

Grundlagen und Forschungsfelder der Webometrie

Mike Thelwall, Wolverhampton (England) und Tina Ruschenburg, Bielefeld

Die Webometrie ist ein Teilbereich der Informationswissenschaft, der zur Zeit auf die Analyse von Linkstrukturen konzentriert ist. Er ist stark von der Zitationsanalyse geprägt, wie der empirische Schwerpunkt auf der Wissenschaftsanalyse zeigt. In diesem Beitrag diskutieren wir die Nutzung linkbasierter Maße in einem breiten informetrischen Kontext und bewerten verschiedene Verfahren, auch im Hinblick auf ihr generelles Potential für die Sozialwissenschaften. Dabei wird auch ein allgemeiner Rahmen für Linkanalysen mit den erforderlichen Arbeitsschritten vorgestellt. Abschließend werden vielversprechende zukünftige Anwendungsfelder der Webometrie benannt, unter besonderer Berücksichtigung der Analyse von Blogs.

Basic principles and research fields of webometrics

Webometrics in information science is currently dominated by link analysis and strongly influenced by citation analysis, being typically applied to scientific data. In this chapter we discuss the use of link count metrics in the broad context of informetrics, assessing methodologies and their potential for general social sciences research. In this context, a generic framework for social science link analysis is presented. The final section comprises promising fields for future research in webometrics, placing a special focus on the analysis of blogs.

Einleitung

Bereits als das World Wide Web erst wenige Jahre alt war, erkannten einige Informationswissenschaftler die strukturelle Ähnlichkeit von Hyperlinks und Zitationen, die beide Verweise zwischen Dokumenten darstellen (Larson, 1996; Rodríguez i Gairín, 1997; Rousseau, 1997). Dies bildete die Grundlage für ein neues Forschungsfeld, die Webometrie (Almind & Ingwersen, 1997). Sie ist definiert als Untersuchung des Webs mit Hilfe quantitativer, aus der Informatik entlehnter Verfahren (Bjørneborn & Ingwersen, 2004).

Das ganze Potential des Webs wurde erst für die Linkanalyse nutzbar, als kommerzielle Suchmaschinen die Option anboten, gezielt nach Links zu suchen (Ingwersen, 1998; Rodríguez i Gairín, 1997). So akzeptierte AltaVista ab 1997 äußerst mächtige Anfragen, beispielsweise nach der Zahl aller Seiten weltweit, die Links auf schwedische Seiten enthalten (Ingwersen, 1998). Dadurch wurde es möglich, mit einem Arbeitsaufwand von wenigen Stunden den „impact“ verschiedener Websites durch geeignete Anfragen zu vergleichen. Grundlegend hierfür ist die Annahme, dass sich die Außenwirkung veröffentlichter Informationen durch Links – ebenso wie durch Zitationen – messen lässt. Während in der Zitationsanalyse zumeist die kostenpflichtigen Datenbanken des Institute for Scientific Information verwendet werden, ist die „Datenbank“ World Wide Web gratis zugänglich, was der Linkanalyse möglicherweise einen größeren Nutzerkreis verschafft. Mit Hilfe kommerzieller Suchmaschinen wurde bereits die Außenwirkung vieler geographischer und organisationaler Einheiten verglichen, darunter Fachzeitschriften, Länder, die Universitäten eines Landes und ihre Fachbereiche sowie Bibliotheken (An & Qiu, 2004; Harter & Ford, 2000; Ingwersen, 1998; Smith, 1999; Tang & Thelwall, 2006; Thomas & Willet, 2000). Bereits die frühen Studien zeigten, dass Linkanalysen besondere Vorsicht und Sorgfalt erfordern, da es zahlreiche Störfaktoren gibt. Zu nennen sind hier z.B. mehrfach existierende Webseiten und Websites, Fehler in der Ausgabe von Suchmaschinen, unvollständige Abdeckung des Webs durch Suchmaschinen, Linkwiederholungen innerhalb einer Website sowie fragliche oder triviale Gründe für das Einrichten von Links (Barllan, 2001; Bjørneborn & Ingwersen, 2001; Egghe, 2000; Harter & Ford, 2000; Smith, 1999; Snyder & Rosenbaum, 1999; van Raan, 2001). Nichtsdestotrotz hat die Linkanalyse interessante und nützliche Ergebnisse hervorgebracht, und verschiedene Forschungsfelder außerhalb der Informationswissenschaft haben ihre Verfahren übernommen (s.u.).

Dieser Artikel bietet einen Überblick über einige gegenwärtig praktizierte und zukünftig vielversprechende Varianten der Linkanalyse. Zunächst folgt jedoch eine kurze Darstellung möglicher Zugänge zu Linkdaten.

Datenquellen

Linkdaten können bei Suchmaschinen abgefragt, aus Linkdatenbanken extrahiert oder direkt mit einem Webcrawler-Programm erhoben werden.

Zum Zeitpunkt der Entstehung dieses Textes kann die Linksuche von AltaVista dazu genutzt werden, die Anzahl der Webseiten in der Datenbank dieser Suchmaschine zu ermitteln, die einen Hyperlink auf eine vorgegebene Seite enthalten. So liefert bspw. der Befehl „link:www.wlv.ac.uk/disclaimer.htm“ die Anzahl von Seiten mit einem virtuellen Verweis auf diese URL. Zusätzlich bietet diese Suchmaschine noch eine allgemeinere Suchoption: Mit dem Kommando „linkdomain:“ lassen sich diejenigen Seiten zählen, die einen Link in die angegebene Domain aufweisen. Das Ergebnis von „linkdomain:www.wlv.ac.uk“ umfasst folglich alle Seiten, die Verweise auf eine beliebige Seite innerhalb der Domain www.wlv.ac.uk enthalten, nicht nur auf die Homepage. Folglich lassen sich die Linkbeziehungen innerhalb einer Gruppe von Webseiten (mit „link:“) oder einer Gruppe von Websites (mit „linkdomain:“) analysieren, indem man die Anzahl der Links zwischen jedem einzelnen Paar von ihnen ermittelt. Google bietet ebenfalls eine Option zur Linksuche an sowie eine Programmierschnittstelle (API), die die Automatisierung von Abfragen ermöglicht (vgl. www.google.com/apis). Allerdings handelt es sich bei Google API um eine Beta-version, die einem empirischen Vergleich zufolge nur auf einen Teil der gesamten Google-Datenbank zugreift und andere Rankingprozeduren als die Standardsuche über www.google.com verwendet (Mayr & Tosques, 2005).

