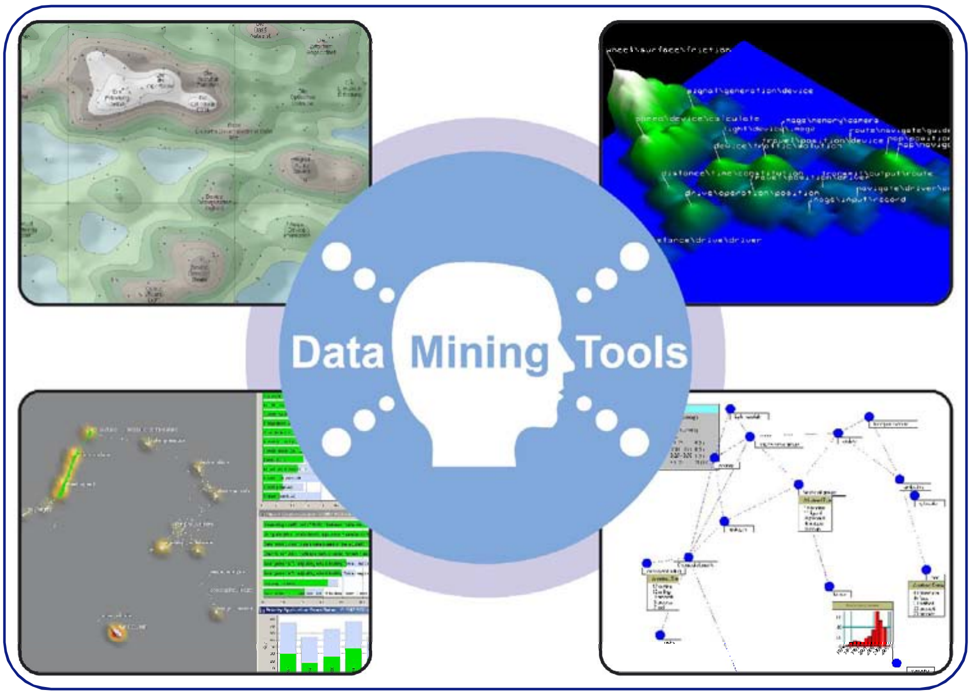
ESPOO 2008

VTT RESEARCH NOTES 2451



Laura Ruotsalainen

Data Mining Tools for Technology and Competitive Intelligence

VTT TIEDOTTEITA – RESEARCH NOTES 2451

Data Mining Tools for Technology and Competitive Intelligence

Laura Ruotsalainen

ISBN 978-951-38-7240-3 (soft back ed.)
ISSN 1235-0605 (soft back ed.)

ISBN 978-951-38-7241-0 (URL: http://www.vtt.fi/publications/index.jsp)
ISSN 1455-0865 (URL: http://www.vtt.fi/publications/index.jsp)

JULKAISIJA – UTGIVARE – PUBLISHER

VTT, Vuorimiehentie 5, PL 1000, 02044 VTT
puh. vaihde 020 722 111, faksi 020 722 7001

VTT, Bergsmansvägen 5, PB 1000, 02044 VTT
tel. växel 020 722 111, fax 020 722 7001

VTT Technical Research Centre of Finland, Vuorimiehentie 5, P.O. Box 1000, FI-02044 VTT, Finland
phone internat. +358 20 722 111, fax + 358 20 722 7001

Cover figure by Tuomo Hokkanen.
Top left: Aureka’s ThemeMap. Registered trademark of Thomson Reuters.
Top right: OmniViz’s ThemeMap. Software developed by Biowisdom.
Bottom left: Visualization made with STN AnaVist. Registered trademark of the American Chemical Society.
Bottom right: Thomson Data Analyzer's FactorMap. Registered trademark of Thomson Reuters.

Technical editing Leena Ukskoski

Edita Prima Oy, Helsinki 2008

Ruotsalainen, Laura. Data Mining Tools for Technology and Competitive Intelligence. Espoo 2008. VTT
VTT Tiedotteita – Research Notes 2451. 63 p.

Keywords patent data, text mining, data mining, patent mining, patent mapping, competitive intelligence, technology intelligence, visualization

Abstract

Approximately 80% of scientific and technical information can be found from patent documents alone, according to a study carried out by the European Patent Office. Patents are also a unique source of information since they are collected, screened and published according to internationally agreed standards. In addition to being an extremely valuable source of technology intelligence, patent documents offer a business competitive intelligence by revealing a competitor's strengths and strategies. Information gained from patents can also help in locating partners for cross-licensing and collaboration.

Since the patent system was established, more than 60 million patent applications have been published. It would be impossible to find and analyze relevant documents manually. The need for analysis and evaluation tools for patents has been acknowledged by many solution providers. New solutions are continuously coming onto the market; tools for reading and evaluating individual patents and tools for analyzing sets of patent documents. Solutions of the latter type can still be roughly divided into two groups: tools for retrieving and preparing basic statistics for patent documents, and tools for visualization and progressive analysis of patents. The former group deals only with data in a structured form, whereas the latter also analyzes unstructured text and other data.

In this study, four efficient tools for analyzing patent documents were tested: Thomson Reuter's Aureka and Thomson Data Analyzer, Biowisdom's OmniViz, and STN's STN AnaVist. All four tools analyze structured and unstructured data alike. They all visualize the results achieved from clustering the text fields of patent documents and either provide basic statistics graphs themselves or contain filters for performing them with other solutions.

The tools were tested with two cases, evaluating their ability to offer technology and business intelligence from patent documents for companies' daily business. Being aware of the state of the art of relevant technology areas is crucial for a company's innovation process. Knowledge of developed techniques and products forestalls overlapping R&D projects and thereby prevents unnecessary investment. Equally important is the recognition of other actors operating in the field. Benchmarking and evaluating a competitor's R&D and market strategies aids in managing one's own processes and locating possible parties for collaboration or cross-licensing.

This study took the point of view of a patent analyst with a basic understanding of patent data but no special knowledge of data mining techniques or the tools tested.

All the tools evaluated are very useful for the task and quite easy to adopt for daily work. All four had some strengths and weaknesses in comparison to each other. As a conclusion it could be stated that OmniViz and Thomson Data Analyzer are tools for sophisticated and diversified mathematical analysis of the data. Aureka and AnaVist are convenient for easily visualizing basic statistics and “top lists” of the data and for making stylish patent maps. The unique features of OmniViz, when compared to the other tools tested, are the possibility to visualize clustered data from many different points of view and the possibility to evaluate some attributes with patent map animations. Thomson Data Analyzer offers efficient tools for comparing different subsets of the data, e.g. for identifying unique values of an attribute. Aureka is the only tool to allow citation analyses and has the most illustrative patent map. STN AnaVist is superior in the possibility to retrieve basic statistics fast and smoothly.

The results obtained with all four tools were very much alike, even though different databases for retrieving the data were used. The top assignees and inventors lists were uniform, as were the year trends and both technological and geographical business areas. Only the reciprocal orders and amounts of documents varied. However, the conclusions drawn from the results, and business decisions made with them, would all be similar regardless of the tool used.

Preface

This research has been carried out as a diploma work for training programme Patentit-Teollisuus-Tekniikka. The training programme dealt with intellectual property rights and was arranged by Helsinki University of Technology's Lifelong Learning Institute Dipoli.

The work was made possible by the support of Thomson Reuters, BioWisdom Ltd. and American Chemical Society by offering free trials for the tools evaluated and access to data used in the study. The offers and assistance are gratefully acknowledged.

October 2008

Laura Ruotsalainen

Contents

Terminology

EPO	European Patent Office
PCT	International patent application system, based on Patent Cooperation Treaty
IPC	International Patent Classification

1. Introduction

Approximately 80% of scientific and technical information can be found from patent documents alone, according to a study carried out by the European Patent Office. They are also a unique source of information since they are collected, screened and published according to internationally agreed standards. In addition to being an extremely valuable source of technology intelligence, patent documents offer a business competitive intelligence by revealing a competitor's strengths and strategies. Information gained from patents can help in locating partners for cross-licensing and collaboration.

Since the patent system was established, more than 60 million patent documents have been published. It would be impossible to find and analyze all relevant documents manually. The need for analysis and evaluation tools for patents has been acknowledged by many solution providers. New solutions are continuously coming onto the market; tools for reading and evaluating individual patents and tools for analyzing sets of patent documents. Solutions of the latter type can still be roughly divided into two groups: tools for retrieving and preparing basic statistics for patent documents, and tools for visualization and progressive analysis of patents. The former group deals only with data in a structured form, whereas the latter also analyzes unstructured text and other data.

In this study, four sophisticated patent analysis and visualization tools were tested, all of which also dealt with treating unstructured text data. This study discusses observations made during the testing and the results retrieved. The study is structured as follows: Chapter 2 discusses the advantages gained from analyzing patents. Chapter 2.1 explains the structure and content of patent documents and shows an example. Chapter 3 describes the study in more detail. First (in Chapter 3.1), it explains the terms used while speaking of analyses. Then (in Chapter 3.2), it introduces the two test cases used, and finally (in Chapter 3.3), it explains the contents of the data sets used for testing. Chapter 3.4 introduces the four tools tested. Chapters 4 and 5 finally present the results with many figures. Chapter 6 describes the conclusions drawn in the study.

2. Technology and Competitive Intelligence from Patent Documents

The patent system is based on the rule that in order to gain a monopoly, the invention has to be explained with enough accuracy for anyone skilled in the art to implement. Patent applications are public 18 months after filing and are available for anyone, nowadays mostly also in electronic form.

The importance of patent data as a source for technology and competitive intelligence has been acknowledged for a long time. Ove Grandstrand [1] has distinguished between four types of technical information carriers: patents, scientific and technical publications, people and products/processes. In his book, Granstrand states, "In this context patent information, despite its many and well-recognized inadequacies, stands out as a unique source of technical information. More than any other source, it is collected, screened and published according to internationally agreed standards. It continually provides an assessment of the state of the art together with at least a rudimentary measure of metric of technological change. It thereby enables a transparent accumulation of knowledge on a global scale."

Due to the remarks stated above, patent data is crucial for research, development and business actions, even for companies without intentions to apply for patents of their own. Analyses of patent documents indicate the present state of the art, as well as aid in locating "white spaces"; technology areas lacking inventions. Due to the requirement to disclose inventions precisely, the compositions of patented products and methods may be examined in their entirety, which is not always possible otherwise. Yearly trends in research and development may be revealed by analyzing the yearly information found in patent documents. Research and development emphases also vary according to the geographical location, which can be examined through patent data. Yearly trends in patenting indicate present "hot areas", i.e. areas with many inventions, as well as declining technology fields. Analysis for technology intelligence should always be done with specialist in the technology area in question in order to achieve the maximum benefit.

Patent documents also offer competitive intelligence to support a company's business decisions. Knowledge of other companies' patent portfolios offers valuable information on competitors and helps to locate possible actors for licensing technologies and collaboration, as well as identifying new entrants in the market.

Knowing another company's or competitor's patenting activities reveals its strengths and business strategies. Yearly trends show the technology areas the company has abandoned and the areas it is currently concentrating on. Application activity for different Patent Organizations, i.e. each country's patent office, reveals a company's geographical business strategy. It has to be kept in mind though, that some information always remains confidential to the company.

In order to receive a patent the invention has to be novel, among other requirements. This means that the patent application must be the first place where the invention is disclosed. This offers a great possibility of revealing new products and techniques long before they enter the market. Analysis of patent assignees and inventors reveals collaboration between other actors and offers information for headhunting.

2.1 Patent Data

Patent documents contain both structured and unstructured data. Figure 1 shows a sample first page of a patent document. The first page consists of bibliographical data which is strictly structured, and of a title and an abstract, which are both unstructured. A patent document also contains a description of the invention, claims (a concise definition of the legal protection of the patent is) and drawings.

Numbering and publication practices vary quite a lot between different Patent Organizations. The basic attributes used for analysis are, however, found in all patent documents, and are to some extent standardized by different database producers. The patent application introduced in Figure 1 is published under the Patent Co-operation Treaty (PCT). Fields found in patent documents are identified with international numbering. The meanings of those fields used in the analysis introduced in Chapters 4 and 5 will be explained below, with numbers corresponding to the figure marked.

Numbers and dates

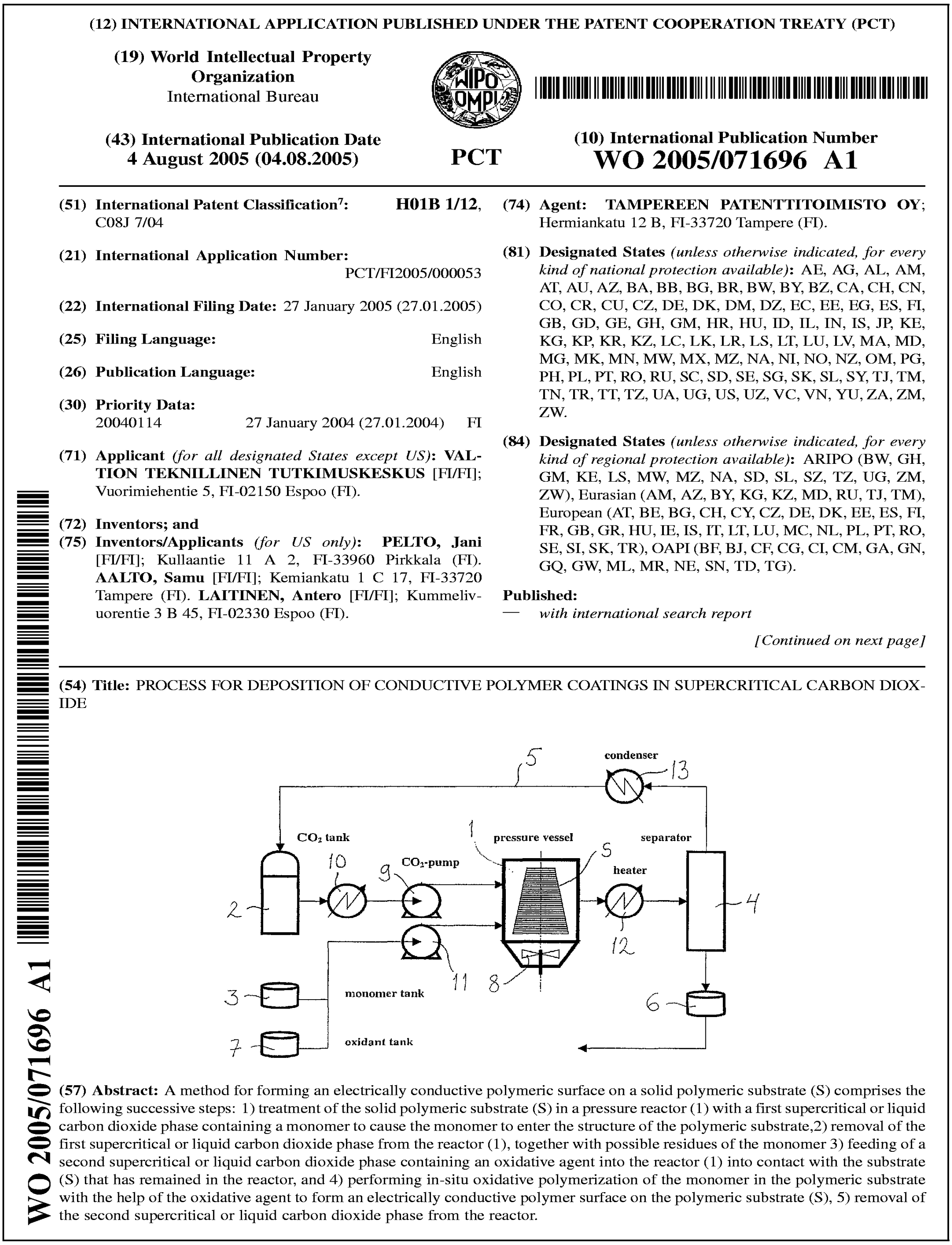
Patent documents contain identifying numbers. Priority data (field 30) consists of a priority number assigned for the first application applying for a patent and the corresponding date. A publication number is given to the document when it is published, 18 months after filing. The publication date refers to the corresponding date. The issue date is the date the patent is granted, usually 3–5 years after filing, depending on the Patent Office.

Assignees

Patent assignees or applicants (71) are the organizations or individuals holding the rights for the invention and applying for the patent. Inventions developed in collaboration are assigned to all the parties involved.

Inventors

Inventors (72, 75) are the researchers who have developed the invention in question.



(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
4 August 2005 (04.08.2005)

PCT

(10) International Publication Number
WO 2005/071696 A1

(51) International Patent Classification[7]: H01B 1/12, C08J 7/04

(21) International Application Number: PCT/FI2005/000053

(22) International Filing Date: 27 January 2005 (27.01.2005)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
20040114 27 January 2004 (27.01.2004) FI

(71) Applicant *(for all designated States except US)*: VALTION TEKNILLINEN TUTKIMUSKESKUS [FI/FI]; Vuorimiehentie 5, FI-02150 Espoo (FI).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: PELTO, Jani [FI/FI]; Kullaantie 11 A 2, FI-33960 Pirkkala (FI). AALTO, Samu [FI/FI]; Kemiankatu 1 C 17, FI-33720 Tampere (FI). LAITINEN, Antero [FI/FI]; Kummelivuorentie 3 B 45, FI-02330 Espoo (FI).

(74) Agent: TAMPEREEN PATENTTITOIMISTO OY; Hermiankatu 12 B, FI-33720 Tampere (FI).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NA, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IS, IT, LT, LU, MC, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *with international search report*

[Continued on next page]

(54) Title: PROCESS FOR DEPOSITION OF CONDUCTIVE POLYMER COATINGS IN SUPERCRITICAL CARBON DIOXIDE

(57) Abstract: A method for forming an electrically conductive polymeric surface on a solid polymeric substrate (S) comprises the following successive steps: 1) treatment of the solid polymeric substrate (S) in a pressure reactor (1) with a first supercritical or liquid carbon dioxide phase containing a monomer to cause the monomer to enter the structure of the polymeric substrate,2) removal of the first supercritical or liquid carbon dioxide phase from the reactor (1), together with possible residues of the monomer 3) feeding of a second supercritical or liquid carbon dioxide phase containing an oxidative agent into the reactor (1) into contact with the substrate (S) that has remained in the reactor, and 4) performing in-situ oxidative polymerization of the monomer in the polymeric substrate with the help of the oxidative agent to form an electrically conductive polymer surface on the polymeric substrate (S), 5) removal of the second supercritical or liquid carbon dioxide phase from the reactor.

WO 2005/071696 A1

Figure 1. An example of the first page of a patent document.

Classifications

Inventions are classified according to the technologies they are related to. The most commonly used is the International Patent Classification (IPC, 51). Many Patent Organizations have created their own classification system, e.g. the United States Patent and Trademark Office (US patent classification) and the European Patent Office (ECLA European Classification). Some database producers have also created their own classification system to make searching and analysis of the relevant documents easier. In this study two such classifications have been used provided by Thomson Reuters: Derwent Classification and Derwent's Manual Codes.

Title and abstract

The title (54) and abstract (57) are descriptions of the invention in natural language. They are unstructured text and the informative manner of them is varies greatly depending on the author.

3. Study

The total amount of patent documents published is 60 million. This also includes applications which have never led to a patent. The amount is increasing at an accelerating rate. It would be impossible to find and analyze all relevant patents manually.

The need for analysis and evaluation tools for patents has been acknowledged by many solution providers. New solutions are coming onto the market continuously; tools for reading and evaluating individual patents (like ScioSphere, STN Viewer and PatBase), and tools for analyzing sets of patent documents. Solutions of the latter type can still be roughly divided into two groups; tools for retrieving and making basic statistics for patent documents (like LexisNexis, QPat and PatBase), and tools for visualization and progressive analysis of patents. The former group deals only with data in a structured form, whereas the latter also analyzes unstructured text data. The latter group is the target of this evaluation.

This study was carried out from the point of view of a patent analyst. Patent analysts, as meant here, offer technology and business information for a company's R&D and business actions. The study wanted to clarify what kind of help the launched solutions would provide for analysts' daily work. Four solutions widely used for patent analysing were tested. These tools are Aureka, Biowisdom's OmniViz, STN AnaVist and Thomson Data Analyzer. This report presents the observations made during the testing and shows the results of two test cases.

In this chapter the research frame and data used in the study are described. In Chapter 3.1 the term Data Mining used in the title is explained. In Chapter 3.2 the two test cases used for testing the tools are introduced, and in Chapter 3.3 the databases used for retrieving the test data are presented.

3.1 Data Mining

The use of terms in the literature related to processing patent-related data is quite confusing. The terms "data mining," "patent mining," "text mining" and "visualization" are employed for the processing of the documents. This chapter will try to give some explanations of the terms and explain why "data mining" was chosen for the title of the study.

Hand, Mannila and Smyth [2] define the term “data mining” as follows:

“Data mining is the analysis of (often large) observational data sets to find unsuspected relationships and to summarize the data in novel ways that are both understandable and useful to the data owner.”

Feldman’s and Sanger’s [3] definition of “text mining”: “Text mining can be broadly defined as a knowledge-intensive process in which a user interacts with a document collection over time by using a suite of analysis tools. In a manner analogous to data mining, text mining seeks to extract useful information from data sources through the identification and exploration of interesting patterns. In the case of text mining, however, the data sources are document collections, and interesting patterns are found not among formalized database records but in the unstructured textual data in the documents in these collections.”

Clustering is a process which groups the objects into groups called clusters. This is done by classifying the objects. The difference between clustering and categorization is, according to Feldman and Sanger [3], that “In categorization problems we are provided with a collection of preclassified training examples, and the task of the system is to learn the description of classes in order to be able to classify a new unlabeled object. In the case of clustering, the problem is to group the given unlabeled collection into meaningful clusters without any prior information. Any labels associated with objects are obtained solely from the data.”

The term “patent mining” is quite widely used in the literature referring to processing patent data with data and text mining techniques (e.g. Jin *et al.* [4] and Kasravi *et al.* [5]). It refers to data mining of patent documents.

The idea of visual data exploration, “visualization,” is, according to Keim [6], “to present the data in some visual form, allowing the human to get insight into the data, draw conclusions, and directly interact with the data.” The development of visualization techniques in the last decade has made it possible to widen the visualization of low-dimensional data, e.g. making histograms of yearly attributes, to create sophisticated visualizations of high-dimensional text data.

Patent documents contain structured and unstructured data alike; they are “semi-structured” as Feldman and Sanger [3] write. The bibliographic information of a patent is structured and follows a strict format. For example, it contains patent assignee and inventor names, different identifiers such as priority and publication numbers, years and classifications. Unstructured data is text explaining the invention and the extent of protection of the patent, e.g. title, abstract and claims. The structure of patents was explained in more detail in Chapter 2.1. According to Tseng *et al.* [7], visualizations of results of patent analysis are called patent graphs if they are prepared from structured data, and patent maps if they are from unstructured texts. However, they admit that, loosely speaking, the term map is used often for both cases.

All four tools tested analyze both structured data and unstructured text data, and visualize the results of both categories. The visualizations are possible for low- and high-dimensional data alike, e.g. bar and pie graphs for structured data and patent landscapes, i.e. maps, for unstructured data. An analysis of text data is carried out by first clustering it according to the most frequent words. In this study the term “data mining” is used as an upper-level term for all the handling of large amounts of data, and “text mining” as a subset of it, referring to the analysis of unstructured data.

3.2 Test Cases

The tools were tested with two cases, introduced below, evaluating their ability to offer technology and business intelligence from patent documents for companies’ daily business. Being aware of the state of the art of relevant technology areas is crucial for a company’s innovation process. Knowledge of developed techniques and products forestalls overlapping R&D projects and thereby prevents unnecessary investments. Equally important is the recognition of other actors operating in the field. Benchmarking and evaluating a competitor’s R&D and market strategies aids in managing one’s own processes and locating possible parties for collaboration or cross-licensing.

The study took the point of view of a patent analyst with a basic understanding of patent data but no special knowledge of data mining techniques or the tools tested. In addition to evaluating the results of the analysis and their value for technology and business intelligence, their usability was compared by answering the following questions:

- How easy is the tool to use? Does it require a lot of reading of manuals before starting the analysis?
- What kinds of data format does the tool support? Is there a need to manipulate the data before it can be analysed by the tool?
- Can the tool be used for improving the search, meaning focusing the results by excluding documents that have come in because of too extensive a search profile?

Case 1: Technology-based data

The first case evaluated the tools with technology-based data. Patent documents are the most comprehensive and accurate source of technology-related information. According to a study carried out by the European Patent Office, 80% of technical information can be found from patent documents alone. In return for a temporary monopoly on an invention, the inventor has to disclose it in sufficient detail.

The first case analyzed patent documents dealing with technologies for measuring friction between a vehicle and the road surface. Any restrictions by year for the documents were not included. The search profile was constructed by restricting the search for relevant documents into one IPC class, G01; measuring and testing. Then patent documents containing the words “vehicle,” “road” and “friction” or “condition,” with synonyms for all the terms were searched for. A more detailed search profile is introduced in Chapter 3.3.

The evaluations were committed by looking for answers to the following questions.

- What are the trends of patenting in the technology area now and how has the patenting changed over time?
- Who are the most active patent assignees and inventors in the area?
- How is it possible to examine a specific subgroup of the technology?
- How do the inventions of company X relate to the inventions of company Y?
- Is it possible to locate a specific document among the mass of patents?

The results of the first case are introduced in Chapter 4.

