TECH Mining

By Alan L. Porter

Tech mining (TM) uses text mining software to exploit science and technology (S&T) information resources. Tech mining is done to inform technology management. In it we combine an understanding of technological innovation processes with software tools to derive vital S&T knowledge.

Traditional means of gathering competitive technological intelligence (CTI) are time-consuming and expensive. One company took six months of work to catalog half of 13,000 potentially relevant patents (Teichert 2002). In addition, all too often managers don't use the resulting intelligence, for a variety of reasons. (Porter 2005). We can solve these problems through integrating of database access, applying TM software, and creating predetermined output forms. Technology managers using this derived S&T knowledge can gain a marked competitive advantage.

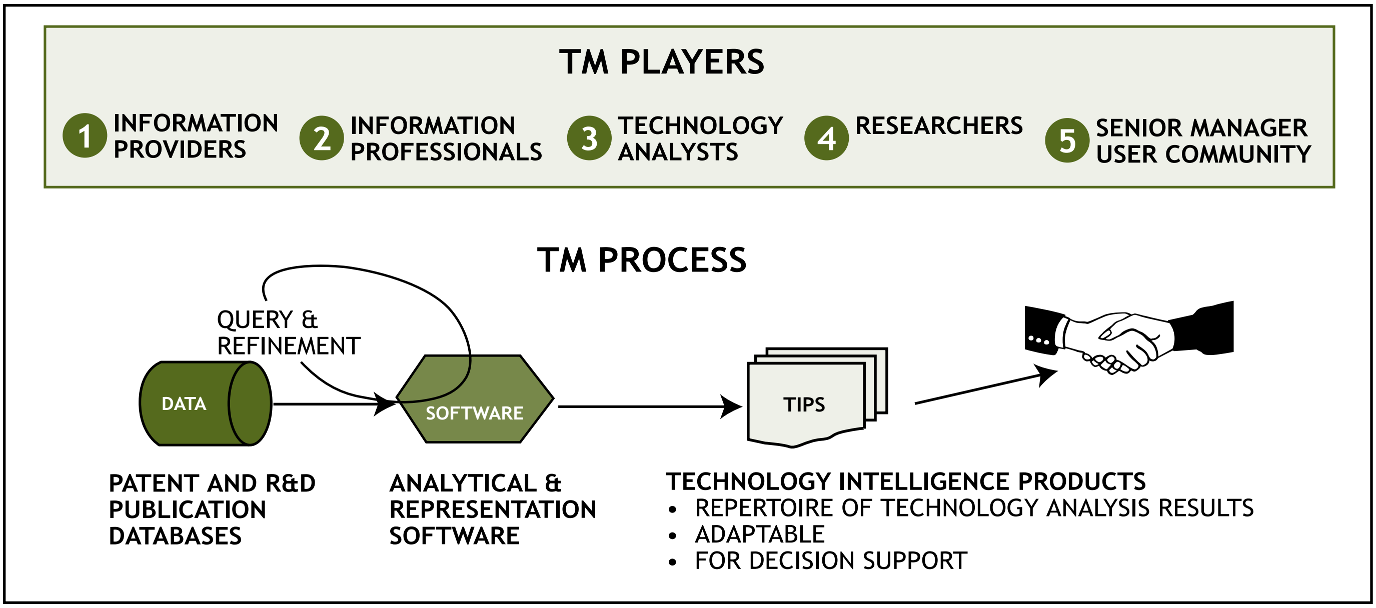

TM PLAYERS
1 INFORMATION PROVIDERS
2 INFORMATION PROFESSIONALS
3 TECHNOLOGY ANALYSTS
4 RESEARCHERS
5 SENIOR MANAGER USER COMMUNITY
TM PROCESS
QUERY & REFINEMENT
DATA
SOFTWARE
TIPS
PATENT AND R&D PUBLICATION DATABASES
ANALYTICAL & REPRESENTATION SOFTWARE
TECHNOLOGY INTELLIGENCE PRODUCTS
• REPERTOIRE OF TECHNOLOGY ANALYSIS RESULTS
• ADAPTABLE
• FOR DECISION SUPPORT

Figure 1: Tech Mining (TM) process and players

THE TECH MINING APPROACH

The cast of characters in tech mining differ in their skills and knowledge, and their priorities. (See Sidebar 1.) Figure 1 sketches what's involved in tech mining. The basic process contains three key elements:

- ***Data*** starts with R&D publications (Science Citation Index, INSPEC, MEDLINE) and patent abstract databases (Derwent World Patent Index, Delphion), complemented by other data on research funding and projects, plus business information, marketing, policy, and popular press resources, topped off by internet searches for current activities.
- ***Software*** from search and retrieval, through cleaning and analysis, to representation and visualization, software tools make analysis of thousands of records practical and informative.
- ***Technology intelligence products*** are the outputs of the empirical analyses target the intended users' information needs.