Kommerzielle Suchmaschinen bieten keine vollständige Abdeckung des Webs, d.h. sie erfassen einige Websites nicht oder nur teilweise. Sie sind deshalb kein ideales Erhebungsinstrument für Linkdaten, wenngleich es sich häufig nicht vermeiden lässt, sie zu nutzen (Thelwall, 2004). Unter <http://cybermetrics.wlv.ac.uk/database> ist eine Zusammenstellung von Online-Linkdatenbanken zu finden, die u.a. die Linkstrukturen der Universitäts-Websites von fünf Ländern enthält. Ihre Nutzung, einschließlich verschiedener Programme zur Analyse der Ergebnisse, ist für Forschungszwecke kostenlos.

Gleiches gilt für einen Webcrawler, der unter <http://linkanalysis.wlv.ac.uk> abrufbar ist und der die Möglichkeit bietet, eigene Linkdaten zu sammeln. Er ist für Websites mit bis zu 5.000 Seiten geeignet, für sehr große Internetauftritte nicht.

Literaturübersicht

Die vergleichende Linkanalyse von Universitätsinstituten

Verschiedene empirische Studien widmen sich dem Vergleich der Websites von Universitätsinstituten (bzw. von Fakultäten oder Fachbereichen). Meist haben sie das Ziel, Linkmaße als Indikatoren für die virtuelle Außenwirkung von Forschungseinrichtungen zu validieren. Eine verbreitete Hypothese ist, dass die Zahl der Links, die auf ein Institut verweisen, mit anderen etablierten Forschungsindikatoren korreliert, z.B. mit Zitationsmaßen. Üblicherweise werden dabei Links innerhalb der untersuchten Einrichtung ausgeschlossen, in der Annahme dass sie größtenteils internen Navigationszwecken dienen. Außerdem werden die Linkwerte häufig standardisiert, indem die Anzahl der Links auf die Website eines Universitätsinstituts durch die Anzahl ihrer einzelnen Seiten oder durch die Mitarbeiterzahl dividiert wird. Es handelt sich dabei um Variationen des „Web Impact Factor“ von Ingwersen (1998).

Nachdem erste Ergebnisse gegen diese Hypothese sprachen (Thomas & Willet, 2000), zeigten spätere Studien tatsächlich eine Korrelation zwischen Forschungsindikatoren und Linkwerten, was für die Nutzung von Links zur Wissenschaftsanalyse spricht (Li, Thelwall, Musgrove & Wilkinson 2003; Li, Thelwall, Wilkinson & Musgrove 2005a und 2005b). Allerdings sollten sie in Evaluationen keinen signifikanten Stellenwert einnehmen, anders als es bei Zitationen zum Teil der Fall ist. Denn nur ein kleiner Anteil der Links spiegelt direkt Forschungsleistungen wider, wie z.B. mit Zitationen vergleichbare Links auf Online-Artikel. Meist besteht nur ein eher indirekter Zusammenhang, z.B. wenn sich die Links auf die Lehre oder die gemeinsame Mitgliedschaft in einer Organisation oder Forschungsgruppe beziehen (Harries, Wilkinson, Price, Fairclough & Thelwall, 2004; Wilkinson, Harries, Thelwall & Price, 2003). Vergleichliche verschiedener Disziplinen auf der Grundlage institutsbezogener Linkanalysen haben erhebliche Unterschiede offenbart, was den Umfang der Internetauftritte betrifft. Während es noch naheliegend scheint, dass in der Informatik aus

fachlichen Gründen eine besondere Affinität zur Gestaltung von Websites besteht, ergab eine Studie über die USA, dass auch Universitätsinstitute im Bereich Physik mindestens tausendmal mehr Webseiten und Links einrichten als solche im Bereich Geschichte (Tang & Thelwall, 2003). Die Studien von Li et al. (2005a; 2005b) zeigten die disziplinären Unterschiede bisher am detailliertesten. Darin wurden Institute mehrerer naturwissenschaftlicher Disziplinen in ähnlichen Ländern (jeweils Physik, Chemie und Biologie in Australien, Kanada und Großbritannien) verglichen. Die Ergebnisse deuten darauf hin, dass selbst ähnliche Fächer das Web in sehr unterschiedlicher Art und Weise nutzen, u.a. im Hinblick auf die Menge der eingerichteten Links. Dieser Befund ist mit Theorien aus der Sozialinformatik (Kling & McKim, 2000) konsistent, nach denen die Nutzung von Kommunikationstechnologien nicht notwendigerweise zu einheitlichen Mustern führt: Die jeweiligen alltagspraktischen Bedürfnisse bestimmen, wie neue Technologien angenommen und eingesetzt werden. Vielleicht überraschender an der Untersuchung von Li et al. sind die internationalen Unterschiede innerhalb der einzelnen Disziplinen. Bei den Biologie-Instituten wiesen Websites in Australien mit 60 Prozent deutlich weniger internationale Links auf als in Großbritannien (74 Prozent) oder Kanada (80 Prozent). Aus einer funktionalistischen Perspektive und unter Berücksichtigung des internationalen Charakters des World Wide Web könnte man eine größere Ähnlichkeiten im Hinblick auf die Webnutzung erwarten. Die gefundenen Differenzen sprechen jedoch für die organisationsbezogenen Ansätze der Wissenschaftssoziologie, die die Bedeutung diverser sozialer Faktoren für die wissenschaftliche Praxis betonen (Fuchs, 1992). Aus einem praktisch-informatrischen Blickwinkel ist das entscheidende Ergebnis, dass Linkdaten wertvolle Hilfe dabei leisten können, unerwartete Unterschiede und – da der Linkkontext auch eingesehen werden kann – ihre Ursachen aufzudecken. Folglich scheint die Linkanalyse ein naheliegendes Werkzeug für die Wissenschaftssoziologie zu sein, ebenso wie die Zitationsanalyse (z.B. bei Merton).

