Visual Analytics for Strategic Decision Making in Technology Management

Kawa Nazemi, Tim Feiter, Lennart B. Sina, Dirk Burkhardt, and Alexander Kock

Abstract Strategic foresight, corporate foresight, and technology management enable firms to detect discontinuous changes early and develop future courses for a more sophisticated market positioning. The enhancements in machine learning and artificial intelligence allow more automatic detection of early trends to create future courses and make strategic decisions. Visual Analytics combines methods of automated data analysis through machine learning methods and interactive visualizations. It enables a far better way to gather insights from a vast amount of data to make a strategic decision. While Visual Analytics got various models and approaches to enable strategic decision-making, the analysis of trends is still a matter of research. The forecasting approaches and involvement of humans in the visual trend analysis process require further investigation that will lead to sophisticated analytical methods. We introduce in this paper a novel model of Visual Analytics for decision-making, particularly for technology management, through early trends from scientific publications. We combine Corporate Foresight and Visual Analytics and propose a machine learning-based Technology Roadmapping based on our previous work.

K. Nazemi (✉) · L. B. Sina · D. Burkhardt
Darmstadt University of Applied Sciences, 64295 Darmstadt, Germany
e-mail: kawa.nazemi@h-da.de

L. B. Sina
e-mail: lennart.b.sina@h-da.de

D. Burkhardt
e-mail: dirk.burkhardt@h-da.de

T. Feiter · A. Kock
Techniche Universität Darmstadt, 64295 Darmstadt, Germany
e-mail: tim.feiter@tu-darmstadt.de

A. Kock
e-mail: kock@tim.tu-darmstadt.de

B. Kovalerchuk et al. (eds.), *Integrating Artificial Intelligence and Visualization for Visual Knowledge Discovery*, Studies in Computational Intelligence 1014,
https://doi.org/10.1007/978-3-030-93119-3_2

1 Introduction

Visual Analytics and information visualization enable, through the combination of automated analysis methods and interactive visualizations, the decision-making process in various domains [29]. Visual Trend Analytics incorporates, in particular, the temporal dimension of data and enables identifying, detecting, and predicting technological trends to support strengthening the competitiveness of firms. Technological developments have an important impact on strategic decision-making. The early awareness of possible upcoming or emerging technological trends could strengthen enterprises' competitiveness and market positioning. The anticipation of trends into the cooperate strategy is called corporate foresight [20], and the application of corporate foresight is positively related to firms' long-term performance [50]. Qualitative methods like technology roadmaps are commonly applied but show severe weaknesses compared to qualitative approaches that are rarely applied. Visual Trend Analytics includes prediction algorithms that facilitate corporate foresight with accessible, emerging trends. This enables profound strategic decision-making and a higher acceptance rate beneath managers. Therefore, we further investigate the conceptual study of Lee et al. [30] and develop a general three-step managing process that investigates technology trends.

There exist a variety of methods for analyzing emerging or decreasing trends and predict possible future scenarios. Existing technologies and approaches commonly focus, in particular, on measuring and computing some values for identifying emerging trends [41] or providing better prediction results [37], the process of strategic decision-making is often not considered in such analytical systems. However, the process of managing information, technologies, innovations, and emerging trends are crucial for decision-making. Another important aspect for identifying emerging technology trends is data. Social media, news, company reports, and blogs refer commonly to those technologies that already reached their climax or are already available at the market. Early technology trends are often propagated first in research and scientific publications. Therefore, these data should be considered early signals and trends [43]. Although scientific publications and their value for identifying early trends are apparent, accurate analysis and identification of emerging trends out of textual scientific publications are rarely proposed. The gathering and analysis of this continuously increasing knowledge pool is a very tedious and time-consuming task and borders on the limits of manual feasibility.

We propose in this paper a new approach for Visual Analytics for decision making by incorporating some main ideas from innovation management. We first give a literature review of existing approaches and systems that are mining and visualizing trends. The literature review reveals the missing inclusion of management and decision-making approaches in such analytical systems. Our general model tries to fill this gap by a first attempt and combines an appropriate model to enable strategic decision-making. The outcome is a model with three main steps for integrating innovations in firms. This model is enhanced by a more technical approach that illustrates the process of Visual Analytics and illustrates the main steps of our approach.

This chapter enhances the three-fold contribution of our previous work [41], consisted of (1) a model for gathering trends from text to visual interactive analysis representations, (2) the identification of upcoming or emerging trends based on text, and (3) an approach for visual interaction through different data models and related interactive visual representations to explore the potentials of technologies and detect new insights, with an advanced Visual Analytics approach and system that enables decision making for technology and innovation management with advanced methods in terms of analysis and integrates the ideas of technology management in a Visual Analytics system.

2 Related Work

The literature review in this paper follows the procedure recommended by Webster and Watson [57]. We performed a "concept centric" review with the main concepts of technology forecasting methodologies, machine learning techniques, and methods for visualizations. Therefore, our literature review is subdivided into two main sections, a review of visualization techniques and a review of visual forecasting methods. The review is complemented with approaches from technology and innovation management to bridge the gap between the disciplines.

2.1 Trend and Text Visualization

Current trend mining methods provide useful indications for discovering trends [1, 14, 15, 19, 32] Nevertheless, the interpretation and conclusion for serious decision making still require a human's knowledge acquisition abilities. Therefore, the representation of trends is one of the most important aspects of analyzing trends. Common approaches often include basic visualization techniques. Depending on the concrete results, line graphs, bar charts, word clouds, frequency tables, sparklines, or histograms convey different aspects of trends. *ThemeRiver* represents thematic variations over time in a stacked graph visualization with a temporal horizontal axis [22]. The variation of the stream width indicates the strength of a specific topic over time. *Tiara* uses a similar approach, with the difference that it includes additional features such as magic lenses and an integrated graph visualization [35]. *ParallelTopics* includes a stacked graph for visualizing topic distribution over time [12]. Although the system was not designed for discovering trends but rather for analyzing large text corpora, it allows users to interactively inspect topics and their strength over time and thus allows the exploration of important trend indicators in the underlying text collection. *Parallel Tag Clouds* (PTC) is based on multiple word clouds that represent the contents of different facets in the document collection [10]. Temporal facets can be used to identify certain keywords' differences over time and infer the dynamics of themes in a text collection. Another extension of word clouds is *SparkClouds* that includes a sparkline for each word [31]. These sparklines indicate each term's temporal distribution and allow conclusions about the topic trends. A user study reveals that participants are more effective with *SparkClouds* compared to three other visualization techniques in tasks related to trend discovery [31]. A similar approach [36] also includes co-occurrence highlighting. In contrast to *SparkClouds*, this technique includes a histogram for representing the temporal relevance of each tag. Additional overlays in the histograms show the co-occurrences over time for a selected word to enable a more comprehensive analysis of trend indicators. Han et al. introduce with *PatStream* a visual trend analysis system for technology management [21]. Their system measures similarity between pairs of patents using the cosine metrics and extends the work of Heimerl et al. [23] in particular in regards to visualization. The evolution and structure of topics that indicate the trends is visualized through a *Streamgraph*, which was already proposed in the previous works of Heimerl et al. [23]. In contrast to this previous work, *Patstream* breaks down the streams into vertical time slices, which represent periods of years. These time slices are based on their introduced concept that uses the term score, the ratio between the radiative frequency of a term in the given patent collection, and its relative frequency in a general discourse reference corpus [21]. Although their concept makes use of term frequencies, title score, and claims score [21], the most useful approach seems to be the term score. Thus it relies on a relative score and investigates the entire document or patent corpus. The topic stream visualization is similar to a stacked graph with the included term (topics) in the area-based visual representation. As they hierarchically cluster patents according to their textual similarities, users are able to zoom in into a cluster through a level-slider. Besides the main visual representation, the stream visualization, they provide four other visual representations, such as a "scatterplot" with brushing and linking [21].

2.2 Technology Forecast Analytics

The literature review for "Technology Forecast" was performed by keyword searches to identify literature and then backward and forward reviews of the identified literature according to Webster and Watson [57]. The initial search was performed by combining keywords from different categories. The search was structured as follows: $S_1 =$ (Technology) AND (Trends OR Forecast OR Foresight OR Intelligence) AND (Methods OR Visualization OR Analysis OR Model OR Discovery). After the first search and analysis, the results were used to search for more individual keywords, which often represented identified elements of the categories. $S_2 =$ Visual Analytics, Text Mining, Trend Mining, NLP, Patent Analysis, Tech Mining, Data Mining, Technology Roadmapping, Patent Network, Technology Radar, Trend Analysis, Bibliometrics, TRIZ. The search terms in S_2 were combined with those in S_1. The search terms were applied to six different databases: IEEE Xplore, ACM Digital Library, SpringerLink, ScienceDirect, CiteSeerX, and GoogleScholar. Overall, 100 publications were chosen for further analysis. A second selection procedure was done by reading the papers and investigating the appropriateness of our research. Out of the 100 publications, 42 were chosen for a deeper investigation. In the third step, we further reduced our review according to the following criteria:

- only papers were investigated that were published after 2016
- papers that used existing frameworks with new data were removed if the used framework was older than five years
- If authors had multiple publications on the same topic, the one with the most citations were used

An et al. proposed a novel approach to derive "technology intelligence" from patents, which can be used to forecast technology [2]. The advancement of text mining techniques has enabled the analysis of the descriptive parts of patent documents and therefore extended the scope of patent analysis [9]. Conventional patent analysis approaches are based on keyword analysis and use specific algorithms to specify keyword sets. The relationship of these keyword sets is analyzed with text mining techniques to describe the technological content of the patent. Previous patent analysis used a subject-action-object (SAO) model to analyze the semantic structure of the keywords. This approach is limited in that it increases the complexity too much and ignores non-functional relationships [2]. As an improvement, the authors propose a preposition-based semantic analysis to develop a technology-relation-technology (TRT) network [2]. This approach uses the fact that the number of prepositions in the English language is limited and can be used to identify the functional and non-functional relationship between technological terms. A network is used to structure the data systematically and to easier visualize and quantify the relationship between different analysis units [46]. A keyword-based network analysis relies on unstructured data and can be used to identify vacant technology and map technology evaluation by analyzing the degree and type of relationship between technological keywords [2, 60]. In the first step of creating a TRT network, the relevant patent data, structured and unstructured, needs to be extracted. Afterward, the data is preprocessed, and the noun-preposition phrases in the abstracts are extracted [2]. With the help of NLP libraries used for text parsing, the target technological keywords are used to filter the noun-preposition phrases. In the third step, using the noun-preposition phrases are used to identify "technological keyword—preposition—technological keyword" structures, where the first keyword is extracted from the noun phrase while the second keyword is extracted from the preposition phrase [2]. These structures are clustered into several groups based on the used preposition, which should define the relationship of the keywords. To verify the relationship, the structures are transformed into TRT structures. In the final step, the TRT network is developed. Domain experts choose the target keywords for the network out of all the keywords from the TRT structures [2]. A relationship matrix for each relationship type is constructed, where the frequency of the TRT based on that type is transformed into binary values. The value 1 indicates a relationship between the keywords, 0 indicates no relationship [2]. This information is used as the input for the network. Trend analysis can be conducted on this network by using the year of application and observing the evolution of technological structures based on the change in technological keywords, and their relationship [2]. This allows forecasting changes in technologies. The network can also be used to forecast possibly new technology by comparing TRT networks with similar structures in different areas and comparing the keyword sets. The framework is tested on electronic vehicles in the paper [2].

