

Persönliche Publikationslisten im WWW – Webometrische Aspekte wissenschaftlicher Selbstdarstellung am Beispiel der Universität Bielefeld

Najko Jahn, Mathias Lösch und Wolfram Horstmann

Universität Bielefeld
Universitätsbibliothek
Universitätsstraße 25
D-33615 Bielefeld

{najko.jahn, mathias.loesch, wolfram.horstmann}@uni-bielefeld.de

Zusammenfassung

Persönliche Publikationslisten sind ein Bestandteil wissenschaftlicher Selbstdarstellung im Web. Der Beitrag stellt sich die Frage, wie sich wissenschaftliche Dokumente aus den persönlichen Webseiten heraus identifizieren und für empirische Untersuchungen im Rahmen von Linkanalysen nutzen lassen. Am Beispiel der Universität Bielefeld wurden 1.358 wissenschaftliche Volltexte eruiert und ihre Rezeption über zwei exemplarische Linkkontexte – Wikipedia und dem politischen Diskurs im Web – exploriert.

Einleitung

Das Web hat nicht nur die Art und Weise verändert, wie Wissenschaftler publizieren, sondern auch wie sie ihre Forschungsergebnisse und mithin sich selbst im Web darstellen. Die persönliche Publikationsliste ist für diese Entwicklung ein altbekanntes Beispiel, das disziplinübergreifend verbreitet ist. Eingebettet in den persönlichen CV einer Wissenschaftlerin oder eines Wissenschaftlers ist die persönliche Publikationsliste eine besonders kondensierte Darstellung wissenschaftlicher Aktivitäten. Neben der klassischen Funktion als Referenz auf die erschienenen Forschungsbeiträge nutzen Forschende zusätzlich die persönliche Publikationsliste, um auf Dokumente zu verlinken, sie verfügbar zu machen oder diese sogar als originären Erscheinungsort für die erste Fassung von zur Veröffentlichung bestimmten Beiträgen zu verwenden.

Eine Vielzahl an Services hat das Potential persönlicher Publikationslisten in den letzten Jahren aufgegriffen. Philpapers.org stellt ein prominentes Beispiel dar, wie aggregierte Volltexte, die von persönlichen Publikationslisten stammen, in ein kollaboratives Angebot innerhalb einer Fachgemeinschaft, hier die Philosophie, aufgenommen und von Fachkollegen erschlossen und diskutiert werden. Ein disziplinübergreifendes Beispiel ist researchGate.net. Der Social Network Service konzentriert sich neben kollaborativen Werkzeugen auf die wirkungsmächtige Außendarstellung eines Forschenden. Und auch Hochschulen greifen zunehmend das Potential der wissenschaftlichen Selbstdarstellung über die persönliche Publikationsliste auf (Horstmann & Jahn, 2010).

Ziel des Beitrages ist die beispielhafte Exploration der Selbstdarstellung wissenschaftlicher Aktivitäten im Web anhand von persönlichen Publikationslisten an der Universität Bielefeld. Insbesondere die Rezeption der Dokumente, die sich auf den persönlichen Publikationslisten befinden, ist von besonderem Interesse, da die freie Verfügbarkeit wissenschaftlicher Veröffentlichungen und ihre Vernetzung im Web ‚prima facie‘ Nachnutzungsszenarien über die Grenzen der persönlichen Fachgemeinschaft hinaus ermöglichen.

Die Exploration geschieht methodisch über eine Linkanalyse. Die Bibliotheks- und Informationswissenschaft erhebt Linkanalysen im Rahmen webometrischer Fragestellungen. Die Webometrie widmet sich beispielsweise den Webpräsenzen von akademischen Einrichtungen und Universitäten, um Linkmaße für die Außenwirkung von Forschenden zu validieren (Thelwall & Ruschenburg, 2006). Aufgrund der Komplexität der universitären Webpräsenzen und der Vielzahl an Linkkontexten nehmen wissenschaftliche Volltexte auf persönlichen Webseiten nur einen geringen Stellenwert in der webometrischen Debatte ein. Untersuchungen legen nahe, dass sich nur ein verschwindend geringer Anteil der Links auf universitäre Webpräsenzen überhaupt als ein intendierter Bezug auf die wissenschaftliche Erkenntnis interpretieren lässt (ebd.).

