



**OAW**

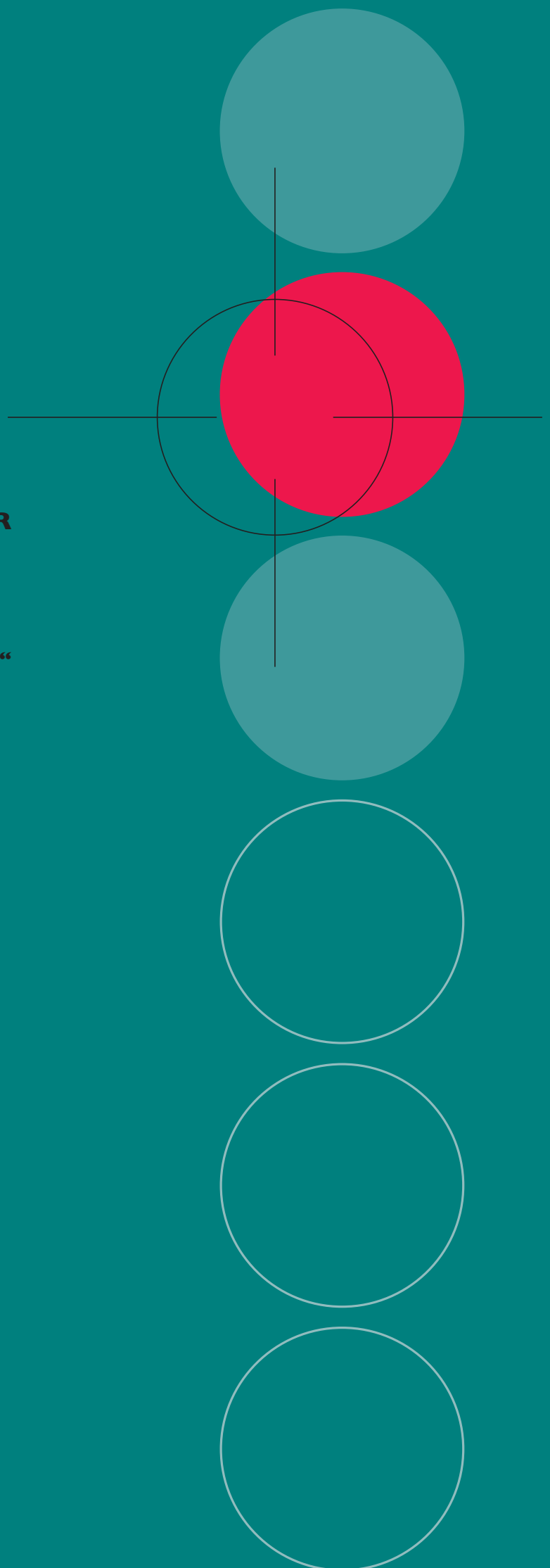
Österreichische Akademie  
der Wissenschaften



INSTITUT FÜR  
TECHNIKFOLGEN-  
ABSCHÄTZUNG

## **GOOGLE, GOOGLE SCHOLAR UND GOOGLE BOOKS IN DER WISSENSCHAFT**

**STECKBRIEF 3 IM RAHMEN DES  
PROJEKTS „INTERACTIVE SCIENCE“**



**ITA-PROJEKTBERICHT NR. A52-3**

**ISSN: 1819-1320**

**ISSN-ONLINE: 1818-6556**





**OAW**

Österreichische Akademie  
der Wissenschaften



INSTITUT FÜR  
TECHNIKFOLGEN-  
ABSCHÄTZUNG

## **GOOGLE, GOOGLE SCHOLAR UND GOOGLE BOOKS IN DER WISSENSCHAFT**

### **STECKBRIEF 3 IM RAHMEN DES PROJEKTS INTERACTIVE SCIENCE**

INSTITUT FÜR TECHNIKFOLGEN-ABSCHÄTZUNG  
DER ÖSTERREICHISCHEN AKADEMIE DER WISSENSCHAFTEN

*Projektleitung:* Michael Nentwich

*Autor:* René König

BEITRAG ZU TEILPROJEKT I „KOLLABORATIVES WISSENSMANAGEMENT UND  
DEMOKRATISIERUNG VON WISSENSCHAFT“ DES VERBUNDPROJEKTS „INTERACTIVE  
SCIENCE – INTERNE WISSENSCHAFTSKOMMUNIKATION ÜBER DIGITALE MEDIEN“,  
GEFÖRDERT DURCH DIE VW-STIFTUNG

WIEN, MAI 2010

## **IMPRESSUM**

### **Medieninhaber:**

Österreichische Akademie der Wissenschaften  
Juristische Person öffentlichen Rechts (BGBl 569/1921 idF BGBl I 130/2003)  
Dr. Ignaz Seipel-Platz 2, A-1010 Wien

### **Herausgeber:**

Institut für Technikfolgen-Abschätzung (ITA)  
Strohgasse 45/5, A-1030 Wien  
<http://www.oeaw.ac.at/ita>

Die ITA-Projektberichte erscheinen unregelmäßig und dienen der Veröffentlichung der Forschungsergebnisse des Instituts für Technikfolgen-Abschätzung. Die Berichte erscheinen in geringer Auflage im Druck und werden über das Internetportal „epub.oeaw“ der Öffentlichkeit zur Verfügung gestellt:  
<http://epub.oeaw.ac.at/ita/ita-projektberichte>

ITA-Projektbericht Nr.: A52-3  
ISSN: 1819-1320  
ISSN-online: 1818-6556  
<http://epub.oeaw.ac.at/ita/ita-projektberichte/d2-2a52-3.pdf>

© 2010 ITA – Alle Rechte vorbehalten

# Inhalt

Zusammenfassung .....	I
Summary .....	II
1 Einführung .....	1
2 Google .....	3
2.1 Was wird gefunden? .....	5
2.2 Was wird wie angezeigt? .....	6
2.3 Einflussmöglichkeiten der NutzerInnen .....	11
2.4 Einsatz in der Wissenschaft .....	14
2.4.1 Suche nach wissenschaftsrelevanten Inhalten .....	15
2.4.2 Suche nach Personen und Institutionen .....	19
2.4.3 Einfluss auf externe Wissenschaftskommunikation (Öffentlichkeitsarbeit) .....	21
3 Google Scholar .....	25
3.1 Was wird gefunden? .....	25
3.2 Was wird wie angezeigt? .....	28
3.3 Einflussmöglichkeiten der NutzerInnen .....	30
3.4 Einsatz in der Wissenschaft .....	32
3.4.1 Suche nach wissenschaftlichen Inhalten .....	32
3.4.2 Nutzung der Zitationsanalyse .....	35
4 Google Books .....	37
4.1 Was wird gefunden? .....	38
4.2 Was wird wie angezeigt? .....	40
4.3 Einflussmöglichkeiten der NutzerInnen .....	45
4.4 Einsatz in der Wissenschaft .....	46
4.4.1 Suche nach wissenschaftlichen Büchern .....	46
4.4.2 Einfluss auf Verlage und AutorInnen .....	48
5 Vorläufige Bewertung und Ausblick .....	51
Literatur .....	55

## Abbildungsverzeichnis

Abbildung 2.1-1: Suchmaschinenmarkt 2008-2009 .....	3
Abbildung 2.1-2: Entwicklung des Suchmaschinenmarktes in Deutschland .....	4
Abbildung 2.2-1: Semantische Interpretation eines Suchergebnisses .....	10
Abbildung 2.3-1: Ergebnisanzeige im „Wunderrad“ .....	14
Abbildung 2.4-1: Einbindung von Ergebnissen aus Google Scholar .....	16
Abbildung 2.4-2: Google-Suche nach dem Begriff „Klimaforschung“ .....	22
Abbildung 3.2-1: Ergebnisanzeige in Google Scholar .....	28
Abbildung 4.2-1: Startseite von Google Books .....	41
Abbildung 4.2-2: Ausschnitt-Ansicht in Google Books .....	42
Abbildung 4.2-3: Beispielsuche in Google Books .....	43

## Tabellenverzeichnis

Tabelle 2.2-1: Abfrageabhängige Faktoren im Ranking .....	7
---	---



# Zusammenfassung

In diesem Bericht wurden drei Dienste des Unternehmens Google Inc. untersucht, die sich zur Textsuche eignen, und es wird versucht, ihre Effekte auf die (in erster Linie interne) Wissenschaftskommunikation abzuschätzen: Die Universalsuchmaschine Google, die akademische Spezialsuchmaschine Google Scholar und die Buchsuche Google Books.

Zunächst werden einführende Informationen zum Unternehmen Google Inc. und seinen Wurzeln gegeben. Anschließend werden die jeweiligen Angebote vorgestellt, indem wichtige Funktionsprinzipien skizziert werden, nämlich die zugrunde liegenden durchsuchbaren Datenbasen, die Art und Weise in der Inhalte angezeigt werden und die Einflussmöglichkeiten der NutzerInnen. Beginnend mit der Universalsuchmaschine Google werden auch einige generelle Aspekte von Suchmaschinen angesprochen.

Es ließen sich für alle drei Dienste Einflüsse auf die (interne) Wissenschaftskommunikation beobachten. Insbesondere in Bezug auf die Suche nach wissenschaftlichen Inhalten erscheinen sie relevant. Da die untersuchten Plattformen anderen Funktionsprinzipien folgen als traditionelle Informationssysteme (z. B. Bibliotheken), werden mögliche Veränderungen skizziert und diskutiert. Zudem werden verschiedene weitere Aspekte der jeweiligen Dienste in den Fokus genommen, z. B. Einflüsse der Google-Websuche auf externe Wissenschaftskommunikation (Öffentlichkeitsarbeit), die Zitationsanalysefunktion von Google Scholar, oder die juristischen Probleme, die sich aus dem Projekt Google Books ergeben.

Es wird gefolgert, dass trotz des vorläufigen Charakters dieses Berichts, der sich aus der Dynamik des untersuchten Feldes ergibt, einige Aspekte bemerkenswert und höchstwahrscheinlich von Dauer sind: Aufgrund der Marktmacht von Google, seiner Popularität und seinen Zielen, die sich teilweise mit denen der Wissenschaft überschneiden (beide sind zentral mit der Sammlung und Organisation von Informationen beschäftigt), ist anzunehmen, dass Google weiterhin großen Einfluss auf die Wissenschaftskommunikation ausüben wird. Während mit einigen dieser Auswirkungen innerhalb des Wissenschaftssystems umgegangen werden kann, werfen andere fundamentale Fragen auf, die auf eine weiter gefasste Perspektive drängen.

## Summary

In this report we examine three Google services that can be used for textual searches and assess their impact on (primarily internal) scholarly communication: The universal web search engine Google, the academic search engine Google Scholar and the book search Google Books.

Firstly, we present basic information about the company Google Inc. and its roots. We then introduce each service by outlining important functional principles, namely the searchable databases, the way content is delivered and the available options for users. Starting with Google web search, we also describe some general aspects of search engines.

We found that all three services have effects on (internal) scholarly communication. They seem to be relevant especially with regard to searches for academic information. Since the observed platforms follow functional principles that differ from traditional information systems (e.g. libraries), possible changes are outlined and discussed. We also focus on various other aspects of the specific services, such as the effects of the Google web search on public relations, Google Scholar's citation analysis functionality, or the legal problems that arise from the Google Books project.

We conclude that despite the preliminary character of this assessment, which results from the dynamics in the field observed, certain aspects seem striking and almost certainly persistent: Due to Google's market power and popularity and the fact that that it shares some goals with academics (both to a large extent deal with gathering and organizing information), it will probably continue to have major effects on scholarly communication. While some of them can be dealt with in the academic system itself, others raise fundamental questions that require a wider perspective.



# I Einführung

Der 1998 gegründete Konzern Google Inc. gehört heute zu den bedeutendsten Firmen der Internetbranche und umfasst neben der bekannten Suchmaschine (siehe 2) eine ganze Reihe von weiteren Diensten. Diese reichen von spezielleren Suchmaschinen, etwa für Bücher (*Google Books*, siehe Abschnitt 4), akademische Publikationen (*Google Scholar*, siehe Abschnitt 3) und Videos (*Google Video*), bis hin zu komplexeren kombinierten Diensten wie der interaktiven Abbildung von geografischen Karten-, Satelliten- und Bilddaten (*Google Maps*, *Google Earth*, *Google Streetview*), sowie weiteren Services wie einem Übersetzungsdienst für Webseiten, einer Plattform zum Schreiben und Publizieren von Sachtexten (*Knol*), Services zur kollaborativen Wissensproduktion (*Google Docs*, *Google Wave*) einem kostenlosen E-Mail-Dienst (*Gmail*, bzw. in Deutschland und Großbritannien *Google Mail*), einem sozialen Netzwerk (*Google Buzz*) und eigener Software wie dem Browser *Chrome* oder dem zur Nutzung von *Google Earth* erforderlichen gleichnamigen Programm. Für 2010 plant Google außerdem ein eigenes Betriebssystem namens *Chrome OS* und trat mit dem *Nexus One* auch in den Smartphone-Markt ein. Zudem akquirierte der Konzern diverse Firmen (zu den populärsten gehören die Videoplattform *YouTube* und der Weblog-Dienst *Blogger.com*) und entwickelt ständig neue Produkte.<sup>1</sup>

Die Weite des Angebots spiegelt die Unternehmensphilosophie des Konzerns wider:

„Das Ziel von Google besteht darin, die auf der Welt vorhandenen Informationen zu organisieren und allgemein zugänglich und nutzbar zu machen.“<sup>2</sup>

Neben dem Bestreben gewissermaßen das Weltwissen greifbar zu machen, wird hier also auch noch auf zwei zentrale Merkmale der Google-Anwendungen hingewiesen: Freie, allgemeine Zugänglichkeit und Funktionalität. Dieser Fokus deutete sich bereits an, als die Unternehmensgründer Sergey Brin und Lawrence Page die namensgebende Websuchmaschine als Projekt an der Stanford University vorstellten: „Another important design goal was to build systems that reasonable numbers of people can actually use“ (Brin/Page 1998, S. 3). Die Verbesserung der Nutzerfreundlichkeit war dabei ein zentrales Ziel (vgl. ebd. 2 f.), das einen entscheidenden Beitrag zum Unternehmenserfolg geleistet haben dürfte. Google-Anwendungen sind gewöhnlich kostenlos zugänglich und leicht zu bedienen.

Seine Einnahmen erzielt Google Inc. durch gezielte Schaltung von Werbung. Dazu wurde das sogenannte *AdWords* entwickelt – ein System, das es InserentInnen ermöglicht, gegen Bezahlung Werbung in Form von Hyperlinks in der Nähe von Suchergebnissen einzublenden, sobald bestimmte frei wählbare Begriffe von NutzerInnen eingegeben werden. Durch seine Effektivität (Googles KundInnen zahlen nur tatsächlich angeklickte Seiten) wurde das System schnell populär und Google wurde zum Marktführer der Internetwerbung<sup>3</sup>. Dazu trägt auch *AdSense* bei, das in ähnlicher Weise Webmastern die Einblendung von Werbelinks auf ihren Seiten erlaubt. Zudem engagiert sich Google

**diverse Angebote abseits der Suchmaschine**

**Unternehmensziel: Weltwissen verfügbar machen**

**freie Zugänglichkeit und Funktionalität**

**Einnahmen durch gezielte Werbung**

<sup>1</sup> Eine umfangreiche Liste der Angebote von Google findet sich hier: [google.de/intl/de/options](http://google.de/intl/de/options) (zuletzt aufgerufen am 25.02.10).

<sup>2</sup> [google.at/intl/de/corporate](http://google.at/intl/de/corporate) (zuletzt aufgerufen am 25.02.10).

<sup>3</sup> [emagazine.credit-suisse.com/app/article/index.cfm?fuseaction=OpenArticle&aoid=261729&lang=DE](http://emagazine.credit-suisse.com/app/article/index.cfm?fuseaction=OpenArticle&aoid=261729&lang=DE) (zuletzt aufgerufen am 25.02.10).

auch in traditionellen Segmenten wie der Print- und Fernsehwerbung, um zusätzliche Märkte zu erschließen (dazu: Kaumanns/Siegenheim 2008; Schießl 2007; Vascellaro 2010).

**„Dienst-gegen-Profil-Modell“**

Funktionalität und Attraktivität der Google-Dienste gründen vor allem auch auf der Strategie, umfangreiche Daten aus ihrer Nutzung zu sammeln und auszuwerten. Die Geschäftsgrundlage ließe sich somit auch als „Dienst-gegen-Profil-Modell“ (Rogers 2009; Elmer 2004) beschreiben: NutzerInnen zahlen zwar für die Dienstleistungen nichts, müssen aber in Kauf nehmen, dass der Konzern umfangreiche Daten erhebt, die u. a. zur Erstellung von Nutzungsprofilen verwendet werden. Dies dient sowohl einer effektiveren Platzierung von Werbung als auch der Verbesserung der Suchfunktionen. Obgleich Google immer wieder den verantwortungsvollen Umgang mit den erhobenen Daten betont, reißt die Kritik und Skepsis gegenüber dieser Unternehmenspolitik nicht ab (siehe z. B. Bredow et al. 2010).

**Open-Source-Strategie, aber auch Geheimhaltung**

Auch auf der Ebene der Produktentwicklung werden Personen von außerhalb des Konzerns hinzugezogen, um diese zu verbessern. So wird häufig mit einer Open-Source-Strategie gearbeitet, etwa beim Browser Chrome. Demgegenüber werden die Algorithmen der Suchmaschine(n) strengstens gehütet und geheim gehalten.