Li et al. proposed an approach that makes use of Twitter data additionally to doing patent analysis for identifying technology trends [34]. It uses data mining to gather Twitter data for sentiment analysis to detect and classify opinions, and emotional reactions to different technology [34]. The main component is the author-topic over time (ATOT) model, a combination of the topic over time (TOT) and author-topic (AT) model, which are based on the Latent Dirichlet Allocation (LDA) [6, 34]. For the ATOT model, the topics, authors, and time information of the documents are combined to form a three-dimensional analysis model to study the evolution of authors' research interests [34]. The approach consists of three processes, patent analysis, Twitter data mining, and the combination of both to identify technology trends. For the first process, a patent database is chosen as a data source. The patents are then converted to raw text format [34]. These are clustered via the "Lingo algorithm" using semantic hierarchical clustering, a vector space model, and singular value decomposition [34, 54]. In the second process, the Twitter data is mined by using the Twitter Search API [34]. This data is also preprocessed, removing duplicates and cleaning redundant information. The tweets are classified according to time. In the third step of the second process, a word document distribution is created [34]. The preprocessed Twitter data is analyzed with the ATOT model, and a topic-feature words probability distribution is obtained. Nouns, adjectives, and adverbs are extracted for keywords for each year to obtain awareness keywords for the sentiment analysis. The profiles of the Twitter users are analyzed, and the ATOT model and experts identify their profession [34]. It is then analyzed which type of professions focused on which topics over time to identify the changing pattern in their topic of interest. In the third process, the results of both processes, the evolution map and trend analysis of the patent analysis and the sentiment and pattern analysis of the Twitter data mining, are combined in a differences analysis [34]. This is done by first doing a comparative analysis of the results, comparing the gaps, and combining the results of changing patterns in the topic of interest to identify developing trends of emerging technologies. The final comparison analysis is also mapped out over time in a combined evolution map, where the results of the patent analysis are in one half, and the results of the Twitter data are one the other half [34].

Li et al. also proposed an approach that makes use of a patent network to forecast through analogy and social network analysis visualization [33]. Forecasting by analogy is transferring previously identified patterns of change in similar technology onto the technology to be forecasted [3]. It is a form of trend analysis that uses historical data in related fields to project the development. The forecasting in the framework is done in five steps [33]. First, a bibliometric analysis is performed on many sources, including scientific publications, patents, and news from the internet. The analysis aims to identify the development trend of the target technology, including a pattern of change. In the second step, time and trends are normalized based on the maximum number of patterns in the technology life cycle and divided into five-year periods.

This is done to facilitate the comparison of technologies developed in different time periods and to identify the characteristics of each technology development period [33]. Afterward, collaboration networks of patents are created, with the companies as nodes and the co-parenting behavior as edges. Sub-networks are created for each period and compared via two centrality metrics, average degree and density. The analogy forecasting is done by analyzing these metrics and using the visualization of the networks to project the future development of the target technology by extrapolating the historical data of the identified similar technology [33].

Yang et al. proposed an approach for semantic analysis of keywords with the "Subject-Action-Object model" [59]. The forecasting is performed by extrapolating technology information in trend analysis and visualizing the output on a technology roadmap [59]. The technology roadmap is a structured and graphical method to explore and communicate the relationships between evolving and developing markets, products, and technologies over time [49]. They use the Web of Science as the data source. The framework is divided into two parts. In the first part, the semantic information is extracted. In the second part, the technology roadmap is constructed. To extract the semantic information, firstly, the principles of semantic information extraction have to be defined. This includes specifying the relevant technology field. After that, a preprocessing of the data is performed. These are annotated, and the SAO structure of the data is extracted algorithm. To create the semantic-based technology roadmap, the technological factors have to be first defined [59]. They use fuzzy matching functions or using the SAO frequency statistics. After the identification of the technological factors, the relationship of the SAO structures is identified. The relationship can be temporal or correlative. A cross-correlation map is used to show the temporal relation or each year, which indicates the order in which the specific technology occurs and can be used for the trend analysis [59]. A factor map is created with a principal component analysis (PCA) to visualize the correlative relations based on SAO co-occurrence. With the help of experts, the technology roadmap is created and used to identify the trends of technology development [59].

Hurtado et al. proposed an approach that makes use of association analysis for discovering topics from a text and time series analysis to forecast their evolving trend [25]. The approach consists of six steps. In the first step, the text corpus is preprocessed, by removing stop words, verbs, using stemming and lemmatization [25]. The sentences are represented as a vector where each dimension corresponds to a keyword, and its value is a binary indicator relating to keyword occurrence. The second step consists of frequent pattern mining by using association analysis on the matrix. The rules discover a set of antecedent items. Because of redundant words, not each association rule (pattern) is treated as a topic [25]. A refinement process is added so that the final topics are all unique. In the third step, a topic incidence matrix is built by collecting the data from each year using the topics and binary incidence matrix. The generated vectors record the number of times a topic appears in each year's paper. A temporal topic correlation analysis is performed on the new matrix. This generates a data set of correlation (correlation co-efficient) among all topics and measures the strength of correlation between two random variables [25]. The correlation coefficient is used to build a network, where each node denotes a topic, and the edge denotes the degree of correlation. To find strongly correlated topics, a threshold value is set to remove edges with weak correlation. A clique percolation method algorithm is used to find groups of topics with strong correlations connected to form a node group. To forecast topic trends, a time series analysis is used [25]. A forecasting model will use historical time series data of a selected field to predict future evolution. It is possible to select multiple fields, which improves forecasting accuracy by having more historical data and making use of the temporal correlation between the different topics [25].

Zhang et al. used text mining for data extraction and processing, machine learning for topic analysis and expert knowledge, and a technology roadmap for forecasting and visualization of the analyzed topics [62]. It is therefore separated into three sectors. In the first sector, the data gathered and preprocessed. An updated Term Clumping process is used to retrieve core terms, and a Term-Record-Matrix is created [61, 62]. A Term Frequency Inverse Document Frequency (TFIDF) analysis is done on these core terms to help the oncoming clustering process [51, 62]. The second section is the topic analysis. The topic analysis is done in four steps, setting up a training set of labeled data for machine learning to use a data-driven K-Means-based clustering method. At first, a cluster validation model is established, which focuses on Total Precision as the target value. In the second step, the features are selected, and a weighting model is created. This is done manually on a multitude of sets. The weighting is only applied to half of the sets in order to calculate similarity [62]. Afterward, a K-Local Optimum algorithm is used as a K-means model for the clustering. It identifies the record with the highest similarity value and uses it as the Centroid of the cluster [62]. The results of this clustering are topics. The topics are weighed using the results of TFIDF analysis, which is also used as the Y-axis of topics in the forthcoming technology roadmap. The semantic relationship between topics is calculated via similarity measure in the K-Means Optimum model. This is used to indicate an evolution in specific fields [62]. In the third section, expert knowledge is used to help with the forecasting of the quantitative results and used to create the technology roadmap [62]. This is done in multiple round interviews, workshops, or seminars. The technology roadmap uses historical data, so both the historical and the possible future evolution of trends are visible. The framework is then applied to research concerning big data technology [62].

Nguyen et al. proposed an approach that makes use of "Term Frequency and Proportional Document Frequency" (TF * PDF) analysis for detecting hot topics and trends from patents with the help of the "International Patent Classification" (IPC) ranking [45, 58]. Another key component is the usage of the Aging Theory to calculate the variation of trends over time [7, 45]. The approach can be divided into four phases [45]. The first phase is the preparation and data collection, where the IPC is defined and the patents are downloaded. In the second phase, the patents are transformed into structured data. Their keywords are extracted with the help of stop word removal, and an NLP Part-of-Speech Tagger [45, 55]. In the third phase, the frequency of the term and the term variation over time are considered the main characteristics. To measure the frequency, the TF * PDF method is used [45]. To measure the term variation over time, a term life cycle model is used. The birth, growth, decay, and death of each term are measured with the help of three mathematical functions. To get weight, the TF * PDF and term variation over time are combined [45]. The terms in a candidate list will then get ranked with the combined weight. The top-ranked terms are identified as "hot terms" that reflect the hot topics in the corpus. In the last step, a patent timeline is created and divided into yearly time slots [45]. For each year, a trend is represented by the normalized weight of occurrence from a term in N documents. These results are visualized by line graphs, which illustrate the change over time for each trend and compare it to other trends. It also can be seen whether a trend is growing or is decreasing. The framework was applied on patents from 1976 to 2005 to identify the ten hottest very stable trends [45].

Cho and Daim integrated a frequency analysis in a technology diffusion model [8]. The Fisher-pry model [17], and growth curves were used for visualization, and forecasting calculation [8]. The idea is to use historical data for technology trajectory utilizing a mathematical model [5]. The authors propose a six-step approach to forecast technology based on growth patterns. The first step identifies technology trends in the market. This is done manually by creating literature reviews, taxonomies, and market reports of the target technology and market [8]. The next step relies heavily on expert knowledge [8]: an expert panel discussion is used to analyze the market structure and provide potential keywords. Data Mining is used to gather patent data from a multitude of sources. This data is discussed by experts again. The third step is a general gathering of data by using bibliometric data. The keywords used to get data from "Web of Science" were applied for research and patents. In the fourth step, the frequency of publications is calculated [8]. This includes a preprocessing of the data, as unrelated publications are removed. The frequency seems to be calculated without weight, though the authors do not mention the specific calculation method. The frequency is calculated for each used database [8]. To calculate growth curves and identify growth patterns, a mathematical model is needed. To simulate market penetration, technology diffusion models are used [18]. The Fisher-Pry model simulates technology advances as competitive substitutions of one method to satisfy the need for another and is similar to biological system growth [8, 17]. The growth of technology is slow in the beginning, then rapid until an upper limit is reached. After that, the growth slows down again. The upper limit is estimated by using historical analogies [8, 17]. It forecasts the growth rate of the substitution technology based on the technological advantage over the old technology. The data for the mathematical model was gathered in the steps before, and the diffusion rate is calculated. In the last step, the growth patterns of each database are consolidated by identifying the time lag. Growth curves are created to forecast and visualize the technology trend [8].

