Preprint – to appear in *Technovation, 2011*

Mining External R&D

Alan L. Porter and Nils C. Newman
Search Technology, Inc.

Open Innovation presses the case for timely and thorough intelligence concerning research and development activities conducted outside one's organization. To take advantage of this wealth of R&D, one needs to establish a systematic "tech mining" process. We propose a 5-stage framework that extends literature review into research profiling and pattern recognition to answer posed technology management questions. Ultimately, one can even discover new knowledge by screening research databases.

Once one determines the value in mining external R&D, tough issues remain to be overcome. Technology management has developed a culture that relies more on intuition than on evidence. Changing that culture and implementing effective technical intelligence capabilities is worth the effort. P&G's reported gains in innovation call attention to the huge payoff potential.

1. Why interest in external R&D?

For decades, R&D was synonymous with internal research and development. However, in recent years attention to external R&D has become significant, largely in conjunction with "**CTI**" – Competitive Technical Intelligence. Large companies, defense agencies, and others have a need to know what technologies their key competitors are pursuing – *competitive intelligence*. Given increasing bottom-line implications of intellectual property management, it is also essential to know what's happening in a given technological arena before one invests heavily in R&D – *technical intelligence*. CTI addresses both the "who" and "what" of external R&D.

In the past few years, the stakes have escalated dramatically, led by the call for "Open Innovation." Chesbrough's (c.f., 2003; 2006) exposition of the merits of two-way interchange of intellectual property to bolster technological innovation has proven a call to action. Huston and Sakkab (2006) describe how Procter & Gamble (P&G) has put this into practice to the tune of 35% of their innovative new product elements deriving from external R&D -- driving home the merits of Open Innovation. Open Innovation is bi-directional. Inbound reflects companies taking advantage of external R&D to innovate in their products and services; conversely, outbound entails working proactively to license one's own research-based knowledge to others. CTI serves both (Porter, 2007).

Several additional innovation approaches reinforce the message that knowledge of external R&D is vital. The notion of the "fuzzy front end" (c.f., Smith and Reinertsen, 1998) places a premium on the early idea stages in innovation processes. We want to know about all of "**ST&I**" – Science, Technology and Innovation (i.e., research underway and research publications, as well as patents and new product introductions). Gibbons et al. (2004) distinguish research activities as Mode 1 vs. the nascent Mode 2. Mode 1 reflects the traditional researcher-initiated pursuit of disciplinary topics of intellectual interest. Mode 2's interdisciplinary, problem-driven research pulls academic researchers toward industrial initiatives and potential university-industry collaboration. Mode 2 aligns well with technology-based (e.g., information technology) innovations, and even more so with science-based innovations (e.g., biotechnology and nanotechnology). The "triple helix" conceptualization (Etzkowits and Leydesdorff, 2000) directs attention to interchange of research knowledge among universities, industry, and government. "Pasteur's Quadrant" (Stokes, 1997) further alerts us to the rich possibilities of use-inspired basic research, as exemplified by Pasteur's work.

Radical innovation is also on the radar screen. The notion of Accelerating Radical Innovation ("ARI") casts a wider reach to exploit ideas arising beyond the immediate technology delivery system (Dismukes et al., 2005; Leifer et al., 2000). All of these emergent concepts – Open Innovation, fuzzy front end, Mode 2, Triple Helix, Pasteur's Quadrant, and ARI – *point to the desirability of companies' awareness and intellectual interchange concerning externally conducted research* (especially in academia). We also note the complementary value in university researchers' knowing about industry's innovation interests. CTI serves these ends (c.f., Kostoff, 2006).

2. Five stages in mining external R&D

So, how does one gain knowledge about R&D conducted outside one's organization? We distinguish five stages to advance thinking about "what" information and "how" to obtain it. To the best of our knowledge, this is a new perspective. As such, these notions warrant examination, experimentation, and refinement. Table 1 aims to frame an energetic debate, not to constitute the final answer.

Table 1. Five Stages in Mining External R&D Knowledge

Stage	Mechanism
1	Literature review (within research community)
2	Research Profiling
3	Tech Mining
4	Structured Knowledge Discovery
5	Literature-Based Discovery (LBD)

Each stage adds a significant degree of richness and value, but without supplanting the previous stages. Research profiling goes beyond the traditional literature review to comprehend what's happening relevant to the target research interests beyond the traditional 20 or so duly noted references of a typical research article. Research profiling summarizes thousands of articles (Porter et al., 2002). But, this does not mean that one need not also read and summarize the most salient sources as per the literature review. Tech Mining changes the scope from understanding the research milieu to answering pointed Management of Technology ("MOT") questions. It calls attention to multiple information resources and ways to expedite delivery of CTI.

