UNIVERSITY OF TWENTE.

Faculty of Behavioural Management
and Social Sciences

The Next Big Thing

The Use of Text Mining Analysis of Crowdfunding Data for Technology Foresight

Master Thesis

M.Sc. Business Administration

M.Sc. Innovation Management & Entrepreneurship

University of Twente:
First supervisor: Prof. Dr. Fons Wijnhoven
Second supervisor: Dr. Erwin Hofman

Technische Universität Berlin:
Supervisor: Jakob Pohlisch

Sebastian Stöckl

Abstract

Technology foresight is the systematic approach to look into the future of technologies and identify emerging fields of economic and social benefits. Foresight methods, such as roadmaps and scenarios, have been developed to structure the foresight process, acquire information and explore novel ideas. Despite the increasing volume of external data sources and data mining techniques, these methods still primarily rely on qualitative approaches that lead to inefficient and non-transparent processes. To address these limitations, this thesis suggests the systematic analysis of crowdfunding data for technology foresight applications. The goal of the research is to examine crowdfunding data in order to increase the knowledge base for foresight activities, by detecting emerging trends, technologies and markets. Six different methods are designed that can be implemented into technology foresight processes: Word clouds, keyword emergence map, market portfolio map, technology risk map, market hype curve, and co-word networks. The applicability and usefulness of the proposed methods are exemplified by a case study, analyzing the development of robot technologies. It is shown that the analysis of crowdfunding data bears the potential to support and improve foresight activities and offers new insights for strategic planning, decision making and product development.

Management summary

Due to rapidly changing market needs and technological progress, it is essential for companies to scan the future of technologies and identify emerging fields of economic benefits. Various foresight methods have been developed that still primarily rely on manual research and the expertise and internal knowledge of experts. This often results in costly, time-consuming and biased processes. To address these limitations and enhance the knowledge base of foresight activities, this study suggests the design of new quantitative research methods that are based on the exploitation of crowdfunding data to derive novel and detailed insights about the development of technologies and markets. This information can be used to support technology foresight experts during different steps of the foresight process.

To systematically analyze the data, text mining methods are combined with contextual data, as well as trend and co-word analysis. Six different methods are designed: Word clouds, keyword emergence map, market portfolio map, technology risk map, market hype curve, and co-word networks that are meant for supporting the traditional foresight process and facilitate the detection of future signals and trends. These methods are aligned with the requirements for technology roadmapping and scenario planning. The applicability and usefulness of the proposed methods are exemplified by a case study, analyzing the development of robot technologies.

Based on the analysis of more than 26,200 crowdfunding campaigns from the years 2009 until 2017, it is shown that the impact of this thesis on technology foresight activities is multifold: First, crowdfunding as a new and high-potential data source is presented and analyzed, and it is shown that novel and highly relevant insights can be derived from it. Second, innovative methods to analyze technologies from various points are presented and successfully applied. Third, new insights about the identification of weak signals have been made by demonstrating the detection of weak signal technologies in crowdfunding datasets and confirming the correlation between term frequencies and investment rates for weak signal technologies. Fourth, guidance on how to support technology roadmapping and scenario processes during different steps are presented. In this way, these foresight processes are expected to become more efficient and transparent and are enhanced through access to additional data-driven knowledge sources.

To verify the practical benefits of this study, the methods have been evaluated in a discussion with data analyst experts at a German innovation consulting company. It is shown that crowdfunding data can be used as a relevant source for the detection of technological trends and opportunities and that the proposed methods can support strategic planning, decision making and product development.

Outline of the thesis

List of figures

List of tables

List of abbreviations

TF: Term Frequency
RTF: Relative Term Frequency
RIR: Relative Investment Rate
IoT: Internet of Things
AI: Artificial Intelligence
B2C: Business-to-Customer

1. Introduction

“When you are running a business, there is a constant need to reinvent oneself.

One should have the foresight to stay ahead in times of rapid change

and rid ourselves of stickiness in any form in the business.”

Shiv Nadar, founder of HCL technologies

1.1. Context of the study and research goal

Today’s economic, scientific and social environments are highly characterized by the ever-increasing speed of technological change and disruptive technologies (Saritas & Burmaoglu, 2015). Apart from the potential creation of new opportunities for strategic investments, this also leads to emerging uncertainties and complexities that call for systematic ways to foresee, predict and shape technological change (Martin, 1995; Miles, 2010). Companies have to cope with their competitive business environment and need to react to these technological changes by the early detection of new trends and opportunities (Slaughter, 1997).

To do so, several foresight techniques have been developed, such as brainstorming analysis, scenario analysis, technology roadmapping, Delphi analysis and patent mining (Coates et al., 2001; Martino, 1993). Foresight methods are designed to support decision-making by analyzing future trends, technologies and innovations. These techniques primarily rely on qualitative approaches, knowledge of experts, as well as extensive desk and literature analysis. This, however, often results in an inefficient process that involves high expenditures of time, resources and costs. Furthermore, researchers criticize that traditional foresight activities are non-transparent, encourage isolated thinking and are limited to the domain-specific knowledge of participating experts (Cachia, Compañó & Da Costa, 2007; Geum, Lee, Lee & Park, 2015). Moreover, the increasing amount of data from various sources has led to information overflow that cannot be processed and analyzed manually by traditional approaches (Yoon, 2012). These points of critique have been particularly expressed about the process of technology roadmapping and scenario analysis (Cachia et al., 2007; Geum et al., 2015; Lee, Lee, Seol & Park, 2008).

The success of foresight activities highly depends on the availability of relevant expertise and information sources. Therefore, the integration of external data sources is needed to enhance the efficiency of the foresight process and broaden the view of foresight analysts. In previous years, patent and publication data has found its way into foresight research and activities (Abbas, Zhang & Khan, 2014). On the other hand, there is little research on the systematic exploitation of other data sources, such as web content. This is remarkable as through the rise of big data, many more options, actors and views exist that might constitute valuable input for analyzing future developments (Yaqoob et al., 2016). Recent studies started to exploit these new data sources, such as social media data and web news articles (Glassey, 2012; Kayser & Blind, 2017; Yoon, 2012). They suggest the use of text mining, the systematic approach to analyze textual data, to extend the knowledge base for foresight activities. It is shown that the systematic examination of web data creates several opportunities to support human foresight practitioners and entails valuable input for strategic choices and decision-making. These insights are based on exploratory research methods analyzing the development of term frequencies during several periods. In this context, Yoon (2012) focused on the detection of early indicators in datasets, so called weak signals, to identify future business opportunities. Futurologists describe the identification of innovation signals and trends as a key to anticipating technological changes (Hiltunen, 2008; Holopainen & Toivonen, 2012).

A data source that has not been addressed yet is crowdfunding data. Crowdfunding platforms, such as Kickstarter or Indiegogo become increasingly popular amongst startups, inventors and investors. For foresight analysis, crowdfunding data is interesting as it contains large datasets of technological ideas and future innovations. These ideas and technologies are already evaluated by investors and consumers, who offer financial contributions in return for non-monetary or monetary givebacks. Based on these unique characteristics, crowdfunding data bears the potential to indicate future innovation opportunities.

Therefore, the purpose of this study is to analyze crowdfunding data by means of quantitative data mining methods to find out which technologies will become 'the next big thing'. The study claims to extend previous approaches that started to analyze the emergence and development of technologies based on keywords' occurrence frequencies. The methods designed in this thesis are meant for examining technologies from multiple views, such as emerging supply, demand, investment risks and technological contexts. Furthermore, it is suggested that they are relevant for supporting different steps of the technology foresight process to enhance the efficiency of traditional foresight techniques by integrating additional information and reducing time expenditures and costs. Therefore, two of the most popular foresight approaches in strategic product planning, technology roadmaps and scenario analysis, are investigated in detail and the methods are embedded into the different phases of the foresight process. To assess and demonstrate the functionality of the proposed methods, a case study analyzing the technological field of robotics is presented.

Research goal: The goal of the research is to examine crowdfunding data to increase the knowledge base for technology foresight activities, by detecting emerging technologies and markets and thereby support the development of technology roadmaps and scenarios.

1.2. Research question

The thesis is organized around the following research question:

How can text mining of crowdfunding data be applied to analyze the detection of emerging technologies, market developments and trends and thereby support the process of technology roadmapping and scenario planning?

Text mining is the empirical approach of evaluating unstructured textual data in a systematic way. Additionally, this thesis integrates the analysis and interpretation of contextual numerical data, such as investment and success rates of crowdfunding campaigns. The study focusses on technology foresight and therefore examines the development of technologies and technological trends. The primer goal is to support foresight analysts and extend the input for existing technology foresight processes by systematically analyzing a new data source and implementing specific data-driven analysis methods.

On the way of analyzing the research question, several sub-questions are being answered. These questions have been formulated to address the potential benefits of crowdfunding data to efficiently support the foresight process of technology roadmaps and scenarios and to fulfill the requirements of successful technology foresight approaches.

Regarding the analysis of crowdfunding data:

- *Which methods can be applied to provide access and information about technological contexts and environments?*
- *Which methods can be applied to identify weak signals, emerging technologies and fields of increasing supply?*
- *Which methods can be applied to identify emerging market demand?*
- *Which methods can be applied to assess the investment risk of a technologic innovation?*

To answer these questions, new text mining based research methods are suggested that are specifically developed and customized for the use in crowdfunding analysis. These methods are designed in the form of visualization maps to make the results accessible for foresight researchers and easily reproducible in firms' technology foresight applications. Furthermore, it is examined how these methods can be integrated into the process of technology roadmapping and scenario planning to extend the knowledge base of foresight activities.

1.3. Theoretical underpinnings and structure of the thesis

The structure of the thesis follows the design science research framework as proposed by Peffers et al. (2007). This model is used to design, develop and demonstrate the application of new methods and research artifacts to resolve existing problems and limitations in the field of technology foresight.

Chapter 2 provides the literature review of the most relevant research for this study. The thesis is based on recent research on technology foresight analysis, technology roadmaps, scenario analysis and text mining. Further important concepts are future signal analysis, crowdfunding, patent analysis, and innovation management. In chapter 3, the methodological approach of the thesis is outlined. The design of the crowdfunding analysis methods is presented in chapter 4. In this part, the text mining process and data analysis methods are described. The applicability of these methods is demonstrated in chapter 5, based on the results of an illustrative case study of robot technologies. In chapter 6, the main research findings and implications are summarized and evaluated with regard to existing theory about text mining based technology foresight analysis. Furthermore, the potential practical benefits of the proposed methods are evaluated in a discussion with an innovation consulting company. Finally, limitations, as well as theoretical and practical implications of the thesis are described and possible future research directions are discussed.

2. Theoretical Framework

The theoretical framework follows a systematic structure (see figure 1). As technology foresight constitutes the overarching concept of this thesis, a profound introduction into the topic is provided and important definitions and methods are presented. The two foresight methods under investigation are technology roadmaps and scenario analysis. Therefore, recent theories about both are discussed. Then, the concept of future signal analysis is described, which is a technique to identify future technologies and emerging trends and therefore plays an important role for strategy making through technology roadmaps, as well as scenarios. In the subsequent chapters, data-driven methods, such as patent and web mining that have already been suggested for the use in technology foresight approaches are discussed. Furthermore, the principles of crowdfunding and important characteristics of crowdfunding data are outlined to show how the analysis of this data source might add relevant knowledge to the field of technology foresight analysis.

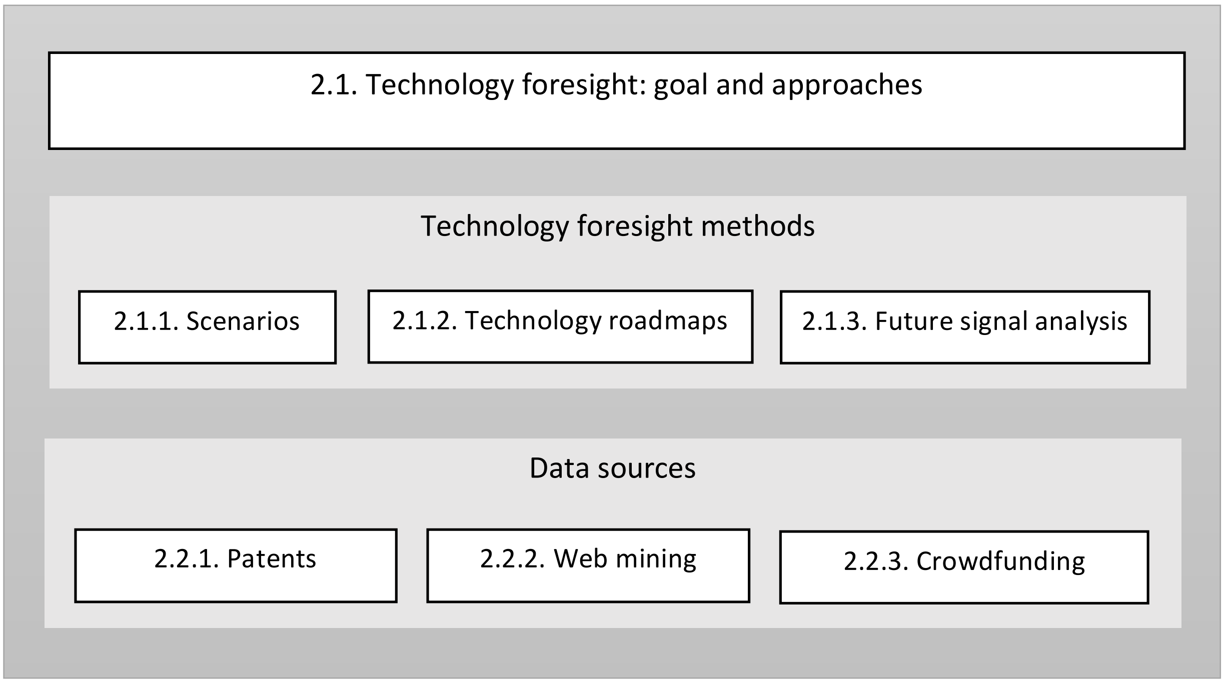

Figure 1. Theoretical framework.

2.1. Technology foresight

The endeavor to analyze future opportunities for strategic planning is of course not new. The concept of technology foresight has its roots in the 1970s, when especially in Japan national technology planning campaigns were introduced (Georghiou et al., 2009; Miles, 2010). Irvine and Martin (1984) introduced the term ‘foresight’ to describe strategic forward-looking technological analysis. The term finally encountered its breakthrough in the 1990s, when governments throughout the world implemented foresight policy tools to cope with scientific, technological and innovation-related issues (Miles, 2010; Miles, Meissner, Vonortas & Carayannis, 2017). In the following years, governmental programs and researchers used the terms ‘foresight’ and ‘forecasting’ more or less as synonyms to describe future oriented activities (Georghiou et al., 2009). This led to considerable definitional confusions as there actually exist important differences between both (Georghiou et al., 2009; Martin, 2010). Technology forecasting is an approach to predict and reflect a single future event, such as future revenues, prices or sales. This thesis focusses on the process of technology foresight. Technology foresight is a less deterministic approach that seeks to explore multiple, plausible and contingent pathways that can shape and elaborate an uncertain future (Saritas & Burmaoglu, 2015). From its beginnings as a policymaking tool, the principles of technology foresight diffused through a wide range of regions, companies and organizations during the last decades (Pietrobelli & Puppato, 2016).

Definition of technology foresight

One of the first clear definitions of foresight was published by Coates (1985, p.30): “Foresight is a process by which one comes to a fuller understanding of the forces shaping the long-term future which should be taken into account in policy formulation, planning and decision making (...) Foresight includes qualitative and quantitative means for monitoring clues and indicators of evolving trends and developments and is best and most useful when directly linked to the analysis of policy implications.“ Since then, a large number of researchers provided various definitions to outline the characteristics of technology foresight and its driving factors within different contexts, organizations or methods. Martin (1995) describes technology foresight as a systematic approach to look into the future of technologies and identify emerging fields of economic and social benefits. According to Cachia et al. (2007) technology foresight is based on three concepts: (1) Foresight considers and develops plausible views of future options; (2)

foresight does not predict the probability of future events, but serves actors as early-warning and action tool to actively shape future outcomes; (3) foresight is an open domain that should not be restricted to a small number of experts. Pietrobelli and Puppato (2016) emphasize that technology foresight supports societies and economies to define strategic areas in which the future of science and technology would lead. Researchers highlight that the participatory approach of technology foresight applications does not only increase the awareness, accountability and transparency of future technologies, but also bears the potential to influence future technological directions (Miles, 2010; Pietrobelli & Puppato, 2016). This enables the extension of strategic insights and facilitates decision making.

Technology foresight process

Voros (2003) created a generic foresight process framework that is based on prior works of Horton (1999) and Slaughter (1989). The process consists of three phases for applying foresight activities. The first phase (input phase) is about gathering information, scanning the environment and setting overall objectives, such as the time horizon and process scope. The second step (foresight work) is about the application of foresight methods, including the analysis and interpretation of the data that has been generated in the input phase. The outputs are then presented in the third phase (output phase) in the form of reports, presentations or workshops to generate an expansion of perceptions and perceived options. Voros (2003) furthermore distinguishes between the output of the foresight process (strategic thinking) and the subsequent interpretation of results that lead to the strategic actions taken by the organization. This implies that the foresight process provides relevant input for the strategic decisions in organizations.

Technology foresight methods

As different actors, such as companies, research institutes or government agencies pursue various motives to identify technological trends, several foresight techniques have been developed over time, such as brainstorming analysis, scenario analysis, national foresight studies, roadmapping methods, Delphi analysis and patent mining methods (Coates et al., 2001; Martino, 1993; Phaal, Farrukh & Probert, 2004). These methods structure the foresight process to acquire information and data, explore novel ideas, clarify situations and negotiate solutions (Saritas, 2013). Saritas and Burmaoglu (2015) examined the evolution of quantitative and qualitative methods used in practical foresight activities. They show that scenario analysis, Delphi analysis and technology roadmapping rank among the key foresight methods. These approaches are primarily based on qualitative (scenarios) or semi-qualitative (roadmaps) techniques, which may cause limitations, such as subjectivity and lack of transparency. To address these limitations, recent foresight studies suggest the implementation of quantitative approaches during the foresight process (Saritas & Burmaoglu, 2015). Therefore, current developments in foresight analysis are increasingly influenced by the integration of new data sources and the rise of web 2.0 applications. This even prompted researchers to develop the term 'foresight 2.0' (Schatzmann, Schäfer & Eichelbaum, 2013). They argue that the rapid increase of content generated by a large number of users can create completely new forms of open and collaborative foresight projects. This leads away from processes that are solely based on subjective expert opinions and results in approaches that combine their expertise with digital-collaborative intelligence. Researchers are convinced that foresight 2.0 applications enable transparent, efficient and rapid foresight approaches (Schatzmann et al., 2013). The systematic analysis of new data sources through text mining might be a building block in the process of new foresight 2.0 applications. Therefore, it will be examined how the methods developed in this thesis can be implemented into the process framework of existing foresight techniques.

2.1.1. Scenario analysis

Scenario analysis is a popular foresight method used in technology strategy development (Schwarz, 2008; Ramirez & Wilkinson, 2014; Tran & Daim, 2008). The scenario process achieved notoriety by its application at Royal Dutch/Shell, developed by Pierre Wack in 1984. Wack (1985) defined scenario planning as a discipline for "rediscovering the original entrepreneurial power of creative foresight in contexts of accelerated change, greater complexity and genuine uncertainty" (Wack, 1985, p. 150). The purpose of the process is to provide strategists with various plausible future scenarios (Mietzner & Reger, 2005). Scenarios are narrative descriptions representing a set of hypothetical future alternatives that result from a combination of trends and policies (Amer, Daim & Jetter, 2013). The goal of this process is to stimulate thinking about the future and challenge prevailing mindsets. Well written scenario descriptions should be plausible, consistent, structurally different, challenging and useful for decision making (Mietzner & Reger, 2005). Van der Duin (2016) outlines that good scenarios lead to new insights, present surprising new future realities and encourage people to step outside usual thought patterns and take action. Various researchers contributed to the literature by formulating a general framework of the scenario process (Burt, Wright, Bradfield, Cairns & van der Heijden, 2006; van der Heijden, 2005; Tapinos, 2012). Tapinos (2012) proposed a scheme for a generic scenario process that consists of eight steps: (1) Defining the scope of the exercise; (2) identifying factors of external uncertainty; (3) reducing or clustering the uncertainties; (4) developing initial scenario themes; (5) checking for internal consistency; (6) expressing scenarios in narratives; (7) assessing the impact of scenarios and (8) developing and selecting potential strategies.

Despite its benefits, the process of scenario development also entails weaknesses (Mietzner & Reger, 2005). Scenario planning is said to be very time-consuming and requires much effort. Large volumes of data and information from different sources are required to study and assess the field of research. Another point of critique is the almost exclusive focus on qualitative data and expert views that leads to non-transparent and subjective results (Hussain, Tapinos & Knight, 2017; Mietzner & Reger, 2005). Data-driven tools bear the potential to address these weaknesses and support foresight practitioners in fulfilling the requirements for well written scenarios. However, studies analyzing the use of text mining for scenario analysis are rare and its particular benefit is not clearly defined yet. This thesis suggests that the use of quantitative crowdfunding based methods might serve as valuable input to develop more plausible and consistent scenarios. Empirical methods could support experts in several stages by providing new insights and facilitating the identification of relevant factors, future signals and alternative technologies. The use of text mining furthermore facilitates the process of desk and literature research which could lead to a substantial reduction of temporal and financial expenditures.

2.1.2. Technology roadmapping

Technology roadmapping is a flexible tool to support strategic and long-range planning (Phaal et al., 2004). Its purpose is to explore and communicate the connections between organizational objectives, technological resources and changing environments. Technology roadmapping serves as a collaborative planning tool that coordinates the identification, selection and development of alternatives for corresponding product needs (van der Duin, 2016). The roadmap is the document that results from this process and represents the connections of technologies and products with market opportunities (Carvalho, Fleury & Lopes, 2013; Moehrle, Isenmann & Phaal, 2013). Researchers argue that roadmapping can lead to better investment decisions of companies or even entire industries, as it provides access to information about the identification of critical product needs and the determination of technological alternatives and milestones (Garcia & Bray, 1997; Kostoff & Schaller, 2001).

