Analysis of Technology Trends Based on Diverse Data Sources

Aviv Segev, *Member, IEEE*, Sukhwan Jung, and Seungwoo Choi

Abstract—The paper suggests a method for analyzing technology trends. The process, which investigates development of technologies over time, identifies main technologies displaying the fastest growth compared to greater influence of new inventions. The method analyzes term frequency and change over time of technological terms in academic articles and patents to identify the prior technologies that lead to a new technology and to detect technologies that have the biggest impact. The analysis was performed on 4,354,054 patents from the US Patent Office dating from 1975 until today. In addition, academic articles were analyzed as a trend forecasting data set to identify patents trends 4-5 years in advance and technology trends up to 9 years in advance. The forecasting method was extensively validated using a large repository of real-world technology terms, and the results were verified against Gartner technology predictions, Web searches, news articles, and book publications. The method shows higher accuracy than existing forecasting methods do. Some correlation is displayed between technology trends and future US stock market performance.

Index Terms—Technology trend, prediction, big data, patents, academic articles

1 Introduction

The topic of identifying new technologies has implementations in the areas of stock prediction, technology venture funds, and government research investment planning. The current work presents a method for analyzing technology trends and identifying the cause and effect of a given technology. The method is based on temporal term frequency analysis compared with first and second derivatives of change of similar technologies. This perspective presents both the growth and the influence of a specific technology in a specific time period. These technologies are compared to identify cause and effect of specific technologies and technology trends that have major influence on innovation over time.

Commonly used methods of technology forecasting [1] include the Delphi method [2], forecast by analogy [3], growth curves [4], and extrapolation [5]. Normative methods of technology forecasting—like decision trees [6], morphological models [7], and mission flow diagrams [8]—are also frequently employed.

The Delphi method, which is a commonly used method, is a structured communication technique, first developed as a systematic and interactive method of forecasting that relies on a panel of experts. In the standard version, the experts fill out questionnaires in two or more rounds. The belief is that during this process the range of the answers will decrease and the group will converge toward the "correct" answer. The process is concluded after a predefined stop criterion (e.g., number of rounds, achievement of consensus, stability of results), and the mean or median scores of the final rounds determine the results. However, the Delphi method requires experts in each field, while the current work aims at automating the process.

Studies of past forecasts have indicated that one of the most common reasons forecast methods fail is that the forecaster ignores related fields [9]. A given technical approach may fail to attain the level of capability predicted for it, because it is replaced by another technical approach the forecaster ignored. Another problem is inconsistency between forecasts. Given these problems, it is frequently necessary to combine forecasts of different technologies. Hence, rather than to attempt to choose the one most appropriate method, it may be better to try to combine the forecasts obtained by different methods. In this combination of forecasts achieved by different methods, the strengths of one method may help counterbalance the weaknesses of another.

The main reason for combining forecasts of the same technology is to try to balance the disadvantages of one forecasting method with the advantages of another [5]. Furthermore, the use of more than one forecasting method frequently provides the forecaster with more insight into the processes at work responsible for the growth of the forecast technology [10].

An often used combination is that of growth curves and a trend curve for some technology. A succession of growth curves each describes the level of functional capability achieved by a specific technical approach, while an overall trend curve reflects the historical data. With growth curves alone, the forecaster cannot say anything about the time at which a given technical approach may be replaced by a successor approach. With the trend curve alone, the forecaster cannot say anything about the ability of a specific technical approach to meet the projected trend or about the need to search for a successor approach. Thus, the need for combining forecasts is evident. The use of growth curves and a trend curve together enables the forecaster to draw conclusions about the future growth of a technology, which might not be possible, were either method used alone.

- *The authors are with the Department of Knowledge Service Engineering, Korea Advanced Institute of Science and Technology (KAIST), Daejeon 305-701, South Korea*
E-mail: aviv@kaist.edu, {raphael, sw.choi}@kaist.ac.kr.

Manuscript received 1 Aug. 2013; revised 12 Apr. 2014; accepted 21 June 2014. Date of publication 13 July 2014; date of current version 11 Dec. 2015.
For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org, and reference the Digital Object Identifier below.
Digital Object Identifier no. 10.1109/TSC.2014.2338855

Another often used combination of forecasts is the trend curve and one or more analogies [8]. The scatter of data points about a trend curve is usually thought to be due to random influences that cannot be controlled or measured. However, consistent deviations may represent something other than just random influences, and when they are identified, there may be an opportunity to apply an analogy. Typical events that cause deviations from a trend are wars and depressions. Hence, the purpose of combining analogies with a trend forecast is to predict deviations from the trend deviations associated with or caused by external events or influences.

A trend analysis method that uses term frequency focuses on the implications of each keyword in a time series [11]. This approach is based on the idea that terms with changes in their appearance frequency, even if transitory, are more important for finding changes in trends than are stable, constantly used terms. In trend analysis, terms in a given time period are classified according to technologies, and trends are discovered based on how these classifications change over each time period.

The current work presents a method of prediction based on term frequency and identifies exponential growths of technology. The method compares different data sources of technology representatives: 4,354,054 patents from the US Patent Office, academic articles, Web searches, news articles, book publications, and Gartner technology predictions. The comparison of the method data sets analyzes which approach can predict the technology with higher proficiency. In addition, each data source is compared to analyze how far into the future, in number of years, the trend can be observed.

The main contributions of this work are as follows:

- On a conceptual level, we introduce a method for predicting technology growth based on term frequency.
- On an algorithmic level, we provide an implementation of the method based on data sources in the fields of academic articles, patents, Web searches, news articles, and book publications and compare the predictions with Gartner technology predictions.
- On a practical level, we analyze each data source of the different method implementations to identify how far into the future the predictions can be performed and the accuracy that can be attributed to each of the method implementation data sources.

The remainder of the paper is organized as follows. Section 2 discusses the related work. Section 3 describes the technology trends analysis method and illustrates each step of the process using an example. Section 4 presents experiments, and Section 5 presents the results of our proposed approach. Finally, Section 6 discusses issues related to the experiments, and Section 7 provides some concluding remarks.

2 Related Work

2.1 Information Retrieval (IR)

Previous work in information retrieval has targeted patent documents. During the NTCIR (NII Test Collection for IR Systems) Workshops, in patent retrieval tasks a test collection of patent documents was produced and used to evaluate a number of participating IR systems. One task analyzed geographic and temporal information retrieval [12], with the focus to perform searches with geographic and temporal constraints. The data collections (Japanese and English news stories) combined geographical IR with time-based search to find specific events in a multilingual collection. Other attempts at patent classification focused on cross-lingual link discovery (CLLD) [13], which sought to automatically find potential links between documents in different languages. The goal of this NCTIR task was to create a reusable resource for evaluating automated CLLD approaches. The goal of the task was to build and refine systems for automated link discovery. The task focused on linking between English source documents and Chinese, Korean, and Japanese target documents. Currently, the NTCIR tasks aim at machine translation of sentences and claims from Chinese to English, Japanese to English, and English to Japanese [14].

The Workshop of Cross-Language Evaluation Forum (CLEF 2009) [15] gave separate topic sets for the language tasks, when the document language of the topics was English, German, and French. CLEF-IP included prior art candidate search task (PAC) and classification task (CLS). Participants in the PAC task were asked to return documents in the corpus that could constitute prior art for a given topic patent. Participants in the CLS task were given patent documents that had to be classified using the International Patent Classification codes. In addition, evaluations were performed on chemical data sets in chemical IR in general and in chemical patent IR in particular. A chemical IR track in TREC (TREC-CHEM) [16] addressed the challenges in chemical and patent IR.

Previous work, as described above, analyzes automatic patent retrieval, while this work describes a method that involves a manual decision process assisted by an automatic suggestion of relevant concepts related to patent technology evolution over time.

2.2 Trend Prediction

A mathematical model developed for the extrapolation of technological performance functions describes the rate of technological progress in the form of a logistic curve [17]. The model included methods developed for estimating uncertainty levels associated with technological figure-of-merit projections and the relationships between technological progress and market substitution.

