

# Understanding Trends in the Patent Domain

## User Perceptions on Trends and Trend Related Concepts

Julia M. Struß  
University of Hildesheim  
Institute for Information  
Science and Natural  
Language Processing  
Marienburger Platz 22  
31141 Hildesheim  
julia.struss@uni-  
hildesheim.de

Thomas Mandl  
University of Hildesheim  
Institute for Information  
Science and Natural  
Language Processing  
Marienburger Platz 22  
31141 Hildesheim  
mandl@uni-  
hildesheim.de

Michael Schwantner  
FIZ Karlsruhe – Leibniz  
Institute for Information  
Infrastructure  
Hermann-von-Helmholtz-  
Platz 1  
76344 Eggenstein-  
Leopoldshafen  
michael.schwantner@fiz-  
karlsruhe.de

Christa Womser-Hacker  
University of Hildesheim  
Institute for Information  
Science and Natural  
Language Processing  
Marienburger Platz 22  
31141 Hildesheim  
womser@uni-  
hildesheim.de

### ABSTRACT

The proceeding globalization in combination with an increasing competition in research conducted at universities and other research institutes as well as in industry, emphasises the necessity of identifying trends at an early stage, not only in industry but by universities and governments. One of the resources to be considered are patents, as most of the information contained therein is not published anywhere else. The existing research focuses on the technical perspective of identifying trends in patents. This work addresses the user perspective of the problem, in particular the user's working environments, understanding of trends, the underlying tasks and the user requirements regarding a trend mining system are examined.

### Categories and Subject Descriptors

D.2.1 [Software Engineering]: Requirements/Specifications;  
H.1.2 [Models and Principles]: User/Machine Systems—  
*Human factors*

### Keywords

trend mining on patents, requirement analysis, semi-structured interviews

### 1. INTRODUCTION

The increasing competition in research conducted at universities and other research institutes as well as in industry, further intensified by the increasing globalization, reinforces the importance of identifying new trends at an early stage. According to a study by Thomson Reuters [13], 70% to 90% of the information covered in patents – depending on the research area – is not published anywhere else. The growth of this huge information resource in terms of filed patents is also increasing faster every year: According to the annual report of the European Patent Office in 2012 new records for the third year in a row have been observed, with the largest growth in patent filings from Asian countries like China, Japan and Korea [3]. And there is also another increase of 2,8% in the number of filed patents in 2013 compared to 2012 [10].

In order to provide a system that supports the above mentioned target audience<sup>1</sup> in planning their research strategies through (semi-) automated trend detection, one needs to understand the information needs and working environments of these user groups, and most important their understanding of trends and requirements regarding the functionality of a trend mining system. This paper reports the findings of a qualitative survey on this subject with both scientists, who are working with patents, and information professionals from the patent domain.

The paper is organized as follows: In the next section related work is presented, before the methodology of this study is described in section 3. The subsequent sections present the

Copyright ©2014 for the individual papers by the papers' authors.  
Copying permitted for private and academic purposes.  
This volume is published and copyrighted by its editors.  
Published at Ceur-ws.org

Proceedings of the First International Workshop on Patent Mining and Its Applications (IPAMIN) 2014, Hildesheim, Oct. 7th, 2014.  
At KONVENS'14, October 8–10, 2014, Hildesheim, Germany.

<sup>1</sup>For a more detailed description of the target audience see section 3

results of the survey, followed by a discussion of the results in section 5 and some concluding remarks.

## 2. RELATED WORK

There have been several papers that address the technical perspective of trend mining in the patent domain. Most of this work concentrates on identifying technology trends retrospectively like [16], [6] and [2]. Other work considers related areas like the identification of patents with high novelty [4, 5] or are engaged with technology monitoring in patents [4].

