

Recommender-System für Projektkollaborationen basierend auf wissenschaftlichen Publikationen und Patenten

Christoph Quix ^{1,2}, Sandra Geisler ¹, Rihan Hai ¹

¹ Informatik 5, RWTH Aachen University

² Fraunhofer Institut für Angewandte Informationstechnik FIT

Kurzfassung

Die erfolgreiche Durchführung von Entwicklungs- und Forschungsprojekten hängt von vielen Faktoren ab. Innovationspotential und Zukunftsorientierung helfen bei der Antragsbewilligung. Doch genauso wichtig ist die Zusammensetzung des Projektteams. Insbesondere bei interdisziplinären Projekten ist man auf ein Team angewiesen, das aus hervorragenden Experten der jeweiligen Teilgebiete besteht. Die Medizintechnik ist ein gutes Beispiel für ein sehr innovatives und gleichzeitig hoch interdisziplinäres Feld. Aber gerade die Interdisziplinarität macht die Suche nach Experten schwierig und langwierig, da man sich erst in fremden Domänen zurechtfinden muss und eventuell nicht zum gewünschten Ergebnis kommt. In diesem Beitrag stellen wir unsere Arbeiten an einem Recommender-System für Projektpartner im Rahmen des mi-Mappa- Projekts vor, das basierend auf Informationen aus Patenten, wissenschaftlichen Publikationen und Produktinformationen Experten für ein Projekt innerhalb eines Innovationsfeldes der Medizintechnik empfehlen kann.

Abstract

Successful research and development projects start with finding the right partners for the venture. Especially for interdisciplinary projects, this is a difficult and tedious task as experts from foreign domains are not known. Furthermore, the transfer of knowledge from research into practice is becoming more important in research projects to enable the quick application of research results. This is in particular relevant for projects in medical engineering. Patents and publications contain technical knowledge which can be exploited to find suitable experts. Patents are usually more product-oriented as the inventors have to describe an application area and products might be protected by patents. On the other hand, scientific publications represent the state-of-the-art in research. The challenge is finding the right mixture of research- or application-oriented experts from different domains. Hence, we propose a recommender system for finding experts for a certain topic based on patent topic clustering, ontologies, and ontology matching, which maps patents to corresponding innovation fields. The medical engineering domain serves as a first test bed, since projects in this area are highly interdisciplinary.

1 Einleitung

Bei der Zusammenstellung eines Konsortiums für ein Forschungs- und/oder Entwicklungsprojekt wird oft auf die existierenden Netzwerke und Kontakte und weniger auf objektive Kriterien bzw. Hilfsmittel für die Auswahl von potenziellen Projektpartnern zurückgegriffen. Bei interdisziplinären Projekten, wie sie in der Medizintechnik üblich sind, müssen aber Experten, die außerhalb bestehender Netzwerke zu suchen sind, aus unterschiedlichen Bereichen zusammengeführt werden. Darüber hinaus spielen die innovativen Fähigkeiten der Projektpartner eine zunehmend größere Rolle, da die Produktentwicklungszyklen immer kürzer werden. Daher ist die Bereitschaft und Fähigkeit der potenziellen Partner zur schnellen Umsetzung der Projektergebnisse in ein Produkt ein wichtiges Kriterium für die Auswahl von Projektpartnern in der Medizintechnik.

Studien haben gezeigt, dass die Kooperation zwischen Forschungseinrichtungen und Unternehmen sowohl Produkt- als auch Prozessinnovation begünstigen können (Robin und Schubert, 2013). Gerade in interdisziplinären Projekten kann die Suche nach geeigneten Partnern sehr zeitaufwändig, komplex und am Ende vielleicht nicht so erfolgreich wie erwartet sein. Daher wäre eine Unterstützung der Suche in Form eines Recommender-Systems für Projektpartner wünschenswert. Ein solches System könnte den Suchprozess erheblich beschleunigen und neue Kooperationsmöglichkeiten eröffnen.

