Preface

Tech Mining (TM) is a special form of large data analytics (LDA). It concentrates on mining global ST&I publication/patent databases normatively by searching on a target emerging technology (or key organization) of interest in global databases, or investigates in trends and developments in STI domains in an explorative way. One then downloads, typically, thousands of field-structured text records (usually abstracts) and analyzes those for useful competitive technical intelligence (CTI) on existing and emerging trends and developments in the areas under investigation. Tech Mining has now been widely recognized by public and private research institutions, policy and strategy makers, universities, as well as corporations identifying future opportunities and threats leading to future innovations. Publication and patent databases are among the most frequently used sources for Tech Mining. Besides these structured data sources, more recently Tech Mining have been extended to analyze semi-structured and unstructured data including social network data, Web sites, blogs, reports, and even speech. New concepts and approaches with the combination of quantitative and qualitative methods are continuously introduced into the field. There are also advancements in computational tools which, besides generating bibliometric/scientometric data in more efficient and visually powerful ways, help to analyze text semantically to obtain useful insights from large volumes. To this end of anticipating future innovation pathways through LDA, the present book draws from authors who presented their cutting-edge approaches in the recent leading conferences including:

- Portland International Conference on Management of Engineering and Technology (PICMET; Aug., 2015).[5] PICMET is the leading conference in engineering and technology management bringing scholars, industry, and government representatives together.

[5]http://www.picmet.org/new/conferences/16/.

- Global Tech Mining Conference (GTM; Sep., 2015). The goal of this conference is to engage cross-disciplinary networks of analysts, software specialists, researchers, policymakers, and managers to advance the use of textual information in multiple science, technology, and business development fields.[6]
- International Conference on Future-Oriented Technology Analysis (FTA; Nov., 2014). FTA brings together those studying foresight and various forecasting and assessment methods, in a policy-oriented context.[7]

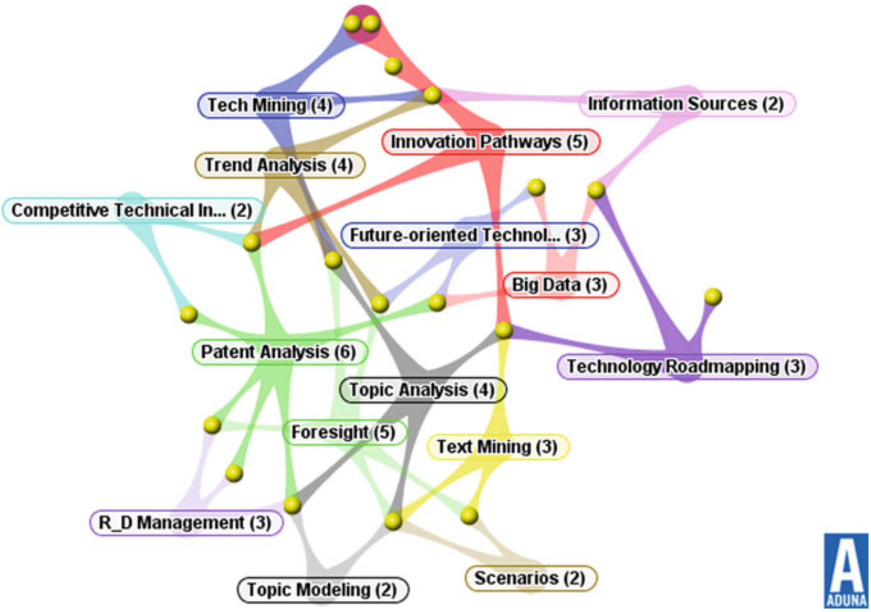
In the following 18 chapters, we are pleased to present advances in:

1. frameworks to consider Tech Mining and related analytical interests,
2. multiple methods able to treat science, technology, and innovation (ST&I) data effectively—plus means to incorporate human expertise and interests,
3. translation of analyses to useful intelligence on likely future developments of particular emerging ST&I targets,
4. informative indicators and compelling visualizations.

