A Methodology for the Identification of Technology Indicators

zur Erlangung des akademischen Grades eines
DOKTORS DER INGENIEURWISSENSCHAFTEN (Dr.-Ing.)
der Fakultät für Maschinenbau
der Universität Paderborn

vorgelegte
DISSERTATION

von
B. Eng. Hua Chang
aus *Qingdao, China*

Acknowledgments

The present work “A Methodology for the Identification of Technology Indicators” presents my research work at the Heinz Nixdorf Institute, University of Paderborn. It summarizes the working experience and the research results that I gained from the research areas of technology planning and product innovation in several research projects.

First of all, I would like to cordially thank Prof. Dr.-Ing. Jürgen Gausemeier for giving me the chance to do my PhD and also for his patient instruction as well as support.

I greatly appreciate the co-supervisorship and advice of Prof. Dr.-Ing. Jörg Wallaschek and Prof. Dr. Wilhelm Schäfer. I am also grateful to Prof. Dr.-Ing. Ansgar Trächtler as the chairman of the board of examiners.

I am also thankful to the International Graduate School Dynamic Intelligent Systems at the University of Paderborn for offering the professional, financial and intercultural support in the three years of my PhD.

Regarding my work, I would like to give the very special thanks to my team leader Dipl.-Wirt. Ing. Christoph Wenzelmann, who has always provided me useful assistance and is obliging to impart me his experience and knowhow. Great thanks should also be forwarded to Dipl.-Wirt. Ing. Stephan Ihmels, Dipl.-Wirt. Ing. Karsten Stoll and M. Sc. Chengyee Low.

I would also like to thank all other current and former members of the team of Innovation Management for having shared a very intense time and for the brilliant teamwork experience: Dr.-Ing. Daniel Steffen, Dr.-Ing. Jan Stefan Michels, Dr.-Ing. Thomas Peitz, Dipl.-Wirt. Ing. Guido Stollt, Dr.-Ing. Arnt Vienenkötter, Dipl.-Wirt. Ing. Ingo Kaiser, Dipl.-Wirt.-Ing. Volker Brink, Dipl.-Wirt.-Ing. Sebastian Deyter, Dipl.-Wirt.-Ing. Sascha Kahl, and Dipl.-Inf. Sebastian Pook. I will never forget to say thanks to our charming and efficient secretaries: Ms. Sabine Illigen and Ms. Alexandra Dutschke; to our “computer guardians”: Dipl.-Ing. Karsten Mette and his apprentices; as well as to my nice workmate and also my last roommate Dr.-Ing. Salvatore Parisi for his great support.

Last but not least, my sincere thanks to my boyfriend Shan He for all his support and patience; to my family for showing me the way of life and always being there; to family Gausemeier for their ungrudging help and care.

Paderborn, February 2008 *Hua Chang*

To my family

List of published partial Results

[CB06] CHANG, H.; BRUESEKE, U.: Einsatz bibliometrischer Analysen in der strategischen Frühaufklärung, 2. Symposium Vorausschau und Technologieplanung, Schloss Neuhardenberg bei Berlin, Deutschland, 9-10 November 2006

[CGI+07a] CHANG, H.; GAUSEMEIER, J.; IHMELS, S.; WENZELMANN, C.: A Technology Management System to foster Product Innovation. 16th International Conference on Management of Technology, (IAMOT 2007), Miami Beach, USA, May 13-17 2007

[CGI+07b] CHANG, H.; GAUSEMEIER, J.; IHMELS, S.; WENZELMANN, C.: "Technology Intelligence with Bibliometrics" Proceedings of the IAENG International Conference on Data Mining and Applications (ICDMA'07). 21-23 March, Hong Kong, 2007

[CGW06] CHANG, H.; GAUSEMEIER, J.; WENZELMANN, C.: Indicator-based Technology Management, 15th International Conference on Management of Technology, (IAMOT 2006), Peking, China, May 22-28. 2006

[CGW07] CHANG, H.; GAUSEMEIER, J.; WENZELMANN, C.: Bibliometrics-based methodology for the identification of technology indicators. International Journal of Technology Intelligence and Planning (IJTIP), Vol. 3, No. 3, 2007

Planned publication in 2008:

[CGI+08] CHANG, H.; GAUSEMEIER, J.; IHMELS, S.; WENZELMANN, C.: Innovative Technology Management System with Bibliometrics in the context of Technology Intelligence. Trends in Intelligent Systems and Computer Engineering (ISCE) Series: Lecture Notes Electrical Engineering, Vol. 6. Castillo, Oscar; Xu, Li; Ao, Sio-Iong (Eds.) Springer, Available: May 2, 2008

Contents

1 Introduction 3

1.1 Problem Analysis 3

1.2 Research Objectives 3

1.3 Research Approach 3

2 Problem Analysis 3

2.1 Product Innovation 3

2.1.1 The Product Innovation Process 3

2.1.2 Technology Roadmap 3

2.2 Technology Database (Heinz Nixdorf Institute) 3

2.2.1 Concept of the innovative Technology Database 3

2.2.2 Information Procurement 3

2.2.2.1 Process of Information Procurement 3

2.2.2.2 Challenges of Information Procurement 3

2.3 Placement of this Dissertation 3

2.4 List of Requirements 3

3 State of the Art 3

3.1 Expert Consultation 3

3.2 Information Retrieval 3

3.3 Artificial Intelligence 3

3.4 Mining Technique 3

3.4.1 Data Mining 3

3.4.2 Text Mining 3

3.5 Ontology 3

3.5.1 Main Elements of Ontology 3

3.5.2 Types of Ontology 3

3.5.3 General Generation Process of Ontology 3

3.5.4 Available Ontologies 3

3.5.5 Ontology Applications 3

3.6 Bibliometric Analysis 3

3.6.1 One-dimensional Bibliometric Analysis 3

3.6.2 Two-dimensional Bibliometric Analysis 3

3.6.3 Patent Analysis 3

3.6.4 Application of Bibliometric Analysis 3

3.7 Call for Action ... 3

4 Methodology for the Identification of Technology Indicators ... 3

4.1 Foundation of the Methodology ... 3

4.1.1 Basic Methods ... 3

4.1.2 Guide to the Interpretation of Knowledge Map ... 3

4.1.2.1 Basic Instruction ... 3

4.1.2.2 General Steps ... 3

4.1.3 The Ontology of Technology Indicators ... 3

4.2 General View of the Methodology ... 3

4.3 Phase of Problem Analysis ... 3

4.4 Phase of Literature Search ... 3

4.4.1 Limitation of Search Area ... 3

4.4.2 Search for Publications ... 3

4.4.3 Pre-processing of the Publications retrieved ... 3

4.5 Phase of the preliminary Identification of Technology Indicators ... 3

4.5.1 Using Publication Analysis ... 3

4.5.2 Using Co-word Analysis ... 3

4.6 Phase of Concretization of Raw Technology Indicators ... 3

4.6.1 Values Assignment to Raw Technology Indicators by interpreting the Publication Diagrams ... 3

4.6.2 Values Assignment to Raw Technology Indicators by interpreting the Knowledge Map ... 3

4.7 Phase of Expert Consultation ... 3

4.7.1 Expert Consultation ... 3

4.7.2 Comparison of Results from Experts and the Complete Technology Indicators ... 3

4.7.3 Regular Update ... 3

4.8 Integration of the Methodology and the Technology Database 3

4.8.1 Technology Indicators as input for the Technology Database ... 3

4.8.2 Visualization of Technology Indicators as output of the Technology Database ... 3

5 Case Studies and Evaluation ... 3

5.1 Case Study of MID Technology ... 3

5.1.1 Phase of Problem Analysis ... 3

5.1.2 Phase of Literature Search ... 3

5.1.3 Phase of preliminary Identification of Raw Technology Indicators ... 3

5.1.4 Phase of Concretization of Raw Technology Indicators3
5.1.5 Phase of Expert Consultation....................................... 3
5.2 Evaluation of the Methodology.. 3

6 Summary and Outlook 3

7 Bibliography 3

1 Introduction

1.1 Problem Analysis

In business today, companies are being increasingly influenced by global competition, always shorter product life cycle, changes of consumer habits, and many other impact factors. Product innovation is no longer an option for growth, but a requirement for survival. On the one hand, new technology or new combination of technologies is the driving force of product innovation (technology push). On the other hand, product innovation is also driven by market demands, which are considered as the catalyst of technological changes (market pull). For technology-intensive companies especially in mechanical engineering industry and electronic industry, technology itself has become a decisive factor because of its significant influence on product development and process optimization. It is important to identify advantages or barriers of technologies, to compare them as well as to analyze the probability of being substituted [WB02].

Therefore, scientific researchers and decision makers in companies address the attention to Technology Intelligence, which is the sum of methods, processes, best tools used to identify sensitive information about technological developments or trends that can influence a company's competitive position. The Technology Intelligence process spans across four levels: Data, Information, Knowledge, and Decisions. **Data** are symbols without meaning. **Information** is data that has been given meaning by way of relational connection. **Knowledge** is the output of information scouting, processing and analyzing. And **Decisions** are made on the base of knowledge [DB95]. Within the framework of Technology Intelligence, the main task is to procure accurate information about performances and developments of technologies, i.e. to identify Technology Indicators.

Technology Indicators are those indices or statistical data, which allow direct characterization and evaluation of technologies throughout their life cycles. "Key player", "technology maturity", and "technology trend" are examples of typical Technology Indicators.

Technology Indicators are represented in the form of published information such as scientific papers, dissertations, product descriptions, technological reports, press releases, etc. In order to extract the Technology Indicator, a corpus of published information should be collected and analyzed. There are two main problems that need to be solved during the process of extraction of Technology Indicators from technology information corpus.

The amount of information is too large.

According to BROCKHAUS, there are 100,000 to 200,000 scientific periodicals at present, while there were only 1,000 in the middle of 19th century. The amount of daily scientific publications is now around 20,000; and it was only 2,000 in 1950. Also the World Patents Index records yearly 1.5 million new patents [BG05], [GHK+06]. It is already well known that the amount of information has increased tenfold or even more with the development of information technology.

> *"The new communications technologies and content services as well as the convergence of the Internet and the mobile phone create freedom and possibilities as never seen before in the world."* *[Ala99]*

People previously read documents one by one and collect key information manually. But nowadays, the volume of information is so high that it is no longer possible to evaluate or characterize technologies by reading documents. Even in a limited area e.g. mechatronics, the number of publications is too huge to be processed manually. Therefore, there is demand on computer-aided information procurement process, which can (semi-)automatically retrieve, filter, analyze, and interpret a mass of information.

Companies lack a guideline to procure the Technology Indicators.

As a matter of fact, there are already some methods that are contributed to automatic information procurement; for example Knowledge Discovery in Databases, Information Retrieval, Artificial Intelligence, Information Mining, Patent Analysis, etc. However, those methods have never been integrated with characterization and evaluation of technologies. There is no systematic approach aiming at knowledge extraction for Technology Indicators. The absence of that systematic guideline has caused the following discomforts for decision makers in companies:

- Normally, the information corpus is collected and analyzed manually. Therefore, the process is time-consuming and the results are subjective. The limited information amount due to manual processing causes the risk that some aspects could be ignored.
- Various procedures lead to waste of manpower and material resources. Every time when a technology needs to be analyzed and evaluated, it has to be thought about where to begin and how to begin.
- Technologies change quickly. Accordingly, the information about Technology Indicators should be updated regularly. The same knowledge extraction process should be run again with reusable and newly

added information. Redundant work is inevitable without standard guidelines.

Based on the former discussion, it can be summarized that technology is an impact factor for most companies. Technological performance, development, and trends can be characterized and evaluated with the aid of Technology Indicators. There are already methods available for knowledge extraction in the context of technology management. However, none of them is devoted to automatic identification of Technology Indicators. Furthermore, there is an absence of a standard process which can be used for all technologies.

1.2 Research Objectives

Regarding the challenges, a methodology for the (semi-) automatic identification of Technology Indicators is urgently needed. Research work in this dissertation aims at developing such a methodology that can cover all the current problems and requirements of decision makers discussed above.

Input of this methodology is a vast amount of information. Through the methodology, information should be rapidly computer-aided analyzed and filtered. As the results of the methodology are the semi-automatically identified Technology Indicators and their values. In order to catch up with the up-to-date technology changes, the methodology also facilitates the regular update of technology information.

In the methodology, intelligent methods are used instead of manual work to speed up the knowledge extraction and decision-making processes. Decision makers can procure the relevant information about targeted technology by following a standardized process introduced within the methodology.

Furthermore, the methodology is integrated in the innovative Technology Database developed by Heinz Nixdorf Institute. They work together as a centralized technology information warehouse that offers relevant information of technologies as well as innovative product and production ideas.

1.3 Research Approach

Based on the research objectives mentioned in section 1.2, the dissertation is structured as follows. The introduction in **chapter 2** explores the product innovation process. In the first cycle of the product innovation process, Technology Roadmaps are applicable. For the purpose of automatic generation of complex Technology Roadmaps, Heinz Nixdorf Institute has developed an innovative Technology Database, whose information procurement process still has to be optimized. A methodology is desired. Concrete requirements are derived from the innovative Technology Database.

In **chapter 3** the existing methods used in information procurement are reviewed. The basic functions and current status of the methods are introduced. It is also evaluated, to what extent the methods fulfill the requirements defined in chapter 2. At the end of chapter 3, the call for action is educed.

Chapter 4 shows the main creation of value of this dissertation. The methodology for the identification of Technology Indicators is explained in detail. First of all, the methodic foundation of the methodology is explicated. Then, an overview of the methodology is given. The methodology is divided into five phases. The phases are described theoretically and separately in turn. At last, the integration of the methodology and the innovative Technology Database is elucidated.

In order to verify the methodology, a case study is demonstrated in **chapter 5**. The case study is focused on the technology MID (Molded Interconnected Devices). It was carried out step by step in terms of the methodology introduced in chapter 4. The case study of MID proves the usability and validity of the methodology. It also highlights the advantages of this methodology by comparing the fulfillments of the requirements. More details are inside the chapter.

Chapter 6 includes the brief summary of the research work and the discussion of future work.

2 Problem Analysis

2.1 Product Innovation

Due to intense competition, companies have high demands on product innovation, which means decision makers in companies should choose the right business strategies, develop new products and services that fulfill the changing users' requirements, and constitute applicable technology plans. All the activities are based on regular, quick information procurement and agile reaction.

2.1.1 The Product Innovation Process

The product innovation process begins with the idea of a product or business and leads to the successful product launch. It incorporates the areas of strategic product planning, product development, and manufacturing process development. The general work flow is shown in the Fig. 2-1. In practice, the product innovation process comprises a number of cycles.

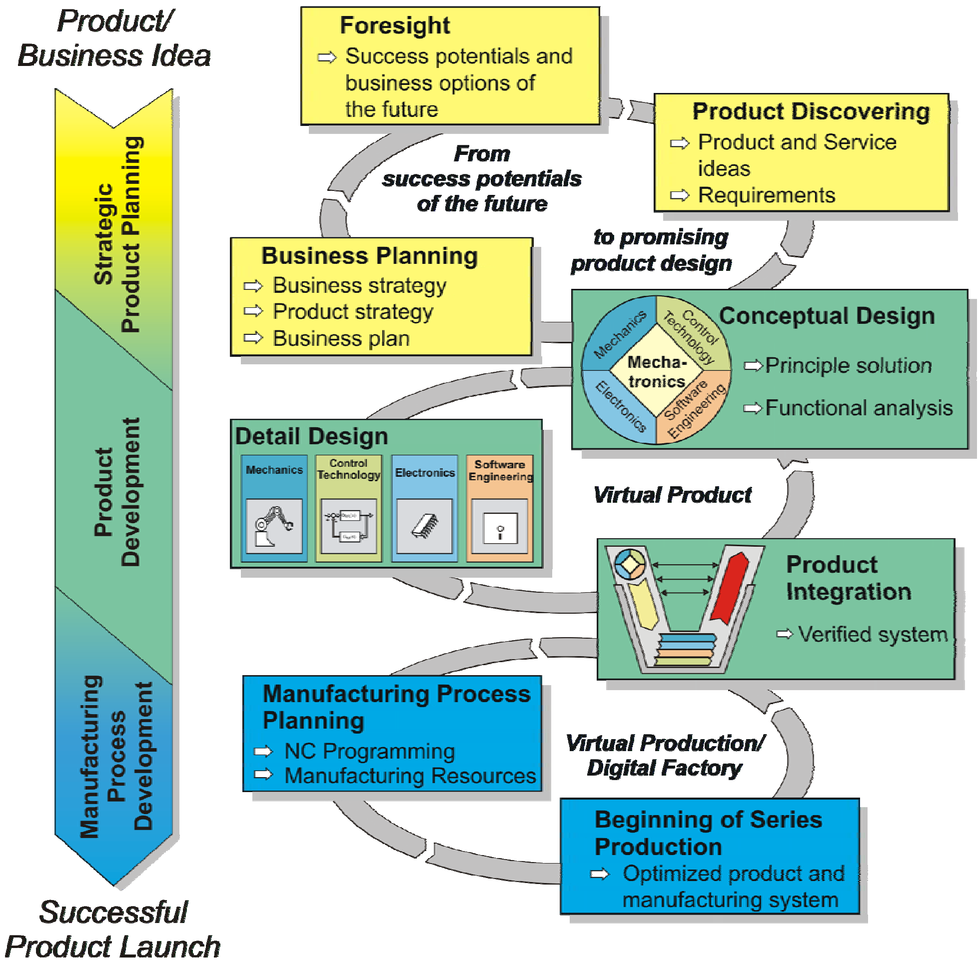


Fig. 2-1: The product development process as a sequence of cycles [GEK01]

The first cycle characterizes the steps from finding the success potentials of the future to create the promising product design, which is the principle solution. There are four major tasks in this cycle:

- foresight,
- product discoverying,
- business planning and
- conceptual design.

The aim of **foresight** is to identify the potentials for future success, as well as the relevant business options. The methods used here are such as the scenario technique, Delphi studies and trend analyses.

The objective of **product discovering** is to find new product ideas. In this phase we apply creativity techniques such as the Lateral Thinking of de Bono or the well-known TRIZ on the one hand. One the other hand we utilize our technology planning concept.

Business planning is the final task in the cycle of strategic product planning. It initially deals with the business strategy, i.e. answering the question like which market segments should be covered, when and how. The product strategy is then elaborated on this basis. It contains information about

- setting up the product program,
- cost-effective handling of the large number of variants required by the market,
- the technologies used and
- updating the program over the product lifecycle.

Additionally, a business plan must be worked out to make sure an attractive return on investment can be achieved.

This first cycle is also concerned with the **conceptual design**, although this area of activity is assigned to product development in the narrower sense. The result of the conceptual design is the principle solution. It is required to estimate the manufacturing costs needed in the business plan. That is the reason why there is a close interaction between strategic product planning and product design linked by conceptual design. Conceptual de-sign is the starting point for the next cycle.

The second cycle corresponds to the actual understanding of product development according to the VDI-Guideline 2206 “Design Methodology for Mechatronic Systems” [VDI04]. The work to achieve this guideline was managed by the Heinz Nixdorf Institute. The essential point here is the refinement of the cross-domain principle solution by the experts from the domains involved, such as mechanical engineering, control technology, electronics and software engineering. The results of the domains elaborated in this cycle must be integrated into an encompassing product specification. This specification has to be verified according to the requirements formulated in the first cycle. This is done in the product integration phase.

The third and last cycle focuses on manufacturing process development and the optimization of the product design with respect to manufacturing.

2.1.2 Technology Roadmap

One of the main topics in the first cycle of strategic product planning is Product Innovation. Market demands change fast, innovative products are required to satisfy new market trends. The key point is to identify the opportunities resulting from technological advancement and those resulting from the development of markets and to balance them. Product innovation is therefore driven by both, Technology Push and Market Pull.

In order to enable product innovation from both sides, the method Technology Roadmap is needed. A **Technology Roadmap** stands for a plan which shows, which technology can be used in which products at what time [Eve02], [WB02]. Fig. 2-2 shows a Technology Roadmap in a simplified form.

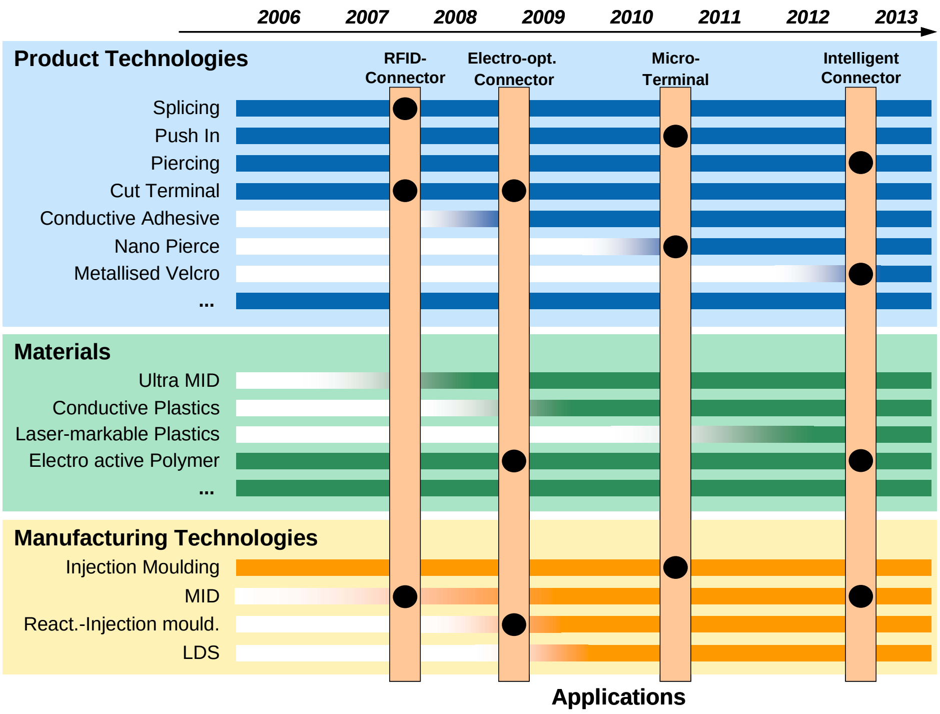


Fig. 1-2: Example of a Technology Roadmap (simplified) [GW05]

In the horizontal row the relevant technologies for the enterprise are specified. It is indicated on the time axis, when the respective technology is mature for employment in a series product. Usually some technologies have to cooperate in order to realize a beneficial application. In Fig. 2-2 four example applications are shown. The black junctions mark the utilized technologies.

Our experience shows that the generation of such roadmaps must be computer aided, on the one hand due to the high number of technologies which can be regarded - it can easily be more than hundred - and on the other hand the often high number of applications can no longer be handled any longer in a manually generated graphics. Such a high number of applications also require a classification of the options for action based on the Technology Roadmap. The classification, which follows the product market matrix of ANSOFF [Ans66], is represented in Fig. 2-3.

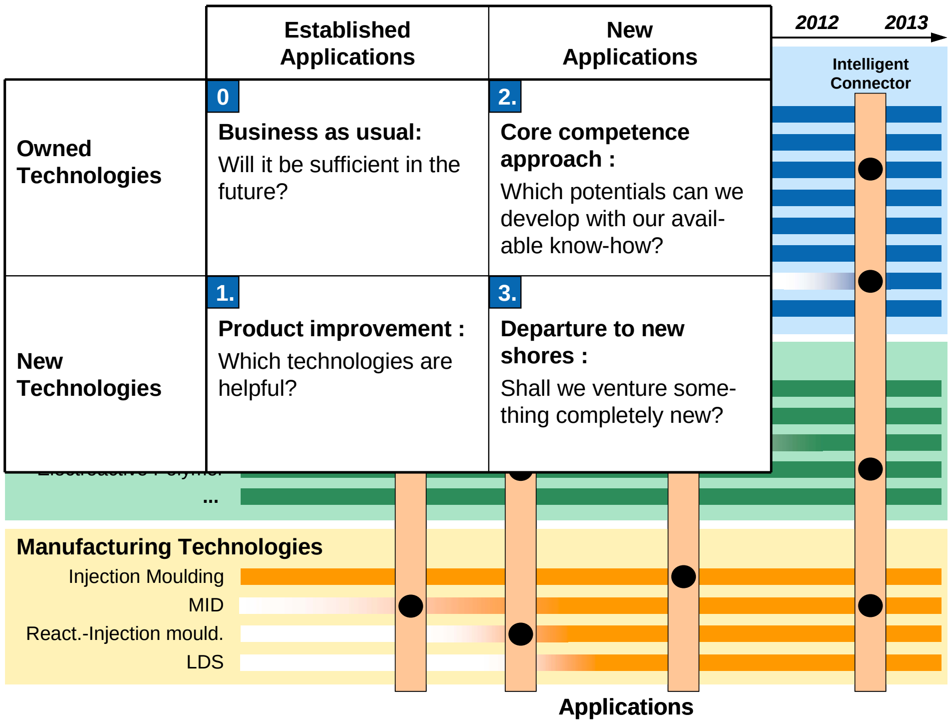


Fig. 2-2: The options for the technology-referred advancement of the business

Afterwards it is determined whether the up-to-date operated business still carries the enterprise or business innovations should be necessary. If business innovations are necessary, the three following classes of options for action are examined in the indicated order, because the uncertainty of success increases accordingly.

Product improvement: this option answers the question: which technologies those are not in possession of the enterprise yet, can improve the cost-performance ratio of the existing products?

Core competence approach: the technologies that are controlled by the enterprise frequently represent competences, and cannot be developed easily by competitors. The question arises: Which new application fields can be developed on the basis of the existing competences in order to generate benefit for the customer and/or to satisfy his needs?

Departure to new shores: a completely new business is required to be established; both the technologies and the customers are new. Naturally this comes along with the highest risk and therefore is usually only considered if the two options mentioned previously do not offer approaches for the advancement of the business.

There are three requirements for using the Technology Roadmap in product innovation.

- How to find out important information about technologies and their applications? The information volume is huge; an intelligent approach of information procurement is urgently required.
- How to identify the interaction of technologies and applications automatically? There are thousands of technologies and applications, the inherent connections should be identified.
- How to generate the Technology Roadmap automatically? It is difficult to generate a Technology Roadmap manually when the numbers of technologies and applications are more than 100. Automatic visualization of Technology Roadmap should be realized.

For the purpose of better utilization of the Technology Roadmap in product innovation, a modern intelligent system is desired, which fulfills all the requirements mentioned above.

2.2 Technology Database (Heinz Nixdorf Institute)

In this context, Heinz Nixdorf Institute has developed an intelligent technology management system – the innovative Technology Database. The Technology Database facilitates collection, storage, analysis, access, and update of technology information. It also allows automatic generation of reports, which help decision makers to characterize and evaluate technologies. The innovative Technology Database offers computer-aided support on product innovation by suggesting possible combinations of technologies and applications. The structure of the innovative Technology Database is introduced in the following paragraphs, so as the information flow, its functions, and advantages. The concept of the innovation Technology Database is shown in Fig. 2-4.

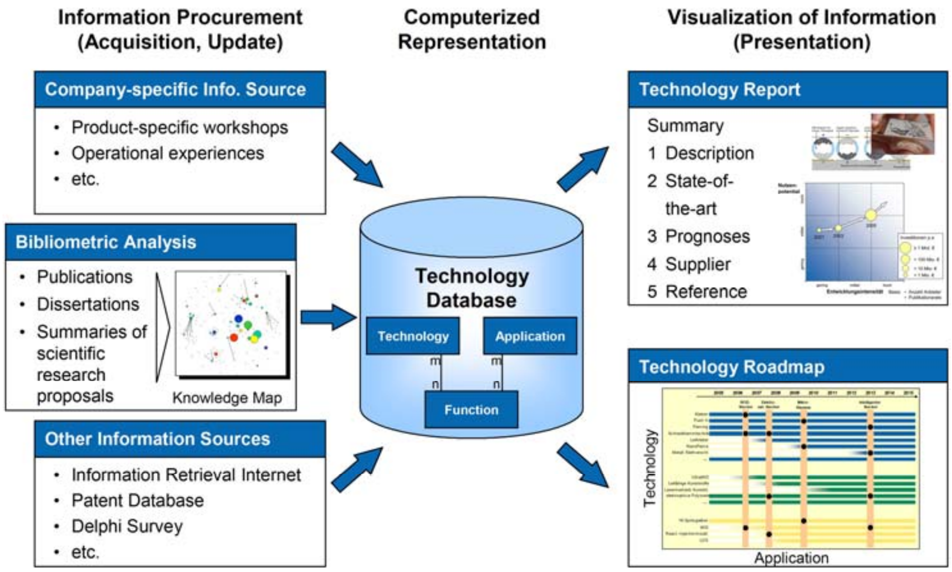


Fig. 2-4: Overview of the innovative Technology Database developed by Heinz Nixdorf Institute [GW05]

2.2.1 Concept of the innovative Technology Database

The core of this system is a relational Database, in which accumulated knowledge and emerging information about technologies and applications are stored. It consists of four main interconnected entities. The close interaction of the four entities described below enables various queries in the Technology Database and also other functions. Their relational model is illustrated in Fig. 2-5. [GHK+06]

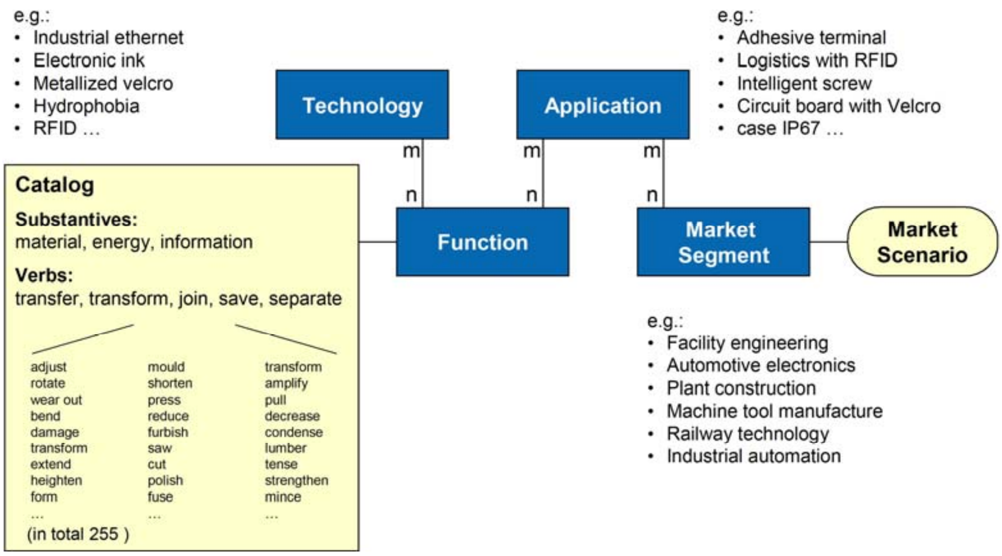


Fig. 2-5: Technology Database, simplified relational data model

- **Technology:** Here the relevant information about a technology, e.g. descriptions, publications, graphics etc., is saved in the database. Instances for a technology are: electronic ink or metallic Velcro. Information is loaded in the form of structured metadata or continuous text.

- **Application:** Applications are practical solutions to problems, such as products or services, which are based on one or more technologies. For example, e-book reader is an application related to the technology electronic ink. Similar to **Technology** above, the necessary information (description, market analysis, supplier etc.) of applications is also available in the Technology Database.

- **Function:** On one side, every technology performs certain functions, e.g. the technology electronic ink can display patterns; and on the other side, each application is based on a group of functions, e.g. the e-book reader holds the functions as read documents, save data, etc. The innovative Technology Database contains a fixed list of general functions based on the corresponding scientific works of BIRKHOFER [Bir80] and LANGLOTZ [Lan00]. Every function is composed by a substantive (optional from material, energy, and information), a main verb, and a corresponding sub verb, e.g. information – transfer – transmit, or material – transform – stretch. A function can be assigned to more than one technology or application; conversely, a technology or an application performs a group of functions. The relations between "**Technology** and **Function"** and "**Application** and **Function**" in the relational database are both m:n relations. When a technology or an application is newly added into the Technology Database, it should be linked to the corresponding functions. It is noticed here, the Technology Database recognizes only standardized expressions of functions, i.e. substantive – main verb – sub verb. Thus, technologies and applications are indirectly connected with each other through functions. [GCI+07]

- **Market Segment:** A market segment is a subgroup of people or organizations sharing one or more characteristics that cause them to behave in the same way or have similar product needs. Examples for market segments are facility engineering, automotive industry, and so on. In our database market segments are described in detail and attached with future-orientated market scenarios [GEK01]. Those scenarios are descriptive reports, which present and help to understand different ways that future market events could unfold. **Market Segment** is directly connected to **Application** with an m:n relation.

Except for the storage and management of information/knowledge about the four entities, the innovative Technology Database also allows various queries and visualization of the outputs in two essential presentation forms, which are shown on the right side of Fig. 2-4: the Technology Report and the Technology Roadmap.

The various retrieval queries are such as “which technologies will be applied in building engineering” or “which applications are realized by the technology metallic Velcro?” In principle, the Technology Database concerns questions in connection with inherent driving approaches Technology Push and Market Pull. For a better understanding, the function model of Technology Push and Market Pull is illustrated in Fig. 2-6.

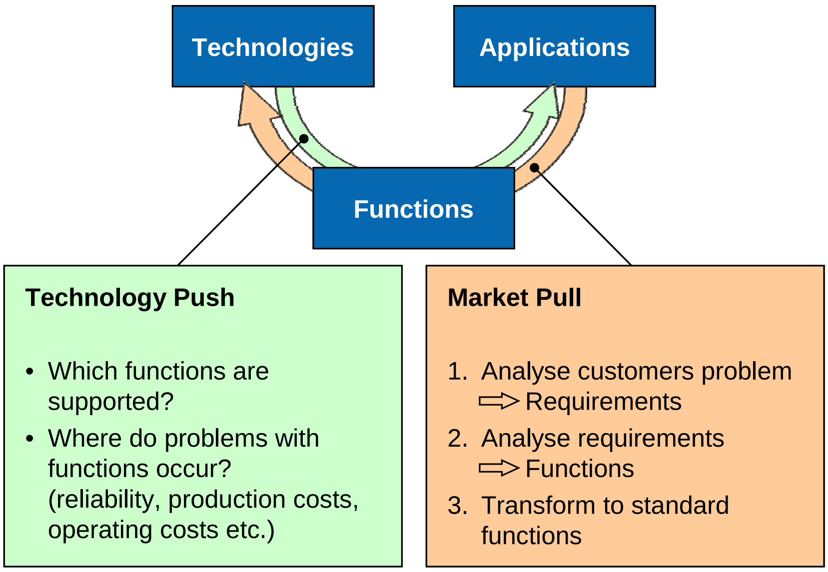


Fig. 2-6: Technology Push and Market Pull based upon the Technology Database [GW05]

The Technology Push approach begins with determination of functions performed by the target technologies. Similar technologies are retrieved by means of matching the same functions. So do the applications. All the existing and potential applications that depend on those technologies are discovered automatically from the database through the determined functions. Then, the current problems are analyzed, e.g. production cost is too high, or production process is unsafe, etc. Pointed to those problems, suitable applications are finally selected. From the other side, the Market Pull approach starts with the analysis of customers’ problems, i.e. demands of the market. Result is a list of all requirements the applications should meet. Secondly, the functions that fulfill the requirements are identified. However, those functions are described with engineer’s language. They should be translated into standard expressions, which match to the function list pre-defined in the database. For example, a lever is used to transmit force. The standard function understandable by the Database is defined with a substantive plus a main verb and a sub verb: “energy – transfer – transmit”. After that, the existing and potential combinations of applications and technologies are established through functions, which is similar to Technology Push approach. The combinations are then evaluated by decision makers according to the realizability and other pre-determined requirements. Final conclusions are drawn on the most practicable technology solutions.

Both approaches are mainly supported by Technology Roadmaps, which show the current and potential combinations of numerous technologies and applications based on the same functions (Fig. 2-2). The innovative Technology Database facilitates the automatic generation of Technology Roadmaps with numerous technologies and applications, which is not possible by manual work.

Furthermore, it is also possible to generate Technology Reports automatically from the Database. The Technology Report is especially attractive for researchers and decision makers because it describes the technologies in detail. The Technology Report is constructed in a default format:

- Summary: a presentation of the substance of the Technology Report in a condensed form by reducing it to its main points. The summary brings decision makers a brief overview including the abstracts of every part listed below.
- Description: this part contains general information about the technology, i.e. history of the development, work principle, typical applications and application fields, main advantages, etc.
- State of the art: the current status of the technology's development is reported here. For instance, the present technological level, the actual market development including accessible economic data, investment volume, recent important patents, technical barriers, market barriers, and so on.
- Prognoses: in this part, the future development direction of the technology is estimated. Further technological progresses are forecasted; potential application fields are predicted; future investment and marketing activities are deduced from the tendency of market data, etc.
- Key players: Leading suppliers, customers, and experts of the technologies constitutes key players. In the Technology Reports, the firm names, websites, and the current contact information (contact person, telephone number, email address, post address, etc.) are documented.
- Information sources: references are listed at the end of the Technology Reports in order to testify the validity of the contents. Besides that, decision makers can check the original documents if they need more information.

Information and knowledge used to construct both of the Technology Reports and the Technology Roadmaps are sorted directly from the innovative Technology Database. That means the related information/knowledge of technologies should be located and loaded previously in the database.

In Fig. 2-4, the left side is reserved for the process of information procurement, search areas, information sources, and methods used. Since the innovative Technology Database is an intelligent technology management system, it is obvious that the information procurement process should be intelligent as well. In other words, the relevant information and knowledge should be systematically, automatically, and procedurally retrieved from the mass of raw information. It is important to define the following issues: what is the information required (input for the Technology Database)? How to define the information procurement process? Which methods are best practicable? Which requirements should be met? The first and foremost issue is to analyze the challenges of information procurement. The next section introduces the current status of information procurement.

2.2.2 Information Procurement

Information procurement aims at the achievement of valuable or desired information patterns from a collection of literature materials. Nowadays, the development of information technologies and communication technologies facilitate storage of electronic data in different forms of warehouses, databases, and other information repositories. The study “How much Information?” in 2003 has shown that over 93 percent of the information produced in 1999 was in digital format [LV03-ol]. Digital technology facilitates easy storage, and fast distribution of information. Additionally, information resources are easily shared by using Internet as well as other information and networking systems. Therefore, modern information procurement means in most cases the acquisition of relevant data and knowledge from electronic information databases or other shared domains.

