

Ben Kaden

Über „Google Scholar“

Verschriftlichung eines Vortrags vom 02. Januar 2006 im Institut für Bibliotheks- und Informationswissenschaft der Humboldt-Universität zu Berlin. Es werden Grundelemente sowie Vor- und Nachteile des aktuellen Stands von Google Scholar betrachtet.

Subsidiary text version of a lecture held at the Department of Library and Information Science/Humboldt-University to Berlin on January 2nd 2006. Basic elements and pros and cons of “Google Scholar” as it is available to date (jan 2006) are delineated.

Der im November 2004 als BETA-Version gestartete Google-Dienst *Google Scholar* stellt ein Nachweissystem für im Internet – sowohl beschränkter wie auch unbeschränkter zugängliche – verfügbare wissenschaftliche Literatur bzw. Dokumente dar. Der generelle Unterschied zur allgemeinen Google-Suche liegt in der Ausrichtung von als wissenschaftlich eingeschätzten Quellen (*scholar* engl = „der Gelehrte“, mlat. *Scholaris* = “zur Schule gehörig“ von lat. *schola* = “Schule“, griech. *σχολή* = „Rast, Muße, Vortrag, Ort des Vortrags, Schule“).

Damit stößt Google offensiv in einen Bereich vor, der traditionell von kommerziellen Datenbank Anbietern und Volltextarchiven abgedeckt wird. Google Scholar ist in gewisser Hinsicht nicht nur Konkurrenz für „herkömmliche“ Literaturnachweissysteme, sondern potentiell auch für das Noch-Monopol von ISI als *Science Citation Index*, da sich mithilfe von Google Scholar Webzitationen nachvollziehen lassen.

In der folgenden Präsentation soll kurz ein Überblick über den Dienst an sich und die mit diesem verbundenen Vor- und Nachteile gegeben werden.¹

Google Scholar – Basics

Keep it Simple – Das Google-Prinzip

Google Scholar² ging im November 2004 – offiziell als Testversion – für die Öffentlichkeit verfügbar online.³ Das Projekt „aims to provide access to scholarly materials via the crisp and familiar Google search box.“⁴ Interessant ist dabei, dass Google den Standard im Interface-Design für Rechercheanwendungen in einer Art und Weise festsetzt, dass dieser mittlerweile bei allen Nutzern als vertraut angesehen wird. Das *Google-Prinzip* besteht aus denkbar wenigen Interaktionselementen und stellt damit einen Gegensatz zu den häufig recht komplex anmutenden Suchmasken von

¹ Stand Anfang Januar 2006

² <http://scholar.google.com/>

³ Butler, 2005

⁴ Banks, 2005

Bibliothekskatalogen und Datenbanken dar, wobei das Prinzip dieser „Simple Search“ auch dort zunehmend Anwendung findet.

Simplizität als dominierendes Positivmerkmal findet sich auch in der Selbstbeschreibung des Produkts: „Google Scholar provides a simple way to broadly search for scholarly literature.“⁵

Insofern verwundert es nicht, dass zur Einführung des Dienstes eine Beschränkung auf die einfache Standardsuche als ausreichend angesehen wurde. Erst einen Monat später trat die „Erweiterte Suche“ dazu.

Im Gegensatz zu Google, Google Books (ehemals Google Print) und anderen Google-Diensten gibt es Google Scholar bislang ausschließlich mit einer englischsprachigen Nutzeroberfläche.

Die Zielgruppe des Angebotes ist weitgehend undefiniert und richtet sich prinzipiell an die gesamte Bandbreite potentieller Nutzer wissenschaftlicher Literatur, wobei davon auszugehen ist, dass das Angebot eher Studierende als Fachwissenschaftler ansprechen wird, da letztere in der Regel weitgehend auf fachlich spezialisierte Angebote zurückgreifen.

