = 0.25 to r = 0.5 (see Figure 5.22). Setting the threshold r=0.3, we can observe a respective time offset for different colour groups in 0, -1, -2 and -4 months. For these four cases, the SC stakeholder would have an economical advantage using the extracted information from the blogs. At this place, it is worthwhile noting that for increasing the reliability of the results, the data pool should be expanded by integrating further years of data into the analysis. The data pool of this research was limited to 12 months of sales data and 24 blog data. In this case, access to more sales data was not possible, as we were only able to purchase the data for 12 months. Including more data will probably also lead to an increase in the significance of the results.

Having the economic value in mind, the question remains how the respective SC stakeholder can use this information in their decision processes. For addressing this question, the potential impact of the availability of the extracted information from the blogs on decision processes of the stakeholders was surveyed by the mean of a questionnaire (see Sections 2.4 and 6.1). From the perspective of the different roles of SC stakeholders, the retailer and manufacturer seem to be most impacted by the extracted information. For this reason, experts from the manufacturer and retailer side are included in the sample for the evaluation. The results presented in Section 6.1 reveal that from the manufacturer perspective, the availability of potentially extracted information from the blogs will impact their decision processes in the case being available 4 months before the start of the selling season. This applies to all surveyed processes, the range building, order confirmation processes as well as garment production processes. The potential impact of information extracted from the blogs on the three decision processes for the four cases is illustrated in Figure 6.2. The intensity of the colour boxes represents the relevance of the assigned impact. It can be observed that the impact of information extracted from the blogs gains increased relevance for the three decision processes when being available earlier in time. Nevertheless, the utility of the process model and the added value is demonstrated as the information has already a value for the decision-makers being available one or two months before the start of the selling season.

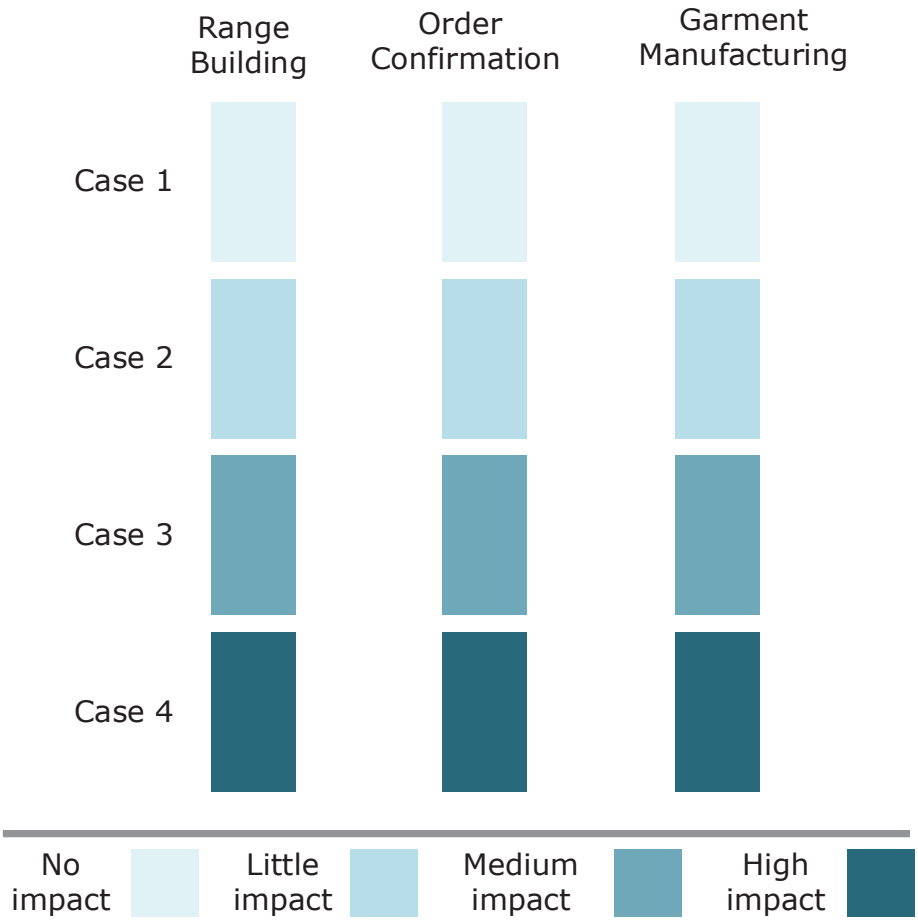


Figure 6.2.: Impact of extracted information from the blogs on decisions in range building, order confirmation and garment manufacturing

As we observed in the application example presented, it is possible to identify trends using textual data before they hit the market, i.e., information extracted from textual data shows a positive correlation with regards to sales data. This connection may be exploited economically by considering observable trends in social media as a prewarn indicator to improve prediction-based decision making within an SC.

7 Conclusion and Outlook

In this chapter a recapitulation of the presented work is provided. Firstly, a short summary of the results and the contribution of this work is presented. Secondly, an outlook and future research perspectives are suggested.

7.1. Summary of Results and Contribution

As initially stated, fashion and apparel SC stakeholders face high uncertainties in decision-making processes due to existing peculiarities of fashion and apparel SCs. In order to respond to them, a range of information sources is typically accessed for widening the decision base. Social media-based sources are considered to have a large potential in the context of fashion. Therefore, the objective of this research was the development of a methodology to exploit social media data to support fashion and apparel SC decisions taking into consideration fashion and apparel SC and social media characteristics as well as decision-makers' needs and requirements. In a number of existing approaches on the exploitation of social media, it is considered as a homogeneous information source or a specific social media channel is considered for illustrating its exploitation. These approaches are often coupled with not differentiating the diversity of social media and the characteristics of the different channels in terms of data formats, volumes, penetration rates and structure. However, for generating an added-value for decision-makers by using social media as an additional source for fashion and apparel SC decisions, it is claimed that not only the characteristics of fashion and apparel SCs but also the perspective of the decision-maker expressed by their needs and requirements must be included in social media exploitation processes in addition to social media characteristics. The importance of the veracity attribute turns, in particular, evident as the extracted information should be used in decision-making processes.

For addressing the outlined objective, this thesis is based on the DSR (Hevner et al., 2004) and applied the DSRM process model proposed by Peffers et al. (2007). The process model developed consists of four layers; the Process Layer, Information Source Layer, Social Media

Layer and Text Mining Layer. The process model was built as a nominal process, starting with the Process Layer involving the decision-makers perspective on the exploitation processes.

The decision makers' perspective is included in the process model by incorporating their information needs and requirements. The Process Layer further establishes the objective of the exploitation process and determines the involvement with other stakeholders. Subsequently, potential information sources are outlined based on the defined objectives and information needs and requirements, and relevant information components are selected based on the previously defined requirements. The Information Source Layer ends by classifying the identified sources according to their types. The exploitation is continued at the Social Media Layer by using social media based sources and enabling an understanding and exploitation of social media data by targeting a systematical selection of social media channels, examining and defining extraction strategies, and examining the availability of relevant information needs. For tackling veracity issues which impacts the quality of data and consequently the quality of decisions, the use of the IQ dimensions, that is, relevance, believability, timeliness, and accessibility, is put forward for the structural selection of the channel on medium level. In particular, an approach is introduced for assessing the believability of social media data that involves both the data provider and the data consumer, for example, an SC manager. Based on the selection of the individual medium, data extraction addresses the examination of the selected social media unit, the selection of relevant information components based on the defined information requirements, the determination of a suitable extraction strategy and generating a sample corpus. The Social Media Layer ends by examining the availability of the defined information needs. The output of this layer goes into the Text Mining Layer, which deals with the collection and exploitation of textual social media data. The design of the Text Mining Layer is based on a generic text mining process model and a text mining model originating from an organizational perspective. The involvement of the latter is crucial as we address the decision makers' perspective on fashion and apparel SCs.

The demonstration of the utility of the process model constitutes a further contribution of this thesis. To this end, a case study to which the process model is applied was elaborated. The case study was conducted as a posteriori analysis, i.e. after the point of sales. We assumed that the utility of the process model could only be shown by illustrating the added value of its use for the decision-maker, e.g. the SC manager. For this purpose, two questions were stated in the context of the case study. While the first targeted the correlations between real-world sales data and extracted data from social media-based sources, the second examined the existence of an economically advantageous time offset between these two types of information sources. Conducting experimental analyses, we observed an economically advantageous time offset for four colour groups. In particular, for "beige" and "pink", the offset is large enough to cover the next stage within an SC and, in our case, may allow not only for operational but also tactical decision support.

The evaluation was put forward aiming to compare the objective of the process model, i.e. exploiting social media data relevant to the SC stakeholders' needs, to actual observed results from the use of the artifact in the demonstration. Involving experts working in fashion and apparel SCs, enabled us to map their judgment on the utility of the process model. According to them, the highest impact is given, if information is available 4 months prior to the season since at that time, there are decisions on sample collection to be made. In the case of confirmation orders between manufacturer and retailer, the 4 months before the selling season can have a high impact as the extracted information can still be incorporated into ongoing decision processes since typically, the order process is not finalized by that time. The same expert assessment applies for decisions related to the garment manufacturing processes 4 months prior to the selling season. By showing the utility of the process model and evaluating its use, we advocate that following the designed process, information extracted from fashion blogs may be suitable to be an addition to existing information sources to support decision makers of fashion and apparel SCs.

In summary, the contribution of this research is read as follows. An artifact, defined as a process model, enabling an understanding of social media data and systematical exploitation of textual social media data to support fashion and apparel SC decisions, was designed and developed based on requirements determined in advance regarding its functionality and architecture. According to the stated objective of this thesis, the development of the process model is an expansion of purely algorithm-centered approaches often applied when dealing with social media. The utility of the articfact was demonstrated in a real-world context using a case study approach and was evaluated, including potential real-world users.

7.2. Outlook and Future Research

In different stages of this research, a range of questions and topics emerged which are worth investigating in future research.

One of these topics is the demonstration of the utility of the process model to further cases. For this purpose, the case study can be expanded addressing different target groups (male/children), product types (e.g. footwear) or product features (e.g. print, silhouette). A fashion trend is often manipulated by different product features. Therefore, it is promising to examine the existence of an economically advantageous time offset between sales and extracted social media data for more than one product feature. A similar question addresses the examination of a specific product item (e.g. dress) with a product feature (e.g.colour). Another topic is to address specific fashion segments and to compare the utility of the process model for these different segments. An interesting question in this context is if the process model generates similar results for different segments. As mentioned before, the fashion and apparel SC was schematically and exemplified sketched in this thesis without examining different SC configurations. As further work, targeting an economically advantageous time

offset between sales data and extracted information from social media, it would be worth examining the applicability and utility of the process model on different SC configurations.

A further crucial topic is the extension of the data pool to include several years of sales and blog data. The presented analysis conducted is limited to a time frame of 12 months of sales data and 24 months of blog data. It is assumed that enhancing the period will lead to an increase in the reliability of the calculated time offsets.

From an information source perspective, further social media sources apart from fashion blogs such as Instagram can be included. In addition to textual data, pictures and videos can be considered using image processing methods. For establishing increased reliability of results, information can be merged extracted from the different social media sources both from textual and multimedia data. This data can be complemented by metadata, statistics about the number of followers, comments or number of likes. Using this variety and richness of social media data in an adequate manner, increases the reliability of the extracted information. This is even more true, in case of targeting an overall data strategy and combining transactional data and other internal company data, on the one hand, with social media, web-based, or consumer data on the other hand.

From a practical point of view, one of the most evident streams is the implementation of the process model into fashion and apparel SCs as a decision support system (DSS). Typically, a range of DSS is used in the different stages of an SC by different stakeholders. These include the use of forecasting tools or textile and inspection. By setting up other case studies and experiments, it would be possible to examine the added-value for other SC stakeholders and stages than those shown in this research. Thus additional implementation options can be created.

In order to make use of such DSS applicable to a range of different SC stakeholders in the fashion and apparel SC, approaches from the field of guided analytics can be applied in the implementation. In this way, the tool is not only usable for stakeholders with a strong analytics background but can also be edited by experts with less analytics skills. This brings us to another future work that focusses on user experience and usability of such a DSS. For this purpose, the requirements of the different roles of the SC stakeholders can be mapped, as a manufacturer may have different needs in terms of usability than a retailer.