Dereli and Durmusoglu [18] developed a trend-based patent alert system (PAS) to find current trends in patents for industrial technologies. The PAS was extended using fuzzy linear regression based on possible deviations where deviations are reflected as the fuzziness of the system [19]. Patent count data based on number of patents filed in a specified time period is considered an indicator defining a current trend. However, this trend extraction algorithm was developed based on linear regression analysis of patent data.

Patent documents are a plentiful source of technical and commercial knowledge, and therefore patent analysis has long been deemed a useful means for research and development management and techno-economic analysis. Citation analysis has been the most commonly adopted tool for patent analysis. Yoon and Park [20] propose a network-based analysis, an alternative method for citation analysis that uses an illustrative data set and describes an overall process of developing a patent network. Furthermore, such new indexes as the technology centrality index, technology cycle index, and technology keyword clusters are proposed for in-depth quantitative analysis. Their method depicts the overall relationship among patents as a visual network and provides richer information, thus enabling deeper analysis since it takes into consideration more diverse keywords and produces more meaningful indexes. These visuals and indexes can be used in analyzing up-to-date trends of high technologies and in identifying promising avenues for new product development. The work collected wavelength division multiplexing (WDM) related patent documents from the U.S. Patent and Trademark Office database. The work used text mining that extracts keywords from patent documents and generated a patent network with nodes (patents) and links (relation among patents). The analysis generated a selective set of influential patents that require more intensive control. In addition, initial work on the topic of trend analysis using patents described the potential of using patents [21]. However, the work did not evaluate the method or present a method that utilizes academic articles, news, books, or Web searches to achieve a much wider perspective of trend technology.

Technology fusion has been advanced as a promising method for the creation of hybrid technologies. No and Park [22] try to further the understanding of the development trajectories of technology fusion in three important aspects. The first aspect is the development of an index that measures the degree of fusion of cross-disciplinary technology at the meso level. The second aspect is the classification of the trajectory patterns of technology fusion in terms of fusion degree. They analyze the fusion mechanism by utilizing citation network analysis. The third aspect is the visualization of the relationship between patents and their backward and forward patent citations, at the patent class level, with their direction on a citation map. The changes in fusion patterns are analyzed using time series comparisons. An empirical analysis in the nanobiotechnology field shows no positive relationship between the inflow and outflow degree of fusion. Changes in the trajectory patterns of fusion over time are observed. This analysis demonstrates that each fusion pattern has evolved in such a way that technologies focus more on their niche technologies and those technologies that cannot incorporate the technology fusion have been eliminated during the development process. Trajectories of technology fusion can be viewed as one element of technology trend prediction.

Kim et al. [23] propose a model that analyzes and forecasts technology trends based on quantitative analysis and several text mining technologies for effective, systematic, and objective information analysis. The work executes a comparative evaluation between the proposed model and Gartner's forecasting model so as to validate the proposed model, because Gartner's model is widely and generally used for information analysis and forecasting. The technology discovery model used consists of three sub-models, technology life cycle discovery (TLCD) model, technology maturity forecasting (TMF) model, and emerging technology discovery (ETD) model, and proposed a technology trends analysis and forecasting model by using a decision tree.

The intention of a study of Web 2.0 articles was to analyze the content of what is written and to develop a statistical model to predict whether authors write about the need for new instructional design strategies and models [24]. Eighty-eight technology articles were subjected to lexical analysis, and a logistic regression model was developed. Lexical analysis was used to determine the specific content of the articles, and logistic regression was used for prediction. The findings consist of a concept map and suggest that Web 2.0 applications are most commonly used in reading, writing, and literacy instruction. The findings also identified two variables (design of information use and design of integration) that can be used to predict whether an author writes about the need for new instructional design strategies to make effective use of Web 2.0 applications.

Lai [25] proposes a feasible appraising structure based on grey theory to predict the trend of IT knowledge items applied in the field of healthcare management (HM). In this study, the Google Scholar search engine is used to collect raw data by keywords of IT knowledge items and HM. The number of search results that include papers and books interpreting knowledge prevalence are evaluated. The forecasting data are produced by grey theory, while the forecasting accuracy is indexed by mean absolute percentage error. The work aims at healthcare managers who might introduce an IT knowledge item according to its development trend and life cycle phase state.

A method for the continuous assessment of major technological advances is presented by the George Washington University (GWU) forecast of emerging technologies [26]. Environmental scanning and trend analysis are used to identify emerging technologies (ETs), and a Delphi-type survey then asks a panel of authorities to estimate the year each advance will occur, its associated probability, the potential size of its market, and the nation that will lead each ET. Eighty-five prominent ETs have been identified and grouped into twelve fields: energy, environment, farming and food, computer hardware, computer software, communications, information services, manufacturing and robotics, materials, medicine, space, and transportation. Results were presented from four survey rounds covering the past eight years and compared longitudinally to estimate the range of variance. The data was also divided into three successive decades to provide scenarios portraying the unfolding waves of innovation that comprise the coming technology revolution.

Andersen [27] quantitatively identifies changes in technological opportunities during the last century. US patent data classified at a very detailed level are used as the source of reference. By analyzing the complexities behind the changing technological opportunities, epochs and typical trajectories are traced empirically. The work analyzes the composition of technological opportunities that have evolved across historical waves. The paper illustrates how technological evolution has become increasingly interrelated and complex and how typical trajectories of individual technologies explain technological evolution better than do conventional aggregate measures. Evidence also suggests how path-dependent technological change is characterized by "creative incremental development".

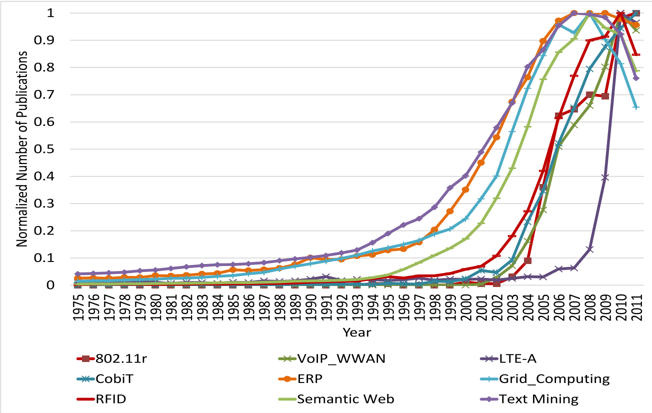


Fig. 1. Technology Example Articles.

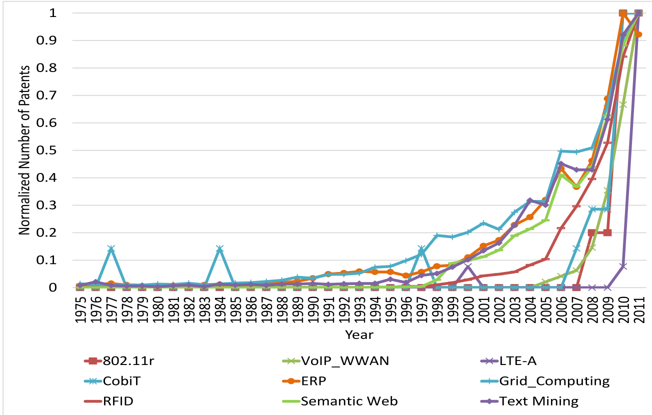


Fig. 2. Technology Example Patents.

However, while prior work focused on predefined technologies in specific domains, it did not attempt to analyze a whole domain market and identify leading technologies. In addition, previous work did not analyze all the leading technologies to understand global market changes that might influence most new technologies. The proposed method allows short term predictions of technologies based on patents and long term predictions based on research articles. The method analyzes domain markets and presents a tool for suggesting relevant leading technologies.

3 Technology Trends Analysis Method

The technology trends analysis method is based on analyzing a large data set of technology-based documents such as patents, research articles, published books, news articles, and Web searches. The data set is assumed to be organized sequentially by date of issue. The method consists of identification of the main terms related to a given technology, extraction of the sequential graph describing the frequency of the terms, followed by elimination of graphs with different behavior, and finally identification of graphs with closest delta distance that represent the cause and effect of the analyzed technology. In addition, the first and second derivatives are evaluated to analyze the magnitude of the effect. The analysis was performed on 4,354,054 patents from the US Patent Office dating from 1975 until today. In addition, an analysis was performed on research articles as a trend prediction of both patents and technology.