Keep it broad – Was und von wem wie erschlossen wird

Laut Selbstbeschreibung erfasst Google Scholar alle digital verfügbar Dokumententypen. Dies sind u.a. „peer-reviewed papers, theses, books, abstracts and articles“.⁶ Die Materialien stammen von Servern von „academic publishers, professional societies, preprint repositories, universities and other scholarly organizations“. Die fachliche Vielfalt wird mit „many disciplines“ umschrieben. Welche Datenbestände genau indiziert werden, ist leider unklar.⁷ Enthalten sind offensichtlich die Volltextangebote aller größeren wissenschaftlichen Verlage, ausgenommen die des Scirus-Anbieters *Elsevier*.⁸ Nicht angegeben wird allerdings wie hoch der Indizierungsgrad ist bzw. an welchen Punkten Indizierungslücken bestehen. Generell erscheint die Erfassung, die zu großen Teilen aus der Erfassung über *Citations* (bzw. Links) erfolgt, relativ unsystematisch, sodass eine Parallelrecherche für den wissenschaftlichen Alltag in der Regel nach wie vor notwendig bleibt.

Keep it BETA – Warum noch nicht alles wie gewünscht passt

Die Qualität sowohl des *Recalls* wie auch der *Precision* ist bislang z.T. noch recht unbefriedigend und kann mit der Ergebnisqualität von separaten Recherchen bei entsprechenden kommerziellen Referenz- und Volltextdatenbanken nicht konkurrieren. Gründe für diese Situation und daraus resultierende Minuspunkte für das System sind z.T. die bereits erwähnte Nichtangabe der erfassten Quellen und die Erschließungstiefe. Google „entschuldigt“ diese Schwächen durch die Platzierung des „BETA-Labels“, was laut Jascó aus der Erfahrung mit anderen Diensten gern auch für Jahre stehen

⁵ <http://scholar.google.com/scholar/about.html>

⁶ <http://scholar.google.com/scholar/about.html>

⁷ vgl. auch Giustini; Barsky, 2005, S. 86 „What is not indexed is more difficult to determine, as Google has been vague at times about GS's content.“

⁸ Jascó, 2005

bleiben kann.⁹ Ob ein BETA-Dienst als „all-in-one“ für die wissenschaftliche Arbeit wirklich angewandt werden sollte, ist – besonders angesichts der ebenfalls unten noch einmal erwähnten „Google-Fixierung“ vieler Studierender – durchaus fragwürdig.

Free Access – die Vorteile

Freier Zugang

Sieht man von der für Google typischen hochoptimierten Nutzungsfreundlichkeit (Usability) des Angebots ab, lässt sich der Hauptvorteil am freien Zugriff auf das Recherchetool festmachen: “Google makes it very easy and free to find scholarly information about any topic – an important service for those who do not have access to the most appropriate fee-based indexing/abstracting databases which traditionally have helped in information discovery.”¹⁰

Volltexterschließung

Ein weiterer Vorteil des Angebotes gegenüber kommerziellen Datenbanken, die häufig auf bibliographische Angaben, Abstracts, Schlagworte und Metainformationen beschränkt sind, ist, dass Google und damit auch Google Scholar generell auch die Volltexte durchsucht.¹¹

Omnisuchmaske

Google Scholar ist damit besonders für Institutionen, Bibliotheken etc. geeignet, die Abonnements der von Google erfassten Datenbanken (inkl. Volltextdatenbanken von Verlagen) haben, da sie nun mittels Google-Funktionalität (d.h. dem simplen Suchfeld) eine breitere Vielfalt von Datenbanken und den Zugang zu Volltexten verknüpfen können. Dies entspricht nicht zuletzt dem Anliegen von Projekten wie Vascoda, wobei letzteres sicherlich als weitaus spezieller und funktional beschränkter zu sehen ist.

Cited-By-Funktion

Ein hochinteressantes Merkmal des Angebotes stellt die *Cited-By*-Funktion dar, wie folgendes Beispiel von Banks illustriert:

“For example, I ran a *Scholar* search for "bioinformatics", which returned "about 62,600" results within 06 seconds. I selected the third citation, which was cited by 88 other resources within *Scholar*. After 5 clicks I was down to one article, standing at the root of one chain of thematic connection.”¹²

Diese ist allerdings für Zitationsanalysen nur eingeschränkt geeignet (sh. Nachteile – Ranking).