2.2.2.1 Process of Information Procurement

Generally speaking, there are two performers in the information procurement process: client and server. The client sends a message to the server. The message includes a request to source information about a certain object or issue. The server receives the message. The message is analyzed and processed by means of different methods used in the processing process, e.g. Information Retrieval, knowledge discovery from database, etc. Significant and useful information with regard to the client’s request is retrieved from a large collection of information resources. Then, the server responds to the client with a message containing information that meets the client's requirements. Fig. 2-7 shows the general process of information procurement. [IP06-ol]

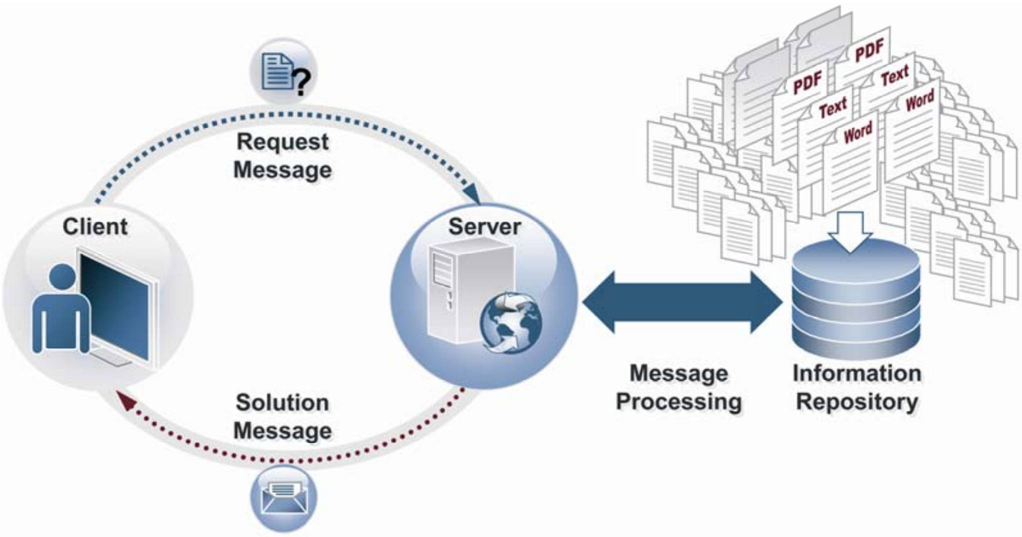


Fig. 2-7: General process of information procurement

In terms of technology monitoring, information procurement is defined in a narrow sense. Client means, in this context, decision makers in organizations. Server refers to software solutions, or search agents, which provide a service on information procurement for many users. Decision makers carry out a problem analysis to identify what their current problems are and what they urgently need. A list of requirements on information procurement is then derived from decision makers based on these problems. Typical situations of decision makers in companies are summarized as follows:

- Decision makers are roughly informed about an emerging technology, which has caught their interests. In order to assess the actual value of using the emerging technology in their own companies, they need to know more about it.

- In case that decision makers develop new applications (products or services) with the existing in-house technologies, they are eager to get the data on applications. First of all, decision makers put their interests to applications that are already released to the market: Which are those applications? Who are the suppliers? Are the applications produced by competitors? etc. Furthermore, the potential usages of the in-house technologies should be investigated. That means to check out which technologies are feasible for which application. Related information is therefore required by decision makers.

- Innovation relates not only to new applications, but also in production processes. For the purpose of saving cost, reducing production time, increasing machining precision, decreasing the complexity of producing process, or improving other impact factors, decision makers usually have to consider various possibilities of changing the production process using other technologies. Process innovation is commonly combined with product design. The current applications can be reconstructed regarding characteristics and functions offered by alternative technologies. For example, cell phones were fixed with an outer antenna previously; while at present, the technology MID (Molded Interconnected Devices) has enabled the installation of an inner antenna within the cell phones. Thus, the product design is improved; and the corresponding manufacturing process is adjusted. It is to decide here: when is the right time to replace the old technology? Which technologies are the substitutes of the old ones? Are those alternative technologies ready for series production? Which substitutes should be chosen to replace the current one? etc. In other words, the precondition for process innovation is to find out the explicit information about the technologies that can substitute those currently used in companies.

- Innovation is also driven by customers' requirements and market development. One of the trends today is that the development of some consumer electronics (e.g. mp3 players, cell phones, e-book readers, etc.) tends to the combination of mini size and multi-functions. This trend has an impact on the development of smart technologies that facilitate the manufacturing of mini electronics. At the same time, those technologies also accelerate the construction of innovative products. Obviously, decision makers should constantly monitor the market demand to keep their agile reaction to the dynamic changes.

- The last decision to be made is the "make or buy". When decision makers get into contact with some out-of-house technologies that are well usable for the in-house production, they usually need to draw the conclusions: should the technology be brought into the company's internal competence to ensure the self-manufacturing of corresponding product parts (make strategy)? Or is it more economic to buy the usage of the technology or product parts manufactured by that technology from external suppliers (buy strategy)? With the "make or buy" strategies, the purchasing and production process should coordinate in such a way that the profit is maximized. The principle rule here is:

  "Wir machen nur das, was wir besser können als andere und vom Markt honoriert wird." (*We make only what we can do better than others and is honored by the market.) [MB04]*

Information such as investment requirements, out-of-house supplier, application fields of the technology, technological complexity, and controllability are especially of interest for decision makers.

To sum up the points that we have just indicated, Fig. 2-8 shows the detailed information related to technologies, which are required by decision makers.

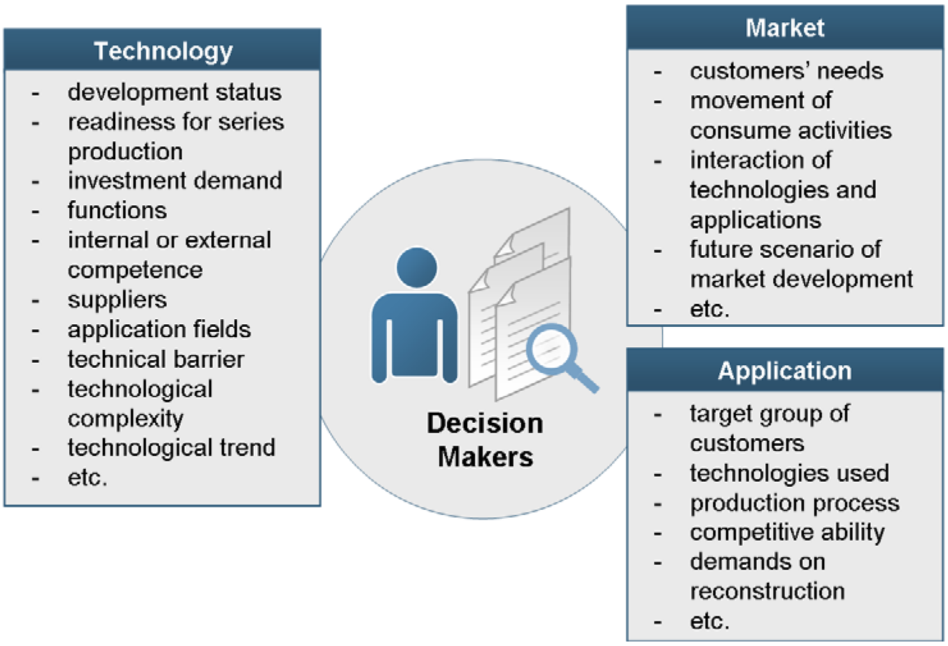


Fig. 2-8: Action fields of information

Decision makers formulate the requirements on information procurement with respect to their problems. After that, they put search requests including their requirements in software agents or other solution systems. The systems accept the search requests and forward them to corresponding search areas. Normally, systems are connected to a huge information base, which can be either local or open source. Enormous documents, diagrams, and metadata are all available in the information base. Software agents and systems communicate with the information base within the requests from decision makers. Through integrated processes, the requests are analyzed and the desired information is extracted from the information base. As a result, the retrieved information is sent back to questioners for the purpose of supporting their decision-making processes through the intelligent systems.

2.2.2.2 Challenges of Information Procurement

Based on the understanding of the general process of information procurement, and the embodiment in context of technology characterization and evaluation, the current challenges within the whole information procurement process can be discussed now.

- **Exact formulation of search requests**

  Business problems are so complex that decision makers do not know how to form their search requests. Simply described search request can lead to extensive hits (search results) with a high rate of inaccuracy. The reason is: it is not clear what to search. Search requests include criteria and limitations for information procurement, e.g. time period of the information, information about a certain technology, etc. Fluff information (waste information) is filtered out in the natural selection process of raw information by matching it to restrictions defined in search requests. The more conditions the request contains, the less the number of hits is; and hence the more concise the results are.

- **Huge amount of information**

  The information amount is too huge. As mentioned previously, we are now in the so-called "Information Age", where the information is being rapidly propagated. The modern information and communication technologies have undoubtedly a profound influence on the fast global distribution of information and the increasing amount of information. The world produces between 1 and 2 exabytes of unique information per year; and people can easily get data without going outside [LV03-ol]. The Internet is one of the youngest and fastest growing media to transfer information in today's world, whose growth is still accelerating, i.e. the Internet has not yet reached its highest expansion period [STI00-ol]. Even though, it has already caused an information expansion. We use World Wide Web as an example. The Web consists of approximately 2.5 billion documents with a growth rate of 7.3 million pages per day. Making an average estimation, the growth rate of new information is 0.1 terabytes per day [Wso00]. Another example is shown in Fig. 2-9. It demonstrates the development of GDSP (Global Disk Storage per Person) over time. GDSP is defined as the amount of digital storage space sold in a year divided by the world population of adults. Based on the fact that most published information is saved digitally, the calculation of GDSP reveals the chronologically changing of information amount.

To sum up the arguments above, we are facing the information explosion. Such a vast amount of information makes it no longer possible to analyze documents and other published information manually. Decision makers normally search a few documents on the topic of certain technology, and read them personally. Important information is then manually collected. This process is no longer suitable to solve contemporary problems. First of all, small groups of publications do not represent the whole field; incomprehensive information base leads to information loss. While on the other side, manual processing of comprehensive information base is not imaginable. A compromising solution is expected.

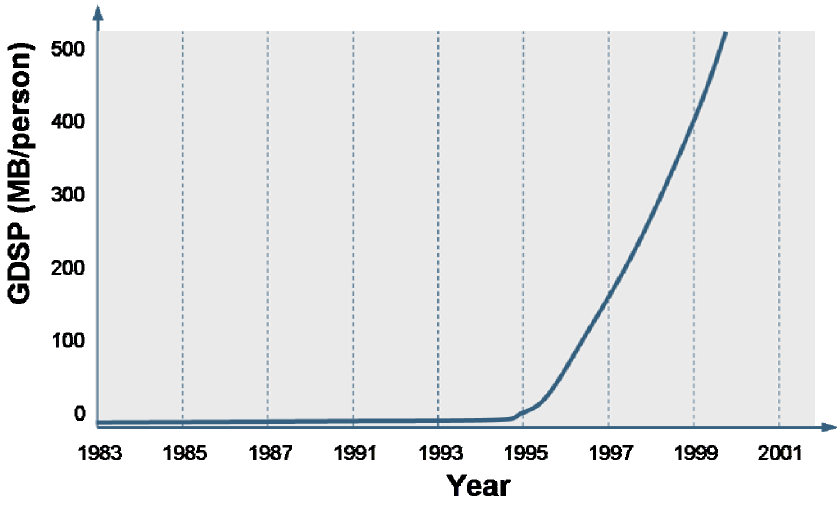


Fig. 2-9: Global disk storage per person over time [Swe01]

- **Information overload**

  As the amount of available data grows, the problem of managing the information becomes more difficult, which can lead to information overload. Information overload, also known as information flood, refers to the state of having too much information about a topic to make a decision [Tof70]. A world-wide survey [Reu96] found out that two thirds of managers suffer from increased tension and one third from illness because of information overload. Other effects caused by too much information include anxiety, poor decision-making, difficulties in memorizing and remembering, and reduced attention span [She97].

- **Duplication of data**

  Duplication of data is also a confusing issue for information procurement. It is difficult to distinguish if the data is “copied” or “original”. That results in information redundancy. The questions here are: what is the influence on information procurement? Should the duplicated information be filtered out in the pre-processing process? How to deal with duplicated information?

- **Accuracy of search results**

  It is difficult to achieve a high accuracy of search results. As discussed above, unclear defined search requests and the huge amount of information have negative effects on the accuracy rate of procured information. However, the correctness of search results also depends on further factors, for instance, the methods selected, the approach used, interpretation of search requests and information, etc.

- **Evaluation of the retrieved information**

  How to evaluate the procured information? We just spoke about the accuracy of results within the fourth point. In order to ensure and measure the accuracy, an evaluation step is required.

- **A lack of standardized process**

  There already exist several methods to simplify and speed up the process of information procurement, which have also proven valid in certain applications. However, these methods have not been systematically compared with each other. It is still hard to determine, which methods should be chosen. Explained in detail, the following questions should be addressed: which methods are supplements for each other? Which methods are best combinable? Which methods are laborsaving? Which methods are reusable? etc. Every decision maker has his own methods, which causes redundant internal work in the companies. There is a lack of standardized processes of information procurement, which aims at searching information relevant to technology planning, production innovation, and also other technology issues.

- **Update**

  Technologies change rapidly. So do the applications, market demands, and production processes. Decision makers require the up-to-date information about technologies and its related fields in order to keep their agile reaction to sudden changes or movements. Current update is either carried out without a standardized process or completely manually repeated, which wastes time and labor source.

2.3 Placement of this Dissertation

As discussed in section 2.2.1 and 2.2.2, relevant information and knowledge about technologies are the foundation of the innovative Technology Database. Therefore, an effective, intelligent information procurement process ensures the normal operation of the Technology Database, i.e. the generation of Technology Reports and Technology Roadmaps, and other search queries. Although the innovative Technology Database has proven successful in several industry projects, there are still some open issues in the part of information procurement to be addressed at the present (see the left side of Fig. 2-4).

Setting down the search targets (search objects)

Before the information is searched, it should be determined what the targets are. In the context of the Technology Database, the so-called Technology Indicators that describe the characters, performances, and trends of the technologies are firstly preferred by decision makers. It is necessary to pre-define all of the general Technology Indicators in order to minimize search focus.

Defining information resources

The large information volume exists within a complex area. Rational selection of information resources and limitation of search areas are the preconditions for the simplification of further information procurement processes. It is helpful to define the information resources for knowledge extraction of technologies: internal or external; internet article or papers from conference proceedings; data patterns or continuous texts; etc. Also rules for pre-filtering are expected.

Choosing intelligent methods instead of manual work

Progresses are requested for the information procurement process in the Technology Database. The traditional approach is based on manual work, which causes waste of time and labor resources. Additionally, the manual work can no longer catch up with the changing of information. However, business does not wait. Methods, which are suitable to automatic processing of vast information, are required to replace the manual work for the purpose of speeding up decision-making processes.

Standardizing the process

In order to save time, speed up decision-making process, and improve decision-making quality, decision makers urgently need a guide to extract desired knowledge from unstructured information. Therefore, to define a standard process of information procurement is one of the most important open issues for the innovative Technology Database.

Simplifying the update

The last open issue is that the information procurement process in the Technology Database should be intelligent enough to regularly re-procure, i.e. to update the information about technologies and its related objects with minimum efforts.

This dissertation is placed exactly to the process of information procurement for the innovative Technology Database, i.e. on the right side of Fig. 2-4. The open issues explored above should be solved and improved with the work in the thesis. The corresponding concrete requirements of the expected process that should be developed in this dissertation are listed below.

2.4 List of Requirements

As discussed before, this dissertation aims at developing an information procurement process, which can be embedded into the innovative Technology Database, for the purpose of supporting decision makers in the characterization and evaluation of technologies and other related activities. To sum up the chapter "Problem Analysis", the upper-class requirements are **large quantity and high quality**. Large quantity is necessary for efficient analysis of vast information; high quality means that the retrieved information should meet the decision makers' requirements. Neither of the two requirements can be dispensed. The **information procurement process** should be enforced from both sides. Furthermore, there are some detailed requirements for the information procurement process, which are listed as follows:

R1: It is capable to deal with a vast amount of information at a time.

It is mentioned in fore-content that the information repository should be big enough in order to offer a comprehensive overview of the target technology. Therefore, the information procurement should be valid for the processing of a huge amount of information.

R2: The information procurement process should facilitate the (semi-) automatic analysis of the information.

An engineer needs approximately 2 hours to read a 12-page scientific article. It can be imagined how long he needs to read 200 scientific articles. Traditional way of manual work is no longer suitable to process a huge amount of information because it is too time-consuming and leads to waste of human and material resources. Modern information technology facilitates computer-aided information search. It is required to select the best practical methods as well as corresponding countermeasures (e.g. to limit the search areas, to filter out the fluff information, etc.) for the information procurement process, which can realize the automatic search, process, analysis, and extraction of information. Based on the automatic process, the efforts needed from the side of decision makers can be reduced, and the process of information procurement and hence the process of decision-making can be accelerated.

R3: The information procurement process should be so effective that the only useful information/knowledge is procured.

Within the process, the needs of decision makers should be precisely analyzed. Only the information and knowledge relevant to target technologies should be extracted, i.e. a high rate of search accuracy is required. The results of the process should meet the decision makers' requirements.

R4: The information procurement process should be standardized and suitable to all the technologies.

Decision makers need a guide to follow within the information procurement process. It is important to develop a standard process that is feasible for the characterization and investigation of all technologies.

R5: The information procurement process is not a one-off. It should be active in the whole process of technology monitoring.

Technologies keep on developing. So do applications, markets and hence the decisions on business strategies, etc. Therefore, the decision makers are rather aggressive to grasp the up-to-date information. Therefore, the information procurement process should be iterative in order to keep on monitoring the dynamic changes of technology information. Some assistance phases should be defined reusable in order to simplify the update process.

Based on the requirements, the existing methods in the field of information procurement are reviewed in the next chapter, and an appropriate solution is proposed.

3 State of the Art

Information procurement is not only important for the innovative Technology Database, but also for other fields of research. Generally speaking, it has been an interesting issue for decades. Many researchers devoted themselves to developing, or optimizing methods that can be used to solve problems in information procurement. For instance, Expert Consultation, Information Retrieval, and Mining techniques are often used methods to extract knowledge from database.

In this chapter, the methods for extraction of knowledge from document collection, and other closely related or supporting methods are reviewed. The methods are then compared and evaluated according to the requirements determined in section 2.4. The options for action are derived from the evaluation of the methods, which are introduced below.

3.1 Expert Consultation

In the past, the most often used way to get relevant information was to seek experts' opinions. Users such as decision makers formulated their questions in questionnaires. Questionnaires were then sent to a small group of experts in a target field. Experts completed the questionnaires according to their experiences, and sent them back to decision makers. Decision makers analyzed the answers of experts. Based on the analysis of experts' opinions, business decisions were made by decision makers. Sometimes, the communication with experts was not a one-time doing, but a more-time iteration of opinion collection and feedback rounds.

For example, a **Delphi Survey** is a structured group interaction process that is directed in "rounds" of opinion collection and feedback. Opinion collection is achieved by conducting a series of surveys using questionnaires. The result of each survey is presented to the group and the questionnaire used in the next round is built upon the result of the previous round [RW78]. Fig. 3-1 shows the principal procedure of a Delphi Survey [GPW07].

The Delphi Survey has already been worldwide widely used. It helps to achieve convincing results especially by long-term and general questioning [Häd02]. For example, in the project "WZM 20XX", Delphi Survey was successfully carried out to ensure the future projections of key factors for the development of scenarios [Kin05].

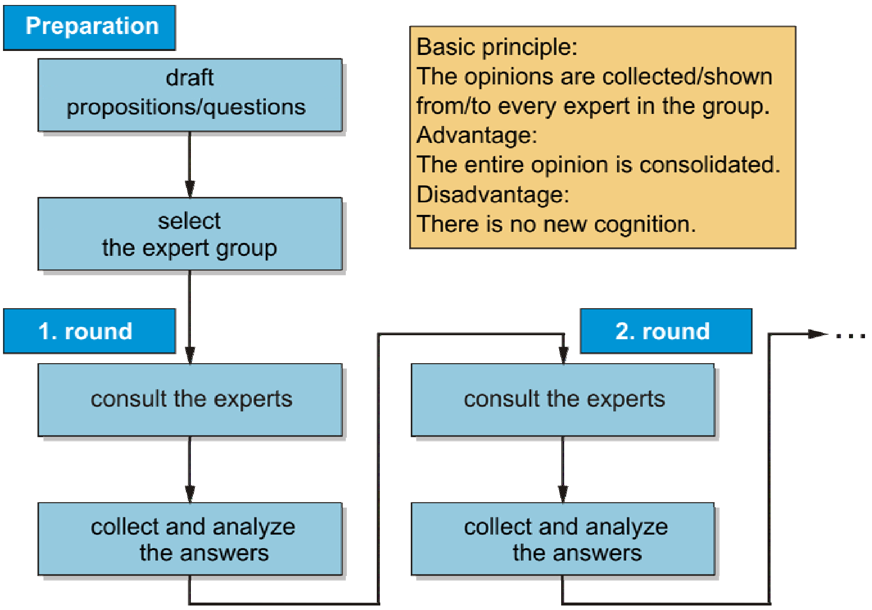


Fig. 3-1: The basic principle of Delphi Survey [GPW07]

Although Expert Consultation is a traditional way to obtain relevant information, it has some drawbacks: Expert Consultation is expensive, time consuming, and normally is carried out among a small number of experts; hence the survey is too small that its representativeness is open to question; furthermore, an inherent problem is that the experts' opinion offers, to a big extent, subjective information.

3.2 Information Retrieval

Knowledge patterns consist in various forms, e.g. text data, news papers, models, or hand writings. Widely and loosely defining, Information Retrieval is the art and science of searching for information in documents. With the spreading of modern digital technology and information median, e.g. Internet, e-journal, etc, most of the information is already digital available. Processing of text-based digital information becomes more and more important. The Information Retrieval introduced in this dissertation is mainly concerned with automatic Information Retrieval systems. [RKR+01]

Fig. 3-2 shows the work principle of Information Retrieval system (IR system), which includes phases of information/document search (D1-D5 on the side of demanders) and information/document preparation (P1-P4 on the side of provider). Information demander enters a request into the IR system. The documents are searched in the indexed data restricted to certain attributes (e.g. author, title, publish year), or a full text search is facilitated. The search requests in this case are often formulated or translated into queries with Boolean Logic (AND, OR, NOT, etc.) [Wil00].

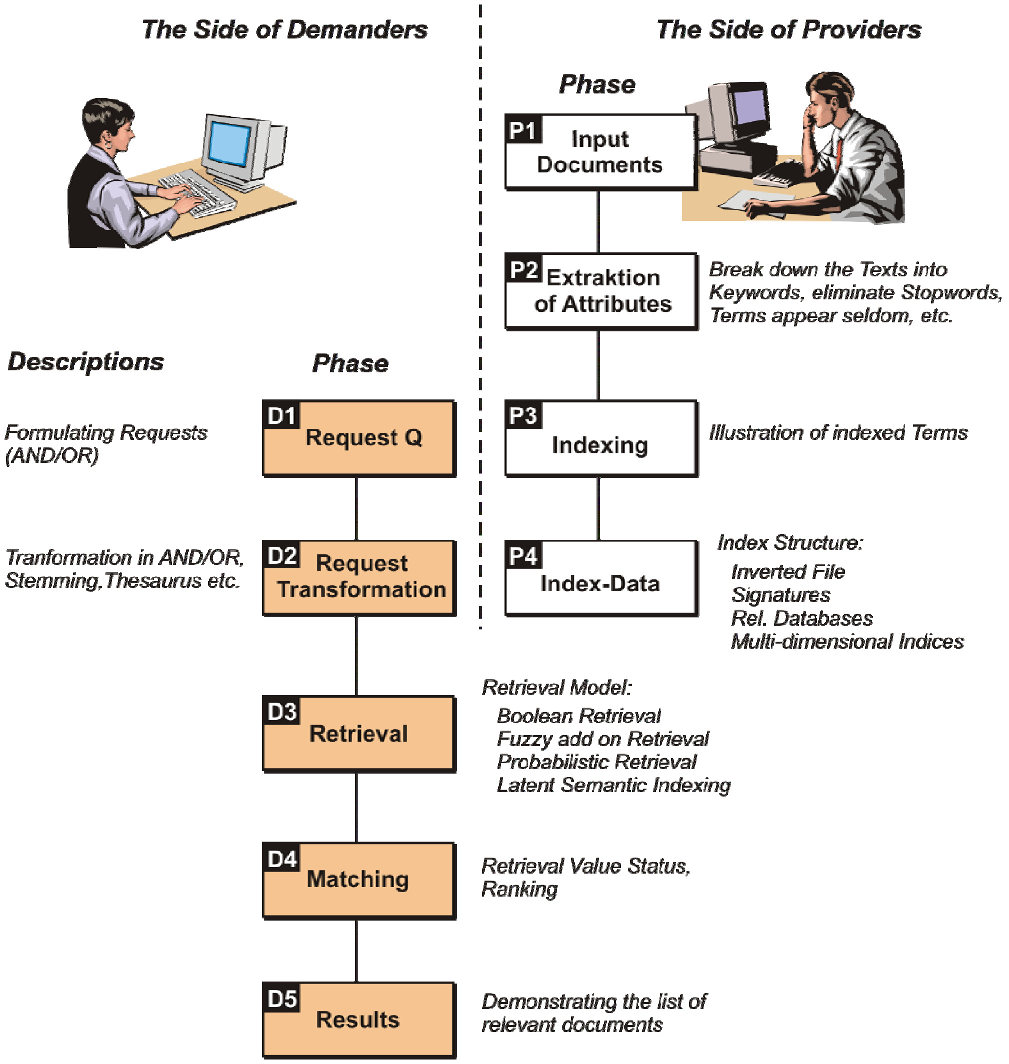


Fig. 3-2: Function model of Information Retrieval system [GHK+06]

IR is interdisciplinary, based on computer science, mathematics, library science, information science, linguistics, statistics, etc. It is used in almost all processes of information search or information procurement. Web search engines such as Google or Yahoo are the most visible IR applications. However, IR only works as document finder in the whole information procurement process. To go deeper to analyze the content of documents, other methods are required. [BR99]

3.3 Artificial Intelligence

The term Artificial Intelligence (AI) was first proposed by McCarthy as "the science and engineering of making intelligent machines" in 1956 [MMR+56-ol]. There also exist other definitions in AI history.

"The automation of activities that we associate with human thinking, activities such as decision-making, problem solving, learning …" by BELLMAN, 1978 [Bel78]

"The study of how to make computers do things at which, at the moment, people are better." by RICH and KNIGHT, 1991 [RK91]

"… the study of the computations that make it possible to perceive, reason, and act." by WINSTON, 1992 [Win92]

"AI … is concerned with intelligent behavior in artifacts." by NILSSON [Nil98]

Simply speaking, AI is the science and engineering of making intelligent machines, especially intelligent computer programs think and act like humans. AI has an interdisciplinary foundation containing philosophy, mathematics, economics, neuroscience, computer engineering linguistics, etc. [RN03]

AI is used in extensive fields like planning and scheduling [JMM+00], natural language, facial recognition, software application, and strategy games like computer chess or other video games. Just few applications of AI are sampled here, which is enumerated below; others appear throughout the dissertation.

Understanding Natural Language

Natural language is a versatile means of communication, but it is not easy for computers to understand. AI facilitates the understanding of natural languages by using intelligent algorithms to parse sentences, analyze sequence of words and their relationships, etc. [LKS99]. Usually, the AI systems are based on a big database of old examples in a certain domain. Therefore, they understand only domain-specific languages.

Expert Systems

An expert system is a computer program, or a knowledge based system, which contains some domain-specific knowledge as well as the knowledge from human experts, and has an analytical skill. An expert system "interviews" human experts in a certain domain, pockets their knowledge in a computer program, and tries to analyze problems, make decisions like a human experts. [FMN88]

Expert systems are already used in the fields of accounting, medicine, process control, financial service, production, human resources etc. For example, one of the first expert systems was Mycin, which was developed in 1974 to diagnose bacterial infections of the blood and suggest treatments. Mycin and several other expert systems have proven applicable in practice. However, an expert system works depends on the current development of AI. Furthermore, the knowledge bases and the rules that are usually defined by the users of the systems also strongly influence the expert systems' efficiency.

Heuristic Classification

One of the most feasible applications of AI is heuristic classification. Normally, there are three steps of heuristic classification process, as described in following:

1. abstraction from a concrete and particular problem description to a problem class,
2. heuristic match of a principal solution (method) to the problem class, and
3. adaptation of the principal solution to a concrete solution for the concrete problem.

Example for heuristic classification is to give suggestions of whether to accept a proposed credit card purchase. The records of payment, past examples of frauds are pre-saved in a database. So does information about the credit card, the owner, the item bought, the purchase system, etc. Information about new purchase activities are compared with the fraud database. If there is a match, system will give a warning that there might be unsafe information for this credit card purchase. [Mcc04]

Except the application fields introduced above, AI can also be used in other fields like robotics, game playing, and so on [WR03] [GK97]. However, AI is not feasible in complicated environments. The computational demands are just too high. In conclusion, AI has made great progress in its short history, but there are still many remains to be done to achieve more success. [RN03]

3.4 Mining Technique

Besides Information Retrieval and Artificial Intelligence, another technique – Mining Approach, which is used for extracting knowledge automatically from the information pool, has gained more and more attention of researchers [Gen99]. Different to other methods of information procurement, Mining Approach does not only find the relevant information, but also analyzes data to extract interesting and useful information (i.e., knowledge) through certain mining processes or algorithms. Basically, there are two Mining Approaches: Data Mining and Text Mining [GG00].

3.4.1 Data Mining

Data Mining aims at extracting novel and useful information from structured data. It works towards finding patterns automatically from huge amount of information and using them to improve decision making [BL97]. The patterns can be categorized into three kinds: strong patterns (regularities for numerous objects); weak patterns (reliable exceptions representing a relatively small number of objects); and random patterns (random and unreliable exceptions). Traditional Data Mining Approaches are able to find the strong patterns, which are more accurate and highly predictive. However, the weak patterns are more interesting for researchers in some cases because they are more unexpected, unknown, and show weak signals. Therefore, most current Data Mining processes focus on reorganization of week patterns [LLF+99].

Data Mining is commonly used in supporting business intelligence, financial analyses, and also increasingly for extracting scientific information from the enormous data sets. An example here is the Market Basket Analysis used in retail sales, which helps to identify which kind of products are most preferred by customers from transaction-based data [SPB07]. Typical applications for Data Mining are characterized with the following four functions [HK06]:

- Association analysis, which discovers interesting relationships hidden in a large database, e.g. Market Basket Analysis.
- Classification, which means the individual items are placed into groups according to quantitative information of one or more inherent characters, and based on a training set of previously labeled items. Filtering of spam emails is an example for classification in practice.
- Prognosis, i.e. to forecast the development trend of objects, e.g. to predict the course of foreign exchange rate.
- Clustering, this refers to normally automatic partitioning of a data set into subsets (clusters) according to similarities of objects. Proximity is an indispensable index in this case, which measures the similarity of objects and is often visualized as distance between objects. The example here is segmenting the market in order to determine the target markets.

3.4.2 Text Mining

Text represents knowledge [Lis01]. Different to data, text is unstructured, which makes the knowledge extraction from text more complicated and has relative high requirements on intelligence processes. Text Mining attributes the extraction of novel, useful, information patterns (i.e. knowledge), and their content-relevant connections from unstructured data sets (texts) [HQW06]. Similar to Data Mining, Text Mining is based on statistical and pattern-based approaches. However, Text Mining usually involves additional processes, such as to structure the text (e.g. to tag the texts in a HTML format), and to extract patterns from the structured data. Furthermore, Text Mining is always combined with visualization of extracted information patterns, for instance, in a Knowledge Map (Fig.4-4). Therefore, it is also necessary to interpret the results by the aid of visualization views.

Text Mining allows innovative applications in knowledge management within extensive fields. Typical Text Mining tasks include text categorization, text clustering, extraction of key concepts or entities, production of granular taxonomies, sentiment analysis, document summarization, and so on. [WIZ+04]

Mining Approaches are commonly used in various fields ranging from analysis of purchase activities of customers to strategic forecasting. They are driven by the emerging and strongly increasing demands of processing and extracting knowledge from large databases. Therefore, the Mining Approaches are basically often combined with databases and data warehouse applications. Mining Approaches should be continued to be developed and optimized. [FS06]

3.5 Ontology

Human and application systems need to communicate with each other and between themselves. However, there can be widely different meaning of the terms due to various background contexts or viewpoints, which may cause misunderstanding, thus complicate the human–machine communication [Poc00] [PRR98]. For example, the word "Amazon" can be understood as a river, a women warrior, or a company (Fig. 3-3).

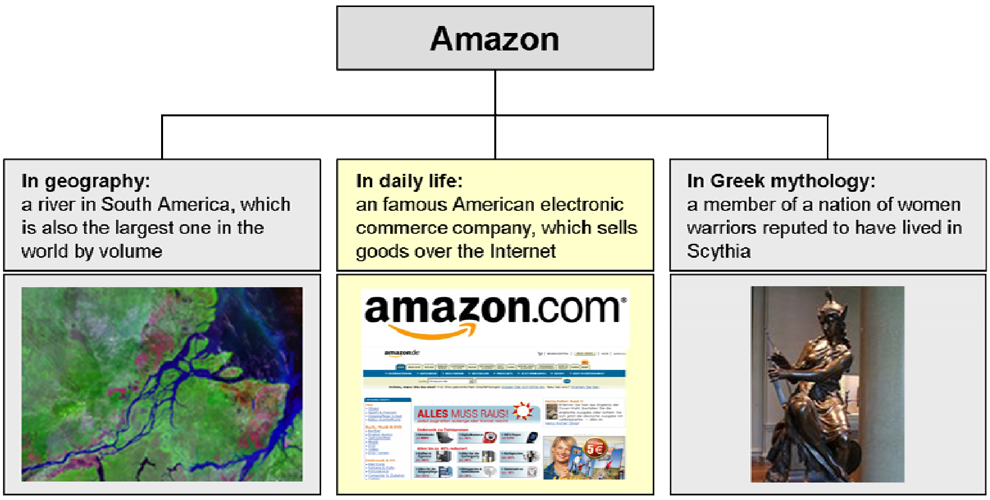


Fig. 3-3: Various understandings of the word "Amazon" in different surroundings

The solution of reducing conceptual and terminological confusion and sharing understanding is to build up ontology [UG96]. Ontology means, in philosophy, the study of the nature of being and the essence of things. In the early 1990s computer scientists took over the term and gave it a new, but related meaning. Therefore, in the knowledge representation community, ontology is the formal specifications of a certain domain, which facilitate the exchange and sharing of knowledge. Today, the most often used and highly cited definition is from Gruber:

> *"An ontology is a formal, explicit specification of a shared conceptualization. 'Conceptualization' refers to an abstract model of phenomena in the world by having identified the relevant concepts of those phenomena. 'Explicit' means that the type of concepts used, and the constraints on their use are explicitly defined. 'Formal' refers to the fact that the ontology should be machine readable. 'Shared' reflects that ontology should capture consensual knowledge accepted by the communities." [Gru93a]*

As the shared and common understanding of the domain, ontology can enhance the communication of human–machine. It optimizes the semantic interpretation of information. Furthermore, ontology is reusable and extendable [SS01].

3.5.1 Main Elements of Ontology

Ontology provides a common vocabulary of an area and defines, with different levels of formality, the meaning of the terms and the relations between them. There are four main elements of Ontology: concepts (classes), relations, properties and instances [Gru93b]. A simplified example of ontology is shown in Fig. 3-4.

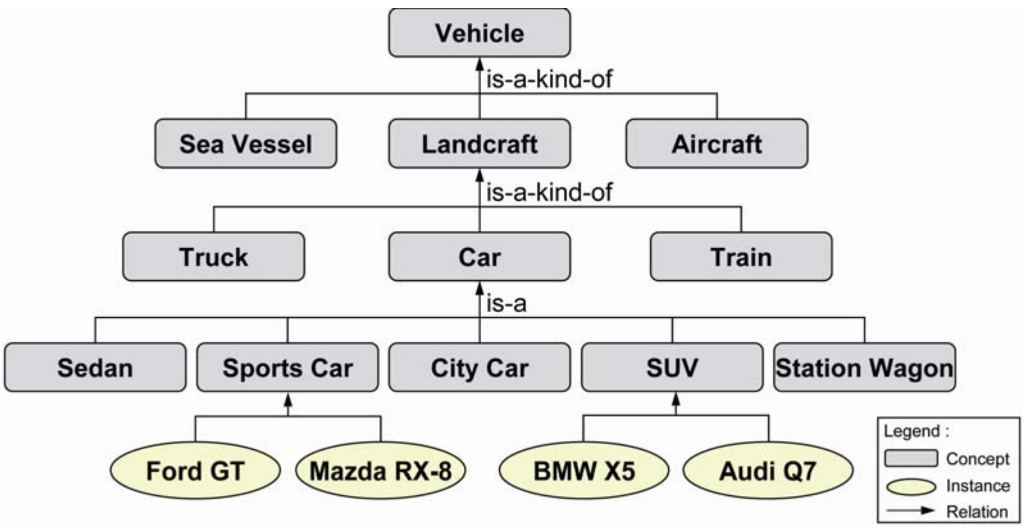


Fig. 3-4: Simplified car ontology [Bre03]

A **concept** represents a set of entities (things) within a domain. It can be abstract or concrete, elementary or composite. Usually, the concepts are organized in taxonomies. That means all concepts of ontology are hierarchically structured. Synonym of concept in the context of ontology is class.

Relations describe the interactions between at least two concepts. A connection between two concepts is called binary relation. Examples of binary relations are: is-a, has-a, subclass-of, or connected-to.

Attributes include various features, properties, characteristics, or parameters that a concept can have or share. For example, color and brand can be two attributes of the concept 'car'.

Instances (also called individuals), which is the basic or "ground level" in an ontology, give intuitional presenters of the concept.

3.5.2 Types of Ontology

Based on argumentations of GUARINO, ontology can be divided into four types according to detailing levels and reusability: upper ontology, domain ontology, task ontology, and applications ontology [Gua97] [Kli03]. The more detailed the ontology is, the less reusable it is. Upper ontology is the most general one, which does not describe the objects in deep details, and is often reused; while application ontology is restricted to a small field, which is more detailed but less reusable (see Fig. 3-5).