Most work addresses the problem by the use of machine learning techniques (e.g. [2, 9, 11, 17]), particularly by employing clustering techniques (e.g. [1, 15]) and network analysis (e.g. [1, 2, 8, 12, 17]). In most works the final decision about the existence of a trend is left to the users, to whom the results are presented by different visualisation techniques (e.g. see [7]).

A wide range of features has been investigated in those works, like terms selected based on their frequency, to mention the most common one, (e.g. [14, 15]), adjective-noun pairs for potential technology features and verb-noun pairs for potential technology functions [17], noun and verb phrases [6], or subjective-action-object-relations (e.g. [2, 4]), but most works don't present a sound evaluation of their approaches or only evaluations on selected steps of the complete process, due to the missing evaluation resources. Instead mostly case studies are performed.

To our knowledge no study on the understanding of trends and the informational background of the potential users of such a system has been conducted so far.

## 3. METHOD

We are interested in getting deeper insights in the users' understanding of trends as well as their requirements towards a trend mining system. Therefore and due to the lack of prior studies in this area, we choose a qualitative approach and conducted semi-structured interviews.

In order to get a better idea of the working environment and the specific needs of information professionals in the patent domain, two pre-interviews were conducted with domain experts from a big information infrastructure institute working with patents and offering software products for information professionals in the patent domain. Due to this pre-interviews the area of interest was narrowed down to the engineering sciences, as patent documents in chemistry-related domains add the additional challenge of handling chemical notations, which is out of the scope of the project in whose context this research is conducted.

Seven interviews have been conducted subsequently. Three interview partners are scientists (SCI1-3) and four interview partners are information professionals (IP1-4), who either have a background as professional patent searchers (IP1, IP3), work in the IP management (IP2) or work in a company offering different patent services to clients (IP4). Figure 1 shows the questions which were asked within the interviews, where the questions were adapted to the respective group of the target audience (scientists and information

professionals). The order of the questions was not necessarily kept during the interviews, but was adapted to the particular interview situation.

The interviews were audio recorded and transcribed afterwards<sup>2</sup>, before they were analyzed with regard to the questions in figure 1.

## 4. RESULTS

In this section the insights from the interviews are presented. First the characteristics of trends as viewed by the interview partners are described, before questions and work tasks in the area of trend analysis as well as strategies for trend analysis are depicted. Section 4.4 takes a closer look at the parts and sections of a patent, which are important for trend mining. The section closes with an overview of the functions a trend mining system should offer according to the interview partners.

### 4.1 Characteristics of Trends

One main factor for recognizing a trend is the increasing number of publications in that area (SCI1, SCI2, IP1, IP3). IP1 points out that there needs to be a critical mass of patents, before you can name it a trend and suggests numbers between 20 and 50 with a stronger tendency towards 50. IP2 also gives some numbers, which range from 10 to 15, likewise with a tendency towards the higher value. These numbers can of course not be taken as strict rules, but they show, that according to different disciplines magnitudes can be quite different. One reason for this can be seen in the size of the research area and another in the understanding of a trend with regards to the content and the granularity of interest.

Other factors for recognizing a trend in the context of the patent domain are the appearance of new IPC-classes or the frequent co-occurrence of IPC-classes from different areas of research assigned to the patents (IP1).

When it comes to time spans of trend evolutions the interview partners mostly agree, that it is a matter of several years. IP2 is giving the smallest time span ranging from several months up to one or two years, IP3 also gives a range from about two years, whereas IP1, SCI1 and IP4 describe longer time periods between five years (IP1) and ten years (IP4), with IP4 emphasising the fact that these numbers can be quite different from discipline to discipline.

According to the granularity of the abstraction level of the content in the context of trend analysis the interview partners are mainly interested in two levels, which are not specific to one group of interview partners: On the one hand trends on the top level of an entire research area, and on the other hand detailed subject-specific or technical developments within a field of interest are mentioned. SCI1 explains for example, that a scientist usually knows the specific developments within the own research area, whereas it would be interesting to see trends of neighboring disciplines, which might inspire the own direction of research. Contrariwise

<sup>2</sup>One interview partner did not allow to audio record the interview, therefore the interview notes were used for further analysis.