Eine interessante Anwendungsmöglichkeit der Linkanalyse ist ihre Nutzung in Kombination mit weiteren Datenquellen, die über die Verbindungen zwischen Forschern Aufschluss geben. Ein Beispiel ist die Verbindung von Linkdaten mit Informationen über die Teilnahme an EU-Projekten zu dem Zweck, europäische Kooperationsmuster in einem bestimmten Forschungsgebiet zu analysieren (Heimeriks, Hörlesberger & van den Besselaar, 2003). Dies zeigt, dass die Linkanalyse in der Szientometrie auch in einem Mehr-Methoden-Design verwendet werden kann.

Soziale Netzwerkanalyse

Ein weiteres Forschungsgebiet, für das Links interessant sind, weil sie dem üblichem Forschungsgegenstand ähnlich sind, ist die soziale Netzwerkanalyse (SNA) (Wasserman/Faust 1994). Dabei geht es um Individuen oder Organisationen sowie vor allem um die Verbindungen zwischen ihnen. Eine bekannte Studie befasst sich bspw. mit dem Fluss mündlich verbreiteter Informationen in einem Dorf, wobei zur Erklärung der gefundenen Kommunikationsmuster Netzwerke sozialer Bekanntschaften herangezogen werden (Rogers & Kincaid, 1981).

Für die Anwendung von Methoden der SNA auf Netzwerke aus Links zwischen Websites wurde der Begriff „Hyperlink Network Analysis“ geprägt (Park, 2003). Die Forschungsfragen dieser Linknetzwerkanalyse konzentrieren sich häufig auf die Netzwerkeigenschaften von Website-Gruppen. Um Eigenschaften wie die Zentralität einzelner Knoten (also Websites oder -seiten) zu bestimmen, werden Maße der SNA auf das Web übertragen. Dazu gehören beispielsweise die Anzahl der Inlinks¹, die Anzahl der Outlinks und die Häufigkeit, mit der ein Knoten Bestandteil der kürzesten Linkverbindung zwischen zwei andern Knoten ist („betweenness centrality“). Ein illustratives Beispiel dieser Art von Forschung ist eine Untersuchung von Garrido und Halavaia (2003) über Websites, die die Zapatistenbewegung in Mexiko unterstützen. Sie zeigte, dass Seiten zur Geschlechterpolitik in diesem Netzwerk erstaunlicherweise den zweitgrößten Stellenwert einnahmen, direkt nach offiziellen zapatistischen Informationsseiten. Ein Kritikpunkt an SNA-basierter Webanalyse ist die Tendenz einiger Studien zu der Annahme, dass Links immer ein Kommunikationsmittel seien, da sie faktisch eine breite Vielfalt von Funktionen erfüllen (Park & Thelwall, 2003).

In der Informationswissenschaft ist das Potential der sozialen Netzwerkanalyse für die Untersuchung von Informationsnetzwerken erkannt worden (Otte & Rousseau, 2002). Björneborn (2004) hat SNA-Maße auf britische Universitätswebsites angewandt und interessante, disziplinenübergreifende Verbindungen gefunden. Es handelt sich hierbei um ein junges Forschungsfeld, das neue Erkenntnisse über Informationsstrukturen und akademische Kommunikationsmuster erschließen und eventuell weitere Inspiration aus soziologischen und mathematischen Netzwerktheorien erhalten kann (Granovetter, 1973; Watts & Strogatz, 1998).

Interessant ist, dass auch einige informatiker Verfahren zur Analyse sozialer Netzwerke unter Verwendung von Links entwickelt haben (Adamic & Adar, 2003). Ihre Neugier im Hinblick auf soziale Gruppen könnte mit dem Bedarf der Informatik an neuen Software-Anwendungen zur Unter-

1 Der Begriff „Inlink“ bezeichnet in der Webometrie die auf eine Website, eine Webseite oder einen anderen definierten Bereich des Webs verweisenden Links, während die darin enthaltenen Links „Outlinks“ genannt werden. Dies entspricht der in der Bibliometrie üblichen Unterscheidung von Zitationen und Referenzen.

stützung von Gruppenaktivitäten wie Online-Zusammenarbeit und mit der Aussicht auf die Nutzung von Gruppenwissen zur Optimierung von Computersystemen (z.B. durch „Collaborative Filtering“) zusammenhängen. Soziale Analysen des World Wide Web scheinen also eine interessante Forschungsrichtung für viele verschiedene Bereiche der Wissenschaft zu sein.

Andere Formen der sozialen Linkanalyse

In den Sozialwissenschaften wurden Links bereits mit vielversprechenden Resultaten in einem außer-informetrischen Kontext ausgewertet. Das „Internet and Elections Project“ (politicalweb.info) ist ein internationaler, kooperativer Versuch, den Einsatz des Webs bei politischen Wahlen auf der ganzen Welt zu vergleichen. Links nehmen bei dieser Analyse eine zentrale Position ein. Zum einen werden sie dazu genutzt, relevante Websites zu identifizieren, zum anderen ist auch die Linkpraxis selbst Forschungsgegenstand. Eine Vorstudie zu einer US-Wahl kam zu dem Schluss, dass das Einrichten von Links nicht auf allen untersuchten Wahlkampfseiten verbreitet war, so dass sich bisher keine allgemeinen Standards etabliert haben. Dennoch zeigte sich eine gewisse Einheitlichkeit in ihrem Gebrauch, und zwar insbesondere eine Tendenz zu Referenzen auf gleichgesinnte Websites und weniger auf solche, die abweichende Meinungen zu einem bestimmten Thema enthielten (Foot, Schneider, Dougherty, Xenos & Larsen, 2003). Das Einrichten von Links ist auch für die qualitativen Sozialwissenschaften von Interesse, was auf seine Bedeutung als eigenständiges Phänomen hinweist. Als Beispiel kann eine Diskussion der Linkpraxis von Webautoren dienen, die Beiträge zu einer öffentlichen Debatte verfassen (Hine, 2000). Auch eine ethnographische Untersuchung der Links, die auf die Online-Publikationsdatenbank „PubMed Central“ verweisen (Beaulieu, 2005), fällt in diesen Bereich. Ihr zufolge sind auf den

