

## Google Scholar und Windows Live Academic Search – aktuelle Entwicklungen bei wissenschaftlichen Suchmaschinen

Konstanze Söllner

### 1. Das WWW als Quelle wissenschaftlicher Information und seine Erschließung durch allgemeine Suchmaschinen

Dass das World Wide Web als geeignete Quelle wissenschaftlicher Information angesehen werden kann, ist für viele Fächer inzwischen unumstritten. Weniger im Bewusstsein der Nutzer verankert ist die Erkenntnis, dass das Web nur in den Ausschnitten und in der Anordnung zur Kenntnis genommen werden kann, wie sie die Suchmaschinen präsentieren. Dieser Aspekt ist vor allem durch die allgegenwärtigen Rankingverfahren der führenden Suchmaschinenanbieter in den Hintergrund getreten. Dagegen ist die Suche in Internetkatalogen, die offensichtlich nur Ausschnitte und eine unmittelbar nachvollziehbare Anordnung präsentieren, seit dem Siegeszug von Googles PageRank unpopulär geworden. Rankingverfahren, Abdeckung und Aktualität eines Suchmaschinenindex' sind nicht nur Geschäftsgeheimnis der Betreiber, sondern angesichts ihres scheinbar reibungslosen Funktionierens für den normalen Nutzer uninteressant. Dass auch Kenntnisse einer strukturierten Suche oder Suchsyntax für das textbasierte Web meistens unnötig oder inadäquat sind, lässt die Websuche besonders einfach erscheinen – im Unterschied zu den Suchmasken und Syntaxkonventionen der Bibliotheks-OPACs und vieler Datenbankanbieter.<sup>1</sup> Diese Einfachheit wird durch installierte Suchwerkzeuge à la Google Toolbar bzw. fest integrierte Suchschlitze wie im Mozilla-Browser Firefox bzw. im Internet Explorer 7.0 zum Standard für die großen Suchmaschinen, und durch entsprechende Such-Plugins auch für andere Suchen.<sup>2</sup> Ein besonderer Wettbewerbsvorteil der Suchmaschinen gegenüber den Datenbankanbietern, der sich auch nicht durch Anpassungen beim Design kompensieren lässt, ist die Kostenfreiheit der Suche. Darüber hinaus erfüllen sie eine wesentliche Anforderung an sehr viele wissenschaftliche Themenstellungen: sie sind strikt interdisziplinär. Damit machen Sie aus der Not eine Tugend, denn eine fachliche Auswahl oder intellektuelle Auswertung findet in der Regel nicht statt. Die Defizite der großen Suchmaschinen liegen aus wissenschaftlicher Sicht betrachtet aber auch genau in diesem Bereich: unspezifische oder thematische

1 Die Generalüberholung der führenden medizinischen Literaturdatenbank Pubmed (National Library of Medicine) im Januar 2006 orientierte sich am Design der Suchmaschinen von MSN, A9 und Google.

2 Auch alle Arten von Toolbars bzw. Suchschlitzen, die als White Label Toolbars für unterschiedlichste Kundenwünsche angepasst werden, können schon aus Platzgründen keine strukturierte Suche anbieten.

Suchen führen zu Massen populärer Websites, nicht aber zum wissenschaftlichen Content, insbesondere etwa in der Medizin. Das Dilemma des häufig schlechten Rankings wissenschaftlicher Inhalte ist einer der Ausgangspunkte für den Ruf nach wissenschaftlichen Suchmaschinen.

