NORTH-HOLLAND

Technological Forecasting & Social Change
69 (2002) 495–506

Technological Forecasting and Social Change

Automated extraction and visualization of information for technological intelligence and forecasting

Donghua Zhu[a], Alan L. Porter[b,c,*]

[a]Institute of Forecasting and Development, Hefei University of Technology, Hefei 230009, Anhui, China
[b]Industrial and Systems Engineering and Public Policy, Georgia Institute of Technology, Atlanta, GA 30332-0205, USA
[c]Technology Policy and Assessment Center, Georgia Institute of Technology, Atlanta, GA 30332-0205, USA

Received 19 June 2001; accepted 5 July 2001

Abstract

Empirical technology forecasting (TF) is not well utilized in technology management. Three factors could enhance managerial utilization: capability to exploit huge volumes of available information, ways to do so very quickly, and informative representations that help manage emerging technologies. This paper reports on efforts to address these three factors via partially automated processes to generate helpful knowledge from text quickly and graphically. We first illustrate a process to generate a family of technology maps that help convey emphases, players, and patterns in the development of a target technology. Second, we exemplify the generation of particular "innovation indicators" that measure particular facets of R&D activity to relate these to technological maturation, contextual influences, and market potential. Both technology mapping and innovation indicators rely upon searches in huge, easily accessible, abstract databases and text mining software. We augment these through "macros" (programming scripts) that automatically sequence the necessary steps to generate particular desired information products. These analytical findings can be tailored to the needs of particular technology managers. © 2002 Elsevier Science Inc. All rights reserved.

Keywords: Competitive technological intelligence; Technology forecasting; Text mining; Innovation indicators; Technology maps

* Corresponding author. Industrial and Systems Engineering and Public Policy, Georgia Institute of Technology, Atlanta, GA 30332-0205, USA. Tel.: +1-404-894-2330; fax: +1-404-894-2301.

E-mail addresses: zhudh111@mail.hf.ah.cn (D. Zhu), alan.porter@isye.gatech.edu (A.L. Porter).

PII: S0040-1625(01)00157-3

1. The challenge

Technology forecasting (TF) activities have rebounded energetically in the 1990s with various emphases—competitive technological intelligence, technology foresight, and technology roadmapping. TF is expanding its tool set to synthesize findings drawn from a range of quantitative and qualitative approaches [1]. Efforts to systematize the TF process hold promise [2].

Tough economic competition is today's primary driver of technological innovation and, hence, the key motivator to conduct TF. Large companies need TF, in its various guises, to prioritize R&D, plan new product development, and make strategic decisions on technology licensing, joint ventures, and so forth. Small companies often depend on technological innovation for their existence, yet have been notably weak at TF. Small firms have been limited by time pressures, lack of information resources, and unfamiliarity with methods. Government agencies also have TF needs as they seek to advance public agendas in the face of increasing rates of technological changes with constrained budgets.

Forecasting a technology's future draws on empirical evidence and expert opinion. We focus on empirical analyses in this paper, recognizing that expert opinion provides essential complementary information. Forecasters have long had complex algorithmic approaches at their disposal, but their ability to effectively execute those approaches has been limited by the availability of information and costs of manual information manipulation and analysis. The situation is changing.

Empirical analysis of emerging technologies poses a number of challenges to analysts. In particular, we note the need to:

1. digest enormous amounts of available information,
2. do so rapidly,
3. present findings vividly and understandably.

Let us consider each of these three challenges in turn.

The defining characteristic of the "information economy" is tremendously enhanced access to information. This offers particular promise to improve TF. In addition to Internet technology information sites worth mining, there are superb database resources for R&D information, including:

- projects (e.g., NSF projects—http://www.nsf.gov)
- research opportunities (e.g., http://www.cos.com)
- publications (e.g., Chemical Abstracts—http://www.cas.org/SCIFINDER/SCHOLAR)
- citations (Social Science Citation Index and Science Citation Index—http://www.isinet.com/isi)
- patents (e.g., Derwent World Patents Index http://www.derwent.com/worldpatentsindex)

Tremendous research and development activity, worldwide, results in explosive growth in the amount of scientific and engineering literature. For instance, *Science Citation Index* contains almost 15 million abstracts of journal and conference papers published since 1987;

MEDLINE contains some 12 million medical research abstracts. In engineering domains, *EI Compendex* includes over 5 million records and *INSPEC* nearly 7 million since 1970. US, Japanese, and European patents are searchable online. Gateway services, such as *Dialog*, link to numerous databases through a standard interface, for ready access. Importantly, many organizations license diverse R&D databases for unlimited searching, e.g., universities for their students and faculty. A major portion of the world's R&D is literally "at our fingertips."

