Matching Information Products to Technology Management Processes

Alan L. Porter, Nils C. Newman, Robert J. Watts, Donghua Zhu, and Cherie Courseault

Alan Porter and Cherie Courseault: Technology Policy and Assessment Center, and the School of Industrial & Systems Engineering, Georgia Tech, Atlanta GA, USA 30332-0205; alan.porter@isye.gatech.edu; cherie@isye.gatech.edu
Nils Newman: IISC, Inc., Atlanta, GA, USA; newman@mindspring.com
Robert Watts: U.S. Army Tank & Automotive Command, Warren, MI, USA; wattsb@cc.tacom.army.mil
Donghua Zhu: The Institute of Forecasting and Development, Hefei University of Technology, Hefei, P.R. China; zhudh@bepc2.ihep.ac.cn

Abstract

Information mining generates many potentially valuable information products. However, there are many difficulties in technology professionals and managers making use of those products. This paper identifies a number of issues and promising steps to remedy them.

Introduction

A recent *Harvard Business Review* article asked "What's your strategy for managing knowledge?" (Hansen 1999). The authors suggested we differentiate codified knowledge (reusable knowledge stored in databases) from personalized knowledge (highly customized solutions delivered through person-to-person communication). In our work, we extract information from collections of research and development (R&D) abstracts to meet distinct technology management needs. We think this represents a hybrid form of information exploitation -- "customized, codified knowledge on demand."

Putting such new forms of knowledge to use poses challenges, notably:

- what is the nature of these information products?
- what factors affect their acceptance?
- how do we produce them effectively?

This position paper addresses these issues, following a short background.

A team including the authors and colleagues at Georgia Tech, Search Technology, Inc., and Texas A&M has been developing a suite of information exploitation tools called the Technology Opportunities Analysis System (TOAS) since 1993 (Porter and Detampel 1995). The tools mine information obtained from large electronic databases to aid technology management processes (Feldman et al. 1998).

TOAS is a specialized form of Knowledge Discovery in Databases (KDD) – a rapidly evolving discipline that uses artificial intelligence, mathematics, and statistics to tease knowledge out of warehoused data (Carlisle et al. 1999; O'Leary 1998; Watts et al. 1997). It can also be characterized as bibliometrics (counting particular aspects of publication-related data sets) or textual data mining (Kostoff and Geisler Forthcoming; Ahonen et al. 1998). TOAS extracts information about particular emerging technologies (Watts, Porter, and Courseault Forthcoming) through a process of:

- search & retrieval from abstract databases maintained by others (e.g., *EI Compendex, MEDLINE, U.S. Patents, Business Index*)
- profiling & analysis of the search sets (typically 200 to 20,000 abstracts)
- representation & interpretation of usable knowledge.

For instance, imagine one's business unit had a need to know which other organizations were actively pursuing research and development of *fuel cells*. One might determine to search for research publications in *EI Compendex* and patent activity in *U.S. Patents* (as well as taping other sources). Analyzing the resulting abstract sets would help profile trends in *fuel cell* development.

Knowledge aims in the use of TOAS include benchmarking the developmental status of alternative technologies, competitive technological intelligence, and technology forecasting & assessment. Potential users range across practicing scientists & engineers, information professionals (e.g., technology licensing specialists), R&D program managers, and strategic planners. Issues in assessing knowledge for this 'management of technology' arena are elaborated elsewhere (Porter et al. 1991).

What is the nature of these Information Products?

Turning to the first of the three challenges, what is special about this knowledge? Figure 1 offers a simple schematic that spotlights main features of the TOAS "KDD" process. Several facets stand out.

First, TOAS-based information products reflect generation of customized information from codified

secondary sources. TOAS exploits large information collections (e.g., *EI Compendex* presently contains some 3,500,000 abstracts of engineering articles or conference papers since 1985 – a major portion of the world's engineering R&D). Widely accessible search and retrieval processes put this information instantly "at one's fingertips" on a given topic. However, just having the information accessible, even "minable," is not enough. Users demand customized analyses that directly address their present needs.