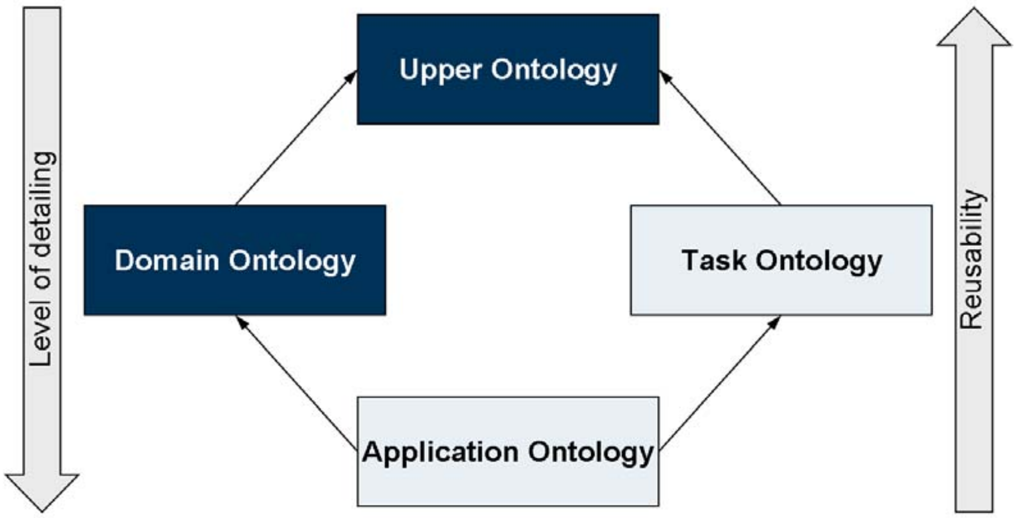


Fig. 3-5: Types of ontology according to GUARINO [Gua97]

Upper ontology, also known as foundation ontology, is a hierarchy of entities and associated rules (both theorems and regulations). It describes those general entities that across a wide range of domains. The aim is to have a large number on ontologies accessible under this upper ontology. Strictly speaking, upper ontology is actually not a real ontology but a unique combination of a taxonomy and a controlled vocabulary.

Domain ontology, also called domain-specific ontology, describes the concepts in a specific domain (a part of the world). The meanings of the concepts in domain ontology are restricted to that particular domain. Here, the concepts in upper ontology are specialized.

Task ontology is at the similar detailing and reusable level as domain ontology. The vocabularies in task ontology are concerned to the generic activities and tasks. It is both domain and application independent. Analogue to domain ontology, the concepts in upper ontology are specialized.

Application ontology focuses on a special, concrete domain or task ontology, which usually particularizes the concepts in a domain or task ontology.

In order to make two parties with different ontologies understand each other, ontology mapping comes into play. Ontology Mapping is the process whereby two ontologies are semantically related at conceptual level, and the source ontology instances are transformed into the target ontology entities according to those semantic relations [Poc00]. Besides mapping, ontologies can also be merged, or aligned. The differences among mapping, emerging, and articulation are shown in Fig. 3-6.

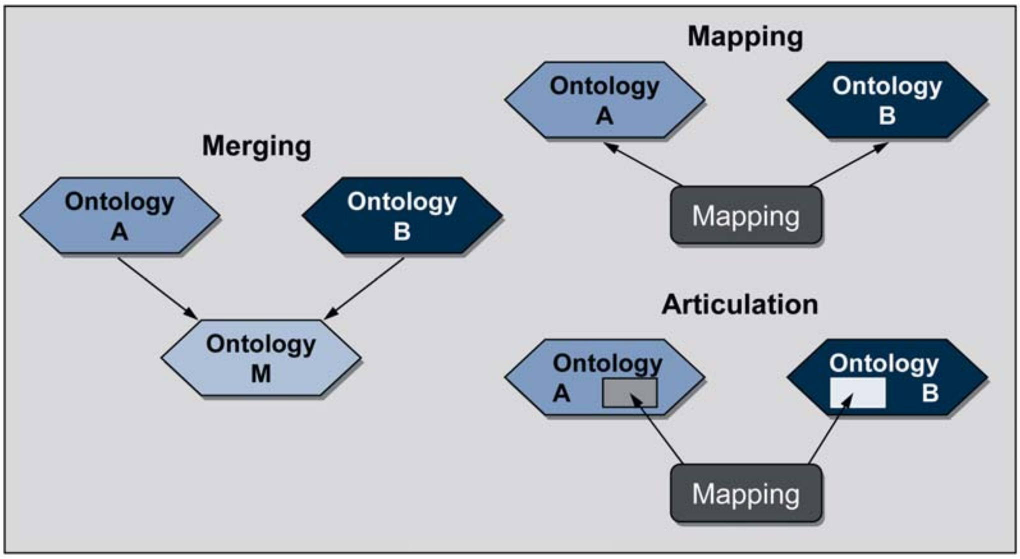


Fig. 3-6: Different ways to combine ontologies: merging, mapping, and articulation [Noy05]

3.5.3 General Generation Process of Ontology

In the literature, there are many processes of generating ontology proposed by ontology experts over a time period. In this dissertation, two basic processes are introduced:

According to GUARINO the basic design principle is embodied in four steps: 1) be clear about the domain; 2) take identity seriously; 3) isolate a basic taxonomic structure; and 4) identify roles explicitly. [Gua98]

USCHOLD and GRUNINGER proposed a purely manual process of building ontologies: 1) identify purpose and scope; 2) build the ontology in three steps: identify the key concepts and relationships, and their corresponding definitions; represent the concepts, relations in ontology language; integrate with existing ontologies; 3) evaluation; 4) documentation; 5) guidelines for each of the previous phases. [UG96]

The final built ontology should have the following characters [Hwa99]:

- clear, i.e. definitions should be maximally clear and unambiguous;
- consistent and coherent, that is to say, an ontology should be both internally and externally consistent;
- extensible and reusable, which means an ontology should be designed in such a way in order to maximize subsequent reuse and extensibility.

3.5.4 Available Ontologies

Within the development of ontology, there are already some ontologies existing and available for suitable usage. In the following texts, three of the most famous and usable ontologies are introduced.

Cyc

Cyc is a well-known and quite comprehensive ontology available today, which is a proprietary system developed since 1985. Cyc consists of an upper ontology and several domain-specific ontologies. A subset of Cyc, called OpenCyc, has been released for free [Cyc07]. And an almost unabridged version with the name ResearchCyc is available for non-commercial use.

Website: http://www.cyc.com/

Basic Formal Ontology (BFO)

The Basic Formal Ontology framework has been developed and formulated by Barry Smith, Pierre Grenon, and the associates. BFO is narrowly focused on the task of providing an upper ontology in order to support domain ontologies developed for scientific research. Thus BFO does not contain physical, chemical, biological or other terms which would properly belong to the domain of the special sciences. [BFO07]

BFO consists in a series of sub-ontologies at different levels of granularity. The most important sub-ontologies are: SNAP and SPAN. SNAP is series of snapshot ontologies, indexed by times. SPAN is a single videoscopic ontology. Each SNAN ontology is an inventory of all entities existing at a time. Each SPAN is an inventory of all processes unfolding through time. Interrelations are defined between the SNAP and SPAN in a way that makes BFO capable to deal with both static/spatial and dynamic/temporal features of reality.

Website: http://www.ifomis.uni-saarland.de/bfo/

WordNet

WordNet is originally designed as a semantic network based on psycholinguistic principles. It is also a freely available database that was expanded by adding definitions. WordNet includes most general concepts as well as some specialized concepts. All the concepts are related to each other by the subsumption relations, part-of, cause relations, and other semantic relations. Those logical relations between the concepts in WordNet have not been precisely defined compared with Cyc. WordNet is now viewed as a dictionary, and has been widely used in Natural Language Processing research. [WN07]

Website: http://wordnet.princeton.edu/

3.5.5 Ontology Applications

Basically, ontology shares the common understanding of the domain, and hence facilitates the human – machine communication. In particular, ontology's applications have been extended to different domains especially like areas dealing with vast amounts of distributed and heterogeneous computer-based information. Examples for those application fields are: World Wide Web or Intranet information systems, complex industrial software applications, knowledge management, electronic commerce, and e-business [DF02]. Furthermore, it has a strong influence on the creation of semantic relationships between various pieces of relevant and useful information for the purpose of enhancement of learning experience in a web-based educational environment.

Although ontology can be already used in many fields like those mentioned above, it is still to notice that, there are some barriers of using ontology [SSV02]:

- It takes time to assess whether an existing ontology is suitable.

- It is difficult to find existing ontologies that exactly satisfy users' needs. Ontology libraries and ontology brokers are still in their infancy.
- It is costly to build an ontology independently.
- It is a risk to commit to an ontology whose stability is not assured.

To sum up, ontology is indispensable to avoid misunderstandings. Anyway, it is not a dominant method, but only a means of support in the process of information procurement. Other methods that can directly extract information are required.

3.6 Bibliometric Analysis

Bibliometrics, just as its name implies: "biblio" means literature; "metrics" means measure. Joining together, it means the measurement of literature. To explain it in detail, Bibliometrics is devoted to quantitative studies of literature [Gor92] [KS95].

The term "Bibliometrics" was firstly introduced by Pritchard in 1969 as "the application of mathematical and statistical methods to books and other media of communication" [Pri69]. Since the eighties, computer and information science & technology have developed fast, which facilitate large bibliographic databases in machine-readable form. Till then, Bibliometrics has evolved into a distinct scientific discipline with a specific research profile, several subfields and the corresponding scientific communication structure. The primary symbols for that development of Bibliometrics are: the international journal Scientometrics was published in 1979 as the first periodical specialized on Bibliometric topics; international conferences on Bibliometrics occurred since 1983; the journal Research Evaluation has been published since 1991 [Gla03].

Bibliometric assessment of research performance is based on one central assumption: scientists do vigorously publish their research output in the form of an open, international publication [Raa01]. Publications are not the only, but certainly very important elements in the process of scientific communication. Therefore, statistical analysis of publications (literature) can reveal and measure the research activities [KMS93]. Anyway, there was a period, when people suspected the validity of Bibliometric methods, because the quality of literature is extensively various. Anyway, some countermeasures, such as normalization process, impact factor, etc., were created to minimize the shortcoming of Bibliometrics analysis. In a research paper dealing with the measurement of scientific activities with Bibliometric indicators, GODIN described the present status of Bibliometrics:

> *There may have been a time when the fact that Bibliometric indicators were standardized limited their usefulness, but this is no longer the case. Furthermore, they are not expensive to produce. They do have their limits, notably because they normally include only the natural sciences, engineering, and the biomedical sciences. There is also an obvious linguistic bias that largely limits the coverage of scientific output to publications in English. Finally, it must be remembered that publishing represents only one of the activities of researchers. In spite of such limits, Bibliometric indicators are one of the principal tools for measuring research output, while providing a very good tool — contrary to popular belief — for research conducted by other types of actors. For this reason, they deserve a place in scientific and technological directories. [God96]*

Due to its significant advantages, Bibliometrics are paid more and more attention to. And it is still considered nowadays as one of the most important methods to measure literature with different objectives. Bibliometrics was used in the field of library cataloging and classification in the past. In recent years, Bibliometrics has been increasingly related to the investigation of scientific excellence and research outcomes. Quantitative analysis and statistics are used to describe patterns of scientific publications within a given field or the body of literature itself.

It is to notice that there are three levels of aggregation in Bibliometrics research, which are described as follows:

- macro-level: to analyze the publication output in a field as a whole, e.g. a whole country
- meso-level: to analyze the research performance of universities or major parts of universities, e.g. faculties or institutions
- micro-level: to analyze the work of individuals, research groups or programs, which are the real "working floor" of research practice

From the viewpoint of Bibliometric methodology, the distinction between three levels of aggregations is important. Each level of aggregation has its own mathematical and statistical background, and requires different Bibliometric processes. [Raa03]

Furthermore, the approaches of Bibliometric Analysis are also divided into two ways: one-dimensional Bibliometric Analysis, and two-dimensional Bibliometric Analysis. In the following sections, the two approaches are explicitly introduced.

3.6.1 One-dimensional Bibliometric Analysis

One-dimensional Bibliometric Analyses include traditional Publication Analysis and Citation Analysis [Kin87] [Gor05-ol]. Publication Analysis deals with counting of publication numbers according to time, region or other criteria. The hypothesis is: the numbers of publications can reveal present and past activities of scientific work. For instance, as shown in the middle of Fig. 3-7: the “Canadian share of world publications” is based on the publication numbers counted by countries, which shows that 31% of the worldwide publications were produced in USA in 1995; Japan and UK were both in place two with 8%; and Canada was the sixth most prolific country in the world; etc. Based on the assumption of Publication Analysis, it is estimated that USA was the most active country in scientific area in 1995, followed by Japan and UK, then Germany, France, Canada, and so on. Similarly, the distribution of Canadian publications by province or sectors reflects the research activities at provincial level or in concrete sectors of Canada in 1995.

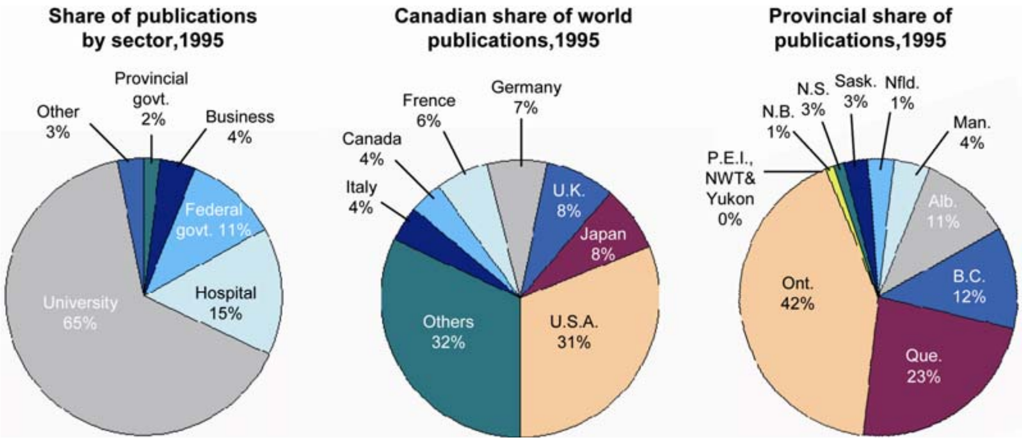


Fig. 3-7: Examples of Publication Analysis – three aspects of Canadian scientific output, 1995 [Gau98]

Fig. 3-8 illustrates another example of relative Publication Analysis. The data is provided by ISI (Information Sciences Institute) product “Life Science Series”, which summarizes information about publications that appeared in 1,374 periodicals devoted to the research field of life sciences. In this case, 248,381 articles from 21 monitored countries appeared in 1998 were focused. The numbers of publications were normalized by inhabitants in a given country. The normalization reduces redundancy of data, and hence ensures the objectivity, accuracy, and impartiality of the analysis results.

Another important one-dimensional Bibliometric Approach is the Citation Analysis, which is the examination of the frequency and pattern of citations in articles and books [Rub04] [LS03]. The numbers of citation indicate the importance of the article on the assumption that the more often the article is cited, the more important it is. That assumption was firstly formulated by PRICE in 1976 as follows:

> *“Success seems to breed success. A paper which has been cited many times is more likely to be cited again than one which has been little cited. An author of many papers is more likely to publish again than one who has been less prolific. A journal which has been frequently consulted for some purpose is more likely to be turned to again than one of previously infrequent use.” [Pri76]*

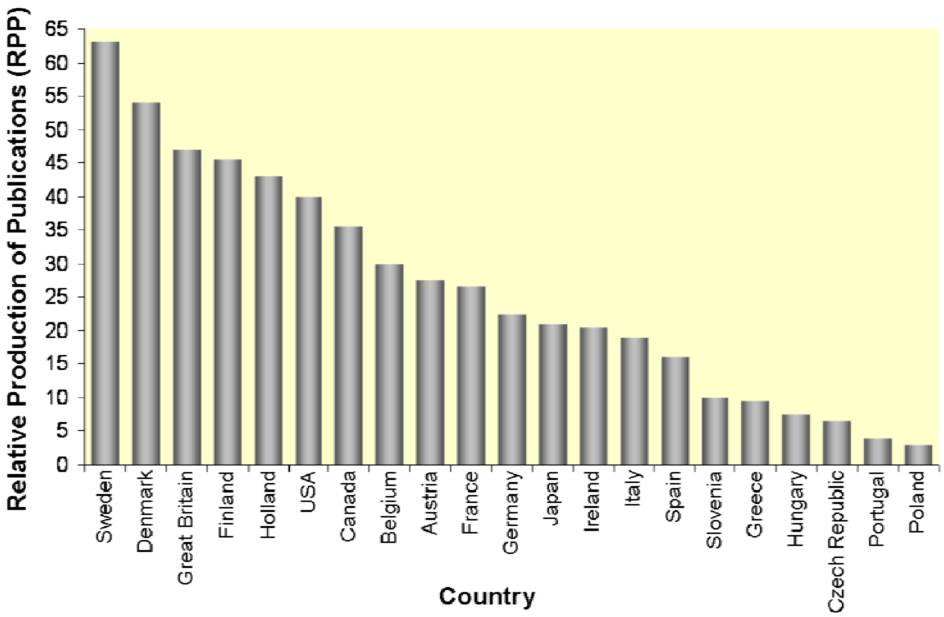


Fig. 3-8: Examples of relative production of publications – numbers of publications per 100,000 inhabitants in a given country in the research field of Current Contents Life Sciences (1998) [KBH99]

A very important application of citation is the Journal Impact Factor (JIF), which is a measurement of the frequency with which the "average article" in a journal has been cited in a particular year or period. JIF is used for ranking, evaluating, categorizing, and comparing journals. The calculation of JFI is defined as dividing the number of current year citations to the source items published in that journal during the previous two years [SCI93]. The following descriptions show the detailed steps to calculate JIF. It is to notice that the 1992 or 1990-91 has no real meaning and stands for publication years.

A= total cites in 1992

B= 1992 cites to articles published in 1990-91 (a subset of A)

C= number of articles published in 1990-91

D= B/C = 1992 Impact Factor

The JIF is useful in clarifying the significance of absolute citation frequencies for the purpose of providing a gross approximation of the prestige of journals. It eliminates some of the bias especially for the following cases: large journals vs. small ones, or frequently issued journals vs. less frequently issued ones, or older journals vs. newer ones.

To sum up, the absolute or relative numbers of publications or citations measure the volume and impact of research work at various levels. That may be an author, an institution, a sector of activity covering several institutions (universities, public laboratories, industries), or even a geographic area (city, province, country). When the number of publications or citations is counted over prolonged periods of time, they provide a means of identifying trends.

3.6.2 Two-dimensional Bibliometric Analysis

Two-dimensional Bibliometric Analysis is related to the measurement of similarity by counting of co-frequency of two publications or two elements of publications. The methods covered here are: Co-publication Analysis, Co-citation Analysis, and Co-word Analysis.

First of all, a brief overview of **Co-publication Analysis**, namely the measurement of co-occurrence of publications according to countries, authors, disciplines, and also other criteria, is introduced. The co-publications reflect the relationships among the investigated objects. For example, co-authorship measures the co-operation activities of researchers in a given time period. As shown in Fig. 3-9, there has been a large increase in multi-national studies. In 1998, the USA was Japan's leading research partner with over 6,000 collaborative papers, which is even larger than with all other Asian nations.

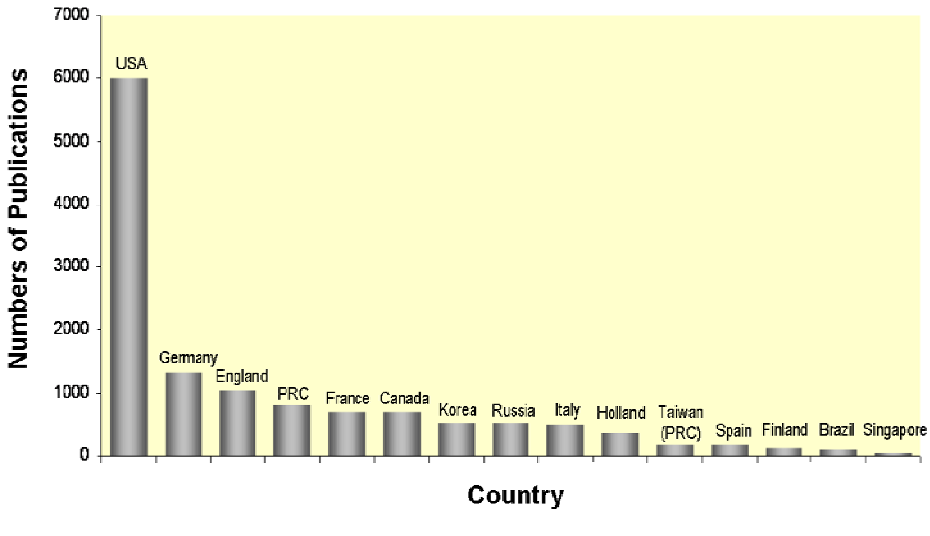


Fig. 3-9: Japanese co-authorship with other countries based on the Japanese publications in 1998 [Gar99]

Co-citation Analysis was defined by Small in 1973 as "the frequency with which two items of earlier literature are cited together by the later literature" [Sma73]. Publications are connected to each other through co-citations. The more frequent the two publications are co-cited by others, the stronger they are related to each other. Fig. 3-20 shows graphically the explanation of co-citation.

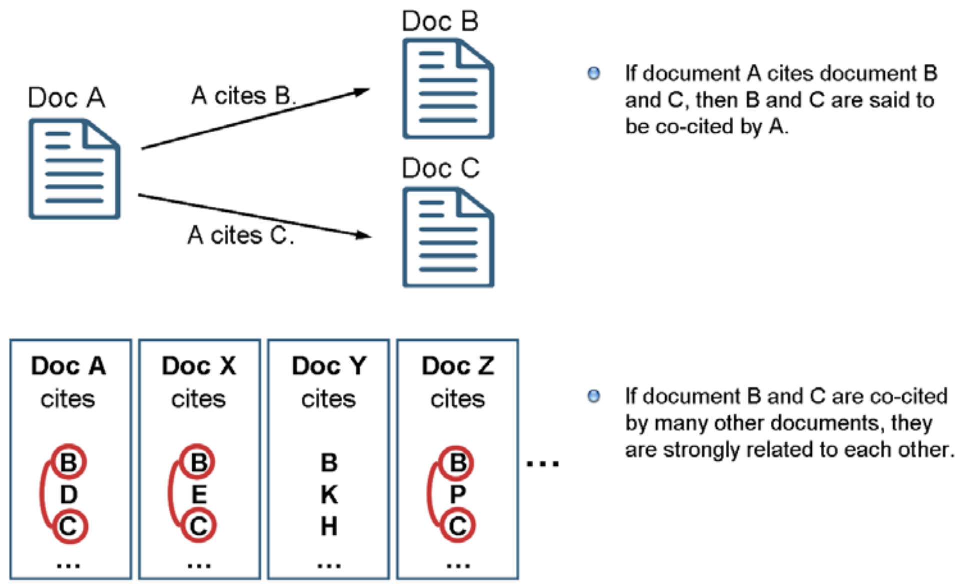


Fig. 3-20: Illustration of co-citation linkage

Co-citation Analyses have been successfully applied to examine the intellectual structure of many disciplines and to show significant clustering of topically related authors [WG81].

Co-word Analysis is the most important method used for content analysis, which counts and analyzes co-occurrences of keywords in the publications on a given subject [KS98] [NFS02]. Co-word Analysis draws upon the assumption that a paper's keywords, which are the important carrier of scientific concepts, ideas and knowledge [Raa93], represent the main topics of the papers and the relationships of the keywords indicate the link of those topics.

The common process to carry out Co-word Analysis is to break down the contents into words. The words are then reduced into keywords, which reflect the main topics of the papers. After that, the absolute text frequencies of keywords as well as the frequencies of co-occurrences are calculated and shown in a matrix [SK95]. Fig. 3-21 shows an example of a co-word Matrix.

Based on the calculation of co-occurrences, the keywords can be located in a Knowledge Map by using MDS (Multi-dimensional Scaling). The Knowledge Map can be read according to the following rules (Fig. 3-22).

Document 1
Augmented Reality
Product Innovation
Automobile Industry

Document 2
Mechatronics
Augmented Reality
Product Innovation
Prototype

Co - Occurrences

Keywords \ Keywords	Product Innovation	Augmented Reality	Automobile Industry	Mechatronics	Prototype
Product Innovation	2	2	1	1	1
Augmented Reality	2	2	1	1	1
Automobile Industry	1	1	1	0	0
Mechatronics	1	1	0	1	1
Prototype	1	1	0	1	1

Fig. 3-21: Calculation of co-occurrences of keywords [BC06]

Every pellet in the map stands for a keyword. The diameter means the text frequency of the keyword which is represented by the pellet. The hypothesis for Co-word Analysis is: the more often the keywords appear together in documents, the more content-similar they are. So the keywords describing similar topic are positioned in the vicinity [Raa04]. For example, the word "mechatronics" is always located in the near of the words "mechanics" and "electronics" because they always appear together in the same documents. The thickness of the lines between the keywords represents the relative co-frequency.

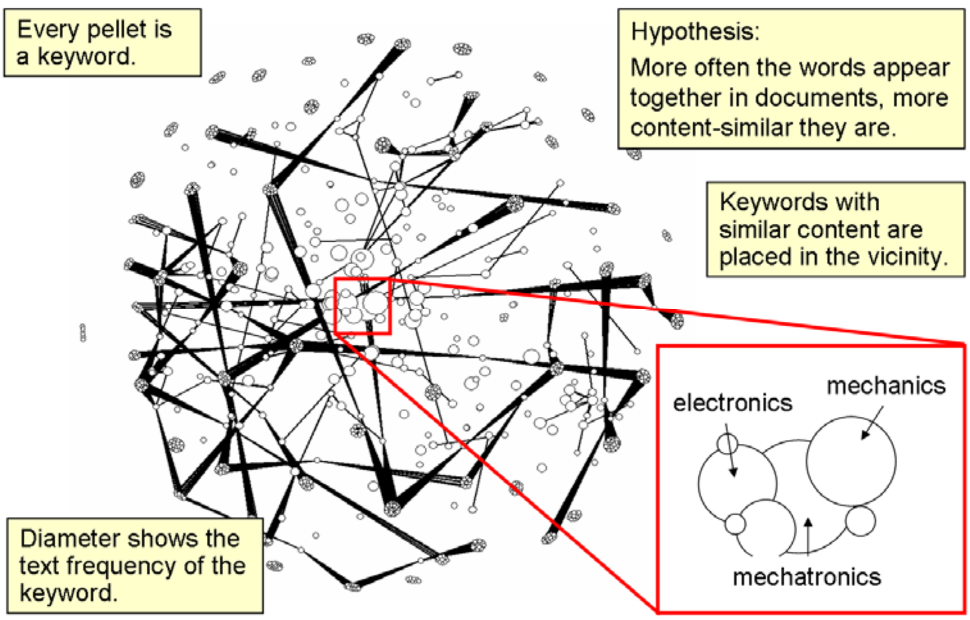


Fig. 3-22: Knowledge Map based on Co-word Analysis

Co-word Analysis has been used as an important method to explore the concept network in different fields, to extract research topics and to trace changes during time periods. The advantages of Co-word Analysis combined with Knowledge Maps are: it reduces and projects the data into a specific visual representation with the maintenance of essential information contained in the data; it enables the structuring of data from various perspectives: main topics of the publications (keywords and their absolute text frequencies); relationships of the publications (keywords networks extracted from the contents of investigated publications); and transformation of keywords networks over time periods. Furthermore, Knowledge Maps facilitate the visualization of interactive keywords networks within a small space, which is easy to understand, and still indicative of interrelated concepts in the literature.

3.6.3 Patent Analysis

Patent Analysis is a sub-area of Bibliometric Analysis. Scientific articles are restricted in this context to patents. Relevant information of patents can be retrieved from the patent databases, most of which are already online available. The databases of the United States Patent and Trademark Office (USPTO) and of the European Patent Office (EPO) are usually the most frequently used databases for analyses of patent-literature linkages. Moreover, subject-related patent information can also be retrieved from the domain-specific bibliographic databases, e.g. Chemical Abstract Society (CAS). Information about patents is structured and saved in patent databases. The most important patent information used in Bibliometrics is shown as follows:

1) Patent identification
2) Names of inventors
3) Assignee
4) Addresses
5) References (patents and other publications)
6) Abstract
7) Classification

Patent Analysis aims at investigating technology development, technology progress, technology trends, and also competitive analysis. Fig. 3-23 gives an example for the Patent Analysis. The amount of invention, level of invention, profitability, and customer benefits are calculated and estimated through counting patents and analyzing their contents. Furthermore, the development course of airbags is derived.

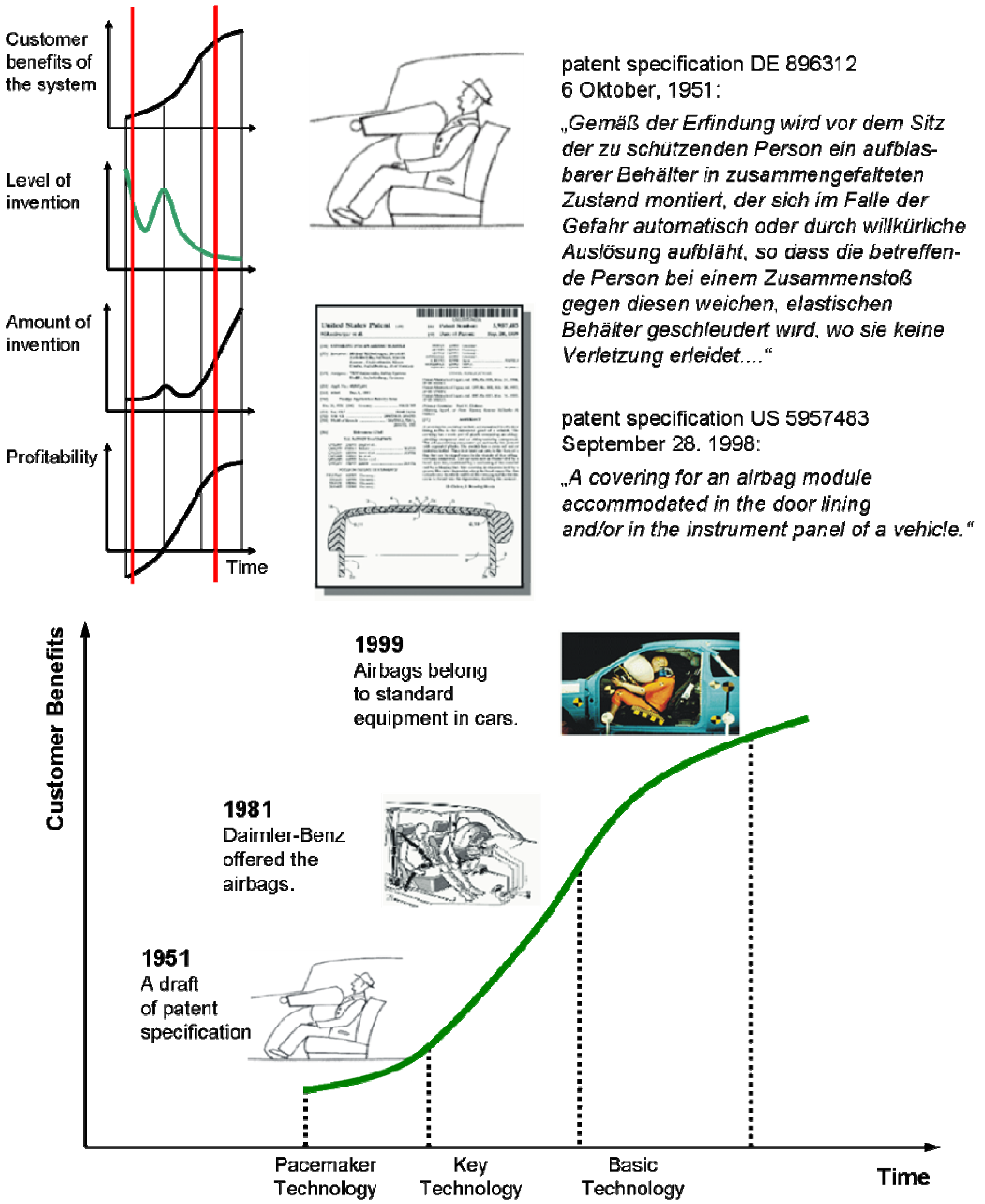


Fig. 3-23: Patent Analysis depicting development of airbags [Lin01]

The advantages of Patent Analysis are: patents are easily computer readable, because they are documented in a fixed structure; patent documents contain over 80% information, which can not be found in other technical literature [RW89]; patent movements of competitors are analyzable; central patent databases are available, also online, e.g. http://ep.espacenet.com (European Patent Office).

However, there are also some shortcomings of Patent Analysis regarding the current situation. Firstly, the international patent databases are no longer uniform. There are patent databases in different regions, different languages, within different domains, etc. It is complicated to collect and process the patent information for the purpose of investigating its comprehensive development.

Secondly, many companies do not apply for patents any more because of the always shorter product life-cycle, and the phenomenon of patent infringement especially in Asia. Thirdly, the time gap between application and approval of patents is too long. Market change can not wait. And furthermore, the patent information seldom describes market development, investments, or other economic data.

3.6.4 Application of Bibliometric Analysis

Originally, Bibliometric Analysis was used in the library management such as cataloging and classification. Nowadays, the Bibliometric methods have been attached importance by more and more researchers. The applications of Bibliometrics are also extended to broad fields. The following texts describe main aspects or target fields, which the Bibliometric Analysis aims at, except library management:

Assessment of research activities at individual, organizational, national or international level

Generally speaking, Bibliometrics is based on analysis of scientific articles, which is not the only one but still a very important symbol of scientific research. The research performances including investigation mainstreams, intensity, research enthusiasm, etc. are investigated through statistical analysis of publication numbers. According to various range of data sources, research outputs at different levels, e.g. in a given country or institute, are investigated. Furthermore, the influence of a single writer, or work, or institute, etc. is determined by ranking their publication amount. To this point, the descriptive indicators like absolute publication numbers, relative publication numbers, absolute citations, and others that derived from one-dimensional Bibliometric methods are used.

At present, a familiar instance about the topic of research evaluation is public policy with Bibliometrics. Here the institutional, regional, and especially national structures of science and technologies are measured through statistical analysis of scientific outputs. The comparison presents the independent position, for example among different countries.

Relationships between authors, organizations, or nations

Statistical analysis of co-occurrence of two items (e.g. publications, authors, or keywords, etc.) measures the similarity, and hence reflects the co-relationships of those two items. Co-authorships reveal the cooperation at different levels. Commonly used method is Co-publication Analysis.

Bibliometrics for scientific disciplines

Co-citations help to classify publications in different areas in order to explore the development of scientific disciplines. The investigation in scientific disciplines is one of the most popular topics related to Bibliometrics, which is interesting to researchers. The dynamic interdisciplinary changes give direction for future research movements. Moreover, ambiguous joint borderland of two or more disciplines implies the potential innovation in interdisciplinary research.

Information mining

As pointed out previously, Co-word Analysis is devoted to content analysis. According to the assumptive rule in the information mining, i.e. the words mentioned frequently in articles are recapitulative for the contents, which should be mined out as relevant information. Logically, the words co-appeared frequently with dominant keywords are representative for the main content, and should be extracted. Co-word Analysis can be applied for information mining.

3.7 Call for Action

The review of those existing methods that are used for information procurement shows there is a dramatically increased demand in that research field in recent years. The information procurement meets not only the researchers' interest, but also the interest of decision makers in industries and companies. As explained in chapter 2, technology plays a vital role for decision making. Decision makers starve for relevant information about technologies in order to develop proper strategies, carry out technology planning, etc. Therefore, they need an automatic, effective, standard, and reusable process, which can guide them to extract desired information about technologies in daily business. To that point, it is necessary to evaluate the methods reviewed in this chapter.

In section 2-4, five requirements for the information procurement are determined based on the decision makers' position. The methods reviewed are compared with each other and evaluated by measuring the fulfillments of the five requirements. The results of evaluation are summarized in Fig. 3-24.

Question: To which extent do the methods reviewed fulfill the requirements on a methodology for the indentification of Technology Indicators?

Requirements are:

- ○ = not fulfilled
- ◔ = partly fulfilled
- ◕ = mostly fulfilled
- ● = fully fulfilled

Methods reviewed	R1: processing of large amount of information	R2: automatic process	R3: high accuracy of the results	R4: standardized and general process	R5: simplified update process
Expert Concultation	○	○	◕	◔	◔
Information Retrieval	◕	◔	◔	◔	◕
Artificial Intelligence	◔	◕	◔	◔	◔
Mining Technique	◕	◔	◔	◔	◔
Ontology	◔	◔	◕	◔	◕
Bibliometric Analysis	◕	◕	◔	◕	◔

Fig. 3-24: Evaluation of the reviewed methods for information procurement

It is concluded from Fig. 3-24, that every method has its own advantages and disadvantages. There is no single method that absolutely fulfills all requirements. Information Retrieval is good at automatic search for relevant documents from a large amount of information, but is totally unfeasible for automatic analysis of information. Although Bibliometrics meets most of the requirements very well, it still needs supports in some aspects. The evaluation inspires that one single method does not solve all problems of information procurements perfectly. However, each method satisfies partial requirements. It is worth selecting the proper existing methods, and combining them in a way that all shortcomings are eliminated and all requirements are the biggest extend satisfied.

Aiming at that, the requirements that are related to solving new problems caused by modern information development are firstly considered (namely R1 and R2): the methods should be able to process large amount of information automatically. Four methods meet the requirements: Information Retrieval, Artificial Intelligence, Mining Approach, and Bibliometrics. Among those methods, Artificial Intelligence and Mining Approaches are left out in the first round of selection. The reasons are as follows: The development of Artificial Intelligence is still in infancy. The status of its development strongly influences the accuracy of information procurement. Data Mining is only available for analyzing structured data, not applicable to unstructured data, e.g. texts. Text Mining is suitable for analyzing unstructured data; but the extracted information patterns are not correlated with each other, which makes the interpretation difficult. Concerning that, the Co-word Analysis is a perfect match, because it is able to analyze the content of unstructured data, and to visualize the extracted information patterns with their relations, e.g. in a Knowledge Map. Another Bibliometric Approach, Publication Analysis, offers a holistic overview of the development intensity of technology by counting publications over time and regions. Publication Analysis is an excellent complement of Co-word Analysis. Other Bibliometric Approaches, like Citation, or Co-citation Analysis, have high demands on management of citations, i.e. collecting, structuring, and transforming of citations. They are too complicated for decision makers to use in their daily business. Therefore, Citation Analysis and Co-citation Analysis are removed from the group of methods candidates.

However, Publication Analysis and Co-word Analysis analyze only the given documents. That means a method that searches for information is still needed. In this dissertation, we focus on Information Retrieval. With the help of Information Retrieval, documents relevant to target technology are retrieved and are ready to be analyzed by Bibliometrics.

Furthermore, with Co-word Analysis, information pattern, namely keywords, are positioned in a Knowledge Map. To simplify the interpretation of the Knowledge Map without misunderstandings, ontology is the best assistant. Ontology supports also the extraction of keywords in the given domain, and the standardization of the keywords, etc.

Till here, the methods selected search and analyze the information only from quantitative perspective. As mentioned in section 2.4, qualitative evaluation is also indispensable. Therefore, another method - Expert Consultation is used to ensure the extracted information from qualitative perspective.

To sum up, four basic methods are selected from the existing methods of information procurement. They are integrated in a comprehensive approach to achieve the goal of extracting relevant technology information and to satisfy the requirements of decision makers. Detailed explanation is in chapter 4.

4 Methodology for the Identification of Technology Indicators

The methodology presented in this chapter aims at the automatic identification of Technology Indicators from a large amount of information. In this chapter, the methodical foundation is firstly presented (section 4.1) and then the process model of the methodology is demonstrated (section 4.2). Subsequently, the phases and milestones of the process model are explained in detail (section 4.3 – 4.7). The chapter ends up with the introduction of the integration of the methodology with the innovative Technology Database (section 4.8).