An R&D information specialist put this challenge in dramatic terms to us. A search on a particular topic in the late 1980s yielded about a dozen pertinent papers; a similar search in the late 1990s yielded over a hundred; she anticipates this heading toward a thousand in a very few years. This challenges all those who would manage technology—from researchers to industrialists—to grasp advances in their own domains and in work potentially relating to their areas. It mandates that we all apply the emerging analytical tools to "mine" this information to inform our analyses.

Our experience has elucidated the diversity of user needs for TF information in various forms, with different time and resource constraints. However, one of the dominant themes—and our second key challenge—is that information is usually needed *fast*. In a small survey of technology professionals and managers, two of three reported they typically needed technology information in a week or less to inform decision making [3]. In contrast, most TF has taken months to generate. A primary objective of our development is to semiautomate analyses so as to generate findings rapidly.

A third hard-earned lesson gained from our developmental experiences with "bibliometrics" (counting bibliographic activity) and "text mining" has been that TF-related results must be easily understood and must directly relate to a user's perceived information needs. A communication from Theresa Gow of DERA (UK) gives the flavor of these concerns:

> However, when trying to present the analyses to 'non-users,' I am aware that many do have difficulty accepting KDD 'knowledge' and would prefer to have experts providing the knowledge. [KDD refers to "knowledge discovery in databases."]

A private sector user of VantagePoint noted the difficulty in learning to devise usable TF findings to which her management could relate. This is counterbalanced by the realization that if you don't use such tools, you risk that your competition will, thereby accruing a competitive advantage [4].

We have found the development of useful TF information extremely challenging. With support from the National Science Foundation (NSF),[1] we have worked with five partner organizations (two companies, two government agencies, one industrial technology consortium) and surveyed others to learn what information is valued. In the process, we have identified a number of critical factors, including [5]:

- user involvement in the formulation of an inquiry and the analytical process,
- the need for "hooks" (vivid representations of key findings),

[1] NSF—Management of Technological Innovation Program Sponsored Project (DMI-9872482), 1998–2000; extended to 2001.

- the nature of technological change under scrutiny (normal, transitional, transformational),
- focus just as the user prefers (proper blend of technological, contextual, and business aspects),
- credibility (of the analyst, methods used, data employed),
- communications (right amount of information in clear and accessible form).

In sum, then, we seek to respond to these challenges—analyzing large text resources, rapidly, to generate compelling findings—to enhance TF (including competitive technological intelligence, technology foresight, etc.). Our approach, called technology opportunities analysis (TOA), seeks to facilitate this process by profiling search sets of bibliographic abstracts on technologies of interest.

2. Technology opportunities analysis (TOA)

TOA has been under development at Georgia Tech since 1990 [6,7]. The premise is that useful information on the prospects of particular technological innovations can be extracted from abstracts collected by searching on the given topic in suitable publication, patent, citation, and/or project databases. That information extraction is enabled by software. Software development was initiated in 1993 and continues under the lead of Search Technology, in collaboration with Georgia Tech, with major support of the US Army Tank-automotive and Armaments Command (TACOM) and the Defense Advanced Research Projects Agency (DARPA). It is commercially available as VantagePoint since 2000 (http://www.theVantagePoint.com).

The TOA process entails these main steps:

1. Search and retrieve text information, typically from large abstract databases.
2. Profile the resulting search set. VantagePoint applies a combination of machine learning, statistics, and natural language processing to yield what bibliometricians call a mix of "one-dimensional" descriptions (lists) and "two-dimensional" relationships (matrices) [8]. Profiling may focus on documents (e.g., "bucketing" documents into related, manageable groups; cf., Refs. [9,10]). Or, it may focus on concepts (e.g., principal components analysis (PCA) to group related terms as conceptual clusters; cf., Refs. [11,12]). A third choice is a combination—seeking to link documents to concepts (e.g., relevance scoring [13]). Conceptual distinctions and methods are discussed further elsewhere [14].
3. Extract latent relationships. VantagePoint applies iterative principal components analyses to uncover links among terms and underlying concepts (cf., examples on the website: http://tpac.gatech.edu [9–12,15]).
4. Represent relationships graphically. Generation of "mapping" and "indicators" are elaborated in the following sections [16].
5. Interpret the prospects for successful technological development. This typically entails integrating the bibliographic search set analyses with expert domain knowledge (interviews) (cf., Ref. [11]; companion workshop paper on "ROI").