Internetauftritten von Forschern oder Laboren erstaunlich selten Verweise auf diese Website zu finden. Dies gilt selbst für Wissenschaftler, deren Artikel in dieser Datenbank enthalten sind und die ihre Forschungsergebnisse folglich durch einen Link leichter zugänglich machen könnten.

Das Verfahren der sozialwissenschaftlichen Linkanalyse

Die folgende Liste von Arbeitsschritten kann als allgemeiner Rahmen für Linkanalysen in der sozialwissenschaftlichen Forschung dienen (Thelwall, 2004, p. 3):

- 1) Eine geeignete Forschungsfrage formulieren, unter Berücksichtigung des vorhandenen Wissens über die betreffenden Webstrukturen.
- 2) Eine Pilotstudie durchführen.
- 3) Webseiten oder Websites identifizieren, die geeignet sind, um die Forschungsfrage zu prüfen.
- 4) Linkdaten mit einer Suchmaschine oder mit einem eigenen Webcrawler erheben. Dabei Sicherheitsmaßnahmen treffen um zu gewährleisten, dass die erzielten Ergebnisse korrekt sind.
- 5) Techniken zur Datenbereinigung auf die Linkdaten anwenden, sofern möglich, und eine geeignete Zählweise wählen.
- 6) Die Ergebnisse der berechneten Linkmaße zum Teil durch Korrelationstests validieren.
- 7) Die Interpretation der Ergebnisse durch eine partielle Klassifikation der Links validieren.
- 8) Einen Bericht über die Ergebnisse verfassen, in dem die Ergebnisse in Übereinstimmung mit der Linkklassifikation interpretiert werden. Darin eine detaillierte Beschreibung der Klassifikation bzw. Beispiele zur Illustration der verwendeten Kategorien einschließen.
- 9) Die Einschränkungen der Untersuchung sowie die Parameter der Datenerhebung und -verarbeitung offenlegen (Schritt 3 bis 5).

Dieser allgemeine Rahmen spiegelt viele der bereits ausgeführten Punkte wider. Besondere Aufmerksamkeit gebührt der Pilotstudie, da das Web in vielfältiger Weise genutzt wird (Burnett & Marshall, 2002) und sich intuitive Vorstellungen über den Einsatz von Links in spezifischen Anwendungskontexten als falsch erweisen können. Die Vorstudie ermöglicht es, ein Forschungsvorhaben rechtzeitig abzubrechen, bevor allzu großer Aufwand betrieben wird, wenn sich eine Linkanalyse als nicht aufschlussreich erweist. Die vielleicht wichtigste Information in Bezug auf die genannten Arbeitsschritte ist jedoch, dass die Klassifikation von Links einen zentralen Stellenwert bei der Interpretation der Ergebnisse einnimmt. Ohne genaues Wissen über die Gründe für die Entstehung der Links lassen sich aus den Daten lediglich abstrakte Schlüsse ziehen.

Zukünftige Forschungsrichtungen der Linkanalyse

Innerhalb der Informationswissenschaft gibt es gegenwärtig mehrere vielversprechende Forschungsaufgaben für die Webometrie. Im Folgenden werden fünf von ihnen kurz abgehandelt, bevor eine sechste, die Linkanalyse von Blogs, detaillierter beschrieben wird:

- Die Untersuchung der Motive für das Einrichten von Links, speziell im akademischen Kontext. Einige solcher Studien gibt es bereits (Bar-Ilan, 2004b, 2004c; Harries et al., 2004; Wilkinson et al., 2003). Ihre Ergebnisse deuten jedoch auf eine große Vielfalt von Motiven hin, welche die Klassifikation der Verweise erschwert. Dennoch wäre es interessant und hilfreich für die webometrische Forschung, mehr über die Unterschiede zwischen den verschiedenen Motiven für das Einfügen von Links zu wissen.



www.knowledgespeak.com
Der tägliche Newsletter für die
Fachinformationsanbieter im Bereich STM



Weitere Dienstleistungen:
blog area, Veranstaltungskalender, Interviews,
white papers, Adressentool im Bereich STM

Gratis! Jetzt testen.

