

Visual Patent Retrieval

Mark Giereth, Steffen Koch, Harald Bosch und Thomas Ertl
Institut für Visualisierung und Interaktive Systeme, Universität Stuttgart
Universitätsstraße 38, 70569 Stuttgart

Schlagnworte: Rechtsvisualisierung, Patentinformation, Visuelle Anfragesprachen, Annotation

Abstract: Dieser Artikel beschreibt das PatViz Framework zur Visualisierung von Suchanfragen und Ergebnismengen im Patentbereich. PatViz erlaubt die Verwaltung komplexer Patentsuchanfragen auf Basis einer visuellen Anfragesprache. Es unterstützt die Analyse von Suchergebnissen hinsichtlich deren zeitlicher, geographischer, klassifikatorischer und stichwortbezogener Verteilung. PatViz baut dabei auf einem semantischen Framework auf. Es integriert die Ergebnisse automatischer semantischer Analysen und ermöglicht eine manuelle semantische Annotation von Patenten auf Basis eines semantischen Wiki.

1. Einleitung

Die Bedeutung geistigen Eigentums und dessen Sicherung durch Patente gewinnt in modernen wissensorientierten Gesellschaften zunehmend an Bedeutung. Bei mehr als 60 Millionen Patentdokumenten weltweit (davon ca. 5,6 Million in Kraft) und einer durchschnittliche Wachstumsrate von jährlich 4.7%¹ werden effiziente Werkzeuge für das Finden, Bewerten und Verstehen patentrelevanter Informationen immer wichtiger. Neben neuen Verfahren zum Information Retrieval, zur Informationsextraktion und zur semantischen Analyse, stellen interaktive Visualisierungstechniken einen immer wichtiger werdenden Faktor für die Verarbeitung großer Datenmengen – in der Rechtsinformatik im Allgemeinen und im Patentwesen im Besonderen – dar.

Im Rahmen des europäischen Forschungsprojekts PatExpert² wurden neue Verfahren in den genannten Bereichen entwickelt, mit dem langfristigen Ziel eines Paradigmenwechsels von textueller Verarbeitung hin zu semantisch visueller Verarbeitung von Patentinformation [1]. In punkto Patentvisualisierung wurden hierbei vier Bereiche untersucht. Der erste Bereich umfasst Patentmetadaten, wie beispielsweise Patentklassifikation, Patentfamilien, Zitierungen, Rechtsstandsinformation, Anmelder oder Erfinder. Der zweite Bereich befasst sich mit Patentdokumenten an sich, d.h. deren Struktur, der Abhängigkeit von Ansprüchen sowie enthaltenen Abbildungsreferenzen. Den dritten Bereich bilden Patentsuchanfragen und deren Ergebnismengen. Einen weiteren vierten Bereich bilden die Ergebnisse semantischer Analysen, beispielsweise in Form domänenspezifischer Ontologien, manuelle semantische Annotationen sowie deren Zuordnung zu Patenttexten. Der Fokus dieses Artikels liegt auf der Beschreibung von Visualisierungen für Patentsuchanfragen, deren Ergebnismengen sowie der Einbettung semantischer Annotationen.

2. Systembeschreibung

PatViz ist eine im Rahmen von PatExpert entwickelte graphische Benutzungsschnittstelle, die den Zugriff auf unterschiedliche Patentinformationsquellen in einer einheitlichen Oberfläche integriert. PatViz unterstützt dabei die wesentlichen Schritte innerhalb eines Patentrechercheprozesses (siehe Abbildung 1). Im Rahmen von PatExpert wurden unterschiedliche Suchverfahren auf unterschiedlichen Patentdatenquellen entwickelt, deren Aufteilung und Zusammenführung durch ein so genanntes Merger-Modul verwaltet wird. Die entwickelten Suchverfahren beinhalten klassisches textbasiertes Information Retrieval, sowie erweiterte Verfahren wie Ähnlichkeitssuche auf Basis vorgegebener Texte, Bildähnlichkeitssuche, semantische Suche oder Relevanzbewertung auf Basis von Benutzer Feedback.