Für ein solches Recommender-System können Patente als sehr wichtige Informationsquelle dienen, da sie sehr viele technische Informationen zur Entwicklung eines Produkts beinhalten. Allerdings ist ihre Analyse schwierig, da die Sprache (Begriffe und Formulierungen) von anderen wissenschaftlichen Texten stark abweichen (Aras et al., 2014; Zhang et al., 2015). Da Autoren von Patenten (Erfinder) nicht nur Experten in ihrem Gebiet sind, sondern auch eine produktorientierte Sicht auf ihr Forschungsfeld haben, stellen Erfinder gerade für innovative Forschungsbereiche wie die Medizintechnik interessante potentielle Projektpartner dar. In diesem Beitrag stellen wir einen Ansatz vor, der Techniken aus den Bereichen Patentanalyse und -Mining, Ontologie-Mappings und -Matching (Shvaiko und Euzenat, 2013) kombiniert und in einem System zur Empfehlung von Kooperationspartnern implementiert wird.

Im Projekt mi-Mappa¹ wird derzeit an geeigneten Methoden zur Suche von Projektpartnern geforscht. Aufgrund des Projektes konzentrieren wir uns auf den Anwendungsbereich Medizintechnik, jedoch sollten die erforschten Methoden letztlich auch auf andere Bereiche übertragbar sein. Eine Voraussetzung für die Anwendung unseres Ansatzes in einem bestimmten Forschungsbereich ist allerdings die Identifikation von Innovationsfeldern, welche bereits in der Medizintechnik stattgefunden hat (Schlötterburg et al., 2008). Nach Schlötterburg et al. (2008) ist ein Innovationsfeld in der Medizintechnik ein Bereich, in dem es signifikante Innovationsaktivitäten, Zukunftspotenzial und eine möglichst vollständige Wertschöpfungskette gibt. Die Hauptinnovationsfelder der Medizintechnik umfassen derzeit (Schlötterburg et al., 2008; Deutsche Gesellschaft für Biomed. Technik im VDE, 2012)

- Bildgebende Verfahren,
- Prothesen und Implantate,
- Telemedizin und modellbasierte Therapien,
- Operative und interventionelle Geräte und Systeme,
- In-Vitro-Diagnostik,
- Regenerative Medizin

und auch Querschnittsthemen wie z.B. Patientensicherheit oder Gebrauchstauglichkeit.

Unser Ansatz kombiniert zwei komplementäre Wege. Einerseits erstellen wir für die Erfinder ein Profil auf Basis ihrer Patente, wissenschaftlichen Publikationen, Webseiten und anderen forschungsrelevanten Informationen. Dies beinhaltet die Identifikation eines Erfinders mit einem Autor einer Publikation. Wenn ein entsprechender Autor gefunden werden kann, können die Publikationen den Innovationsfeldern zugeordnet und somit das Spezialgebiet des Erfinders bzw. Autors bestimmt werden. Zur Abbildung von Publikationen auf Innovationsfelder nutzen wir Klassifikationsschemata von Publikationsdatenbanken und semiautomatische Ontology-Matching-Verfahren. Andererseits, sollte der Erfinder nicht als Autor wissenschaftlicher Publikationen identifiziert werden können, gruppieren wir die Patente eines Erfinders nach Themen und bilden diese Themen auf die

¹ <http://www.dbis.rwth-aachen.de/mi-Mappa>

Innovationsfelder ab. Auch für diese Verknüpfung zwischen Themen und Innovationsfeldern nutzen wir Ontology- Matching-Verfahren.

Im folgenden Abschnitt diskutieren wir zunächst verwandte Arbeiten im Bereich Patentanalyse und Recommender-Systeme für Forschungspartner. Unser Ansatz wird in Abschnitt 3 beschrieben. In Abschnitt 4 geben wir eine Zusammenfassung und einen Ausblick auf zukünftige Arbeiten.