- Please give some details on your personal background and your working environment\*/ on your research area\*\*.
- Do you selectively conduct trend searches / analysis or trend observations?
- Do you include patents in this search or analysis process?\*\*\*
- What kind of questions do you want to answer by these trend searches / analysis?
- How would you characterise such a trend or what makes a trend a trend in your working environment?
  - What kind of shapes with regard to the trend curve are of interest?
  - At which time points are those curves interesting?
  - What are the time periods we are talking about (months, years)?
  - Which time related fields should be used for measuring a trend?
  - What is the subject of a trend in terms of content (the granularity level of the content)?  
Could you give an example?
  - How would you measure such a trend?
  - Where can one see a trend at first (what kind of publications)?
- How do you realize, that a trend is developing?
- What does the result of a trend analysis look like?\*
- What strategy do you pursue, when you do a trend analysis and what steps can you identify in the process?
- Which parts of a patent are most applicable or effective in this context?
- Which functions should a trend mining system offer?

\* information professionals \*\* scientists

**Figure 1: Questions for the semi-structured interviews**

SCI3 focuses on the more subject-specific type of trends. As mentioned before the information professionals are also interested in both types of trends. IP1 explains, that customers who want to use a specific technology (e.g. SMEs) are more interested in IPC-class level trends, whereas enterprises wanting to control a commercialization process or to get full market coverage are interested in more fine grained information, like on substance or technology level, when it comes to trend analysis.

## 4.2 Information Needs in Trend Analysis

Trend searches or analysis are conducted with different aims or objectives and are guided by different questions. One question coming up in both groups of interview partners is concerned with finding out if it is worthwhile to engage oneself with a specific research topic (SCI1, IP1, IP3), although there are different reasons behind this question. SCI1 is interested in knowing if there is a possibility of funding, that is worth the effort of preliminary work and writing an application, as this process takes approximately 1.5 years. IP1 constitutes the importance of knowing if the area is already covered by patents and IP3 expresses the situation, that the existing patents mean, that competitors have been working for more than 1.5 years in an area, once the patents are available to the public, due to the 18 month delay in publication.

Another question in the context of trend analysis regards the persons, research teams and companies already engaged in the area of interest. On the one hand the interview partners are interested in knowing how many of them are there (SCI1), on the other hand they are specifically interested in observing the competitors (IP2) or finding out how big the development team of a specific competitor is, as this is an indicator of how important a topic is to that competitor (IP3).

Other questions have a broader focus, e.g. ask about the development of new technical fields (IP1) or the direction the development in a technical field is taking (IP1, IP4). There are also questions which are dealing with possible markets (IP1).

## 4.3 Points of Interest in the Trend Evolution

The above presented characteristics and information needs do have an influence on the points of interest within the development of a trend. Most interview partners agree, that the beginning of a trend is a point in time, when a trend becomes interesting (SCI1, IP1–4). This is especially the case, if the reason for the analysis is to get involved in a specific area of research.

The information professionals also consider other points in the evolution of a trend as interesting and stress the depen-

dence on the requests of the clients and customers (IP1, IP3, IP4). Some customers are interested in licencing a specific technology, which means it needs to be functional already, and therefore a later point in the evolution of the trend is interesting (IP1). IP4 describes a similar scenario and assigns descending trends to those customers. A descending trend curve with regard to patent applications does not mean, that a trend is ending, but that the technology has reached a certain degree of maturity.

#### 4.4 Applicable Sections of a Patent for Trend Mining

The question about applicable sections for trend mining on the one hand aimed at clarifying which date related fields should be used for trend mining and on the other hand which content related sections of a patent are best suited for trend mining.