Case 2: Company-based data

The second case evaluated the patent portfolio of a specific company. Knowing another company’s or competitor’s patenting activities reveals its strengths and business strategies. Yearly trends show the technology areas the company has abandoned and the areas it is concentrating on now. Applying this activity to different Patent Organizations, i.e. different countries’ patent offices, reveals a company’s geographical business strategy. Patent analysis offers a great possibility to reveal new products and techniques long before they come onto the market. Analysis of patent assignees and inventors reveals collaboration between other actors and offers information for headhunting.

Fraunhofer-Gesellschaft was chosen to be the company for the second test case. Fraunhofer-Gesellschaft is Europe’s largest organization for applied research, based in Germany. All its patents and applications filed since 1995 were searched for. A more detailed search profile is introduced in Chapter 3.3.

The second case tries to evaluate how the following questions are answered by using the tools.

- What are the key technologies the company is concentrating on now?
- How has the focus of its business changed in the past ten years?
- What is the geographical area of operation for the company?
- What kind of co-operation does the company have and with whom?

The results of the second case are introduced in Chapter 5.

3.3 Databases and Data Sets Used for Testing

The data was retrieved from four different commercial databases: MicroPatent, Derwent World Patent Index (DWPI), USPatfull and PCTFull, and three different data sets were made. The content of the data sets is constructed with quite similar profiles, introduced below.

A combination of DWPI, USPatfull and PCTFull data was used for testing STN AnaVist.

DWPI data was used for testing Thomson Data Analyzer and MicroPatent's data for testing Aureka and OmniViz.

The utilization of the tools in real life has been imitated by using different databases and datasets. Aureka uses the MicroPatent database, and STN AnaVist uses four different databases named in Chapter 3.4. Thomson Data Analyzer analyzes data from different sources, but has special filters for handling DWPI data. OmniViz uses data in any format and offers a wizard for importing data in Microsoft Excel format easily, which is why MicroPatent data was used to test it.

The analysis was carried out using the title, abstract and bibliographic data of the documents.

MicroPatent's PatentWeb

MicroPatent's PatentWeb is an online repository produced by Thomson Reuters. It contains more that 50 million full text and front page records. PatentWeb contains documents from six Patent Organizations and PCT applications; full text documents from USPTO (the United States Patent and Trademark Office), EPO (the European Patent Office), Germany's and Great Britain's Patent Offices and PCT applications, and front page documents from Japan's Patent Organization.

Derwent World Patent Index (DWPI)

The Derwent World Patent Index (DWPI) is a database produced by Thomson Reuters. It is called “value added,” which means that the patent documents have rewritten titles and abstracts in English to be more informative. DWPI has patent documents from more than 40 patent authorities around the world. Data from all members of the patent family has been incorporated into one document. The documents also have DWPI’s own classifications, Derwent classification codes and Derwent manual codes, to improve searching for and handling relevant documents.

USPatfull

USPatfull is a database produced by the U.S. Patent and Trademark Office (USPTO). It contains full text patent documents from USPTO since 1971.

PCTFull

PCTFull covers the full text of Patent Co-operation Treaty (PCT) published applications of the World Intellectual Property Organization (WIPO) since 1978.

Search profiles for each database combinations

Since the data consists of two different cases and three combinations of databases, this resulted in six data sets. All sets contain patent and patent application documents. The MicroPatent data was saved in Microsoft Excel format, and Derwent World Patent Index data in text format. Below detailed search profiles for all six data sets will be given with the number of patents they contained. Patents filed with different Patent Organizations for the same invention constitute a patent family. In order not to bias the analyses with several occurrences of the same invention, the documents were restricted to one from each family.

Explanations of used symbols:

*	: Any number of marks
nN	: Words appear within n words or less
IPC	: International Patent Classification
PA	: Patent assignee.

Technology-based data

The search was restricted to documents classified into IPC class G01; Measuring and Testing.

Documents had to contain one of the words:

Vehicle or *Car* or *Automotive*

and either of the following combinations of words:

*(Skid** or *Friction)(5N)(Road* or *Highway* or *Freeway* or *Roadway)*

or

(Condition)(3N)(Road or *Highway* or *Freeway* or *Roadway).*

The search produced 646 documents from MicroPatent and 1081 from the Derwent World Patent Index. In STN the documents retrieved from DWPI were merged with documents retrieved from USPATFULL and PCTFULL. The documents were searched by limiting the fields searched for title, abstract and claims to make the search more accurate. This yielded 1343 documents.

Company-based data

The search was restricted to applications and patents filed after 1994:

FRAUNHOFER/PA.

The search produced 4513 documents from the Derwent World Patent Index and 4628 documents from MicroPatent. The difference in numbers is due to the few weeks' time lapse between the retrievals.

3.4 Tools for Analysis

There are numerous tools and solutions for analysing patents and new ones are coming onto the market continuously. There are, for example, tools for reading and evaluating individual patents and tools for analyzing sets of patent documents. Solutions of the latter type can still be roughly divided into two groups; tools for retrieving and preparing basic statistics for patent documents, and tools for visualization and progressive analysis of patent. The former group deals only with data in a structured form, whereas the latter also analyzes unstructured text and other data.

Analysis tools have been tested and introduced earlier by many writers (e.g. Eldridge [8] and Yang *et al.* [9]). In this study the tools were tested with two real-life cases to see how they fit to true cases present in a company's daily operations.

Four extensive analysis tools were tested. The purpose of this study was to evaluate how easy the tools are to use and how useful and informative analyses performed with them are. The study went over two cases, one with technology-based data and the other with one company's patent portfolio. The outputs of the analysis are presented in the next two chapters. Chapter 4 introduces graphics and remarks of the analysis performed with technology-based data, and Chapter 5 the one with company-based data. This chapter will present information about the tools and observations about their usability.

All the tools had processes for clustering unstructured text, in this case titles and abstracts of patent documents, and visualizing the clusters. They all also provided tools for retrieving basic statistics about different attributes found in the data.

The tools tested may be roughly divided into two groups. The first group consists of Aureka and STN AnaVist, which are easy to use and offer basic analysis with little effort and studying needed from the user. The data available is limited to a few patent databases; Aureka uses data retrieved from the MicroPatent database, and STN AnaVist from four STN patent databases. Along with highly usable interfaces, the ease of their use comes from restricting the user from influencing the algorithms and methods used for handling the data.

The tools in the other group, OmniViz and TDA VantagePoint, are tools empowering very sophisticated statistical analysis performed on almost any kind of data. The use of these tools requires some learning. Due to the possibility to analyse data from almost unrestricted sources, some preparation before committing the analysis is needed. Both tools provide filters and wizards to help with the importing of data. Most analyses offered by these tools are easy to render by using the default values. For a power user there are wide-ranging possibilities to choose from, e.g. clustering algorithms.

Preparing the right search profile, especially for retrieving technology-related patent documents, is often quite difficult. The searcher has to find a balance between leaving some relevant documents out and including irrelevant ones. The relevancy of the analysis, however, depends on the validity of the data.

All the four tools tested enabled the re-defining of the data by making subsets of it and handling them as bases for new analyses. Visualizing the data made it easy to locate documents that didn't relate to the subject and exclude them.

Aureka

Aureka is Thomson Reuters' tool for analysing and visualizing patent data and sharing it effectively inside the company. The best features of MicroPatent are its representative visualizations, basic statistics and interface.

Aureka is integrated for the MicroPatent database, which was introduced in more detail in Chapter 3.3. It is flexible and allows broadening and modifying of the data set even in the middle of the analysis, though it limits the analysis to only full text patent documents from a few of the world's biggest patent offices. Working with Aureka needs very little preparation of the data or learning of the tool.

Aureka emphasizes the sharing of information inside a company and representation of the results of analysis. Basic statistics can be viewed with ready-made reports or by exporting data to Microsoft Excel. Aureka also offers Excel macros to ease the analysis.

Aureka doesn't provide interactivity at the same level as the other tools tested, but individual patent documents may be located from the visualizations and viewed in detail.

OmniViz

OmniViz is BioWisdom's powerful data mining tool. It is designed mainly for analyzing biological data, but is well suited to treating patent documents from other technology fields as well. The best features of OmniViz are its flexibility, efficiency, high degree of interactivity and supply of many different visualization techniques.

OmniViz is great tool for a user who is familiar with data mining methods and algorithms and is willing to influence their utilization. While preparing the data, the user can choose, for example, the stop words and algorithm used for clustering (K-Means, Hierarchical etc.). OmniViz is made to be easy to use with the default values and is suited to making basic visualizations and statistics of patent data.

OmniViz emphasizes visual presentations and analysis of data. There are eight different visualizations for taking different perspectives of the data, looking at clusters made from it, the associations of terms in them, the correlation between numerical attributes and special visualizations for network relationships in biological data.

Any format of data can be treated with OmniViz. It offers filters for importing the data and for a capable user there are almost no limits. The data used for analysis may also be a combination of different data sets, e.g. patents and publications. Because the format of the data is not predetermined, it needs some preparation before clustering. Cleaning of data may be done before clustering by editing auxiliary files, or afterwards using a tool in OmniViz.

OmniViz focuses on visual analysis and is quite complex just for producing basic statistics. It offers filters for exporting the relevant sets of data to other tools, e.g. Microsoft Excel.

OmniViz offers a very high degree of interactivity between tools used for analyzing the data. Selections in one tool immediately affect everything open in the workspace. It is possible to see animations of numerical or categorical data, e.g. patent filing years and related records in one of the visualizations.

STN AnaVist

STN Anavist is the American Chemical Society's tool for flexible analysis of data retrieved from the STN data bank. The best features of STN AnaVist to be its representative visualizations, seamless interaction between different analyses, the ease with which various statistics could be prepared, and the user-friendly interface.

AnaVist analyzes data retrieved from four STN databases (five since June 2008). Two of them are full text patent databases (USPatfull and PCTfull), and two are value added databases (DWPI and CAplus, which contains patent and scientific references for chemistry and biochemistry). In June 2008 the number of databases increased by one, when the EPFULL database, containing full text European patent documents filed since 1978, was included. The data may be imported easily from STN and exported back for further processing. The converse of this ease though, is that analysis is restricted to patents in other fields than chemistry and biochemistry.

Creating a visualization of the data with AnaVist is quite fast, as is preparing statistics from different points of view. AnaVist also has a high degree of interactivity between different analyses.

Thomson Data Analyzer – VantagePoint

Thomson Data Analyzer is an analysis tool from Thomson Reuters which uses Search Technology's VantagePoint data mining software for the analysis. The best features of TDA VantagePoint were its flexibility and efficiency, as well as its macros for creating different reports and tools for comparing different groups made of the data.

Thomson Data Analyzer analyzes data in almost any format. It provides filters to help with importing and special tools for handling and processing data from the Derwent World Patent Index.

Thomson Data Analyzer emphasizes the processing of the data and has quite modest-looking visualizations, although with powerful calculations backing it up. It offers different kinds of analysis tools, lists and matrices for basic statistics, and maps for evaluating relationships between terms and clustering text. The tools allow the user to easily reveal records in the data set that are similar or assignees that have common documents, i.e. patent applications filed due to collaboration. These methods are very useful for identifying statistical divergence from the data, i.e. finding technology areas that are highly patented, “hot areas” as well as companies that are remarkably active in some specific technology area.

Thomson Data Analyzer offers three different types of predefined reports. The reports contain basic as well as more processed information on the subject. They all are Microsoft Excel files containing tables and graphs. The reports may not be upgraded. The three types of reports are: Company Report, for retrieving information on one special company, Company Comparison Report, for comparing two to five companies and Technology Report containing variety of metrics regarding a technology area.

4. Analysis with Technology-based Data

The analysis tools were tested with two cases, introduced in Chapter 3.2. This chapter introduces the results of the analyses with technology-based data. Some figures have been excluded because they showed similar results to the ones introduced in other chapters. Chapter 4.1 introduces patent landscapes made for getting a general view of the data. Chapter 4.2 deals with closer analysis of an interesting technological area discovered from the landscape. Chapter 4.3 looks at yearly trends in patenting and Chapter 4.4 compares the patent portfolios of two companies. Chapter 4.5 looks at patenting around one specific invention, i.e. one especially interesting patent document.

4.1 Landscape

Evaluation of the data is best begun by creating a general view of it. Different specific technologies patented and their frequency in the data may be evaluated by clustering the terms appearing in documents. In this study titles and abstracts were used for clustering. The tools visualize results by mapping the documents and clusters in proportion to each other, i.e. creating patent maps. Documents with similar subjects appear close to each other in maps. This makes it very easy to locate the most developed areas in the technology. It also shows outliers in the data, documents that don't have much to do with the subject but are in the data by accident. All the tools tested enable closer evaluation of individual documents and clusters.

Some basic statistical analysis was also performed of the data in order to get a better understanding of it. Answering the questions "Who did?", "What did?" and "When did?" gives a general idea of the field. These questions were answered by looking at patent assignees and inventors, application or priority years and different classifications given for the documents.

Figure 2 shows the visualization made with Aureka. Aureka's representation was found to be the clearest and most illustrative. The frequency of the documents is shown with contour lines and colours. Every document is represented with a dot. Three most important terms in the cluster differentiating the documents from other documents are shown. The greatest frequency of documents was of techniques related to the brakes of a vehicle, described with words (brake, braking force and system) in the visualization. The cluster is shown on left middle and is coloured with white.

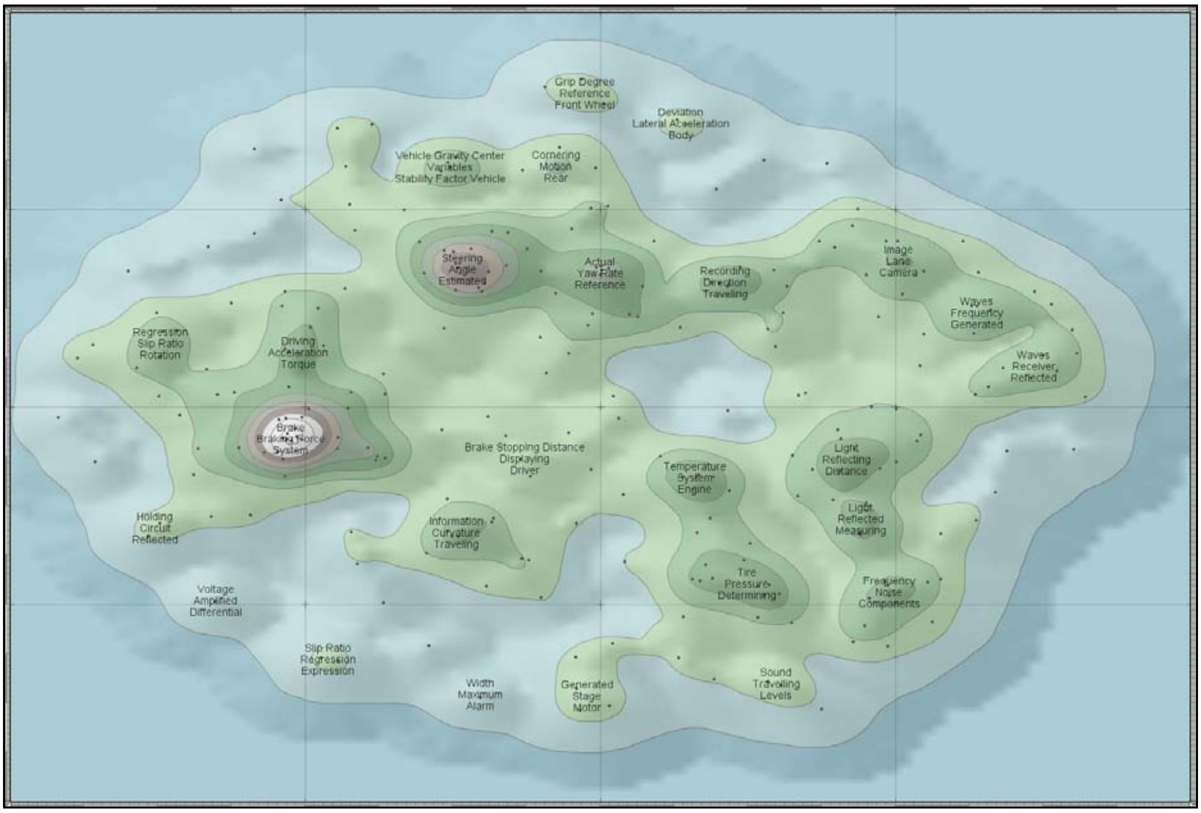
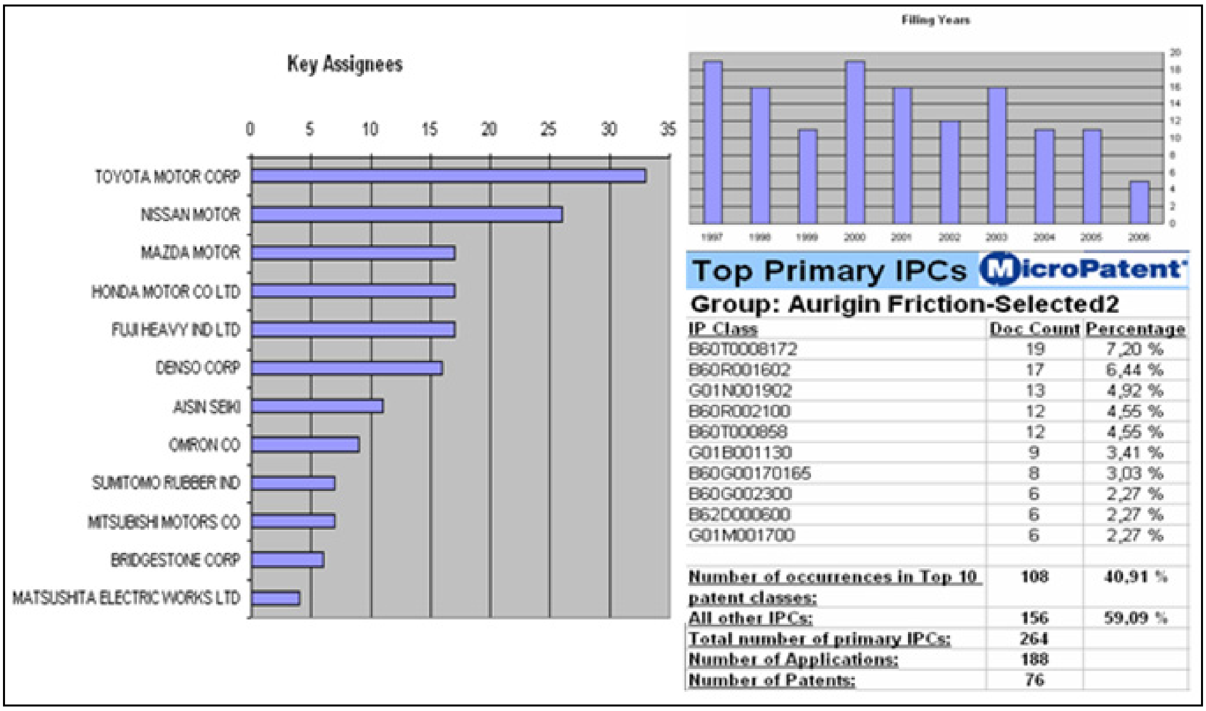

Grip Degree Reference Front Wheel
Deviation Lateral Acceleration Body
Vehicle Gravity Center Variables Stability Factor Vehicle
Cornering Motion Rear
Steering Angle Estimated
Actual Yaw Rate Reference
Recording Direction Traveling
Image Lane Camera
Waves Frequency Generated
Waves Receiver Reflected
Regression Slip Ratio Rotation
Driving Acceleration Torque
Brake Braking Force System
Brake Stopping Distance Displaying Driver
Temperature System Engine
Light Reflecting Distance
Light Reflected Measuring
Holding Circuit Reflected
Information Curvature Traveling
Tire Pressure Determining
Frequency Noise Components
Voltage Amplified Differential
Slip Ratio Regression Expression
Width Maximum Alarm
Generated Stage Motor
Sound Travelling Levels

Figure 2. Aureka: Visualization of patents related to measuring friction on the road surface made with Aureka's ThemeMap.

Figure 3 shows statistics prepared from data with Aureka. Aureka offers two tools for retrieving the information; predefined reports and filters for exporting the data to Microsoft Excel. Predefined reports are offered for all attributes found in the data. The reports are impossible to edit, which is often needed for getting the data into a displayable form. Due to this Aureka offers tools for importing the data into Microsoft Excel and macros for making proper presentations of the statistics. The top patent assignees and the amount of documents they have filed related to the technology in question are shown on the left side of the figure. Toyota, Nissan and Mazda are the most active companies patenting in the area, which is not a surprise when dealing with techniques related to roads. Yearly trends in filing are shown at the top right. Patenting activity varied quite a lot from 1997 to 2006, and seems to have fallen nowadays. The chart at the bottom right shows the top International Patent Classifications given by Patent Organizations. They show a more detailed division of the different techniques among the documents than the clustering does. The IPCs in the predefined report are not informative because they have only the numeric codes and not the verbal explanations. All statistics in the figure were prepared with Microsoft Excel and Aureka's macros.



Top Primary IPCs

Group: Aurigin Friction-Selected2

IP Class	Doc Count	Percentage
B60T0008172	19	7,20 %
B60R001602	17	6,44 %
G01N001902	13	4,92 %
B60R002100	12	4,55 %
B60T000858	12	4,55 %
G01B001130	9	3,41 %
B60G00170165	8	3,03 %
B60G002300	6	2,27 %
B62D000600	6	2,27 %
G01M001700	6	2,27 %
Number of occurrences in Top 10 patent classes:	108	40,91 %
All other IPCs:	156	59,09 %
Total number of primary IPCs:	264	
Number of Applications:	188	
Number of Patents:	76	

Figure 3. Aureka: Analysis of patents related to measuring friction on the road surface. The figure shows basic statistics of the data_ top patent assignees on the left, yearly trends in filing from 1997 to 2006 at the top right and the top International Patent Classifications (IPC) at the bottom right. The charts were prepared with Microsoft Excel.

Figure 4 shows visualizations of the data from different perspectives made with OmniViz. The similarity between the documents can be seen from the graphics on the left, made with the Galaxy visualization. OmniViz has clustered documents containing equal terms and positioned similar clusters close to each other. Each individual document is shown with a dot and squares representing centroids of the clusters. The three most frequent terms can be shown for all clusters, but they have been restricted to only a few to clarify the image. The graphic on the right was made with the ThemeMap visualization. It shows the same patent landscape but from a different angle. It is suitable for estimating the amounts of documents in each cluster. The graphics can be rotated and zoomed in OmniViz to look at it from all directions. The cluster with the terms *wheel, surface, friction* has the highest frequency of documents. Most of the clusters appear near each other at the top of the representation. There are three clusters at the bottom far apart from the others and having only a few documents. There is a possibility that they are not relevant to the subject. A closer investigation of the clusters was performed. It revealed that they discuss navigation techniques, not evaluating the condition of road surface, and are therefore outliers. A subset analysis was carried of the data by leaving them out in order to get more accurate data for further analysis.

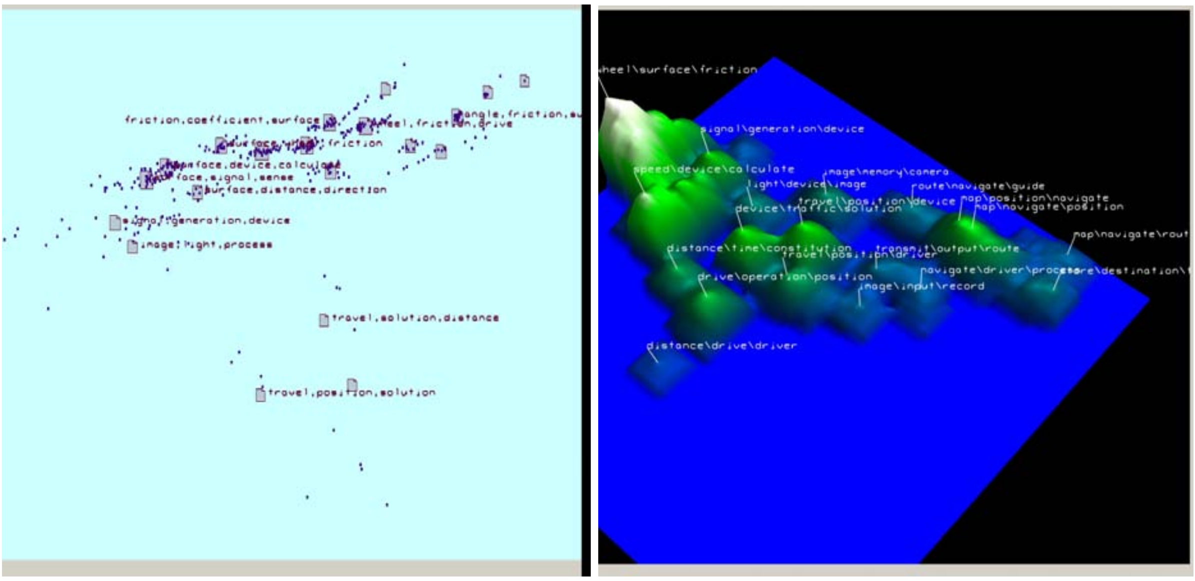


Figure 4. OmniViz: Visualization of patents related to measuring friction on the road surface made with two visualizations, Galaxy on the left and ThemeMap on the right.

Figure 5 shows basic statistics of the data made in OmniViz. OmniViz offers filters for exporting data to other applications for further analysis. At the bottom is a histogram prepared with Microsoft Excel of the top patent assignees, using data imported from OmniViz. In the upper-left corner is OmniViz's group tool showing the document counts of the most active assignees, and on the right a visualization with Galaxy showing the assignees coloured with corresponding colours. In the middle of the figure is OmniViz's Dynamic Query tool showing the yearly trends of patent application years.