There are two distinct orientations to analysis of R&D information. Data mining recognizes the rich data resources and digs in, generating scads of analyses. The mindset is often "whee – look at all the neat figures and charts we can create! Surely, these results will help you manage better." We make fun of ourselves as it took a lot of disappointment for us to understand "why don't managers use our analyses?" (Porter 2001).

With support from the National Science Foundation (Project DMI-9872482) and the Center for Innovation Management at North Carolina State University, our team reviewed our own experiences in some 100 technology analyses and the literature on research utilization, and compiled 32 case studies.

Technology management issues

Tech mining starts with technology manager needs rather than with the data. We work back from those needs to generate well-targeted technology intelligence products. We do so by considering technology management issues, questions, and pertinent innovation indicators.

We began with the notion that one general set of tech mining analyses could serve all user needs. We no longer think so. Instead, whoever is performing TM ought to first interact thoroughly with the target users to understand what technological intelligence they want, and how they want it delivered. To help kick off this process, we identified 13 technology management issues:

- R&D portfolio selection
- R&D project initiation
- Engineering project initiation
- New product development
- New market development
- Mergers
- Acquisitions of intellectual property
- Exploiting one's intellectual assets
- Collaborative technology development
- Assessing competing organizations
- Forecasting opportunities and threats
- Strategic technology planning
- Technology roadmapping

SIDEBAR 1: TM CAST OF CHARACTERS

Information providers at patent offices and database companies become increasingly involved in fostering application of their information by facilitating data access for mining and by linking with software tools.

Information professionals are typically most knowledgeable concerning the information sources, including tradeoffs in coverage and costs, searching, and cleaning.

Technology Analysts are adept at analyzing the data, but often need to work on communicating results effectively to the technology managers.

Researchers are an often overlooked group and include engineers, inventors, and project managers. They may include occasional users, but are also power users who become sources of TM expertise.

Manager-users are a heterogeneous mix of professionals and managers who can benefit from TM and run across a gamut of technology management domains, such as R&D, new product development, process engineering, and strategic planning.

Multiple questions

Each issue poses multiple questions. Tech mining can answer many but not all of those questions. The good news here is that many questions relate to more than one issue; we don't have a giant hierarchy. The bad news is that questions don't map neatly to issues. Instead, we have arrayed 39 questions, subsets of which relate to each of the 13 issues. For instance, with regard to the issue of engineering project initiation, we spotlight 13 questions:

- What's hot?
- Fit into tech landscape?
- Drivers?
- Competing technologies?

- Development prospects? Likely development paths?
- Component tech maturity?
- Systems maturity?
- Match to our interests?
- Our opportunities here?
- Needs addressed?
- Our strengths and gaps?
- Commercialization prospects?

Empirical indicators

In turn, each question can be addressed through many empirical indicators. We approach these from both ends of the data/needs spectrum. From the data end, consider the available data (a function of which data resources are being tapped). We inventory possible measures.

From the needs end, we consider specific objectives for the tech mining activity. A model of technological innovation processes helps us identify measures. We call them innovation indicators and they speak to the prospects for successful innovation. Three general types (Watts 1997) are:

- *technological maturation* – how far and how fast is the technology in question progressing toward commercialization or other implementation?