Second, this KDD is a new form of knowledge based on patterns extracted from large data compilations. It relies on statistical analyses and text mining that are not likely to be familiar to the target business users. In our case, the target users are typically either scientists or engineers -- used to intimate familiarity within a narrow domain, or managers – used to relying on people as the ultimate guarantors of the value of bits of "knowledge." KDD knowledge derives directly from processing the information resources per se – not from experts digesting the raw information to generate the new knowledge. This is not so reassuring to either the technical professionals nor their managers.

A third characteristic of this new type of knowledge, TOAS derives knowledge products focused on a particular domain -- technological change. We seek empirically based measures to assess technological maturation, contextual influences, and market opportunities (Watts and Porter 1997). For instance, we compute the percentage of publishing and/or patenting by industry as a commercial-readiness indicator [TOA analysis of 'KDD': http://tpac.gatech.edu]. Our term for these heavily processed measures is "innovation indicators." The hope is that these will be directly informative to the users . Such derived knowledge is intended to support decision processes.

Fourth, this derived information can be "packaged" in quite different forms. We can consolidate the content into one comprehensive report or packetize into sequential issue analyses. We can vary the media used, emphasizing text interpretations, numerical presentations (e.g., lists of "Top 10" organizations, authors, or journals), and/or graphical representations (e.g., "technology maps"). We have experimented with providing reports together with CD's containing the topical abstracts and a simplified version of the TOAS software so users can hyperlink to check out the underlying information (e.g., to read, say, IBM's abstracts on 'intelligent agents' on the spot). Some users are ready for such a "hands-on" form of information product delivery; others are not.

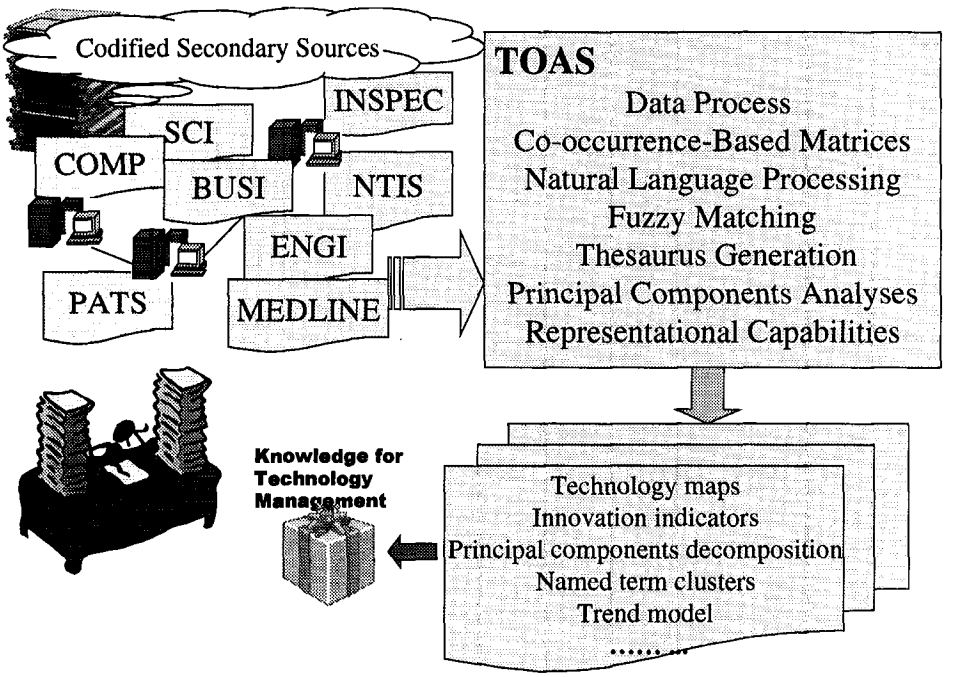

Figure 1 TOAS--Generation of Customized Information Products for Technology Management

What factors affect acceptance?

Acceptance of non-traditional information products and tools by knowledge users entails meeting their manifold requirements (Simpson, Kingston, and Molony 1999).