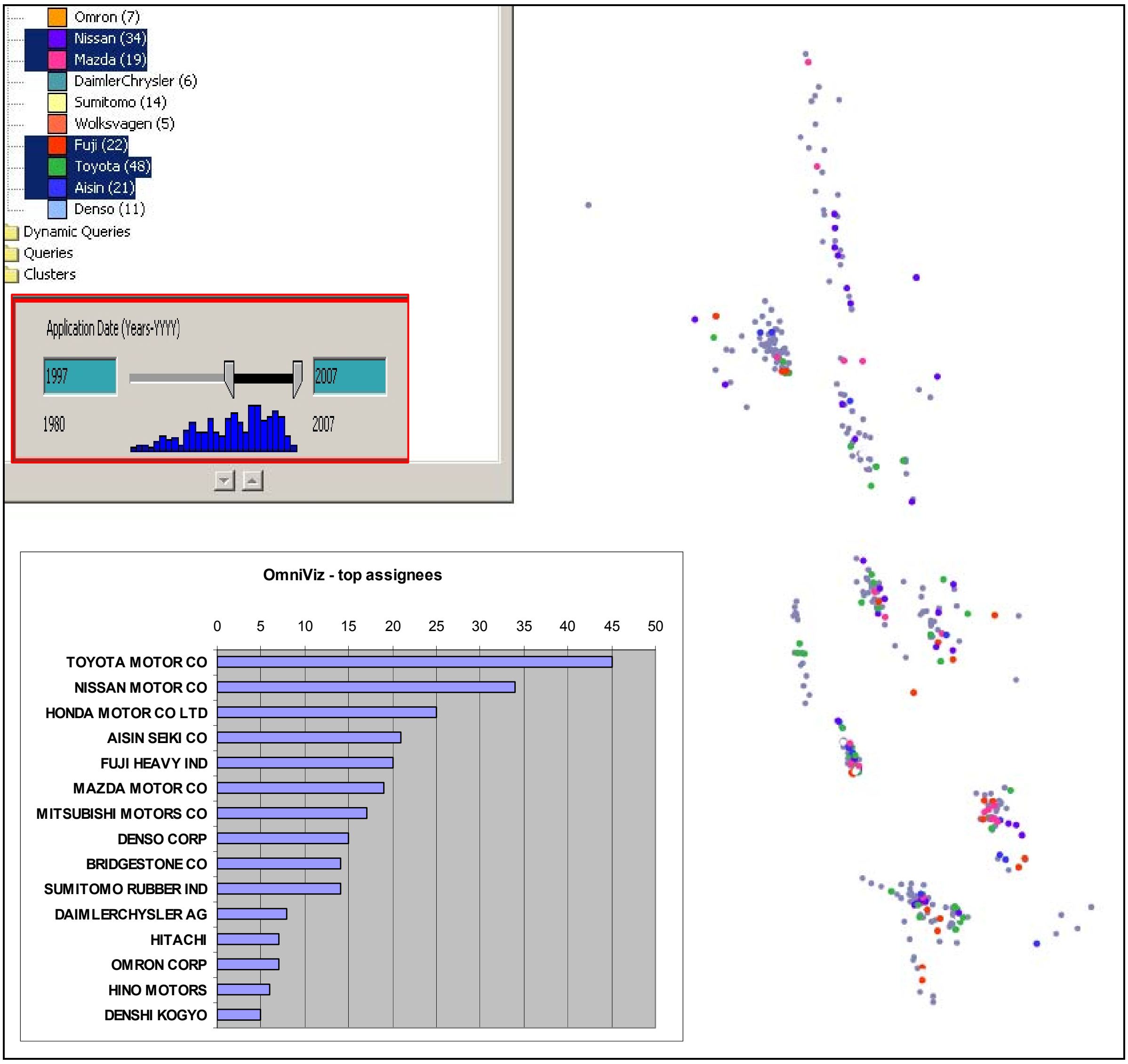


Figure 5. OmniViz: Galaxy landscape with patent documents of most active assignees coloured. In the upper-left corner is OmniViz's group tool showing the document counts of the most active assignees and the colour corresponding to the assignee. In the middle of the figure is OmniViz's Dynamic Query tool showing the yearly trends of patent application years. At the bottom is a histogram prepared with Microsoft Excel of the top patent assignees, using data imported from OmniViz.

Figure 6 and Figure 7 show the analysis made with STN AnaVist. The visualization of the patent landscape can be seen on the left in Figure 6. Light blue dots in the landscape represent individual documents. The coloured areas are clusters and the more red the cluster is, the higher its frequency of documents is. The red cluster at the bottom has the highest frequency in the data set. The two words beside the clusters indicate the two most frequent words in the cluster. A list of the ten most frequent terms may be seen by moving the cursor on the landscape. The figure may also be rotated for a better view.

STN AnaVist enables the fluent creation of basic statistics. Yearly trends in patenting may be seen from the bar graph in the lower-right corner representing the priority years. The graph in the upper-right corner shows the top patent assignees, with Toyota, Nissan and Honda leading. The graph in the middle represents geographical areas of protection. Most of the patents have been filed in the US, Japan being the second.

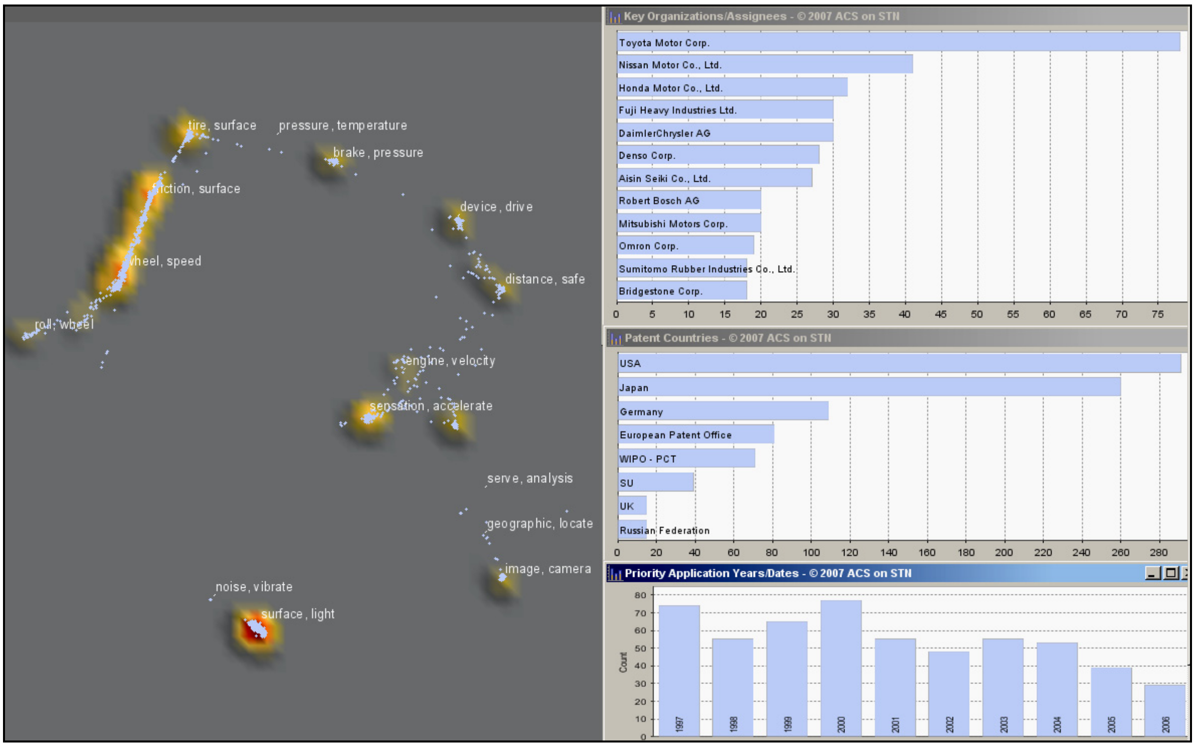
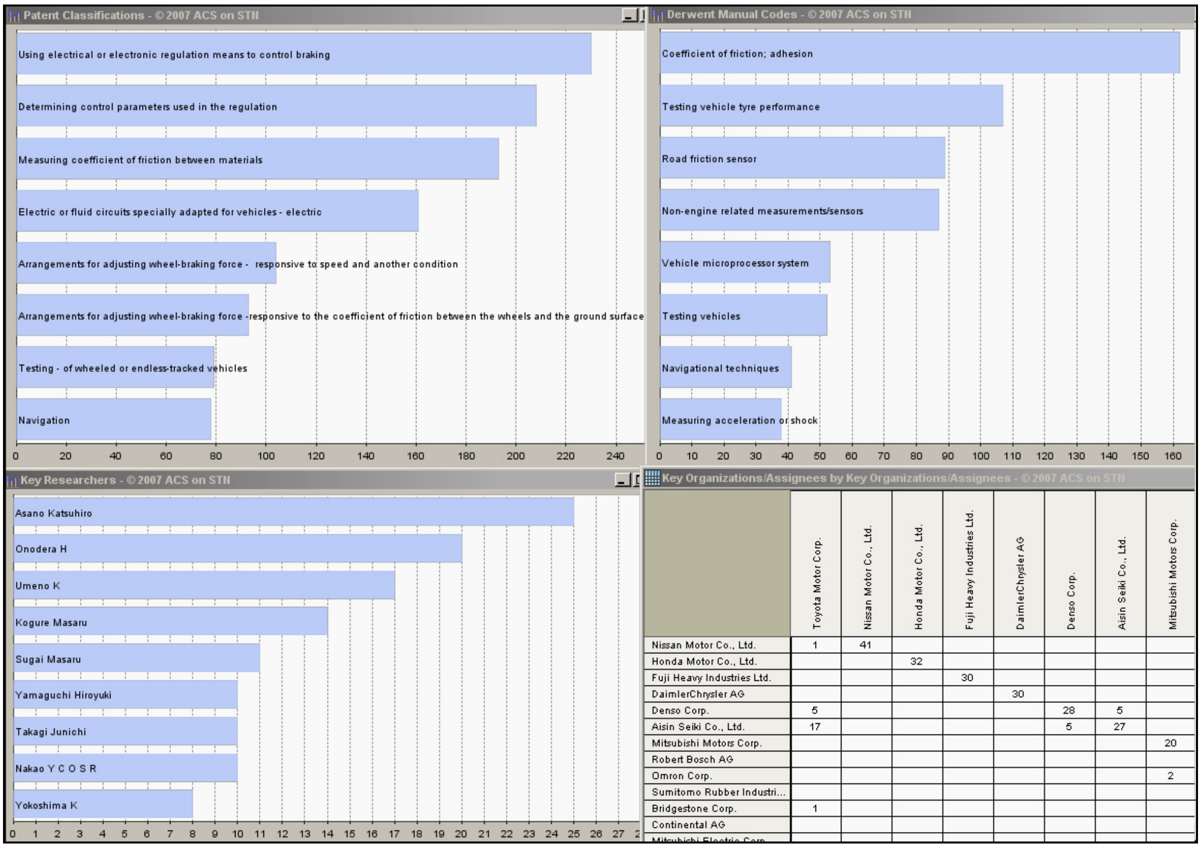


Figure 6. STN AnaVist: Visualization of patents related to measuring the friction on the road surface made with STN AnaVist. The landscape of clustered documents is on the left, yearly trends in patenting at the bottom left, the top assignees are in the upper-right corner and statistics of the countries where the applications have been made are in the middle.

Patent applications have been classified to describe their technological focus. The International Patent Classification (IPC) is an international classification provided by patent authorities. Some database producers also maintain their own classifications to help in the discovery of relevant patent documents. At the top of Figure 7 two graphs of the top classification codes are shown: IPC on the left and Derwent's Manual Codes on the right. The most used IP Classes are technologies related to controlling braking and measuring friction. The Derwent Manual Codes show more detailed technologies, in the lead are technologies related to measuring friction on the road surface. At the bottom left of the figure is a list of the top inventors, obtained from the applications. That there are two inventors with more than 20 patent applications in the area is worthy of consideration. The bottom-right corner shows a co-operation matrix of the patent assignees. The diagonal of the matrix (shifted up by one row for representational purposes) shows the number of patents for each assignee. Other cells show the number of patent applications resulting from co-operation between the assignees in corresponding rows and columns. This is an excellent way to reveal collaboration among the companies patenting on the subject.


Patent Classifications - © 2007 ACS on STN
Using electrical or electronic regulation means to control braking
Determining control parameters used in the regulation
Measuring coefficient of friction between materials
Electric or fluid circuits specially adapted for vehicles - electric
Arrangements for adjusting wheel-braking force - responsive to speed and another condition
Arrangements for adjusting wheel-braking force -responsive to the coefficient of friction between the wheels and the ground surface
Testing - of wheeled or endless-tracked vehicles
Navigation
0
20
40
60
80
100
120
140
160
180
200
220
240
Derwent Manual Codes - © 2007 ACS on STN
Coefficient of friction; adhesion
Testing vehicle tyre performance
Road friction sensor
Non-engine related measurements/sensors
Vehicle microprocessor system
Testing vehicles
Navigational techniques
Measuring acceleration or shock
0
10
20
30
40
50
60
70
80
90
100
110
120
130
140
150
160
Key Researchers - © 2007 ACS on STN
Asano Katsuhiro
Onodera H
Umeno K
Kogure Masaru
Sugai Masaru
Yamaguchi Hiroyuki
Takagi Junichi
Nakao Y C O S R
Yokoshima K
0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
Key Organizations/Assignees by Key Organizations/Assignees - © 2007 ACS on STN

	Toyota Motor Corp.	Nissan Motor Co., Ltd.	Honda Motor Co., Ltd.	Fuji Heavy Industries Ltd.	DaimlerChrysler AG	Denso Corp.	Aisin Seiki Co., Ltd.	Mitsubishi Motors Corp.
Nissan Motor Co., Ltd.	1	41						
Honda Motor Co., Ltd.			32					
Fuji Heavy Industries Ltd.				30				
DaimlerChrysler AG					30			
Denso Corp.	5					28	5	
Aisin Seiki Co., Ltd.	17					5	27	
Mitsubishi Motors Corp.								20
Robert Bosch AG								
Omron Corp.								2
Sumitomo Rubber Industri...								
Bridgestone Corp.	1							
Continental AG								
Mitsubishi Electric Corp.								

Figure 7. STN AnaVist: Analysis of patents related to measuring friction on the road surface made with STN AnaVist. The graphs show the results of basic statistics: top International Patent Classification in the top-left corner, Derwent Manual Codes on the right. The graph showing the most active inventors is in the bottom-left corner. The matrix in the bottom-right corner reveals co-operation between companies. The numbers on the diagonal (shifted up by one row) show the patents of each actor; the other numbers show the patent applications filed together by the corresponding companies.

Figure 8 shows the results of clustering prepared with Thomson Data Analyzer's Factor Map. It is a graphical representation of the results of clustering by Principal Component Analysis (PCA) used for finding frequently occurring terms in the dataset. The clusters are presented with blue circles with the most frequent term shown. The lines between the clusters represent a measure of similarity between them. The legend in the top-left corner displays information about the analysis. The number of clusters is 27 and the data coverage is 88% (meaning the percentage of documents included in any of the clusters).

Thomson Data Analyzer allows closer evaluation of the clustering by moving the cursor over the map. By selecting one cluster, the titles of all documents in it are shown. Basic statistics about the corresponding cluster may also be seen. The largest number of documents can be found in the "wet" cluster at the middle left. The boxes next to the clusters show the most frequent terms in the clusters.

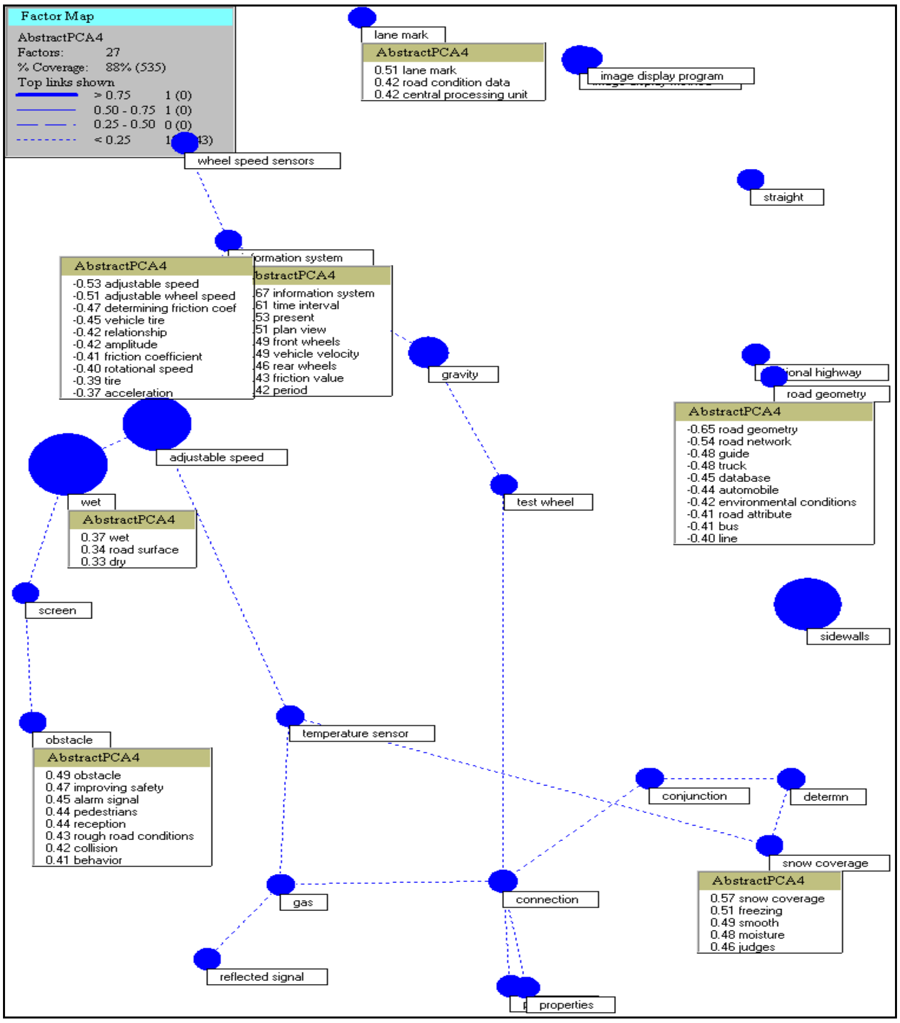


Figure 8. TDA VantagePoint: Analysis of patents related to measuring the friction on the road surface made with Thomson Data Analyzer. The figure shows a FactorMap with blue dots representing the clusters. The size of the dot corresponds to its frequency of documents. The words describing the clusters are the most frequently occurring terms in each cluster.

Figure 9 shows the top patent assignees graph and yearly trends made with Thomson Data Analyzer's Technology Report. Technology Report is a predefined report containing many basic and value-added statistics on the subject of technology. Some of the information offered by the reporting tool would be quite hard to get by making the analysis manually, i.e. unique technology indicators (classifications for top assignees). At the top is a graph representing the number of applications filed by the top patent assignees, and at the bottom are yearly trends in patenting.

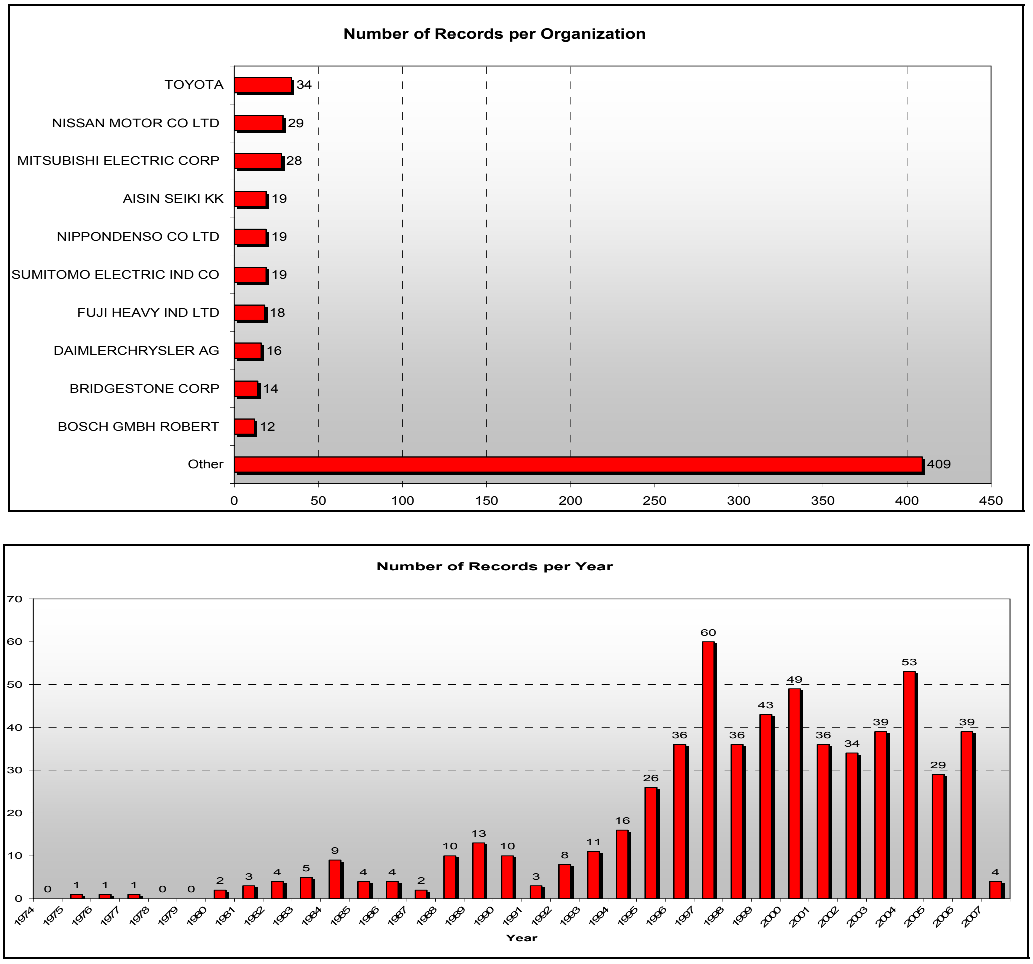


Figure 9. TDA VantagePoint: Analysis of patents related to measuring friction on the road surface made with Thomson Data Analyzer. The figure shows the top patent assignees and yearly trends in the field with priority years.

Summary

The results obtained from analyses made with all the tools were very similar. They gave almost the same top 12 assignees; only their reciprocal order varied. Toyota and Nissan seem to be the most significant actors in the field. The only deviation was that Honda was missing from Thomson Data Analyzer's top companies list. The yearly trend graphs show accelerated patenting since the mid-1990s, peaks in the late 1990s and early 2000s, and a slight decrease at the present moment.

4.2 Closer Inspection of Specific Technology

Visualizations often reveal areas of technology that appear more interesting than others and need closer investigation. Aureka allows the user to select an area and save the documents in it into a new dataset for further analysis. In Thomson Data Analyzer, documents having some interesting common feature may be incorporated into a group and then analyzed as a subset. STN AnaVist and OmniViz enable the selection of documents by defining the relevant clusters straight from the visualizations. All the same operations as presented before may then be performed on the analyzed subset.

Figure 10 shows OmniViz's CoMet visualization for evaluating which areas companies are focused on. CoMet is a tool for analyzing associations between entities. The columns of the visualization in the figure represent data about each patent assignee, and the rows are the topics occurring in the data. Documents related to measuring the surface of the road with light were looked at more closely. By choosing the term "light" (shown with the row rounded with green lines) assignees who have filed for patents related to the subject (shown with yellow) could be seen. Due to the interactivity the documents containing the term are coloured with red in the Galaxy visualization shown on the right.