Date related fields for patents include application dates, priority dates and publication dates. The application date refers to the date of the application at the patent office, whereas the publication date denotes the date, when the patent was made available to the public, which can be up to 18 month after the application was handed in. If there are multiple applications to different patent offices for an invention, these patents form a patent family<sup>3</sup>. The earliest application date of a patent family is denoted as the priority date.

Related work in trend mining on patents uses different date related fields to explore temporal developments. Some works choose the application date (e.g. [6]) while others prefer the publication date of a patent (e.g. [4,7]). The interview partners mostly agreed that for trend mining the priority date would be the date related field of choice. Although some acknowledge, that one could use the application date (IP2, IP3). According to IP1 the publication date could be useful, if the impact of an invention on an industrial sector is of interest.

With respect to the content related sections, a wide variety has been used in prior research: title and abstract have been used as well as claims and descriptions and varying combinations of these (e.g. see [2,5,12,14,17]). The same variety is also found in the interviews. Table 1 lists the content related sections suggested or excluded by the individual interview partners.

Especially when it comes to titles and abstracts the opinions diverge. IP1 explains, that it depends on the database whether these two fields could be used for determining the content of a patent: Some providers of patent information offer added values like manually rewritten titles and abstracts according to the contents of a patent and therefore make these a good data resource, while titles and abstracts taken directly from the patent application often form a bad base for analysis (IP1) as the applicants try to conceal the content and claim of a patent, in order to keep it as broad as possible.

<sup>3</sup>For further details on patent families see for example [http://www.intellogist.com/wiki/Patent\\_Families](http://www.intellogist.com/wiki/Patent_Families)

scientists	information professionals
first main claim, main claims (SCI1) claims (SCI3)	claims (IP4) perh. claims (IP3)
description (SCI2, SCI3)	first page of the description (IP3) the replication of contents in the description dilute the results (IP1)
figures (SCI3)	figures (IP2) perh. figures (not for informatics or telecommunications)
	edited / enhanced titles (IP1) titles (IP3)
	edited / enhanced abstracts (IP1) abstract (IP2, IP3) abstracts are too general (IP4)
	introduction, especially the task description (IP3)

**Table 1: Content related sections of a patent (not applicable for trend mining)**

#### 4.5 Trend Analysis Strategy

Besides the information needs and their understanding of a trend the interview partners were also asked for their strategies with regard to trend searches and analysis.

IP1 gives descriptions of strategies for both of the above mentioned trend types. When the interest is primarily on the first type of trends e.g. within an IPC-class, he first creates a basic set of documents and then aggregates the patents with regard to their respective patent families in order to avoid duplicate counting of the same invention. If necessary the document set is further aggregated according to national patent families and then the number of patents per year based on the priority date are calculated and visualized. The last step would be to select technology areas with growth above average and if necessary conduct further analysis.

For the second trend type IP1 proposes an iterative approach, involving the client at every stage of the process. Especially at the beginning, according to IP1 clients are not always able to explain their objectives or questions explicitly. Another point is, that concept names used within one company might be different from those commonly used in patents, or there might as well be some variety in the concept names found in the respective patents. Therefore as a first step a patent landscape of the domain of interest needs to be generated and then explored together with the client. This serves the goal of getting a common understanding of the task at hand and identifying aspects of a topic which are of special interest to the client. These identified areas are then further analyzed with text mining techniques like

clustering.

IP3 gives a description of how to get the basic document set for the analysis. He starts off with known competitor names and their publications and then looks at the IPC classes and might take those into consideration as well.

## 4.6 Functions of a Trend Mining System

At the end of the interview the scientists and information professionals were asked what kind of functions a trend mining system should possess. These range from possibilities to drill down within a research area to more specific areas and explore trends at every stage, to having an alert function informing about changes in a predefined area of interest (SCI1).

IP1 describes the ideal trend mining system as a system possessing two modes, one standard mode and one advanced mode for experts. Both modes should be transparent to the user and make interim results accessible in order to make the process comprehensible. The advanced mode should additionally give the possibility of taking actions at various steps during the process, like incorporating additional knowledge about the domain in question or defining the number of clusters that should be build during a clustering step.