Technology roadmapping ranks among the most frequently used foresight techniques (Saritas & Burmaoglu, 2015). This popularity can be traced back to the extensive research that exists about different application fields (Barker & Smith, 1995; Battistella, De Toni & Pillon, 2015; Carvalho et al., 2013; Kostoff & Schaller, 2001; Moehrle et al., 2013; Phaal et al., 2004; Phaal & Muller, 2009; Saritas & Aylen, 2010).

As roadmapping can be conducted to address various organizational goals, several forms exist. Phaal et al. (2004) identified eight types of roadmaps: Product, service, strategic, long-range, knowledge asset, program, process and integration planning. This thesis focusses on product planning technology roadmaps. This type of roadmap is typically used to structure the development of new products and incremental innovations in firms. According to Phaal et al. (2004) product planning roadmaps constitute the most common type of technology roadmaps. The roadmap design is commonly based on the three main layers, market, product and technology, as well as their interrelations and time-based linkages. To successfully implement a roadmapping process, foresight experts have to explore and discuss the relationships between and within these layers (Phaal et al., 2004). Based on Garcia and Bray (1997) the systematic roadmapping process consists of three phases. In phase 1 (preliminary activity) the scope and boundaries for the roadmap have to be defined and a perceived need has to be identified. Phase 2 is about the development of the technology roadmap. This includes the identification of the relevant product, its related needs and critical requirements, the specification of major technological areas, as well as the identification of technology alternatives. In phase 3 (follow-up activity) the created roadmap is validated and reviewed and its implementation is developed.

Next to the advantages for strategic foresight and decision making, technology roadmapping also comprises several challenges (Yoon, Phaal & Probert, 2008). The entire process can be very time-consuming and costly. Furthermore, researchers criticize that it encourages linear and isolated thinking as it often solely relies on subjective expert opinions (Saritas & Aylen, 2010). In recent years, researchers proposed to link technology roadmapping with other foresight methods to minimize its limitations and overcome its normative and isolated character (Mietzner & Reger, 2005; Phaal & Muller, 2009; Porter et al., 2004). Particular attention was paid at combining technology roadmapping with scenario planning (Hussain et al., 2017; Kajikawa, Kikuchi, Fukushima & Koyama, 2013; Phaal & Muller, 2009; Saritas & Aylen, 2010). Researchers claim that next-generation roadmapping tools need to support experts through adding intelligence to the process (Carvalho et al., 2013; Kayser & Blind, 2017; Yoon et al., 2008). The successful creation of technology roadmaps is highly dependent on the available information about the development of markets, products and technologies (Saritas & Aylen, 2010). Four major requirements can be defined to support this process: (1) extending the knowledge base through additional input, (2) integrating information about multiple layers and their relations (marketing, product, technology), (3) facilitating the identification of trends and alternatives, (4) providing impactful input for decision making. This thesis strives to contribute to these four requirements. The use of text mining based analytics allows to access external data and integrate a huge amount of additional information that could not be processed manually. Impactful methods need to provide information about emerging technologies and market opportunities and indicate the relationships within and between the different roadmap layers. The examination of trends and emerging signals can be used to discuss and present the analyzed data and thus facilitate decision making.

2.1.3. Future signal analysis

Another important concept of foresight analysis is the detection of future signals that serve as indicators for the emergence of future events (Hiltunen, 2008). One of the most popular methods to identify these signals are weak signal analyses. Weak signals are “current oddities, strange issues that are thought to be in key position in anticipating future changes in organizational environments” (Hiltunen, 2008, p. 247). They can be understood as pre-indications and early warnings of future change. The term ‘weak signal’ was originally proposed by Ansoff (1975) who developed weak signal analysis as an alternative strategic planning tool for companies. Since then, several researchers contributed to the specification of weak signal applications (Hiltunen, 2008; Holopainen & Toivonen, 2012; Kuosa, 2010). Hiltunen (2008) introduced the term ‘future sign’ to develop a deeper understanding for weak signals. The determination of future signs depends on the visibility of signals, the number of events and the interpretation of a future sign’s meaning. While weak signals are considered as early indicators for potential trends, strong signals are characterized by an increased visibility, predictability and probability of realization (Holopainen & Toivonen, 2012). Weak signals turn into strong signals when they are supported by dedicated actors that strengthen their influence until they reach a critical mass. However, the detection of weak signals is not easy. Already Ansoff (1975) noticed that their identification requires sensitivity, expertise and creativity. Since for a long time, all suggested approaches to examine future signals relied on qualitative methods, such as interview techniques (Geerlings & Rienstra, 2003; Toivonen, 2004), discussions about the reliability of future signal analysis as a foresight method evolved. Popper (2008) reveals that due to missing guidance, the use of weak signal analysis in technology foresight is rare and quantitative empirical methods are still underdeveloped. To address these limitations, Yoon (2012) introduced a method to detect weak signals based on text mining of web news articles. By analyzing the development of occurrence frequencies of pre-defined keywords, he created keyword portfolio maps that serve as indicators for weak and strong signal terms. The analysis is based on two major propositions: (1) Keywords of many occurrences in a collection are important and (2) recent appearances of keywords are more important than past appearances. The determination of weak signals is straightforward: Keywords that have a low term frequency, but a high growth rate can be classified as weak signals, keywords with large term frequency and high increasing rate as strong signals. Park & Cho (2017) used Yoon's (2012) approach to investigate upcoming trends in the smart grid industry. Both are convinced that this method is capable of complementing interview-based approaches and can be implemented into long-term business planning processes.

2.2. Integrating external data sources into technology foresight

The creation of roadmaps and scenarios, as well as the detection of future signals is highly dependent on the integration of domain-specific expertise and knowledge. To complement and extend the knowledge base of experts, researchers started to analyze the exploitation of external data sources for technology foresight, such as patent data and web content. A detailed overview of foresight studies using external data sources is provided in Appendix 1.

2.2.1. Patent analysis

To reveal R&D trends, portend future developments and identify product opportunities, the analysis of patent data through text mining became a popular field of research in recent years (Abbas et al., 2014; Bonino, Ciaramella & Corno, 2010; Yoon, Park & Kim, 2013). The analysis of patents as a data source for technology foresight seems obvious: Patents represent inventions in a particular field of technology and are often used as indicator for the innovativeness of organizations or countries (Abbas et al., 2014). Patent databases cover a great variety of innovations over periods of time (Yoon & Park, 2004). Researchers emphasize the advantages of patent data over other data sources in terms of scope, uniqueness, objectivity, recentness, market relevance, availability, detailedness, standardization and analyzability (Gerken, Moehrle & Walter, 2010; Tseng, Lin & Lin, 2007). As of common standards, all patents consist of textual content that includes the patent title, abstract, claims, and description (Bonino et al., 2010).

The evaluation of patents is a demanding task that requires much effort and expertise (Tseng et al., 2007). The utilization of automated tools can support experts to extract and analyze the information and speed up the analysis process (Abbas et al., 2014). To process these large volumes of data, different techniques evolved over time. Abbas et al. (2014) classified common patent analysis techniques into two major approaches: Text mining and visualization techniques. While text mining is applied to extract the information, visualization methods are used to represent the extracted output, simplify the handling of relevant data and enable faster interpretations for decision-making (Keller & Tergan, 2005). Meanwhile, there exist a number of studies that deal with the use of text mining of patent data for technology foresight (Choi, Kim, Yoon, Kim & Lee, 2013; Jin, Jeong & Yoon, 2014; Lee et al., 2008; Park, Ree & Kim, 2013; Tseng et al., 2007; Yoon et al., 2008; Yoon et al., 2013).

Particularly noteworthy is the study of Lee et al. (2008) who used keyword-based text mining to derive ideas and development paths for creating new products and technologies. They suggest the use of patent maps that are embedded into a firm’s technology roadmapping process. The keyword portfolio map shows the occurrence frequency of technological keywords and their increasing rate over time. Keywords are classified into four different types: Core, emerging, established and declining keywords. While core keywords are characterized by a high frequency of appearance and a high rate of increase, the most interesting category for future opportunities is the group of emerging keywords that occur moderately often at a high increasing rate. In fact, this approach is very similar to Yoon's (2012) method for detecting weak signals. Another approach, the keyword relationship map, is based on a co-word analysis of keywords and shows the relations and co-occurrences between different attributes. Technologies that are highly related to each other could be considered simultaneously in future product designs. The applicability of the methods has been confirmed in a case study, leading to more efficient and cost-effective insights for companies.

The experiences that have been made through patent mining methods serve as an important basis for this study. The evaluation of crowdfunding data can draw upon the same underlying methodological principles, such as the application of text mining-based visualizations that have been suggested by Lee et al. (2008) and Tseng et al. (2007).

2.2.2. Web mining

Next to the analysis of patent data, first approaches exist that make use of web content to extend the input for foresight analysis. Yoon (2012), as well as Park and Cho (2017) analyze web news articles to identify weak signals. They argue that web news cover a wide range of political, economic, social and technological topics and represent a reliable and most often factual source of information (Yoon, 2012). However, as news articles do not contain standardized technological descriptions, a lot of redundant data has to be excluded from the analysis to detect technological keywords. Therefore, predefined term dictionaries are necessary to analyze the dataset and an unsupervised analysis process is rather inefficient. Since only textual data can be analyzed, results from the analysis of web news are limited to the evaluation of term occurrence frequencies.

Another promising data source is social media data which is based on user generated content on social network platforms such as Facebook or Twitter. Through the use of social media data, a large number of actors and views can be integrated in the analysis of future events. The content can either be analyzed through text mining or the application of sentiment analysis. Sentiment analysis allows to evaluate whether a certain topic is connoted with positive or negative sentiments. Cachia et al. (2007) introduce the idea to examine social media data for technology foresight. They come to the conclusion that online social networks can contribute to foresight analysis as they indicate emerging changes in social behavior and enhance collaborative intelligence and creativity. Uhl, Kolleck and Schiebel (2017) discuss the analysis of Twitter data for strategic foresight exercises. They conclude that Twitter can serve as a useful tool for gathering information in the beginning of the foresight analysis and also complement the scanning and monitoring during the ongoing process. Kayser and Blind (2017) explore the potential of text mining for foresight by considering different data sources, text mining approaches, and foresight methods. Their results show that text mining facilitates the detection and examination of emerging topics and technologies by extending the knowledge base of foresight. On the other hand, Kayser and Blind (2017), as well as Uhl et al. (2017) note that the analysis of Twitter data also entails certain drawbacks. The data cannot be retrieved retrospectively which impedes the identification of longer term trends. Furthermore, Twitter data is not representative of a society or population.

In recent years, an increasing number of studies focused on the analysis of web data to predict future outcomes or sales. Table 6 (see Appendix 1) provides an overview about current research that makes use of web-content to support foresight techniques or conduct predictive analytics. Therein, the difference between technology forecast and technology foresight becomes evident. The goal of this thesis is to provide quantitative backgrounds for strategic options to support decision making. Therefore, this study ties in with existing research about the use of web data to extend technology foresight methods. Results show that the analysis of web and patent data can lead to important implications. However, these studies also obtain several limitations, such as missing time horizons or generalizability. Thus, researchers postulate the establishment of further techniques and methods that can be embedded into the foresight process (Cachia et al., 2007; Kayser & Blind, 2017). It also becomes evident that none of these studies makes use of crowdfunding data.

2.2.3. Crowdfunding data

The development of web 2.0 applications did not only offer access to global data sources but also brought about the evolution of new business and financing models that build upon the interaction of digital users. One of these emerging approaches is crowdfunding. In recent years, crowdfunding became a serious alternative for external financing of entrepreneurial activities (Gierczank, Bretschneider, Hass, Blohm & Leimeister, 2015). Crowdfunding can be described as “an open call, mostly through the Internet, for the provision of financial resources either in form of donation or in exchange for some form of reward and/or voting rights” (Belleflamme, Lambert & Schwienbacher, 2014, p. 589). It is based on the overarching concept of crowdsourcing that is commonly used by companies to obtain ideas, solutions or feedback from a large number of people (the ‘crowd’) using information technologies (Leimeister, 2012). Crowdfunding involves the participation of three different stakeholders: Project founders, crowd funders (investors) and crowdfunding platform providers. Project founders are private persons, start-ups, SMEs, non-governmental organizations or established companies. Their main motivation is the acquisition of capital to fund projects, but also to attract attention or receive feedback for their products (Belleflamme, Lambert, & Schwienbacher, 2013). Other motives are the speed and flexibility of funding, the possibility to test products on the market, low formal obligations and to use the wisdom of the crowd

(Hienerth & Riar, 2013; Macht & Weatherston, 2014). To receive funds, founders have to explain their idea and define the scope of their project, including the target funding amount, the duration of the campaign and givebacks for investors. On the other hand, investors offer financial contributions in return for non-monetary or monetary givebacks. They are not only financially motivated, but also driven by intrinsic and social motives to be part of new, exciting and promising projects (Allison, Davis, Short & Webb, 2015; Lin, Boh & Goh, 2014; Ordanini, Miceli, Pizzetti & Parasuraman, 2011). Crowdfunding platforms serve as intermediaries between founders and investors. Platforms determine different funding rules and mechanisms to provide both sides with necessary information to reduce the risks of investments. Today, there exist various crowdfunding platforms that are specialized on different branches and industries. By 2012, there existed more than 800 active crowdfunding platforms (Crowdsourcing.org, 2012). The total funding volume of crowdfunding increased from 2.9 billion dollars in 2012 to more than 34 billion dollars per year and is expected to grow further (CrowdExpert, 2016). The total number of funding campaigns is estimated to amount to more than 13.7 million by 2021. In terms of technological innovations, alone in 2016, 14,267 technology related campaigns were successfully funded. The leading crowdfunding platform is Kickstarter, followed by Indiegogo, crowdfunder.co.uk and Fundrazr (The Crowdfunding Center, 2016). Kickstarter, which is used as data source in this thesis, was founded in 2009. Since then, over 380,000 campaigns in 14 different categories have been launched. As this thesis focusses on technology foresight, only campaigns related to this category are relevant to the data analysis process. More than 32,000 technology-related campaigns were published on Kickstarter until 2018, with a total funding value of more than 600 million dollars (Kickstarter, 2018).

Next to its growing popularity as an alternative financing tool, crowdfunding is also gaining increasing attention in research. Moritz and Block (2016) identified 127 studies that focus on different crowdfunding related topics. While a number of researches dealt with the motivations of either investors (Allison et al., 2015; Lin et al., 2014; Ordanini et al., 2011) or founders (Belleflamme et al., 2013) to participate in crowdfunding, others focused on the determination of factors for crowdfunding success (Belleflamme et al., 2014; Greenberg, Hariharan, Gerber & Pardo, 2013; Yuan, Lau & Xu, 2016). Until now, there has been no study that associates crowdfunding with technology foresight. Feldmann, Gimpel, Kohler and Weinhardt (2013) compare crowdfunding platforms with prediction markets which are simulated markets to trade and assess ideas represented by securities. They suggest the use of crowdfunding inside organizations for idea assessment as it provides better access to collective intelligence than surveys or open discussion forums. Although their study pursues a different motive, some interesting implications can be derived: (1) Crowdfunding platforms are marketplaces for ideas and products. (2) Crowdfunding serves as an effective tool for idea assessment and idea generation.

It can be concluded that the emergence of crowdfunding brought about fundamental changes to the funding of ideas, projects and startups and created various possibilities to test the demand for products on a market. After having analyzed the principles of crowdfunding, it is proposed that crowdfunding data also offers additional opportunities for technology foresight: First, crowdfunding platforms are pre-mass markets on which consumers invest in products before they are available on the actual market. Second, they contain large datasets of technological ideas and innovations. Third, investment and success rates can be evaluated. This reveals not only technological but also market and consumer perspectives for detecting future opportunities. It is assumed that these characteristics make crowdfunding particularly interesting to foresee market opportunities and adds additional knowledge to the process of technology foresight applications.

2.2.4. Selecting data sources for foresight applications

As outlined in the previous sections, recent literature proposed the use of external data sources and quantitative analysis techniques for technology foresight (see Appendix 1). It has to be noted that the output of a foresight approach highly depends on its input and that each data source is also subject to limitations. The possible use cases of the different data sources, patents, social media data, web news articles and crowdfunding are discussed in the following with regard to their specific characteristics and limitations that have been outlined in the preceding chapters.

Table 1 shows the advantages and limitations of these data sources for technology foresight. It can be concluded that patent data is especially useful for long-term technology foresight and to compare the emergence of technological inventions. Social media data is useful for short-term foresight and to observe social behavior. Furthermore, the evaluation of social networks fosters creativity and collaborative intelligence (Cachia et al., 2007). Web news are a more factual and reliable data source that can be used to analyze emerging topics of a certain technology. Finally, crowdfunding data might be the most flexible data source, as it allows to analyze textual descriptions and consumer behavior. Therefore, it can be used to conduct analyses of emerging technologies from different angles and not only compare the importance of technologies, but also examine the differences in supply and demand. This might be especially useful for strategic investment decisions. It might be beneficial to analyze various different data sources during the foresight process to include multiple views and information. This, however, would also be very time and resource consuming.

Data source	Potential fields of analysis	Advantages	Limitations
Patent data	Textual data	Standardized, objectivity, detailedness, uniqueness, technology focus	Only textual data, recent market trends difficult to observe
Social media data	Textual data and sentiments	Social behavior and changes, collaborative intelligence, recentness	Only short-term foresight, retrospective analysis not possible, not focused on technologies
Web news articles	Textual data	Objectivity, reliability, covers wide range of topics	Only textual data, not focused on technologies, only supervised learning
Crowdfunding data	Textual data and numeric values	Standardized, uniqueness, market relevance, numeric parameters, technology focus	Rather short-term foresight, restricted to certain technologies (rather B2C)

Table 1. Data sources analyzed in technology foresight.

3. Research methodology

The research methodology is based on the design science research paradigm as proposed by Hevner, March, Park and Ram (2004), Peffers et al. (2007), as well as Cleven, Gubler and Hüner (2009). In this chapter, the fundamental principles of design science research are presented and it is shown how the structure of this thesis is aligned with the design science research process.

3.1. Design science

According to Hevner et al. (2004) design science is used to extend the boundaries of human and organizational capabilities by creating new and innovative artifacts. Therefore, it focusses on the creation and development of applicable technology-based solutions for practical problems (Peffers et al., 2007). Hevner et al. (2004) defined several guidelines for conducting design science research in information service disciplines. Design science research requires the creation of innovative and purposeful artefacts that address a specified problem. The artifact itself must be relevant and useful for the solution of an unsolved and important business problem. Thereby, it draws from existing theories and knowledge. Finally, its utility, quality and efficacy must be demonstrated and evaluated.

This thesis follows the design science research approach and seeks to resolve existing limitations in the field of technology foresight analysis. To produce explicitly applicable research solutions, new exploratory quantitative data mining methods are designed that systematically analyze crowdfunding data to identify future signals, trends and opportunities. The validity, reliability and reproducibility of these methods is demonstrated by the application of illustrative case studies.

3.2. Design science research process

To provide guidance on structuring design science research, Peffers et al. (2007) created a methodology framework for design science research in information systems. The process consists of the following steps: Problem identification and motivation, definition of the objectives for a solution, design and development, demonstration, evaluation, as well as communication. After having outlined the main theoretical concepts for this thesis, the motivations and drivers can be summarized and the research model can be specified, including the planned design, demonstration and evaluation of the research artefacts to be developed in this thesis.

Problem identification and motivation

Foresight analysis and the identification of technological trends is a field of major interest for various stakeholders, such as startups, SMEs and large companies. It became evident that foresight methods foremost rely on subjective perceptions of experts and manual literature analysis. This contradicts with the aspiration of technology foresight to involve the participation and input of various views and leads to costly, non-transparent and inefficient foresight processes. Therefore, voices have been raised to integrate systematic methods and external data sources into traditional foresight activities. First approaches attempt to improve the efficiency of foresight through quantitative data. However, there is still a lack of clear guidance for effective empirical foresight methods. Furthermore, existing research approaches are limited to the analysis of textual data (Lee et al., 2008; Yoon, 2012; Kim & Lee, 2017).

Definition of the objectives for a solution

To enhance the efficiency and transparency of foresight processes and serve managers with relevant information, the main requirements to support technology foresight approaches have been examined. Successful technology roadmapping and scenario applications require profound insights about emerging technologies, markets and trends, as well as knowledge about alternative technologies and fields of uncertainty. Methods have to be developed that support foresight experts in each of these steps and data sources have to be analyzed that allow to draw conclusions on these topics. These methods are ideally built on systematic quantitative analyses to analyze large volumes of data and thereby enhance the efficiency and transparency of foresight approaches. Foresight methods need to be easily understandable, valid, reliable and reproducible. Furthermore, clear guidance has to be provided on how to implement these methods during the foresight process.

Design and development (see section 4)

Based on the different variables that can be accessed through crowdfunding data, this thesis suggests the design of new quantitative approaches that combine the analysis of textual and numerical data. These methods are aligned with the just mentioned objectives. The procedure of applying text mining algorithms to crowdfunding data is thoroughly described in the subsequent chapter. The systematic analysis of the different data points does not only reveal technological, but also consumer perspectives for detecting future opportunities. After systematically analyzing the data, specific methods are used to make the results accessible and interpretable for foresight analysts and managers.

Demonstration (see section 5)

The functioning and applicability of the proposed methods is demonstrated by examining an illustrative case about the development of robot technologies. According to Cleven et al. (2009) and Peffers et al. (2007) case studies are a valid research method to demonstrate the reliability and validity of designed methods. Therefore, first the application and results for each method are presented in section 5. Then, it is shown how they can be used during the process of technology roadmapping and scenario planning. This is achieved through the presentation of two narrative case studies to capture the essential meaning of the suggested methods.

Evaluation and communication (see section 6)

The usefulness, applicability, as well as theoretical and practical implications of the new methods are discussed with regards to existing theories. To evaluate the benefits of the proposed analyses of crowdfunding data, the methods are discussed with data analyst experts at a German innovation consulting company. Communication finally refers to the overall statements and conclusions that are made available through this thesis.

4. Design of a crowdfunding data analysis for technology foresight

The following sections describe the design of a systematic crowdfunding data analysis to be applied in technology foresight activities. The chapter is separated into five major parts: First, the collection and characteristics of the underlying crowdfunding data is described. Then, the overall data analysis process is outlined. This study suggests a systematic text mining procedure to process the unstructured datasets that is a necessary prerequisite for the application of specific foresight methods. In section 4.4., the possible impact of crowdfunding based methods on foresight activities is illustrated and the design of a data-driven technology foresight process is proposed. Finally, the design and development of six different crowdfunding foresight methods are presented.