4.1 Foundation of the Methodology

As discussed in section 3.8, the following four methods were selected as the suitable basic components for the methodology for the identification of Technology Indicators: Information Retrieval, Bibliometric Analysis (Publication Analysis and Co-word Analysis), Ontology and Expert Consultation. The methods interact with each other and hence constitute the whole methodology.

A key point of using Co-word Analysis is to interpret the Knowledge Map correctly. In section 4.1.2, a guide for the interpretation of Knowledge Map is proposed.

The central task of the methodology is to identify Technology Indicators from the information collection. In order to automate the identification, a Technology-Indicator-Ontology (TI-Ontology) is built up based on the experience of case studies (see section 4.1.3). The TI-Ontology together with the four basic methods and the guide to the interpretation of the Knowledge Map compose the methodical foundation of the methodology.

4.1.1 Basic Methods

The methodology proposed for the identification of Technology Indicators is based on four basic methods. All four methods were introduced previously in chapter 3. In this section, the functions of the four methods in the process of identification of Technology Indicators are explained.

Information Retrieval (IR)

IR is used to automatically search for raw data sets that are relevant to a given subject. With the help of IR, the desired information can be efficiently searched and automatically separated into relevant and irrelevant documents. The documents relevant to the examined topics are retrieved. [BR99]

Bibliometric Analysis

Only two Bibliometric methods are used in the methodology proposed in this dissertation. One is the traditional Publication Analysis; the other is Co-word Analysis (Fig. 4-1). The two methods help to draw conclusions about the development status and trend of technologies by analyzing empirical data. Publication Analysis deals with the calculation of absolute and relative publication numbers; while Co-word Analysis takes contents of publications into consideration. Identification and concretization of Technology Indicators by using Publication Analysis and Co-word Analysis are highlights of the methodology. The information is broken down into a network of keywords. Noisy information, which does not make sense, or disturbs the analysis, is filtered out. The condensed data are reconstructed within semantic context into valuable knowledge. Publication Analysis and Co-word Analysis play important roles in condensing, processing, and analyzing raw data sets from quantitative aspect.

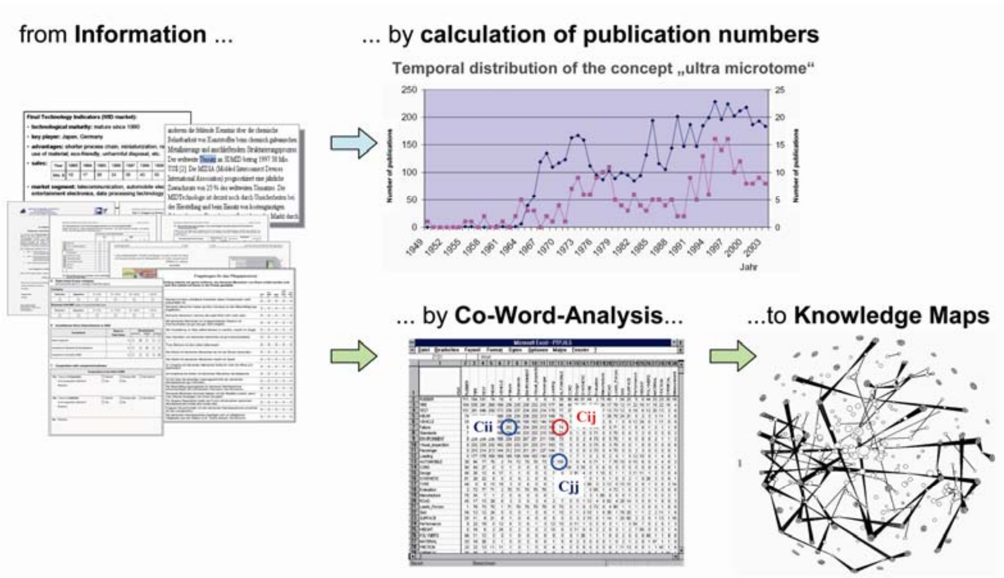


Fig. 4-1: Bibliometric Methods used in the methodology: Publication Analysis and Co-word Analysis

The refined keywords are located in a two-dimensional Knowledge Map according to the results of Co-word Analysis. Compared with two-dimensional tables and one-dimensional word list, the Knowledge Map visualizes not only the essential contents of the articles but also their relationships, which facilitates the knowledge extraction.

Ontology

As mentioned in section 3.5, ontology is contributed to share common understanding of domains. It facilitates human–machine communication. In the methodology for the identification of Technology Indicators, ontologies offer domain-specific semantic context and therefore support the interpretation of

Knowledge Map. Furthermore, a general ontology for Technology Indicators is explicated in section 4.1.3, which helps to identify the Technology Indicators.

Expert Consultation

Expert Consultation is used in a small group of experts to seek their professional but subjective opinions. It evaluates the Technology Indicators from the qualitative aspect.

Interaction of the four Methods

According to FAYYAD, the typical process of Knowledge Discovery in Databases (KDD) is arranged into the following stream of steps as shown in Fig. 4-2: selection of data; pre-processing of data; transformation of data into forms appropriate for the mining procedure; extraction of potentially useful patterns (Data Mining); interpretation and visualization of mining procedure [FPS96].

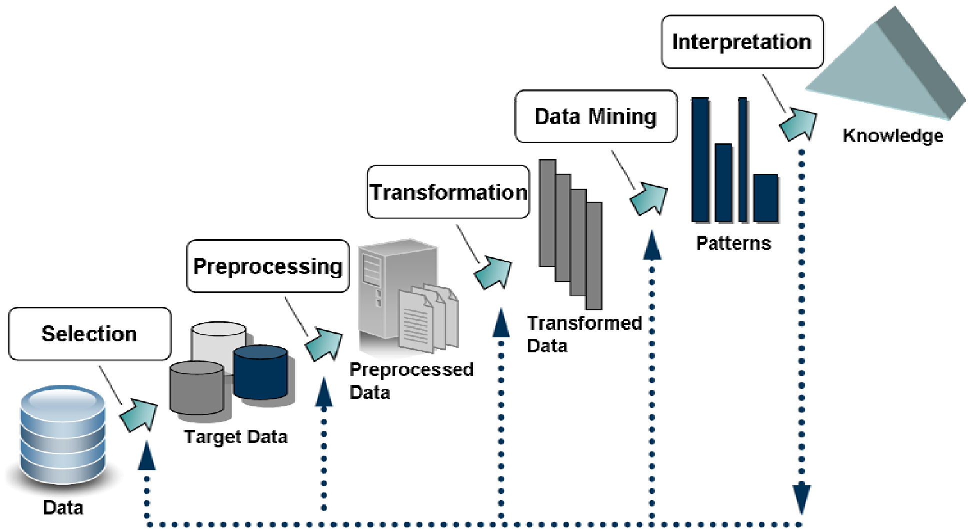


Fig. 4-2: Process of Knowledge Discovery, according to FAYYAD

In Fig. 4-2, the process leading from raw data to knowledge in the methodology for the identification of Technology Indicators is defined as the following four steps (Fig. 4-3):

- Search: the suitable sources of searching (e.g. databases) are selected and then the relevant information is collected.
- Mining process: the mining process involves transforming, cleaning, pre-processing, structuring, statistically analyzing, and extracting of data.
- Interpretation: using visualization and other techniques to help users understand, interpret the Data Mining results. (knowledge extraction)

- Evaluation: in this step, it is necessary to confirm, modify, and supplement the extracted knowledge from the qualitative aspect.

Each method used in the methodology plays its own role. They offer complementary advantages and thus work smoothly with each other. Fig. 4-3 gives an overview of the integration of methods into KDD process within the framework of the methodology proposed in this dissertation.

	Knowledge Discovery Process			
Methods	Search	Mining Process	Interpretation	Evaluation
Information Retrieval	✓			
Bibliometric Analysis		✓		
Ontology		✓	✓	
Expert Consultation				✓

Fig. 4-3: Overview of the functions of the methods in Knowledge Discovery Process of the methodology presented in this dissertation

4.1.2 Guide to the Interpretation of Knowledge Map

As mentioned previously, keywords can be positioned in a spatial coordinate system according to their text frequencies and co-occurrences. With help of MDS, the keywords are located in a two-dimensional Knowledge Map. Fig. 4-4 shows an example of a typical Knowledge Map.

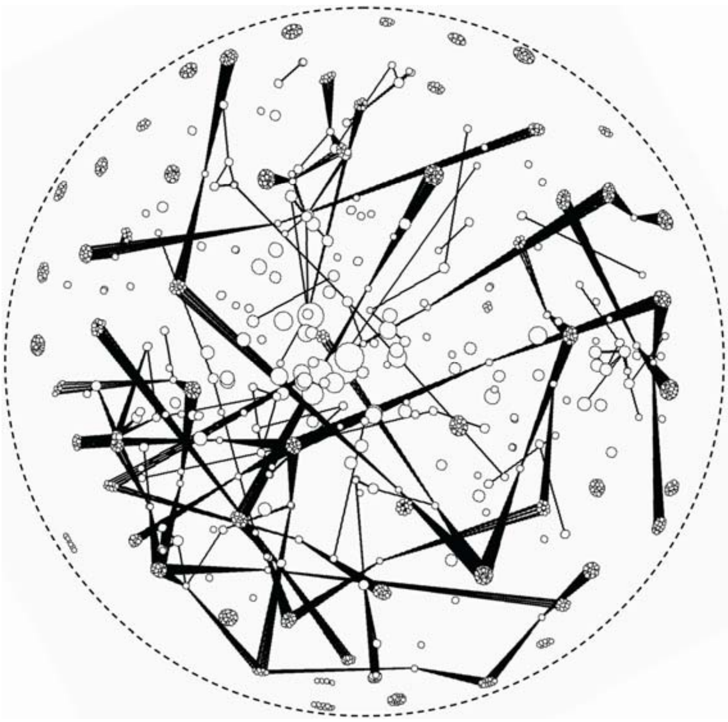


Fig. 4-4: Example for a Knowledge Map

The contents of a text corpus are visualized by means of showing co-related keywords within a Knowledge Map instead of using a hierarchical (one-dimensional) keyword list. The Knowledge Map makes it possible to see different thematic clusters of the contents at a glance. However, without a guideline, it is difficult to understand the Knowledge Map. So it is important to compile a guide about how to interpret a Knowledge Map. In 4.1.2.1 and 4.1.2.2, the guide to the Interpretation of Knowledge Maps is proposed and explained in detail.

4.1.2.1 Basic Instruction

There are some basic understandings of the patterns shown in Knowledge Maps. By the aid of Fig. 4-5, the basic rules are explained as follows:

- A **circle** represents a keyword. For instance, the first circle on the left side stands for the word "assembling"; the circle at the top right corner represents the word "marketing".

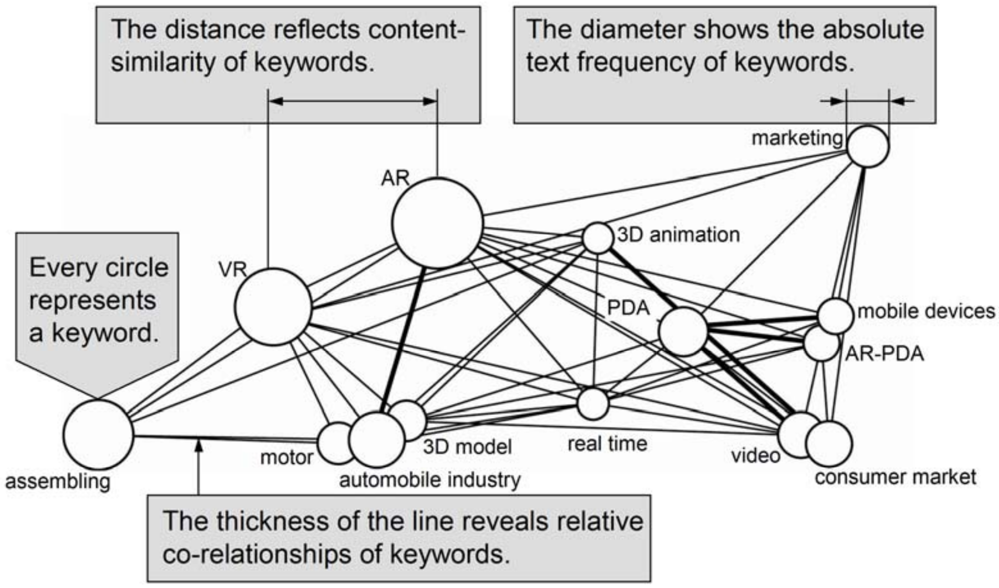


Fig. 4-5: Basic legends of Knowledge Map

- The **diameter** of a circle shows the absolute text frequency of that keyword, which means to show how often the keyword appears in the whole text corpus. The bigger the circle is; the more often it appears in the text corpus. As shown in Fig. 4-5, the biggest circle is of the word "AR", i.e. AR is most frequently mentioned in the texts. On the right side of AR, there is a word "3D animation", whose circle is much smaller than AR's. That means "3D animation" seldom appears in the text corpus.

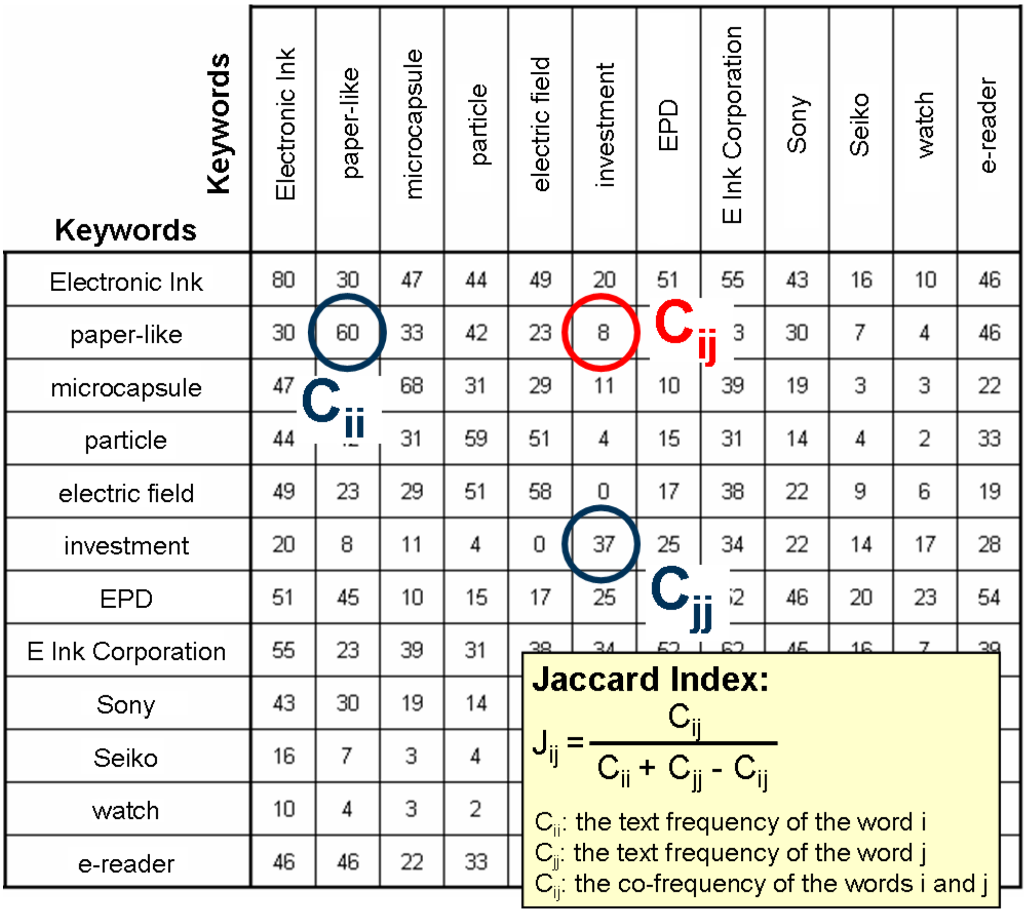
- The **distance** of two circles reflects the similarity of those two keywords. The nearer the circles are, the more content-similar the keywords represented by the circles are. The distances are based on the absolute co-occurrence of keywords, which are calculated by Co-word Analysis. The hypothesis is, the more often the words appear together in documents; the more content-similar they are. For example, the keywords "motor" and "automobile industry" are closely located together. It is estimated that "motor" and "automobile industry" have a very close relationship. They can be clustered together.

- The **thickness** of the lines reveals relative co-relationships of keywords, i.e. the thicker the line is; the more inevitable the keywords appear together. The thickness is calculated by Jaccard Index. The Jaccard Index, also known as Jaccard Similarity Coefficient, is used to measure the similarity of sample sets. It is defined as the size of the intersection divided by the size of the union of the sample sets (see the formula 4.1).

  $$J(A,B) = \frac{|A \cap B|}{|A \cup B|} \qquad (4.1)$$

  In Co-word Analysis, the Jaccard Index is defined as the co-occurrence of two keywords divided by the union of those two keywords (Fig. 4-6). C_{ii} stands for the total text frequency of the word i. Similarly C_{jj} means the total text frequency of the word j. And C_{ij} represents the co-occurrence of the words i and j. The Jaccard Index J_{ij} ranges from 0 to 1. The closer J_{ij} is to 1, the thicker is the line between the words i and j. On the contrary, if J_{ij} is equal to 0, there is no connection between i and j. For instance, there is a thick line between "3D animation" and "PDA" in Fig. 4-5. It can be estimated that, if "3D animation" appears in the document, it always appears together with "PDA". The relationship between them is very strong.

The four rules mentioned above help us to basically understand the Knowledge Maps. How to read a Knowledge Map in detail? What should be firstly considered? Are the bigger circles more meaningful than the small ones? What is the difference between circles in the middle and circles at the edge? To answer those questions, a concrete guide to interpret the Knowledge Map is required. In the next section, the guide will be given step by step.



Keywords	Electronic Ink	paper-like	microcapsule	particle	electric field	investment	EPD	E Ink Corporation	Sony	Seiko	watch	e-reader
Electronic Ink	80	30	47	44	49	20	51	55	43	16	10	46
paper-like	30	60	33	42	23	8	[illegible]	[illegible]	30	7	4	46
microcapsule	47	[illegible]	68	31	29	11	10	39	19	3	3	22
particle	44	[illegible]	31	59	51	4	15	31	14	4	2	33
electric field	49	23	29	51	58	0	17	38	22	9	6	19
investment	20	8	11	4	0	37	25	34	22	14	17	28
EPD	51	45	10	15	17	25	[illegible]	[illegible]	46	20	23	54
E Ink Corporation	55	23	39	31	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Sony	43	30	19	14	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Seiko	16	7	3	4	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
watch	10	4	3	2	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
e-reader	46	46	22	33	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

Fig. 4-6: Jaccard Index in the context of Co-word Analysis

4.1.2.2 General Steps

1. **Main topics**

   Most of the bigger circles are placed in the middle. They represent the keywords appearing most frequently. It is estimated that the keywords, which appear most frequently reflect the main topics of the text corpus. For instance, if the text corpus deals with a certain technology, the bigger circles always show the name, the main characters of the technology, the important applications using the technology, and the basic application fields [Gau98].

2. **Cluster-oriented observation**

   The keywords can be clustered according to organizations, regions, authors, topics, and other criteria. Each cluster is labeled according to the keywords appearing most frequently in the cluster. Different clusters can be marked with different colors. The Knowledge Map makes the clustering of keywords, i.e. the content of text corpus, more transparent and clearly visible. Then the boundary and the overlapping keywords of two clusters are analyzed in order to estimate the relationship between those two clusters. In some cases, the boundary and the overlapped keywords indicate the potential of cooperation.

As shown in Fig. 4-7, information base is the publications of the year 2003 in workgroup Gausemeier at the Heinz Nixdorf Institute. The keywords are clustered according to teams. Grayish circles stand for the keywords from IM team; words in dark gray circles belong to VR team; and the black ones are from EIS team. The main topics of IM team are represented by big circles from number one to number six. They are “innovation management”, “strategic product planning”, “forecasting”, “self-optimizing system”, “mechatronics”, and “methodology”. Similarly, the main topics of VU team are labeled by the most frequently appeared keywords: “virtual reality”, “augmented reality”, and “tracking system”. EIS team is engaged in “product data management”. The overlapped keyword between VU team and EIS team is “simulation of material flow”. It is concluded that VU team and EIS team both need the topic simulation of material flow. They can cooperate in that field.

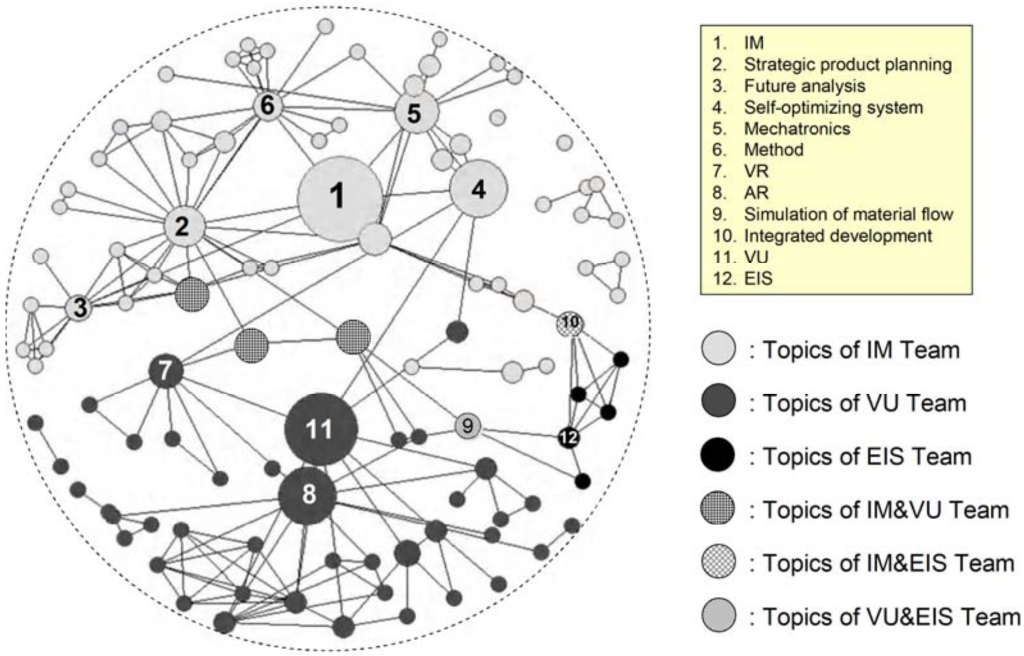


Fig. 4-7: Cluster-oriented observation of keywords in Knowledge Map (Information base is the publications of the year 2003 in workgroup Gausemeier at the Heinz Nixdorf Institute)

3. **Micro-analysis**

It is necessary to observe targeted keywords for the purpose of extracting knowledge in detail from Knowledge Map. As mentioned previously, the thickness of lines in Knowledge Map is calculated according to the Jaccard Index, which measures the relative similarity of keywords. The distance between circles on Knowledge Map represents co-occurrence of the keywords, which measures their content-similarity. Therefore, not only the target keyword should be the focus, but also its strongly co-related keywords, i.e. the keywords with the thickest or shortest lines to the target keyword. The relationships between those keywords should be interpreted in a logical way. A domain-specific ontology is needed to avoid the misunderstanding of the keywords. The target keyword is thus characterized and described by interpreting the most frequently co-occurred keywords and their relationships.

4. **Comparison of different time periods**

   Knowledge Map also helps to trace changes during different time periods. Based on the whole period, the keywords are analyzed for the purpose of grasping the overall status in the whole period. Then the whole period is divided into several small time periods (a time period can be one month, one year or even 10 years, according to research demand.). Keywords in every small time period are separately visualized in periodical Knowledge Maps. Comparing those periodical Knowledge Maps, the following questions can be answered:

   - Which keywords appear newly in which time period? What are the connections with the other words?
   - Which keywords disappear in which time period?
   - Whose diameter has grown up or reduced? When?
   - Which linkages are becoming thicker or thinner? When?

   By comparing those periodical Knowledge Maps, the dynamic changes during those periods, e.g. outdated topics, new topics, trend of research activities, etc. are identified.

5. **Analysis of marginal keywords**

   Marginal keywords are those keywords near the circumference of the Knowledge Map. Most of them are small circles, and do not have a lot of connections to the other keywords. It means that marginal keywords seldom appear in the text corpus. If they appear recently, they indicate emerging research directions. If the marginal keywords appeared in old texts, they indicate potential research directions, which might be referred long time ago, but not paid great attention by researchers. By analyzing marginal keywords, we can draw conclusions about research vacancy, potential research directions, new research trend, etc.

4.1.3 The Ontology of Technology Indicators

The methodology aims at the extraction of technology-relevant knowledge from a collection of raw information. The concept “**Technology Indicators”** is proposed in this dissertation, which represent the essential knowledge of technologies.

Technology Indicators are those indices or statistical data, which allow direct characterization and evaluation of technologies throughout their whole life cycles. For example,

- **Technology maturity:** Has this technology been used for series production? Is it a pacemaker technology, key technology or basic technology? Is it just a prototype? Or is it already obsolete?
- **Market segment:** In which fields is this technology applicable?
- **Key player:** Which country, company or expert is the most active in this technological field?

These Technology Indicators describe technologies concisely. Combinations of Technology Indicators reveal technological and market development. With the help of Technology Indicators, the decision makers can easily master the features of the target technology and hence compare it with other technologies. The process of decision making is speeded up and the quality of decision-making is improved.

After three case studies, it is noticed that the identified Technology Indicators have similar titles with different contents and values. It proves that there is a list of Technology Indicators generally valid for all technologies. Enlightened by that, an ontology of Technology Indicators is built up for the purpose of simplifying and automating the process of identification of Technology Indicators. The specific ontology is named “Technology-Indicator-Ontology” (TI-Ontology), which is built according to the following steps.

First of all, all the available information about Technology Indicators is summarized, based on experiences of several case studies. It is like a computer-aided brainstorming: all the Technology Indicators are enumerated. Then the list of Technology Indicators is mended, explored, and structured by a small group of experts. As demonstrated in Fig. 4-8, which is a strongly simplified segment of the whole TI-Ontology, the main elements for TI-Ontology comprise concept, relation, and instance. Concepts are equal to Technology Indicators. In this context, instances stand for the contents, and values of Technology Indicators.

The overall TI-Ontology is divided into two parts. One of them is called technological development; and the other part is market development. Under the technological and market development, there are several sub-Technology Indicators. The hierarchical structure extends downward until there are no sub-indicators. For example, the Technology Indicator "key player" on the right side of Fig. 4-8 is on the one side a sub-indicator of "market indicators"; while on the other side, it is also super-class of "supplier", "customer", and "expert". Instances for the lowest class "Supplier" are e.g. Firm X and Firm Y.

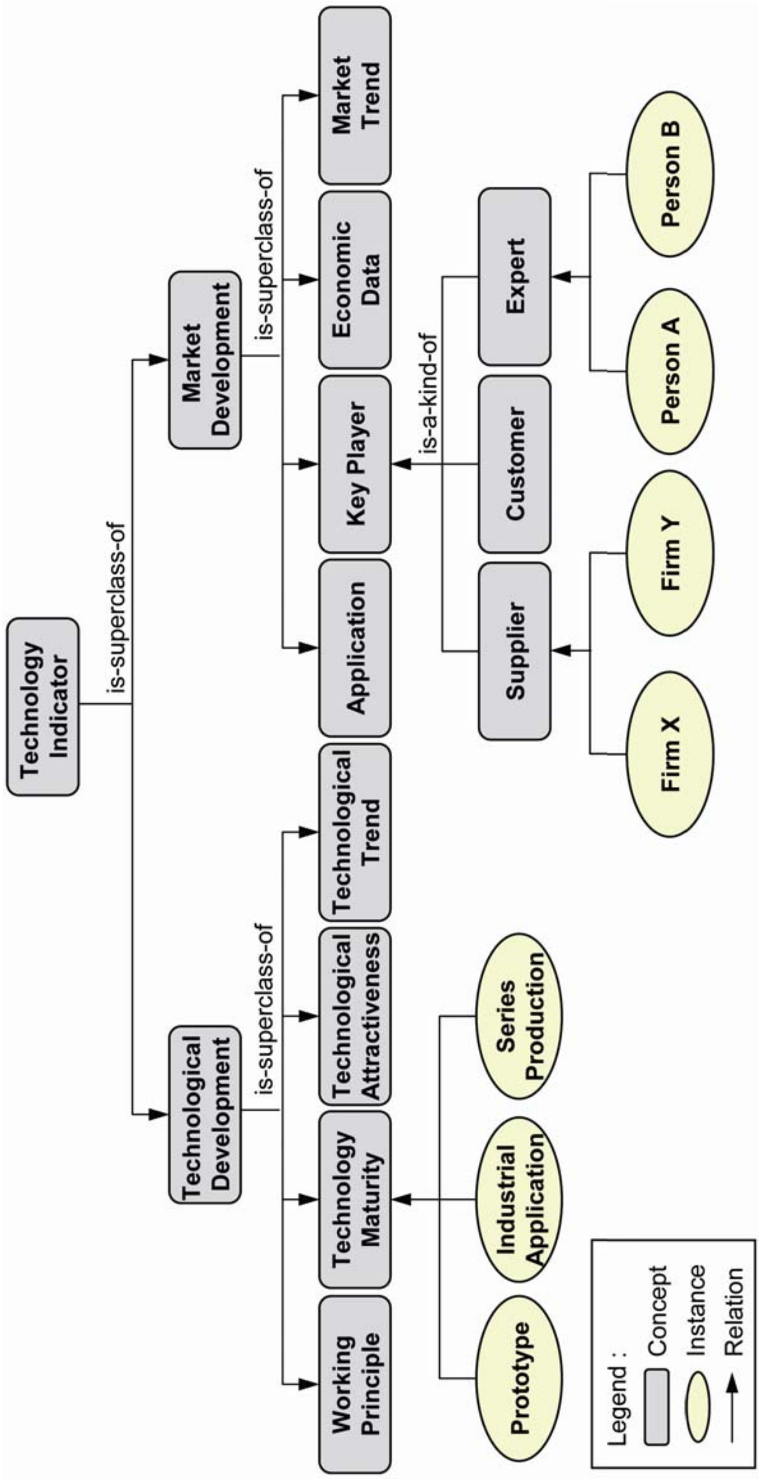


Fig. 4-8: Simplified overview of the TI-Ontology

Fig. 4-8 only shows the principle concept networking of TI-Ontology. It should be conceptualized with the definitions of Technology Indicators (concepts), synonyms, instances, and so on (Fig. 4-9).

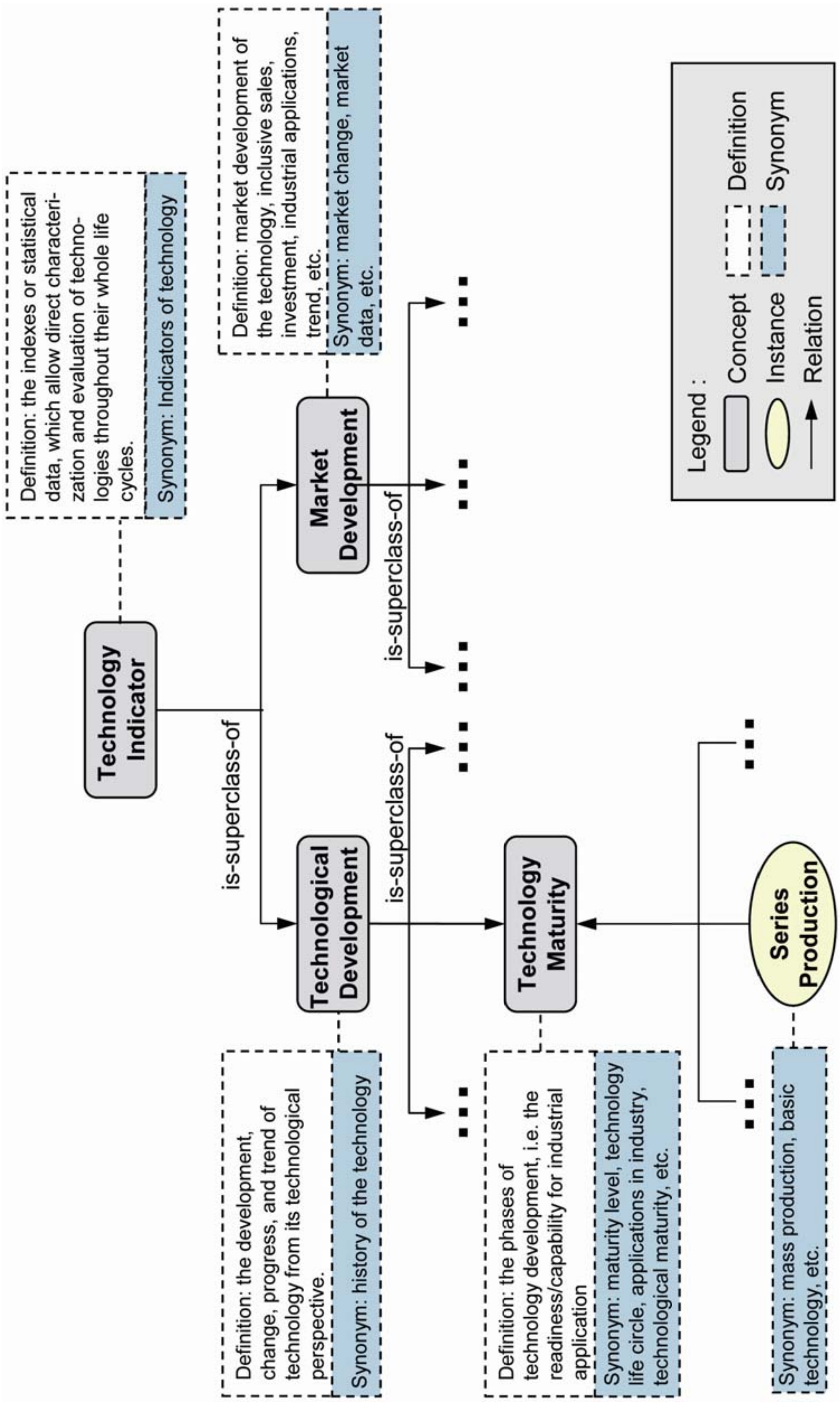


Fig. 4-9: Segment of the TI-Ontology with definitions and synonyms

Definitions help to restrict the meaning of the concept in the domain of Technology Indicators in order to avoid misunderstandings. For instance, “maturity level” has different meanings in different language environments. It can be a factor to evaluate wines; it can also characterize the process capability of an organization in the Capability Maturity Model[1] [SR00]. In order to distinguish the concept in the field of Technology Indicators from other fields, “technology maturity” is defined as the phases of technology development in the TI-Ontology. It means the readiness/capability for industrial application in the context of technology development. Normally, there are three technology maturity levels, i.e. prototype, industrial application, and series production. The definition ensures that users fully understand “technology maturity” related to Technology Indicators. Each concept has one or more synonyms, which are equivalent terms to concepts. A concept and its synonyms can be exchanged without changing the concept's meaning. Synonyms play an important role in ontologies because they increase the probabilities of identifying the concepts by checking various styles of expression, for example, “technology maturity” can be expressed as “level of maturity”, “degree of maturity” or even “degree of ripeness”. A concept (raw Technology Indicator) with its instances (contents and values of Technology Indicator) constitutes a complete Technology Indicator.

As shown in Fig. 4-9, the TI-Ontology is filled with details and available for the employment in the methodology. In accordance with the distinct characteristic of the ontology, the TI-Ontology is unambiguous, reusable, and optimizable. It supports the process of identifying Technology Indicators by offering a clear definition of the Technology Indicators, their synonyms and relations (which one is sub-indicator or which one belongs to the superclass). The TI-Ontology can be repeatedly used, because it is generally valid for all technologies. It conceptualizes the Technology Indicators and hence simplifies their identification process. A lot of time is saved by using TI-Ontology instead of manual work. Furthermore, the TI-Ontology can be changed by adding new important concepts (Technology Indicators), eliminating the timeworn concepts, optimizing the definition, or completing synonyms.

---

1 Capability Maturity Model (CMM) is a model to help software organizations improve the maturity of their software processes in terms of an evolutionary path from ad hoc, chaotic processes to mature, disciplined software processes. [SEI-ol]

4.2 General View of the Methodology

Based on the methodical foundation introduced above, the methodology for the identification of Technology Indicators is developed. The iterative process model of the methodology is divided into five phases. As illustrated in Fig. 4-10, the phases and milestones are demonstrated on the left side of the figure. The tasks and the methods used in every phases are listed in the middle. The results for every phase are shown on the right side.

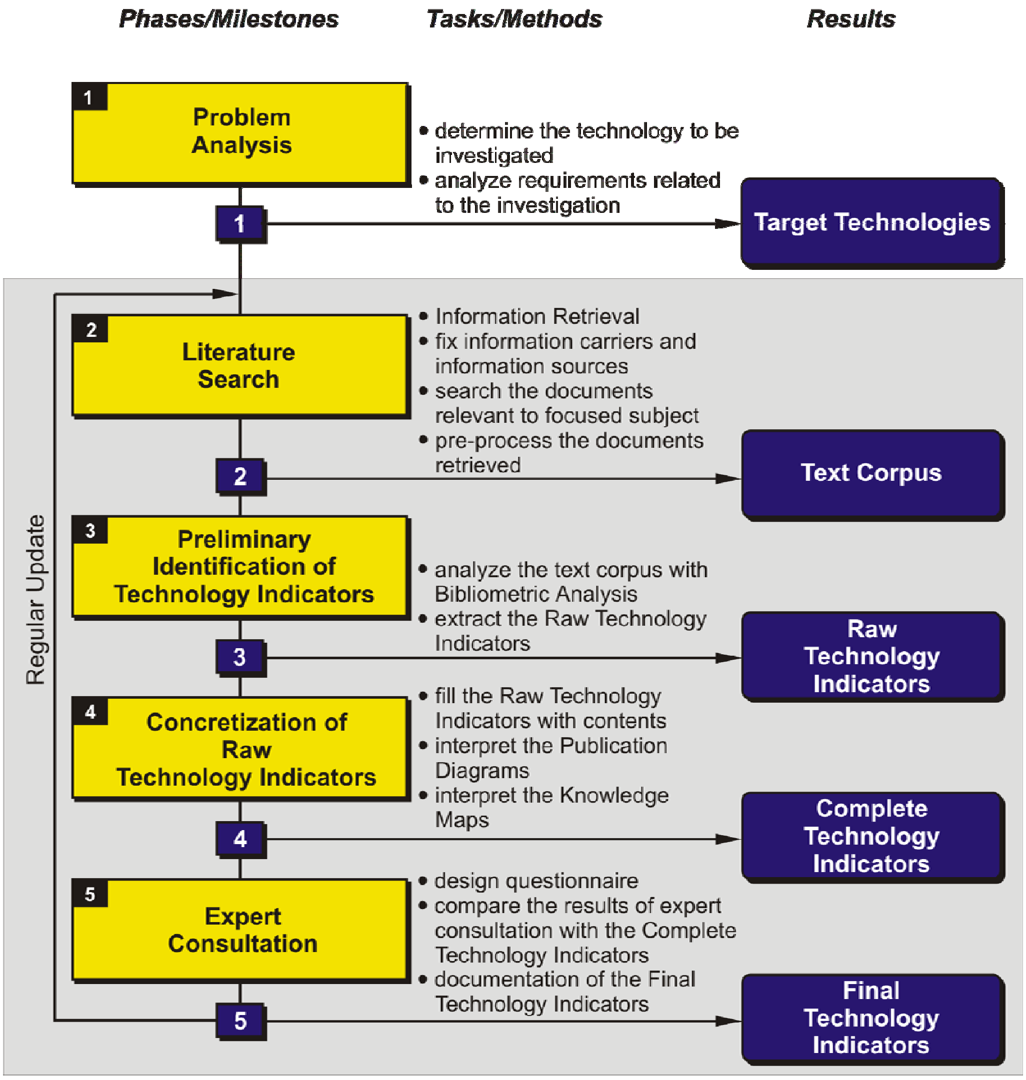


Fig. 4-10: Process model of the methodology for the identification of Technology Indicators

Phase 1: Problem Analysis

The first step is to determine research objectives, i.e. to answer the question "Who wants to know what in which areas?" The results of phase 1 are a list of investigation requirements and the target technologies, which will be investigated one by one in the follows phases (section 4.3).