3.1 Extracting Related Terms

The first step identifies all the terms related to the technology analyzed. A method to extract the relevant terms can include extracting all the terms that appear in the technology term description in Wikipedia. The technology term list can be extracted automatically using methods such as context recognition [28], and additional terms can be added manually. The list of terms can be revalidated by comparing with Wikipedia terms or a domain specific term list.

The number of terms can be expanded based on extracting all Web pages linked from the current website recursively. This allows an unlimited expansion of the number of terms collected. Possible limitations to the number of terms can be based on a predefined value or the depth of webpages search.

An example of technology growth is displayed in Fig. 1 based on published articles and Fig. 2 based on patents. The method allows identification of technologies that are related to the fast growth. However, not all these technologies can be used as predictive technologies.

3.2 Extracting Relevant Technologies

The second step involves extracting values that represent term frequency in a large data set of documents that can represent the different technologies. An example of such data sets can be patents or academic articles. The term frequency uses simple keyword search in either the subject, abstract, full description of the document, or all of these options. The time slot analyzed usually is a year, since smaller time slots can entail high incidents of seasonal noise. The term frequency has to be weighted, since the extraction searches for an increase in term frequency rather than just elevated values. The weight method analyzed used $\max_j(tf_j)$ value on all technology term frequencies. Other terms such as $(tf_i - tf_{i-1})/\max_j(tf_j)$ were also evaluated. Although text analysis usually uses the method of Term Frequency/Inverse Document Frequency to eliminate irrelevant terms, in our case all terms are considered relevant as possible technology trends since only relevant terms were selected in the previous step.

3.3 Identification Process

The identification process includes identifying all the graphs that represent exponential growth of a new technology. The following types of regression functions were analyzed to identify the best fitting function for technology growth including: linear, quadric, cubic, quadratic, exponential, and mixed functions.

The best matching function to describe fast technology growth was exponential growth. The following generic exponential function (1) was analyzed on a predefined set of existing technologies which were identified by Gartner as "Hype Technologies" at different stages of their life cycle.

$$y = Ae^{Bx}, \quad 0 < A, \quad B \leq 1 \tag{1}$$

Fig. 1 presents a normalized number of academic articles published from 1975 - 2011 of technology terms including: 802.11r, VoIP WWAN, LTE-A, CobiT, ERP, Grid Computing,

TABLE 1
Technology Exponential Growth Correlation

Technology	Articles			Patents		
	y	R^2	p	y	R^2	p
802.11r	$9E\text{-}07e^{0.3847x}$	0.8104	0.0	$5E\text{-}11e^{0.6438x}$	0.8	0.106
VOIP WWAN	$0.0027e^{0.1409x}$	0.5712	0.003	$2E\text{-}11e^{0.6748x}$	0.9909	0.0
LTE-A	$0.0045e^{0.0885x}$	0.63	0.0	$0.0019e^{0.1386x}$	0.3243	0.614
CobiT	$0.0001e^{0.2301x}$	0.8603	0.0	$0.0914e^{0.0419x}$	0.4207	0.082
ERP	$0.0148e^{0.1194x}$	0.9639	0.0	$0.003e^{0.1465x}$	0.9155	0.0
Grid Computing	$0.0101e^{0.1288x}$	0.9784	0.0	$0.0049e^{0.1379x}$	0.9698	0.0
RFID	$0.0008e^{0.1831x}$	0.9336	0.0	$6E\text{-}07e^{0.3954x}$	0.9724	0.0
Semantic Web	$0.0016e^{0.177x}$	0.9472	0.0	$3E\text{-}06e^{0.361x}$	0.8941	0.0
Text Mining	$0.0275e^{0.1008x}$	0.9587	0.0	$0.0023e^{0.147x}$	0.8554	0.0

RFID, Semantic Web, and Text Mining. All these technologies display exponential growth between 2000 and 2009.

Fig. 2 presents the normalized number of patents dating from 1975 to 2011 extracted from the US Patent Office for the similar technology terms. All these technologies also present an exponential growth of technology. However, here the exponential growth appears between the years 2006-2011. Some technologies seem to have sudden peaks over the years, because a similar term is used in a different context. However, these terms are not related to the technology we are seeking and can be viewed as noise.

Table 1 displays the exponential growth function using least squares model fitting of non-zero values of technologies based on academic articles (Fig. 1) and patents (Fig. 2). For all technologies, the coefficient of determination R^2 was calculated as the square of the sample correlation coefficient between the outcomes and their predicted values in the matching function. In addition, p-value for each function is displayed.

Most technologies show high correlation to exponential growth either in published articles or in patents. In addition, comparing the exponential functions to their corresponding figures, we can identify that as A in function 1 is smaller the time of the beginning of the technology trend growth is delayed. In addition, the greater the value of B, the faster the growth of the technology in a short period of time.

3.4 Cause-Effect and Impact of Technologies

Once all the technologies with exponential growth have been identified, the next step includes classifying technologies by cause, effect, and how much impact each technology has. The coefficient of determination is used again to identify the distance between the technology being analyzed and all other technologies. A similar process is used based on predefined Δt time difference. The Δt represents time for one technology to be influenced by the other. If the majority of the data samples are before a specific technology, then the current technology might be a predictive cause of the specific future technology. If the majority of samples are after the technology, then the current technology could be a cause of the new technology, or an effect of the analyzed technology.

3.5 Extent of Technology Growth

The first derivative $y'_i = \Delta y_i / \Delta t$ describes the extent of the growth of a specific technology. Comparison of technologies with similar graph behavior (growth at specific time periods) shows the main technologies that might contribute to a specific trend. For example, analyzing influential technologies based on articles in Fig. 3 shows that technologies such as 802.11r display similar behavior and even precede technologies such as VoIP WWAN and LTE-A. In addition, some similarity can be seen between CobiT and Semantic Web technologies. However, viewing the same influential technologies graph based on patent (Fig. 4) shows more similar behavior between these technologies and less sharp changes in the technologies impact until a later date.

3.6 Change of Technology Trend

The second derivative $y''_i = \Delta(\Delta y_i / \Delta t)/\Delta t$ displays the acceleration of change of technology trends in the market based

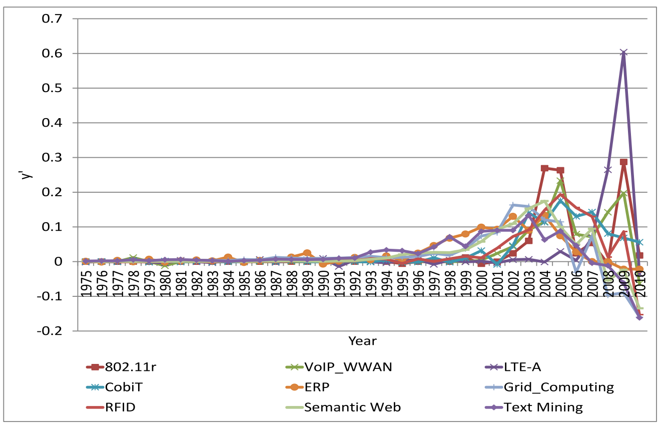


Fig. 3. Influential Technologies—Articles.

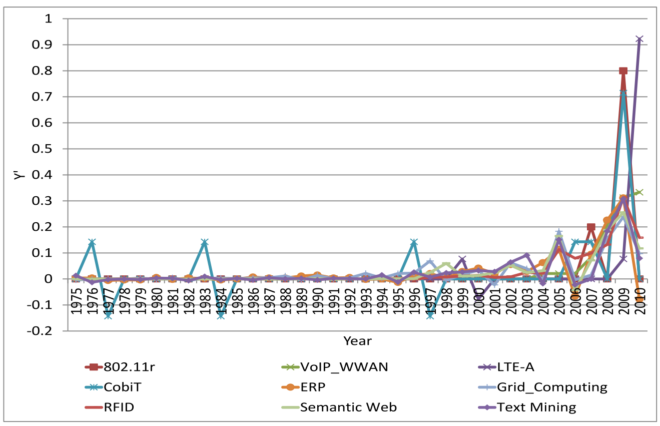


Fig. 4. Influential Technologies—Patents.

