TECH MINING
EXPLOITING NEW TECHNOLOGIES FOR COMPETITIVE ADVANTAGE

ALAN L. PORTER
SCOTT W. CUNNINGHAM

WILEY-INTERSCIENCE
A JOHN WILEY & SONS, INC., PUBLICATION

TECH MINING

TECH MINING

EXPLOITING NEW TECHNOLOGIES FOR COMPETITIVE ADVANTAGE

ALAN L. PORTER
SCOTT W. CUNNINGHAM

WILEY-INTERSCIENCE
A JOHN WILEY & SONS, INC., PUBLICATION

Published by John Wiley & Sons, Inc., Hoboken, New Jersey.
Published simultaneously in Canada.

Library of Congress Cataloging-in-Publication Data:
Porter, Alan L.
Tech mining : exploiting new technologies for competitive advantage / by Alan L. Porter and Scott W. Cunningham.
p. cm.
Includes bibliographical references and index.
ISBN 0-471-47567-X (cloth)
1. Data mining. 2. Technological innovations—Economic aspects. 3. Research, Industrial. I. Cunningham, Scott W. II. Title.

QA76.9.D343P67 2005
005.74—dc22

2004042241

Printed in the United States of America

10 9 8 7 6 5 4 3 2 1

Table of Contents

List of Figures

Preface

In the "information economy," we recognize the increasing availability of information. On the one hand, we can be intimidated by the overwhelming amount of information bearing down on us. On the other hand, we now have tools to enable us to garner great value from that information quite readily. New information products can better inform decision processes. As businesses are making decisions under tremendous competitive pressures, they increasingly seek better information.

This book addresses how to inform technology management by mining a particularly rich information resource—the publicly accessible databases on science and technology. These include amazing compilations of the world's open R&D literature, patents, and attendant business and public aspects. This information, when integrated with other data sources (the Internet) and expert review, can improve decisions concerning development, licensing, and adoption of new technology.

"Tech mining" presents particular challenges. Most fundamentally, it uses information resources in unfamiliar ways. In the past, we searched abstract databases to find a few articles worth reading. However, when there are literally thousands of relevant articles or patents, we also need to present the "big picture." This book helps understand the value in "profiling research domains," mapping topic relationships, and discerning overall trends. This is a qualitatively different way to use technology information.

We wrote *Tech Mining* for those whose jobs engage emerging technologies. This includes two groups. Part I addresses those who use such studies, rather than perform them. We seek to help such professionals and managers become better informed consumers of tech mining. We inform engineers, researchers, product developers, business analysts, marketing professionals, and various technology managers on effective ways they can exploit these information resources. Part II adds "how to" details for those who analyze, or directly manage the analysis of, changing technologies. This includes information professionals, patent analysts, competitive intelligence specialists, R&D managers, and strategic planners.

This book is a primer. It sets forth the basic objectives and tools of tech mining. Chapters 1–5 aim to provide conceptual bases for practical tech mining actions. The conceptual foundations reside in understanding of how science and technology leads to successful technology commercialization (the innovation process) more than in information science. Chapters 6–16 provide practical advice on performing tech mining. These treat basic and advanced analyses but also process management considerations vital to effective implementation. We carry through to point to products of tech mining analyses and indicate how they can serve particular technology management functions. Chapter 13 arrays technology management issues and questions along with candidate "innovation indicators" to answer them.

Each chapter focuses on a particular aspect of tech mining. It explains the relevant aims, presents the basic steps in accomplishing those aims, and provides pointers to those who want further details. We illustrate the content with experiential cases slanted toward practical implementation issues and how results can be used. Some chapters work through a "chapter challenge" to think through application of the concepts presented.

Chapters 4 and 16 together step through a concrete analytical example. This applies *VantagePoint* software to actual abstract records obtained from three databases (Derwent World Patent Index, INSPEC, and Web of Science) on the topic of "fuel cells." Chapter 4 spotlights sample tech mining results to get you thinking of ways you could gain value from tech mining. Chapter 16 illustrates the analytical progression and notes pitfalls. The Wiley website ftp://ftp.wiley.com/public/sci_tech_med/technology_management offers a sample data set in *VantagePoint Reader* to experience the tech mining analyses directly.