Spamming and fake content pose significant challenges for exploiting social media data. Therefore, the believability assessment approach suggested in the Social Media Layer, can be expanded. Implementing strong machine learning algorithms, will enable us to deal in advance with this critical problem and lead to more reliable and valid extracted information from social media-based sources. Therefore, it is worth examining the potential for the identification of believable data to improve results and provide decision makers with more accurate information.

When implementing the Text Mining Layer, it is necessary to consider existing legal regulations, particularly in terms of data privacy. With the General Data Protection Regulation

law, the EU has established a strong law for data protection. Any implementing approach needs to ensure ”privacy by design”. In this regard, it would be worth researching and implementing approaches compliant with data protection.

Meeting the preferences of the customer more accurately means that it is less likely that companies will outpace the market needs. Products have not to be put on reduction in order to be saleable. Therefore, overproduction and disposal of unsold inventory can be reduced by more accurate forecasts (Singh et al., 2019). With this in mind, it is worthwhile expanding this research towards fashion sustainability approaches. Though, further research is necessary to find empirical evidence that production planning using social media based sources actually produces more accurate forecasts.

Examining the transferability of the process model beyond fashion and apparel SCs, other consumer-oriented chains seem promising. This includes, for example, consumer electronics or home goods. A great amount of content is available on social media on smartphones or home decorations. Similar to apparel items, involving customer feedback in an early development stage is also highly crucial for these two markets. Therefore, it is worth examining the existence of an economically advantageous time offset of the data for both chains following the process model.

Finally, following a long-term aim to introduce social media as a complementary source of information to triangulate and assess respective risks as a pre-warning indicator in forecasting, i.e. an a priori usage of information extracted from social media data should be followed for this target. This implies the expansion of the process model for forecasting purposes.

A APPENDIX

A.1. ADDITIONAL MATERIAL CHAPTER 5

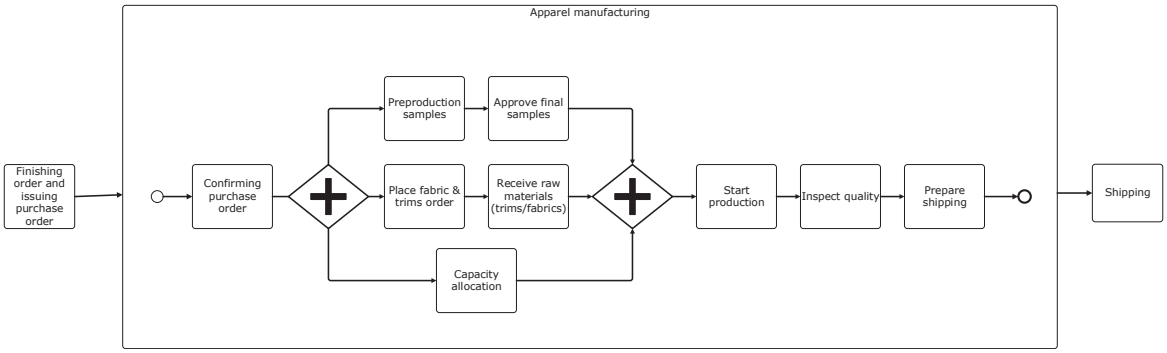


Figure A.1.: Fashion and apparel SC in BPMN 2.0 - sub-process apparel manufacturing expanded-

Blog Name	Blog URL
Fashion Passion Love	www.fashionpassionlove.de/
andysparkles	www.andysparkles.de/
bekleidet	www.bekleidet.net/
BILDHÜBSCH	www.bildhuebschfashion.com/
CATS & DOGS	www.wie-hund-und-katze.com/de/
C'est Lali	www.cestlali.com/
Cheana	www.cheana.de/
die edelfabrik	www.die-edelfabrik.blogspot.com/
Ein Zimmer voller Bilder	www.einzimmervollerbilder.com/
Euglis fabelhafte Welt	www.euglisfabelhaftewelt.de/
Fashion Kitchen	www.fashion-kitchen.com/
Fashioncarpet	www.fashiioncarpet.com/wordpress/
Fafine	www.fafine.de/
Fashionvictress	www.fashionvictress.com/
Fashionzauber	www.fashionzauber.com/
glam up your lifestyle	www.glamupyourlifestyle.com/
Glamoursisters.com	www.glamoursister.com/
Goldencherry	www.goldencherry.de/
heylilahey.	www.heylilahey.com/
how I met my outfit	www.howimetmyoutfit.de/
featheranddress.com	www.featheranddress.com/
I♡PONYS	www.iloveponysmag.com/
In love with	www.weareinlovewith.com/
Joleena	www.joleena-fashion.com/
Josie Loves	www.josieloves.de/
Kleidermädchen	www.kleidermaedchen.de/
Klischee	www.klischees.blogspot.com/
Lady Blog	www.lady-blog.de/
Lady of Style	www.ladyofstyle.com/
Leonie Loewenherz	www.leonie-loewenherz.com/
Lilly Calling	www.lillycalling.com/
LoveandFashion	www.loveandfashion.net/
Masha Sedgwick	www.masha-sedgwick.com/de/
Priscillaa.com	www.priscillaa.com/
RAUSCHgiftENGEL	www.rauschgiftengel.com/
Shades of Ivory	www.shades-of-ivory.de/
STYLE SHIVER	www.styleshiver.com/
Sue Loves NYC	www.suelovesnyc.com/
teetharejade.com	www.teetharejade.com/
Tensia Beauty	www.tensia.de/
The Random Noise	www.therandomnoise.com/
The world was a mess but...	www.wunderhaftig.net/
ToBeYoutiful	www.tobeyoutiful.com/
VITA	www.vitacorio.com/
Katolies-Welt	www.katolies-welt.de/
zauberhafte-elv.de	www.zauberhafte-elv.de/

