

Exploration wissenschaftlicher Netzwerke und Publikationen

Wolfgang Reinhardt, Muneeb I. Ahmad, Pranav Kadam, Ksenia Khadrachieva, Jan Petertonkoker, Amit Shrestha, Pragati Sureka, Junaid Surve, Kaleem Ullah, Tobias Varlemann, Vitali Voth

Institut für Informatik, Universität Paderborn

Zusammenfassung

In diesem Artikel beschreiben wir Ansätze für eine kollaborative Multitouch-Anwendung, die es Forschern ermöglicht wissenschaftliche Netzwerke zu explorieren, denen sie angehören oder an deren Themen sie interessiert sind. Wir beschreiben vier verschiedene Zugänge zu vorhandenen Informationen und unterschiedliche Nutzungsszenarien für die PUSHPIN_{MT} Anwendung.

1 Einleitung

Die Anwendung von Web 2.0 Techniken, Methoden und Werkzeugen führt nicht nur dazu, dass sich die Art, wie heutzutage Forschung stattfindet geändert hat. Vielmehr führt Forschung 2.0 auch dazu, dass mehr potentiell relevante Publikationen entstehen, über die ein Forscher gewahr sein sollte. Renear und Palmer (2009) stellen dabei fest, dass heutige Forscher etwa 50% mehr Publikationen lesen als ihre Kollegen in den 1970er Jahren, dabei allerdings weniger Zeit mit jeder Publikation verbringen. Die enorme Zahl publizierter wissenschaftlicher Arbeiten führt dabei oft auch zur Überlastung von Begutachtungskommissionen in Konferenzen, Journalen und Verlagen sowie zu Awareness-Problemen der einzelnen Forscher (Priem und Hemminger 2010, Reinhardt 2012). Thematische Zusammenhänge und ähnliche Forschungsschwerpunkte sind auf Grund der Datenmengen kaum noch manuell zu erfassen und oft wissen Forscher nichts über relevante Publikationen oder Kollegen, die an ähnlichen Fragestellungen forschen. Dies führt oft zu fragmentierten Forschungsgemeinschaften, welche durch Awareness-Unterstützung in Form von Empfehlungssystemen, Visualisierungen und explorativ-scientrometrischen Informationssystemen profitieren können. Hier setzt das PUSHPIN (*Supporting Scholarly Awareness in Publications and Social Networks*) Projekt an und hat die direkte Awareness-Unterstützung für Forscher durch aktive Soziale Netzwerke und automatische Publikationsanalysen zum Ziel. Forscher sollen PUSHPIN nutzen, um Zusammenhänge und Ähnlichkeiten zwischen Publikationen zu erkennen und ihre Verbindungen zu anderen Forschern besser verstehen zu können. Neben

einem webbasierten und einem mobilen Zugang zu PUSHPIN existiert auch die Möglichkeit zur Exploration der Daten in einer Multitouch-Anwendung, welche wir in diesem Artikel präsentieren.

2 Datenquellen, -analyse und -abfrage

Scientometrie ist die Wissenschaft der Analyse wissenschaftlicher Publikationen, Kooperationen und Wissenschaftskommunikation (Weingart und Winterhager 1984). Für unsere Arbeit sind vor allem die Zitations- und Publikationsanalyse sowie die Analyse von Kooperationsnetzwerken relevant. Dazu müssen wissenschaftliche Publikationen automatisiert analysiert werden, um Metadaten wie Titel, Autoren, Schlüsselworte, Zitationsdaten etc. zu extrahieren. Im Rahmen des PUSHPIN Projekts setzen wir dazu die Bibliotheken ParsCit und GROBID ein, welche recht exakte Daten aus verschiedenformatigen Eingabedaten liefern. Die extrahierten Daten werden persistent in einer Datenbank abgelegt und sind so für die weitere Analyse zugreifbar. Besonders herausfordernd ist in diesem Zusammenhang die eindeutige Identifikation von Autoren, da insbesondere in den Referenzlisten Autorennamen nur abgekürzt genannt werden. Weiterhin ist die Qualität der automatisch extrahierten Daten mittels der o.g. Werkzeuge oft nur gering, so dass die extrahierten Daten miteinander kombiniert werden, um die gesamte Datenqualität zu erhöhen.

2.1 Datenanalyse

Die extrahierten Rohdaten werden in anschließenden Analyseschritten miteinander verrechnet, um spezifischere Aussagen treffen zu können. So werden beispielsweise Koautorenschaften extrahiert und separat gespeichert. Auf Grund der Zitationsdaten werden Werte für bibliografische Kopplung (Kessler 1963) und Kozitation bestimmt und gespeichert. Darüber hinaus werden die Inhalte der jeweiligen Publikationen analysiert, das Dokument beschreibende Schlüsselworte und Eigennamen extrahiert und separat gespeichert. Benutzer von PUSHPIN können jedoch auch manuell Tags zur Klassifizierung der vorhandenen Personen und Publikationen hinzufügen. Schließlich wird im Rahmen der automatischen Datenanalyse auch eine Autornamen-Disambiguierung durchgeführt, um möglichst sicherzustellen, dass identifizierte Autorennamen in der Datenbank eindeutig einer Person zugeordnet werden (vgl. Smalheiser und Torvik 2009).