Stages 1, 2, and 3 all seek to locate and utilize existing research-based knowledge. Stages 4 and 5 strive to create new knowledge – i.e., discover what is not yet known. Our section on Stage 4 relates a recent such discovery based on mining genomics information. Stage 5 carries discovery a notch further to generate new knowledge possibilities based on mining research literature.

3. Stage 1: The Literature Review

Stage 1 merits a bit of attention, even if our excitement lies in the additional stages. Let's begin with the classic stereotype (whether or not it ever was accurate). Research thrives within disciplines – actually within narrower, highly specialized, even isolated, research streams (Chubin and Connolly, 1982). Within a given research area, one knows "everyone" who is working on its core challenges. Students follow their mentors, citing the well-known research immediately pertinent to their study. How does this research become well known to them? They read it, of course.

Porter and colleagues (2007) have been studying research interdisciplinarity. In broad samples of research papers, what is most striking is how interdisciplinary research has become. For instance, in some 2200 papers published in 2005 across 22 Subject Categories,[1] fewer than 10% are single-authored. The average number of authors per paper is increasing over time – e.g., in Biotechnology, doubling from 2.5 (1975) to 4.8 (2005) One indicator of cross-disciplinary knowledge transfer is whether papers published in a journal associated with one Subject Category reference papers published in others. To an amazing degree they do. Our measure of the diversity of knowledge resources upon which a paper draws -- "integration" (Porter et al., 2007) -- is increasing over time (Porter and Rafols, 2009). In 1975, the average Biotechnology paper referenced 6.8 Subject Categories; in 2005, 12.4. Just to give the flavour, one highly integrative paper in the Subject Category, "Materials Science, Biomaterials" (that title alone suggests how cross-disciplinary research is becoming!) references 17 Subject Categories:

BIOPHYSICS
MATERIALS SCIENCE, MULTIDISCIPLINARY
PHYSICS, ATOMIC, MOLECULAR & CHEMICAL
CHEMISTRY, PHYSICAL
PHARMACOLOGY & PHARMACY
POLYMER SCIENCE
CARDIAC & CARDIOVASCULAR SYSTEMS
PATHOLOGY
CHEMISTRY, ANALYTICAL
ENGINEERING, BIOMEDICAL
MULTIDISCIPLINARY SCIENCES
BIOTECHNOLOGY & APPLIED MICROBIOLOGY
PHYSICS, APPLIED
MATERIALS SCIENCE, BIOMATERIALS
ELECTROCHEMISTRY
COMMUNICATION
PHYSICS, MULTIDISCIPLINARY

Who can keep up with research advances by reading the literature over such a span? As Harmston et al. (2010) put it, "we may waste research time and resources on reinventing the wheel simply because we can no longer maintain a reliable grasp on the published literature."

How much scholarly attention is devoted to the literature review process? Not very much, but we find indications of merit in capturing more information, with pointers toward literature categorization (Hart, 1998), systematic review of wider bodies of literature (Stiles and Mick, 1994), and structured, empirical review (counting activity levels) (c.f., Ottenbacher, 1983). Most importantly for our interests, *meta-analysis* has emerged as a critical way to consolidate knowledge from disparate research studies on a given topic (Glass et al., 1981; Lipsey and Wilson, 2001; Hedges, 1998). Harris Cooper's methodological monograph on *Integrating Research: A Guide for Literature Reviews* (1989) makes a strong case for systematizing literature review.

4. Stage 2: Research Profiling

"Research profiling" brings to bear modern search engine and text mining tools to characterize entire bodies of related literature (Del Rio et al., 2001; Porter et al., 2002). This is done with computer-based tools, rather than by reading. It exploits the wealth of accessible information in electronic abstract databases such as *MEDLINE* and *Science Citation Index*. When searches return hundreds or thousands, even tens of thousands, of relevant research,

[1] Thomson Scientific's Web of Science distinguishes some 244 Subject Categories of research concentration. These provide the most widely used means to operationalize the notion of cross-"discipline."

reading is not a sensible option. Given this wealth of information, one has to rely on techniques such as research profiling to help spotlight especially salient pieces worth reading in depth.

We promote the use of such aids by researchers, R&D portfolio managers, research funders, and technology managers. We seek knowledge about the full "body" of relevant literature beyond that obtainable by digesting individual pieces. That is, by analyzing the whole, patterns (trends, networks) emerge that could never be grasped, even by reading the thousands of papers or patents one at a time.

