

Dynamisches Relevanz-Feedback im Patent-Retrievalsystem PatentAide

René Hackl

Universität Hildesheim
31141 Hildesheim, Deutschland
rene.a.hackl@gmx.de

Abstract

Im Patent Retrieval haben sich Rankingverfahren und Methoden wie Relevanz-Feedback noch nicht etabliert. An Ranking Systemen wird vor allem die mangelnde Transparenz für den Benutzer bemängelt. Das System PatentAide versucht, aufbauend auf einer Analyse der Rechercheprozesse im Patent Retrieval, ein Ranking-System zu implementieren. PatentAide unterstützt wichtige Techniken im Patent-Retrieval Prozess wie Term-Erweiterung, bietet ein geranktes Ergebnis und erlaubt darüber hinaus dynamisches Relevanz-Feedback.

1 Einleitung

Patente befinden sich an der Schnittstelle von Recht, Technik und Wirtschaft. Sie sind exklusive Schutzrechte, die dem Patentinhaber für einen begrenzten Zeitraum – maximal 20 Jahre – das alleinige Verwertungsrecht gewähren. Im Gegenzug zur Gewährung dieser umfangreichen Schutzrechte¹ legt der Patentanmelder die Erfindung offen. Es wird geschätzt, dass über 80% des technischen Wissens allein in Patenten vorliegt. Damit sind Patente eine der wichtigsten Quellen detaillierter technischer Information. Für Unternehmen nehmen Patente in Bereichen wie Forschung und Entwicklung, Marketing und Intellectual Property Rights Management die Position einer Schlüsselressource ein. Nicht nur wird der „internationale Kampf um Marktanteile [...] zunehmend über den Innovationsvorsprung entschieden“ [Weckend u. Wurzer, 2004], auch strategische Verfahren wie der Austausch von Patentlizenzen sind zu wichtigen Instrumenten im Unternehmensalltag geworden.

2 Patent Retrieval

Die Patentrecherche ist ein komplexes, zeitaufwändiges Arbeitsgebiet, das weitgehende Anforderungen an Rechercheure stellt. Rechercheergebnisse müssen höchsten Ansprüchen in Bezug auf Gründlichkeit, Qualität und Vollständigkeit genügen. Häufig hängen wichtige Forschungs- und Geschäftsentscheidungen von Patentanalysen ab. Zentrale Bestandteile der Patentrecherche umfassen die Entwicklung von Suchstrategien, die Auswahl von Informationsquellen und

den Umgang mit zur Verfügung stehenden Werkzeugen.

Der Zugriff auf und Umgang mit Informationsquellen wird allerdings immer selbstverständlicher, so dass nicht nur Berufsrechercheure als Patentsucher auftreten, sondern auch und in zunehmendem Maße Wissenschaftler, Manager, Ingenieure und Patentanwälte. Daher ist eine Hauptforderung der Nicht-Profis, die Produkte und Werkzeuge mögen sich ihrem Niveau anpassen und sie in der Suche besser unterstützen [Mendelsohn, 2000]. Analog dazu argumentiert [Poynder, 2000], die Usability der vorhandenen Systeme müsse sich verbessern, die Gewinnung von Mehrwert müsse einfacher zu erreichen sein.

Im Information Retrieval etablierte unterstützende Techniken wie Ranking und Relevanz-Feedback [s. Womser-Hacker, 1997] kommen in der Patentdomäne allenfalls sporadisch und nicht den Anforderungen entsprechend zum Einsatz. IR-Systeme, die auf der booleschen Logik aufsetzen, bilden seit Jahrzehnten die Grundlage der Patentrecherche. Neuere Werkzeuge und Produkte verfehlen mehrheitlich die speziellen Anforderungen. Insbesondere wird fehlende Kontrollierbarkeit bemängelt.

3 Gestaltungsprinzipien des Prototypen

Das System PatentAide wurde in einem Kooperationsprojekt zwischen der Informationswissenschaft der Universität Hildesheim und dem Fachinformationszentrum Karlsruhe Gesellschaft für wissenschaftlich-technische Information mbH (FIZ Karlsruhe)² entwickelt. Das Projekt geht der Frage nach, inwiefern die unterstützenden Information – Retrieval – Techniken Ranking und Relevanz-Feedback in der Patentrecherche zum Einsatz kommen können. Beiden Techniken ist normalerweise ein ausgesprochener *black box* Charakter gemein. Diese Eigenschaft macht sie ungeeignet für den Patentbereich. Das Hauptziel des Prototypen besteht daher darin, zu prüfen, wie transparente Bedingungen für die Komponenten geschaffen werden können. Dabei werden die Angabe statistischer Werte, der Gebrauch von Visualisierungen und verschiedene Interaktionsformen vor dem Hintergrund typischer Arbeitsabläufe thematisiert.

