Visual analytics for technology and innovation management

An interaction approach for strategic decision making

Kawa Nazemi[1] · **Dirk Burkhardt[1]** · **Alexander Kock[2]**

Abstract

The awareness of emerging trends is essential for strategic decision making because techno-logical trends can affect a firm's competitiveness and market position. The rise of artificial intelligence methods allows gathering new insights and may support these decision-making processes. However, it is essential to keep the human in the loop of these complex analyti-cal tasks, which, often lack an appropriate interaction design. Including special interactive designs for technology and innovation management is therefore essential for successfully analyzing emerging trends and using this information for strategic decision making. A combination of information visualization, trend mining and interaction design can support human users to explore, detect, and identify such trends. This paper enhances and extends a previously published first approach for integrating, enriching, mining, analyzing, iden-tifying, and visualizing emerging trends for technology and innovation management. We introduce a novel interaction design by investigating the main ideas from technology and innovation management and enable a more appropriate interaction approach for technology foresight and innovation detection.

✉ Kawa Nazemi
kawa.nazemi@h-da.de

Dirk Burkhardt
dirk.burkhardt@h-da.de

Alexander Kock
kock@tim.tu-darmstadt.de

[1] Human-Computer Interaction and Visual Analytics, Darmstadt University of Applied Sciences, Darmstadt, Germany

[2] Technische Universität Darmstadt, Technology and Innovation Management, Darmstadt, Germany

1 Introduction

Technological developments have an essential impact on strategic decision making in companies [16]. The early awareness of possible emerging technological trends strengthens competitiveness and market position of companies [45]. If innovation-driven companies ignored emerging technological developments. However, they may not tap the full potentials of their own products or technologies. The majority of companies detect early signals of emerging trends by chance rather than through systematic processes [43]. Larger enterprises already consider this essential aspect and create foresight units that try to anticipate future developments and upcoming innovations [46]. But even in larger high-tech companies the foresight process is usually manual, labor-intensive and error-prone. A combination of a humans' intelligence with machine learning methods is rarely implemented in this process, although Visual Analytics provides exactly this combination [23]. Currently, existing information systems apply different information retrieval methods in combination with different algorithms to extract trends from text. However, the related visual representations do not enable the exploration, identification and detection of significant trends for inferring future technological developments. The interactive overview on data, the continuous changes in data and the ability to explore data and gain insights are essential to identify upcoming trends. Another important aspect for supporting the process of analytics in technology and innovation management is the interaction design by incorporating the human in the entire analytical process [45]. It is therefore necessary to provide an appropriate interaction design that covers the main processes and ideas of technology and innovation management.

We start our paper with a literature review investigating the works in the areas of trend mining from text and visualization of trends. These aspects are crucial to understand the background of the introduced analytics system for technology management and foresight. Thereafter, based on our previous work [36], we introduce the transformation process of raw data to interactive visualizations that enables the process of decision making through detection of unknown patterns that is the main foundation of this work. Introducing this approach is crucial to understand the interaction design that is a dedicated approach for gathering innovations and technologies in early stages. Thereafter, we introduce our interaction design approach specific for the application context technology and innovation management. The interaction design approach is our main contribution and is a dedicated analytical interaction design approach for technology and innovation management.

2 Related work

The approach presented in this paper uses trend mining approaches from text, and trend and text visualizations. Therefore, we introduce the related work in trend mining and trend and text visualization and try to correlate technology and innovation aspects to bridge the areas of analytics, visualization and technology management to gather a more appropriate picture of the interaction design in this context.

2.1 Trend mining from text

One of the most important tasks in the early works of text mining was the discovery of trends from text, also known as Knowledge Discovery from Text (KDT) [13]. The problem of identifying trends in these early states was described as the task of discovering deviations from expected models that are constructed from past data. Different approaches toward

trend mining arose that aimed at identifying key topics and discovering their relevance over time. Lent et al. [27] proposed one of the first approaches for discovering trends in text. They defined a trend as a specific phrase's sequence of frequencies. The document corpus was divided in several temporal sets for identifying these kinds of trends. The key phrases were extracted with Sequential Pattern Mining (GSP-Algorithm) [49] after this segmentation, and a history of each phrase was generated as a sequence of their occurrences per time interval. This representation could identify trends by means of shape queries [1]. Users were able to formulate a trend's specific shapes and receive the corresponding phrases that comply with these shapes. The *Trend Graphs* by Feldman et al. [12] provide an overall picture of all major trends and focus on concept relations and their evolution. Feldman et al. define a *trend* as a change of the relations between the terms in the corpus in a specific context, rather than sequences over time, for which they introduce the notion of a *Context Graph*. The vertices of a *Context Graph* correspond to terms found in the documents that are connected with an edge if both terms co-sufficiently occur in the given "context". In respect of this notion, a *Trend Graph* is generated from documents in a certain time interval, and the trends are indicated by means of different edge representations according to the predecessor graph and the type of the identified trend. Their approach incorporates first techniques of information visualizations for representing trends in a temporal manner. It is to distinguish between changing and stable trends as Montes-Gomez et al. [32] propagate. They further introduced analysis methods for identifying the key topics that contribute to a trend between two time intervals in a document collection. A normalized topic vector, which was extracted from the documents of each time interval, is the starting point for the trend discovery. They defined different metrics for discovering key Change and Stability Factors based on this representation and a symmetric similarity measure for comparing the topic distributions. Furthermore, they introduced the concept of Topic Deviations that allows identifying of anomalous instances that do not fit in the standard case. Glance et al. introduced *BlogPulse* as a toolkit for analyzing online collections of time-stamped documents in order to detect trends automatically [14]. The system harvested daily blog articles and transferred the text documents in sets of tokens that include different types of annotations (e.g., part-of-speech-tags, sentence boundaries) that facilitate further processing steps. The system extracted key paragraphs that include the majority of phrases of identified key topics for indicating the current trend in the blogging community, beside the extraction of key phrases and key persons per day. An also included trend search tool visually showed the hits of a search query over time. Although these relative counts could provide a temporal overview of keywords for identifying trends, Glance et al. argue that these counts are not indicative in every domain and propose an approach of automatically categorizing trends for further exploitation [14].

Two methods for discovering evolutionary theme patterns in text are introduced by Mei and Zhai [31]. Their methods base on a set of salient themes, which are extracted from temporal sets of documents using a probabilistic mixture model [55]. They present two different methods for identifying latent trends based on these themes per time interval. *Theme Evolution Graphs* is the first method and represents the change of themes over time in a graph constructed with the themes as vertices and the relations determined using the Kullback-Leibler divergence. Based on two different datasets, they could illustrate that this method is suitable for discovering how themes in one time period influence other themes in later periods and providing a temporal overview. Theme *Live Cycles*, the second method, is based on Hidden Markov Models (HMM) and aims at discovering globally interesting themes. HMM is used as a generative model for determining the strength of trans-collection themes over time. Theme Live Cycles allow users not only to see the trends of strength variation but also provide a method for comparing the relative strengths of different themes over time, as

the evaluation reveals. Mei and Zhai also argue that temporal text mining and trend mining have not been well addressed in existing related work and that further research is needed to develop trend mining systems helping users to navigate through large text collections based on temporal trend mining. In contrast, Viermetz et al. [51] utilized temporal granularity and a density-based clustering algorithm for extracting short- and long-term topics as keyword vectors. Each topic is described as a keyword vector that is transformed to a representative keyword vector by contrasting the foreground model corpus to a background model. The method allows to discover topic trends as the evolution of long-term topic clusters, expressed by the emergence and disappearance of short-term topics by linking short-term to long-term topics.

An approach for discovering technology trends from patent texts [24] was presented by Kim et al. They defined a *technology trend* as several salient technologies sharing the same problem or solution. Their approach is divided into the two steps of (1) semantic key-phrase extraction and (2) technological trend discovery. In the first step, key-phrases in each patent are identified that are classified either as a solution or a problem using a learned Support Vector Machine (SVM). In the second step, time spans and their most salient trends are identified for discovering technology trends. However, their approach is primarily applicable in the specific domain of patent texts.

An approach that discovers emerging trends in text collections by identifying significant phrases associated with user-defined known products, companies or people of interest is presented by Goorha and Ungar [15]. The system extracts key-phrases found in the context of user-specified keywords and ranks the interestingness of these key-phrases with an empirically verified formula. A scatterplot that represents the significance of emerged trends over a specific time period visualizes the results.

An approach that allows the identification of topic-related trends is *Tiara*. Tiara utilizes the *Latent Dirichlet Allocation* (LDA) for generating a topic model [28]. The topics' strength is calculated over time that is visualized in a stacked graph for identifying peaks and slopes of each topic. A two-step approach for detecting hot topic and technology trend tracking in patent data is proposed by Nguyen et al. [41]. As a first step, the terms are extracted through TF*PDF [6] and, in a second step, the variation of the extracted terms are measured over time. The works of Chen et al. [8], who proposed an aging theory with the four cycles of birth, growth, decay, and death, builds the fundament of the second step. The proposed calculations of Chen et al. [8] was adapted by Nguyen et al. [41] to determine the Energy of a topic, which is defined by a term's frequency in a specific time slot and indicates if the term is hot followed by the energy function that converts a term energy value into a life support value.

An approach for topic discovery and future trend forecasting in text documents using sentence level pattern mining is proposed by Hurtado et al. [20]. They introduce a finely granular process for converting sentences into a transaction format and association rule mining to discover frequent patterns in the document for topics discovery. They use association rules for correlating topics to each other and Pearson's correlation to correlate the topics with the temporal dimension. The foundation of their forecasting is linear regression with the assumption that all topics correlate. The results are displayed as nodes and edges, whereas shaded regions indicate communities with strongly correlated topics. Their visual representation is not suitable to gather the trend of the extracted topics. Thus, the temporal dimension is not visualized.

Mullroth and Grottke [34] proposed an approach that investigates the entire transformation process through natural language process, topic modeling, emergence detection and visualization. Their model contains query generation, data collection, data pre-processing,

topic modeling, topic analysis and visualization, whereas only the query generation is assigned to be performed partially by humans [34]. The proposed process is similar to our early work [40] that investigates (without natural language processing) the same transformation process that was refined and enhanced with the emergence of trends [36]. Although one main outcome of their approach is visualization, a real analytical interaction design for visual representations is not introduced. Even their systematic literature review [33] neither investigates our previous work nor any kind of analytical interaction with such complex systems.