Um somit ein aussagekräftiges Bild der Rezeption wissenschaftlicher Volltexte auf persönlichen Publikationslisten zu erlangen, bedarf es des dezidierten Fokus auf persönliche Listen und die in ihnen enthaltenen wissenschaftlichen Dokumente. Und um Linkkontexte festzustellen, wird wiederum eine Gruppierung der Links benötigt. Somit stellen wir uns zunächst die Frage, wie sich wissenschaftliche Dokumente aus den persönlichen Webseiten heraus identifizieren lassen. Anschließend aggregieren wir die Links, die auf die Dokumente verweisen, und stellen exemplarisch zwei Linkkontexte der Selbstdarstellung wissenschaftlicher Veröffentlichungen auf persönlichen Webseiten vor: die Wikipedia und der politische Diskurs im Web.

Datengewinnung

Datengrundlage unserer Untersuchung ist ein Korpus aus wissenschaftlichen Publikationen, das wir von persönlichen Publikationslisten Bielefelder Wissenschaftler für eine frühere Studie zusammengestellt hatten (Jahn, Lösch & Horstmann, 2010). Ausgangspunkt war ein zuvor manuell erstelltes Verzeichnis der Publikationslisten von 1.214 Bielefelder Wissenschaftlern. Das Verzeichnis wurde insofern eingeschränkt, als wir nur Personen berücksichtigten, die zum Zeitpunkt der Erhebung promoviert waren. Das Korpus wurde dann erstellt, indem frei zugängliche PDF-Dokumente automatisiert von den persönlichen Publikationslisten der Wissenschaftler aggregiert und intellektuell als Publikationen klassifiziert wurden. Letzteres war nötig, weil nicht alle PDF-Dokumente, die auf solchen Webseiten veröffentlicht werden, notwendigerweise Publikationscharakter haben – man findet dort auch Lebensläufe, Seminarunterlagen, Präsentationen und andere, nicht wissenschaftlich relevante Dokumente. Die Beschränkung auf PDF-Dokumente wurde pragmatisch aus der Erfahrung der Autoren begründet, dass die meisten digitalen Volltexte als PDF veröffentlicht werden. Das Ergebnis dieser Prozedur ist ein Korpus der frei zugänglichen Dokumente der Bielefelder Forschenden. Zusätzlich zu den eigentlichen Dokumenten wurden auch deren URLs gespeichert. Diese bilden den Ausgangspunkt der vorliegenden Studie.

Um herauszufinden, welche Webseiten Hyperlinks auf die von uns als wissenschaftliche Publikationen verifizierten Dateien enthalten, verwendeten wir deren URLs als Eingabe für das API von Yahoo! Site Explorer (<http://siteexplorer.search.yahoo.com/>). Dabei handelt es sich um einen Service, der es erlaubt, Informationen über Webseiten aus dem Index der Suchmaschine Yahoo! auszulesen und insbesondere welche Seiten auf die Datei mit einer gegebenen URL zeigen (Backlinks). Wir verwendeten ein eigens entwickeltes Python-Skript, um die Backlinks für alle Dokumente in unserem Korpus zu aggregieren. Das Skript ordnet jedem Dokument seinen Autor und die entsprechenden Backlinks zu.