¹ Quelle: WIPO Patent Report 2007

² Webseite: <http://www.patexpert.org>

Der erste Rechenschritt umfasst die Formulierung einer Suchanfrage, bzw. die Modifikation einer bestehenden Anfrage, auf Basis einer visuellen Anfragesprache, die im nächsten Abschnitt vorgestellt wird. Die Ergebnisse einer Patentsuchanfrage durchlaufen eine Visualisierungspipeline [7], die als Resultat unterschiedliche Sichten auf eine Ergebnismenge liefert. Auf die Visualisierung und Filterung von Patentergebnismengen wird in Abschnitt 4 näher eingegangen. Ausgehend von einer (in der Regel gefilterten) graphischen Repräsentation einer Ergebnismenge, ermöglicht PatViz die Exploration einzelner Patenttexte in so genannten Detailansichten. In Detailansichten lassen sich Patentinhalte mit Hilfe automatischer semantischer Analyse bzw. mit manuellen semantischen Annotationen graphisch anreichern.

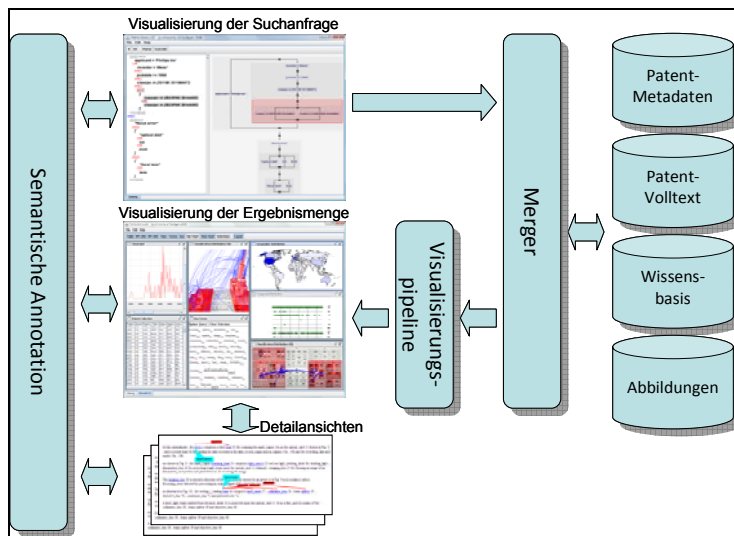


Abbildung 1: PatViz Systemübersicht

3. Visualisierung von Patentsuchanfragen

Die Visualisierung von Patentsuchanfragen in PatViz basiert auf einer eigenen Anfragesprache. Sie unterstützt die Formulierung einer Anfrage in dreierlei Hinsicht:

1. Verdeutlichen der logischen Struktur der Anfrage
2. Unterstützen der Formulierung der elementaren Filterausdrücke
3. Beliebige Kombination unterschiedlicher Suchfunktionen

Die Struktur der Anfrage wird dem Benutzer verdeutlicht indem sie in einer Graphik angezeigt wird, welche sich an klassischen Flussdiagrammen orientiert. Es wird die Metapher eines Flusses gewählt, um Oder-Verknüpfungen als parallele Wege und Und-Verknüpfungen als sequentielle Wege darstellen zu können (siehe Abbildung 2). Dadurch erhält der Benutzer die Möglichkeit die Gültigkeitsbereiche von Verknüpfungen besser zu erfassen und die Positionen, an denen er seine Anfrage verfeinert muss, schneller aufzufinden.

Die Unterstützung für die Formulierung einzelner Filterausdrücke erfolgt über spezielle Editoren. Diese zeigen dem Benutzer je nach Typ des Filters seine möglichen Optionen auf, so dass syntaktische Eingabefehler ausgeschlossen werden können und der Nutzer kognitiv entlastet wird. Eine Besonderheit der PatViz-Suchanfragen ist, dass Filterausdrücke verschiedenartigen Suchfunktionen angehören können. So ist die Verknüpfung einer Bildähnlichkeitssuche mit einer Suche nach speziellen semantischen Konzepten innerhalb derselben Oberfläche möglich.

Zusätzlich zur Graphik wird die Anfrage weiterhin textuell angezeigt, so dass den mit der Anfragesprache vertrauten Benutzern die Möglichkeit zur Direkteingabe vorhanden bleibt. Durch Einrückungen und Farbmarkierungen wird auch hier die Struktur hervorgehoben.