The works of Mullroth and Grottke [33, 34] clearly illustrates that there is a gap between the different disciplines. While they focus on works from the area of economics and management, these issues and process models were proposed much earlier. An investigation of interaction approaches for analytical reasoning was not performed at all.

2.2 Trend and text visualization

For discovering trends, current trend mining methods provide useful indications. Nevertheless, human's knowledge acquisition abilities are still required in the interpretation and conclusion for serious decision making. Hence, the representation of trends is one of the most important aspects for analyzing trends. Basic visualization techniques are often included in common approaches. Line graphs, bar charts, word clouds, frequency tables, sparklines or histograms convey different aspects of trends, depending on the concrete results. Thematic variations over time in a stacked graph visualization with a temporal horizontal axis are represented with *ThemeRiver* [18]. The strength of a specific topic over time is indicated by the variation of the stream width. *Tiara* uses a similar approach, with the difference that it includes additional features such as magic lenses and an integrated graph visualization [28]. *ParallelTopics* includes a stacked graph for visualizing topic distribution over time [10]. The system 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, although the system was not designed for discovering trends but rather for analyzing large text corpora. *Parallel Tag Clouds* (PTC) is based on multiple word clouds that represent the contents of different facets in the document collection [9]. To identify the difference of certain keywords over time and to infer the dynamics of themes in a text collection, temporal facets can be used. *SparkClouds* are another extension of word clouds that includes a sparkline for each word [26]. Each term's temporal distribution and allow conclusions about the topic trends is indicated by these sparklines. A user study reveals that participants are more effective with *SparkClouds* compared to three other visualization techniques in tasks related with trend discovery [26]. Co-occurrence highlighting was also included in a similar approach [29]. This technique includes a histogram for representing the temporal relevance of each tag, which was a contrast to *SparkClouds*. To enable a more comprehensive analysis of trend indicators, additional overlays in the histograms show the co-occurrences over time for a selected word. *PatStream*, introduced by Han et al., is a visual trend analysis system for technology management [17]. Based on the cosine metrics and extends of the work of Heimerl et al. [19] in particular in regards of visualization, their system measures similarity between pairs of patents. *Streamgraph* was already proposed in the previous works of Heimerl et al. [19] and visualizes the evolution and structure of topics that indicates the trends. In contrast to this previous work, *Patstream* breaks down the streams into vertical time slices, which represent periods of years. 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 [17], is used in their introduced concept that

build the base of these time slices. The most useful approach seems to be the term score, although their concept makes use of term frequencies, title score and claims score [17], thus it relies on a relative score and investigates the entire document or patent corpus. A stacked-graph with included term (topics) in the area-based visual representation is similar to the topic stream visualization. Users are able to zoom in into a cluster through a level-slider as they hierarchically cluster patents according to their textual similarities. Beside the stream visualization, the main visual increase, they provide four further visual representations, such as a scatterplot with brushing and linking [17].

PatStream is the most advanced interactive visual representation. It provides more than one view, makes use of relative scores and co-occurrences and visualizes the temporal spread of the topics with the related categories. However, the approach does not really visualize emerging trends, provide an overview of upcoming trends in a certain field, or support an in-depth analysis through users' interaction.

A number of algorithms for gathering trends from text and different approaches to visualize the extracted terms and trends were revealed in the literature review in both fields. We could outline that the most common indicator is the terms' frequency to define a trend as hot or emerging based on the reviewed existing approaches and there used algorithms and approaches. An analytical visualization system that enables the human to analyze trends through different data models and visual structures could not be found. It can be further summarized that although the introduced approaches and systems aims at supporting the main ideas of technology and innovation management, but the interaction design does not provide a deep analysis for management of technology or innovation management, such as an analysis on extracted topics and technologies together with highly interactive analytical visualizations.

3 Process model of visual trend analytics

3.1 General process

The goal is the realization of an exploratory visualization approach that enables the identification of trends and its probable future potentials in a graphical manner. The main feature is the inclusion of users' interaction in visual structures, which enables the ability to view the data from different perspectives. The user should be enabled to gather an overall trend evolution and different perspectives (e.g., geographical or temporal). Thus, the focus lies on answering the questions [40] toward the analysis of potential technological trends: "(1) *when* have technologies or topics emerged and when were they established, (2) *who* are the key players, (3) *where* are the key players and key locations, (4) *what* are the core topics (5) *how* will the technologies or topics probably evolve, and (6) *which* technologies or topics are relevant for a certain enterprise or application area?" [36, 40] The question where basically introduced by Marchionini [30] toward exploratory searches. However, we expanded the question space and adopted them to the specific characteristics of technology and innovation management that includes early trend detection. The analysis based on the questions provides an overview of core topics that are currently relevant, next to the navigation through the different perspectives, a result analysis and probable reasons about evolving trends. Our approach was developed based on these requirements with the following steps (see Fig. 1).
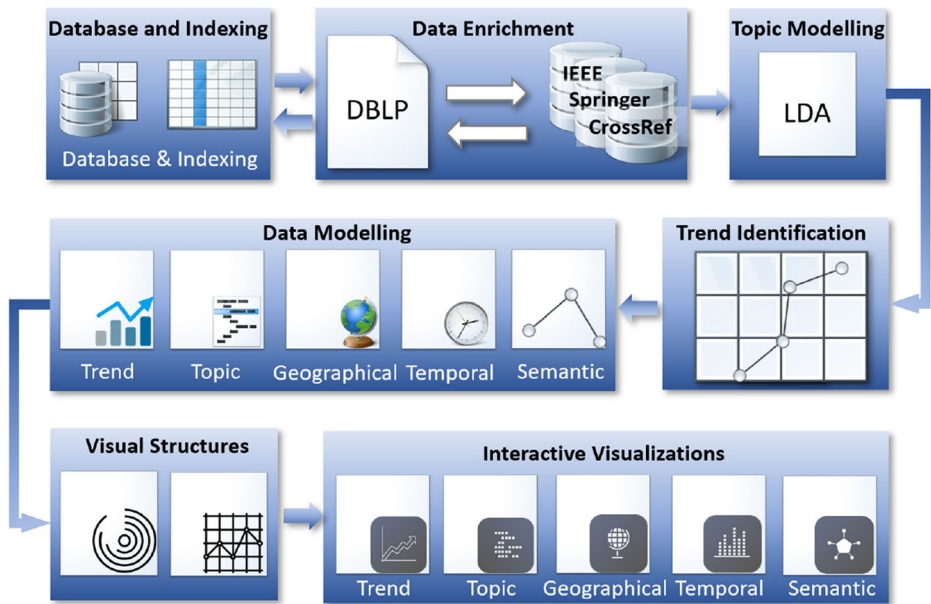
Fig. 1 Our transformation process from raw data to interactive visual representations consisting of seven main steps (adapted from [36]). The transformation process is described in the following section more detailed

In this section, we explain the processing exemplary based on the DBLP database. We chose the BPLB indexing database, since it does not provide any abstracts or full-texts and makes the data gathering process more difficult. The DBLP is a research paper index for computer science related research. Commonly, patents or web news are considered to identify technological trends, which both have the downside that they often indicate new technologies when they are already market-ready. In particular patents are highly established for this purpose, but it is to consider that patent registrations usually take between one and two years, which counts also for identified trends based on patent data. And web news usually appear when a solution is already market-ready. In contrast to these sources, research publications introduce new technologies usually in an early prototype stage, so that at its detection there is still enough time to react to those developments.

3.2 Data indexing

We use an indexing database on the server side. The initial data source for the transformation process is DBLP, which provides rudimentary metadata in the area of computer science and related areas. Each document can have a unique id, the *Document Object Identifier* (DOI), through these identifiers, the data-entities can be identified and enriched from several additional data sources [36, 40]. The initial data is stored in relational database. For the identification of trends, especially emerging ones, we use enriched data (abstracts and full-text articles) to extract and model topics. The enriched data and the identified trends data model are stored in the database after the extraction process. According to Card et al. [7],

the data models build the foundation for choosing visual structures and enable in the last step to either choose appropriate interactive visualization or a juxtaposed visual dashboard for the initially mentioned tasks and questions [36].

For this, each DOI is sent to all publishers, e.g. "ACM DL", "IEEE XPlore", "Springer" etc. If a certain document has no DOI, the title of the document in combination with authors' names is used to identify the document on Web.

3.3 Data enrichment

A proper analysis requires sufficient data quality. As first step of transformation, we use *Data Enrichment* techniques to gather additional data from other web sources to enhance data quality. The data collection used as basis is a combination of multiple datasets. The different considered individual datasets offer a varying quality and varying number of available meta-information. As initial dataset we use DBLP in our approach with about six million entries [36]. The DBLP dataset entries are without text (e.g., abstracts or full-text articles). Since topic modeling is only possible with appropriate text documents, we compensate the limitation of the original DBLP dataset by augmenting each publication entry with additional information [36].

To enrich the data, the system figures out, where the particular data resources are located on the web, and where further information about a certain publication is available. For this, each DOI is sent to all publishers, for example, "ACM DL", "IEEE Xplore", "Springer", or "CrossRef". If a certain document has no DOI, the title of the document in combination with authors' names is used to identify the document on Web. The information for the publication enrichment can be either gathered through a web-service or through crawling techniques. The response of results of a web-service is well-structured and contains commonly all required information, while in contrast crawling techniques require a conformation of robot policies and the results have to be normalized. However, it should be considered that the retrieved data may contain duplicates, missing or faulty data. Hence, common data cleansing techniques are applied [36]. As a result, with this step we further enrich the data of DBLP with metadata, such as abstracts and full-text articles from the publisher and citation information through *CrossRef* that allows to identify the most relevant papers in a field with regards to citation count. A detailed description of the data enrichment would go beyond the scope of this paper. The process is described in a replicable way in [39].