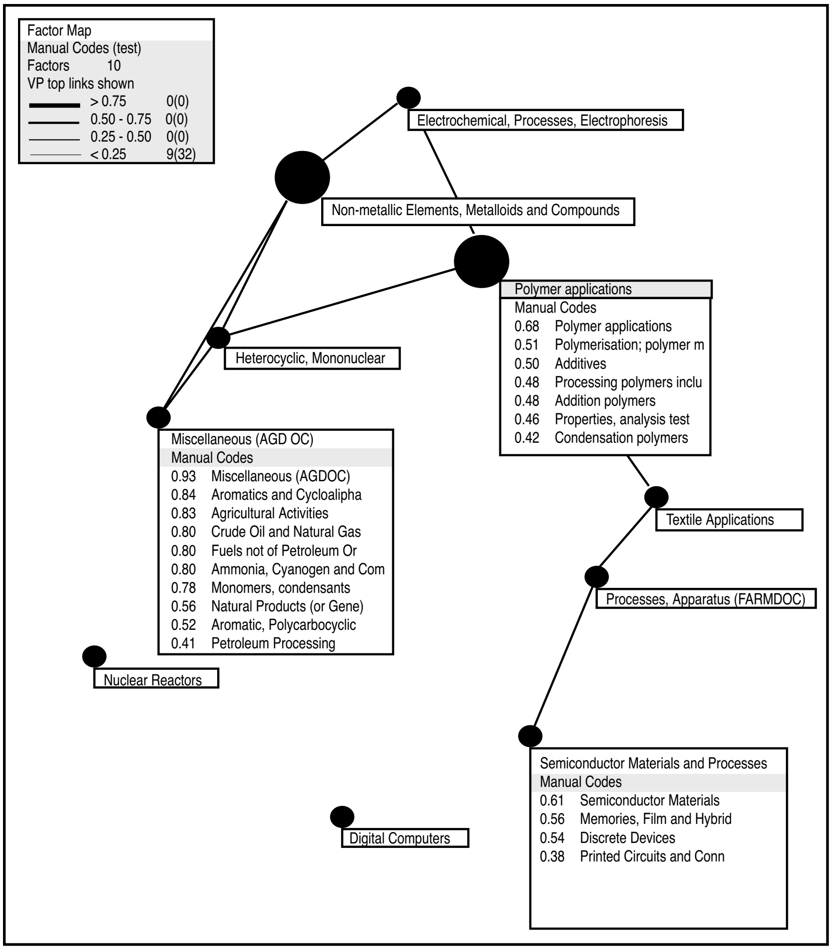

Factor Map
Manual Codes (test)
Factors 10
VP top links shown
> 0.75 0(0)
0.50 - 0.75 0(0)
0.25 - 0.50 0(0)
< 0.25 9(32)
Electrochemical, Processes, Electrophoresis
Non-metallic Elements, Metalloids and Compounds
Polymer applications
Manual Codes
0.68 Polymer applications
0.51 Polymerisation; polymer m
0.50 Additives
0.48 Processing polymers inclu
0.48 Addition polymers
0.46 Properties, analysis test
0.42 Condensation polymers
Heterocyclic, Mononuclear
Miscellaneous (AGD OC)
Manual Codes
0.93 Miscellaneous (AGDOC)
0.84 Aromatics and Cycloalipha
0.83 Agricultural Activities
0.80 Crude Oil and Natural Gas
0.80 Fuels not of Petroleum Or
0.80 Ammonia, Cyanogen and Com
0.78 Monomers, condensants
0.56 Natural Products (or Gene)
0.52 Aromatic, Polycarbocyclic
0.41 Petroleum Processing
Textile Applications
Processes, Apparatus (FARMDOC)
Nuclear Reactors
Semiconductor Materials and Processes
Manual Codes
0.61 Semiconductor Materials
0.56 Memories, Film and Hybrid
0.54 Discrete Devices
0.38 Printed Circuits and Conn
Digital Computers

Figure 2: Map of fuel cell patents based on class codes

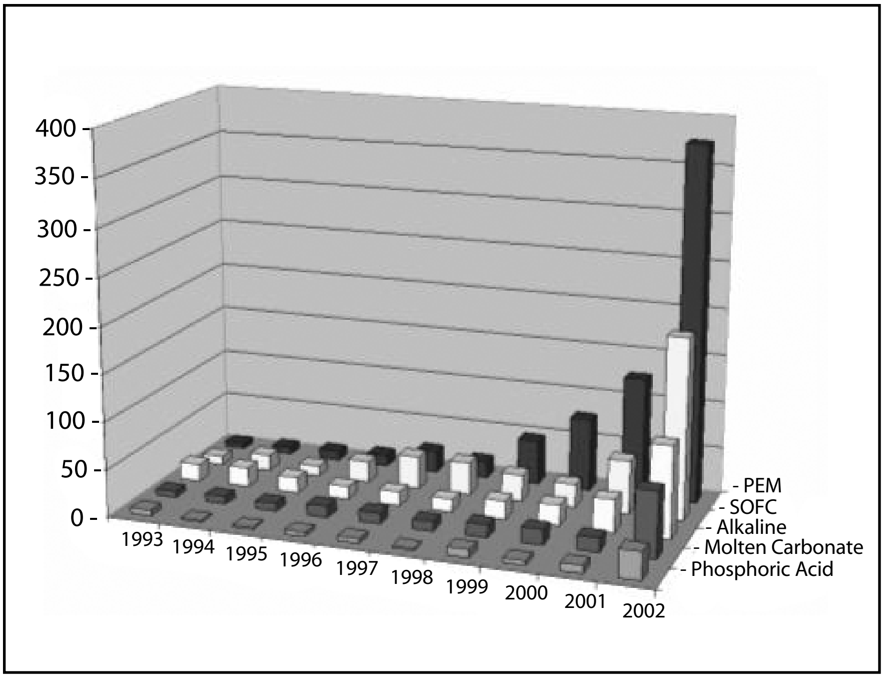


Figure 3: Patent activity over time

- *contextual influences* – how do various influences on innovation stack up?
- *market opportunities* – what are the prospects?