We seek knowledge from a "body" of literature beyond that obtainable by digesting individual pieces. We *treat retrieved text as data* [17] to parse text into informative units, count those units, and uncover patterns that can speak to TF interests. Work on text mining is extremely active. For our purposes, this draws on efforts under several labels, including KDD (cf., www.cs.cmu.edu/~dunja/WshKDD2000.html; www.cs.biu.ac.il/~feldman/ijcai-workshop%20cfp.html) and bibliometrics (counting of bibliographic activity—cf., sistm.web.unsw.edu.au/conference/issi2001).

We seek empirical measures to help gauge development progress and prospects. In particular, we focus on three sets of "innovation indicators" to get at life cycle status, contextual influences (support factors), and market prospects [12,18]. A keen objective in "text mining for TF" [19] is to develop an automated sequence of steps that generates such indicators rapidly (i.e., in one day). We believe such a process can dramatically improve the ability of companies and other organizations to forecast the progress of technology to improve their ability to manage emerging technologies

The following sections focus on two types of TOA-based knowledge representations—technology maps and innovation indicators. They draw upon illustrative cases, based on simple search in the *INSPEC* database. *INSPEC* is a widely available R&D publication database abstracting some 300,000 journal articles and conference papers from select technical domains annually. It is produced by IEE and available various ways (e.g., through "Dialog" or by subscription). A "nanotechnology" search provided 3552 abstract records. A thorough analysis of this topic would certainly warrant more extensive review of nanotechnology R&D, as well as expert perspectives.

3. Mapping

As noted, we seek to identify and represent relationships inherent in sets of abstracts resulting from a database search. This inductive approach does not impose groupings, but instead elicits them from the data. We have developed a partly automated process to do so based on "co-occurrence" information. Co-occurrence is based on the pattern of terms occurring together in the records. If two terms occur together in the records more frequently than expected, there is a presumption of relationship between them. Terms can include authorship (also organizational affiliation, nationality) or "keywords" (subject index terms), or noun phrases generated from titles or abstracts using our natural language processing (NLP) routine (cf., Refs. [18,20]).

Such relationships can be straightforward—for instance, showing which topics particular organizations mention most frequently in their writing—the basis of Fig. 1. Relationships can also be more subtle. PCA is a useful technique for extracting the main relationships implicit in a data set. A PCA-based approach called latent semantic indexing (LSI; cf. Refs. [14,21]) generates conceptual indices instead of individual words to improve information retrieval. LSI is based on co-occurrence information from large text sources, such as collections of abstract records. Interesting issues concern the use of grouping techniques such as PCA and LSI, but those are not central to our focus here (cf., Refs. [8–10,12,15,20])—which concerns how to represent relationships quickly and effectively.