3.4 Topic modelling

We gathered abstracts for the majority of "DBLP" entries and some open access full text for some entries from general public sources such as "CEUR-WS" or the "Springer" database in the previous processing stage. We are now able to perform information extraction from text to generate topics based on the previously enriched data. As topic classification, learned probabilistic topic models are applied for topic generation. This approach is a viable alternative and can even outperform subject heading systems when evaluating similarities between documents clustered by both systems [42], as studies have shown. Consequently, we implemented the Latent Dirichlet Allocation (LDA) algorithm [3]. The main advantage of LDA is its fully automatic topic classification and assignment capability. The classification is performed consistently on all publications, based on classifiers that controls the assignment of all the topics. The accuracy of the resulting model strongly depends on the number of documents. After the processing, each document is assigned to one or multiple topics. A topic

is typically represented by the top 20 used words withing the topic. We also generate most of the used phrases for each topic in form of N-Grams (similar to [54]) additional to the uni-grams. In the current setting [52] we generate 500 topics with 20 words and 20 phrases through 4000 iteration of the LDA-algorithm. During the analysis, the topics can be used to filter search results in the shape of facets. Further, they are used to create the *Topic Model*. We have evaluated the Latent Semantics Indexing (LSI) and LDA with and without lemmatization for abstracts and full-texts and found that in the majority of cases the generation of topics with LDA and without lemmatization provides a good coherency [39].

3.5 Trend identification

The topic modeling uses the Latent Dirichlet Allocation (LDA) proposed by Blei et al. [3] and is the foundation for identifying trends in the analysis process. But if we tried to identify trends based on the topics' frequency over years, the retrieved trends would not be appropriate. The nearly increasing numbers of all topics through the years would let any topic look like a trending topic. Furthermore, the number of publications have increased dramatically in the last years. This is illustrated in Table 1, which shows the real number of publications for every four years.

So, the normalization of topic frequencies is the first step to get the real trends over time. Hence, we calculate the normalized number of documents containing a topic for each year. Let d_y be the total number of documents in a year y, and t_y is the number of documents in year y that contain a certain topic t [36]. Then, $\tilde{t}_y$ is the normalized topic frequency in the given year y, and is computed as

$$\tilde{t}_y = \frac{t_y}{d_y} \tag{1}$$

After having the normalized frequency of documents containing the topic, the entire years with documents with a certain $\tilde{t}$ are split into periods of a fixed length $x > 1$, limiting the length of the period to the time of the topic's first occurrence, if necessary. So at the current year y_c, each period p_k covers the previous years $[y_c - x \cdot (k + 1), y_c - x \cdot k]$. For example, in the year 2019, for $x = 5$, we have the periods $p_0 = [2015, \ldots, 2019]$, $p_1 = [2010, \ldots, 2014]$, $p_2 = [2005, \ldots, 2009]$, up to the period where the topic appeared for the first time [36].

For each period, we calculate the regression of the normalized topic frequencies, and take the gradient (slope) as indicator for the trend. The following (2) 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 [36].

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

Table 1 Number of publications in DBLP [36]

	Documents in DBLP						
Years	1998	2002	2006	2010	2014	2018	2019
Documents	58.875	89.547	165.596	218.991	275.790	308.448	387.853

Each calculated slope $b_{\tilde{t},k}$ is weighted through two parameters. The first parameter is the regression's coefficient of determination R_k^2. The second parameter is a weight ω_k that is determined with a function that decreases for earlier periods.

For example, the weight ω_k that is used for one period can be defined using a linearly decreasing function:

$$\omega_k = max(0, 1 - \frac{k}{4}) \tag{3}$$

This means that the weight is 1 for the most recent period p_0, decreases linearly about 0.25 for each earlier period, and becomes 0 for period p_4 and beyond.

Alternatively, the weight can decrease exponentially:

$$\omega_k = \frac{1}{2^k} \tag{4}$$

In this case, the weight is 1 for the most recent period p_0, then 0.5 for period p_1, and 0.25 for period p_2.

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:

$$\omega = \frac{1}{K} \cdot \sum_{i=1}^{K} b_{\tilde{t},k} w_k R_k^2 \tag{5}$$

To identify the best measurement, we integrated the linear and the exponential measurements for the weight ω_k and evaluated those through two different systems. The more appropriate results seems to be achieved with the linear calculation, due to the overall 20 years fixed time periods.

The weighting of the trends, the slopes in different time periods and the regression allow us to identify trends with better results compared to trend identification methods described and illustrated in the literature review, although the method is quite simple [36].

3.6 Data modelling

The creation of data models in our approach in the *Data Modeling* stage is realized according to Card et al. [7] for different aspects of the data that are relevant in the analysis process. The interaction with our system should lead to answer the questions mentioned in Section 3.1. To answer those questions with particular given aspects in the data, we considered aspect-oriented data models [36]. With five data models, the *Semantics Model*, *Temporal Model*, *Geographical Model*, *Topic Model* and *Trend Model*, we enable a refined data structuring [36]. The *Enriched Data* and the *Trend Identification* build the basis for the creation of these models. An exposed position has the generation of the semantic data model [38], which serves as the primary data model to hold all information. For an easier extraction of needed information in order to create the visual representations, structure and semantics is added to the data. Particularly the textual list presentation makes mostly use of the semantics model, where all available information about every publication needs to be shown, beside the generation of facet information for filtering purpose.

Several temporal visualizations make use of the temporal data model. Here several aspects of the information in the data collection must be accessible based on the temporal property. The temporal model must map the publication years to the set of publications in a given year, to create an overview of the entire result set in a temporal spread. This temporal analysis is not only necessary for the entire available result set. Furthermore, it is also necessary to analyze specialized parts of faceted aspects. Based on these faceted attributes,

detailed temporal spreads for all attributes of each facet type must be part of the temporal model. The temporal spread analysis must be available for each facet in the underlying data. The temporal visualizations can be created more easily with this information. These can show a ranking over time or show comparisons of popularity over time [36].

The geographical aspect of the available data is represented in the geographical data model. The complexity of the geographic data model is lower than the temporal model, because the geographic visualization only needs quantity information for each country. The data in this model provides the information about the origin country of the authors' affiliations. Also the data is enriched with information from various additional databases, many data entities lacks of the information about the country. To face this problem, we integrated two approaches: (1) we use the country of the author's affiliation and (2) we consider the publications from the same author and field based on the extracted topics and within a certain time range (plus and minus one year) to assume the country. Due the fact that authors can change affiliation and thus country, the year of publication is important to respect [36].

The topic model contains detailed information about the generated probabilistic topic model. The topic model supplements other data models, but in particular the semantics data model, by providing insights into the assigned topics.

The inclusion of the most frequently used phrases can help the user immensely when reformulating the search query to find additional information on interested topics. However, the main purpose of the topic model is gathering relevant information about technological developments and the approaches used within those implementations. To provide the temporal spread of topics, the topic model is commonly correlated to the temporal model. Figure 9 demonstrates the temporal spread of topics toward the exemplary search "Information Visualization". The trend model is generated as combination of the trend recognition process described in Section 3.5 and the temporal model. The combination enables to illustrate the main trends either as an overview of the "top trends" identified by the described weight calculation or after a query has been performed. The same procedure is applied, but with the difference that only with the results that relate to a queried term instead of the entire database, as second case.

3.7 Interactive visualizations

The generation of interactive visualizations has two main processing stages: (1) Visual Structure, and (2) Visual Representations, which will be described more detailed in this section.

3.7.1 Visual structure

The "visual structure" of our approach enables an automatic generation and selection of visual representations based on the underlying data model. We applied the three steps model by Nazemi [35, p. 256], consisting of *semantics*, *visual layout* and *visual variable*, which was originally created for the procedure of visual adaptive applications. The model starts the visual transformation to generate the "visual structure" with the semantics layer. Thus, our system is not yet adaptive, we investigated the data characteristics for choosing appropriate "visual layouts". Afterward, we defined a number of "visual variables" according to Bertin [2] that are applied to a certain "visual layout". The inclusion of this model allows us to enhance the system with an adaptation capabilities and reduces the complexity of integrating

new visualizations. Thereby the system performs adaptations more likely as recommendation that the user can neglect at any time. The users can therewith decide to compose their own analysis data view or to consider the adaptation recommendations from the system.

3.7.2 Visual representations

We integrated several "data models" (as described in Section 3.6) that allow the users to interact with different aspects of the underlying data. We provide several interactive visual layouts based on visual structures to enable information gathering from different perspectives, which create the integrated data models. For the analysis processes in technology and innovation management, we identified overall five different visual representations as necessary. Most of the integrated visual representations are based on temporal data, since they are most common in visual data analysis. These temporal data allow not only to visualize the temporal spread of certain data entities over time but also provide forecasting and foresight based on statistical and learning methods. A simple temporal spread of a certain search term is illustrated in Fig. 2. Thereby the right visual representation includes some statistical values, for example, regression line, maximum and minimum.

A main aspect of the temporal visual representation is to get an insight of the temporal spread of certain automatically extracted topics. The temporal spread of the highest weighted topics over time is exemplary illustrated in Fig. 3. It shows pretty well that "neural networks" gained more attention in "artificial intelligence" research - the analysis was initiated via the search for "machine learning".

Another well-established visual layout for temporal data are "stacked charts". However, one should consider that the visual perception of the underlying information might become difficult, if more information entities are illustrated. Stacked visualization makes it sometimes difficult to identify differences between multiple datasets or changes withing the same dataset over time. As a consequence we integrated the temporal river layout, which in contrast to stacked layouts separates all the topics and trends for a more comprehensible view.

Figure 4 illustrates two visualizations of the same data, on the left a river chart and on the right a stacked chart. Each river has a center line and a uniform expansion to each side based on frequency distribution over time. The placement of multiple rivers beside each other makes it easier to spot differences in the temporal datasets and to compare the impact of various authors, topics, or trends on a search-term.