For example, we have identified a set of indicators in response to the first question listed above – *What's hot* with respect to the target technology:

WHAT?

- mapping of topic clusters within the technology
- 3-D trend charts for topic clusters
- ratio of conference to journal papers (benchmarked)
- scorecard rate-of-change metrics for topic clusters
- time slices to show evolution of topical emphases
- topic growth modeling (S-curve) fit & extrapolation

TABLE 1: SAMPLE DERWENT WORLD PATENT INDEX RECORD FIELDS

Fields	No. of Items
Raw Record	9,724
Abstract Phrases	118,683
Derwent Classifications (Cleaned)	278
Family Member Countries (Cleaned)	42
Family Member Years	39
Inventors (Cleaned)	10,112
Patent Assignees (Cleaned)	3,311

WHO?

- pie chart – company vs. academic vs. government publishing
- topical main players' profiles
- spreading (or constricting) # of players by topic

Two key points: First, such a list is just a starter. It aims to stimulate thinking about what technology intelligence products can help reach the necessary decisions. Second, we have a potential explosion of information. Suppose that you were deciding whether to initiate a new engineering project. You might well have a stage-gate process that posed specific questions. Imagine this consisted of the 13 questions listed above and that we could generate about nine indicators to respond to each question. No manager wants 13 x 9 = >100 charts to digest! We'll return to this concern.

TECH MINING ILLUSTRATIONS

Let's illustrate. We've done sample analyses on *fuel cell* data. A March 2003 search of the Derwent World Patent Index located almost 24,000 patent records. We focused on 9,724 patent families that contained at least one non-Japanese patent. 45 fields of information available in these abstract records, deriving from patent front page information clarified and classified by Derwent.

We also compiled fuel cell R&D publication abstracts from the *Science Citation Index* and *INSPEC* via Dialog. We combined those searches and removed duplicates to yield 11,764 records. These illustrative analyses used simple searches without the iteration and refinement warranted for specialized CTI purposes.

Vantage Point/Derwent Analytics

Data from the searches noted were imported into the *VantagePoint/DerwentAnalytics* software (a version called TechOASIS is available for US Government use; a commercial version tailored to Derwent WPI data is available as Derwent Analytics) to create two abstract record files, one on fuel cell patents and the other on publications (journal articles and conference papers). *VantagePoint* is MS Windows text mining software developed by Georgia Tech with extensive support from the US government, including the Defense Advanced Research Projects Agency, Army Tank and Automotive Command, Office of Naval Research, and the National Science Foundation.

VantagePoint helps clean the data through a suite of tools including data fusion, fuzzy matching, thesaurus building and application, list comparison, and so forth. The software manipulates text to facilitate tabulation and analysis. For example, the 'Abstract Phrases' reflects the application of natural language processing (NLP) to the abstracts. Text mining software helps discover relationships based on co-occurrence of terms in records – e.g., one can explore *knowledge networks* based on authors or inventors collaborating, or just using similar terminology.

We divide innovation indicators into two general types – *what and who*? *What* measures can be very straightforward. For instance, we could list the number of patents addressing each of the five main fuel cell types. Let's examine two *what*? examples of the six indicators suggested above.

Clustering topics

Text mining uses statistical tools to cluster topics based on co-occurrence across the records. One can do this using keywords or abstract phrases, for instance, but Figure 2 illustrates using Derwent patent classifications. In this visualization, nodes reflect the number of patent records containing any of the high-loading classes that co-occur frequently. Pull-downs in Figure 2 illustrate high-loading classes for three of the nodes.