2.2 Datenzugriff

Der Zugriff auf die analysierten Daten erfolgt über eine gesicherte REST-API, die Autorisierung erfordert. Hier kann man verschiedene Datensätze anfragen, die dann in JSON Repräsentation zurückgeliefert werden.

3 Exploration von Wissenschaftsnetzwerken

Prinzipiell unterscheiden wir vier mögliche Zugänge zur scientometrischen Exploration wissenschaftlicher Netzwerke: 1) den Themenzentrierten, 2) den Personenzentrierten, 3) den Eventzentrierten und 4) den Trendzentrierten Zugang.

Im *Themenbasierten Zugang* zu Wissenschaftsnetzwerken erfolgt die Exploration der analysierten Daten ausgehend von den Inhalten der Publikationen. Der Benutzer kann dazu eine Freitextsuche verwenden oder Klassifikationen durchsuchen. Die Ergebnisse werden in Listen und Netzwerken angezeigt, wobei die Kanten in den Netzwerkdarstellungen Informationen über die Verbindung zwischen zwei Publikationen beinhalten.

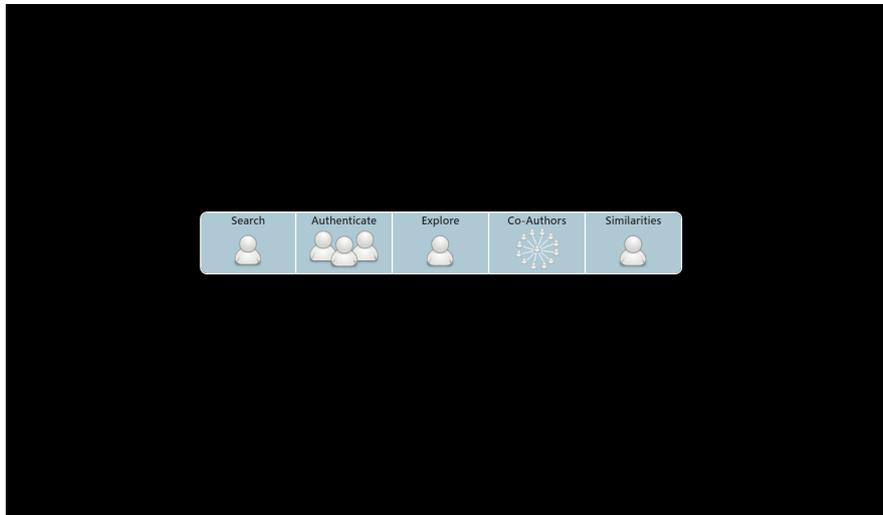
Beim *Personenbasierten Zugang* erfolgt die Exploration ausgehend von Autorenprofilen. Auch hier kann eine Suche gestartet werden und ausgehend von Autorenprofilen können die Ko-Autorennetzwerke exploriert werden. In einem solchen Netzwerk repräsentieren Kanten das gemeinsame Schreiben einer Publikation (vgl. Abbildung 1b).

Der *Eventbasierte Zugang* bietet sich vor allem für die Exploration einer Konferenz oder Konferenzserie an. Dazu wird die Multitouch-Anwendung in einem öffentlichen Raum bei einer Konferenz zugänglich gemacht und den Teilnehmern der Konferenz wird es ermöglicht die thematischen Zusammenhänge und Ähnlichkeiten zwischen Publikationen und Autoren der Konferenz(-serie) zu explorieren (siehe Abbildung 1c).

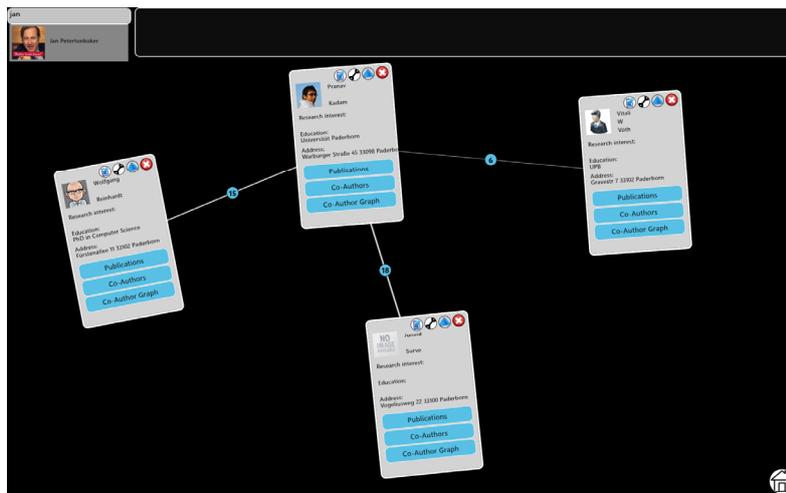
Der *Trendbasierte Zugang* zu Wissenschaftsnetzwerken schließlich ermöglicht es dem Benutzer Trends in Publikationsthemen zu explorieren. Dazu kann beispielsweise dargestellt werden, wie sich die publizierten Themen innerhalb eines Forschungsbereichs oder einer Konferenzserie in einem Zeitintervall verändert haben.