**starker Einfluss auf gesellschaftliches Wissen**

Google Inc. ist nicht nur in wirtschaftlicher Hinsicht erfolgreich, sondern ist auch in seinem hochgesteckten Ziel der Organisation und Nutzbarmachung des Weltwissens durchaus ernst zu nehmen. Schon jetzt lässt sich feststellen, dass der Konzern erheblich zu einem strukturellen Wandel gesellschaftlichen Wissens beigetragen hat. Das umgangssprachliche „Googeln“ wird im Duden geführt und einige rufen bereits die „Google-Gesellschaft“ (Lehmann/Schettsche 2007) aus, oder skizzieren eine „Vergoogelung“ (Rogers 2009), bzw. sogar eine „Googlization of everything“<sup>4</sup>. Auch die Wissenschaftskommunikation bleibt von dieser Entwicklung nicht unberührt. Insbesondere die Anwendungen zur Textsuche erscheinen in diesem Zusammenhang von Bedeutung.<sup>5</sup> Daher soll in diesem Bericht die allgemeine Websuchmaschine Google (Abschnitt 2), die akademische Suchmaschine Google Scholar (Abschnitt 3) und die Buchsuche Google Books (Abschnitt 4) näher beleuchtet werden und hinsichtlich ihrer wissenschaftskommunikativen Bedeutung vorläufig eingeschätzt werden.

<sup>4</sup> Der Ausdruck stammt von Siva Vaidhyanathan, der derzeit ein Buch mit dem Titel „The Googlization of Everything“ schreibt: [googlizationofeverything.com](http://googlizationofeverything.com) (zuletzt aufgerufen am 20.04.2010). Zuvor tauchte der Begriff auch bereits in einem Artikel auf (Vaidhyanathan 2007).

<sup>5</sup> Selbstverständlich sind dies nicht die einzigen Anwendungen der Firma Google Inc., die für die Wissenschaftskommunikation interessant und relevant erscheinen. So wird etwa *Google Earth* bereits im Bereich der Medizinischen Geographie eingesetzt (Kamadjeu 2009). Eine vollständige Betrachtung aller Google-Anwendungen würde jedoch den Rahmen dieser Untersuchung sprengen.

## 2 Google

Die große Popularität der Websuchmaschine Google bringt es mit sich, dass nicht viele einleitende Worte nötig sind, um diese vorzustellen. Weltweit wurden zwischen März 2008 und März 2009 62 % der Suchanfragen an Google gestellt (Abbildung 2.1-1). Laut der Webanalyse-Firma *Net Applications* liegt der globale Marktanteil im Jänner 2010 sogar bei fast 85 %.<sup>6</sup>

**(Quasi-) Monopol im Suchmaschinenmarkt**

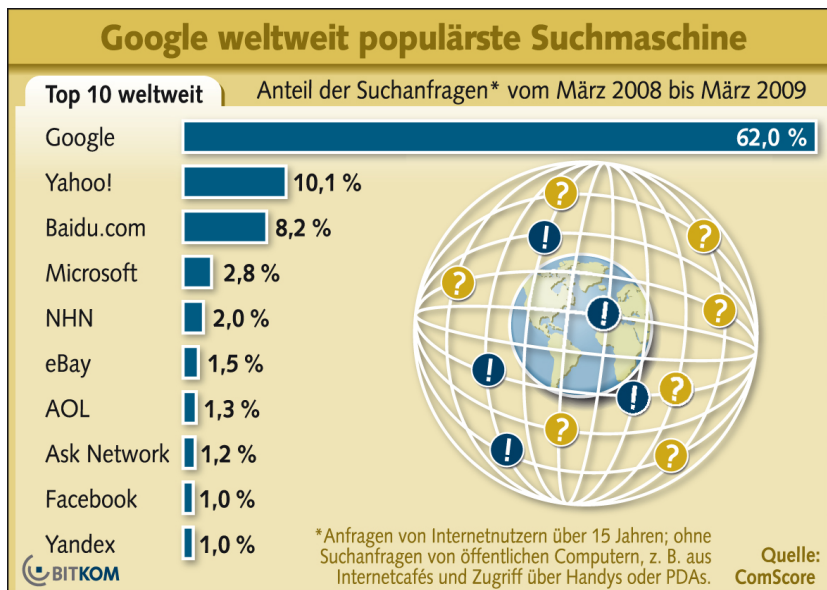


Abbildung 2.1-1: Suchmaschinenmarkt 2008-2009  
 Quelle: [bitkom.org/de/presse/8477\\_59404.aspx](http://bitkom.org/de/presse/8477_59404.aspx)  
 (zuletzt aufgerufen am 28.01.2010).

In einigen Ländern besitzt Google seit Jahren ein Quasi-Monopol im Suchmaschinenmarkt. Dazu gehört etwa Deutschland, wo der Marktanteil mitunter bei annähernd 90 % liegt (Abbildung 2.1-2, die x-Achse zeigt die Jahreszahlen, die y-Achse den Marktanteil in Prozent). Trotz immer wieder neu aufkommender Alternativen wie etwa Microsofts *Bing*, scheinen kurz- bis mittelfristig auch keine tiefgreifenden Änderungen bei dieser Tendenz erwartbar zu sein (vgl. Maaß et al. 2008).

Während die Suchmaschine Google also als bekannt vorausgesetzt werden kann und offenbar ein recht hohes Ansehen genießt, sind sich viele ihrer NutzerInnen nicht über die hintergründigen Funktionsprinzipien im Klaren und pflegen einen eher naiven Umgang mit Suchmaschinen (Fallows 2005). Dieser zeigt sich bereits am Terminus des „Googelns“, der gewissermaßen den Erfolg der Suche gleich mittransportiert:

„‘Googeln’ bezeichnet also die ebenso optimistische wie operationale Variante des Suchens, die Annahme nämlich, dass die Informationen, die gesucht werden, (erstens) im Netz vorhanden sind, sie (zweitens) gefunden werden können und dass dies (drittens) mit Hilfe von Google oder einer

**Funktionsprinzipien für viele NutzerInnen unklar**

**Google als „Findemaschine“ ...**

<sup>6</sup> [marketshare.hitslink.com/search-engine-market-share.aspx?qprid=4](http://marketshare.hitslink.com/search-engine-market-share.aspx?qprid=4)  
 (zuletzt aufgerufen am 28.01.10).

google-artigen Suchmaschine geschieht. In diesem Sinne sind Google & Co. keine Such- sondern Findemaschinen.“ (Schetsche et al. 2007, S. 30)

... und schwer zu durchschauende „Black Box“

Dabei wird gewissermaßen suggeriert, die Suchergebnisse seien erschöpfend: „Web search engines suggest that all information available can be searched within just one system“ (Lewandowski/Mayr 2006, S. 1). Dem Status einer Kulturtechnik entsprechend (vgl. Schetsche et al. 2007, S. 31), wird diese Handlung in der Regel im Alltag nicht hinterfragt. Der Umgang mit Suchmaschinen ist „(...) durch Kritiklosigkeit und fehlendes Problembewusstsein geprägt“ (Machill et al. 2007, S. 29 f.), sie gleichen einer „Black Box“ (Lewandowski 2005, S. 97). Diese Naivität und auch der Erfolg Googles, dürften darauf zurückzuführen sein, dass sich diese Technik schlichtweg in vielen Kontexten bewährt hat. Google führt uns derart oft zu den Ergebnissen, die wir uns wünschen, dass weder das Konsultieren anderer Informationsquellen, noch das Hinterfragen dieser Suchpraxis nötig erscheint.

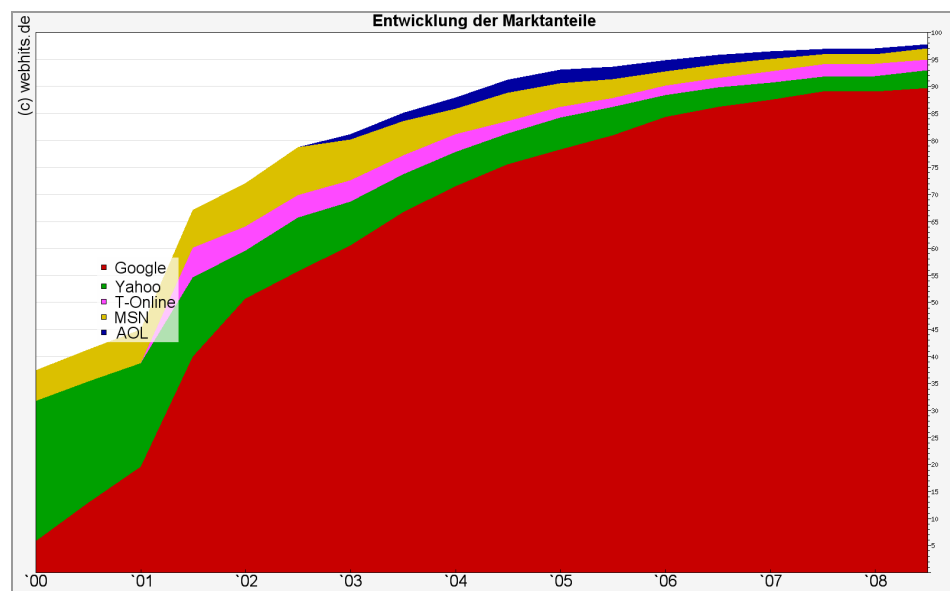


Abbildung 2.1-2: Entwicklung des Suchmaschinenmarktes in Deutschland

Quelle: [webhits.de/artwork/ws\\_engines\\_historical\\_druck.png](http://webhits.de/artwork/ws_engines_historical_druck.png)  
(zuletzt aufgerufen am 25.02.2010).

tiefergehende Kenntnisse über die Funktionsweise jedoch wichtig

Gehen wir jedoch davon aus, dass auch über alltagspraktische Banalitäten hinaus „gegoogelt“ wird, offenbart sich der Einfluss und die Problematik dieser Suchmaschine – vor allem wenn man ihre dominante Marktrolle mitbedenkt. Was gefunden wird und wie es angezeigt wird, gerät bei gesellschaftspolitischen Fragen zum Politikum, bei gesundheitlichen Fragen zum Risiko und bei wissenschaftlichen Fragen zur Gefahr für Validität und Qualität. Daher soll im Anschluss zunächst geklärt werden, was Google überhaupt finden kann bzw. nicht finden kann (2.1). Daraufhin beleuchten wir, wie Google das Gefundene darstellt (2.2) und schließlich inwiefern NutzerInnen in diese Prozesse eingreifen können (2.3).<sup>7</sup> Diese Informationen werden im Folgekapitel hinsichtlich ihrer Bedeutung für die Wissenschaftskommunikation beleuchtet (2.4).

<sup>7</sup> Dabei handelt es sich um analytische Trennungen, die aber bei Suchmaschinen stark miteinander verwoben sind. So beeinflussen etwa die NutzerInnen mit ihrem Suchverhalten, das von Google ausgewertet wird, die Ergebnisanzeige massiv.

## 2.1 Was wird gefunden?

Um eine schnelle Suche zu gewährleisten, durchsucht Google (wie auch andere Suchmaschinen) nicht alle verfügbaren Webseiten, sondern lediglich komprimierte Versionen, die auf den Servern des Unternehmens gespeichert sind. Google kann also nur die Webseiten finden, die hier gespeichert und deren Inhalte anschließend *indiziert* wurden. Dazu werden sogenannte *Crawler* (teilweise auch *Robots/Spider*) verwendet, die ständig das Web nach neuen Inhalten durchsuchen. Wie voraussetzungsreich und sensibel dieser Vorgang ist, stellen die Google-Gründer bereits in einem frühen Artikel heraus:

„Running a web crawler is a challenging task. There are tricky performance and reliability issues and even more importantly, there are social issues. Crawling is the most fragile application since it involves interacting with hundreds of thousands of web servers and various name servers which are all beyond the control of the system.” (Brin/Page 1998, S. 10)

Dadurch werden viele Webseiten bei der Indexierung nicht berücksichtigt. Bergman prägte für diesen Zusammenhang die Metapher des *Deep Web* (häufig werden auch die Begriffe *Invisible/Dark/Hidden Web* synonym verwendet), in dessen Tiefen gewöhnliche Suchmaschinen nicht vordringen können:

„Searching on the Internet today can be compared to dragging a net across the surface of the ocean. While a great deal may be caught in the net, there is still a wealth of information that is deep, and therefore, missed.” (Bergman 2001)

Die Hürden, die sich den Crawlern bei der Erfassung des Deep Web in den Weg stellen, sind vielfältig: Passwortgeschützte Inhalte, Webseiten, auf die nicht durch andere mittels Hyperlink verwiesen wird, inkompatible Dateitypen, mangelnde schriftliche Informationen usw. Diese Hürden sind zum einen rein technischer Art, zum anderen aber auch sozial erwünscht. So wollen sich bestimmte Anbieter etwa aus Gründen des Daten- oder Urheberrechtsschutzes vor Suchmaschinenindexierungen schützen. Webmaster können über entsprechende Befehle (in der sogenannten *robot.txt*) eine „Unsichtbarkeit“ vor Google et al. problemlos umsetzen. Umgekehrt versucht Google selbst gewisse Inhalte, wie etwa Spam oder teilweise auch jugendgefährdende Inhalte, nicht zu indexieren bzw. diese zu filtern (siehe auch 2.2).

Die Größe dieses für Suchmaschinen unsichtbaren Bereichs des Webs wird unterschiedlich eingeschätzt und ist abhängig von der Definition des Deep Web und der angewandten Methode. Bergman (2001) kam zu dem Schluss es müsse etwa 500 Mal größer sein als das sichtbare Web. Diese Zahl wurde jedoch verschiedentlich in Frage gestellt. Lewandowski und Mayr (2006, S. 5 ff.) merken etwa an, Bergman habe die Zahl viel zu hoch angesetzt, da er die Datengröße und nicht die Dokumentenanzahl als Grundlage seiner Berechnungen verwendet. Dadurch, dass weite Teile des Deep Web aus Datenbanken mit großen Dateien wie etwa hochauflösenden Satellitenbildern bestehen, schließen Lewandowski und Mayr, das unsichtbare Web müsse wesentlich kleiner sein.

So oder so ist davon auszugehen, dass das Verhältnis zwischen sichtbarem und unsichtbarem Web einer gewissen Dynamik unterliegt:

„A definition of what constitutes a truly invisible resource must necessarily be somewhat fluid, since the engines are constantly improving and adapting their methods to embrace new types of content.” (Sherman/Price 2003, S. 297)

**Schnelligkeit durch Crawling und Indexierung**

**nicht erfasste Seiten im „Deep Web“**

**viele Hürden für Crawler**

**Größe des Deep Web unklar**

**dynamisches Verhältnis zwischen sichtbarem und unsichtbarem Web**

Tatsächlich stellen etwa PDF-Dateien, die 2003 von Sherman und Price noch in den Bereich des „truly invisible web“ (ebd.) fielen, für Google mittlerweile kein Hindernis mehr dar. Zum Teil begegnet der Konzern dem Problem des Deep Web auch durch Entwicklung spezieller Suchmaschinen (z. B. Google Scholar, Abschnitt 3), die sozusagen in das tiefere Web eindringen können. Dennoch bleiben erhebliche Teile des Webs vor Googles Crawlern verschlossen und die oben beschriebene Suggestion einer allumfassenden Suche kann nicht erfüllt werden. Hinzu kommt, dass aktualisierte Webseiten nicht immer unmittelbar neu indexiert werden, weshalb diese Aktualisierungen unter Umständen erst mit Verzögerung von Suchmaschinen berücksichtigt werden. Auch diesem Problem begegnet man teilweise mit Spezialsuchmaschinen wie *Google News*.

**praktisches Problem:**  
**eher zu viele als zu**  
**wenig Suchergebnisse**

Eine Feststellung dieser Grenzen von Google ist für die Einschätzung der Suchmaschine sicherlich von großer Bedeutung. Im Alltag dürfte jedoch von den NutzerInnen eher ein konträres Problem wahrgenommen werden, nämlich nicht ein Informationsdefizit, sondern ein Überangebot. Dabei stellt sich eine Frage, der im folgenden Kapitel nachgegangen wird: Wie selektiert und hierarchisiert Google Informationen und macht diese in Form von Suchergebnissen zugänglich?

## 2.2 Was wird wie angezeigt?

**Hierarchisierung mit**  
**PageRank**

Viele Suchanfragen führen zu Ergebnissen mit mehreren Millionen Hyperlinks. Eine solche Fülle von Informationen wäre für NutzerInnen ohne zusätzliche Ordnung wenig hilfreich. Google begegnet diesem Problem, indem es die Ergebnisse in einer Rangfolge hierarchisiert. Diese ergibt sich aus zahlreichen Faktoren, von denen der bekannteste wohl das sogenannte *PageRank*-Verfahren ist, das patentiert wurde und den Grundstein für den Erfolg der Suchmaschine legte. Benannt nach dem Google-Mitbegründer und Entwickler Lawrence Page (Page et al. 1998; vgl. Brin/Page 1998), bewertet es Webseiten, indem ihnen ein Rang zugewiesen wird. Dieser *PageRank* zwischen 0 und 10 (früher zwischen 0 und 7) basiert auf der Popularität der Seite, gemessen an der Anzahl der hyperlinkförmigen Verweise, die sie erhalten hat. Dabei erhalten Verweise von Seiten mit einem höheren *PageRank* auch ein höheres Gewicht. Somit ähnelt dieses Prinzip dem akademischen Zitieren, worauf auch die Entwickler bei ihrer Beschreibung des Verfahrens hinweisen:

„Academic citation literature has been applied to the web, largely by counting citations or backlinks to a given page. This gives some approximation of a page’s importance or quality. PageRank extends this idea by not counting links from all pages equally, and by normalizing by the number of links on a page.” (Brin/Page 1998, S. 4)

Webseiten, die zahlreiche und möglichst „hochwertige“ Verweise erhalten haben, werden also von Google als besonders relevant angesehen und erscheinen in den Suchergebnissen weiter vorne.

**abfrageabhängige**  
**Faktoren**

Zu diesen „abfrageunabhängigen Faktoren“ (Lewandowski 2005, S. 90) kommen noch diverse weitere „abfrageabhängige Faktoren“ (ebd.), wie etwa die Position des gesuchten Begriffs auf der jeweiligen Seite. Ein Überblick über die gängigsten Kriterien dieser Art stammt von Lewandowski (2005, S. 90 ff.) und wird in Tabelle 2.2-1 zusammenfassend dargestellt. Eine detailliertere Beschreibung des von Google verwendeten Ordnungsprinzips würde den Rah-

men dieses Berichtes sprengen<sup>8</sup> und könnte auch nicht vollends erschöpfend sein – schließlich ist der zugrunde liegende Algorithmus ein gut gehütetes Firmengeheimnis.