Depending on one's intents, you could separate out the patents relating to one node (e.g., the semiconductor topics) for further investigation. Location of nodes in the map is based on multi-dimensional scaling to reflect relationship. However, this is a weak indicator of relationship, so a

TABLE 2: "TOP 10" EUROPEAN, AUTOMOTIVE-ORIENTED, FUEL CELL PATENT

Patent Assignees	#
Xcellsis GMBH	49
Daimler Chrysler	40
Siemens-Westinghouse	22
Mannesmann	20
Volkswagen	18
Emitec	17
Renault	16
DBB Fuel Cell Engines	15
Valeo Klimasysteme	7
Bosch	6

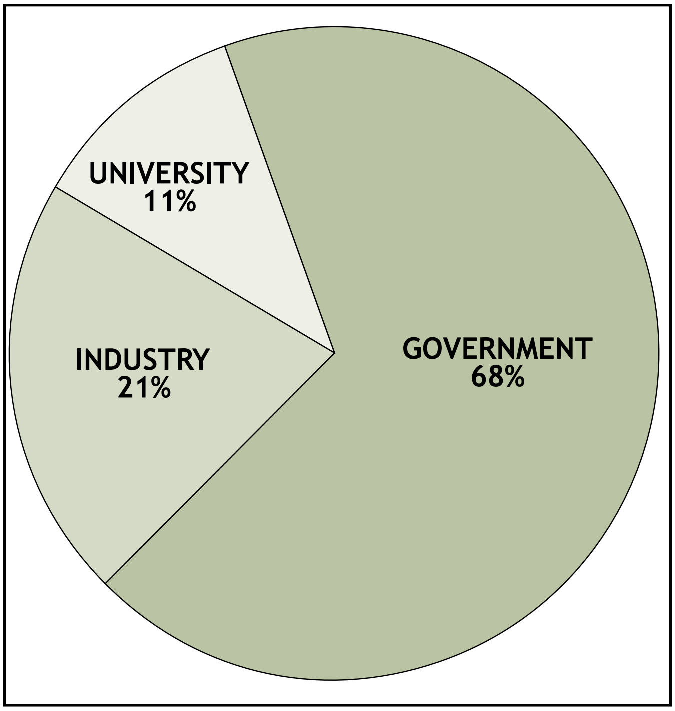


Figure 4: Percentages of publications

path-erasing algorithm is reflected by the strength of the interconnecting lines. In the figure, *nuclear reactors* appear relatively distinct from other fuel cell applications.

Text mining software also enables us to examine two fields of information together. For example, Figure 3 shows the number of patents mentioning particular fuel cell types, over time. This particular date measure indicates ongoing interest. It suggests dramatically different interest levels by fuel cell type, led by PEM (proton exchange membrane). But all the types show strong current activity – fuel cells are hot!

The who of R&D

Our intent here is just to convey the flavor of tech mining. To illustrate further, here are two of many possible ways to get at *who* is doing all this R&D. One simple, but powerful, indicator of commercial interest derives from publications. Figure 4 shows what portion of these come from university, corporate, or government organizations. Fuel cell papers show a very active involvement by industry – a strong indicator of commercial promise.

Table 2 suggests how one can probe further by subdividing the dataset. This tallies 278 of the total 9724 fuel cell patents that are automotive-related (identified by searching for particular terms in certain record fields), patent assignee in Western Europe, and recent (priority patent dating 2000-2003).

Much of CTI is *playing detective*. This tabulation provides an intriguing starting point for further investigation. We might initially wonder – who are Xcellsis and DBB? A quick examination of joint patent assignment finds that Ballard Power Systems, a leading fuel cell company, is linked to both. With a little help from Google on the internet, we find that Daimler Chrysler and Ballard collaborated on fuel cell development, 1993-97. This blossomed into a jointly owned company – Xcellsis, formerly named DBB. Further, we spot that Ford Motor Company joined their alliance in 1997, investing in both Xcellsis and Ballard.

Note that care is required in interpreting simple activity counts. For instance, Daimler Chrysler patenting appears to drop sharply after 1999 (not shown), whereas, in fact, their commitment to fuel cell development escalates via these highly active joint ventures. It is advisable to seek complementary, expert information to verify CTI observations.