Considering the Visual Analytics approaches, the most advanced approach is the work of Heimerl and colleagues [23]. It provides more than one view, uses relative scores and co-occurrences, and visualizes the temporal spread of the topics with the related categories. Furthermore, it provides a kind of process of functionalities to support trend analysis and technology management, particularly for patents. They propose a five-step approach derived from the works of Ernst [13], and Joho et al. [26] that starts with (1) obtaining an overview of different technology topics in a given field, (2) identifying relevant trends according to individual information needs, (3) evaluating the importance of technological trends, (4) observing the behavior and productivity of different players relevant to a specific trend, and (5) spotting new technologies related to a trend. Although this work was the only one, by our best of knowledge, that considers human tasks and is propagating a kind of procedure, the system itself does not really support the users in the proposed way. The approach and the according system "*PatStream*" just provides a dashboard of four static visualizations. The interaction capabilities are limited, a real overview is not given, and changing visual representations to gather aspect-oriented visualizations is not provided.

As the literature review revealed, several algorithms exist for gathering trends from text, forecasting technologies, topics and trends, and various approaches to visualize the extracted terms, trends, and forecasting results. From the visualization point of view, the systems are commonly designed to illustrate the trend or frequency of terms (even without the temporal dimension). An analytical visualization system that enables humans to analyze the innovation and technology management process through different data models and selectable and appropriate visual structures could not be found.

3 General Method

Corporate foresight is a dynamic capability describing an organization's ability to anticipate and proactively adapt to future developments [16, 20] and thus an organization's innovative capabilities. A company's proficiency in corporate foresight significantly affects its success [50]. Corporate foresight comprises many activities, but the most common approach is applying methods such as scenario planning, technology roadmapping, or the Delphi approach, creating an understanding of the future for specific objectives. In recent years, foresight methods are commonly adopted and integrated into the organizational strategy formulation. Usually, these methods aim at detecting discontinuities or future projections [16]. A common problem with these approaches is the dependency on expert knowledge and opinions. Consequently, several studies try to apply big data and modern technology to anticipate future trends. Recent research discusses how to support the identification of technological trends, a foundation of corporate foresight, by text mining methods [27, 38].

Only a few studies rely on specific corporate foresight methods that are known in an organizational context. The improvement of existing corporate foresight methods based on automated and data-driven approaches enables easy integration into existing organizational processes and reduces managers' adoption barriers. This can increase the acceptance of surprising results in an organizational context. This opportunity is rarely used, and further research is necessary [27]. Subsequently, we propose a method for technological trend prediction based on an established corporate foresight method—technology roadmapping.

Technology roadmapping is a method for technology planning that aims to align an organization to technology developments strategically. The roadmaps, which have diverse forms and visualizations, can be used to explore and discuss upcoming relational aspects between technologies [48]. The most common visualizations show the technological maturity (e.g., technology, product, and market) on the vertical axis and the time-dependent progress on the horizontal axis [30]. The different technologies can be linked based on improvements or similarities. The visualization can change based on the specific purpose [30, 48]. Consequently, technology trends show a significant impact on these visualizations' characteristics. Internal emergent technologies will impact future products, and external research achievements may threaten existing solutions. Consequently, organizations should foster awareness about these topics and create mechanisms to efficiently map these trends and include them in their technology roadmaps.

In the field of technology road-mapping using visual analytics, only two approaches can be identified. Pepin et al. [47] develop a dynamic topic extraction and create a visualization based on a Sankey diagram. Based on a Twitter data set, they divide the tweets into explicit phases and extract topics for each period. The relational processes between the technologies are calculated based on topic similarity and a specific threshold. In contrast, Kayser et al. [28] calculate several visualizations supporting individual steps in the roadmap creation and miss the objective of a comprehensive method.

In our study, we develop a roadmapping process based on the customization model of Lee, and Parker [30] who describe different types and use cases of roadmaps. They define three phases—classification, standardization, and modularization—to structure existing roadmapping approaches. We classify our visualization methodology for trend identification based on their framework to enhance the adoption willingness in an organizational context. This helps to choose targeted methods in an organizational context for corporate foresight activities. The classification phase defines the roadmap's functional purpose. They differentiate forecasting, planning, and administration as managerial use cases of technology roadmapping. Forecasting is thereby the most plausible activity because the assessment of technologies' future-readiness as roadmapping's main objective relies mainly on predicting technological trends. Second, a technology roadmap is a foundation for strategic planning activities. It reduces uncertainties and creates clear objectives. Third, better communication based on visualization and a feeling for the overall vision support administrative processes. Lee and Parker propose several roadmap types in the standardization phase based on the differentiation of products and technologies. The final modularization phase aims to match the objectives of the initial phase to the standardized visualization methods.

Considering the technological capabilities of automation and possible data sources, we reduce Lee and Parkers' [30] framework and ensure the compatibility of the automation with the managerial strategy alignment through a defined objective and standardized output. The application of automated methods is primarily beneficial if individuals cannot perceive and process a high amount of information, for example, a data set of many text documents. Up-to-date language processes are especially useful for latent topics that are not explicitly prevalent in the data set. Especially product and strategically relevant topics are not ubiquitous in such organizational data sets.

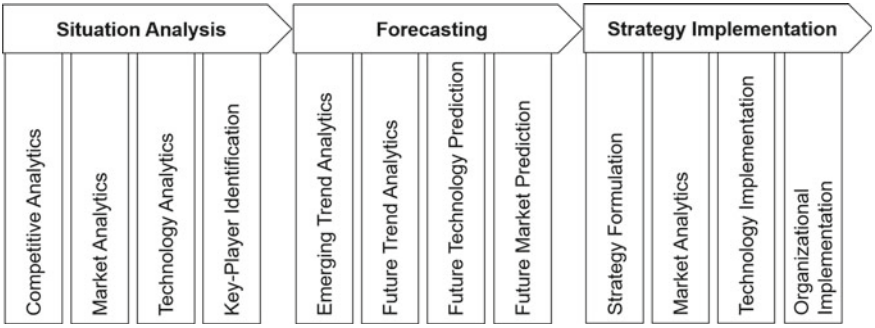

Fig. 1 The general model of innovation management through Visual Analytics

The knowledge about strategic objectives and products is usually defined in specific documents, and there is no need for further specifications based on language processing. Hence, the administrative objective is not adequately addressed with automated processes. Additionally, product knowledge is difficult to separate from incorporated technologies, and product-specific content cannot be visualized. The main objectives for automated technology roadmapping are forecasting and planning, focusing on technology rather than concrete products. Lee and Parker [30] propose a single possible customized roadmap considering the mentioned requirements: The technology trend roadmap visualizes technology trends over time in different technology areas. In this visualization, each technology is assessed for its importance and technological coverage in the organization. The technology trend map can be created fully automated, while the technology portfolio map improves in quality considering the objective data from trend analysis.

To create a technology trend map and use the information for the technology portfolio, we follow the following steps: Situation Analysis, Forecasting, and Strategy Implementation (see Fig. 1).

In the first step, the current situation is analyzed. This includes competitive analytics, market analytics, technology analytics, and key-player analytics. These steps require automated methods that allow the analysis of the market, competitors, key players, and technologies.

In the second step, forecasting of possible future scenarios should be performed. This step includes emerging trend analytics, future trend analytics, future technology prediction, and future market prediction. For this, enhanced predictive analytics methods are used based on historical data to gather a probability of future scenarios. The probabilities allow at least to validate the hypotheses from step one or even gathering information about probable future scenarios.

The last step of the model focuses on the "strategy implementation". The implementation process is based on the gathered information through the first two steps and is performed in a more "organizational" way, which commonly leads to "strategy formulation". Based on the formulation, a comprehensive market analysis can be performed, technology implementation through technology development, the organizational implementation that leads to an ideation process that can be supported through the exploratory character of the Visual Analytics system. The tasks can be combined and are primarily performed by humans. The ideation process as part of "strategy formulation" can be supported through Visual Analytics.

4 Visual Analytics Approach

We introduced in Sect. 3 the general method and will investigate in this section the according to Visual Analytics approach. Our approach includes six main steps as illustrated in an abstract way in Fig. 2 [41]. We describe these steps assuming that scientific data are used for early trend identification, and these have to be crawled through the Internet. Thus, we focus our attention primarily upon enabling users to interactively gather an overall topic trend evolution and different perspectives (e.g., geographical or semantic) on data to inspect and analyze potential technological trends.

In this section, we explain the processing exemplary based on the DBLP database. We chose the DPLB indexing database since it does not provide any abstracts or full-texts and makes the data gathering process more difficult so that the process

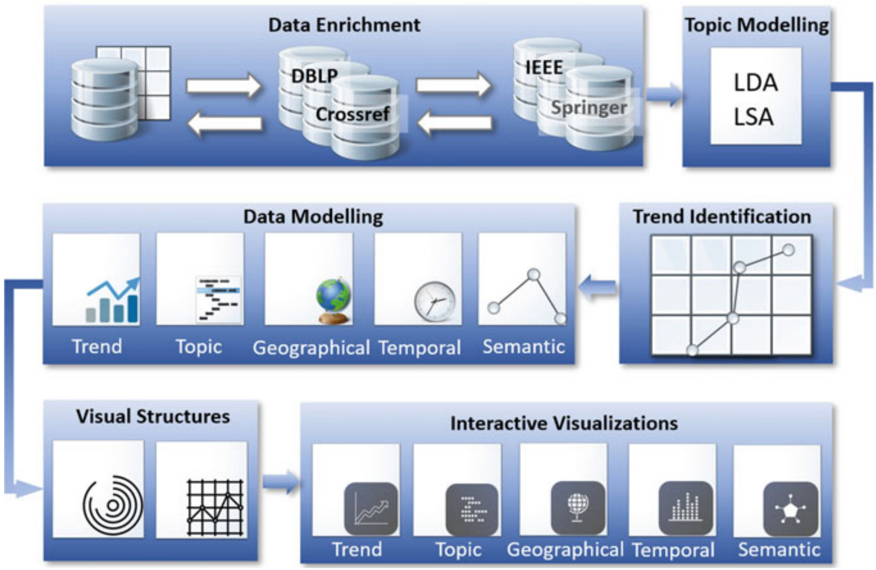

Fig. 2 Our transformation process from raw data to interactive visual representations consisting of six main steps (adapted from [41])

of data gathering can be illustrated too. The DBLP is a research paper index for computer science-related publications.