Figure A.2.: Selected Blogs

A.2. KNIME Workflows

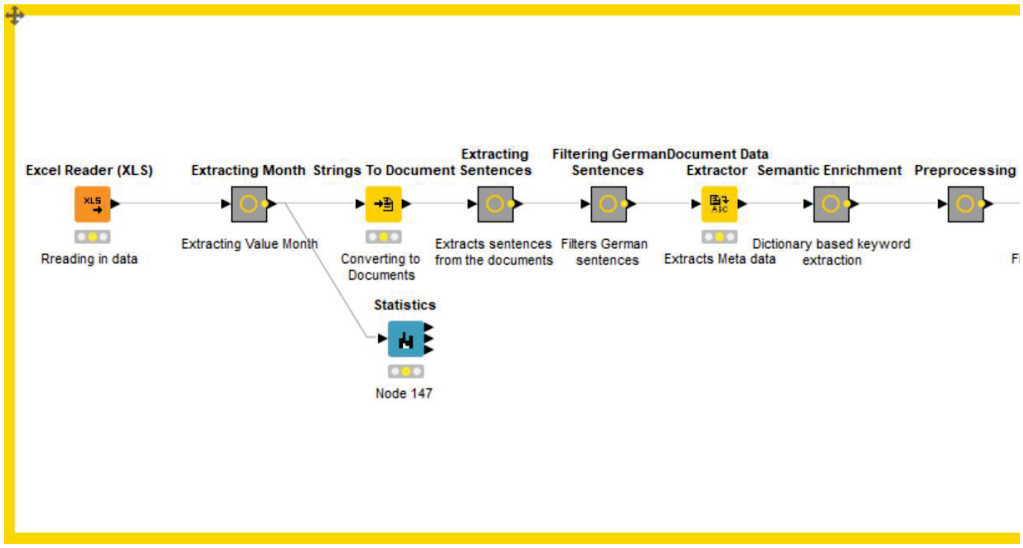


Figure A.3.: KNIME main workflow part 1

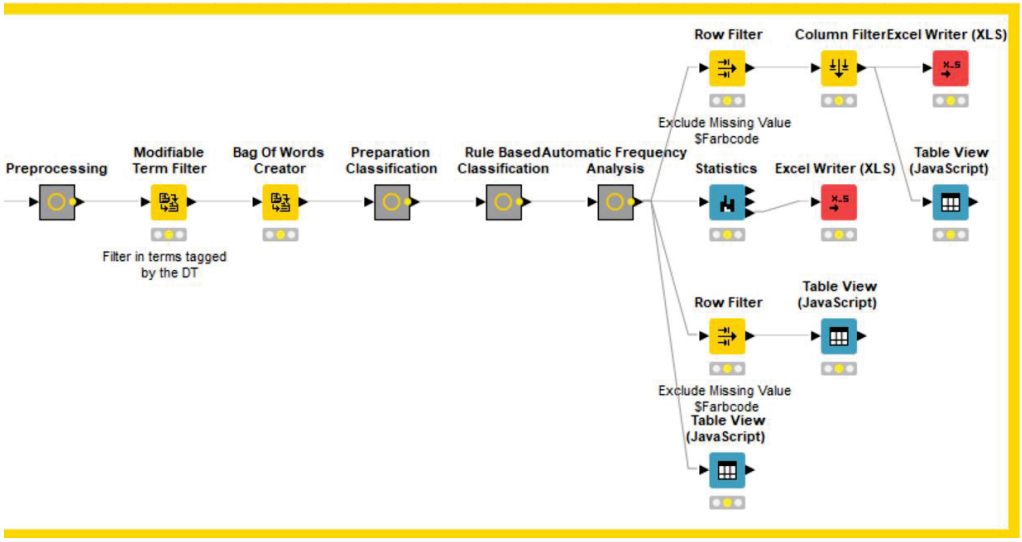


Figure A.4.: KNIME main workflow part 2

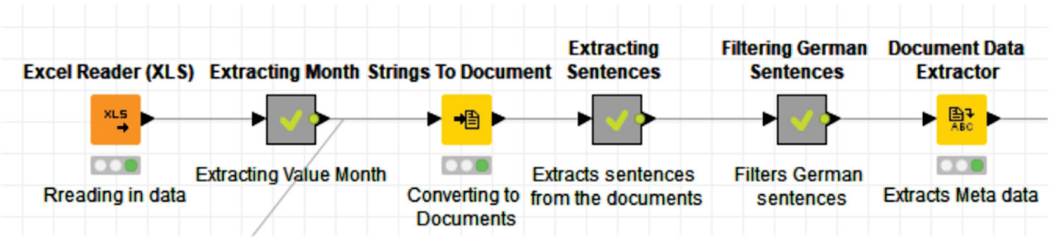


Figure A.5.: KNIME workflow for Read in blog posts

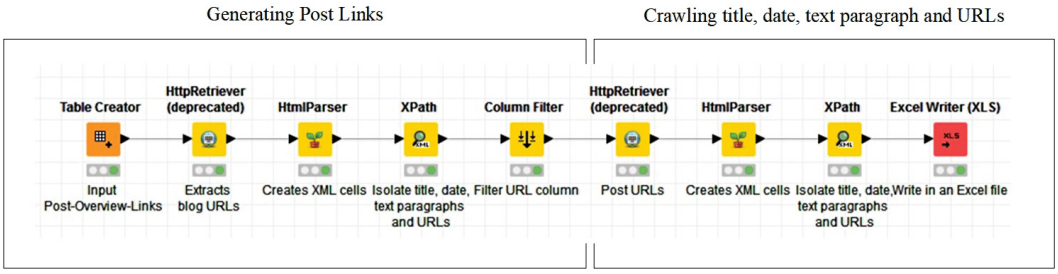


Figure A.6.: KNIME workflow for crawling date, title and message

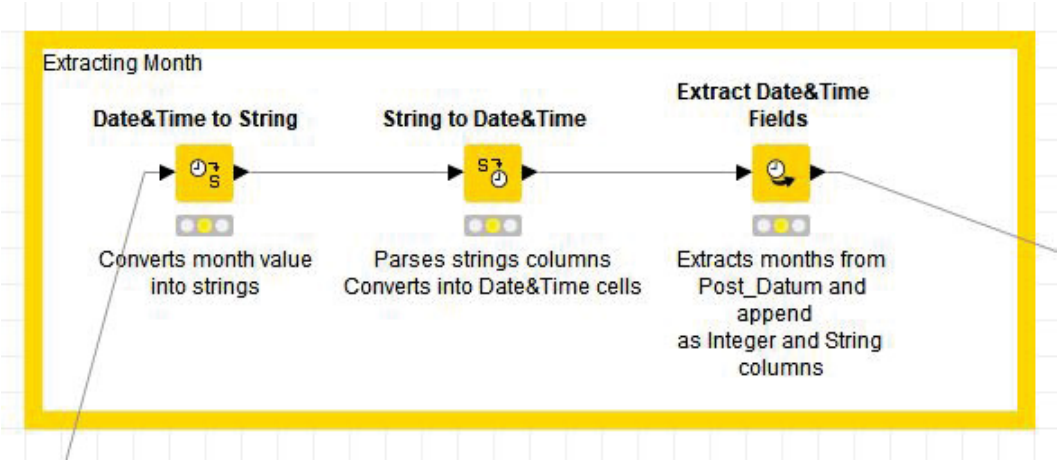


Figure A.7.: KNIME workflow for extracting month

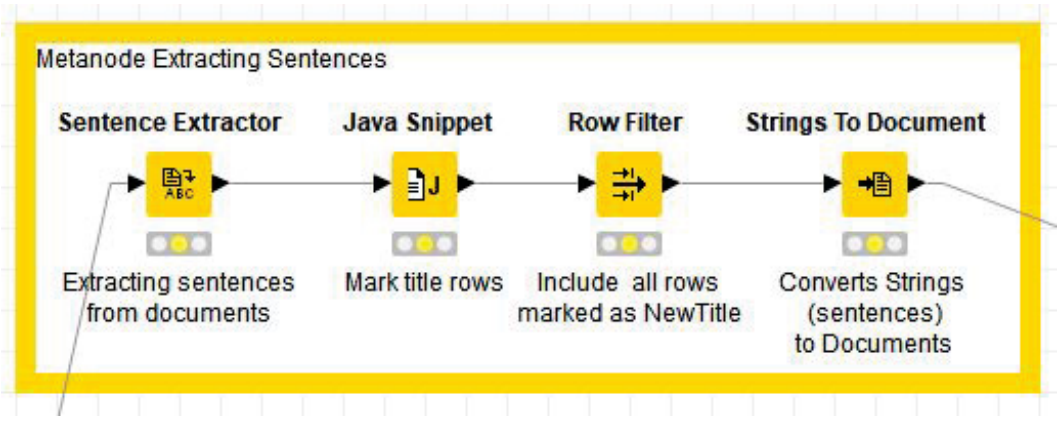


Figure A.8.: KNIME workflow for extracting sentences

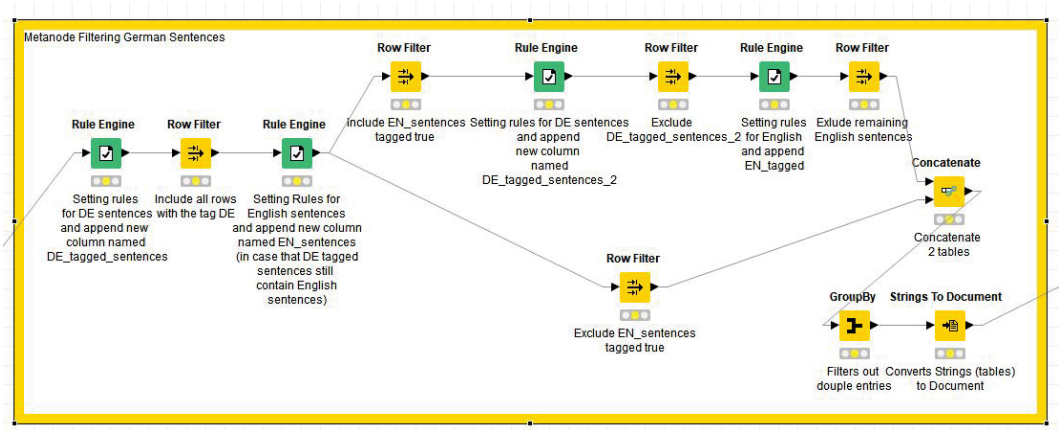


Figure A.9.: KNIME workflow for filtering sentences

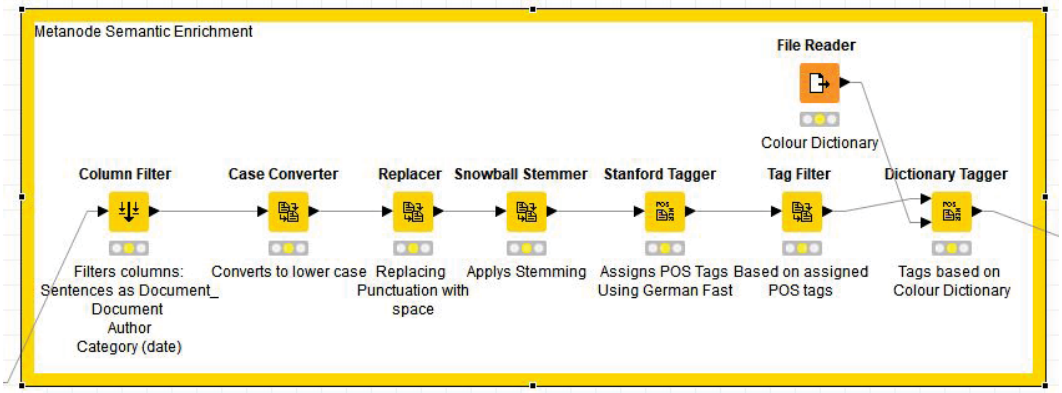


Figure A.10.: KNIME workflow for semantic enrichment

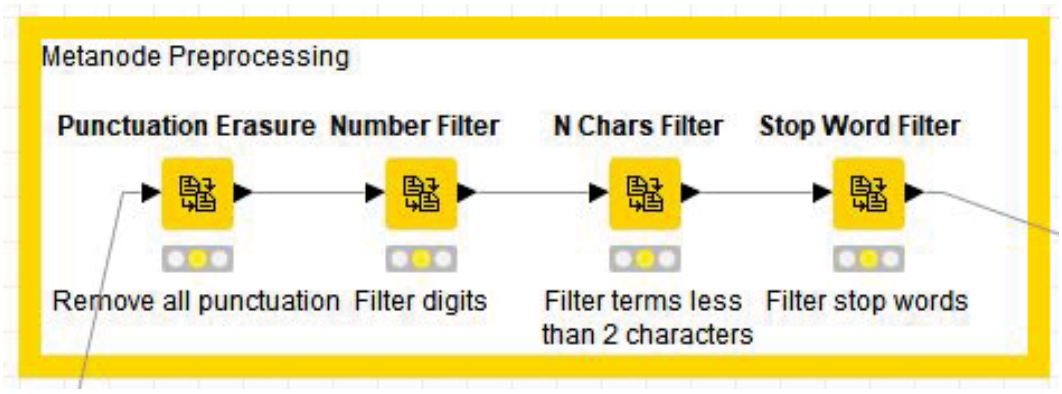


Figure A.11.: KNIME workflow preprocessing

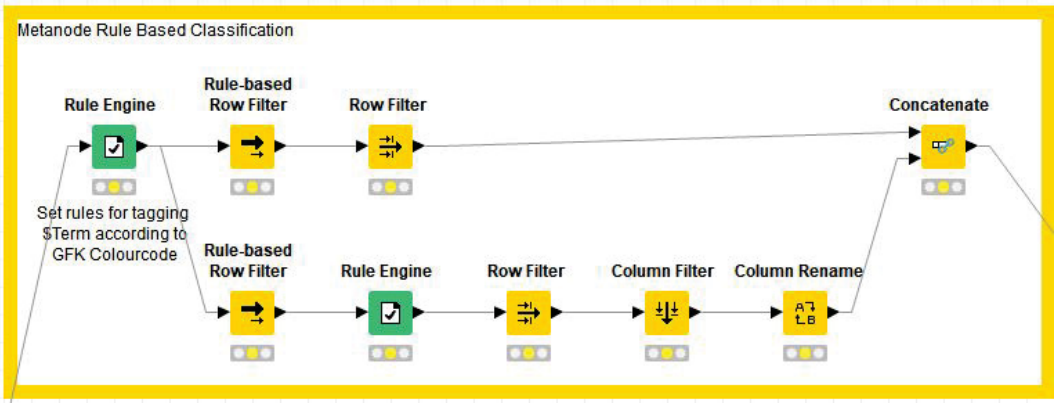


Figure A.12.: KNIME workflow for rule-based-classification

A.3. Results Text Mining Layer

Table A.1.: Monthly colour distribution 2016

Blog 2016	beige (%)	blue (%)	brown (%)	yellow (%)	gold (%)	grey (%)	green (%)	purple (%)	orange (%)	pink (%)	red (%)	silver (%)
January	13,99	41,25	0,70	0,70	5,59	0,00	5,59	1,40	0,00	15,38	15,38	0,00
February	3,63	23,74	0,72	1,44	6,47	0,72	4,32	0,72	2,18	20,85	25,90	4,32
March	19,61	29,41	0,00	0,00	7,84	1,96	3,92	0,00	2,94	20,59	12,75	0,98
April	10,34	28,25	0,00	0,00	6,90	3,45	7,59	0,00	2,07	22,76	18,62	0,00
May	7,87	34,83	0,00	2,25	7,87	1,12	8,99	0,00	2,25	11,24	22,47	1,12
June	14,55	32,73	0,00	0,00	8,18	0,91	7,27	1,82	1,82	12,73	20,00	0,00
July	9,23	30,77	2,31	1,54	6,92	3,08	0,77	0,00	0,77	23,08	21,54	0,00
August	3,67	32,11	0,00	0,00	7,34	1,83	5,50	2,75	2,75	15,60	26,61	1,83
September	19,51	26,83	0,00	0,00	14,63	7,32	6,10	6,10	0,00	4,33	14,63	0,00
October	3,33	23,33	0,00	0,00	10,83	1,67	10,53	1,67	0,83	25,83	20,83	0,00
November	8,39	27,41	1,48	0,00	14,81	2,95	5,19	0,00	0,00	14,81	22,22	2,22
December	12,61	16,22	0,00	0,00	20,72	1,80	4,50	0,90	0,00	21,62	17,12	3,60
Yearly distribution	10,74	28,98	0,57	0,57	9,61	2,12	5,80	1,13	1,27	18,02	20,00	1,20

Table A.2.: Monthly colour distribution 2015 (without black and white)

Blog 2015	beige (%)	blue (%)	brown (%)	yellow (%)	gold (%)	grey (%)	green (%)	purple (%)	orange (%)	pink (%)	red (%)	silver (%)
January	13,08	23,04	0,00	0,00	3,41	3,74	8,41	2,80	0,00	7,48	23,04	0,00
February	15,58	27,92	0,00	0,65	5,84	0,65	1,95	1,30	1,30	14,94	29,87	0,00
March	15,86	27,59	0,00	2,07	7,59	2,07	3,45	0,69	6,90	15,36	17,93	0,00
April	14,56	40,78	0,00	1,94	6,80	1,94	4,85	4,85	3,88	11,36	8,74	0,00
May	13,51	33,33	0,90	0,00	12,61	1,80	6,31	2,70	0,00	11,71	17,12	0,00
June	4,39	44,74	0,00	1,75	3,51	0,88	7,02	0,00	4,39	13,16	20,18	0,00
July	3,54	26,55	0,00	0,00	13,27	1,77	9,73	0,00	5,31	12,39	26,55	0,88
August	10,93	30,05	1,16	0,00	10,93	0,00	6,94	3,47	3,47	15,61	16,18	1,16
September	10,00	37,86	0,00	0,00	8,11	2,70	6,76	0,71	0,00	10,00	25,00	0,00
October	10,14	33,11	1,35	0,00	8,11	2,70	6,76	0,00	1,35	10,81	25,00	0,68
November	15,67	30,60	2,99	0,00	11,94	0,75	8,21	2,24	0,75	9,70	16,42	0,75
December	13,86	21,78	0,00	0,00	10,89	1,98	3,91	0,99	0,99	13,86	26,73	0,00
Yearly distribution	11,85	31,76	0,58	0,52	8,68	1,56	6,74	1,62	2,40	12,44	21,52	0,32