Over the past six years we have used TOAS and its predecessor "TOA Knowbot" software to perform, perhaps, 100 analyses for different governmental and business users [c.f., http://tpac.gatech.edu]. Since early

1999, TOAS software has become available to some of those users themselves. This offers a nice menu of options to the user, ranging from "do it yourself" to having an outside analyst deliver various forms of reports. What works?

The experiences have been sobering – these new information products are not easily adopted in support of business processes. In general, successful managerial applications of "management science," including text mining, are few and far between (Kostoff and Geisler Forthcoming). Bibliometrics fare awkwardly in R&D evaluation processes. Detail-oriented scientists don't know what to make of profiles of entire research domains. Focusing on textual data mining, analysts are often decoupled from the target users (Kostoff and Geisler Forthcoming; Kostoff Submitted for publication). With support from the National Science Foundation (NSF) (Porter, Carlisle, and Watts 1998), we have reviewed experiences and experimented with alternative ways to provide information to business and government. We have identified a number of issues – here are five major ones.

First, target users need the right information for the task at hand. Too often, our cleverly crafted reports miss the mark. There appear to be deeply rooted, "cultural" preferences among knowledge workers for either *reports* or *answers*. Some organizations emphasize reports – e.g., governmental R&D evaluation to meet Government Performance Review Act (GPRA) requirements generates evaluative reports (Kostoff Submitted for Publication). We have recently been working with a National Institute of Occupational Health lab on use of TOAS in support of GPRA reporting. On the one hand, profiling a biomedical research domain (e.g., *asthma*), provides valuable benchmarking (how does this lab stack up against others?), trends (what topics are increasing in emphasis recently?), and applications (what results are linked with what problems?). Yet, such tabulations are often greeted with suspicion – why was this paper missed? Are citations biased by friendships and rivalries? At the National Science Foundation, mention of "bibliometrics" generates a special form of apoplexy (not universally).

Technology management reports vary in emphases, including alerting to emerging technology opportunities, benchmarking organizational performance, or forecasting technological prospects. On the other hand, many knowledge workers seek answers to more specific questions. For instance, we have been asked which, if any, alternative technologies could compete with a proposed new structural material, to help the client decide whether to build a production plant. Posing the questions so TOAS can provide answers takes considerable effort. Providing the results in the form of user-friendly answers takes even more effort. This often pulls the TOAS analyst far outside her or her "comfort zone" of analyzing abstracts.

A second major issue -- users need the information on a timely basis. A survey of 26 technology professionals and managers collaborating in our NSF project found them typically needing results fast:

- within a day (21%)
- within a week (45%)
- within a month (24%)

But our analyses have typically taken a month or more. The mismatch in temporal orientation is particularly severe with academic researchers trying to meet the needs of industrial technology managers.

Third, most would-be users have a difficult time accepting KDD-derived knowledge as credible. The ready availability of these electronic databases is relatively new. Understanding how they consitute valid bases of the derived knowledge is difficult, posing credibility concerns for TOAS-based information products. Managers prefer to have experts provide the requisite knowledge. Experts constitute a more familiar font of knowledge. They also offer "blamability" advantages should decisions go awry. Of course, combining KDD analyses with expert opinion is highly sensible. Again, this stresses the analysts who now have to both analyze these database-derived information sets (requiring certain forms of technical skills) and obtain expert opinion (requiring distinctly different skills), then integrate these two forms of knowledge effectively.

A fourth issue -- we believe that involving the user in the KDD process really facilitates acceptance (Kostoff and Geisler Forthcoming). However, this often proves hard going. Scientists and engineers tend to be uncomfortable with profiling entire R&D domains. But, if they can be engaged, they develop valuable perspective on how their research links with others', fostering new contacts and collaborations. For managers, digging into these information products entails unfamiliar skill sets and significant time demands. They tend to designate a junior associate or an information specialist to learn TOAS. This imposes an intermediary in the process, making acceptance of the information products that much more difficult.