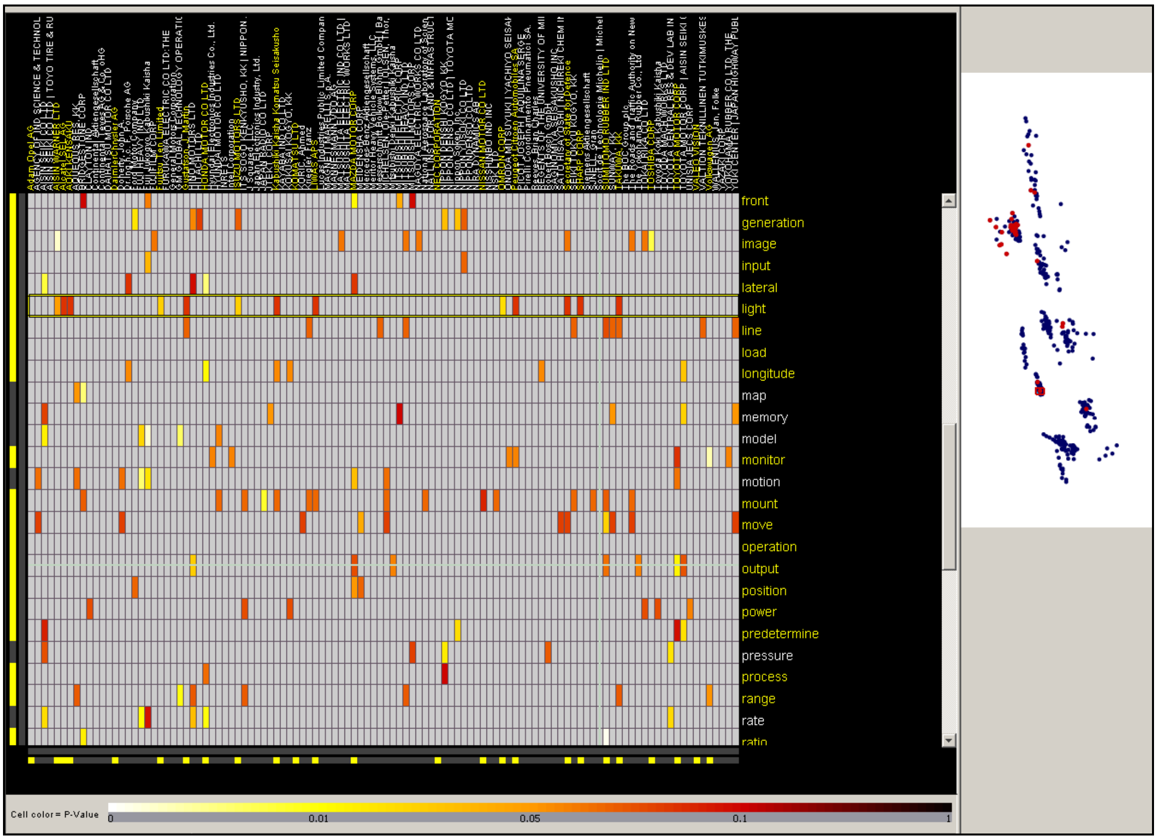


Figure 10. OmniViz: Closer investigation of specific technology area. The rows show the major terms appearing to be significant during the clustering and the columns are patent assignees. By highlighting the term "light" patent assignees having applications related to the subject are coloured with yellow. A Galaxy visualization showing the documents containing the terms is on the right.

Figure 11 shows a closer investigation of an interesting technology area made with STN AnaVist. Three clusters seeming to be the most relevant to the subject were selected. The clusters can be found in the top-left corner of the landscape and are defined with terms "tire,surface"; "friction,surface" and "wheel,speed". The frequencies of documents in the selected area are shown in green in the statistics graphs and other documents are in light blue. The statistics have been ordered by highlighted documents. The graph at the top right shows the top assignees in the specific area, with top IPCs in the middle and yearly trends in patenting at the bottom. Yearly trends in patenting in the specific area have followed technology-wide patenting, which can be clearly seen in the priority year graph.

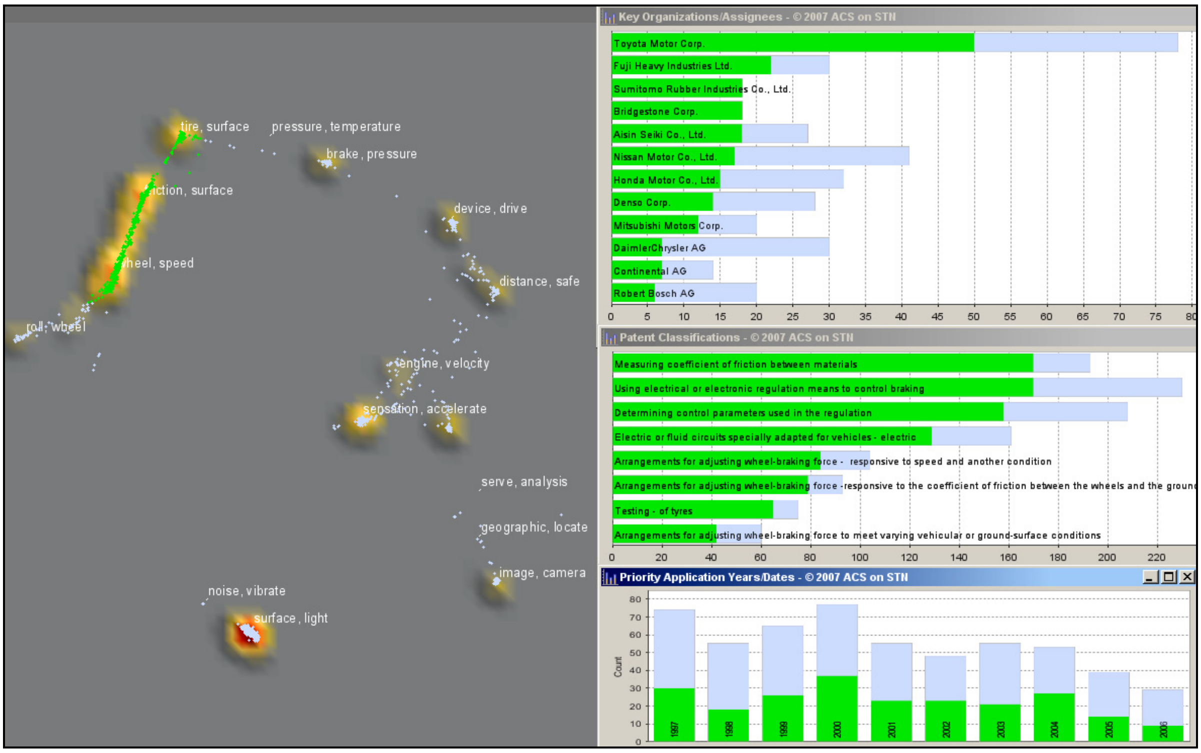
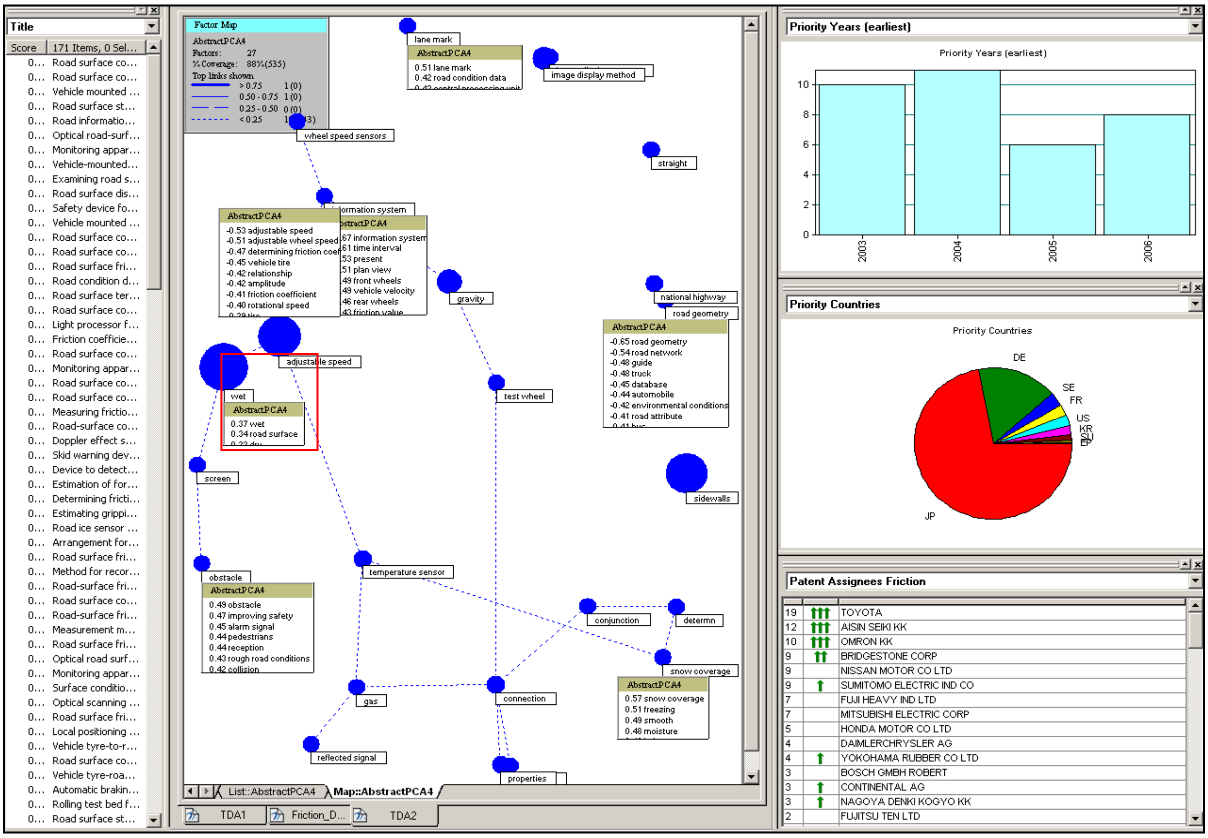

tire, surface
pressure, temperature
brake, pressure
iction, surface
heel, speed
roll, wheel
device, drive
distance, safe
engine, velocity
sensation, accelerate
serve, analysis
geographic, locate
image, camera
noise, vibrate
surface, light
Key Organizations/Assignees - © 2007 ACS on STN
Toyota Motor Corp.
Fuji Heavy Industries Ltd.
Sumitomo Rubber Industries Co., Ltd.
Bridgestone Corp.
Aisin Seiki Co., Ltd.
Nissan Motor Co., Ltd.
Honda Motor Co., Ltd.
Denso Corp.
Mitsubishi Motors Corp.
DaimlerChrysler AG
Continental AG
Robert Bosch AG
0 5 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80
Patent Classifications - © 2007 ACS on STN
Measuring coefficient of friction between materials
Using electrical or electronic regulation means to control braking
Determining control parameters used in the regulation
Electric or fluid circuits specially adapted for vehicles - electric
Arrangements for adjusting wheel-braking force - responsive to speed and another condition
Arrangements for adjusting wheel-braking force -responsive to the coefficient of friction between the wheels and the ground
Testing - of tyres
Arrangements for adjusting wheel-braking force to meet varying vehicular or ground-surface conditions
0 20 40 60 80 100 120 140 160 180 200 220
Priority Application Years/Dates - © 2007 ACS on STN
Count
0 10 20 30 40 50 60 70 80
1997 1998 1999 2000 2001 2002 2003 2004 2005 2006

Figure 11. STN AnaVist: Closer investigations of specific technology area done with STN AnaVist. Patent map, on left with the clusters of the examined area selected and coloured in green. Corresponding occurrences are coloured with green in the other graphs also, with top assignees in the top-right corner, Patent Classification Codes in the middle and document frequencies by priority years at the bottom.

Figure 12 shows an analysis of the cluster “wet” in FactorMap by Thomson Data Analyzer, introduced in more detail in Chapter 4.1. By clicking on the relevant cluster, statistics from the corresponding documents may be withdrawn. Document titles can be seen and selected for closer examination on the left side of the map. The right side of the map shows some basic statistics of the data. The bar graph in the upper corner shows the frequency of documents by priority year. The first patents included in the cluster were filed in 2003. The graph shows some decrease in interest in the technology area. The pie chart in the middle shows the geographical area of protection, with Japan seeming to be the most important market area. The list in the lower corner introduces the most active assignees in the area. The green arrows after assignee names represent surprisingly high frequencies of the item, i.e. strong items. The three green arrows in TDA’s lists show that the number of occurrences for the corresponding actor is much higher than the expected value. The arrows shown for Toyota mean that Toyota is a very strong actor in techniques related to measuring the wet surface of the road.



Score	171 Items, 0 Sel...
0...	Road surface co...
0...	Road surface co...
0...	Vehicle mounted ...
0...	Road surface st...
0...	Road informatio...
0...	Optical road-surf...
0...	Monitoring appar...
0...	Vehicle-mounted...
0...	Examining road s...
0...	Road surface dis...
0...	Safety device fo...
0...	Vehicle mounted ...
0...	Road surface co...
0...	Road surface co...
0...	Road surface fri...
0...	Road condition d...
0...	Road surface ter...
0...	Road surface co...
0...	Light processor f...
0...	Friction coefficie...
0...	Road surface co...
0...	Monitoring appar...
0...	Road surface co...
0...	Road surface co...
0...	Measuring frictio...
0...	Road-surface co...
0...	Doppler effect s...
0...	Skid warning dev...
0...	Device to detect...
0...	Estimation of for...
0...	Determining fricti...
0...	Estimating grippi...
0...	Road ice sensor ...
0...	Arrangement for...
0...	Road surface fri...
0...	Method for recor...
0...	Road-surface fri...
0...	Road surface co...
0...	Road-surface fri...
0...	Measurement m...
0...	Road surface fri...
0...	Optical road surf...
0...	Monitoring appar...
0...	Surface conditio...
0...	Optical scanning ...
0...	Road surface fri...
0...	Local positioning ...
0...	Vehicle tyre-to-r...
0...	Road surface co...
0...	Vehicle tyre-roa...
0...	Automatic brakin...
0...	Rolling test bed f...
0...	Road surface st...

19	↑↑↑	TOYOTA
12	↑↑↑	AISIN SEIKI KK
10	↑↑↑	OMRON KK
9	↑↑	BRIDGESTONE CORP
9		NISSAN MOTOR CO LTD
9	↑	SUMITOMO ELECTRIC IND CO
7		FUJI HEAVY IND LTD
7		MITSUBISHI ELECTRIC CORP
5		HONDA MOTOR CO LTD
4		DAIMLERCHRYSLER AG
4	↑	YOKOHAMA RUBBER CO LTD
3		BOSCH GMBH ROBERT
3	↑	CONTINENTAL AG
3	↑	NAGOYA DENKI KOGYO KK
2		FUJITSU TEN LTD

Figure 12. TDA VantagePoint: Visualization of patents related to measuring friction on the road surface made with Thomson Data Analyzer. Examination of the cluster “wet” at the left middle of the FactorMap. The left side shows the titles of documents included in the cluster; the top right corner shows the frequency of documents for each priority year present; in the middle is a pie graph of patent authorities and in the bottom corner is a list of the top assignees. Strong assignees in the area are marked with green arrows.

Summary

All four tools enabled closer investigation and retrieval of basic statistics of a specific technology area. In Aureka this was done by preparing a new visualization of the chosen documents while the others enabled it interactively by only outlining the area in question.

4.3 Yearly Trends in Patenting

Discovering trends in filing for patents reveals new "hot areas" in R&D and shows technologies that have become less interesting or have been abandoned.

Figure 13 shows a patent landscape visualized with Aureka. A group of documents filed from 2004 to 2005 was made with Aureka's *Time Slice* tool and coloured them in red. The figure reveals that there are quite few patent documents in the above-mentioned application years though no strong statistical conclusions may be drawn. Two technologies are shown to have clumping of documents: braking techniques and determining tire pressure.

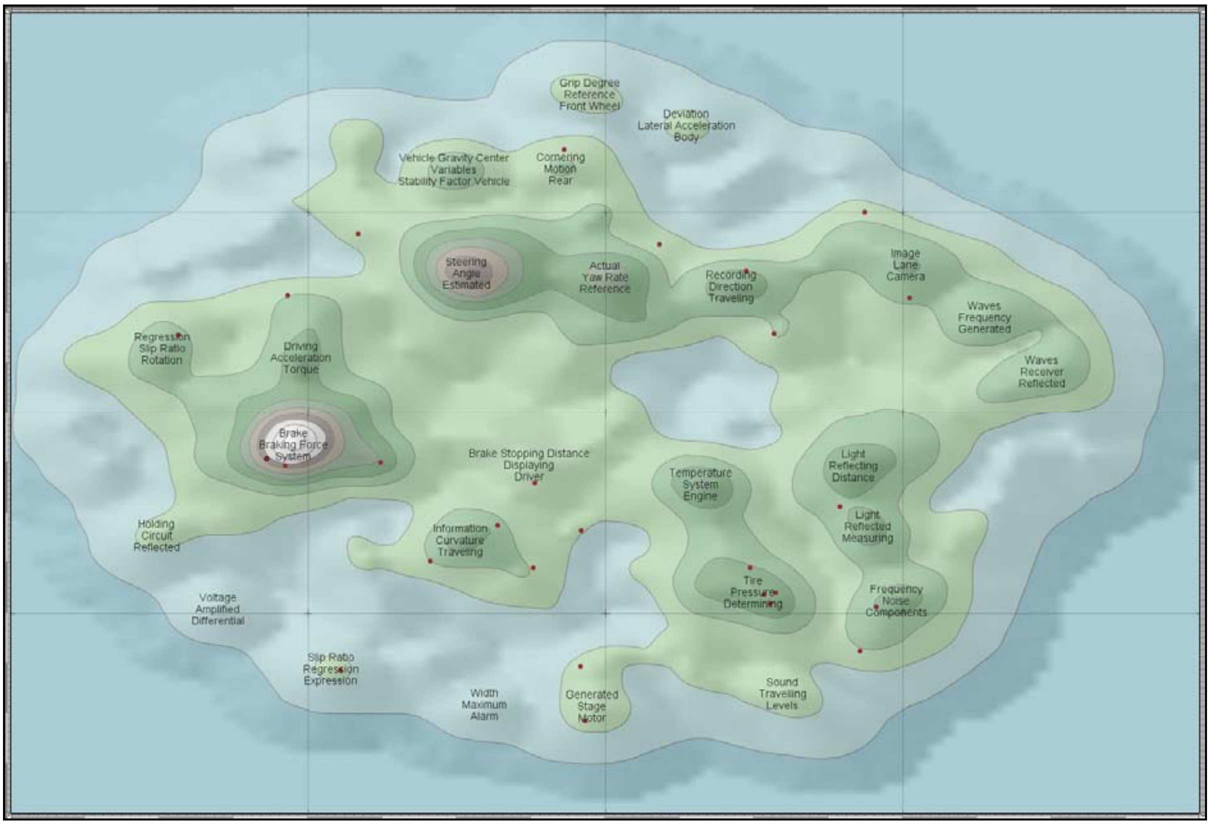

Grip Degree Reference Front Wheel
Deviation Lateral Acceleration Body
Vehicle Gravity Center Variables Stability Factor Vehicle
Cornering Motion Rear
Steering Angle Estimated
Actual Yaw Rate Reference
Recording Direction Traveling
Image Lane Camera
Waves Frequency Generated
Waves Receiver Reflected
Regression Slip Ratio Rotation
Driving Acceleration Torque
Brake Braking Force System
Brake Stopping Distance Displaying Driver
Temperature System Engine
Light Reflecting Distance
Light Reflected Measuring
Holding Circuit Reflected
Information Curvature Traveling
Tire Pressure Determining
Frequency Noise Components
Voltage Amplified Differential
Slip Ratio Regression Expression
Width Maximum Alarm
Generated Stage Motor
Sound Travelling Levels

Figure 13. Aureka: Visualization of patents related to measuring friction of the road surface made with Aureka's ThemeMap. Patent applications filed during 2004 and 2005 are colored in red.

Figure 14 shows an OmniViz Galaxy visualization with the documents distributed into three groups by the priority year. The labels on the left show the colours used for each group and the visualization on the right shows the distribution of corresponding documents. The cluster in the middle of the visualization named “tire,surface,device” shows some piling up of new documents which might be considered for closer evaluation.

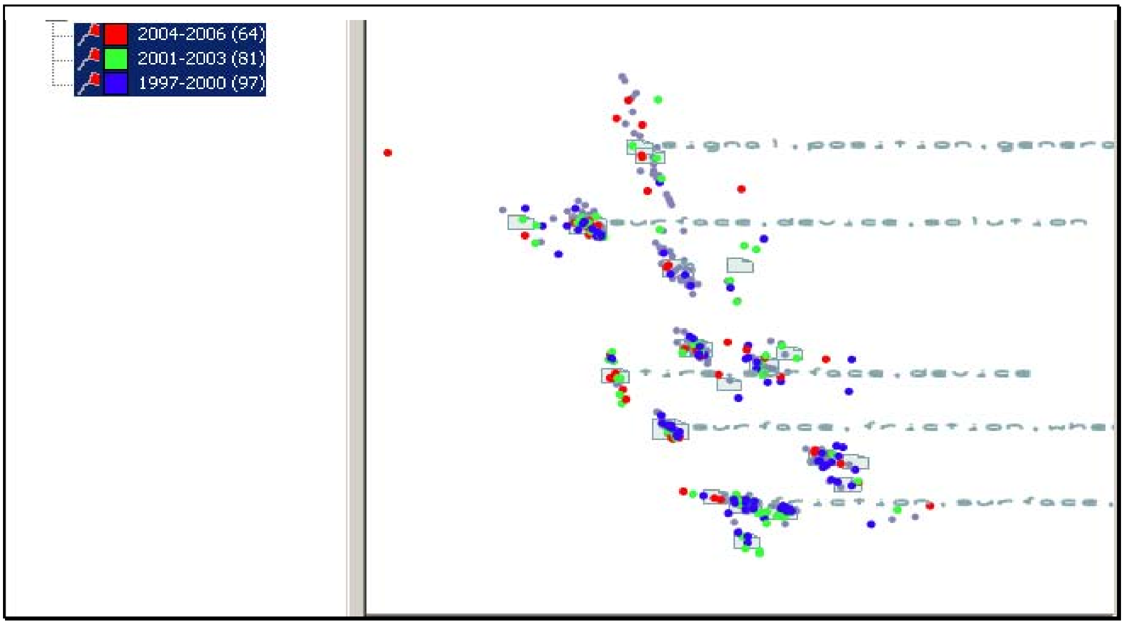


Figure 14. OmniViz: Visualization of patents related to measuring friction on the road surface. Patents included in one of three time periods coloured with a colour corresponding to the period.

Figure 15 shows an analysis conducted with STN AnaVist. Three time periods were selected from the priority year graph at the bottom for closer study: 1997 to 2000, 2001 to 2003 and 2004 to 2006. The documents in the patent landscape on the left are automatically coloured with the corresponding colours. The top patent assignee and IPC graphs on the right side also express the spread of the documents during the time periods. The top patent assignees nowadays seem to be the same as at all times previously. The assignee graph identifies newcomers to the market: General Motors and Alpine. It also shows fading development in the area from most of the companies, especially Fuji and Omron.

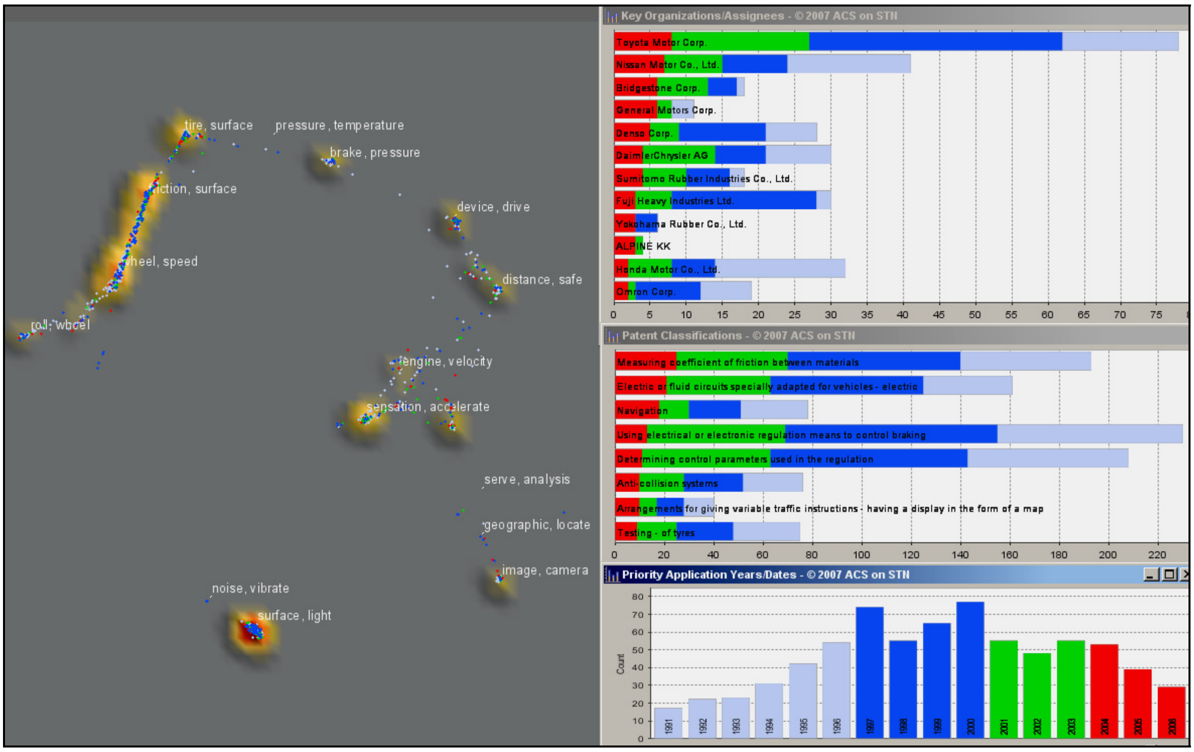


Figure 15. STN AnaVist: Visualization of patents related to measuring friction on the road surface. The figure represents the segmentation of documents according to the priority year periods shown in the graph in the lower-right corner. On the left is a patent landscape with documents coloured with colours corresponding to the year periods; the top right shows patent assignees and in the middle are International Patent Classifications.

Thomson Data Analyzer allows the user to create groups according to the years found in documents. Basic statistics of assignees, inventors etc. may be drawn, as well as a completely new analysis using the groups as data sets, as in other tools too. Thomson Data Analyzer also provides special tools for making a more progressive analysis, such as revealing "hot areas" of patenting. The means and results of such an analysis are presented for the Fraunhofer data in the next chapter.

Summary

None of the four tools revealed any distinct trends in patenting. All clusters made by the tools had a uniform spread of documents; there were no clearly abandoned technology areas or new "hot areas".

