Trend Mining on Patent Information for Sciences

Project cooperation between the FIZ Karlsruhe and the University of Hildesheim

Funded by the Leibniz Association

Outline

- Aim of the project

- Motivation

- Related Work

- Project outline

Aim of the project

- Developing and testing semantic and statistic methods for (semi-) automatic trend detection in patent documents

- Development of an easy-to-use prototype, that supports scientists in planning their research activity.

Motivation

- Results of technical sciences are often only published as patents (70-90% of the contents in patents is not published anywhere else (cf. The Thomson Corporation 2007: 5)

- Strong growth in patent filing: EPO reports new records for the third year in a row (cf. European Patent Office 2013)

Visualisierung

	Visualisierung								semi-automatisch	
	Häufigkeitstabellen	Histogramme	weighted ratios	log-log graphs	Fisher-Pry curves	technology map	Timeline	Timeriver	subject-action-object model	Generalized Sequential Patterns
Allan et al. 1998*	X						X			
Havre et al. 2002*	X							X		
Swan et al. 2000*	X						X			
Shih et al. 2008										
Lent et al. 1997*	X		X						X	X
Chang et al. 2010										
Yoon et al. 2011										
Nanba et al. 2010										
Porter et al. 1995*	X	X	X	X	X	X	X			
Pottenger et al. 2001*	X									
Choi et al. 2011									X	
Blank et al. 2001*										

* nach Kontostathis, Galitsky et al. 2004

Project outline

- **Requirements Analysis**
  - Get to know the working environment and the information behavior of the users (scientists and information professionals)
  - What do users regard as a trend?
  - What are the areas of interest when it comes to trend mining?

- **Corpus Development**
  - Based on EPFULL (3.2 million patents; 76% in English)
  - Identify important fields of the database
  - Determine size of the corpus

Project outline II

- Text Mining and Trend Analysis
  - Identification of relevant document subsets for given topics via state-of-the-art searches
  - Implementation of topic detection and tracking (TDT) methods in order to build topic clusters as basis for the trend analysis
  - Identification of patterns based on variable time lines that help detecting trends (e.g. regarding the appearance and development of new terms or term combinations)
  - Allow for three dimensional trend detection by incorporating geographic data
  - Development of an automatic and an interactive approach

- **Development of the Prototype**
    - Iterative development
    - user-centered design
    - Based on Solr/Lucene and UIMA
    - Web interface for user interaction containing various visualization approaches
- **Evaluation (formative and summative)**
    - Development of a benchmark in cooperation with scientists and information professionals to evaluate the efficiency of the system
    - User studies accompanied by interviews about the system's efficiency as well as expert interviews

References

- Allan, James; Papka, Ron; Lavrenko, Victor (1998): On-line New Event Detection and Tracking. In : Proceedings of the 21st annual international ACM SIGIR conference on Research and development in information retrieval. ACM SIGIR. New York, NY, USA: ACM, pp. 37–45. Available online at http://doi.acm.org/10.1145/290941.290954, checked on 2/09/2013.

- Blank, Glenn David; Pottenger, William M.; Kessler, G. Drew; Herr, Martin; Jaffe, Harriet; Roy, Soma (2001): CIMEL: Constructive, Collaborative Inquiry-based Multimedia E-Learning. (poster session). In : Proceedings of the 6th Annual Conference on Innovation and Technology in Computer Science Education. ITiCSE. Canterbury, UK, June 25-27, 2001. New York, NY, USA: ACM, p. 179. Available online at http://doi.acm.org/10.1145/377435.377692, checked on 2/09/2013.

- Chang, Pao-Long; Wu, Chao-Chan; Leu, Hoang-Jyh (2010): Using Patent Analyses to Monitor the Technological Trends in an Emerging Field of Technology: a Case of Carbon Nanotube Field Emission Display. In *Scientometrics* 82 (1), pp. 5–19.

- Choi, Sungchul; Yoon, Janghyeok; Kim, Kwangsoo; Lee, Jae Yeol; Kim, Cheol-Han (2011): SAO Network Analysis of Patents for Technology Trends Identification: a Case Study of Polymer Electrolyte Membrane Technology in Proton Exchange Membrane Fuel Cells. In *Scientometrics* 88 (3), pp. 863–883.