4.1 Data Enrichment

In the first step of "data enrichment," we use web-based resources as a baseline to gather initial information. Our Visual Analytics system gathers data from different resources, e.g. "DBLP", "Crossref", "Springer", "ACM" or "IEEE". The most complex procedure is gathering data from "DBLP", since it is an indexing database and does not contain any abstracts, full-texts, or further meta-data, e.g., geographical information of the authors. The "DBLP" data are stored first in a database and provides at least "title", "year", and authors' names of all indexed publications. In more and more cases, an "Document Object Identifier" (DOI) can be gathered too that allows a unique identification of certain publications.

For a proper analysis of the given data, enhancements of data quality are necessary. To enhance the quality of data, we gather additional data from the Web. The data collection used as a basis is a combination of multiple different data sets. The individual data sets offer data of varying quality and content. We, therefore, balance out the limitation of the original data basis of "DBLP" by augmenting the available data with additional information for each publication. For this purpose, the system has to figure out where data resources are located on the Web or which online digital library has more information about a particular publication. We integrated the publisher named above. The primary data collection contains a link to the publisher's resource and is used to identify the digital library and the location of additional information. This information can be gathered either through a web service or crawling techniques. The resulting web services response is well structured and commonly contains all required information, while crawling techniques require confirmation of robot policies, and the results have to be normalized. Nevertheless, the data may contain duplicates, missing, or inaccurate data. Therefore, standard data cleansing techniques are applied. With this step, we enrich the data of DBLP with additional metadata, including abstracts and text directly from the publishers, and include some citation information through "CrossRef" that should enable the identification of the most relevant papers in a field with regards to citation count.

4.2 Topic Modelling

In the previous step, we gathered at least abstracts for a major part of the "DBLP" entries and some open access full text for some entries from a general public source like CEUR-WS or the *Springer* database. Based on these enriched data, we are able to perform information extraction from text to generate topics. We conducted a preliminary study with 2.670 full-text articles and their corresponding abstracts and used the Latent Dirichlet Allocation and Latent Semantic Analysis (LSA) [11]. Both algorithms were used with and without lemmatization for full-text articles and abstracts. The best results could be gathered through LDA without lemmatization even though topic extraction from abstracts of the publications [42]. We applied the Latent Dirichlet Allocation for topic extraction according to Blei et al. [6]. The topic extraction through LDA was performed with single words and n-grams that consisted of two or more terms (n-grams). Overall, 500 topics were generated automatically in a data-set of more than seven million documents. For each document, we set a number of 20 for words and 20 for generating n-grams. The according labels of the topic-based trends were generated through the highest scored n-gram, word, or word-combination as the label for a particular topic. A word combination consists of two words, where the score value of each of the words is significantly higher than the scores of the first three n-grams together. The score itself is the value for the distribution of a topic in a finite set of documents. This could be the result-set of a search query or even the entire data set. This kind of labeling has the advantage that reliable and sense-full topic-based trends can be generated with statistical methods. It is not language-dependent (like LDA), and the generation of the labels is fast and easy. This could be enhanced with semantic approaches and linguistic corpora and would provide far better accuracy and sense in terms of semantics, at least for the most prominent languages. A topic-based trend (element) has one more and crucial dimension: time. Temporal information of the topics and their distribution is extracted through the publication date of a document or the dates of a set of documents. With this procedure, we generate labeled and time-stamped topics that can be used to identify trends. Figure 3 illustrates on the left side the measured score of the n-gram "big data". Thereby, the score of the n-gram is higher than the sum of the scores of the second and third words. In such a case, the according n-gram is used as a label since only the first word has a significantly higher score but not the second word. The document set consists here of the search results of the term "Visual Analytics". On

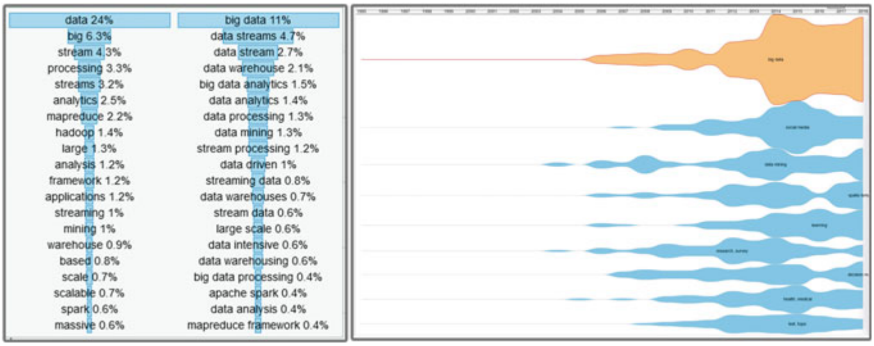

Fig. 3 Topic labeling and topic-based trends: The labeling of topics is performed through the distribution score of words and n-grams in a finite set of documents (left) and the topic-based trends can be gathered through the distribution of the topics over the years in finite set of documents (right)

the right side of Fig. 3 the temporal spread of topics related to the same document-set is visualized according to temporal values and temporal spread. Thereby "big data" is prevalent and has the highest score of all topics related to the finite document set.

4.3 Trend Identification and Forecasting

We extracted topics through LDA, labeled them through a statistical method, and enriched those topics with temporal data. If we tried to identify trends based on the frequency of the topics over the years, we would not get any appropriate trends. Nearly the number of all topics will increase through the years. This is because the number of publications increased in the last years dramatically. Table 1 illustrates the actual number of publications at the time of writing this paper in the DBLP database for every five years.

We worked out in our previous work [41] a five-year periodic model that measures through the slopes of the regression probabilities for the subsequent five periods (25 years) and validated the results with historical data. We started with the normalization of the topic frequencies. We, therefore, calculate for each year the normalized number of documents containing a topic. The normalized topic $\tilde{t}_y$ for a particular year y was calculated by dividing the occurrence of a topic t_y by the number of documents d_y in that year [41]. After having the normalized frequency of documents containing the topic, the entire years with documents with a specific $\tilde{t}$ are split into periods of a fixed length $x > 1$, limiting the length of the period to the time of the first occurrence of the topic, if necessary. So at the current year y_c, each period p_k covers the previous years $[y_c - x \cdot (k+1), y_c - x \cdot k]$ [41]. For each period, we calculate the regression of the normalized topic frequencies and take the gradient (slope) as an indicator for the trend. Equation (1) calculates the slope for a topic t in a period p_k, based on the normalized topic frequencies $\tilde{t}_y$, where $\bar{t}$ is the mean of the normalized topic frequencies and $\bar{y}$ is the mean of years in the time period.

$$b_{\tilde{t},k} = \frac{\sum_{y \in p_k} (y - \bar{y}) \cdot \left(\tilde{t}_y - \bar{t}\right)}{\sum_{y \in p_k} (y - \bar{y})^2} \tag{1}$$

Each calculated slope $b_{\tilde{t},k}$ is weighted through two parameters. The first parameter is the coefficient of determination R_k^2 of the regression. The second parameter is a weight ω_k that is determined with a function that decreases for earlier periods [41].

Table 1 Number of publications in DBLP every five year

	Documents in DBLP					
Years	1995	2000	2005	2010	2015	2020
Documents	13,775	38,908	111,022	201,245	365,426	507,375

This parameter was calculated through a linear function $\omega_k = max\left(0, 1 - \frac{k}{4}\right)$ and through an exponential function $\omega_k = \frac{1}{2^k}$ [41], whereas we found out that the linear function provides more reliable emerging trends.

The final weighting for a topic t is then computed from the slopes $b_{\tilde{t},k}$, the coefficients of determination R_k^2, and the weights ω_k of each of the K periods as follows:

$$\omega = \frac{1}{K} \cdot \sum_{i=1}^{K} b_{\tilde{t},k} w_k R_k^2 \tag{2}$$

Besides identifying emerging trends, we tested different methods to forecast topic-based trends based on historic trend evolution. The historical evolution was tested through various statistical methods, e.g., regression or ARIMA, and various machine learning methods to get the best mean prediction probability (MPP). We integrated "Dense" models, regression models, and a variety of neural networks (NN), including "Graph Neural Networks" [52]. The most appropriate results could be determined through LSTM-based [24] Recurrent Neural Networks (RNN) to forecast long periods. Thereby the "labeled topic" was used as an input variable to determine the forecasting quality. The forecasting is still work in progress. Through different parametrization, the quality could be significantly improved.

4.4 Data Modeling

The analysis process according to our general method requires the identification of various factors, e.g. key-player, competitors, or geographic spread. To meet these requirements, we integrated "aspect-oriented data models" that focus on certain aspects, e.g. temporal or geographical that are given in the data or are gathered through the step of "data enrichment". We generated five data models, "*Semantics Model*", "*Temporal Model*", "*Geographical Model*", "*Topic Model*," and "*Trend Model*" [41]. Thereby, the "semantic data model" [39] serve as the primary data model for storing all information. It adds structure and relation between data elements to generate graphs from the data and visualize relations in the data set. This model is used in visual layouts of semantics structures, textual list presentation, and facet generation. A graph representation of data is generated containing all publications with their attributes and relations to accomplish this.

Multiple temporal visualizations use the temporal data model. Here, multiple aspects of the information in the data collection need to be accessible based on the time property. For the overview of the whole result set in a temporal spread, the temporal model needs to map publication years to the number of publications in a particular year. This temporal analysis is not only necessary for the entirety of the available result set, but it is also vital to analyze specialized parts of faceted aspects. Based on these faceted attributes, detailed temporal spreads need to be part of the temporal model for all attributes of each facet type. The temporal spread analysis needs to be available for each facet in the underlying data. With this information, temporal visualizations can be built more easily. These are then able to show a ranking over time or demonstrate comparisons of popularity over time. The temporal model allows us to measure trends and forecast possible future evolutions of trends through time-series analysis and machine learning approaches. The geographical data model contains geographical information of the available data. The complexity of this model is lower than that of the temporal model, as the geographical visualization only needs quantity information at the country level. This data model provides information about the origin country of the authors of publications. Although the data are enriched with information from different databases as described, there are many data entities without the country information. To face this problem, we introduced two approaches, if no country information could be gathered: (1) we take the affiliation of the authors for gathering the country, and (2) we take publications from the same author from the same discipline based on the extracted topics and the same year plus and minus one to estimate the country. The year of publication is important since many researchers change the affiliation and with the affiliation the country. The topic model contains detailed information about the generated topics as described above. The semantic model contains publications with all assigned properties and relations, including topics, and the topic model supplements this data by offering insights into the assigned topics. Like already mentioned in Sect. 4.2, the information about each topic contains the top 20 most used words and phrases with the assigned probability of usage for each word (see Fig. 3). The inclusion of the most used phrases can help the user immensely in the reformulation of the search query to find additional information on topics of interest. Nevertheless, the primary purpose of the topic model is to gather relevant information about technological developments and the used approaches within a development. The topic model is commonly correlated to the temporal model and also provides the temporal spread of topics. Figure 3-right illustrates the temporal spread of topics related to the search term "Visual Analytics". The trend model is generated through the trend identification process described in Sect. 4.3 in combination with the temporal model. It illustrates the main trends either as an overview of "top trends" identified through the described weight calculation or after a performed query. In the second case, the same procedure is applied with the difference that the document corpus is not the entire database but only the results referring to the queried term.