Phase 2: Literature Search

The second phase aims at thematically searching for literature relevant to target technologies. The method used for literature search is Information Retrieval. Aiming at different research directions, it is necessary to select appropriate information carriers (e.g. books) and information sources (e.g. databases). Then the relevant documents are retrieved and pre-processed. The detailed information is introduced in section 4.4. The pre-processed documents are saved in a text corpus.

Phase 3: Preliminary Identification of Technology Indicators

The goal of the third phase is to automatically identify raw Technology Indicators from the text corpus collected in phase 2. Bibliometric Analysis is used. As discussed in section 4.1.1, the chosen Bibliometric methods are Publication Analysis and Co-word Analysis. Firstly, the number of publications is statistically analyzed and visualized in diagrams. From each diagram, a corresponding Technology Indicator is identified. Then, the content of the text corpus is broken down into words. The words are cleaned, standardized, statistically calculated, and connected (section 4.5). The contents of retrieved documents are condensed into a network of keywords. The keywords are visualized in a Knowledge Map. With the aid of the TI-Ontology introduced in section 4.1.3, the potential Technology Indicators are extracted. The result of this phase is a list of Technology Indicators with names and definitions, which are defined as the Raw Technology Indicators.

It is to be noted that an ontology of the investigated technological field can be built up, based on the network of keywords mentioned above. That technology-specific ontology can support analyses and interpretation in the following phases.

Phase 4: Concretization of Raw Technology Indicators

In the fourth phase, it is to fill raw Technology Indicators with contents and to assign values to them. The main tasks here are the interpretation of publication diagrams and Knowledge Maps generated in Phase 3. Concrete procedures are explained in section 4.6.

Documenting all the interpretation as values assigned to Raw Technology Indicators, the results at the end of this step are the Complete Technology Indicators with names, definitions and values.

Phase 5: Expert Consultation

All the analyses above are based on statistics. In Phase 5, it is necessary to ask the experts' opinion from qualitative perspective. Within the Expert Consultation, the definitions, values etc. of Technology Indicators are evaluated and supplemented by experts. After integrating the results of qualitative and quantitative analyses, the final Technology Indicators are identified and documented (section 4.7).

Remarkably, technology is changing fast and the related information updates quickly. Decision makers always need the firsthand information to keep agile reaction to sudden change of technologies. It is indispensable to update the information in the Technology Database regularly. The process model of the methodology is iterative.

4.3 Phase of Problem Analysis

As a key factor, technology influences product development, production processes, and the company's competitiveness. Therefore, successful technology planning is indispensable for keeping companies in the leading position. In the process of technology planning, an all-around comprehension of the technological and market development plays an important role. So does the accurate prediction of the future trend about technologies. On most occasions, there are more than one candidate technologies. In order to select the most beneficial technology from several candidates, the technologies should be compared and evaluated. As explained in section 4.1.3, Technology Indicators characterize the performance of technologies and indicate future trends. The comparison of Technology Indicators contributes to the evaluation of similar technologies. In the case of investigating a given technology or comparing several technologies, the methodology for the identification of Technology Indicators presented in this dissertation comes into play.

The starting point of the methodology is to ensure the investigation objectives concerning the performance of technology planning. If there is a given technology to be investigated, it is important to determine who wants to be informed about which aspect. If there is no given technology, the following questions should be replied to support the determination of target technologies:

- Where do problems exist, in product development or production processes?
- Which technologies can solve those problems?
- Who (which department) is the decision maker (user of the methodology), e.g. research center, strategy team?
- Which kind of decision will be made by the decision maker? For example, is it a buy-in or self-develop decision? Or is it a selection decision?
- Are there any emphases of investigation, e.g. current technology maturity, prediction of market trend, or in general?

- At which time period do decision makers aim at?

After answering the questions, investigating objectives are determined as the result of Phase 1. Technologies to be investigated are targeted. Decision makers and their requirements are figured out. Focuses of the investigation are defined. It is noticed that the following phases are able to identify Technology Indicators only for one technology at a time. Therefore, the target technologies have to go through the process model one by one.

4.4 Phase of Literature Search

Similar to other normal knowledge discovery processes, an information base is necessary. The aim of this phase is to search for information relevant to target technologies in order to build up an information base available for statistical analysis. Starting with this phase, the target technologies are investigated one by one as mentioned above.

4.4.1 Limitation of Search Area

The amount of published information has been dramatically increased with the rapid development of information technology since the 1970s. Although the modern information technology such as World Wide Web and database techniques facilitates a rapid and relatively effective process of information search, it is helpful to limit the search area before the information is retrieved. Generally speaking, the information must be digitally available because the methodology demands highly on automatic process. However, there is still a vast amount of digital information, which needs to be pre-filtered by reducing the search areas. Therefore, the following measures are taken to minimize search area of information: determination of the forms of publications and selection of information repositories.

Determination of the Forms of Publications

Information about technologies is structured and represented in publications. According to different writing structures, means of expression, and information media, the publications are classified in various forms (the written forms are only discussed. Videos, animations, and album of figures are not taken into consideration):

- scientific article;
- book;
- press release, advertising article;
- market study, final project report, enterprising financial report;

- user handbook, white book;
- patent;
- monograph, dissertation, habilitation treatise, research application;
- other forms of publications

Publications in different forms focus on different aspects of technology. Scientific articles demonstrate technologies mainly from technological perspective; while in patents, the functionality and work principle are explicitly described. Press releases, advertising articles, and financial reports in companies introduce the advantages of technologies, show their market development, enumerate economic data, etc. in order to propagandize the technologies. More details are in Fig. 4-11:

Legend:
● = full hit
◕ = frequently related
◑ = half hit
◔ = less related
○ = no connection

	work principle	state of the art	technology trend	advantage/ disadvantage	applications	market data	market trend
scientific article	●	◕	◕	◔	◑	◔	◔
book	●	●	◑	◑	◕	◔	◔
press release; advertising article	◔	◑	○	●	◕	◑	◕
market study; final project report; enterprising financial report	◑	◑	◕	◕	●	◕	◕
user handbook; white book	●	◔	○	◔	◕	○	○
patent	●	●	◑	◑	◑	○	○
monograph; habilitation treatise; dissertation; research application	●	●	◑	◕	◕	◔	◔

Fig. 4-11: Overview of emphases on different aspects of technology due to different forms of publication

As discussed in Phase 1, the investigation objective including the emphases of investigation is determined. Consulting Fig. 4-11, the suitable publication forms that are appropriate to investigation emphases are selected. For instance, if the investigation focuses on market development, the press releases, advertising articles, financial reports of companies, and market-oriented studies are most preferred. If the investigation aims at characterization of the technology from its technical aspect, the premier selection is scientific articles, patents, and research applications. The advisable selections of publication forms according to different investigation emphases are summarized in Fig. 4-12.

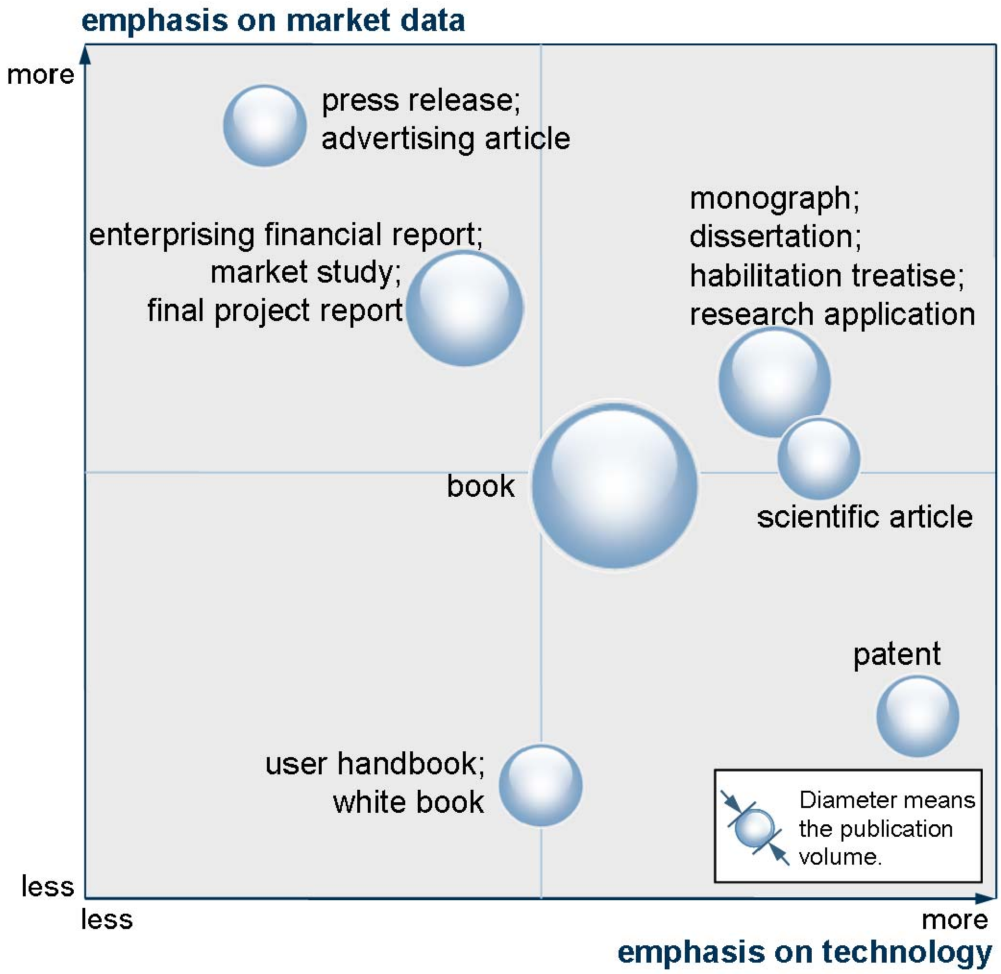


Fig. 4-12: Overview of premier selection of publication forms according to different investigation emphases

Furthermore, the complexities of publication forms are dissimilar. In this context, complexity means the access availability, publication volume, and workload required for pre-processing (discussed in section 4.4.2). Books are the most complex publication form. It is of big volume, normally not available in digital format. Patents need small effort to pre-process because they are written in a default structure. Press releases and advertising articles are commonly short and easy to access. Due to the expenditure of investigation, the search area of information can be limited by selecting publication forms with applicable complexity.

Selection of Information Repository

After the selection of publication forms concerning the investigation objectives, it is time to choose concrete information repositories. Driven by the force of modern information technology, the definition of information repository is no longer limited to physical locations where the public may view files containing current and historical information, technical reports, and reference documents. Information technology has enabled the digital collection, storage and access of a huge amount of information. Therefore, the definition of information repository extends to data warehouses for electronic information, which facilitates interactive communication between human and machinery.

In broad sense, information repositories appear in the forms of libraries, databases, web-based search engines, internet platforms, and so on. One of the important factors for information search is the accessibility of publications. There are three levels of accessibility to digital information in repositories

- **Level 1:** The detailed information of publication including full text is freely accessible for all users.
- **Level 2:** Partial information of publications is freely accessible. Users have to register and pay fees for complete information including full text.
- **Level 3:** Publications are unable to subscribe and hence not in full text available. Only titles, authors and often abstracts are freely accessible.

Furthermore, there are other criteria that facilitate the selection of information repository suitable to investigation objectives, as shown in Table 4-1.

Table 4-1: Classification of information repositories according to different criteria with corresponding examples

Criteria	Description	Examples
language	Which is the preferred language: English, German or Chinese?	• English: ISI Web of Knowledge powered by Thomson Scientific • Chinese: China Knowledge Internet http://www.cnki.net/index.htm
region	Which regions are focused: Asia, Europe or somewhere else?	• Europe: European Database on Logistics http://www.logistik.tu-berlin.de/ • Asia: Asia Pacific Tech Monitor http://www.techmonitor.net/techmon/home.htm
publication form	Information repositories are different according to different publication forms.	• Patent: European Patent Office http://www.epo.org/ • Scientific article: Google Scholar http://scholar.google.de/
domain	Information of different technology domains is saved in corresponding information repositories.	• MID: http://www.3dmid.de/ • Pharmaceutical & Biotechnology: Investigational Drugs database (IDdb) powered by Thomson Current Drugs

Requirements of Investigation listed in Phase 1 lead to restrictions on languages, regions, domains, and publication forms. Consulting Table 4-1, concrete information repositories, i.e. databases, e-journals are selected.

It is to note that every company has its own data collection. The internal database is also a candidate information repository, which should be taken into consideration.

In summary, the two measures discussed above: determination of publication forms and selection of information repositories are effective to restrict the literature search to a small area. They save time and make the search results more accurate.

4.4.2 Search for Publications

Information relevant to target technology is searched for within the reduced search area. The method used here is Information Retrieval (introduced in section 3.2). The chosen search technique is keyword search.

First of all, a group of phrases is defined, which can briefly and concisely describe the target technology. Every phrase is then supplemented by their synonyms and different expressions. Concerning investigation objectives, it is also necessary to translate the phrases into other languages. For example, the target technology is E Ink. According to the basic knowledge of E Ink, it is described in the phrases: electronic ink, e-ink, e ink, and a technology for EPD, technology used in electronic paper display, elektronische Tinte (in Germany), etc. The sequence of those phrases is not important.

Then, the phrases are combined with Boolean Operations: conjunction (AND), disjunction (OR), negation (NOT). Synonyms are connected with "OR" logic. Two phrases that are both indispensable for search process are combined with "AND" logic. And conflict phrases are linked with "NOT" logic with other phrases. The combinations of phrases are used as search queries in the Technology Databases, internet platforms, intranet or other information repositories that are chosen in section 4.4.1. As a result, the documents relevant to the target technology are retrieved.

Although keyword search is a simple way compared with other search techniques, it meets the requirement of Bibliometric Analysis: high integrity of the information base. The more qualifications the search queries have, e.g. with Boolean Logic, the more accurate is the search result. The more complete the search queries are, the more comprehensive is the search result [Eib99].

4.4.3 Pre-processing of the Publications retrieved

The publications retrieved in section 4.4.2 are collected and ready for pre-processing. Pre-processing means in the context of this methodology to bring the publications into computer-easily-readable format. The following steps give a detailed description to pre-processing:

- Transformation of formats: publications are written in various document formats, e.g. PDF, Word, Text, HTML, XML, etc. There are few software products, which can read all the formats. In order to facilitate automatically computer analyzing huge amounts of publications, the different document formats have to be transformed into a certain format. In this dissertation, we use the software BibTechMon[2] to analyze publications, which can only understand Text format. Therefore, a computer program is written to automatically transform all publications into Text format.

- Cleaning up: The formulas and images are removed because those are difficult patterns for computer to recognize and analyze. It is noticed that, the main contents of the figures are expressed by legends. As a result, the figure legends should be kept in text. The special characters (e.g. €, $) are substituted by letters (euro, dollar etc.) for the purpose of better extraction. Furthermore, the references, bibliographies in articles are eliminated. There are two reasons for removing the references. Firstly, Citation Analysis has shortcomings: e.g. excessive self-citation may lead to inflated impact rankings of authors or papers; the content of citations should be examined, and so on. The second reason is that it is time-consuming to analyze references. References are presented in different structures and formats, which cause confusion and difficulties during the processing and cognition by computers. However, it is difficult and very taxing for researchers to process them into a unified form. If the references are kept in texts, they will be extracted into keywords within Co-word Analysis and influence the result of content analysis. Therefore, the references are removed from original publications and neglected for Bibliometric Analysis.

- Structuring: the publications are transformed into a machinable structure. Each publication is marked with an ID. The tags of important parts of publications are inserted, i.e. names of journal, publication year, organizations of authors, etc. Moreover, the scientific articles are structured in default sections such as Abstract, Introduction, Methods, Results, Discussion and Future Work. The default sections will be tagged according to investigation demand and expenditure.

After the pre-processing, the publications are condensed and structured as the example shown below (Fig. 4-13):

[2] BibTechMon (BTM): is a software product developed at Seibersdorf, which analyzes and visualises proximity relations between keywords and documents [KS01]. BTM is used in this dissertation as a tool for Bibliometric Analysis.

ID Number: 0039
Title: CITIZEN AND E INK DEMONSTRATE BENDABLE CLOCK
Subtitle: Clock to be showcased at the Eco-Products 2005 exhibition in Tokyo, Japan
Publication form: press article
Publication year: 2005
Region: Japan
Keywords: E Ink, Citizen, clock, electronic paper display
Content: Citizen Watch Co., Ltd. (CEO: Makoto Umehara, Japan) its consolidated subsidiary, Citizen T.I.C. Co., Ltd. (CEO: Toshio Takahashi; Koganei, Tokyo; Capital: 100 million yen) and E Ink (CEO: Russ Wilcox, USA), today announced the demonstration of the world's first flexible clock using an electronic paper display (EPD). This unique clock offers a thin, lightweight, fully flexible form factor combined with substantially lower power consumption over traditional displays through its use of E Ink Imaging Film TM. With a thickness of 3.0 mm, a weight of 1.5 kg and a battery life that is 20 times longer than conventional digital clocks, this revolutionary clock also can be bent significantly while operating. Furthermore, the use of electronic ink in the clock display enables a wide viewing angle of approximately 180 degrees with a bright, high contrast look, allowing for high visibility in either a dim room or under direct sunlight. These benefits allow the clock to be installed in locations that would otherwise be difficult with other technologies.
This product will be presented at the Citizen Booth at "Eco-Products 2005", which will be held at Tokyo Big Site for the 3 days from December 15 to 17.
Information Source: E Ink Corporation, Press Release

Fig. 4-13: Example of a pre-processed text

The pre-processed publications are collected and digitally saved in a text corpus. The result at the end of this phase is the text corpus, which is available for the following analysis and investigation.

4.5 Phase of the preliminary Identification of Technology Indicators

The tasks of this phase are to analyze the text corpus with Bibliometric Analysis and hence preliminarily identify the Raw Technology Indicators (Technology Indicators with names, definitions but no contents). The Bibliometric methods used here are traditional Publication Analysis and Co-word Analysis.

4.5.1 Using Publication Analysis

First of all, the Publication Analysis is used. In order to make the approach more standard, it is suggested to follow the steps introduced below as the user guide:

- **Counting the numbers of publications according to time**

  The temporal statistical series achieved are visualized in a diagram, from which we could estimate when the technology has been firstly referred to, or in which time period the number of publications about this technology has dramatically increased, etc. The temporal distribution of publications indicates the temporal development of the technology investigated. The result here is defined as the Raw Technology Indicator **“intensity of development”**.

- **Counting the numbers of publications according to locations**

  The achievement is regional distribution of publications on the given technology. Visualizing the statistical series by means of histogram, the most active countries are concluded. The result here is the identified Raw Technology Indicator **“key player regions”**.

- **Counting the numbers of publications according to organizations**

  Analogously, a histogram of organizational distribution of publications is demonstrated. The top 10 active institutes, firms, or universities are figured out, i.e. the Raw Technology Indicator **“key player organizations”** is identified.

- **Counting the publications according to authors**

  The active authors are shown, based on the statistical calculation of numbers of publications per authors. Similarly, the counting gives birth to the Raw Technology Indicator **“key player experts”**.

There are many other counting criteria within Publication Analysis. In the methodology presented in this thesis, only the points mentioned above are involved. The easiest way is always the best way.

4.5.2 Using Co-word Analysis

After the Publication Analysis, it is time to analyze the contents of the text corpus for the purpose of identifying other Raw Technology Indicators. The method used here is Co-word Analysis. This approach is also standardized with the following steps:

1. Extracting the texts corpus into keywords

The Co-word Analysis draws upon the assumption that a paper’s keywords constitute an adequate description of its contents. So, analysis of keywords is equal to the analysis of the contents represented by those keywords. The first step of keywords extraction is to break down the contents of publications into words and phrases. There is a pre-defined list of phrases in common use. Those phrases are identified by matching the pre-defined list with content of the text corpus. The rest part of the content is decomposed into words by e.g. cognition of blank characters. Subsequently, the stop words are eliminated. Stop words, also called fluff words, occur frequently as insignificant words. The common stop words include “the, a, by, and, of, this, then, which, that, such, are, at, if, it, can” and so on. They are normally unmeaningful and count for little for the main content of the text. The elimination of stop words does not influence the condensation of main meaning. The words excluding are potential keywords. The absolute text frequencies of the potential keywords are calculated, i.e. how often the potential keywords appear in texts is counted. The calculation is automatically done by computer programs.

2. Standardizing keywords

After the condensation of content, there is still a vast amount of keywords because of the redundancy. Therefore, standardization of keywords is required [HP04]. The following points illustrate the rules of standardizing keywords:

- The keywords are eliminated, whose text frequencies are less than two in order to simplify the subsequent work of standardization. The text frequency of the keyword remained f_k is more than one: $f_k \geq 2$
- Misspellings are corrected, e.g. planing $\Rightarrow$ planning. British English is transformed into American English, e.g. colour $\Rightarrow$ color.
- Plurals are transformed into singulars, e.g. models $\Rightarrow$ model, because plurals and singulars are doubleganger.
- Synonyms are brought into consolidation, e.g. Bibliometrics \ Bibliometric Analysis \ Bibliometric methods $\Rightarrow$ Bibliometric Analysis.
- Keywords with narrow sense are switched to keywords with broad sense, e.g. end users \ users \ a group of users $\Rightarrow$ users.

After the standardization, the absolute text frequencies are recalculated, i.e. to add the text frequencies of concerned keywords together. The result here is a list of strongly reduced keywords with their absolute text frequencies.

3. Calculating the co-occurrences and Jaccard Index

Generally speaking, terms, keywords, and concepts that co-occur frequently tend to be related. For instance, when the term "computer" is mentioned, the frequently co-occurred terms “hardware” or “software” are immediately thought about, not the terms “hardwood” of “softhead”. Co-occurrences of keywords indicate the semantic proximity or an idiomatic expression. One of the tasks in this step is to count co-occurrences of standardized keywords. That means how often the standardized keywords appear together in publications is counted. The results are shown in a Co-word matrix.

Based on the absolute text frequency (f_i, f_j) and co-occurrence of keywords (f_{ij}), the Jaccard Index (J_{ij}) is calculated according to formula 4.2.

$$J_{ij} = \frac{f_{ij}}{f_i + f_j - f_{ij}} \qquad (4.2)$$

$$i, j \in (0, n], \; Jij \in [0,1]$$

where:

- $J_{ij} = 0$ when $f_{ij} = 0$; i.e., keyword *i* and keyword *j* do not co-occur (keywords are mutually exclusive).
- $J_{ij} > 0$ when $f_{ij} > 0$; i.e., keyword *i* and keyword *j* co-occur (keywords are not mutually exclusive).
- $J_{ij} = 1$ when $f_{ij} = f_i = f_j$; i.e., keyword *i* and keyword *j* co-occur whenever either keyword occurs.

The Jaccard Index measures the similarity of the keywords set. Results are shown in the Jaccard matrix.

4. Visualizing the keywords and their linkages in Knowledge Map

Based on the calculation of absolute text frequencies of keywords, co-occurrences of keywords, and the Jaccard Index, the keywords are positioned in a two-dimensional Knowledge Map by using MDS. As its name implies, the Knowledge Map is a map that contains knowledge. Connected with this methodology, the Knowledge Map visualizes relevant information of the given technology, its Technology Indicators and the contents of those indicators.

5. Extracting Raw Technology Indicators with the aid of TI-Ontology

As discussed in section 4.1.3, the TI-Ontology that demonstrates the knowledge of Technology Indicators and their connections in the context of technology characterization and comparison, is built up. The TI-Ontology is based on the experience of earlier case studies. Therefore, it is commonly suitable for all technologies. Terms in this ontology are i.e. *cost, sales, turn-over, market share, trend, advantage, business segment, technology maturity, series production, market barrier* etc.

The keywords in the Knowledge Map are compared with those Technology Indicators and their synonyms in the TI-Ontology. The matched keywords are extracted. Logically, those matches are the required Technology Indicators.

Documenting all the keywords matched with TI-Ontology, we get the result of a list of Raw Technology Indicators, whose contents and values are filled in the next phase.

It is to note that a technology-specific ontology is built in this phase during the first investigation of the given technology. The keywords from the Knowledge Map offer a computer-based brainstorming of the essential concepts (one of the main elements of ontology, see section 3.5.1) in the domain of the given technology. The linkages based on co-occurrences of keywords indicate the relationships (another main element of ontology) of those concepts. The keywords and their linkages are evaluated by experts in this field. Based on the keywords, a taxonomic structure is built; concepts are defined; and the roles (relationships) are explicitly identified. Consequently, the technology-specific ontology is constructed, which conceptualizes the common understanding of the domain. It describes the semantics within the domain in both a human-understandable and computer-processable way. The technology-specific ontology built in this phase aims at supporting the interpretation of Knowledge Map in the next phase. It can also be used for standardization of keywords in the future investigations due to reusability of ontology.

4.6 Phase of Concretization of Raw Technology Indicators

There are two kinds of Technology Indicators. The first kind of Technology Indicator is assigned with qualitatively represented values, e.g. advantage, technological barrier, market segment. The other kind of Technology Indicator is filled with numerical values, e.g. sales, price, investment. In Phase 3, a list of Raw Technology Indicators is achieved. Filling them with contents and values, the Raw Technology Indicators become Complete Technology Indicators, which are the target outcomes of this phase. Assignment of the contents and values depends on the interpretation of publication diagrams and the Knowledge Map. Corresponding to the preliminary identification of Raw Technology Indicators in Phase 3, the results of Publication Analysis and the Knowledge Map are interpreted in Phase 4.

4.6.1 Values Assignment to Raw Technology Indicators by interpreting the Publication Diagrams

As discussed in section 4.5.1, the result of Publication Analysis is visualized in four diagrams. There are also four Raw Technology Indicators defined corresponding to the diagrams. General interpreting processes are illustrated as follows.

An interpretation model for the Technology Indicator **“intensity of development”** is built as a guide. As shown in Fig. 4-14, there are generally four stages to be observed.

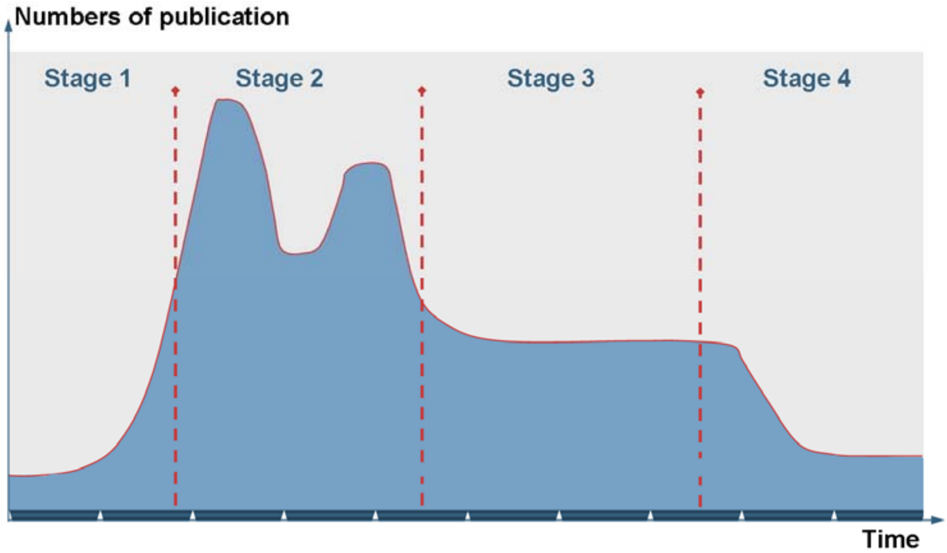


Fig. 4-14: Interpretation model for the Technology Indicator “intensity of development”

Stage 1 — early development: the number of publications of the given technology is small; however, there is a slight upward trend of the publication amount. Based on the assumptions that publications are the outcomes of research activities and the publication amount reflects research intensity, it is estimated that only a few researchers are working in the technological field in that time period; the technology is situated in the early stage of development.

Stage 2 — intensive development: in this stage, the temporal distribution of publications is not stable. The curve goes up and down many times, and is always accompanied by at least one steep upward slope. Every peak of publication amount indicates that the technology has attracted vast interest of researchers or companies at that time. A lot of time and energy are invested in its corresponding research. The technology is intensively researched and quickly developed in that time period.

Stage 3 — stable development: the third stage is stable. If there are any changes, they are all slight waves. This stage appears after the intensive development. The amount of publications is between wave crest and the bottom. Normally, the time period is longer than that of Stage 2. The stable distribution of publications indicates that the research activities are continuous, or technology is cited constantly as status quo in publications. Therefore, it is estimated that the technology has been used in industries and may become a basic technology.

Stage 4 — regression: the curve in Fig. 4-14 toboggans to the bottom and then levels off. It means that the technology is seldom cited in publications, and much less researched. The Stage 4 reflects a regression time period. The technology loses its significance; and might be substituted.

To assign the content to the Raw Technology Indicator **"key player regions"**, the histogram based on regional distribution of publications is analyzed. As shown in Fig. 4-15, the publication amounts are ranked with respect to regions. The hypothesis is that the more publications produced in that country, the more active is the country in that technological field. Therefore, the countries with top 10 publication amount are identified as those countries with the most influence in the technological domain.

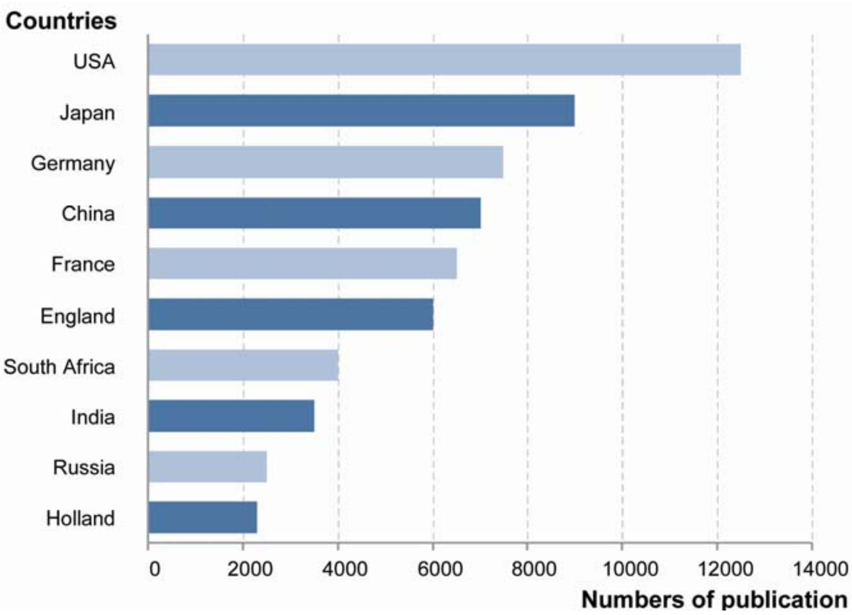


Fig. 4-15: Illustration of the assignment to the Raw Technology Indicator "key player regions"

By the same means, the contents for **"key player organizations"** and **"key player experts"** are educed from the publication distributions according to organizations and authors.

4.6.2 Values Assignment to Raw Technology Indicators by interpreting the Knowledge Map

Values assignment to those Raw Technology Indicators extracted by Co-word Analysis is implemented by the interpretation of the Knowledge Map. As introduced in section 4.1.2, a guide is constructed for the purpose of standardizing interpretation steps of Knowledge Maps. With the guide and TI-Ontology, the values are assigned to Raw Technology Indicators following the steps written below:

Focusing on the Raw Technology Indicators extracted directly from the Knowledge Map

As mentioned previously, Raw Technology Indicators are directly extracted from the Knowledge Map by matching keywords with the concepts within TI-Ontology. The TI-Ontology conceptualizes Technology Indicators in a hierarchical structure. Values are assigned to Raw Technology Indicators in a bottom-up way, i.e. the lowest sub-indicators are filled first and then the focus moves upwards to upper layers. The reason for bottom-up approach is that sub-indicators deal with only small area of keywords in the Knowledge Map. Co-relationships in small area are much easier to interpret.

As introduced in the guide in section 4.1.2, the thickness of lines in the Knowledge Map reveals the relative similarity of keywords. The distance between circles on the Knowledge Map reflects their content-similarity. The Raw Technology Indicator is focused together with its most frequently and strongly co-occurred keywords. The relationships between those keywords should be interpreted by the aid of the domain-specific ontology, which is built in Phase 3. The interpretations of the networked keywords and their relationships constitute the values of Raw Technology Indicators.

Fig. 4-16 illustrates the value assignment to the Raw Technology Indicator "product" in the investigation of E Ink technology. Product, in the context of Technology Indicator, means the applications of the technology, i.e. the products produced by using that technology. The strongly co-related keywords to "product" were focused.

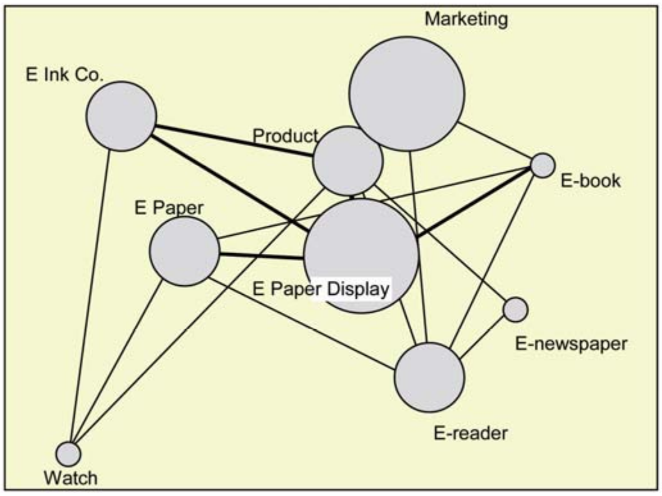


Fig. 4-16: Illustration of the assignment to the Raw Technology Indicator "product" (case study of E Ink technology)

Within the domain of E Ink technology, the connections between those keywords were understood as follows:

- One of the most important applications of E Ink technology is E Paper Display, which is developed by E Ink Co.
- E-reader is also an important application of E Ink technology. The e-book and e-newspaper are associated products of e-readers.
- E Ink technology might also be used in Watch. E Ink Co. might take part in the development of the watch with E Ink technology.
- E Paper is the basic elements for all applications of E Ink technology.

In respect the assignment of numerical values, the relationships between keywords are confusing and the domain-specific ontology does not help further. The original segments of texts need to be retrieved.

For instance, the targeted Raw Technology Indicator in Fig. 4-17 is “investment”. The keywords appearing most frequently with “investment” are selected. However, it is still difficult to conclude the logic relationships between those keywords. Therefore, the original documents are recalled with the “investment” highlighted. With the aid of the original text, the segment of the Knowledge Map in Fig. 4-17 is interpreted as: E Ink Co. has announced 15.8 million dollars of venture investment in 1998.

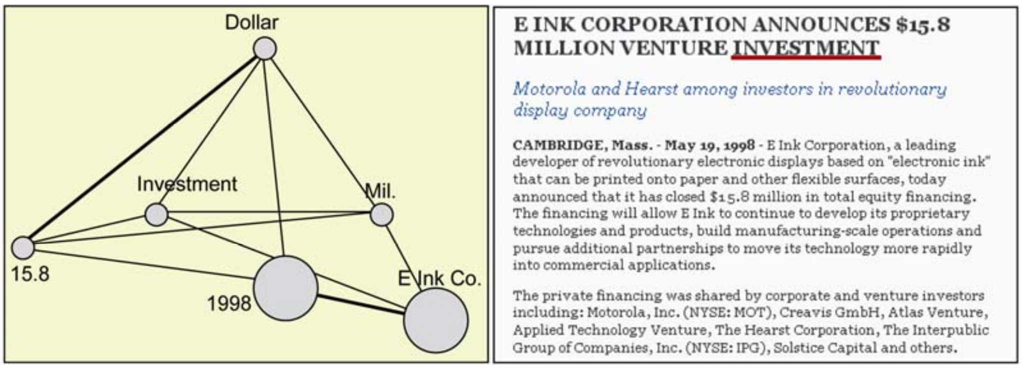


Fig. 4-17: Illustration of the assignment to the Raw Technology Indicator “investment” (case study of E Ink technology)

Analyzing dynamic changes

According to the guide to interpretation of the Knowledge Map (section 4.1.2), changes due to different time periods are traced after the values assignment to pre-extracted Raw Technology Indicators. The whole empirical time period is divided into several small time periods. Consequently, different Knowledge Maps are built by visualizing keywords of different time periods. The periodical Knowledge Maps are compared with each other. The following points are observed.

1. What are the differences of the values of Technology Indicators in different time periods? The changing of indicator values reflects the de-

velopment course of the investigated technology. Moreover, the tendency direction based on empirical data predicts future trend.

2. Which keywords appear newly? What are their relations with Technology Indicators? Keywords appeared in recent publications imply new development directions, new application fields of this technology, research progress, new key players, and so on. Knowledge extracted by analyzing newly occurred keywords is added to the values of corresponding Technology Indicators.
3. Keywords representing years, e.g. 1885, 2015, are picked out. Normally, years and their frequently co-related keywords link the historical developments of the technology together. The coming years, like 2080 compared with today, are connected consequentially to future trend. Thus, the historical development is disclosed by analyzing timing keywords. The results supplement the Technology Indicators and their values.

In one word, dynamic changes in the technology-specific domain are traced by analyzing periodical Knowledge Maps. Those changes are described by Technology Indicators and their values.