- European Patent Office (2013): Annual Report 2012. Statistics and trends. Total European patent filings. Available online at http://www.epo.org/about-us/annual-reports-statistics/annual-report/2012/statistics-trends/patent-filings.html, updated on 6/03/2013, checked on 16/09/2013.

- Havre, S.; Hetzler, E.; Whitney, P.; Nowell, L. (2002): ThemeRiver: Visualizing Thematic Changes in Large Document Collections. In *IEEE Trans. Visual. Comput. Graphics* 8 (1), pp. 9–20.

- Huang, L. C.; Li, Y. (2010): Research on Technology Trend Based on Patent Information. In : Proceedings of the 5th IEEE International Conference on Management of Innovation and Technology. ICMIT. Singapore, 2 - 5 June 2010. Piscataway, N.J.: IEEE, pp. 209–213.

- Kontostathis, April; Galitsky, Leon M.; Pottenger, William M.; Roy, Soma; Phelps, Daniel J. (2004): A Survey of Emerging Trend Detection in Textual Data Mining. In Michael W. Berry (Ed.): Survey of Text Mining. Clustering, Classification, and Retrieval. New York, London: Springer, pp. 185–224.

- Lent, Brian; Agrawal, Rakesh; Srikant, Ramakrishnan (1997): Discovering Trends in Text Databases. In David Heckerman (Ed.): Proceedings of the Third International Conference on Knowledge Discovery & Data Mining. KDD. Newport Beach, California, August 14 -17, 1997. Menlo Park, Calif: AAAI Press, pp. 227–230.

References

- Nanba, Hidetsugu; Fujii, Atsushi; Iwayama, Makoto; Hashimoto, Taiichi (2010): Overview of the Patent Mining Task at the NTCIR-8 Workshop. In : Proceedings of the 8th NTCIR Workshop Meeting on Evaluation of Information Access Technologies: Information Retrieval, Question Answering, and Cross-Lingual Information Access. NTCIR. Tokyo, Japan, June 15-18, 2010. National Institute of Informatics, pp. 293–302. Available online at http://research.nii.ac.jp/ntcir/workshop/OnlineProceedings8/NTCIR/01-NTCIR8-OV-PATMN-NanbaH.pdf, checked on 4/09/2013.

- Porter, Alan L.; Detampel, Michael J. (1995): Technology opportunities analysis. In *Technological Forecasting and Social Change* 49 (3), pp. 237–255.

- Pottenger, William M.; Yang, Ting-Hao (2001): Detecting Emerging Concepts in Textual Data Mining. In Michael W. Berry (Ed.): Computational Information Retrieval. Philadelphia, PA, USA: Society for Industrial and Applied Mathematics, pp. 89–105. Available online at http://dl.acm.org/citation.cfm?id=762544.762552.

- Shih, Meng-Jung; Liu, Duen-Ren; Hsu, Ming-Li (2008): Mining Changes in Patent Trends for Competitive Intelligence. In Takashi Washio, Einoshin Suzuki, Kai Ming Tin, Akihiro Inokuchi (Eds.): Advances in Knowledge Discovery and Data Mining. 12th Pacific-Asia Conference, PAKDD 2008 Osaka, Japan, May 20-23, 2008 Proceedings, vol. 5012. Berlin, Heidelberg: Springer (Lecture Notes in Computer Science, 5012), pp. 999–1005.

- Swan, Russell; Jensen, David (2000): TimeMines: Constructing Timelines with Statistical Models of Word Usage. In : Proceedings of the Sixth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM SIGKDD. Boston, MA, USA, August 20-23, 2000. Available online at http://www.cs.cmu.edu/~dunja/KDDpapers/Swan_TM.pdf, checked on 2/09/2013.

- The Thomson Corporation (2007): Global Patent Sources. An Overview of International Patents. 6th ed. London: Thomson Scientific. Available online at http://ip-science.thomsonreuters.com/m/pdfs/mgr/global_patent_sources.pdf, checked on 16/09/2013.

- Wang, Ming-Yeu; Chang, Dong-Shang; Kao, Chih-Hsi (2010): Identifying Technology Trends for R&D Planning Using TRIZ and Text Mining. In *R&D Management* 40 (5), pp. 491–509.

- Yoon, Janghyeok; Choi, Sungchul; Kim, Kwangsoo (2011): Invention Property-function Network Analysis of Patents: a Case of Silicon-based Thin Film Solar Cells. In *Scientometrics* 86 (3), pp. 687–703.