4.5 Visual Structure

The visual structure enables a fully automatic selection of visual representations based on the underlying data model. We applied the procedure of visual adaptation according to our previews work [44] with the three steps of *semantics*, *visual layout*, and *visual variable*. As proposed in [44], we start the visual transformation for generating a visual structure with the semantics layer. Thus our system is not yet adaptive. We investigate the data characteristics for choosing an appropriate visual layout. Based on the chosen visual layout, we identified visual variables according to Bertin [4] that are applied to a certain visual layout. This procedure allows us to enhance the system with adaptive behavior and reduces the complexity of integrating new visualizations.

5 Visual Analytics for Decision Making

We described in the previous section the overall procedure of gathering data, extracting topics and trends, and modeling data to enable an interactive visual approach for decision making. We will introduce the user interface of our system in this section, including some visualizations that enable the process of decision-making and the interaction design of our visual analytics system.

5.1 User Interface Design

The user interface (UI) of our visual analytics system consists of four areas (see Fig. 4). All areas are dynamic, particularly in terms of data and data entity selection. The top area (1) provides search functionalities, including an advanced search for dedicated search in certain fields and "assisted search". The "assisted search" enhances users' query based on the resulted top five phrases and words of the top-ranked topic [43]. This allows extending the search with similar and related topics to the formulated query. The top area provides, besides search functionalities, the choice of all databases and color schemes for the entire user interface and all visual-

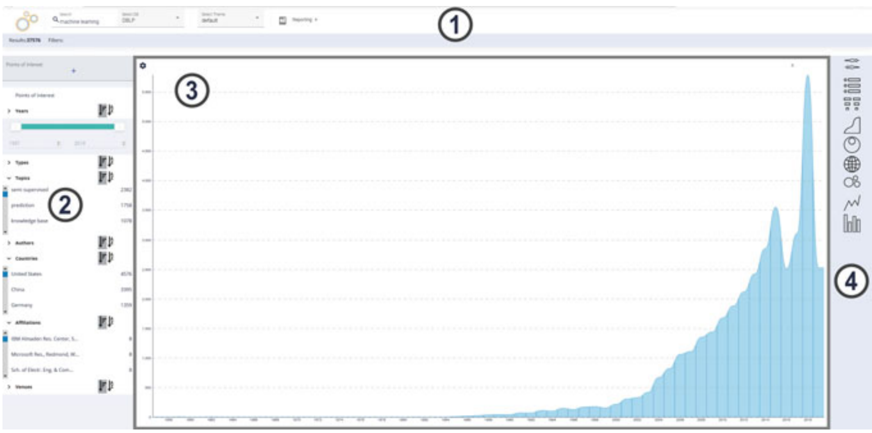

Fig. 4 The UI of our system with its four areas

izations. This is enhanced with a "reporting functionality" to enable the generation of reports for decision-makers who do not interact with the system.

At the left (2), the facets of the underlying data are generated and illustrated automatically. This area also includes the number of results that are automatically adapted to the selected facets, a logical facet selection, and the "graphical search" functionality (see Fig. 12). The logical facet selection allows users to reduce the amount of the results to get the most appropriate documents for a specific task and provide beside a visual interaction an overview of data entities for reducing or enhancing the visualized results.

In the center area (3), the main visualization(s) are placed that are either automatically selected by the type of data, the search query (see Sect. 4.4), or by the user himself in the right area (4), where a dynamic set of visualization are available based on the data and their structure. In Fig. 4 the temporal overview of the entire data is visualized. This area allows the placement of more than one visualization to generate visual dashboards.

The right area (4) provides the functionality to choose either one visualization as illustrated in Fig. 4 or create through drag and drop an arrangement of several visualizations. This area shows icons of visualizations that are supported through the underlying data. For example, if geographic data are provided through country names or longitudes and latitudes, an icon for geographical visualization appears. The according visualizations are related to the data that should be visualized. In the case of a query, these data are the results of the query. If the result-set does not provide any geographical or semantic information, the according visual layout disappears from that area.

5.2 Visual Representations and Visual Interaction

We integrated into our general approach various data models that enable users to interact with different aspects of the underlying data (see Sect. 4.4,). These data models allow us to provide several interactive visual layouts that enable information gathering from different perspectives and support decision-making. We applied two different approaches for interacting with visualizations. Thereby, two complementary approaches were integrated to support users' information acquisition and analysis process. First, the *information-seeking mantra* by Shneiderman [53] is applied to provide an overview followed by zoom and filter and then details on demand. This procedure allows seeing either emerging trends or the most recent search terms typed by other users as a "word cloud" to gather first information of the underlying data and interact until they reached their intended goals. Second, the approach proposed by van Ham and Perer [56] was applied that enables searching the database, getting the context, and get more detailed information. This approach was designed for graph exploration. We think that the entire process of interacting with visualizations can profit from this approach due to its complementary interacting process compared to the overview-first approach. Figure 5 illustrates the start screen of our system

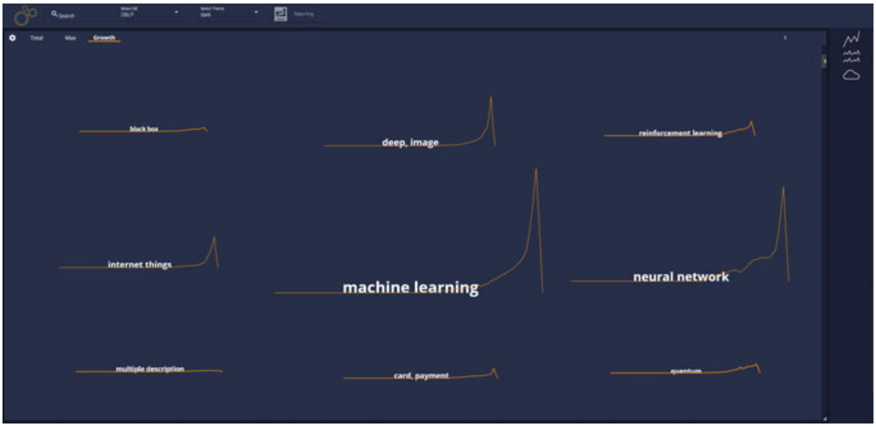

Fig. 5 The initial starting screen of our system applying "SparkClouds" for emerging trends in the entire data-base. Each item is selectable and leads to a search

with an overview on emerging topics as "SparkClouds". Thereby a different color scheme was chosen. The starting screen after a search or choice of the initial screen is illustrated in Fig. 4.

A simple temporal visual layout for overview purposes that visualizes the number of documents over the years of the entire search results is illustrated in Fig. 4. The visualization can be enhanced with several statistical values to allow a more detailed view of the data. These include linear and Loess regression (locally estimated scatterplot smoothing), neighborhood-weighting through color and values, and minimum, average, and maximum values. Figure 6 illustrates such a temporal visualization. Thereby the term "machine learning" was chosen as a search term through the initial start screen, and the mouse hovers the year 2017.

Temporal visual layouts that illustrate the temporal topic distribution of the according documents over the years instead of the numbers of documents use the two data models, the topic model, and the temporal model. Our temporal topic river is such a visualization and separates all the topics and trends for a more comprehensible view. Instead of layering (stacking) the items on top of each other with no space between them, we represent each facet item with a "river". Each river has a center-line and a uniform expansion to each side based on frequency distribution over time. Additionally, placing multiple rivers next to each other makes spotting differences in temporal data sets straightforward. Tasks like comparing the impact of various authors, topics, or trends on a search term become easier. Figure 7 illustrates our topic river for two different data sets of the same database. We arranged the visual layouts on top of each other. The above topic river illustrates the temporal spread of the topics, and the river on the bottom illustrates the temporal spread of publications of those countries that published the most works in the area of machine learning.

For analyzing trends, it is crucial to gather the knowledge of the underlying topics, technologies, etc. emerged during the time or lost their relevance. To enable a fast

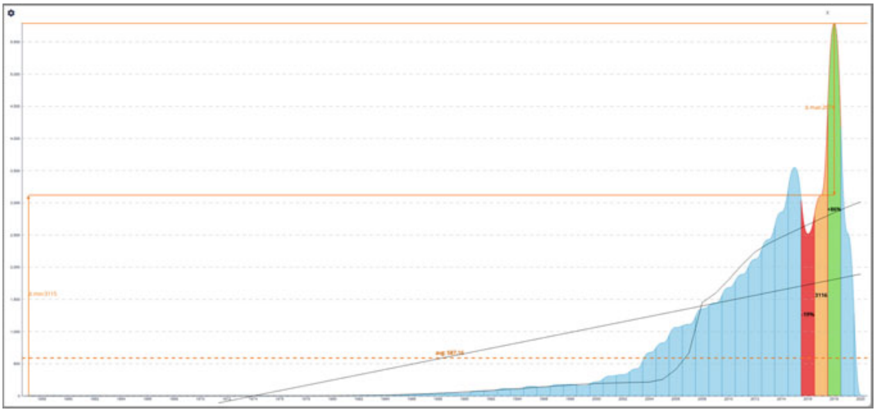

Fig. 6 Statistical values in simple visual layouts to gain more insights. A chosen year is illustrated in orange, increased numbers of publications in green and decreased in red. The illustration contains further statistical values, e.g., the LOESS-Regression, the linear regression etc.