Tabelle 2.2-1: Abfrageabhängige Faktoren im Ranking

Kriterium	Erläuterung
Dokumentspezifische Wortgewichtung (WDF)	Relative Häufigkeit des Vorkommens eines Worts in einem Dokument.
Wortabstand	Bei Anfragen mit mehreren Suchbegriffen wird der Abstand der Suchbegriffe voneinander berücksichtigt.
Position der Suchbegriffe	An markanten Stellen des Dokuments vorkommende Suchbegriffe werden höher bewertet. Zum Beispiel Vorkommen im Titel, in den Überschriften, in der URL.
Reihenfolge der Suchbegriffe in der Anfrage	In der Anfrage zuerst stehende Begriffe werden als bedeutender angesehen.
Metatags	Vorkommen der Suchbegriffe in den Metatags
Stellung der Suchbegriffe innerhalb des Dokuments	Vorkommen der Suchbegriffe am Beginn des Dokuments wird höher gewertet als späteres Auftreten.
Betonung von Begriffen durch HTML-Elemente	Hervorgehobene Begriffe (fett, kursiv) werden höher bewertet.
Groß-/Kleinschreibung	Dokumente, in denen die Suchbegriffe in exakt der eingegebenen Form vorkommen, werden bevorzugt.
Inverse Dokumenthäufigkeit (IDF)	Relative Häufigkeit des Vorkommens eines Wortes in Dokumenten der gesamten Datenbank; seltene Begriffe werden bevorzugt.
Ankertext	Vorkommen der Suchbegriffe im Linktext eines Dokuments, welches auf das Zieldokument verweist.
Sprache	Dokumente, die in der Sprache des benutzten Länderinterface verfasst sind, werden höher bewertet.
Geo-Targeting	Seiten, die ihren „Standort“ in der Nähe des Benutzers haben, werden bevorzugt.

Quelle: Lewandowski 2005, S. 91

Festzuhalten ist an dieser Stelle zunächst, dass Google durch die Selektion und Hierarchisierung von Webseiten massiv auf die Zugänglichkeit von Inhalten und die Wahrscheinlichkeit ihres Abrufs einwirkt. Da Suchmaschinen im Allgemeinen, und Google durch seine marktbeherrschende Position im Besonderen, als der „zentrale Einstiegspunkt ins Netz“ (Schetsche et al. 2007, S. 20) gelten, üben sie „(...) eine den klassischen Gatekeepern vergleichbare Funktion aus“ (Machill et al. 2007, S. 7). Neuartig an diesem speziellen Typus des Gatekeepers („Schlüsselworts“) ist freilich, dass dieser zu einem großen Teil automatisiert agiert. Dadurch entzieht er sich zumindest teilweise der menschlichen Kontrolle.

**Google als  
Gatekeeper**

<sup>8</sup> Google gibt an, dass die Suchergebnispositionierung von über 200 Faktoren abhängig sei (Quelle: [google.com/support/webmasters/bin/answer.py?answer=70897#3](https://support.google.com/webmasters/bin/answer.py?answer=70897#3); zuletzt aufgerufen am 08.07.09).

**Einflussnahme durch  
Suchmaschinen-  
optimierung**

Beide Aspekte – sowohl die zentrale Bedeutung der Google-Positionierung als auch die unzureichende Kontrollierbarkeit – zeigen sich an der Entstehung eines neuen Marketingbereichs: der Suchmaschinenoptimierung (häufig auch *SEO – Search Engine Optimization*). Dabei versuchen kommerzielle Anbieter die bekannten Funktionsprinzipien von Suchmaschinen dahingehend auszunutzen, dass ihre Webseite in den Ergebnislisten möglichst weit oben erscheint. Google kommt dieser Praxis einerseits entgegen indem entsprechende Empfehlungen bereitgestellt werden (Google 2008), andererseits stellt das Unternehmen auch einschränkende Richtlinien<sup>9</sup> auf und wehrt sich gegen Verstöße mit verschiedenen Maßnahmen. Für Aufsehen sorgte in diesem Zusammenhang etwa der vorübergehende Bann von einigen BMW-Seiten, nachdem der Autohersteller unerwünschte Suchmaschinenoptimierung betrieben hatte (Spiegel Online 2006).

**Probleme maschineller  
Relevanzherstellung**

Im Gegensatz zu traditionellen Gatekeepern können bei Suchmaschinen mitunter Relevanzen maschinell erzeugt und auf diese Weise manipuliert werden. Zwar wurde gegen diese Praxis in dem erwähnten Fall vorgegangen, jedoch erst im Nachhinein. Offenkundig erscheint zumindest, dass Suchmaschinen durch ihre automatisierte Funktionsweise verletzlich gegenüber manipulativen Eingriffen sind, die Inhalten eine in erster Linie maschinelle Relevanz verleihen. Das bedeutet auch umgekehrt, dass qualitativ hochwertige und reputable Inhalte durch Google schlecht positioniert werden können, wenn sie für das Crawling- und Rankingverfahren „nicht gut aufgestellt“ sind. NutzerInnen von Google (d. h. sowohl Suchende als auch Webmaster, die sich eine Auffindbarkeit ihrer Inhalte bei Google wünschen) unterwerfen sich zwangsläufig diesen technischen Regeln. Schetsche charakterisiert sie daher auch als „algorithmische Normen“ (2006, Absatz 12).

**„algorithmische  
Normen“ geben  
Handlungsmöglich-  
keiten vor**

**Intransparenz trotz  
beschränkter  
Einflussmöglichkeiten**

Allerdings sollte man dieses Argument nicht überstrapazieren. So können sich NutzerInnen von derartigen algorithmischen Normen emanzipieren, indem sie auf externe Quellen zurückgreifen. Auch ergeben sich gewisse Einflussmöglichkeiten durch Verfahren wie die oben beschriebene Suchmaschinenoptimierung oder auch Hacker-Aktivitäten (siehe z. B. Calishain/Dornfest 2003). Doch selbst wenn diese Maßnahmen mögliche Lücken des Systems ausnutzen mögen, akzeptieren sie doch letztlich bereits durch dessen bloße Nutzung den zugrunde liegenden Algorithmus. Dadurch, dass dieser als Firmengeheimnis gehütet wird und seine Funktionsweise extrem komplex ist, kommt es zu einer sehr weit gehenden Intransparenz der Suchergebnisse. Wieso welche Ergebnisse von Google als relevant erachtet werden und vor allem, welche als irrelevant nach hinten verschoben oder gar komplett entfernt werden, ist für NutzerInnen kaum ersichtlich.

**unzureichende  
Anpassung an  
Nutzerbedürfnisse**

Problematisch ist dies insofern, als Google hier für seine NutzerInnen entscheidet, was für diese relevant ist, ohne deren Bedürfnisse zu kennen. Zudem berücksichtigt die statistische Messung von Relevanz nur unzureichend Möglichkeiten zur Bestimmung qualitativer Faktoren, so dass selbst bekannte Bedürfnisse nicht effektiv berücksichtigt werden können. Das zeigt sich z. B. hinsichtlich der Schwierigkeitsgrade von Dokumenten:

„Da keine Unterscheidung des Schwierigkeitsgrads der erfassten Dokumente stattfindet, tauchen Dokumente unterschiedlicher Verständlichkeit in den Trefferlisten auf. Während die Suche nach einer Krankheit in einer medizinischen Fachdatenbank nur Fachartikel hervorbringt und in einer Magazin-Datenbank nur für den Laien geschriebene Artikel, kann eine Internet-Recherche neben weiteren Formen beide genannten Arten von Artikeln hervorbringen.“ (Lewandowski 2005, S. 97)

<sup>9</sup> [google.com/support/webmasters/bin/answer.py?hl=de&answer=35769](http://google.com/support/webmasters/bin/answer.py?hl=de&answer=35769) (zuletzt aufgerufen am 17.05.2010).



Somit erfordert der Umgang mit Google mehr Medienkompetenz als es die Einfachheit des Suchvorgangs suggeriert.

Gravierender fallen natürlich die Einflussnahmen ins Gewicht, bei denen Google Ergebnisse komplett ausselektiert. Häufig handelt es sich dabei lediglich um Duplikate oder Spam. Allerdings wird Google auch immer wieder bezichtigt, Zensurmaßnahmen zu unterstützen. Für Aufsehen sorgte in diesem Zusammenhang insbesondere der Start der chinesischen Version der Suchmaschine, *Google.cn*. Diese wurde so konzipiert, dass sie der restriktiven Informationspolitik der chinesischen Regierung entgegen kommt und dem Regime unliebsame Ergebnisse filtert (Grossman 2006). Google reagierte damit auf die Blockade der chinesischen Sprachversion von *Google.com*, die diesen Wünschen nicht entsprach. Hierin sahen viele eine Umkehrung des inoffiziellen internen Mottos „Don't be evil“, und warfen dem Konzern eine Priorisierung ökonomischer Interessen vor: „Did the virtuous Google just sell out its honor?“ (ebd.), fragte etwa das Magazin *Time* kritisch. Offensichtlich scheint Googles Entgegenkommen von der Regierung jedoch nicht immer als ausreichend empfunden worden zu sein, weshalb es auch zu Blockierungen von *Google.cn* kam (Kremp 2009). Inzwischen werden BesucherInnen von *Google.cn* auf die Hongkonger Version *Google.com.hk* weitergeleitet, die nicht den rigiden Zensurmaßnahmen der chinesischen Regierung unterworfen ist. Allerdings bleiben viele der Google-Dienste im chinesischen Festland weiterhin blockiert<sup>10</sup>.

### **Filter und Zensur**

Auch in westlichen Ländern konnten Zensurmaßnahmen beobachtet werden. So werden z. B. vom deutschen *Google.de* einige Inhalte mit vermeintlich nationalsozialistischem Gedankengut entfernt, die in anderen Sprachversionen enthalten sind (Schröder 2002). „Googelt“ man etwa bei *Google.at* den Begriff „stormfront“, erhält man als ersten Treffer einen Verweis auf die populäre nationalsozialistische Webseite *stormfront.org*. Führt man diese Suche auf *Google.de* aus, erscheint als erster Treffer ein englischer Wikipedia-Artikel über diese Seite und darunter zudem der unauffällige Hinweis: „Aus Rechtsgründen hat Google 14 Ergebnis(se) von dieser Seite entfernt.“<sup>11</sup>

### **Geotargeting und lokale Unterschiede**

So extrem diese Beispiele sind, verdeutlichen sie doch, dass auch der Standort des verwendeten Computers ein nicht zu vernachlässigender Faktor bei der Zusammenstellung von Googles Ergebnislisten ist. Diese kann von Google anhand der IP-Adresse des Rechners durch sogenanntes *Geotargeting* festgestellt werden.

Die beschriebenen Eingriffe Googles zur Herstellung von vermeintlich relevanten Ergebnissen bieten sicherlich allerhand Anlass zu Kritik. Dazu führt auch eine weitere Maßnahme des Konzerns, die dazu dienen soll, möglichst relevante Ergebnisse zu liefern: Die Speicherung und Auswertung von Nutzerdaten (dazu z. B. Reischl 2008, S. 43 ff.). In einem offiziellen Google-Blogeintrag mit dem Titel „The future of search“ wird die hintergründige Strategie beleuchtet:

### **Speicherung und Auswertung von Nutzerdaten**

„Maybe the search engines of the future will know where you are located, maybe they will know what you know already or what you learned earlier today, or maybe they will fully understand your preferences because you have chosen to share that information with us. We aren't sure which personal signals will be most valuable, but we're investing in research and

<sup>10</sup> Eine Übersicht über den aktuellen Status der Verfügbarkeit von Google-Diensten kann hier aufgerufen werden [google.com/prc/report.html](http://google.com/prc/report.html) (zuletzt aufgerufen am 16.05.2010).

<sup>11</sup> Die Suche wurde am 10.07.2009 durchgeführt. Die verschiedenen Sprachversionen sind unter [google.com/language\\_tools](http://google.com/language_tools) erreichbar.

experimentation on personalized search now because we think this will be very important later.”<sup>12</sup>

**interpretative Eingriffe –  
Elemente semantischer  
Suche**

Wie oben herausgestellt, sind einige dieser Zukunftsentwürfe bereits Realität. Schon jetzt erkennt Google den Standort einer Suchanfrage recht präzise: Eine in Wien gestellte Suchanfrage nach dem Wort „Pizza“ listet etwa „Lokale Branchenergebnisse für Pizza im Umkreis von Wien“, inklusive Verweis auf den Kartenservice *Google Maps*. Es ist damit zu rechnen, dass derart interpretative Eingriffe auf Basis von Nutzerdaten in Zukunft eine noch größere Rolle bei der Websuche spielen werden. Unter dem Begriff des *Semantischen Webs*, der auf einen Vorschlag des Web-Pioniers Tim Berners-Lee (1998) zurückzuführen ist, wird dieser Trend seit Jahren diskutiert. Die bisherigen Möglichkeiten semantischer Suche erscheinen zwar einerseits noch eher begrenzt, andererseits werden sie aber auch bereits praktisch angewendet. Neben den typischerweise als semantische Suchmaschinen beschriebenen Anbietern wie *Wolfram Alpha*, übt auch Google sich mitunter bereits in semantischer Interpretation. Gibt man etwa „Was ist die Hauptstadt von Frankreich?“ (ohne Anführungszeichen) in die Suchmaske ein, erscheint als erstes Ergebnis nicht wie üblich ein Link zu einer Webseite mit den angefragten Begriffen, sondern bereits die Antwort (Abbildung 2.1-1; zu ähnlichen Funktionen siehe auch 2.3, S. 12). Auch erkennt Google etwa, dass bei der Eingabe von „Gandhi bio“ „bio“ für „biography“ steht, während es bei „bio warfare“ als „biological“ interpretiert wird (Levy 2010). Die Auswertung von Suchanfragen ermöglicht ebenso die Korrektur von falsch geschriebenen Wörtern. Generell stellt die semantische Interpretation jedoch eine Herausforderung für jede Suchmaschine dar, da es häufig schwierig ist, die „richtigen“ Zusammenhänge automatisiert zu erkennen.



Abbildung 2.2-1: Semantische Interpretation eines Suchergebnisses  
Anmerkung: Suche durchgeführt am 12.07.2009.

<sup>12</sup> [googleblog.blogspot.com/2008/09/future-of-search.html](http://googleblog.blogspot.com/2008/09/future-of-search.html)  
(zuletzt aufgerufen am 12.07.2009).

Zusammenfassend lässt sich feststellen, dass der gewöhnlich nur wenige Zehntelsekunden dauernde, unscheinbare Suchvorgang ein komplexes Zusammenspiel unzähliger Faktoren ist.<sup>13</sup> Diese sind für durchschnittliche NutzerInnen weder hintergebar, noch deutlich erkennbar. Somit stellt sich eine Frage der im Anschluss nachgegangen wird: Welche Einflussmöglichkeiten bleiben den NutzerInnen im Umgang mit der „Black Box“ Google?

## 2.3 Einflussmöglichkeiten der NutzerInnen

So gravierend und mitunter auch kritikwürdig die in 2.2 beschriebenen Einflussnahmen Googles auch erscheinen mögen, stellen sie letztlich wohl einen kaum zu unterschätzenden Erfolgsfaktor der Suchmaschine dar. Denn Universalsuchmaschinen wie Google „(...) bedienen die Informationsbedürfnisse von in der Regel ungeschulten Nutzern und müssen sich auf diese spezielle Nutzergruppe einstellen“ (Lewandowski 2005, S. 34). Google leistet mit seiner Selektion und Hierarchisierung von Webseiten genau dieses. Wie sich bereits an der weiten Akzeptanz und Nutzung zeigt, scheinen diese Bedürfnisse durchaus befriedigt zu werden – zumindest besser als von der Konkurrenz.

Bisherige Analysen des Suchverhaltens (z. B. Machill et al. 2003; Fallows 2005; weitere überblickartig bei Lewandowski 2005, S. 34 ff.) ergeben ein recht einheitliches und klares Bild von durchschnittlichen Suchmaschinen-NutzerInnen: In der Regel geben diese nur wenige Begriffe ein, nutzen keine erweiterten Suchfunktionen und schauen sich allenfalls die ersten 10 Resultate der Ergebnisse an (d. h. die erste Bildschirmseite). Zwar bietet auch Google diverse Möglichkeiten zum Stellen differenzierter Suchanfragen, wodurch sich die in 2.2 problematisierten Einflussnahmen minimieren ließen. Jedoch scheinen diese manipulierenden Maßnahmen von den meisten NutzerInnen weniger als Problem wahrgenommen zu werden, denn als Lösung:

„Interestingly, search engine users are easily satisfied. When asked about the quality of the results they achieve, they usually express satisfaction with their searching strategies (...). Therefore, they don't see why they should apply sophisticated searching strategies to other search systems. If Google is able to return relevant results for a one-word query, then every other system should be able to do the same.“ (Lewandowski 2008, S. 262)

Bevor hier weiter auf die Einflussmöglichkeiten von NutzerInnen eingegangen wird, muss daher zunächst betont werden, dass es offenbar gerade die Unnötigkeit der eigenen Einflussnahme ist, die geschätzt wird. Google nimmt seinen NutzerInnen diese Tätigkeit ab, bietet Orientierung im Informationsdickicht des World Wide Webs, indem Nicht-Wissen einen eigenen Wert erhält (vgl. Degele 2007, S. 63): „Selektion und Aussonderung durch Abwehr und Vermeidung“ (ebd., S. 64) sind in diesem Kontext hochgeschätzte Dienstleistungen, die von Google effektiv geliefert werden. Diese Effektivität und

**Einflussnahme Googles  
als Erfolgsfaktor**

**Studien zum  
Suchverhalten: Einfache  
Suchen bevorzugt**

**Relevanzherstellung  
durch Google erwünscht**

<sup>13</sup> Diese Komplexität macht sich auch durch eine energieaufwendige Infrastruktur bemerkbar. Harvard-Physiker Alex Wissner-Gross sorgte mit einer Studie für Aufsehen, wonach angeblich zwei Google-Suchanfragen mit sieben Gramm etwa so viel CO<sub>2</sub> verursachen, wie das Kochen einer Tasse Tee. Google dementierte rasch und korrigierte den Wert auf 0,2 Gramm pro Suchanfrage. Zudem zeigte sich, dass Wissner-Gross falsch wiedergegeben wurde. Der Fall ist inklusive weiterführender Verweise nachzulesen im offiziellen Google-Blog: [googleblog.blogspot.com/2009/01/powering-google-search.html](http://googleblog.blogspot.com/2009/01/powering-google-search.html) (zuletzt aufgerufen am 12.07.2009).