## 2. Das Academic Invisible Web

Das so genannte Academic Invisible Web ist Teil des „Deep Web“. Es umfasst die wissenschaftlichen Inhalte, die von regulären Suchmaschinen nicht erfasst werden können – in erster Linie die Inhalte, die in über das Web erreichbaren Datenbanken gespeichert sind, und nur als Ergebnis interaktiver Suchabfragen dynamisch erzeugt werden. Der Hauptteil dieser Inhalte stammt aus Verlagsdatenbanken, die nur Lizenznehmern zugänglich sind, und nicht aus Open-Access-Sammlungen. Der Begriff „Invisible Web“ wurde von Sherman und Price geprägt<sup>3</sup> und mehrfach von dem Suchmaschinen-Spezialisten Dirk Lewandowski als „Academic Invisible Web“ in die Diskussion gebracht<sup>4</sup>. Als „Academic Invisible Web“ formuliert er den Anspruch, auch die hinter Verlagsportalen verborgenen wissenschaftlichen Inhalte unbeschränkt recherchierbar und auffindbar zu machen. Dieser Nachweis sollte nicht wie in Bibliotheks-OPACs auf bestimmte Dokumentarten oder den Nachweis von Sucheinstiegen beschränkt sein, sondern ebenso wie die allgemeine Websuche potentiell jedes Dokument auffindbar machen, selbstständige ebenso wie unselbstständige Veröffentlichungen, Working Papers, Preprints, Datensammlungen usw. Und dies auch in neu entstehenden Veröffentlichungsformen wie Science-Wikis oder Science-Blogs. Die Erschließung des Invisible Web gewinnt für das Academic Web daher so besondere Bedeutung, dass nicht nur sehr viele wissenschaftlichen Inhalte nur im Deep bzw. Invisible Web auffindbar sind, sondern dass es sich bei den wissenschaftlichen Anteilen des Invisible Web meist um qualitativ besonders hochwertige Datenbankinhalte handelt.

## 3. Die Erschließung des Academic Web durch wissenschaftliche Suchmaschinen

Da wissenschaftliche Inhalte besondere Anforderungen an Suchmaschinen in Bezug auf Erschließung des Deep Web und Ergebnisranking stellen, entwickeln nun auch große Anbieter allgemeiner Web-Suchdienste wissenschaftliche Suchmaschinen. Seit kurzer Zeit ist der Markt wieder in Bewegung geraten durch die

---

3 Sherman, Chris; Price, Gary: The Invisible Web. Medford NJ : Information Today, 2001.

4 Lewandowski, Dirk; Mayr, Philipp: Exploring the Academic Invisible Web. In: Library Hi Tech 24 (2006) 4 [Erscheinen angekündigt]. Preprint unter [http://www.durchdenken.de/lewandowski/doc/LHT\\_Preprint.pdf](http://www.durchdenken.de/lewandowski/doc/LHT_Preprint.pdf) [Stand: 8.6.2006]

von Microsoft präsentierte Beta-Variante seiner Windows Live Academic Search<sup>5</sup>. Die konkurrierende akademische Suche Google Scholar (ebenfalls Beta) ging zwar schon am 18.11.2004 an den Start, erhielt aber auf der Google-Hauptseite bislang keine eigene Präsenz, weder durch eine Ankündigung noch etwa einen eigenen *tab* oberhalb des Suchfeldes, und wurde recht versteckt unter einer eigenen URL <http://scholar.google.com> präsentiert. Von vielen Informations- und Bibliotheksspezialisten wurde Google Scholar im Gegensatz zur Google Book Search kaum wahrgenommen, und damit einmal mehr das Academic Web als Quelle wissenschaftlicher Information<sup>6</sup> ignoriert. Die Begeisterung der Open-Access-Aktivisten war naturgemäß sehr viel größer, doch konnten die hohen Erwartungen an eine weitgehende Erschließung der auf Open-Access-Servern vorgehaltenen Literatur von Google Scholar bis jetzt nicht eingelöst werden.<sup>7</sup> Dies gilt in noch stärkerem Maße für die Windows Live Academic Search, wobei Microsoft im Unterschied zu Google erklärtermaßen seine Quellen offen legen will.<sup>8</sup> Hier war die Euphorie nach den Erfahrungen mit dem mangelhaften Nachweis von Open-Access-Dokumenten allerdings von vornherein gedämpfter.

### 3.1 Google Scholar

Google Scholar hat ebenso wie die Windows Academic Live Search das CrossRef Search Pilot Project der Verlagsorganisation CrossRef<sup>9</sup> zum Ausgangspunkt, die zugleich auch offizielle Registrierungsstelle für Digital Object Identifiers (DOI) ist. Nach Aussage des bei Google für Verlags- und Bibliothekskooperationen und „Content acquisition“ zuständigen Managers John Lewis Needham basiert Goog-