Analyzing marginal keywords

After the analysis of dynamic changes, the marginal keywords are taken into consideration. Marginal keywords as introduced in the guide (section 4.1.2) are the keywords at the edge of the Knowledge Map. Generally, marginal keywords have small text frequencies and fewer connections to other keywords. If the marginal keywords appeared recently, they may indicate the emerging development directions in the investigated technology domain. Otherwise, they imply vacancies in the technology and hence give suggestions to potential research directions. Sometimes, the marginal keywords are possibly from other disciplines, which are totally different from the investigated technology. In that case, they may predict new interdisciplinary development. Results achieved by analysis of marginal keywords are assigned as values to correlative Technology Indicators.

At the end of this phase, all Raw Technology Indicators are assigned with qualitative or quantitative values. They build the Complete Technology Indicators, which are based on the statistical analyses with Bibliometrics. The achieved Complete Technology Indicators are documented together and are ready for the evaluation.

4.7 Phase of Expert Consultation

The aim of this phase is to evaluate Complete Technology Indicators by experts. Expert consultation is the traditional way used to extract relevant information of technology. In the methodology proposed in this thesis, empirical information base of a given technology is retrieved, statistically analyzed, and interpreted by the aid of ontology in order to discover knowledge, i.e. Technology Indicators and their values. The results of Bibliometric Analysis are based on a semi-automatic process. It could happen that some misunderstandings appear in the processes of keywords extraction, standardization, or interpretation. Therefore, it is necessary to let a small group of experts evaluate the results from qualitative respect.

4.7.1 Expert Consultation

The Complete Technology Indicators simplify the construction of questionnaires because the main content of the questionnaire is nothing else but the definitions of Complete Technology Indicators, their values and the dynamic changes. The questionnaires are sent to a small group of experts (at least 6 feedbacks from experts are required) in the focused technology domain. Experts can be selected from the Technology Indicator “key player experts”. The values of Technology Indicators achieved from Bibliometric Analysis are verified by experts. Blanks are reserved for experts to write their different opinions, corresponding reasons and if possible, also references.

4.7.2 Comparison of Results from Experts and the Complete Technology Indicators

The questionnaires completed by experts are sent back. Contents of the questionnaires are collected and statistically analyzed. Supplements and comments from experts are summarized. Then the results are compared with the Complete Technology Indicators elicited from Bibliometric Analysis. There are three possibilities after the comparison:

- Experts agree with some of the Complete Technology Indicators and their values. If the statistical analysis matches the experts’ experience, those Complete Technology Indicators do not need to be reprocessed.
- Experts do not agree with some values of the Complete Technology Indicators based on Bibliometrics. The discrepancies may be caused by wrong explanation of the Knowledge Map or incorrect consolidation of synonyms during the keywords standardization, etc. The discrepancies are then traced again in text corpus. If the experts’ opinions are found correct in texts, those Complete Technology Indicators are changed ap-

propriately to qualitative assertion by experts. Otherwise, the different opinions are sent again to experts with the original references, which have been used for Bibliometric Analysis. The experts give their opinions concerning the original documents again. The iteration process stops until a common agreement is reached. Final results are documented.

- The open values of the Complete Technology Indicators, i.e. those can not be achieved from Bibliometric Analysis, are filled in by experts. It is expected that the missing values of certain Technology Indicators can be supplemented by experts with related references. Therefore, the Complete Technology Indicators are more comprehensive by adding contents offered by experts.

After the comparison and modification of the results from Expert Consultation and Bibliometric Analysis, the Complete Technology Indicators are corrected, refined, complemented, and finalized. Documenting all the modified and confirmed Technology Indicators and their corresponding values, the final output of the methodology, i.e. a list of so-called Final Technology Indicators is received, which is exactly required by decision makers in companies.

4.7.3 Regular Update

As discussed previously, technology develops quickly to catch up with the fast change of products, customer requirements, manufacturing processes, etc. So does the information about technology. Therefore it is therefore necessary for the methodology to search, monitor, and statistically analyze the up-to-date publications. By doing that, the text corpus is updated regularly. And the whole process is carried out again to achieve the recent development of technologies. The update step is shown with an arrow in Fig. 4-10. The frequency of the update process depends on users' demands.

No more additional effort is needed in the update and the iterative process for the following reasons:

- Information search (update search) and collection is carried out automatically.
- Keyword extraction is computer-aided. The keywords standardization implemented previously is saved and still valid now. Only the newly appearing keywords need to be standardized, which can be aided by the technology-specific ontology generated during the last analysis.
- The TI-Ontology stays unchanged. It supports the information search, identification of the Raw Technology Indicators, and the interpretation

of the Knowledge Map. The TI-Ontology is predefined, which is stable and suitable for all technologies.

- The technology-specific ontology is reusable, which is built up in previous implementation of the methodology. The technology-specific ontology assists the keywords standardization process and the interpretation of Knowledge Map by offering a common understanding of the terms used in the given technological domain. During the update process, it is not necessary to build up a new ontology. If necessary, the existing ontology is improved by analyzing only the newly appearing keywords.
- The interpretation of the Knowledge Map is based on former experience. Only the newly appearing keywords need to be interpreted. Similar in Expert Consultation, only the current changes of Technology Indicators and their values are sent to experts for evaluation.

4.8 Integration of the Methodology and the Technology Database

As mentioned previously in section 2.2, the Heinz Nixdorf Institute has developed an innovative Technology Database, which aims at supporting the product innovation process by managing, monitoring, and reporting technologies and their applications. The Technology Database is already internally applicable at Heinz Nixdorf Institute (Fig. 4-18) and also successfully used in several industry companies in Germany[3]. The methodology introduced in this dissertation is integrated into the information procurement process of the innovative Technology Database. In the following paragraphs, the connections, interfaces, and workflow of the methodology and the innovative Technology Database are discussed in detail.

[3] The innovative Technology Database has been customized according to user requirements and successfully implemented in the several industry companies in Germany.

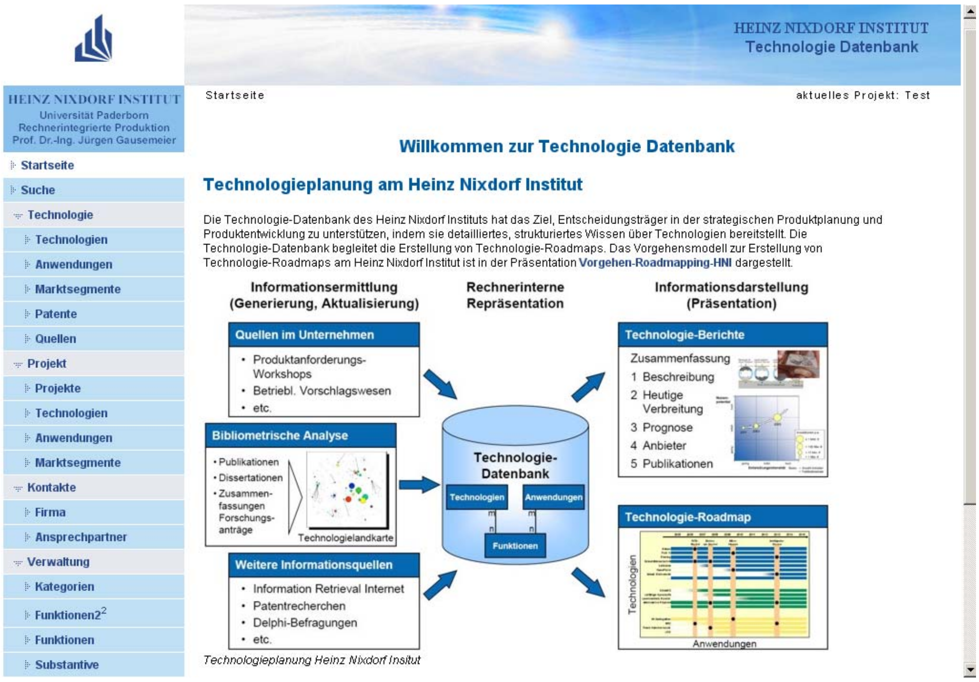


Fig. 4-18: Screenshot of the Heinz Nixdorf Institute Technology Database

4.8.1 Technology Indicators as input for the Technology Database

The methodology for the identification of Technology Indicators is applied in the process of information procurement on the input side of Technology Database (Fig. 4-19). The objectives here are to collect, analyze and select the relevant information related to technologies.

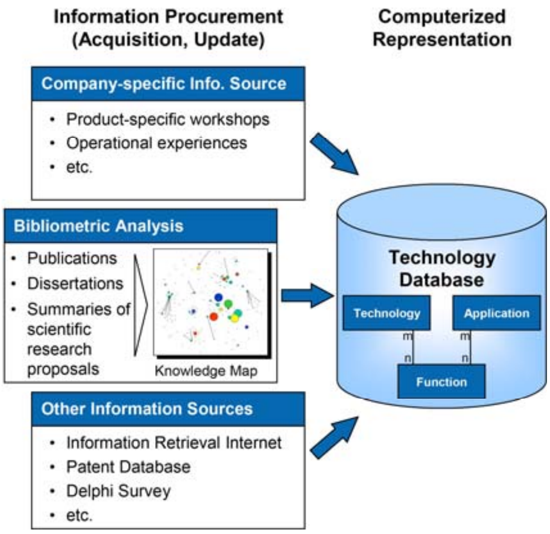


Fig. 4-19: Input side of the innovative Technology Database

Except for some figures and some general information written in continuous text, the inputs for the Technology Database are mainly those Technology Indicators with their values obtained through the methodology. For each Technology Indicator defined in the TI-Ontology, there is a corresponding data field constructed in the user interface in order to enter the Technology Indicator and its values into the Technology Database. For instance, the input interfaces for the Technology Indicators "function" (Fig. 4-20) and "technology maturity" (Fig. 4-21) are illustrated as follows:

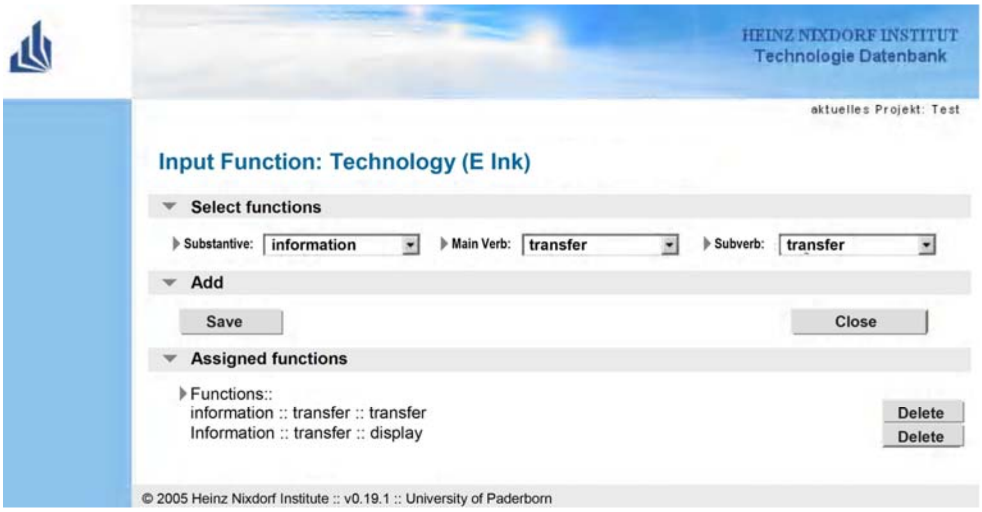


Fig. 4-20: Input interface for the Technology Indicator "function" (screenshot from the Technology Database used at Heinz Nixdorf Institute)

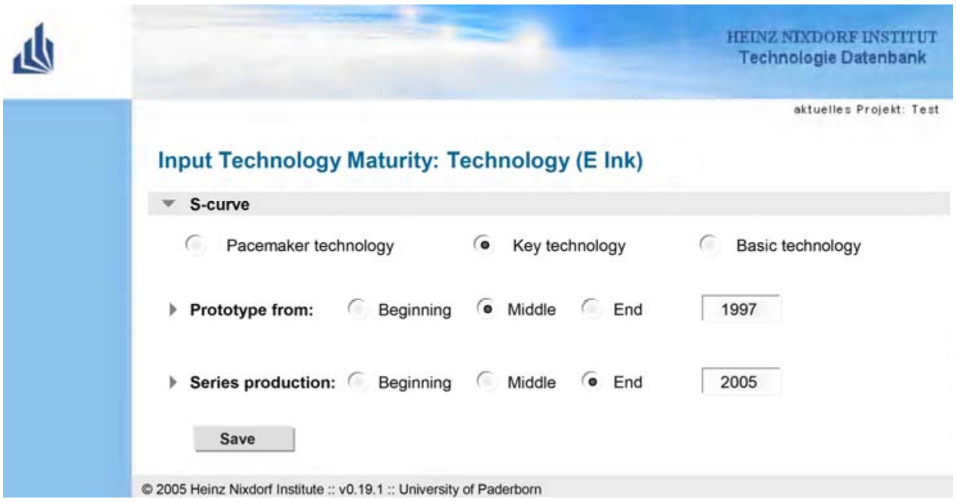


Fig. 4-21: Input interface for the Technology Indicator "technology maturity" (screenshot from the HNI Technology Database)

Similarly, all the Technology Indicators, their values, and their updated information achieved from the Methodology are input into and saved in the Technology Database.

4.8.2 Visualization of Technology Indicators as output of the Technology Database

As mentioned previously, decision makers can be informed quickly about the key information of target technologies with the aid of those Technology Indicators. Technology Indicators facilitate the evaluation and comparison of technologies. The use of Technology Indicators is realized by visualizing them in different report formats. As shown on the output side of Technology Database (Fig. 4-22), there are two kinds of formats: Technology Reports and Technology Roadmaps.

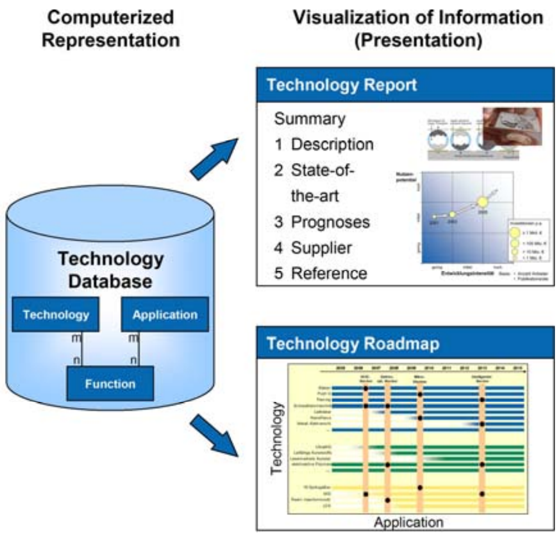


Fig. 4-22: Output side of the innovative Technology Database

Technology Reports offer decision markers in companies the essential information, key points of focused technology in detail by describing Technology Indicators in continuous texts and visualizing them in various portfolios. The Technology Report is constructed in a fixed structure of five sections: summary, general description, state of the art, prognoses, and experts. For instance, in the section of general description, the Technology Indicators “working principle”, “technology advantages”, etc. are characterized in continuous text. The texts in the section of state of the art contain the Technology Indicators such as “current applications”, “current market development” etc.

Some Technology Indicators are combined in portfolios in order to show the characteristics, dynamic changes, and trends of technologies clearly at a glance. In the following two figures (Fig. 4-23, Fig. 4-24), two examples of portfolios are illustrated.

This S-Curve shows three Technology Indicators: “customer benefits”, “accumulated R&D cost” and the “technology maturity”, which interact closely with each other. The technology is positioned in the S-Curve by measuring the values of the three Indicators.

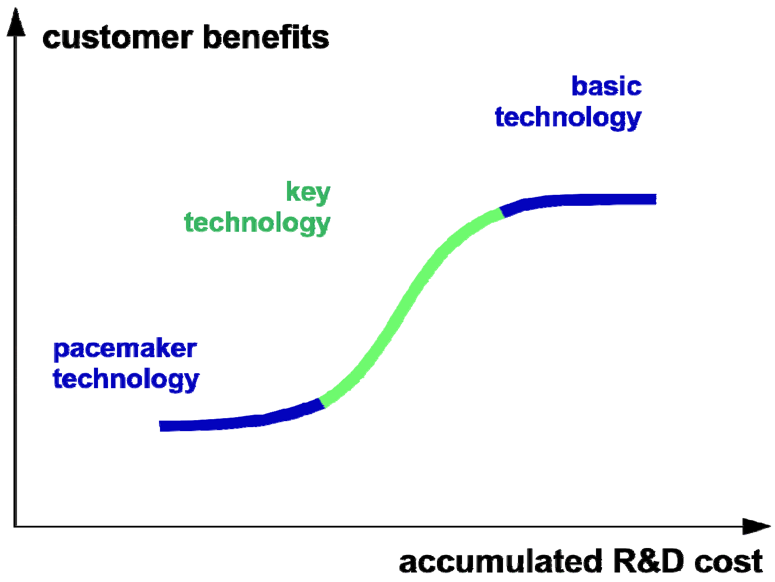


Fig. 4-23: S-Curve

The current technology maturity is plainly visible in the S-Curve; the further movement of the technology can be estimated from the trend of S-Curve. For example, when a technology is mature enough as a basic technology, the customer benefits are saturated, i.e. it won't increase a lot with the enhancement of accumulated R&D cost. It is then predictable that a substitute technology will occur on demand. A well-known example is: telephone has switched from analog to digital.

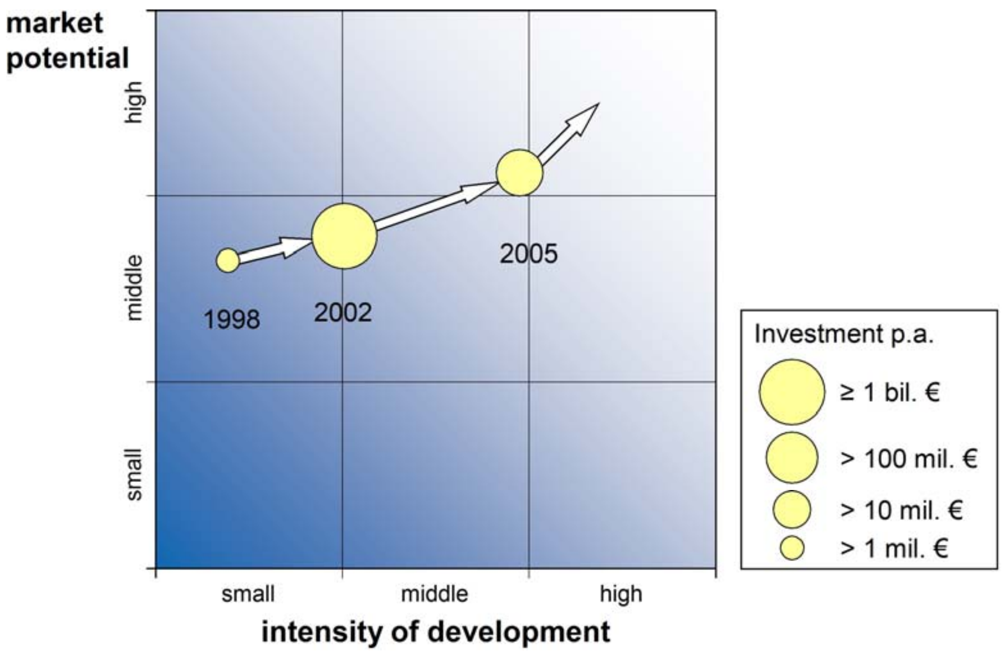


Fig. 4-24: Portfolio of developing intensity and market potential

In the portfolio shown in Fig. 4-24, three Technology Indicators are visualized simultaneously: “intensity of development”, “market potential”, and “investment”. The “intensity of development” is measured by temporal distribution of publications and the growth of experts in the focused technology domain. The "market potential" is estimated by current economic data, technology attractiveness, market trend, etc. As demonstrated in that portfolio, the dynamic movements of the technology are easily observed. The last arrow logically indicates the future developing direction of the technology.

Comprising all the Technology Indicators, Technology Reports satisfy basic requirements of decision makers on evaluation of technologies.

Technology Roadmap is a plan that shows which technology has been or could be used for which applications at what time (see section 2.1.2). The generation of Technology Roadmap is realized automatically in the innovative Technology Database and is based on the values of three Technology Indicators: "application", "function", and "technology maturity". As illustrated in Fig. 4-25, numerous technologies and applications are connected with each other through functions. The gradually changed color bands reflect technology maturity, i.e. if the technology is already used in series production. Technology Roadmaps spontaneously offer a large number of optional combinations of technologies and applications. With the aid of Technology Indicators, decision makers evaluate all the possible combinations in the Technology Roadmap and finally select the finest solutions according to their own demands.

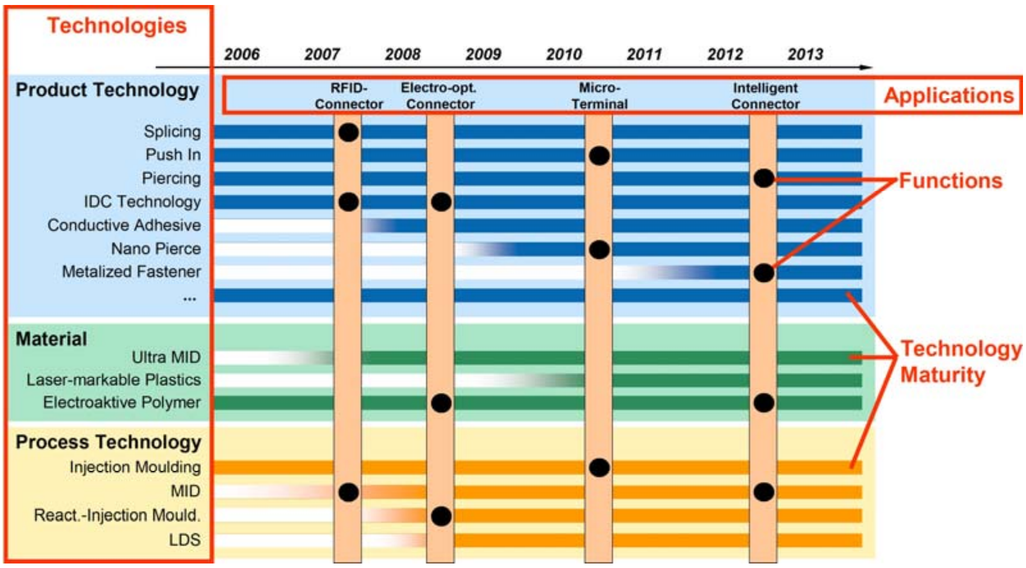


Fig. 4-25: Simplified Technology Roadmap

To summarize chapter 4, the methodology expatiated here realizes semi-automatic extraction of Technology Indicators (knowledge) from text corpus focusing on a given technology (huge amount of information). The methodology is embedded in the innovative Technology Database developed by Heinz Nixdorf Institute. Technology Indicators procured through the methodology are saved in the Technology Database and used in forms of Technology Report and Technology Roadmap in order to inform decision makers about the current status of technology development, trends and hence to support them in technology planning and product innovation.

5 Case Studies and Evaluation

As mentioned in chapter 2, Heinz Nixdorf Institute has developed an innovative Technology Database, which works as an intelligent technology management system to support the product innovation processes. The methodology for the identification of Technology Indicators is proposed in chapter 4 to optimize the information procurement process of the Technology Database. The methodology helps decision makers in companies to effectively extract relevant information about technologies and hence speed up their decision-making process. In order to prove its feasibility and validity in practice, some case studies were carried out. One of the case studies is described in this chapter, which aims at the characterization and evaluation of the technology Molded Interconnected Devices (MID). The case study strictly followed the process defined in the methodology. The implementation is introduced here step by step with illustrations. Furthermore, with the case study, the methodology proposed in this dissertation is evaluated according to the requirements declared in section 2.4.

5.1 Case Study of MID Technology

One of the most comprehensive and successful case studies to evaluate the methodology proposed in this dissertation is the investigation of MID technology. The MID technology is an emerging technology that enables the manufacturing of spatial circuit carriers. Three-dimensional plastic parts are manufactured and their surface is partially metallized. Different manufacturing technologies are available. Electronic parts can be soldered or glued with conductive glue on the emerging electrical conductive strips and contacts. This can happen under utilization of all spatial degrees of freedom. Thus MID integrates mechanical and electronic functions like “carrying and electrically connecting parts” in a single entity. Moreover, the metallized surfaces can be arranged in such a way that shieldings, cold bridges or antennas emerge. Fig. 5-1 shows the application of MID in the robot’s housing. Substantial advantages of MID are miniaturization, reduction of parts and the freedom of spatial configuration. [RA304] [KFG+06]

Since the case study exemplifies the methodology, it was carried out in a five-phase process as defined in chapter 4. The whole process of the case study is introduced detailedly in the following paragraphs.

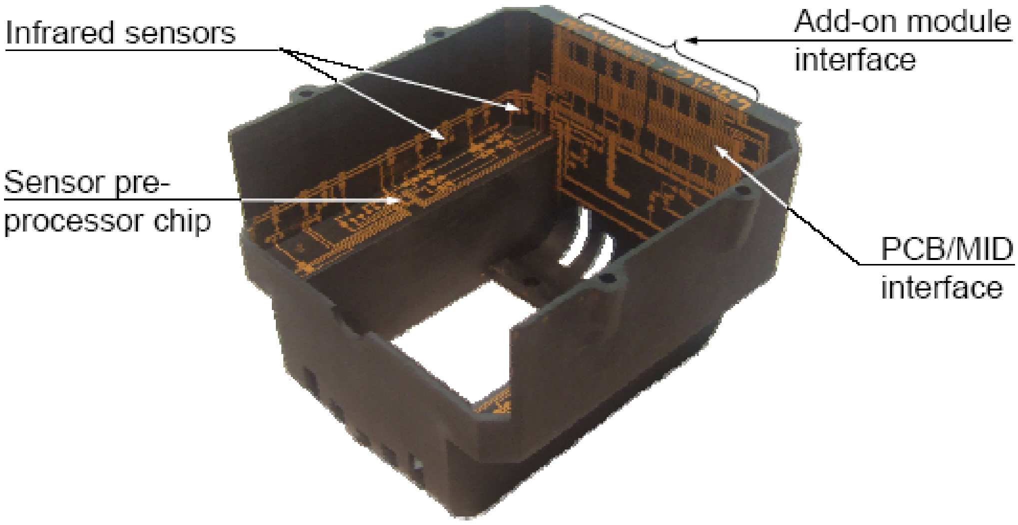


Fig. 5-1: MID circuits on the robot housing, connecting electronic devices

5.1.1 Phase of Problem Analysis

In our case study, the investigation objective is not so complicated. Our workgroup at Heinz Nixdorf Institute get in touch with various technologies within the daily research work. One emerging technology called Molded Interconnected Devices has especially caught our attention. Our workgroup decided to investigate MID technology, and to check if there are any possibilities to use this technology for our projects. Furthermore, the investigation focused on the market development of MID. The reason was that we considered MID as a promising technology. The factual evidences were required to prove that argument, and to help us to make the decision: if we should choose the MID technology for our research projects.

The result of this phase was the investigation objective. Target technology was the given technology MID. There was no other candidate or similar technologies. In this case, decision maker was the research workgroup at Heinz Nixdorf Institute. The requirements for the investigation are listed as follows:

- With the Technology Indicators, the MID technology should be characterized and evaluated.
- The investigation emphasis is the market development of MID.
- The investigation does not focus on any time period. The comprehensive history of MID technology should be observed.
- The development of MID in Germany should be paid special attention to.

- It should be figured out whether the MID technology was applicable for the research projects in the workgroup.
- Will MID technology become more important? Or will it become less important?
- What is the future trend of MID technology?
- Does MID technology attract industries' attention?
- etc.

Aiming at the objectives, the MID technology was investigated by identifying its Technology Indicators from a document collection. The next step was to search relevant documents about MID.

5.1.2 Phase of Literature Search

In this phase, the documents relevant to MID technology were retrieved. The Method used was Information Retrieval. As introduced in chapter 4, there is a preparation process, i.e. to limit the search area. In our case study, the following measures were taken according to the methodology (section 4.4.1).

1. Determination of the Forms of Publications

There were two investigation directions in the case study. One direction was to investigate the current status, and trend of MID market development; the other one was to evaluate if the MID technology was suitable to be used in our current research projects. For the latter one, information about technology development of MID was necessary. Guided by the Fig. 4-12, attention should be paid firstly to the following publication forms during the information search:

- press release; advertising article
- enterprising financial report; market study; final project report
- patent
- scientific article

Although the other publication forms were not excluded, they were backup choice because the resource assigned to the case study was relatively small.

2. Selection of Information Repository

After the selection of publication forms, the corresponding information repositories were taken into consideration. Guided by the Table 4-1, several information repositories were determined according to various criterions.

Firstly, since our research group is in Germany, the language preferred by the research assistants and industrial partners is German. Based on this criterion, documents, articles written in German deserved prior attention. The Japanese publications, Chinese, Indian, etc. were automatically filtered out.

Secondly, the case studies focused on the research development or market development of MID technology in Germany. Therefore, the German as well as the Europe databases were mainly searched. The high priorities were given to the articles published by German research centers or corporation press department.

The third criterion is publication form. As mentioned above, four kinds of publications should be firstly collected. The press releases, advertising articles, enterprising financial reports, market studies, final project reports are normally easy to get from the Internet, e.g. homepages of the companies, press releases, newsletters on the websites of companies. Online search engines “Google” and “Yahoo search” were determined in order to find easy entrance to those publication forms. Considering patent search in German language area, or Europe regions, the online patent database “European Patent Office” (http://ep.espacenet.com) was chosen. And for scientific articles, “Google Scholar” (http://scholar.google.de) was selected. Google Scholar is a freely-accessible web search engine that indexes the full-text of scholarly literature across an array of publishing formats and disciplines. The Google Scholar index includes most peer-reviewed online journals of the world's largest scientific publishers [UGS05-ol].

The last criterion in Table 4-1 is domain. Since the target technology was fixed to MID technology, the MID-specific database was obviously ideal. Through our partnership with Research Association Molded Interconnect Devices 3-D MID e.V., we got access to their 3D-MID database. There are also a lot of worthy information about MID technology on the website of 3-D-MID e.V. (http://www.3dmid.de), which is only accessible with valid user accounts.

The search areas were limited, and the corresponding information repositories were selected. It was time to search documents, articles, etc. relevant to MID technology.

3. Search for Publications

Guiding by the methodology (see section 4.4.2), the relevant documents were retrieved with following steps. The chosen search technique is keyword search.

First of all, aiming at searching literature relevant to MID technology, a group of phrases were defined: such as Molded Interconnected Devices, or Moulded Interconnected Devices (America English), 3D-MID technology, MID, integration of mechanics and electronics, 3D substrate, electrical circuit, also Räumliche spritzgegossene Schaltungsträger (German Version). Those phrases were formulated with Boolean Operations, i.e. conjunction (AND), disjunction (OR), and negation (NOT). Examples of the combinations using Boolean Operations were: MID "AND" electrical circuit "AND" 3D substrate; MID "AND" integration of mechanics and electronics; Molded Interconnected Devices "or" MID technology. Those combinations of phrases were used as search queries to search documents in the databases, or search engines determined above, e.g. Midis Database (3D-MID e.V.), Google Scholar, etc. The documents that matching with those search queries were listed as search results. Fig. 5-2 shows an example in the case study. On the left side in Fig. 5-2, search query used was "MID". 465.000.000 documents were retrieved; while on the right side, search query was optimized to "MID" and "Molded Interconnected Devices". And the search results were reduced to 67 documents. It is verified that the more appropriate are the search queries, the less and more essential are the search results.

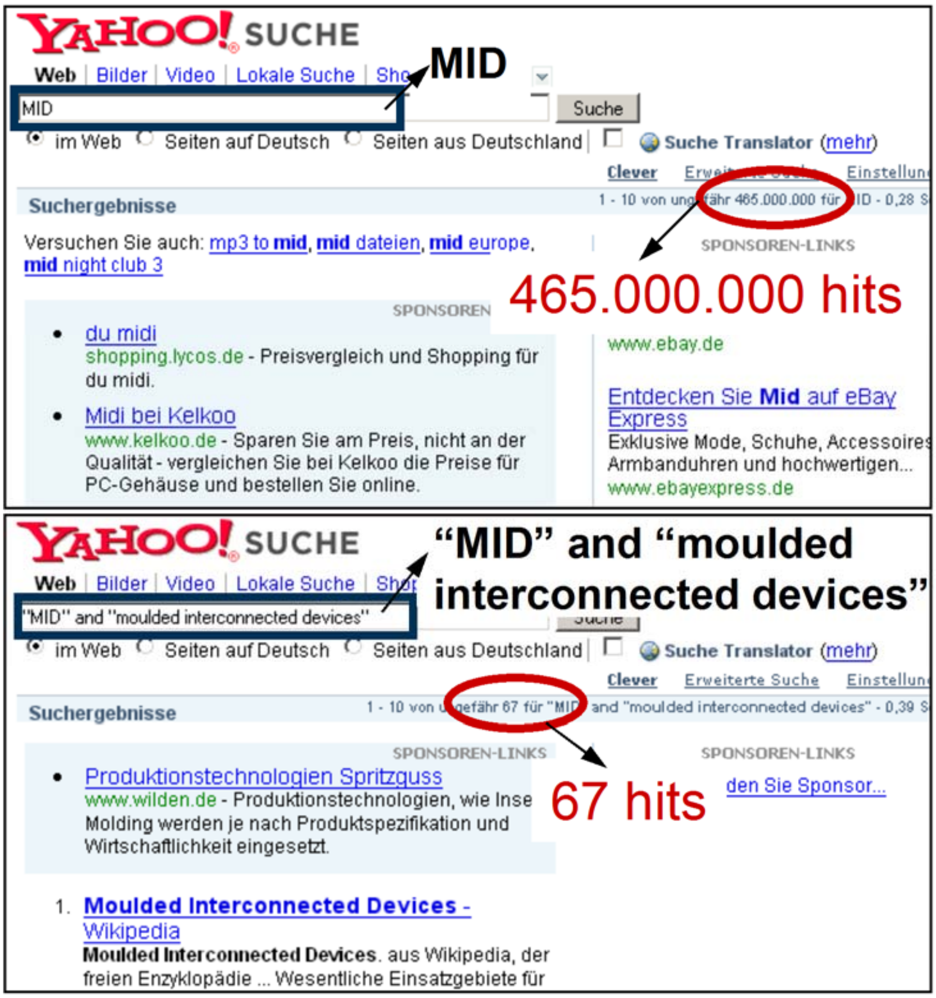


Fig. 5-2: Comparison of the search results related to different search queries (search engine used: Yahoo search, http://de.yahoo.com)

It is to notice that the MID-relevant information was also searched in the internal database of the workgroup. One market study on MID technology was retrieved internally. As a result, 489 documents thematically relevant to MID were retrieved from different information repositories. As shown in Table 5-1, the retrieved documents were composed of press articles, patents, scientific articles, etc. The total publication volumes are 4668 pages, which can not be read one by one manually.

Table 5-1: Overview of the retrieved documents relevant to MID technology

publication form	press release, advertising article	scientific article	handbook	patent	research application	market study, final project reports	the rest publication forms	Sum
quantity (numbers of publications)	62	127	1	136	14	46	103	**489**
publication volume (in pages)	188	1857	301	136	114	1623	449	**4668**

The 489 documents retrieved from various databases were collected to form a text corpus. In the next step, the text corpus was pre-processed in order to simplify the further analysis.

Pro-processing of the Text Corpus

As guided by the methodology, there are three steps of changing the retrieved documents into a computer-easily-readable format. In our case study, the documents were pre-processed according to the following steps as proposed in section 4.4.3.

- **Transformation of formats:** all documents retrieved in various formats, e.g. PDF, Doc, HTML, PPT, were transformed into Text format, because the software chosen for Bibliometric Analysis in Phase 3 can only read Text format. A small program was written to implement the transformation automatically, which has replaced the manual work, and saved a lot of time as well as human power.

- **Cleaning up:** in this step, the noisy information was eliminated. Noisy information is the information that interrupts normal analysis, or is the unimportant and characterless information. In the case study, the formulas were firstly removed. The reason is that the numbers decomposed from the formulas disturb normal analysis of sales amount, production volume, etc.; and the symbols decomposed from the formulas are noisy (meaningless) information. Besides formulas, the images in documents were also eliminated because the methods so far can not process images. Moreover, the references in articles were removed as explained in section 4.4.3. Some special characters, e.g. €, $, were replaced by words (euro, dollar) to facilitate their automatic extraction by software. After the cleaning up, the information in text corpus was filtered and reduced. The next step was to structure the texts.

- **Structuring:** in the case study, all texts were transformed into a unified structure. That means every text was marked with an ID number, title, subtitle, publication form, publication year, author, organization, region, keywords, abstract, content, and information source (see Fig. 5-4). The aim of the structuring is to facilitate the search in determined fields. For instance, the information analysis can be restricted to the fields of keywords and abstract in case the contents are not interesting to decision makers. Another example is that the structured texts could be classified according to publication year, which simplified the comparison of dynamic changes of the MID technology over time.

An example of pre-processing is shown in Fig. 5-3 (the original document in PDF format with images and so on) and Fig. 5-4 (the pre-processed Text format). The relevant publications about the MID technology were retrieved, transformed, cleaned up, and structured. The Phase 2 of literature search was completed. The result was a structured text corpus specific to the MID technology, which is ready for Bibliometric Analysis in Phase 3.