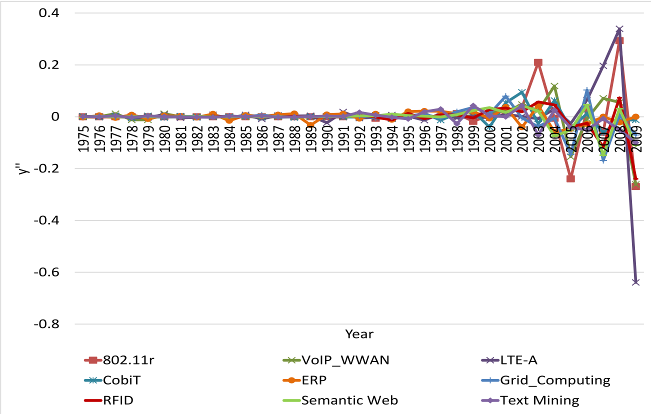


Fig. 5. Acceleration/Decceleration of Technology Trends Articles.

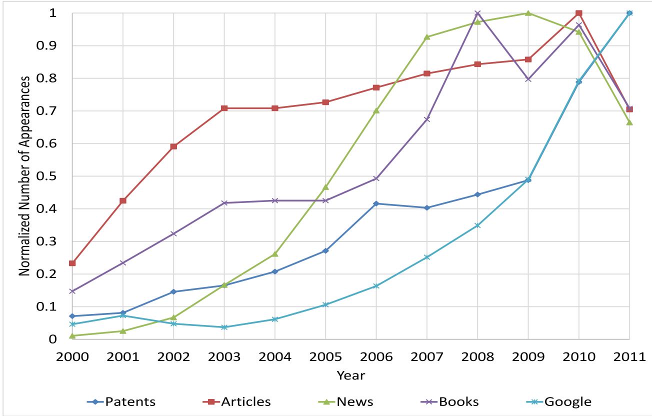


Fig. 7. Technology Example—MPEG-4.

on articles (Fig. 5) and patents (Fig. 6). The results are more visible in the patents and show some relation to economic market changes with emphasis on the timeline of emerging technologies. The results display four major time periods with a drop, where three are related to a drop in stock market performance that might be attributed to technology: 1999 (2000 dot-com bubble), 2007 (2008 financial crisis), and 2011 (August 2011 stock market fall). Although the stock market prediction shows promising results, a few issues should be considered. The year 2005 also shows a drop in technology patents. However, the stock market did not show a similar fall the year after. A more important issue is that the data used were granted patents. In other words, the information is based on patents that were submitted three-four years before to the patent office.

3.7 Academic Articles versus Patents

Based on the assumption that patents can predict technology in the near future of one-two years, it would be interesting to view predictions of patents that will enable an increase in the scope of technology prediction. One of the more intuitive bases for patents is academic articles. Analysis of academic articles according to specific terms and comparison to patents could show a lead time of potential technologies.

Fig. 7 displays a comparison of MPEG-4 technology terms between academic articles, patents, Google search, and news items. All values are normally-distributed between 0 and 1 for comparison. The example displays that between 2000 and 2003 the number of academic articles grew exponentially, compared to the number of patents that grew similarly between 2004 and 2007. Finally, the number of Google searches and the number of news articles grew between 2009 and 2011. The overall comparison between these sets of technologies suggest that the number of academic articles can predict patents by a lead time of up to four-six years and widely used technologies by up to nine years. This assumption is analyzed next in the experiments section.

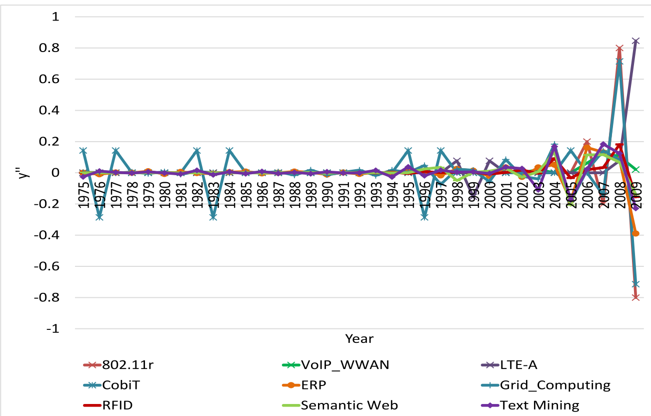


Fig. 6. Acceleration/Decceleration of Technology Trends Patents.

4 Experiments

4.1 Experimental Data

The data for the experiments were taken from an existing technology list identified by Gartner. An example of a Gartner technology list from Gartner Hype Cycles 2012 is presented in Fig. 8. The list is divided into technologies at different stages of development: On the Rise, At the Peak, Sliding into the Trough, Climbing the Slope, and Entering the Plateau.

Gartner Hype Cycles provide a graphic representation of the maturity and adoption of technologies and applications and their potential relevance to solving real business problems and exploiting new opportunities.

On the Rise—Technology Trigger. A potential technology breakthrough kicks things off. Early proof-of-concept stories and media interest trigger significant publicity. Often no usable products exist and commercial viability is unproven.

At the Peak—Peak of Inflated Expectation: Early publicity produces a number of success stories—often accompanied by scores of failures. Some companies take action; many do not.

Sliding into the Trough—Trough of Disillusionment: Interest wanes as experiments and implementations fail to deliver. Producers of the technology improve or fail. Investments continue only if the surviving providers improve their products to the satisfaction of early adopters.

Climbing the Slope—Slope of Enlightenment: More instances of how the technology can benefit the enterprise start to crystallize and become more widely understood. Second- and third-generation products appear from technology providers. More enterprises fund pilots; conservative companies remain cautious.

- On the Rise
  - In-Vehicle Ethernet
  - In-Vehicle Health Monitoring
  - Real-Time Parking
  - Individual Mobility Services
  - Car Connectivity Consortium
  - Autonomous Vehicles
  - Internet Radio
  - Mood Recognition
  - OLED Displays
- At the Peak
  - Mobile Applications in Automobiles
  - Big Data
  - Car-Sharing Services
  - Electro Mobility
  - HTML5
  - Genivi Alliance
  - LBSs in Automotive
  - Vehicle Information Hub
  - Augmented Reality
- Sliding into the Trough
  - Mobile Device Integration into Automobiles
  - Eye Tracking Automotive
  - NFC
  - Cloud Computing
  - Gesture Control
  - Car-to-Infrastructure Communications
  - Automotive HMI Technologies
  - NGTP
  - Car-to-Car Communications
  - Electric Vehicles
- Climbing the Slope
  - Automotive Speech Recognition
  - Long Term Evolution
- Entering the Plateau
  - Bluetooth in Automobiles
  - CRM in Automotive
  - Commercial Telematics

Fig. 8. Gartner Technology List 2012.

Entering the Plateau - Plateau of Productivity: Mainstream adoption starts to take off. Criteria for assessing provider viability are more clearly defined. The technology's broad market applicability and relevance are clearly paying off.

Fifty-eight technologies were selected from the years 2006 - 2012. The technologies were selected according to the following criteria: Technology needs to appear in a Gartner Hype Cycle over a long period of time, idioms were preferred (harder to be mixed up with other keywords). Technologies should appear in only one technological field on the hype cycle (for example, WDM shows different cycles at different fields). Keywords should be unique (unlike RFID, which is associated with many different keywords). A single word keyword was preferred (avoiding the search engine 'OR' option).