Within the abstract (or full text) records compiled in ST&I databases, one finds information on: authors, their institutions, their country, article or conference paper content (via title and abstract phrases, keywords, and classification coding), and dates. The content descriptors can be tabulated just as one would count numeric instances – i.e., text mining. Indeed, simple counts associated with such *lists* provide the first order of useful information – e.g., which authors and institutions are most actively publishing on your research topic?

Most any two such lists can be combined to create a co-occurrence *matrix*. For instance, we could cross author organizations by keywords to see what sub-topics each researcher emphasizes.

Further, relationships among terms based on their tendency to "co-occur" in the research records can be mapped. Such visualizations help see shared interests and associated topics. For example, one can map who collaborates with whom in one's research arena (social network analyses – c.f., Borner et al., 2004; Bozeman et al., 2001; Deroian, 2002). Or, one can map whether aspects of a proposed dissertation topic have been studied heavily to help discern promising novelty.

Payoffs from "research profiling" include:

- *What* issues are central? What techniques are emphasized?
- *Who* constitutes the scholarly community engaged in this particular research domain?
- *When* -- how is the research domain evolving over time? (c.f., Bordons et al., 2004)
- *Combinations* – what new techniques have recently been brought to use? On which problems?

In sum, research profiling provides a broader perspective. It locates one's research within the contextual landscape (c.f., Borner et al., 2003). This can help in understanding the larger research community to identify additional journals and conferences, and peers, who may not be part of the already-known, immediate research stream. Research profiles can elucidate promising topics and point to novel methods.

5. Stage 3: Tech Mining

"Tech Mining" refers to text mining of ST&I information resources. As mentioned, Tech Mining recasts the approach. Instead of starting with the data and trying out various data mining approaches to see if anyone finds the results useful, begin with the Management of Technology (MOT) questions. Porter and Cunningham (2005) frame Tech Mining in terms of 13 recurring MOT issues, that spawn some 40 more explicit questions, which can be addressed via ~200 or so potential empirical "innovation indicators." Others have run with this concept to explore how best to gain value by it (Cunningham et al., 2006). We see this approach as a prime vehicle to enable Open Innovation processes (Porter, 2007).

Several analytical approaches come to bear in striving to answer various MOT questions:

- *Inductive Patterning* – one isn't limited to pre-specified categorizations. Instead, cluster and factor analyses can elucidate underlying groupings that speak to competitive concerns -- or collaborative opportunities.
- *Pattern Analyses* – one can see associations. For instance, one may be able to infer patent strategy by analyzing a company's patent activity in a target arena (e.g., "picket-fencing" as contrasted to active product innovation aims).
- *Trend Projection* – by extrapolating time series into the future based on alternative growth models, one can initiate technology forecasting analyses.
- *Blackspace Mapping* – by graphing patent (or other R&D indicator) activity, one may be able to discern areas of special opportunity in between intellectual property concentrations.
- *One-pagers* – one can compose compact and highly visual representations of key findings for easy managerial uptake.
- *Scripting* – an enabling technology. Analyses can be made faster and more efficient by writing simple programs to execute repetitive steps involving the text mining and other software. That can provide CTI for more MOT decision processes quickly – e.g., within a day. In turn, that means more decisions can be empirically based.

The Tech Mining perspective reminds us that there are multiple ST&I information resources to be tapped. Table 2 (from Porter, 2007) distinguishes six types. The columns distinguish between basic technical content and contextual content. Technical focuses on scientific & technical research & development activities. Given that the objective is to foster innovation, attention to socio-economic influences and success factors is also essential – the "contextual content." These can help spot impediments to the development and delivery of a commercial product or service. They can also help discern leverage mechanisms to overcome those blocks (e.g., sources of governmental funding, regulatory and standards bodies open to novel approaches).