Änderungen sind sowohl in der textuellen als auch in der graphischen Repräsentation möglich und spiegeln sich umgekehrt in beiden Ansichten wieder. Bei der graphischen Repräsentation erfolgen die Änderungen über ein einfaches Kontextmenü und eingebundene Editoren.

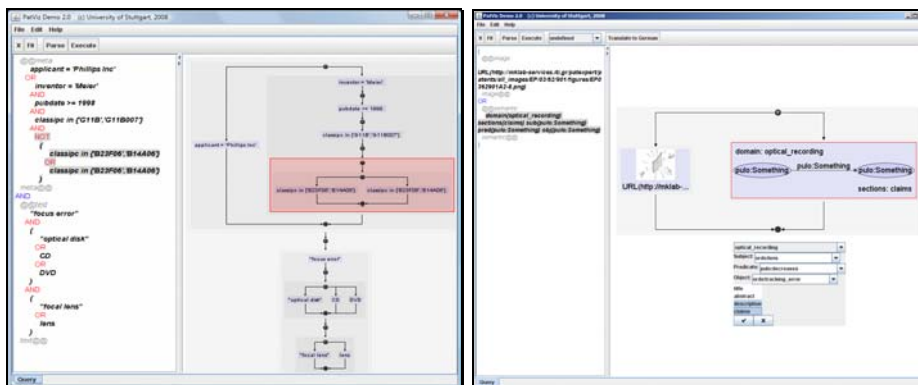


Abbildung 2: PatViz Anfragevisualisierung

4. Visualisierung von Patentergebnismengen

Während innerhalb dieses Artikels eine konzeptuelle Trennung von Suchanfrage und Ergebnisvisualisierung vorgenommen wird, gibt es diese scharfe Trennung bei der Durchführung von Patentrecherchen nicht. Zum einen können abhängig vom Ziel einer Analyse unterschiedliche Visualisierungen der Problemstellung dazu beitragen komplexe Sachverhalte, wie zum Beispiel die Entwicklung eines technologischen Gebiets über die Zeit hinweg, besser zu verdeutlichen. Zum anderen lassen sich geeignete Visualisierungen von (Zwischen-)Ergebnismengen bereits in vorgelagerten Schritten dazu nutzen, Suchanfragen in iterativen Schritten zu verfeinern und zu verbessern.

Im Gegensatz zur rein textbasierten Darstellungen von Ergebnismengen, wie dies von der Websuche her bekannt ist, stehen dem Ersteller der Anfrage mit interaktiven Darstellungen Möglichkeiten zur Verfügung, die eine schnellere und zielgerichtete Einschränkung oder Erweiterung von Suchanfragen ermöglichen. Der Übergang von der Anfrageverfeinerung zur tatsächlichen Analyse erfolgt dabei fließend. Sowohl bei der Anfrageerstellung als auch bei der Analyse von Ergebnismengen spielt insbesondere die Bereitstellung geeigneter Interaktionsmechanismen eine herausragende Rolle. Erst diese ermöglichen die Selektion von dargestellten „Patentobjekten“ in unterschiedlichen Granularitätsstufen und deren anschließende Verwendung bei der Anfragemodifikation oder weiteren Analyseschritten.

Ein weiterer wichtiger Aspekt des im Rahmen von PatExpert entwickelten Ergebnisvisualisierungsmoduls besteht in der Möglichkeit eine Patentmenge in unterschiedlichen Visualisierungen gleichzeitig anzeigen zu können. Diese Nutzung eines „Multiple View“-Konzepts gibt Patentrechercheuren die Kontrolle über die Anzahl und Art der für ihr Analysevorhaben am besten geeigneten Darstellungsformen.

Nachfolgend werden einige Visualisierungen genauer vorgestellt. Weitere Patentvisualisierungen sind in [2] beschrieben. Abbildung 3 zeigt im linken Teil eine so genannte Tag-Cloud [8], welche die relevanten Begriffe innerhalb der aktuellen Ergebnismenge darstellt. Hierbei gilt, dass die Schriftgröße eines Begriffs mit seiner Relevanz³ korreliert. Selektierte Begriffe können nun wiederum genutzt werden um die ursprüngliche Suchanfrage zu verfeinern oder für Analysezwecke alle Patentdokumente hervorzuheben, die den Begriff enthalten.