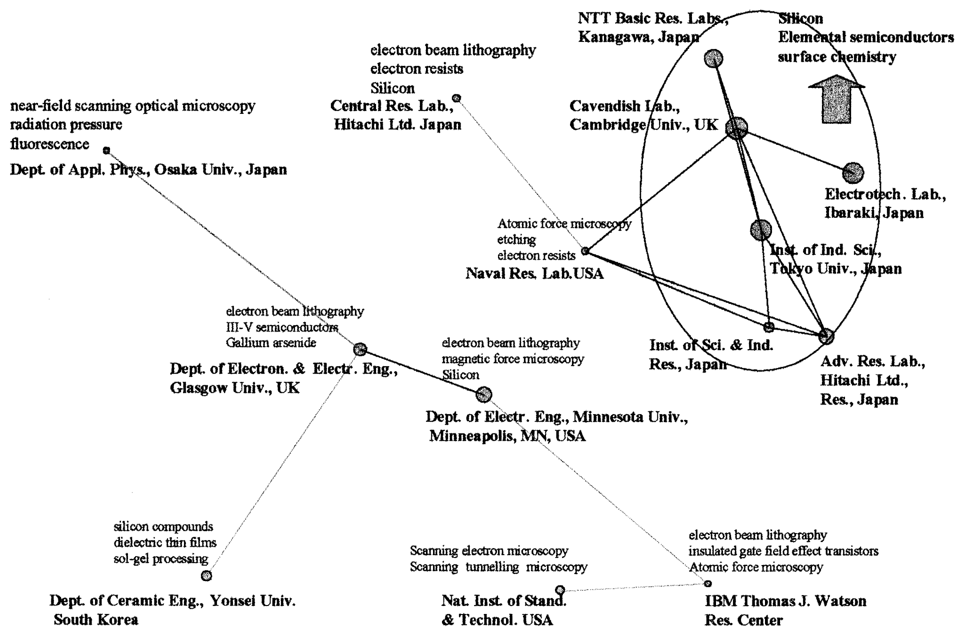

Fig. 1. Nanotechnology—affiliations map.

Effective visualization of the basic co-occurrence and correlation matrix information entails a sequence of analyses:

- a new two-step multidimensional scaling (MDS) algorithm,
- an improved path-erasing algorithm,
- a routine to determine and display size (relative frequency of occurrence),
- macros to create maps in VantagePoint, Microsoft Word or MS PowerPoint,
- a routine to consolidate duplicate principal components (in the mapping process),
- an algorithm to automatically name principal components,
- an algorithm to cut off principal components to just include high-loading terms (the last three steps are needed for principal components maps; cf., Refs. [16,18]),

Our routine generates various maps, such as:

1. principal components map [represents the relationships among conceptual clusters];
2. keywords map [represents the relationships among frequently occurring subject index terms, title phrases, or whatever terms are chosen];
3. affiliations map [represents the relationships of affiliations' research topics, based on terms they use in their documents—see Fig. 1];
4. authors map [analogous to affiliations map, but for individual researchers];
5. countries map [analogous to affiliations map];
6. sources (e.g., journals) map [analogous to affiliations map].

Fig. 1 shows an affiliations (organizations) map for the "Nanotechnology" topic. Displayed are the most prolific publishers abstracted in *INSPEC* for 1998. Along with the organizational name are shown the three keywords most frequently used in its publications in the search set. The size of a node reflects the number of publications. Positioning is determined using our MDS and path-erasing algorithm

Maps such as Fig. 1 are built from a similarity matrix. In essence, the challenge is to reduce n-dimensional (in this case, n equates to 40-dimensional since there are some 40 affiliations' similarity being represented) to 2-D or 3-D. MDS is the generally favored approach to accomplish this. In MDS, an important parameter called stress is used to control its procedures. The process of generating a MDS map seeks the optimum location for each element in the map by minimizing the stress. Traditionally, the "steepest descent" algorithm is employed in most MDS applications (e.g., SPSS uses this). We have found that the "steepest descent" algorithm is not very effective in a number of text mapping cases, especially for 3-D solutions. The algorithm can often be trapped in a local minimum of "stress space," never reaching the global minimum.

We have devised a "step-by-step" search algorithm. This algorithm is effective at finding the global stress minimum, although it usually consumes more CPU time than the "steepest descent" algorithm. In our mapping algorithm, we bind the two "step-by-step" and "steepest descent" algorithms in the MDS iteration process. Comparison of a number of cases shows our MDS solution to yield substantially better visual representations than other MDS solutions.

As noted, MDS tries to represent high-dimensional spatial relations by displaying the elements in 2-D or 3-D spaces. The resulting distortions tend to become problematic when many elements (hence many dimensions) are involved. Therefore, we have added an additional representational element, connecting links, based on a "path-erasing" algorithm. This is built on a proximity matrix among the elements (in Fig. 1, the affiliations). Its logic is as follows:

1. connect all elements in the proximity matrix together,
2. set a series of thresholds to erase the connecting lines one by one,
3. devise a suitable stop criterion.

Our experience is that the links are easily perceived as dominant proximity representations, with the MDS-based location taken as secondary. The MDS axes are essentially arbitrary so our routine provides four alternative axial perspectives for 3-D views. We find the resulting representations superior to others in capturing the conceptual entities and visualizing them (cf., Ref. [22]). However, we note that the representations are not singular—different numbers of entities mapped and different views give quite different results. We are presently conducting a series of experiments to devise preferred algorithms in terms of conceptual clarity (e.g., how many principal components to extract from how many terms to achieve robust results?) and visual clarity (e.g., how many organizations to include in Fig. 1?).