4.2 Trend Prediction Data Sources

The method analyzes term frequency on data sets from different sources in each year as described in Section 3.2. The following data sets are used to compare the prediction results:

- *Patents*. The analysis was performed on 4,354,054 patents from the US Patent Office dating from 1975 until today.
- *Articles*. Each term is analyzed with Google Scholar to identify the number of academic articles appearing with each term.
- *News*. Number of news articles retrieved that include the term.
- *Books*. Number of books published that include the search term according to the Amazon search engine.
- *Web search*. Number of results returned using a simple Google search by year with each term.

4.3 Evaluation Comparison Methods

The trend prediction evaluation is based on Java code software which was developed for the experiments and can be accessed as open source (http://kse.kaist.ac.kr/~aviv/Technology_Trends.zip). Evaluation was performed on each of the trend prediction data set graphs. For each technology term the graph was evaluated against each of the other predicting data sets. In addition, each data set was compared to the Gartner technology hype cycle. The graph comparison was performed manually.

A comparison was performed to four other trend prediction methods:

- *Growth/trend curves*. An often used combination of growth curves and a trend curve for some technology. The method is a succession of growth curves, each describing the level of functional capability achieved by a specific technical approach, while an overall trend curve reflects the historical data. The method is implemented by Google Trends evaluating the number of web searches for a specific technology.
- *Extrapolation*. The method performs extrapolation on growth curves to try to predict technological progress based on historical data. The method is implemented based on performing extrapolation to Google Trends growth curves.
- *Analogy*. Combination of forecasts based on the trend curve and one or more analogies. The implementation is based on Google Correlate.
- *Morphological models*. Keyword suggestion method used mainly for advertising based on an enhanced search function that uses a predictive model to display popular search queries in YouTube. The implementation is based on Youtube Keyword Tool.

The Growth/Trend Curves and Extrapolation prediction methods are analyzed using each of the data sets. The Analogy and Morphological methods comparison is based on analyzing how many of the keyword analogies generated are related to the relevant trend topic.

5 Results

Fig. 9 shows the Wireless Power technology change over time, and Fig. 10 displays a similar change for Web Analytics technology. The X-axis represents the years from 2000 to 2011, and the Y-axis represents the number of appearances in logarithmic scale. Both Figs. 9 and 10 display constant increase over the years in all data sets: number of patents, number of academic articles, news items, book publications, and Web searches.

Fig. 11 displays the Wireless Power technology change over time in normal distribution, and Fig. 12 displays results for Web Analytics. The normal distribution is based on dividing the number of term appearances by maximum

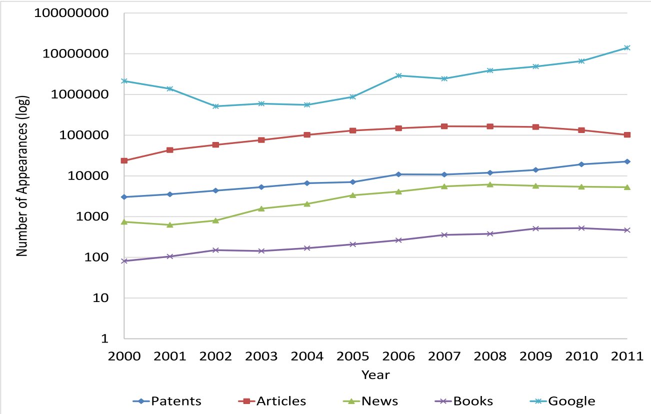


Fig. 9. Wireless Power Technology Timeline.

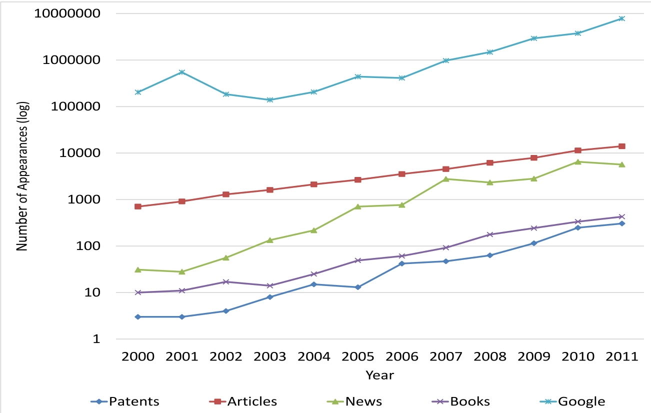


Fig. 10. Web Analytics Technology Timeline.

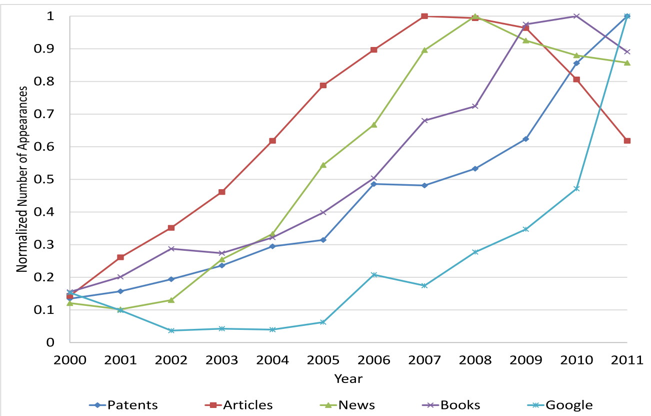


Fig. 11. Wireless Power Technology Normal Distribution.

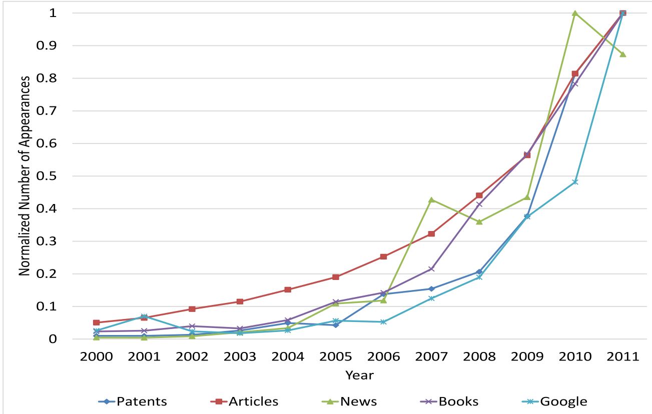


Fig. 12. Web Analytics Technology Normal Distribution.

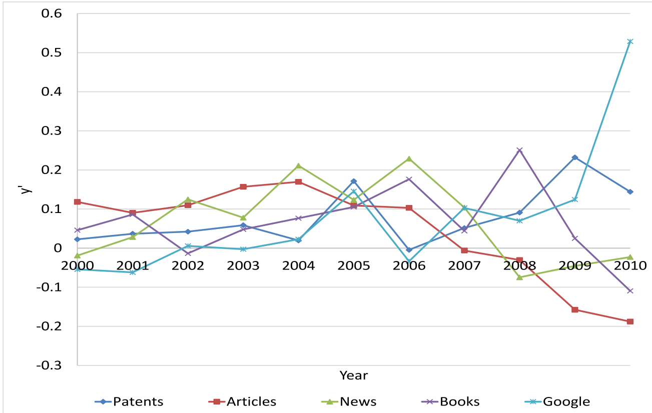


Fig. 13. Wireless Power Technology First Derivative.

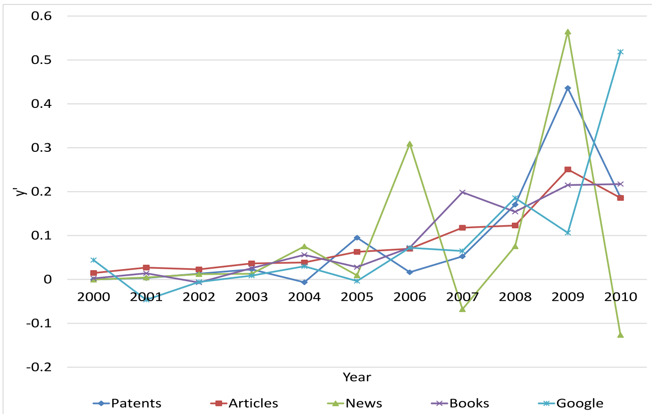


Fig. 14. Web Analytics Technology First Derivative.

value of appearances. Fig. 11 shows that the exponential growth of the academic articles begins in 2001 and ends in 2007, while the exponential growth of the patents begins in 2005 and ends in 2009. The normal distribution term comparison between academic articles and patents shows that academic articles consistently give a lead of two-four years over patents. In addition, in Fig. 12, academic articles display a six year lead over patent publications and a seven year lead over market penetration, as seen by Web searches. News publications on both technology topics do not display direct correlation: in both Fig. 11 the news topics relating to the Wireless Power and in Fig. 12 Web Analytics news reports displayed are not consistent with the behavior of any of the other data sets. The increase of the book publications data set appears to be the average of the increases of the academic articles, patents, and Web searches. When more academic articles are published, shortly after the number of books grows faster.