### **Mitgestaltung qua Nutzung**

damit auch der Erfolg von Google bestehen gerade darin, die gewünschten Informationen bereitzustellen, *ohne* nähere Kenntnisse über strategisch sinnvolle Suchbegriffe, komplexe Filterfunktionen und ihre Anwendung zu erfordern. Der Trend zu automatisch ergänzten Sucheingaben (wie er auch von Google bereits verwendet wird), stellt ein Beispiel dafür dar, wie Suchmaschinen ihren ungeschulten Usern entgegenkommen (Griffiths/Brophy 2005, S. 541). Daran zeigt sich auch beispielhaft, wie qua Nutzung Google von den AnwenderInnen mitgestaltet wird.

### **auch diverse erweiterte Suchoptionen verfügbar**

Die vergleichsweise wenigen NutzerInnen, die zusätzliche Eingriffsmöglichkeiten wünschen, erhalten über die *Erweiterten Suchoptionen*, diverse Möglichkeiten ihre Suche selbständig einzuschränken und zu präzisieren. So lassen sich anstatt der standardmäßigen „Und“-Suche ebenso eine „Oder“-Suche, genaue Wortfolgen, oder auch gezielt ausschließende Begriffe eingeben. Der sprachliche, regionale und zeitliche Suchraum kann eingegrenzt werden und auch spezielle Webseitenbereiche (z. B. Titel oder URL) oder Domains können gezielt durchsucht werden. Seit Mai 2010 rückt Google einige erweiterte Suchoptionen mehr in den Vordergrund, indem sie am Seitenrand der Ergebnisliste platziert werden und so eine nachträgliche Umsortierung der Ergebnisse ermöglichen. Zudem können nun auch u. a. gezielt Blogs und Microblogs (z. B. *Twitter* und *FriendFeed*) durchsucht werden, wobei die Ergebnisse für Microblogs in Echtzeit aktualisiert werden.

### **gezielte Suchen mit Hilfe von Operatoren**

Wie bereits in 2.2 (S. 10) für das Beispiel einer Hauptstadtsuche beschrieben, lassen sich auch einige Anfragen so stellen, dass Google nicht lediglich nach den eingegebenen Wörtern sucht, sondern diese bereits interpretiert und somit mitunter präzisere Ergebnisse liefert. Bislang sind diese Möglichkeiten jedoch noch recht begrenzt. So lassen sich z. B. aktuelle Wetter-, Börsen- und Sportdaten direkt anzeigen, oder Rechenaufgaben und Währungsumrechnungen werden bei entsprechender Eingabe unmittelbar von Google übernommen.<sup>14</sup> Eine Suche nach „10+5“ führt z. B. zunächst nicht zur Verlinkung von Webseiten mit diesem Inhalt, sondern zum Ergebnis der mathematischen Aufgabe. Dabei ist es natürlich hilfreich, die zugrundeliegenden Befehle (die sogenannten Operatoren) zu kennen. Gibt man etwa „define:“ vor dem Suchwort ein, versucht Google gezielt Definitionen des Begriffes anzuzeigen.<sup>15</sup>

### **Einflussmöglichkeiten für Webmaster**

Weiterhin gibt es nicht nur für Inhaltssuchende, sondern auch für Inhaltsanbieter beschränkte Möglichkeiten auf die Google-Ergebnisse einzuwirken, vor allem durch das gegenseitige Verlinken zur Erhöhung des PageRanks. Professionell geschieht das im Marketingbereich der Suchmaschinenoptimierung (vgl. S. 8), zudem versuchen auch Hacker Lücken des Systems für ihre Zwecke auszunutzen (vgl. S. 8). Durch so genannte *Google-Bomben* lassen sich etwa durch Setzen mitunter zahlreicher Links Webseiten bei bestimmten Suchanfragen gezielt in die oberen Ergebnisse befördern. Ein prominentes Beispiel dafür konnte lange Zeit beim Eingeben des Wortpaares „miserable failure“ beobachtet werden: Als erstes Ergebnis wurde die offizielle Biographie des damals amtierenden US-Präsidenten George W. Bush angezeigt – freilich ohne, dass diese weder das eine noch das andere Wort enthielt.<sup>16</sup> Laut Google sind solche Manipulationen allerdings nur in sehr eingeschränktem Maße möglich: „Googlebombs very rarely happen for common queries, because the lack of any relevant results for that phrase is part of why a Googlebomb can work“ (Moulton/Carattini 2007).

<sup>14</sup> [google.com/intl/de/help/features.html#essentials](http://google.com/intl/de/help/features.html#essentials) (zuletzt aufgerufen am 14.07.2009).

<sup>15</sup> Einige dieser Befehle finden sich hier: [google.de/help/cheatsheet.html](http://google.de/help/cheatsheet.html) (zuletzt aufgerufen am 14.07.2009).

<sup>16</sup> Siehe dazu z. B.: [heise.de/newsticker/Google-ueberrascht-bei-der-Suche-nach-erbaermlichen-Versagern--/meldung/42679](http://heise.de/newsticker/Google-ueberrascht-bei-der-Suche-nach-erbaermlichen-Versagern--/meldung/42679) (zuletzt aufgerufen am 14.07.2009).

Letztlich stellen derartige Versuche der Einflussnahme auch kein Umgehen oder Manipulieren des Google-Algorithmus dar, sondern viel mehr seine besonders effektive Anwendung. In gravierender Weise macht dieser sich jedoch weniger bei exotischen Begriffssuchen bemerkbar als bei äußerst populären, die eine entsprechend extreme Selektion und Hierarchisierung durch Google erfordern. Anbieter von Themen, zu denen es viele konkurrierende Webseiten gibt, sind somit in besonderem Maße von dieser Ordnungslogik abhängig und auf Suchmaschinenoptimierung angewiesen. Oder negativ formuliert: Tun sie dies nicht, bleiben sie für Google-NutzerInnen praktisch unsichtbar. Durch die exponierte Position, die Google im World Wide Web einnimmt, kann dies von Anbietern, die auf Auffindbarkeit ihrer Inhalte wert legen, kaum ignoriert werden. Das gilt auch umgekehrt: Möchte man, dass Inhalte z. B. aus urheber- oder datenschutzrechtlichen Gründen *nicht* gefunden werden, sollte die betreffende Webseite entsprechend gestaltet sein. Gegebenenfalls ermöglicht Google hierzu auch die nachträgliche Entfernung von indizierten Seiten.<sup>17</sup>

Für alle beschriebenen Funktionsprinzipien ist zu beachten, dass sie im ständigen Wandel sind. Google arbeitet fortlaufend an der „Verbesserung“ seiner Dienste und betont dies in seiner Unternehmensphilosophie:

„Google sieht seine Spitzenstellung als Ausgangspunkt und nicht als Endpunkt. Eine gut funktionierende Sache wird von Google durch Innovationen und schrittweises Vorgehen auf unerwartete Weise immer weiter verbessert.“ (Google 2009)

„Verbesserung“ bedeutet in diesem Kontext natürlich eine effektivere Websuche mit einer entsprechend höheren Chance die gewünschten Informationen durch „Googeln“ zu erhalten. Dazu trägt etwa auch die Möglichkeit bei, durch Anmeldung mit einem Google-Account seine Suchanfragen speichern und von Google auswerten zu lassen. Derartige „Verbesserungen“ bedeuten also meist auch weitere selektierende und hierarchisierende Eingriffe der Suchmaschine. Bei der im Trend liegenden semantischen Websuche (vgl. 2.2) werden etwa Begriffe von Suchmaschinen maschinell interpretiert – selbstverständlich ohne Garantie, dass sich diese Interpretation mit der tatsächlichen Intention der NutzerInnen deckt. Geschieht dies zudem mit der derzeit herrschenden Intransparenz der Rankings bei gleichzeitig geringem Wissen über diese Prozesse seitens der User, intensiviert sich diese Problematik noch weiter. In jedem Fall stellt jede zusätzliche ordnende Maßnahme eine Schwächung der Eingriffsmöglichkeiten und eine Stärkung der Hierarchisierungs-, Selektions- und somit auch der Gatekeeper-Funktion der Suchmaschine dar.

Wie die Änderungen der Ergebnisanzeige im Mai 2010 zeigen, können jedoch auch neue Ordnungsfunktionen zu zusätzlichen Eingriffsmöglichkeiten führen. So kann nun etwa durch das sogenannte *Wunderad* eine Art Kategorienbaum aus verwandten (jeweils anklickbaren) Suchbegriffen angezeigt werden (Abbildung 2.3-1) und die Ergebnisse lassen sich durch einen *Zeitstrahl* ordnen. Es bedarf weiterer Untersuchungen, um die Effektivität dieser Möglichkeiten einzuschätzen und Nutzungsstudien müssen klären, ob und wie sie von den AnwenderInnen angenommen werden. Es ist anzunehmen, dass sie in absehbarer Zukunft keinen grundlegenden Effekt auf die Dominanz der hierarchischen Ergebnisstrukturen haben werden. Ohnehin werden auch hier die Ergebnisse letztlich hierarchisiert dargestellt und intransparente Algorithmen wirken hier ebenso, so dass nur bedingt von einer Stärkung der Eingriffsmöglichkeiten gesprochen werden kann.

**Zwang zur Berücksichtigung des Google-Algorithmus**

**Funktionen im ständigen Wandel**

**Tendenz zu interpretierenden Eingriffen**

**Schwächung der Eingriffsmöglichkeiten**

**teilweise neue Such- und Ordnungsfunktionen**

<sup>17</sup> [google.com/support/webmasters/bin/answer.py?hl=de&answer=35301](http://google.com/support/webmasters/bin/answer.py?hl=de&answer=35301) (zuletzt aufgerufen am 14.07.2009).

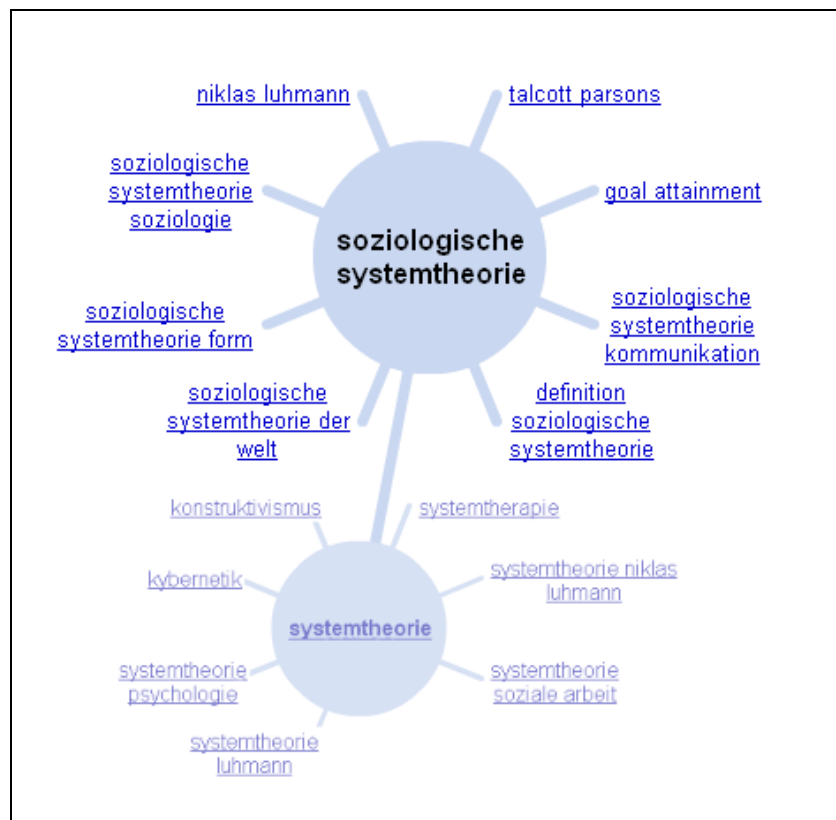


Abbildung 2.3-1: Ergebnisanzeige im „Wunderrad“

Anmerkung: Suche „Systemtheorie“ mit Öffnung des vorgeschlagenen Begriffs „Soziologische Systemtheorie“; durchgeführt am 14.05.2010.

## 2.4 Einsatz in der Wissenschaft

Inwiefern ist nun die populäre Suchmaschine Google mit ihren oben beschriebenen Funktionsprinzipien relevant für die interne bzw. externe Wissenschaftskommunikation?

Dieser Frage muss differenziert nachgegangen werden, da zu berücksichtigen ist, wer hier was, wie und zu welchem Zweck sucht. Wie oben beschrieben, greift Google zwar durch Hierarchisierungs- und Selektionsmechanismen beträchtlich in den Suchvorgang ein, es bleibt aber letztlich den NutzerInnen überlassen, wofür sie die Suchmaschine einsetzen, wie sie suchen, und vor allem, wie sie mit den Ergebnissen umgehen. Generell erscheinen hier vor allem zwei Einsatzgebiete von Bedeutung: Die Suche nach wissenschaftsrelevanten Inhalten (2.4.1) und die nach Personen und Institutionen (2.4.2). Zudem stellt sich die Frage, welche Auswirkungen Google auf die externe Wissenschaftskommunikation (Öffentlichkeitsarbeit) hat, bzw. wie diese auf die Wissenschaft rückwirken (2.4.3).

### 2.4.1 Suche nach wissenschaftsrelevanten Inhalten

Auf dem ersten Blick ließe sich argumentieren, eine Universalsuchmaschine habe keine signifikante Bedeutung für die meist extrem spezialisierten Informationsbedürfnisse in den Wissenschaften. Tatsächlich wurde bereits oben darauf hingewiesen, dass bestimmte Inhalte von Universalsuchmaschinen gar nicht indexiert und somit auch nicht in den Ergebnissen aufgelistet werden. Es ist davon auszugehen, dass dies auch für zahlreiche akademische Publikationen gilt, etwa weil sie häufig passwortgeschützt sind (Lewandowski/Mayr 2006).

Diesem Problem wird teilweise entgegen gewirkt, indem bei einigen Suchanfragen auch Ergebnisse aus der akademischen Suchmaschine Google Scholar (siehe auch Abschnitt 3) in den regulären Google-Suchergebnissen gelistet werden, etwa wenn ein Fachbegriff wie „Adenosin-3',5'-monophosphat“ eingegeben wird (Abbildung 2.4-1 auf der nächsten Seite, rote Rahmen).<sup>18</sup> Eine vergleichende Studie (Brophy/Bawden 2005) kam sogar zu dem Ergebnis, dass Google relevante Inhalte quantitativ besser abdeckt als die untersuchte Bibliotheksdatenbank, wobei aber die inhaltliche Qualität der Ergebnisse bei Google schlechter abschnitt.

Dies ist auch an dem oben aufgeführten Beispiel erkennbar: Zwar werden durchaus Fachtexte angezeigt, das erste Suchergebnis (grüner Rahmen) verweist jedoch auf Wikipedia, also einer Enzyklopädie, die sich vor allem an ein Laien- und nicht an ein Fachpublikum richtet.<sup>19</sup> Es lassen sich durch „Googeln“ also durchaus wissenschaftliche Publikationen finden, bereits aufgrund der Vermengung solcher Veröffentlichungen mit verschiedensten anderen Webinhalten erscheint die Suchmaschine für diesen Zweck allerdings wenig effektiv.

Das Hierarchisierungsprinzip (das tendenziell populäre Webseiten bevorzugt) bewirkt zudem, dass in der Regel beliebte Inhalte vorne in den Ergebnissen positioniert werden, die selbstverständlich nicht unbedingt wissenschaftlichen Qualitätskriterien entsprechen müssen.<sup>20</sup> Dies wird insbesondere dann zum Problem, wenn nach Begriffen gesucht wird, die entgegen dem oben gewählten Beispiel auch für viele LaiInnen von Interesse sind. Eine Google-Recherche nach technischen Papieren zum Einsturz des New Yorker World Trade Centers kann beispielsweise schnell zu einer Reihe von Papieren und Webseiten führen, die eine Sprengung der Gebäude als Ursache nahe legen – was freilich nicht dem fachlichen Kanon entspricht (dazu auch: König 2009). Aufgrund des Charakters einer Universalsuchmaschine räumt Google auch Inhalten Relevanz ein, die in der *scientific community* über keine oder nur marginale Bedeutung und Reputation verfügen (siehe auch 2.4.3/2.4.3). Zusammen mit der nur unvollständigen Indexierung akademischer Publikationen, muss die Suchmaschine daher für die Recherche nach einschlägigen wissenschaftlichen Veröffentlichungen als tendenziell disfunktional bewertet werden.

**akademische  
Publikationen nur  
begrenzt auffindbar**

**Einbindung von  
Ergebnissen aus Google  
Scholar**

**Vermengung  
wissenschaftlicher  
Publikationen und  
anderer Webseiten**

**Google-Ranking  
entspricht nicht  
unbedingt akademischer  
Relevanz**

<sup>18</sup> Erkennbar ist dies anhand typischer Zusatzinformationen, die Google Scholar bereit stellt, etwa Autorennamen und Erscheinungsjahr, sowie teilweise Verlinkungen zu Zitationen des jeweiligen Textes.

<sup>19</sup> Dies schließt selbstverständlich nicht aus, dass auch Wikipedia in mancherlei Hinsicht wissenschaftsrelevant ist (König/Nentwich 2009).

<sup>20</sup> Brooks (2004) bezeichnet das Ordnungsprinzip von Google daher auch als „Lay Indexing“.

Google Adenosin-3',5'-monophosphat  Suche [Erweiterte Suche](#)

Suche:  Das Web  Seiten auf Deutsch  Seiten aus Deutschland

Web [+ Optionen anzeigen...](#) Ergebnisse 1 - 10 von ungefähr 875.000 für Adenosin-3',5'-monophosphat. (0,29 Sekunden)

Tipp: Anstatt auf "Suche" zu klicken, können Sie auch die Eingabetaste drücken, um Zeit zu sparen.