Table 2. Competitive Technical Intelligence Resources

Source	Technological Content	Contextual Content
Databases (empirical knowledge)	**a**) Compiled, filtered & organized R&D publication, patent, etc. information	**b**) Compiled, filtered & organized business & socio-economic Information
Internet (empirical knowledge)	**c**) Diffuse, up-to-the-minute, ill-structured technical information	**d**) Diffuse, up-to-the-minute, ill-structured business & socio-economic Information
Human (tacit knowledge)	**e**) Technical Expertise (tacit knowledge)	**f**) Business/Context Expertise (tacit knowledge)

The rows partition information resources into three types. Each has strengths and weaknesses, so mining multiple sources strengthens the analyses. Top row: Electronic databases are limited to selected sources, usually with some lag time (e.g., S&T article abstracting and indexing takes time). But databases filter for quality control and standardize format for easy mining. Middle row: Internet sources (e.g., identified by Google search) are amorphous, inconsistently formatted, and lacking in quality standards. But they are incredibly rich and timely. Bottom row: Tacit expert knowledge is biased by a given person's limited information sources, interests, personal stakes, and memory. On the other hand, experienced technical and business experts have pronounced abilities to integrate and interpret disparate bits of information.

6. Tech Mining Case Example

Unfortunately, many fascinating cases of successful conduct of tech mining pertain to CTI and cannot be openly shared. We relate highlights of one performed by the US Army Tank and Automotive Command (TACOM) (Watts and Porter, 1997; with succeeding events treated in Porter and Cunningham, 2005, p. 284-285).

In the 1980's TACOM had pursued ceramics research, but then abandoned it as a failure. In the mid-1990's, Watts set out to explore whether the technology had matured sufficiently to warrant renewed Army pursuit. His R&D profiling included the compelling analyses condensed into Figure 1. These time slices of structural ceramics R&D associated with automotive engine applications illuminated a distinct maturational change. The back row shows that research rose strongly from 1985-86 to 1987-88, then plummeted. By 1993-95 it had rebounded only slightly. The middle row shows a similar pattern of only slight re-engagement of the topic by research organizations. The front row shows something quite different. This derived measure indicates that the complexity of the research jumped dramatically in the 1993-95 period, far exceeding that of the heyday of 1987-88. Breakout of the types of keywords (e.g., different materials, process attributes) showed a pattern of specialization. Experts confirmed that, indeed, pursuit of ceramic applications was within reach.

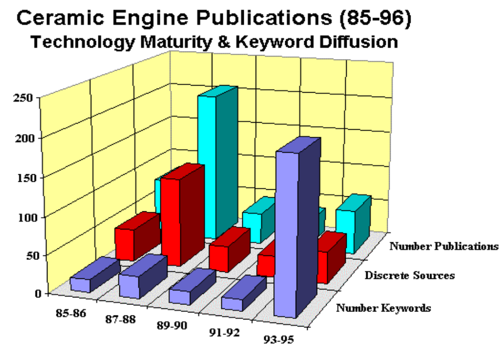

Figure 1. Ceramics R&D Pertaining to Engines: Changes over Time

This interpretation spurred a second stage of analyses to ascertain who was conducting R&D that had significant potential to remedy Army engine issues. Profiling of "who is doing what research" located truly cutting-edge work with significant potential to increase engine operating temperatures, and, consequently, efficiency, while reducing emissions. Surprising to the automotive engineers, this promising research was being conducted in an unrelated field. Microelectronics researchers were pursuing thin-film ceramic coating of metallic surfaces. This Tech Mining led to Army partnerships with a national lab and a specialty company to explore engine environment applications (which were totally new to those R&D units focusing on semiconductors). The ceramics performed very well. By 2005, a large processing facility was in operation to coat military engine parts for extended life and improved performance.

7. Stage 4: Structured Knowledge Discovery

This section might have been called, "doing research without a lab." The key concept is to mine a substantial set of information resources for patterns that lead to discovery. Note the ambitiousness here – we are saying that one can discover a new finding via pattern recognition – a finding that is not itself present in those data.

Bioinformatics is at the forefront of this approach. Instead of conducting experiments in the lab, researchers collect, collate and analyze existing research results to find new discoveries. An example of this approach was recently profiled in *The Economist* ("Going by the book," 12 January 2008). As outlined in the article, researchers in China analyzed over 1,000 research articles to isolate five main genes involved in addiction, their networks, and their possible feedback loops. A closer look at the article reveals that the information source used for the study was PubMed (Li, 2008), the medical research database maintained by the US National Library of Medicine.

This article speaks volumes to the notion of Open

Innovation. A research team in China searches a US maintained database to isolate research results from over 1,000 addiction articles from around 30 different countries. They then use bioinformatic (text mining) techniques to extract genomics data from those articles to build a database. That database is used to model biological processes. Finally they provide free access to their database so that others can build on their work (Li, 2008).