Fig. 13 presents the first derivative for the Wireless Power technology and Fig. 14 for Web Analytics. In Fig. 13 the results of the method investigating academic articles begin to rise from 2002 until 2004 and the patents from 2007 to 2009. The first derivative allows insight five years into the future, while the direct regular timeline analyzed with

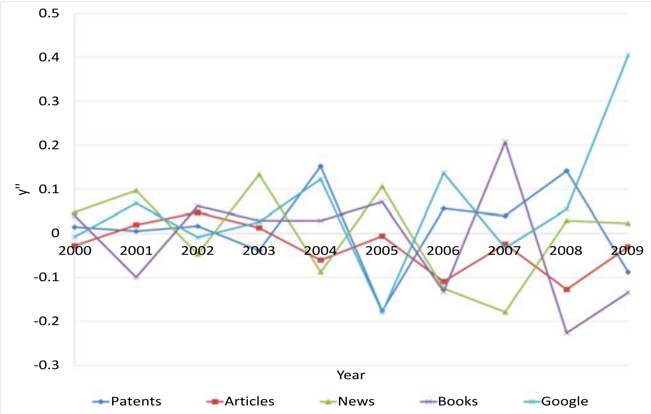


Fig. 15. Wireless Power Technology Second Derivative.

logarithmic scale allows five years and the normal distribution allows two-four years. The first derivative also allows the identification of changes in the term, which could mean a more mature technology in the research community and time to start producing products with the technology. Conversely, in Fig. 14 Web Analytics display a consistently growing research field, while patents display exponential growth since 2007. An interesting issue is that the patents and the Web search results display opposite results: while one increases, the other decreases, and vice versa. This could be attributed to products appearing in the market or advertised a year later based on Web Analytics.

Fig. 15 displays the second derivative for the Wireless Power technology and Fig. 16 for Web Analytics. Looking at a specific technology, we can view what might be a result of market change. Academic articles and patent technology in Wireless Power display results that increase and decrease at opposing times, which could be attributed to market behavior. The opposite behavior between articles and patents throughout shows a year delay, suggesting there is a high correlation between them. In Web Analytics the behavior displayed is more consistent, is less influenced by external influences such as financial crisis, and shows consistent growth.

Table 2 displays the Gartner Hype Cycles forecast for the technologies 802.11r, VOIP WWAN, LTE-A, CobiT, Semantic Web, Grid Computing, and Wireless Power for the years 2006-2012. An arrow going upwards (↗) represents technology *On the Rise*, a parallel arrow (⟶) identifies technology *At the Peak*, an arrow going downwards (↘) is *Sliding into the Trough*, and (×) represents *Off the Hype Cycle*. In comparison to the Wireless Power Technology Timeline displayed in Fig. 9, there is no direct correlation between multiple years. However, the Wireless Power Technology Normal Distribution (Fig. 11) shows direct correlation of the years appearing in the Gartner forecast and is able to predict two years in advance based on publication of articles. Fig. 13, which presents the first derivative of the change between every two years of the technology trend, also shows similarity between Wireless Power academic articles and the Gartner forecast and even predicts a decline.

Summarizing the results of all the analysis, we can see that if we take the Gartner forecast as the "gold standard", then the analysis of patents supplies similar results. In addition, if we look at the forecast based on academic publication, then we can have a prediction of up to four or five years in advance.

Table 3 displays the percent of successful forecast of each of the method data sets—article publications, patents, news, book publications, and Web searches—against each other and against the Gartner forecast. Each column presents the data set analyzed, and each row presents the success of predicting the other data sets. The table presents the highest results of prediction for all data sets identifying the Gartner forecast, implying that the Gartner forecast is not unique. As expected, the data set with the lowest ability to predict is the research topics. The interesting result is the ability of Gartner to predict with low accuracy 10 percent of research topics. However, this could be attributed to ongoing research fields that are still "hot" over many years. The highest overall performance is attributed to academic research as an earliest indicator of expected technology. However, to analyze best prediction, it is necessary to evaluate the prediction time. For example, the prediction time of

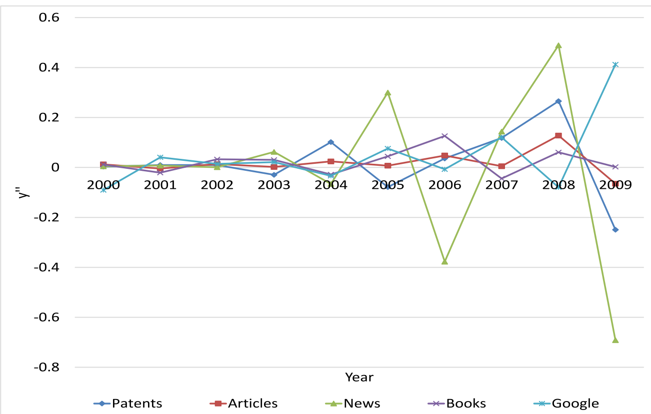


Fig. 16. Web Analytics Technology Second Derivative.

TABLE 2
Gartner Forecast

Technology	2006	2007	2008	2009	2010	2011	2012
802.11r	↗	⟶	⟶	↘	↘	↘	↘
VOIP WWAN	↗	↗	↗	⟶	⟶	⟶	⟶
LTE-A			↗	↗	↗	↗	↗
CobiT	↗	↗	↗	⟶	⟶	⟶	⟶
Semantic Web	↗	↗	↗	↗	↗	↗	↗
Grid Computing	↗	↘	↘	↗	⟶	×	
Quantum Computing		↗	↗	↗	↗	↗	↗
Web Analytics	↗	↗	↗	↗	↗	↗	⟶
Wireless Power				⟶	⟶	⟶	⟶

TABLE 3
Forecast Accuracy Comparison

Forecaster	Patent	Articles	News	Google	Books	Gartner
Patent	-	83%	62%	31%	69%	45%
Articles	9%	-	17%	12%	24%	10%
News	16%	59%	-	17%	48%	17%
Google	47%	86%	72%	-	78%	10%
Books	43%	67%	59%	28%	-	24%
Gartner	48%	83%	71%	84%	71%	-

TABLE 4
Average Forecast Time (Years)

Forecaster	Patent	Articles	News	Google	Books	Gartner
Patent	-	3.02	2.03	2.67	2.13	1.58
Articles	2	-	1.7	2.86	1.93	1.67
News	2	2.62	-	1.8	1.71	1.4
Google	2.3	3.46	2.31	-	2.24	1
Books	1.96	3.03	2.21	2.75	-	1.29
Gartner	1.86	4.27	2.34	2.69	2.34	-

academic articles predicting Web searches, which is the highest at 86 percent, might not be equal to the prediction time of the Gartner technology, 83 percent, because technology prediction can have a longer time horizon than Web search, which represents user search interests. Therefore, viewing the results with another dimension of time displays how successful the results are and how far into the future the method can predict.

Table 4 displays the time dimension of the prediction for each of the data sets. Similarly, each column presents the data set analyzed, and each row presents the number of years into the future that this data set predicts the other data sets. Again, the article publications data set supplies an average forecast of 4.27 years of future technologies before the Gartner prediction. Based on the assumptions that the Gartner forecast predicts at least one year in advance, the published articles supply an overall prediction of technology of more than five years. The consistency of the results can be seen from the combination of academic articles predicting patents and the patents predicting Gartner, and this combination is similar to just the academic articles predicting Gartner. Another interesting observation is that news, Web searches, and books all supply a prediction of at least two years ahead of the Garner technology list. This could be attributed to development time required to create products from technology that has reached maturity. Furthermore, the relatively lower time periods predicted by patents could suggest that patents are used more as a tool to limit competitors than as a tool to develop technology. Otherwise, the patent would be expected to display a longer prediction horizon than mass media knowledge represented by news, books, and Web searches.