2 Verwandte Arbeiten

Recommender-Systeme zur Suche von Experten bzw. Kooperationspartnern

Die Suche von Kooperationspartnern beinhaltet üblicherweise manuelle Schritte. Zum Beispiel müssen festgelegte Kriterien analysiert, bewertet und gewichtet werden (Geum et al., 2013; Awasthi et al., 2015). Systeme zur Suche von Experten für ein bestimmtes Thema basieren häufig auf Informationen, die von den Experten selbst veröffentlicht und verwaltet werden, auf vom Experten veröffentlichten Dokumenten oder auf Informationen aus sozialen Netzwerken (Wang et al., 2013). Diese Systeme erstellen entweder Expertenprofile (Expert Profiling) oder unterstützen den Suchprozess (Expert Finding) (Balog und De Rijke, 2007). Beispiele aus dem wissenschaftlichen Bereich sind ResearchGate², Google Scholar³ oder AMiner⁴. Aktuelle Arbeiten nutzen Algorithmen für die Analyse von sozialen Netzwerken (z.B. PageRank oder HITS (Rafiei und Kardan, 2015; Wang et al., 2013)) und graphbasierte Algorithmen (Rani et al., 2015). Wir konzentrieren uns in dieser Arbeit auf Systeme zur automatischen Expertensuche und -profiling auf Basis von veröffentlichten Dokumenten (z.B. Patenten und Publikationen). Viele der bestehenden dokumentbasierten Experten- Recommender-Systeme (ERS) beschränken ihren Suchbereich auf Dokumente und Personen innerhalb eines Unternehmens. Im Gegensatz dazu schlagen wir ein System vor, das Informationen aus beliebigen Datenquellen integrieren kann (Hai et al., 2016). Collaboration Spotting⁵ ist ein Projekt des CERN-Instituts, das basierend auf Patenten und Publikationen die Analyse eines wissenschaftlichen Feldes zu ermöglichen.

² <https://www.researchgate.net>

³ <https://scholar.google.com>

⁴ <https://aminer.org>

⁵ <http://collspotting.web.cern.ch>

Das Projekt verbindet sowohl die Analyse von Dokumenten als auch die soziale Netzwerkanalyse Skogstad et al. (2013). Es kann anhand von Begriffen gesucht werden und man kann sich verschiedene Netzwerke, z.B. von Unternehmen in einem Feld, visualisieren lassen. Ein Profiling der Experten wird jedoch nicht durchgeführt. Der DEMOIR-Ansatz (Yimam-Seid und Kobsa, 2003) nutzt ebenso wie unser Ansatz Ontologien, jedoch werden damit lediglich die Expertise der Experten und die Anwendungsdomäne modelliert. Ein 'Matching' der Ontologien findet nicht statt.

Patentanalyse mit Ontologien Die Nützlichkeit von Ontologien als formalisierte Domänenmodelle wurde auch für die Patentanalyse erkannt, insbesondere für die Patentsuche (Bonino et al., 2010). Beispielsweise nutzt das System PatExpert ein Netzwerk aus Ontologien und Wissensdatenbanken um Patentsuche, -klassifikation und -Clustering zu unterstützen. Trappey et al. (2009) stellen ein System vor, das Textfragmente mit einem Konzept in einer Ontologie verknüpft. Die Ähnlichkeit von zwei Patenten hängt dann von der Anzahl der Konzepte ab, die beide Patenten gemeinsam haben. Dieser Ansatz ist aber nur auf die Konzepte beschränkt, die innerhalb der Ontologie vorkommen.

Patent-Clustering Eine (graphische) Übersicht einer Menge von Patenten kann mithilfe von Clustering-Techniken realisiert werden (Tseng et al., 2007). Den berechneten Clustern kann ein Titel zugewiesen werden, der sich aus den häufigsten Begriffen der Patente innerhalb des Clusters zusammensetzt. Moguee und Kolar (1999) präsentieren einen bibliometrischen Ansatz, in dem Ko-Zitationsanalyse verwendet wird, um ko-zitierte Dokumente zu verknüpfen. Hierbei wird angenommen, dass diese Dokumente das gleiche Thema behandeln. Allerdings kann dieser Ansatz zu oberflächlichen Ergebnissen führen, da Details der Patente nicht berücksichtigt werden. Trappey et al. (2010) beschreiben eine Methodik für das Patent-Clustering, die zunächst die wichtigsten Phrasen aus einem Patent extrahiert (mithilfe von ontologiebasierten statistischen Methoden). Anschließend werden diese Phrasen zu Technologie-Clustern ('Technology Clusters') zusammengefasst und schließlich werden mit der gleichen Methode aus den Patentdokumenten Cluster gebildet.

Zusammenfassend kann man sagen, dass zwar einzelne Aspekte unseres Ansatzes bereits betrachtet wurden, die Kombination von verschiedenen Techniken zur Patent-Analyse, Clustering, Ontologie-Entwicklung und -Matching aber einen innovativen wissenschaftlichen Beitrag leistet.