INFORMATION PRESENTATION

As mentioned, one could generate lots of measures from these data. That's why we believe it's vital to focus on what information can best answer the technology questions management posed. We recommend that the technical intelligence provider interact directly and extensively with the target users to learn what they need to know for the matter at hand. Then learn how the intended users like that information presented:

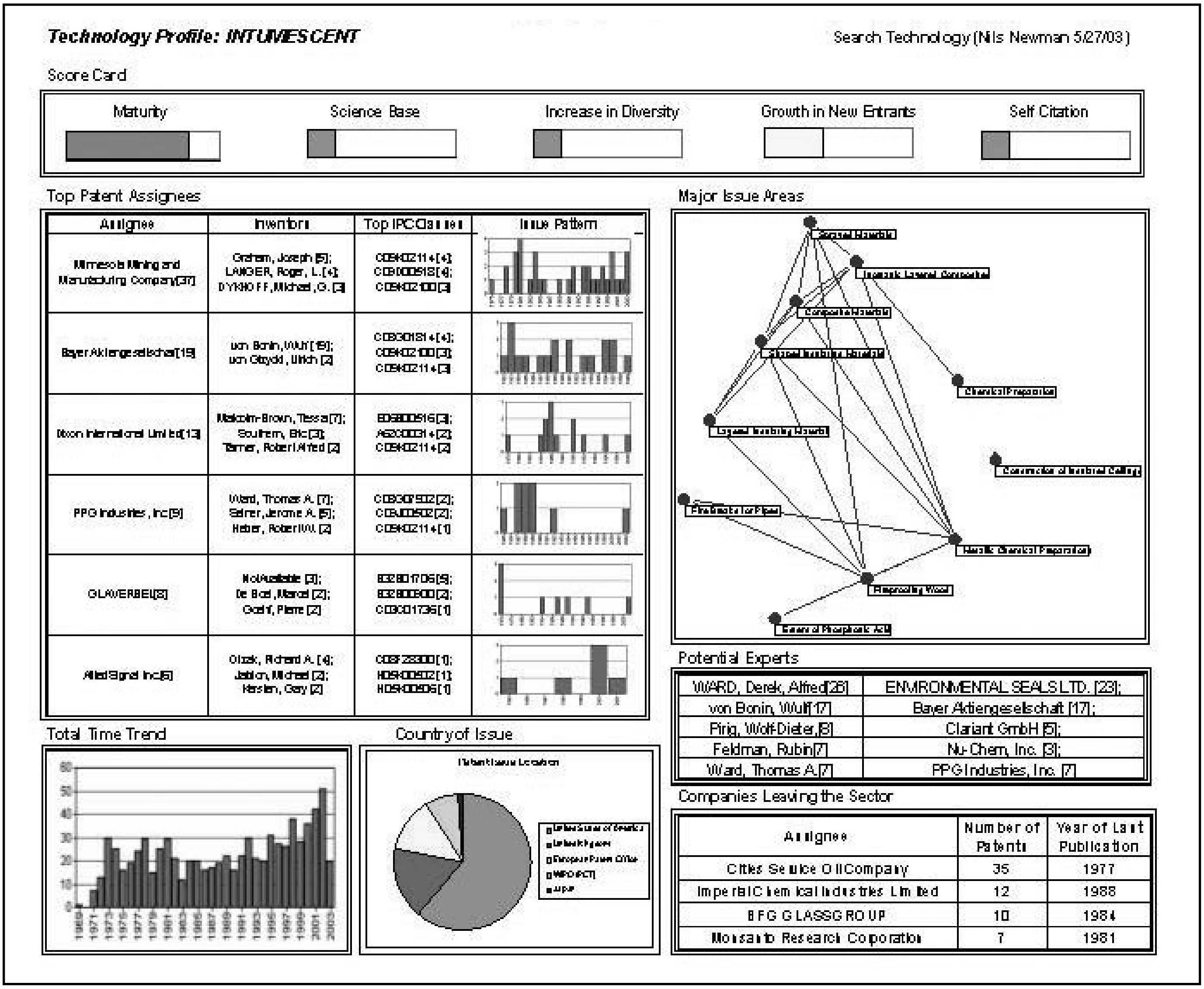
- in what manner – we suggest interactive, face-to-face, whenever possible
- in what form – what balance of visual, numerical, and text (interpretation)
- how much – with the general target of layering, showing senior managers the key findings that point toward actions, backed up by suitable auxiliary information to be perused only as needed

One-pagers offer a nice presentation target. Obviously, this should be modified to fit the circumstances. Most importantly, the information presented should be tailored to answer the prime question at hand.

Figure 5 illustrates the notion of a one-pager. This is not for fuel cells. To present a bonafide topical composite presentation, we really need a driving technology management question, for well-specified decision circumstances, to determine what content is appropriate. That would likely integrate findings from patent, publication, and other information sources.