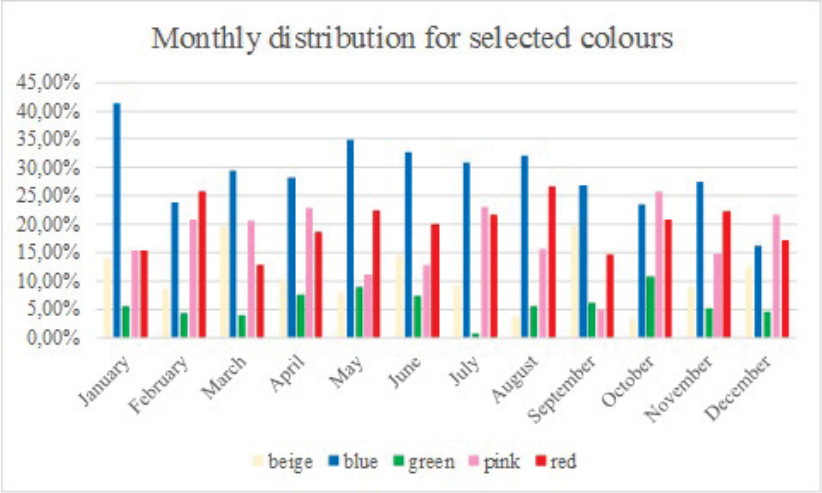


Figure A.13.: Extracted colours 2016 - monthly distribution - (without black-white)

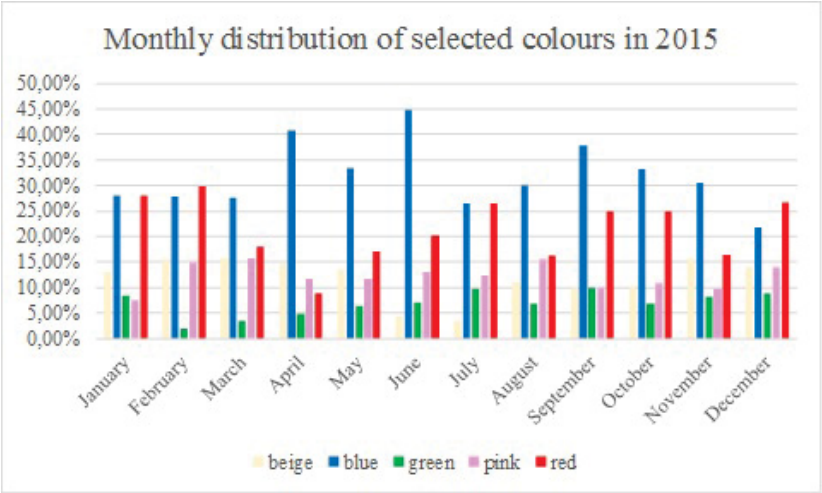


Figure A.14.: Extracted colours 2015 - monthly distribution - (without black-white)

Table A.3.: Rank correlation coefficients

(a) For colour blue

Offset	Period 1	Period 2	Period 3	Period 4	Period 5	Period 6	Period 7
0	0,714285714	0,942857143	1	0,828571429	0,142857143	0,371428571	-0,028571429
-1	0,771428571	0,771428571	0,257142857	-0,371428571	-0,657142857	-0,714285714	-0,085714286
-2	0,542857143	0,142857143	-0,485714286	-0,6	0,085714286	0,428571429	0,428571429
-3	0,371428571	-0,085714286	-0,485714286	-0,314285714	-0,6	-0,6	-0,828571429
-4	-0,6	-0,828571429	-0,085714286	-0,085714286	0,428571429	0,257142857	-0,085714286
-5	-0,828571429	0,028571429	-0,257142857	0,085714286	0,314285714	0,028571429	0,028571429

(b) For colour green

Offset	Period 1	Period 2	Period 3	Period 4	Period 5	Period 6	Period 7
0	- 0,2571429	-0,2571429	-0,1428571	0,37142857	0,2	0,2	0,37142857
-1	0,08571429	0,08571429	0,77142857	0,82857143	1	0,82857143	0,25714286
-2	-0,0285714	0,77142857	0,42857143	0,48571429	0,14285714	-0,1428571	-0,5428571
-3	0,77142857	0,37142857	0,02857143	-0,7142857	-0,8285714	-0,7142857	-0,6
-4	0,08571429	0,0857143	-0,6571429	0,7714286	-0,6571429	-0,6571429	0,8285714
-5	-0,0285714	-0,7142857	-0,3142857	0,2	0,31428571	-0,5428571	-0,4285714

(c) For colour pink

Offset	Period 1	Period 2	Period 3	Period 4	Period 5	Period 6	Period 7
0	0,2	0,37142857	0,65714286	0,42857143	0,25714286	0,25714286	-0,3142857
-1	0,08571429	0,08571429	-0,0285714	-0,1428571	-0,1428571	-0,7714286	-0,7714286
-2	-0,8285714	-0,5428571	-0,8285714	-0,7142857	-0,6571429	-0,4285714	-0,4285714
-3	0,42857143	0,02857143	0,08571429	0,31428571	0,14285714	0,31428571	-0,4285714
-4	0,6	0,48571429	0,77142857	0,77142857	0,65714286	0,48571429	-0,1428571
-5	-0,4285714	0,14285714	0,08571429	0,42857143	0,94285714	0,77142857	0,77142857

(d) For colour red

Offset	Period 1	Period 2	Period 3	Period 4	Period 5	Period 6	Period 7
0	-0,6571429	-0,1428571	-0,3142857	-0,2	0,25714286	0,14285714	-0,2571429
-1	0,25714286	-0,2571429	-0,0285714	0,77142857	0,48571429	0,25714286	0,54285714
-2	0,14285714	0,42857143	0,82857143	0,25714286	0,31428571	0,48571429	0,31428571
-3	-0,3142857	0,54285714	-0,0285714	0,02857143	0,08571429	0,08571429	0,08571429
-4	0,82857143	-0,0285714	0,25714286	-0,0857143	-0,2	-0,3714286	0,0579771
-5	-0,02857143	0,25714286	-0,3142857	-0,3714286	-0,4857143	0,0579771	-0,3478626

A.4. Questionnaire

Social Media Data for supporting decisions in the Fashion and Apparel Supply Chain

PART 1: Demographic Information

Please indicate your background:

Organization: ___Manufacturer ___Retailer

Focus of Products: ___Woman ___Men ___Children

Number of employee: ___Over 500 ___251-500 ___51-250 ___11-50 ___Less than 10

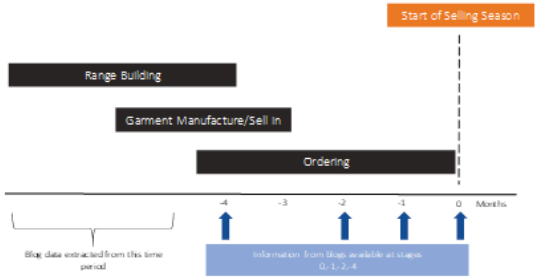
Name of company, name, individual position and main responsibilities:

PART 2: Description of research

For the purpose of the research on the exploitation of textual data from social media for decision support for the fashion and apparel supply chain, fashion blogs have been analysed. Colour occurrences are extracted from the blogs and matched to real world purchasing information for women's wear.

PART 3: Description of Scenario

Imagine the analyses of the blogs reveal that a certain colour trend being emerged in the last months in the blogs. For instance, pink colours have been emerged increasingly. This information is available 0 months, 1 months, 2 months and 4 months prior to the start of selling season (see figure).



Having the information on this particular colour trend from the blogs, how would you utilize this information for decisions in the range building (sample collection), confirmation of orders with the retailer/manufacturer, production processes, and in further processes which you think this information may have an impact. In cases that you do not see impacts, please also indicate.

Time of availability of blog insights before start of selling season	Impact on decisions in Range Building (sample collection)
0 month	
1 months	
2 months	
4 months	

Time of availability of blog insights before start of selling season	Impact on decisions in confirmation of orders with the retailer/manufacturer
0 month	
1 months	
2 months	
4 months	

Time of availability of blog insights before start of selling season	Impact on decisions in garment production processes
0 month	
1 months	
2 months	
4 months	

Figure A.15.: Questionnaire

	Manufacturer				
Time of availability of blog insights before start of selling season	**M 1**	**M 2**	**M 3**	**M 4**	**M 5**
0 month	At this time there would be no impact. As this time represents when the whole sample collection would have been made and important decisions regarding range building would have been finalised	Its already too late to adapt to a new decision about the product aspects, since all the decisions have been already taken for manufacturing. But definitely the trend collection can be summed up in the new range for the upcoming inquiries.	We started reading blogs and social websites to check the trend of styling and colouring of garments and it gave us some idea in this regard	As a manufacturer, this will have no impact because by this time goods will be at warehouses of retailer.	Very low impact. That particular trend can be tried in next season but considering the pace of the industry it won't be fruitful as there would be another trend ready.
1 months	Impact would be little as these will mark the final stages of sample collection owing to decisions made in previous months	Same as above	During 2nd month our study was further imp[roved and we helped in building our database more efficiently	As a manufacturer, this will also have no impact because by this time, goods would be in transit or delivered at warehouses.	Lower impact, as timeline to manufacture the order before selling period would be very small.
2 months	Considerable impact, at this point in time we would have completed information gathering required to build the sample collection. Range building would also be started during this time	Swatches can be shared against trend shades information and confirmation in order to replicate particular styles. Though it is huge risk for such productions and it is not always possible to handle such bold steps.	During this month we were in a position to set our goals regrading the colour and styling generation which we will be working and which we will target	Firstly, we will check the authenticity of blogs and the blogger then we will definitely review the collection we have already made to check if we have anything similar to the blog insights. If don't then will share the same with retailers/buyers to check their interest for quick sampling if required.	Medium impact, Samples can be shared with client for quick delivery orders.
4 months	Most important phase. During this phase all preliminary decisions will be taken on what range and sample collection will be made, what colors it will be made in, styles . Selling plan and marketing efforts will also be decided during this phase	Assurance of quality goods supplies against the trend requirement using the cheapest resources. Particular styles from the range can be replicated or collected with the trend color for marketing purpose.	In this month we finalized our study and defined our targets on which our whole team was on line and we started working on it	As manufacture we always look forward for new inspirations, buyers shared trend boards and blogs. 4 months will give us a lot of time for selections, discussions, buyers need review, and making collection in right direction with all the flavours.	Highest impact, information shared in this timeline will help in directing the coming season's collection. With a clear trend information, pitching styles to customers is relatively easier and client would also be comfortable in decision making.

Figure A.16.: Overview of manufacturer's replies- range building-

	Manufacturer				
Time of availability of blog insights before start of selling season	**M 1**	**M2**	**M 3**	**M4**	**M5**
0 month	No Impact. Everything required would be in house by now and contracts would be signed and stamped.	No impact	We started our working regarding styling so at this stage we were not targeting customers as we have to set our goals in	As manufacturer, no impact – goods would be delivered to retailer.	Low to medium impact
1 months	Little impact, most of the things required including raw materials and contracts would be already finalised or in the final phases.	No impact	In this month we took help from previous working and started targeting customers on the basis of our vague assumption / hypothesis	As manufacturer, no impact – as goods would either be in transit or delivered.	Medium impact
2 months	Most important phase. Order placements, contracts drafting, deadlines agreement, all will be done during this phase.	Generally, no impact but in case the trend color is same as of production then an added order quantity can be expected.	In this stage, our previous working helped us a lot as we already have a baseline for our garments styling so we started target customers on this basis.	As manufacturer, no impact as by this time orders would be in production. Yes, if retailer need that particular trend so bad then we can always try to be flexible and come with more options. e.g. if sky blue is in and there is mid blue denim article in production we can send couple of wash options and agree the lighter shade. Or if fabric is available, we will try to book top up quantity for new shade to manage the needs of retailer	High impact
4 months	Considerable impact as this will be the planning phase for orders placement with manufacturers/retailers	The addition of trend colors can leave a good impression on buyer and there is a good probability of change in order quantities in a positive manner.	At this stage our previous efforts helped us a lot and we succeeded in capturing customers in styling and fashion which agreed and locked at 1st stage	If the blog insights are available 4 months before then definitely we would add this in sample collection for retailer selection for ordering. If selected, we will check the lead times and after agreeing terms with retailer, will proceed fabric and trims buying.	High impact

Figure A.17.: Overview of manufacturer's replies-order confirmation-

	Manufacturer				
Time of availability of blog insights before start of selling season	**M1**	**M2**	**M3**	**M 4**	**M 5**
0 month	Almost zero impact	No impact on running orders as far as manufacturing is concerned. Though the manufacturing facility can take steps to get acquainted with the new trend for synchronization and develop cheapest/safest methods to materialize product.	We already have set our lines with basic to semi fashion operators. So we already were prepared for 80% of styling and orders. So the orders which we locked and finalized, we already were ready for this.	No impact as manufacturer goods would have been dispatched to retailer by then.	No impact
1 months	Little or no impact, only chronic issues will be addressed during this phase.	No impact		No impact as manufacturer goods would have been dispatched to retailer by then.	Very low impact
2 months	Considerable importance. Minor changes can be accommodated during this phase and production can be updated with important insights received through blogs.	No impact		As mentioned above, if retailer agrees to change the wash level then we will definitely try to adjust and cooperate.	Medium impact
4 months	Most important phase as garment production process is pre-planned, mostly a couple of months ago before delivery.	No impact		This is will be real helpful, for example if the new trend is something like embellishment or embroideries then we would make samples keeping in view our production feasibility and efficiencies. We will try to make samples production friendly but if we would feel any limitation then we will have meetings to foresee problem if we get orders and try to have back up plans	High impact

Figure A.18.: Overview of manufacturer's replies-garment production-

	Retailer		
Time of availability of blog insights before start of selling season	**R1**	**R2**	**R3**
0 month	Will not impact my decision.	Cannot make a decision	This will not impact as the range can't be built in such short period of time.
1 months	Can only analyse data of competitive brand	Cannot make a decision	Same as above
2 months	Will start collecting new samples based on available data	Right time	Here it can be used for some specific range of samples.
4 months	Will incorporate the information in upcoming collection	Too long	There is the most important time for availability of such information as this provides adequate for the designer to build the range according to the customer requirement extracted from the blogs.