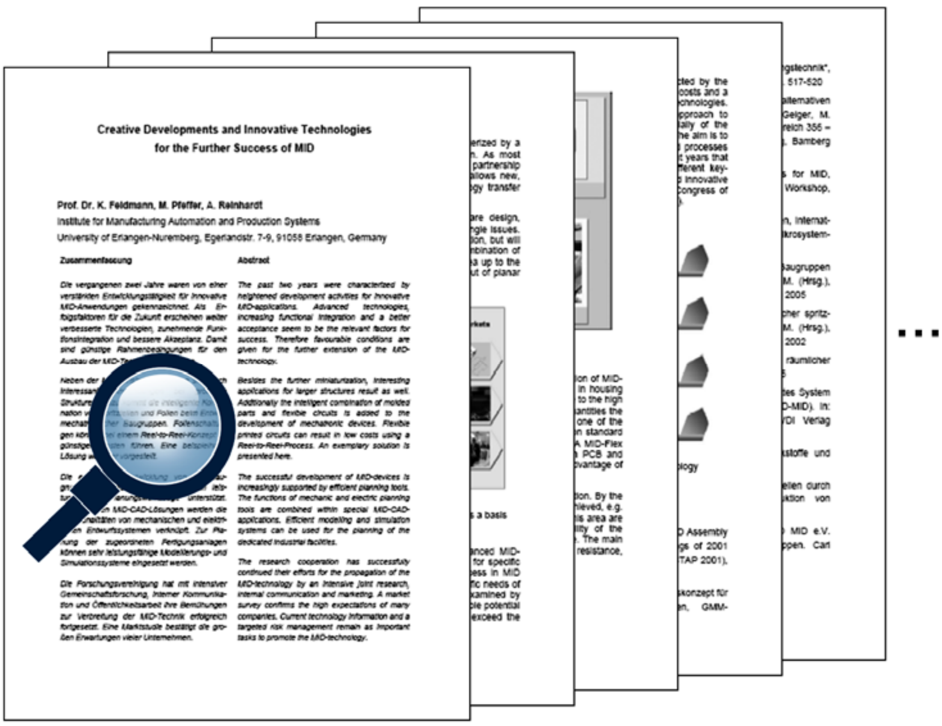


Fig. 5-3: An example of raw information – a scientific article about MID in PDF format (information source: 7. International Congress MID 2006 in Fuerth, 27.-28. September 2006)

ID Number: 002
Title: Creative Developments and Innovative Technologies for the Further Success of MID
Subtitle:
Publication form: scientific article
Publication year: 2006
Author: K. Feldmann, M. Pfeffer, A. Reinhardt
Organization: Institute for Manufacturing Automation and Production Systems, University of Erlangen-Nuremberg
Region: Germany
Keywords:
Abstract: The past two years were characterized by heightened development activities for innovative MID-applications. Advanced technologies, increasing functional integration and a better acceptance seem to be the relevant factors for success. Therefore favourable conditions are given for the further extension of the MID-technology. Besides the further miniaturization, interesting applications for larger structures result as well. Additionally the intelligent combination …
Content: 1. Potential and Development Fields The development of innovative and successful MID-products is characterized by a combination of different key competencies within the production chain. As most companies do not cover all aspects of MID realization steps, a close partnership between industrial development and research activities of universities allows new, powerful products. The 3-D MID e.V. is the platform for the technology transfer between research and production (figure 1). Three main issues for opening up new markets for MID-products are design, production and marketing. None of these aspects can be handled as single issues. This would only result in another optimizing routine for standard production, but will not break the technical limitations of conventional manufacturing. A combination of these factors along the whole product development from the product idea up to the manufacturing is the basis for the implementation of new functionality out of planar processing. Fig.1: 3-D MID e.V. in a central position coordinates the MID-activities as a basis for successful technology Successful MID-parts can be identified by the implementation of advanced MID-technology and by solving technical challenges. Today this is possible for specific applications …
Information Source: 7. International Congress MID 2006 in Fuerth, 27.-28. September 2006

Fig. 5-4: A segment of the pre-processed article shown in Fig. 5-3, in Text format

5.1.3 Phase of preliminary Identification of Raw Technology Indicators

Here, the text corpus retrieved in the last phase was analyzed by using the Bibliometric Analysis to extract Raw Technology Indicators and their values with respect of MID technology. Technology Indicators are defined in this dissertation as features or values that can characterize and evaluate technologies. The ontology of all general Technology Indicators was built and introduced in section 4.1.3. As guided by section 4.5, the extraction of those Raw Technology Indicators in the case study was realized by carrying out the traditional Publication Analysis, and the Co-word Analysis.

Carrying out Publication Analysis

In this step, the publications retrieved were calculated over time, according to regions, organizations, and authors. The results were four diagrams that represent separately four corresponding Raw Technology Indicators, as defined in section 4.5.1. Two diagrams of the case study were selected as examples to demonstrate the process of identifying Raw Technology Indicators by using Publication Analysis. Fig. 5-5 and Fig. 5-6 illustrate respectively the Raw Technology Indicators “**intensity of development**” and “**key player organizations**”.

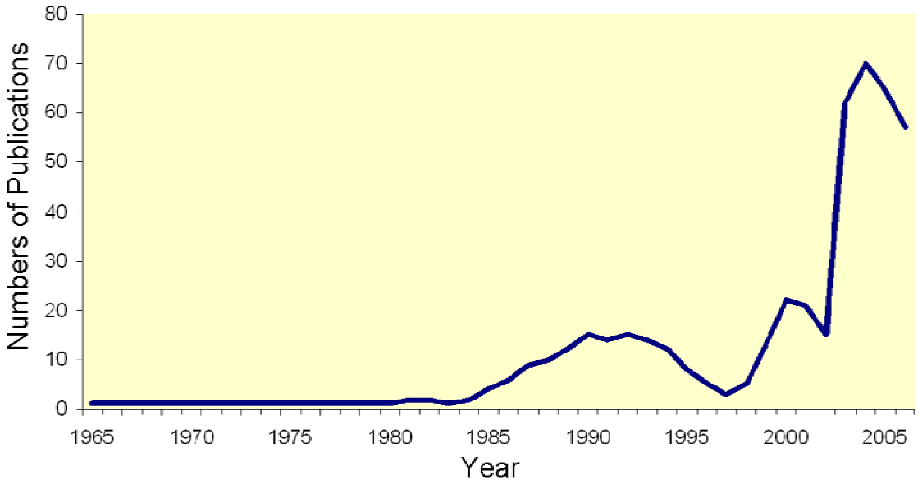


Fig. 5-5: The Raw Technology Indicator "intensity of development" identified by counting the publication amount according to time

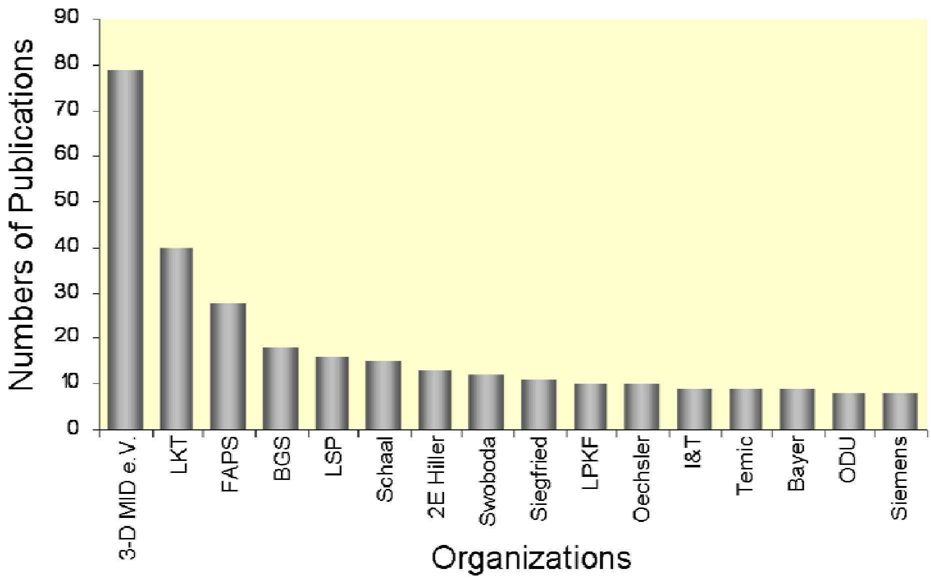


Fig. 5-6: The Raw Technology Indicator "key player organizations" identified by counting the publication amount according to organizations

Carrying out Co-word Analysis

Following Publication Analysis, the contents of the text corpus were analyzed by using Co-word Analysis. The software used in this dissertation is BibTech-Mon, which is developed by ARC systems research GmbH. With the help of BibTechMon, the contents of the pre-processed texts were decomposed into words. The keywords were extracted by filtering out the default stop words, which were defined by developers of BibTechMon. Fig. 5-7 enumerates the default English stop words and German stop words.

English Stopwords	German Stopwords
accommodate	abgeführte
accommodating	abgegebene
accompanied	abgegrenzt
according	abgegrenzten
achieve	abgelegt
achieved	abgelegten
achieving	abgelehnt
acknowledge	abgeleitet
acknowledges	abgeleitete
actual	abgeleiteten
additional	abgeschätzt
addressed	abgeschiedenen
adequate	abgeschirmt
adjust	abgeschlossenen
adjusted	abgeschwächte
adopt	abgeschwächter
adopting	abgesehen
affected	abgesichert
affects	abgesicherten
after	abgesteckt
again	abgestimmt

Fig. 5-7: Default stop words in BibTechMon - © 2007 systems research

After the filtering of stop words, there were 10943 keywords extracted from the text contents (Fig. 5-8). Guided by section 4.5.2, the keywords were standardized with the following steps.

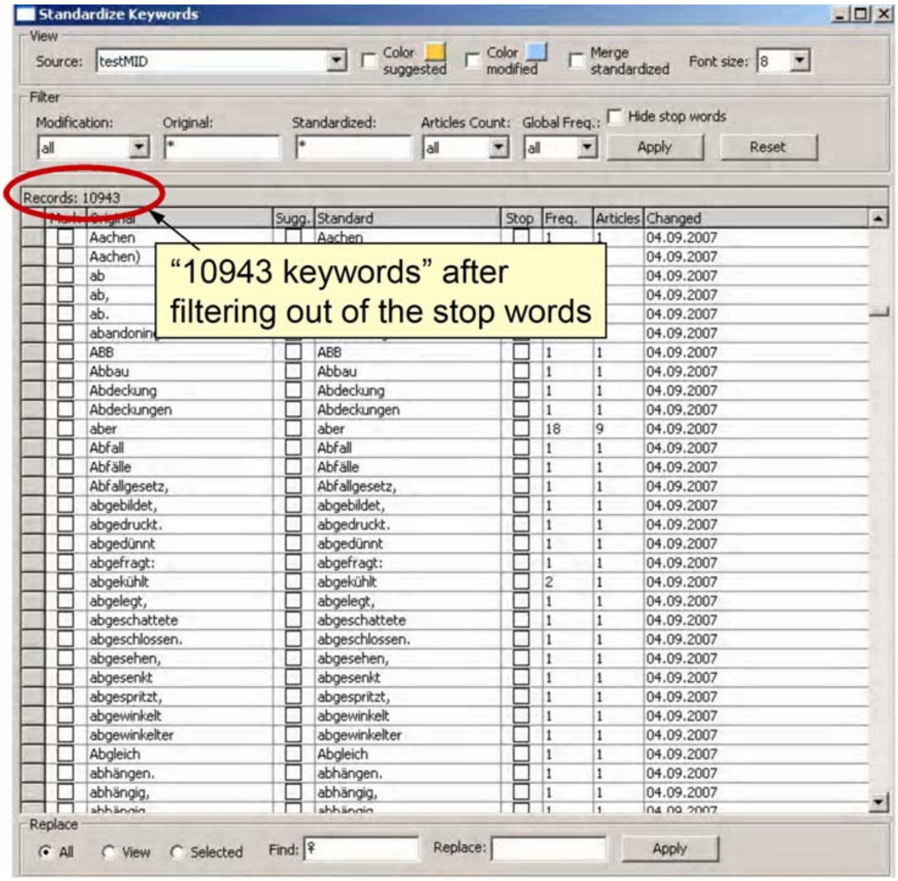


Fig. 5-8: The keywords extracted by filtering the default stop words

- The keywords, whose text frequencies were less than two, were eliminated. The results were 2559 keywords left (Fig. 5-9).

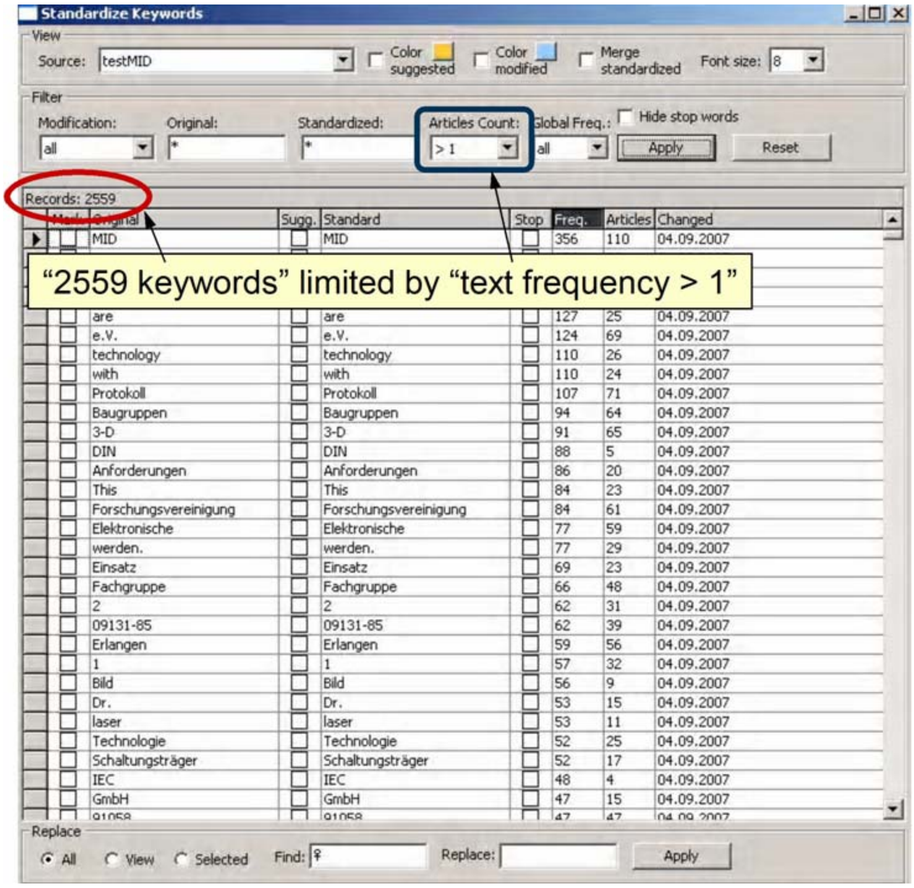


Fig. 5-9: The keywords extracted by eliminating those, whose text frequencies were less than two

- Misspellings were corrected. In the case study of MID, there is a company called Lisa Dräxlmaier GmbH. It was written in some publications as Lisa “Dräxlmeier” GmbH. During the standardization of keywords, the misspelling was corrected.

- All plurals were transformed into singulars. For example, in the case study, “trends” was transformed to “trend”; “applications” was transformed to “application”; “functions” was transformed to “function”.

- Synonyms or different formulations were unified. Furthermore, the English keywords were translated into German language because the investigation focused on Germany, and most of the documents retrieved were written in German. Instances in the case study are: the keyword “Molded Interconnected Devices” was standardized as “MID”; the keywords “copper” and “Cu” were brought into consolidation; “advantage” was translated into “Vorteil” (German word for advantage).

- Some keywords with narrow sense were merged into their upper class concepts. For instance, in the case study, the keyword “Herstellung-sunternehmen” was changed to “Hersteller”.

The standardized keywords were merged. The global frequency and text frequency of the keywords extracted were calculated again. There were only “323 keywords” left after the standardization. The result was shown in Fig. 5-10.

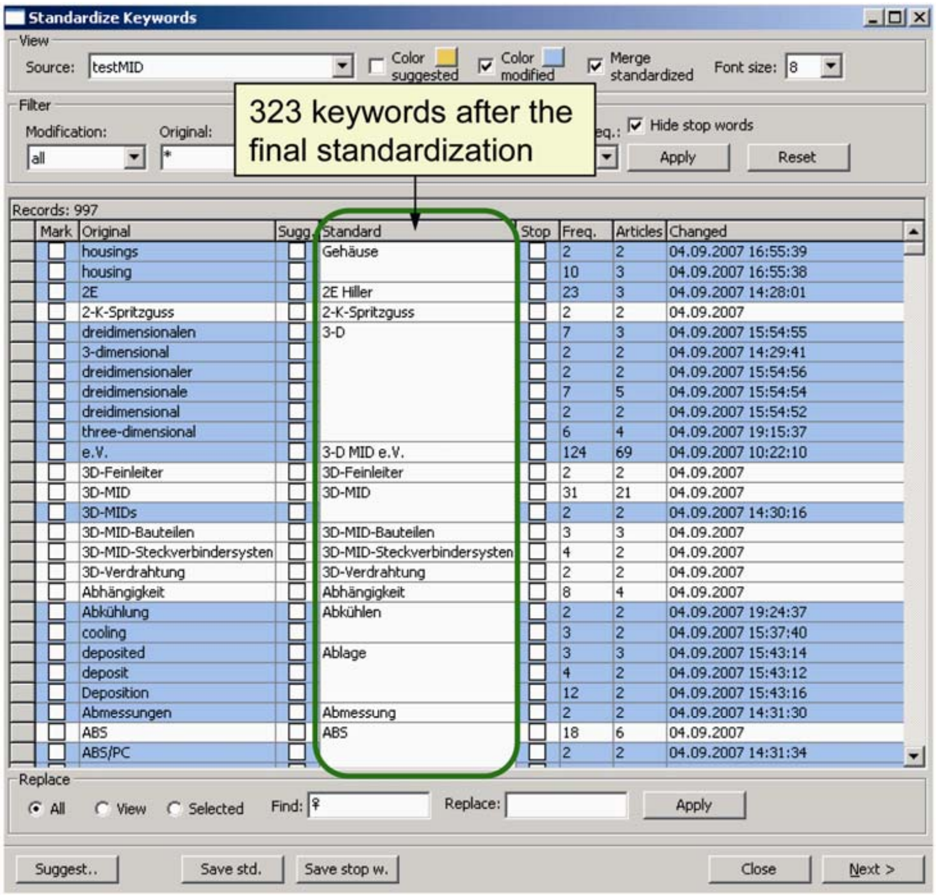


Fig. 5-10: Result after the standardization of keywords

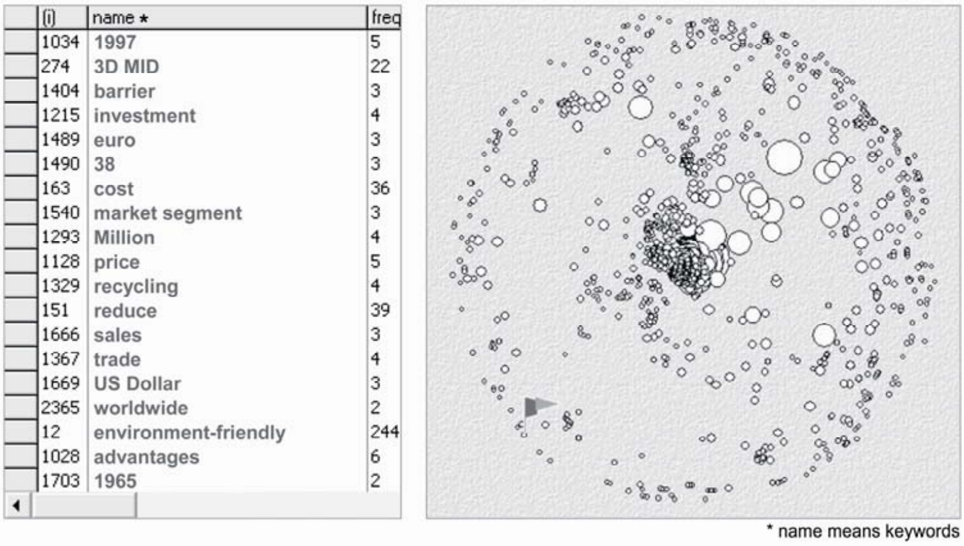
With the aid of BibTechMon, the co-occurrences of keywords and the Jaccard Index were calculated. Both of the matrixes are illustrated in Fig. 5-11.

Global Text Frequencies and Co-Frequencies of Keywords	Gehäuse	2E Hiller	2-K-Spritzguss	3-D MID e.V.	3D-MID-Bauteilen	3D-MID-Steckverbindersystems	3D-Verdrahtung
Gehäuse	4	1	2	4	3	0	0
2E Hiller	1	3	1	2	3	0	1
2-K-Spritzguss	2	1	2	2	1	0	1
3-D MID e.V.	4	2	2	69	3	2	1
3D-MID-Bauteilen	3	3	1	3	3	1	1
3D-MID-Steckverbindersystems	0	0	0	2	1	2	0
3D-Verdrahtung	0	1	1	1	1	0	2

Jaccard Index	Gehäuse	2E Hiller	2-K-Spritzguss	3-D MID e.V.	3D-MID-Bauteilen	3D-MID-Steckverbindersystems	3D-Verdrahtung
Gehäuse	4	0.167	0.5	0.058	0.75	0	0
2E Hiller	0.167	3	0.25	0.457	1	0	0.25
2-K-Spritzguss	0.5	0.25	2	0.029	0.25	0	0.333
3-D MID e.V.	0.058	0.457	0.029	69	0.043	0.029	0.014
3D-MID-Bauteilen	0.75	1	0.25	0.043	3	0.25	0.25
3D-MID-Steckverbindersystems	0	0	0	0.029	0.25	2	0
3D-Verdrahtung	0	0.25	0.333	0.014	0.25	0	2

Fig. 5-11: Upper: matrix of co-word analysis; lower: matrix of Jaccard Index

Based on the text frequencies, co-occurrences, and Jaccard Index calculated, the keywords were visualized by BibTechMon in a Knowledge Map (see Fig. 5-12).



(i)	name *	freq
1034	1997	5
274	3D MID	22
1404	barrier	3
1215	investment	4
1489	euro	3
1490	38	3
163	cost	36
1540	market segment	3
1293	Million	4
1128	price	5
1329	recycling	4
151	reduce	39
1666	sales	3
1367	trade	4
1669	US Dollar	3
2365	worldwide	2
12	environment-friendly	244
1028	advantages	6
1703	1965	2

Fig. 5-12: The general Knowledge Map of MID case study

Following the methodology proposed in chapter 4, the next step was to compare the keywords with the TI-Ontology. In the case study of MID, 15 Raw Technology Indicators were extracted after the Co-word Analysis: e.g. application, function, cost, customer, supplier, barrier, and trend. The Raw Technology Indicators extracted from both Co-word Analysis and Publication Analysis were documented and should be assigned with contents and values in the next phase.

5.1.4 Phase of Concretization of Raw Technology Indicators

As defined in the methodology, the Technology Indicators should be completed with contents and values in this phase. The assignments of contents and values depend on the interpretation of publication diagrams and the Knowledge Map achieved above.

Interpretation of Publication Diagrams

In the case study, the diagrams achieved from Publication Analysis were firstly interpreted in order to assign values and contents to those four identified Raw Technology Indicators. We clarify this process by explaining two diagrams as examples. The Raw Technology Indicator “intensity of development” was identified from Fig. 5-5, which was compared with the interpretation model shown in Fig. 4-14 in this phase. The features of curve in Fig. 5-5 matched the statements of Stage 1 and Stage 2 (see section 4.6.1). Therefore, the contents were assigned to the “intensity of development” as follows (Fig. 5-13): the MID technology was firstly mentioned in 1965. During the 1990s, it was firstly attached with importance. From 2000 until now, the MID technology is at the stage of intensive development. It has attracted the interests of many researchers and companies. A lot of time and energy have been invested in its corresponding development. The MID technology is intensively researched and is developing quickly.

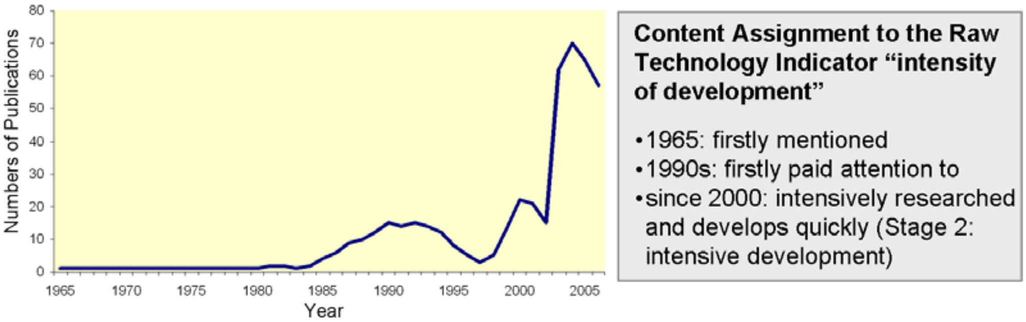


Fig. 5-13: Content assignment to the Raw Technology Indicator "intensity of development"

Fig. 5-14 shows the interpretation of Fig. 5-6, from which the Raw Technology Indicator "key player organizations" was identified. As mentioned previously, the distribution of publication amount according to organizations indicates the organizational research performance. Therefore, the most active organization in the field of MID technology in Germany is 3-D MID e.V., which has the maximum number of publications related to MID. After 3-D MID e.V., the second most active organization of MID is the workgroup LKT from the University of Erlangen-Nuremberg, and then, the workgroup FAPS, the firm BGS, LSP, Schaal, and so on (see Fig. 5-14).

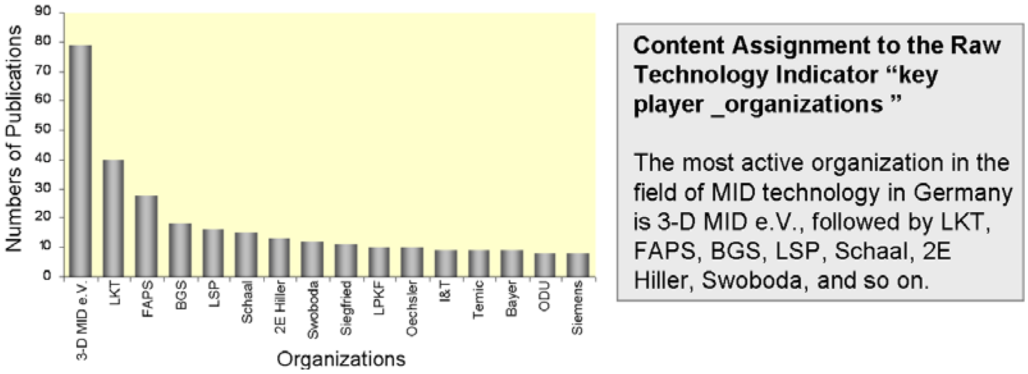


Fig. 5-14: Content assignment to the Raw Technology Indicator "key player organizations"

Interpretation of the Knowledge Map

After the interpretation of publication diagrams, it was time to explain the Knowledge Map of MID achieved by using Co-word Analysis. As discussed before, the contents of the Technology Indicators are represented either qualitatively or quantitatively. In order to verify the process proposed in the methodology in section 4.6.2, two examples are introduced in the following paragraphs: “advantages” with qualitative contents and “sales” with quantitative values.

Guided by the methodology, the Raw Technology Indicator “advantages” that was extracted from the Co-word Analysis was focused. All the strongly co-related keywords to “advantages” were selected and highlighted (see Fig. 5-15). With the help of MID-Ontology (the domain-specific ontology of MID), the connections between keywords on the map were interpreted as follows:

- MID integrates electrical and mechanical parts. The product size is miniaturized. MID enhances agility of construction.
- With MID technology, the number of parts is reduced and hence the material used is reduced. So the second advantage of MID is rationalization.
- The material used by MID are recyclable. MID is an environment-friendly technology.

The interpretations above were assigned to the Raw Technology Indicator “advantages” as its qualitatively represented contents.

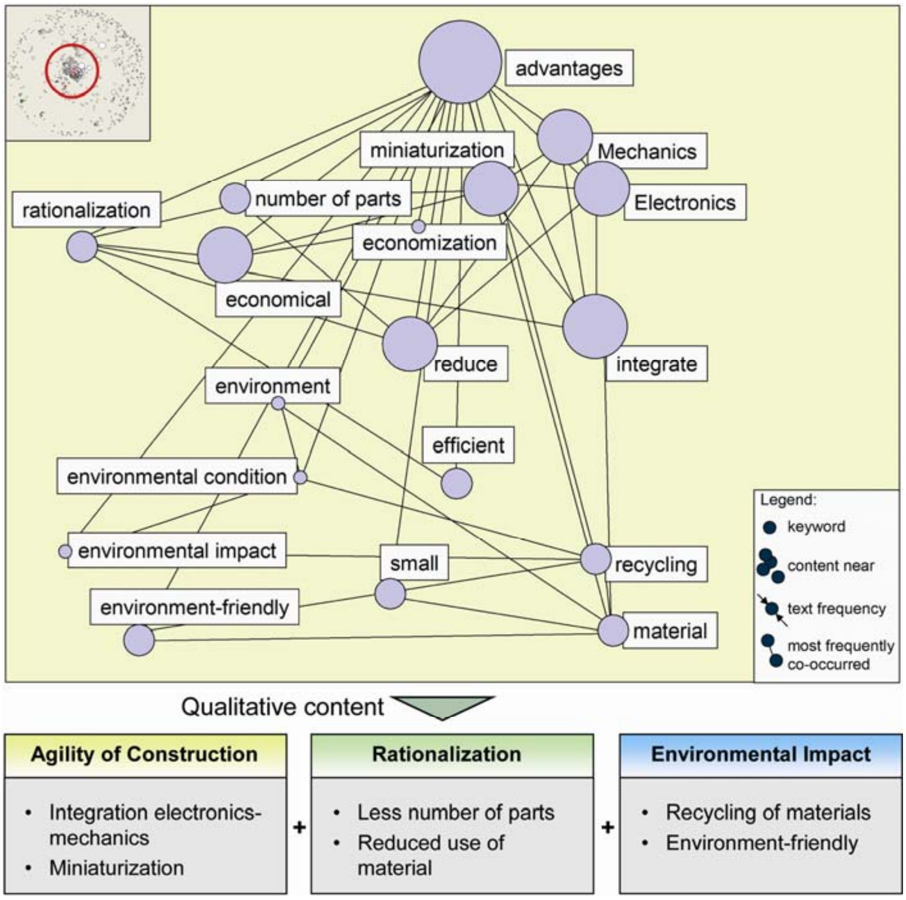


Fig. 5-15: Illustration of the assignment to the Raw Technology Indicator “advantages”

Another example is shown in Fig. 5-16. The targeted Raw Technology Indicator was “sales”. As displayed, the keywords appearing most frequently with “sales” included year, currency units, numbers, etc. However, it was difficult to estimate the numerical values of sales. Therefore, the corresponding segments of texts were recalled with the “sales” highlighted. The texts recalled helped to understand the connections among “sales” and its co-related keywords. With respect to the original texts, the connections were interpreted. Summarizing the interpretations, “sales” was described as follows:

- Sales on MID in 1997 was 38 million dollars world wide.
- There are only few companies, whose sales are more than one million euros.

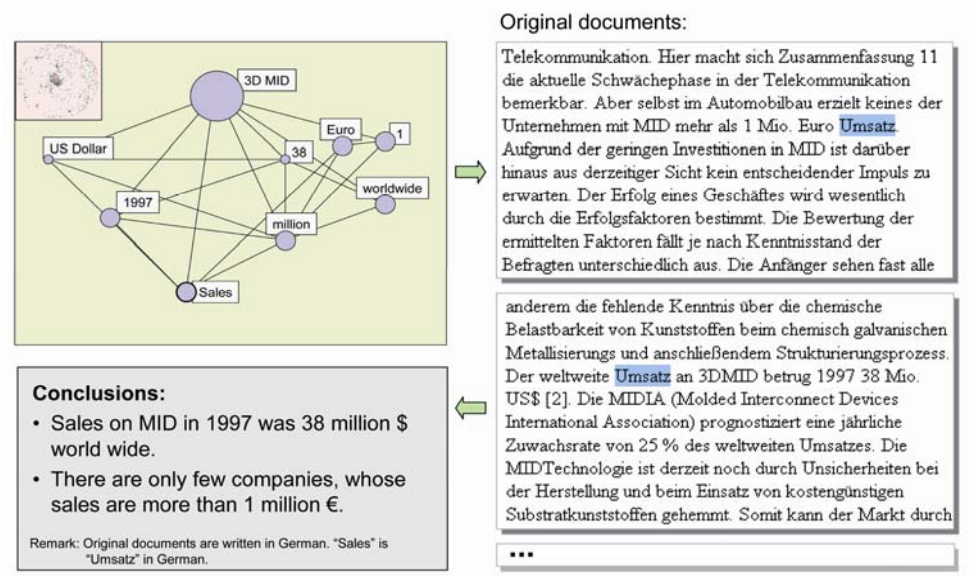


Fig. 5-16: Illustration of the assignment to the Raw Technology Indicator “Sales”

Similarly, the other Raw Technology Indicators identified by using Co-word Analysis were also completed with contents and values by interpreting the co-relations in the Knowledge Map. According to the methodology, the next step was to analyze the dynamic changes of MID.

For instance, the Knowledge Map was divided into 2 parts: one with all keywords that appeared from 1965 to 1999; the other one with all keywords that appeared in the last six years, i.e. from 2000 to 2006. The two parts of Knowledge Map were compared with each other. The keywords newly appearing were identified. For instance, in the case study of MID, the keyword “handy” (cell phone in English) appeared separately in 2005 and 2006. The keyword “Antenne” (antenna in English) appeared firstly in the MID publication in 2002. Furthermore, the two keywords were positioned near each other. It was estimated that MID technology has been recently used for antenna of the cell phone.

In addition to the analysis of dynamic changes, the marginal keywords were also taken into account. Fig. 5-17 gives an example from the cast study of MID. The marked keyword is “3D-MID-Steckverbindersystem” (connector system with 3D-MID). As directed by the methodology, the marginal keywords may indicate the emerging research direction or research vacancy. With respect to Fig. 5-17, it is estimated that the connector system may be a new application field for MID technology, which will be paid more attention to in the next few years.

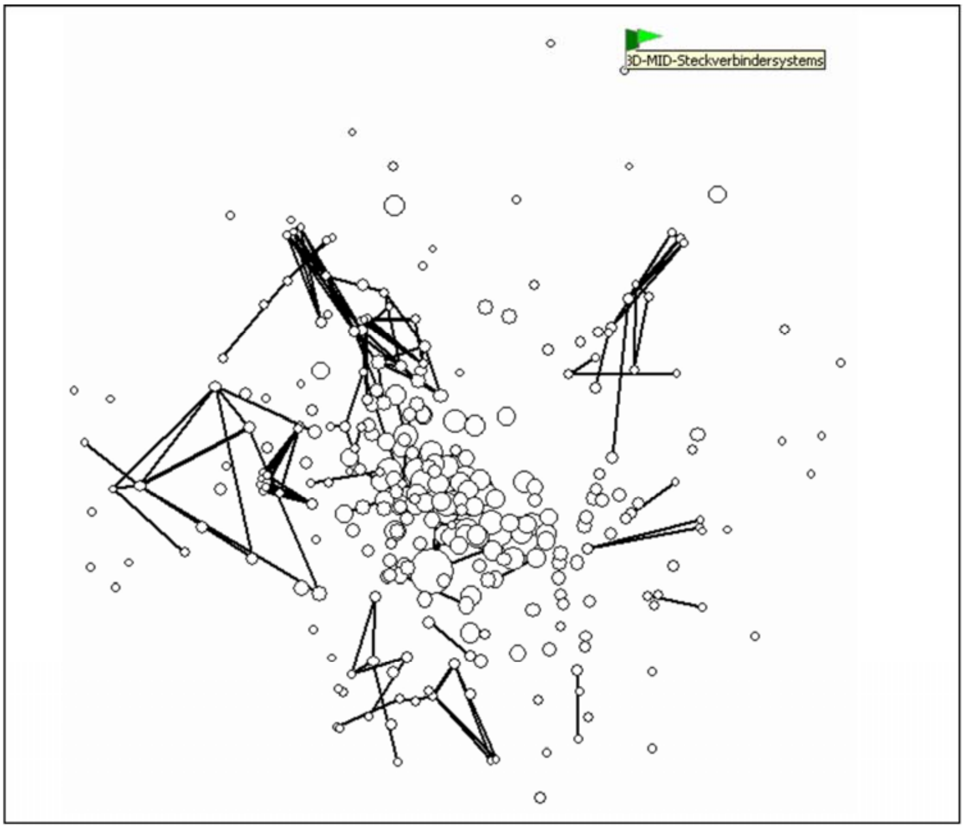
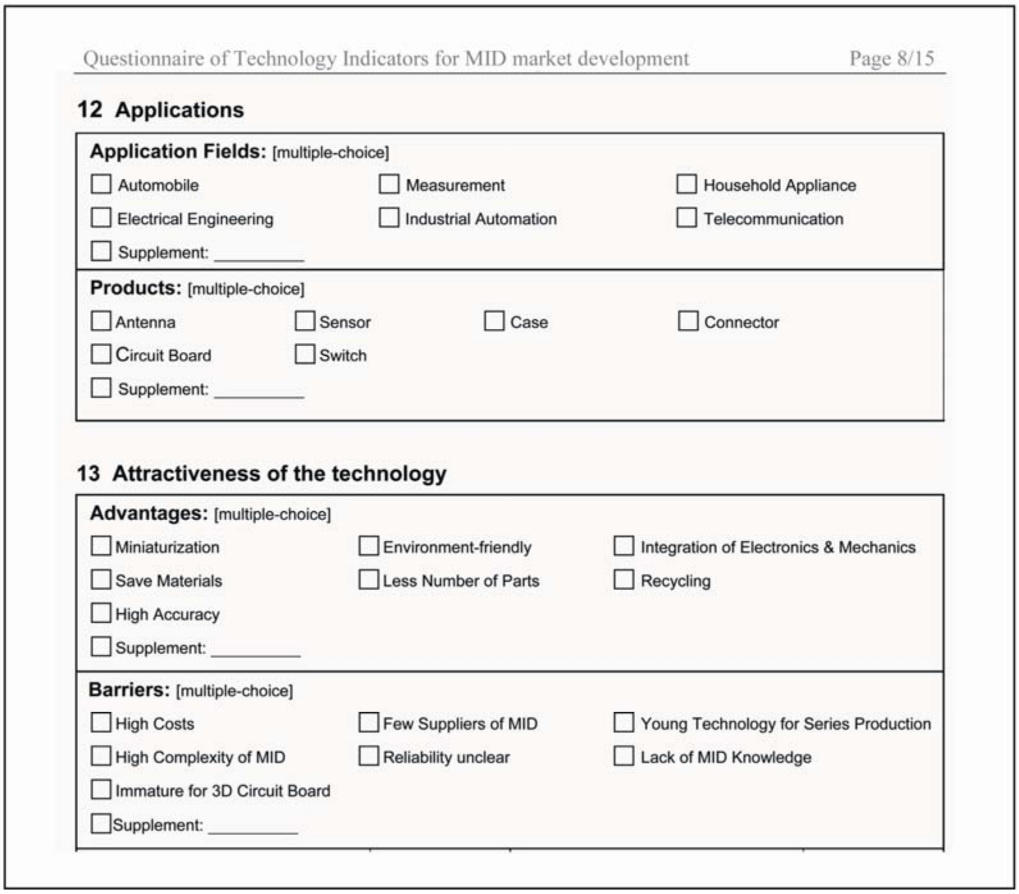


Fig. 5-17: Illustration of the analysis of marginal keywords within the Knowledge Map generated from the keywords appeared from 2000 to 2006

After the interpretations about publication diagrams and the Knowledge Map, the Raw Technology Indicators identified in the last phase were completed with contents and values. The result of this phase was a list of the Complete Technology Indicators with MID-specific contents and values.

5.1.5 Phase of Expert Consultation

Following the methodology, the Technology Indicators extracted by using quantitative analyses should be evaluated in this phase from the qualitative perspective. The method chosen was expert consultation. Based on the results of Complete Technology Indicators of MID, the questionnaires were constructed in a short order. Fig. 5-18 gives an overview of a segment from the questionnaires used in the case study of MID.