Applying the practice of generating indicators and visualizations to our own content, we imported the keywords accompanying chapter abstracts into the VantagePoint text mining software to generate an Aduna map of chapter–keyword relationships. Each of the 18 yellow dots represents a chapter, and the “arms” of each yellow dot extend out to their chapter’s primary keywords (number in parentheses indicates the number of chapters with keyword). We can see from this visualization that topical attention varying from patent/publication databases to scenarios indeed provides an integrated discussion on large data and innovation in terms of uses, methods, processes, and outputs.

[6]http://www.gtmconference.org/index.html.

[7]https://ec.europa.eu/jrc/en/event/site/fta2014/about.



This book has three parts. The first part reviews the fundamental concepts. The second part provides a set of emerging methods. The final part demonstrates the concepts through applied cases.

Part I Data Science/Technology Review

Chapters in the review part are mainly concerned with the use of data and Tech Mining for the purpose of future-oriented technology analysis (FTA), as well as the evolution of the FTA and Tech Mining concepts over time. In the first paper, entitled "How can Future-oriented Technology Analysis (FTA) be improved to take advantage of emerging information resources?", Loveridge and Cagnin suggest lessons to be garnered from business "due diligence" processes. Foresight and FTA strive to identify future possibilities. Those reflect complex balances of emerging capabilities and needs, informed by a mix of human knowledge and perceptions—plus data. Big data and analytics could serve FTA aims. The results suggest FTA-transforming potentials of diverse data resources and interconnections, such as autonomous systems, energy management, logistics, social networking, and forecasting institutions. The combination of accessible data and potent algorithms can foster FTA "due diligence" spanning a wide spectrum of potential factors (e.g., social, economic, ecological, political). It also poses staunch challenges re: balance between automated decision systems and roles for human judgments.

Following the opening paper, Li presents a more specific conceptual framework to consider text analyses in support of FTA—expressly, forecasting innovation pathways (FIP). "Tech Mining"—text analyses in support of FIP and FTA—is empirical in nature and has not been informed by a clear knowledge structure concerning objectives and methods. Li addresses objectives in terms of engineering processes, set in a broader management/policy context of attendant issues concerning development. He turns an engineering perspective to reflect on the elements of a systematic Tech Mining process. He divides the Tech Mining process into three connected components: models and tools; processes and technical standards; and process management factors. Drawing on analogy with software engineering, Li offers a top-down model to manage Tech Mining engineering systematically via needs assessment, strategic planning, and implementation planning. He provides a "big picture" perspective on using Tech Mining in support of innovation processes, with quality controls.

The next three papers use Tech Mining to investigate the evolution of Foresight and Tech Mining itself. Andersen and Alkaersig analyze changes in FTA and Foresight empirically through a bibliometric analysis of journal papers. They analyze journal special issues with publications emanating from FTA conferences in 2004, 2006, 2008, and 2011, along with abstracts submitted for the 2014 conference. Consideration of topical emphases in the FTA conferences over time finds increasing attention to "systems thinking." Hypothesized increases in emphasis on participatory methods and regional (vs. national) issues were not supported (rather, these seemed stable over time). The data indicate that the special issue publications exert comparable influence to other papers in the key FTA and Foresight journals.

Similarly, Mikova uses "Tech Mining" to examine trends in the content of the Global Tech Mining (GTM) conferences, 2011–2015. She benchmarks that activity against general trends of literature concerning Tech Mining (TM). Main streams of TM research are led by methodological development (e.g., bibliometrics, citation analyses, clustering, and network analyses), followed by consideration of uses in FTA. Special interests include focus on research and technology profiling, devising technology life cycles, and development of indicators. Visualization is noted as a keen interest. Mikova catalogues emerging methods, data types, and software tools used in TM; these can stimulate ideas to enrich established users' approaches. Semi-automated TM algorithms may combine with alternative Web-based data to open new possibilities.

In the final chapter of this part, Efimenko, Khoroshevsky, and Noyons offer a novel approach to anticipate future science, technology, and innovation pathways called Map of Science (Sect. 5.2). They perform extensive searches in Web of Science and Scopus on Scientometrics and related terms (e.g., science indicators) and on Tech Mining and related topics (e.g., technical intelligence). They analyze which R&D fields are addressed in those papers to map them, with co-occurrence providing a basis for establishing associations (i.e., fields being analyzed together in scientometric or TM studies). The name Map of Science (Sect. 5.2) reflects that these are visualizations based on the analyses of research fields (in Scientometrics and TM). They examine such VOSviewer maps over time to show evolving cross-disciplinary connection shifts. They also report topical evolutions in Scientometrics and in TM over time. This is an interesting hybrid approach using text and statistical analyses to extract information on R&D concentration, connection, and movement.