The book does not require any statistics or artificial intelligence background. It is not specific to a particular technology domain (e.g., information technology). In addition to practitioners and managers, we believe it can benefit technology analysis workshops and graduate courses.

Acknowledgments

We expressly thank IEEE and Thomson Scientific for their permission to utilize their data in our sample analyses reported in the book and illustrated via sample data on the web at ftp://ftp.wiley.com/public/sci_tech_med/technology_management. We thank Wiley for taking this venture to publication. Our series editor, Andy Sage, has exerted major influence in forming the content toward a meaningful composition.

We thank our kind colleagues who took the time and energy to review one or more of the draft chapters. Some were kindly and said nice things; we liked that. Others were mean and found lots of problems; we didn't enjoy those nearly as much, but we and you owe them a hearty "Thanks!" for making the book better.

Erik Ayers
Kevin Boyack
Tony Breitzman
Merrill Brenner
Linda Carton
Joe Coates
Patrick Duin
Paul Frey
Arnaud Gasnier
Luke Georghiou
Russ Heikes
Leon Hermans
Diana Hicks
Katherine Jakielski
Sylvan Katz
Alisa Kongthon
Ron Kostoff
Loet Leydesdorff
Hal Linstone
Vincent Marchau
Brian Minsk
Nils Newman
Doug Porter
Scott Radeker
David Roessner
Fred Rossini
Phil Shapira
Robert Tijssen
Tony Trippe
Robert Watts
Julie Yang

Acronyms & Shorthands—Glossary

(We also try to define these on first use in a chapter; this is just for ready reference.)

AI	Artificial intelligence
ALP	Alan Porter, author
CTI	Competitive technological intelligence
DWPI	Derwent World Patent Index, or "Derwent" for short
EPO	European Patent Office
GTEL	Georgia Tech Electronic Library
INSPEC	A major database (covering physical and information sciences and engineering, especially electrical engineering; provided by IEE)
IP	Intellectual property
IT	Information technology
JPO	Japanese Patent Office
KDD	Knowledge discovery in databases
NLP	Natural language processing
PCA	Principal components analysis
PCD	Principal components decomposition
PCT	Patent Cooperation Treaty
R&D	Research and development
S&T	Science and technology
SCI	Science Citation Index (a database covering fundamental research, found at Thomson Scientific's Web of Knowledge)
SWC	Scott Cunningham, author

TIPs	Technology information (or intelligence) products (outputs of tech mining analyses)
Tech mining	Our shorthand for text mining applied to records, particularly science and technology abstracts retrieved from databases
USPTO	United States Patent and Trademark Office
VP	*VantagePoint* tech mining software; also available under the names TechOasis or Derwent Analytics.
WIPO	World Intellectual Property Organization

Part I

Understanding Tech Mining

Chapter **1**

Technological Innovation and the Need for Tech Mining

"Tech mining" is our shorthand for exploiting information about emerging technologies to inform technology management (see the Preface). This chapter anchors tech mining to technological innovation processes and payoffs. It keys on the two contextual forces that drive the book: "emerging technologies" and the "information economy." Chapter 2 builds on this to explain what tech mining entails and to the describe the book's organization.

1.1. WHY INNOVATION IS SIGNIFICANT

We use "innovation" to mean technological change. We are concerned with technological change resulting in practical implementation or commercialization, not just idea generation. This section addresses the importance of technological innovation to today's competitive economy and polity.

Today's worldwide economy depends on technology and technological innovation to an extraordinary degree:

- We perform a lot of research—for one thing, American companies spend over $100 billion annually on R&D; for another data point, the Organisation for Economic Cooperation and Development (OECD) countries spent over $550 billion in 1999 (about 70% by companies, 30%

Tech Mining: Exploiting New Technologies for Competitive Advantage, Edited by Alan L. Porter and Scott W. Cunningham.
ISBN 0-471-47567-X

by government).* That research pays off—participating companies in the U.S. Industrial Research Institute estimate average new sales ratio—the percentage of sales attributable to products newly designed in the past five years—at roughly 35%. In other words, $1 of every $3 in their revenue comes from recent innovations.