- 5 <http://academic.live.com> leitet derzeit noch auf <http://www.live.com> weiter. Live.com nutzt derzeit noch das alte Design-Konzept, soll aber schon bald den neuen Header „Windows Live Flare“ erhalten. Dieser wird auch in der aktuellen Betaversion des Windows Live Messengers eingesetzt.
- 6 Lossau Norbert: Search Engine Technology and Digital Libraries, Libraries Need to Discover the Academic Internet. In: D-Lib Magazine 10 (2004) 6.
- 7 Mayr, Philipp; Walter, Anne-Kathrin: Google Scholar – Wie tief gräbt diese Suchmaschine? In: In die Zukunft publizieren: Herausforderungen an das Publizieren und die Informationsversorgung in den Wissenschaften. Bonn : Informationszentrum Sozialwissenschaften, 2005, S. 214–262.  
Mayr, Philipp; Walter, Anne-Kathrin: Abdeckung und Aktualität des Suchdienstes Google Scholar. In: Information – Wissenschaft und Praxis 57 (2006) 3.
- 8 Die wegen fehlender Normierung an Mehrfachnennungen reiche Liste ist zu finden unter <http://academic.live.com/journals>. Da von der Domain <http://academic.live.com> aktuell ein Redirect auf <http://www.live.com> besteht, nur konsultierbar über den Google Cache [Stand: 8.6.2006]
- 9 <http://www.crossref.org/>

le Scholar auf folgenden Grundlagen: „GS principles: inclusive, indexes all forms of articles/information, ranks works not instances, ranks in the way that researchers do (on the basis of relevancy, currency, authority), and automates citation extraction and normalization (process not perfect but enables a powerful scale for users)“.<sup>10</sup> Im April 2006 reagierte Google auf die Einführung der Microsoft-Suche mit Regionalausgaben von Google Scholar (darunter auch <http://scholar.google.de/>), der Sortierfunktion „recent articles“ und Exportmöglichkeiten für Literaturverwaltungsprogramme.

### 3.1.1 Abdeckung

Google Scholar erhebt einen gewissen Anspruch auf Vollständigkeit, d.h. die Suchmaschine will wenigstens einen Verweis oder ein Zitat, möglichst ein Abstract liefern. Deshalb werden Treffer in vier Qualitäten angeboten: *PDF/PS/DOC* = Volltext, *Link* = i.d.R. Abstract-Level (für Lizenznehmer der Zeitschrift oder bei Open Access auch Volltext), *Citation* = Offline-Nachweis von Aufsätzen und Büchern, *Books* = Treffer aus der Google Book Search (z.T. Offline-Nachweis). Der inhaltliche Schwerpunkt liegt auf den Zeitschriften der großen Verlage, also in den STM-Fächern und bei englischsprachigen Dokumenten. In den Geisteswissenschaften trifft man auf sehr viele *Citation*-Treffer, die aus dem Volltext anderer Veröffentlichungen extrahiert wurden. Diese lediglich zitierten Dokumente sind folglich auch nicht volltextindiziert. Die Größe des Google-Scholar-Index' ist unbekannt. Schätzungen belaufen sich auf bis zu 500 Millionen Objekte. Von Seiten des Unternehmens gibt es auch keinerlei Informationen zu den abgedeckten Verlags- bzw. Open-Access-Servern, wie es bei Datenbank Anbietern üblich ist.<sup>11</sup> Die Kritik daran verkennt allerdings, dass mit zunehmender Integration der Sozial- und Geisteswissenschaften und der Treffer aus der Buchsuche die Dimensionen einer Liste ohnehin gesprengt wären. Zudem finden Inhalte nicht beteiligter Verlage wie etwa Elsevier auf anderen Wegen in den Index, etwa über die Pubmed-Version der Datenbank Medline. In einer aktuellen Untersuchung wird zur Wissenschaftlichkeit der Dokumente festgestellt: „the large of overlapping material (mit der Elsevier-Datenbank Scopus und ISI-Web of Science, K.S.) gives credence to the scholarly nature of the Google Scholar Database“<sup>12</sup>. In zeitlichen Abständen

10 Needham, John Lewis: Innovative search systems. ICSTI: 2006 Winter Meeting Workshop, [http://www.icsti.org/winter\\_mtg\\_2006/presentations/summaries/needham.html](http://www.icsti.org/winter_mtg_2006/presentations/summaries/needham.html). [Stand: 9.6.2006]

11 Eine grundlegende Untersuchung zum Abdeckungsgrad von Google Scholar findet sich in Mayr u. Walter 2005. Die Daten sind mittlerweile aber mehr als ein Jahr alt.