In working with some fifty private and public organizations, we have come upon a special fragility (our fifth issue). Expertise in dealing with these electronic information resources and KDD processes is still relatively rare. Strenuous organizational learning is needed to adapt the information products to decision needs. In a couple of instances, we have rejoiced prematurely at reaching this plateau. This learning can be instantly undone by reassignment of one or more key persons, because substitutes with comparable KDD understanding are not readily available. This supports Kostoff's suggestion that KDD and related analytical tools must become an integral part of business decision processes to attain high value (Kostoff and Geisler Forthcoming). They are not there yet.

How can we produce the right Information Products when needed?

In our work, we begin with the TOAS tool suite. It provides list processing, co-occurrence-based matrices, natural language processing to parse texts into analyzable phrases, fuzzy matching, thesaurus generation and enhancement via experience, principal components analyses, and representational capabilities (combining multi-dimensional scaling with a path erasing algorithm). TOAS is thus general purpose KDD or text mining software.

We have devised a set of algorithms that build upon this general analytical functionality. These combine conceptual understanding (about technological innovation processes) with empirical experiences to greatly facilitate generation of specific outputs. Namely, we can semi-automatically generate six different technology maps, two innovation indicators, helpfully named term clusters, four trend model fits, and two versions of a proprietary principal components decomposition. Focusing tool enhancement on one domain enables quick generation of particular management-oriented information products.

The need to enable lightly trained analysts (e.g., students) to generate particular useful information products has prompted us to go beyond the algorithms. We are building more detailed "templates" combining tacit analytical knowledge and reusable resources (Kerber et al. 1998; Staab and Schnurr 1999). Sequences of analytical steps are scripted for quick production of rather sophisticated technology management information products.

Thus, by elaborating upon the TOAS software through special algorithms and processing scripts, we can respond more quickly to sequential inquiries from a manager. We see this as essential to fulfilling the user needs – some 2/3 of whom want results in a week or less.

Further steps would entail combining such database mining with exploitation of other resources, such as websites and on-line expert Delphi surveys to enrich findings. That implies better rounded information products that integrate multiple forms of information. However, the tradeoff is time and comfort level – this is a demanding extension of analyst requirements.

Conclusion

Rapidly developing analytical capabilities enable generation of customized information products from codified information resources. These have tremendous potential to aid real-time decision processes. However, successful adoption of these new forms of knowledge requires pointed attention to the capabilities and needs of the intended users.

We point to a number of specific considerations in striving to deploy our TOAS-based information products for technology knowledge management (Taylor 1999). We continue our National Science Foundation supported research to improve the matching of KDD capabilities to technology management needs (Porter, Carlisle, and Watts 1998).

More generally, we suggest that various information mining developers need to direct significant attention to usability. Rather than testing our ideas out on each other (namely, fellow KDD innovators), we need to test them out on the ultimate users. We assert that this is likely to show the need for highly tailored information products and for detailed "process engineering" to enable timely delivery of the exact products wanted, when they are wanted. We need to emulate the customer-driven model learned in the quality movement if we aspire to see our KDD and associated approaches join the managerial information mainstream.

As noted earlier, it's awfully hard to break into that mainstream. In a "Catch—22" bind, at present the small numbers of knowledgable data miners (KDD experts and users) makes for extreme fragility. Real-life cases make advocates cringe. In one organization experimenting with KDD-based information products, we found a lamentable trail of difficulties. One needs access to data to do data mining. In this organization, they had subscribed at considerable cost for years to databases, but had lost the capability to actually obtain the data. We found the requisite software safely stored in a box, unused for years. The KDD champion fought through that hurdle to reacquire effective access and accomplished several marked successes, resulting in definitive organizational action. Not enough – his colleagues had not learned the analytical techniques themselves. So, when he moves on to another position (likely at this moment), the investment in KDD capabilities will atrophy.