For the analysis of trends, it is important to gather the knowledge of the underlying topics (e.g., technologies) that emerged or possibly lost relevance over time. We have integrated a number of temporal visual structures, which can be combined into analysis dashboards
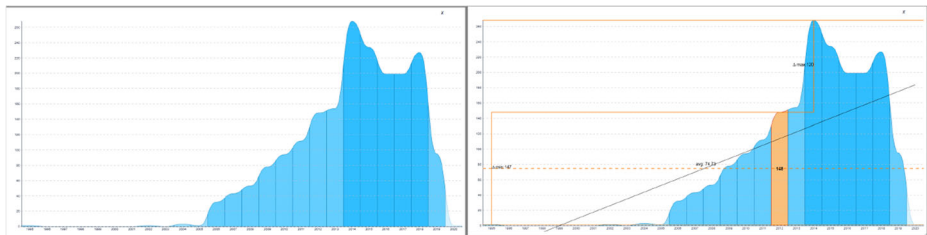

Fig. 2 Temporal visualization: left without any statistical values and right with a regression line, minimum and maximum
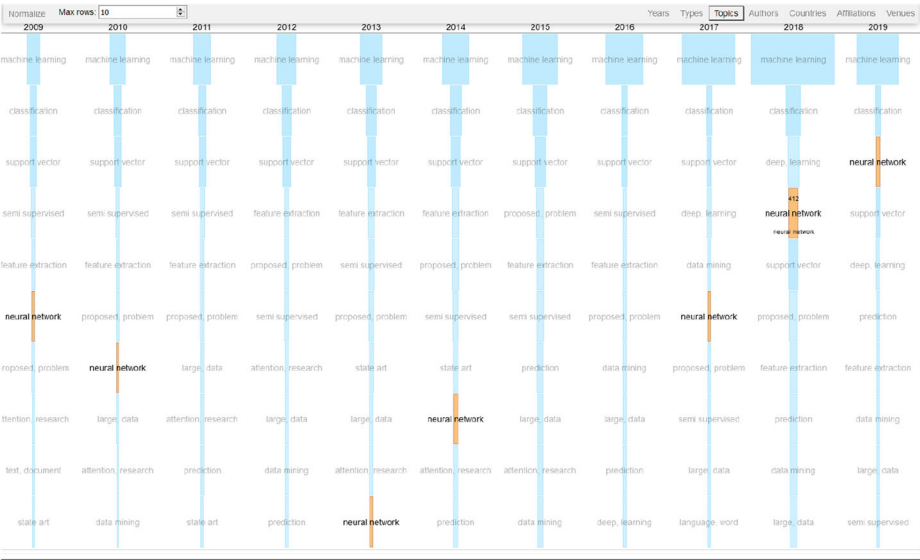
Fig. 3 Temporal visualization: temporal spread of topics in the area of machine learning. Neural networks gained more attention over the last years [36]

to enable a fast and comprehensible analysis. It is even more important to gather different correlations through the semantic data model, geographic spread, topics and temporal spreads and especially the trends, which are modeled through the described procedure (see Section 3.5). A small set of visual representations implemented in our system is shown in Fig. 5. The practical usage in form of the interaction behavior while analyzing trends, technological advancement, and correlations will be described in the next section.

4 Interaction design for technology and innovation management

In the past years, a number of different approaches from Visual Analytics, information Visualization, search behavior and the area of technology and innovation management have
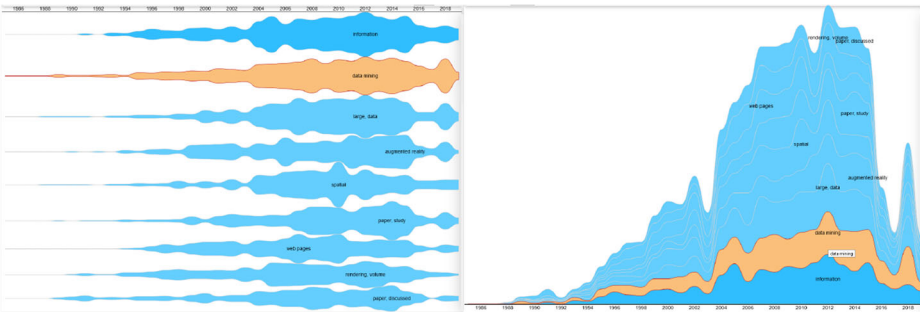
Fig. 4 Temporal visualization: left a river chart with a highlighted topic and right a stack river chart with the same highlighted topic. Both visual representation are using the same data
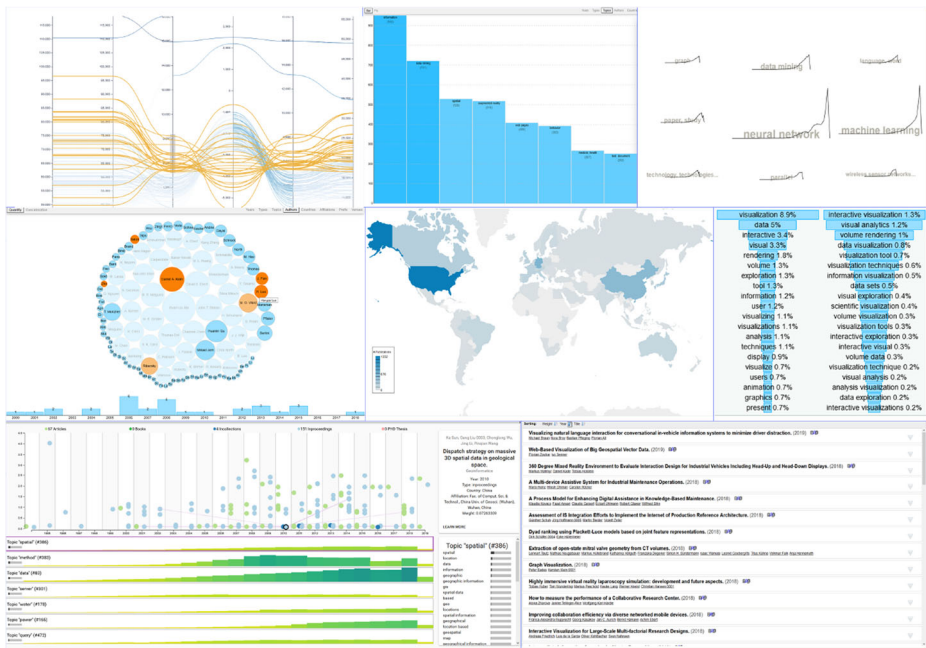
Fig. 5 Visual Representation: A set of different interactive visual representations modeled through the data and the visual structure

arisen to support the analytical process, especially through "visual interactive systems" [37]. Users should be enabled to gather required information through interaction with the system regardless of their domain knowledge or their knowledge about the system. Hence, two complementary approaches have been applied, which support the users' information acquisition and analysis process: (I) Shneiderman's model [47] and (II) van Ham and Perer's model [50]. The model of Shneiderman proposes a top-down interaction following his "Visual Information Seeking Mantra": an interactive visualization's basic principle should be to give an "overview" first, then "zoom and filter", and finally provide "details on demand" [47]. The model of van Ham and Perer proposed a complementary "bottom-up approach", which was designed for the visual exploration in large graphs to support users with their "degree of interest": it starts with "search", then "shows the context of a certain graph", and finally "expands this on demand" [35, 50]. Even the work was designed for large graph exploration, the principal idea can be adapted to a more user-centered approach and therewith complement the basic principle of Shneiderman in particular by the "search" task. Both approaches are shown in an abstract way in Fig. 6.

Marchionini [30] proposed an "exploratory search approach" based on Bloom's taxonomy [4]. He proposed three overlapping types of search activities, where searchers furthermore may be involved in more than one activity at the same time: (I) *Lookup*, (II) *Learn*, and (III) *Investigate*. The lowest level of search activity and thereby the basic step is *lookup*. This activity leads to discrete and well-structured information and enables to answer the questions of who, where, and when in contrast to why or how. The query formulation of this activity premises domain knowledge of searchers [30, 53]. *Learn* as the next step of the
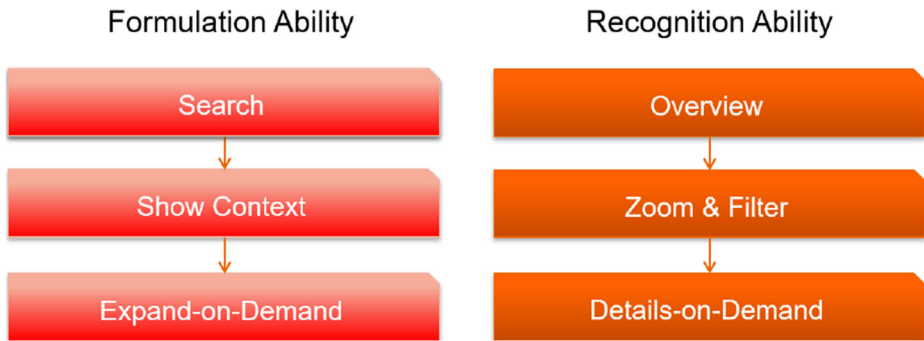
Fig. 6 The complementary interaction approaches applied in our system [35]

search process model, is an exploratory activity. It involves multiple iterations of searching and result evaluations to enhance the knowledge about a certain domain or topic. In comparison to the already introduced search activities, *investigate* is the most complex cognitive activity. It includes tasks as analysis, synthesis, and evaluation. Enhanced knowledge about the topic or domain of interest is required to successfully solve this search activity. This search activity includes not only finding and acquiring new knowledge and information, it also involves analytical tasks such as discovering gaps in the knowledge domain. Marchionini further proposes that serendipitous browsing is a kind of investigative search, which stimulates analogical thinking [30], where users relate their experiences and internalized knowledge from one knowledge task to a related one [30, 53]. Since the process of knowledge discovery constructs knowledge by investigating various sources and ideas [21], it is therefore important to support the process of search and learning by maximizing the number of possible relevant objects (recall) rather than minimizing the number of possible irrelevant information (precision).

4.1 Visual-interactive analysis model for trend analytics

A number of processes arose that should strengthen companys' market position and support technology and innovation management through a well-defined analytics process. For this purpose, patents are commonly used as data and as technology trigger. Bonino et al. [5] proposes that the tasks pursued by patent information users can be subdivided into the three main classes of *search*, *analysis* and *monitoring*. They correlated five search tasks to three main questions of *when*, *why* and *what*, and a "focus" that can be either specific or broad. As a résumé of the elaborated works, the broad-focus can be seen as exploratory search. Such a correlation matrix could be *portfolio survey* as search task, "when" is set to *business planning*, "why" to *identifying the technical portfolio of different players*, "what" to *patents, scientific and technical publications in a given technology area*, and the focus is *broad*. Bonino et al. [5] subdivided *Analysis* into *micro-* and *macro-analysis*. Thereby micro-analysis focuses just on a single patent document or research paper and macro-analysis involves a portfolio of documents (patents or research papers in their case) [5]. The analysis in perspective of patents is usually performed to evaluate and assess the Intellectual Property (IP), to map and chart the IP, to identify trends and competitors and also to identify new areas of potential to exploit. Particularly the trend and competitor identification as well as the identification of news area is also able on behalf of research papers. From its

participial character, the analysis task and and Marchionini's definition of the exploratory investigation step [30] are pretty similar. To keep the user up to date about upcoming (patent or in general research) information in the specified domain of interest, the *monitoring* task is aimed for. This task is commonly performed by the system according to a properly defined interest profile. However, the monitoring task can also be performed manually [5]. Based on the classification of Bonino et al., [22] investigated in an empirical study the questions how search tasks are performed, which search functionalities are perceived as important, and what the ideal (patent or research paper) search system is. The first derived "important system features" from these questions could roughly be grouped into (1) *query formulation*, (2) *result assessment and result navigation*, and (3) *search management, organization and history*. Aspects such as Boolean search, query expansion, field operators etc. are included in the *query formulation*, while *result assessment and result navigation* focused on highlighting, navigation and relevance score, and aspects of combining queries, search or navigation history and timelines are included in *search management, organization and history*. In their evaluation they found that in contrast to web searchers, users of such analytics systems are willing to adopt and leverage functionality [22].