3 Lösungsansatz

In der Patentanalyse werden quantitative und qualitative Ansätze unterschieden (Hong, 2004). Qualitative Ansätze extrahieren Metadaten aus den Patenten (z.B. Erfinder, Literaturreferenzen, Anmelder), während quantitative Ansätze die Häufigkeit von Begriffen im Patenttext zählen. Das Ziel unseres Ansatzes, dargestellt in Abb. 1, ist die Verknüpfung von Erfindern und ihren Patenten zu Innovationsfeldern. Wie bereits in der Einleitung dargestellt wurde, verfolgen wir dafür zwei parallele Ansätze:

- Verknüpfung der Erfinder mit Autoren von wissenschaftlichen Publikationen oder
- Clustering der Patente und Zuordnung eines oder mehrere Themen zu einem Cluster (Topic Clustering).

In beiden Ansätzen werden nur die Patente betrachtet, die das Ergebnis einer initialen Stichwortsuche sind. Die Stichworte sollten aus der Beschreibung des geplanten Projekts extrahiert werden. Die Ergebnisse beider Ansätze (Publikationen bzw. Themen) werden dann Konzepten aus existierenden medizinischen Ontologien zugeordnet. Die medizinischen Ontologien sind umfangreiche Beschreibungen der Domäne und können daher auch für die Charakterisierung der Innovationsfelder der Medizintechnik genutzt werden, d.h. den Innovationsfeldern werden also Konzepte der Ontologien zugeordnet. Schließlich können durch die Verkettung der Ergebnisse Erfinder Innovationsfeldern zugeordnet werden.

In den folgenden beiden Abschnitten werden wir die beiden Ansätze detailliert beschreiben. Im Anschluss daran werden wir auf den Entwurf der Ontologien und die Matching- Techniken eingehen.

3.1 Verknüpfung von Erfindern mit Autoren

Grundlage für die Verknüpfung ist die Annahme, dass Erfinder auch als Forscher tätig sind und wissenschaftliche Publikationen zum gleichen Thema publizieren. Zu den Publikationen existieren häufig Klassifikationsterme oder andere manuell festgelegte Schlagworte, die sich einfacher mit Ontologien verknüpfen lassen als die eher abstrakten Klassifikations terme der 'International Patent Classification' (IPC).