To close on a more optimistic note, we have done a bit of self-study, namely performing a TOAS review of the KDD domain. One of the intriguing side-notes was that the most prolific research group, located at Caltech, suddenly had a number of its core members move to Redmond, WA. Indeed, we have confirmed that Microsoft is actively developing information mining. Additionally note the advent of IBM's *Intelligent Miner* software introduced in 1998 to facilitate mining application development. The involvement of these two leading computing companies implies a high likelihood of further tool development and commercialization over the coming few years. That sort of impetus should really aid the various forms of information mining to contribute effectively to mainstream business practices.

Acknowledgements

The authors appreciate funding for this research provided by the National Science Foundation – project DMI-9872482, managed by Joseph Hennessey.

References

Ahonen, H., Heinonen, O., Klemettinen, M., and Verkamo, A. I. 1998. Applying Data Mining Techniques for r Descriptive Phrase Extraction in Digital Document Collections. 2-11. *Proceedings, IEEE International Forum on Research and Technology: Advances in Digital Libraries (ADL'98).*

Carlisle, J. P., Cunningham, S. W., Nayak, A., Porter, A. L., and Sprague, Jr., R. H. 1999. Related Problems of Knowledge Discovery. Hawaii International Conference on Systems Sciences (HICSS-32), Proceedings on CD – Modeling Technologies and Intelligent Systems Track; Data Mining and Knowledge Discovery Mini-track.

Feldman, R., Fresko, M., Hirsh, H., Aumann, Y., Liphstat, O., Schler, Y., Rajman, A., and Reimer, U. 1998. Knowledge Management: A Text Mining Approach. *Practical Aspects of Knowledge Management* 9: 1-10.

Hansen, M. T., Nohria, N., and Tierney, T. 1999. What's Your Strategy for Managing Knowledge. *Harvard Business Review* 77: 106.

Kerber, R., Beck, H., Anand, T., Smart, B., Agrawal, R., and Stolorz, P. 1998. Active templates: Comprehensive support for the knowledge discovery process, *Proceedings, Fourth International Conference on Knowledge Discovery and Data Mining*, 244-248.

Kostoff, R. N. Network-centric Peer Review. Submitted for publication.

Kostoff, R. N., and Geisler, E. F. Strategic Management and Implementation of Textual Data Mining in Government Organizations, *Technology Analysis and Strategic Management.* Forthcoming.

O'Leary, D. E. 1998. Enterprise Knowledge Management. *Compute* 31 (3): 54-61.

Porter, A. L., Carlisle, J. P., and Watts, R. J. 1998. *Mining Bibliographic Information on Emerging Technologies*, Technical Report, DMI-9872482, U.S. National Science Foundation, Management of Technological Innovation Project.

Porter, A. L., and Detampel, M. J. 1995. Technology Opportunities Analysis. *Technological Forecasting & Social Change* 49: 237-255.

Porter, A. L., Roper, A. T., Mason, T. W., Rossini, F. A., and Banks, J. 1991. *Forecasting and Management of Technology.* New York, NY, Wiley.

Simpson, J., Kingston, J., and Molony, N. 1999. Internet-based Decision Support for Evidence-based Medicine, *Knowledge-Based Systems* 12: 247-255.

Staab, S. and Schnurr, H-P. 1999. Knowledge and Business Processes: Approaching an Integration, *OM '99 - Proceedings of the International Workshop on Knowledge Management and Organizational Memory (IJCAI-99).* Stockholm, Sweden.

Taylor, R. M. 1999. Steps on the Path to Knowledge, *Knowledge Management Information Management Event Proceedings,* 33-36.

Watts, R. J., and Porter, A. L. 1997. Innovation Forecasting, *Technological Forecasting & Social Change* 56: 25-47.

Watts, R. J., Porter, A. L., Courseault, C. Functional Analysis: Deriving Systems Knowledge from Bibliographic Information Resources. *Information, Knowledge, Systems Management.* Forthcoming.

Watts, R. J., Porter, A. L., Cunningham, S. W., and Zhu, D. 1997. TOAS Intelligence Mining: Analysis of Natural Language Processing and Computational Linguistics, in J. Komorowski and J Zytkow (eds.), *Principles of Data Mining and Knowledge Discovery, First European Symposium, PKDD'97,* 323-335. Trondheim, Norway: Springer.