- National economies depend critically on technology. "High-Tech Indicators" (http://tpac.gatech.edu) show that the U.S. once dominated technology-based export competition. Then Japan raced up to become a staunch competitor. Now, other countries are advancing dramatically. Tiny Singapore now exports technology-based products at the level of the European powers. China is advancing dramatically in technology-based exports, but also in R&D that will drive future generations of products and services. And they are not the only ones looking ahead. The 371 expert panelists anticipate that, in another 15 years, essentially all of the 33 countries tracked will be significant high-tech competitors (Fig. 1-1).

Technological innovation impacts our lives in many ways, some direct and some not so direct. High-technology companies are a significant and growing component of the economy, contributing over 20 million jobs in the U.S. (Hecker, 1999). The competitiveness of those companies depends on innovation, credited with being the main economic growth factor in the western world.

Innovation delivers substantial public and private returns. Mansfield's classic survey (1982) on 37 innovations concluded that the firm's median return on investment was close to 25 cents on every dollar. And the public benefits of innovation far outweigh the firm's benefits—70 cents on every dollar spent on R&D is returned to society. Despite these rosy average returns, Mansfield and others find that innovation is highly risky, and failure can be immensely costly. In some cases, companies bet their existence on the success of an innovation.

Innovation is improving our standard of living. Developments in medical and pharmaceutical technologies have delivered extensive returns in health and life span. The toddler of today can expect 25 more years of life than the newborn of the year 1900. Death rates from infectious disease have been reduced 10-fold over the course of the previous century. However, we remain engaged in an evolutionary war with infectious disease and continue to struggle with cancer and vascular diseases (Lederberg, 1997). Our health and welfare is intimately linked to innovation.

Without belaboring it, innovation is vitally important to scientists and engineers, private and public organizations, and society. Our key underlying premise is that tech mining facilitates innovation. To accomplish this, tech

*OECD Science, Technology and Industry Scoreboard 2001: Towards a knowledge-based economy [http://www1.oecd.org/publications/e-book/92-2001-04-1-2987/]

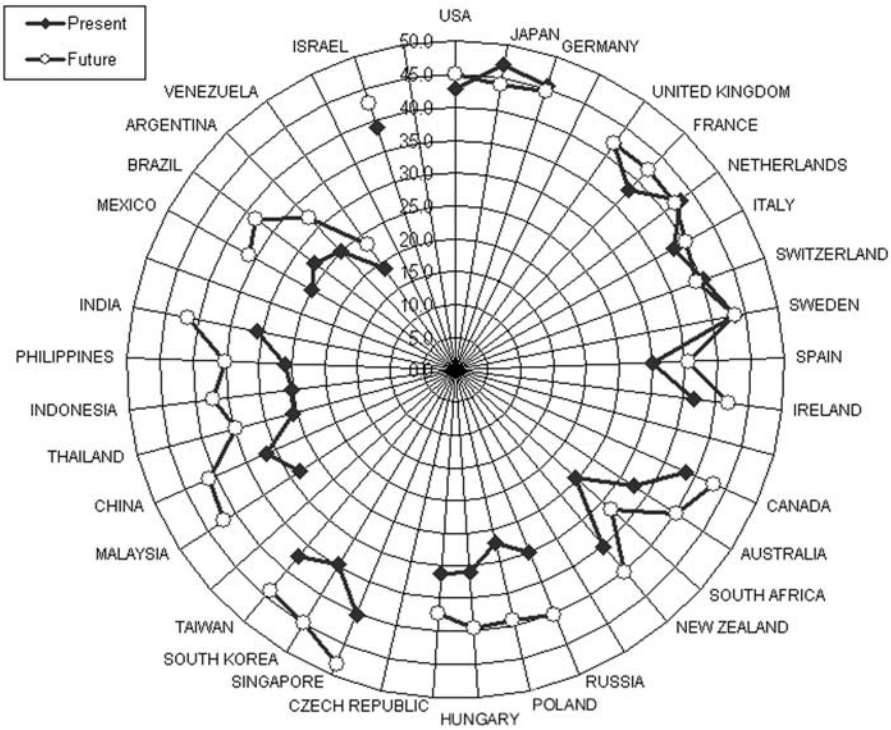

Figure 1-1. Increasing national technological competitiveness. This radar chart summarizes scaled opinions of knowledgeable observers on the relative ability of each country to compete in high-tech-based exports. Scores farther out from the center reflect relatively stronger competitiveness. Note the almost universal expectation of increasing competitiveness. For details see "High-Tech Indicators" at //tpac.gatech.edu.

mining relies on understanding technological innovation processes to track them effectively and to inform decisions about R&D and subsequent implementation and adoption choices.