- Zeitreihenanalysen. Ein der Linkanalyse immanentes Problem ist, dass sich das Web kontinuierlich weiterentwickelt und jede akademische Studie darüber folglich bereits zum Zeitpunkt ihrer Publikation veraltet sein kann. Daher ist es äußerst wichtig zu wissen, wie die Ergebnisse aller Arten von Linkanalysen über die Zeit variieren. Eine niedrige Variation würde die „Haltbarkeit“ webometrischer Ergebnisse erhöhen.
- Die Anwendung von Maßen der „Social Network Analysis“ auf Informationssammlungen. Björneborn (2004) zufolge kann diese Art von Forschung auf Informationen im Web (z.B. Newsgroups oder Gruppen von Websites) angewendet werden, um Erkenntnisse über die Struktur dieser Informationen und über die Gruppierungen zu gewinnen, die sie zusammenstellen (wie z.B. „invisible colleges“; Caldas, 2003). Dabei ist zu berücksichtigen – möglicherweise in jeder Studie auf spezifische Weise –, dass das Einrichten von Links nicht obligatorisch ist: Das Fehlen von Verweisen zwischen zwei Websites oder zwei einzelnen Seiten bedeutet nicht notwendigerweise, dass sie keinen gemeinsamen Bezug aufweisen.
- Die Analyse wiki-basierter Informationssammlungen. Die Enzyklopädie Wikipedia ist ein Beispiel für eine web-spezifische kooperative Sammlung von Informationen. Ihre Websites in verschiedenen Sprachen weisen interessante interne und externe Linkstrukturen auf. Ihre Erforschung erscheint vielversprechend, steht jedoch erst am Anfang (vgl. Voß, 2005).
- Die Unterstützung breiter sozialwissenschaftlicher Forschung. Da das Web kein exklusiv akademischer Raum ist, kann es bei einem breiten Spektrum sozialwissenschaftlicher Fragestellungen herangezogen werden, und zwar sowohl als eigenständiger Forschungsgegenstand (z.B. zur Untersuchung von Online-Communities) als auch als leicht zugängliche Informationsquelle für Offline-Phänomene, die sich im Web widerspiegeln.

Linkanalysen von Blogs

Als Weblogs (kurz: Blogs) werden Online-Tagebücher bezeichnet, die Millionen von Autoren im Web verfassen (BBC, 2005; Nardi, Schiano, Gumbrecht & Swartz, 2004). Sie zeichnen sich besonders durch ihren einfachen Gebrauch aus: Auch Personen, die wenig Erfahrung mit dem Erstellen von Websites besitzen und nicht über Detailkenntnisse verfügen, können ein attraktives Blog einrichten und führen. Diese Tagebücher sind enorme Speicher von Informationen unterschiedlicher Qualität. Häufig werden sie von Webnutzern gepflegt, die auf diesem Weg regel-

mäßig aktualisiertes Expertenwissen zu einem bestimmten Thema weitergeben wollen. Solche Blogs nehmen die Rolle spezialisierter Newsletter ein (Bar-Ilan, 2004a). Ein Beispiel dafür ist „The Shifted Librarian“, das unter www.theshiftedlibrarian.com zahlreiche Fakten mit Bezug zu Bibliotheken liefert. Sowohl innerhalb als auch zwischen Blogs gibt es intensive Linkstrukturen (Kumar, Novak, Raghavan & Tomkins, 2004; Marlow, 2004). Viele von ihnen bieten Besuchern die Option, Anmerkungen zu den Einträgen abzugeben. Außerdem ist es anderen Verfassern (sogenannten „Bloggern“) leicht möglich, Kommentare dazu in ihrem eigenen Tagebuch abzugeben und einen Link auf die URL des ursprünglichen Beitrags zu setzen. Eine weitere in diesem Zusammenhang interessante Besonderheit ist die sogenannte „Blogroll“, eine Linkliste mit ähnlichen oder empfohlenen Blogs. Da das Einrichten von Links zwischen den Tagebüchern so einfach und alltäglich ist, könnte es ein besonders vielversprechendes Medium für die Linkanalyse sein. Tatsächlich wird bereits anhand von Linkzahlen bereits eine täglich aktualisierte Liste der hundert beliebtesten Blogs ermittelt (unter www.blogstreet.com/top100.html). Es erscheint nicht unwahrscheinlich, dass diese Linkdaten auch analog zur Zitationsanalyse verwendet werden können, da Blogger („Autoren“) viele Beiträge ins Netz stellen und Links die individuellen Beiträge („Dokumente“) in Blogs verbinden können. Allerdings gibt es auch Unterschiede: die fehlende Qualitätskontrolle der veröffentlichten Mitteilungen; die Tatsache, dass sie – anders als wissenschaftliche Publikationen – vermutlich selten einen zentralen Stellenwert im Berufsleben ihrer Autoren einnehmen; sowie der Mangel eines thematischen Rahmens, wie ihn Zeitschriften ihren Artikeln geben. Dennoch sind die folgenden blogspezifischen Anwendungen der Linkanalyse zu erwarten:

- Listen der beliebtesten Blogs (die sinngemäße Entsprechung der meist zitierten Autoren)
- Listen der beliebtesten einzelnen Blog-Einträge (meist zitierte Artikel)
- Beziehungsanalyse/Netzwerkdiagramme von Links zwischen Blogs (autorenbezogene Zitations-/Kozitationsdiagramme)
- Beziehungsanalyse/Netzwerkdiagramme von Links zwischen einzelnen Einträgen (artikelbezogene Zitationsdiagramme)

Abgesehen davon, was Blog-Linkanalysen in der Zukunft leisten können, ist es wichtig zu erörtern, zu welchen Zwecken sie sinnvollerweise eingesetzt werden können und wie sich die Möglichkeiten der Linkanalyse dabei am besten nutzen lassen. So besteht kein dringender Bedarf an einer evaluativen Linkanalyse von Blogs. Sie be-

sitzt definitiv nicht die Bedeutung, die die Auswertung von Zitationen für die Bewertung wissenschaftlicher Produktivität innehat: Zwar könnte sie in gewisser Hinsicht nützlich sein, sie kann jedoch keinesfalls dazu beitragen, die Verteilung staatlicher Forschungsgeldern zu steuern. Auswertungen der Linkstrukturen von Blogs werden in erster Linie der sozialwissenschaftlichen Forschung weiterhelfen, indem sie Informationen zum Phänomen des Bloggens und Daten zur Verbreitung einzelner Themen liefern. Beispielsweise könnten Politikwissenschaftler die Perzeption von Kanzlerduellen in Blogs verfolgen. Noch allgemeiner können Linkanalysen dazu beitragen, die Diffusion von Informationen im Blograum zu analysieren, indem sie sich über verschiedenste Themen hinweg ausbreitende Muster aufdecken (Gill, 2004; Gruhl, Guha, Liben-Nowell & Tomkins, 2004). Für die themenzentrierte Linkanalyse von Blogs ist vermutlich eine Art von Textanalyse erforderlich, um thematisch relevante Tagebücher oder Einträge zu identifizieren, sowie ein Programm zum teilautomatischen Sammeln und Filtern von Blogdaten.

