Technological Forecasting & Social Change 116 (2017) 208–215

Contents lists available at ScienceDirect

Technological Forecasting & Social Change

Extending the knowledge base of foresight: The contribution of text mining

Victoria Kayser [a,b,*], Knut Blind [b,c]

[a] Fraunhofer Institute for Systems and Innovation Research, Breslauer Strasse 48, 76131 Karlsruhe, Germany
[b] Technische Universität Berlin, Chair of Innovation Economics, Marchstraße 23, D-10587 Berlin, Germany
[c] Fraunhofer Institute for Open Communication Systems and Rotterdam School of Management, Chair of Standardisation, Erasmus University, Burgemeester Oudlaan 50, 3062 PA Rotterdam, The Netherlands

ARTICLE INFO

Article history:
Received 9 February 2016
Received in revised form 23 September 2016
Accepted 18 October 2016
Available online 2 December 2016

Keywords:
Foresight
Text mining
Data analysis
Roadmapping
Scenario development
Big data

ABSTRACT

While the volume of data from heterogeneous sources grows considerably, foresight and its methods rarely benefit from such available data. This work concentrates on textual data and considers its use in foresight to address new research questions and integrate other stakeholders. This textual data can be accessed and systematically examined through text mining which structures and aggregates data in a largely automated manner. By exploiting new data sources (e.g. Twitter, web mining), more actors and views are integrated, and more emphasis is laid on the analysis of social changes. The objective of this article is to explore the potential of text mining for foresight by considering different data sources, text mining approaches, and foresight methods. After clarifying the potential of combining text mining and foresight, examples are outlined for roadmapping and scenario development. As the results show, text mining facilitates the detection and examination of emerging topics and technologies by extending the knowledge base of foresight. Hence, new foresight applications can be designed. In particular, text mining provides a solid base for reflecting on possible futures.

1. Introduction

The volume of data from heterogeneous sources has considerably grown (Ortner et al., 2014) and the scientific output is constantly increasing (see, e.g., Bornmann and Mutz, 2014). Identifying the relevant data from the huge quantity of available information is challenging and more effort is needed in monitoring thematic fields. Furthermore, (textual) data from websites or social media could be analyzed to address new aspects and research questions (see e.g., Boyd and Crawford, 2012; Kitchin, 2014). For example, the user-generated content on the web may be interesting in the context of foresight, for examining social perspectives and the user's perception of current developments. However, the web data is at present rarely considered for a systematic examination (Yoon, 2012; Cachia et al., 2007; Glassey, 2012). However, new indicators could be established to extend the scheme of present indicators with their focus on science and technology through patent and publication analysis (Kostoff, 2012; Abbas et al., 2014). Today, many relevant information sources are left out, though this data could be used to perceive ongoing changes and make more precise statements about possible future developments and emerging technologies. Therefore, new methods and tools for processing and integrating data for foresight are required.

This article focuses on textual data (e.g., reports, blog entries, or Twitter data) that can be accessed and systematically examined through text mining (Berry, 2004; Feldman and Sanger, 2008). By integrating text mining into foresight, other data sources are accessible to be considered in a comprehensive way, especially unstructured and large datasets. Therefore, the objective of this article is to identify and elaborate the potential of text mining for foresight and its methods. One aspect is to consider data sources such as Twitter or websites and new techniques for data retrieval such as web mining. This article examines the extent to which foresight and its methods can be improved through the results of text mining. Therefore, applications are presented wherein text mining is combined with foresight methods such as roadmapping (e.g., Möhrle et al., 2013; Phaal et al., 2010) or scenario development (e.g., Reibnitz, 1991; van der Heijden, 2005; O'Brien and Meadows, 2013). By additional data and stakeholder views, it is expected to enhance the detection and examination of emerging themes and technologies and to provide a solid base for decision making.

This article begins with the fundamentals of foresight and the basic principles of text mining in Section 2. Section 3 addresses the use of text mining for foresight. Different data sources are described, the state-of-the-art concerning existing implementations is summarized, and further applications are outlined. Finally, the results are discussed in the framework of foresight and a conclusion is drawn in Section 4.

* Corresponding author.
E-mail address: mail@vkayser.de (V. Kayser).

http://dx.doi.org/10.1016/j.techfore.2016.10.017
0040-1625/

2. Foresight and text mining

This section introduces the two main components of this article—foresight and text mining—and gives an overview of the recent debate.