4.4 Comparing the Patent Portfolio of Two Companies in the Technology Field

Comparing patent portfolios of two actors gives valuable information of their differences and similarities in R&D and business. This method could also be employed for technological benchmarking by comparing a company's own patent portfolio to somebody else's. The analyses presented in this subchapter introduce the means offered from the tools tested by comparing the portfolios of two assignees, Toyota and Bridgestone, both of which have many patents in the data set.

Figure 16 shows a patent landscape made with Aureka. Toyota's patents are coloured with green and Bridgestone's with red. Aureka offers filters for exporting data from analysis to Microsoft Excel and macros for creating graphs there. The bar graph in the upper-right corner presents the yearly trends for both companies. Bridgestone has entered the area quite recently and has concentrated mostly on one technology area, which can be seen from the red dots on the lowest part of the landscape.

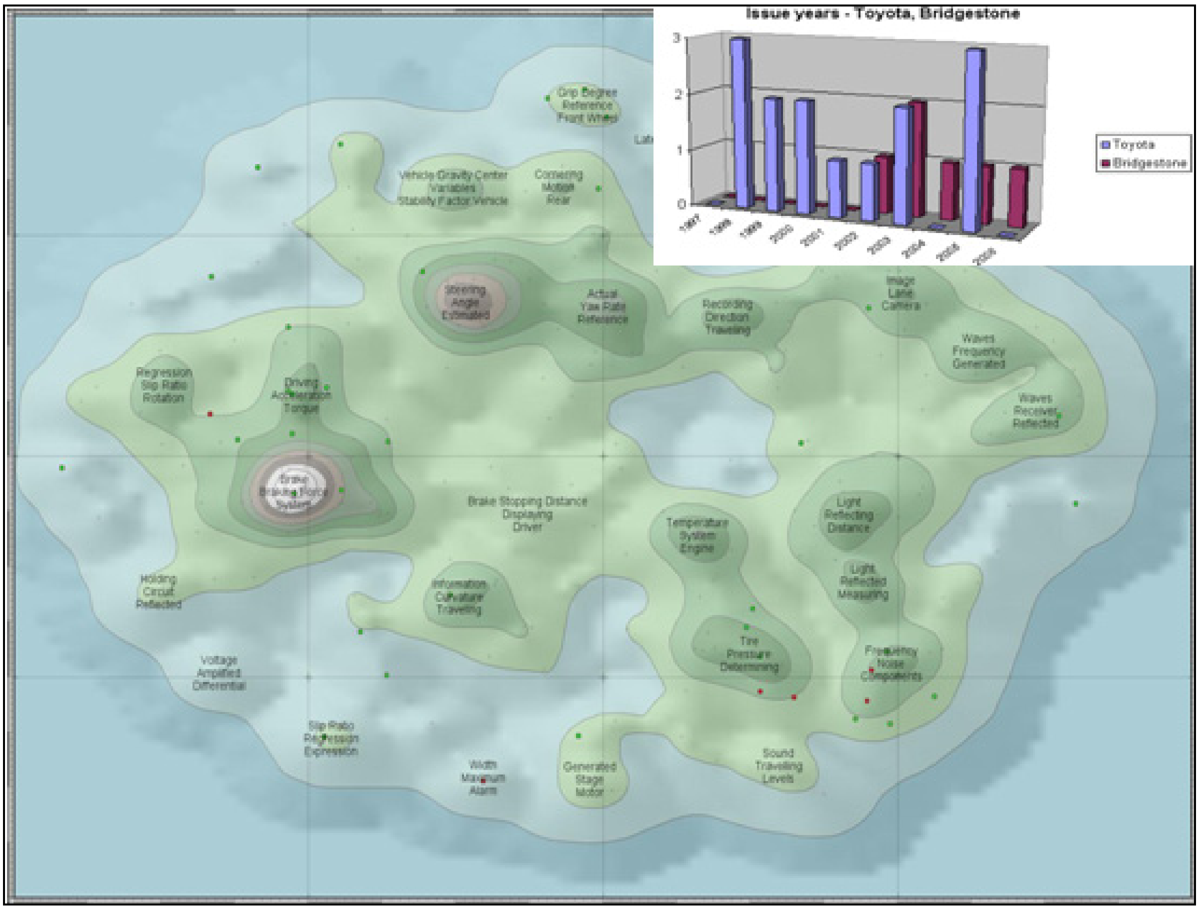
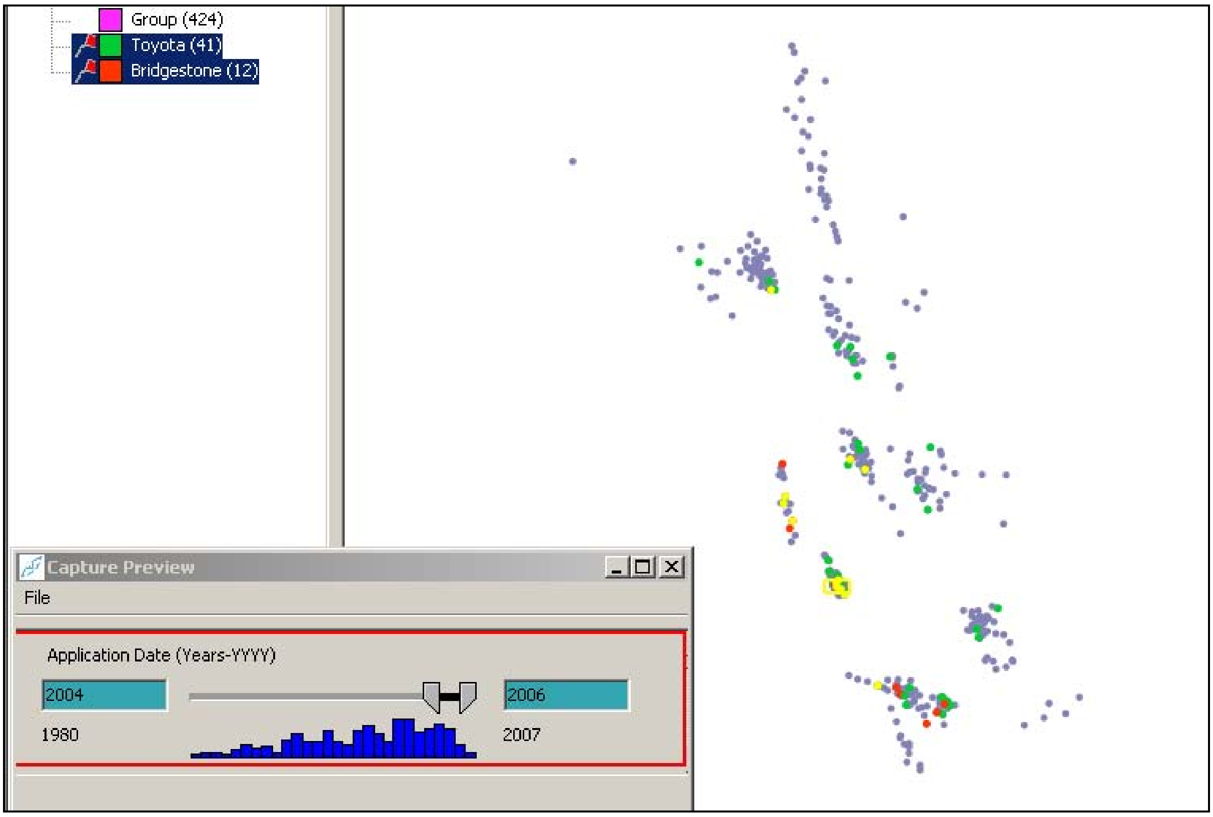

Issue years - Toyota, Bridgestone
3
2
1
0
1997
1998
1999
2000
2001
2002
2003
2004
2005
2006
Toyota
Bridgestone
Grip Degree Reference Front Wheel
Lat
Vehicle Gravity Center Variables Stability Factor Vehicle
Cornering Motion Rear
Steering Angle Estimated
Actual Yaw Rate Reference
Recording Direction Traveling
Image Lane Camera
Waves Frequency Generated
Waves Receiver Reflected
Regression Slip Ratio Rotation
Driving Acceleration Torque
Brake Braking Force System
Brake Stopping Distance Displaying Driver
Temperature System Engine
Light Reflecting Distance
Light Reflected Measuring
Holding Circuit Reflected
Information Curvature Traveling
Tire Pressure Determining
Frequency Noise Components
Voltage Amplified Differential
Slip Ratio Regression Expression
Width Maximum Alarm
Generated Stage Motor
Sound Traveling Levels

Figure 16. Aureka: Visualization of patents related to measuring friction on the road surface. Patent landscape with Toyota's patents coloured with green and Bridgestone's red. The graph in the upper-right corner shows the patent issue yearly trends for both companies.

Figure 17 shows OmniViz's Galaxy visualization with Toyota's and Bridgestone's patents collected in their own groups and the documents in both groups coloured with corresponding colours. The group tool in the upper-left corner shows the colours used, green for Toyota and red for Bridgestone, and the number of documents in both groups. Toyota has filed 41 patents in this technology area over the years and Bridgestone 12.

Application years from 2004 to 2006 were selected with the dynamic query tool shown in the lower-left corner. This caused the corresponding Toyota and Bridgestone documents to be coloured with yellow. This method loses some information when the image is printed, though it works well when using the animation feature online.



🟪 Group (424)
🟩 Toyota (41)
🟥 Bridgestone (12)

Application Date (Years-YYYY)
2004 – 2006
1980 – 2007

Figure 17. OmniViz: Visualization of patents related to measuring friction on the road surface. First the patents assigned by Toyota are coloured in green, and those by Bridgestone in red. Then the documents filed between 2004 and 2006 have been coloured in yellow.

Figure 18 represents a comparison of Toyota's and Bridgestone's patents made with STN AnaVist. The two assignees have been selected by clicking their names in the Key Organizations/Assignee graph in the top-right corner. The patent applications filed by Toyota are shown in green, those by Bridgestone in red and one document of an invention made in collaboration in1998 in blue. The key assignees chart also shows co-operation between Toyota and other assignees, shown with green bars. Bridgestone's patents cover the same technological area, while Toyota is patenting widely different techniques, as can be seen from the landscape. The priority year graph in the figure tells us that the co-operation has been going on for quite a long time and is still continuing. Patent country statistics reveal the geographical area of protection and the International Patent Classifications at the bottom show more specific areas of research, development and collaboration.

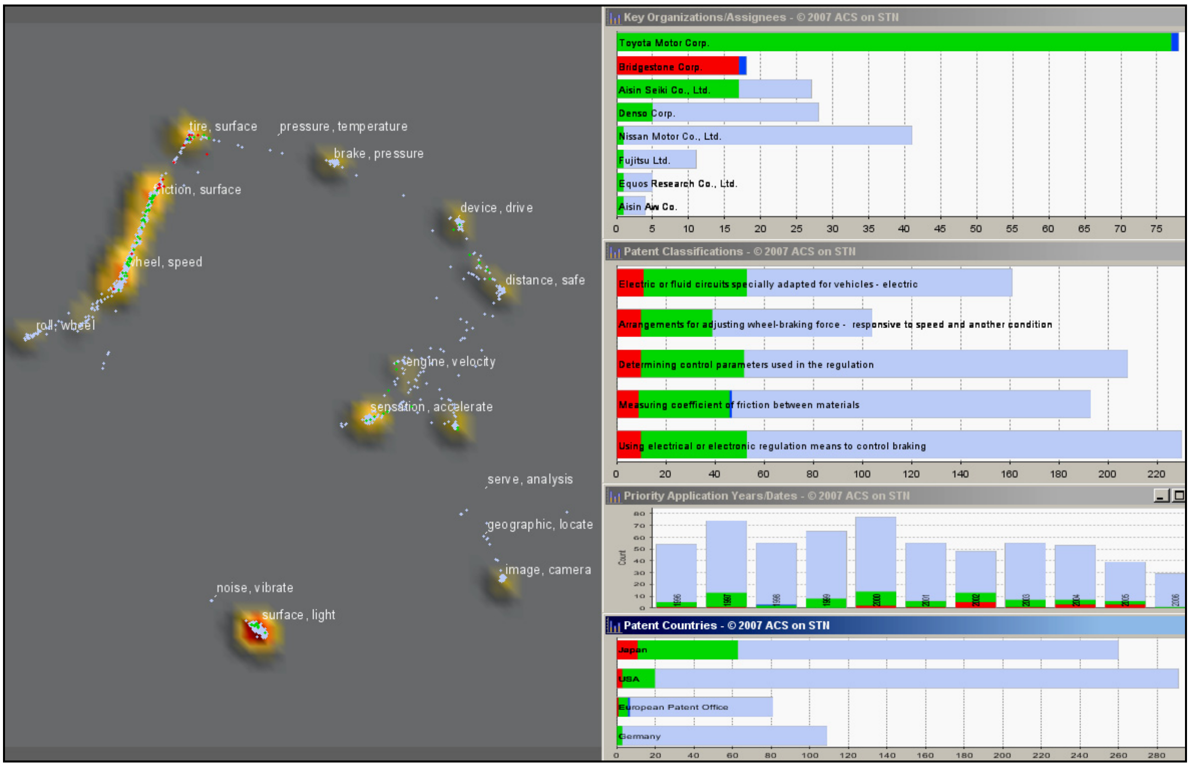
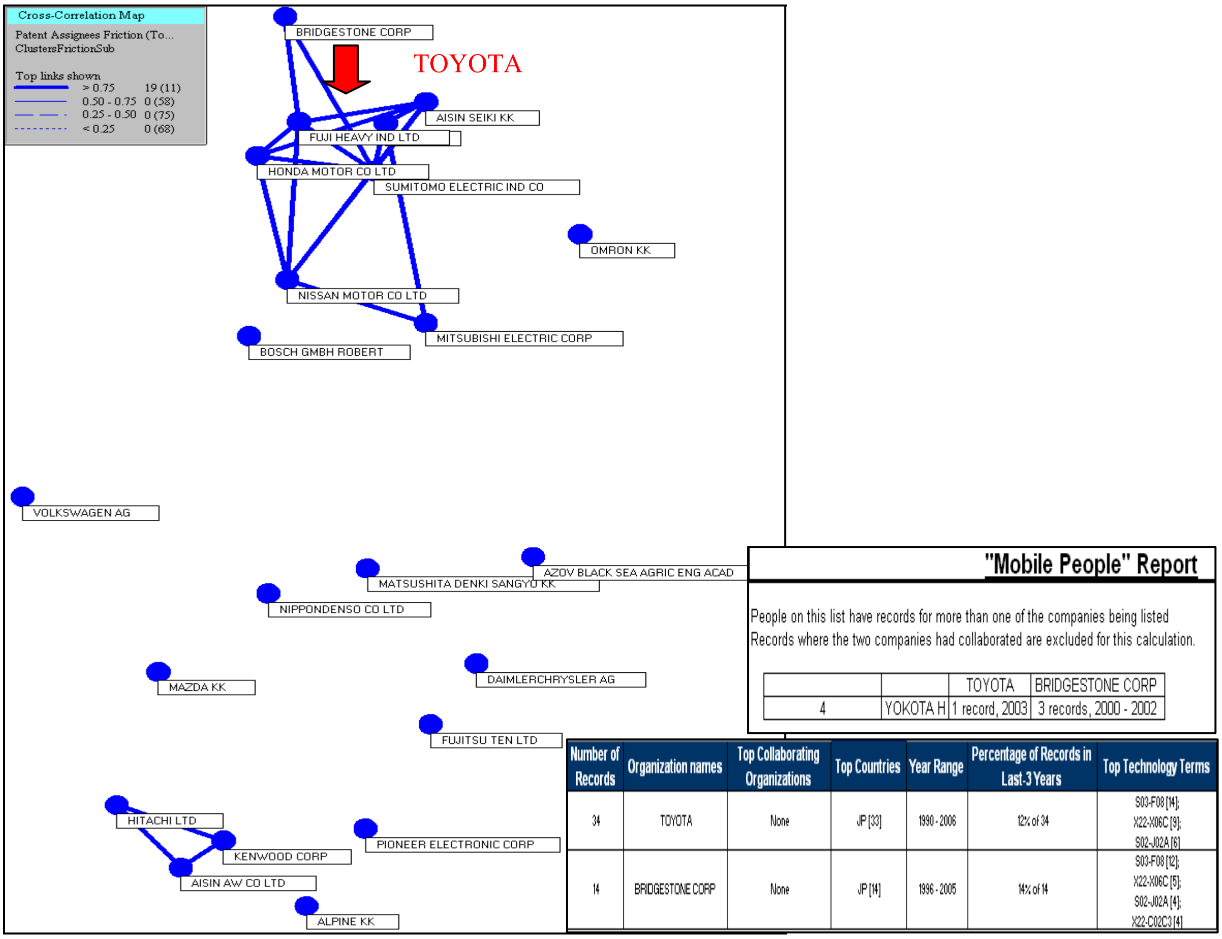


Figure 18. STN AnaVist: Visualization of patents related to measuring friction on the road surface. Comparison of Toyota's (in green) and Bridgestone's (in red) patents by colouring patent applications filed in collaboration in the upper-right corner. The corresponding documents are coloured in the landscape on the left, International Patent Classifications under patent assignees, priority years below, and patent countries at the bottom.

Figure 19 shows the results of different analyses made with Thomson Data Analyzer. On the left is a Cross-Correlation Map representing relationships among the most active assignees over the patent data set. It reveals which companies are working on similar areas by the terms used in patent applications. Strong equivalencies are shown with lines connecting clusters. Accumulation of clusters can be seen at the top of the map. Closer investigation of the assignees might reveal if they are competitors acting in the same area or companies with strong collaborations. Thomson Data Analyzer also offers tools for preparing different reports. One of them is a Company Comparison report providing a variety of statistics for up to 5 companies. At the bottom right is a table taken from the report, showing some basic metrics for both companies under examination. It gives information about the number of patent documents found in the data set, priority year range, percentage of new applications filed and top technology terms from Derwent Classifications. Above it is a table revealing the names of inventors who have worked for both companies; applications made in collaboration are excluded.



"Mobile People" Report

People on this list have records for more than one of the companies being listed
Records where the two companies had collaborated are excluded for this calculation.

		TOYOTA	BRIDGESTONE CORP
4	YOKOTA H	1 record, 2003	3 records, 2000 - 2002

Number of Records	Organization names	Top Collaborating Organizations	Top Countries	Year Range	Percentage of Records in Last-3 Years	Top Technology Terms

34	TOYOTA	None	JP [33]	1990 - 2006	12% of 34	S03-F08 [14];
						X22-X06C [9];
						S02-J02A [6]

14	BRIDGESTONE CORP	None	JP [14]	1996 - 2005	14% of 14	S03-F08 [12];
						X22-X06C [5];
						S02-J02A [4];
						X22-C02C3 [4]

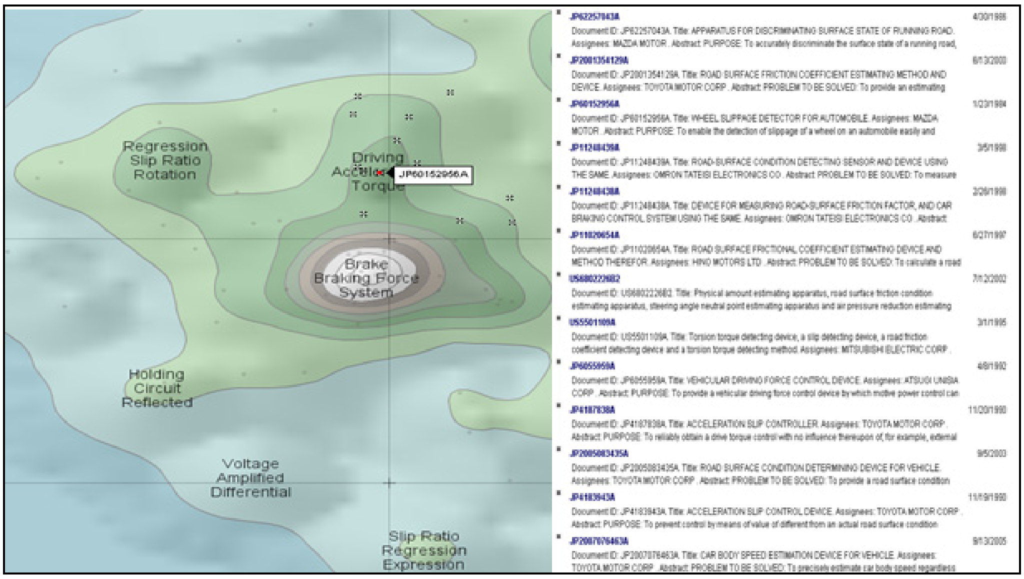
Figure 19. TDA VantagePoint: Analysis of patents related to measuring friction on the road surface. The CrossCorrelation Map on the left represents relationships between the most active assignees in the patent data set. On the right are tables from the Company Comparison report showing statistics of Toyota's and Bridgestone's patents in this area. The Mobile People report shows the names of inventors employed by both companies and the table below reveals basic information about the patent portfolios of each company.

Summary

Comparison of the two companies showed that Bridgestone has concentrated on narrower areas of technology than Toyota, which was found by all four tools. The year type used in the analysis with MicroPatent was the issue year, whereas with the others it was the application or priority year. The patent is usually issued (and receives the issue year) from three to five years after filing the application, when with some terms the application or priority year is given. This implies parallel results from comparison of two actors with all four tools. Unique information was obtained with STN AnaVist and Thomson Data Analyzer, too. STN AnaVist revealed one patent filed in collaboration and Thomson Data Analyzer identified a researcher who has worked for both companies, first for Bridgestone and then for Toyota.

4.5 Patenting Around One Significant Invention

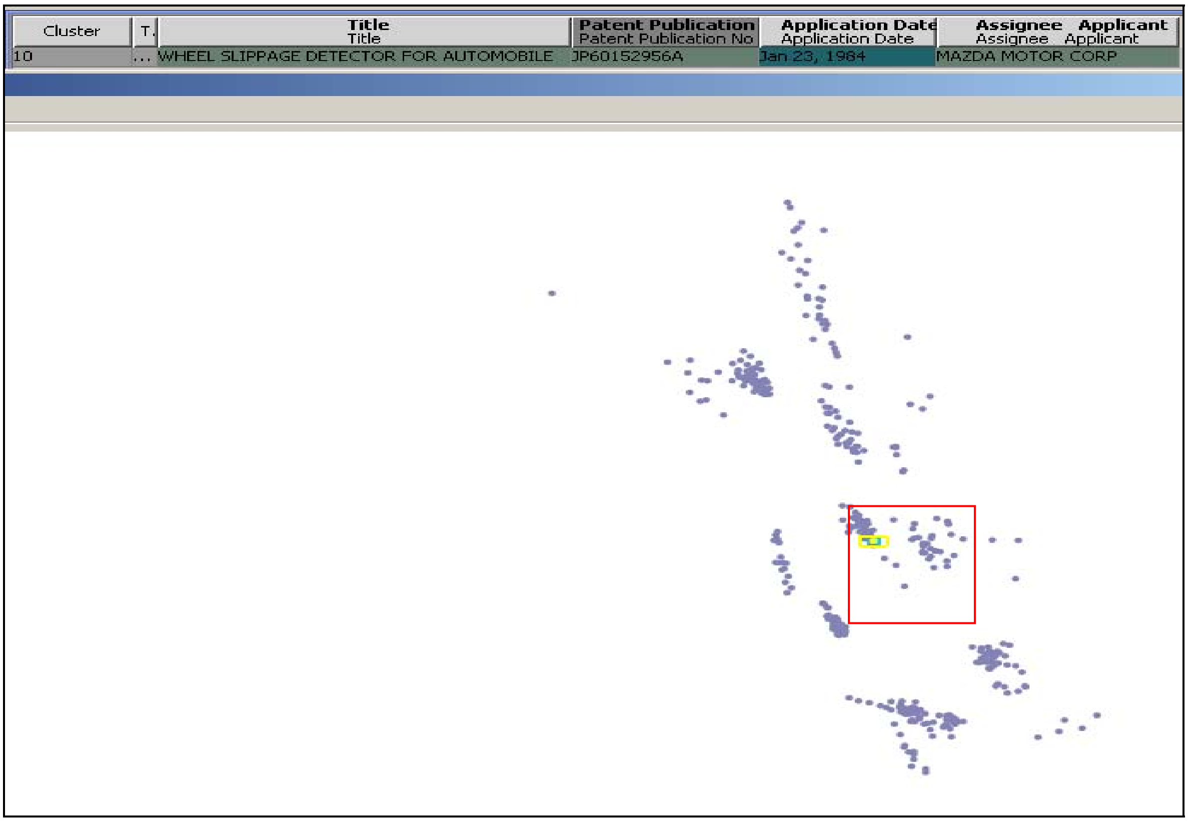
Sometimes it is of great importance to evaluate how a specific document relates to other patents in the same technology area. This might be the case, for example, when planning on further developing an already invented technique and evaluating how many patents there are in the area, or trying to locate partners for collaboration in a specific technological field. The ability of the tools to locate a specific patent application known by title was tested. Figure 20 shows a part of the Aureka landscape on left and the document window on the right. The document window shows details of patent documents from an area selected in the landscape. The location of a specific patent document is shown with an arrow and patent number on Landscape by clicking the corresponding patent in the Document window.