Another important aspect are interactive visualisations of the results, enabling the user for example to zoom in for more details (IP3). IP1 also remarks that visualisations that help to understand the contents of a set of documents is a desirable feature and make it possible to explore results together with costumers.

## 5. DISCUSSION

As this study has the character of an exploratory study and only a small sample is involved, the findings of this study can only give first insights into the domain and a starting point for further research, but the variety of information needs and understandings of trends within just the field of engineering sciences emphasises the necessity of incorporating the target audience in the development process of a trend mining system.

The presented results show that there are quite a few differences in the understanding of trends or the characteristics that make a trend interesting to the target audience, although the interview partners mostly had a background in engineering.

Mainly two types of trends, that are interesting to the target audience, could be identified: Trends at the top level of an entire research area or domain and subject-specific or technical developments within a specific area of interest. The results also show, that the time spans encompassing a trend can be quite different according to the content granularity of interest and the domain of interest.

Additionally the results of the interviews show, that not only emerging trends are of interest to the target audience, but also trends which have reached their height or are even on a decreasing path, as this denotes, that a technology has reached a stage, where it can be used, and licenced by other organisations to incorporate them in their own products.

The interest on trends at this stage are mainly ascribed to SMEs.

The study also shows that research is needed with regard to the question of which content related sections of a patent are best applicable for trend mining, due to the fact that almost every content related section has been named by at least one interview partner.

The findings show as well, that at least for some of the patents searchers it is important to integrate their customers and clients in the trend mining process. Therefore a system with such a target audience should also incorporate visualisation techniques, that allow for exploring analysis results together with clients and make it easy for a non-patent specialist to understand the results shown by the trend mining system.

## 6. CONCLUSIONS

This paper gives first insights into the user perspective of trend analysis in the patent domain. Besides showing different perspectives and understandings of trends as well as pointing out characteristics making a trend interesting to the target audience within the area of engineering sciences, the study gives first insights into the underlying tasks and information needs of the target audience and some requirements regarding the functionality of a trend mining system in the patent domain.

The study also shows the necessity for further research when it comes to the question of which content related sections of a patents are applicable for trend mining, as there is neither a clear picture on this aspect from the interviews, nor is there in related research.

## 7. ACKNOWLEDGMENTS

This work was conducted as part of the project "Trendmining für die Wissenschaft"<sup>4</sup> (T4P), which is a joint project of FIZ Karlsruhe – Leibniz Institute for Information Infrastructure and the Institute for Information Science and Natural Language Processing at the University of Hildesheim and is funded by the Leibniz Association in the context of the Leibniz Competition.

## 8. REFERENCES

- [1] P.-L. Chang, C.-C. Wu, and H.-J. Leu. Using Patent Analyses to Monitor the Technological Trends in an Emerging Field of Technology: a Case of Carbon Nanotube Field Emission Display. *Scientometrics*, 82(1):5–19, 2010.
- [2] S. Choi, J. Yoon, K. Kim, J. Y. Lee, and C.-H. Kim. SAO Network Analysis of Patents for Technology Trends Identification: a Case Study of Polymer Electrolyte Membrane Technology in Proton Exchange Membrane Fuel Cells. *Scientometrics*, 88(3):863–883, 2011.
- [3] European Patent Office. Annual Report 2012: Statistics and trends: Total European patent filings, 2013. Online available at: <http://www.epo.org/about-us/>