12 Bakkalbasi, Nisa; Bauer, Kathleen; Glover, Janis; Wang, Lei (2006): Three options for citation tracking: Google Scholar, Scopus and Web of Science, S. 13f. <http://eprints.rclis.org/archive/00006080/> [Stand: 9.6.2006]

wiederholte Suchen zeigen außerdem, dass der Google-Scholar-Index dramatisch im Wachsen begriffen ist.

### 3.1.2 Aktualität

Aktualitätsfaktoren spielen bei der Relevanzermittlung wissenschaftlicher Dokumente eine besondere Rolle. Die Indexinhalte von Google Scholar werden crawlerbasiert erhoben, und dies in fortlaufenden Update-Prozessen. Es konnte allerdings nachgewiesen werden, dass Google bei der Aktualisierung seines Hauptindex' zwar weit überwiegend aktueller ist als andere Suchmaschinen, dass es aber einen bestimmten Prozentsatz von Seiten gibt, die sehr lange nicht aktualisiert werden.<sup>13</sup> Dies trifft mit einiger Sicherheit auch auf den Google-Scholar-Index zu.

### 3.1.3 Such-, Treffer- und Rankingqualität, Usability

Google Scholar unterstützt wie Google keine konventionellen Wildcard-Suchen.<sup>14</sup> Über die einschlägigen Kategorien der Google-Kommandosprache hinaus (z.B. intitle:, inurl:, site: oder filetype:) ermöglicht Google Scholar strukturierte Suchen nach Autor, Quelle und Veröffentlichungsdatum, allerdings nicht nach dem Titel. Die wissenschaftliche Variante der Suchmaschine Google arbeitet offensichtlich mit einer Erweiterung des Google-PageRank: „Google Scholar ordnet Ihre Suchergebnisse nach Relevanz an. So wie bei der Webseitensuche mit Google werden die nützlichsten Verweise oben auf der Seite angezeigt. Die Ranking-Technologie von Google berücksichtigt den vollständigen Text eines Artikels, den Autor, wo der Artikel veröffentlicht wurde und wie oft der Text in der wissenschaftlichen Literatur zitiert wurde.“<sup>15</sup> Die Such-, Treffer- und Rankingqualität steht und fällt mit der Qualität der automatisch erzeugten Metadaten. So lassen nicht nur Extraktion und Normierung der Autorennamen bisweilen zu wünschen übrig, weswegen offenbar bis jetzt auf die Verlinkung der Autoren verzichtet wurde. Die Google-Scholar-Suche reagiert auf diese Fehler aufgrund der obligatorischen Volltextindizierung aber ausgesprochen robust. Die Gruppierung der Versionen eines Artikels ist häufig mangelhaft, wenn eine Version nur eine Offline-Citation mit besonders fehlerbehafteten Metadaten ist. Für die Suche nach einem kostenlosen Preprint oder einer Open Access-Version ist die Gruppierungsfunktion trotzdem in jedem Fall sehr nützlich. Fehlerhafte Metadaten sind aber dann besonders ärgerlich,

13 Lewandowski, Dirk: Aktualität als erfolgskritischer Faktor bei Suchmaschinen. In: IWP-Information: Wissenschaft und Praxis 57 (2006) 3, 141–148, S. 142f.

14 abgesehen vom \*-Platzhalter, der für ein vollständiges Wort zwischen zwei anderen Stichwörtern steht

15 <http://scholar.google.de/intl/de/scholar/about.html> [Stand: 9.6.2006]

wenn der Download in Zitierungsmanager wie BibTeX, EndNote oder RefWorks angeboten wird.

Die Integration in die Websuche erfolgt durch Übergabe des ersten Autorennachnamens und einer Wildcard-Phrase aus den ersten drei bedeutungstragenden Titel-Stichwörtern. Aus der Websuche heraus wird mit einem so genannten One-Box-Ergebnis aus der Scholar-Suche reagiert, wenn sich ein Treffer nicht im Hauptindex befindet.<sup>16</sup>

Die deutschsprachige Variante von Google Scholar ist im April mit dem Anspruch an den Start gegangen, deutschsprachige Inhalte vollständiger zu indexieren. Ein Unterschied in den Suchergebnissen, der auf getrennte Indices oder ein abweichendes Ranking hinweisen könnte, ist allerdings bisher nicht festzustellen.