Im mittleren Teil von Abbildung 3 werden Patentdokumente entsprechend ihres Publikationsdatums sowie nach Ländern gruppiert dargestellt. Zusätzlich lassen sich

³ Die Relevanz kann hierbei über die „term frequency-inverse document frequency (TF-IDF)“ bestimmt werden, um die für eine Unterscheidung der Menge signifikantesten Begriffe auszufiltern.

Informationen zu den enthaltenen Prioritätsbeziehungen anzeigen, indem der Benutzer den Cursor der Maus auf ein Patentknoten in der Visualisierung richtet. Weitere Details lassen sich innerhalb eines eingeblendeten Infobereichs (Tooltip) erkennen. Außer für beliebige Patentmengen eignet sich diese Darstellung besonders gut für die Anzeige von Patentfamilien⁴.

Die Graphik im rechten Teil von Abbildung 3 stellt die Internationale Patent Klassifikation (IPC) hierarchisch in Form einer Treemap [6] dar und zeigt in der Ergebnismenge enthaltene Co-Klassifizierungsbeziehungen als Kantenbündel. Co-Klassifizierungsbeziehungen entstehen auf Grund der Tatsache, dass Patente in der Regel mehreren Klassen zugeordnet werden.

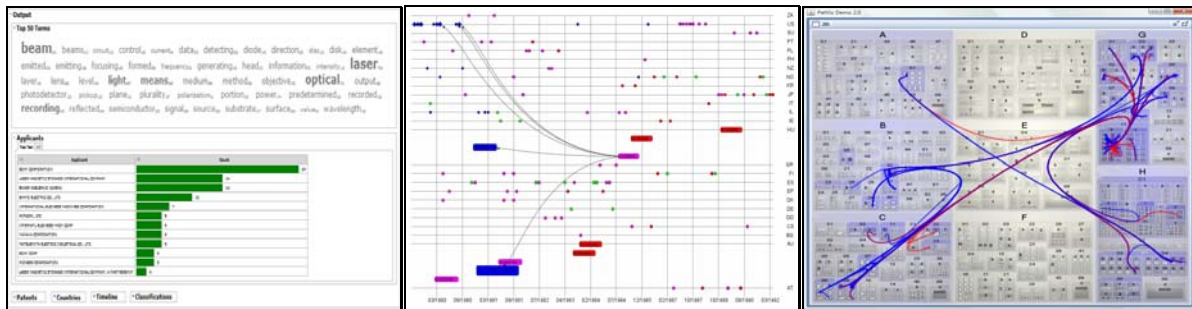


Abbildung 3: Termhäufigkeit (Tag Cloud) und Verteilung der Anmelder (links), zeitliche Bezüge gruppiert nach Land überlagert mit Prioritätsbeziehungen (Mitte), Patentklassifizierung mit gebündelten Co-Klassifizierungsbeziehungen (rechts)

5. Semantische Annotation

Für die semantische Repräsentation von Patenten ist es sinnvoll diese als „Wissensobjekte“ aufzufassen und durch entsprechende Konzepte, Eigenschaften und Beziehungen zu beschreiben. In PatExpert werden hierzu Standards wie RDF (Resource Description Framework) und OWL (Web Ontology Language) eingesetzt. Die Definition von Konzepten, Eigenschaften und Beziehungen ist auf unterschiedliche Ontologiemodule aufgeteilt. Eine genauere Beschreibung des PatExpert Repräsentationsframeworks findet sich in [1,4].

Eine wesentliche Motivation für Verwendung eines semantikbasierten Repräsentationsformalismus ist die Möglichkeit sowohl automatische Extraktions- und Analyseverfahren als auch manuelle Annotationen einheitlich zu behandeln und für unterschiedliche Aufgaben, wie beispielsweise für in die Visualisierung, zu verwenden. Eine detaillierte Beschreibung der in PatExpert entwickelten automatischen Extraktionsverfahren findet sich in [5]. Der weitere Fokus dieses Artikels liegt auf der manuellen Annotation.