- JP62257043A 4/30/1986
  Document ID: JP62257043A. Title: APPARATUS FOR DISCRIMINATING SURFACE STATE OF RUNNING ROAD. Assignees: MAZDA MOTOR . Abstract: PURPOSE: To accurately discriminate the surface state of a running road,
- JP2001354129A 6/13/2000
  Document ID: JP2001354129A. Title: ROAD SURFACE FRICTION COEFFICIENT ESTIMATING METHOD AND DEVICE. Assignees: TOYOTA MOTOR CORP . Abstract: PROBLEM TO BE SOLVED: To provide an estimating
- JP60152956A 1/23/1984
  Document ID: JP60152956A. Title: WHEEL SLIPPAGE DETECTOR FOR AUTOMOBILE. Assignees: MAZDA MOTOR . Abstract: PURPOSE: To enable the detection of slippage of a wheel on an automobile easily and
- JP11248439A 3/5/1998
  Document ID: JP11248439A. Title: ROAD-SURFACE CONDITION DETECTING SENSOR AND DEVICE USING THE SAME. Assignees: OMRON TATEISI ELECTRONICS CO . Abstract: PROBLEM TO BE SOLVED: To measure
- JP11248438A 2/26/1998
  Document ID: JP11248438A. Title: DEVICE FOR MEASURING ROAD-SURFACE FRICTION FACTOR, AND CAR BRAKING CONTROL SYSTEM USING THE SAME. Assignees: OMRON TATEISI ELECTRONICS CO . Abstract:
- JP11020654A 6/27/1997
  Document ID: JP11020654A. Title: ROAD SURFACE FRICTIONAL COEFFICIENT ESTIMATING DEVICE AND METHOD THEREFOR. Assignees: HINO MOTORS LTD . Abstract: PROBLEM TO BE SOLVED: To calculate a road
- US6802226B2 7/12/2002
  Document ID: US6802226B2. Title: Physical amount estimating apparatus, road surface friction condition estimating apparatus, steering angle neutral point estimating apparatus and air pressure reduction estimating
- US5501109A 3/1/1995
  Document ID: US5501109A. Title: Torsion torque detecting device, a slip detecting device, a road friction coefficient detecting device and a torsion torque detecting method. Assignees: MITSUBISHI ELECTRIC CORP .
- JP6055959A 4/8/1992
  Document ID: JP6055959A. Title: VEHICULAR DRIVING FORCE CONTROL DEVICE. Assignees: ATSUGI UNISIA CORP . Abstract: PURPOSE: To provide a vehicular driving force control device by which motive power control can
- JP4187838A 11/20/1990
  Document ID: JP4187838A. Title: ACCELERATION SLIP CONTROLLER. Assignees: TOYOTA MOTOR CORP . Abstract: PURPOSE: To reliably obtain a drive torque control with no influence thereupon of, for example, external
- JP2005083435A 9/5/2003
  Document ID: JP2005083435A. Title: ROAD SURFACE CONDITION DETERMINING DEVICE FOR VEHICLE. Assignees: TOYOTA MOTOR CORP . Abstract: PROBLEM TO BE SOLVED: To provide a road surface condition
- JP4183943A 11/19/1990
  Document ID: JP4183943A. Title: ACCELERATION SLIP CONTROL DEVICE. Assignees: TOYOTA MOTOR CORP . Abstract: PURPOSE: To prevent control by means of value of different from an actual road surface condition
- JP2007076463A 9/13/2005
  Document ID: JP2007076463A. Title: CAR BODY SPEED ESTIMATION DEVICE FOR VEHICLE. Assignees: TOYOTA MOTOR CORP . Abstract: PROBLEM TO BE SOLVED: To precisely estimate car body speed regardless

Figure 20. Aureka: Analysis of patents related to measuring friction on the road surface. A specific document has been located on the patent map. The document window on the right shows the contents of documents surrounding the specific document.

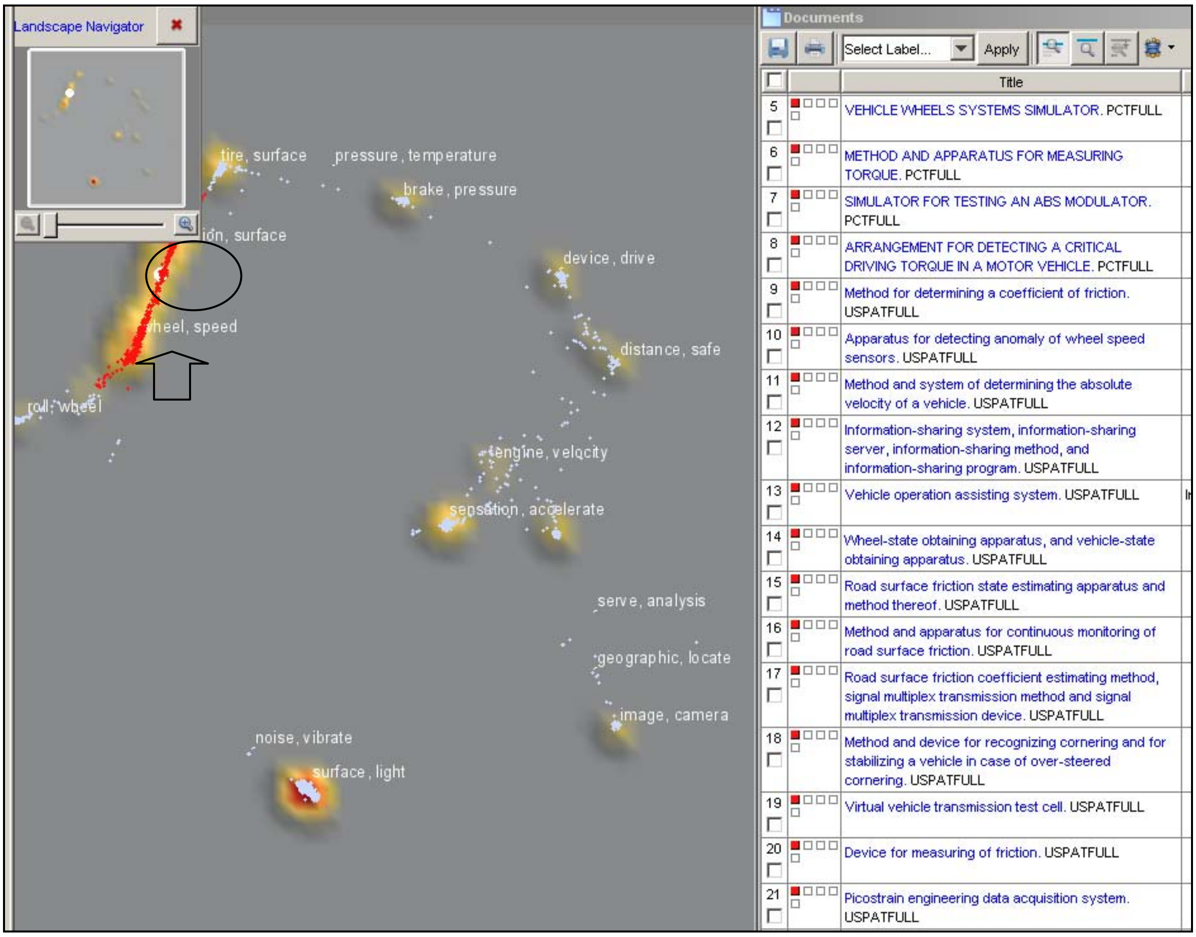
Figure 21 shows the record selected from the record viewer in OmniViz's Galaxy, coloured in yellow. Details of the document in question are shown at the top of the visualization.



Cluster	T.	Title Title	Patent Publication Patent Publication No	Application Date Application Date	Assignee Applicant Assignee Applicant
10	...	WHEEL SLIPPAGE DETECTOR FOR AUTOMOBILE	JP60152956A	Jan 23, 1984	MAZDA MOTOR CORP

Figure 21. OmniViz: Analysis of patents related to measuring friction on the road surface. A specific document located from the patent map and coloured in yellow. The record window at the top shows details of the document.

Figure 22 shows the locating tools of STN AnaVist. While selecting the document from the document display on the right, a white spot blinks around the document and the landscape navigator shows the specific space on the landscape.



☐		Title
5 ☐	■□□□ □	VEHICLE WHEELS SYSTEMS SIMULATOR. PCTFULL
6 ☐	■□□□ □	METHOD AND APPARATUS FOR MEASURING TORQUE. PCTFULL
7 ☐	■□□□ □	SIMULATOR FOR TESTING AN ABS MODULATOR. PCTFULL
8 ☐	■□□□ □	ARRANGEMENT FOR DETECTING A CRITICAL DRIVING TORQUE IN A MOTOR VEHICLE. PCTFULL
9 ☐	■□□□ □	Method for determining a coefficient of friction. USPATFULL
10 ☐	■□□□ □	Apparatus for detecting anomaly of wheel speed sensors. USPATFULL
11 ☐	■□□□ □	Method and system of determining the absolute velocity of a vehicle. USPATFULL
12 ☐	■□□□ □	Information-sharing system, information-sharing server, information-sharing method, and information-sharing program. USPATFULL
13 ☐	■□□□ □	Vehicle operation assisting system. USPATFULL
14 ☐	■□□□ □	Wheel-state obtaining apparatus, and vehicle-state obtaining apparatus. USPATFULL
15 ☐	■□□□ □	Road surface friction state estimating apparatus and method thereof. USPATFULL
16 ☐	■□□□ □	Method and apparatus for continuous monitoring of road surface friction. USPATFULL
17 ☐	■□□□ □	Road surface friction coefficient estimating method, signal multiplex transmission method and signal multiplex transmission device. USPATFULL
18 ☐	■□□□ □	Method and device for recognizing cornering and for stabilizing a vehicle in case of over-steered cornering. USPATFULL
19 ☐	■□□□ □	Virtual vehicle transmission test cell. USPATFULL
20 ☐	■□□□ □	Device for measuring of friction. USPATFULL
21 ☐	■□□□ □	Picostrain engineering data acquisition system. USPATFULL

Figure 22. STN AnaVist: Analysis of patents related to measuring friction on the road surface. A specific document is located from the patent map by choosing it in the Documents window. The related document blinks on the landscape and the Landscape Navigator in the top-left corner helps to locate the site.

Summary

Aureka, OmniViz and STN AnaViz provided easy ways to locate a specific document in the visualization. Thomson Data Analyzer doesn't locate specific documents in the maps, but details of all records in each cluster may be seen. All four tools enabled access to the whole document for closer examination.

5. Analysis with Company-based Data

This chapter introduces the results of analyses made with company related data, introduced in Chapter 3.2. The analyses evaluate the patent portfolio of Fraunhofer, which is Europe's largest organization for applied research, based in Germany. Data set with all of Fraunhofer's patents and patent applications filed since 1995 was used.

Knowing another company's or competitor's patenting activities reveals its strengths and business strategies. Yearly trends show the technology areas the company has abandoned and the areas it is concentrating on now. The filing activity for different Patent Organizations, i.e. each country's patent office, reveals a company's geographical business strategy. The protection of an invention is sought only for countries considered to be important markets. The inventor information of patent applications reveals key researchers in specific techniques, which is a valuable feature for headhunting. Collaboration with other actors in the field can also be seen from patents.

Chapter 5.1 introduces Fraunhofer's patent portfolio, or "patent landscape". Chapter 5.2 shows the yearly trends in its patenting and Chapter 5.3 gives information about Fraunhofer's co-operation with Nokia-Siemens-Network.

5.1 Landscape

Achieving an overview of the data is also important for competitive analysis. The Fraunhofer data retrieved from MicroPatent and used for testing Aureka and OmniViz includes documents written in German. Both tools allow the cleaning of data before clustering; Aureka even has a "use English data only" option. However all the data was used. This was done in order to evaluate how the clustering algorithms manage the situation of handling documents in different languages and also having as much data as possible for the basic analysis. Both tools managed the task well. Documents in German are in their own clusters.

Figure 23 visualizes Fraunhofer's patent portfolio of patent applications filed since 1995 with Aureka. The highest frequencies of documents are for clusters related to optics in the middle, signals and audio techniques at the centre-right and documents written in German at the top.

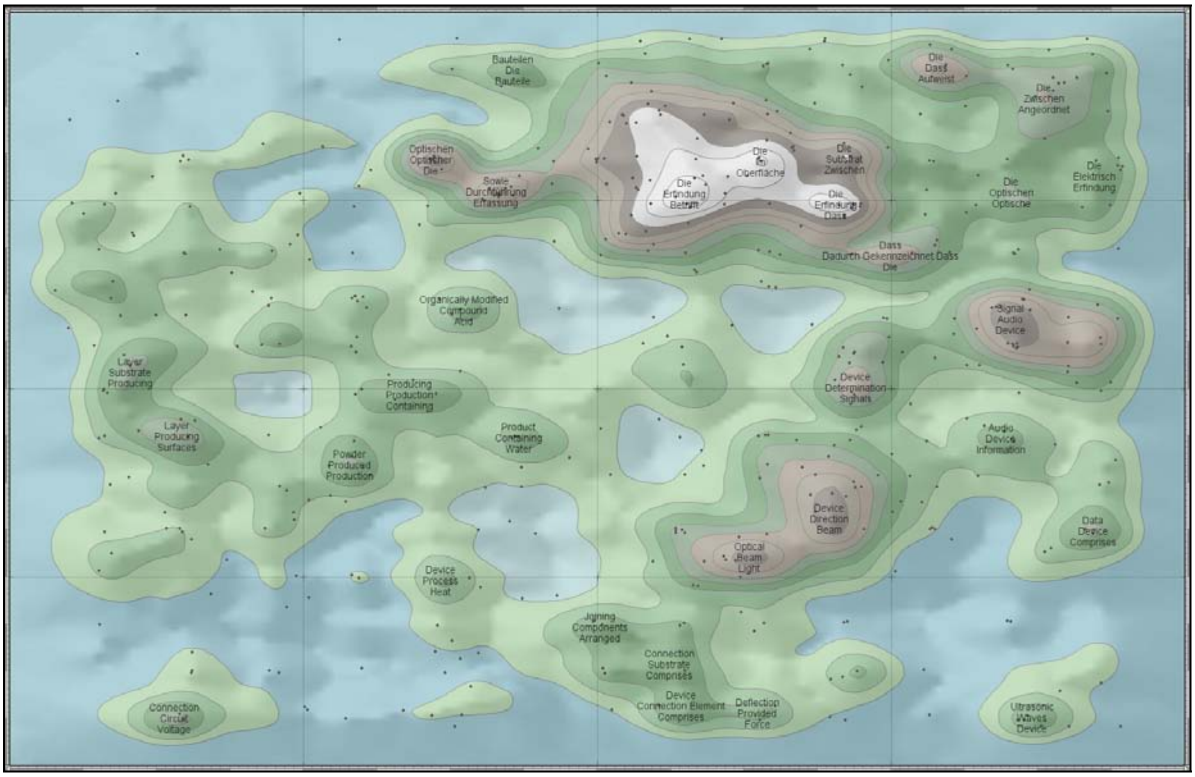
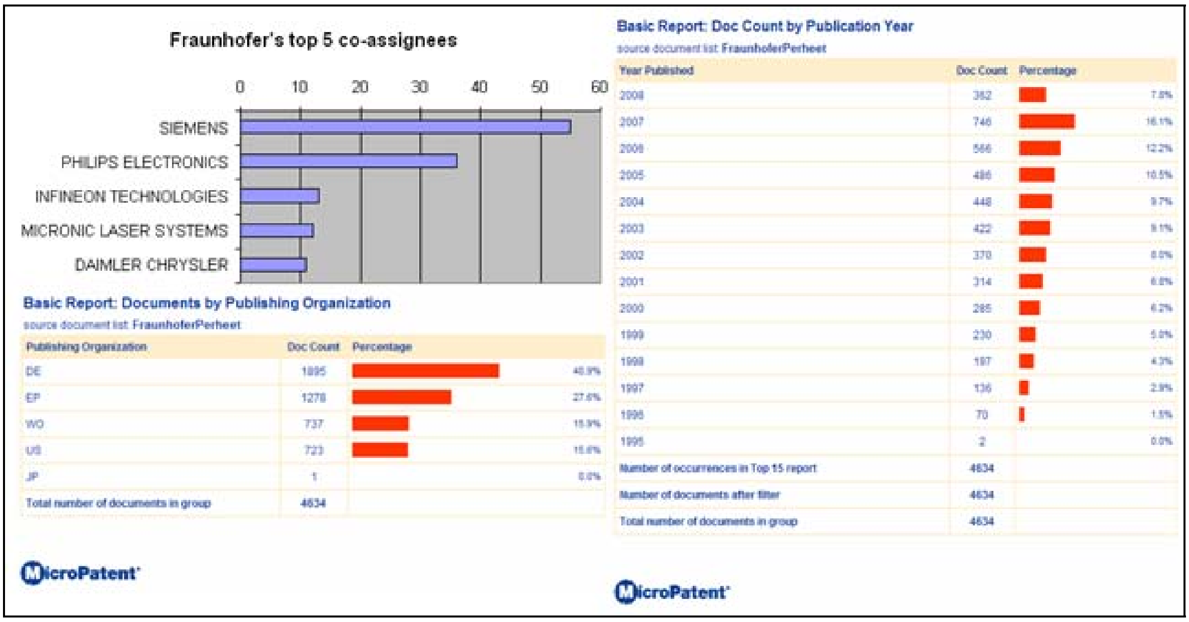


Figure 23. Aureka: Visualization of Fraunhofer's patent portfolio. Data includes all patents filed since 1995. The visualization was prepared with Aureka's Thememap.

Figure 24 shows basic statistics created with Aureka. Aureka provides two different ways to achieve top occurrence lists of different attributes. It has basic reports in standard formats and filters and macros for creating the graphs in Microsoft Excel. The bar graph in the top-left corner shows Fraunhofer's most important co-assignees. The formats of the names of assignees are often quite varied. The name of one company can be written in many different ways, e.g. Nokia Oy, Nokia Corp., etc. This is why the analysis with Aureka's Basic report led to the top assignees list having many entries for Fraunhofer's different departments. The analysis had to be performed by importing the data into Microsoft Excel and combining the related assignee names manually. Siemens and Philips seem to be the most active collaborators with Fraunhofer over the whole time period. The report in the lower left shows the top frequencies of patent applications published by Patent Organizations. Since 1995, Fraunhofer has filed most of its patents, 40%, in Germany only. The second-largest amount has been filed with the European Patent Office, and the third through the international application procedure; PCT (documents with application numbers having the two-letter prefix "WO"). The basic report on the right shows the yearly trends in publication years. Fraunhofer seems to continuously increase its patenting activity. Attention should be paid to 2008, as documents were retrieved in April and 2008 is therefore incomplete.



Basic Report: Documents by Publishing Organization

source document list: FraunhoferPerheet

Publishing Organization	Doc Count	Percentage
DE	1895	40.9%
EP	1278	27.6%
WO	737	15.9%
US	723	15.6%
JP	1	0.0%
Total number of documents in group	4634	

Basic Report: Doc Count by Publication Year

source document list: FraunhoferPerheet

Year Published	Doc Count	Percentage
2008	362	7.8%
2007	746	16.1%
2006	566	12.2%
2005	486	10.5%
2004	448	9.7%
2003	422	9.1%
2002	370	8.0%
2001	314	6.8%
2000	285	6.2%
1999	230	5.0%
1998	197	4.3%
1997	136	2.9%
1996	70	1.5%
1995	2	0.0%
Number of occurrences in Top 15 report	4634	
Number of documents after filter	4634	
Total number of documents in group	4634	

Figure 24. Aureka: Basic statistics of Fraunhofer's patent portfolio. The data includes all patents and applications filed since 1995. The graphics show the frequency of documents per top five patent assignees, publishing patent organization, and per publication year.

Figure 25 is a visualization made with OmniViz's Galaxy. There seem to be some clusters with a high frequency of documents; the cluster related to fibres at the top, materials and gas in the middle, and clusters related to electronics at the bottom.

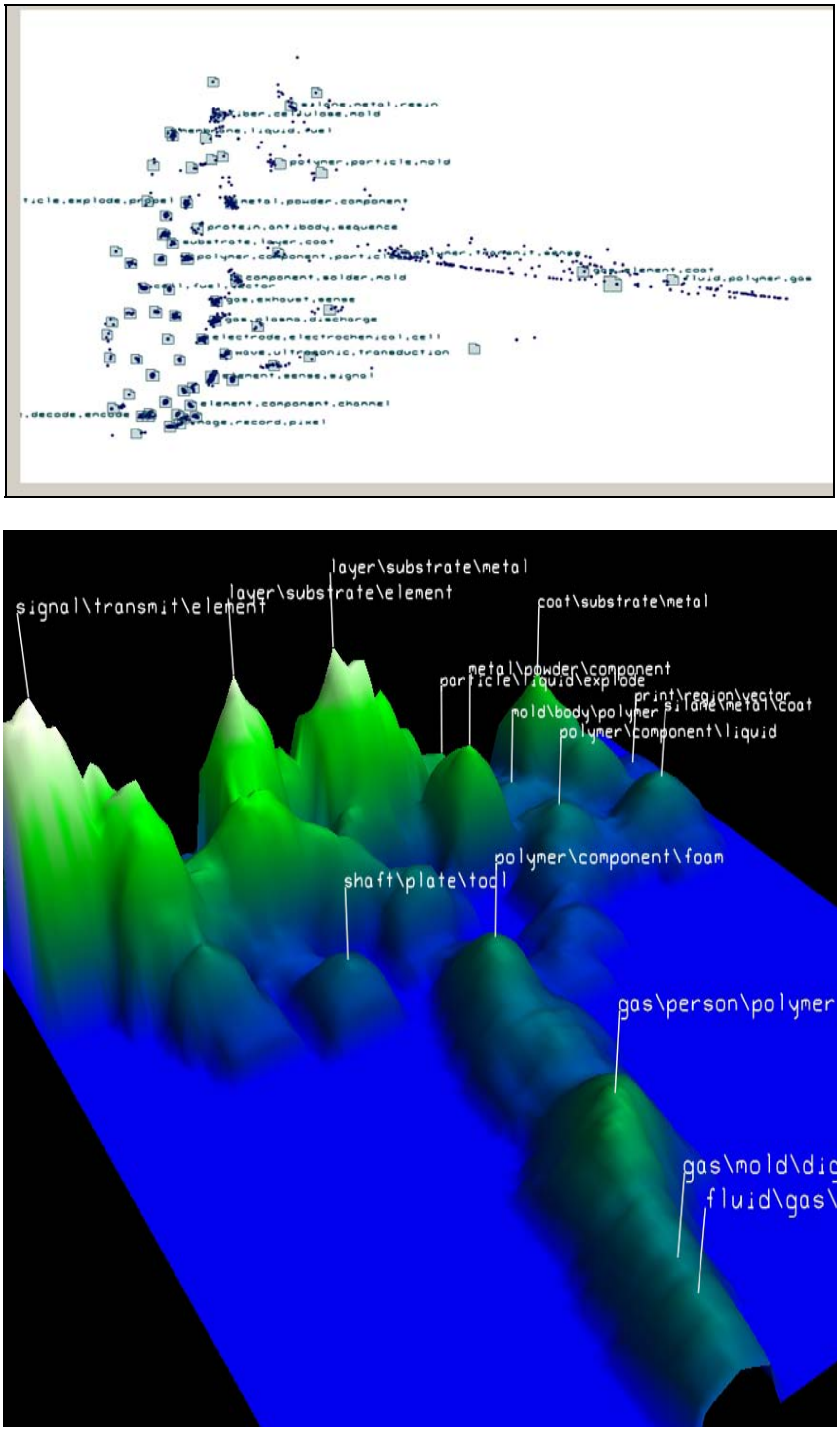


Figure 25. OmniViz: Visualization of Fraunhofer's patent portfolio. The data includes all patents filed since 1995. The visualizations were created with OmniViz's Galaxy and ThemeMap tools.

Figure 26 shows a visualization of Fraunhofer's patent portfolio and basic statistics made with STN AnaVist. On the left are the landscape and priority yearly trends. The highest frequency of documents is in the cluster related to layers and substrates. The graph at the top right expresses co-operation with other companies, below are patent countries and key researchers.