Google Scholar wurde häufig angelastet, dass neueste Literatur, die naturgemäß weniger häufig gelinkt oder zitiert wird, im Ranking zu schlecht platziert ist. Das seit April mögliche Umsortieren der Treffer nach dem Kriterium „recent“ schafft allerdings kaum Abhilfe, da es offenbar das Suchergebnis nur auf die letzten fünf Jahre einschränkt, das Ranking aber beibehält. Damit werden zwei bis drei Jahre alte Publikationen als allerneueste Literatur präsentiert, ganz abgesehen davon, dass das Umsortieren per Hand der Google-Philosophie widerspricht, die auf Verbesserung des Rankings setzt.

### 3.1.4 Mehrwert

Google Scholar zeigt die Anzahl der Referenzpublikationen an. Dies eröffnet die Möglichkeit, im Schneeballsystem nach neuesten Publikationen zu suchen. Darüber hinaus scheint Google Scholar die ISI-Zitierungsindices wegen der Äquivalenz der Zitationszahlen ersetzen zu können.<sup>17</sup> Dank seiner potentiell breiteren Datenbasis könnte Google Scholar sogar belastbarere Werte liefern.

Eine Vielzahl von *Book*- Treffern aus Google Scholar sind inzwischen mit der Book Search verlinkt, obwohl Metadaten- und OCR-Qualität oft sehr mangelhaft sind. Da Google Inklusivität anstrebt, ist mit einer weiteren Integration von Google Scholar und Google Book Search zu rechnen.

An Google Scholars „Library Links Program“ zur Integration von Bestandslinks über bibliothekseigene Linkresolver in die Google-Anzeige nehmen in Deutschland nur eine Handvoll Bibliotheken teil. Seitdem Bestandsdaten zum regelmäßigen Harvesting durch Google Scholar angemeldet werden können, ist die

16 Dies funktioniert regelmäßig allerdings nur bei Dokumenten, die im Google-Scholar-Index, nicht aber im Hauptindex enthalten sind.

17 Pauly, Daniel; Stergiou, Konstantinos I.: Equivalence of results from two citation analyses: Thomson ISI's Citation Index and Google's Scholar Service. In: *Ethics in Science and Environmental Politics* 2005, S. 33–35.

umständliche Offline-Meldung der Holdings Geschichte, aber es bestehen auch grundsätzliche Bedenken, einer kommerziellen Suchmaschine Wettbewerbsvorteile zu verschaffen, und womöglich die eigene Klientel an einen Mitbewerber zu verlieren. Häufig wird zudem argumentiert, dass Googles schlechte Metadatenqualität Linkings verhindert, die lizenzrechtlich sehr wohl möglich wären. Diese Fehler würden der besitzenden Bibliothek angelastet werden, die jedoch keinen Einfluss auf die Qualität der Metadaten hat.

### 3.2 Windows Live Academic Search

Die im April von Microsoft in einer Beta-Version lancierte Windows Live Academic Search (WLAS) ist eine Komponente des designierten MSN-Nachfolgers Windows Live. Windows Live erscheint zur Zeit noch als Variante der verbreiteten Browserdesktops, vergleichbar Pageflakes oder Google Personalized Desktop. Die akademische Suche erfolgt über die Option „wissenschaftlich“ – von <http://academic.live.com> wird auf die Website <http://www.live.com> umgeleitet. Auch WLAS greift auf die CrossRef-Daten zurück.

#### 3.2.1 Abdeckung

Nach eigenen Aussagen umfasst der WLAS-Index 8 Mio. Artikel.<sup>18</sup> Microsoft hat die Abdeckung seiner akademischen Suche im Bereich der Verlagspublikationen beim Start im April offengelegt, eine Aktualisierung der Liste ist seitdem nicht erfolgt.<sup>19</sup> Angaben über die betroffenen Jahrgänge der Zeitschriften gibt es nicht. Darüber hinaus wird auch [arxiv.org](http://arxiv.org) indiziert. Inhaltlich liegt der Schwerpunkt derzeit auf Informatik, Elektrotechnik und Physik. Schon die einfache Stichwortsuche *theodor+hänsch* erzielt bei WLAS allerdings 83% weniger Treffer als bei Google Scholar – und dies, obwohl viele Artikel mehrfach gelistet sind (Stand: 10.6.2006).