8. Stage 5: Literature-Based Discovery (LBD)

The basic LBD notion is simple (Figure 2). We explore the literature within research focus "A," thereby uncovering related literature "B," within which certain interesting elements appear. We then do a fresh search on one or more of those "B" elements, thereby uncovering related elements "C." We seek to discover promising "C" elements that have heretofore not been directly related to "A" (Kostoff, 2008).

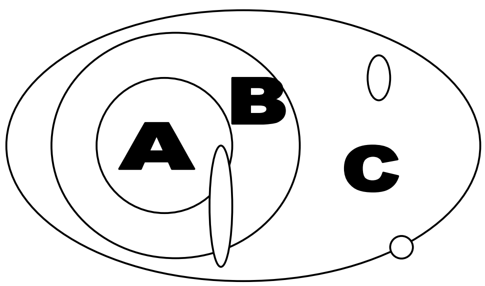

Figure 2. Literature-Based Discovery

In the initial LBD exercise, Swanson's (1986) research focus (A) was to find a treatment for Raynaud's disease (a disorder of narrowing of the blood vessels, usually in the extremities). A pertinent component of the vascular system is blood viscosity. So, a new literature search (B) centers on blood viscosity (not specifying Raynaud's). This uncovers, among others, an agent (C) – eicosapentaenoic acid (fish oil) – that lowers blood viscosity. Checking the literature finds no mention of fish oil as a treatment for Raynaud's. So, Swanson poses this solution to biomedical researchers to asses its efficacy.

LBD has been pursued since Swanson's pioneering work (re-examined thoughtfully by Gordon and Lindsay, 1996), through further biomedical studies (c.f., Swanson and Smalheiser, 1997) and tools, especially Arrowsmith (c.f., Swanson and Smalheiser, 1998). Kostoff and colleagues have demonstrated LBD via a series of case studies that include looking for treatments for cataracts and multiple sclerosis, and for new means for water purification (Kostoff, 2008). Petric et al. (2009) extend the methodology.

Others are increasingly pursuing variations on this theme. "SciMiner"was applied to assess genes and proteins apt to link to diabetic complications (Hur et al., 2010). Zhu et al. (2010) present a tool for finding non-obvious relationships between compounds (drugs) and biological properties (e.g., diseases). Ozgur et al. (2010) combine natural language processing and network centrality analyses to get at gene interactions. Others pursue a variety of potentially valuable links (c.f., Ahlers et al., 2007; Hettne et al., 2007; van Haagen et al., 2009; Baker and Hemminger, 2010).

LBD offers significant promise to multiply the value of research knowledge by identifying heretofore unexplored associations. But conducting LBD is challenging work that can benefit from Research Profiling and Tech Mining approaches. We want to develop better tools to facilitate LBD process by:

- *Selecting outreach arms* – by profiling the literature retrieved in search "B" to help separate and identify promising factors
- *Filtering* – by profiling the next wave of results – the solution literature ("C") to help identify discovery candidates
- *Vetting* – Kostoff (2008) makes the case that one needs to ascertain whether C has been associated with A in other literatures, projects, or patents. Cross-database list comparison is needed.
- *Representing* – multi-tier literature capture (A to B to C) escalates, potentially generating thousands of articles and many more conceptual "chunks." How can these be summarized and visualized so that a researcher can screen for potential discoveries?

We have composed ideas on how to frame discovery possibilities of different types. That is, looking for a disease solution differs in nature from searching for innovation prospects. However our research in this area is just beginning.

9. Discussion

Is your organization fully exploiting multiple information resources? Table 2 distinguished six types; each poses challenges, and those differ among organizations.

To exploit those rich repositories of knowledge, one needs a systematic process. To simplify, we can point to:

A. *MOT Decision Processes* – specifying in just what ways CTI can contribute
B. *Data Access* – identifying key sources and obtaining access
C. *Tools* – learning to use a suitable suite of analytical and visualization tools
D. *Incorporating CTI into Decision*-making – piloting, adapting, and standardizing technology intelligence products.

Step A may be the toughest. Most American technology managers favour intuition over empirical information as the essential basis for decision-making. Accepting and effectively utilizing empirical resources, together with (not in place of) human-based intelligence, is the root challenge.

Step B – obtaining Data -- does not come cheaply, and issues vary. Every leading American research university licenses the Web of Science for use by faculty, staff, and students, and, importantly, not charged on a per-record basis, so ideal for Tech Mining. Companies pay dearly for such data access. Based on limited sampling, fewer US universities license any major patent database. Most have a number of contextual databases (e.g., Lexus-Nexus). Google provides some exciting tools that help gather internet information, but with restrictions. We experimented with web site mining to compare the strategies used by some 400 highly innovative small firms to exploit their intellectual property (Porter et al., 2007).