**Cyclic adenosine monophosphate - Wikipedia, the free encyclopedia** - [ [Diese Seite übersetzen](#) ]  
 Cyclic **adenosine monophosphate** (cAMP, cyclic AMP or 3',5'-cyclic **adenosine monophosphate**) is a second messenger important in many biological processes. ...  
[Synthesis and decomposition](#) - [Functions](#) - [Pathology](#) - [See also](#)  
[en.wikipedia.org/.../Cyclic\\_adenosine\\_monophosphate](http://en.wikipedia.org/.../Cyclic_adenosine_monophosphate) - [Im Cache](#) - [Ähnlich](#)

**Effect of adrenocorticotropin or adenosin-3' 5'-monophosphate on ...**  
 Scriba, Peter Christian (1964): Effect of adrenocorticotropin or **adenosin-3',5'-monophosphate** on in vitro adrenal synthesis. ...  
[epub.ub.uni-muenchen.de/8777/](http://epub.ub.uni-muenchen.de/8777/) - [Im Cache](#) - [Ähnlich](#)

**Cyclisches Adenosin-3' 5'-monophosphat in ...**  
 von D Becker - 1973 - [Zitiert durch: 7](#) - [Ähnliche Artikel](#)  
 Cyclic **Adenosine-3':5'-monophosphate** in Translocation Tissues of Plants ? .... tion in barley endosperm by cyclic 3',5'-adenosine **monophosphate** and adeno- ...  
[www.springerlink.com/index/G668270064P1561J.pdf](http://www.springerlink.com/index/G668270064P1561J.pdf) - [Ähnlich](#)

**Wirkung von cyclischem Adenosin-3', 5'-monophosphat und cyclo (Lys ...**  
 von G LINDNER - 1987  
 Wirkung von cyclischem **Adenosin-3', 5'-monophosphat** und cyclo (Lys-Pro) +HCL als neuronotrophe Faktoren in der Kultur. G LINDNER, G GROSSE, M WENZEL ...  
[cat.inist.fr/?aModele=afficheN&cpsidt=8350998](http://cat.inist.fr/?aModele=afficheN&cpsidt=8350998) - [Ähnlich](#)

**Changes in Adenosine 3', 5'-Monophosphate Level and Protein Kinase ...** - [ [Diese Seite übersetzen](#) ]  
 von H Higashino - 1974 - [Zitiert durch: 4](#) - [Ähnliche Artikel](#)  
**adenosine 3', 5'-monophosphate**. in liver nuclei. The nuclei were isolated with non-aqueous solvents, to eliminate that com- ...  
[j.b.oxfordjournals.org/cgi/reprint/75/1/189.pdf](http://j.b.oxfordjournals.org/cgi/reprint/75/1/189.pdf) - [Ähnlich](#)

**Cyclisches Adenosin-3' 5'-monophosphat**  
 Cyclisches **Adenosin-3',5'-monophosphat**. Akronym. cAMP. Testparameter. 2113/CAMP · cAMP (Plasma). 2115/CAMPU. cAMP (Urin). Material ...  
[www.labseelig.de/analysenspektrum/DOCS/00/000065.html](http://www.labseelig.de/analysenspektrum/DOCS/00/000065.html)

**Cyclic adenosine 3' 5'-monophosphate response element binding ...** - [ [Diese Seite übersetzen](#) ]  
 von TE Meyer - 1993 - [Zitiert durch: 341](#) - [Ähnliche Artikel](#) - [Alle 3 Versionen](#)  
 Cyclic **adenosine 3',5'-monophosphate** response element binding protein (CREB) and related transcription-activating deoxyribonucleic acid-binding proteins. ...  
[www.ncbi.nlm.nih.gov/pubmed/8319595](http://www.ncbi.nlm.nih.gov/pubmed/8319595) - [Ähnlich](#)

**Elevated Cyclic Adenosine 3', 5' Monophosphate Enhances Lactic ...**  
 von TL Ratliff - 1981 - [Ähnliche Artikel](#)  
 Elevated Cyclic **Adenosine 3', 5' Monophosphate** Enhances. Lactic Acid Production by Streptococcus lactis. TIMOTHY L. RATLIFF 1 and DWIGHT E. TALBURT ~ ...  
[jds.fass.org/cgi/reprint/64/3/391.pdf](http://jds.fass.org/cgi/reprint/64/3/391.pdf) - [Ähnlich](#)

**Adenosine 3' 5'-Monophosphate Phosphodiesterase**  
 von AW Murray - 1971 - [Zitiert durch: 21](#) - [Ähnliche Artikel](#)  
**Adenosine 3',5'-Monophosphate** Phosphodiesterase in the Growth Medium of Physarum polycephalum. Abstract. The acellular slime mold Physarum polycephalum ...  
[www.sciencemag.org/cgi/reprint/sci;171/3970/496.pdf](http://www.sciencemag.org/cgi/reprint/sci;171/3970/496.pdf) - [Ähnlich](#)

Abbildung 2.4-1: Einbindung von Ergebnissen aus Google Scholar  
 Anmerkung: Suche durchgeführt am 11.12.2009.

**jedoch auch hilfreiche  
 Einsatzmöglichkeiten**

Nun beschränkt sich der wissenschaftliche Informationsbedarf natürlich nicht auf Fachartikel. Häufig werden Themen nur tangiert, oder Begriffe müssen lediglich kurz nachgeschlagen werden. Entsprechend sind hier kurze, prägnante Informationen erforderlich. Universalsuchmaschinen können in solchen Fällen äußerst hilfreich sein und ersparen den Gang in die Bibliothek oder andere aufwendige Recherchemethoden, wie auch Lewandowski bestätigt:



„Research shows that with limited time, even for scientific information needs, search engines could be a better choice for the searcher than the library’s offerings.” (Lewandowski 2008, S. 262)

Tatsächlich muss aufgrund der hohen Anwenderfreundlichkeit davon ausgegangen werden, dass „Googeln“ längst ein Teil der wissenschaftlichen Alltagspraxis geworden ist. Eine Studie, für die ForscherInnen an drei schwedischen Universitäten beobachtet wurden, stellt fest:

„For many researchers, especially in the sciences, Google is the first choice for information – all kinds of information. The researchers use Google for scientific information, looking for everything from methodological information to ISSNs, and some even state having moved from subject specific databases to Google (and Google Scholar).” (Haglund/Olsson 2008, S. 55)

Es wäre nach unserer Auffassung zu leichtfertig, dies pauschal als Gefahr für die wissenschaftliche Qualität zu bewerten – schließlich wird durch die Beschleunigung von Recherchevorgängen auch wertvolle zusätzliche Zeit verfügbar. Viel mehr hängt das Für und Wider des Suchmaschineneinsatzes von der konkreten Nutzungsweise ab, d. h. vor allem auch von den NutzerInnen. Z. B. scheint sich Google hervorragend dafür zu eignen, Fachzeitschriften im Netz zu finden (Thelwall 2002).

Die oberflächliche Anwenderfreundlichkeit kann sich jedoch durchaus auch als „false friend“ erweisen. Denn während das Eingeben der Suchanfrage zwar kinderleicht ist, erfordert die Qualitätsbewertung der Ergebnisse ein beträchtliches Maß an Medien- und Fachkompetenz auf der Nutzerseite:

„The difficulty is that information – through Google – is seen to be both abundant and cheap. Because of this rapid ranking and return, ‘anyone’ can manage it. Actually, the abilities required to assess information are difficult and costly to obtain.” (Brabazon 2006, S. 163)

In den Ergebnislisten mischen sich unter Umständen hochgradig renommierte Inhalte mit irrelevanten, oder gar schlichtweg falschen Informationen. Die Hyperlinkförmigkeit der Ergebnisse löst diese dabei aus ihrem ursprünglichen Zusammenhang. Eine solche Rekontextualisierung erschwert die Einschätzung der Qualität zusätzlich, da Quellen nicht immer ohne weiteres eindeutig erkennbar sind und somit häufig als entsprechende Indikatoren wegfallen, bzw. nur durch Umwege erreichbar sind.

Diese Kombination aus scheinbarer Einfachheit der Recherche bei gleichzeitig schwer kalkulierbarer Qualität erweist sich insbesondere für den akademischen Nachwuchs mitunter als Fallstrick. Das macht Informationswissenschaftler Michael Zimmer (2009) an einem Beispiel aus seiner Lehrtätigkeit deutlich: Ein Student sollte über die Funktionsweise von Google referieren und fiel auf einen Aprilscherz herein, den er ironischerweise durch „Googeln“ fand und der auch von dem Suchmaschinenanbieter selbst in die Welt gesetzt wurde. Unter der durchaus seriös wirkenden URL [www.google.com/technology/pigeonrank.html](http://www.google.com/technology/pigeonrank.html) (zuletzt aufgerufen am 18.01.2010) wird behauptet, Googles grundlegendes Funktionsprinzip sei nicht der *PageRank*, sondern der *PigeonRank*, bei dem spezielle Tauben eingesetzt würden, um die Relevanz von Webseiten zu bestimmen. Der Student gelangte an diese Information, indem er die Frage „how does google work?“ an Google stellte und dem ersten Link in der Ergebnisliste folgte. Dabei übersah er offensichtlich den Hinweis am Ende der Seite, der diese als Aprilscherz ausweist.

Freilich handelt es sich bei dieser Anekdote um ein Extrembeispiel. Jedoch spiegelt sie eine Grundproblematik wider, die sich auch in einigen Studien zum Suchverhalten akademischer NutzerInnen abzeichnet. Befragte BibliothekarInnen der amerikanischen Cornell University gaben an, sowohl Studen-

**„Googeln“ als wissenschaftliche Alltagspraxis**

**Nutzen abhängig von AnwenderInnen**

**Qualitätsbewertung erfordert Kompetenz**

**Kontext schwer einschätzbar**

**Risiko falscher Informationen insbesondere für Nachwuchs hoch**

**Studien: Großes Vertrauen in Google**

tInnen als auch Lehrkräfte seien nicht nur überaus zufrieden mit Google-Ergebnissen, sondern es herrsche auch ein „blind trust and an increasing reliance on search results, especially on whatever appears on the first couple of screens“ (Rieger 2009). Zudem äußerten sie Besorgnis darüber, dass besonders unter Studierenden offenbar Zweckmäßigkeit und Bequemlichkeit dominierende Faktoren im Suchverhalten darstellen. Ein Eindruck, der sich in nachfolgenden Befragungen unter Studierenden und Lehrenden bestätigte:

„Users find Google very intuitive and easy to use and believe that it represents the information space in which they are interested with excellent breadth and depth. Our informants, especially the undergraduate and graduate students, consistently made remarks that reflected their trust in and loyalty to Google.“ (ebd.)

**beliebte  
Informationsquelle unter  
Studierenden**

Eine Studie an einer britischen Universität kam zu dem Ergebnis, dass 45 % der 27 teilnehmenden Studierenden unterschiedlicher Disziplinen Google als erste Informationsquelle nutzten (Griffiths/Brophy 2005). An zweiter Stelle folgte mit nur 10 % der Online-Bibliothekskatalog der Universität. Auch Becker (2003) berichtet von der Dominanz Googles beim Suchverhalten von Studierenden, ebenso wie eine internationale repräsentative Studie, nach der 89 % der befragten Studierenden zunächst eine Suchmaschine zur Informationssuche verwenden, wobei auch hier meist Google gewählt wird (OCLC 2005, S. 1-17 f.).

**jedoch nicht unbedingt  
funktional**

Dabei wären Bibliothekskataloge unter Umständen die bessere Wahl:

„While the preference for very simple search engine approaches is prevalent, it is important to note that this does not mean that students are necessarily best served by this approach. Indeed, it may be that students would get better results using specialist subject gateways, but most students do not take this approach. Exclusive use of any commercial SE [search engine] coupled with a lack of awareness and understanding of peer-reviewed, quality resources is not in the best interest of students or academic staff.“ (ebd., S. 552)

**vereinzelt wird auch von  
kritischem Umgang mit  
Suchmaschinen berichtet**

Demgegenüber nennt eine Studie am Saint Mary's College of California Kursmaterialien als erste Informationsquelle der teilnehmenden Studierenden, die eine durchaus kritische Haltung gegenüber Suchmaschinen einnehmen (Head 2007). Dies führt Head vor allem darauf zurück, dass Studierende unsicher bei der Einschätzung der Qualität externer Ressourcen sind.

**ernstzunehmende  
Konkurrenz für  
Bibliotheken**

So lässt sich zusammenfassend folgern, dass bei aller Vorläufigkeit, unzureichender Generalisierbarkeit und nur teilweise möglicher Vergleichbarkeit bestehender Studien zum Nutzungsverhalten, Suchmaschinen offensichtlich eine Herausforderung für ihre akademischen VerwenderInnen darstellen – nicht nur, aber besonders für den Nachwuchs. Gleichzeitig erscheint auch evident, dass Google längst zu einer ernstzunehmenden Konkurrenz für Bibliotheken im Bereich der Informationssuche geworden ist. Das zeigt sich auch daran, dass diese zunehmend mit eigenen Suchmaschinen auf diese Entwicklung reagieren, wobei teilweise auch Google-Dienste in diesen Suchmaschinen integriert werden.<sup>21</sup>

<sup>21</sup> Z. B. *BASE* (Bielefeld Academic Search Engine), die auch auf Ergebnisse von Google Scholar zugreift.

## 2.4.2 Suche nach Personen und Institutionen

Eine bislang weniger intensiv diskutierte und untersuchte Frage ist, wie Google sich auf die Suche nach Personen und Institutionen auswirkt. Die Dominanz von Google für die schnelle Informationssuche legt nahe, dass die Suchmaschine auch für diese Zwecke zum Einsatz kommt. Beobachtet man die beliebtesten Suchbegriffe, die regelmäßig von Google veröffentlicht werden<sup>22</sup>, fällt auf, dass viele von ihnen eine navigierende Funktion erfüllen. D. h. anstatt eine URL in die Adressleiste des Browsers einzugeben, wird einfach nach der jeweiligen Seite „gegoogelt“ (häufige Suchbegriffe sind etwa „wikipedia“, „facebook“ oder „youtube“).<sup>23</sup>

Dies lässt vermuten, dass auch akademische Personen und Institutionen auf diese Weise gesucht werden, anstatt etwa durch Eingabe der offiziellen URL, oder einer Personensuche über institutionelle Suchmasken. Auch setzen Personalabteilungen Google zur Suche nach Informationen über ihre BewerberInnen ein. Wie WissenschaftlerInnen oder Institute bei Google positioniert werden, dürfte daher durchaus von Bedeutung für diese sein.

Dabei geht es nicht nur um die Frage, ob und wo die gewünschten Seiten in den Ergebnissen auftauchen, sondern vor allem auch darum, welche zusätzlichen – eventuell eben auch unerwünschten – Informationen von Google zugänglich gemacht werden. Denn selbstverständlich garantiert das hier aktive Ordnungsprinzip nicht, dass eine positive Selbstdarstellung auch am Anfang der Ergebnisse auftaucht.

Besonders Personen mit ungewöhnlichen Namen müssen aufgrund der leichteren Auffindbarkeit via Google damit rechnen, dass neben einem offiziellen Profil unter Umständen auch private Informationen bei Google angezeigt werden. Gerade im Zusammenhang mit einer vermehrten aktiven Nutzung des Netzes durch WissenschaftlerInnen, müssen diese davon ausgehen, dass auch ihre Datenspuren gespeichert und für andere zugänglich werden (Nentwich 2009, S. 23). Auch Halavais weist darauf hin, dass Suchmaschinen unsere verschiedenen (Teil-) Identitäten allgemein sichtbar machen, ähnlich wie es bei ländlichen Gesellschaften der Fall ist:

„Public identities are often constructed out of what may be discovered via a search engine, and this affects how we view our colleagues, our friends, and our family members. Drawing heavily on the burgeoning exhibitionist technologies like blogs, social networking sites and photo-sharing sites, search engines help us to create portfolios of the people we interact with. In many cases, this means that they provide us with information on parts of their lives that would otherwise remain hidden, at least to most of the world.” (Halavais 2009, S. 139)

Personen mit einem häufig vergebenen Namen sehen sich dagegen mit dem Problem konfrontiert, dass möglicherweise Informationen über einen Namensvetter erscheinen, die dann eventuell der falschen Person zugeordnet werden. So schildert etwa Halavais den Fall einer Studentin, die ihren Namen mit einer Pornodarstellerin teilt und daher ständig befürchtet, für diese gehalten zu werden (ebd., S. 142).

**Suchanfragen mit navigierender Funktion verbreitet**

**auch relevant für akademische Personen und Institutionen**

**(ungewünschte) negative Informationen ebenso auffindbar**

**verschiedene (Teil-)Identitäten werden sichtbar**

**Verwechslungsgefahr**

<sup>22</sup> [google.com/intl/en/press/zeitgeist/yearend.html](http://google.com/intl/en/press/zeitgeist/yearend.html) (zuletzt aufgerufen am 22.01.2010).

<sup>23</sup> In der Literatur werden derartige Suchanfragen auch als „navigational queries“ bezeichnet (Rieger 2009; Broder 2002).