Figure A.19.: Overview of retailers' replies- range building-

	Retailer		
Time of availability of blog insights before start of selling season	**R1**	**R2**	**R3**
0 month	Less number of orders expected	No impact	At this stage it can only be used by the for order confirmation rather such information at this stage can only b sued for store merchandising or visual outlook arrangement according to the colours .
1 months	Less leadtime will be allowed for order	No impact	No impact as order will be at production stages.
2 months	Pressure situation to meet the new order timelines	Three months	At this stage it can only be utilised for placement of some urgent orders.
4 months	Will be incorporated in the current orders	Too late	Again at this stage this will be useful at this stage as it will provide enough time to decide the which styles and colour need to be produced for next season.

Figure A.20.: Overview of retailers' replies -order confirmation-

	Retailer		
Time of availability of blog insights before start of selling season	**R1**	**R2**	**R3**
0 month	Will not impact my decision	Garment production should not be impacted by these decisions.	No impact as production general garments production is 90 -120 days
1 months	Can only analyse data of competitive brand and can go into product development	Garment production should not be impacted by these decisions.	No impact
2 months	Will speed up the production process to be able to make product with new information	Garment production should not be impacted by these decisions.	No impact
4 months	Will incorporate the new available information into production process	Garment production should not be impacted by these decisions.	At this stage the garments manufacture product development team can make a collection as per the data in order to support their sales team and the production team can align their production lines as per required product.

Figure A.21.: Overview of retailers' replies -garment production-

B PUBLICATIONS

The following list gives an overview of related publications sorted chronologically :

1. Beheshti-Kashi, S. and Kinra, A. (2020). Process Model for the Selection of Social Media Channels by Means of Information Quality Assessment of Fashion Blogs for Apparel Supply Chains. In: Proceedings of the 2020 CERR Conference, Valencia, Spain, pp. 143-159.

2. Beheshti-Kashi, S. (2020). Development of a social media process model for fashion and apparel supply chain decisions. Cyber-Physical Systems, 6(2):76–95.

3. Beheshti-Kashi, S., Pannek, J., and Kinra, A. (2019). Complementing decision support and forecasting risk in supply chain with unstructured data. IFAC-PapersOnLine, 52(13), pp. 1721 – 1726. 9th IFAC Conference on Manufacturing Modelling, Management and Control MIM 2019.

4. Beheshti-Kashi, S., Lütjen, M., and Thoben, K.-D. (2018). Social Media Analytics for Decision Support in Fashion Buying Processes. In Artificial Intelligence for Fashion Industry in the Big Data Era, Springer, pp. 71–93.

5. Beheshti-Kashi, S.; Hribernik, K. A.; Lützenberger, J.; Arabsolgar, D.; Thoben, K.-D. (2017). Fashion Supply Chains and Social Media: Examining potential relationships of social media text data and decision making processes in fashion supply chains. In: Bandinelli, R.; Rinaldi, R. (Eds.): Business Models and ICT Technologies for the Fashion Supply Chain; Proceedings of the IT4Fashion 2016; Springer International Publishing, pp. 271-281.

6. Beheshti-Kashi, S. (2017). Fashion Trend Concept and its Applicability to Fashion Markets and Supply Chains In: Freitag, M.; Kotzab, H.; Pannek, J. (eds.): Dynamics in Logistics. Proceedings of the 5th International Conference LDIC, 2016 Bremen, Germany. Springer International Publishing, 2017, pp. 163-172

7. Beheshti-Kashi, S.; Thoben, K.-D. (2016). The Usage of Social Media Text Data for the Demand Forecasting in the Fashion Industry In: Kotzab, H.; Pannek, J.; Thoben, K.-D. (eds.): Dynamics in Logistics. Proceedings of the 4th International Conference LDIC, 2014 Bremen, Germany. Springer International Publishing, 2016, pp. 723-727.

8. Teucke, M.; Ait Alla, A.; El-Berishy, N.; Beheshti-Kashi, S.; Lütjen, M. (2016). Forecasting of Seasonal Apparel Products In: Kotzab, H.; Pannek, J.; Thoben, K.-D. (eds.): Dynamics in Logistics. Proceedings of the 4th International Conference LDIC, 2014 Bremen, Germany. Springer International Publishing, 2016, pp. 633-642

9. Beheshti-Kashi, S., Lütjen, M., Stoever, L., and Thoben, K.-D. (2015). TrendFashion – A Framework for the Identification of Fashion Trends. In Cunningham, D., Hofstedt, P., Meer, K., Schmitt, I., editor, INFORMATIK 2015, Lecture Notes in Informatics (LNI), Bonn 2015.

10. Beheshti-Kashi, S., Karimi H.R., Thoben K.D., Lütjen, M. & Teucke, M. (2015). A survey on retail sales forecasting and prediction in fashion markets, Systems Science & Control Engineering, 3:1, 154-161.

11. Beheshti-Kashi, S. (2015). Fashion Forecasting in Social Media: A qualitative comparison of relevant channels In: LogDynamics Research Report, (2015), International Graduate School for Dynamics in Logistics, Bremen.

Bibliography

Abdous, M. and He, W. (2011). Using text mining to uncover students' technology-related problems in live video streaming. British Journal of Educational Technology, 42(1):40–49.

Abts, D. and Muelder, W. (2013). Grundkurs Wirtschaftsinformatik: eine kompakte und praxisorientierte Einführung. Lehrbuch. Springer Vieweg, Wiesbaden, 8. überarbeitete und erweiterte Auflage edition.

af Rosenborg, K., Christina, D., Buhl-Andersen, I., Nilsson, L. B., Rebild, M. P., Mukkamala, R. R., Hussain, A., and Vatrapu, R. (2017). Buzz vs. sales: Big social data analytics of style icon campaigns and fashion designer collaborations on h&m's facebook page. In 50th Hawaii International Conference on System Sciences (HICSS), pages 1861–1870.

Agarwal, N. and Yiliyasi, Y. (2010). Information quality challenges in social media. In International Conference on Information Quality.

Akundi, A., Tseng, B., Wu, J., Smith, E., Subbalakshmi, M., and Subbalakshmi, M. (2018). Text Mining to Understand the Influence of Social Media Applications on Smartphone Supply Chain. Procedia Computer Science, 140:87–94.

Allgöwer, F., Badgwell, T. A., Qin, J. S., Rawlings, J. B., and Wright, S. J. (1999). Nonlinear Predictive Control and Moving Horizon Estimation: An Introductory Overview. In Frank, P. M., editor, Advances in Control, pages 391–449. Springer London.

Allweyer, T. (2015). BPMN 2.0-Business Process Model and Notation: Einführung in den Standard für die Geschäftsprozessmodellierung. BoD–Books on Demand.

Archak, N., Ghose, A., and Ipeirotis, P. G. (2011). Deriving the pricing power of product features by mining consumer reviews. Management science, 57(8):1485–1509.

Armstrong, J. S. (2001). Standards and Practices for Forecasting, pages 679–732. Springer US, Boston, MA.

Ashwin Kumar, T., Kammarpally, P., and George, K. (2016). Veracity of information in twitter data: A case study. In 2016 International Conference on Big Data and Smart Computing (BigComp), pages 129–136.

Atkinson-Abutridy, J., Mellish, C., and Aitken, S. (2004). Combining information extraction with genetic algorithms for text mining. IEEE Intelligent Systems, 19(3):22–30.

Au, K.-F., Choi, T.-M., and Yu, Y. (2008). Fashion retail forecasting by evolutionary neural networks. International Journal of Production Economics, 114(2):615–630.

Awwad, M., Kulkarni, P., Bapna, R., and Marathe, A. (2018). Big Data Analytics in Supply Chain : A Literature Review. In International Conference on Industrial Engineering and Operations Management, pages 418–425.

Bahl, L. R. and Mercer, R. L. (1976). Part of speech assignment by a statistical decision algorithm. In IEEE International Symposium on Information Theory, pages 88–89.

Barnes, L. and Lea-Greenwood, G. (2006). Fast fashioning the supply chain: shaping the research agenda. Journal of Fashion Marketing and Management: An International Journal, 10(3):259–271.

Beheshti-Kashi, S. (2017). The fashion trend concept and its applicability to fashion markets and supply chains. In Freitag, M., Kotzab, H., and Pannek, J., editors, Dynamics in Logistics, pages 163–172. Springer International Publishing.

Beheshti-Kashi, S. (2020). Development of a social media process model for fashion and apparel supply chain decisions. Cyber-Physical Systems, 6(2):76–95.

Beheshti-Kashi, S., Hribernik, K., Lützenberger, J., Arabsolgar, D., and Thoben, K.-D. (2016). Fashion Supply Chains and Social Media: Examining the Potential of Data Analysis of Social-Media Texts for Decision Making-Processes in Fashion Supply Chains. In Workshop on Business Models and ICT Technologies for the Fashion Supply Chain, pages 271–281. Springer.

Beheshti-Kashi, S., Karimi, H. R., Thoben, K.-D., Lütjen, M., and Teucke, M. (2015a). A survey on retail sales forecasting and prediction in fashion markets. Systems Science & Control Engineering, 3(1):154–161.

Beheshti-Kashi, S. and Kinra, A. (2020). A process model for the selection of social media channels by means of information quality assessment of fashion blogs for apparel supply chains. In 5th Colloquium on European Research in Retailing. Proceedings of the 2020 CERR Conference, Valencia, Spain, 31.08.-02.09.2020., pages 143–159.

Beheshti-Kashi, S., Lütjen, M., Stoever, L., and Thoben, K.-D. (2015b). TrendFashion – A Framework for the Identification of Fashion Trends. In Cunningham, D., Hofstedt, P., Meer, K., Schmitt, I., editor, INFORMATIK 2015, Lecture Notes in Informatics (LNI), Bonn 2015.

Beheshti-Kashi, S., Lütjen, M., and Thoben, K.-D. (2018). Social Media Analytics for Decision Support in Fashion Buying Processes. In Artificial Intelligence for Fashion Industry in the Big Data Era, pages 71–93. Springer.

Beheshti-Kashi, S., Pannek, J., and Kinra, A. (2019). Complementing decision support and forecasting risk in supply chain with unstructured data. IFAC-PapersOnLine, 52(13):1721 – 1726. 9th IFAC Conference on Manufacturing Modelling, Management and Control MIM 2019.

Bell, E., Bryman, A., and Harley, B. (2018). Business research methods. Oxford university press.

Belvedere, V. and Goodwin, P. (2017). The influence of product involvement and emotion on short-term product demand forecasting. International Journal of Forecasting, 33(3):652–661.

Berger, J. and Milkman, K. L. (2012). What Makes Online Content Viral? Journal of Marketing Research, 49(2):192–205.

Bermingham, A. and Smeaton, A. (2011). On using twitter to monitor political sentiment and predict election results. In Proceedings of the Workshop on Sentiment Analysis where AI meets Psychology (SAAIP 2011), pages 2–10.

Berry, M. W. and Kogan, J. (2010). Text Mining: Applications and Theory. John Wiley & Sons.