We derived a more general approach for investigating the entire process of visual trend analytics based on the above explained approaches and models. Our model illustrated in Fig. 7 consists of the four main steps of "Overview", "Search", "Visualization" and "Tasks". As the initial steps a user performs during the analysis process, the first two steps of *overview* and *search* can be assigned. So, the first steps are either searching for a certain term or getting an overview of the data or the sub-set of data. The following visualization of the results enable to solve the different tasks. Our approach combines "*analysis*" or "*discovery*" as "more abstract tasks" [37] with specific tasks that are commonly performed during the analysis process, for example "result reduction" or "comparison". Although these tasks are not at the same level of "analysis" or "discovery", dedicated functionalities for the analysis process are required as the introduced models showed. In the following, we outline the
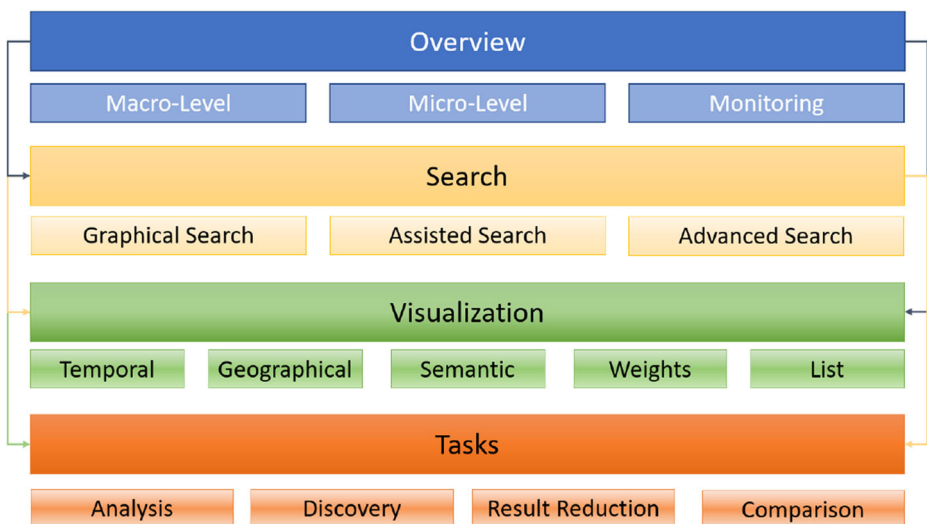

Fig. 7 General interaction design approach in technology and innovation management [37]

different steps and describe how we applied interactive visualizations and further interaction approaches to support the process.

It is to outline, that the implemented system in general supports multiple textual data sources and enables to solve also other, majorly simpler tasks than decision support in technology and innovation management. But we carefully investigate the requirements and evaluate the progress together with analysts to optimize the system of the trend detection for technology and innovation management (see Section 5).

4.2 Overview

Shneiderman's work "Visual Information Seeking Mantra" [47] builds the foundation of our *overview* step that gives an overview to the entire dataset or a sub-set of data. Following Bonino et al. [5], we integrated the three overview levels, namely overview on the "macro level", on the "micro level" and for "monitoring". The overview on the "macro level" gives the initial overview of emerging trends in perspective of trend analytics [36], as illustrated on Fig. 8. The emerging trends gathered through topic modeling are thereby visualized as *SparkCloud* [26], and the users are able to see the most emerging trends at a glance. The right side shows how other topics of interest can be selected. Furthermore, the user can choose an overview of the most frequent topics in the entire dataset or the topics with the highest climax. The interactively designed overview on the "macro level" enables the user to get details-on-demand with one click on certain SparkCloud.

The temporal spread of topics (trends) for a certain key-term that is either searched or selected from the "macro level" overview is in trend analytics part of the overview on micro-level [37]. It gives an overview of all related topics in a temporal manner and insights of the main technologies and approaches for a certain key-term. One example of a micro-level overview is shown in Fig. 9, where the user chose the term "Information Visualization" from the macro-level overview and gets the temporal spread of related technologies and approaches ranked based on the frequency of documents' related topics [37].
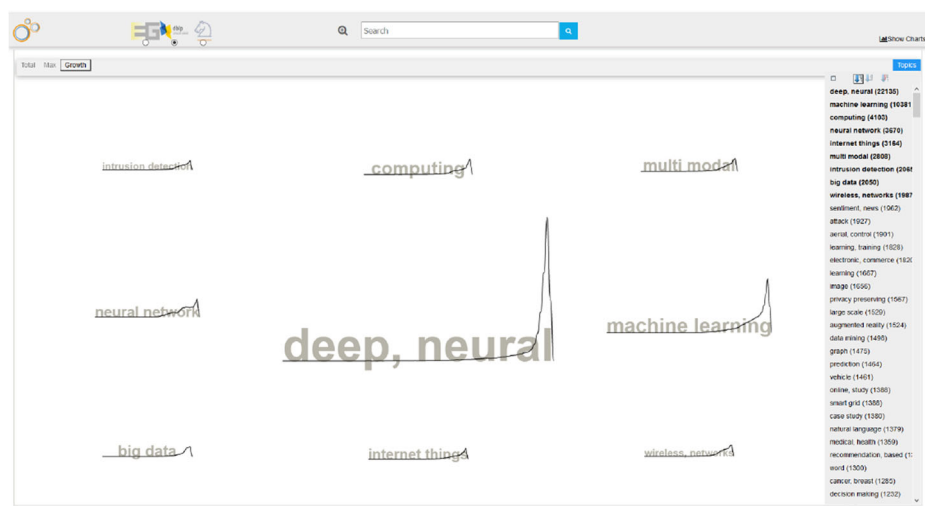

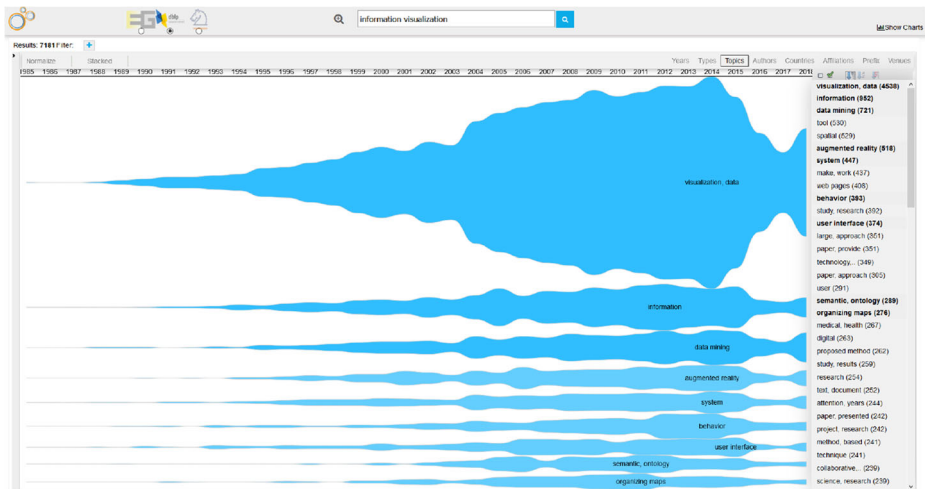**Fig. 8** Overview on Macro-Level: emerging trends of the entire database [37]

Fig. 9 Overview on Micro-Level: temporal spread of related topics for a single key-term [37]

We implemented a personalized word cloud in monitoring as overview, which illustrates a person's most searched or selected terms. The users can select terms they want to monitor, which is implemented via a user model with a *bag-of-words* approach. Figure 10 illustrates the word cloud for two different persons and clearly illustrates that the number of monitored terms can vary significantly.

4.3 Search

Particularly with large amounts of data, search functionalities are essential. To assist the users in the search process, we integrated different search approaches. The integrated "graphical search" function is a novel approach, whereby the initial search term (or selected term) is visualized as a "circle in the center of the screen" [37]. The users can define further search terms as "points of interest" (POI), whereas each POI is displayed as a small circle in a certain color. When the POIs are dragged into the initial search circle, the results including both search terms are illustrated. A nested search is enabled with the graphical search,
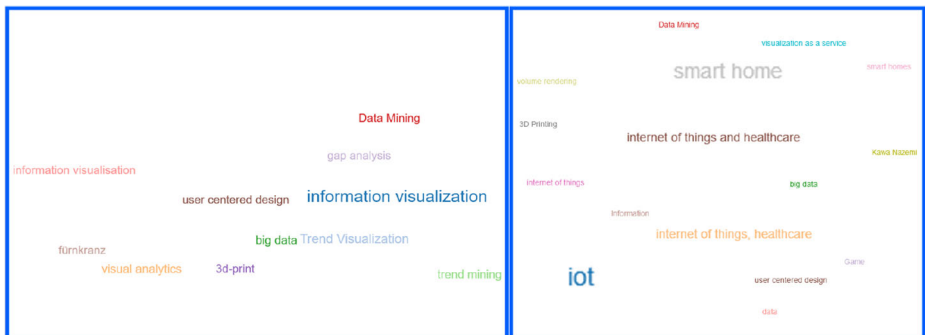
Fig. 10 Overview - Monitoring: personalized word-clouds of two different persons [37]
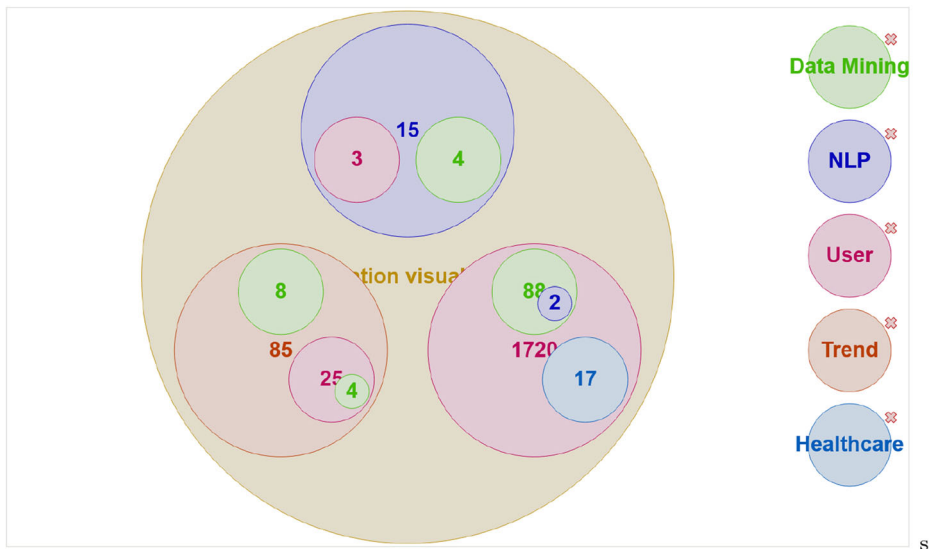
Fig. 11 Search - Graphical Search: the initial search-term is represented as a circle, users can define further search terms and drag them into the initial search-term to get a nested number of combined search [37]