1.2. INNOVATION PROCESSES

Our colleague Mary Mogee defines innovation (1993) as "the process by which technological ideas are generated, developed and transformed into new business products, processes and services that are used to make a profit and establish marketplace advantage." Let's explore this *process* to figure out empirical measures deriving from innovation activities to generate actionable technological intelligence (tech mining).

We briefly scan the rich history of *models* of technological innovation processes (Fig. 1-2). Dating from the 1950s, the *technology push* model focused

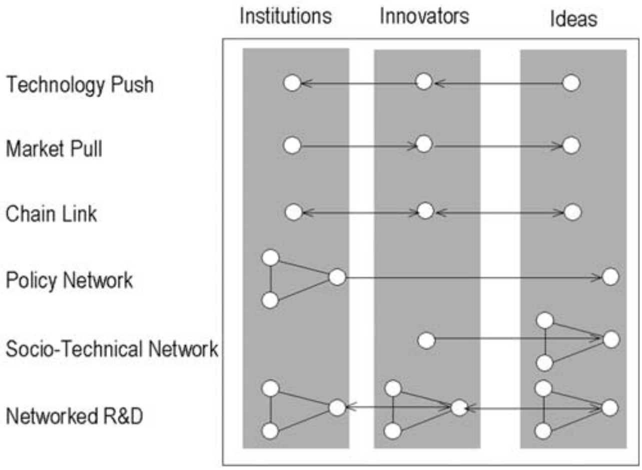

Figure 1-2. Comparison of R&D models

on R&D as generating the essential push that prompts new product development, which the marketplace then accepts. The realization that many innovators and institutions deliberately frame R&D to meet perceived market opportunities suggested the *market pull* model. This reversed the main influence pathway to begin from the customer end. The *chain link* model offered a compromise between the two, acknowledging that flows between technology and the marketplace are iterative and multidirectional. This first class of models is basically singular in nature—one organization generates new technology and takes it to market.

A second class of models recognizes interplay among institutions in generating and acting on science and technology. The *policy network* approach acknowledged that institutions exist in a framework of competitive and collaborative relationships. As the great central R&D laboratories (e.g., Bell Labs, IBM) shrank and distributed activities among operating divisions, companies turned outward for science and technology inputs. Governmental and academic R&D eased from the isolated, single-investigator model of science. Organized research units fostered interdisciplinary and interinstitutional collaborations. Institutions share and compete for the R&D findings of innovators, with significant knowledge spillovers. Notions such as regional innovation centers emerged to bolster purposeful interchange of science and technology approaches and results. A complimentary perspective, the *socio-technical systems* approach, examined how different innovators link and unify ideas. This evolution points toward *networks* of concepts and material objects ("artifacts") forming a stratum for the creation and dissemination of new science and technology knowledge, resulting in technological innovation (Fig. 1-2). (Chapter Resources adds pointers to continuing refinement of such networking models.)