⁹ Jascó, 2005, 214

¹⁰ Jascó, 2005

¹¹ dies allerdings nicht komplett, siehe Nachteile

¹² Banks, 2005

Weak Structure – die Nachteile

Unvollständigkeit und Unübersichtlichkeit

Die inhaltliche Abdeckung ist, unter Beachtung der Zielsetzung, „scholarly materials“ zu erfassen, unvollständig zu nennen: Google Scholar erschließt nur einen Teil der Datenquellen und verzerrt damit das tatsächliche Bild der wissenschaftlichen Literatur. Jascó berichtet von großen Lücken, speziell auch im OA-Bereich, vor allem bei *delayed open access* Produkten.¹³ Google hält die reale Abdeckung unter Verschluss, so dass es schwierig ist, den Fehler zu kalkulieren. Allgemein wird ein Mangel an Informationsqualität postuliert.¹⁴

Als ausgesprochen unpraktisch erweist sich die Nichtsortierbarkeit der Treffer (nach Dokumententyp, Aktualität etc.). Hier bietet z.B. Scirus mit der Differenzierbarkeit nach „journal results“, „preferred web results“ und „other web results“ sowie „relevance“ und „date“ als wesentlich praktischer.

Die „Advanced Search“¹⁵ ermöglicht hier über bestimmte „presearch“-Einstellungen eine gewisse Kompensation.

Heterogenität

Da Google Scholar auf diverse, in Strukturierung, Umfang und inhaltlicher Ausrichtung recht heterogene Datenquellen zurückgreift, bekommt man ganz unterschiedliche Darstellungen der einzelnen Ergebnisse (teilweise mit Abstractausschnitten, z.T. mit Volltextsnippets, häufig nur den Seitentitel...), d.h. der Aussagegehalt der einzelnen Trefferangaben ist häufig bei weitem nicht so aussagekräftig (aufgrund der fehlenden, identisch strukturierten Metainformation) wie bei Datenbanken.

Ranking

Das an sich für den Wissenschaftsbetrieb hochinteressante Ranking der Artikel – das theoretisch auf das Web bezogen sozusagen als webometrisch potentiell korrespondierend zum szientometrischen Gehalt des ISI Science Citation Index anwendbar wäre – wird durch die weitgehende Undurchsichtigkeit der Rankingkriterien leider verwässert. Die Angabe

„Google Scholar orders your search results by relevance. As with Google web search, the most useful references appear at the top of the page. Google ranking technology considers the full text of each

¹³ Jascó, 2005

¹⁴ Gardner, Eng, 2005

¹⁵ http://scholar.google.com/advanced_scholar_search

article, the author, the publication in which the article appeared, and how often the piece has been cited in other scholarly literature.“¹⁶

listet zwar einige Faktoren auf, wie sehr diese jedoch relativ zu einander gewichtet werden, wird nicht mitgeteilt. Giustini und Barsky schreiben in diesem Zusammenhang knapp: “However, keep in mind that PageRank in GS is not the same as ISI’s bibliometric tools, a distinction that must be iterated to users.”¹⁷

“File-Size does matter”

Ein zusätzliches Manko ist die häufiger erwähnte Beschränkung der Erschließungsgröße auf maximal 100-120 kB pro Dokument, wobei für elektronisch vorliegende wissenschaftliche Aufsätze nicht selten eine Größe von 1 MB anzunehmen ist. Suchmaschinenkonkurrent Yahoo erschließt immerhin bis zu auf 500 kB.

Auch wird nicht erwähnt, wie oft der Datenbestand von Google Scholar aktualisiert wird. Péter Jascó berichtet für den Zeitraum von Mitte November 2004 – Ende Januar 2005 von einer Stagnation der Resultate.¹⁸

Teaching “google literacy” – Bibliothekarische Perspektive:

Mit dem Freischalten von Google Scholar wurde nach der Kontroverse um Google Books quasi die zweite große Herausforderung für die Bibliotheken durch Google postuliert.¹⁹ So wird befürchtet, dass Bibliotheken durch das Projekt Nutzer verlieren sowie, dass sich bei Studierenden der Trend der weitgehenden Google-Fixierung fortsetzt.²⁰ Gerade bei dieser Zielgruppe wird eine Abwendung von den Bibliotheken und ihren Dienstleistungen hin zu Google (Print und Scholar) angenommen. Ein tatsächliche Konkurrenz stellt Google Scholar objektiv aufgrund seiner nach wie vor eingeschränkten Inhaltsqualität für (auf der Höhe der Zeit befindliche) Bibliotheksdienstleistungen nicht dar. So findet sich z.B. bei Willinsky die durchaus nachvollziehbare Aussage „[...] Google still lacks the basic features of a good subject index.“²¹

Das Hauptproblem der Bibliotheken in Bezug auf Google Scholar wird die sinnvolle Integration des Angebotes in den Dienst für den Nutzer darstellen.²² Hier scheint es besonders gegeben, die Grenzen und Möglichkeiten des Angebots in Form einer spezifischen Nutzungskompetenz zu vermitteln.