so that more than two circles can be dragged. The number of the results after the interaction with the graphical search is always the number of all nested POIs. The graphical search approach is depicted in Fig. 11. Thereby the initial search is information visualization and the user has created five POIs and dragged them into the initial search term, represented as a circle. The number of results corresponds to the shown number. The circles are interactive and show the underlying results when clicked [37].

The search function was extended to the assisted search, according to our previous work [40], which extends the search functionalities of traditional linguistic methods with a topic-based approach. The approach incorporates the information of the generated topics to enhance the query's search-terms. The topics consists of N-Grams and represent the most frequent phrases within the corresponding topic. The topics consists of N-Grams and represent the most used phrases within the corresponding topic. These phrases often represent different ways to articulate the topic's idea and can be used as a key-phrase to represent the topic. We use data from the phrases and choose the top five most used phrases as additional search-terms to extend the result set. Based on the initial users' search, the most dominant topic in the result set is identified.

In the next step, semantically similar phrases are extracted based on the identified dominant topic. The system performs additional searches using the semantically similar phrases as search-terms. To activate the assisted search, the user has to enable it next to the search field. The advanced search automatically detects all available data properties and provides a dedicated search in those. Both functionalities are shown in Fig. 12.

4.4 Visualization

With a set of different integrated visualization, we aim to respond to the above-mentioned questions [40, p. 4]. The integrated different data models are the foundations of the underlying visual structures that address the stated questions. The visualizations automatically
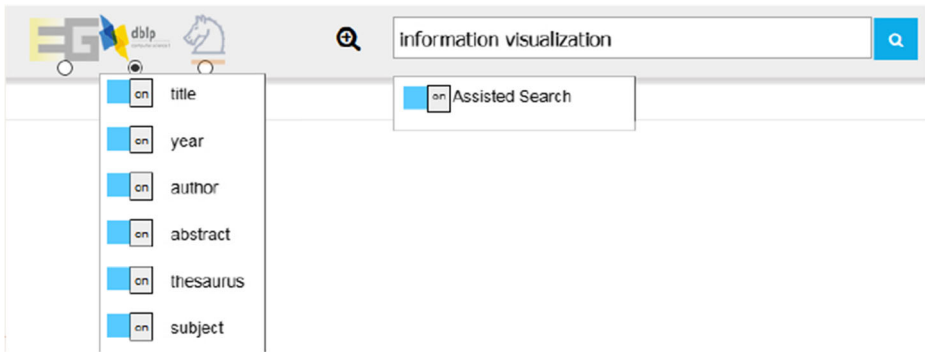
Fig. 12 Search - assisted search & advanced search

detect the supported and required data models and visualize the data. The semantic data model serves as the primary model from which the other data models are forked. The yet integrated set of visualizations cover temporal, geographical, semantic and topic (weights) visual layouts that are all interactive and enable solving different tasks. To select and view a single entity, a simple list-view is also part of the visualization set. Some examples of the integrated visualizations that enable the solving of different analytical tasks are illustrated in Fig. 5. The visualizations can be used as single interactive visual interface or be combined in an interactive dashboard.

4.5 Tasks

The visualizations and the overall user interface enable users to do different tasks. The tasks vary in every domain and are strongly dependent on the visual system's users. The analysis tasks in our approach respond to the questions posed in Section 4.4: Aspects of predicting the trajectory of technologies [48], detecting emerging trends [36] and strengthening the market position of enterprises play an essential role [44]. The analysis tasks incorporate solving a complex analytical task with a specific goal in order to strengthen a company's or other institution's potentials [45]. The discovery tasks aim at detecting unexpected patterns, topics, technologies or correlations in data [33]. These tasks can be best supported by aspect-oriented visualizations, where the user gets insights from different perspectives and detects a new correlation. Such a correlation in Trend Analytics may be that a technology, material or approach can be used in a totally different application, which can be a driver of a product's innovativeness and eventually its commercial success [25]. Besides these two more abstract tasks, we defined in our approach result reduction and comparison as main tasks. Result reduction is essential for every research task, for example researching for related technologies or competitors. To support these tasks, we integrated a dynamic faceting that is further supported by each of the visual layouts. User can start exploring a huge amount of data related to a certain field and then reduce the amount based on their requirements.

Figure 13 illustrates the user interface of our application including the faceting and visual reduction. The user started with a search-term with about 60,000 results (top left of the screen), used the graphical search, faceting and visual interactions to reduce the amount to only six relevant papers, and then chose four visualization to see the results. The faceting is visualized on the left and the result reduction is on top (highlighted with a blue rectangle).
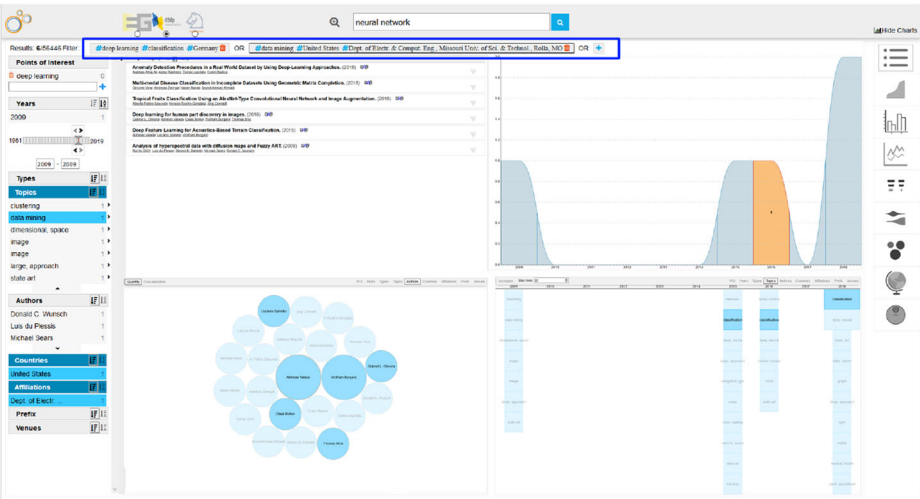
Fig. 13 Tasks: result reduction - The facets outlined on top can be combined with logical operators and refine the amount of the results

The comparison tasks can be performed on two different levels: Comparison of data-subset with two same visual layouts or comparison of different databases. This differentiation is similar to the proposed micro-level and macro-level overview. Based on the requirements, a user is able to set up a campaign and compare it to the entire data of a different database. For example, this enables the comparison of company-data with scientific publications. Figure 14 illustrates a comparison on the macro-level through different databases and the visual layout is a macro-level overview too, so that the user is able to see the emerging trends in two different databases.
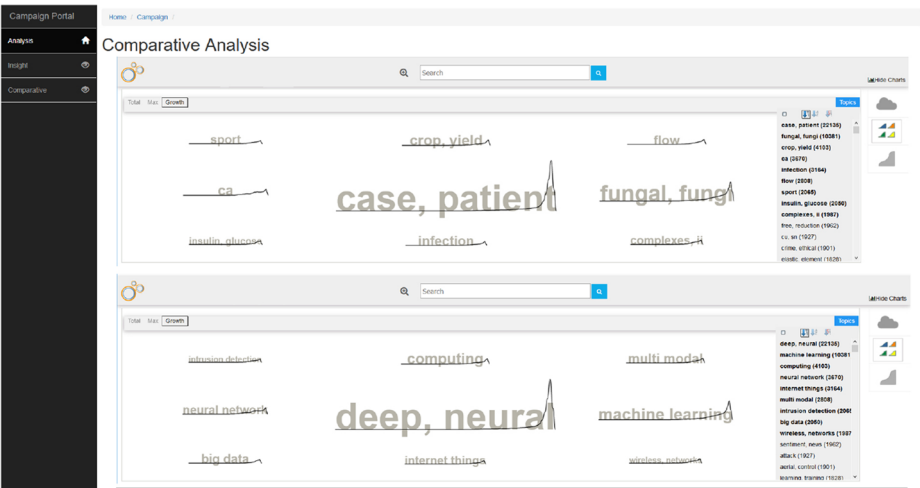
Fig. 14 Tasks: Comparison on macro-level by using macro-level overview visual layouts
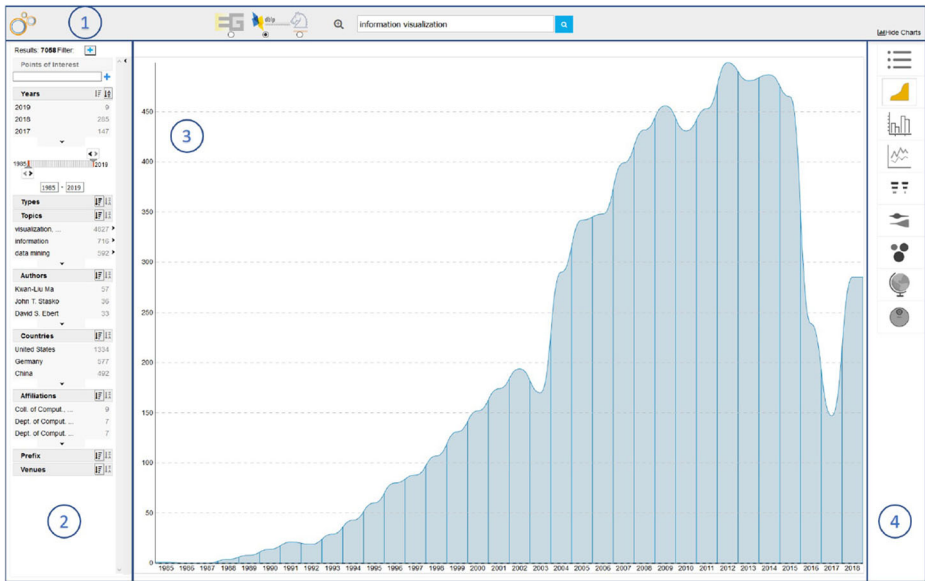
Fig. 15 The user interface of our system with its four areas

With this procedure, users are able to start their analytical tasks and receive —either after querying a search term or after clicking on a trend or a search phrase— a temporal overview of the database's entire documents. Figure 15 illustrates one initial screen with the entire user interface and the temporal overview of the underlying documents after a query for "Information Visualization".