Fig. 17 compares the accuracy of the results to the prediction horizon for each of the data sets: patents, academic articles, news, books, Web searches, and Gartner predictions. Each data set is compared to all other data sets. The X-axis displays the number of years into the future of the prediction. The Y-axis displays the percent of success for each year. The comparison only evaluates successful predictions and evaluates the dispersion of the total successful results over the years. Therefore, the evaluation identifies how far into the future the peak of each data source is. The best results can be attributed to the academic articles, for which 12.5 percent of the predictions are as far as eight years in the future and 4.2 percent for nine years of the Gartner list. The second best results are attained by the news, 4.9 percent, seven years in the future, and Web search, 2 percent, eight years in the future. Both the news and the Web search results can be attributed to the academic articles results, and thus the one year delay. The peak that is highest and furthest in years is that of the research articles, which predict three years for most of the other data sets. Other data sets peak at two years, showing the average predictive advantage of the academic articles over other data sets.

Table 5 and 6 present forecasting comparison with other methods. Table 5 shows that predicting based on articles outperforms existing methods based on combination of growth curves and trend curve or growth curves and extrapolation. The proposed method outperforms for each of the data sets. The difference between the prediction accuracy of the different methods decreases for the data set of news items and can be attributed to shorter forecasting time periods and information already available on the Web. The difference increases for the longer technology trends forecasting time periods appearing in patents or reported by Gartner. The lowest performance values are achieved by predicting the existing Web pages technology topics (Google) based on the current number of Web searches on specific technology terms using growth/trend curves or extrapolation. For the Web search pages the proposed method achieves the highest value based on articles and has the biggest difference compared to the other two methods. The difference can be attributed to reasons such as: similar syntax and different semantic meaning are used in Web searches, Web searches are usually performed on past events and not possible future events, and large volume of Web searches compared to article publications makes it more difficult to identify small and unique changes.

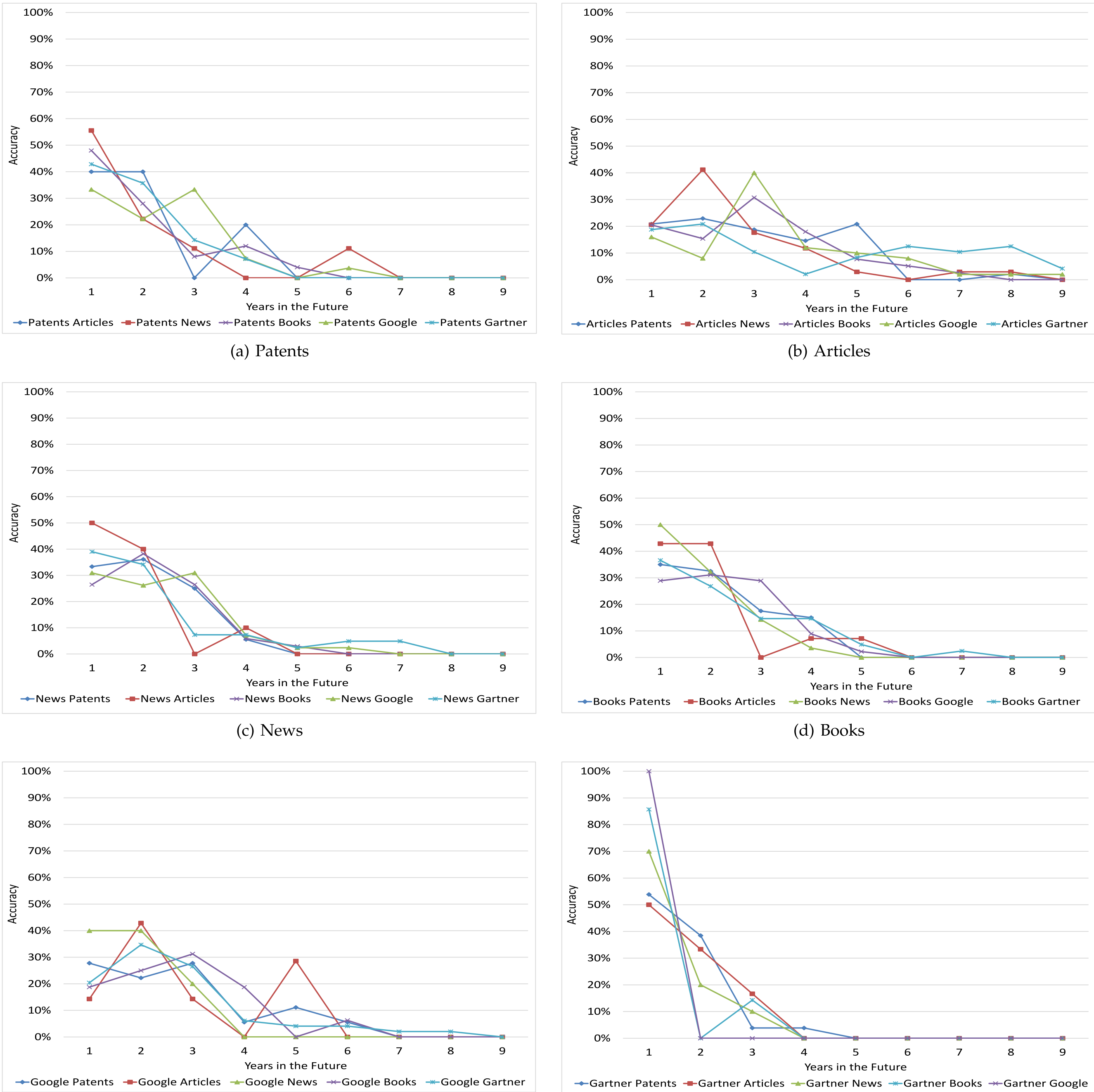
Table 6 compares another two methods: keyword extraction by analogy and keyword extraction morphological models. These methods do not compare the prediction based on the data sets since the methods themselves do not have time dimensions. Both methods show overall low numbers of accurate related extracted words. Integrating the analogy keywords set or the morphological models keywords set with growth curves methods for trend predictions would yield even lower prediction results.

6 Discussion

One of the issues that influence the results is the years where there are no patents, after a surge in the number of patents the year before. In such a case, it is hard to analyze whether the technology is not ready to go to market or the life span of the technology is relatively short. In addition, patents are used many times to achieve marketing advantage and not necessarily technological advantage in a specific area.

Another issue that influences the results is technologies that have a surge in the number of articles or patents toward the end of the time line analyzed. For example, the topic of gamification only displays sudden growth in the years 2011-2012, and thus it is hard to analyze whether this is a trend that presents extremely fast growth or just an extension of another trend using different terms.

A comparison of the analysis method data sets shows that the normalized values display the easiest way to compare the results. The logarithmic-based method displays parallel lines in many of the results, while the normalized method identifies the exponential growth. The first derivative helps identify the magnitude of the change in the trend but is too sensitive to minor trend changes. The second



(a) Patents

(b) Articles

(c) News

(d) Books

(e) Google

(f) Gartner

Fig. 17. Overall Results—Accuracy Versus Time.

derivative helps identify overall market effects but requires a comparison of multiple trends.

The complexity of the trend extraction is $n \cdot t$, where n is the number of trends analyzed and t is the number of years. The main consideration is the limitations set by service providers, such as Google, which restricts the number of service requests per user. Using a time delay to match the service limitation, each trend can be extracted for each of the five data sets in fifteen minutes on a standard computer.