4.1. Data collection

The research is based on crowdfunding data from Kickstarter, which is one of the most famous and biggest crowdfunding platforms. The data used in this study has been crawled and made publicly available by the data analytics company Webrobots using a Python web-crawler via the Kickstarter API (Webrobots, 2018). From these datasets, only those campaigns that are assigned to the category 'technology' have been filtered out. The final dataset entails data from 2009 until November 2017, in total more than 26,200 campaigns. The dataset contains diverse information about the single campaigns in this period: The title of the campaign, a short description, number of supporters, starting and end date, target funding amount, the amount actually funded and the link to the campaign page. To conduct trend analysis, the data has been chronologically sorted and semi-annual datasets have been created.

4.2. Data analysis

Although not only textual data is analyzed, the main focus of the thesis is a text mining analysis of the crowdfunding campaigns. The text mining process is needed as a preliminary work to process the unstructured data and enable the implementation of further analysis techniques. There exist a number of different tools to conduct text mining analysis, such as SAS, KNIME, Weka or RapidMiner (Chen, Mao & Liu, 2014). In this thesis, the text mining software RapidMiner is applied. RapidMiner is an open source software used for data mining, machine learning, and predictive analytics. According to Chen et al. (2014) it ranks among the five most widely used data mining and big data software tools. The software has been chosen as it provides the necessary tools to conduct advanced text mining tasks of large datasets and offers various options to pre-process the input as well as evaluate its output. The results from the text mining process are then analyzed in a specifically developed data dashboard. There, the analyzed term frequencies are combined with the contextual data, such as the amount of investments and the success rates of campaigns. This allows to examine and calculate the development of different parameters related to a certain technology. To facilitate and enable the causal interpretation of results, this thesis suggests the application of different visualization methods. Visualizations support analysts to deal with large volumes of data and have proven their use in strategy making (Card, Mackinlay & Shneiderman, 1999; Keller & Tergan, 2005). This thesis draws upon the findings of Yoon (2010), who suggests the use of visualization methods, such as maps, curves and networks for technology foresight. These methods have been selected for three reasons: (1) They facilitate the interpretation of results to be applied in foresight applications and strategy development. (2) They enable the presentation of available indicators of crowdfunding data. (3) They can be implemented and adjusted to the requirements of technology roadmaps and scenario analysis. A simplified overview of the analysis framework can be seen in figure 2.

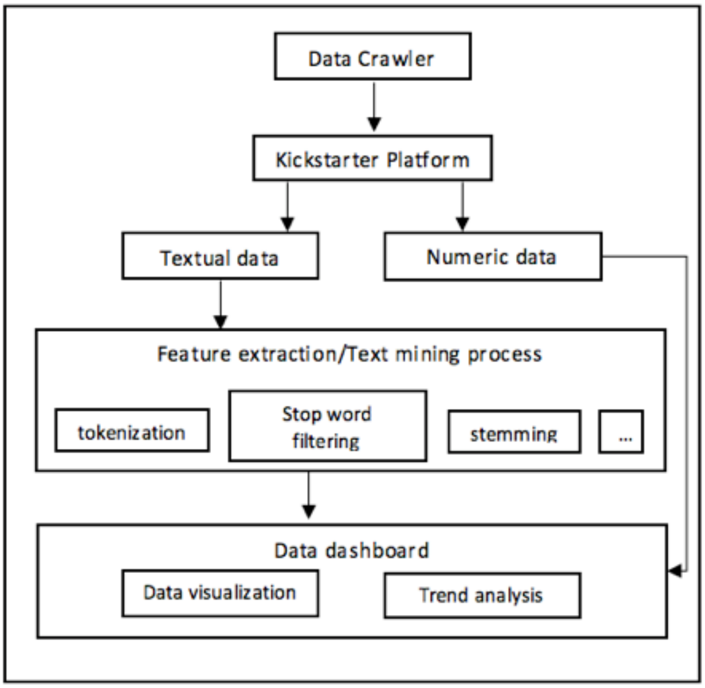

Figure 2. Data analysis framework.

4.3. Text mining of crowdfunding campaigns

Text mining is a method to analyze and process textual data (Sanger & Feldman, 2007). It combines techniques from data mining, machine learning, natural language processing, information retrieval and knowledge management (Berry & Kogan, 2010; Sanger & Feldman, 2007). In recent years, text mining methods are increasingly used to process the growing volume of unstructured data from heterogeneous sources, commonly known as big data. Text mining approaches follow the common process framework of data mining applications, consisting of four main steps: Pre-processing tasks, core mining operations, presentation of layer components and refinement techniques (Sanger & Feldman, 2007; Weiss, Indurkhya, Zhang & Damerau, 2005). As results of text mining processes are often limited and not self-explanatory, their interpretation is crucial for the use in foresight analysis. Therefore, they should be conducted by domain experts that are able to embed the findings into the context of the foresight process (Kayser & Blind, 2017).

This chapter provides a detailed description about the text mining method applied in this thesis. First, the data has been manually prepared and chronologically sorted in CSV documents. Then, the different documents were imported into the RapidMiner software to conduct the systematic extraction of relevant keywords. As crowdfunding campaigns are already assigned to a certain subcategory, such as robots or software, the type of campaigns to be analyzed can already be determined upfront. The processing of the document consists of seven steps that convert the text files into a structured keyword-document matrix (see figure 3). The matrix indicates how often a keyword occurs in the crowdfunding dataset. First, the textual parts of each campaign have to be tokenized. Tokenization is the process of dividing textual documents into the individual elements. These tokens constitute attributes in the final example set. To refine the results, several structuring operators have been executed. All tokens have been adjusted to lower cases through the transform cases operator. Then, stopword filters and manually defined stopword lists have been applied that filter out irrelevant and unwanted tokens from the example set, such as articles or prepositions. Subsequently, a stemming operator has been used to simplify the conversion of unstructured text into structured data. Stemming is the process of reducing words to their basic roots, as often words occur that have the same meaning and word stem, but are used in different variations. In this study, the Porter stemming method has been applied that is based on the premise to exclude the suffix of words. It is the most common stemming method in text mining analytics and proved to be very efficient (Hofmann & Chisholm, 2016). Finally, the generate n-gram operator has been executed. This operator enables the aggregation of words that typically occur together, such as for example smartphone and app. The use of n-grams, typically bigrams or trigrams, offers new and useful insights for statistical evaluations.

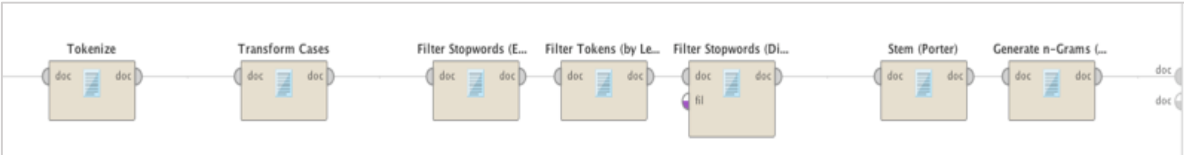

Figure 3. Text mining process in RapidMiner.

The output of this process leads to a document vector file that indicates the occurrence frequency of terms. The results from the text mining process were then analyzed in a specifically developed data dashboard. There, the determined word frequencies have been combined with the contextual data of the respective campaigns, such as the amount of investments and success rates. The values of different periods were used to calculate the increasing rates of term frequencies (TF) and investment rates. The generation of co-word networks is based on the applications of TF-related correlation matrices using RapidMiner.

To make the analyzed values accessible and interpretable for foresight practitioners, they were processed and aggregated into the corresponding methods that are presented in the following sections.

	Term frequency (TF)	TF/ number of Campaigns	$ invested (I)	Investment goal	I/total investments	Investors count	Success-rate (inverse risk-rate)	Spending/ investor
Android	47	0.018	618,541	1,718,550	0.0364	9,472	0.234	65
App	541	0.206	3,857,636	32,084,471	0.2272	29,476	0.114	131
Arduino	37	0.014	860,926	670,129	0.0507	9,080	0.702	95
artificial intelligence	7	0.003	22,714	1,191,500	0.0013	465	0.285	49
Battery	48	0.018	3,807,072	3,496,380	0.2243	10,396	0.416	366
blockchain	2	0.001	6,808	3,210,000	0.0004	75	0	91
bluetooth	28	0.011	4,375,674	1,049,500	0.2577	33,653	0.5	130
Camera	53	0.020	3,977,357	2,466,610	0.2343	23,667	0.471	168

Table 2. Data dashboard example.

4.4. Design of a data-driven technology foresight process

Data mining analysis of web data and the subsequent visualization and interpretation of results can provide several insights for experts about future technological trends. To support strategic planning in organizations, data-driven methods have to be embedded into structured foresight planning processes in which decision makers discuss the overall strategic orientation. Existing text mining research approaches are predominantly limited to supporting technology foresight processes in a specific stage, most commonly in the beginning of the process to support desk and literature research. This thesis strives to follow a different, more extended approach. Since crowdfunding data offers many data points for analyzing emerging technologies from different views, this thesis pursues the claim to implement these findings into structural foresight processes to assist foresight practitioners during different steps. The results of the proposed methods require contextual interpretation and should be understood as supportive tools for the analysis of external information. Therefore, this thesis suggests the design of a data-driven foresight process that combines qualitative thinking and quantitative methods.

Figure 4 shows how this foresight process is built up and how new data-driven methods are implemented during the different steps. Based on the conceptualizations of Voros (2003), Tapinos (2012), Garcia and Bray (1997), as well as Kayser and Blind (2017) the general foresight process is divided into four main phases: (1) Preliminary activities/definition of scope and search field, (2) identification of market, product and technology developments, (3) generation of the technology roadmap. The fourth (4) phase, labeled as follow-up activity, involves the validation of the results. During this process, analyses of the retrieved documents are conducted and internally discussed to identify the key factors and finally concretize these findings through the creation of roadmaps or scenario stories. These steps are supported by the data-driven analysis methods proposed in this thesis. For every phase a method is developed to add relevant information that facilitates strategic decisions and supports experts in doing their work.

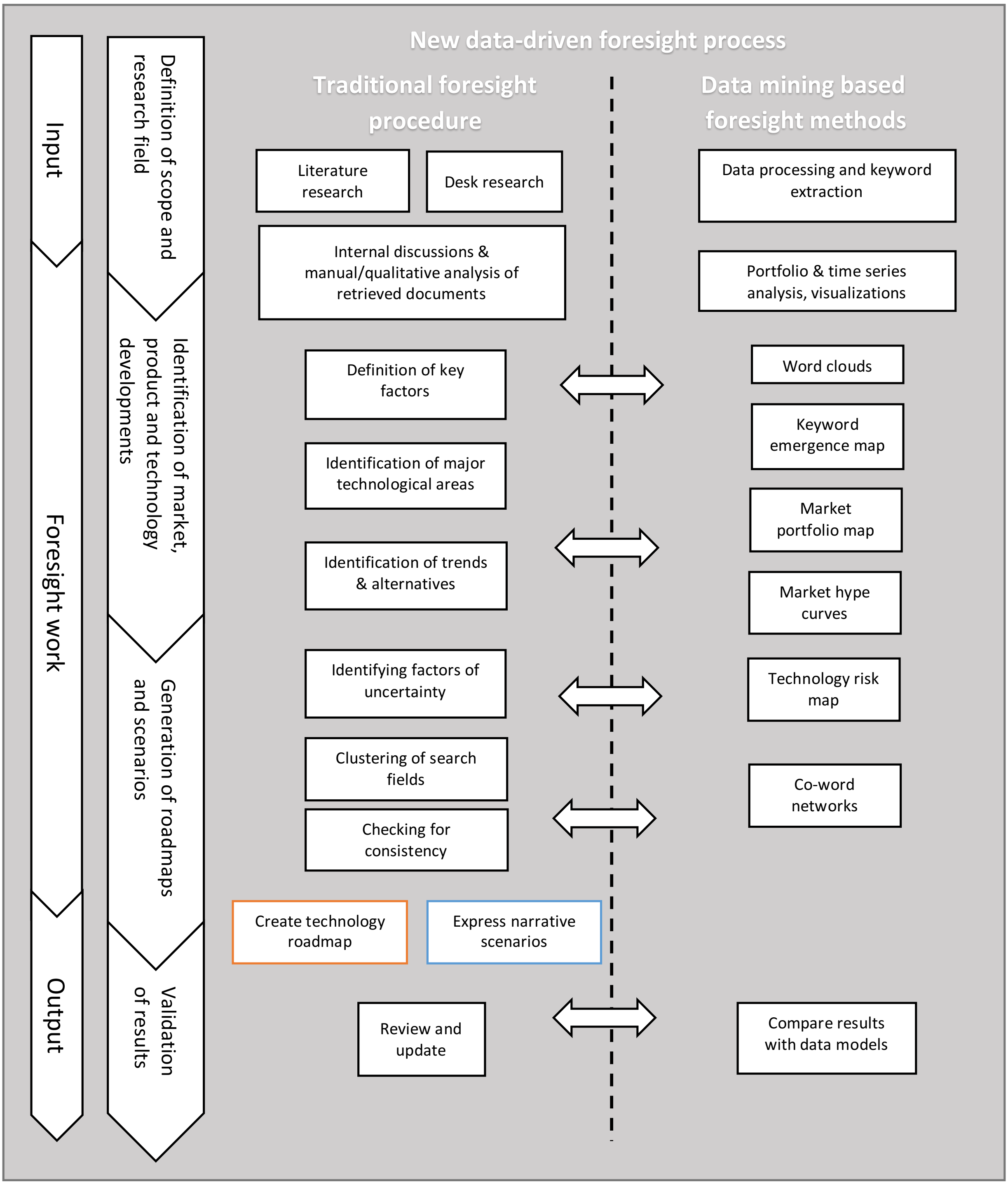

Figure 4. Data-driven technology foresight process.

4.5. Design and development of effective crowdfunding based foresight methods

The following sections describe the development of six methods that are designed to derive useful results from the crowdfunding datasets to support technology foresight research. The analysis methods are inspired by the propositions of Lee et al. (2008), Yoon (2010), Yoon (2012) and Kayser and Blind (2017), who conducted first research on the support of technology roadmapping and scenario planning through text mining analysis. The methods are based on different variables that can be extracted from crowdfunding data (see table 2). Next to term frequencies (TF), the amount of investments (I) and success rates are particularly examined. Through the application of different analysis techniques such as co-word analysis, networks and the conjunction of textual and numerical data, specific methods are developed that are aligned with the requirements of foresight planning as proposed in section 4.4.. The methods are based on a general proposition regarding the quantitative analysis of crowdfunding campaigns.

Proposition 1. *Campaigns that contain a certain technological keyword of interest primarily deal with the development and planned commercialization of this specific technology.*

4.5.1. Word clouds

After the decision of a company has been made to conduct foresight activities, the scope and boundaries for the respective situation have to be set. To clarify the topic under investigation, extensive desk research and brainstorming sessions have to be conducted. Explorative text mining methods are capable of processing large volumes of data within a short time frame that also lead to new and relevant insights. A simple and fast approach to get a first understanding of the research topic and identify relevant requirements is the creation of word clouds that have been already proposed by Kayser (2016), as well as Hofmann and Chisholm (2016) as an easy to handle visualization tool for textual data. Word clouds visualize the most important topics, represented by the occurrence of words. They can be directly derived from the text mining wordlists and are created by online tools, such as wordle.net (Hofmann & Chisholm, 2016). The creation of word clouds does not require the use of predefined word dictionaries, as the goal is to include the entire information available and obtain insights from various views, such as technology, business, marketing and manufacturing.

4.5.2. Keyword emergence map

As data from several periods is analyzed, trend analysis can be conducted that is based on the methods of Yoon (2012) and Lee et al. (2008). They argue that the evaluated increasing rates of term frequencies allow to draw conclusions on the detection of strong and weak signals: Based on the theories of Hiltunen (2008), topics that have a high increasing rate but a low absolute occurrence frequency possibly relate to weak signals. This thesis reincorporates these findings in the context of crowdfunding campaigns. The crowdfunding data from the years 2009 until 2017 is divided into 15 time periods. Each period lasts 6 months, except for the years 2009 and 2010, which are aggregated to one period, as the amount of data has been very rare in the early phase of Kickstarter. To control for the varying number of total campaigns in different periods, the term frequency (TF) is divided by the total number of campaigns to calculate the relative term frequency (RTF). Then, the average increasing rate of the RTF in the different periods is calculated.

$$RTF = \left(\frac{TFij}{NNj}\right)$$

TF: term frequency; NN: total number of campaigns; i: certain term; j: point of time

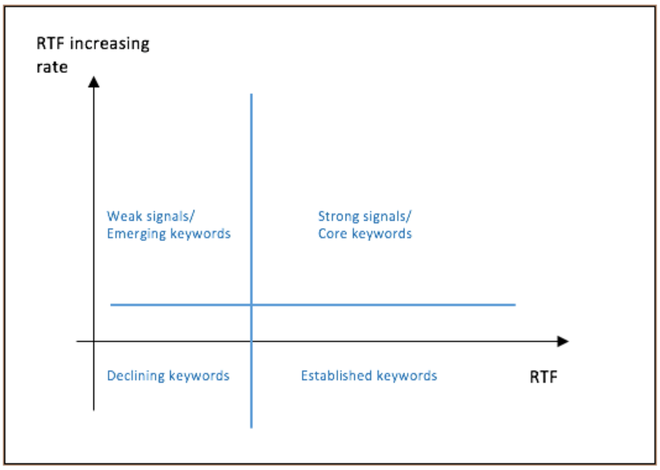

Figure 5. Keyword emergence map.

The results can finally be visualized trough the keyword emergence map to simplify their interpretation and show the different developments of the analyzed technologies (see figure 5). In accordance with Lee et al. (2008), it is argued that a high occurrence frequency of a technological keyword suggests a high innovativeness in this field of research. Simultaneously, a significant increase of keywords indicates that this technology is becoming increasingly important. The keyword emergence map hence represents a method that indicates if a certain technology is in an emerging, core, established or declining state.

Proposition 2. *Technological keywords of many occurrences in a dataset indicate high innovativeness and R&D activities in this technological field.*

4.5.3. Market portfolio map

The marketplace character of crowdfunding platforms provides the opportunity to analyze demand-specific economic measures, such as investment rates. The market portfolio map is designed to serve as an indicator for the demand for technologies. The total amount of investments related to a technology keyword are plotted on the x-axis and the increasing rate of investments during the considered periods is plotted on the y-axis. Thereby, the market portfolio map indicates how strong the demand for a certain technology might be and how it evolves over time (see figure 6). This can be especially useful for the market perspective of technology foresight activities.

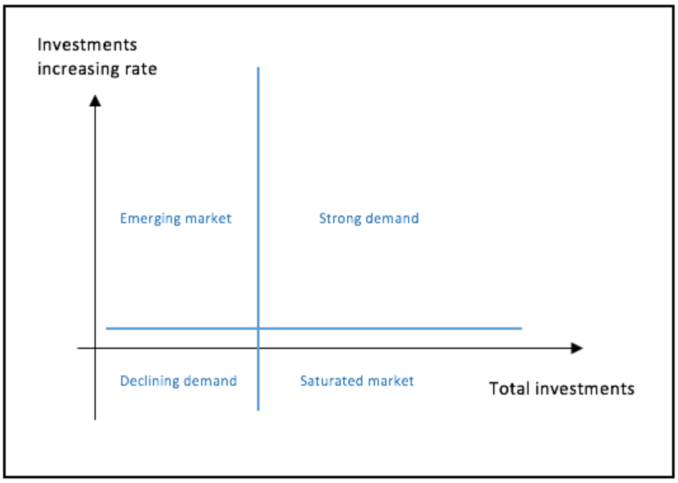

Figure 6. Market portfolio map.

Proposition 3. *The amount of investments and the investment increasing rates related to a technological keyword indicate the market demand for the respective technology.*

4.5.4. Technology risk map

To assess the risk involved with the development of technologies, the technology risk map is designed. It makes use of the success rates (inverse risk rate) of campaigns associated with the occurrence of a certain keyword. For each keyword, the corresponding success rate is plotted on the y-axis, while the RTF is plotted on the x-axis (see figure 7).

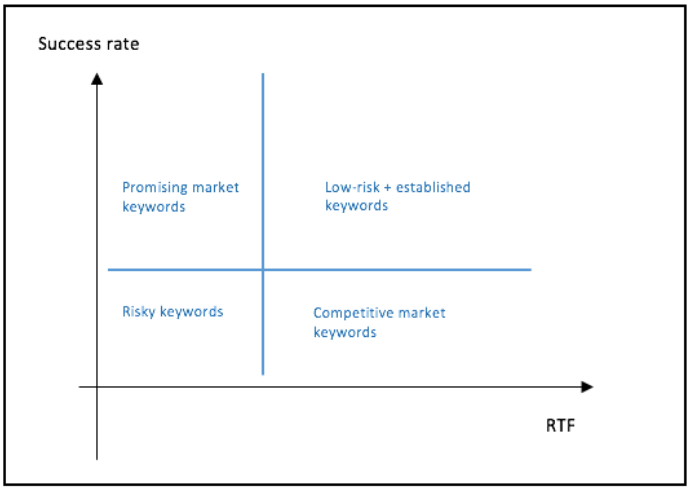

Figure 7. Technology risk map.

The technology keywords can then be classified into four different types: Risky, competitive market, promising and low-risk/established keywords. If a technology keyword appears very frequently and its related projects tend to succeed, an investment in this technology is associated with relatively low risk. On the other hand, if a keyword appears very frequently but the average success rate of related campaigns is very low, this indicates that the market is very competitive and the probability to fail is comparatively high. The map provides relevant support for investment decisions and is especially useful for the business perspective of technology foresight.

Proposition 4. *The higher the success rate of campaigns that are related to a certain technological keyword and the more campaigns exist for this keyword, the lower is the risk to develop or invest in this technology.*

4.5.5. Market hype curve

To evaluate the development of a specific technology in more detail, the market hype curve approach is developed that integrates the analysis of both, term frequencies and investment rates of a single technological keyword over time. The market hype curve is used to visualize meaningful relations between the RTF and the relative amount of money that is invested in a certain technology. The relative investment rate (RIR) is calculated by dividing the amount of investments of a single technology by the total amount of investments in this period.

$$RIR = \frac{I_{ij}}{TI_j}$$

I: amount of investments; TI: total amount of investments; i: certain term; j: point of time

The approach is based on a time series analysis to examine the evolution of term frequencies and investments. The analysis of historical data points can be used by managers to make decisions, based on the assumption that previous developments will continue into the future. Figure 8 shows that the time is plotted on the x-axis, while the y-axis indicates the values for RIR and RTF.