<b>Identitätsdiebstahl</b>	Zudem ist natürlich auch denkbar, dass negative Informationen von Widersachern bereit gestellt werden, bzw. mittels sogenannten „Identitätsdiebstahls/Nicknapping“ andere sich als die betroffene Person präsentieren (vgl. Nentwich 2009, S. 23). Ebenso kann Google dafür genutzt werden, Informationen zu Personen für entsprechende Missbrauchszwecke zu erlangen.
<b>Datenschutz wird immer mehr zur Herausforderung</b>	Derartige Probleme beschränken sich natürlich nicht auf Google. Die herausragende Stellung, die der Suchmaschine beim Auffinden von Informationen im Web zukommt, gibt ihr aber auch in dieser Hinsicht eine besondere Bedeutung. Dazu trägt auch bei, dass Google unter Umständen noch Kopien von Dokumenten ausfindig macht, die anderswo gelöscht wurden, oder diese selbst noch im Cache aufführt. Mit zunehmendem Alter des noch jungen Webs, dürften derartige Probleme an Sichtbarkeit gewinnen und verstärkt auch in das Bewusstsein von WissenschaftlerInnen gelangen. Dies wird zudem dadurch verstärkt, dass in Zukunft immer mehr „Offline-Objekte“ digital erfassbar werden und auch Audio-, Bild- und Videodaten für Suchmaschinen analysier- und zurechenbar werden – was die Trennung verschiedener Teil-Identitäten weiter erschwert (Čas 2005). Google trägt mit seiner Strategie der breiten Erfassung und Verknüpfung von Informationen besonders zu dieser Problematik bei (Bredow et al. 2010).
<b>auch Chancen für NachwuchswissenschaftlerInnen</b>	Neben solchen Risiken, bringt die Möglichkeit, Personen und Institutionen auch abseits der etablierten Pfade zu „ergoogeln“, unter Umständen auch neue Chancen auf Rezeption und Relevanz. Die traditionellen Strukturen der Reputationsverteilung im Wissenschaftssystem (d. h. vor allem viel zitierte Publikationen in renommierten Fachzeitschriften und -büchern) werden durch Google sicherlich nicht grundsätzlich angegriffen. Google gibt jedoch auch weniger etablierten WissenschaftlerInnen zumindest theoretisch die Chance, ihre Inhalte auch ohne dass diese in Zeitschriften mit rigidem Peer-Review-Verfahren erschienen wären, zugänglich zu machen (z. B. über Blogs). Das gilt insbesondere für die Kommunikation mit LaiInnen, die eher geneigt sein dürften, wissenschaftliche Inhalte über Google zu suchen, als in den einschlägigen Organen (dazu auch 2.4.3).
<b>institutioneller Einsatz der „Benutzerdefinierten Suche“</b>	Ferner sei darauf verwiesen, dass viele wissenschaftliche Institutionen Google durch die Funktion der <i>Benutzerdefinierten Suche</i> unmittelbar in ihre Webseiten einbinden. Dadurch können nicht nur einzelne Webseiten von wissenschaftlichen Einrichtungen durchsucht werden: Durch Sammlung diverser Webseiten lassen sich auch ganze Themenbereiche zielgerichtet erschließen. Dies bietet alternative Zugangs- und Vernetzungsformen gegenüber Formaten wie Linksammlungen. So findet sich beispielsweise auf der Webseite des <i>Netzwerks Technikfolgenabschätzung (NTA)</i> <sup>24</sup> die Möglichkeit, 41 institutionelle Domains der NTA-Mitglieder im Volltext zu durchsuchen. Neben diesen neuen Optionen entstehen hier freilich auch Abhängigkeiten, da man sich qua Nutzung zwangsläufig den intransparenten Google-Algorithmen unterwirft.

---

<sup>24</sup> [netzwerk-ta.net/suche.htm](http://netzwerk-ta.net/suche.htm) (zuletzt aufgerufen am 21.05.2010).

### 2.4.3 Einfluss auf externe Wissenschaftskommunikation (Öffentlichkeitsarbeit)

Selbst wenn WissenschaftlerInnen Google gar nicht nutzen würden (was wie oben beschrieben nicht der Fall ist), bliebe immer noch die gesamtgesellschaftliche Popularität der Suchmaschine zu berücksichtigen, die auch auf das Wissenschaftssystem zurückwirkt. LaiInnen durchsuchen gewöhnlich keine akademischen Datenbanken, um an wissenschaftliche Informationen zu gelangen – aus gutem Grund, schließlich würden diese kaum angemessenes Material für ihren Wissensstand beinhalten, abgesehen davon, dass diese in der Regel kostenpflichtig sind. Traditionell wurden entsprechend für LaiInnen aufbereitete wissenschaftliche Inhalte vor allem durch den Wissenschaftsjournalismus, PR-Abteilungen und andere populärwissenschaftliche Publikationen bereitgestellt. Wie alle traditionellen Wissensautoritäten bekommen diese nun durch das Internet Konkurrenz und Google spielt als Gatekeeper eine entscheidende Rolle für den Zugang zu diesen Informationen (Halavais 2009, S. 97 ff.).

Die Favorisierung populärer Webseiten in den Suchergebnissen kann dazu führen, dass vom Wissenschaftssystem marginalisierte, jedoch für die Öffentlichkeit attraktive Inhalte mit neuer Relevanz versehen werden. Oder umgekehrt betrachtet: Renommierte wissenschaftliche Inhalte (und somit auch ihre AutorInnen und die dahinter stehenden Institutionen) können durch Google marginalisiert werden, wenn diese nicht ausreichend mit hyperlinkförmigen Verweisen ausgestattet sind oder gar nicht erst indiziert werden.

„Googelt“ man etwa nach dem Begriff „Klimaforschung“, erscheint als erster Treffer der „ungefähr“ 220.000 Ergebnisse ein Link zur Seite *klimaforschung.net* (Abbildung 2.4-2). Diese lässt sich grob als eine Webseite der sogenannten „Klimaskeptiker“ einordnen, die eine menschlich verursachte globale Erwärmung für unwahrscheinlich halten – womit sie im Widerspruch zur etablierten Klimaforschung stehen, in der darüber weitgehend Konsens herrscht.

Google bevorzugt hier also nicht den Wissensstand renommierter Institutionen, sondern im Gegenteil: Die Suchmaschine stattet vom Wissenschaftssystem ansonsten marginalisierte Inhalte mit besonderer Bedeutung aus. Es ließen sich weitere Beispiele für eine derartige Relevanzumkehrung finden, etwa alternative Erklärungen zum Einsturz des New Yorker World Trade Centers (König 2009). Dass in Google-Rankings nicht unbedingt Mainstream-Positionen repräsentiert werden, ist bekannt. Halavais (2009, S. 86) schildert etwa, dass in der US-Version eine Suchanfrage nach Martin Luther King lange Zeit eine rassistische Seite als ersten Treffer listete. Soziale Netzwerke, die sich im Web besonders intensiv einem speziellen Thema widmen, haben durch Googles PageRank-Architektur gute Chancen auf ein hohes Ranking:

„Once cliques of clusters of websites are established, they tend to reinforce themselves. A reader might find site A by following a link from site B, and decide to link to both of them.“ (Halavais 2009, S. 60)

Auf diese Weise entstehen Wissensnetzwerke bzw. verstärken sich gegenseitig, was zu einer „Cyberbalkanization“<sup>25</sup> führen könnte. So nennt Putnam den Effekt, wenn sich immer stärker ausdifferenzierte Interessensgruppen im Internet herausbilden (Putnam 2000, S. 177). Personen und Institutionen, die bislang von den etablierten Wissensautoritäten verdrängt wurden, können sich

**Konkurrenz für traditionelle externe Wissenschaftskommunikation**

**wissenschaftliche Reputation keine Garantie für Relevanz bei Google**

**Chancen für gesellschaftlich/wissenschaftlich marginalisierte Inhalte**

**Tendenzen zur „Cyberbalkanization“?**

<sup>25</sup> Kritische Anmerkungen zum Balkanisierungs-Begriff finden sich bei Nentwich (2003, S. 230).

in dieser Umgebung gegenseitig finden, vernetzen und so zu neuer Geltung kommen. Das tun sie besonders dort erfolgreich, wo sie sich vehement einem Thema widmen, das von etablierteren Wissenssautoritäten eher am Rande behandelt wird. Es würde zu weit gehen, aus derartigen Überlegungen und Einzelfallbeispielen den Schluss zu ziehen, marginalisierte Wissensbestände würden generell durch Google bevorzugt<sup>26</sup>. Erst recht sollten keine vorschnellen Schlüsse auf die Akzeptanz solcher Inhalte gezogen werden. Es erscheint jedoch evident, dass Google die Exklusivität wissenschaftlicher Autoritäten unterminiert und zumindest in einigen Fällen randständige WissenschaftlerInnen und ihre Inhalte neue Chancen zur Rezeption erhalten.

The image shows a Google search interface for the term "Klimaforschung". The search bar contains the text "Klimaforschung" and the search button is labeled "Suche". Below the search bar, there are radio buttons for "Das Web", "Seiten auf Deutsch", and "Seiten aus Österreich". The search results are displayed below, starting with a sponsored advertisement for "Klimaforscher" from "www.gastreferenten.de". The first organic result is a news article titled "Klimaforschung: Klimaerwärmung nicht durch Industrieabgase !?" from "www.klimaforschung.net". Other results include "Max-Planck-Gesellschaft - Geo- und Klimaforschung", "Welt der Physik: Die Klimaforschung ÜBERSICHT", and "Video-Ergebnisse zu Klimaforschung" which includes two video thumbnails: "Klimawandel" (8 Min. - 6. Sept. 2008) and "ZDF Joachim Bublath - Klima Teil 1" (10 Min. - 27. Juni 2007). Further down, there are results for "Klimaforschung - Schweiz - Informationen" and "Klimaforschung: Das Wetter vor fünfzig Jahren ...".

Abbildung 2.4-2: Google-Suche nach dem Begriff „Klimaforschung“  
Anmerkung: Suche durchgeführt am 22.01.2010.

<sup>26</sup> So kam eine Untersuchung zu medizinischen Informationen im Netz zu dem Ergebnis, dass kommerziell-schulmedizinisch orientierte Webseiten besser bei Google positioniert waren als Anbieter aus dem alternativen Bereich (Mager 2010).

Während „googelnde“ WissenschaftlerInnen meist wohl noch über die Expertise verfügen dürften, um von ihrem Fach marginalisierte Positionen als solche zu identifizieren, vermittelt Google für LaiInnen unter Umständen schnell ein irreführendes Bild vom tatsächlichen Stand der Wissenschaft, bzw. von ihren relevanten VertreterInnen. Zusammen mit der gesamtgesellschaftlichen Bedeutung der Suchmaschine ergibt sich daraus eine Herausforderung für die externe Wissenschaftskommunikation. Sollen wissenschaftliche Inhalte erfolgreich an die Öffentlichkeit kommuniziert werden, kann Google kaum ignoriert werden. Das gilt umso mehr vor dem Hintergrund, dass außerwissenschaftliche Einflüsse seit einigen Jahren zunehmend auf das Wissenschaftssystem wirken (Weingart 2001).

**Herausforderung für externe Wissenschaftskommunikation**

In der Medizin wurde dieses Problem längst erkannt und thematisiert. Das Fachurteil von ÄrztInnen wird inzwischen von den PatientInnen häufig durch Zuhilfenahme von Internetquellen hinterfragt. Untersuchungen zeigen, dass Google bei solchen Informationssuchen die erste Wahl ist (Mager 2009, S. 1134). Dabei wird die Qualität der Inhalte vor allem über ihre übergreifende Konsistenz eingeschätzt und weniger über deren institutionelle Reputation:

**Beispiel Medizin: Konsistente Informationen bei Google als Qualitätskriterium**

„(...) end-users hardly recognize the website provider at all, least of all its identity. Instead of trusting sites as a whole, they rather validate pieces of information by recurrence and how they fit into the overall information assemblage with which they are provided.” (ebd., S. 1138)

Dies steigert die Akzeptanzchancen weniger reputabler WissenschaftlerInnen in der Öffentlichkeit. Oder umgekehrt: Eine guter akademischer Ruf ist im Zusammenhang solcher LaiInnen-Recherchen nur bedingt hilfreich. Untereinander stark vernetzte Anbieter mit ähnlichen Inhalten haben beste Chancen das von LaiInnen offenbar als wichtig eingeschätzte Kriterium konsistenter Informationen zu erreichen. Die entsprechenden Webseiten-Betreiber sind sich dessen in der Regel durchaus bewusst, gehen jedoch unterschiedlich damit um, wie Mager weiter berichtet. Demnach sind kommerzielle Anbieter stärker darauf bedacht, ihre Seiten für Google zu optimieren als nicht-kommerzielle Dienste, die darin häufig ein manipulatives Vorgehen sehen (ebd., S. 1133).

In jedem Fall obliegt es aber den einzelnen NutzerInnen, Ergebnisse zu interpretieren und zu bewerten. Sie erhalten durch Google also nicht nur neue Möglichkeiten, sondern auch zusätzliche Verantwortung:

**zusätzliche Verantwortung für NutzerInnen**

„(...) users are not only provided with the possibility of individually assembling web information, but also with the duty to recontextualize and make sense of it.” (ebd., S. 1139)

Entsprechend könnte man zusammenfassend feststellen, dass externe Wissenschaftskommunikation zunehmend auf Google reagieren muss, indem entsprechende Web-Inhalte zur Verfügung gestellt und optimiert präsentiert und vermarktet werden. Gleichzeitig übernehmen LaiInnen zwangsweise Aufgaben, die traditionell Gatekeepern wie Redaktionen zukamen.





## 3 Google Scholar

Wie der Name bereits verrät, handelt es sich bei Google Scholar um eine speziell akademisch ausgerichtete Suchmaschine mit dem Ziel, wissenschaftliche Publikationen auffindbar zu machen. Sie unterscheidet sich jedoch nicht nur hinsichtlich ihrer Zielgruppe von der Universalsuchmaschine Google, sondern auch in ihrer technischen Konzeption: Um auch Inhalte des für Crawler häufig nicht zugänglichen akademischen Webs verfügbar zu machen, wird hier auf zusätzliche Methoden zur Indexierung der gewünschten Inhalte zurückgegriffen (3.1). Gleichzeitig werden auch bestimmte Funktionsprinzipien wie die Anzeigeweise der Treffer (3.2) und einige Suchoptionen (3.3) aus der Universalsuchmaschine übernommen bzw. adaptiert und erweitert.

Google Scholar ging Mitte November 2004 online, wobei auf Daten des Vorgängerprojekts *CrossRef Search* zurückgegriffen werden konnte. Im April 2006 folgten weitere Sprachversionen (u. a. eine deutsche). Wie viele Google-Projekte, befindet sich auch diese akademische kostenfreie Suchmaschine noch immer offiziell in der Beta-Phase, d. h. in einem Teststadium. Dennoch wird sie bereits intensiv genutzt und stellt eine ernstzunehmende Konkurrenz zu herkömmlichen wissenschaftlichen Informationsdiensten dar (3.4.1). Dazu tragen auch zusätzliche Leistungen bei: Insbesondere die enthaltenen Ansätze der Zitationsanalyse machen Google Scholar gegenüber anderen kostenpflichtigen Angeboten (z. B. *Science Citation Index*, *CiteSeer*) attraktiv (3.4.2).

**akademische  
Spezielsuchmaschine**

**angepasst an spezielle  
Erfordernisse**

**Erfolg trotz Mängel**

**u. a. durch kostenlose  
Zusatzangebote**

### 3.1 Was wird gefunden?

Wie oben bereits angedeutet, liegt die Kernfunktion von Google Scholar darin, dass diese Spezielsuchmaschine auch akademische Inhalte indexiert, die für übliche Web-Crawler nicht zugänglich sind. Dabei werden vielfach (im Gegensatz zu herkömmlichen Systemen) auch die Volltexte durchsucht und nicht nur Titel, Schlagwörter und Zusammenfassungen. Dies wird vor allem durch Kooperationen mit Verlagen und Bibliotheken erreicht, die dem Dienst die nötigen bibliographischen Informationen und mitunter eben auch Zusammenfassungen und Volltexte zur Verfügung stellen. Dementsprechend ist der Umfang und die Qualität der über Google Scholar zugänglichen Dokumente in starkem Maße von diesen Kooperationen abhängig. So wurde gerade zu Beginn noch beklagt, die Abdeckung wissenschaftlicher Publikationen durch Google Scholar sei teilweise unbefriedigend (Mayr/Walter 2006; Neuhäus et al. 2006; Lewandowski 2007), u. a. weil der äußerst bedeutsame *Elsevier*-Verlag nicht an der Kooperation teilnahm (Jacsó 2005). Diese Zurückhaltung dürfte damit zusammenhängen, dass Google Scholar eine unmittelbare Konkurrenz für den verlagseignen Suchdienst *Scopus* darstellt. Seit 2007 öffnet Elsevier seine Inhalte jedoch im Zuge des Erfolgs von Google Scholar, der auch Publikationen des Verlags eine höhere Sichtbarkeit einbringt.

Durch Verlinkung von Beständen aus der Buchsuche Google Books (siehe auch Abschnitt 4) konnte der Umfang der von Google Scholar indexierten Inhalte signifikant ausgebaut werden (Jacsó 2008, S. 104 f.), da auf diese Weise mehr Bücher samt ihrer Volltexte erfasst werden können.

**Kooperationen mit  
Verlagen und  
Bibliotheken**

**Verbindung zu  
Google Books**

**bibliographische  
Daten aus Zitaten**

Dadurch dass auch bibliographische Daten aus Zitaten extrahiert werden, können auch Quellen indexiert werden, auf die Google Scholar selbst nicht unmittelbar aus eigenen Datenbeständen zugreifen kann. Auf diese Weise wird die Abdeckung der Suchmaschine sukzessive erweitert. Entsprechend lobt Jacsó, der sich diesbezüglich in der Frühphase Google Scholars noch kritisch äußerte (2005), 2008 die „broad content coverage of Google Scholar“ (S. 106). Gleichzeitig betont er jedoch auch, dass es noch immer „huge gaps in the full-text indexing of the most important serial publications“ (ebd.) gäbe und illustriert dies an einem Beispiel:

„For example, less than 17 per cent of the 430,500 documents at the nature.com web site were indexed by Google Scholar directly from that site (which includes not only Nature magazine but also many other journals of the Nature Publishing Group). True, many more than 17 per cent of them have a record in Google Scholar, but many of these are just citation records with minimal information.“ (Jacsó 2008, S. 106)

Auch Mayr und Walter (2007) weisen in ihrer Studie auf ein solches Übergewicht von Minimalinformationen hin.