In viewing Fig. 1, note that commonality of interests should reflect in connections. For instance, at the right are a group of academic departments and libraries with shared

Silicon, Elemental semiconductors, and Surface chemistry interests (as estimated by keyword usage in their publications). Other nanotechnology maps (principal components, terms, authors, countries, and sources) are displayed on our website: http://tpac.gatech.edu/~donghua/nanotechnology.

Fig. 1 is badly cluttered. VantagePoint provides a cleaner, interactive version just showing the nodes and names, but allowing the user to "pull down" other desired information concerning a node. For instance, one might want to focus on the IBM node (lower right, Fig. 1). One could pull down short lists of the leading IBM authors, the dates of their publications, or even read their full abstract records (http://www.theVantagePoint.com).

4. A composite indicator

The maps just discussed represent co-occurrence and correlative information gathered within the dataset. We believe that additional insightful representations can be produced by adding external information. Toward that end, we have explored a number of candidate innovation indicators (Fig. 2). These empirically measure various aspects of the abstracts set and auxiliary information to get at different aspects of technological innovation, e.g., trends in research activity in particular domains, relative prevalence of industrial versus academic research activity. To produce these indicators presently, the analyst manually combines the requisite information.

We have taken an initial step to automate this process, generating an "indicator mapping" test program. This uses a Perl script to automatically extract information from a source database accessed through the Georgia Tech Electronic Library (such as *INSPEC*). It

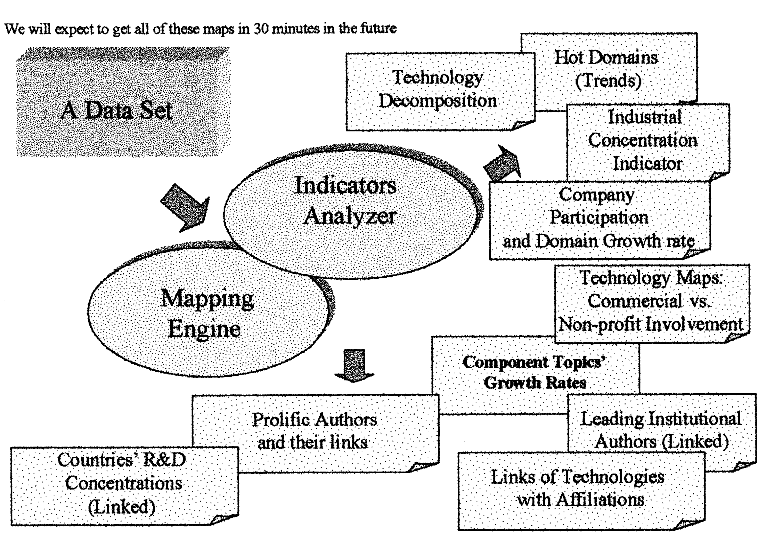

Fig. 2. Toward automated generation of innovation indicator and maps.

produces an indicator—"term specificity"—that measures terms' relative association with a given data set.

Term specificity is calculated as the ratio of a keyword's frequency of occurrence in the search set to the keyword's occurrences in the overall source database. A high ratio implies that the term is relatively particular to that data set; a low ratio implies that the term is relatively "universal." Fig. 3 plots the logs to yield an informative two-dimensional indicator—in this case for "Internet" topics. Terms that lie below the diagonal and are very large—such as "information technology" or "software tools"—can be taken as noise terms from a semantic view. On the other hand, terms that lie above the diagonal and exhibit high term specificity—such as "online front-ends" or "security of data"—are relatively particular to the topic under scrutiny. Fig. 3 points the analyst to these terms as strong candidate topics to analyze in seeking to understand complementary technologies and contextual influences on the development in question.

Rapid access to such innovation indicators should enrich the analytical process. We are working to expand the set of automatically generated innovation indicators. We expect that soon an analyst will be able to create a set of technology maps and innovation indicators by just striking a few keys and making a few choices [18].