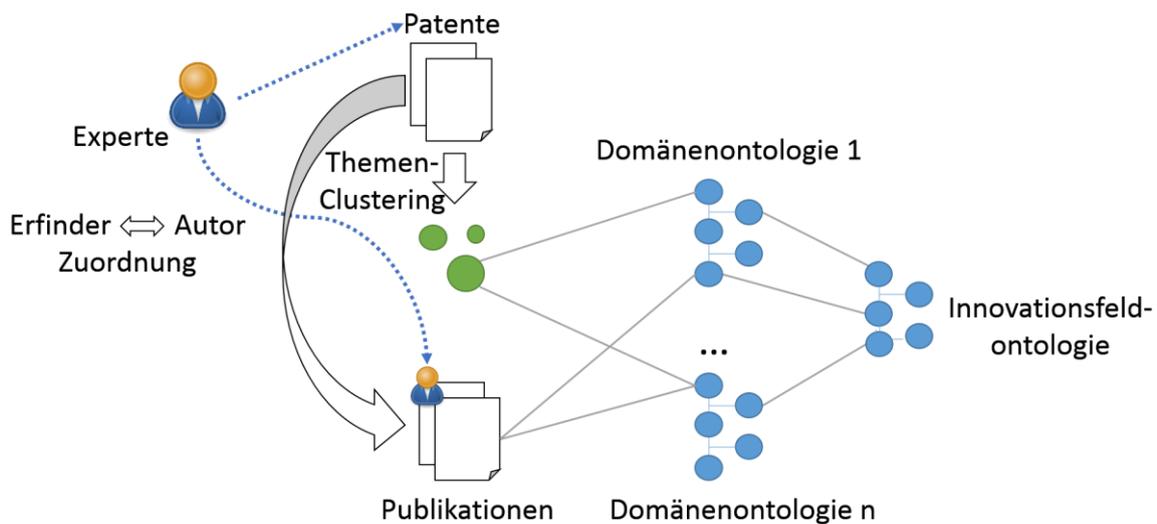


Abbildung 1: Architektur zur Abbildung von Patenten auf Innovationsfelder

In unserem ersten Prototyp haben wir Publikationsdatenbanken wie 'Web of Science' oder PubMed eingebunden (Hai et al., 2016), die Artikel anhand weitverbreiteter medizinischer Taxonomien klassifizieren, z.B. MeSH (Medical Subject Headings). Aus den Datenbanken werden alle Publikationen eines Autors abgefragt (ein Erfinder der im initialen Schritt gefundenen Patente). Da keine eindeutigen Identifikatoren für Erfinder bzw. Autoren zur Verfügung stehen, sondern nur Namen als Zeichenketten, ist ein mehrstufiger Prozess zur Identifizierung von Erfindern und Autoren notwendig. Hierfür extrahieren wir diverse Metadaten der Erfinder eines Patents (neben Name z.B. Firmenname, Adresse) und nutzen diese Informationen zusammen mit den inhaltlichen Informationen eines Patents zum Clustering der Erfinder. Im Idealfall sollten alle Patente eines Erfinders zu einem Cluster zugeordnet werden können, der dann eindeutig auf einen Autor von wissenschaftlichen Publikationen abgebildet werden kann.

3.2 Topic Clustering

In verschiedenen Studien wurde belegt, dass 70-90% des technischen Wissens nur in Patenten veröffentlicht werden (The Thompson Corporation, 2007; Stiftung Universität Hildesheim, 2014). Daher ist es wahrscheinlich, dass Erfinder nur Patente publizieren und bisher nicht als Autoren wissenschaftlicher Publikationen in Erscheinung getreten sind. Aus diesem Grund müssen wir einen alternativen Weg für die Zuordnung von Erfindern zu Innovationsfeldern zur Verfügung stellen. Für diesen zweiten Weg unseres Ansatzes nutzen wir Topic Clustering, das gleichzeitig auch für die Verifikation der ersten Lösungsmöglichkeit genutzt werden kann.

Das IPC-Klassifikationsschema für Patente ist zwar umfassend, aber für eine gute Patentanalyse nicht detailliert genug (Tseng et al., 2007). Genauere Kategorisierungen der Patente, wie sie mit Topic Clustering erstellt werden können, sind wünschenswert. Clustering gruppiert ähnliche Objekte miteinander, wobei die Ähnlichkeit durch die Distanz zwischen den Objekten in einem n-dimensionalen Vektorraum bestimmt wird.

Für die Merkmalsextraktion aus Textdokumenten nutzen wir die üblichen Vorverarbeitungsmethoden, wie Stoppwortentfernung, Wortstammbildung und Tokenisierung (Gonçalves et al., 2010). Dabei werden nur die wichtigsten Begriffe betrachtet, die durch die Berechnung der 'Inverse Document Frequency' (IDF) und 'Term Frequency' (TF) ermittelt werden. Den erkannten Clustern wird ein Titel bestehend aus den 3-5 wichtigsten Begriffen zugeordnet. Da die Merkmalsextraktion aus Textdokumenten aufwändig ist, betrachten wir nur den ersten Teil der Textabschnitte. Dies verbessert nicht nur die Verarbeitungsgeschwindigkeit sondern auch die Ergebnisqualität (Fall et al., 2003).

3.3 Auswahl, Entwurf und Matching der Ontologien

Auswahl der Ontologien: Unser Ansatz basiert auf der Verfügbarkeit von medizinischen Ontologien und deren Verknüpfung mit der Ontologie für Innovationsfelder. Da eine Vielzahl von Ontologien im biomedizinischen Bereich existieren, müssen zunächst eine oder mehrere Ontologien ausgewählt werden, die unsere Domäne am besten abdecken.

Dazu haben wir zunächst über Bioportal⁶ (eine Suchmaschine für biomedizinische Ontologien) nach passenden Ontologien gesucht. Des Weiteren nutzten wir den 'Ontology Lookup Service'⁷ und Ontobee⁸, um einen umfassenden Überblick über relevante Ontologien zu bekommen. Für die Suche extrahierten wir 174 Begriffe aus den Beschreibungen der sechs Innovationsfelder (Schlötelburg et al., 2008; Deutsche Gesellschaft für Biomed. Technik im VDE, 2012). Die relevantesten Ontologien waren.

- National Cancer Institute Thesaurus (NCIT),
- Systematized Nomenclature of Medicine - Clinical Terms (SNOMEDCT),
- Medical Subject Headings (MeSH) und
- die 'Robert Hoehndorf Version of MeSH' (RHMeSH).