Bhardwaj, V. and Fairhurst, A. (2010). Fast fashion: response to changes in the fashion industry. The International Review of Retail, Distribution and Consumer Research, 20(1):165–173.

Bird, S., Klein, E., and Loper, E. (2009). Natural language processing with Python: analyzing text with the natural language toolkit. O'Reilly Media.

Birtwistle, G., Siddiqui, N., and Fiorito, S. S. (2003). Quick response: perceptions of uk fashion retailers. International Journal of Retail & Distribution Management, 31(2):118–128.

Bode, C., Wagner, S. M., Petersen, K. J., and Ellram, L. M. (2011). Understand responses to supply chain disruptions: insights from information processing and resource dependence perspectives. The Academy of Management Journal, 54(4):833–856.

Boldt, L. C., Vinayagamoorthy, V., Winder, F., Schnittger, M., Ekran, M., Mukkamala, R. R., Lassen, N. B., Flesch, B., Hussain, A., and Vatrapu, R. (2016). Forecasting nike's sales using facebook data. In 2016 IEEE International Conference on Big Data (Big Data), pages 2447–2456.

Bollen, J., Mao, H., and Zeng, X. (2011). Twitter mood predicts the stock market. Journal of Computational Science, 2(1):1–8.

Borchers, A. (2009). A Multidisciplinary View of Data Quality, pages 2741–2747. IGI Global, Hershey.

Boyd, D. M. and Ellison, N. B. (2007). Social Network Sites: Definition, History, and Scholarship. Journal of Computer-Mediated Communication, 13(1):210–230.

Breitner, M. H. (2012). Vorgehensmodelle. In Kurbel, K., Becker, J., Gronau, N., Sinz, E., and Suhl, L., editors, Enzyklopädie der Wirtschaftsinformatik. Oldenbourg Wissenschaftsverlag, München.

Broniatowski, D. A., Paul, M. J., and Dredze, M. (2013). National and local influenza surveillance through twitter: an analysis of the 2012-2013 influenza epidemic. PloS one, 8(12):e83672.

Bruce, M. and Daly, L. (2006). Buyer behaviour for fast fashion. Journal of Fashion Marketing and Management: An International Journal, 10(3):329–344.

Bruns, A. (2006). Towards produsage: Futures for user-led content production. In Ess, C., Sudweeks, F., and Hrachovec, H., editors, Proceeding of the 5th International Conference on Cultural Attitudes towards Technology and Communication, pages 275–284. School of Information Technology, Australia.

Byström, K. and Järvelin, K. (1995). Task complexity affects information seeking and use. Information Processing & Management, 31(2):191–213.

Cai, L. and Zhu, Y. (2015). The Challenges of Data Quality and Data Duality Assessment in the Big Data Era. Data Science Journal, 14(2):1–10.

Case, D. and Spink, A. (2012). Looking for Information: A Survey of Research on Information Seeking, Needs and Behavior. Library and Information Science. Emerald Group Publishing Limited.

Case, D. O. and Given, L. M. (2016). Looking for Information: a Survey of Research on Information Seeking, Needs, and Behavior. Studies in information. Emerald, Bingley.

Chakraborty, G., Pagolu, M., and Garla, S. (2014). Text Mining and Analysis: Practical Methods, Examples, and Case Studies using SAS. SAS Institute.

Chan, H. K., Lacka, E., Yee, R. W., and Lim, M. K. (2017). The role of social media data in operations and production management. International Journal of Production Research, 55(17):5027–5036.

Chaniotakis, E., Antoniou, C., and Pereira, F. (2016). Mapping social media for transportation studies. IEEE Intelligent Systems, 31(6):64–70.

Chau, R. and Yeh, C.-H. (2004). A multilingual text mining approach to web cross-lingual text retrieval. Knowledge-Based Systems, 17(5):219–227. Special Issue: Web Intelligence.

Chen, G. J., Wiener, J. L., Iyer, S., Jaiswal, A., Lei, R., Simha, N., Wang, W., Wilfong, K., Williamson, T., and Yilmaz, S. (2016). Realtime data processing at facebook. In Proceedings of the 2016 International Conference on Management of Data, pages 1087–1098. ACM.

Cheng, T. E. and Choi, T.-M. (2010). Innovative quick response programs in logistics and supply chain management. Springer Science & Business Media.

Choi, T.-M. (2018). Incorporating social media observations and bounded rationality into fashion quick response supply chains in the big data era. Transportation Research Part E: Logistics and Transportation Review, 114:386–397.

Choi, T.-M., Hui, C.-L., Liu, N., Ng, S.-F., and Yu, Y. (2014). Fast fashion sales forecasting with limited data and time. Decision Support Systems, 59:84–92.

Choi, T.-M., Hui, C.-L., Ng, S.-F., and Yu, Y. (2012). Color trend forecasting of fashionable products with very few historical data. IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), 42(6):1003–1010.

Chou, C.-H., Sinha, A. P., and Zhao, H. (2008). A text mining approach to internet abuse detection. Information Systems and e-Business Management, 6(4):419–439.

Chou, C.-H., Sinha, A. P., and Zhao, H. (2010). A hybrid attribute selection approach for text classification. Journal of the Association for Information Systems, 11(9):491–518.

Choudhary, A., Oluikpe, P., Harding, J., and Carrillo, P. (2009). The needs and benefits of text mining applications on post-project reviews. Computers in Industry, 60(9):728–740.

Christopher, M. (2000). The agile supply chain: competing in volatile markets. Industrial marketing management, 29(1):37–44.

Christopher, M., Lowson, R., and Peck, H. (2004). Creating agile supply chains in the fashion industry. International Journal of Retail & Distribution Management, 32(8):367–376.

Chung, J. and Mustafaraj, E. (2011). Can collective sentiment expressed on twitter predict political elections? In Proceedings of the Twenty-Fifth AAAI Conference on Artificial Intelligence, AAAI'11, page 1770–1771. AAAI Press.

Crawford Camiciottoli, B., Ranfagni, S., and Guercini, S. (2014). Exploring brand associations: an innovative methodological approach. European Journal of Marketing, 48(5/6):1092–1112.

Cui, R., Allon, G., Bassamboo, A., and Mieghem, J. A. V. (2015). Information sharing in supply chains: An empirical and theoretical valuation. Management Science, 61(11):2803–2824.

Cui, R., Gallino, S., Moreno, A., and Zhang, D. J. (2018). The operational value of social media information. Production and Operations Management, 27(10):1749–1769.

Curtin, R. T. (1982). Indicators of consumer behavior: The University of Michigan Surveys of Consumers. Public Opinion Quarterly, 46(3):340–352.

Da Silveira, G., Borenstein, D., and Fogliatto, F. S. (2001). Mass customization: Literature review and research directions. International Journal of Production Economics, 72(1):1–13.

Dasgupta, S. and Sengupta, K. (2016). Analyzing Consumer Reviews with Text Mining Approach: A Case Study on Samsung Galaxy S3. Paradigm, 20(1):56–68.

Davydenko, A. and Fildes, R. (2014). Measuring forecasting accuracy: Problems and recommendations (by the example of sku-level judgmental adjustments). In Choi, T.-M., Hui, C.-L., and Yu, Y., editors, Intelligent Fashion Forecasting Systems: Models and Applications, pages 43–70. Springer Berlin Heidelberg, Berlin, Heidelberg.

Demchenko, Y., Grosso, P., de Laat, C., and Membrey, P. (2013). Addressing big data issues in scientific data infrastructure. In 2013 International Conference on Collaboration Technologies and Systems (CTS), pages 48–55.

Dhar, V. and Chang, E. A. (2009). Does chatter matter? The impact of user-generated content on music sales. Journal of Interactive Marketing, 23(4):300–307.

Dijkman, R. M., Ipeirotis, P. G., Aertsen, F., and van Helden, R. (2015). Using twitter to predict sales: A case study. Computing Research Repository, abs/1503.04599.

Duan, J., Zhang, M., Jingzhong, W., and Xu, Y. (2011). A hybrid framework to extract bilingual multiword expression from free text. Expert Systems with Applications, 38(1):314–320.

Esteban-Santos, L., García Medina, I., Carey, L., and Bellido-Pérez, E. (2018). Fashion bloggers: communication tools for the fashion industry. Journal of Fashion Marketing and Management, 22(3):420–437.

Fan, W., Wallace, L., Rich, S., and Zhang, Z. (2006). Tapping the Power of Text Mining. Communications of the ACM, 49(9):76–82.

Feldman, R. and Sanger, J. (2007). The Text Mining Handbook: Advanced Approaches in Analyzing Unstructured Data. Cambridge University Press.

Fernie, J. (2014). Relationships in the Supply Chain, pages 35–57. Kogan Page Publishers.

Fettke, P. (2014). Eine Methode zur induktiven Entwicklung von Referenzmodellen. In Kundisch, D., Suhl, L., and Beckmann, L., editors, Tagungsband Multikonferenz Wirtschaftsinformatik 2014 (MKWI 2014), pages 1034–1047.

Fiddis, C. (1997). Manufacturer-retailer relationships in the food and drink industry: strategies and tactics in the battle for power. FT Retail & Consumer Publishing Pearson Professional London.

Fildes, R., Goodwin, P., Lawrence, M., and Nikolopoulos, K. (2009). Effective forecasting and judgmental adjustments: an empirical evaluation and strategies for improvement in supply-chain planning. International Journal of Forecasting, 25(1):3–23.

Filß, C., Höhn, R., Höppner, S., and Wetzel, H. (2005). Arbeitsbericht der Gl Fachgruppe WI-VM Arbeitskreis Vorgehensmodelltypen: Rahmen zur Auswahl von Vorgehensmodellen. In Petrasch, R., editor, Entscheidungsfall Vorgehensmodelle : 12. Workshop der Fachgruppe WI-VM der Gesellschaft für Informatik e.V. (GI), pages 183–227, Aachen. Shaker.

Fischer, T., Biskup, H., and Müller-Luschnat, G. (1998). Begriffliche Grundlagen für Vorgehensmodelle. In Kneuper, R., Müller-Luschnat, G., and Oberweis, A., editors, Vorgehensmodelle für die betriebliche Anwendungsentwicklung, pages 13–31. Vieweg+Teubner Verlag.

Fisher, C., Lauria, E., Chengalur-Smith, I., and Wang, R. (2011). Introduction to information quality. AuthorHouse.

Fisher, M. L. (1997). What is the right supply chain for your product? Harvard business review, 75:105–117.

Flynn, B. B. and Lu, G. (2016). On Theory in Supply Chain Uncertainty and its Implications for Supply Chain Integration. Journal of Supply Chain Management, 52(3):3–27.

Fogg, B. and Tseng, H. (1999). The elements of computer credibility. In Proceedings of the SIGCHI conference on Human Factors in Computing Systems, pages 80–87. ACM.

Fosso Wamba, S., Gunasekaran, A., Papadopoulos, T., and Ngai, E. (2018). Big data analytics in logistics and supply chain management. The International Journal of Logistics Management, 29(2):478–484.

Fulcher, A. J. and Hills, P. (1996). Towards a Strategic Framework for Design Research. Journal of Engineering Design, 7(2):183–193.

Gaur, V., Giloni, A., and Seshadri, S. (2005). Information Sharing in a Supply Chain Under ARMA Demand. Management Science, 51(6):851–1013.

Gereffi, G. and Memedovic, O. (2003). The global apparel value chain: What prospects for upgrading by developing countries. United Nations Industrial Development Organization Vienna.

Ghemawat, P., Nueno, J. L., and Dailey, M. (2003). ZARA: Fast fashion, volume 1. Harvard Business School, Boston, MA.

Giri, S. and Rai, S. S. (2013). Dynamics of garment supply chain. International Journal of Managing Value and Supply Chains, 4(4):29.

Goepel, K. D. (2018). Implementation of an online software tool for the analytic hierarchy process (AHP-OS). International Journal of the Analytic Hierarchy Process, 10(3):469–487.

Grob, H. L., Reepmeyer, J. A., and Bensberg, F. (2004). Einführung in die Wirtschaftsinformatik. Vahlen, München, 5., vollst. überarb. und erw. aufl. edition.

Gu, W. and Liu, X. (2010). Computer-Assisted Color Database For Trend Forecasting. In 2010 International Conference on Computational Intelligence and Software Engineering, pages 1–4. IEEE.