Step C -- Tool access -- in our experience, the situation within Academia and Industry is inverted. Taking the case of our software, *VantagePoint*, large companies can handily afford it when compared to price they paid for data. However price is a significant consideration for academia since the cost of text mining tools is usually the burden of individual researchers or research centers.[2] Also Industry tends to have centralized business units engaged in tasks such as CTI which focuses and facilitates learning the use of text mining and visualization tools, even when tools are widely distributed within the organization (eg.. bench scientist)

Step D -- Incorporating CTI into MOT decision processes -- can be tackled in various ways. Air Products has shared its compelling experiences in mining multiple data sources, with multiple tool sets, for multiple payoffs (Brenner, 2005). A key leverage point is standardization. Tech Mining (Porter and Cunningham, 2005) asserts that a finite set of core MOT issues recur and can thus be effectively addressed via standardizing questions and tuning empirical indicators to each. In the Air Products model, their stage-gate process calls for explicit questions to be addressed at given stages, using prescribed data sources, and generating particular charts. Two advantages accrue from this:

1. *Familiarization* – senior researchers and executives come to recognize the CTI report elements and know how to interpret them effectively
2. *Scripting* – analyses can be expedited by writing mini computer programs (macros) to generate the desired charts and other outputs (not to supplant human analyst tuning and interpretation of same).

The elements just addressed contribute to all five stages of mining external R&D (Table 1). Managing those well takes focus on the value of the resulting CTI to further organizational interests. Based on our experiences, an organization's interest in CTI leans toward the immediate. Competitive intelligence tends to emphasize what one's key competitors are up to lately. Technical intelligence leans more toward the tactical needs of a current development rather than strategic, long term explorations.

A far bolder model is provided by a Korean institution. At Georgia Tech we have worked on several research projects with KISTI – the Korea Institute of Science and Technology Information. We clearly recall a particular front end meeting at which Porter asked his KISTI colleagues "on which research areas are we focusing?" and their reply – "all of them." And they meant it, and had figured out clever ways to go about this. They were scanning the world's research to identify research fronts that held particular promise for Korea. Such strategic outreach can welcome the discovery aspects of Stages 4 and 5 that we have introduced here.

We hope to initiate discussion, particularly, of the discovery possibilities in ST&I literature and patent resources. Open Innovation especially challenges us to more fully exploit externally generated research knowledge. Approaches rooted in meta-analysis are now enhanced by text mining tools and "at your fingertips" access to millions of R&D publication and patent abstract records (plus corresponding contextual, business-side resources too). So, we can do it, if we have the will.

To conclude, we pose three questions to consider:

- What discoveries from well-formulated research literature scans would offer potentially high payoff for your organization?
- Who in your organizations could organize and support both Tech Mining and discovery processes?
- What would it take for them to be successful?

[2] Search Technology is trying to facilitate access via special academic licensing – see www.theVantagePoint.com.

10. References

Ahlers C.B., Hristovski D., Kilicoglu H., Rindflesch T.C. (2007): Using the literature-based discovery paradigm to investigate drug mechanisms. *AMIA Annual Symposium Proceedings*, Vol. 11, p. 6-10.

Baker, N.C., Hemminger, B.M. (2010): Mining connections between chemicals, proteins, and diseases extracted from Medline annotations. *Journal of Biomedical Informatics*, Vol. 43 (4), p. 510-519.

Bordons, M., Bravo, C.; Barrigon, S. (2004): Time-Tracking of the Research Profile of a Drug Using Bibliometric Tools. *Journal of the American Society for Information Science and Technology*, Vol. 55 (5), p. 445-461.

Börner, K., Chen, C.; Boyack, K. W. (2003): Visualizing Knowledge Domains, *Annual Review of Information Science and Technology*, Vol. 37, p. 179-255

Borner, K., Maru, J.T.; Goldstone, R.L. (2004): The Simultaneous Evolution of Author and Paper Networks, *Proceedings National Academies of Science USA* (April 6), 101 Suppl 1: p. 5266-73.

Bozeman, B., Dietz, J.; Gaughan, M. (2001): Models of Scientific Careers: Using Network Theory to Explain Transmission of Scientific and Technical Human Capital, *International Journal of Technology Management*, Vol. 22, p. 716-740.

Brenner, M.(2005): Technology intelligence at Air Products: Leveraging analysis and collection techniques, *Competitive Intelligence Magazine*, Vol. 8 (3), p. 6-19.