<sup>4</sup>translation: trend mining for sciences

- annual-reports-statistics/annual-report/2012/statistics-trends/patent-filings\_de.html, last accessed: 2013-09-16.
- [4] J. M. Gerken. *PatMining – Wege zur Erschließung textueller Patentinformationen für das Technologie-Monitoring*. PhD thesis, Universität Bremen, Bremen, 2012.
- [5] P. Hu, M. Huang, P. Xu, W. Li, A. K. Usadi, and X. Zhu. Finding Nuggets in IP Portfolios: Core Patent Mining Through Textual Temporal Analysis. In X. Chen, editor, *Proceedings of the 21st ACM International Conference on Information and Knowledge Management*, pages 1819–1823, New York and NY and USA, 2012. ACM.
- [6] Y. Kim, Y. Tian, Y. Jeong, R. Jihee, and S.-H. Myaeng. Automatic Discovery of Technology Trends from Patent Text. In *Proceedings of the 2009 ACM Symposium on Applied Computing*, pages 1480–1487, New York and NY and USA, 2009. ACM.
- [7] C. Lee, J. Jeon, and Y. Park. Monitoring Trends of Technological Changes Based on the Dynamic Patent Lattice: A Modified Formal Concept Analysis Approach. *Technological Forecasting and Social Change*, 78(4):690–702, 2011.
- [8] H. Park, K. Kim, S. Choi, and J. Yoon. A Patent Intelligence System for Strategic Technology Planning. *Expert Systems with Applications*, 40(7):2373–2390, 2013.
- [9] W. M. Pottenger and T.-H. Yang. Detecting Emerging Concepts in Textual Data Mining. In M. W. Berry, editor, *Computational Information Retrieval*, pages 89–105. Society for Industrial and Applied Mathematics, Philadelphia and PA and USA, 2001.
- [10] O. Schröder. Facts and Figures 2014. European Patent Office, 2014. Online available at: [http://documents.epo.org/projects/babylon/eponet.nsf/0/125011cc1d9b8995c1257c92004b0728/\\$FILE/epo\\_facts\\_and\\_figures\\_2014\\_en.pdf](http://documents.epo.org/projects/babylon/eponet.nsf/0/125011cc1d9b8995c1257c92004b0728/$FILE/epo_facts_and_figures_2014_en.pdf), last accessed: 2014-06-24.
- [11] M.-J. Shih, D.-R. Liu, and M.-L. Hsu. Mining Changes in Patent Trends for Competitive Intelligence. In T. Washio, E. Suzuki, K. M. Tin, and A. Inokuchi, editors, *Advances in Knowledge Discovery and Data Mining*, volume 5012 of *Lecture Notes in Computer Science*, pages 999–1005. Springer, Berlin and Heidelberg, 2008.
- [12] J. Tang, B. Wang, Y. Yang, P. Hu, Y. Thao, X. Yan, B. Gao, M. Huang, P. Xu, Li, Weichang, and A. K. Usadi. PatentMiner: Topic-driven Patent Analysis and Mining. In *Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, New York and NY and USA, 2012. ACM.
- [13] The Thomson Corporation. *Global Patent Sources: An Overview of International Patents*. Thomson Scientific, London, 6 edition, 2007. Online available at: [http://ip-science.thomsonreuters.com/m/pdfs/mgr/global\\_patent\\_sources.pdf](http://ip-science.thomsonreuters.com/m/pdfs/mgr/global_patent_sources.pdf), last accessed: 2013-09-16.
- [14] M.-Y. Wang, D.-S. Chang, and C.-H. Kao. Identifying Technology Trends for R&D Planning Using TRIZ and Text Mining. *R&D Management*, 40(5):491–509, 2010.
- [15] B. Yoon and Y. Park. A Text-mining-based Patent Network: Analytical Tool for High-technology Trend. *The Journal of High Technology Management Research*, 15(1):37–50, 2004.
- [16] B. Yoon and Y. Park. A Systematic Approach for Identifying Technology Opportunities: Keyword-based Morphology Analysis. *Technological Forecasting and Social Change*, 72(2):145–160, 2005.
- [17] J. Yoon, S. Choi, and K. Kim. Invention Property-function Network Analysis of Patents: a Case of Silicon-based Thin Film Solar Cells. *Scientometrics*, 86(3):687–703, 2011.