Guellil, I. and Boukhalfa, K. (2015). Social big data mining: A survey focused on opinion mining and sentiments analysis. In IEEE 12th International Symposium on Programming and Systems (ISPS), pages 1–10. IEEE.

Halvorsen, K., Hoffmann, J., Coste-Manière, I., and Stankeviciute, R. (2013). Can fashion blogs function as a marketing tool to influence consumer behavior? Evidence from Norway. Journal of Global Fashion Marketing, 4(3):211–224.

Hammond, J. (1999). Managing the apparel supply chain in the digital economy.

Hayes, C. (2007). Using Tags and Clustering to Identify Topic-Relevant Blogs. AAAI International Conference on Weblogs and Social Media (ICWSM-07).

Herring, S. C. (2010). Web content analysis: Expanding the paradigm. In J. Hunsinger, M. Allen, &. L. K., editor, The International Handbook of Internet Research, pages 233–249. Springer Verlag, Berlin.

Hevner, A. R., March, S. T., Park, J., and Ram, S. (2004). Design Science in Information Systems Research. Management Information Systems Quarterly, 28(1):75–105.

Hines, T. (2007). Supply Chain Strategies, Structures and Relationships. In Fashion Marketing, pages 51–77. Routledge, London.

Hines, T. and Bruce, M. (2007). Fashion Marketing: Contemporary issues. Routledge, London.

Hippner, H. and Rentzmann, R. (2006). Text Mining. Informatik-Spektrum, 29(4):287–290.

Hofmann, E. (2017). Big data and supply chain decisions: the impact of volume, variety and velocity properties on the bullwhip effect. International Journal of Production Research, 55(17):5108–5126.

Hofmann, E. and Rutschmann, E. (2018). Big data analytics and demand forecasting in supply chains: a conceptual analysis. The International Journal of Logistics Management, 29(2):739–766.

Holsapple, C., Hsiao, S., and Pakath, R. (2014). Business Social Media Analytics: Definition, Benefits, and Challenges: Completed Research Paper. 20th Americas Conference on Information Systems, AMCIS 2014, Savannah, Georgia, USA.

Hotho, A., Nürnberger, A., and Paaß, G. (2005). A Brief Survey of Text Mining. Journal for Computational Linguistics and Language Technology, 20(1):19–62.

Hu, X. and Liu, H. (2012). Text analytics in social media. In Aggarwal, C. C. and Zhai, C., editors, Mining Text Data, pages 385–414, Boston, MA. Springer US.

Hui, P. and Choi, T.-M. (2016). 5 - Using artificial neural networks to improve decision making in apparel supply chain systems. In Choi, T.-M., editor, Information Systems for the Fashion and Apparel Industry, Woodhead Publishing Series in Textiles, pages 97–107. Woodhead Publishing.

Iftikhar, R. and Khan, D. M. (2019). Social media big data analytics for demand forecasting: Development and case implementation of an innovative framework. Journal of Global Information Management, 28:103–120.

Ivanov, D., Tsipoulanidis, A., and Schönberger, J. (2017). Demand forecasting. In Ivanov, D., Tsipoulanidis, A., and Schönberger, J., editors, Global Supply Chain and Operations Management: A Decision-Oriented Introduction to the Creation of Value, Springer Texts in Business and Economics, pages 301–315. Springer International Publishing.

Jackson, T. (2007). The Process of Trend Development Leading to a Fashion Season. In Hines, T. and Bruce, M., editors, Fashion Marketing: Contemporary Issues, pages 168–188. Routledge, London.

Jackson, T. and Shaw, D. (2000). Mastering Fashion Buying and Merchandising Management. Macmillan International Higher Education.

Jain, C. L. (2007). Benchmarking forecasting software and systems. The Journal of Business Forecasting, 26(4):30.

Kahn, K. B. (2014). Solving the problems of new product forecasting. Business Horizons, 57(5):607–615.

Kalampokis, E., Tambouris, E., and Tarabanis, K. (2013). Understanding the predictive power of social media. Internet Research, 23(5):544–559.

Kaplan, A. M. and Haenlein, M. (2010). Users of the world, unite! The challenges and opportunities of Social Media. Business Horizons, 53(1):59–68.

Kasiviswanathan, S., Melville, P., Banerjee, A., and Sindhwani, V. (2011). Emerging topic detection using dictionary learning. In CIKM'11 - Proceedings of the 2011 ACM International Conference on Information and Knowledge Management, International Conference on Information and Knowledge Management, Proceedings, pages 745–754.

Kim, Y., Huang, J., and Emery, S. (2016). Garbage in, garbage out: data collection, quality assessment and reporting standards for social media data use in health research, infodemiology and digital disease detection. Journal of medical Internet research, 18(2).

King, J. (2012). Colour trends and selection in fashion design. In Best, J., editor, Colour Design, Woodhead Publishing Series in Textiles, pages 535–550. Woodhead Publishing.

Kobayashi, V. B., Mol, S. T., Berkers, H. A., Kismihók, G., and Hartog, D. N. D. (2018). Text mining in organizational research. Organizational Research Methods, 21(3):733–765.

Kozinets, R. V., de Valck, K., Wojnicki, A. C., and Wilner, S. J. (2010). Networked narratives: Understanding word-of-mouth marketing in online communities. Journal of Marketing, 74(2):71–89.

Krcmar, H. (2015). Informationsmanagement. Springer Berlin Heidelberg.

Krippendorff, K. (1980). Content analysis : an introduction to its methodology. Sage Publications, Beverly Hills.

Kulmala, M., Mesiranta, N., and Tuominen, P. (2013). Organic and amplified eWOM in consumer fashion blogs. Journal of Fashion Marketing and Management: An International Journal, 17(1):20–37.

Lam, J. K. and Postle, R. (2006). Textile and apparel supply chain management in Hong Kong. International Journal of Clothing Science and Technology, 18(4):265–277.

Lamb, A., Paul, M. J., and Dredze, M. (2013). Separating fact from fear: Tracking flu infections on Twitter. In Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, pages 789–795, Atlanta, Georgia. Association for Computational Linguistics.

Laplante, P. A. (2013). Requirements Engineering for software and systems. Auerbach Publishers Inc, London.

Lassen, N. B., Madsen, R., and Vatrapu, R. (2014). Predicting iphone sales from iphone tweets. In 2014 IEEE 18th International Enterprise Distributed Object Computing Conference, pages 81–90. IEEE.

Le Pechoux, B., Little, T. J., and Istook, C. L. (2007). Innovation management in creating new fashions. In Fashion marketing: Contemporary issues, pages 188–2016. Routledge, London, 2 edition.

Lee, J. E. and Watkins, B. (2016). Youtube Vloggers' influence on consumer luxury brand perceptions and intentions. Journal of Business Research, 69(12):5753 – 5760.

Lee, S., Song, J., Baker, J., Kim, Y., and Wetherbe, J. C. (2011). The commoditization of it: Evidence from a longitudinal text mining study. Communications of the Association for Information Systems, 29(1):12.

Lee, S.-E. and Chen, J. C. (1999). Masscustomization methodology for an apparel industry with a future. Journal of Industrial Technology, 16(1):1–8.

Leeflang, P. S. H., Vries, L. D., Gensler, S., and Lee, P. S. H. (2012). Popularity of brand posts on brand fan pages : An investigation of the effects of social media marketing. Journal of Interactive Marketing, 26(2):24–28.

Li, G. and Liu, F. (2012). Application of a clustering method on sentiment analysis. Journal of Information Science, 38(2):127–139.

Lin, J. (2015). On building better mousetraps and understanding the human condition: Reflections on big data in the social sciences. The ANNALS of the American Academy of Political and Social Science, 659(1):33–47.

Liu, B. et al. (2010). Sentiment analysis and subjectivity. In Handbook of Natural Language Processing, volume 2, pages 627–666. Taylor and Francis Group, Boca.

Liu, T., Ding, X., Chen, Y., Chen, H., and Guo, M. (2016). Predicting movie box-office revenues by exploiting large-scale social media content. Multimedia Tools and Applications, 75(3):1509–1528.

Liu-Thompkins, Y. and Rogerson, M. (2012). Rising to Stardom: An Empirical Investigation of the Diffusion of User-generated Content. Journal of Interactive Marketing, 26(2):71–82.

Lokot, T. and Diakopoulos, N. (2016). News Bots: Automating news and information dissemination on Twitter. Digital Journalism, 4(6):682–699.

Lomborg, S. and Bechmann, A. (2014). Using APIs for data collection on social media. The Information Society, 30(4):256–265.

Lowson, R. H. (2003). Apparel sourcing: assessing the true operational cost. International Journal of Clothing Science and Technology, 15(5):335–345.

Lukoianova, T. and Rubin, V. (2014). Veracity Roadmap: Is Big Data Objective, Truthful and Credible? Advances in Classification Research Online, 24(1):4–15.

Marr, B. (2010). The intelligent company. Wiley Online Library.

Marr, B. (2015). Why only one of the 5 Vs of big data really matters| IBM Big Data Analytics and Hub. URI= http://www. ibmbigdatahub. com/blog/why-only-one-5-vs-big-data-really-matters.

Meredith, J. (1998). Building operations management theory through case and field research. Journal of Operations Management, 16(4):441–454.

Metzger, M. J. and Flanagin, A. J. (2013). Credibility and trust of information in online environments: The use of cognitive heuristics. Journal of Pragmatics, 59:210–220.

Miner, G., Elder, J., Fast, A., Hill, T., Nisbet, T., and Delen, D. (2012). Practical Text Mining and Statistical Analysis for Non-Structured Text Data Applications. Elsevier Science Publishing Co Inc.

Mitchell, R. (2018). Web Scraping with Python: Collecting More Data from the Modern Web. O'Reilly Media, Inc.

Mostard, J., Teunter, R., and De Koster, R. (2011). Forecasting demand for single-period products: A case study in the apparel industry. European Journal of Operational Research, 211(1):139–147.

Natarajan, M. (2005). Role of text mining in information extraction and information management. DESIDOC Journal of Library & Information Technology, 25(4):31–38.

Nayak, R. and Padhye, R. (2015). Introduction: The apparel industry. In Garment Manufacturing Technology, pages 1–17. Elsevier.

Ngai, E., Peng, S., Alexander, P., and Moon, K. K. (2014). Decision support and intelligent systems in the textile and apparel supply chain: An academic review of research articles. Expert Systems with Applications, 41(1):81–91.

Nilsson, M. and Alserud, F. (2017). Who can an organization believe in social media? exploring the process of believability assessment. http://lup.lub.lu.se/student-papers/record/8910720.

Nunamaker Jr, J. F., Chen, M., and Purdin, T. D. (1990). Systems development in information systems research. Journal of Management Information Systems, 7(3):89–106.

OECD (2007). Participative web and user-created content: Web 2.0, wikis, and social networking .

Olston, C., Najork, M., et al. (2010). Web crawling. Foundations and Trends® in Information Retrieval, 4(3):175–246.

O'Reilly, T. (2004). The architecture of participation.

Page, R., Barton, D., Unger, J. W., and Zappavigna, M. (2014). Researching language and social media: A student guide. Routledge, London.

Pagolu, V. S., Reddy, K. N., Panda, G., and Majhi, B. (2016). Sentiment analysis of twitter data for predicting stock market movements. In 2016 International Conference on Signal Processing, Communication, Power and Embedded System (SCOPES), pages 1345–1350. IEEE.

Peffers, K., Tuunanen, T., Rothenberger, M., and Chatterjee, S. (2007). A Design Science Research Methodology for Information Systems Research. Journal of Management Information Systems, 24(3):45–77.

Peng, W. and Sun, T. (2012). Method and system for identifying a key influencer in social media utilizing topic modeling and social diffusion analysis. US Patent 8,312,056.

Peters, K., Chen, Y., Kaplan, A. M., Ognibeni, B., and Pauwels, K. (2013). Social Media Metrics — A Framework and Guidelines for Managing Social Media. Journal of Interactive Marketing, 27(4):281–298.

Phillips, L., Dowling, C., Shaffer, K., Hodas, N., and Volkova, S. (2017). Using social media to predict the future: a systematic literature review. Computing Research Repository (CoRR).

Piller, F. T. and Müller, M. (2004). A new marketing approach to mass customisation. International Journal of Computer Integrated Manufacturing, 17(7):583–593.

Porter, M. F. (1980). An algorithm for suffix stripping. Program: Electronic Library and Information Systems, 14(3):130–137.