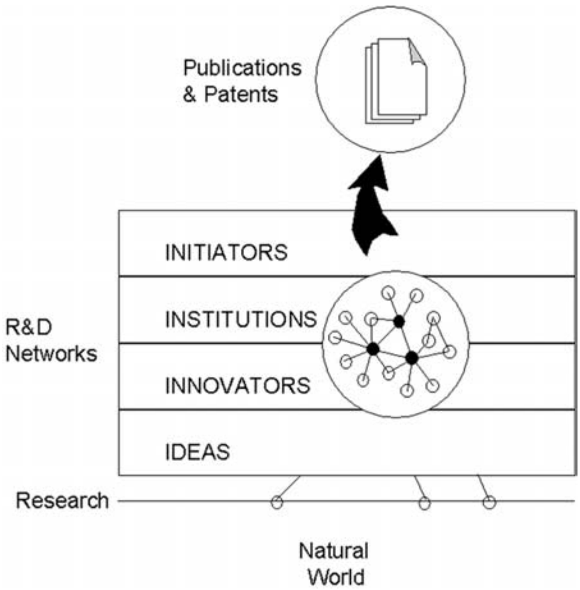

Figure 1-3. A networked, instrumented model of innovation

We see networks again and again—networks of researchers and networks of ideas. These "knowledge networks" are woven by many individuals—certainly by scientists and engineers—and also by the many institutions that support and fund new R&D activity—"initiators." Like the webs of knowledge they create, individuals and institutions find themselves in complex and interwoven relationships with other innovators. We distinguish four layers of networking activity (Fig. 1-3). *Ideas* compete and become interlinked. *Innovators* select, vary, and propagate the successful ideas. *Institutions* construct teams of innovators and cooperate and compete with other institutions. *Initiators* fund the research and development activities of institutions. At the foundation of the system lies the *natural world.* Ideas are tested constantly against the facts and needs of the real world.

This networking interchange provides the essential opportunity for tech mining. The various exchanges of science and technology information effectively *instrument* (document) knowledge at all four levels. How so? Innovators (scientists and technologists) produce findings. Institutions provide incentives for innovators to publish or patent those findings. The ideas used by the innovators are reflected in their publications and patents. Relationships among innovators can also be discerned from papers (journals and conferences) and patents. Also, the institutional arrangements, in funding, conducting, and disseminating R&D, often are reflected in the details of those publications and patents. So, publications and patents—as by-products of the exploitation and exploration of science and technology—provide a lot of insight into actual practices leading to technological innovation.

Innovation is significant! But how can tech mining assist in the innovation process? In the next two sections we examine innovators and innovative institutions in society. Our brief survey suggests challenges and needs faced by these groups and individuals.

1.3. INNOVATION INSTITUTIONS AND THEIR INTERESTS

Let's examine the institutions that fund and perform research, with an eye toward how tech mining can further their interests. At least five sources fund research—industry, government, education, nonprofit, and cross-national funding. Recent sources place industrial funding of R&D at more than 63 percent of the total (OECD, 2003). In the United States, the Federal government is the largest single source of R&D funding. "High-technology" manufacturers fund the highest portion of industrial R&D activity. Service-related R&D spending is much smaller, but a rapidly rising proportion of the total. Most industrial R&D focuses on "development"-related efforts. Notably, industry is the largest performer of R&D. Most of that is done by the largest companies (NSF, 2000). Data for 1997 show five leading U.S. companies contracting for $3 billion or more: GM, Ford, IBM, Lucent, and H-P.

Companies face multiple challenges in making those huge R&D investments (Tassey, 1999). Technology investment is inherently risky, and one's R&D often results in spillovers whereby others accrue benefits from it. So, before diving into an R&D program, the company needs to ascertain what existing knowledge might be capitalized upon. Tech mining can uncover external research results to save rediscovering that wheel. It can identify intellectual property ("IP") land mines before a substantial technology development program finds itself blocked.

If new development activities need to be initiated, one method of reducing risk is strategic partnership. This allows individual partners to leverage their resources, reduce costs, and enable activities that might not otherwise have been possible. Additional benefits for corporations may include speeding up development and reduced competition when the developed product reaches the marketplace (NSF, 2003). Tech mining can find out what R&D others are pursuing and pertinent IP so that you can determine the best route to your goals, possibly via partnering in some form.

Academia is a significant source of public science and the dominant generator of basic research findings. Government also plays a very substantial role—particularly through defense funding—in the support of new technology development. Industry is often involved in carrying technological developments to fruition, so it pays to keep tabs on university and governmental lab research activities.

Significant issues for innovators and their institutions include:

- How can we recognize and reward new and innovative ideas in our organization?
- How do we capitalize on the strengths of our knowledge to attract new funding?
- Can we attain new knowledge before our competitors?
- Can existing, publicly available knowledge provide us with needed solutions?