**Informationen aus  
dem Web**

Neben diesen speziellen Zugängen bezieht Google Scholar auch Informationen aus dem öffentlich zugänglichen Web ein, was die potenziell erreichbaren Quellen weiter erhöht. Entsprechend breit beschreibt Google Scholar selbst sein Angebot:

„(...) von Kommilitonen bewertete Seminararbeiten, Magister-, Diplom- sowie Doktorarbeiten, Bücher, Zusammenfassungen und Artikel, die aus Quellen wie akademischen Verlagen, Berufsverbänden, Magazinen für Vorabdrucke, Universitäten und anderen Bildungseinrichtungen stammen.“ (Google Scholar 2010c)

**Kriterien zur Aufnahme  
von Webseiten**

Damit Web-Dokumente von den Crawlern von Google Scholar erfasst werden können, müssen bestimmte Kriterien erfüllt sein. Es ist möglich auch private Webseiten gezielt für die Crawler der Suchmaschine anzumelden (sofern sie nicht ohnehin bereits automatisch gefunden wurden). Dabei wird jedoch ausdrücklich darum gebeten, nur wissenschaftliche Inhalte bereitzustellen:

„Der auf Ihrer Website gehostete Content muss wissenschaftlicher Natur sein. Content wie beispielsweise Buchrezensionen und Leitartikel ist für Google Scholar nicht geeignet.“<sup>27</sup>

**automatisierte  
Bewertung von  
Wissenschaftlichkeit**

Wie Google Scholar selbst die Wissenschaftlichkeit und somit auch ein (dem eigenen Anspruch nach) entscheidendes Indexierungskriterium abschätzt, ist dabei unklar. Auf Anfrage erklären die Betreiber, es kämen dabei „in erster Linie automatische heuristische Prinzipien“ (E-Mail-Korrespondenz vom 08.03.2010) zum Einsatz. Ein Faktor sei dabei die Formatierung als wissenschaftliche Arbeit. Die Crawler von Google Scholar bewerten also automatisiert anhand spezifischer Dokumentenstrukturen, ob es sich um wissenschaftliche Texte handelt.

**formale Anforderungen  
durch Crawl-Vorgänge**

Wie bei allen Crawl-Vorgängen müssen daher gewisse technische Standards eingehalten werden, damit diese erfolgreich sein können. Inhalte auf Flash- oder Javascript-Basis bereiten Google Scholar dabei ebenso Probleme, wie Artikel, die auf verschiedene Dateien verteilt sind, oder mehrere Artikel, die in ein und derselben Datei enthalten sind. Auch sollte keine Registrierung auf der Webseite erforderlich sein. Schließlich wird gefordert, es müsse mindestens eine Zusammenfassung oder der Volltext verfügbar sein:

<sup>27</sup> [google.com/support/scholar/bin/request.py](http://google.com/support/scholar/bin/request.py) (zuletzt aufgerufen am 26.02.2010).

„Bibliografische Informationen ohne Kurzfassungen bzw. Volltextartikel werden von unseren Crawlern nicht für die Aufnahme erfasst.“<sup>27</sup>

Da Google Scholar durch das Crawling im allgemein zugänglichen Web diverse frei verfügbare Dokumente aufspürt, eignet es sich prinzipiell gut zum Finden von Open-Access-Publikationen – sogar weitaus besser als die auf Open-Access-Dokumente spezialisierten Dienste *OAIster* und *OpenDOAR* (Norris et al. 2008a). Mayr und Walter (2007) kamen in einer früheren Studie noch zu dem konträren Ergebnis, Google Scholar würde hinsichtlich der Abdeckung von Open-Access-Publikationen massive Mängel aufweisen. Diese scheinen jedoch inzwischen deutlich verbessert worden zu sein, wie Mayr später berichtete (Mayr 2009).

Durch die Websuche ist Google Scholar nicht nur hinsichtlich seines Umfangs breiter aufgestellt als traditionelle Fachdatenbanken, sondern es besteht auch ein fundamentaler Unterschied auf inhaltlicher Ebene hinsichtlich der Fächergrenzen: Die Suche mit Google Scholar ist – mit Ausnahme der englischen Sprachversion, die auch fachspezifische Suchen ermöglicht – per se multidisziplinär. Denn die Ergebnisse werden nicht nach Fachgruppen geordnet und lassen sich entsprechend auch nicht mit einem disziplinären Fokus durchsuchen (dazu auch 3.3). Gleichzeitig wurde darauf hingewiesen, dass die verschiedenen Disziplinen von Google Scholar unterschiedlich gut abgedeckt werden. So stellten Neuhaus et al. (2006) eine sehr gute Abdeckung von naturwissenschaftlichen Publikationen und eine schlechtere bei den Geistes- und Sozialwissenschaften fest.

Während im frühen Stadium noch ein anglophoner Bias beobachtet wurde (Neuhaus et al. 2006; Mayr/Walter 2007), schätzt Jacsó (2008, S. 105) inzwischen auch die sprachliche und lokale Abdeckung als sehr gut ein, wobei natürlich auch hier die ohnehin generell bestehende Dominanz englischsprachiger Publikationen nicht wettgemacht werden kann. Diese fällt jedoch im Vergleich mit anderen Suchdiensten eher gering aus, so Jacsó. Dazu trägt auch die Möglichkeit sprachspezifischer Suchen bei, über die Google Scholar seit April 2006 verfügt.

Wie bei anderen Websuchmaschinen kann die Aktualität der indexierten Inhalte je nach Crawl-Häufigkeit variieren. Während einige Seiten sehr häufig von Crawlern aufgesucht werden und ihre Inhalte entsprechend auch in Google Scholar aktuell gehalten werden, kann es durchaus auch vorkommen, dass einige Webseiten über einen längeren Zeitraum nicht erfasst werden (Falagas et al. 2008, S. 341; Söllner 2006, S. 832; Mayr/Walter 2007).

Durch die Verwendung unterschiedlichster Quellen bei gleichzeitig unbekannter Datenbasis (Google hält sich darüber wie üblich sehr bedeckt), ist es unmöglich, den tatsächlichen Umfang der durch Google Scholar indexierten Dokumente exakt zu bemessen (vgl. Jacsó 2008, S. 105; Söllner 2006, S. 831). Es ist davon auszugehen, dass die Abdeckung mit zunehmendem Reifegrad der Plattform immer weiter verbessert wird. Dazu tragen nicht zuletzt die innovativen Methoden der automatischen Auslesung bibliographischer Informationen aus unterschiedlichsten Quellen bei. Eben hier liegt jedoch auch die Achillesferse der akademischen Suchmaschine: Falsch ausgelesene Daten ziehen den Wert der damit arbeitenden Zusatzfunktionen in Zweifel (3.4.2) und auch die Relevanz vieler gefundener Dokumente ist nicht immer unbestritten (dazu 3.2).

**gute Abdeckung von  
Open-Access-  
Publikationen**

**keine disziplinären  
Differenzierungs-  
möglichkeiten**

**Abdeckungsunterschiede  
zwischen Disziplinen**

**gute sprachliche/lokale  
Abdeckung**

**variierende Aktualität**

**genauer Umfang  
unbekannt**

**innovative Methoden  
der Datenerfassung  
führen zu Stärken und  
Schwächen der  
Plattform**

## 3.2 Was wird wie angezeigt?

<b>Orientierung an Google</b>	Auf den ersten Blick erinnert die Ergebnisanzeige in Google Scholar an die der Universalsuchmaschine Google (Abbildung 2.1-1): Die Ergebnisse werden hierarchisch gelistet, wobei standardmäßig zehn Ergebnisse auf einer Seite angezeigt werden. Auch die einfache Suchmaske und die Geschwindigkeit mit der die Anfragen bearbeitet werden, gleichen dem Vorläufer. Auf den zweiten Blick zeigen sich bereits einige Unterschiede: Anstatt Quellen als URL unter dem Textausschnitt anzuzeigen, werden hier die Namen der AutorInnen und das Erscheinungsjahr der jeweiligen Publikationen unterhalb des Titels angezeigt. Zudem werden zusätzliche Optionen angeboten – weitere zitierende Quellen und ähnliche Artikel können angezeigt werden, und auch andere Versionen lassen sich abrufen bzw. werden unter einem Treffer subsumiert. Dabei werden gewöhnlich Verlagsversionen als Haupttreffer aufgeführt, während kostenlose andere Versionen z. B. aus Preprint-Archiven oder privaten Homepages, Universitätsservern usw. stammen (Lewandowski 2007, S. 165). Auch die Verfügbarkeit in lokalen Bibliotheken kann erfragt werden, sofern diese mit dem Dienst kooperieren. Dazu wird der jeweilige Bibliothekskatalog über sogenannte <i>Link-Resolver</i> für Google Scholar zugänglich gemacht, wobei dieser auf bestimmte Computer beschränkt werden kann (z. B. auf Universitätsrechner oder Anschlüsse mit speziellen IP-Adressen). <sup>28</sup> Auch der Export in Programme zur Literaturverwaltung (BibTex, EndNote, RefMan, RefWorks, WenXianWang) wird – wie die anderen Dienste – kostenlos angeboten.
<b>Unterschiede</b>	
<b>verschiedene Artikelversionen</b>	
<b>Zusatzfunktionen</b>	

The screenshot shows the Google Scholar interface. At the top, the search bar contains 'google scholar' and a search button. Below the search bar, there are options for 'Web-Suche' (selected) and 'Suche Seiten auf Deutsch'. The main content area displays search results for 'Mycorrhizal symbiosis' and 'Seventh report of the Joint National Committee on prevention, detection, evaluation ...'. Each result includes the title, authors, journal information, and citation counts. The first result is by SE Smith, DJ Read, published in the Soil Science Society of America Journal. The second result is by AV Chobanian, GL Bakris, HR Black, WC Cushman, published in the Am Heart Assoc journal. The third result is a citation for 'Effects of drugs on clinical laboratory tests' by DS Young, DW Thomas, RB Friedman, LC ... in Clinical Chemistry.

Abbildung 3.2-1: Ergebnisanzeige in Google Scholar  
Anmerkung: Suche durchgeführt am 25.02.2010.

<sup>28</sup> Weitere Informationen zu diesem von Google Scholar „Bibliotheks-Links-Programm“ genannten Verfahren finden sich hier: [scholar.google.de/intl/de/scholar/libraries.html#start](http://scholar.google.de/intl/de/scholar/libraries.html#start) (zuletzt aufgerufen am 11.03.2010).

Verweise in den Ergebnissen, bei denen es sich um Zitationen, direkte Links zu PDFs oder Einträge in Google Books handelt, werden mit entsprechenden Hinweisen versehen. Die Ergebnislisten können nachträglich gefiltert werden, indem nur noch Quellen ab einem bestimmten Jahrgang angezeigt werden und/oder Zitationen und Treffer, die nicht mindestens über eine Zusammenfassung verfügen ausgeschlossen werden.

Da abgesehen von diesen nachträglichen Eingriffsmöglichkeiten keine Optionen zum Sortieren der Suchergebnisse geboten werden (im Gegensatz zu vielen herkömmlichen Systemen, die etwa eine Sortierung nach Autorennamen erlauben), sind NutzerInnen in starkem Maße von der durch Google Scholar erzeugten Rangordnung abhängig. Wie bei der Universalsuchmaschine, bleiben die Informationen darüber, wie diese zustande kommt, eher diffus und Google hält sich dazu bedeckt. Der Konzern beschreibt die Faktoren so:

„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.“ (Google Scholar 2010c)

Ob dieser Mechanismus immer so funktioniert wie hier beworben und tatsächlich zum „Ausfindigmachen der wichtigsten Arbeiten auf einem beliebigen Forschungsgebiet“ (ebd.) dient, ist jedoch fraglich. Die Beispielsuche in Abbildung 3.2-1 ist bereits symptomatisch für ein typisches Problem: Die Abfrage „Google Scholar“ (ohne Anführungszeichen) führt eben nicht zu den einschlägigsten Texten über die Suchmaschine. Der erste Treffer verlinkt zu einem Buch über Symbiosen zwischen Pilzen und Pflanzen, der Zweite zu einem Bericht über die Behandlung von Bluthochdruck und an dritter Stelle folgt die Zitation zu einem 1972 erschienenen Text mit dem Titel „Effects of drugs on clinical laboratory tests“. Dieses Ergebnis entsteht offenbar dadurch, dass die verlinkten Webseiten auch den Begriff „Google Scholar“ enthalten und die Suchmaschine diesen irrtümlich als Teil der Publikationen interpretiert.

Dieses Beispiel ist symptomatisch, weil es auf ein Grundproblem der Funktionsweise Google Scholars zurückzuführen ist: Die Suchmaschine ist in hohem Maße abhängig von den Entscheidungen ihrer Algorithmen, die nicht nur automatisiert darüber urteilen, was indexiert wird (3.1), sondern auch, wie es angezeigt wird, also in welcher Rangordnung. Da wenig darüber bekannt ist, wie die unterschiedlichen von Google genannten Faktoren gewichtet werden, kann an dieser Stelle nur darüber spekuliert werden.

Die Beispielsuche deutet darauf hin, dass die Anzahl der Zitate, die ein Dokument erhalten hat (die aufgeführten Dokumente wurden laut Google Scholar zwischen 1365 und 13.360 mal zitiert), offenbar einen stärkeren Einfluss auf die Positionierung hat als andere Faktoren (wie z. B. die Lage des Suchbegriffs im Dokument). Erst an neunter (!) Stelle folgte ein Dokument, das tatsächlich Google Scholar behandelt und den Begriff auch im Titel führt (Jacsó 2005), jedoch nach der Suchmaschine „lediglich“ 107 mal zitiert wurde. Zudem können auch andere Faktoren wie etwa der Publikationskontext zu diesem ineffektiven Suchergebnis führen.

Vor dem Hintergrund, dass auch die Universalsuchmaschine Google Dokumente vor allem aufgrund ihrer erhaltenen Verweise bewertet (2.2), ist anzunehmen, dass Google Scholar diesem Faktor ebenfalls hohes Gewicht beimisst. Im Unterschied zur Universalsuchmaschine, werden hierzu jedoch nicht hyperlinkförmige Verweise genutzt, sondern Zitate.

**verschiedene  
Ergebnistypen**

**nachträgliche  
Filteroptionen**

**Abhängigkeit von  
Rangordnung**

**Ranking-Kriterien  
nicht klar**

**nicht immer relevante  
Ergebnisse**

**durch Abhängigkeit von  
automatisierten  
Entscheidungen**

**Zitationszahlen als  
wichtiger Faktor**

**teils zweifelhafte  
Rankings durch  
fehlerhafte  
Zitationsanalyse**

Während Hyperlinks durch Crawler ohne weiteres als solche zu identifizieren sind, müssen Zitate erst als solche erkannt werden. Google Scholar vertraut dabei auf einen eigenen Mechanismus der Zitationsextraktion- und analyse, dessen Funktionalität jedoch umstritten ist, da die erforderlichen Metadaten nicht immer korrekt ausgelesen werden (3.4.2). Entsprechend geben diese Zahlen nicht unbedingt reale Zitationsraten wieder, geschweige denn die tatsächliche Relevanz von Publikationen. In diesem Kontext wurde Google Scholar auch dafür kritisiert, dass aktuellere Veröffentlichungen benachteiligt würden, da diese naturgemäß über weniger Zitationen verfügen (Söllner 2006, S. 833). Inzwischen begegnet man diesem Problem, indem sich die Ergebnisse auf aktuellere Publikationen reduzieren lassen.

**kann sich auch  
negativ auf weitere  
Zusatzfunktionen  
auswirken**

Das fehlerhafte Extrahieren von Metadaten kann sich auch problematisch auf die Nutzung der von Google Scholar angebotenen Zusatzfunktionen auswirken: Denn natürlich können die Daten nur korrekt in ein Programm zur Literaturverwaltung oder mit einer Suche in einem örtlichen Bibliothekskatalog verknüpft werden, wenn sie zuvor richtig extrahiert wurden. Immerhin kam eine quantitative vergleichende Studie (Robinson/Wusteman 2007), zu dem Ergebnis, dass Google Scholar bei wissenschaftlichen Suchanfragen insofern präziser operiert als Universalsuchmaschinen, als Google Scholar in höherem Ausmaß relevante wissenschaftliche Publikationen findet.

### 3.3 Einflussmöglichkeiten der NutzerInnen

**teils einfache, aber auch  
erhöhte Anforderungen  
durch NutzerInnen**

Wie im Fall der Universalsuchmaschine (2.3), deuten einige Studien darauf hin, dass auch NutzerInnen von Google Scholar häufig nicht an erweiterten Suchoptionen interessiert sind. Allerdings ergeben sich durch das akademische Zielpublikum auch besondere Ansprüche, die von der Suchmaschine nur zum Teil erfüllt werden (dazu 3.4.1, 3.4.2).

**Parallelen und  
Unterschiede zu  
Google-Suchoptionen**

Das Fenster zur Basissuche gleicht dem von Google. Auch hier können prinzipiell beliebig viele Suchbegriffe eingegeben werden (im Gegensatz zu vielen herkömmlichen Literaturdatenbanken). Auch einige *Erweiterte Suchoptionen* entsprechen denen der Universalsuchmaschine. So sind hier ebenfalls „Oder“-anstatt „Und“-Suchen möglich und auch genaue Wortfolgen, oder gezielt ausschließende Begriffe lassen sich eingeben. Interessanterweise bietet Google Scholar jedoch in diesem Fenster insgesamt *weniger* Suchoptionen als die allgemeine Suchmaschine. So lassen sich z. B. keine regionalen Vorgaben machen oder bestimmte Domains durchsuchen. Stattdessen werden einschränkende Suchen mit Fokus auf spezifische Fachzeitschriften, AutorInnen und Erscheinungszeiträume erlaubt.

**Operatoren**

Entsprechend werden einige Operatoren, die auch auf Google verfügbar sind unterstützt (z. B. „OR“, „intitle:“), während andere nicht verwendet werden können (z. B. „filetype:“, „allinurl:“), jedoch auch zusätzliche hinzukommen (z. B. „autor:“). Jacsó (2008, S. 106 ff.) meldet allerdings Zweifel an ihrer Funktionsfähigkeit an, da u. a. der „OR“-Operator zwischen zwei Suchwörtern teilweise zu weniger Ergebnissen führe, anstatt – wie es bei einer „Oder“-Suche eigentlich der Fall sein sollte – mehr Treffer zu erzielen. Natürlich ist die Funktionsfähigkeit von einigen erweiterten Suchoptionen auch abhängig von der Korrektheit der Metadaten, die jedoch oft mangelhaft sind und zu entsprechend verfälschten Ergebnissen führen (dazu auch 3.4.1).