Rao, T. and Srivastava, S. (2012). Analyzing stock market movements using twitter sentiment analysis. In Proceedings of the 2012 International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2012), pages 119–123. IEEE Computer Society.

Redman, T. C. (1998). The impact of poor data quality on the typical enterprise. Communications of the ACM, 41(2):79–82.

Rickman, T. A. and Cosenza, R. M. (2007). The changing digital dynamics of multichannel marketing: The feasibility of the weblog: Text mining approach for fast fashion trending. Journal of Fashion Marketing and Management: An International Journal, 11(4):604–621.

Russom, P. et al. (2011). Big data analytics. TDWI best practices report, fourth quarter, 19(4):1–34.

Salton, G., Wong, A., and Yang, C.-S. (1975). A vector space model for automatic indexing. Communications of the ACM, 18(11):613–620.

Sanders, N. R. and Ganeshan, R. (2015). Call for Papers Special Issue of Production and Operations Management on “ Big Data in Supply Chain Management ”. Production and Operations Management, 24(7):1193–1194.

Sanders, N. R. and Manrodt, K. B. (2003). The efficacy of using judgmental versus quantitative forecasting methods in practice. Omega, 31(6):511–522.

Schaer, O., Kourentzes, N., and Fildes, R. (2019). Demand forecasting with user-generated online information. International Journal of Forecasting, 35(1):197–212.

Schieber, A. and Hilbert, A. (2014a). Entwicklung eines generischen Vorgehensmodells für Text Mining. Dresdner Beiträge zur Wirtschaftsinformatik, 69(14).

Schieber, A. and Hilbert, A. (2014b). Process model for content extraction from weblogs. International Journal of Intelligent Information Technologies (IJIIT), 10(2):20–36.

Scully, K. and Cobb, D. J. (2012). Color forecasting for fashion. Laurence King Publishing Ltd.

Sebei, H., Hadj Taieb, M. A., and Ben Aouicha, M. (2018). Review of social media analytics process and big data pipeline. Social Network Analysis and Mining, 8(1):1–28.

See-To, E. W. and Ngai, E. W. (2018). Customer reviews for demand distribution and sales nowcasting: A big data approach. Annals of Operations Research, 270(1-2):415–431.

Şen, A. (2008). The US fashion industry: a supply chain review. International Journal of Production Economics, 114(2):571–593.

Seram, N., Julian, N., and Lanarolle, G. (2016). Apparel product development : An overview of existing models. International Journal of Product Development, 21(5/6).

Shankar, G. and Watts, S. (2003). A Relevant, Believable Approach for Data Quality Assessment. In Proceedings of 8th International Conference on Information Quality, pages 178–189.

Shankaranarayanan, G. and Blake, R. (2017). From content to context: The evolution and growth of data quality research. Journal of Data and Information Quality (JDIQ), 8(2):9.

Shankaranarayanan, G., Iyer, B., and Stoddard, D. (2012). Quality of Social Media Data and Implications of Social Media for Data Quality. In ICIQ, pages 311–325.

Shannon, C. E. (1948). A mathematical theory of communication. The Bell System Technical Journal, 27:379–423.

Shen, B., Qian, R., and Choi, T.-M. (2017). Selling luxury fashion online with social influences considerations: Demand changes and supply chain coordination. International Journal of Production Economics, 185(C):89–99.

Singh, P., Gupta, Y., Jha, N., and Rajan, A. (2019). Fashion retail: Forecasting demand for new items. ArXiv, abs/1907.01960.

Smith-Yoshimura, K. and Shein, C. (2011). Social Metadata for Libraries, Archives and Museums Part 1: Site Reviews. OCLC Research, Dublin, Ohio.

Souza, G. C. (2014). Supply chain analytics. Business Horizons, 57(5):595–605.

Stahlknecht, P. and Hasenkamp, U. (2005). Einführung in die Wirtschaftsinformatik. Springer-Lehrbuch. Springer, Berlin [u.a.], 11., vollst. überarb. aufl., 185. - 200. tsd. edition.

Sterne, J. (2010). Social media metrics how to measure and optimize your marketing investment. New Rules Social Media Series. Wiley, Hoboken, N.J.

Stevenson, M. and Spring, M. (2007). Flexibility from a supply chain perspective: definition and review. International Journal of Operations & Production Management, 27(7):685–713.

Stieglitz, S. and Dang-Xuan, L. (2013). Social Media and Political Communication: a Social Media Analytics Framework. Social Network Analysis and Mining, 3(4):1277–1291.

Stieglitz, S., Dang-Xuan, L., Bruns, A., and Neuberger, C. (2014a). Social Media Analytics. Business & Information Systems Engineering, 6(2):89–96.

Stieglitz, S., Dang-Xuan, L., Bruns, A., and Neuberger, C. (2014b). Social Media Analytics. Wirtschaftsinformatik, 56(2):101–109.

Stieglitz, S., Mirbabaie, M., Ross, B., and Neuberger, C. (2018). Social Media Analytics–Challenges in Topic Discovery, Data Collection, and Data Preparation. International Journal of Information Management, 39:156–168.

Subhashini, R. and Kumar, V. J. S. (2010). Evaluating the Performance of Similarity Measures Used in Document Clustering and Information Retrieval. 2010 First International Conference on Integrated Intelligent Computing, pages 27–31.

Sun, Z.-L., Choi, T.-M., Au, K.-F., and Yu, Y. (2008). Sales forecasting using extreme learning machine with applications in fashion retailing. Decision Support Systems, 46(1):411–419.

Swain, A. and Cao, R. (2019). Using sentiment analysis to improve supply chain intelligence. Information Systems Frontiers, 21(2):469–484.

Tayi, G. K. and Ballou, D. P. (1998). Examining Data Quality. Communications of the ACM, 41(2):54–57.

Taylor, R. S. (1968). Question-negotiation and information seeking in libraries. College & Research Libraries, 29(3):178–194.

Teucke, M., Ait-Alla, A., El-Berishy, N., Beheshti-Kashi, S., and Lütjen, M. (2016). Forecasting of seasonal apparel products. In Dynamics in Logistics, pages 633–642. Springer.

Thomassey, S. (2010). Sales forecasts in clothing industry: The key success factor of the supply chain management. International Journal of Production Economics, 128(2):470–483.

Thomassey, S. (2014). Sales Forecasting in Apparel and Fashion Industry: A Review. In Choi, T.-M., Hui, C.-L., and Yu, Y., editors, Intelligent Fashion Forecasting Systems: Models and Applications, pages 9–27. Springer Berlin Heidelberg, Berlin, Heidelberg.

Thomassey, S. and Happiette, M. (2007). A neural clustering and classification system for sales forecasting of new apparel items. Applied Soft Computing, 7(4):1177–1187.

Tiwari, S., Wee, H. M., and Daryanto, Y. (2018). Big data analytics in supply chain management between 2010 and 2016 : Insights to industries. Computers & Industrial Engineering, 115:319–330.

Tokar, T., Aloysius, J., Williams, B., and Waller, M. (2014). Bracing for demand shocks : An experimental investigation. Journal of Operations Management, 32(4):205–216.

Tumasjan, A., Sprenger, T. O., Sandner, P. G., and Welpe, I. M. (2010). Predicting elections with twitter: What 140 characters reveal about political sentiment. In Proceedings of the International AAAI Conference on Web and Social Media, volume 4, pages 178–185.

Ur-Rahman, N. and Harding, J. A. (2012). Textual data mining for industrial knowledge management and text classification: A business oriented approach. Expert Systems with Applications, 39(5):4729–4739.

Uzunoğlu, E. and Misci Kip, S. (2014). Brand communication through digital influencers: Leveraging blogger engagement. International Journal of Information Management, 34(5):592–602.

van Noort, G., Voorveld, H. A., and van Reijmersdal, E. A. (2012). Interactivity in Brand Web Sites: Cognitive, Affective, and Behavioral Responses Explained by Consumers' Online Flow Experience. Journal of Interactive Marketing, 26(4):223–234.

Varol, O., Ferrara, E., Davis, C. A., Menczer, F., and Flammini, A. (2017). Online Human-Bot Interactions: Detection, Estimation, and Characterization. In Proceedings of the Eleventh International Conference on Web and Social Media, ICWSM 2017, Montréal, Québec, Canada, May 15-18, 2017, pages 280–289.

Venable, J., Pries-Heje, J., and Baskerville, R. (2012). A Comprehensive Framework for Evaluation in Design Science Research. In Peffers, K., Rothenberger, M., and Kuechler, B., editors, Design Science Research in Information Systems. Advances in Theory and Practice, pages 423–438, Berlin. Springer Berlin Heidelberg.

Verlage, M. (1998). Vorgehensmodelle und ihre Formalisierung. In Kneuper, R., Müller-Luschnat, G., and Oberweis, A., editors, Vorgehensmodelle für die betriebliche Anwendungsentwicklung, pages 60–75. Vieweg+Teubner Verlag, Wiesbaden.

Waller, M. A. and Fawcett, S. E. (2013). Data science, predictive analytics, and big data: A revolution that will transform supply chain design and management. Journal of Business Logistics, 34(2):77–84.

Walls, J. G. and Sawy, O. A. E. (1992). Building an Information System Design Theory for Vigilant EIS. Information Systems Research, 3(1):36–59.

Wang, G., Gunasekaran, A., Ngai, E. W., and Papadopoulos, T. (2016). Big data analytics in logistics and supply chain management: Certain investigations for research and applications. International Journal of Production Economics, 176:98–110.

Wang, R. Y. and Strong, D. M. (1996). Beyond accuracy: What data quality means to data consumers. Journal of Management Information Systems, 12(4):5–33.

Webb, B. (2007). Retail brand marketing in the fashion industry. In Fashion Marketing, pages 131–153. Routledge.

Webster, J. J. and Kit, C. (1992). Tokenization as the initial phase in NLP. In The 15th International Conference on Computational Linguistics.

Weimann, G. (1994). The Influentials: People Who Influence People. State University of New York Press, Albany.

Wenzel, R. and Van Quaquebeke, N. (2018). The double-edged sword of big data in organizational and management research: A review of opportunities and risks. Organizational Research Methods, 21(3):548–591.

Wickett, J. L., Gaskill, L. R., and Damhorst, M. L. (1999). Apparel Retail Product Development: Model Testing and Expansion. Clothing and Textiles Research Journal, 17(1):21–35.

Wiswedel, B., Kötter, T., and Silipo, R. (2013). Analyzing the web from start to finish. https://www.knime.com/sites/default/files/inline-images/knime_web_knowledge_extraction.pdf.

Wojaczek, B. (1996). Koordinationsorientiertes Logistik-Management in der Textilwirtschaft Ein Beitrag zur ganzheitlichen Optimierung der Logistik aus Sicht der Bekleidungsindustrie. Frankfurt am Main.

Wong, W. and Guo, Z. (2010). A hybrid intelligent model for medium-term sales forecasting in fashion retail supply chains using extreme learning machine and harmony search algorithm. International Journal of Production Economics, 128(2):614–624.

Wong, W. K. (2013). Understanding key decision points in the apparel supply chain. In Optimizing Decision Making in the Apparel Supply Chain Using Artificial Intelligence (AI), pages 1–12. Woodhead Publishing Series in Textiles.

Wong, W. K., Zeng, X. H., and Au, K. F. (2013). Selecting the location of apparel manufacturing plants using neural networks. In Optimizing Decision Making in the Apparel Supply Chain Using Artificial Intelligence (AI), pages 41–54. Woodhead Publishing Series in Textiles.

Yeung, H.-T., Choi, T.-M., and Chiu, C.-H. (2010). Innovative mass customization in the fashion industry. In Innovative quick response programs in logistics and supply chain management, pages 423–454. Springer.

Yin, S. and Kaynak, O. (2015). Big data for modern industry: Challenges and trends [point of view]. Proceedings of the IEEE, 103(2):143–146.

Zeng, D., Chen, H., Lusch, R., and Li, S.-H. (2010). Social media analytics and intelligence. IEEE Intelligent Systems, 25(6):13–16.

Zhang, X., Fuehres, H., and Gloor, P. A. (2011). Predicting stock market indicators through twitter "i hope it is not as bad as i fear". Procedia-Social and Behavioral Sciences, 26:55–62.

Zülch, G., Koruca, H. I., and Börkircher, M. (2011). Simulation-supported change process for product customization–a case study in a garment company. Computers in Industry, 62(6):568–577.