Chesbrough, H.W. (2003): The Era of Open Innovation, *MIT Sloan Management Review*, Vol. 44 (9), p. 35-41.

Chesbrough, H.W. (2006): *Open Innovation: The New Imperative for Creating and Profiting from Technology*, Harvard Business School, Cambridge, MA (paperback edition).

Chubin, D.E.; Connolly, T. (1982): Research Trails and Science Policies: Local and Extra-local Negotiations of Scientific Work, in N. Elias et al. (eds.), *Scientific Establishments and*

Hierarchies. Sociology of the Sciences, Yearbook Vol. 6, Dordrecht: D. Reidel, p. 293-311.

Cooper, H.M. (1989): *The Integrative Research Review: Moving Beyond Meta-Analysis*. Newbury Park, CA: Sage.

Cunningham, S.W., Porter, A.L.; Newman, N.C. (2006): Tech Mining Special Issue, *Technology Forecasting and Social Change,*Vol. Vol. 73 (8).

Del Rio, J.A., Kostoff, R.N., Garcia, E.O., Ramirez, A.M.,;Humenik, J.A. (2001): Citation mining: Integrating text mining and bibliometrics for research user profiling, *Journal of the American Society for Information Science and Technology,* Vol. 52(13), p. 1148–1156.

Deroian, F. (2002): Formation of social networks and the diffusion of innovations, *Research Policy,*Vol. 31(5), p. 835-846

Dismukes, J., Miller, L.K.; Bers, J.A. (2005): Technologies of Thinking Seen Key to Accelerating Radical Innovation, *Research Technology Management*, Vol. 48 (July-August), p. 2-4.

Etzkowitz, H.; Leydesdorff, L. (2000): The Dynamics of Innovation: from National Systems and "Mode 2" to a Triple Helix of University-industry-government Relations. *Research Policy,* Vol. 29 (2), p. 109-123.

Gibbons, M., Limoges, C., Nowotny, H., Schwartzman, S., Scott, P.; Trow, M. (1994): *The New Production of Knowledge: the Dynamics of Science and Research in Contemporary Societies,* Sage, London.

Glass, G., McGaw, B.; Smith, M. (1981): *Meta-analysis in Social Research*. Beverly Hills, CA: Sage.

Going by the book. (12 January 2008): *The Economist*, p. 71.

Gordon, M. D.; Lindsay R.K. (1996): Toward Discovery Support Systems: A Replication, Re-Examination, and Extension of Swanson's Work on Literature-Based Discovery of a Connection between Raynaud's Disease and Fish Oil. *JASIS*, Vol. 47(2), p.116-28.

Harmston, N., Filsell, W., Stumpf, M.P. (2010): What the papers say: Text mining for genomics and systems biology. *Human Genomics*, Vol. 5 (1), p. 17-29.

Hart, C. (1998): *Doing a Literature Review*. London/Thousand Oaks, CA: Sage.

Hedges, L.V. (1998): *Statistical Methods for Meta-Analysis*. Orlando, FL: Academic Press.

Hettne K.M., Weeber M., Laine M.L., ten Cate H., Boyer S., Kors J.A., Loos B.G., (2007): Automatic mining of the literature to generate new hypotheses for the possible link between periodontitis and atherosclerosis: lipopolysaccharide as a case study. *Journal of Clinical Periodontology*, Vol. 34 (12), p. 1016-1024.

Hur J., Sullivan K.A., Schuyler A.D., Hong Y., Pande M., States D.J., Jagadish H.V., Feldman E.L. (2010): Literature-based discovery of diabetes- and ROS-related targets. *BMC Medical Genomics*, Vol. 3, 49.

Huston, L; Sakkab, N. (2006): Connect and Develop, *Harvard Business Review*, March, 58-66.

Kostoff, R.N. (2006): Systematic Acceleration of Radical Discovery and Innovation in Science and Technology. *Technological Forecasting and Social Change.* Vol. 73 (8), p. 923-936.

Kostoff, R.N. (2008): Literature-Related Discovery (Special Issue), *Technological Forecasting and Social Change*, Vol. 75 (2).

Leifer, R., McDermott, C.M., O'Connor, G.C., Peters, L.S., Rice, M.P.; Veryzer, R.W. (2000): *Radical Innovation: How Mature Companies Can Outsmart Upstarts,* Harvard Business School Press, Boston.

Li C-Y, Mao X; Wei L (2008): Genes and (Common) Pathways Underlying Drug Addiction. *PLoS Computational Biology* 4(1): e2. doi:10.1371/journal.pcbi.0040002.