My colleague, Nils Newman, generated Figure 5 to demonstrate technology intelligence presentation possibilities:

- Scorecard indicators across the top are non-numerical presentations, particularly helpful in comparing multiple technologies 'at a glance.'
- Profiling in the upper left section focuses on a set of leading entities (either who or what categories work)



Technology Profile: INTUMESCENT

Search Technology (Nils Newman 5/27/03)

Score Card

Top Patent Assignees

Assignee	Inventors	Top IPC Classes	Issue Pattern

Minnesota Mining and Manufacturing Company[37]	Graham, Joseph [5];
	LANGER, Roger, L. [4];
	DYKHOFF, Michael, G. [3]
		C09K02114[4];
		C09D00518[4];
		C09K02100[3]	

Bayer Aktiengesellschaft[19]	von Bonin, Wulf [19];
	von Gizycki, Ulrich [2]
		C08G01814[4];
		C09K02100[3];
		C09K02114[3]	

Dixon International Limited[13]	Malcolm-Brown, Tessa [7];
	Southern, Eric [3];
	Turner, Robert Alfred [2]
		E06B00516[3];
		A62C00314[2];
		C09K02114[2]	

PPG Industries, Inc[9]	Ward, Thomas A. [7];
	Seiner, Jerome A. [5];
	Heber, Robert W. [2]
		C08G07502[2];
		C09J00502[2];
		C09K02114[1]	

GLAVERBEL[8]	Not Available [3];
	De Bast, Marcel [2];
	Goelff, Pierre [2]
		B32B01706[5];
		B32B00900[2];
		C03C01736[1]	

AlliedSignal Inc[6]	Olszak, Richard A. [4];
	Jablon, Michael [2];
	Kersten, Gary [2]
		C08F28300[1];
		H05K00502[1];
		H05K00506[1]	

Major Issue Areas

Potential Experts

WARD, Derek, Alfred[26]	ENVIRONMENTAL SEALS LTD. [23];
von Bonin, Wulf[17]	Bayer Aktiengesellschaft [17];
Pirig, Wolf-Dieter,[8]	Clariant GmbH [5];
Feldman, Rubin[7]	Nu-Chem, Inc. [3];
Ward, Thomas A.[7]	PPG Industries, Inc. [7]

Total Time Trend

Country of Issue

Companies Leaving the Sector

Assignee	Number of Patents	Year of Last Publication
Cities Service Oil Company	35	1977
Imperial Chemical Industries Limited	12	1988
BFG GLASSGROUP	10	1984
Monsanto Research Corporation	7	1981

Figure 5: Tech mining one-pager

– in this case, patent assignees. We break out additional information for each assignee. Here, we see their leading inventors, patent classes, and temporal pattern.

- Trend plots in the lower left quickly convey change in activity. These can be elaborated, using scripts (macro's) to fit S-shape or other growth models and to extrapolate trends into the future.
- Patenting location or other geographic breakouts may be useful (here we present a pie chart by patent authority).
- Maps of various kinds show key activity concentrations and interactions. This map shows active R&D topics within the domain, and linkages among them.
- Special interest tabulations can be effective. In the lower right we see candidate experts (active inventors, not associated with large companies) and companies that appear to have exited this domain. (Another instance where TM invites further detective work – we could pursue why Cities Service ceased patenting in this technological domain?)

CONCLUSIONS

This short paper illustrates how vast science and technology (S&T) information resources can be mined to generate effective competitive technical intelligence. We've approached this in terms of technology management issues that generate questions, and innovation indicators help answer; a forthcoming book pursues these in detail (Porter 2005). We've also applied *VantagePoint/Derwent Analytics* software to derive the desired empirical indicators.

Tech mining takes advantage of several developments to make this work better:

- S&T data providers are working to make the data more accessible through suitable licensing options and coordination with text mining software.
- Use of scripting automates repetitive steps in data cleaning, analysis, and presentation. This drastically speeds up and reduces the cost of tech mining.
- Deriving innovation indicators from the data to get at key technology commercialization influences.
- Composite, tailored technology intelligence products (one-pagers) answer key questions to support decision-making.

This last point can be extended toward standardized technology intelligence products. They can be tailored to specific requirements of strategic business decision processes. Standardization can make a dramatic leap forward in managerial familiarity with TM and its results, thereby fostering utilization. Search Technology, with National Science Foundation support, worked with Merrill Brenner of Air Products to explore ways to enable quick technology intelligence processes.

Text mining can and will play an increasing role in technology management because it provides competitive advantage. Many management specializations have become increasingly data-driven over the past decades. For instance, traditional manufacturing process management relied on floor supervisors' tacit knowledge to determine if things were working well. As this intuition came to be augmented by empirical information and statistical analyses, quality control leaped forward. There would be no six sigma quality without this enhanced knowledge.