Part II Text Analytic Methods

The second part of the book focuses on the methods of Tech Mining and text analysis. There are a variety of approaches proposed. These range from the innovative approaches based on the use of new automated processes, algorithms, and ontologies, as well as the combination of Tech Mining approaches with other frequently used FTA methods to connect intelligence into decision-making processes.

The chapter by De Spiegeleire, van Duijne, and Chivot presents the metafore approach developed by The Hague Centre for Strategic Studies. Foresight 3.0, as the authors call it, is an attempt to distill more insights into future by combining all available data systematically. The chapter presents the main steps in the metafore research protocol that was used in a recent project for the Dutch government's "Strategic Monitor," which tries to anticipate the future in the area of foreign, security, and defense policy. The project had partners from multiple countries.

Benson and Magee introduce a new method called the classification overlap method (COM) which provides a reliable and an automated way to divide the patent database into understandable technological domains where progress can be measured. The authors conclude that there is now an overall objective method named Patent Technology Rate Indicator (PTRI) for using just patent data to reliably estimate the rate of technological progress in a technological domain. Thus, the first link between the patent database information and the dynamics of technological change is now firmly established; robustness and back-casting tests have shown that the assertion of reliability is meaningful and that the estimate has predictive value.

Courseault Trumbach, McKesson, Ghandehari, DeCan, and Eslinger introduce ontologies which are used in text mining processes to better understand text from a specific domain. Authors present a broad ontology for the innovation and design process. Through an example within the shipbuilding domain, the authors take steps toward building an innovation and design process ontology which can be applied to the forecasting innovation pathways (FIP) framework as a means of capturing and understanding the influences on the technology delivery system.

Huang, Zhang, Ma, Porter, Wang, and Guo combine topical analysis, patent citation analysis, and term clumping analysis to gather technology intelligence. The method identifies key subdomain patents, associated with particular component technology trajectories and then extracts pivotal patents via citation analysis. The authors compose evolutionary pathways by combining citation and topical intelligence obtained through term clumping. The case of dye-sensitized solar cells (DSSCs) is used to demonstrate the approach.

Huang, Shang, Wang, Porter, and Zhang use patent information and semantic analysis to identify targets for technology mergers and acquisitions. A case of China's cloud computing industry is analyzed to demonstrate the approach.

Chen, Zhang, and Zhu propose a topic change identification approach based on latent Dirichlet allocation, to model and analyze topic changes and topic-based trend with minimal human intervention. After textual data are cleaned, underlying semantic topics hidden in large archives of patent claims are revealed automatically. The results of a case study show that the proposed approach can be used as an automatic tool to provide machine-identified topic changes for more efficient and effective R&D management assistance.

The two remaining papers in the chapter describe the integration of Tech Mining with well-known FTA methods. Zhang, Chen, and Zhu attempt to develop technology roadmaps (TRM) semi-automatically through multiple data sources. The authors apply the fuzzy set to transfer vague expert knowledge to defined numeric values for automatic TRM generation. They present a case study on computer science-related R&D and show that the approach can assist in the description of computer science macro-trends for R&D decision makers.

In the final chapter of this part, Kayser and Shala propose developing scenarios using text mining. In their approach, text mining automatically processes texts and summarizes the topic. Two different approaches: Concept mapping and speech tagging are applied on two different scenarios which were developed through a Europe-wide project.

Part III Anticipating the Future—Cases and Frameworks

Following the discussion on the concepts and methods in the first two parts, the third and final part of the book demonstrates the ways of putting ideas into practice through case studies in selected high-technology domains. The first paper, a case study of LDA employed in CTI for national technology strategy purposes, is provided by Salvador, Menendez, and Novillo on the field of additive manufacturing. Their assessment of the potential for additive manufacturing begins with a global technological and market landscape, which is then compared with actual development in Latin America. Market research and expert interviews are rounded out with patent and scientific literature analyses to determine the degree to which Latin America is behind other regions of the world in technological and market development. In such a position, such analyses are useful to pinpoint leading research organizations, research focus, key research networks, key patent holders, and private players investing in process commercialization in order to streamline decisions on where best to enter technology and market development.