To facilitate the interpretation of time series, linear trend curves are calculated based on the development of RTF and RIR that indicate whether these values are generally increasing or decreasing. Since the development of RTF and RIR on crowdfunding platforms is often volatile, linear trend curves serve as indicators for the overall long-run developments of technologies. They are calculated through the following equation:

$$Y = m \times t + a$$

Y: projected y value for the selected t value
t: selected period of time
m: slope of the line, average change in Y for each time period t
a: estimated value of y when t=0

Based on proposition 2, term frequencies are associated with the degree of innovativeness in a technological field. This is also an indicator for technology supply and market competition. On the other hand, high investments indicate a strong market demand for technologies (see proposition 3). This leads to the proposition that the development of RTF and RIR displayed through market hype curves serve as indicators for technology supply and market demand of technologies (see proposition 5).

Since the values for RTF and RIR do not always evolve in parallel, it is proposed that there are phases in which a market hype or oversupply for technologies exists. In phases of market hype, the RIR exceeds the RTF (see figure 8). This means that consumers invest a relatively high amount of money in the respective technologies, while the amount of innovations in this technological field is comparatively low. Thus, the market hype curve does not only indicate if a technology is becoming more relevant in terms of innovativeness and market demand, but also if there currently exists a hype or oversupply for the respective technology.

Concerning the linear trend curves, the most interesting technologies are those that show a high increasing rate in terms of RIR and a moderate increasing rate in RTF, while the investment curve surpasses the term frequency curve (see figure 9). In this case, the supply and demand for a technology are increasing. However, since the RIR curve exceeds the RTF values, this indicates that there is still potential for further innovations. On the other hand, declining term frequency and investment curves imply that a certain technology is disappearing from the market (see figure 10).

Proposition 5. *The developments of relative term frequencies and relative investment rates serve as an indicator for technology supply and market demand.*

Proposition 5.1. *A period in which the relative investment rate is higher than the relative term frequency indicates a market hype (market demand > technology supply).*

Proposition 5.2. *A period in which the relative investment rate is lower than the relative term frequency indicates a surplus in supply (market demand < technology supply).*

Proposition 5.3. *An overall decrease in relative term frequency and relative investment rate implies that a technology is disappearing from the market.*

Proposition 5.4. *An overall increase in relative term frequency and relative investment rate implies that a technology is becoming more important at the market.*

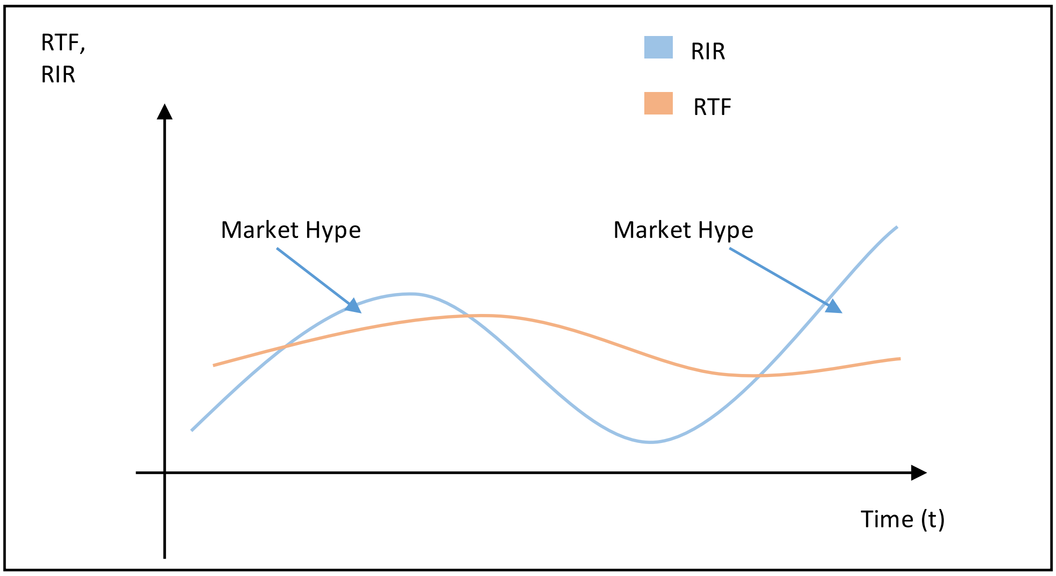

Figure 8. Market hype curve.

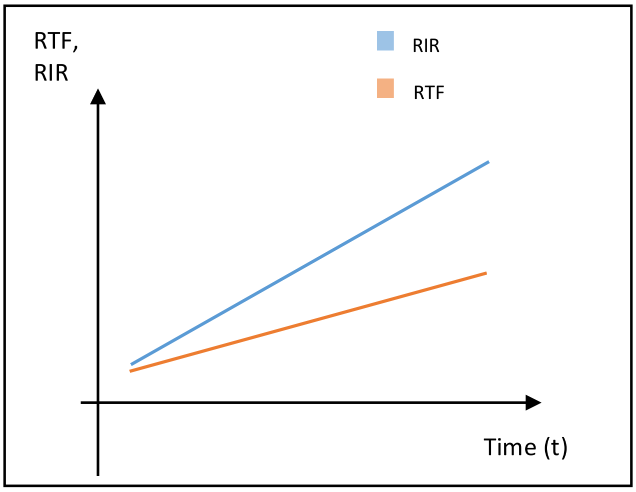

Figure 9. Market hype curve showing an emerging technology.

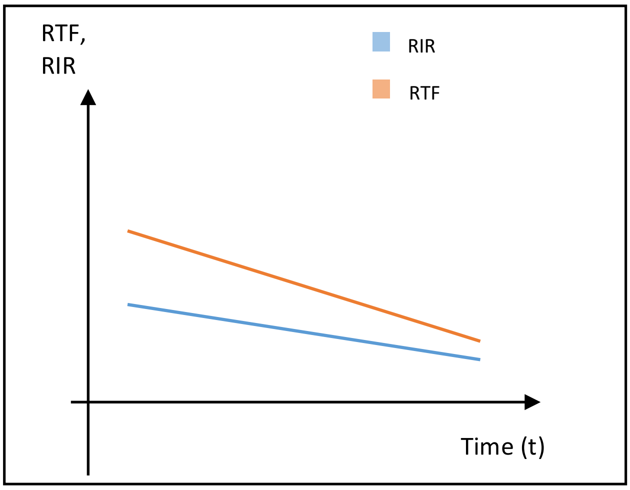

Figure 10. Market hype curve showing a declining technology.

4.5.6. Co-word networks

After the main trends, drivers and topics have been identified, the technology roadmaps or scenarios have to be created. In this phase, the relations within and between these attributes play a major role. For technology roadmapping, the different layers have to be connected and dependencies between objects need to be identified. For scenario analysis, the information has to be aggregated and combined to formulate future alternatives. In both cases, co-word analysis of the identified keywords can be used to visualize meaningful relations and dependencies in the form of networks. Term networks display the relations (edges) between several terms (nodes). The structure of the network indicates the importance of a certain term (by the number of connections), and also whether a node connects a pair of other nodes and therefore plays an interconnecting role in the network. The number and positioning of nodes and edges implies the density of a technological field (Yoon, 2010). This thesis proposes the creation of co-word networks to identify and present links between emerging technologies and related attributes. The goal is to use the information from crowdfunding data and create networks that indicate the correlations between technologies, products and market/application attributes. Figure 11 shows how these three-layered technology networks can be visualized to provide relevant insights in a structured and efficient way.

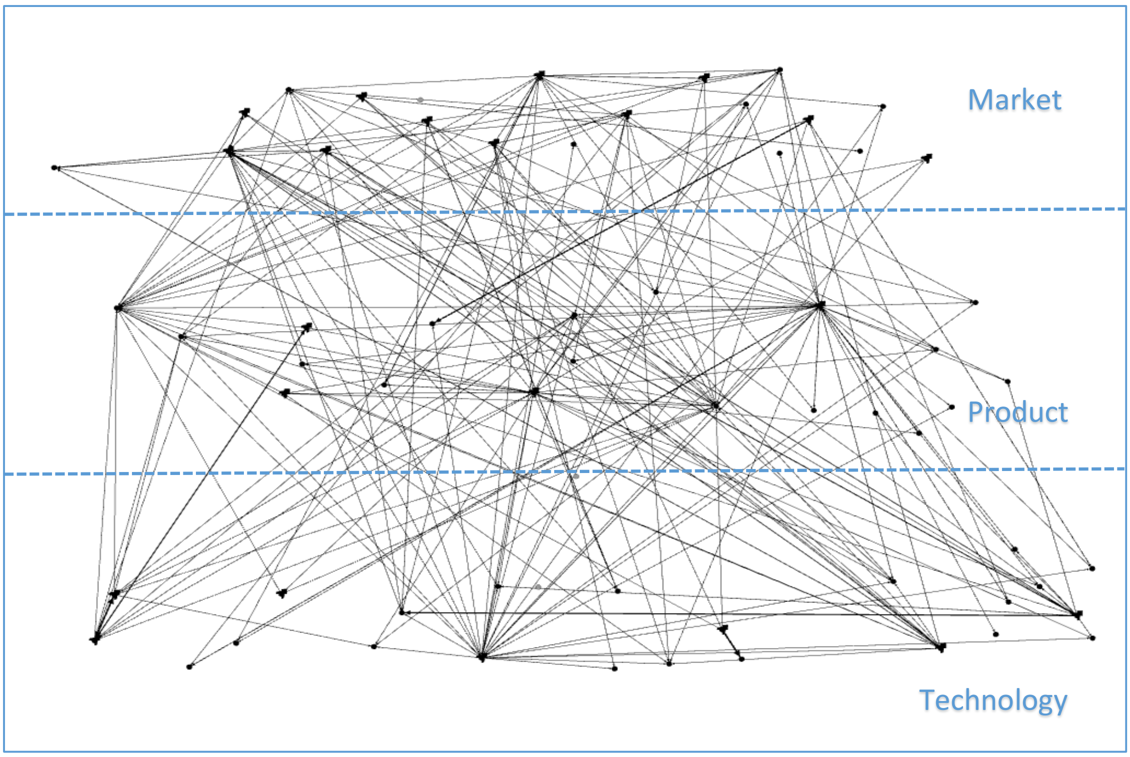

Figure 11. Technology network with three layers: Market, product and technology.

5. Demonstration of crowdfunding foresight methods

In this chapter, the functionality and applicability of the proposed crowdfunding foresight methods is demonstrated. The technology ‘robot’ has been selected as an example case to show how the suggested methods provide different insights to the evaluation of a specific technological field. The methods are described in the chronological order of the technology foresight process that has been presented in section 4.4.: (1) Definition of search fields, (2) identification of technology, product and market trends, (3) creation of roadmaps or future scenarios. Therefore, first the results of the consecutive methods are presented, then it is summarized in section 5.4. how these results can be practically implemented during the foresight process to create technology roadmaps and scenarios.

5.1. Definition of search fields

To get an initial overview of the different aspects of a certain topic, the word cloud approach has been applied. As an example case, the superordinate topic ’robots’ has been chosen. Therefore, only campaigns that belong to the category robot have been analyzed through text mining. Furthermore, data from the years 2016 and 2017 has been evaluated to analyze current topics. As a result, a list of 115 keywords has been created that was used as input for the creation of word clouds. The term robot itself has been excluded from the list as it occurred disproportionally often and would have distorted the final visualization results. The font size of each term in the cloud displays the frequency of terms in the analyzed dataset and therefore indicates its overall importance for the topic. The final result is illustrated in figure 12. It shows that the topic ‘robot’ embraces a wide field of heterogeneous subtopics. Especially prominent is the term ‘programming’. This seems to be a key activity in the development of robots. Furthermore, several other technologies can be identified, such as drones, engines, laser, sensors, cameras, 3D printing, apps or artificial intelligence. The word cloud also shows different application fields of robots. Topics such as education, assistance, toy, game, children or work become apparent and specific product characteristics such as precision, interaction or modularity are mentioned. From a market perspective, affordability seems to play a major role.

It can be concluded that the word cloud provides a broad overview for different fields that play a major role in the current development of robots. Alternative technologies, potential application fields and product characteristics are displayed, which are important perspectives that have to be considered in experts' discussions of foresight applications.

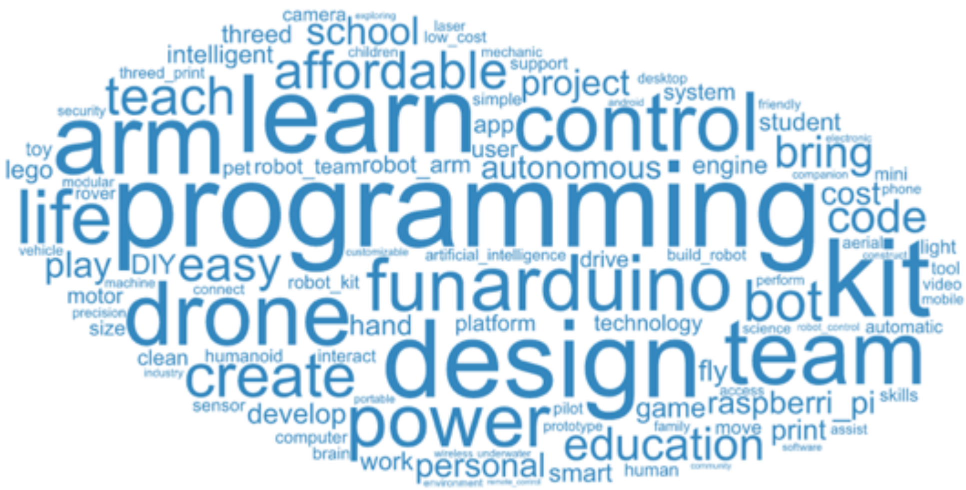

Figure 12. Word cloud for the term robot.

5.2. Identification of technology, product and market trends

The second phase of the foresight activities involves the analysis of technology and market developments. This includes the specification of major technological areas and drivers and the identification of potential alternatives. To support these tasks, five different analysis methods are applied: Keyword emergence map, market portfolio map, technology risk map, market hype curve and co-word networks.

5.2.1. Emerging technologies

In this part, the application of keyword emergence maps is demonstrated and results for the identification of weak signals and emerging technologies are presented. The map is based on a predefined list of technologies that were derived from the previous creation of the word cloud (e.g. robots, drones, 3D printing, cameras, apps), but also contains further recurrent technological innovations derived from the text mining process, such as wearables, tablets or smartphones for a more comprehensive assessment. To show how different innovations evolved over time, the entire time period was divided into three sub-periods, from 2009 until 2013 (see figure 13), from 2013 until 2015 (see figure 14) and from 2015 until 2017 (see figure 15). The allocation into these periods has been chosen as it shows developments and changes in innovation activities more clearly than yearly or half-yearly evaluations. For the classification into the four different categories (weak signals, strong signals, declining keywords, established keywords), the following parameters have been set: Keywords of less than 1% occurrence frequency and more than 10% increasing rate in a period are considered as weak signals, keywords of more than 1% occurrence frequency and more than 10% increasing rate as strong signals.

Period 1 (2009-2013)

The results show that due to the number of campaigns and the term increasing rate of nearly 16%, robots are considered as a strong signal technology (see figure 13). Among these are also technologies such as cameras, smartphones, tablets and the open source microcontroller Arduino. Particularly interesting is the group of emerging keywords in the upper left corner of the map. There, technologies that are rarely exposed but experience a high increasing rate are listed, such as wearables, virtual reality, streaming and drones. According to the word cloud approach, drones are nowadays among the most important robot-related technologies. A few years ago, the number of drone innovations was still very small but increasing, which are indicators for a weak signal technology. On the other hand, laser technologies or desktop computer rank among the declining keywords.

Period 2 (2013-2015)

In the following period the total number of robot innovations is still comparatively high (see figure 14). However, the innovativeness tends to decline, indicated by a negative growth rate (- 20%). Technological innovations such as AI (artificial intelligence) and IoT (Internet of things) appeared and show a very high rate of increase (190% and 75%). Particularly interesting is the development of wearables over the periods 1 and 2. While wearables counted to the group of weak signals in the previous period, they became a very strong signal in period 2.

Period 3 (2015-2017)

Period 3 is the most relevant period from today's perspective, as decision makers are usually interested in technologies that are emerging in present periods. During the years 2015 to 2017, the innovativeness in the field of robots again increased at a rate of almost 25% (see figure 15). Other robot-related technologies that came up in these years are autonomous technologies and augmented reality. Particularly noteworthy is the rise of the blockchain technology. There is still a very low amount of blockchain innovations on crowdfunding platforms, however, the increasing rate is very high (310%). These characteristics are associated with an emerging technology/weak signal.

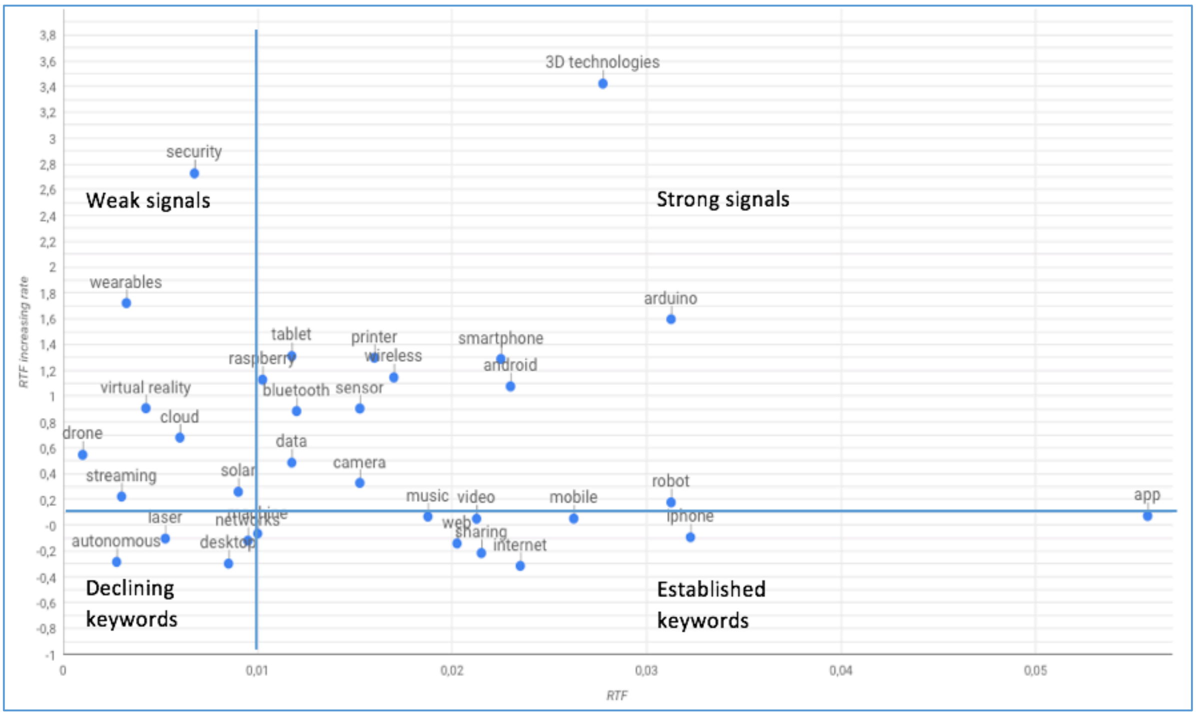

Figure 13. Keyword emergence map, years 2009-2013.

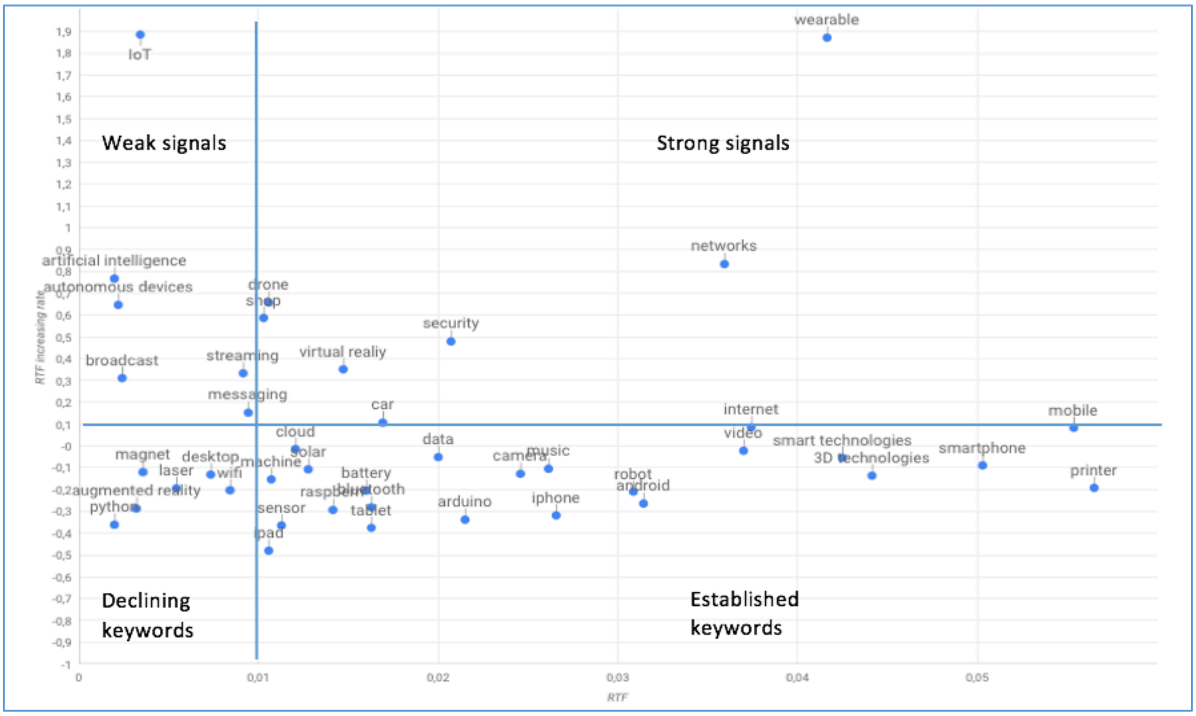

Figure 14. Keyword emergence map, years 2013-2015.