Für diese Ontologien analysierten wir die Abdeckung der Suchbegriffe durch einzelne Ontologien (vgl. Abb. 2). Die Y-Achse stellt die Prozentzahl der in der Ontologie vorkommenden Suchbegriffe dar.

Offensichtlich ist keine Ontologie in allen Bereichen besser als die anderen Ontologien. Generell ist die Abdeckung auch sehr niedrig. Daher versuchten wir die Abdeckungsrate zu verbessern, indem mehrere Ontologien betrachtet werden (vgl. Abb. 3). Durch die Verwendung von allen vier Ontologien kann die Abdeckungsrate im Schnitt um etwa 10 Prozentpunkte verbessert werden. Auch das 'Ontology Recommender Tool' des Bioportal kommt zu einem ähnlichen Ergebnis. Zusätzlich werden wir die gefundenen Ontologien noch mit Hilfe von Qualitätskriterien für Ontologien bewerten (Vrandečić, 2009; Gomez-Perez, 2004).

⁶ <http://bioportal.bioontology.org>

⁷ <http://www.ebi.ac.uk/ontology-lookup>

⁸ <http://www.ontobee.org>

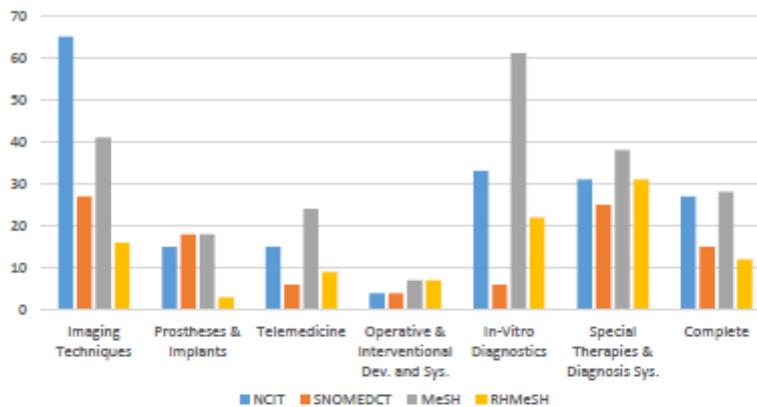


Abbildung 2: Abdeckung der Suchbegriffe durch einzelne Ontologien

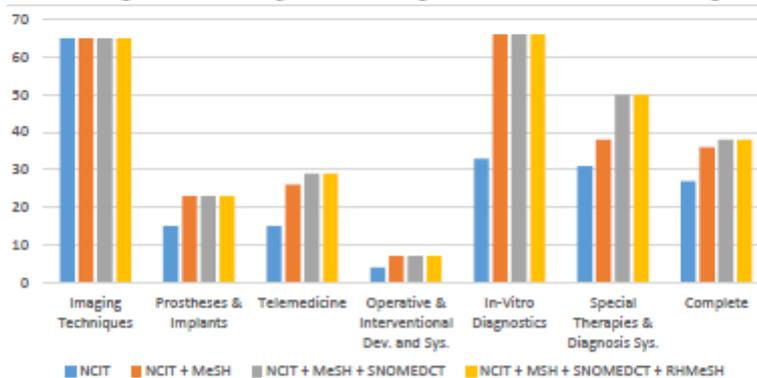


Abbildung 3: Abdeckung der Suchbegriffe durch Kombinationen von Ontologien

Entwurf der Ontologie für Innovationsfelder Um Publikationen oder Patent-Cluster mit Innovationsfeldern zu verknüpfen, ist eine Modellierung der Innovationsfelder durch Ontologien erforderlich. In dieser Ontologie sollten nicht nur die Innovationsfelder beschrieben sein, sondern auch wichtige Konzepte, die die wesentlichen Elemente eines Innovationsfelds darstellen. Dazu wurden Interviews mit Domänenexperten durchgeführt, existierende Ontologien analysiert und Fachliteratur zu den Innovationsfeldern untersucht. Der Entwurfsprozess orientiert sich an der NeOn-Methodik (Suárez-Figueroa, 2010). Die in der Ontologiesuche verwendeten 174 Begriffe stellen auch den Startpunkt für die Ontologie-modellierung dar. Die erste Version der Innovationsfeldontologie wurde durch Interviews mit Domänenexperten evaluiert und wird nun kontinuierlich auf Basis der Evaluierungsergebnisse weiterentwickelt.