Lipsey, M.W.; Wilson, D.B. (2001): *Practical Meta-Analysis.* Thousand Oaks, CA: Sage.

Ottenbacher, K. (1983): Quantitative Reviewing: The Literature Review as Scientific Inquiry, *American Journal of Occupational Therapy*, Vol. 37 (5), p. 313-319.

Ozgür A., Xiang Z., Radev D.R., He Y. (2010): Literature-based discovery of IFN-gamma and vaccine-mediated gene interaction networks. *Journal of Biomedical Biotechnology*, Epub 2010, p. 426479.

Petric I., Urbancic T., Cestnik B., Macedoni-Luksic M. (2009): Literature mining method RaJoLink for uncovering relations between biomedical concepts. *Journal of Biomedical Informatics*, Vol. 42 (2), p. 219-227.

Porter, A.L. (2007): Tech Mining to Drive Open Innovation, in Li, J-L, Zhu, D., Porter, A.L.; Wu, D.D., *Proceedings of the First International Conference on Technology Innovation, Risk Management and Supply Chain Managemen*t (TIRMSCM 2007, Beijing), Universe Academic Press, Toronto, ISBN 978-0-9783484-4-1, p. 1-13.

Porter, A.L., Cohen, A.S., Roessner, J.D.; Perreault, M. (2007): Measuring Researcher Interdisciplinarity, *Scientometrics,* Vol. 72(1), p. 117-147.

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

Porter, A.L., Kongthon, A.; J-C Lu (2002): Research profiling: Improving the literature review, *Scientometrics*, Vol. 53, p. 351-370.

Porter, A.L.; Rafols, I. (2009) Is science becoming more interdisciplinary? Measuring and mapping six research fields over time, *Scientometrics*, Vol. 81(3), p. 719-745.

Porter, A.L., Schoeneck, D.J., Frey, P.R., Hicks, D.M.; Libaers, D.P. (2007): Mining the Internet for Competitive Technical Intelligence, *CI Magazine*, Vol. 10 (5), Sep/Oct, p. 24-28.

Smalheiser, N.R.; Swanson, D.R. (1998): Using ARROWSMITH: a computer-assisted approach to formulating and assessing scientific hypotheses, *Computer Methods and Programs in Biomedicine*, Vol. 57 (3): p. 149-153.

Smith, P.G.; Reinertsen, D.G. (1998): *Developing Products in Half the Time: New Rules, New Tools,* John Wiley, New York (2d edition).

Stiles, R.A.; Mick, S.S. (1994): Classifying Quality Initiatives: A Conceptual Paradigm for Literature Review and Policy Analysis, *Hospital & Health Services Administration,* Vol. 39 (3), p. 309-326.

Stokes, D.E. (1997): *Pasteur's Quadrant: Basic Science and Technological Innovation*, Brookings Institution Press, Washington, D.C.

Swanson, D.R. (1986): Fish oil, Raynauds syndrome, and undiscovered public knowledge, *Perspectives in Biology and Medicine,* Vol. 30 (1): p. 7-18.

Swanson, D. R.; Smalheiser, N.R. (1997): An Interactive System for Finding Complementary Literatures: a Stimulus to Scientific Discovery. *Artificial Intelligence,* Vol. 91 (2), p. 183-203.

Swanson, D.; Smalheiser, N. (1998): Using Arrowsmith: a Computer-assisted Approach to Formulating and Assessing Scientific Hypotheses. *Computer Methods and Programs in Biomedicine*, Vol. 57, p. 149-53. [http://arrowsmith.psych.uic.edu/arrowsmith_uic/index.html]

van Haagen H.H., 't Hoen P.A., Botelho Bovo A., de Morrée A., van Mulligen E.M., Chichester C., Kors J.A., den Dunnen J.T., van Ommen G.J., van der Maarel S.M., Kern V.M., Mons B., Schuemie M.J. (2009): Novel protein-protein interactions inferred from literature context. *PLoS One,* Vol. 18:4(11):e7894.

Watts, R.J.; Porter, A.L. (1997): Innovation Forecasting," *Technological Forecasting and Social Change*, Vol. 56, p. 25-47.

Zhu, Q., Lajiness M.S., Ding Y., Wild D.J. (2010): WENDI: A tool for finding non-obvious relationships between compounds and biological properties, genes, diseases and scholarly publications. *J. Cheminformatics* Aug. 26, p. 2-6.