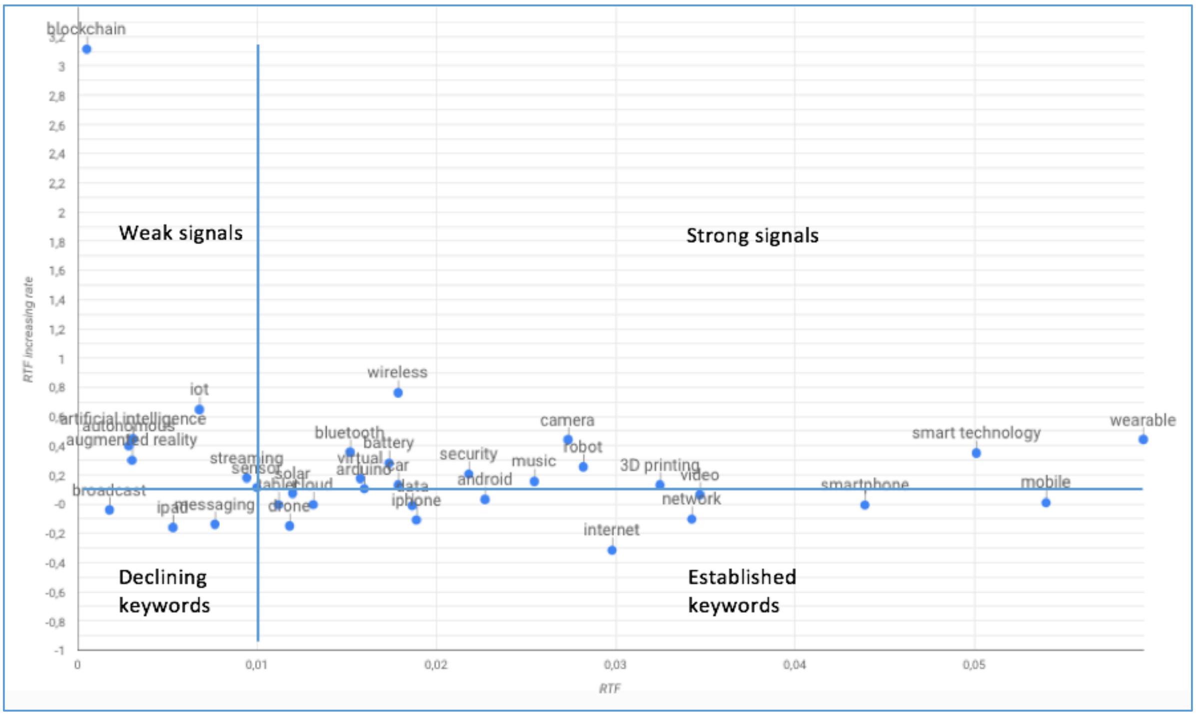

Figure 15. Keyword emergence map, years 2015-2017.

5.2.2. Emerging markets

The market portfolio map is used to display the market demand for technologies. The entire period under review has been divided into the same three sub-periods as the keyword emergence maps to simplify the interpretation and enable comparisons between both methods. Figure 16 presents the results for the market portfolio map for the years 2015 until 2017. Technology keywords of more than 10% increasing rate and less than 5 million investments are associated with emerging markets, keywords of more than 10% increasing rate and more than 5 million investments as fields of strong demand. The two market portfolio maps from 2009 until 2015 can be found in the appendix (see Appendix 2).

The map shows that the total amount of investments for robots is comparatively high, while the increasing rate of investments is very high (63%). Robot-related technologies that also experience a very high increase in investments are IoT (168%) and AI (2,462%), which is excluded from the map due to distorting effects. Surprisingly, also tablets are among those increasing markets. This stands in opposition to the results of the keyword emergence map that indicate that tablets have recently become a declining technology (see figure 15). A possible interpretation is that the amount of innovations for tablets declined, while the demand for this technology increased.

It has to be noted that the investment increasing rates, especially for emerging technologies, such as AI, are very high compared to real market values. This is due to the fact that the total amount of campaigns and investments is comparatively small and already a small number of campaigns can have a high impact on the overall development and increasing rates of investments. The numbers should not be understood as real market values, but as rough indicators for emerging market demands.

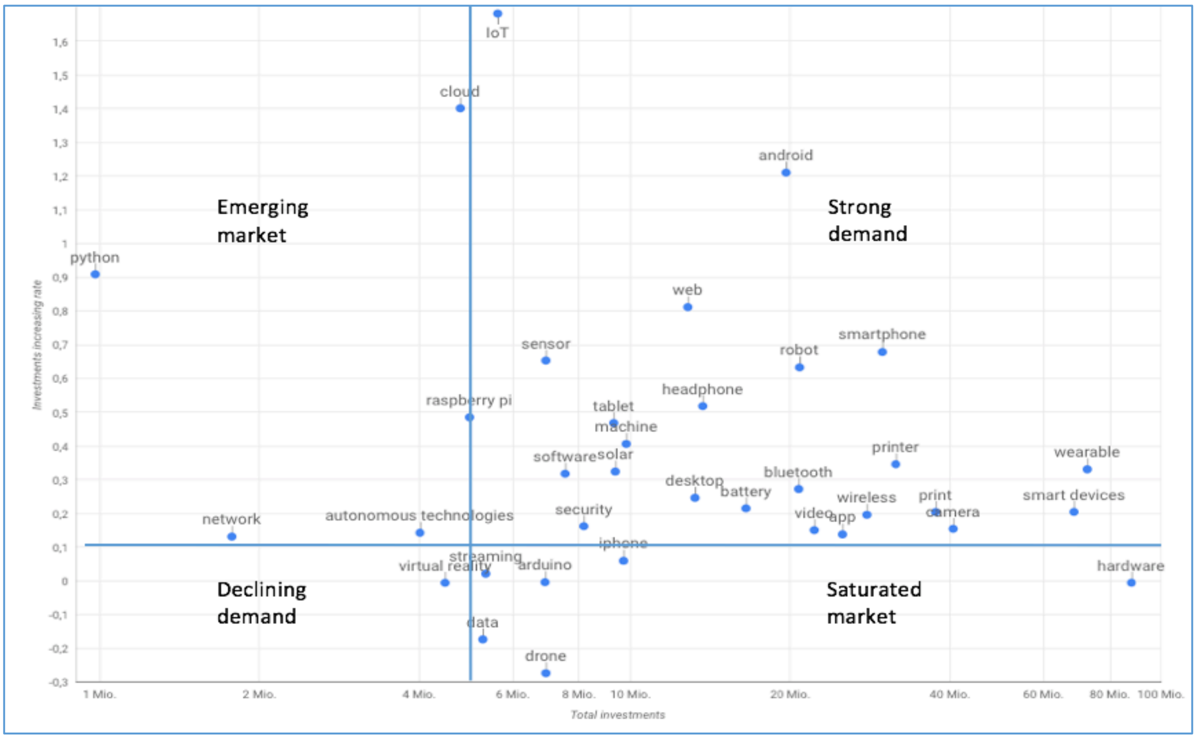

Figure 16. Market portfolio map, years 2015-2017.

5.2.3. Identification of fields of uncertainty and risk

To analyze the failure risk involved with the development of technologies, the application of the technology risk map is demonstrated. It is suggested that keywords with a success rate of more than 33% are classified as low risk technologies. The results imply that the development of robots is associated with a relatively low amount of risk, as over 40 percent of campaigns are successful (see figure 17). In contrast, autonomous technology or augmented reality projects are characterized by a relatively high failure rate, indicating that the development and commercialization of these technologies is denoted with high risk. Apps for example are a very frequently developed technology, however, the success rate for app projects stands at only 12%. This means that the market entry in this technology is very competitive and risky.

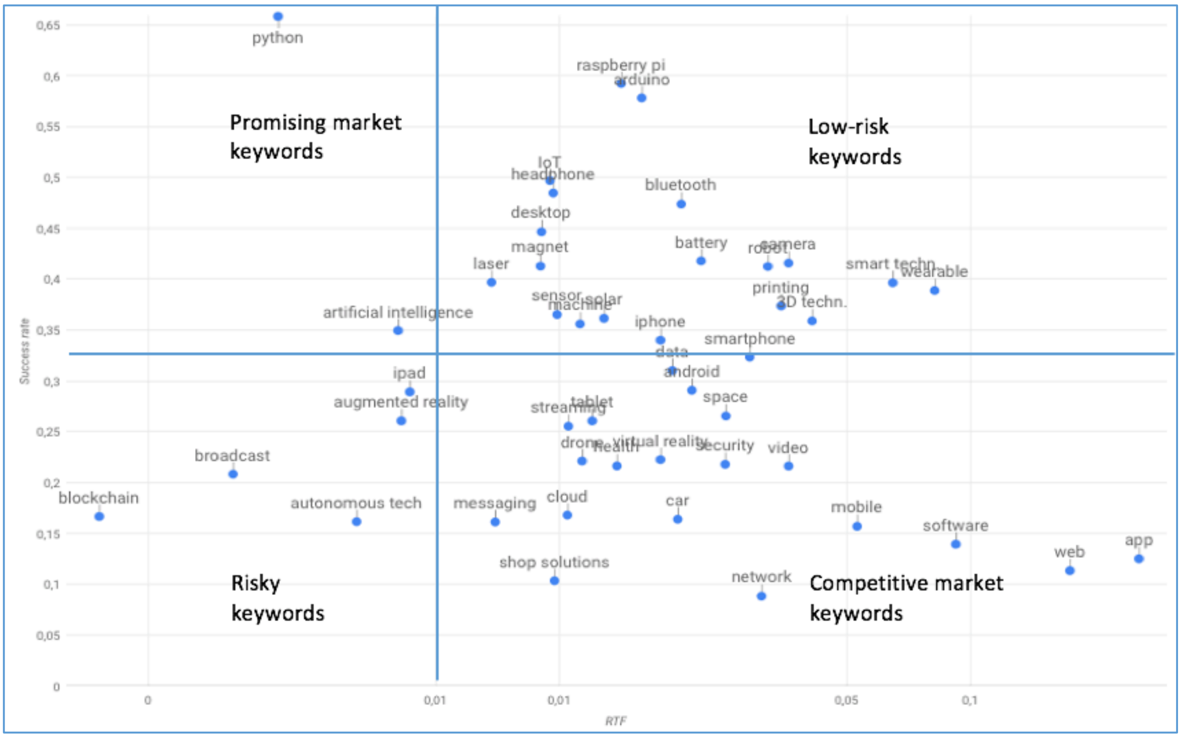

Figure 17. Technology risk map.

5.2.4. Identification of hypes and oversupply

The market hype curve is used to analyze the development of single technologies in more detail. Based on linear trend curves, it can be observed that the relative number of robot technology projects decreases over the period as a whole (see figure 18). On the other hand, the relative investments show an increasing trend. In the years 2015, 2016 and 2017, the investment curve surpasses the term frequency curve, implying that there is a market hype for robot technologies (see figure 18). As it has been shown, also drones play a major role in the technological field of robotics. As seen in figure 13 and 14, drones developed from an emerging technology to an established technology. The market hype curve provides more detailed insights into this development. According to the linear trend curves, the demand for drones increases in parallel with the amount of innovations, while the overall level of demand exceeds the level of supply (see figure 19). However, since end of 2016, it can be observed that the investment rates significantly dropped. This might indicate that the major hype for drones is already over.

It is noted that especially the RIR curve tends to be volatile. This might be due to the fact that the number of campaigns of a single technology is often relatively small. When certain crowdfunding projects go viral and become a huge success, this leads to significant upswings in the demand curve. One example is the peak in demand for drones in 2014 and 2016 (see figure 19). In 2014, the drone project “Zano – Autonomous. Intelligent. Swarming. NanoDrone” became a huge success and raised over 2.5 million dollars. In the beginning of 2016, “PowerUp FPV – Paper AirplaneVR Drone” raised nearly 500,000 dollars. As the linear trend curve is based on the historic data of 15 time periods, it balances these outliers and might be a better indicator for the overall demand for a technology. However, it also shows that a closer look on the specific campaigns of a technology can be very informative and provides more detailed insights about the development of technological innovations.

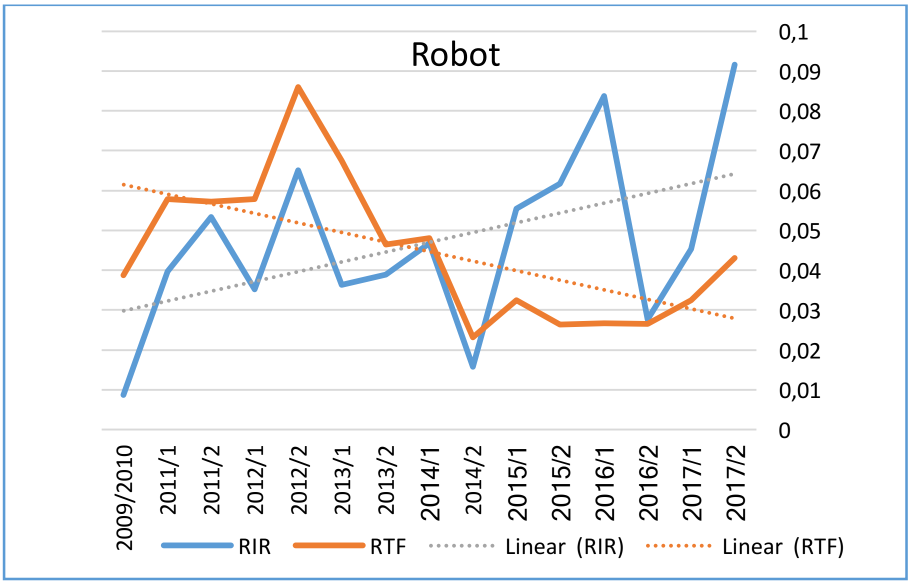

Figure 18. Market hype curve for robots.

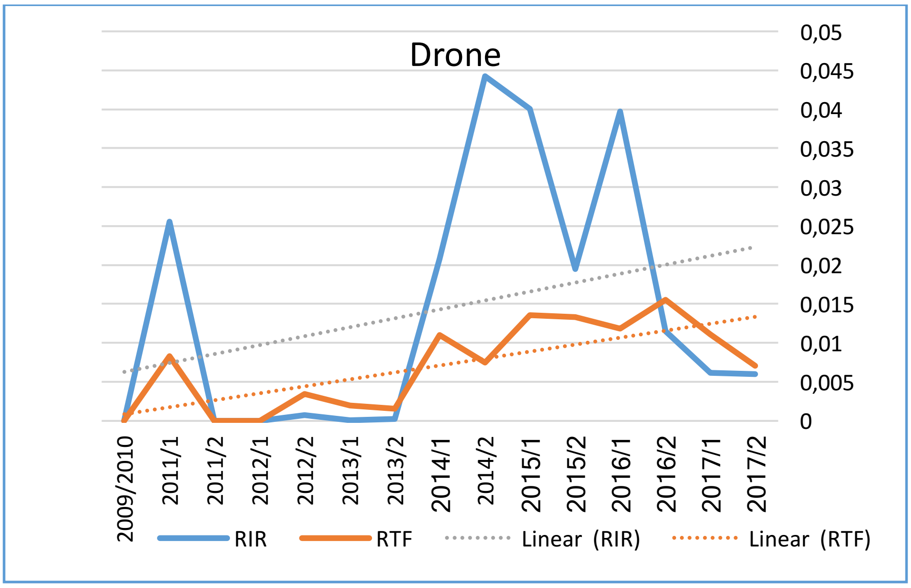

Figure 19. Market hype curve for drones.

5.2.5. Analysis of technological contexts and relations

The creation of co-word networks through text mining can be used to present the correlations between related keywords. A co-word analysis of relevant keywords has been applied using RapidMiner. The output is used to create a correlation matrix for the term robot (see table 3).

	affordable_ robot	app	arduino_ robot	robot_arm	artificial_ intelligence
affordable_robot	1,000000	0,000221	0,000001	0,065947	0,000004
app	0,000221	1,000000	0,000295	0,001596	0,000099
arduino_robot	0,000001	0,000295	1,000000	0,027420	0,000005
robot_arm	0,065947	0,001596	0,027420	1,000000	0,000031
artificial_intelligence	0,000004	0,000099	0,000005	0,000031	1,000000

Table 3. Example output of the correlation matrix for the term robot.

To keep the structure of the network as clear as possible, the terms with the highest correlations to each other have been selected to create the co-word network. The network graph has been built using the open source software Gephi, which is an interactive software tool to visualize and explore networks. Gephi has been applied as it provides simple access to network data and allows for creating, filtering, navigating and clustering all types of networks (Bastian, Heymann & Jacomy, 2009). For the development of the co-word network, the correlation matrix has to be imported to the Gephi interface. The network has been clustered using the force atlas algorithm. This algorithm follows the principle that linked nodes attract each other, while non-linked nodes are pushed apart. This layout facilitates the interpretation of closely connected topics.

Figure 20 shows an unstructured co-word network for the technological field of robots. It is shown that the term robot tightly correlates with the development of robot arms, humanoid robots or product aspects such as connectivity and customizability. To structure the graph, the keywords have to be assigned to different layers (see figure 21). Therefore, the co-word network has to be divided into the three main layers, market, product and technology, and each term is manually assigned to the respective category. Terms such as education robot, child or low cost are assigned to the market layer and indicate common target groups or price settings. The product layer shows products that should be considered in the strategy making process, such as drones, smartphone app or product features such as speed or Wi-Fi. The technology layer displays alternative technologies that could play an important role for the development of new product innovations. Furthermore, the network shows the linkages between and within the different layers that can foster creativity and innovativeness to create new compositions and product concepts. Cars for example are related to the technology layer keyword autonomous, indicating the upcoming trend of autonomous driving. Another example is Arduino that is strongly related to the market group of do-it-yourself (DIY) users. This shows that the development of Arduino robotics is primarily addressed to the target group of DIY users.

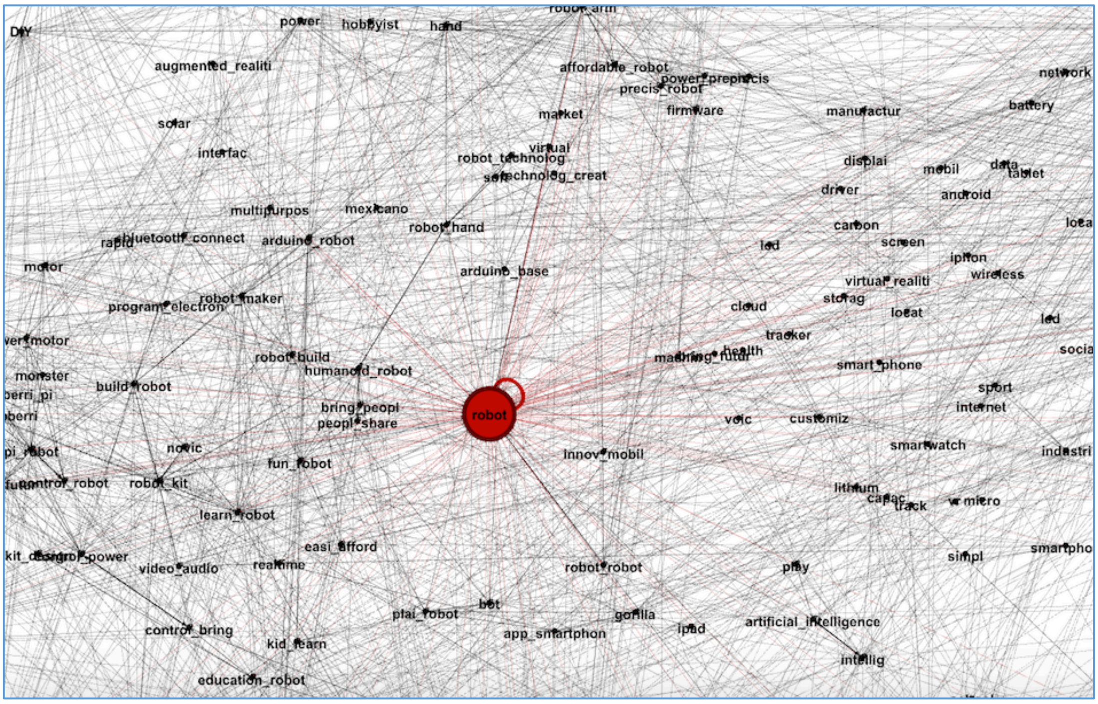

Figure 20. Unstructured co-word network for the term robot.

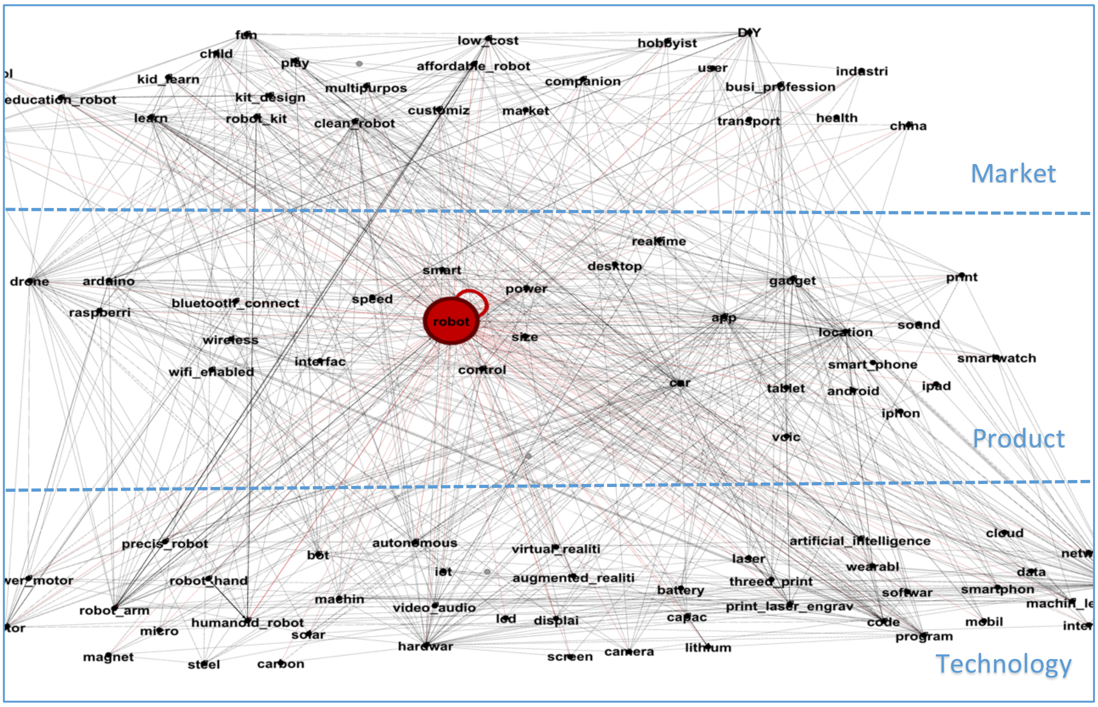

Figure 21. Co-word network with three layered structure for the term robot.