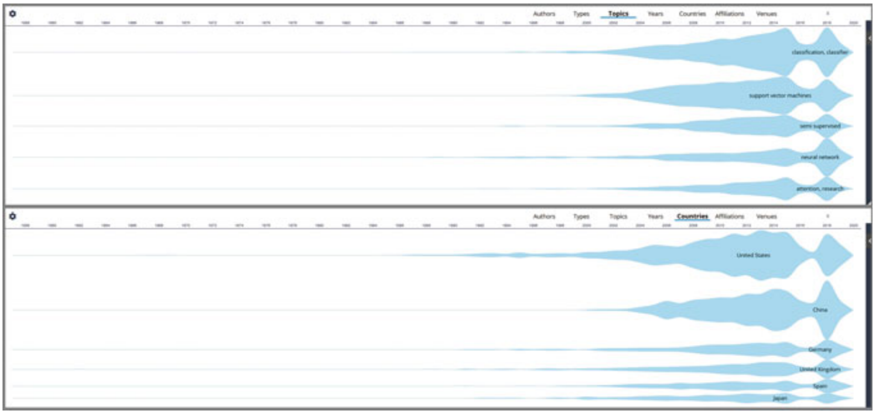

Fig. 7 Stacked river: on top, a topic river illustrates the temporal topic spread and the temporal spread of publications per country on the bottom. Thereby, the number of topics published in a particular year or from a specific country is taken for measuring the spread

and comprehensible analysis, we integrated a temporal ranking (see Fig. 8). Besides the introduced configuration areas, this visual layout offers the ability to specify the number of rows to be visualized. The visual layout is divided horizontally into columns for each year of the analyzed time period. The arrangement is based on the number of publications having a facet item as a property of the selected facet type, sorted in descending order from top to bottom.

The order only represents the ranking. The width of each rectangle represents additional, more concrete information about the relative amount. With these position and form indicators, the user can quickly determine facet items with high influences

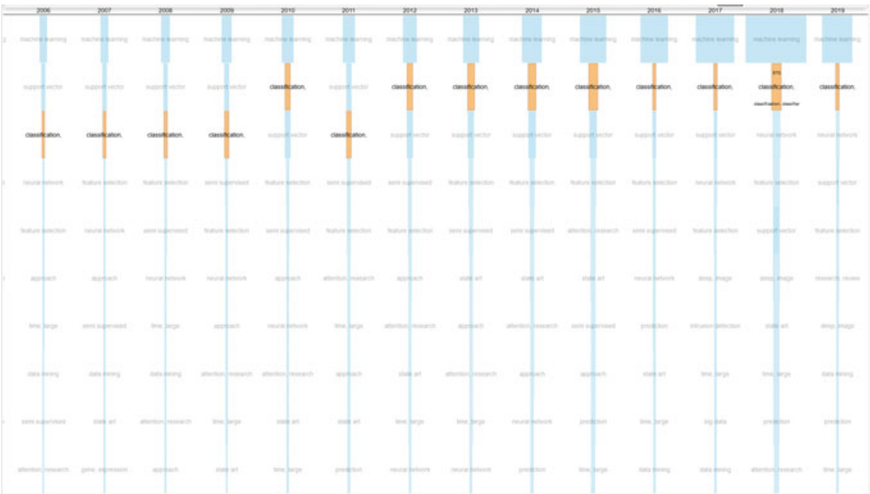

Fig. 8 Temporal spread of topics and the related words and phrases as columns

each year. In Fig. 8 the temporal data model, the topic model, and the semantic model are merged to get the visualized information. Thereby, the top related topics of documents from the result set of the query "machine learning" are visualized, and by selecting one topic, the temporal ranking for each year is highlighted.

For a non-temporal visualization of topic distribution, we have integrated a set of various visualizations. This set should allow to figure out which topics are prevalent in a particular database or a data-set as a result of a search query. The main focus lies on the general occurrence of topics instead of the temporal spread. We reduce the temporal information to enable a more focused view of the topics and their distribution. Figure 9 illustrates a dashboard of two identical topic distributions, left as a pie-chart and right as a bar chart. The simple views allow gathering non-temporal information faster if this is necessary for a certain analytical task.

Our system provides a semantic visual layout that offers the ability to understand relations between authors (co-author information) and between topics (which topic relations) and the semantic correlation between the information entries. Commonly semantic relations are visualized with node-link graphs, leading to complex visualizations and reducing the analysis capability. We integrated beside such node-link visualizations a circle layout that arranges the entities as a spiral starting from the center of the screen. The size of each element indicates the amount of publication per author, whereas the degree option indicates the number of distinct relation targets within the facet type.

The semantic data model provides detailed relational information about individual facet items, which can be accessed through user interaction. After selecting a circle, all relational information within the same facet type is highlighted. This leads to a real-time loading of all co-authors in the result set. Further, users are able to get

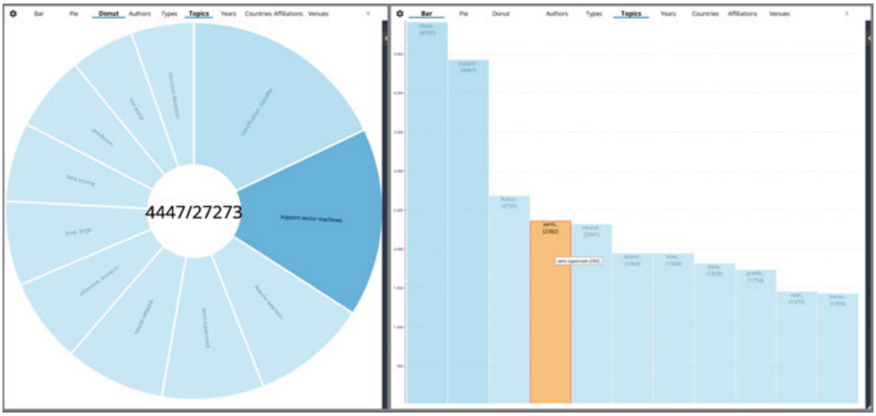

Fig. 9 Examples for non-temporal topic-distribution visualizations

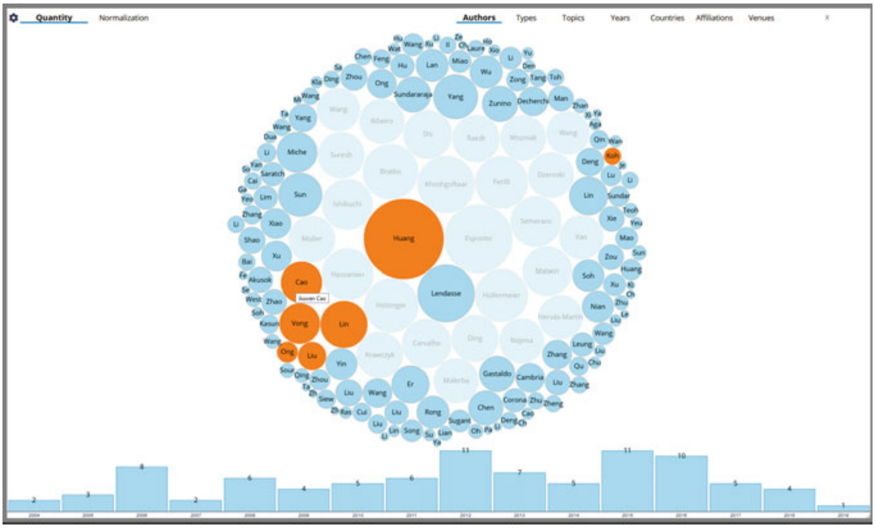

Fig. 10 Co-authorship with relations temporal information

an insight about correlations within the semantic relations through mouse-over that refines the in the figured cased the co-authorship of certain authors through the color. We integrated the disambiguation of authors' names since authors from the same discipline could have the same surname. To distinguish the authors, we compare the first names, the affiliations (with relation to the year), the country, and the co-authorship. Figure 10 illustrates the co-authorship relations based on the search term "machine learning". At first glance, the users can see the most publishing visual layout that visualizes topics related to search terms based on the frequency of their appearance.

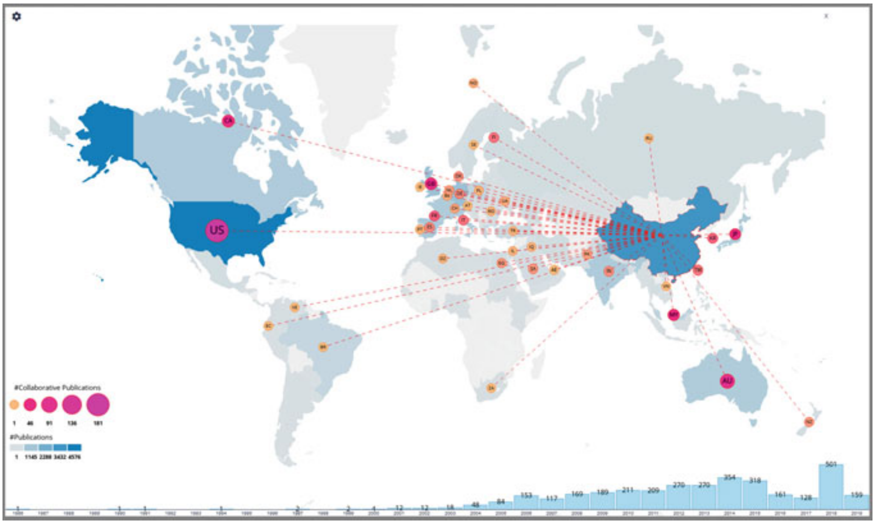

Fig. 11 Geographical search based on the geographical, semantic and temporal data models

Besides temporal, semantic, and topic visualizations, we integrated an advanced geographic visualization that illustrates the number of publications on certain search-term on the country level through saturation. Besides this, the visual interface makes use of the semantic and temporal data models. Figure 11 illustrates the geographical result-set of the search-term “machine learning”. Thereby the user clicked on China to see the temporal spread of publications from China and the relations to other countries. This relation is measured through the co-authors of the publications of Chinese authors and the origin countries of the co-authors. Figure 11 illustrates on the bottom the temporal spread of publications of which at least one author’s origin country is China. The legend of the visualization shows that the United States published the highest number, and China is primarily collaborating with the United States on “machine learning”.