Fallstudie: „The Shifted Librarian“

Das Blog „The Shifted Librarian“ ist das Ziel zahlreicher Links: AltaVista meldete am 27. April 2006 186.000 Inlinks auf beliebige Seiten der gesamten Website, wenn man interne Links – also solche, die aus dem Blog selbst stammen – ausschließt („linkdomain:www.theshiftedlibrarian.com NOT domain:theshiftedlibrarian.com“). Zum Vergleich: Die Website der Universität Cambridge zog AltaVista zufolge zum gleichen Zeitpunkt nur 93.000 externe Links auf sich („linkdomain:www.cam.ac.uk NOT domain:cam.ac.uk“). Die hohe Zahl von Verweisen auf diese Bibliothekars-Website resultiert aus der Struktur von Blogs und aus der Tatsache, dass sie auf einer Datenbank beruhen, einer Art einfachem Content-Management-System. Dies bedeutet, dass sich Links leicht über alle Seiten des Blogs hinweg replizieren lassen. Wenn der „Shifted Librarian“ zu einer der Lieblingsseiten eines produktiven Bloggers avanciert, könnte er seiner „Blogroll“ einen Link darauf hinzufügen, der dann automatisch auf allen weiteren Seiten des Tagebuchs erscheint. Auf diese Weise entstehen unter Umständen zehntausende von Links mit einem einzigen Tastendruck. Ähnliche Phänomene wurden bereits in anderen Bereichen des Webs beobachtet, z.B. bei der automatischen Wiederholung einer Linkleiste auf allen Seiten einer Website (Thelwall, 2002). Im Blogbereich scheint dies jedoch wesentlich verbreiteter zu sein. Dies ist für die Linkanalyse insofern problematisch, als die Linkwerte in der Folge eher die Wiederholungsmuster ihrer Quell-Blogs als das Ansehen des Ziel-Blogs wiedergeben.

Die Lösung könnte sein, die betreffenden Internetauftritte mit Hilfe eines speziellen Webcrawlers zu erfassen und die gefundenen Links nach Websites zu aggregieren, so dass jeweils nur ein Link zwischen zwei Blogs gezählt wird und alle weiteren ignoriert werden.

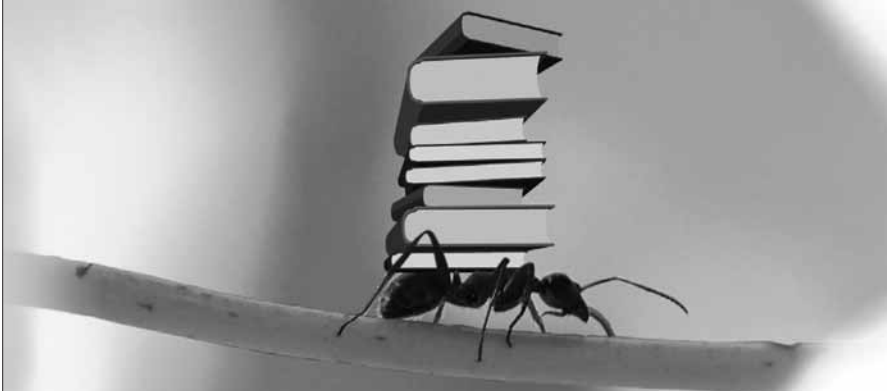
Schlussfolgerungen

Es gibt noch einen großen Spielraum für weitere methodologische Forschung zur Klärung der Fragen, wie Links am besten zu analysieren sind und in welchem Zusammenhang ihre Auswertung am nützlichsten ist. Diese Art von Studien sollten von Wissenschaftlern durchgeführt werden, deren fachliches Fundament im Bereich Information und speziell in der Zitationsanalyse liegt.

Die größte Stärke der Linkanalyse ist die breite Vielfalt von Kontexten, in denen sie verwendet werden kann, d.h. sie kann nicht nur in ihrem fachlichen Herkunftsbereich, der Informationswissenschaft, sondern auch in den Sozialwissenschaften im allgemeinen helfen, das World Wide Web zu reflektieren und webbezogene Probleme zu lösen. Die Anwendung der Linkanalyse auf Blogs erscheint in dieser Hinsicht besonders vielversprechend, da der soziale Hintergrund von Bloggern vermutlich sehr vielfältig ist. Eine zweite Stärke ist, dass sowohl die Daten als auch die Erhebungsinstrumente (Suchmaschinen oder Webcrawler) kostenlos verfügbar und einfach zu handhaben sind. Diese beiden Vorzüge machen die Linkanalyse zu einem praktischen neuen Forschungswerkzeug.

Eine grundsätzliche Schwäche der Linkanalyse ist, dass das Einrichten von Links ein unsystematisches Phänomen ist, das zum Teil von Faktoren abhängt, die für die meisten Forschungsfragen nicht von Interesse sein dürften (z.B. von dem jeweiligen Design, das einzelne Autoren für ihre Website wählen). In größeren Untersuchungen gleichen sich diese Art von Entscheidungen häufig aus, was bei kleineren Studien jedoch nicht notwendigerweise der Fall ist. Folglich müssen, sofern nicht die Verweise selbst der Forschungsgegenstand sind, Linkdaten von kleineren Websites meist per Methodentriangulation mit anderen Datenquellen, wie z.B. Zitationen, kombiniert werden. Die Nutzung der Linkanalyse erscheint insbesondere in interdisziplinäre Projekten zwischen Informations- und anderen Sozialwissenschaftlern naheliegend: Die Erschließung webbasierter Informationen im Kontext verschiedenster sozialwissenschaftlicher Fragestellungen ist eine interessante neue Rolle für die Informationswissenschaft (Beaulieu, 2004; Wouters, 2000).