5.3. Analyzing the impact of weak signal detections in crowdfunding data

As initially described, the identification of emerging technologies and weak signals is one of the major interests in technology foresight and future signal analysis (see section 2.1.3.). Detecting these technologies at a very early stage might set organizations in a superior position in competition. This is especially the case when these technologies also tend to result in high payoffs in future periods. The systematic analysis of technology crowdfunding campaigns in conjunction with the use of the keyword emergence map allows to draw conclusions on these emerging technologies (weak signals). This resulted in the identification of 14 weak signal technologies during the three periods under investigation (see table 4). Following the development of these technologies during subsequent periods, it becomes evident that they either stay a weak signal, or turn into a strong signal, an established technology, or also a declining technology.

Weak signals identified during text mining analysis					
	Newly identified Weak signals	Weak signals	Strong signals	Established technology	Declining technology
Period 1	7	-	-	-	-
Period 2	5	6	4	2	-
Period 3	2	**6**	**3**	**3**	**2**
Total	**14**				

Table 4. Development of weak signal technologies identified in this study.

Table 4 indicates that 6 of the 14 weak signal technologies turned into a strong signal or an established technology during these three periods, while 6 technologies are still considered as weak signals and 2 technologies are considered as declining technologies.

It is now interesting to analyze whether the occurrence frequency of weak signal technologies also has an effect on increasing investments and increasing demand. Therefore, the RTF and RIR values of these weak signal technologies have been analyzed through a correlation analysis and a linear regression model, to examine whether there exists a positive correlation between RTF and RIR of weak signal technologies and whether there is also a significant effect of RTF on RIR in different periods.

Table 7 (see Appendix 3) shows the corresponding correlation matrix. It can be observed that there is a very high correlation between the RTF and the RIR of weak signal technologies.

For the linear regression calculations, only those 7 weak signal technologies are included that have been detected during the first time period, to involve the values for 8 consecutive years. This leads to a total amount of 56 observations in the panel data. The linear regression has been conducted including two-year time lags. It is shown that RTF and RIR are positively correlated to each other and that there is a very significant influence of RTF on RIR, significant at $p < 0.001$, and a high R^2 of 0.7917, which means that 79.17 % of the variation in RIR is explained by the RTF (see table 5). The within R^2 value indicates how much of the variance within the panel units the model accounts for, while the between R^2 refers to the variance between separate panel units. The overall value is a weighted average of these two. Fixed effects (within estimators) are used, as the focus lies on analyzing the impact of variables that vary over time. The coefficient for the RTF value is 3.477, indicating that an increase by one unit in RTF causes a 3.477-units increase in RIR. Particularly noteworthy is that this effect even increases during the subsequent two years (see Appendix 4). With a two-year lag the coefficient for RTF is 4.226. This shows that the effect of RTF on RIR is significantly high and is becoming even higher after one, respectively two years. The results, however, have to be considered with caution, since the total number of analyzed observation units is very small.

Fixed-effects (within) regression		Number of obs =	56
Group variable: Tec_No		Number of groups =	8
R-sq:		Obs per group:	
within	= 0.7917	min =	7
between	= 0.9346	avg =	7.0
overall	= 0.8422	max =	7
		F(7,41) =	22.26
corr(u_i, Xb)	= -0.3404	Prob > F =	0.0000

RelInvRate_L0	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
RelTermFreq_L0	3.477021	.3196151	10.88	0.000	2.831545	4.122497

Table 5. Linear regression model, RTF and RIR for weak signal technologies.

5.4. Creation of technology roadmaps and scenarios for robot technologies

To demonstrate the creation of data-driven technology roadmaps and scenarios, two narrative example cases are presented. The relevance of narrative case studies as a research instrument has been highlighted by several researchers (Brandell & Varkas, 2001; Gilgun, 1994). Narrative case studies are used to investigate problems or methods within their environmental context, for the purpose of making complex fields of research more understandable and accessible and to capture the essential meaning of a study (Brandell & Varkas, 2001). The following cases describe the fictional situation of a robotic company to illustrate the application of the previously presented foresight methods and to provide additional understanding on how to implement the proposed methods and results into the foresight process.

5.4.1. Example case 1: Development of a technology roadmap

A company that is specialized on robotic solutions for children considers the development of new products. To examine fields of opportunities, the management suggests the application of a technology roadmap. To support the work of internal experts, the management decides to systematically analyze crowdfunding data. In a first step, the thematic field has to be defined. The management is interested in alternative technologies and application fields that are currently developed in the field of robotics. After the crowdfunding data has been crawled and filtered for the category robots, text mining analysis is conducted. To get a first overview over trending topics, the most important terms are visualized through a word cloud. It becomes evident that several alternative technologies play a major role in the field of robotics, such as drones, engines, laser, sensors, cameras, 3D printing, AI and apps (see section 5.1.). The management decides to further examine the development of these technologies. Through the evaluation of the keyword emergence map, it can be concluded that AI is an emerging topic that becomes increasingly important (see figure 15). Drones and 3D printing have been weak signals a few years ago. Now, there already exist a broad number of innovations for these technologies. The experts discuss that AI could be the next big thing for technological innovations. Since AI is connected with objects, they decide to consider drones as a possible AI application field. The market portfolio map indicates that also the demand for AI is rapidly growing (see figure 16). While the managers are still uncertain about the success of an AI product, the technology risk map shows that more than 35% of AI projects succeed, which is a relatively high value (see figure 17). To take a closer look on the development of drones, the management uses a market hype curve (see figure 19). Based on the data of previous years, the market for drones, as well as the number of innovations is increasing. However, in 2017, the interest in drones seems to have drastically declined. Particularly conspicuous are two phases of market hypes in 2014 and 2016. Therefore, the management takes a closer look at the specific campaigns that were responsible for the sudden increase in consumer interest: A drone project equipped with many intelligent techniques, such as sensors, cameras and connectivity solutions, as well as another project that focused on affordability (see section 5.2.4.). The management decides that it might be beneficial to combine both approaches. To create the technology roadmap and support the work of technology, product and market experts, they conduct a co-word analysis to visualize the technology network for drones (see figure 23). The final roadmap looks as follows (see figure 22):

From the starting point of building robotic devices for children, the company started to consider AI and came up with the idea to equip drones with AI solutions. This requires several complementing elements for connectivity such as Bluetooth and Wifi. Since the goal is to build a low cost device, they decided not to develop an own remote control but to use a specific smartphone app to control the functions of the drone. Since children usually do not have smartphones, the management decides for young adults and hobbyists as a new target group for their product.

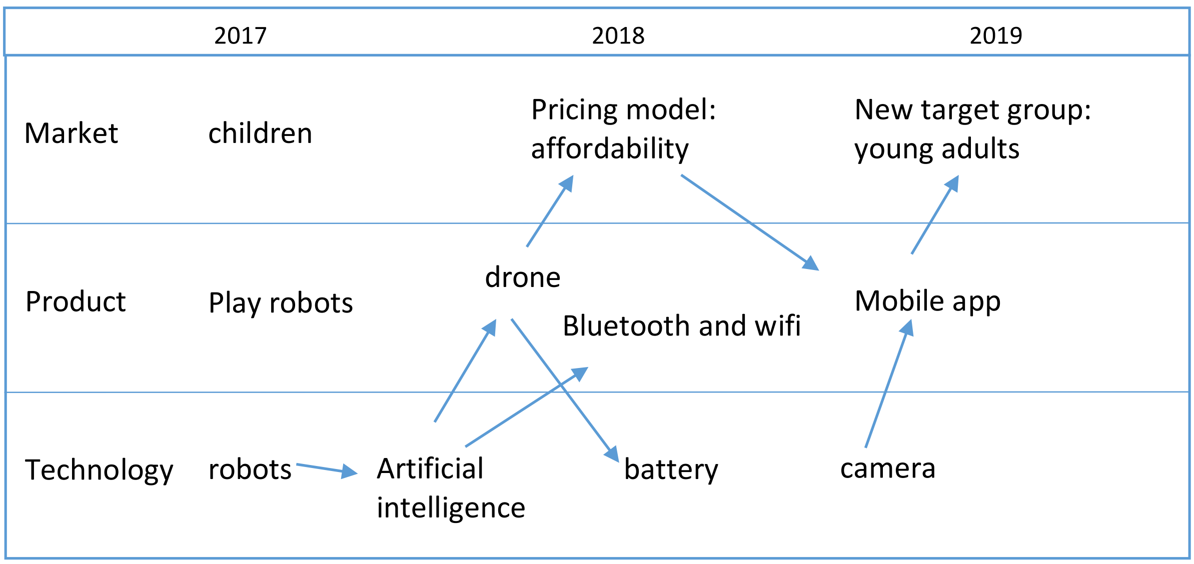

Figure 22. Example of a technology roadmap for drone development.

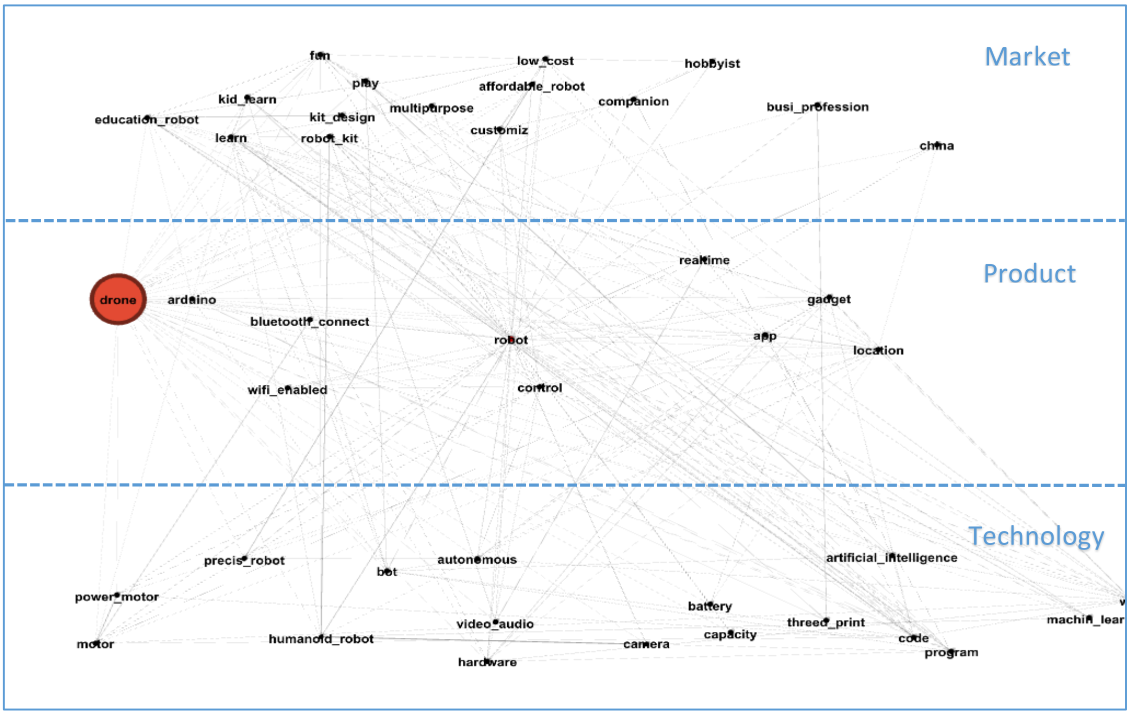

Figure 23. Co-word network for the term drone.

5.4.2. Example case 2: Development of scenarios

Scenario planning is an explorative approach. In this case, the goal is not to develop a new product, but to come up with different future projections for the development of robotic devices. The word cloud is again used to get an initial overview of the topic. For scenarios it is especially interesting to see which technologies are currently emerging and might shape the future. Therefore, keyword emergence map (see figure 15) and market portfolio map (see figure 16) are analyzed to identify emerging technologies and markets. Three technological fields are selected that might play an important role in the future of robotics: Augmented reality, drones and AI. To support experts in analyzing each technology, market hype curves (see figure 19) and co-word networks are applied (see figure 21 & figure 23). These indicate which alternative technologies, products and market factors have an influence on the development of the respective technologies.

Scenario 1. Intelligent drones (see figure 23)

Intelligent drones are becoming a major field of robotic technologies. Future drones will be able to fly autonomously and are equipped with smart systems and AI. The drones are continuously learning about their environment through machine learning algorithms. Furthermore, they can be customized by hobbyists through adding 3D printed parts that are shared on web platforms by robot companies. New business models for robotic firms will be software solutions that can be accessed via app or web interface in real-time to customize and enhance the functioning of drones.

Scenario 2. Virtual reality to control healthcare robots (see figure 21)

Virtual reality glasses are used to control robots from a distance. This can for instance be applied in the healthcare sector to do urgent surgeries in remote places and precisely control robot hands and tools.

Scenario 3. Artificial intelligence robots for nursing care (see figure 21)

Robots are becoming more intelligent and are used in elderly care. Humanoid robots will be able to learn and absorb any information like a real human can. These robots also learn how to interact socially, recognize body language and voice inflection.

5.5. Summarized findings

The results show that the proposed methods provide various relevant and valid information to support the foresight process and decision making. For the term robot it has been shown that there is a wide heterogeneity of related technologies, product attributes and fields of application. Beside the field of robotics, also other technologies can be analyzed using the proposed analysis methods. A good investment case could have been the market for wearables a few years ago. The keyword emergence map shows how wearables arose in the years 2009 until 2015 from a weak signal to a strong signal technology (see figures 13, 14, 15). Furthermore, the market demand was continuously growing in these years. Also the success rate of wearable projects is relatively high. From today's perspective, AI and IoT are becoming increasingly important and can be regarded as promising technologies. Also the blockchain technology recently came up as weak signal (see figure 15). This technology however is still associated with high risk (see figure 17). It becomes apparent that the analysis revealed many technologies and topics that are currently popular in the innovation landscape and also correlate with the technologies mentioned in the annually published Gartner Hype Cycle for emerging technologies, such as wearables or virtual reality (Gartner, 2017). This can be seen positively and speaks for the validity of the results. While many innovation reports and articles rely on expert interviews, the results from this study are based on extensive quantitative research that reflect the number of innovations in a technological field and the actual consumer demand on a pre-mass market.

6. Discussion and evaluation

This thesis pursues the goal to examine how technology foresight approaches can be supported through the use of quantitative analysis methods of crowdfunding data. The results show that crowdfunding data offers different possibilities to address this goal. Since the application of crowdfunding has not been discussed yet in technology foresight literature, the results will be evaluated and compared to previous findings about the use of quantitative data in technology foresight. In each subchapter, the impact of this study on existing knowledge in the field of technology foresight will be examined. Furthermore, it is shown how the 'old' expert based foresight processes are extended and improved by the new quantitative techniques. The methods are finally evaluated in a discussion with data analyst experts of the German innovation consulting company HYVE, to validate the practical benefits and opportunities of the proposed approach and methods.

6.1. The use of crowdfunding data in technology foresight

It has been shown that crowdfunding data adds a new dimension to the analysis of future technological trends. Unlike with patent, social media or web news data, crowdfunding data entails further variables and values that can lead to additional insights and relevant technology-specific knowledge. The analysis allows to derive insights about the emergence of technologies, as well as market relevance, which is an important factor in foresight analysis. Text mining of crowdfunding campaigns reveals many market and application keywords as the textual descriptions are originally addressed to consumers and investors. This is different to patent documents that are written from technology-based perspectives and therefore predominantly reveal technology keywords (Lee et al., 2008). Furthermore, it becomes apparent that relevant data is very concentrated, which makes crowdfunding a very efficient source for analyzing technologies and facilitates holistic analyses of technological landscapes. On the other hand, also crowdfunding data has limitations. Crowdfunding data is not suitable for all types of technologies and products. Campaigns are often addressed to the end consumer. This implies that the data is more suitable for the analysis of B2C technologies. Furthermore, inventors usually try to commercialize their inventions in the foreseeable future. Therefore, these inventions are usually in a more advanced stage compared to patent data. This means that crowdfunding data is rather useful for short-term foresight approaches.

6.2. Revised technology foresight framework

This study has not only been conducted to introduce new data-driven methods to access additional foresight information, but to provide guidance on how to make foresight processes more efficient, factual and transparent. Therefore, a new data-driven technology foresight process is suggested. The thesis introduces six different methods that serve as decisive tools to make sense of the analyzed datasets and constitute main building blocks in increasing the knowledge base and efficiency of foresight activities. Furthermore, they serve as orientation for future technology foresight applications that consider the use of crowdfunding data. Some of these methods, word clouds, co-word networks and the keyword emergence map have already been discussed in a similar way in previous studies (Lee et al., 2008; Yoon, 2010; Yoon, 2012). In this thesis, the application of these methods has been brought into the context of crowdfunding analysis and the implementation into the different steps of the foresight process. With regard to the evaluation of the proposed methods, the new data-driven foresight process can be finally outlined.

Phase 1 – Definition of scope and research field

In phase 1, experts need to explore and define their field of research and identify important and relevant topics. This step is typically characterized by extensive desk research to acquire necessary information about the extent of the research field. Through the use of text mining, much more data can be analyzed compared to a manual literature analysis. The most common topics and terms are visualized to allow a quick overview over the different potential topics of a research field. It has been shown that through the analysis of crowdfunding data, meaningful word clouds can be created in a non-supervised approach. It has to be acknowledged that further clustering of the data might be beneficial to structure the findings. However, clustering requires additional manual adaptations. In this case, the goal was to present an easy to implement and reproducible tool that provides experts with initial insights about the extent and diversity of the research field under investigation.

Phase 2 – Identification of market, product and technology developments

In phase 2, experts discuss about emerging trends on market, technology and product level. Therefore, they make use of internal knowledge, which is often limited, subjective and biased (Geum et al., 2015; Hussain et al., 2017; Saritas & Aylen, 2010). The methods proposed in this thesis assist this process by adding additional external and data-based information to the discussions that is presented through easy to interpret maps. Furthermore, the tools support the detection of developments and trends on the different levels.

The keyword emergence map has already been suggested in a similar form in previous research (Kim & Lee, 2017; Lee et al., 2008; Yoon, 2012). It has been shown that with a crowdfunding dataset, it is possible to evaluate the emergence of different complementary or even competing technologies. This offers new possibilities and constitutes a new major finding in the research of keyword emergence maps. Furthermore, this thesis makes an important differentiation in the interpretation of emerging signals from keyword-based text mining approaches by distinguishing between supply and demand for an innovation. While previous research of Yoon (2012) describes the detected signals as "impending important events or trends" (Yoon, 2012, p. 12543), this thesis pleads for a more precise definition. It is proposed that the results from the keyword emergence map serve as an indicator for the innovativeness and R&D activities in this technological field. This specification has also been made with reference to the results of further analysis methods, such as the market portfolio map.

The analysis of investment rates through market portfolio maps can be used by experts to discuss whether an emerging technology will also be adopted by consumers and achieve greater market penetration. Therefore, it is highly recommended to analyze and interpret the conformities and discrepancies of both, keyword emergence and market portfolio map to make strategic decisions. With previous approaches using patent, social media or web news data, market and investment data has not been accessible.

The assessment of risk is an important factor for strategic decision making and plays an important role in technology foresight (van der Heijden, 2005). However, to the knowledge of the author, there exists no standalone foresight tool that indicates how risky the investment in a certain technology is. The analysis of success rates of crowdfunding campaigns offers the opportunity to display and compare success probabilities that are used as a measure of risk involved with the development of technologies.

Since crowdfunding is the only yet analyzed foresight data source that includes both, textual and numeric data, the relation of innovation supply and market demand for a certain technology has not been analyzed before in the context of technology foresight. The market hype curve refers to the immanent economic interplay of supply and demand (Smith, 2002). It has been shown that the method provides valuable insights for managers about the development of single technologies.

Phase 3 – Creation of roadmaps and scenarios

Phase 3 is about creating the roadmaps or scenarios and connecting the different information streams and perspectives. By conducting co-word analysis with previously identified terms and topics, it is possible to visualize the linkages between these topics and identify important relationships. Already Choi et al. (2013), Lee et al. (2008) and Tseng et al. (2007) conducted co-word analysis in their studies. As they examined patent data, they had no access to consumer related descriptions and therefore did not include the market layer into the analysis. In this thesis, the network analysis has not only been conducted to visualize the correlations between technologies, but also within and between product attributes and market application fields. It has been demonstrated that this provides additional and relevant information for scenarios and technology roadmaps.

Phase 4 – Validation of results

Finally, the results have to be validated and checked. The integration of external data allows to compare the final foresight results with the topics and trends that have been identified through the data analysis process. As already proposed by Kayser and Blind (2017), continuous feedback loops between data-driven methods and expert discussions are suggested to balance internal expert knowledge and external data-based information.

6.3. Impact of the study on weak signal research

With the extraction of technological keywords and the application of the keyword emergence map, this thesis suggests a method to detect weak signal technologies through analyzing crowdfunding data. Since crowdfunding platforms offer dense information about technological inventions, the diversity of weak signal identifications is not limited to a certain technological field. The correlation between the occurrence frequency of weak signal technologies and their investment rates has been analyzed through a correlation matrix, as well as a linear regression model. It is shown that there exists a high correlation between the term frequencies and investment rates of the examined weak signal technologies. It is particularly noteworthy that the effect even increases with a two-year time lag. This study is the first one that confirms that the detection of weak signals through text mining reveals technologies that are also correlated with increasing investment rates and future demand and therefore represent high-potential fields of investment. This constitutes a major finding in the theories about weak signal detections through data mining processes.

6.4. Discussion and validation of results at HYVE AG

To verify the practical benefits of the proposed approaches, they have been evaluated in a discussion with data analysts at HYVE AG. HYVE is a German innovation and consulting company founded in 2000. Since then, they established as a leading international innovation company. The company is specialized on the identification of new fields of growth and the development of future product generations. HYVE works across industries and supports international companies as well as SMEs, political institutions or research centers in their innovation processes. Meanwhile more than 70% of the DAX companies are among HYVE's customers (HYVE, 2018). One of the core competencies of HYVE is their market research department that focusses on the identification of technological trends and opportunities. HYVE systematically scouts for external experts, monitors international startup activities and started the exploitation of various data sources such as patent databases and social media.