Analogously, we foresee tech mining advancing technology management by bringing to bear better knowledge of R&D advances. In an age where companies cannot perform all technological development in-house, this knowledge is essential to good technology management. Professionals and managers who take advantage of this better, quicker, and richer CTI will outperform their peers.

[We gratefully acknowledge research support from the National Science Foundation (Project DMI 0872482, 1998-2001) and The Center for Innovation Management Studies at North Carolina State University (2001-2002).]

REFERENCES

Teichert, T. and Mittermayer, M-A, (2002) 'Text mining for technology monitoring,' *IEEE IEMC*, p596-601.

Porter, A.L. et al. (2004). 'Getting what you need from technology information products,' *Research Technology Management*.

Porter, A.L. and Newman, N.C. (2001) 'Why don't managers want our technological intelligence? And what can we do about it?' SCIP annual conference, Seattle.

Watts, R.J. and Porter, A.L. (1997). 'Innovation forecasting,' *Technological Forecasting and Social Change*, v56 p25-47.

Porter, A.L. and Cunningham, S.W. (2005). *Tech Mining: Exploiting New Technologies for Competitive Advantage.* Wiley.

Alan L. Porter's major concentration is technology intelligence, forecasting and assessment. He has led development of "technology opportunities analysis" -- mining electronic, bibliographic data sources to generate intelligence on emerging technologies. Dr. Porter is Director of R&D for Search Technology, Inc., Norcross, GA. He is also Professor Emeritus of Industrial & Systems Engineering, and of Public Policy at Georgia Tech, where he remains with the Technology Policy and Assessment Center. He is author of some 200 articles and books, including Tech Mining, due in late 2004 from Wiley.

Let *VantagePoint* help you find

Who?

What?

Where?

When?

… so you can focus on

Why?

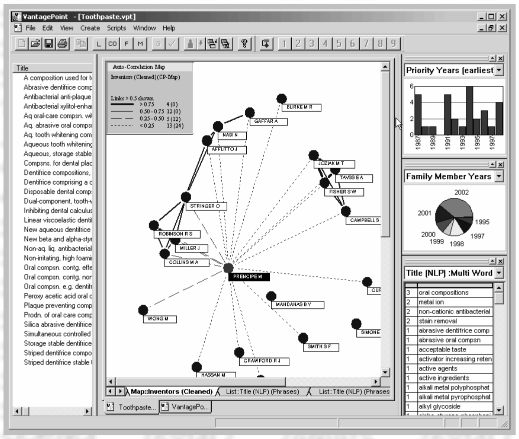
How?

… and most importantly

What to do next!

Extract Patent Data in a Variety of Useful Ways: VantagePoint is adaptable to virtually all patent data sources and formats your data the way *you* want it. ***Clean up Text Data:*** Let's face it – patent data are not perfect. VantagePoint will help you conquer text shortcomings so you can analyze your data with confidence. ***View Your Data from Many Vantage Points:*** Use graphs, tables, matrices and maps to guide your analysis in the most productive directions. ***Automate Your Analysis Process:*** Save time, improve consistency, and reduce human error by encoding your repetitive analyses in Java or VBScript.

vantage point



Title
A composition used for t
Abrasive dentifrice comp
Antibacterial anti-plaque
Antibacterial xylitol-enha
Aq oral-care compsn. wi
Aq. abrasive oral compsi
Aq. tooth whitening com
Aqueous tooth whitening
Aqueous, storage stable
Compsns. for dental plac
Dentifrice compositions,
Dentifrice comprising a c
Disposable dental comp
Dual-component, tooth-v
Inhibiting dental calculus
Linear viscoelastic denti
New aqueous dentifrice
New beta and alpha-styr
Non-aq. liq. antibacterial
Non-irritating, high foami
Oral compsn. contg. effe
Oral compsn. contg. nor
Oral compsn. e.g. dentifr
Peroxy acetic acid oral c
Plaque preventing comp
Prodn. of oral care comp
Silica abrasive dentifrice
Simultaneous controlled
Storage stable dentifrice
Striped dentifrice compo
Striped dentifrice stable

Title (NLP) :Multi Word

3	oral compositions
2	metal ion
2	non-cationic antibacterial
2	stain removal
1	abrasive dentifrice comp
1	abrasive oral compsn
1	acceptable taste
1	activator increasing reten
1	active agents
1	active ingredients
1	alkali metal polyphosphat
1	alkali metal pyrophosphat
1	alkyl glycoside

Search Technology

770.441.1457

VantagePoint @ searchtech.com

www.theVantagePoint.com