Die Entwicklung des Systems ging von Anfang an unter partizipatorischen Gesichtspunkten vorstatten. Durch das frühzeitige Hinzuziehen von Experten und iteratives Vorgehen wurde gewährleistet, dass

¹ §§ 9-14 PatG

² <http://www.fiz-karlsruhe.de>

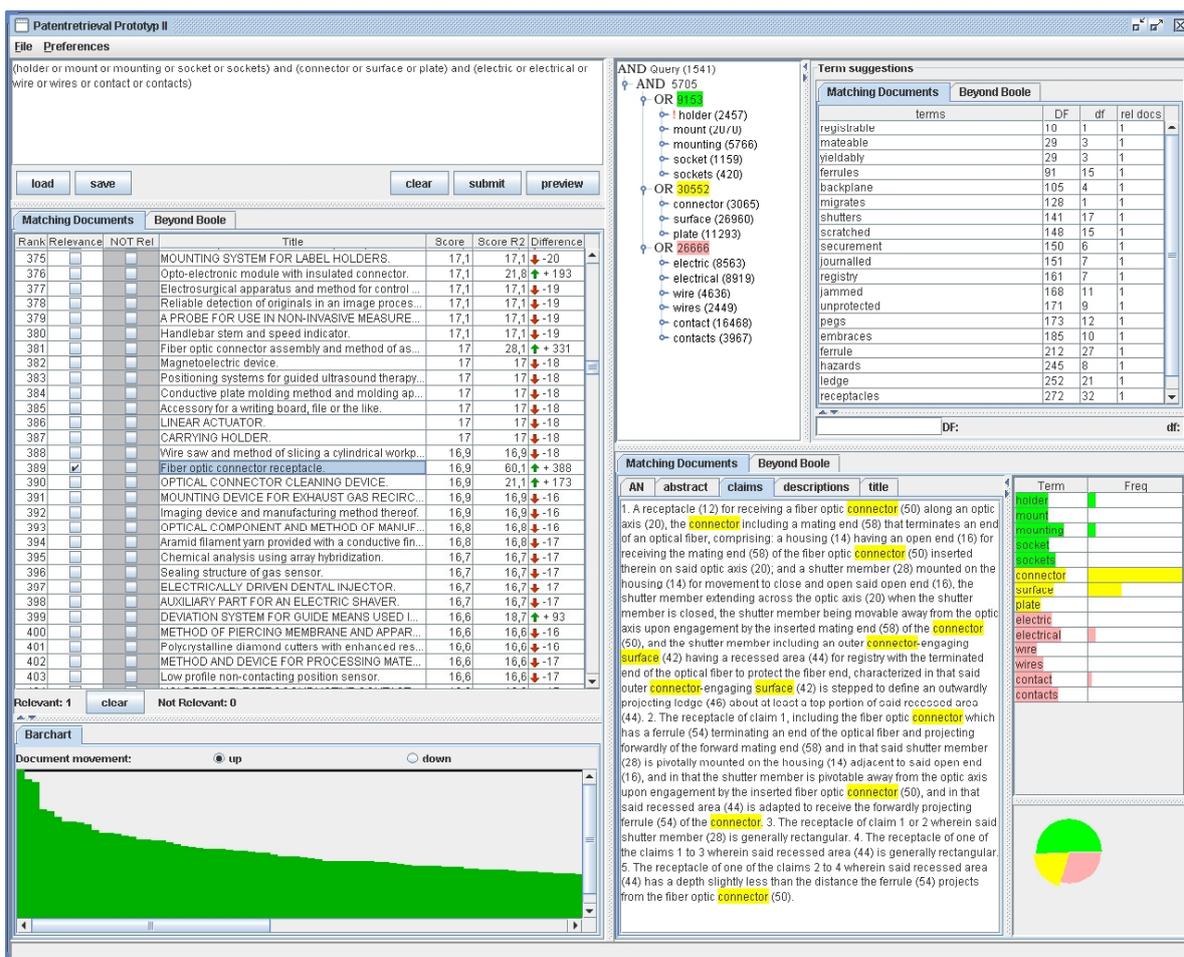


Abbildung 1: Benutzungsoberfläche des PatentAide

gewonnene Erkenntnisse nicht nur von akademischen Interesse sind. Die Patentexperten des FIZ Karlsruhe wurden bereits in der planerischen Phase einbezogen. Für die Relevanz-Feedback Komponente wurde als Grundlage auf ein an der Universität Hildesheim entwickeltes System zurückgegriffen, das bereits mehrfach evaluiert wurde [s. Hackl und Mandl, 2006]. Nach den ersten Entwicklungszyklen wurden immer weiter entwickelte Versionen einem breiteren Publikum und externen Fachleuten vorgestellt und sind auf positive Resonanz gestoßen.