Two main topics were discussed: First, the usefulness and applicability of crowdfunding as a data source for technology foresight and indicator for future opportunities and trends. Second, the potential benefits of the specific methods presented in this study.

Crowdfunding as a data source for technology foresight

During the discussion, it turned out that HYVE experts already considered the use of crowdfunding as a data source for analyzing technological activities and emerging trends. In accordance with the propositions made in this thesis, HYVE experts see valuable potential in the special characteristics and principles of crowdfunding. They argue that crowdfunding is the best known indicator for market reactions as consumers invest real money on the presented products. Therefore, HYVE experts share the opinion that crowdfunding platforms can be considered as 'markets before the mass market' and might constitute a relevant, efficient and promising source for foresight activities. Moreover, crowdfunding platforms offer early feedback mechanisms for inventors, who first present their ideas to the market and then produce products in case their campaigns have been successful. Therefore, the platforms can be an interesting source for new innovative ideas and idea generation.

HYVE recently started to analyze crowdfunding data in external consulting applications. They examined for example emerging fields of smart home technologies, such as smart security or smart lightning and compared the different investment and success rates.

On the other hand, also limitations of crowdfunding analysis have been discussed. The types and the number of different product categories on crowdfunding platforms is limited. Therefore, the analysis can only be used for some technologic fields. Moreover, generalizability has to be questioned. It is to some extent unclear whether crowdfunding investors are early adopters, which would be a very interesting market group, or rather hobbyists and gadget lovers who are only interested in very specific niche products and constitute a questionable indicator for general market demand. Furthermore, the success of crowdfunding campaigns is not only based on the underlying technological idea, but also on various additional factors, such as marketing skills of the campaign creators.

To obtain detailed views of the technologic landscape, HYVE suggests to monitor and systematically explore all types of social media platforms, to which also crowdfunding can be counted. Social media mining offers a very powerful source to explore and verify innovations and ideas. In terms of generalizability, the analysis of Facebook is a relevant data source in Germany, while Twitter is rather limited as it is mainly used by journalists and media professionals. Particularly interesting is the analysis of web blogs and specific internet forums, which are valuable and representative sources to examine consumer needs in several branches.

Potential benefits of the proposed text mining methods

HYVE distinguishes between two main perspectives to monitor and analyze technologic developments, the telescopic and the microscopic perspective. While the telescopic perspective strives to examine general trends and developments in the technologic landscapes, the microscopic view focusses on the detailed examination of a specific technological field. Also the tools presented in this thesis can be classified into these two perspectives. The keyword emergence, market portfolio and technology risk map provide an overview of the general development of the technologic landscape, emerging technologies, trends and potential opportunities. This information is especially interesting for research centers, political institutions or companies that are interested in various technologies. HYVE and HYVE’s business clients are particularly interested in microscopic perspectives and the detailed examination of a specific technology. For these purposes, the market hype curve and co-word networks are highly suitable. The approach of the market hype curve raised interest as it constitutes a dynamic analysis, incorporating the development of the occurrence frequency and investment rates over time. Concerns were expressed regarding the impact of outliers. Since the results are sometimes based on relatively small quantities, single crowdfunding campaigns might have a distorting effect on the overall visualized development of a technology. On the other hand, it might be particularly interesting to take a closer look at the projects and innovations that were responsible for these peaks. Also co-word networks are considered to be a relevant method to show the relations between technologies, products and application fields. Especially the linkages between technologies and possible markets/application fields are interesting in order to discover future innovation opportunities.

7. Conclusion

This study proposes a new quantitative research approach to analyze crowdfunding data for technology foresight applications. The research methodology follows Peffers' et al. (2007) design science research process. It focusses on the design of innovative, reliable and reproducible analysis methods and demonstrates their usability and applicability to resolve existing limitations in foresight approaches. The study is organized around the following research question:

How can text mining of crowdfunding data be applied to analyze the detection of emerging technologies, market developments and trends and thereby support the process of technology roadmapping and scenario planning?

The research focus is twofold: (1) Analyzing a new data source by designing new methods to examine emerging technologies and trends. (2) Implementing these methods into technology roadmaps and scenarios to increase the transparency and efficiency of foresight processes. Several variables, such as term frequencies, investment and success rates are derived from the crowdfunding datasets and have been analyzed to examine the research question and sub questions:

- *Which methods can be applied to provide access and information about technological contexts and environments?*
- *Which methods can be applied to identify weak signals, emerging technologies and fields of increasing supply?*
- *Which methods can be applied to identify emerging market demands?*
- *Which methods can be applied to assess the investment risk of a technologic innovation?*

Answering these questions plays a major role in today's foresight applications. This thesis is based on the assumption that crowdfunding data offers new and unexplored opportunities for technology foresight activities. To systematically analyze the data, text mining methods have been combined with contextual data, as well as trend and co-word analysis. Six different methods have been designed that are meant for supporting the traditional foresight process and facilitate the detection of future signals and trends: (1) word clouds, (2) keyword emergence map, (3) market portfolio map, (4) technology risk map, (5) market hype curve and (6) co-word networks. These methods have been aligned with the different phases of the foresight process and the requirements for technology roadmapping and scenario planning. To demonstrate the functionality and usability of the proposed methods, a case study about robot technologies has been conducted by analyzing more than 26,200 crowdfunding campaigns from the years 2009 until 2017. The methods and results have been evaluated in a discussion with data analyst experts at an innovation consulting company.

To conclude, the main contributions of this thesis can be summarized as follows:

The applicability and meaningfulness of examining crowdfunding data for technology foresight is demonstrated. New methods are established that provide novel, data-driven insights for analyzing technological fields. Furthermore, it is shown that data-driven methods provide guidance and support during different steps of the foresight process. They may therefore improve the efficiency and strengthen the knowledge content of traditional technology roadmapping and scenario applications.

The main benefits for foresight processes are manifold:

First, time consuming effort for literature and desk research is reduced through the systematic and automatized text mining process. Second, the identification of key technological and environmental factors is supported through the extraction of relevant keywords. Third, the detection of technology and market trends is enhanced through the identification of emerging signals and the implementation of time series analysis. Fourth, the examination of uncertainty factors is assisted through the analysis of success rates. Fifth, the interrelations between relevant components is exemplified through the visualization of term correlations.

7.1. Theoretical contributions

The application of advanced data mining tools for technology foresight is a young field of research. Existing studies come to the conclusion that the increasing amount of data can be used to improve and support technology foresight analysis (Kayser & Blind, 2017; Lee et al., 2008). As these approaches are still in an early stage, researchers postulate that further investigations should be carried out and different data sources and methods should be integrated into the foresight process (Kayser & Blind, 2017). This thesis ties in with these first approaches that suggest the use of text mining of web data for technology foresight and strives to extend and optimize previous findings by analyzing a new and promising data source.

To combine the specifications of crowdfunding data with the need to develop more efficient, precise and transparent ways for foresight analysis, new methods are suggested that are inspired by previous research (Choi et al., 2013; Geum et al., 2015; Lee et al., 2008; Yoon, 2012). The evaluation of textual and numeric data leads to new insights that cannot be addressed through the analysis of solely textual data from patents or web news articles. Furthermore, the thesis adds knowledge to research on technology roadmaps and scenario analysis. It is shown that the systematic analysis of unstructured data can not only play an important role at the beginning of the foresight process, but also during the different phases of technology roadmapping and scenario planning. Therefore, the impact and contributions of this thesis on technology foresight research can be summarized as follows:

First, a new, high-potential data source is presented and analyzed, and it is shown that novel and relevant insights can be derived from it. Second, new data analysis methods that allow to analyze technologies from various points are presented and successfully applied. Third, new insights about weak signal research have been made by demonstrating the identification of weak signal technologies in crowdfunding datasets and confirming the correlation between term frequencies and investment rates for weak signal technologies. Fourth, new methods and guidance on how to support the technology roadmapping and scenario processes during different steps have been presented.

7.2. Practical and managerial contributions

Technology foresight is important for companies, whether large or small, in order to stay competitive by analyzing future developments and innovations. Coates et al. (2001) highlight the ever increasing importance of technology analyses to aid decision makers. They outline the need for the development of “easily comprehensible, timely, and cheap sources of foresight” (Coates et al., 2001, p. 15). Furthermore, they postulate the integration of market exploitation insights as “the technical side presents only part of the answer” (Coates et al., 2001, p. 8). This thesis strives to take into account both postulations. First, the proposed methods serve as comprehensible, timely and easy to reproduce foresight tools. Second, the analysis involves insights about market developments and opportunities.

The systematic examination of large datasets creates several options to support foresight practitioners and entails valuable input for strategic choices and decision-making. The potential practical benefits of crowdfunding data for innovation management and foresight analysis have been confirmed in the discussion with data analysts of the innovation consulting company HYVE. The study provides clear advice on how to integrate the proposed methods into the overall strategic foresight activities of companies. It is recommended to use the data-driven foresight tools as an additional information source during the foresight process to enrich discussions, balance internal views and verify expert opinions. It is expected that this data-driven process reduces time and effort in building a profound knowledge base and that the foresight process become more transparent and factual. The insights about future trends and emerging technologies can also be used as a source for idea creation and assessment, which is a key activity in organizations' innovation management (Feldmann et al., 2013).

7.3. Limitations and future research

Despite its theoretical and practical implications, the presented methods and results also entail limitations. First, it has to be noted that the process of text mining has constraints as it is primarily based on the count of term occurrence frequencies. This means that supplementary information for further analysis is missing. Furthermore, text mining output is not self-explanatory and requires subsequent interpretation from domain experts to provide useful input for foresight analysis. This thesis strives to address these limitations, as it is not solely based on the use of textual data, but also integrates numeric information that offers additional insights. However, this also requires further interpretations and analytical capabilities.

As with the use of every data source, also crowdfunding data entails limitations. The identification of trends and attributes through the analysis of crowdfunding data might not be applicable for all types of technologies and innovations. Inventions published on crowdfunding platforms are typically addressed to the end costumer and user. As the results from data mining processes highly depend on the quality and characteristics of their input, it can be concluded that the analysis of B2C oriented crowdfunding platforms is more appropriate for mass market compatible technologies. Furthermore, innovations published on crowdfunding campaigns are usually in an advanced stage and meant to be commercialized in the near future. This indicates that the time scope of foresight results is limited. The goal of this study, however, is not to present long-term predictions, but to provide managers and experts with indications about emerging technological trends to formulate consistent future strategies. This thesis analyzes the crowdfunding platform Kickstarter. Today, many different crowdfunding platforms exist that should be taken into account by future studies and applications. The presented methods are based on the assumptions that behavior on crowdfunding platforms serves as indicator for real market performances. It has to be considered that there is no general validity for this interrelation and that the results have to be seen as reference points instead of universal solutions.

To demonstrate the functioning of crowdfunding methods the study focusses on the topic of robotics. However, the proposed tools are not restricted to the use in robotics. Further technological areas should be analyzed by future studies using the presented evaluation tools. In future applications, the methods have to be customized to the specific needs. The conditions determined for the proposed maps are suggested values and can be adapted by foresight experts to their individual concepts.

This thesis is the first that suggests the use of crowdfunding data for technology foresight. Additional effort is required to validate the results and conduct case studies with companies to analyze how the use of data-driven tools can contribute to the success of their foresight actions. The presented techniques constitute starting points for the systematic analysis of textual and numeric data from crowdfunding platforms. Some evaluations still require manual adaptations, such as the clustering of co-word networks. Future research should refine the suggested methods and for example integrate machine learning algorithms to further automatize the data processing and clustering tools.

The study shows how to identify weak signal technologies from crowdfunding data and examines that the emergence of these technologies highly correlates with investment rates. It has to be acknowledged that these results have to be considered with caution, since the total number of analyzed observation units is very small and limited to the number of weak signals identified in this thesis. Future studies might extend these findings and build upon the proposed approaches to analyze the relevance of weak signal identifications through text mining techniques.

In general, text and data mining analysis for technology foresight is still a young field of research. Studies can contribute to the field by examining additional data sources, such as social media platforms, web blogs or forums, as well as new data mining methods to exploit the potential of big data for detecting technological trends. Furthermore, future work should investigate the interrelation of quantitative and qualitative foresight methods to enhance the validity, reliability and efficiency of future foresight applications.

References

Abbas, A., Zhang, L., & Khan, S. U. (2014). A literature review on the state-of-the-art in patent analysis. *World Patent Information*, *37*, 3–13.

Allison, T. H., Davis, B. C., Short, J. C., & Webb, J. W. (2015). Crowdfunding in a prosocial microlending environment: Examining the role of intrinsic versus extrinsic cues. *Entrepreneurship: Theory and Practice*, *39*(1), 53–73.

Amer, M., Daim, T. U., & Jetter, A. (2013). A review of scenario planning. *Futures*, *46*, 23–40.

Ansoff, H. I. (1975). Managing Strategic Surprise by Response to Weak Signals. *California Management Review*, *XVIII*(2), 21–33.

Asur, S., & Huberman, B. A. (2010). Predicting the future with social media. *Proceedings - 2010 IEEE/WIC/ACM International Conference on Web Intelligence, WI 2010*, *1*, 492–499.

Barker, D., & Smith, D. J. H. (1995). Technology foresight using roadmaps. *Long Range Planning*, *28*(2), 21–28.

Barreira, N., Godinho, P., & Melo, P. (2013). Nowcasting unemployment rate and new car sales in south-western Europe with Google Trends. *NETNOMICS: Economic Research and Electronic Networking*, *14*(3), 129–165.

Bastian, M., Heymann, S., & Jacomy, M. (2009). Gephi: An Open Source Software for Exploring and Manipulating Networks. *Third International AAAI Conference on Weblogs and Social Media*, 361–362.

Battistella, C., De Toni, A. F., & Pillon, R. (2015). The Extended Map methodology: Technology roadmapping for SMES clusters. *Journal of Engineering and Technology Management - JET-M*, *38*, 1–23.

Belleflamme, P., Lambert, T., & Schwienbacher, A. (2013). Individual crowdfunding practices. *Venture Capital*, *15*(4), 313–333.

Belleflamme, P., Lambert, T., & Schwienbacher, A. (2014). Crowdfunding: Tapping the right crowd. *Journal of Business Venturing*, *29*(5), 585–609.

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

Bonino, D., Ciaramella, A., & Corno, F. (2010). Review of the state-of-the-art in patent information and forthcoming evolutions in intelligent patent informatics. *World Patent Information*, *32*(1), 30–38.

Brandell, J., & Varkas, T. (2001). The Handbook of Social Work Research Methods. Thousand Oaks, California: SAGE Publications, Inc.

Burt, G., Wright, G., Bradfield, R., Cairns, G., & van der Heijden, K. (2006). The Role of Scenario Planning in Exploring the Environment in View of the Limitations of PEST and Its Derivatives. *International Studies of Management and Organization*, *36*(3), 50–76.

Cachia, R., Compañó, R., & Da Costa, O. (2007). Grasping the potential of online social networks for foresight. *Technological Forecasting and Social Change*, *74*(8), 1179–1203.

Cameron, M. P., Barrett, P., & Stewardson, B. (2016). Can Social Media Predict Election Results? Evidence From New Zealand. *Journal of Political Marketing*, *15*(4), 416–432.

Card, S. K., Mackinlay, J. D., & Shneiderman, B. (1999). *Readings in information visualization : using vision to think*. San Francisco: Morgan Kaufmann.

Carvalho, M. M., Fleury, A., & Lopes, A. P. (2013). An overview of the literature on technology roadmapping (TRM): Contributions and trends. *Technological Forecasting and Social Change*, *80*(7), 1418–1437.

Chen, M., Mao, S., & Liu, Y. (2014). Big data: A survey. *Mobile Networks and Applications*, *19*(2), 171–209.

Choi, H., & Varian, H. (2012). Predicting the Present with Google Trends. *Economic Record*, *88*(SUPPL.1), 2–9.

Choi, S., Kim, H., Yoon, J., Kim, K., & Lee, J. Y. (2013). An SAO-based text-mining approach for technology roadmapping using patent information. *R and D Management*, *43*(1), 52–74.

Cleven, A., Gubler, P., & Hüner, K. M. (2009). Design alternatives for the evaluation of design science research artifacts. *Proceedings of the 4th International Conference on Design Science Research in Information Systems and Technology*, (October 2016), 1–8.

Coates, J. (1985). Foresight in Federal Government Policymaking. *Futures Research Quarterly*, *1*(2), 29–53.

Coates, V., Farooque, M., Klavans, R., Lapid, K., Linstone, H. A., Pistorius, C., & Porter, A. L. (2001). On the future of technological forecasting. *Technological Forecasting and Social Change*, *67*(1), 1–17.

CrowdExpert. (2016). • Money raised by crowdfunding globally 2012-2015 | Statistic. *Statista - The Statistics Portal*. Retrieved March 17, 2018, from https://www-statista-com.ezproxy2.utwente.nl/statistics/620952/total-crowdfunding-volume-worldwide/

Crowdsourcing.org. (2012). Crowdfunding industry report - market trends, composition and crowdfunding platforms. *Research report*, (May), 1–30.

Dijkman, R., & Ipeirotis, P. (2015). Using Twitter to predict sales: a case study. *arXiv preprint arXiv: 1503.04599.*, 1–14.

van der Duin, P. (2016). *Foresight in organizations: Methods and tools*. *Foresight in Organizations: Methods and Tools*. New York, NY: Routledge.

Elshendy, M., Colladon, A. F., Battistoni, E., & Gloor, P. A. (2017). Using four different online media sources to forecast the crude oil price. *Journal of Information Science*, 0165551517698298.

Feldmann, N., Gimpel, H., Kohler, M., & Weinhardt, C. (2013). Using crowd funding for idea assessment inside organizations: Lessons learned from a market engineering perspective. *Proceedings - 2013 IEEE 3rd International Conference on Cloud and Green Computing*, (May 2015), 525–530.

Garcia, M. L., & Bray, O. H. (1997). *Fundamentals of technology roadmapping*. Albuquerque, NM.

Gartner. (2017). Top Trends in the Gartner Hype Cycle for Emerging Technologies, 2017 - Smarter With Gartner. Retrieved March 17, 2018, from https://www.gartner.com/smarterwithgartner/top-trends-in-the-gartner-hype-cycle-for-emerging-technologies-2017/

Geerlings, H., & Rienstra, S. (2003). Exploring “weak signals” as potential challenges in transport: An inventory of potential future technology developments. *Transportation Planning and Technology*, *26*(6), 469–489.

Georghiou, L., Harper, J. C., Keenan, M., Miles, I., Popper, R., & Elgar, E. (2009). *The Handbook of Technology Foresight*. *Technological Forecasting & Social Change* (Vol. 76). Aldershot: Edward Elgar.

Gerken, J., Moehrle, M., & Walter, L. (2010). Patents as an information source for product forecasting: Insights from a longitudinal study in the automotive industry. *The R&D Management Conference*.

Geum, Y., Lee, H. J., Lee, Y., & Park, Y. (2015). Development of data-driven technology roadmap considering dependency: An ARM-based technology roadmapping. *Technological Forecasting and Social Change*, *91*, 264–279.

Gierczank, M. M., Bretschneider, U., Hass, P., Blohm, I., & Leimeister, J. M. (2015). Crowdfunding: Outlining the new era of fundraising. *Crowdfunding in Europe: State of the Art in Theory and Practice* (pp. 7–23). Cham: Springer.

Gilgun, J. F. (1994). A Case for Case Studies in Social Work Research. *Social Work*, *39*(4), 371–380.

Glassey, O. (2012). Folksonomies: Spontaneous crowd sourcing with online early detection potential? *Futures*, *44*(3), 257–264.

Greenberg, M. D., Hariharan, K., Gerber, E., & Pardo, B. (2013). Crowdfunding Support Tools: Predicting Success & Failure. *CHI 2013, Changing Perspectives, Paris, France*, 1815–1820.

van der Heijden, K. (2005). *Scenarios: The Art of Strategic Conversation*. *Journal of the Operational Research Society* (2nd ed.). John Wiley & Sons.

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

Hienerth, C., & Riar, F. (2013). The Wisdom of the Crowd vs. Expert Evaluation: A Conceptualization of Evaluation Validity. *35th DRUID Celebration Conference 2013*, 1–31.