#### 3.2.2 Aktualität

Was die Aktualität des Index' angeht, hält sich Microsoft ebenso bedeckt wie Google. Obwohl beide Anbieter mit CrossRef kooperieren, finden sich trotzdem aktuelle Artikel aus Verlagszeitschriften, die jeweils nur von einem der beiden

18 Weitere Informationen in den detailreichen Artikeln von Barbara Quint:  
Microsoft Offers Alternative to Google Scholar: Windows Live Academic Search  
<http://www.infotoday.com/newsbreaks/nb060417-1.shtml> [Stand: 9.6.2006]  
Windows Live Academic Search: The Details  
<http://www.infotoday.com/newsbreaks/nb060417-2.shtml> [Stand: 9.6.2006]  
Beide in: NewsBreaks Weekly News Digest vom 17.4.2006

19 <http://academic.live.com/journals>, wegen Redirect auf <http://www.live.com> nur über den Google Cache konsultierbar.

Dienste indiziert sind. Diese Lücke kann Google durch die integrierte Websuche wegen seines aktuelleren Hauptindex' besser schließen.

### 3.2.3 Such-, Treffer- und Rankingqualität, Usability

WLAS bietet zwar die Möglichkeiten, nach bestimmten Suchkategorien (Autor, Aktualität, Quelle) umzusortieren, aber keine strukturierte Suche. In Zusammenarbeit mit der wissenschaftlichen Suchmaschine und Zitationsdatenbank CiteSeer (für Informatik und Informationswissenschaft) wurden die Autorennamen und zitierten Dokumente verlinkt. Allerdings gibt es im Unterschied zu CiteSeer keine Vorwärts-/Rückwärtsverlinkung, d.h. es können weder die zitierenden Publikationen noch deren Anzahl bestimmt werden. Eine Extraktion der Metadaten der zitierten Dokumente wie bei Google Scholar (*Citation*) findet nicht statt.

Dafür wartet WLAS mit dem typischen Web 2.0-Design von Windows Live auf. Dank Ajax-Technologie (Asynchronous JavaScript and XML) können Ergebnislisten endlos gescrollt werden. Microsoft hat außerdem die Entwicklung von so genannten *gadgets* freigegeben. Solche gadgets können u.a. auch beliebige Suchmakros sein, die unter <http://gallery.microsoft.com/> eingestellt und abgerufen werden können. Diese Suchmakros sind mit einem eigenen *tab* analog zur Websuche ausführbar. Bisher können die auf Makros basierenden Suchaufträge allerdings noch nicht zufrieden stellend auf dem Windows Live Browserdesktop als Alert abgelegt werden, ebenso wenig wie wissenschaftliche Suchaufträge. Mit dieser vielversprechenden Funktionalität, die für die Web-, News- und Feed-Suche bereits aktiviert ist, könnten künftig beliebige Suchaufträge aus dem Academic Web abonniert werden, wie es jetzt schon Standard für viele Datenbanken und OPACs ist.

### 3.2.4 Mehrwert

WLAS bietet wie beschrieben potentiell die Möglichkeit, über das übliche Abonnieren fertiger RSS-Feeds hinaus suchbasierte Feeds aus dem Academic Web zu erhalten. Diese weitgehende Personalisierung kann aber nur dann zu einem echten Mehrwert werden, wenn auch ausreichende Abdeckung und hohe Aktualität des Index gewährleistet sind.

Analog zu Google Scholar plant auch Microsoft die Integration von Bibliotheks-Links über die marktüblichen oder selbstentwickelten Linkresolver. Und ebenso wie Google hat auch Microsoft ein Books Publisher Program<sup>20</sup> gestartet und plant eine Windows Live Book Search. Partner beim Books Publisher Program sind die Bibliotheken der University of California und University of Toronto. Auch erklärte Microsoft, der Open Content Alliance beitreten zu wollen, die als Konkurrenzun-  
20 <http://publisher.live.com/>

ternehmen zur Google Book Search gesehen werden kann. Damit signalisierte Microsoft, bei der Digitalisierung im Unterschied zu Google die Rechte der Urheber strikt wahren zu wollen.