The user interface is built with four main areas. At the top (1) of Fig. 15, users are able to search (including advanced search formulation capabilities), activate assisted search, select a database or hide the visualization selection area. The assisted search is implemented according to our previous work [40] that enhances users' performed query based on the resulted top five phrases of the top ranked topic. At the left (2), the underlying data's facets are dynamically generated and visualized. This area also includes the number of results, which is automatically adapted to the selected facets, a logical facet selection, and the search-in-search functionality (see Fig. 11). The logical facet selection allows users to reduce the amount of the results to get the most appropriate documents for a certain task. In Fig. 16 the user has selected after the query for "Information Visualization" the facets *data mining*, *social networks* and *government, public* and combined them with a logical OR-operator with other facets. The number of the results are reduced from 7058 to 8. The list-view shows the final result set of documents.

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 Section 3.6) or by the users themselves in the right area (4), where a dynamic set of visualization are available based on the data and their structure. In Fig. 15 the temporal overview of the entire data is visualized while Fig. 16 illustrates a list-view with a small number of documents that are refined through the facets.

The graphical search functionality allows users to formulate relevant terms, create visual *points-of-interests* (POI) and see at a glance the number of documents that contain the created points of interest (Fig. 17). In Fig. 11, the user searched for "data mining" and created
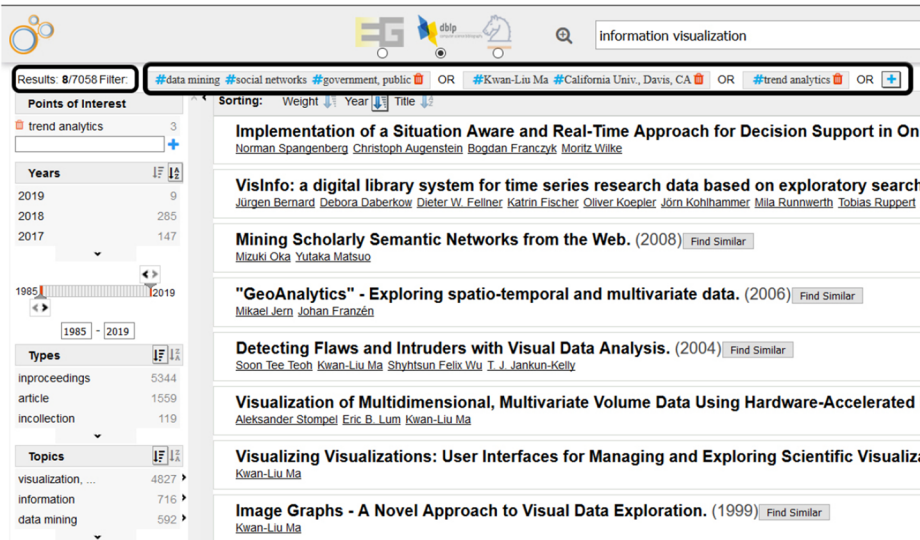
Fig. 16 Facets with the list-view. With highlighted facets chosen by the user and the number of results

a number of visual POIs. The defined POIs are visualized at the right side and can be included into the main search term "data mining" per drag and drop. The color is the indicator for a certain POI and allows users to see how many publications are in the data base with the created POIs. The number represents the search result quantity within the search result set, so users are able to define and redefine such POIs for their purposes. In Fig. 11, the main search term is "data mining". Users are able to see at a glance the results for data mining documents containing the text, image, pattern, clustering or health. With a nested method they are able to see that there are 938 publications containing the phrase "Image" withing the result corpus "Data Mining" (circle at right-bottom) with 17 publications about data mining of images for the domain health, 120 results containing pattern and within this search 30 documents that are using clustering methods.

The described visual interfaces and interaction designs are just a set of possible interactive visualization that should enable solving analytical tasks with regards to technology and
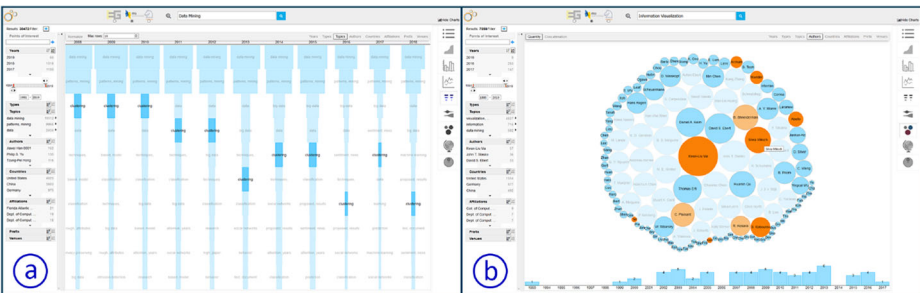

Fig. 17 The visual POIs for enabling a search within the result set [36]

Fig. 18 Example of an alternative visual interface

innovation management and related questions as described in Section 3.7. With our proposed approach varieties of visual interfaces can be integrated for various tasks. Figure 18 illustrates some more examples designed for different tasks.

5 Preliminary evaluation

A formal evaluation was not conducted yet to validate the advancement on gathering new insights through the introduced approach and system. However, we performed with four experts from the area of "strategic decision making" an informal user study by combining the "Thinking Aloud" method combined with interviews in two iterations. The second iteration includes some of their improvement suggestions. All four participants worked for consulting companies in the domain of strategic decision making. All had no or less experience with visual analytics systems.

All had no or little experience with visual analytics systems. All were male with an average age of 43. The subjects received a short explanation of the system and were asked to create a scenario for their own business. Three subjects described a current strategic decision-making task, one subject an exploratory search task. The subjects were asked to explain their actions and thoughts while they interacted with the system. Through this method, we were able to find some interaction errors that were eliminated in the second iteration. The interviews were about 30 minutes long, whereas the subject who decided for an exploratory search worked with the system about one hour in both sessions. The main improvement suggestions of all four in the first iteration were (1) the "graphical search", (2) "the overview with the top- ics of the entire database" and (3) the "geographical visualization that should include a temporal view" too. In the second iteration, three subjects asked for a reporting functionality, which is not included in the introduced system and is

now under investigation, two subjects asked for uploading own data to analyze these to, one subject criticized the number of visualizations and wanted to have more choices for visual representation. All criticized the long response time of the system, which is in the current version about 5 seconds for huge queries (above 100,000 results). Three subjects said that they were able to get new insights and rated the system as "very useful". One subject ranked the system as "useful", whereas this person just searched exploratorily and not with a dedicated task. All subjects ranked the interaction design in the second iteration with "graphical search" and "statistical values" as very appropriate. All participants said that they would use the system in their daily work. This informal and formative interview lead to some changes that were partially included into the system. A formal evaluation will follow.

6 Conclusions

Technological developments have an essential impact on strategic decision making in companies [16]. The early awareness of possible emerging technological trends strengthens a firm's competitiveness and market position [45]. If innovation-driven companies ignored emerging technological developments, however, they may not tap the full potentials of their own products or technologies. The majority of companies detect early signals of emerging trends by chance rather than through systematic processes [43]. Larger enterprises already consider this essential aspect and create foresight units that try to anticipate future developments and upcoming innovations [46].

To face this problem, this paper introduced the entire process for an integrated and interactive Visual Analytics System. We first described an approach for integrating, enriching, mining, analyzing, identifying and visualizing emerging trends from scientific text collections based on our previous work [36] that included the state-of-the-art analysis in trend mining and trend and text visualization. A revised description of our approach for integrating, enriching, mining, analyzing, identifying and visualizing emerging trends for decision making [36] was the baseline for our main contribution, an approach for interaction design with the main goal to keep the human in the loop of visual interactive systems by investigating the scientific approaches from technology and innovation management [11, 16, 33, 45]. To achieve this goal, we first introduced a general interaction design approach for technology and innovation management and illustrated with different examples the interaction design on macro- and micro-level, assistance in Visual Analytics systems and graphical search approaches.