5.3 *Visual Search*

We integrated a “visual search” or graphical search to enable a more advanced search and analysis process. The visual search functionality allows users to formulate terms that are relevant for them, create so-called visual *points-of-interests* (POI) and see at a glance the number of documents that contain the created points of interest. In Fig. 12 the user searched for “machine learning” and created several visual POIs. The defined POIs are visualized on the right side and can be included in the main search term “machine learning” per drag and drop. The color is the indicator for a certain POI and allows users to see how many publications are in the database with

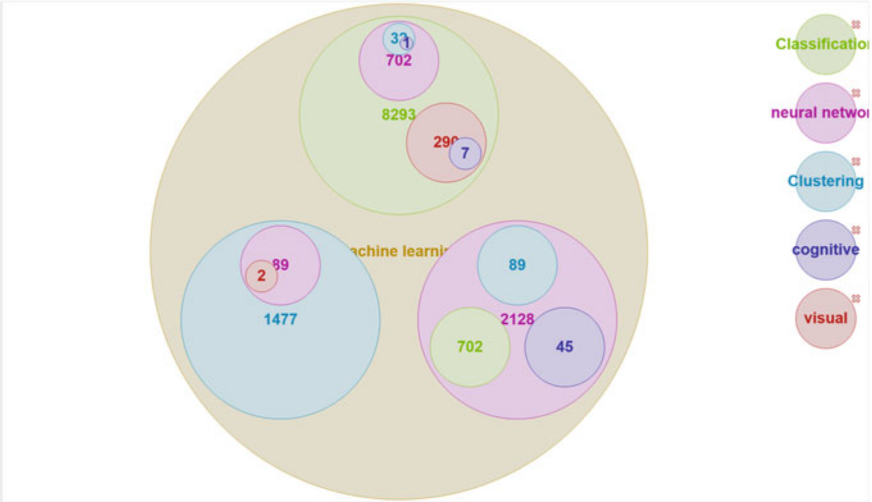

Fig. 12 The visual search for enabling a search within the result set

the created POIs. The number represents the search result quantity within the search result set, so the user is able to define and redefine such POIs for his purposes. In Fig. 12 the main search term is "machine learning". The user is able to see at a glance the results for machine learning documents containing the "classification", "neural network", "clustering", "cognitive" or "visual". With a nested method, they can see that 8293 publications contain the phrase "classification" 702 of these publications include neural networks. Within this search result, 32 publications contain the term "clustering", and one of these publications is related to visual. The users are able just to double click within the circles and get the list of the documents. In the case of machine learning combined with classification, neural networks, clustering, and visual, there is just one publication, so that the list will provide just one publication. This way of interaction allows not only reduces the search result set, but it also gives an overview of the number of related topics, and the number of those relations [40].

5.4 Reporting

We integrated into our system a reporting functionality that enables users to generate reports for non-analysts. This can be performed in two different ways. Each visualization can be saved with the parameters defined by users in our integrated reporting tool. These visualizations are stored permanently to enable the generation of a report. Users might want to integrate the entire dashboard for a certain report. For that reason, they are able to capture the entire dashboard for reporting issues. The reporting functionality aims at creating reports for presentations or reports for

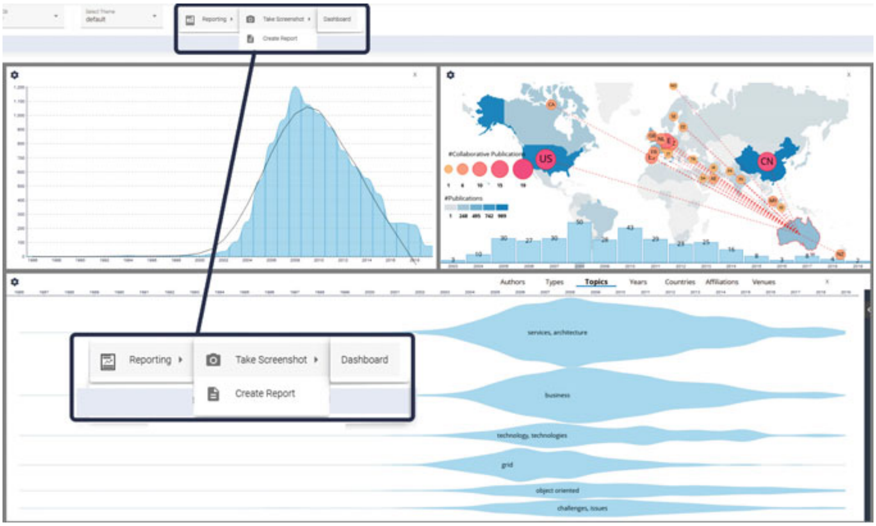

Fig. 13 Capturing the entire dashboard for reporting

decision-makers on the management level. There is no need for decision-makers to interact with the system. Besides this, the reporting functionality creates a “snapshot” at a specific time. Since the data are changing, the reports can be used to validate hypotheses.

For creating a report, first, the figure is stored either through the user interface for the entire screen, regardless if there are one or N visualizations placed on the screen, or through the reporting button on each visualization. Figure 13 illustrates the creation of such a dashboard-capture. The user searched for the term “service oriented architecture” and placed three visualizations on the dashboard. Through the capturing functionality, this dashboard is saved as an image.

After capturing all required visualizations, the users are able to generate a report through the reporting tool. Thereby a web-based HTML editor is provided with predefined text snippets, e.g., for timestamps, search queries, or parameterizations of the visualization. All saved images and text snippets can be found in the reporting tool of our system.

Figure 14 illustrates the reporting tool of our system. On the right side, the HTML editor allows creating the reports and write comments. On the left side, there are the captured figures and text snippets. With just one click, the users are able to generate a report. The dashboard in Fig. 14 illustrates clearly that there is a decreasing trend of using such kinds of web service architectures. Particularly grids were not used in last years. We currently provide exports as PDF, Word documents, and images. We will integrate the raw data used in the reports as JSON and CSV to enable a simple exchange of the data for several systems.

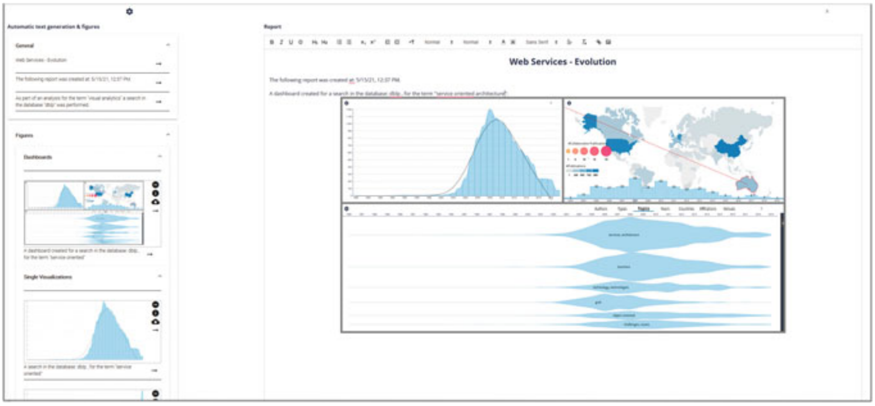

Fig. 14 Generating a report

6 Conclusions

We proposed in this paper a new approach for Visual Analytics for decision making by incorporating some main ideas from innovation management. We first gave an extensive literature review of existing approaches and systems that are mining and visualizing trends. The literature review revealed the missing inclusion of management and decision-making approaches in such analytical systems, particularly for technology management. Our general model filled this gap by a first attempt and combined an appropriate model to enable strategic decision-making. The outcome is a model with three main steps for integrating innovations in firms. This model is enhanced by a more technical approach that illustrated the process of Visual Analytics and the main steps of our approach. Our main contribution is an advanced Visual Analytics approach and system based on our previous work [41] that enables decision making for technology and innovation management with advanced methods in terms of analysis and integrates the ideas of technology management in a Visual Analytics system.

Acknowledgements This work was conducted within the research group on Human-Computer Interaction and Visual Analytics (https://vis.h-da.de).

References

1. Agrawal, R., Srikant, R.: Mining sequential patterns. In: Proceedings of the Eleventh International Conference on Data Engineering (1997)
2. An, J., Kim, K., Mortara, L., Lee, S.: Deriving technology intelligence from patents: preposition-based semantic analysis. J. Inf. **12**(1), 217–236 (2018). https://doi.org/10.1016/j.joi.2018.01.001

3. Armstrong, J.S.: Forecasting by extrapolation: conclusions from 25 years of research. Interfaces **14**(6), 52–66 (1984). https://doi.org/10.1287/inte.14.6.52
4. Bertin, J.: Semiology of graphics. University of Wisconsin Press (1983)
5. Blackman, A.W.: A mathematical model for trend forecasts. Technol. Forecast. Soc. Change **3**, 441–452 (1971). https://doi.org/10.1016/s0040-1625(71)80031-8
6. Blei, D.M., Ng, A.Y., Jordan, M.I.: Latent dirichlet allocation. J. Mach. Learn. Res. **3** (2003). http://www.jmlr.org/papers/v3/blei03a.html
7. Chen, C.C., Chen, Y.T., Sun, Y., Chen, M.C.: Life cycle modeling of news events using aging theory. In: Machine Learning ECML 2003, pp. 47–59. Springer, Berlin (2003)
8. Cho, Y., Daim, T.: OLED TV technology forecasting using technology mining and the fisher-pry diffusion model. Foresight **18**(2), 117–137 (2016). https://doi.org/10.1108/fs-08-2015-0043
9. Choi, S., Kim, H., Yoon, J., Kim, K., Lee, J.Y.: An SAO-based text-mining approach for technology roadmapping using patent information. R&D Manag. **43**(1), 52–74 (2012). https://doi.org/10.1111/j.1467-9310.2012.00702.x
10. Collins, C., Viegas, F., Wattenberg, M.: Parallel tag clouds to explore and analyze faceted text corpora. In: VAST 2009 (2009). https://doi.org/10.1109/VAST.2009.5333443
11. Deerwester, S., Dumais, S.T., Furnas, G.W., Landauer, T.K., Harshman, R.: Indexing by latent semantic analysis. J. Am. Soc. Inf. Sc. **41**(6), 391–407 (1990)
12. Dou, W., Wang, X., Chang, R., Ribarsky, W.: Paralleltopics: a probabilistic approach to exploring document collections. In: VAST 2011 (2011). https://doi.org/10.1109/VAST.2011.6102461
13. Ernst, H.: Patent information for strategic technology management. World Pat. Inf. **25**(3), 233–242 (2003). https://doi.org/10.1016/S0172-2190(03)00077-2
14. Feldman, R., Dagan, I.: Knowledge discovery in textual databases (KDT). In: Proceedings of the First International Conference on Knowledge Discovery and Data Mining (1995)
15. Feldman, R., Aumann, Y., Zilberstein, A., Ben-Yehuda, Y.: Trend graphs: visualizing the evolution of concept relationships in large document collections. In: Zytkow, J.M., Quafafou, M. (eds.) Principles of Data Mining and Knowledge Discovery, pp. 38–46. Springer, Berlin (1998)
16. Fergnani, A.: Corporate foresight: A new frontier for strategy and management. Acad. Manag. Perspect. **0**(0) (2020). https://doi.org/10.5465/amp.2018.0178
17. Fisher, J., Pry, R.: A simple substitution model of technological change. Technol. Forecast. Soc. Change **3**, 75–88 (1971). https://doi.org/10.1016/s0040-1625(71)80005-7
18. Fourt, L.A., Woodlock, J.W.: Early prediction of market success for new grocery products. J. Mark. **25**(2), 31 (1960). https://doi.org/10.2307/1248608
19. Glance, N.S., Hurst, M., Tomokiyo, T.: Blogpulse: automated trend discovery for weblogs. In: In WWW 2004 WS on Weblogging. ACM (2004)
20. Gordon, A.V., Ramic, M., Rohrbeck, R., Spaniol, M.J.: 50 years of corporate and organizational foresight: looking back and going forward. Technol. Forecast. Soc. Change **154**, (2020). https://doi.org/10.1016/j.techfore.2020.119966
21. Han, Q., Heimerl, F., Codina-Filba, J., Lohmann, S., Wanner, L., Ertl, T.: Visual patent trend analysis for informed decision making in technology management. World Pat. Inf. **49**, 34–42 (2017). http://www.sciencedirect.com/science/article/pii/S0172219017300455. https://doi.org/10.1016/j.wpi.2017.04.003
22. Havre, S., et al.: Themeriver: visualizing thematic changes in large document collections. IEEE TVCG **8**(1), 9–20 (2002)
23. Heimerl, F., Han, Q., Koch, S., Ertl, T.: Citerivers: visual analytics of citation patterns. IEEE Trans. Vis. Comput. Graph. **22**(1), 190–199 (2016). https://doi.org/10.1109/TVCG.2015.2467621
24. Hochreiter, S., Schmidhuber, J.: Long short-term memory. Neural Comput. **9**(8), 1735–1780 (1997). https://doi.org/10.1162/neco.1997.9.8.1735
25. Hurtado, J.L., Agarwal, A., Zhu, X.: Topic discovery and future trend forecasting for texts. J. Big Data **3**(1), 7 (2016)