#### **4. Bibliotheken als Mitbewerber bei der Erschließung des Academic Invisible Web**

Mit der Öffnung ihrer Server für systematisches Crawling haben die großen Wissenschaftsverlage den wissenschaftlichen Suchmaschinen Zugang zu Volltexten und Metadaten verschafft, die Bibliotheken und Forschungseinrichtungen samt proprietärer Suche teuer bei Datenbank Anbietern einkaufen. Zwar bieten die Suchmaschinen noch nicht sämtliche Funktionalitäten, wie man sie von den Datenbanken gewohnt ist, aber sie sind ihnen in Einfachheit und Schnelligkeit häufig überlegen.

Da das Interesse der Nutzer an einem sofortigen Volltextzugang zunehmend Vorrang gewinnt gegenüber dem Bedarf nach systematischer Suche, gibt es für die Suchmaschinenanbieter noch sehr viel zu tun bei der Erhöhung der Open-Access-Quote. Dies betrifft vor allem die Vollständigkeit beim Crawling von Open-Access-Servern und die Qualität der Metadatenextraktion. Hier bieten Bibliotheksentwicklungen wie die Bielefelder Suchmaschine BASE qualitativ mehr. Google & Co. zeigen, dass die Beschränkung von Bibliothekskatalogen auf bestimmte Dokumenttypen und reine Sucheinstiege obsolet wird. Der OPAC als „Bibliotheks-Suchmaschine“<sup>21</sup> muss zentrales Nachweisinstrument für alle relevanten Dokumente sein. Dabei ist die Realisierung eines sinnvollen Rankings anstelle der bisherigen Ausgabe nach isolierten Kriterien ein wichtiges Ziel.

Die Zurückhaltung der Bibliotheken bei der Unterstützung von Bestandsnachweisen in den Suchmaschinen hat ihren Grund sicherlich auch in der mangelhaften Sichtbarkeit der Bibliothek als Volltextanbieterin. Für die Bibliotheken wäre es durchaus von Vorteil, mit ihren Beständen auch im Web sichtbar zu sein, da in vielen Fächern zuerst Suchmaschinen und erst im Anschluss oder gar nicht Bibliothekskataloge konsultiert werden. Die schwer auffindbaren Google-Scholar-Einstellungen und die sehr schlichte Klarschrift-Angabe „Bibliothek xyz“ unterstützen dieses Anliegen ebenso wenig wie ein Button im Design des Linkresolver-Anbieters. Interessant für Bibliotheken wäre ein Design, das stärker in Richtung einer White-Label-Lösung gestaltet ist, so dass über das bibliothekseigene Design die Wiedererkennbarkeit bei den Bibliothekskunden gesichert wird. Dies wäre ein wichtiger Vorteil für Bibliotheken, die häufig nicht als Anbieter der von ihnen teu-

21 Lewandowski, Dirk: Suchmaschinen als Konkurrenten der Bibliothekskataloge: Wie Bibliotheken ihre Angebote durch Suchmaschinentechnologie attraktiver und durch Öffnung für die allgemeinen Suchmaschinen populärer machen können  
In: Zeitschrift für Bibliothekswesen und Bibliographie 53 (2006) 2, 71–78, S. 77.

er gekauften und aufwendig administrierten elektronischen Dokumente erkannt werden, wie Hochschulrankings immer wieder zeigen. Wenn sich Bibliotheken entschließen, beim Bestandsnachweis mit den Suchmaschinen zu kooperieren, könnten Sie zudem neue Nutzerkreise auf ihre Angebote aufmerksam machen. Ein weiterer wichtiger Nebeneffekt wissenschaftlicher Suchmaschinen ist, dass sie das Bedürfnis nach mehr Open Access wecken. Und last but not least: Wissenschaftliche Suchmaschinen bringen bibliothekarische Standards ins Web. Denn eine Qualitätssteigerung bei Such-, Treffer- und Rankingqualität können sie nur durch verbesserte Metadatenqualität und strukturierte Suchen erreichen.