Abbildung 1 zeigt die zweite Version des Prototypen von PatentAide. Eine Anfrage wurde zunächst über das Anfragefenster links oben in einen speziell entwickelten Preview-Modus überführt. Dieser Modus, das schmale, mittlere Fenster, repräsentiert die Anfrage in einer Baumstruktur und gibt Hinweise auf die Zusammensetzung der Treffermenge. Weiterhin unterstützt er verschiedene Arten der Anfragemodifikation wie z.B. der Erweiterung um Thesauruseinträge. Im weiteren wurde die modifizierte Anfrage abgeschickt und in der resultierenden, gerankten Trefferliste (links in der Mitte) ein Dokument angesehen. Die Dokumentansicht befindet sich in der rechten unteren Hälfte. Es ist anhand der verschiedenen Graustufen zu erahnen, dass die Terme aus der Anfrage, die im Textabschnitt

„claims“ zu sehen sind, farblich hervorgehoben sind. Außerdem werden die Vorkommen aller Anfrageterme im Gesamtdokument durch farbige Balken angezeigt. In der rechten unteren Ecke ist eine Studie zur Darstellung der Termgewichte auf einen Blick zu sehen. Das betrachtete Dokument wurde für relevant befunden und entsprechend markiert. Zum einen sind daraufhin die Dokumentgewichte neu berechnet worden. Das Ergebnis dieses Prozesses wird sowohl in einer Spalte der Ergebnisliste wie auch in dem Balkendiagramm links unten dargestellt. Zum anderen wurden Termvorschläge auf Grundlage der Relevanzinformation ermittelt. Diese potentiellen Erweiterungsterme werden in der Tabelle rechts oben aufgeführt. Sie können per Drag&Drop in der gewünschten Position der Baumdarstellung eingefügt werden.

Diese Term-Erweiterung zur Optimierung einer Anfrage ist ein typisches Szenario im Patent-Retrieval. Transparenz und Kontrollierbarkeit werden durch die sofortige Anzeige der Änderungen im Preview-Modus erreicht. Die Treffermenge, auf der gearbeitet wird, wird bis zu der expliziten Anweisung nicht verändert.

Auf diese Weise können Anfragen schrittweise verfeinert und Antwortmengen explorativ erschlossen werden. Durch bereitgestellte Konfigurationsmöglichkeiten kann der Prototyp individuell angepasst werden.

4 Ausblick

Der vorgestellte Stand besitzt die angestrebten Kernfunktionalitäten. In weiterhin iterativen Entwicklungsschritten werden jetzt einige Feinheiten und Optimierungen getestet. Anschließend erfolgt eine Evaluierung mit Patentexperten, in der insbesondere untersucht wird, inwiefern die aufgezeigten Möglichkeiten zur Unterstützung der täglichen Recherchearbeit beitragen.

Literatur

- [Hackl u. Mandl, 2006] René Hackl and Thomas Mandl. Bilingual Retrieval Experiments with Social Science Documents. In *Accessing Multilingual Information Repositories: 6th Workshop of the Cross-Language Evaluation Forum, CLEF 2005, Vienna, Austria, Revised Selected Papers*. Berlin et al.: Springer [Lecture Notes in Computer Science 4022]
- [Mendelsohn, 2000] Susan Mendelsohn. Patterns formed by a single shot of malt. In *Information World Review*, Juni 2000.
- [Poynder, 2000] Richard Poynder. Web-challenged. In *Information World Review*, Juli/August 2000.
- [Weckend u. Wurzer, 2004] Edwin Weckend and Alexander Wurzer: Patentrecherche als Produktionsprozess: Elemente eines Qualitätsmanagements. In *PATINFO 2004*, 2004, S. 151-176
- [Womser-Hacker, 1997] Womser-Hacker, Christa: *Das MIMOR-Modell. Mehrfachindexierung zur dynamischen Methoden-Objekt-Relationierung im Information Retrieval*. Habilitationsschrift. Universität Regensburg, Informationswissenschaft, 1997.