Hiltunen, E. (2008). The future sign and its three dimensions. *Futures*, *40*(3), 247–260.
Hofmann, M., & Chisholm, A. (2016). *Text Mining And Visualization: Case Studies Using Open-Source Tools*. *Taylor & Francis Group*.
Holopainen, M., & Toivonen, M. (2012). Weak signals: Ansoff today. *Futures*, *44*(3), 198–205.
Horton, A. (1999). A simple guide to successful foresight. *Foresight*, *1*(1), 5–9.
Hussain, M., Tapinos, E., & Knight, L. (2017). Scenario-driven roadmapping for technology foresight. *Technological Forecasting and Social Change*, *124*, 160–177.
HYVE. (2018). HYVE - the innovation company. Retrieved March 17, 2018, from https://www.hyve.net/en/
Irvine, J., & Martin, B. R. (1984). *Foresight in Science: Picking the Winners*. London: Pinter.
Jin, G., Jeong, Y., & Yoon, B. (2014). Technology-driven roadmaps for identifying new product/market opportunities: Use of text mining and quality function deployment. *Advanced Engineering Informatics*, *29*(1), 126–138.
Kajikawa, Y., Kikuchi, Y., Fukushima, Y., & Koyama, M. (2013). Utilizing risk analysis and scenario planning for technology roadmapping. *Strategic Planning Decisions in the High Tech Industry* (pp. 231–244). London: Springer.
Kayser, V. (2016). *Extending the Knowledge Base of Foresight: The Contribution of Text Mining*. Technische Universität Berlin.
Kayser, V., & Blind, K. (2017). Extending the knowledge base of foresight: The contribution of text mining. *Technological Forecasting and Social Change*, *116*, 208–215.
Keller, T., & Tergan, S.-O. (2005). Knowledge and Information Visualization: An Introduction. *Knowledge and information visualization: Searching for synergies* (pp. 1–23).
Kickstarter. (2018). Kickstarter in Zahlen. *2018*. Retrieved March 17, 2018, from https://www.kickstarter.com/help/stats?lang=de
Kim, J., & Lee, C. (2017). Novelty-focused weak signal detection in futuristic data: Assessing the rarity and paradigm unrelatedness of signals. *Technological Forecasting and Social Change*, *120*, 59–76.
Kostoff, R. N., & Schaller, R. R. (2001). Science and technology roadmaps. *IEEE Transactions on Engineering Management*, *48*(2), 132–143.
Kuosa, T. (2010). Futures signals sense-making framework (FSSF): A start-up tool to analyse and categorise weak signals, wild cards, drivers, trends and other types of information. *Futures*, *42*(1), 42–48.
Lee, S., Lee, S., Seol, H., & Park, Y. (2008). Using patent information for designing new product and technology: keyword based technology roadmapping. *R&D Management*, *38*(2), 169–188.
Leimeister, J. M. (2012). Crowdsourcing: Crowdfunding, Crowdvoting, Crowdcreation. *Zeitschrift für Controlling und Management (ZFCM)*, *56*(2012), 388–392.
Lin, Y., Boh, W. F., & Goh, K. H. (2014). How Different are Crowdfunders? Examining Archetypes of Crowdfunders and Their Choice of Projects. *Academy of Management Proceedings*, *2014*(1), 13309–13309.
Macht, S. A., & Weatherston, J. (2014). The Benefits of Online Crowdfunding for Fund-Seeking Business Ventures. *Strategic Change*, *23*(1–2), 1–14.
Martin, B. R. (1995). Foresight in Science and Technology. *Technology Analysis & Strategic Management*, *7*(2), 139–168.
Martin, B. R. (2010). The origins of the concept of “foresight” in science and technology: An insider’s perspective. *Technological Forecasting and Social Change*, *77*(9), 1438–1447.
Martino, J. P. (1993). Technological Forecasting for Decision Making. *Interfaces*, *25*(1), p137--139.
Mietzner, D., & Reger, G. (2005). Advantages and disadvantages of scenario approaches for strategic foresight. *Int. J. of Technology Intelligence and Planning*, *1*(2), 220–239.
Miles, I. (2010). The development of technology foresight: A review. *Technological Forecasting and Social Change*, *77*(9), 1448–1456.
Miles, I., Meissner, D., Vonortas, N. S., & Carayannis, E. (2017). Technology foresight in transition. *Technological Forecasting and Social Change*, *119*, 211–218.
Moehrle, M. G., Isenmann, R., & Phaal, R. (2013). *Technology roadmapping for strategy and*

innovation: Charting the route to success. Berlin: Springer.
Moritz, A., & Block, J. H. (2016). Crowdfunding: A Literature Review and Research Directions. *Crowdfunding in Europe*, *2*, 1–251.
Oliver Schwarz, J. (2008). Assessing the future of futures studies in management. *Futures*, *40*(3), 237–246.
Ordanini, A., Miceli, L., Pizzetti, M., & Parasuraman, A. (2011). Crowd-funding: transforming customers into investors through innovative service platforms. *Journal of Service Management*, *22*(4), 443–470.
Pagolu, V. S., Challa, K. N. R., Panda, G., & Majhi, B. (2016). Sentiment Analysis of Twitter Data for Predicting Stock Market Movements. *2016 International Conference on Signal Processing, Communication, Power and Embedded System (SCOPES)*, 1345–1350.
Park, C., & Cho, S. (2017). Future sign detection in smart grids through text mining. *Energy Procedia*, *128*, 79–85.
Park, H., Ree, J. J., & Kim, K. (2013). Identification of promising patents for technology transfers using TRIZ evolution trends. *Expert Systems with Applications*, *40*(2), 736–743.
Peffers, K., Tuunanen, T., Rothenberger, M. A., & Chatterjee, S. (2007). A Design Science Research Methodology for Information Systems Research. *Journal of Management Information Systems*, *24*(3), 45–77.
Phaal, R., Farrukh, C. J. P., & Probert, D. R. (2004). Technology roadmapping - A planning framework for evolution and revolution. *Technological Forecasting and Social Change*, *71*(1–2), 5–26.
Phaal, R., & Muller, G. (2009). An architectural framework for roadmapping: Towards visual strategy. *Technological Forecasting and Social Change*, *76*(1), 39–49.
Pietrobelli, C., & Puppato, F. (2016). Technology foresight and industrial strategy. *Technological Forecasting and Social Change*, *110*, 117–125.
Popper, R. (2008). How are foresight methods selected? *Foresight*, *10*(6), 62–89.
Porter, A. L., Ashton, W. B., Clar, G., Coates, J. F., Cuhls, K., Cunningham, S. W., Ducatel, K., et al. (2004). Technology futures analysis: Toward integration of the field and new methods. *Technological Forecasting and Social Change*, *71*(3), 287–303.
Ramirez, R., & Wilkinson, A. (2014). Rethinking the 2×2 scenario method: Grid or frames? *Technological Forecasting and Social Change*, *86*, 254–264.
Sanger, J., & Feldman, R. (2007). *The text mining handbook: Advanced approaches in analyzing unstructured data*. Cambridge university press.
Saritas, O. (2013). Systemic foresight methodology. *Science, Technology and Innovation Policy for the Future: Potentials and Limits of Foresight Studies* (pp. 83–117).
Saritas, O., & Aylen, J. (2010). Using scenarios for roadmapping: The case of clean production. *Technological Forecasting and Social Change*, *77*(7), 1061–1075.
Saritas, O., & Burmaoglu, S. (2015). The evolution of the use of Foresight methods: a scientometric analysis of global FTA research output. *Scientometrics*, *105*(1), 497–508.
Schatzmann, J., Schäfer, R., & Eichelbaum, F. (2013). Foresight 2.0 - Definition, overview & evaluation. *European Journal of Futures Research*, *1*(1), 15.
Slaughter, R. A. (1989). Probing beneath the surface. Review of a decade's futures work. *Futures*, *21*(5), 447–465.
Slaughter, R. A. (1997). Developing and applying strategic foresight. *ABN Report*, *5*(10), 13–27.
Smith, A. (2002). The Theory of Moral Sentiments. *Library*, *1*(1790), 1–405.
Tapinos, E. (2012). Perceived Environmental Uncertainty in scenario planning. *Futures*, *44*(4), 338–345.
The Crowdfunding Center. (2016). Leading crowdfunding platforms in the technology sector worldwide in 2016, by value of funds raised (in 1,000 U.S. dollars). *Statista - The Statistics Portal*. Retrieved March 17, 2018, from https://www-statista-com.ezproxy2.utwente.nl/statistics/758768/leading-technology-crowdfunding-platforms-by-funds-raised/
Toivonen, M. (2004). *Expertise as Business. Long-Term Development and Future Prospects of Knowledge-Intensive Business Services (KIBS). Laboratory of Industrial Management* (Vol. 2).

Tran, T. A., & Daim, T. (2008). A taxonomic review of methods and tools applied in technology assessment. *Technological Forecasting and Social Change*, *75*(9), 1396–1405.
Tseng, Y.-H., Lin, C.-J., & Lin, Y.-I. (2007). Text mining techniques for patent analysis. *Information Processing & Management*, *43*(5), 1216–1247.
Uhl, A., Kolleck, N., & Schiebel, E. (2017). Twitter data analysis as contribution to strategic foresight-The case of the EU Research Project “Foresight and Modelling for European Health Policy and Regulations” (FRESHER). *European Journal of Futures Research*, *5*(1).
Voros, J. (2003). A generic foresight process framework. *Foresight*, *5*(3), 10–21.
Wack, P. (1985). Scenarios: Shooting the rapids. How medium-term analysis illuminated the power of scenarios for Shell management. *Harvard Business Review*, *63*(6), 139–150.
Webrobots. (2018). Kickstarter Datasets – Web Scraping Service. Retrieved March 17, 2018, from https://webrobots.io/kickstarter-datasets/
Weiss, S. M., Indurkhya, N., Zhang, T., & Damerau, F. J. (2005). *Text mining: Predictive methods for analyzing unstructured information*. Springer Science & Business Media.
Wijnhoven, F., & Plant, O. (2018). Sentiment Analysis and Google Trends Data for Predicting Car Sales. *ICIS 2017: Transforming Society with Digital Innovation*, (2014), 1–16.
Yaqoob, I., Hashem, I. A. T., Gani, A., Mokhtar, S., Ahmed, E., Anuar, N. B., & Vasilakos, A. V. (2016). Big data: From beginning to future. *International Journal of Information Management*, *36*(6), 1231–1247.
Yoon, B. (2010). Strategic visualisation tools for managing technological information. *Technology Analysis and Strategic Management*, *22*(3), 377–397.
Yoon, B., & Park, Y. (2004). A text-mining-based patent network: Analytical tool for high-technology trend. *Journal of High Technology Management Research*, *15*(1), 37–50.
Yoon, B., Phaal, R., & Probert, D. (2008). Morphology analysis for technology roadmapping: application of text mining. *R&D Management*, *38*(1), 51–68.
Yoon, J. (2012). Detecting weak signals for long-term business opportunities using text mining of Web news. *Expert Systems with Applications*, *39*(16), 12543–12550.
Yoon, J., Park, H., & Kim, K. (2013). Identifying technological competition trends for R&D planning using dynamic patent maps: SAO-based content analysis. *Scientometrics*, *94*(1), 313–331.
Yuan, H., Lau, R. Y. K., & Xu, W. (2016). The determinants of crowdfunding success: A semantic text analytics approach. *Decision Support Systems*, *91*, 67–76.

APPENDIX

APPENDIX 1 – Theoretical framework: The use of web and patent data in technology foresight and technology forecasting

Data Input	Name of authors	Focus	Purpose of the study	Results	Limitations & remarks
Patents	Lee et al. (2008)	Foresight	Using text mining of patent data to support technology roadmapping	Three types of maps can be applied to specific roadmapping steps and improve the effectiveness of the technique	The proposed methods require additional efforts. Patent documents may not be applicable in all fields. The study is restricted to a certain area.
	Abbas et al. (2014)	Foresight	Literature review on patent analysis	Presents state of the art in patent analysis and a taxonomy of patent analysis trends.	Certain areas (such as SAO extraction) still need improvements. Approaches could also be used for documents other than patents. Approaches should offer multiple suggestions for devising strategies.
	Yoon et al. (2013)	Foresight	Identifying technological competition trends using dynamic patent maps: SAO-based content analysis	The proposed patent maps can support experts in understanding technological competition trends in the process of formulating R&D strategies.	The used methods (MDS and k-means clustering) can cause information loss and ineffectiveness.
	Jin et al. (2015)	Foresight	Using text mining of patent data for technology-driven roadmaps	The study suggests two matrices to be used in technology roadmapping in order to find new business opportunities.	The research cannot be applied to all new technology. The study is highly dependent on experts.
	Choi et al. (2013)	Foresight	Using text mining of patent data to support technology roadmapping	The proposed approach provides a quantitative method that uses patent information and represents the new type of TRM diagram as a function layer.	The process of SAO extraction is very time consuming and inefficient and requires many manual adaptations.
	Kim & Lee (2017)	Foresight	Detecting weak signals through the use of text mining of websites and patent data	signal-portfolio maps are developed to identify the patterns of signal representations	research of extracting meaningful intelligence from futuristic data is in its infant stage and is in need of development. there is still room for incorporating more characteristics.

Web news articles	Yoon (2012)	Foresight	Detecting early indicators in web news articles to identify future business opportunities	The paper shows the possibility of quantifying the detection of weak signals by text mining of Web news. The proposed method can detect weak signals in large datasets more efficiently than human experts.	Only web news articles are used and only for a specific technological field. The method should be applied using other data sources and investigate further technologies.
	Park & Cho (2017)	Foresight	Detecting future signs in smart grids through text mining.	Based on Yoon's (2012) proposed methods they were able to examine weak signals.	The study helps to promote the development of research in future sign analysis through text mining, however does not come up with new research methods.
Social media data	Cachia et al. (2007)	Foresight	Discussing the relevance of online social networks (OSN) for foresight methods	OSN can contribute to foresight analysis since they indicate emerging changes in social behavior and enhance collaborative intelligence and creativity	The paper only explains the theoretical benefits of OSN for foresight and does not present concrete tools. An analytical tool to conduct foresight through social networks should be established.
	Glassey (2012)	Foresight	analyzing the potential contribution of online metadata (in the form of folksonomies) to early trend detection	Folksnonomies can contribute to mitigate filters hindering weak signals detection and processing	Folksnonomies are a young field of research and its ongoing evolution has to be explored and documented. Their use contains several limitations.
	Kayser & Blind (2017)	Foresight	Exploring the potential of text mining for foresight by considering different data sources, text mining approaches, and foresight methods	text mining can facilitate the detection and examination of emerging topics and technologies by extending the knowledge base of foresight	foresight only functions as a combination of qualitative and quantitative thinking. different combinations of data sources, text mining approaches, requirements and foresight methods should be established in future work.
	Uhl et al. (2017)	Foresight	Twitter data analysis as contribution to strategic foresight	Twitter can be regarded as useful tool for gathering current, to complement the scanning process, and to support the monitoring during the foresight process. The evaluation of web-links in the dataset lead to current information about the development of certain technologies.	The limited time frame makes it impossible to make assumptions about topical trends, there are certain limitations associated with hashtag-based approaches, Twitter data is not representative of a population
	Elshendy, Colladon, Battistoni, & Gloor (2017)	Prediction	Using online media data to forecast oil prices	Social parameters from different media platforms show a high correlation with the oil price movements.	The collected data sources have different limitations in terms of structure and user activities (Twitter, Wikipedia, Google trends, GDELT).

	Asur & Huberman (2010)	Prediction	Conducting sentiment analysis of Twitter data to forecast box-office revenues for movies	Social media can be utilized to forecast future outcomes.	Study focusses on predicting box office revenues
	Pagolu, Challa, Panda, & Majhi (2016)	Prediction	Predicting a company's stock price movements through sentiment analysis of Twitter data	It is shown that a strong correlation exists between the rise and falls in stock prices with the public sentiments in tweets.	Only Twitter data is analyzed which may be biased.
	Dijkman & Ipeirotis (2015)	Prediction	Using Twitter data to predict sales	The paper shows how relations between Twitter activities and sales can be identified by classifying Tweets. Positive Tweets by persons can be used to forecast sales.	Further investigations about the forecasting of peaks in sales is necessary and whether there is a causal relation between Tweets and sales
	Cameron, Barrett, & Stewardson (2016)	Prediction	Predicting election results through social media	There is a significant relationship between the size of online social networks and election voting and election results	The shown effect is small, data could be combined with poll data in order to improve predictions
Google trends	Wijnhoven & Plant (2018)	Prediction	Exploring the usefulness of sentiment analysis and Google trends data for car sales forecasting	Google trends could be used by the car industry to effectively predict car sales while social media sentiment have little predictive power	The results are restricted to the Dutch car industry. Since this study provided unexpected results regarding the predictive power of sentiment analysis, further research in this direction has to be conducted
	Barreira, Godinho, & Melo (2013)	Prediction	Nowcasting car sales and unemployment rates with Google trends	Google Trends data usually leads to improvements in unemployment predictions in three of the considered countries, but little support is given that these data may improve car sales forecasts.	More sophisticated benchmark models should be used in future studies.
	Choi & Varian (2012)	Prediction	Predicting the present with Google trends	Google trends data can be used to predict near-term values of economic indicators such as car sales, unemployment claims.	Further and more detailed research using Google trends data has to be conducted.

Table 6. Existing studies using web and patent data for technology foresight/forecasting.

APPENDIX 2 – Market portfolio maps

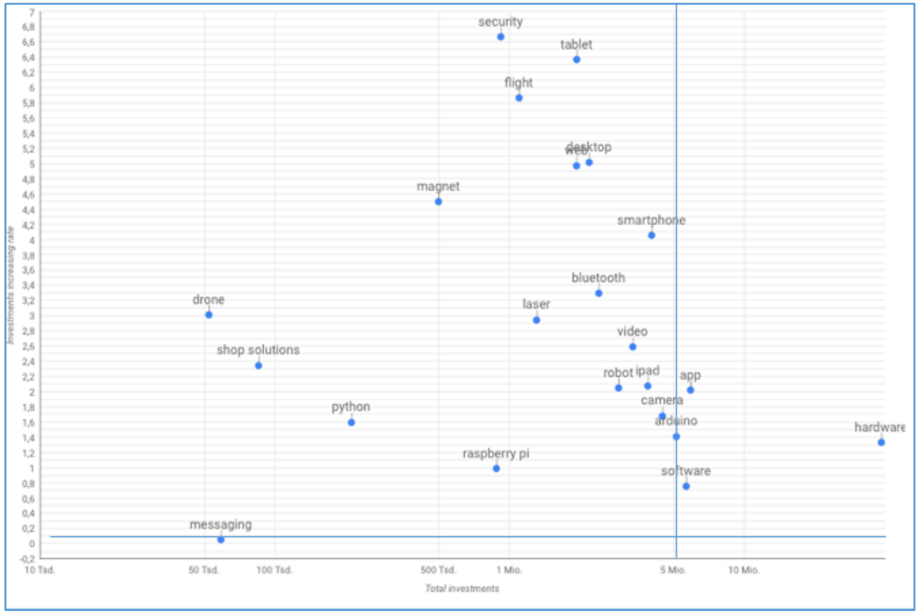

Figure 24. Market portfolio map, years 2009-2013.

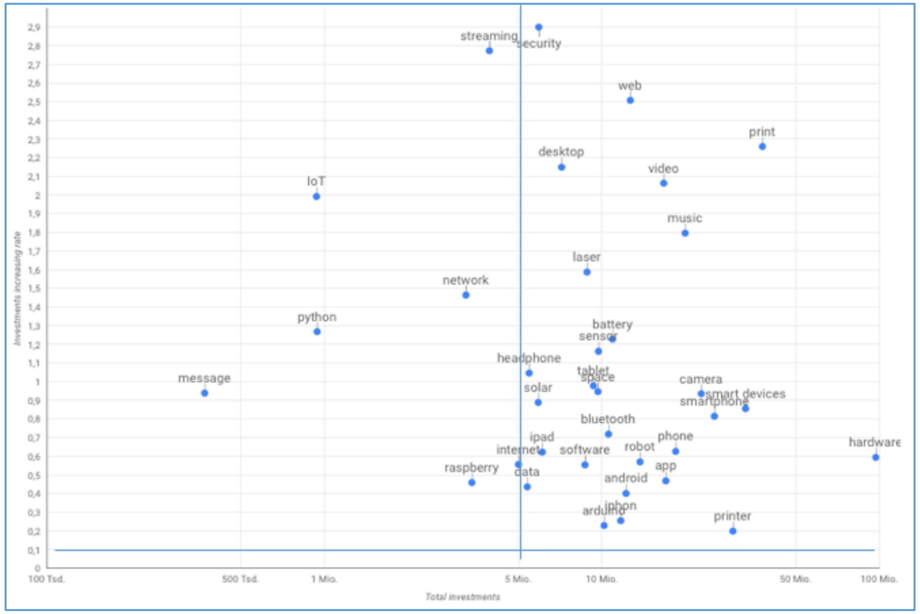

Figure 25. Market portfolio map, years 2013-2015.

APPENDIX 3 – correlation matrix RTF and RIR values (6 years)

	RelTer~0	RelTer~1	RelTer~2	RelTer~3	RelTer~4	RelTer~5	RelTer~6	RelInv~0	RelInv~1	RelInv~2	RelInv~3	RelInv~4	RelInv~5	RelInv~6
RelTermFre~0	1.0000													
RelTermFre~1	0.9956	1.0000												
RelTermFre~2	0.9723	0.9869	1.0000											
RelTermFre~3	0.9296	0.9494	0.9766	1.0000										
RelTermFre~4	0.3904	0.4108	0.4888	0.6357	1.0000									
RelTermFre~5	-0.1641	-0.1756	-0.0985	-0.0621	0.4786	1.0000								
RelTermFre~6	0.7882	0.7870	0.7683	0.7884	0.4518	-0.0475	1.0000							
RelInvRate~0	0.9878	0.9756	0.9398	0.8813	0.3367	-0.1144	0.8103	1.0000						
RelInvRate~1	0.9862	0.9760	0.9362	0.8619	0.2690	-0.1830	0.7496	0.9917	1.0000					
RelInvRate~2	0.9695	0.9783	0.9651	0.9151	0.3756	-0.1202	0.8230	0.9721	0.9649	1.0000				
RelInvRate~3	0.9256	0.9455	0.9594	0.9051	0.3333	-0.0982	0.7916	0.9185	0.9120	0.9706	1.0000			
RelInvRate~4	0.1453	0.1525	0.1061	0.2384	0.3989	-0.2685	0.4735	0.1401	0.0844	0.1716	0.0310	1.0000		
RelInvRate~5	-0.2506	-0.2752	-0.2274	-0.2204	0.1918	0.8641	0.0583	-0.1711	-0.2449	-0.2018	-0.1491	-0.2857	1.0000	
RelInvRate~6	0.5281	0.5922	0.6584	0.6139	0.2180	-0.0240	0.4853	0.5159	0.5196	0.6925	0.7867	-0.0336	-0.1180	1.0000

Table 7. Correlation matrix, RTF and RIR values (6 years).

APPENDIX 4 – linear regression RTF and RIR values, including two-year time lag

LINEAR REGRESSSIONS

	(1)	(2)	(3)	(4)
	Relative Investment Rate	Relative Investment Rate	Relative Investment Rate	Relative Investment Rate
RELTERMFREQ_L0	3.2558***	3.4770***		
	(0.000)	(0.000)		
RELTERMFREQ_L1			3.5751***	
			(0.000)	
RELTERMFREQ_L2				4.2260***
				(0.000)
CONSTANT	-0.0117**	-0.0164	-0.0204	-0.0170
	(0.007)	(0.068)	(0.123)	(0.272)
TECHNOLOGY FE	*NO*	*YES*	*YES*	*YES*
YEAR FE	*NO*	*YES*	*YES*	*YES*
N	56.0000	56.0000	48.0000	40.0000
R2	0.7566	0.7917	0.5867	0.4784
R2_O	0.8214	0.8422	0.7109	0.6026
R2_B	0.9346	0.9346	0.8841	0.7661
R2_W	0.7566	0.7917	0.5867	0.4784
SIGMA_U	0.0050	0.0111	0.0168	0.0308
SIGMA_E	0.0243	0.0241	0.0351	0.0391
RHO	0.0407	0.1750	0.1861	0.3829
P-VALUES IN PARENTHESES				
="* P<0.05	** p<0.01	*** p<0.001"		

Table 8. Linear regression RTF and RIR values, including two-year time lag.