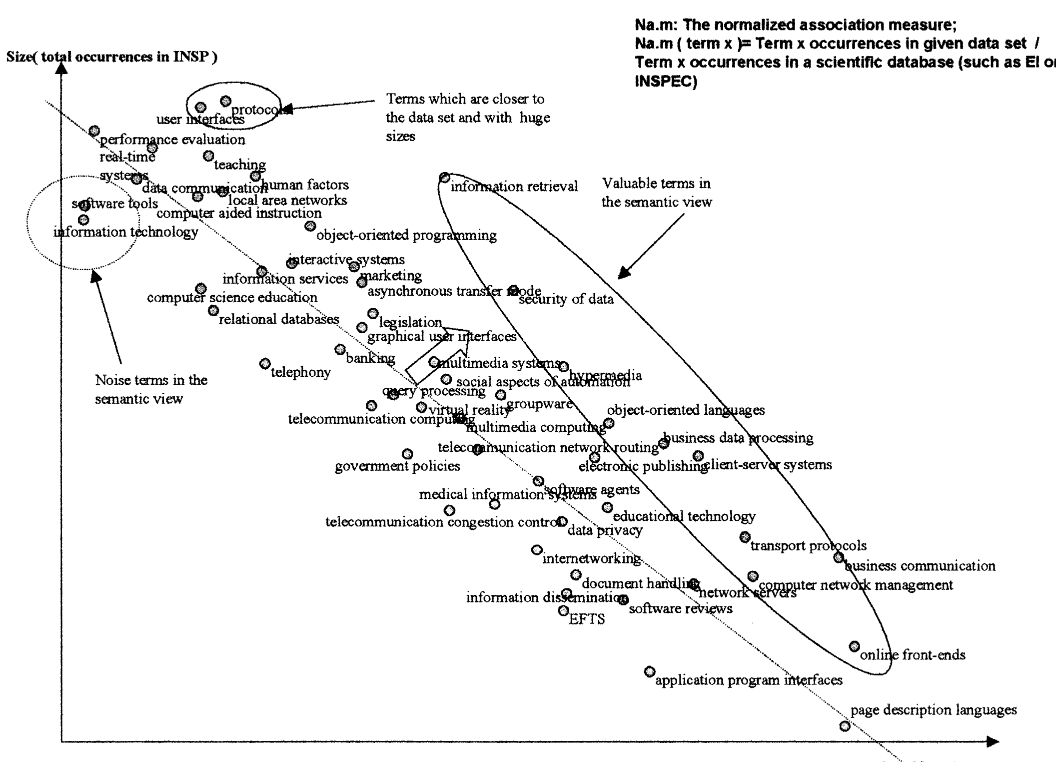

Fig. 3. Internet—indicator map, Na.m. versus size.

5. Observations

The partially automated processes presented provide "value-added" knowledge from bibliographic text mining. The family of maps allows a user to gain an intuitive feel for R&D activity. This offers potential routes to professional and managerial action:

- facility at digging down to initiate professional networking, e.g., follow the leads in Fig. 1 that the Department of Ceramic Engineering of Yonsei University in South Korea is publishing on silicon compounds, dielectric thin films, and sol–gel processing by checking their website;
- perspective on the dispersion of activity on a topic of interest (e.g., medical applications of nanotechnology R&D) to consider one's R&D priorities and commercialization prospects;
- indications of latent relationships, to identify links among topics (as per TOA principal components maps, not shown here (cf., [9]).

The innovation indicator (Fig. 2) illustrates how conceptual understanding (the technology innovation process) can guide creation of empirical indicators. In this case, the indicator combines internal information from the Nanotechnology topic search results with external database information to offer managerial insights.

Blending various forms of extracted information can significantly enhance the utility of text mining in areas such as technology management. Indeed, there appears to be a threshold phenomenon—if the results are rich enough, and helpful enough, then users will devote the energy to deal with unfamiliar information sources, analyses, and representations. If not, they prefer the familiar old ways. We are researching what factors make the real difference in getting various technology managers to use such text mining analyses [23].