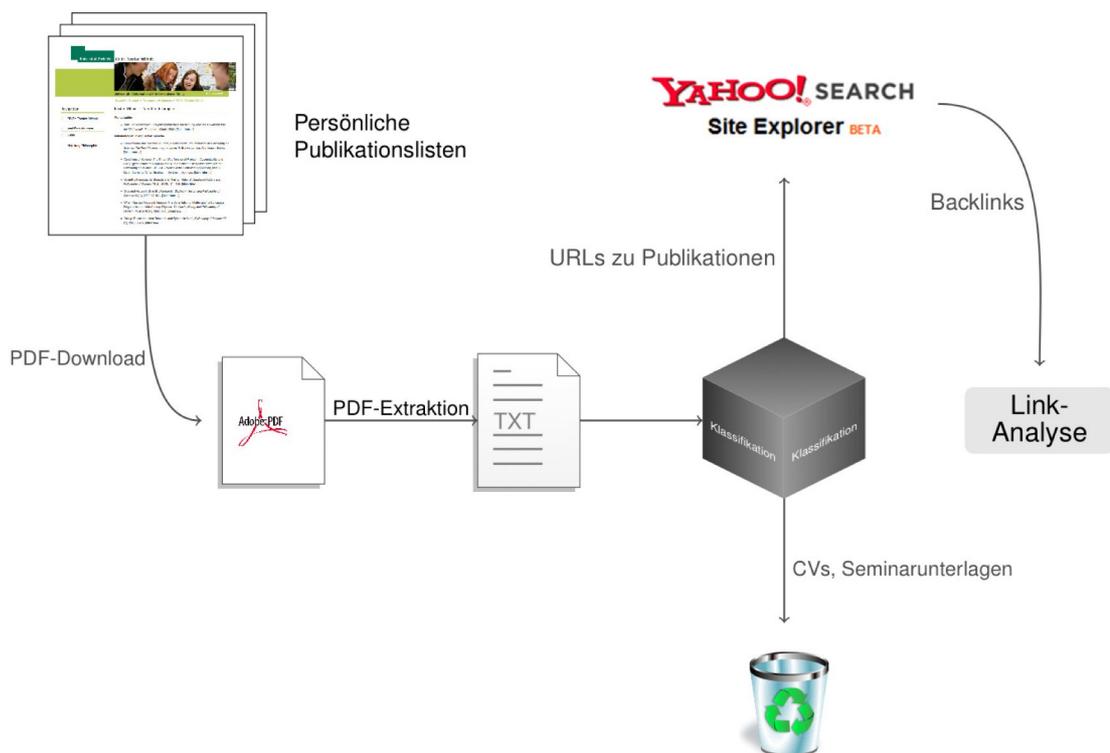


Abbildung 1: Aggregation von Backlinks auf wissenschaftliche Volltexte

Link-Netzwerke

Insgesamt konnten wir 1.358 wissenschaftliche Dokumente aggregieren, die sich auf 138 Wissenschaftler an der Universität Bielefeld verteilen. Auf die wissenschaftlichen Dokumente wiederum entfallen in der Summe 1.175 Backlinks, die über Yahoo! Site Explorer identifiziert wurden. Die weitere Exploration erfolgt mit Methoden der Sozialen Netzwerkanalyse (SNA). Die Visualisierungen wurden mithilfe der Software visone realisiert (Baur, 2008).

Zwei Aggregationsstufen standen hierbei im Vordergrund, um den Datenbestand stärker zu strukturieren. Zunächst lässt sich die persönliche Publikationsliste als Ganzes analog zu einer Forscher-Bibliographie auffassen. Die Backlinks können dann als ein Indikator für die thematische Ausrichtung eines Forschenden herangezogen werden. Des Weiteren ist es möglich, die Backlinks unter ihrer Domain zusammenzufassen. Damit lassen sich wiederum Domains, von denen die Links auf die wissenschaftlichen Volltexte stammen, als Indikator für die thematische Rezeption eines Dokuments heranziehen.

Es gehen daher in beide Untersuchungen jeweils zwei Mengen an Untersuchungsgegenständen ein, die bipartit angeordnet werden. Eine bipartite Anordnung bedeutet grob, dass Beziehungen im Netzwerk über Knoten einer Menge über die Verbindung zu einem Knoten der anderen Menge instantiiert werden. Das bipartite Arrangement wird auch für Zitationsstudien und die Untersuchung von Kollaborationsbeziehungen über Koautorschaften in der Bibliometrie verwendet (Havemann, 2009).

Wikipedia

Unser erstes Netzwerk, das Backlinks auf wissenschaftliche Volltexte als ein Indikator für die thematische Ausrichtung der Forschenden präsupponiert, basiert auf Backlinks, die aus der Wikipedia stammen. Die Wikipedia ist in der Vielzahl der Linkkontexte, die wir eruiert haben, aufgrund ihres enzyklopädischen Charakters und ihrer großen Popularität ein gutes Indiz für die themenbezogene Sichtbarkeit eines Forschenden.

Die Visualisierung des Netzwerks (Abbildung 2) offenbart die Eigenschaft der bipartiten Anordnung. Die grünen Knoten repräsentieren die persönliche Publikationsliste eines Forschenden, die violetten Knoten stehen für die Wikipedia-Artikel. Da die Wikipedia nach Sprachen unterteilt ist, fassen wir Backlinks aus Artikeln in mehreren Sprachen zusammen. So finden sich Werke eines Forschenden in Wikipedia-Beträgen zu „Xenoturbella“, eine Wurmart, die besonders für Genomforscher interessant ist. Publikationen dieses Autors werden in sechs Sprachversionen des Wikipedia-Artikels referenziert. Analog gehen wir vor, wenn ein Wikipedia-Beitrag mehrere Beiträge eines Forschenden zitiert.