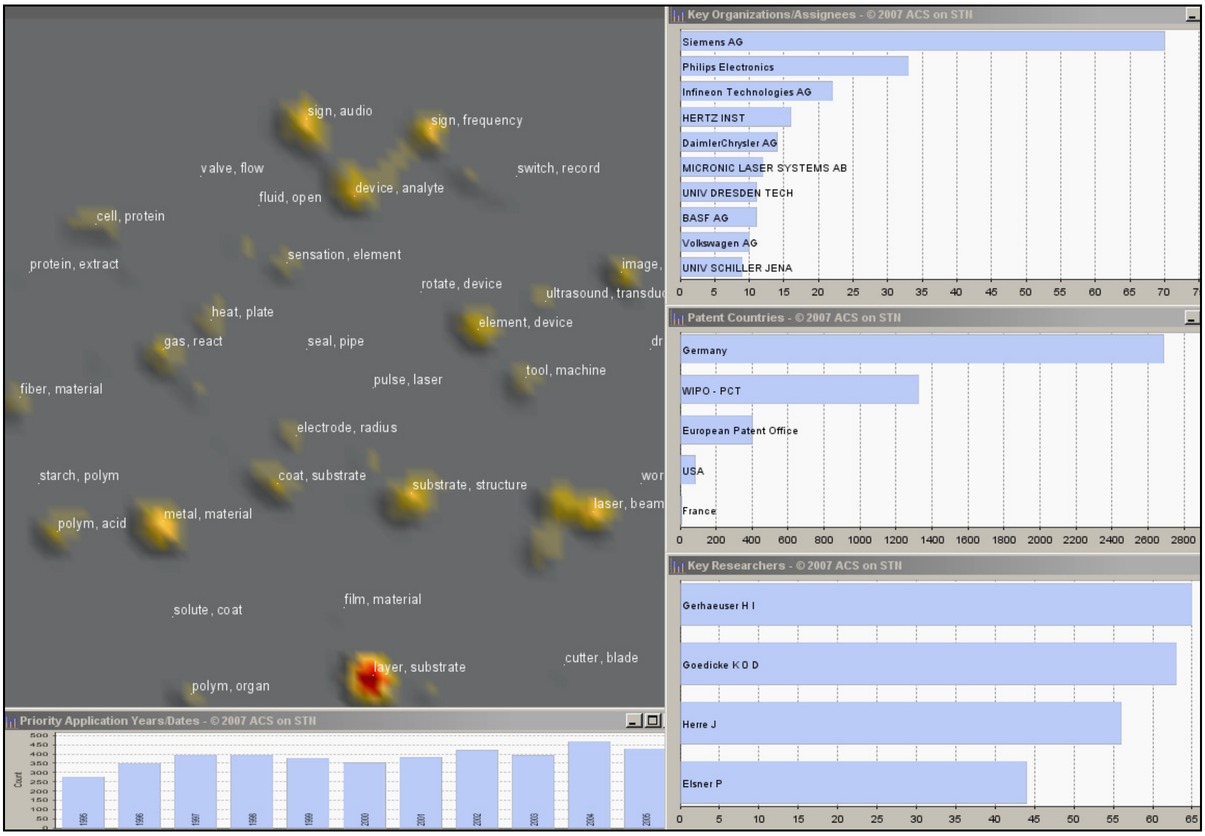
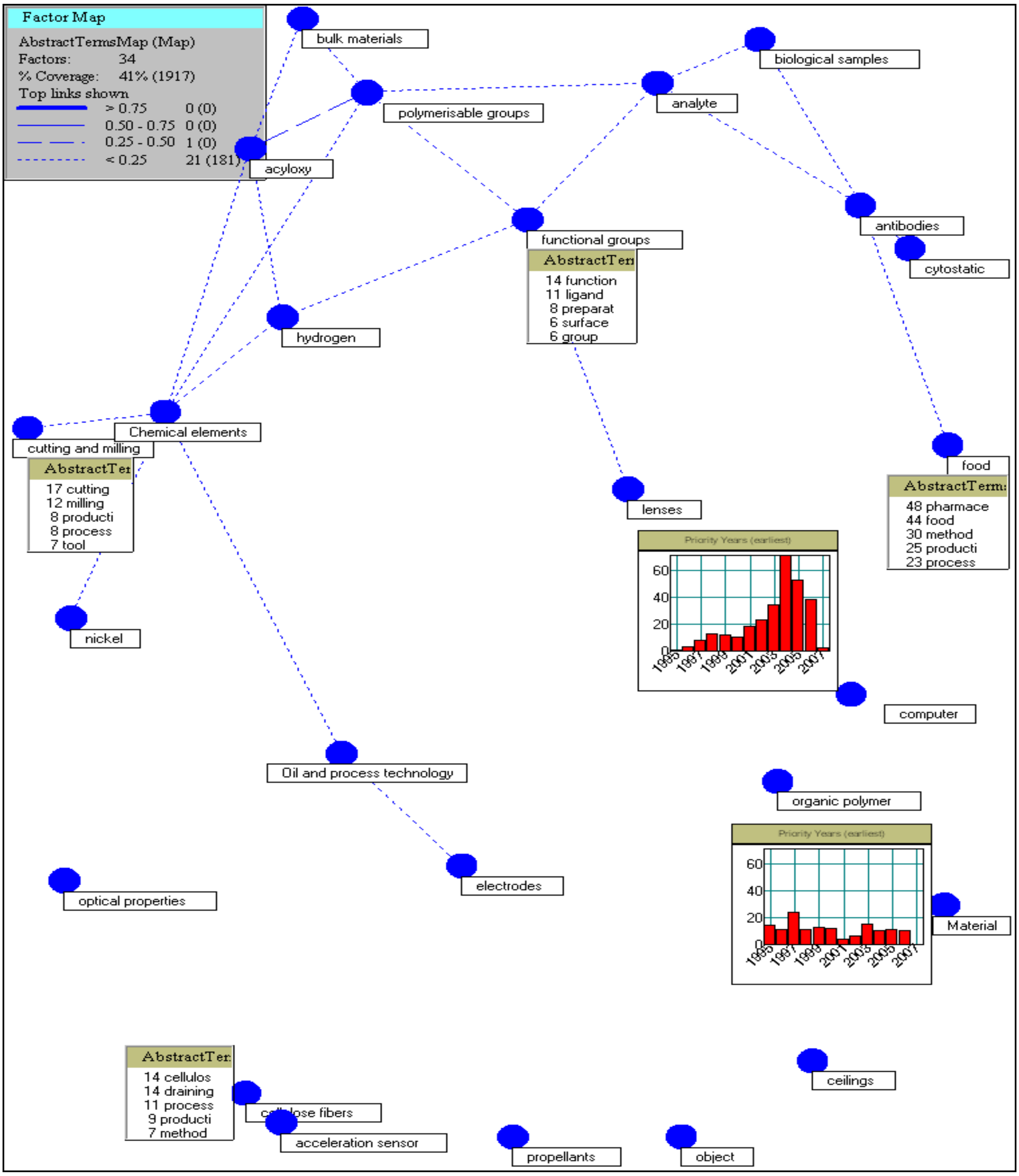


Figure 26. STN AnaVist: Visualization of Fraunhofer's patent portfolio. The data includes all patents filed since 1995. The visualizations were created with STN AnaVist. On the left are landscape and priority yearly trends. The graph at the top right expresses co-operation with other companies, below are patent countries and key researchers.

Figure 27 represents a visualization of Fraunhofer's patent portfolio created with Thomson Data Analyzer's Factor Map. The features of the Factor Map are explained in Chapter 4.1. According to the priority year trend boxes shown in the figure, Fraunhofer has remarkably increased patenting related to computer technologies and respectively decreased those related to materials. Thomson Data Analyzer has distributed the documents into clusters quite evenly.



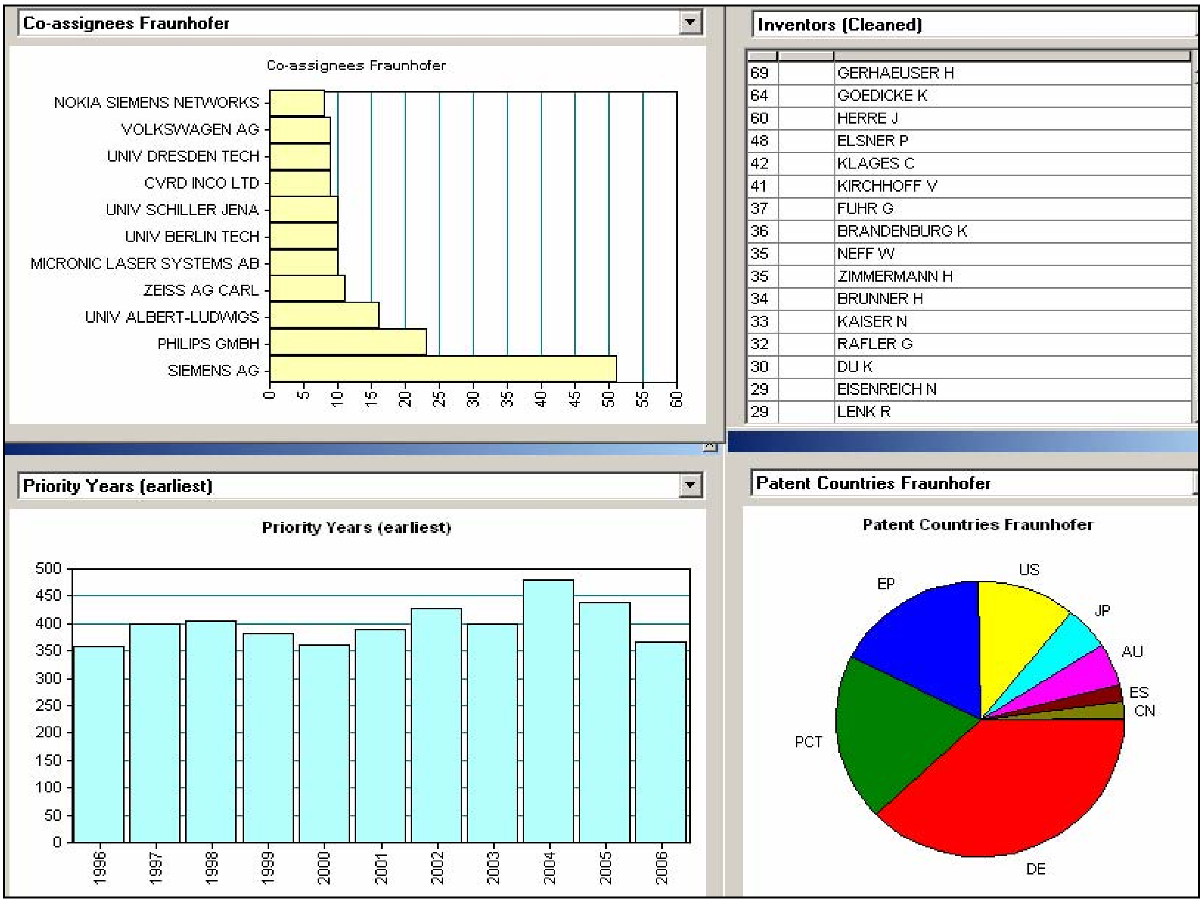
Factor Map

AbstractTermsMap (Map)
Factors: 34
% Coverage: 41% (1917)
Top links shown

━━━	> 0.75	0 (0)
───	0.50 - 0.75	0 (0)
— — -	0.25 - 0.50	1 (0)
- - - -	< 0.25	21 (181)

Figure 27. TDA VantagePoint: Visualization of Fraunhofer's patent portfolio. The data includes all patents filed since 1995. Yearly trends in patenting are shown for two document clusters, "computer" and "material".

Figure 28 shows basic statistics prepared with Thomson Data Analyzer. The list of top assignees is in the upper-left corner and priority yearly trends are below them; the top inventors are in the top-right corner, with patent countries below.



69	GERHAEUSER H
64	GOEDICKE K
60	HERRE J
48	ELSNER P
42	KLAGES C
41	KIRCHHOFF V
37	FUHR G
36	BRANDENBURG K
35	NEFF W
35	ZIMMERMANN H
34	BRUNNER H
33	KAISER N
32	RAFLER G
30	DU K
29	EISENREICH N
29	LENK R

Figure 28. TDA VantagePoint: Analysis of Fraunhofer's patent portfolio. The data includes all patents filed since 1995. Basic statistics of patents filed with co-assignees are in the top-left corner, the most active inventors are on the right, frequencies of applications by priority years are at the bottom left and geographical activity in patenting is in the bottom-right corner.

Summary

The clustering made by the tools was most illustrative with OmniViz and STN AnaVist. The technology areas of digital techniques, chemistry and materials could be distinguished. Identifying the most patented areas was most difficult with Thomson Data Analyzer. All tools gave similar top collaborators and yearly trends information. Fraunhofer's top collaborators are Siemens and Philips, and its patenting has been quite steady, though slightly increasing over time. Fraunhofer concentrates mostly on European markets; its three mostly used patenting practices are: filing the patent straight with Germany's patent office, with the European Patent Office, or through PCT.

5.2 Yearly Trends in Patenting

Evaluating yearly trends in a competitor’s patenting may offer valuable information about its business decisions. Increasing patenting in one area may indicate a competitor’s intention to concentrate on that technology, or an old patent portfolio may relate to a decision to change the business focus. Due to the slow product launch process, the appearance of new products and techniques may be seen in advance from new patent applications.

Figure 29 shows a visualization of Fraunhofer’s patents made with Aureka. On the left is the visualization with documents filed between 2005 and 2008 coloured in red, and on the right those from 1995 to 1998 coloured in green. Aureka allows the colouring of different time slices in the same figure too. For further analysis of the time slice groups a new data set containing the corresponding documents would have to be made.

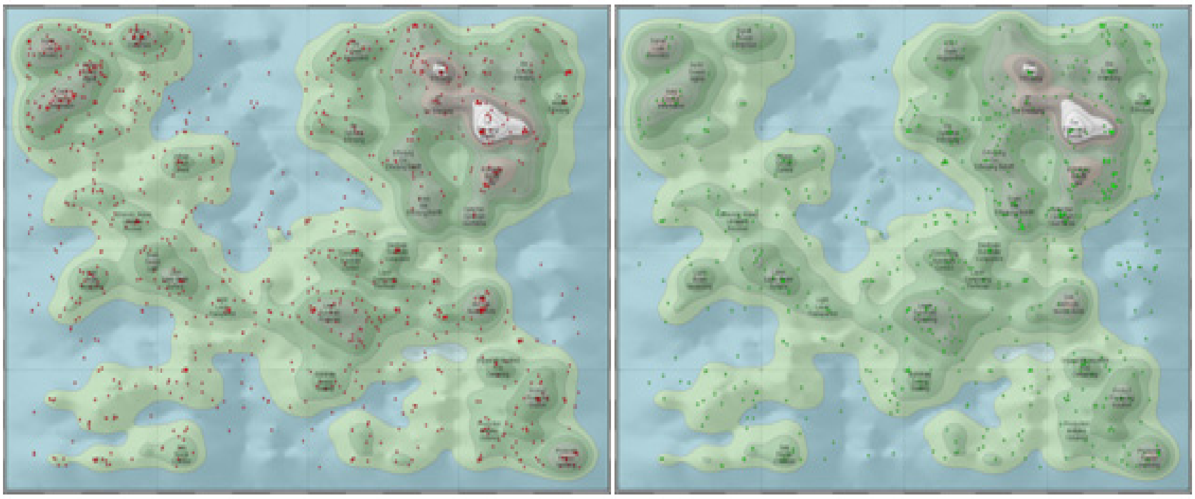


Figure 29. Aureka: Analysis of Fraunhofer’s patent portfolio. Two patent maps made with Aureka; the left one has patent documents from 2005 to 2008 coloured in red, and on the right are documents from 1995 to 1998 in green.

Figure 30 shows an analysis of Fraunhofer's patent portfolio made with STN AnaVist. Documents with priority years from 1995 to 1996 are colored in blue and those from 2005 to 2006 are in red. The landscape on the left represents the distribution of documents with only colored documents shown. The patent country bars in the top-right corner show that Fraunhofer has recently filed just as many patent applications in Germany as in earlier years, and has increased slightly the amount of PCT applications. The top assignees graph in the middle indicates long lasting co-operation with Siemens. The top Derwent Manual codes are presented in the graph in the bottom-right corner showing the domination of computer-related techniques.

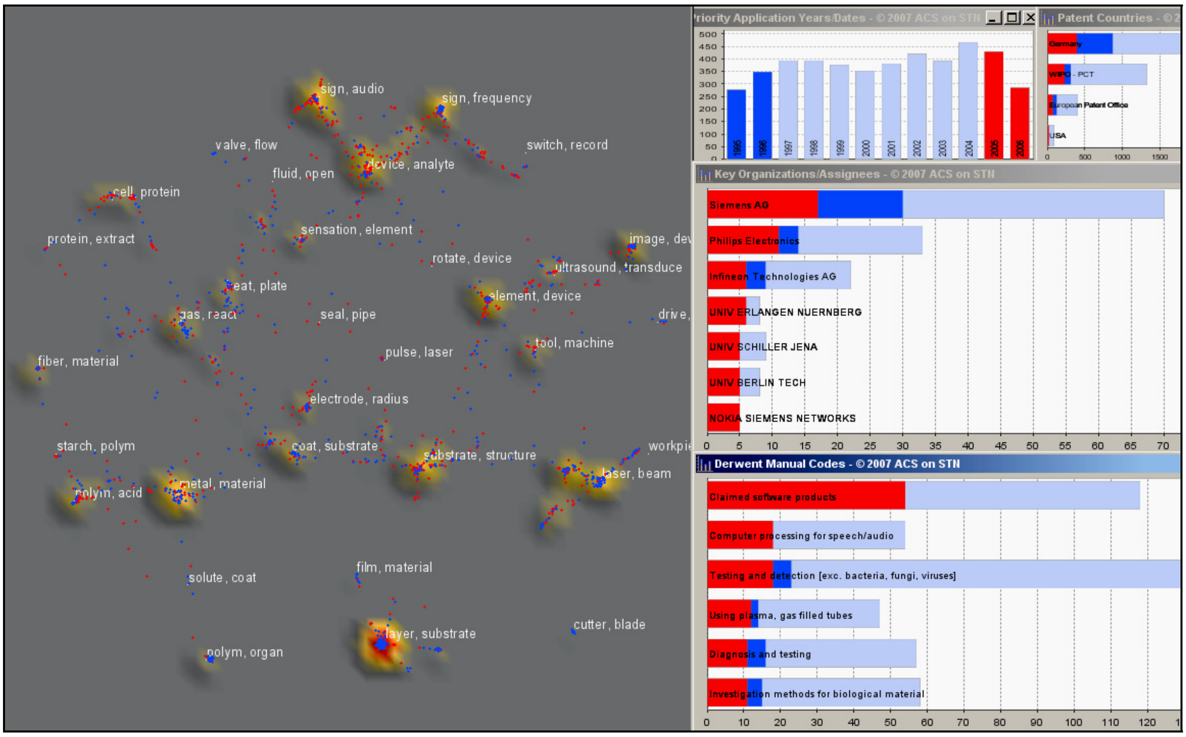
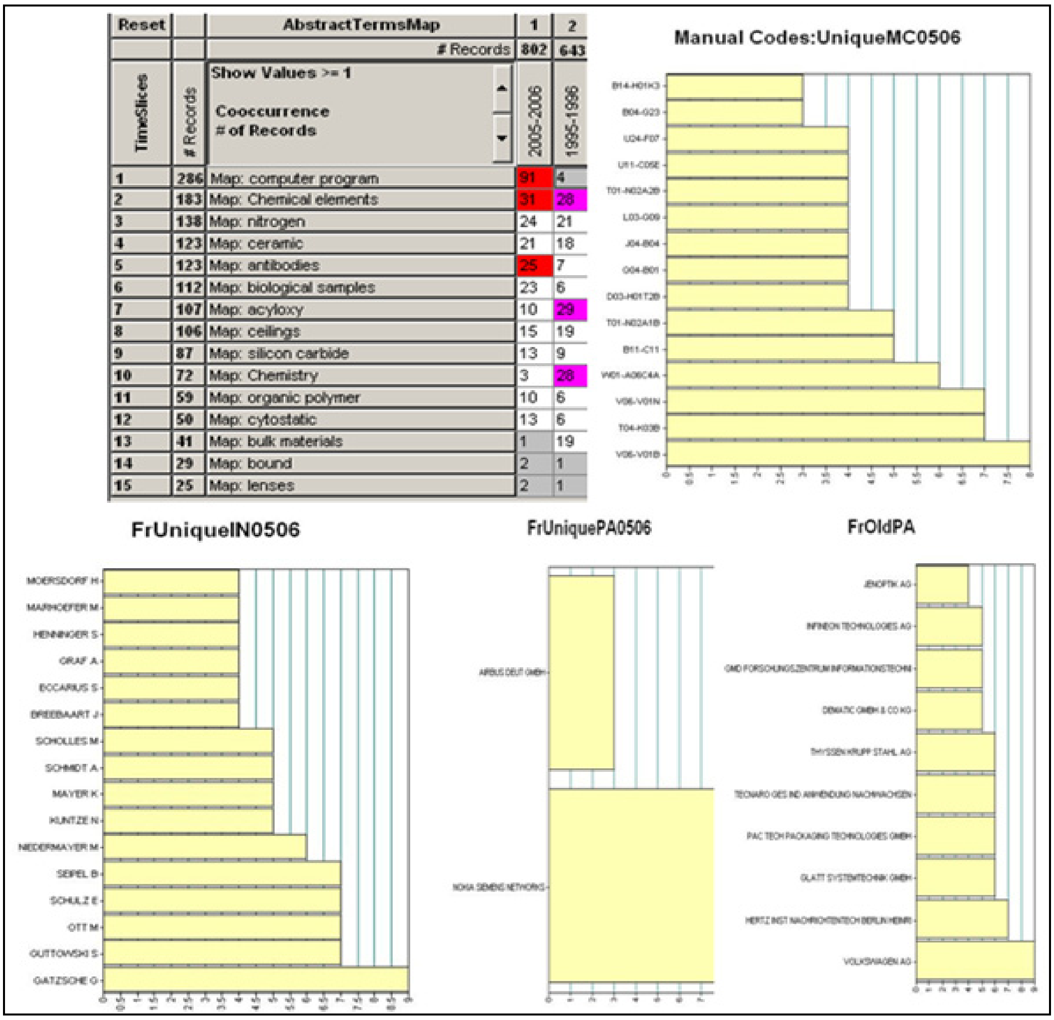


Figure 30. STN: Analysis of Fraunhofer's patent portfolio made with STN AnaVist. Documents with priority years from 1995 to 1996 are colored in blue, and those from 2005 to 2006 are in red. The landscape on the left shows the distribution of documents. Patent countries are in the top-right corner, top assignees in the middle, and top Derwent Manual codes in the bottom-right corner.

Figure 31 shows yearly trend analysis made with Thomson Data Analyzer. A group containing documents filed between 1995 and 1996 was created, and another with documents from 2005 to 2006. The matrix in the upper-left corner represents the "hot area" analysis for both groups. The analysis is committed by first creating a Factor Map to reveal the most important terms over the whole data set and then a Co-occurrence Matrix to evaluate which are the most patented subjects for either group; in other words the "hot areas". The matrix indicates that Fraunhofer's latest hot areas are technologies related to computer programs, chemical elements and antibodies, whereas in the early years of the explored patents they were technologies related to acyloxy radicals, chemical elements and chemistry.

Other statistics in the figure were prepared with Thomson Data Analyzer's list comparison tool. Unique records for either set may be evaluated by comparing the group with the rest of the data set with Thomson Data Analyzer's List Comparison tool. The bar graph in the top-right corner shows Derwent Manual Codes for new inventions. They are unique for the data group consisting of patent applications filed during 2005 and 2006, i.e. they don't appear in other documents in the test data. The graph in the lower-left corner shows unique inventors for the latest applications. This analysis could be used to identify new up-and-coming researchers in the area. The Unique Patent Assignees graph at the bottom middle contains only two companies, Airbus and Nokia Siemens. This indicates a quite recently begun collaboration between Fraunhofer and them. The graph in the lower-right corner shows patent assignees unique to the former data group, though with whom collaboration has ended.



Reset		AbstractTermsMap	1	2
		# Records	802	643
TimeSlices	# Records	Show Values >= 1 Cooccurrence # of Records	2005-2006	1995-1996
1	286	Map: computer program	91	4
2	183	Map: Chemical elements	31	28
3	138	Map: nitrogen	24	21
4	123	Map: ceramic	21	18
5	123	Map: antibodies	25	7
6	112	Map: biological samples	23	6
7	107	Map: acyloxy	10	29
8	106	Map: ceilings	15	19
9	87	Map: silicon carbide	13	9
10	72	Map: Chemistry	3	28
11	59	Map: organic polymer	10	6
12	50	Map: cytostatic	13	6
13	41	Map: bulk materials	1	19
14	29	Map: bound	2	1
15	25	Map: lenses	2	1

Figure 31. TDA VantagePoint: Analysis of patents assigned by Fraunhofer in 1995–1996 and 2005–2006. The graph in the top-left corner shows the "hot areas" of patenting currently and earlier. Statistics of inventions made by inventors who had not submitted applications before 2005 are at the bottom left, manual codes for new technology areas are in the top-right corner, new partners are at the bottom middle and partners with whom there has been no co-operation during recent years are at the bottom right.

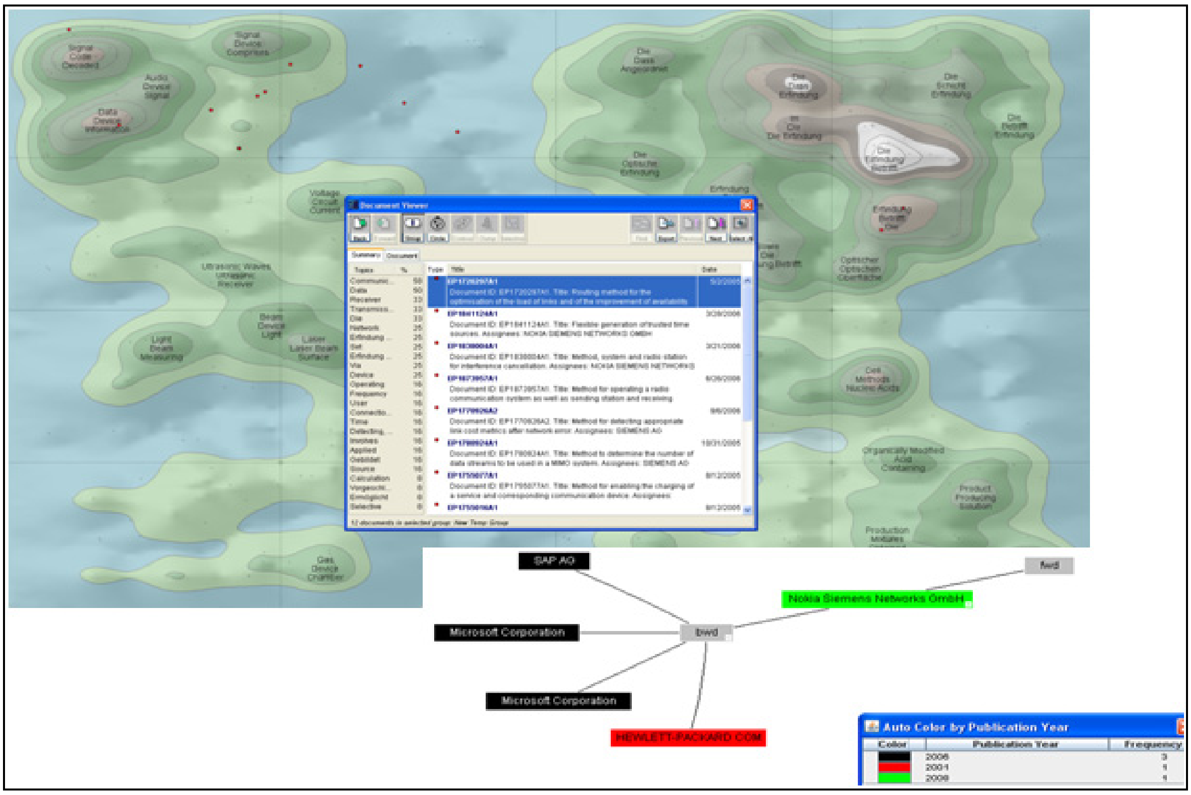
Summary

All four tools enabled the comparison of documents of different time periods. STN AnaVist and Thomson Data Analyzer easily revealed the new “hot areas” in Fraunhofer’s patenting. They both named it to be computer software. Comparison of different time periods was most difficult with Aureka, since it only shows interactively the landscape with selected documents showing, but not other statistics. This could, however, be created by committing a new analysis with the selected data set. Graphs taken of the comparison in OmniViz are not right for the tool, since one of its strengths is the possibility to see the change with animations. The possibility to obtain different comparative statistics of time periods with Thomson Data Analyzer was highly valuable.