Produktivität die begeistert!



LIDOS
Der Name
für produktive
Literaturarbeit.

**Einzelplatz,
Netzwerk, Intranet
und Internet**

Literatur und ähnliche
Dokumente erfassen,
downloaden,
archivieren, verwalten,
auswerten und nutzen,
dokumentieren und
publizieren.

Infos im Netz: www.land-software.de oder bei
LAND Software-Entwicklung,
Postfach 1126, 90519 Oberasbach,
Fax 0911-695173, info@land-software.de



Literatur

Adamic, L. A. & Adar, E. (2003): Friends and neighbors on the Web. *Social Networks*, 25(3), 211-230.

Almind, T. C. & Ingwersen, P. (1997): Informetric analyses on the World Wide Web: Methodological approaches to 'webometrics'. *Journal of Documentation*, 53(4), 404-426.

An, L.; Qiu, J. P. (2004): Research on the relationships between Chinese journal impact factors and external web link counts and web impact factors. *Journal of Academic Librarianship*, 30(3), 199-204.

Bar-Ilan, J. (2001): Data collection methods on the Web for informetric purposes - A review and analysis. *Scientometrics*, 50(1), 7-32.

Bar-Ilan, J. (2004a): Blogarians - A new breed of librarians. In: Proceedings of the 67th Annual Meeting of the ASIST, pp.119-128, Medford: Information Today.

Bar-Ilan, J. (2004b): A microscopic link analysis of academic institutions within a country - the case of Israel. *Scientometrics*, 59(3), 391-403.

Bar-Ilan, J. (2004c): Self-linking and self-linked rates of academic institutions on the Web. *Scientometrics*, 59(1), 29-41.

BBC (2005): Blog reading explodes in America. Available at: <http://news.bbc.co.uk/1/hi/technology/4145191.stm> (letzter Zugriff: 11.10.06).

Beaulieu, A. (2004): From brainbank to database: the informational turn in the study of the brain. *Studies in History and Philosophy of Science Part C: Studies in Biological and Biomedical Sciences*, 35(2), 367-390.

Beaulieu, A. (2005): Sociable hyperlinks: An ethnographic approach to connectivity. In: C. Hine (Hrsg.): *Virtual Methods: Issues in Social Research on the Internet*, pp. 183-197. London: Berg.

Björneborn, L. (2004): Small-world link structures across an academic Web space - a library and information science approach. Kopenhagen: Royal School of Library and Information Science.

Björneborn, L. & Ingwersen, P. (2001): Perspectives of webometrics. *Scientometrics*, 50(1), 65-82.

Björneborn, L. & Ingwersen, P. (2004): Towards a basic framework for webometrics. *Journal of the American Society for Information Science and Technology*, 55(14), 1216-1227.

Burnett, R. & Marshall, P. (2002): *Web theory: An introduction*. London: Routledge.

Caldas, A. (2003): Are newsgroups extending 'invisible colleges' into the digital infrastructure of science? *Economics of Innovation and New Technology*, 12(1), 43-60.

Egghe, L. (2000): New informetric aspects of the Internet: Some reflections – many problems. *Journal of Information Science*, 26(5), 329-335.

Foot, K.; Schneider, S.; Dougherty, M.; Xenos, M. & Larsen, E. (2003): Analyzing linking practices: Candidate sites in the 2002 US electoral web sphere. *Journal of Computer Mediated Communication*, 8(4). Available at: <http://jcmc.indiana.edu/vol8/issue4/foot.html> (letzter Zugriff: 11.10.06).

Fuchs, S. (1992): *The professional quest for truth: A social theory of science and knowledge*. Albany (NY): SUNY Press.

Garrido, M. & Halavais, A. (2003): Mapping networks of support for the Zapatista movement: Applying Social Network Analysis to study contemporary social movements. In: M. McCaughey & M. Ayers (Hrsg.): *Cyberactivism: Online activism in theory and practice*, pp.165-184. London: Routledge.

Gill, K. E. (2004): How can we measure the influence of the blogosphere? Vortrag im WWW2004 Workshop on the Blogging Ecosystem: Aggregation, Analysis and Dynamics. Available at: http://faculty.washington.edu/kegill/pub/www2004_blogosphere_gill.pdf (letzter Zugriff: 11.10.06).

Granovetter, M. (1973): The strength of weak ties. *American Journal of Sociology*, 78(6), 1360-1380.

Gruhl, D.; Guha, R.; Liben-Nowell, D. & Tomkins, A. (2004): Information diffusion through blogspace. *Proceedings of the 13th International World Wide Web Conference* (pp. 491-501). New York: ACM.

Harries, G.; Wilkinson, D.; Price, E.; Fairclough, R. & Thelwall, M. (2004): Hyperlinks as a data source for science mapping. *Journal of Information Science*, 30(5), 436-447.

Harter, S. & Ford, C. (2000): Web-based analysis of e-journal impact: Approaches, problems, and issues. *Journal of American Society for Information Science*, 51(13), 1159-1176.

Heimeriks, G.; Hörlesberger, M. & van den Besselaar, P. (2003): Mapping communication and collaboration in heterogeneous research networks. *Scientometrics*, 58(2), 391-413.

Hine, C. (2000): *Virtual Ethnography*. London: Sage.

Ingwersen, P. (1998): The calculation of Web Impact Factors. *Journal of Documentation*, 54(2), 236-243.

Kling, R. & McKim, G. (2000): Not just a matter of time: Field differences and the shaping of electronic media in supporting scientific communication. *Journal of the American Society for Information Science*, 51(14), 1306-1320.

Kumar, R.; Novak, J.; Raghavan, P. & Tomkins, A. (2004): Structure and evolution of blogspace. *Communications of the ACM* 47(12), 35-39.