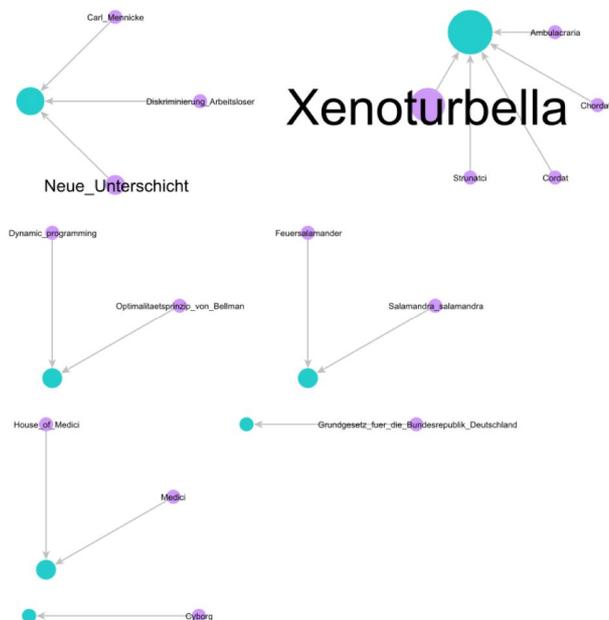


Abbildung 2: Linkkontext Wikipedia

Die vorliegende Unterrichtung weist die formalen Merkmale einer wissenschaftlichen Arbeit auf, um den parlamentarischen Gesetzgebungsprozess zu informieren. Ihre Rezeption ist dabei besonders spannend, da sie sowohl von akademischen Einrichtungen im Umfeld der Rechts- und Politikwissenschaft als auch von Medien, in privaten Blogs und auf Seiten von Nichtregierungsorganisationen rezipiert wird.

Im Ausschnitt wird noch eine weitere Netzwerkeigenschaft deutlich: Domains, die auf die wissenschaftlichen Dokumente der persönlichen Seiten verweisen, koppeln diese Dokumente und bilden dadurch eine weitere Ebene der Darstellung wissenschaftlicher Veröffentlichungen an der Universität Bielefeld, die die Selbstdarstellung des einzelnen Forschenden ergänzt.

Fazit

Persönliche Publikationslisten im Web sind ein wichtiger Aspekt in der Selbstdarstellung wissenschaftlicher Aktivitäten. Verlinkte oder selbstarchivierte wissenschaftliche Volltextdokumente offenbaren eine Vielzahl an Untersuchungs- und Interpretationsmöglichkeiten. Die in der Arbeit explorierten Teilgraphen weisen Beziehungsgefüge auf, die sowohl eine thematische Einordnung wissenschaftlicher Autoren als auch die Bestimmung von Kontexten ihrer Rezeption unterstützen.

Insgesamt erscheinen Linkanalysen, die ihre Daten über verschiedene Aggregationsstufen gewinnen, ein vielversprechender methodischer Ansatz für das wachsende Feld der Digitalen Wissenschaft. Linkanalysen bieten Daten und Modelle für ihre interdisziplinären Fragestellungen und haben zugleich das Potential, die Entwicklung neuer Dienste durch Bibliotheken und Informationsdienstleister zu begleiten und anzuregen.

Literatur

Baur, Michael (2008). visone – Software for the Analysis and Visualization of Social Networks. Universität Karlsruhe. urn:nbn:de:swb:90-108975 .

Havemann, Frank (2009). Einführung in die Bibliometrie. Berlin: Ges. für Wiss.-Forschung c/o Inst. für Bibliotheks- und Informationswiss. urn:nbn:de:kobv:11-10097376 .

Horstmann, Wolfram und Najko Jahn (2010). Persönliche Publikationslisten als hochschulweiter Dienst – Eine Bestandsaufnahme. BIBLIOTHEK Forschung und Praxis 34, no. 2, 185-193. doi:10.1515/bfup.2010.032 .

Jahn, Najko, Mathias Lösch und Wolfram Horstmann (2010). Automatic Aggregation of Faculty Publications from Personal Web Pages. The Code4Lib Journal, September 21. <http://journal.code4lib.org/articles/3765> .

Thelwall, Mike und Tina Ruschenburg (2006). Grundlagen und Forschungsfelder der Webometrie. Information – Wissenschaft und Praxis 57, no. 8, 401-406.