Questionnaire of Technology Indicators for MID market development Page 8/15

12 Applications

Application Fields: [multiple-choice]

☐ Automobile ☐ Measurement ☐ Household Appliance

☐ Electrical Engineering ☐ Industrial Automation ☐ Telecommunication

☐ Supplement: _________

Products: [multiple-choice]

☐ Antenna ☐ Sensor ☐ Case ☐ Connector

☐ Circuit Board ☐ Switch

☐ Supplement: _________

13 Attractiveness of the technology

Advantages: [multiple-choice]

☐ Miniaturization ☐ Environment-friendly ☐ Integration of Electronics & Mechanics

☐ Save Materials ☐ Less Number of Parts ☐ Recycling

☐ High Accuracy

☐ Supplement: _________

Barriers: [multiple-choice]

☐ High Costs ☐ Few Suppliers of MID ☐ Young Technology for Series Production

☐ High Complexity of MID ☐ Reliability unclear ☐ Lack of MID Knowledge

☐ Immature for 3D Circuit Board

☐ Supplement: _________

Fig. 5-18: A segment of the questionnaire for MID investigation

The questionnaires were sent to 22 MID experts in Germany, of which 12 experts filled in the questionnaires and sent them back to us. The opinions of the experts were statistically analyzed. The results of the questionnaires were compared with those Complete Technology Indicators. Some of the Complete Technology Indicators and their contents were confirmed; some of them were corrected; and some of them were supplemented. Fig. 5-19 illustrates the comparison results of the Complete Technology Indicator "advantages". Experts agreed with the advantages of MID technology that were extracted by using Co-word Analysis. Furthermore, they gave another six points to supplement the advantages. In summary, the Technology Indicator "advantages" was finalized with MID-specific contents. The result of this phase was a list of the Final Technology Indicators, which can describe, characterize, and evaluate the MID technology. A segment of the list of Final Technology Indicators identified in the case study of MID is demonstrated in Fig. 5-20.

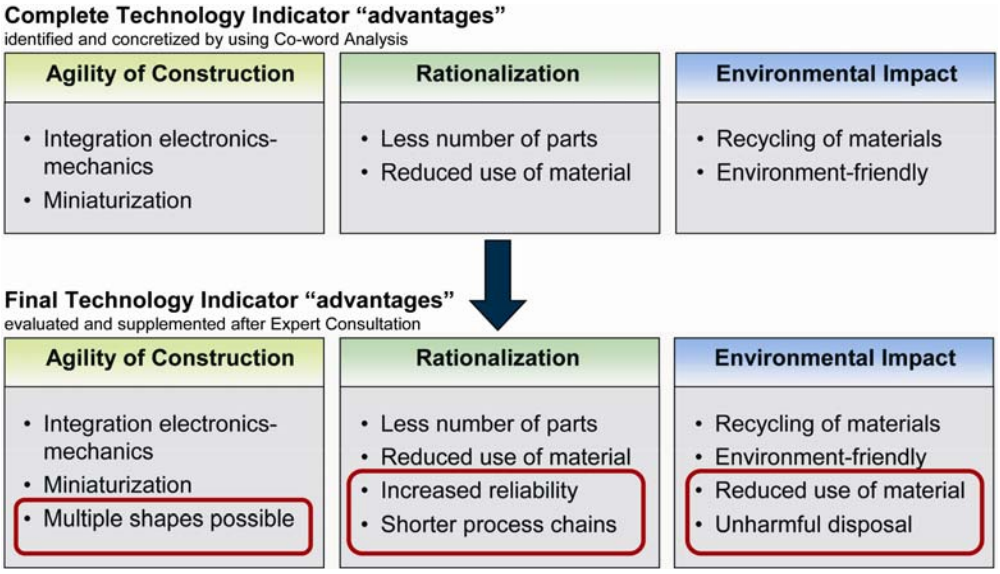


Fig. 5-19: Finalization the Technology Indicator "advantages" with respect to the opinions of MID experts

Documentation of Final Technology Indicators for MID technology

- **Intensity of development:** the MID technology was firstly mentioned in 1965. During the 1990s, it was firstly attached with importance. From 2000 until now, the MID technology is at the stage of intensive development. It has caused the interests of many researchers and companies. A lot of time and energy have been invested in its corresponding development. The MID technology is intensively researched and develops quickly.
- **Key player_regions:** Japan is the most active countries in the MID history, followed by Germany and USA.

...

8. **Advantages:** integrate electronics and mechanics; save construction space; possible for multiple shapes; reduce the number of parts; save materials; increase the reliability; speed up the production chains; environment-friendly; use the recyclable materials; realize unharmful disposal
9. **Barriers:** immature for the manufacturing of 3D circuit board; is an emerging technology, there is a lack of technical know-how; only a few suppliers of MID available; less applications; …

...

Fig. 5-20: A segment of the list of Final Technology Indicators for MID technology

Regular Update

In the case study, the regular update was not tested. However, the documents relevant to MID technology were saved in the innovative Technology Database. The stop words used and the modifications of keywords during the standardization of keywords were saved. The periodical Knowledge Maps and all the interpretations also were recorded. It is understandable and imaginable that those documentations will support and hence simplify the further investigation of MID technology.

To sum up the case study of MID technology, the methodology for the identification of Technology Indicators has proven feasible. In the whole case study, the quantitative analysis only has taken several days; while the qualitative analysis took at least three weeks from the construction of questionnaires to the achievement of experts' feedbacks. The case study totally followed the processes of the methodology, and hence a lot of time and human resources have been saved. The final results of the case study were that the Final Technology Indicators specific for MID technology. The Final Technology Indicators were input into the innovative Technology Database. Based on those indicators, the Technology Report of MID was automatically generated from the database (see Fig. 5-21). Furthermore, through querying the function connections between technologies and applications, MID was found as a potentially suitable technology for an internal project "mini robot" in our workgroup. The investigation objectives of the case study were basically fulfilled.

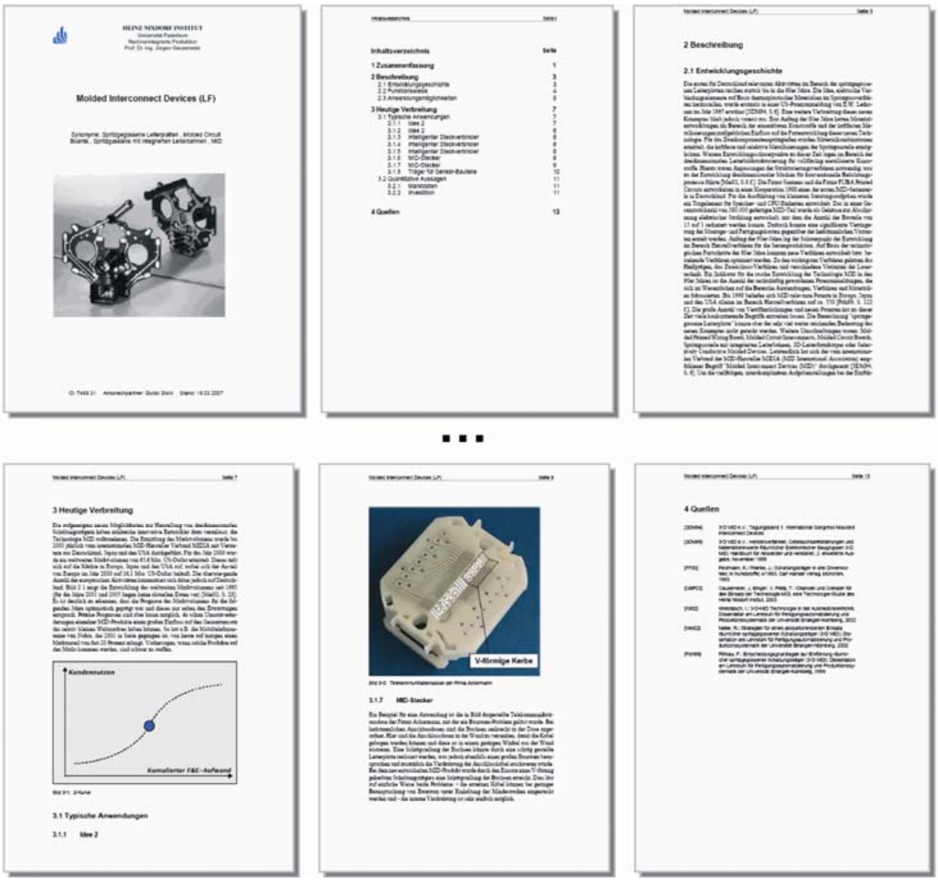


Fig. 5-21: The Technology Report of MID, generated automatically from the innovative Technology Database

5.2 Evaluation of the Methodology

In this section, the methodology proposed in the dissertation is evaluated according to the requirements enumerated in section 2.4. It is commented, to what extent every single requirement is satisfied. All the following comments and evaluations are based on the case studies for the methodology.

R1: capability of dealing with a vast amount of information

As explained previously, the raw information is plenty enough to guarantee a comprehensive analysis of the target technology. In the methodology for the identification of Technology Indicators, the methods of Information Retrieval and Bibliometric Analysis facilitate the search and quantitative analyses for a large amount of information. As demonstrated in the case study of MID technology, infinite online databases were searched and 489 documents with 4668 pages relevant to MID technology were retrieved. The documents were then statistically analyzed by using Bibliometrics in one turn. It is verified that the first requirement is totally fulfilled.

R2: (semi-)automatic analysis of the information

Based on the first requirement, the amount of information is so huge that it is no longer possible to process it manually. Therefore, the second requirement is to use (semi-)automatic process instead of the traditional manual work. In the methodology proposed, most of the processes are realized automatically. For instance, the search for raw information relevant to target technology is computer-aided; the transformation from various document formats to text format is automated; the decomposing of texts into words carries out mechanically; the text frequencies, co-occurrences of keywords, as well as the Jaccard Index are calculated by computer programs; and the visualization of keywords in the Knowledge Map is also automatically done. The points mentioned above were proved in the case study of MID. 70% of the processes of the information procurement in the methodology have been automatized. It is almost realized to know without reading. The second requirement is to a great extent fulfilled.

R3: high efficiency and accuracy

First of all, the realization of automatic information search and analysis of a vast amount of information saves a lot of time and human resources, which verifies that the methodology proposed in the dissertation is very efficient. The following corresponding measurements are taken in the methodology in order to ensure and increase the accuracy of the results achieved:

- select publications forms according to investigation objectives;
- reduce the search areas by limiting suitable information repositories;
- use Boolean Logic to optimize the search queries;

- use ontology to avoid misunderstanding by the interpretation of the Knowledge Maps;
- use Expert Consultation to ensure the results from qualitative perspective.

Furthermore, the extracted Technology Indicators and their contents and values support and simplify the construction of the questionnaires, which reduces the time of normal process of the Expert Consultation. Therefore, it is concluded that the methodology has reached a relative high efficiency and accuracy compared with the other traditional methods.

R4: standardized process to procure information relevant to technologies

There is a lack of methods that can guide a decision maker in companies or researchers to discover the knowledge of technologies. A standard guide is irresistible. The methodology for the identification of Technology Indicators introduces a systematical way from problem analysis to the update of the Technology Indicators. Every single step is defined and explained explicitly (chapter 4). Moreover, the fixed TI-Ontology offers a standard reference to the investigation directions. The decision makers or other users need only to follow the guide of the methodology step by step. Hence, the fourth requirement is fully satisfied.

R5: simple update process of information procurement

The methodology aims at simplifying the whole process of information procurement. Logically, it requires a simple update process. Documentation is an important factor to this point, which makes the most useful processing, raw technology information and Technology Indicators extracted reusable. The access to the past data facilitates the update process by reducing the workload. To give an example, the process of standardization of keywords is saved. If new documents are added, only the newly extracted keywords need to be standardized. Furthermore, the update of Technology Indicators follows the processes defined in the methodology. The fifth requirement is satisfied.

The general requirements

As discussed in section 2.4, the information should be analyzed from both quantitative and qualitative perspectives. The usage of Information Retrieval and Bibliometrics ensures the quantitative analysis. The implementation of Expert Consultation evaluates the results after quantitative analysis from the qualitative aspect. Therefore, it is commented that the methodology satisfies the general requirement of information procurement.

6 Summary and Outlook

Summary

Technologies influence the product development and the production process significantly. For technology-intensive companies, especially in automobile industry and mechanical engineering branch, technologies have become a key factor for competitiveness. Therefore, it is important to identify the key characteristics of technologies such as advantages or barriers, to compare them as well as to analyze the technology trend in order to grasp the current development of technologies, and to develop suitable technology strategies, which leads to an urgent demand of researchers and decision makers on the extraction of technology-relevant knowledge.

There is a vast amount of information available. However, decision makers are not well informed about technologies. There are three basic problems of the information procurement in the context of technology monitoring. First of all, there is no clear definition of the technology-relevant information. Secondly, the development of communication and information technology has resulted in a dramatic increase of the amount of information in recent years, most of which is digitally available. It is no longer possible to process that much information manually. The last problem is the lack of methodology, which can guide decision makers in companies to search and analyze the technology information that they desire.

Considering the first challenge mentioned above, a new term “Technology Indicator” is defined. Technology Indicators are those indexes or statistical data, which allow direct characterization and evaluation of technologies throughout their whole life cycles. For example, technological maturity, market segment, degree of innovation, key player (country, company…). Those Technology Indicators offer a direct view of technologies to decision makers. To cover the second and the third open issues, a methodology for the (semi-)automatic identification of Technology Indicators is proposed in this dissertation.

The proposed methodology is based on the combination of four basic methods: Information Retrieval, Bibliometric Analysis, Ontology, and Expert Consultation. The four methods chosen are combined in such a standard process to give a guide to decision makers for the information procurement. The start point is to analyze the requirements of technology investigation. Input of this methodology is a large amount of raw information retrieved by using Information Retrieval. Then the information is pre-processed, decomposed into words, standardized, and statistically analyzed. A Technology Indicator Ontology is developed in this dissertation, with which the Raw Technology Indicators can be easily identified. The Raw Technology Indicators are then concretized with contents and values by interpreting the co-relationships of the keywords and other results after the statistical analysis. Finally, the Technology Indicators are evaluated by experts through Expert Consultation. The result of the methodology is a list of identified Technology Indicators with their technology-specific contents and values. It is noticed that the methodology also facilitates a simple process of regular update in order to catch up with the current changes of technology.

The methodology is integrated in the innovative Technology Database, which is developed by Heinz Nixdorf Institute and aims at supporting the product innovation process (see section 2.2). The Technology Indicators extracted are input into the database together with other relevant information (e.g. figures, or other general information written in continuous text). On the output side of the database, those indicators are visualized in formats of Technology Reports and Technology Roadmaps, which are automatically generated from the Technology Database. The Technology Report is constructed in a fixed structure and represents detailed information of technologies. The Technology Roadmap is a plan that shows which technology can be used in which products at what time. Both of them help decision makers to know technologies better and to speed up their decision-making process.

The methodology proposed for identification of Technology Indicators has proven feasible in several case studies. It combines quantitative analysis and qualitative analysis to make the results more reliable and accurate. It standardizes the procedure of information procurement and consequently guides decision makers to simplify information processing processes. With the methodology, it is possible to search, process, and analyze a huge amount of information in one turn. Furthermore, the methodology realizes semi-automatic analysis of literature for the purpose of investigation of technologies and facilitates a trivial update process.

To sum up, the methodology fulfils the requirements of the information procurement to a great extent. Based on the case study, it is also convincible that the methodology is suitable for practical application.

Future Work

Although the methodology has been proved efficient and reasonable, there are still some works to do to optimize it. The future work orients itself to the following points:

- Pre-processing of the original documents: it is noticed in the case study that the pre-processing of the documents retrieved is one of the most taxing tasks. One of the solutions is to build databases, in which the documents are already edited in a fixed meta-structure and therefore directly ready for Bibliometric Analysis.

- Optimizing the interaction of ontology and the methodology: the case study shows that the accuracy of the investigation results depends strongly on the extraction of keywords and the interpretation of the Knowledge Maps. As discussed in this dissertation, ontology helps to avoid the misunderstanding of domain-specific information. Therefore, the future work should focus on the development of a better interface between ontology and the methodology in order to simplify the process of knowledge extraction, and to increase the accuracy of the results by eliminating confusing understandings.
- Analysis of figures and tables: the current methods only facilitate text analysis. However, most of the numeric data such as sales, investments, or important technological data, are covered by figures and tables. It is expected to add corresponding methods to the methodology, which realize the knowledge extraction from other information format.

The employment of the methods-combination of Information Retrieval, Bibliometrics, Ontology, and Expert Consultation is a very promising beginning. It is worthy of being further researched.

7 Bibliography

[Ala99] ALAHUHTA, M.: Wireless Information Society. The speech of the President of Nokia Mobile Phones, Oulu, Finland, 1999

[Ans66] ANSOFF, H. I.: Management-Strategie. Verlag Moderne Industrie AG, München, 1966

[BC06] BRUESEKE, U.; CHANG, H.: Einsatz bibliometrischer Analysen in der strategischen Frühaufklärung. In: GAUSEMEIER, J. (Hrsg.): Vorausschau und Technologieplanung. 2. Symposium für Vorausschau und Technologieplanung Heinz Nixdorf Institut, 9. - 10. November 2006, Schloß Neuhardenberg, HNI-Verlagsschriftenreihe, Band 198, Paderborn, 2006

[Bel78] BELLMAN, R.E.: An Introduction to Artificial Intelligence: Can Computers Think? Boyd & Fraser Publishing Company, San Francisco, USA, 1978

[BFO07] BFO: Basic Formal Ontology. Online under: http://www.ifomis.uni-saarland.de/bfo/, Institute for Formal Ontology and Medical Information Science, University of Saarland, Germany, 2007

[BG05] BRUESEKE, U.; GAUSEMEIER, J.: Employment of Bibliometric Analysis in the Strategic Early-Warning. In: 1ST IFIP TC5 Working Conference on Computer aided Innovation, Ulm, Germany, p. 31-41, November 14-15 2005

[Bir80] BIRKHOFER, H.: Analyse und Synthese der Funktionen technischer Produkte. Dissertation, Fakultät für Maschinenbau und Elektrotechnik, TU Braunschweig, VDI-Verlag, Düsseldorf, 1980

[BL97] BERRY, M. J. A.; LINOFF, G. S.: Data Mining Techniques. Wiley, New York, USA, 1997

[BR99] BAEZA-YATES, R.; RIBEIRO-NETO, B.: Modern Information Retrieval. Addison-Wesley, 1999

[Bre03] BREAUX, T.: Information Analysis Using Upper Ontologies. Oak Ridge National Laboratory in Oak Ridge, Tennessee, USA, 2003

[Cyc07] OpenCyc Selected Vocabulary and Upper Ontology. Online under: http://www.cyc.com/cyc/opencyc/, Cycorp, Inc., USA, 2007

[DB95] DAVIS, S.; BOTKIN, J.: The monster under the bed - How business is mastering the opportunity of knowledge for profit. New York at al.: Simon & Schuster, 1995

[DF02] DING, Y.; FOO, S.: Ontology Research and Development Part 1 – A Review of Ontology Generation. Journal of Information Science, Vol. 28, No. 2, 123-136, 2002

[Eib99] EIBL, M.: Recherche in elektronischen Bibliothekskatalogen. Spektrum der Wissenschaft Dossier 2/99 „Software“, Spektrum der Wissenschaft Verlagsgesellschaft, Heidelberg, S. 76-79, 1999

[Eve02] EVERSHEIM, W. (Hrsg.): Innovationsmanagement für technische Produkte. Springer-Verlag, Berlin, Heidelberg, 2002

[FMN88] FEIGENBAUM, E. A.; MCCORDUCK, P.; Nii, H. P.: The Rise of the Expert Company: How Visionary Companies are Using Artificial Intelligence to Achieve Higher Productivity and Profits. Times Books, New York, 1988

[FS06] Feldman, R.; Sanger, J.: The Text Mining Handbook: Advanced Approaches in Analyzing Unstructured Data. Cambridge University Press, USA, 2006

[Gar99] GARFIELD, E.: A Citation Analyst's Perspective on Japanese Science, ISI Symposiums in Osaka and Tokyo, 1999

[Gau98] GAUTHIER, E.: Bibliometric analysis of scientific and technological research: a user's guide to the methodology. No. 8 Science and Technology Redesign Project Statistics, Canada, September 1998

[GEK01] GAUSEMEIER, J.; EBBESMEYER, P.; KALLMEYER, F.: Produktinnovation - Strategische Planung und Entwicklung der Produkte von morgen. Carl Hanser Verlag, München, Wien, 2001

[Gen99] GENTSCH, P.: Wissen managen mit innovativer Informationstechnologie – Strategien, Werkzeuge, Praxisbeispiele. Gabler Verlag, Wiesbaden, 1999

[GCI+07] GAUSEMEIER, J.; CHANG, H.; IHMELS, S.; WENZELMANN, C.: A Technology Management System to foster Product Innovation. 16th International Conference on Management of Technology, (IAMOT 2007), Miami Beach, May 13-17 2007, USA

[GG00] GROTHE, M.; GENTSCH, P.: Business Intelligence – Aus Informationen Wettbewerbsvorteile gewinnen. Addison-Wesley Verlag, München, 2000

[GHK+06] GAUSEMEIER, J.; HAHN, A.; KESPOHL, H.D.; SEIFERT, L.: Vernetzte Produktentwicklung- Der erfolgreiche Weg zum Global Engineering Networking. Carl Hanser Verlag, Munich, 2006

[GK97] GOODMAN, D.; KEENE, R.: Man versus Machine: Kasparov versus Deep Blue. H3 Publications, Cambridge, Massachusetts, USA, 1997

[GPW07] GAUSEMEIER, J.; PLASS, C.; WENZELMANN, C: Strategisches Produktionsmanagement. Carl Hanser Verlag, München, Germany, 2007

[Gla03] GLAENZEL, W.: Bibliometrics as a research field. A course on theory and application of Bibliometric indicators, 2003

[God96] GODIN, B.: The state of science and Technology Indicators in the OECD countries, research paper, Science and Technology Redesign Project, Statistics Canada, 1996

[Gor92] GORRAIZ, J.: Zitatenanalyse - Die unerträgliche Bedeutung der Zitate. In: Biblos. Jg. 41, H. 4, S.193-204, 1992

[Gor05-ol] GORRAIZ, J.: Szientometrie: Zitatenanalyse. In: Skriptum für Informationswissenschaft und -theorie, Teil II, Fachhochschulstudiengänge Burgenland: Fachhochschulstudiengang - Informationsberufe. 2005, unter http://www.zbp.univie.ac.at/gj/citation/skriptum2neu.htm, im Januar 2006

[Gru93a] GRUBER, T.R.: Towards Principles for the Design of Ontologies used for Knowledge Sharing. In: Formal Ontology in Conceptual Analysis and Knowledge Representation. Hrsg. Guarino, N.; Poli, R., Deventer, Kluwer Academic Press, 1993

[Gru93b] GRUBER, T.R.: A translation approach to portable ontology specifications. Knowledge Acquisition 5, 1993

[Gua97] GUARINO, N.: Understanding, building, and using Ontologies. (A commentary to "Using Explicit Ontologies in KBS Development" by van Heijst, Schreiber and Wielinga) In: International journal of human – Computer studies Elsevier, Amsterdam, 1997

[Gua98] GUARINO, N.: Some Ontological Principles for Designing Upper Level Lexical. Resources: Proc. of the First International Conference on Lexical Resources and Evaluation, Granada, Spain, 28-30 May 1998

[GW05] GAUSEMEIER, J.; WENZELMANN, C.: Auf dem Weg zu den Produkten für die Märkte von morgen. In: GAUSEMEIER, J. (Hrsg.): Vorausschau und Technologieplanung. 1. Symposium für Vorausschau und Technologieplanung Heinz Nixdorf Institut, 3. - 4. November 2005, Schloß Neuhardenberg, HNI-Verlagsschriftenreihe, Band 178, Paderborn, 2005

[Häd02] HÄDER, M.: Delphi-Befragungen. Westdeutscher Verlag, Wiesbaden, Germany, 2002

[HK06] HAN, J.; KAMBER, M.: Data mining – concepts and techniques. 2nd edition, Morgan Kaufmann Verlag, 2006

[HP04] HELLER-SCHUH, B.; PRETSCHUH, J.: Zur Funktion der Wissensorganisation bei der Auswahl nachhaltiger Entwicklungsstrategien in Kompetenznetzwerken (Powerpoint Präsentation). 9. Tagung der Deutschen ISKO, Wissensorganisation 2004, Duisburg, 5-7. 11. 2004

[HQW06] HEYER, G.; QUASTHOFF, U.; WITTIG, T.: Text Mining: Wissensrohstoff Text – Konzepte, Algorithmen, Ergebnisse. W3L-Verlag, Herdecke, Bochum, Germany, 2006

[Hwa99] HWANG, C. H.: Incompletely and imprecisely speaking: Using dynamic ontologies for representing and retrieving information. In: Proceedings of the 6th International Workshop on Knowledge Representation meets Databases (KRDB'99), Linköping, Sweden, July 29-30, 1999

[IP06-ol] INFORMATION PROCUREMENT. Free Patent Online, under: http://www.freepatentsonline.com/20060112078.html, 2006

[JMM+00] JONSSON, A.; MORRIS, P.; MUSCETTOLA, N.; RAJAN, K.; SMITH, B.: Planning in interplanetary space: Theory and practice. In: Proceedings of the 5th International Conference on Artificial Intelligence Planning Systems (AIPS-00), p. 177-186, Breckenridge, Colorado. AAAI Press, 2000

[KBH99] KUKAL, Z.; BLAŽKA, M.; HRONEK, F.; et al: Analysis of previous trends and existing state of research and development in the Czech Republic and a comparison with the situation abroad. © Ministry of Education, Youth and Sport and Research and Development Council of the Government of the Czech Republic, Prague, Czech Republic, May 1999

[KFG+06] KAISER, I.; FRANK, U.; GAUSEMEIER, J.; POOK, S.: Design of a spatial electronic circuit carrier by the example of a miniature robot. In: Proceedings of IMECE 2006, Paper 14531. November 5-9th, Chicago, Illinois, 2006

[Kin87] KING J.: A review of bibliometric and other science indicators and their role in research evaluation. In: J. inform Sci. 13:261-76, 1987

[Kin05] KINKEL, S.: Zukünftige Herausforderungen für die deutsche Werkzeugmaschinenindustrie – Ergebnisse einer Mini-Delphi-Studie. In: Newsletter Nr.3 Ergebnisse des Mini-Delphi. Das Begleitvorhaben "WZM-Initiative 20XX" wird mit Mitteln des BMBF im Rahmenkonzept „Forschung für die Produktion von morgen" gefördert und vom Projektträger Produktion und Fertigungstechnologien, Forschungszentrum Karlsruhe, betreut. Germany, 2005, online under: http://www.wzm-initiative.de/public/?seite=veroeffentlichungen&us=projektdoku&id=11&PHPSESSID=3e846e8e59ac8fc0f7befdca26b29738

[Kli03] KLIMESCH, C.: Ein Beitrag zur prozessgetriebenen Informationslogistik durch kontextorientiertes domänenübergreifendes Wissensmanagement. Shaker Verlag, Aachen, 2003

[KMS93] KRYSTEK, U.; MÜLLER-STEWENS, G.: Frühaufklärung für Unternehmen. Schäffer-Poeschel Verlag, 1993

[KS95] KOPCSA, A.; SCHIEBEL, E.: Methodisch-theoretische Abhandlung über bibliometrische Methoden und ihre Anwendungsmöglichkeiten in der industriellen Forschung und Entwicklung. Endbericht zum Projekt Nr. 3437 im Auftrag des Bundesministeriums für Wissenschaft, Forschung und Kunst. Juli 1995

[KS98] KOPSCA, A.; SCHIEBEL, E.: Science and Technology Mapping: A New Iteration Model for Representing Multidimensional Relationships. In: Journal of the American Society for Information Science, Volume 49, Nr.1. American Society for Information Science, 1998

[KS01] KOPCSA, A.; SCHIEBEL, E.: Contentvisualisierung durch BibTechMon. OCG-Österreichische Computer Gesellschaft, Informatik 2001/Network Economy-Visionen und Wirklichkeit, 2001

[FPS96] FAYYAD, U. M.; PIATETSKY-SHAPIRO, G.; SMYTH, P.: From Data Mining to Knowledge Discovery: An Overview. Advances in Knowledge Discovery and Data Mining, 1996

[Lan00] LANGLOTZ, G.: Ein Beitrag zur Funktionsstrukturentwicklung innovativer Produkte. Dissertation, Rheinisch Westfälische Universität Aachen, Shaker Verlag, Aachen, 2000

[Lin01] LINDEMANN, U.: Methoden in der Produktentwicklung. Konstruktion 53, 1/2, 2001

[Lis01] LISSMANN, U.: Inhaltsanalyse von Texten: Ein Lehrbuch zur computerunterstützten und konventionellen Inhaltsanalyse, 2. aktual. und erweiterte, Landau, 2001

[LKS99] LITTMAN, M.L.; KEIM, G. A.; SHAZEER, N. M.: Solving crosswords with PROVERB. In: Proceedings of the Sixteenth National Conference on Artificial Intelligence (AAAI-99), p. 914-915, Orlando, Florida, USA, AAAI Press, 1999

[LLF+99] LIU, H.; LU, H. J.; FENG, L.; HUSSAIN, F.: Efficient search of reliable exceptions. Methodologies for Knowledge Discoveries and Data Mining, Third Pacific-Asia Conference, PAKDD 99, Beijing, China, April 1999.

[LS03] LITZENBERGER, T.; STERNBERG, R.: Die Forschungsleistung der Wirtschafts- und Sozialwissenschaftlichen Fakultät der Universität zu Köln – ein bibliometrischer Vergleich von Fächern, Fächergruppen und Fakultäten. Working Paper No. 2003-03, Universität Köln, ISSN 1434-3746, 2003

[LV03-ol] LYMAN, P.; VARIAN, H. R.; Dunn, J.; STRYGIN, A.; SWEARINGEN, K.: How Much Information? (Study 2003) University Of California, Berkeley, USA, 2003, under: http://www.sims.berkeley.edu/research/projects/how-much-info/

[MB04] Make or Buy. Beschaffungsmarketing, McGrip Webdesign & Online Marketing, McGrip Webdesign Unternehmen, 2004, online under: http://www.mcgrip.de/0-web/wissen/beschaffungsmarketing/02-1-make-or-buy.htm

[Mcc04] MCCARTHY, J.: What is artificial intelligence? Computer Science Department, Stanford University, USA, November 24, 2004

[MMR+56-ol] MCCARTHY, J.; MINSKY, M. L.; ROCHESTER, N.; SHANNON, C. E.: A proposal for the Dartmouth summer research project on Artificial Intelligence. Retrieved from the World Wide Web: http://www-formal.stanford.edu/jmc/history/dartmouth/dartmouth.html, USA, 1956

[NFS02] NOLL, M.; FRÖHLICH, D.; SCHIEBEL, E.: Knowledge Maps of Knowledge Management Tools – Information Visualization with BibTechMon™. In: Karagiannis, D.; Reimer, U. (Eds.): PAKM 2002, LNAI 2569. Springer Verlag, Heidelberg, 2002

[Nil98] NILSSON, N.J.: Artificial Intelligence: A New Synthesis. Morgan Kaufmann, San Mateo, California, USA, 1998

[Noy05] NOY, N.: Ontology Mapping and Alignment, Stanford University, 2005

[Poc00] POCSAI, Z.: Ontologiebasiertes Wissensmanagement für die Produktentwicklung. Dissertation, Institut für Rechneranwendung und Konstruktion, Universität Karlsruhe, Shaker Verlag, Aachen, 2000

[Pri69] PRITCHARD, A.: Statistical bibliography or Bibliometrics? In: Journal of Documentation 24, p. 348-349, 1969

[Pri76] PRICE, D. d. S.: A general theory of Bibliometric and other cumulative advantage processes. In: Journal of the American Society for Information Science, 27, p. 292–306, 1976

[PRR98] PROBST, G.; RAUB, S.; ROMHARDT, K.: Wissen managen, Frankfurt a.M., 1998

[RA304] Research Association 3-D MID e.V.: 3D-MID Technologie – Räumliche elektronische Baugruppen. Carl Hanser Verlag, München, 2004

[Raa93] VAN RAAN, A.F.J.: Advanced bibliometric methods to assess research performance and scientific development: basic principles and recent practical applications; Paper published in: Research Evaluation Vol 3, No. 3, Seiten 151-166, Beech Tree Publishing, 1993

[Raa01] van RAAN A. F. J.: Bibliometric Analysis as an Instrument for Research Evaluation. Platform Technologie Evaluierung, Nr. 12, Austria, April 2001

[Raa03] van RAAN A. F. J.: The use of Bibliometric Analysis in research performance assessment and monitoring of interdisciplinary scientific developments. "Technikfolgenabschätzung", Nr. 1, 12. p. 20-29, University of Leiden, Netherland, 2003

[Raa04] VAN RAAN A.F.J.: Discovery of patterns of scientific and technological development and knowledge transfer. unter: http://www.cwts.nl/TvR/documents/AvR-CRIS2002Paper.pdf, 27. September 2004

[Reu96] REUTERS, P. J.: Dying for information? An investigation into the effects of information overload in the USA and worldwide, based on research conducted by Benchmark Research. Reuters Limited. London, 1996

[Rij79] VAN RIJSBERGEN, C. J.: Information Retrieval. Information Retrieval Group, University of Glasgow, Butterworths, the second edition, London, 1979

[RK91] RICH, E.; KNIGHT, K.: Artificial Intelligence (second edition). McGraw-Hill, New York, USA, 1991

[RKR+01] RADERMACHER, F. J.; KÄMPKE, T.; ROSE, T.; TOCHTERMANN, K.; RICHTER, T.: Management von nicht-explizitem Wissen: Noch mehr von der Natur lernen – Abschlussbericht Teil 2 Wissensmanagement: Ansätze und Erfahrungen in der Umsetzung. FAW, Ulm, 2001

[RN03] RUSSELL, S.; NORVIG, P.: Artificial Intelligence – A Modern Approach. Prentice Hall Series in Artificial Intelligence, Prentice Hall, Pearson Education International, the second edition, USA, 2003

[Rub04] RUBIN, R. E.: Foundations of Library and Information Science 2nd Edition. Neal-Schuman, New York, 2004

[RW78] RAUCH, W.; WERSIG, G.: Delphi-Prognose in Information und Dokumentation. Verlag Dokumentation Saur KG, München, Germany, 1978

[RW89] RAFFEE, H.; WIEDMANN, K.: Strategisches Marketing. 2. Auflage, C. E. Poeschel Verlag, 1989

[SCI93] SCI Journal Citation Reports: a Bibliometric Analysis of science journals in the ISI database. Philadelphia: Institute for Scientific Information, Inc., 1993

[SEI-ol] SOFTWARE ENGINEERING INSTITUT: http://www.sei.cmu.edu/

[She97] SHENK, D.: Data smog: Surviving the information glut, p. 31, Abacus, London, 1997

[SK95] SCHIEBEL, E.; KOPCSA, A.: Methodisch-theoretische Abhandlung über bibliometrische Methoden und ihre Anwendungsmöglichkeiten in der industriellen Forschung und Entwicklung. Endbericht des Projekts 3437, 1995

[Sma73] SMALL, H.: Cocitations in the scientific literature: a new measure of the relationship between two documents. Journal of the American Society for Information Science 24, p. 265-269, 1973

[SPB07] SHMUELI, G.; PATEL, N.R.; BRUCE, P.C.: Data mining for business intelligence – concepts, techniques, and applications in Microsoft Office Excel with XLMiner. John Wiley & Sons, Inc., Hoboken, New Jersey, USA, 2007

[SR00] SCHUPPAN, V.; RUSSWURM, W.: A CMM-Based Evaluation of the V-Model 97. 7th European Software Process Workshop (EWSPT'7), Kaprun, Austria, 21-25 Feb. 2000

[SS01] STUDER, R.; STAAB, S.: Intelligente (Symbolische) Methoden für das Wissensmanagement. In: RADERMACHER, F. J. et al. (Hrsg.): „Management von nicht-explizitem Wissen: Noch mehr von der Natur lernen“, Abschlussbericht Teil 3: Die Sicht verschiedener akademischer Fächer zum Thema des nicht-expliziten Wissens. Forschungsinstitut für anwendungsorientierte Wissensverarbeitung (FAW), Ulm, S. 165-180, 2001

[SSV02] STUDER, R.; SURE, Y.; VOLZ, V.: Managing focussed access to distributed information. I-KNOW’02 International Conference on Knowledge Management 2002, Graz, Austria, July 11-12, 2002

[STI00-ol] SIZING THE INTERNET. In the E-Journal: Cyveillance, June 2000, under: http://www.cyveillance.com/resources/library.asp

[Swe01] SWEENEY, L.: Information Explosion. In: ZAYATZ, L.; DOYLE, P.; THEEUWES, J.; LANE J. (ed.): Confidentiality, Disclosure, and Data Access: Theory and Practical Applications for Statistical Agencies, Urban Institute, Washington DC, USA, 2001

[Tof70] TOFFLER, A.: Future Shock. Random House, USA, 1970

[UG96] USCHOLD, M.; GRUNINGER, M.: Ontologies: Principles, Methods and Applications. To appear in Knowledge Engineering Review Volume 11, Number 2, June 1996

[UGS05-ol] Ueber Google Scholar. Google Scholar, under http://scholar.google.de/intl/de/scholar/about.html, 2005

[VDI04] Verein Deutscher Ingenieure (VDI): Entwicklungsmethodik für mechatronische Systeme. VDI-Richtlinie 2206, Beuth-Verlag, Berlin, 2004

[WB02] WESTKÄMPER, E.; BALVE, P.: Technologiemanagement in produzierenden Unternehmen. In: BULLINGER, H.-J.; WARNECKE, H.-J.; WESTKÄMPER, E. (Hrsg.): Neue Organisationsformen im Unternehmen. Springer Verlag, Berlin, 2002

[WG81] WHITE, H. D.; GRIFFITH B. C.: Core journal networks and co-citation maps in the marine sciences: Tools for information management in interdisciplinary research. In: Journal of the American Society for Information Science, 32(3):163-171, 1981

[Wil00] WILLE, R.: Boolean concept logic. Technische Universität Darmstadt, Fachbereich Mathematik, Darmstadt, 2000

[Win92] WINSTON, P.H.: Artificial Intelligence (third edition). Addison-Wesley, Reading, Massachusetts, USA, 1992

[WIZ+04] WEISS, S. M.; INDURKHYA, N.; ZHANG, T.; DAMERAU, F. J.: Text Mining – Predictive Methods for Analyzing Unstructured Information. Springer Verlag, Berlin, Germany, 2004

[WN07] WordNet. Online under: http://wordnet.princeton.edu/, Cognitive Science Laboratory, Princeton University, USA, 2007

[WR03] WITKOWSKI, U.; RÜCKERT, U.: Development and Incorporation of Elementary Soccer Strategies for the Khepera Mini Robot. In Proc. of the FIRA Robot World Congress 2003, Vienna, Austria, October 2003

[Wso00] Web Surpasses One Billion Documents. Technical report, Inktomi Press Release, California, USA, 2000