Next, Daim, Khammuang, and Garces apply social network analysis (SNA) to the technology area of smart roofing to identify the dynamics of expert networks, so researchers gain a better understanding of the current state of smart roofing research and development programs. Using bibliographic data from Web of Science, the authors generate the SNA attributes of degree of centrality, betweeness, closeness, and number of citations, to pinpoint the top 11 experts in the field as well as their collaborative influence on the R&D network. This in-depth knowledge of experts and networks is particularly useful in such management of technology processes as technology roadmapping (TRM), R&D portfolio selection, R&D project initiation, and strategic technology planning.

New drug development can take a decade or longer, with most compounds not proving fruitful. Analyzing primary patent applications in a therapeutic class can give a picture of which drugs may reach the market in the future, though this picture is severely muddled as at early patenting stages, therapeutic uses are unclear, and multiple patents filed at different stages may include the same compound with different claims. Mendes and Antunes address this matter with a method for gathering and analyzing the primary patent applications for new antibiotics using Derwent Innovations Index Database and VantagePoint text mining software. Employing IPC codes, Derwent Manual Codes, the World Health Organization's Anatomical Therapeutic Chemical (ATC) classification system, and MEDLINE's Medical Subject Headings (MeSH) as resources in the text-mining software, 32,068 antibiotic patents where reduced to a set of 1333 primary patents. These patent applications can be analyzed to show whether antibiotics are from old or new chemical classes, which bacteria they act against, mechanisms of action, and what strategy was used to discover the compounds. Further study can point out antibiotics expected in future markets and if these antibiotics will meet projected critical bacterial disease priorities.

Liu, Sun, Xu, Jia, Wang, Dong, and Chen describe successful institutionalization of the use of text analytic-generated CTI at the National Science Library (NSA), Chinese Academy of Sciences (CAS), for Chinese government and enterprise MOT decision making. By tailoring specific combinations of text analytic-generated indicators from the science literature and patents (bibliometrics, patent metrics, text mining) and expert review, the NSA-CSA is able to address three levels of information needs—micro level (specific technology), meso level (technology field), and macro level (industry). Examples of the CTI products include technology novelty reviews, innovation pathway selections, product development evaluations, competitor monitoring, R&D partner identification, and industry–technology field analyses to support industrial technology and development strategies. The authors provide case studies of a micro level analysis of hydrodynamic cavitation technology for wastewater processing, meso-level analysis of swine vaccine technologies, and a macro-level strategic intelligence analysis of the ionic rare earth industry. The authors also share insights into the specific feedback clients have provided as well as quality control measures adopted over the years.

Finally, the chapter by Bakhtin and Saritas introduces a methodology for the identification of trends through a combination of "thematic clustering" based on the co-occurrence of terms and "dynamic term clustering" based on the correlation of their dynamics across time. In this way, it is possible to identify and distinguish four patterns in the evolution of terms, which eventually lead to (i) emerging, (ii) maturing, and (iii) declining trends, as well as (iv) weak signals of future trends.

Key trends identified are then further analyzed by looking at the semantic connections between terms identified through Tech Mining. This helps to understand the context and further features of the trend. The proposed approach is demonstrated in the field photonics as an emerging technology with a number of potential application areas.

Tugrul U. Daim
Department of Engineering and Technology Management
Portland State University, Portland, OR, USA

Denise Chiavetta
Search Technology, Inc., Norcross, GA, USA

Alan L. Porter
Georgia Institute of Technology, Roswell, GA, USA

Ozcan Saritas
National Research University Higher School of Economics
Moscow, Russia

http://www.springer.com/978-3-319-39054-3

Anticipating Future Innovation Pathways Through Large Data Analysis
Daim, T.U.; Chiavetta, D.; Porter, A.L.; Saritas, O. (Eds.)
2016, XVIII, 360 p. 141 illus., 108 illus. in color.,
Hardcover
ISBN: 978-3-319-39054-3