References

1. Agrawal R, Psaila G, Wimmers EL, Zaït M (1995) Querying shapes of histories. In: Proceedings of 21th international conference on very large data bases (VLDB'95). Morgan Kaufmann
2. Bertin J (1983) Semiology of graphics. University of Wisconsin Press
3. Blei DM, Ng AY, Jordan MI (2003) Latent dirichlet allocation. Journal of Machine Learning Research 3. http://www.jmlr.org/papers/v3/blei03a.html
4. Bloom BS (1956) Taxonomy of educational objectives. David McKay Co. Inc, New York
5. Bonino D, Ciaramella A, Corno F (2010) Review of the state-of-the-art in patent information and forthcoming evolutions in intelligent patent informatics. World Patent Inf 32(1):30–38. https://doi.org/10.1016/j.wpi.2009.05.008
6. Bun KK, Ishizuka M (2002) Topic extraction from news archive using tf*pdf algorithm. In: Proceedings of the third international conference on web information systems engineering, 2002. WISE 2002, pp 73–82. https://doi.org/10.1109/WISE.2002.1181645
7. Card SK, Mackinlay JD, Shneiderman B (1999) Readings in information visualization: using vision to think, 1st edn. Morgan Kaufmann, San Mateo
8. Chen K-Y, Luesukprasert L, Chou S-CT (2007) Hot topic extraction based on timeline analysis and multidimensional sentence modeling. IEEE Trans Knowl Data Eng 19(8):1016–1025. https://doi.org/10.1109/TKDE.2007.1040
9. Collins C, Viegas FB, Wattenberg M (2009) Parallel tag clouds to explore and analyze faceted text corpora. In: VAST 2009. https://doi.org/10.1109/VAST.2009.5333443
10. Dou W, Wang X, Chang R, Ribarsky W (2011) Parallel topics: a probabilistic approach to exploring document collections. In: VAST 2011. https://doi.org/10.1109/VAST.2011.6102461
11. Eggers JP, Park KF (2018) Incumbent adaptation to technological change: the past, present, and future of research on heterogeneous incumbent response. Academy of Management Annals 12(1):357–389. https://doi.org/10.5465/annals.2016.0051
12. Feldman R, Aumann Y, Zilberstein A, Ben-Yehuda Y (1998) Trend graphs: visualizing the evolution of concept relationships in large document collections. In: Żytkow JM, Quafafou M (eds) Principles of data mining and knowledge discovery. Springer, Berlin, pp 38–46
13. Feldman R, Dagan I (1995) Knowledge discovery in textual databases (kdt). In: Proceedings of the first international conference on knowledge discovery and data mining
14. Glance NS, Hurst M, Tomokiyo T (2004) Blogpulse: automated trend discovery for weblogs. In: WWW 2004 WS on weblogging. ACM
15. Goorha S, Ungar L (2010) Discovery of significant emerging trends. In: Proceedings of the 16th ACM SIGKDD. https://doi.org/10.1145/1835804.1835815
16. Gordon AV, Ramic M, Rohrbeck R, Spaniol MJ (2020) 50 years of corporate and organizational foresight: looking back and going forward. Technol Forecast Soc Chang 154:119966. https://doi.org/10.1016/j.techfore.2020.119966
17. Han Q, Heimerl F, Codina-Filba J, Lohmann S, Wanner L, Ertl T (2017) Visual patent trend analysis for informed decision making in technology management. World Patent Inf 49:34–42. https://doi.org/10.1016/j.wpi.2017.04.003
18. Havre S, Hetzler E, Whitney P, Nowell L (2002) The meriver: visualizing thematic changes in large document collections. IEEE TVCG 8(1):9–20. https://doi.org/10.1109/2945.981848
19. Heimerl F, Han Q, Koch S, Ertl T (2016) citerivers: visual analytics of citation patterns. IEEE TVCG 22(1):190–199. https://doi.org/10.1109/TVCG.2015.2467621
20. Hurtado JL, Agarwal A, Zhu A (2016) Topic discovery and future trend forecasting for texts. Journal of Big Data 3(1):7. https://doi.org/10.1186/s40537-016-0039-2
21. Jerome S (1961) Bruner: the act of discovery. Harv Educ Rev 31:21–32
22. Joho H, Azzopardi LA, Vanderbauwhede W (2010) 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. ACM, New York, pp 13–24. https://doi.org/10.1145/1840784.1840789
23. Keim D, Kohlhammer J, Geoffrey E, Mansmann F (eds) (2010) Mastering the information age : solving problems with visual analytics. Eurographics Association, Goslar
24. Kim Y, Tian Y, Jeong Y, Jihee R, Myaeng S-H, Shin SY, Ossowski S (2009) Automatic discovery of technology trends from patent text. In: Proceedings of the 2009 ACM symposium on applied computing (SAC), Honolulu, Hawaii, USA, March 9-12, 2009. ACM. https://doi.org/10.1145/1529282.1529611
25. Kock A, Gemünden HG, Salomo S, Schultz C (2011) The mixed blessings of technological innovativeness for the commercial success of new products. J Prod Innov Manag 28(s1):28–43. https://doi.org/10.1111/j.1540-5885.2011.00859.x

26. Lee B, Riche NH, Karlson AK, Carpendale S (2010) Sparkclouds: visualizing trends in tag clouds. IEEE TVCG 16. https://doi.org/10.1109/TVCG.2010.194
27. Lent B, Agrawal R, Srikant R (1997) Discovering trends in text databases. In: Proceedings of KDD '97
28. Liu S, Zhou MX, Pan S, Song Y, Qian W, Cai W, Lian X (2012) Tiara: interactive, topic-based visual text summarization and analysis. ACM Trans Intell Syst Technol 3(2):25: 1–25: 28. https://doi.org/10.1145/2089094.2089101
29. Lohmann S, Burch M, Schmauder H, Weiskopf D (2012) 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. ACM, New York, pp 753–756. 10.1145/2254556.2254701
30. Marchionini G (2006) Exploratory search: from finding to understanding. Commun ACM 49(4):41–46. https://doi.org/10.1145/1121949.1121979
31. Mei Q, Zhai C (2005) Discovering evolutionary theme patterns from text: an exploration of temporal text mining. In: Proceedings of ACM SIGKDD
32. Montes-y-Gomez M, Gelbukh A, Lopez-Lopez A (2001) Mining the news: trends, associations, and deviations. Computacion y Sistemas 5(1):14–24
33. Mühlroth C, Grottke M (2018) A systematic literature review of mining weak signals and trends for corporate foresight. Journal of Business Economics 88(5):643–687. https://doi.org/10.1007/s11573-018-0898-4
34. Mühlroth C, Grottke M (2020) Artificial intelligence in innovation: how to spot emerging trends and technologies. IEEE Trans Eng Manag, pp 1–18. https://doi.org/10.1109/TEM.2020.2989214
35. Nazemi K (2016) Adaptive semantics visualization. Studies in Computational Intelligence 646. Springer International Publishing, Studies in Computational Intelligence 646. https://doi.org/10.1007/978-3-319-30816-6
36. Nazemi K, Burkhard D (2019) Visual analytics for analyzing technological trends from text. In: 2019 23rd international conference information visualisation (IV). IEEE, pp 191–200. https://doi.org/10.1109/IV.2019.00041
37. Nazemi K, Burkhardt D (2019) A visual analytics approach for analyzing technological trends in technology and innovation management. In: Bebis G, Boyle R, Parvin B, Koracin D, Ushizima D, Chai S, Sueda S, Lin X, Lu A, Thalmann D, Wang C, Xu P (eds) Advances in visual computing, lecture notes in computer science, vol 11845. Springer International Publishing, Cham, pp 283–294. https://doi.org/10.1007/978-3-030-33723-0_23
38. Nazemi K, Burkhardt D, Retz R, Kuijper A, Kohlhammer J (2014) Adaptive visualization of linked-data. In: Advances in visual computing. Springer, pp 872–883
39. Nazemi K, Klepsch M, Burkhardt D, Kaupp L (2020) Comparison of full-text articles and theircorresponding abstracts for visual trend analytics. In: Proceedings of the 24rd international conference information visualisation. IEEE, IEEE. To appear
40. Nazemi K, Retz R, Burkhardt D, Kuijper A, Kohlhammer J, Fellner DW (2015) Visual trend analysis with digital libraries. In: Proceedings of the 15th international conference on knowledge technologies and data-driven business - i-KNOW 2015. ACM Press. https://doi.org/10.1145/2809563.2809569
41. Nguyen K-L, Shin B-J, Yoo SJ (2016) 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. https://doi.org/10.1109/BIGCOMP.2016.7425917
42. Noh Y, Hagedorn K, Newman D (2011) Are learned topics more useful than subject headings. In: Proceedings of the 11th ACM/IEEE JCDL
43. Rohrbeck R, Battistella C, Huizingh E (2015) Corporate foresight: an emerging field with a rich tradition. Technol Forecast Soc Chang 101:1–9. https://doi.org/10.1016/j.techfore.2015.11.002
44. Rohrbeck R, Gemünden HG (2011) Corporate foresight: its three roles in enhancing the innovation capacity of a firm. Technol Forecast Soc Chang 78(2):231–243. https://doi.org/10.1016/j.techfore.2010.06.019
45. Rohrbeck R, Kum ME (2018) Corporate foresight and its impact on firm performance: a longitudinal analysis. Technol Forecast Soc Chang 129:105–116. https://doi.org/10.1016/j.techfore.2017.12.013
46. Schoemaker PJH, Tetlock PE (2017) Building a more intelligent enterprise. MIT Sloan Management Review Spring 2017(58):28–37
47. Shneiderman B (1996) The eyes have it: a task by data type taxonomy for information visualizations. In: VL, pp 336–343
48. Sinan Erzurumlu S, Pachamanova D (2020) Topic modeling and technology forecasting for assessing the commercial viability of healthcare innovations. Technol Forecast Soc Chang 156:120041. https://doi.org/10.1016/j.techfore.2020.120041

49. Srikant R (1997) Mining sequential patterns. In: Proceedings of the eleventh international conference on data engineering
50. van Ham F, Perer A (2009) Search, show context, expand on demand: supporting large graph exploration with degree-of-interest. IEEE Trans Vis Comput Graph 15:953–690
51. Viermetz M, Skubacz M, Ziegler CNZ, Seipel D (2008) Tracking topic evolution in news environments. In: 10th IEEE conference on e-commerce technology, pp 215–220. https://doi.org/10.1109/CECandEEE.2008.112
52. Wallach HM, Mimno DM, McCallum A (2009) Rethinking lda: why priors matter. In: Bengio Y, Schuurmans D, Lafferty JD, Williams CKI, Culotta A (eds) Advances in neural information processing systems, vol 22. Curran Associates, Inc, pp 1973–1981. http://papers.nips.cc/paper/3854-rethinking-lda-why-priors-matter.pdf
53. White RW, Roth RA (2009) Exploratory search: beyond the query-response paradigm. 1947-945X 1. https://doi.org/10.2200/s00174ed1v01y200901icr003
54. Yu Z, Johnson TR, Kavuluru R (2013) Phrase based topic modeling for semantic information processing in biomedicine. In: Proceedings of the international conference on machine learning and applications, vol 2013. International Conference on Machine Learning and Applications, pp 440–445. https://doi.org/10.1109/ICMLA.2013.89
55. Zhai CX, Velivelli A, Yu B (2004) A cross-collection mixture model for comparative text mining. In: Proceedings of the tenth ACM SIGKDD, KDD '04. ACM, New York, pp 743–748. https://doi.org/10.1145/1014052.1014150