Matching der Ontologien Um Erfinder oder Autoren Innovationsfeldern zuzuordnen, müssen die bisher gewonnenen Informationen miteinander verknüpft werden. Die Titel der Patent-Cluster und die Themengebiete der Publikationen müssen auf die Innovationsfeldontologie abgebildet werden. Die Verknüpfung der existierenden medizinischen Ontologien (z.B. NCIT) mit unserer Ontologie für Innovationsfelder wird manuell durchgeführt, da schon bei der Modellierung auf eine Verknüpfung mit existierenden Ontologien geachtet wurde. Da die Themenbereiche der gefundenen Publikationen bzw. Patente von der anfangs eingegebenen Beschreibung abhängen, können für diesen Fall die Verknüpfungen nicht vorab definiert werden. Hierzu wollen wir Methoden aus dem Ontology-Matching (Shvaiko und Euzenat, 2012) einsetzen, die z.B. in unserem GeRoMeSuite-System (Kensche et al., 2007) implementiert wurden. BioPortal bietet darüber hinaus auch bereits Verknüpfungen zwischen verschiedenen Ontologien an, die für diesen Fall genutzt werden können. Die gefundenen Verknüpfungen könnten von einem Experten verifiziert werden. Da aber in einem Recommender-System Fehler toleriert werden können, ist hier kein perfektes Ergebnis erforderlich. Daher könnte der Ontology-Matching-Schritt automatisch durchgeführt werden.

4 Fazit

In diesem Beitrag haben wir einen innovativen ontologiebasierten Ansatz vorgestellt, der die Expertensuche für Forschungsprojekte in der Medizintechnik unterstützt. In unserem Ansatz verwenden wir sehr viele Methoden aus dem Ontology-Engineering, z.B. Analyse, Erstellung und Matching von Ontologien, Definition von Anforderungen für Ontologien und schließlich Evaluierung der Ontologien. Wir verwenden zusätzlich aber auch Technologien aus anderen Bereichen, wie Text Mining und Patentanalyse. Derzeit konzentriert sich unsere Arbeit auf die Datenintegration, Modellierung und Auswahl der Ontologien und der Evaluierung der Clustering-Methoden. Für die bisher verwendete Clustering-Methode arbeiten wir derzeit an Verbesserungen der Performanz, da diese bei komplexen Text-Clustering-Aufgaben eine Herausforderung darstellt und wir ein interaktives System mit geringen Antwortzeiten benötigen.

Die Arbeiten sind derzeit noch nicht abgeschlossen, aber wir denken, dass die Integration von Text Clustering, Topic Modeling, Patentanalyse und Ontologie-Matching gut funktionieren wird. Die verschiedenen Methoden stellen an sich gesehen schon große Herausforderungen dar, aber deren Kombination könnte eine innovative Möglichkeit sein, sich unbekannte Forschungsfelder zu erschließen. Unser Ansatz ist nicht auf die Medizintechnik beschränkt - er kann auch auf andere Forschungsgebiete angewandt werden. Da für die Medizin schon viele Ontologien und Taxonomien verfügbar sind, ist dieses Gebiet für unseren ontologiebasierten Ansatz natürlich ideal.

Literatur

Aras, H., Hackl-Sommer, R., Schwantner, M., und Sofean, M. (2014). Applications and challenges of text mining with patents. In Proc. of 1st Intl. Workshop on Patent Mining and its Applications. Stiftung Universität Hildesheim.

Awasthi, A., Adetiloye, T., und Crainic, T. G. (2015). Collaboration partner selection for city logistics planning under municipal freight regulations. Applied Mathematical Modelling.

Balog, K. und De Rijke, M. (2007). Determining expert profiles (with an application to expert finding). In IJCAI, volume 7, pages 2657–2662.

Bonino, D., Ciaramella, A., und Corno, F. (2010). Review of the state-of-the-art in patent information and forthcoming evolutions in intelligent patent informatics. World Patent Information, 32(1):30–38.

Deutsche Gesellschaft für Biomed. Technik im VDE (2012). Empfehlungen zur Verbesserung der Innovationsrahmenbedingungen für Hochtechnologie-Medizin. Technical report, VDE.

Fall, C. J., Töröcsvári, A., Benzineb, K., und Karetka, G. (2003). Automated categorization in the international patent classification. In ACM SIGIR Forum, volume 37, pages 10–25. ACM.