4 Die PUSHPIN_{MT} Anwendung

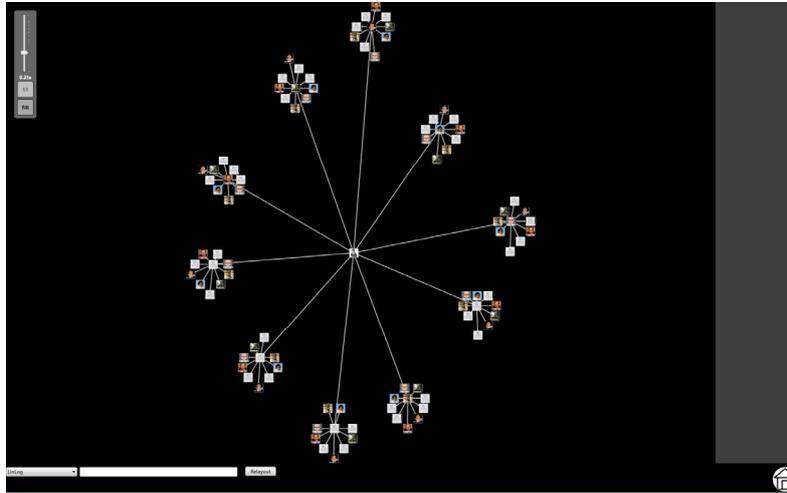
Die Multitouch-Anwendung unterstützt den Anwender durch geeignete Menüführung in der Auswahl des jeweiligen Zugangs (Abbildung 1a). Neben der Suche nach Autoren ist auch das Scannen von QR Codes (zum Beispiel von Konferenz-Badges oder aus der mobilen PUSHPIN Anwendung) möglich, um eine Repräsentation des entsprechenden Forschers oder einer Publikation in PUSHPIN_{MT} zu erhalten.



1a): Auswahl der Explorationszugangsart



1b): Visualisierung von Relationen zwischen Autoren



1c): Visualisierung von Ko-Autorennetzwerken

Abbildung 1: Screenshots der PUSHPIN_{MT} Multitouch-Anwendung

Die PUSHPIN_{MT} Anwendung wurde unter Verwendung von WPF 4.0 und dem Microsoft Surface SDK 2.0 realisiert. Für die Darstellung der Forschernetzwerke wurde die Graph# Bibliothek verwendet.

5 Zusammenfassung und Ausblick

In diesem Artikel haben wir vier Zugänge zur Exploration wissenschaftlicher Netzwerke auf Personen- und Publikationsebene vorgestellt. Innerhalb des PUSHPIN-Projekts wird die Datenanalyse und –bereitstellung realisiert und eine Multitouch-Anwendung entwickelt, die Benutzern die verschiedenen Explorationsmöglichkeiten anbietet. Geeignete Evaluationen müssen zeigen, ob eine derartige Darreichungsform in der Tat zur Awareness-Unterstützung von Forschern beiträgt und inwieweit sie einen Beitrag zur stärkeren Kooperation in Wissenschaftsnetzwerken leisten kann.

Derzeit arbeiten wir an Authentifizierungsmethoden, die es ermöglicht sich an der Multitouch-Anwendung via Bluetooth mittels Smartphone zu identifizieren. Der Benutzer erhält dann eine Repräsentation seiner selbst in der Anwendung und kann ausgehend von sich selbst mit der Personen- oder Themenbezogenen Exploration der Daten beginnen.

Literaturverzeichnis

- Kessler, M.M. (1963). *Bibliographic coupling between scientific papers*, American Documentation, Bd. 14, S. 10–25.
- Priem, J. & Hemminger, B. (2010). Scientometrics 2.0: New metrics of scholarly impact on the social web. *First Monday*, 15(7).
- Reinhardt, W. (2012). *Awareness Support for Knowledge Workers in Research Networks*. Dissertation, Open University of the Netherlands. Online zugänglich <http://bit.ly/PhD-Reinhardt>
- Renear, A. H. & Palmer, C. L. (2009). Strategic reading, ontologies, and the future of scientific publishing. *Science*, 325(5942).
- Smalheiser, N. R. & Torvik, V. I. (2009). *Annual Review of Information Science and Technology*, 43(1), S. 1-43.
- Weingart, P. & Winterhager, M. (1984). *Die Vermessung der Forschung: Theorie und Praxis der Wissenschaftsindikatoren*. Campus: Frankfurt a.M.

Kontaktinformationen

Universität Paderborn
Institut für Informatik
Dr. Wolfgang Reinhardt
Fürstenallee 11
33102 Paderborn
wolle@upb.de