Der Ansatz in PatViz zur manuellen Annotation ist die Verwendung eines semantischen Wiki. Vorteil eines wikibasierten Ansatzes ist die Möglichkeit zur kollaborativen Annotation. Die softwaretechnische Umsetzung eines visuellen semantischen Wiki für Patentinformation ist in [3] näher beschrieben. Grundlegende Idee hierbei ist es semantische Annotation entweder auf Basis einer Wiki-Sprache durchzuführen oder mit Hilfe spezieller semantischer Plug-ins graphisch zu integrieren. Anwendungsgebiete sind beispielsweise Patentbewertung, Literaturangaben zum Stand der Technik oder gemeinschaftliches Indexieren in Form einer freien Verschlagwortung (Folksomony).

Abbildung 4 zeigt ein einfaches Beispiel eines Bewertungsmoduls, mit dem die Faktoren Wichtigkeit, Neuheit und Anwendbarkeit bewertet werden können. Bewertungen werden anschließend in der Wissensbasis als RDF-Ausdrücke gespeichert und stehen für Suche, Ranking oder andere Auswertungen zur Verfügung.

⁴ Patentedokumente die über Prioritätsbeziehungen miteinander verknüpft sind bilden sogenannte Patentfamilien.



Abbildung 4: Einbettung eines Bewertungsmoduls

6. Zusammenfassung und Ausblick

Das PatViz-Framework ermöglicht die Visualisierung patentbezogener Suchanfragen und Ergebnismengen. Außerdem unterstützt es die Analyse von Suchergebnissen hinsichtlich deren zeitlicher, geographischer, klassifikatorischer und stichwortbezogener Verteilung. Es baut auf einem im Rahmen von PatExpert entwickelten semantischen Modell auf und integriert die Ergebnisse automatischer semantischer Analysen sowie manueller semantische Annotation.

PatViz liegt in einer prototypischen Implementierung auf Basis des Visualisierungsframeworks prefuse vor [7]. Weitere Ziele sind zunächst die Durchführung von Benutzerstudien, die zum einen die Effektivität und Effizienz der entwickelten Visualisierungen und zum anderen den Aspekt der kollaborativen Annotation von Patenten untersuchen sollen.

Literatur

- [1] L. Wanner, S. Brüggmann, J. Codina, B. Diallo, E. Escorsa, M. Giereth, Y. Kompatsiaris, S. Papadopoulos, E. Pianta, G. Piella, I. Puhmann, G. Rao, M. Rotard, P. Schoester, L. Serafini und V. Zervaki, "Towards Content-Oriented Patent Document Processing", in: World Patent Information Journal, 30(1):21-33, Elsevier, 2008
- [2] M. Giereth, S. Koch, M. Rotard und T. Ertl, "Web Based Visual Exploration of Patent Information", 11th International Conference on Information Visualization, S. 150-155, IEEE Computer Society, 2007
- [3] M. Giereth und T. Ertl, "Visualization Enhanced Semantic Wikis for Patent Information", 12th International Conference on Information Visualization, IEEE Computer Society, 2008
- [4] M. Giereth, S. Koch, Y. Kompatsiaris, S. Papadopoulos, E. Pianta, L. Serafini, L. Wanner und T. Ertl, "A Modular Framework for Ontology-based Representation of Patent Information", in: Legal Knowledge and Information Systems - JURIX 2007, Frontiers in Artificial Intelligence and Applications, Vol. 165, S. 49-58, IOS Press, 2007
- [5] A. Potrich und E. Pianta, "Learning Domain Specific Isa-Relations from the Web", in: 6th International Conference on Language Resources and Evaluation (LREC-2008), 2008
- [6] B. Shneiderman, "Tree Visualization with Tree-Maps: A 2-D Space-Filling Approach", in: ACM Transactions on Graphics (TOG), 11(1): 92 - 99, ACM Press, 1992
- [7] J. Heer, S. Card, und J. Landay, "Prefuse: A Toolkit for Interactive Information Visualization", in: Human Factors in Computing Systems (CHI), 2005
- [8] O. Kaser, D. Lemire, "Tag-Cloud Drawing: Algorithms for Cloud Visualization", Workshop on Tagging and Metadata for Social Information Organization, 2007
- [9] M. Völkel, M. Krötzsch, D. Vrandeic, H. Haller und R. Studer, "Semantic Wikipedia", in: 15th International Conference on World Wide Web (WWW'06), ACM Press, 2006