26. Joho, H., Azzopardi, L.A., Vanderbauwhede, W.: A survey of patent users: an analysis of tasks, behavior, search functionality and system requirements. In: Proceedings of the Third Symposium on Information Interaction in Context, IIiX '10, pp. 13–24. ACM, New York, USA (2010). https://doi.org/10.1145/1840784.1840789
27. Kayser, V., Blind, K.: Extending the knowledge base of foresight: the contribution of text mining. Technol. Forecast. Soc. Change **116**, 208–215 (2017). https://doi.org/10.1016/j.techfore.2016.10.017
28. Kayser, V., Goluchowicz, K., Bierwisch, A.: Text mining for technology roadmapping—the strategic value of information. Int. J. Innov. Manag. **18**(03), 1440004 (2014). https://doi.org/10.1142/S1363919614400040
29. Keim, D., Kohlhammer J., Ellis G., Mansmann, F. (eds.): Mastering the Information Age: Solving Problems with Visual Analytics. Goslar, Eurographics Association (2010)
30. Lee, S., Park, Y.: Customization of technology roadmaps according to roadmapping purposes: overall process and detailed modules. Technol. Forecast. Soc. Change **72**(5), 567–583 (2005)
31. Lee, B., Riche, N.H., Karlson, A.K., Carpendale, S.: Sparkclouds: visualizing trends in tag clouds. IEEE TVCG **16** (2010). https://doi.org/10.1109/TVCG.2010.194
32. Lent, B., Agrawal, R., Srikant, R.: Discovering trends in text databases. In: Proceedings of KDD '97 (1997)
33. Li, S., Garces, E., Daim, T.: Technology forecasting by analogy-based on social network analysis: the case of autonomous vehicles. Technol. Forecast. Soc. Change **148**, (2019). https://doi.org/10.1016/j.techfore.2019.119731
34. Li, X., Xie, Q., Jiang, J., Zhou, Y., Huang, L.: Identifying and monitoring the development trends of emerging technologies using patent analysis and twitter data mining: the case of perovskite solar cell technology. Technol. Forecast. Soc. Change **146**, 687–705 (2019). https://doi.org/10.1016/j.techfore.2018.06.004
35. Liu, S., et al.: Tiara: interactive, topic-based visual text summarization and analysis. ACM Trans. Intell. Syst. Technol. **3**(2), 1–28 (2012). https://doi.org/10.1145/2089094.2089101
36. Lohmann, S., Burch, M., Schmauder, H., Weiskopf, D.: Visual analysis of microblog content using time-varying co-occurrence highlighting in tag clouds. In: Proceedings of the International Working Conference on Advanced Visual Interfaces, AVI '12, pp. 753–756. ACM, New York, USA (2012). https://doi.org/10.1145/2254556.2254701
37. Muhlroth, C., Grottke, M.: Artificial intelligence in innovation: how to spot emerging trends and technologies. IEEE Trans. Eng. Manag. 1–18 (2020). https://doi.org/10.1109/tem.2020.2989214
38. Mühlroth, C., Grottke, M.: A systematic literature review of mining weak signals and trends for corporate foresight. J. Bus. Econ. **88**(5), 643–687 (2018). https://doi.org/10.1007/s11573-018-0898-4
39. Nazemi, K., Burkhardt, D., Retz, R., Kuijper, A., Kohlhammer, J.: Adaptive visualization of linked-data. In: Advances in Visual Computing, pp. 872–883. Springer (2014)
40. Nazemi, K., Burkhardt, D.: A visual analytics approach for analyzing technological trends in technology and innovation management. In: Advances in Visual Computing, pp. 283–294. Springer International Publishing (2019)
41. Nazemi, K., Burkhardt, D.: Visual analytics for analyzing technological trends from text. In: 2019 23rd International Conference Information Visualisation (IV). IEEE (2019). https://doi.org/10.1109/iv.2019.00041
42. Nazemi, K., Klepsch, M., Burkhardt, D., Kaupp, L.: Comparison of full-text articles and their-corresponding abstracts for visual trend analytics. In: Proceedings of the 24rd International Conference Information Visualisation. IEEE (2020). (To appear)
43. Nazemi, K., Retz, R., Burkhardt, D., Kuijper, A., Kohlhammer, J., Fellner, D.W.: Visual trend analysis with digital libraries. In: Proceedings of the 15th International Conference on Knowledge Technologies and Data-Driven Business—i-KNOW'15. ACM Press (2015). https://doi.org/10.1145/2809563.2809569
44. Nazemi, K.: Adaptive semantics visualization. In: Studies in Computational Intelligence, vol. 646. Springer International Publishing (2016). http://www.springer.com/de/book/9783319308159. https://doi.org/10.1007/978-3-319-30816-6

45. Nguyen, K.: Hot topic detection and technology trend tracking for patents utilizing term frequency and proportional document frequency and semantic information. In: 2016 International Conference on Big Data and Smart Computing (BigComp), pp. 223–230. IEEE (2016). https://doi.org/10.1109/BIGCOMP.2016.7425917
46. Otte, E., Rousseau, R.: Social network analysis: a powerful strategy, also for the information sciences. J. Inf. Sci. **28**(6), 441–453 (2002). https://doi.org/10.1177/016555150202800601
47. Pépin, L., Kuntz, P., Blanchard, J., Guillet, F., Suignard, P.: Visual analytics for exploring topic long-term evolution and detecting weak signals in company targeted tweets. Comput. Ind. Eng. **112**, 450–458 (2017)
48. Phaal, R., Farrukh, C.J., Probert, D.R.: Technology roadmapping–a planning framework for evolution and revolution. Technol. Forecast. Soc. Change **71**(1), 5–26 (2004)
49. Phaal, R., Farrukh, C.J., Probert, D.R.: Technology roadmapping–a planning framework for evolution and revolution. Technol. Forecast. Soc. Change **71**(1–2), 5–26 (2004). https://doi.org/10.1016/s0040-1625(03)00072-6
50. Rohrbeck, R., Kum, M.E.: Corporate foresight and its impact on firm performance: a longitudinal analysis. Technol. Forecast. Soc. Change **129**, 105–116 (2018). https://doi.org/10.1016/j.techfore.2017.12.013
51. Salton, G., Buckley, C.: Term-weighting approaches in automatic text retrieval. Inf. Process. Manag. **24**(5), 513–523 (1988). https://doi.org/10.1016/0306-4573(88)90021-0
52. Scarselli, F., Gori, M., Tsoi, A.C., Hagenbuchner, M., Monfardini, G.: Computational capabilities of graph neural networks. IEEE Trans. Neural Netw. **20**(1), 81–102 (2009). https://doi.org/10.1109/TNN.2008.2005141
53. Shneiderman, B.: The eyes have it: a task by data type taxonomy for information visualizations. In: VL, pp. 336–343 (1996)
54. Stefanowski, J., Weiss, D.: Carrot2 and language properties in web search results clustering. In: Advances in Web Intelligence, pp. 240–249. Springer, Berlin (2003)
55. Toutanova, K., Manning, C.D.: Enriching the knowledge sources used in a maximum entropy part-of-speech tagger. In: EMNLP '00: Proceedings of the 2000 Joint SIGDAT Conference on Empirical Methods in Natural Language Processing and Very Large Corpora: Held in Conjunction with the 38th Annual Meeting of the Association for Computational Linguistics, EMNLP '00, p. 63-70. Association for Computational Linguistics, USA (2000). https://doi.org/10.3115/1117794.1117802
56. Van Ham, F., Perer, A.: Search, show context, expand on demand: supporting large graph exploration with degree-of-interest. IEEE Trans. Vis. Comput. Graph. **15**, 953–960 (2009)
57. Webster, Watson: Analyzing the past to prepare for the future: writing a literature review. MIS Q. **26**(2), 13–24 (2002)
58. World Intellectual Property Organization: Guide to the International Patent Classification. Electronic Print (2019). (Version 2019)
59. Yang, C., Zhu, D., Zhang, G.: Semantic-based technology trend analysis. In: 2015 10th International Conference on Intelligent Systems and Knowledge Engineering (ISKE). IEEE (2015). https://doi.org/10.1109/iske.2015.43
60. Yoon, B., Park, Y.: A text-mining-based patent network: analytical tool for high-technology trend. J. High Technol. Manag. Res. **15**(1), 37–50 (2004). https://doi.org/10.1016/j.hitech.2003.09.003
61. Zhang, Y., Porter, A.L., Hu, Z., Guo, Y., Newman, N.C.: Term clumping for technical intelligence: a case study on dye-sensitized solar cells. Technol. Forecast. Soc. Change **85**, 26–39 (2014). https://doi.org/10.1016/j.techfore.2013.12.019
62. Zhang, Y., Zhang, G., Chen, H., Porter, A.L., Zhu, D., Lu, J.: Topic analysis and forecasting for science, technology and innovation: methodology with a case study focusing on big data research. Technol. Forecast. Soc. Change **105**, 179–191 (2016). https://doi.org/10.1016/j.techfore.2016.01.015