7 Conclusion

The work proposes a method of predicting technology trends based on term frequency. The results showed that more research-oriented data sets such as academic articles and patents have a longer predictive time window and higher accuracy than do data sets based on news, books, and Web searches. The results for the method based on the data source of academic articles displayed very high accuracy in technology prediction and up to five years forecast. The patents-based data source displays lower accuracy and a shorter time forecast. The method implemented based on published books provides an intermediate result compared to the method implemented on academic articles and patents. The news and search results are less successful with the technology prediction.

Areas of further research include analysis of technology prediction in multiple languages and expansion of the time line of the predictions of technologies that last for more

TABLE 5
Forecasting Comparison with Other Methods

Forecaster	Growth/Trend Curves	Articles	Extrapolation
Patent	41%	83%	31%
Articles	45%	-	45%
News	55%	59%	51%
Google	28%	86%	22%
Books	43%	67%	41%
Gartner	45%	83%	47%

TABLE 6
Relevant Keywords Comparison with Other Methods

Forecaster	Analogy	Morphological Models
Average Keyword Accuracy	29.6%	41.9%
Overall Accuracy	30.8%	36.2%

than a decade. Additional research can be used to emphasize market performance versus specific technology trends displayed in this research. Another interesting approach would be analysis of results in areas such as pharmaceuticals and medicine.

ACKNOWLEDGMENTS

This work was supported by the IT R&D program of MSIP/KEIT. [10044494, WiseKB: Big data based self-evolving knowledge base and reasoning platform]

REFERENCES

[1] S. Makridakis, S. C. Wheelwright, and R. J. Hyndman, *Forecasting: Methods and Applications*, 3rd ed. Hoboken, NJ: Wiley, 1998.
[2] N. Dalkey and O. Helmer, "An experimental application of the Delphi method to the use of experts," *Manage. Sci.*, vol. 9, no. 3, pp. 458–467, April 1963.
[3] K. C. Green and J. S. Armstrong, "Structured analogies for forecasting," *Int. J. Forecasting*, vol. 23, no. 3, pp. 365–376, Jul.–Sep. 2007.
[4] N. Meade and T. Islam, "Forecasting with growth curves: An empirical comparison," *Int. J. Forecasting*, vol. 11, no. 2, pp. 199–215, Jun. 1995.
[5] J. S. Armstrong, "Forecasting by extrapolation: Conclusions from twenty-five years of research," *Interfaces*, vol. 14, pp. 52–66, 1984.
[6] N. N. Atanackov and J. E. Boylan, "Decision trees for forecasting trended demand," in *Service Parts Management*, N. Altay and L. A. Litteral, Eds., London. U.K.: Springer, 2011, pp. 53–87.
[7] F. Zwicky, *Discovery, Invention, Research-Through the Morphological Approach*. Toronto, ON, Canada: Macmillian, 1969.
[8] J. P. Martino, *Technological Forecasting for Decision Making*, 3rd ed. New York, NY, USA: McGraw-Hill, 1993.
[9] J. S. Armstrong and F. Collopy, "Error measures for generalizing about forecasting methods: Empirical comparisons," *Int. J. Forecasting*, vol. 8, pp. 69–80, 1992.
[10] M. J. Morris, "Forecasting the sunspot cycle," *J. Royal Statist. Soc.: Series A*, vol. 140, no. 4, pp. 437–468, 1977.
[11] R. Saga, M. Terachi, and H. Tsuji, "FACT-Graph: Trend visualization by frequency and co-occurrence," *Electron. Commun. Japan*, vol. 95, no. 2, pp. 50–58, 2012.
[12] F. Gey, R. Larson, J. Machado, and M. Yoshio, "NTCIR9-GeoTime overview-Evaluating geographic and temporal search: Round 2," in *Proc. NTCIR-9 Workshop*, 2011, pp. 9–17.
[13] L. Tang, S. Geva, A. Trotman, Y. Xu, and K. Itakura, "Overview of the NTCIR-9 crosslink task: Cross-lingual link discovery," in *Proc. NTCIR-9 Workshop*, 2011, pp. 437–463.
[14] I. Goto, B. Lu, K. Chow, E. Sumita, and B. Tsou, "Overview of the patent machine translation task at the NTCIR-9 workshop," in *Proc. NTCIR-9 Workshop*, 2011, pp. 559–578.
[15] G. Roda, J. Tait, F. Piroi, and V. Zenz, "CLEF-IP 2009: Retrieval experiments in the intellectual property domain," in *Proc. 10th Workshop Cross-Lang. Eval. Forum*, 2010, pp. 385–409.
[16] M. Lupu, J. Huang, J. Zhu, and J. Tait, "TREC-CHEM: Large scale chemical information retrieval evaluation at TREC," *SIGIR Forum*, vol. 43, no. 2, pp. 63–70, 2009.
[17] A. W. Blackman, "A mathematical model for trend forecasts," *Technol. Forecasting Social Change*, vol. 3, pp. 441–452, 1972.
[18] T. Dereli and A. Durmusoglu, "A trend-based patent alert system for technology watch," *J. Sci. Ind. Res.*, vol. 68, pp. 674–679, 2009.
[19] T. Dereli and A. Durmusoglu, "Application of possibilistic fuzzy regression for technology watch," *J. Intell. Fuzzy Syst.*, vol. 21, no. 5, pp. 353–363, 2010.
[20] B. Yoon and Y. Park, "A text-mining-based patent network: Analytical tool for high-technology trend," *J. High Technol. Manage. Res.*, vol. 15, pp. 37–50, 2004.
[21] A. Segev, J. Kantola, C. Jung, and J. Lee, "Analyzing multilingual knowledge innovation in patents," *Expert Syst. Appl.*, vol. 40, no. 17, pp. 7010–7023, 2013.
[22] H. J. No and Y. Park, "Trajectory patterns of technology fusion: Trend analysis and taxonomical grouping in nanobiotechnology," *Technol. Forecasting Social Change*, vol. 77, pp. 63–75, 2010.
[23] J. Kim, M. Hwang, D. H. Jeong, and H. Jung, "Technology trends analysis and forecasting application based on decision tree and statistical feature analysis," *Expert Syst. Appl.*, vol. 39, pp. 12618–12625, 2012.
[24] L. Liu and C. D. Maddux, "Web 2.0 articles: Content analysis and a statistical model to predict recognition of the need for new instructional design strategies," *Comput. Schools*, vol. 25, no. 3–4, pp. 314–328, 2008.
[25] H. F. Lai, "Predict the trend of information technology knowledge applied in healthcare management," in *Proc. IEEE Int. Conf. Ind. Eng. Eng. Manage.*, 2009, pp. 1885–1889.
[26] W. E. Halal, M. D. Kull, and A. Leffmann, "The George Washington University forecast of emerging technologies-A continuous assessment of the technology revolution," *Technol. Forecasting Soc. Change*, vol. 59, pp. 89–110, 1998.
[27] B. Andersen, "The evolution of technological trajectories 1890-1990," *Structural Change and Economic Dynamics*, vol. 9, pp. 5–34, 1998.
[28] A. Segev, "Circular context-based semantic matching to identify web service composition," in *Proc. WWW Int. Workshop Context Enabled Source Serv. Sel., Integration Adaptation*, 2008.

Aviv Segev received the PhD degree in technology and information systems from Tel-Aviv University in 2004. He is currently an associate professor at the Knowledge Service Engineering Department, Korea Advanced Institute of Science and Technology (KAIST). His research interests include classifying knowledge using the web, context recognition and ontologies, knowledge mapping, and implementations of these areas in the fields of web services, medicine, and crisis management. He is the author of more than 50 publications. He is a member of the IEEE and the ACM.

Sukhwan Jung received the BSc degree in computer science from the University of Canterbury and the master's degree in knowledge service engineering from the Korea Advanced Institute of Science and Technology (KAIST). He is currently working toward the PhD degree at the Department of Knowledge Service Engineering, KAIST. His research interests include human computer interaction, knowledge discovery, and community detection.

Seungwoo Choi received the BS degree in computer science and engineering in 2013, from Sejong University, Korea. He is currently working toward the MS degree at the School of Knowledge Service Engineering, Korea Advanced Institute of Science and Technology (KAIST). His current research interests include information retrieval and knowledge engineering, and social computing systems.