Larson, R. R. (1996): *Bibliometrics of the World Wide Web: an exploratory analysis of the intellectual structure of cyberspace*. Vortrag auf dem 59th Annual Meeting of the American Society for Information Science. Available at: <http://sherlock.berkeley.edu/asis96/asis96.html> (letzter Zugriff: 11.10.06).

Li, X.; Thelwall, M.; Musgrove, P. B. & Wilkinson, D. (2003): The relationship between the WIFs or in-links of computer science departments in UK and their RAE ratings or research productivities in 2001. *Scientometrics*, 57(2), 239-255.

Li, X.; Thelwall, M.; Wilkinson, D. & Musgrove, P. B. (2005a): National and international university departmental web site interlinking. Part 1: Validation of departmental link analysis. *Scientometrics*, 64(2), 151-185.

Li, X.; Thelwall, M.; Wilkinson, D. & Musgrove, P. B. (2005b): National and international university departmental web site interlinking. Part 2: Link patterns. *Scientometrics*, 64(2), 187-208.

Marlow, C. (2004): Audience, structure and authority in the weblog community. Vortrag auf der International Communication Association Conference. Available at: <http://web.media.mit.edu/~cameron/cv/pubs/04-01.pdf> (letzter Zugriff: 11.10.06).

Mayr, P. & Tosques, F. (2005): Webometrische Analysen mit Hilfe der Google Web APIs. *Information – Wissenschaft und Praxis*, 56(1), 41-48.

Nardi, B. A.; Schiano, D. J.; Gumbrecht, M. & Swartz, L. (2004): Why we blog. *Communications of the ACM*, 47(12), 41-46.

Otte, E. & Rousseau, R. (2002): Social network analysis: a powerful strategy, also for the information sciences. *Journal of Information Science*, 28(6), 441-453.

Park, H. W. (2003): Hyperlink network analysis: A new method for the study of social structure on the web. *Connections*, 25(1), 49-61.

Park, H. W. & Thelwall, M. (2003): Hyperlink analyses of the world wide web: A review. *Journal of Computer-Mediated Communication*, 8(4). Available at: <http://jcmc.indiana.edu/vol8/issue4/park.html> (letzter Zugriff: 11.10.06).

Rodriguez i Gairin, J. M. (1997): Valorando el impacto de la información en Internet: AltaVista, el 'Citation Index' de la Red. *Revista Española de Documentación Científica*, 20(2), 175-181.

Rogers, E. M. & Kincaid, D. L. (1981): *Communication networks: Toward a new paradigm for research*. New York: Free Press.

Rousseau, R. (1997): Situations: an exploratory study. *Cybermetrics* (1). Available at: www.cindoc.csic.es/cybermetrics/articles/v1i1p1.html (letzter Zugriff: 11.10.06).

Smith, A. G. (1999): A tale of two web spaces: Comparing sites using Web Impact Factors. *Journal of Documentation*, 55(5), 577-592.

Snyder, H. W. & Rosenbaum, H. (1999): Can search engines be used for Web-link analysis? A critical review. *Journal of Documentation*, 55(4), 375-384.

Tang, R. & Thelwall, M. (2003): US academic departmental web-site interlinking in the United States disciplinary differences. *Library and Information Science Research*, 25(4), 437-458.

Tang, R. & Thelwall, M. (2006) (im Erscheinen): A hyperlink analysis of US public and academic libraries' Web sites. *Library Quarterly*.

Thelwall, M. (2002): Conceptualizing documentation on the Web: An evaluation of different heuristic-based models for counting links between university web sites. *Journal of American Society for Information Science and Technology*, 53(12), 995-1005.

Thelwall, M. (2004): *Link analysis: An information science approach*. San Diego et al.: Elsevier Academic Press.

Thomas, O. & Willet, P. (2000): Webometric analysis of departments of librarianship and information science. *Journal of Information Science*, 26(6), 421-428.

van Raan, A.F.J. (2001): *Bibliometrics and Internet: Some observations and expectations*. *Scientometrics*, 50(1), 59-63.

Voß, J. (2005): Measuring Wikipedia. In: P. Ingwersen & B. Larsen, eds., *Proceedings of ISSI 2005* (pp. 221-231). Stockholm: Karolinska University Press.

Wasserman, S. & Faust, K. (1994): *Social network analysis: Methods and applications*. Cambridge et al.: Cambridge University Press.

Watts, D.J. & Strogatz, S. H. (1998): Collective dynamics of 'small-world' networks. *Nature*, 393, 440-442.

Wilkinson, D.; Harries, G.; Thelwall, M. & Price, E. (2003): Motivations for academic Web site interlinking: Evidence for the Web as a novel source of information on informal scholarly communication. *Journal of Information Science*, 29(1), 49-56.

Wouters, P. (2000): *Cyberscience: The informational turn in science*. Vortrag an der Freien Universität Amsterdam.

Messung, Informationswissenschaft, Analyse, Verknüpfungsindikator, Bewertungsfaktor, Zitierhäufigkeit, Informationsnetz, Sozialwissenschaften

DIE AUTOREN

Mike Thelwall



ist Leiter der Statistical Cybermetrics Research Group an der University of Wolverhampton in Großbritannien. Sein Forschungsschwerpunkt liegt auf der Linkana-

lyse, über die kürzlich ein Buch von ihm erschienen ist, sowie auf anderen Formen der Webometrie.

University of Wolverhampton
School of Computing and Information Technology
Wulfruna Street
Wolverhampton WV1 1SB
Großbritannien
m.thelwall@wlv.ac.uk

Tina Ruschenburg



ist Diplom-Soziologin und Doktorandin am Institut für Wissenschafts- und Technikforschung (IWT) der Universität Bielefeld. Sie beschäftigt sich mit der Frage, wel-

chen Beitrag die Linkanalyse zur Erforschung internationaler Strukturen der Wissenschaft leisten kann.

Universität Bielefeld
Institut für Wissenschafts- und Technikforschung
Postfach 100131
33501 Bielefeld
tina.ruschenburg@uni-bielefeld.de