Geum, Y., Lee, S., Yoon, B., und Park, Y. (2013). Identifying and evaluating strategic partners for collaborative r&d: Index-based approach using patents and publications. Technovation, 33(6):211–224.

Gomez-Perez, A. (2004). Ontology evaluation. In Staab, S. und Studer, R., editors, Handbook on Ontologies, Intl. Handbooks on Information Systems, pages 250–273. Springer.

Gonçalves, C. A., Gonçalves, C. T., Camacho, R., und Oliveira, E. C. (2010). The impact of pre-processing on the classification of medline documents. In PRIS, pages 53–61.

Hai, R., Geisler, S., und Quix, C. (2016). Constance: An intelligent data lake system. In Proc. SIGMOD. ACM. to appear.

Hong, S. (2004). The magic of patent information. World Intellectual Property Organization. http://www.wipo.int/sme/en/documents/patent_information_fulltext.html.

Kensche, D., Quix, C., Li, X., und Li, Y. (2007). GeRoMeSuite: A system for holistic generic model management. In Proc. VLDB, pages 1322–1325.

- Mogee, M. E. und Kolar, R. G. (1999). Patent co-citation analysis of eli lilly & co. patents. *Expert Opinion on Therapeutic Patents*, 9(3):291–305.
- Rafiei, M. und Kardan, A. A. (2015). A novel method for expert finding in online communities based on concept map and pagerank. *Human-centric Computing and Information Sciences*, 5(1):1–18.
- Rani, S. K., Raju, K., und Kumari, V. V. (2015). Expert finding system using latent effort ranking in academic social networks. *Intl. J. of Information Technology and Computer Science*, 2:21–27.
- Robin, S. und Schubert, T. (2013). Cooperation with public research institutions and success in innovation: Evidence from france and germany. *Research Policy*, 42(1):149–166.
- Schlötelburg, C., Weiß, C., Hahn, P., Becks, T., und Mühlbacher, A. C. (2008). Identifizierung von Innovationshürden in der Medizintechnik. Technical report, Bundesministeriums für Bildung und Forschung.
- Shvaiko, P. und Euzenat, J. (2013). Ontology matching: State of the art and future challenges. *IEEE Transactions on Knowledge and Data Engineering*, vol. 25, no. 1, pp. 158-176.
- Skogstad, E., Le Goff, J.-M., Fragkiskos, S., und Agocs, A. (2013). Visualizing collaborations and technology landscapes with interactive sociograms. In *Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC), 2013 IEEE*, pages 1–6. IEEE.
- Stiftung Universität Hildesheim (2014). Womit beschäftigen sich Erfinder? <http://www.uni-hildesheim.de/en/fb3/institute/iwist>.
- Suárez-Figueroa, M. C. (2010). NeOn Methodology for building ontology networks: specification, scheduling and reuse. PhD thesis, Universidad Politecnica de Madrid.
- The Thompson Corporation (2007). Global patent sources - an overview of international patents.
- Trappey, A. J., Trappey, C. V., Hsu, F.-C., und Hsiao, D. W. (2009). A fuzzy ontological knowledge document clustering methodology. *IEEE Trans. on Systems, Man, and Cybernetics, Part B*, 39(3):806–814.
- Trappey, C. V., Trappey, A. J., und Wu, C.-Y. (2010). Clustering patents using non-exhaustive overlaps. *System Science and System Engineering*, 19(2):162–181.
- Tseng, Y.-H., Lin, C.-J., und Lin, Y.-I. (2007). Text mining techniques for patent analysis. *Information Processing & Management*, 43(5):1216–1247.
- Vrandečić, D. (2009). Ontology evaluation. In Staab, S. und Studer, R., editors, *Handbook on Ontologies*, chapter 13, pages 293–313. Springer.
- Wang, G. A., Jiao, J., Abrahams, A. S., Fan, W., und Zhang, Z. (2013). Expertrank: A topic-aware expert finding algorithm for online knowledge communities. *Decision Support Systems*, 54(3):1442–1451.
- Yimam-Seid, D. und Kobsa, A. (2003). Expert-finding systems for organizations: Problem and domain analysis and the demoir approach. *J. of Organizational Computing and Electronic Commerce*, 13(1):1–24.
- Zhang, L., Li, L., und Li, T. (2015). Patent mining: A survey. *ACM SIGKDD Explorations Newsletter*, 16(2):1–19.