In conjunction with demonstrable utility of its results, the text mining process needs to be readily accessible. This begins with database access licensed for unlimited usage—no charges per record retrieved (increasingly the case in the past year or so through Internet access and gateways, such as Dialog). Text mining for TF is facilitated by automated processes, so that an analyst can produce desired results quickly and correctly. Such automation also implies reproducibility, so that another analyst can replicate a representation to see how activities in a topical domain are changing over time, or to examine another domain in a comparable fashion. The two representations illustrated in this paper can be generated in less than 15 minutes from initiation of a search in any of a dozen or so technical or business text databases to which one has access (e.g., *Science Citation Index*, *Medline*, *US Patents*, *Business Index*). To reiterate, such empirical analyses demand expert review to refine searches, note gaps, and interpret insightfully. We have found that generating draft empirical analyses can stimulate expert involvement by providing handy summaries and fresh perspectives.

Reproducibility of representations holds special promise in facilitating knowledge management as organizations develop "standards," i.e., familiar, widely used forms that enhance information exchange and warehousing. For example, if an organization determined that a particular innovation indicator were valuable, then that can become part of its standard analytical forms. Fig. 3 (term specificity) is one of many candidate indicators we have been exploring. (Others are suggested in Fig. 2 and demonstrated on our website http://tpac.gatech.edu. Look under Technology Opportunities Analysis at the "Fuel Cells" and "KDD" examples).

Text mining tools, such as those described herein, can be applied to diverse information sources. Our emphasis lies in knowledge discovery in large, bibliographic abstract databases, but the tools can be adapted to internal organizational databases of various types. An exciting target is to collect topical information from a wide spectrum of websites—then filter, format, and analyze. The National Natural Science Foundation of China recently supported a 3-year key project for this (http://tpac.gatech.edu/~donghua/index.files/dmmmti.htm). Text mining can help integrate across multiple information resources too. For instance, a US Office of Naval Research research intelligence operation that stimulates cooperation between US and foreign researchers has sought to combine (1) ONR technology priority descriptions, (2) internal trip reports, and (3) external R&D publication database resources to help its experts locate promising research activities.

This paper emphasizes representation of text-derived knowledge. In closing, we note that text mining draws upon notably different approaches—NLP and computational linguistics, machine learning, and statistics (cf., Refs. [24–26]). Numerical data mining and text mining also employ some similar and some distinct approaches. Text mining emphases vary somewhat between mining ill-structured sources (e.g., lengthy documents, websites) and those addressing more structured texts (e.g., field-structured, abstract records). There truly is exciting potential for cross-fertilization among these approaches. So also with the sorts of representations illustrated here. We suggest that development of routines to generate particular representations—technology maps and innovation indicators—*automatically* can enhance the applicability of text mining and bibliometrics to TF. [Note that this does not mean automated TF-expert inputs and analyst art in focusing, sifting, and interpreting such representations is essential.] However, scripting the production of these visualizations can facilitate provision of empirically based, vivid TF findings, in a timely manner, to inform decision making. That could dramatically increase the utilization of TF in management of technology.

References

[1] V. Coates, M. Faroque, R. Klavins, K. Lapid, H.A. Linstone, C. Pistorius, A.L. Porter, On the future of technological forecasting, Technol. Forecast. Soc. Change 67 (1) (2001) 1–17.

[2] M.R. Kirby, D.N. Mavris, A method for technology selection based on benefit, available schedule and budget resources, 2000 World Aviation Conference, American Institute of Aeronautics and Astronautics, 370 L'Enfant Promenade, W., Washington, DC 20024, pp. 1–15, Oct. 10–12, 2000.

[3] A.L. Porter, N.C. Newman, R.J. Watts, D. Zhu, C. Courseault, W. Myers, E. Yglesias, Why Don't Technology Managers Want our Knowledge? International Association for Management of Technology, Miami, 2000.

[4] A.L. Porter, N.C. Newman, Why don't managers want our technological intelligence, and what can we do about it? SCIP, Alexandria, 22314 Society for Competitive Intelligence Professionals Annual Conference, Seattle, 2001.

[5] A.L. Porter, Enhancing the utilization of technology analyses, Center for Innovation Management Studies (CIMS), Raleigh, 2001 (June).

[6] A.L. Porter, X-Y. Jin, J.E. Gilmour, S.W. Cunningham, H. Xu, C. Stanard, L. Wang, Technology opportunities analysis: Integrating technology monitoring, forecasting, and assessment with strategic planning, SRA J. (Society of Research Administrators) 21 (2) (1994) 21–31.

[7] A.L. Porter, M.J. Detampel, Technology opportunities analysis, Technol. Forecast. Soc. Change 49 (1995) 237–255.