¹⁶ <http://scholar.google.com/scholar/about.html>

¹⁷ Giustini; Barsky, 2005, S. 87

¹⁸ Jascó, 210

¹⁹ Kesselman; Watstein, 2005 formulieren dies folgendermaßen: „Google Scholar has set the academic library world on edge, provoking much hand wringing about the role and future of libraries.“, 380

²⁰ vgl. ebd.

²¹ vgl. Willinsky, 2006, 178

²² Ein Beispiel dafür, wie so etwas aussehen könnte, bieten die MIT-Libraries: <http://libraries.mit.edu/help/google-scholar/>

Fazit

Natürlich muss man Google-Scholar (genau wie Google Books und andere BETA-Dienste) als ein sich in Auf-, Ausbau und Entwicklung befindliches Projekt ansehen, das verständlicherweise im aktuellen Stadium zahlreiche Schwächen und Fehler in der Funktionalität aufweist. Gerade deshalb muss man angesichts des „Scholar buzz“ (Banks) durchaus skeptisch sein, inwieweit sich hier kurzfristig eine tatsächliche Konkurrenz für die klassischen Anbieter darstellt, die ja erstaunlicherweise weitgehend freimütig kooperieren. Nicht unterschätzen sollte man zusätzlich den Erfahrungsvorsprung direkter Konkurrenzprojekte wie z.B. Scirus bzw. kommerzieller Angebote wie z.B. Scopus. Auf von vornherein spezialisierten Feldern, wie z.B. der Medizin, sind und bleiben spezialisierte und auf das Fach zugeschnittene Angebote sicherlich das Maß der Dinge.²³ Hier wird Google Scholar allenfalls eine ergänzende Funktion übernehmen können.

Generell bleibt eher fraglich, ob Google-Scholar in der vorliegenden BETA-Version ein angemessenes Werkzeug für „ALPHA“-Wissenschaft ist.

Berlin, 02. Januar 2006

Literatur

Banks, Marcus A. (2005) The excitement of *Google Scholar*, the worry of *Google Print*. In: *Biomedical Digital Libraries* 2005; 2: 2.. URL: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1079792>

Butler, Declan (2004) Science searches shift up a gear as Google starts Scholar engine. In: *Nature* 432, 423 (25 November 2004) (doi: 10.1038/432423a)
URL: <http://www.nature.com/nature/journal/v432/n7016/full/432423a.html>

Gardner, Susan; Eng, Susanna (2005) Gaga over Google? Scholar in the Social Sciences. In: *Library Hi Tech News incorporating Online and CD Notes*, Volume 22, Number 8, August 2005, pp. 42-45(4)

Giustini, Dean; Barsky, Eugene (2005) A look at Google Scholar, PubMed, and Scirus: comparisons and recommendations. In: *Journal of the Canadian Health Libraries Association / Journal de l'Association des Bibliothèques de la Santé du Canada* 26: 85–89 (2005)

Henderson, Jim (2005) Google Scholar: A source for clinicians? In: *Canadian Medical Association Journal*, 2005 Jun 7;172(12):1549-50.

Jascó, Péter (2005) Google Scholar: The Pros and the Cons. In: *Online Information Review*. Vol. 29 No. 2, 2005 pp. 208-214

Kesselman, Martin; Watstein, Sarah Barbara (2005) Google Scholar and libraries: point/counterpoint. In: *Reference Services Review* Vol. 33 No. 4, 2005 pp. 380-387

Willinsky, John (2006) *The Access Principle: the Case for Open Access to Research and Scholarship*. Cambridge, London: MIT

²³ vgl. Henderson, 2005