5.3 Co-operation

Fraunhofer have collaborated recently with Nokia, as was seen from former analyses. Looking at the patents with both named as patent assignees reveals more closely the nature of the co-operation. Patent classifications and yearly trends in the documents give detailed information about the techniques developed and the time-frame of the collaboration.

Figure 32 shows a landscape of Fraunhofer’s patents. The ones filed with Nokia are coloured in red. Corresponding documents may be examined with Document Viewer, shown in the middle. At the bottom is a graph showing the results of a citation analysis carried out for one patent assigned with Nokia. It indicates that the invention relates to techniques developed by SAP, Microsoft and Hewlett Packard. Citation analysis reveals other actors working on the same subject.



Document Viewer

Topics	%
Communic...	50
Data	50
Receiver	33
Transmiss...	33
Die	33
Network	25
Erfindung ...	25
Set	25
Erfindung ...	25
Via	25
Device	25
Operating	16
Frequency	16
User	16
Connectio...	16
Time	16
Detecting...	16
Involves	16
Applied	16
Gebildet	16
Source	16
Calculation	8
Vorgeschl...	8
Ermöglicht	8
Selective	8

Type	Title	Date
●	EP1720297A1 Document ID: EP1720297A1. Title: Routing method for the optimisation of the load of links and of the improvement of availability	5/2/2005
●	EP1841124A1 Document ID: EP1841124A1. Title: Flexible generation of trusted time sources. Assignees: NOKIA SIEMENS NETWORKS GMBH	3/28/2006
●	EP1838004A1 Document ID: EP1838004A1. Title: Method, system and radio station for interference cancellation. Assignees: NOKIA SIEMENS NETWORKS	3/21/2006
●	EP1873957A1 Document ID: EP1873957A1. Title: Method for operating a radio communication system as well as sending station and receiving	6/26/2006
●	EP1770926A2 Document ID: EP1770926A2. Title: Method for detecting appropriate link cost metrics after network error. Assignees: SIEMENS AG	9/6/2006
●	EP1780924A1 Document ID: EP1780924A1. Title: Method to determine the number of data streams to be used in a MIMO system. Assignees: SIEMENS AG	10/31/2005
●	EP1755077A1 Document ID: EP1755077A1. Title: Method for enabling the charging of a service and corresponding communication device. Assignees:	8/12/2005
●	EP1755016A1	8/12/2005

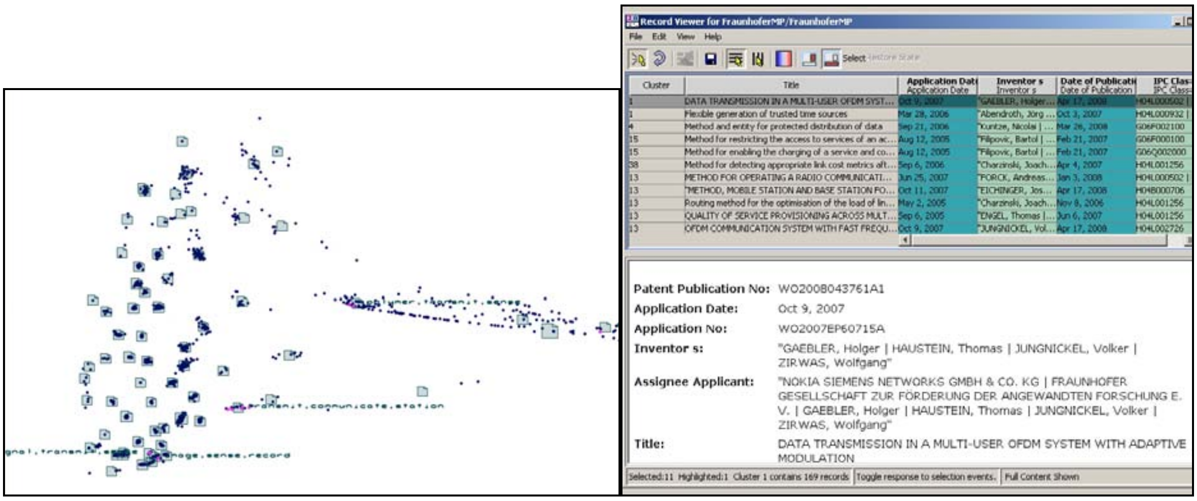
12 documents in selected group: New Temp Group

Auto Color by Publication Year

Color	Publication Year	Frequency
■	2006	3
🟥	2001	1
🟩	2008	1

Figure 32. Aureka: Analysis of Fraunhofer's patent portfolio. Inventions developed in co-operation with Nokia are coloured in red. The Document Viewer in the middle shows details of the corresponding documents and a graph representing the results of citation analysis at the bottom.

Figure 33 shows the division of Fraunhofer's inventions made in collaboration with Nokia in OmniViz's Galaxy visualization. The documents are marked with purple dots. On the right is the Record Viewer for a closer examination of the documents.


signal.transm...
...rahmen.t.communicate.station
...image.sense.record
Record Viewer for FraunhoferMP/FraunhoferMP
File
Edit
View
Help
Select
Restore State

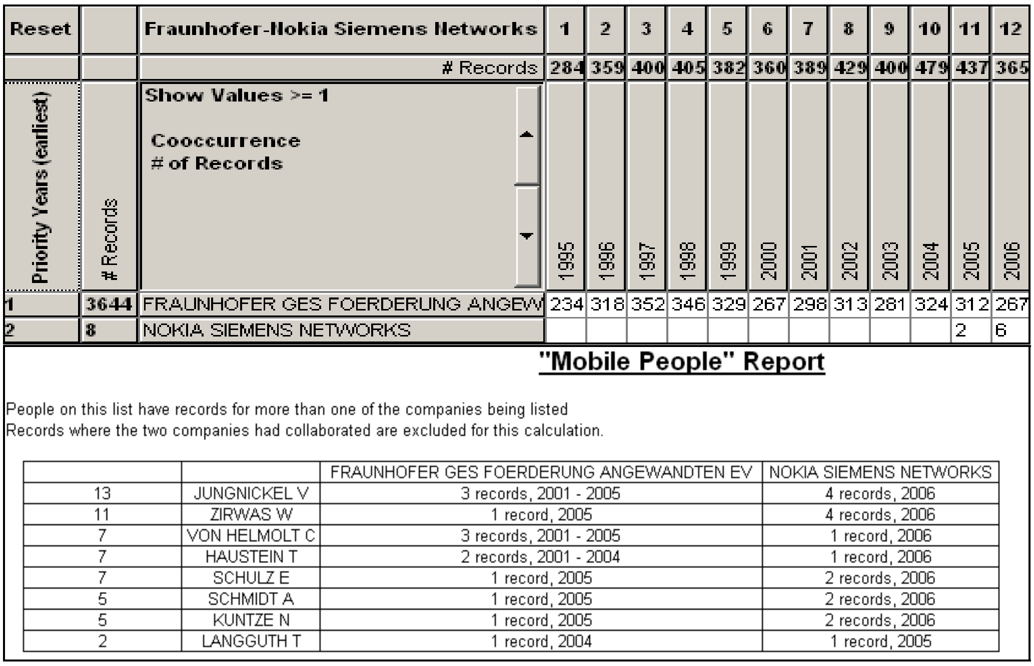
Cluster	Title	Application Date Application Date	Inventor s Inventor s	Date of Publication Date of Publication	IPC Class IPC Class
1	DATA TRANSMISSION IN A MULTI-USER OFDM SYST...	Oct 9, 2007	"GAEBLER, Holger...	Apr 17, 2008	H04L000502 |
1	Flexible generation of trusted time sources	Mar 28, 2006	"Abendroth, Jörg ...	Oct 3, 2007	H04L000932 |
4	Method and entity for protected distribution of data	Sep 21, 2006	"Kuntze, Nicolai | ...	Mar 26, 2008	G06F002100
15	Method for restricting the access to services of an ac...	Aug 12, 2005	"Filipovic, Bartol | ...	Feb 21, 2007	G06F000100
15	Method for enabling the charging of a service and co...	Aug 12, 2005	"Filipovic, Bartol | ...	Feb 21, 2007	G06Q002000
38	Method for detecting appropriate link cost metrics aft...	Sep 6, 2006	"Charzinski, Joach...	Apr 4, 2007	H04L001256
13	METHOD FOR OPERATING A RADIO COMMUNICATI...	Jun 25, 2007	"FORCK, Andreas...	Jan 3, 2008	H04L000502 |
13	"METHOD, MOBILE STATION AND BASE STATION FO...	Oct 11, 2007	"EICHINGER, Jos...	Apr 17, 2008	H04B000706
13	Routing method for the optimisation of the load of lin...	May 2, 2005	"Charzinski, Joach...	Nov 8, 2006	H04L001256
13	QUALITY OF SERVICE PROVISIONING ACROSS MULT...	Sep 6, 2005	"ENGEL, Thomas |...	Jun 6, 2007	H04L001256
13	OFDM COMMUNICATION SYSTEM WITH FAST FREQU...	Oct 9, 2007	"JUNGNICKEL, Vol...	Apr 17, 2008	H04L002726

Patent Publication No:	WO2008043761A1
Application Date:	Oct 9, 2007
Application No:	WO2007EP60715A
Inventor s:	"GAEBLER, Holger | HAUSTEIN, Thomas | JUNGNICKEL, Volker | ZIRWAS, Wolfgang"
Assignee Applicant:	"NOKIA SIEMENS NETWORKS GMBH & CO. KG | FRAUNHOFER GESELLSCHAFT ZUR FÖRDERUNG DER ANGEWANDTEN FORSCHUNG E. V. | GAEBLER, Holger | HAUSTEIN, Thomas | JUNGNICKEL, Volker | ZIRWAS, Wolfgang"
Title:	DATA TRANSMISSION IN A MULTI-USER OFDM SYSTEM WITH ADAPTIVE MODULATION

Selected:11 Highlighted:1 Cluster 1 contains 169 records | Toggle response to selection events. | Full Content Shown

Figure 33. OmniViz: Analysis of Fraunhofer's co-operation with Nokia-Siemens-Networks (NSN). On the left side is a Galaxy landscape with Fraunhofer's patents assigned with NSN coloured in purple. On the right side are corresponding patent documents.

Figure 34 shows a co-occurrence matrix created with Thomson Data Analyzer for patents assigned by Fraunhofer and Nokia-Siemens-Networks. The matrix shows the number of patents filed each year. It reveals that they have filed eight patents together during 2005 and 2006. The mobile people report below shows the names and years of employment of inventors that have worked for both companies. Co-operation applications have been excluded from the Mobile People report so as to avoid biasing the analysis.



Reset		Fraunhofer-Nokia Siemens Networks	1	2	3	4	5	6	7	8	9	10	11	12
		# Records	284	359	400	405	382	360	389	429	400	479	437	365
Priority Years (earliest)	# Records	Show Values >= 1 Cooccurrence # of Records	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006
1	3644	FRAUNHOFER GES FOERDERUNG ANGEW	234	318	352	346	329	267	298	313	281	324	312	267
2	8	NOKIA SIEMENS NETWORKS											2	6

"Mobile People" Report

People on this list have records for more than one of the companies being listed
Records where the two companies had collaborated are excluded for this calculation.

		FRAUNHOFER GES FOERDERUNG ANGEWANDTEN EV	NOKIA SIEMENS NETWORKS
13	JUNGNICKEL V	3 records, 2001 - 2005	4 records, 2006
11	ZIRWAS W	1 record, 2005	4 records, 2006
7	VON HELMOLT C	3 records, 2001 - 2005	1 record, 2006
7	HAUSTEIN T	2 records, 2001 - 2004	1 record, 2006
7	SCHULZ E	1 record, 2005	2 records, 2006
5	SCHMIDT A	1 record, 2005	2 records, 2006
5	KUNTZE N	1 record, 2005	2 records, 2006
2	LANGGUTH T	1 record, 2004	1 record, 2005

Figure 34. TDA VantagePoint: Co-occurrence matrix of Fraunhofer's and Nokia's patents at the top, mobile report of inventors who have worked for both companies the bottom.

Summary

All the tools found some new patents filed in collaboration between Nokia and Fraunhofer, although any explanation for Thomson Data Analyzer finding only eight of them couldn't be found. Aureka's citation analysis offers valuable information about other actors interested in the same technologies.

6. Conclusions

All the tools evaluated were found to be very useful for the task and quite easy to adopt for daily work. All four had some strengths and weaknesses compared to each other.

Aureka and STN AnaVist are easy to use and analysis can be committed with almost no preliminary preparation. Aureka's strengths are its visually impressive representation of results and clear user interface. The availability of citation analysis also gives added value to the analysis. STN AnaVist also gets credit for its user interface. The best features of AnaVist are the possibility to easily see the results of an analysis from many different points of view at the same time and the high degree of interactivity of the tool.

Though Biowisdom's OmniViz and VantagePoint with Thomson Data Analyzer need more preparation and learning before the analysis can be made, they compensate for the effort by offering sophisticated analysis and a high degree of decision-making power for the user. They are the right tools for a "power user," but are effective and useful for basic analysis too. Their strength is also the possibility to use data from almost any source and in almost any format, and they offer filters to aid importation and allow data to be combined in different formats. TDA VantagePoint also has special tools for handling value-added patent data retrieved from Derwent's World Patent Index, which is the most important unique database of patent information. OmniViz has invested in developing many different solutions for visualization of the data. The high degree of interactivity offered, e.g. animations of yearly trends, is one of its best features. Both tools have many other good analysing instruments which are not introduced in this study, e.g. Thomson Data Analyzer's auto-correlation matrices and maps for evaluating the relatedness of items within a field and OmniViz's analysis of numerical and biological data.

The limited availability of data sources are Aureka's and STN AnaVist's weaknesses. However, Aureka has relieved this by allowing the data analysed with Thomson Data Analyser to be imported into it for visualization. That raises the value of Aureka highly, although the use of MicroPatent data only would limit the scope of the analysis. STN AnaVist offers five databases for analysis, and fortunately two of them are highly value-added databases: Derwent's World Patent Index and Chemical Abstract Service's Chemical Abstracts Plus (CAplus). CAplus contains bibliographical data from patents and scientific publications in biochemistry, chemistry and chemical engineering.

The weaknesses of OmniViz were the amount of preparation needed to produce basic statistics and the need to understand sophisticated statistical methods, e.g. correlation matrices. Thomson Data Analyzer's weakness was the stiffness of its visualizations. Basic statistics were easily retrieved but the forming of their presentations was almost impossible, e.g. restricting the number of records shown. The simplified visualization format reduces their clarity.

All the tools had possibilities to make groups for closer evaluation of the data. They also provided tools for re-labelling cluster names and other attributes in the visualizations.

Table 1 shows a comparison of the features of the tools. Each feature has been given a score ranging from none to three plus signs. Three signs mean that the tool answers the need remarkably well and no sign means that the tool lacks the feature. The total number of the tools cannot be compared because the features and the measures are not compatible. The comparison doesn't try to indicate the how good each tool is; it seeks to help in evaluating which is the best tool for the specific needs of each reader.

Table 1. Comparison of the features of the tools. The number of plus signs doesn't indicate how good the tool is, as the measures are not compatible. The comparison is made to make it easier to evaluate the tool for the specific needs of the reader.

Comparison of the features	Tools			
Features	Aureka	OmniViz	STN AnaVist	Thomson Data Analyzer
Fast adoption of the tool	+++	+	+++	+
Use of varying data, including other than patent data	+ [1]	+++	+ [2]	+++
Flexibility	+	+++	+	+++
Visual representation of results	++	++	+++	+
Ease with which basic statistics can be created	++	+	+++	++
Need for more sophisticated analysis	+	+++	+	+++
Possibility to have influence on data mining algorithms	+	+++		++
Ability to edit the terms used in presentations	+	+	++	+
Process execution time	+	+++	++	+

[1] When importing data from Thomson Data Analyzer

[2] Non-patent data available from CAplus from fields of chemistry

In conclusion it could be stated that OmniViz and Thomson Data Analyzer are tools for sophisticated and diversified mathematical analysis of the data. Aureka and AnaVist are convenient for easily visualizing basic statistics and "top lists" of the data and for making stylish patent maps. The unique features of OmniViz, when compared to the other tools tested, are the possibility to visualize clustered data from many different points of view and the possibility to evaluate some attributes with patent map animations. Thomson Data Analyzer offers efficient tools for comparing different subsets of the data, e.g. for identifying the unique values of an attribute. Aureka is the only tool allowing citation analyses and has the most illustrative patent map. STN AnaVist is superior in the possibility to retrieve basic statistics quickly and smoothly.

The results obtained with all four tools were very much alike, even though different databases for retrieving the data were used. The top assignees and inventors lists were uniform, as were the yearly trends and both technological and geographical business areas. Only the reciprocal orders and amount of documents varied. However, the conclusions drawn from the results, and business decisions made with them, would all be similar regardless of the tool used.

References

1. Granstrand, O. 2000. The Economics and Management of Intellectual Property. Edward Elgar Publishing.

2. Hand, D., Mannila, H., Smyth, P. 2001. Principles of Data Mining. Massachusetts Institute of Technology.

3. Feldman, R., Sanger, J. 2007. The Text Mining Handbook. Cambridge University Press.

4. Jin, B., Teng, H.-F., Shi, Y.-J., Qu, F.-Z. 2007. Chinese Patent Mining based on sememe statistics and key-phrase extraction. Advanced Data Mining and Applications 01/01/2007, pp. 516–23.

5. Kasravi, K., Risov, M. 2007. Patent mining – discovery of business value from patent repositories. Proceedings of the 40th Annual Hawaii International Conference on System Sciences 01/01/07. 10 p.

6. Keim, D. A. 2002. Information Visualization and Visual Data Mining. IEEE Transactions on Visualization and Computer Graphics, Vol. 8, No. 1, January–March.

7. Tseng, Y.-H., Lin, C.-J., Lin, Y.-I. 2007. Text mining techniques for patent analysis. Information processing and management 43, pp. 1216–1247.

8. Eldridge, J. 2006. Data Visualization tools – a perspective from the pharmaceutical industry. World Patent Information 28, pp. 43–49.

9. Yang, Y. Y., Akers, L., Klose, T., Barcelon, Y. C. 2008. Text Mining and Visualization tools – Impressions of emerging capabilities. World Patent Information.

Series title, number and
report code of publication

VTT Research Notes 2451
VTT-TIED-2451

Author(s)
Ruotsalainen, Laura

Title

Data Mining Tools for Technology and Competitive Intelligence

Abstract

Approximately 80% of scientific and technical information can be found from patent documents alone, according to a study carried out by the European Patent Office. Patents are also a unique source of information since they are collected, screened and published according to internationally agreed standards. In addition to being an extremely valuable source of technology intelligence, patent documents offer a business competitive intelligence by revealing a competitor's strengths and strategies. Information gained from patents can also help in locating partners for cross-licensing and collaboration.

Since the patent system was established, more than 60 million patent applications have been published. It would be impossible to find and analyze relevant documents manually. The need for analysis and evaluation tools for patents has been acknowledged by many solution providers. New solutions are continuously coming onto the market; tools for reading and evaluating individual patents and tools for analyzing sets of patent documents. Solutions of the latter type can still be roughly divided into two groups: tools for retrieving and preparing basic statistics for patent documents, and tools for visualization and progressive analysis of patents. The former group deals only with data in a structured form, whereas the latter also analyzes unstructured text and other data.

In this study, four efficient tools for analyzing patent documents were tested: Thomson Reuter's Aureka and Thomson Data Analyzer, Biowisdom's OmniViz, and STN's STN AnaVist. All four tools analyze structured and unstructured data alike. They all visualize the results achieved from clustering the text fields of patent documents and either provide basic statistics graphs themselves or contain filters for performing them with other solutions.

The tools were tested with two cases, evaluating their ability to offer technology and business intelligence from patent documents for companies' daily business. Being aware of the state of the art of relevant technology areas is crucial for a company's innovation process. Knowledge of developed techniques and products forestalls overlapping R&D projects and thereby prevents unnecessary investment. Equally important is the recognition of other actors operating in the field. Benchmarking and evaluating a competitor's R&D and market strategies aids in managing one's own processes and locating possible parties for collaboration or cross-licensing.

This study took the point of view of a patent analyst with a basic understanding of patent data but no special knowledge of data mining techniques or the tools tested.

All the tools evaluated are very useful for the task and quite easy to adopt for daily work. All four had some strengths and weaknesses in comparison to each other. As a conclusion it could be stated that OmniViz and Thomson Data Analyzer are tools for sophisticated and diversified mathematical analysis of the data. Aureka and AnaVist are convenient for easily visualizing basic statistics and "top lists" of the data and for making stylish patent maps. The unique features of OmniViz, when compared to the other tools tested, are the possibility to visualize clustered data from many different points of view and the possibility to evaluate some attributes with patent map animations. Thomson Data Analyzer offers efficient tools for comparing different subsets of the data, e.g. for identifying unique values of an attribute. Aureka is the only tool to allow citation analyses and has the most illustrative patent map. STN AnaVist is superior in the possibility to retrieve basic statistics fast and smoothly.

The results obtained with all four tools were very much alike, even though different databases for retrieving the data were used. The top assignees and inventors lists were uniform, as were the year trends and both technological and geographical business areas. Only the reciprocal orders and amounts of documents varied. However, the conclusions drawn from the results, and business decisions made with them, would all be similar regardless of the tool used.

ISBN
978-951-38-7240-3 (soft back ed.)
978-951-38-7241-0 (URL: http://www.vtt.fi/publications/index.jsp)

Series title and ISSN VTT Tiedotteita – Research Notes 1235-0605 (soft back ed.) 1455-0865 (URL: http://www.vtt.fi/publications/index.jsp)	Project number

Date October 2008	Language English	Pages 63 p.

Name of project Patentit-Teollisuus-Tekniikka	Commissioned by Thomson Reuters, BioWisdom Ltd., American Chemical Society
Keywords patent data, text mining, data mining, patent mining, patent mapping, competitive intelligence, technology intelligence, visualization	Publisher VTT Technical Research Centre of Finland P.O. Box 1000, FI-02044 VTT, Finland Phone internat. +358 20 722 4520 Fax +358 20 722 4374

Approximately 80 % of scientific and technical information can be found from patent documents alone, according to a study carried out by the European Patent Office. Patents are also a unique source of information since they are collected, screened and published according to internationally agreed standards. In addition to being an extremely valuable source of technology intelligence, patent documents offer a business competitive intelligence. Being aware of the state of the art of relevant technology areas is crucial for a company's innovation process. Knowledge of developed techniques and products forestalls overlapping R&D projects and thereby prevents unnecessary investment. Equally important is the recognition of other actors operating in the field. Benchmarking and evaluating a competitor's R&D and market strategies aids in managing one's own processes and locating possible parties for collaboration or cross-licensing.

Since the patent system was established, more than 60 million patent applications have been published. It would be impossible to find and analyze relevant documents manually. This publication describes the results and observations obtained in a study testing four sophisticated patent analysis and visualization tools. The tools were tested with two cases, evaluating their ability to offer technology and business intelligence from patent documents for companies' daily business.

VTT RESEARCH NOTES 2451 Data Mining Tools for Technology and Competitive Intelligence

Julkaisu on saatavana	Publikationen distribueras av	This publication is available from
VTT PL 1000 02044 VTT Puh. 020 722 4520 http://www.vtt.fi	VTT PB 1000 02044 VTT Tel. 020 722 4520 http://www.vtt.fi	VTT P.O. Box 1000 FI-02044 VTT, Finland Phone internat. + 358 20 722 4520 http://www.vtt.fi

ISBN 978-951-38-7240-3 (soft back ed.)
ISSN 1235-0605 (soft back ed.)

ISBN 978-951-38-7241-0 (URL: http://www.vtt.fi/publications/index.jsp)
ISSN 1455-0865 (URL: http://www.vtt.fi/publications/index.jsp)