[8] A.F.J. Van Raan, Advanced bibliometric methods to assess research performance and scientific development: Basic principles and recent practical applications, Res. Eval. 3 (3) (1992) 151–166 (See also website: http://sahara.fsw.leidenuniv.nl/cwts/nnmap0.html).

[9] R.J. Watts, A.L. Porter, S.W. Cunningham, D. Zhu, VantagePoint intelligence mining: Analysis of natural language processing and computational linguistics, in: J. Komorowski, J. Zytkow (Eds.), Principles of Data Mining and Knowledge Discovery (First European Symposium, PKDD'97, Trondheim, Norway), Springer, Dusseldorf, Germany, 1997, pp. 323–335.

[10] R.J. Watts, A.L. Porter, C. Courseault, Functional analysis: Deriving systems knowledge from bibliographic information resources, Inf. Knowl. Syst. Manage. 1 (1) (1999) 45–61.

[11] R.J. Watts, A.L. Porter, N.C. Newman, Innovation forecasting using bibliometrics, Compet. Intell. Rev. 9 (4) (1998) 1–9.

[12] R.J. Watts, A.L. Porter, Innovation forecasting, Technol. Forecast. Soc. Change 56 (1997) 25–47.

[13] S. Deerwester, S.T. Dumais, G.W. Furnas, T.K. Landauer, D. Harshman, Indexing by latent semantic analysis, J. Am. Soc. Inf. Sci. 41 (1990) 391–407.

[14] J.P. Carlisle, S.W. Cunningham, A. Nayak, A.L. Porter, Related problems of knowledge discovery, Hawaii International Conference on System Sciences [HICSS] Proceedings on CD-Modeling Technologies and Intelligent Systems Track; Data Mining and Knowledge Discovery Mini-track, 1999 (January). Accesible on line at http://computer.org/proceedings/hicss/001/00016/0001 toc.html.

[15] D. Zhu, A.L. Porter, S. Cunningham, J. Carlisle, A. Nayak, A process for mining science and technology documents databases, illustrated for the case of knowledge discovery and data mining, Cienc. Inf. 28 (1) (1999) 1–8.

[16] D. Zhu, A.L. Porter, S.W. Cunningham, J. Carlisle, A. Nayak, A process for mining science and technology documents databases, illustrated for the Case of 'knowledge discovery and data mining,' Internal TOA Paper #94 [available on request].

[17] P. Losiewicz, D.W. Oard, R.N. Kostoff, Textual data mining to support science and technology management, J. Intell. Inf. Syst. 15 (2) (2000) 99–119.

[18] D. Zhu, and A.L. Porter, TOA: Illustrated for the case of knowledge discovery in databases and data mining; http://tpac.gatech.edu.

[19] A.L. Porter, Text mining for technology foresight, in: J. Glenn, T. Gordon (Eds.), Futures Research Methodology, AC/UNU, in press.

[20] R.N. Kostoff, Various reports on bibliometrics, http://www.scicentral.com/G-scipol.html# reports; http://www.dtic.mil/dtic/kostoff/index.html.

[21] M.W. Berry, S.T. Dumais, T.A. Letsche, Computational methods for intelligent information access, http://www.cs.utk.edu/~berry/sc95/sc95.html, February 1995.

[22] R.W. Schvaneveldt (Ed.), Pathfinder Associative Networks: Studies in Knowledge Organization, , Intellect Ltd., PO Box 862, Bristol BS99 10E, England, 1990 (ISBN 0-89391-624-2).

[23] A.L. Porter, J.P. Carlisle, R.J. Watts, Mining Bibliographic Information on Emerging Technologies, US National Science Foundation, Management of Technological Innovation Project (DMI-9872482), 1998–2000.

[24] H. Mannila, Data mining: machine learning, statistics, and databases, Eighth International Conference on Scientific and Statistical Database Management, Stockholm. Institute of Electrical and Electronics Engineers (IEEE) Computer Society, Piscataway, NJ, (1996) 1–8 (June).

[25] K.M. Decker, S. Focardi, Technology overview: a report on data mining, Technical Report (CSCS TR-95-02), Swiss Scientific Computing Center, 1995.

[26] C. Glymour, D. Madigan, D. Pregibon, P. Smyth, Statistical themes and lessons for data mining, Data Min. Knowl. Discovery 1 (1996) 25–42.