2.1. Foresight

Foresight is a systematic process of looking into the long-term future of science, technology, and innovation (e.g., Martin, 1995; Cuhls, 2003). One definition of foresight is *"opening to the future with every means at our disposal, developing views of future options, and then choosing between them"* (Slaughter, 1995). Foresight thereby considers possible and plausible futures, since there is not only one future. In principal, the future cannot be predicted but is shaped by the decisions and actions of today. Foresight helps to assess the consequences and implications of present actions, early warnings, thinking about desirable futures, and implications of possible future events. Therefore, it is an action-oriented decision support that brings together the relevant stakeholders for an open discourse about possible futures. Foresight builds on a set of different methods (e.g., Popper and Butter, 2008) such as roadmapping (Barker and Smith, 1995; Möhrle et al., 2013) or scenario development (Reibnitz, 1991; van der Heijden, 2005). The set of methods to be selected for application depends on the scope and focus of the foresight exercise and has to be decided from case to case. Foresight, futures studies, and future technology analysis are not further distinguished in the course of this article due to their commonalities.

Foresight is modular and a sequence of steps. Depending on the objectives and application level, different methods and tasks are combined. Building on previous studies (see, e.g., Martin, 1995; Horton, 1999; Voros, 2003; Da Costa et al., 2008; de Miranda Santo et al., 2006), foresight exercises might be categorized into three phases as illustrated in Fig. 2-1—input, process, and output.

2.1.1. Input

Besides some overall objectives, a process scope is defined, a time horizon is set, and information about recent trends and developments is gathered with regard to the considered field. At the beginning of almost every process, the state-of-the-art has to be summarized. Therefore, this first step relates to collecting and summarizing the available information to get an overview of the present situation (Horton, 1999).

2.1.2. Process

Future technology analysis might be seen as a process of knowledge creation (Eerola and Miles, 2011). Specific foresight methods are applied according to the scope and process objectives. By this, important information about the future and possible future developments is gathered and knowledge is generated, which later serves as decision support.

2.1.3. Output

The results are assessed, priorities are set, and strategies are formulated (de Miranda Santo et al., 2006). This phase is about taking action (Horton, 1999). Diverse interests or expectations related to foresight outcomes exist. One intention of foresight is to support the design of future-oriented strategies. Furthermore, political or governmental actors expect recommendations for planning or setting priorities for research programs (Havas et al., 2010; Könnölä et al., 2011; de Smedt, 2013).

2.2. Text mining

Due to the increasing volume of data from heterogeneous sources, the effort needed to study thematic fields and developments and to read the published studies and literature has increased (Ortner et al., 2014). In times of big data, techniques are required that can automatically process this data and be adapted to varying requirements and data sets. Text mining is suitable for this purpose as it offers methods to access and analyze these textual data sources (Weiss, 2010; Feldman and Sanger, 2008) as introduced in the following.

2.2.1. The process of text mining

Text mining processes unstructured textual data into a structured format for further analysis. Text mining can be summarized in three steps, as indicated in Fig. 2-2. First, a data source is selected. Then this data is preprocessed (Step 2) and analyzed (Step 3). Finally, the results are interpreted.

2.2.1.1. Text selection. The selected data source should be able to answer the research question (e.g. social media, patents, standards). For searching data, at least some principal knowledge of the subject or technology under consideration is necessary. While some data is retrieved from databases (e.g. patent, standards, scientific publications), other data has to be manually gathered (e.g. reports).

2.2.1.2. Text preprocessing. Text has to be structured and transformed into a machine- readable format for further processing. Therefore, the text is divided into its individual elements as words (tokenization). To extract the relevant terms, two different approaches are distinguished—working with stopwords and working with grammatical instances. When using the grammatical instances, speech tags are assigned to each word, such as verb, article, or noun. From this, relevant phrases or chains of words are extracted. Alternatively, stopwords are used to remove irrelevant terms and function words (articles, conjunctions, pronouns, etc.). Further techniques like stemming (which reduces each word to its basic form) or

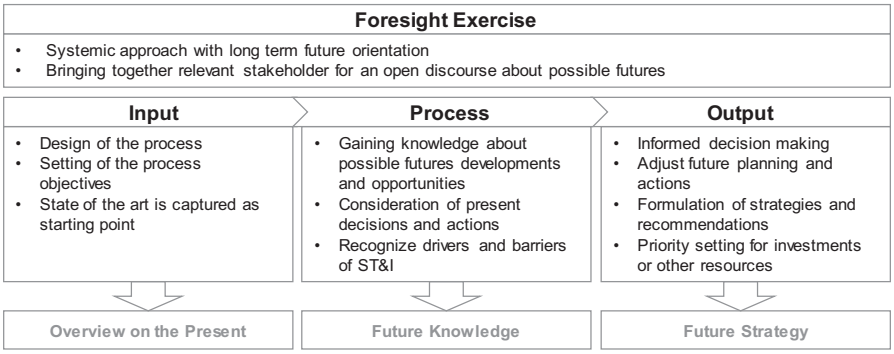

Fig. 2-1. The process of foresight.