Note the extent to which these issues demand knowledge of others' science and technology activities—and tech mining can provide this.

The recognition that new products and processes are central to corporate renewal underlies these issues and why we care about them (Danneel, 2002). A careful balance must be sought between exploiting existing competencies and discovering and developing new competencies. New products close to existing core capabilities have a greater chance of success. Unfortunately, however, existing competencies can crowd out opportunities for growth—resulting in inflexibility and missed opportunities. Positive feedback causes innovative "path dependencies," that is, technological choices that lock a firm, agency, or academic unit in or out of specific development trajectories. With respect to tech mining, we need to track both internal and external technological capabilities. Procter & Gamble tells a story on itself. After submission of a patent application, they got back good news and bad. The bad—a patent had already been issued. The good—they held it. It's hard to keep track of your own technology, much less everyone else's.

The customer is also critical to new product development. New products build on a match of new ideas to existing competencies. Successful products stem from the intersection of customer need and technological competencies. Institutions that are successful in new product development must understand their customer—building upon the existing customer base and learning about new customers, or formerly unrecognized needs of old customers. This is another form of intelligence essential to successful innovation.

March (1991) characterizes the options of building upon old knowledge or reaching out to find new knowledge as "*exploitation*" versus "*exploration*." Exploitation is a process of linking—integrating existing knowledge, combining and recombining core competencies to meet market need. Exploration leverages what is known to gain new knowledge. Tech mining seeks to contribute to both.

1.4. INNOVATORS AND THEIR INTERESTS

Where do you find the innovators (largely scientists and engineers)? The United States has the greatest concentration. Interestingly, nearly one in eight U.S. scientists or engineers was born abroad, coming particularly from Asia.

The European Union and Japan also have large scientific and technical workforces. An estimated two out of three scientists and engineers today are men; however, this situation is rapidly changing. By the year 2020, there may be as many women as men engaged in innovative activities. Most of today's technology innovators work in industry, taking on R&D and various other roles. Although hard to categorize (much is not considered R&D), innovative activities of the advanced service sectors are ascending rapidly.

In tech mining we often want to know "who's doing what." Tracking ideas and individuals provides vital intelligence that serves various innovators (and technology managers).

The output of successful industrial innovation is reflected in new products, processes, and services. The IP involved may be protected through patenting. Some sectors patent more than others as discussed in Chapter 12. On average, an industrial scientist innovator patents once every six years. As we shall see, however, averages can be misleading—the majority don't patent at all, whereas a very few patent a very lot. Industrial researchers are also avid consumers of science and technology information.

Academic innovators contribute to public knowledge mainly through publication (journal and conference papers). An academic scientist or engineer is 45 times more likely to publish his/her research than an industrial counterpart. Not surprisingly, academic researchers contribute about three-quarters of all publicly available R&D. And, although comparatively infrequent, academic patenting is rapidly growing (Hicks et al., 2001). Academic innovators have traditionally also been the greatest consumers of science and technology information.

Too often, innovators reproduce solutions that are known elsewhere. One Russian researcher, Genrich Altshuller, established that roughly 25% of all patents solved problems well known in other disciplines or industries. Another 35% are merely minor extensions to established technologies. Less than 1% of all patents involve creation of foundational knowledge. Altshuller developed tools to stimulate invention—"TRIZ"—that we introduce in Chapter 12.

Finding the science and technology information needed to inform R&D presents a vexing challenge to innovators. Innovations are increasingly dependent on new science, and new sciences such as nanotechnology are increasingly multidisciplinary and therefore distributed across multiple fields. Observers have called for a device called the "memex"—a machine that could link and correlate ideas, finding connections and recommending relevant resources. Today, we have the Internet, but the vision of a seamless network of knowledge seems as distant as it was in the 1930s when the memex notion was conceived (discussed further in Chapter 2).

The nature of the emerging technologies of interest is itself changing. Many of our technology analysis tools were generated for an era of industrial (manufacturing) technologies dominated by defense interests (during the Cold War era) (Technology Futures Analysis Methods Working Group, 2004). In addition to information technologies, we see emergence of the "molecular tech-