In den *Einstellungen* können sowohl die Sprache der Benutzeroberfläche angepasst werden, als auch Suchen auf Sprachräume begrenzt werden. Dabei werden derzeit folgende Sprachen angeboten: Chinesisch (traditionell), Chinesisch (vereinfacht), Deutsch, Englisch, Französisch, Italienisch, Japanisch, Koreanisch, Niederländisch, Polnisch, Portugiesisch, Spanisch und Türkisch (Stand: März 2010). Zudem lässt sich hier auswählen, welches der unterstützten Programme zur Literaturverwaltung genutzt wird und welcher der verfügbaren Bibliothekskataloge in den Ergebnissen verlinkt werden soll (dazu auch 3.2).

#### **verfügbare Einstellungen**

In der englischen Version werden zusätzlich auch disziplinäre Differenzierungsmöglichkeiten für die Suche angeboten:

#### **disziplinäre Differenzierungs- möglichkeiten in der englischen Version**

- Biology, Life Sciences, and Environmental Science
- Medicine, Pharmacology, and Veterinary Science
- Business, Administration, Finance, and Economics
- Physics, Astronomy, and Planetary Science
- Chemistry and Materials Science
- Social Sciences, Arts, and Humanities
- Engineering, Computer Science, and Mathematics.

Allerdings ist davon auszugehen, dass viele Texte keiner dieser Fachgruppen zugeordnet sind, so dass selbst bei Auswahl aller Fachgruppen nur ein Bruchteil der Texte gefunden wird, die in einer einfachen Suche mit demselben Begriff gelistet werden. So schildert Jacsó (2008, S. 108) etwa, dass eine Beispielsuche nach dem Wort „Vietnam“ im Titel in allen aufgeführten Disziplinen nur 15 % der Texte findet, die ohne Auswahl dieser Fächerspezifizierungen gefunden werden. D. h. 85 % der Texte dieses Beispiels sind keiner Fachgruppe zugeordnet.

Weiters bietet die englische Sprachausgabe von Google Scholar Fokussierungen auf spezielle juristische Texte. So lassen sich beispielsweise gezielt Rechtsgutachten eines bestimmten amerikanischen Bundesstaates auffinden. Außerdem kann (bereits in der einfachen Oberfläche) ausgewählt werden, ob auch nach Patenten gesucht werden soll.

#### **weitere Optionen der englischen Version**

Dadurch dass zu einzelnen Dokumenten auch ähnliche und zitierende Artikel gesucht werden können, erlaubt Google Scholar zusätzlich das „Browsen“ in den Ergebnissen.

#### **„Browsen“ in Ergebnissen**

Neben diesen Einflussmöglichkeiten für NutzerInnen der Suchfunktion, ergeben sich auch gewisse Optionen für Anbieter von Inhalten. Zum einen sind das die kostenlosen Kooperationen für Verlage und Bibliotheken, die ihre Angebote für die Indexierung in Google Scholar freigeben können. Auch Webmaster können Inhalte ihrer Internetseiten gezielt für die Crawler von Google Scholar anmelden.<sup>29</sup> Da das Ranking in Google Scholar auf anderen Prinzipien beruht als das der Universalsuchmaschine Google (3.2), entfallen hier viele der üblichen Strategien der Suchmaschinenoptimierung. Doch die hintergründige Idee, Daten suchmaschinengerecht aufzubereiten, kann natürlich auch auf Google Scholar angewendet werden. Wenn beispielsweise eine Zeitschrift PDF-Dateien auf einem Server anbietet, die nur über spezielle interne Suchfunktionen auffindbar sind, werden sie von Crawlern nicht erfasst. Abhilfe schafft hier eine Verlinkung mit Inhaltsverzeichnissen, die ggf. auch bei Google Scholar angemeldet werden können (Lewandowski 2007, S. 168).

#### **Einflussmöglichkeiten für Content-Anbieter**

#### **besondere Anforderungen für eine ...**

<sup>29</sup> [google.com/support/scholar/bin/request.py?hl=de](http://google.com/support/scholar/bin/request.py?hl=de)  
(zuletzt aufgerufen am 12.03.2010).

Neben derartigen Vorgehensweisen, die zunächst lediglich die prinzipielle Auffindbarkeit von Texten innerhalb von Google Scholar ermöglichen (dazu auch: 3.1), kann auch bedingt auf das Ranking eingewirkt werden. Da dies in hohem Maße vom Auslesen der Metadaten abhängt (3.2), empfiehlt Google für ein optimales Ranking diese korrekt bereitzustellen:

„Unsere Indizierungsalgorithmen extrahieren automatisch Metadaten, Zitate und andere Informationen aus Artikeln und verwenden sie für das Ranking. Durch das Bereitstellen verbindlicher Metadaten über Ihre Artikel können Sie diesen Vorgang unterstützen und die Wahrscheinlichkeit, dass alle Erwähnungen Ihres Artikels erkannt werden, erhöhen.“

(Google Scholar 2010b)

**... akademische  
Suchmaschinen-  
optimierung**

Allerdings können fehlerhafte Metadaten unter Umständen auch zu vorteilhaften Positionierungen führen (siehe 3.4.2). Somit lassen sich auch für wissenschaftliche Inhalte spezielle Prinzipien akademischer Suchmaschinenoptimierung identifizieren und durchführen (Beel et al. 2010).

## 3.4 Einsatz in der Wissenschaft

Für die interne Wissenschaftskommunikation spielt Google Scholar vor allem in zweierlei Hinsicht eine Rolle: In erster Linie natürlich für den Hauptzweck der Suchmaschine: die Suche nach wissenschaftlichen Inhalten (3.4.1). In zweiter Linie ist auch zu diskutieren, wie sich die von Google Scholar angebotenen Funktionen der Zitationsanalyse auf das Wissenschaftssystem auswirken (3.4.2). Durch den Fokus auf wissenschaftliche Publikationen entfallen gleichzeitig einige Einflussfaktoren, die bei der Universalsuchmaschine zum Tragen kommen.

### 3.4.1 Suche nach wissenschaftlichen Inhalten

**neues System zur  
wissenschaftlichen Suche**

Nach eigenem Anspruch sucht Google Scholar zunächst ausschließlich nach wissenschaftlichen Publikationen (bzw. Verweisen). Wie oben dargelegt, unterscheidet sich diese Art der Literatursuche von traditionellen Systemen wie Bibliothekskatalogen und Datenbanken vor allem durch die auffindbaren Daten (3.1) und wie diese angezeigt (3.2) und durchsucht (3.3) werden können.

**positive und negative  
Effekte**

Es lassen sich in diesem Rahmen für die noch junge Plattform keine zuverlässigen Prognosen hinsichtlich der langfristigen Auswirkungen auf das Wissenschaftssystem anstellen. Je nach Perspektive, lassen sich jedoch eine Reihe von positiven und negativen Effekten aufzeigen, die sich bereits aus den bisherigen Beobachtungen abzeichnen.

**kostenfrei und  
umfangreich**

Aus der NutzerInnen-Perspektive lässt sich positiv bemerken, dass Google Scholar einen kostenfreien und durchaus umfangreichen Zugang zu wissenschaftlichen Veröffentlichungen bietet, wobei diese häufig sogar im Volltext durchsuchbar und zugänglich sind. Damit hebt sich der Dienst von den meisten Konkurrenten ab. Das gilt auch für die speziell akademisch ausgerichtete Websuche, die viele Dokumente zugänglich macht, die etwa über Bibliothekskataloge nicht auffindbar sind, gleichzeitig aber relevantere Treffer erzielt als Universalsuchmaschinen wie Google.



Der freie Zugang wurde nicht zuletzt von der Open-Access-Bewegung positiv aufgenommen. Zwar handelt es sich auch bei vielen der in Google Scholar indexierten Dokumente lediglich um Links zu kostenpflichtigen Anbietern, jedoch sind eben oftmals auch kostenfreie Alternativversionen (z. B. aus Preprint-Archiven, oder privaten Homepages) verfügbar. Es wäre möglich, dass die Sichtbarkeit derartiger Papiere via Google Scholar diesen zu neuer Relevanz verhilft. Schließlich konnte in einigen Bereichen eine häufigere Zitationsrate von Open-Access-Publikationen beobachtet werden (Norris 2008).<sup>30</sup> Ob dies einen demokratisierenden Effekt auf die interne Wissenschaftskommunikation ausübt (wie häufig erhofft wird), bleibt abzuwarten. So geben etwa Norris et al. (2008b, S. 342) zu bedenken, dass Entwicklungsländer aufgrund mangelhafter IKT-Infrastruktur von derartigen online frei verfügbaren Dokumenten nicht unbedingt profitierten und daher eher geneigt sein dürften, die geringeren Ressourcen in wenige renommierte gedruckte Zeitschriftenausgaben zu investieren. Auch in Ländern mit gut ausgebauter IKT-Infrastruktur ist ein solcher Demokratisierungseffekt nicht per se zu erwarten. Zwar erlaubt Google Scholar den Zugriff auf ansonsten unter Umständen schwer erreichbare Dokumente, welche Publikationen und Personen jedoch nun davon effektiv profitieren, ist bis dato ungeklärt. Aufgrund der Mängel der Suchmaschine lässt sich urteilen, dass Google Scholar zwar eine Ergänzung zu kostenpflichtigen Zugängen darstellt, jedoch eben keinen Ersatz. So bleibt abzuwarten, ob langfristig WissenschaftlerInnen mit geringeren Ressourcen von diesem Angebot profitieren, oder ob es einen zusätzlichen Vorteil für ohnehin begünstigte Personen und Institutionen darstellt.<sup>31</sup> Kurzfristig ist hierin jedoch zunächst ein Vorteil für WissenschaftlerInnen ohne entsprechende Zugänge zu sehen.

**neue Relevanz von  
Open-Access-Artikeln?**

**demokratisierender  
Effekt unklar**

Damit in Verbindung ist zu fragen, inwiefern die Spezialsuchmaschine effektiv und gewinnbringend für die Suche nach wissenschaftlichen Publikationen einsetzbar ist. Gerade unter Studierenden scheint Google Scholar beliebt zu sein, da es ähnlich einfach zu bedienen ist, wie die Universalsuchmaschine. So geben es zumindest schwedische Studierende in einer vergleichenden Studie zu Protokoll, in der Google Scholar dem Suchdienst *MetaLib* gegenübergestellt wurde:

**effektiv und  
gewinnbringend  
einsetzbar?**

„Students consistently wrote that they liked the fact that Google Scholar was simple to use and in some cases added that one of the reasons it was so simple to use was that they were already familiar with the interface from using regular Google. That Google Scholar has a broad scope and typical leads to many results was also viewed positively by some.”  
(Nygren et al. 2006, S. 43)

Gleichzeitig beklagten demnach jedoch einige auch „too many irrelevant results“ (ebd.). So ergibt sich eine ähnliche Problemlage wie bei der Google-Websuche: Die Einfachheit der Sucheingabe und der Ergebnisanzeige wird vielfach positiv aufgenommen, täuscht aber auch über die hintergründige Komplexität des Suchvorgangs hinweg. Durch die Durchmischung der Ergebnisse mit gleichermaßen relevanten wie irrelevanten, qualitativ hochwertigen wie fragwürdigen, fertigen wie unfertigen Texten, kann es gerade für unerfahrene Suchende schwierig werden, die benötigten Veröffentlichungen mit Google Scholar aufzufinden. Dazu tragen auch die unklare Datenbasis und das intransparente Ranking bei, wodurch Suchende nie mit Sicherheit wissen, welche Bereiche abgedeckt werden, bzw. ob die jeweilige Rangfolge den ei-

**ähnlich wie Google:  
Einfachheit täuscht**

<sup>30</sup> Der Zusammenhang zwischen Open-Access-Publikationen und ihren Zitationsraten wird derzeit mit nicht eindeutigen Schlussfolgerungen untersucht und debattiert. Ein Überblick dazu findet sich in der annotierten Bibliographie von A. Ben Wagner unter [istl.org/10-winter/article2.html](http://istl.org/10-winter/article2.html) (zuletzt aufgerufen am 16.05.2010).

<sup>31</sup> Prominent hat Merton (1985) ähnliche Prozesse als Matthäus-Effekt beschrieben.

<b>kein vollwertiges Recherchetool</b>	genen Relevanzen entspricht. Daher lässt sich urteilen, dass Google Scholar derzeit nicht als einziges und vollwertiges Recherchetool geeignet ist.
<b>besonders nicht für unerfahrene NutzerInnen</b>	Allerdings dürfte dies auch auf viele traditionelle Systeme zutreffen, zumal selbst ein präziser und erschöpfender Suchdienst keineswegs auch gleichermaßen effektiv von seinen NutzerInnen bedient wird. Das gilt umso mehr für ungeschulte Personen, die ihr Suchverhalten an der Suchmaschine orientieren, die ihnen hinlänglich bekannt ist – Google. So beobachteten Nygren et al. (2006, S. 55 ff.) bei ihrer Untersuchung des Suchverhaltens von Studierenden u. a., dass diese Suchvorgänge mit MetaLib abbrachen, wenn diese ihnen zu lang vorkamen, bei der korrekten Eingabe von Suchbegriffen nachlässig waren, Stichworte eingaben, die von MetaLib aber als Satz interpretiert wurden etc.. Dies führt sie zu einer grundsätzlichen Frage hinsichtlich des Designs von Suchmaschinen: „Should you hide the inherent complexity to the user or let the user control it?“ (Nygren et al., S. 56). Google hat sich klar zum Verstecken der Komplexität und gegen die Kontrolle durch die NutzerInnen entschieden (sowohl mit Blick auf Google Scholar als auch nahezu aller Dienstleistungen des Unternehmens). Das mag – gerade unter BibliothekarInnen – Anlass zu durchaus gerechtfertigter Kritik sein. Allerdings sollte auch berücksichtigt werden, dass nicht jede Suchanfrage auf eine umfangreiche Literaturrecherche abzielt und in einigen Fällen Schnelligkeit und Einfachheit eher gefragt sind als Präzision und Vollständigkeit. Aus dieser pragmatischen Perspektive ließe sich etwas milder Urteilen – demnach ist Google Scholar „right enough often enough“ (White 2006, S. 22).
<b>manchmal geht Schnelligkeit und Einfachheit vor Präzision und Vollständigkeit</b>	So oder so muss anerkannt werden, dass Google Scholar längst zu einem wichtigen und intensiv genutzten Konkurrenten zu etablierten akademischen Suchdiensten geworden ist. Es mag kritische Debatten dazu geben, doch Pomerantz bringt die Problematik auf den Punkt, wenn er feststellt: „Despite this debate, however, users are using Google Scholar“ (2006, S. 55). Er geht sogar noch weiter und legt dar, Google Scholar und Bibliotheken seien weniger Konkurrenten, sondern würden vielmehr voneinander profitieren:
<b>wichtiger Konkurrent etablierter Systeme</b>	„By enabling users to come into contact with library resources in the course of their ordinary Web searches, Google Scholar has the potential to ensure that libraries remain a critical part of the user’s information-seeking process. It benefits Google when a library participates with Google Scholar, but it also benefits the library and the library’s users: the library is able to provide users with a familiar and easy-to-use path to materials.“ (Pomerantz 2006, S. 55)
<b>gegebenenfalls jedoch auch Kooperationspartner</b>	De facto kooperieren auch bereits zahlreiche Bibliotheken durch Öffnung ihrer Kataloge mit Google Scholar und die bibliothekseigene Suchmaschine <i>BASE (Bielefeld Academic Search Engine)</i> der Universität Bielefeld bindet beispielsweise auch Ergebnisse aus Google Scholar ein. Es bedarf weiterer Analysen, um die mittel- bis langfristigen Effekte solcher Praxen auf die Wissenschaftskommunikation beurteilen zu können. So wäre etwa zu fragen, ob die interdisziplinäre Gestaltung von Google Scholar begrüßenswert ist, etwa weil sie Publikationen fächerübergreifend sichtbar machen kann (vgl. Lewandowski 2007, S. 168), oder ob man diese „Aufweichung“ von Fächergrenzen negativ beurteilt. Ähnlich lässt sich die durch Google Scholar vermittelte neue Sichtbarkeit randständiger Publikationen als erfolgreichen und richtigen Schritt zu einer „Demokratisierung“ des Wissenschaftssystems, oder als Gefahr für seine Qualität, deuten. Daran anknüpfend, stellt sich auch die Frage, welche Auswirkungen das durch Google Scholar vorgenommene Ranking haben wird. Das Bemühen sich an diese Ordnungslogik in Form von „Academic Search Engine Optimization (ASEO)“ (Beel et al. 2010) anzupassen, kann als Zeichen für eine „Googlization“ <sup>32</sup> des Wissenschaftsbetriebs verstanden werden.
<b>langfristige Wirkungen unklar</b>	
<b>weiterer Analysebedarf</b>	
<b>„Googlization“ und andere Trends?</b>	

<sup>32</sup> Der Begriff stammt von Siva Vaidhyanathan; vgl. Fußnote 4, S. 2.

So zwingt Google Scholar das Wissenschaftssystem sich gegenüber der Suchmaschine zu positionieren und ein Ignorieren scheint schlechterdings unmöglich:

„In some ways, the introduction of Google Scholar has forced academic librarians to analyze the relationships among publishers, vendors, free search engines, and research-level libraries. The relationship between libraries and Google Scholar, unprecedented in that it does not depend on sales, has implications for the library Web site. It is an example of a new type of collaboration between a commercial enterprise and academia.”

(Mullen/Hartman 2006, S. 107)

**Ignorieren  
unmöglich**

### 3.4.2 Nutzung der Zitationsanalyse

Die in die Suchmaschine implementierte Zitationsanalyse stellt eine wesentliche Grundlage für die Funktionsweise von Google Scholar dar: Sie extrahiert zunächst Metadaten (AutorIn, Erscheinungsjahr etc.) aus den erfassten Dokumenten, die nicht nur für die vordergründig erkennbaren Zitationsergebnisse und die „zitiert durch“-Funktion verwendet werden, sondern auch hintergründig als Bewertungskriterium für das Ergebnisranking eine Rolle spielen. Zudem werden diese Daten ggf. für weitere Funktionen (wie etwa autorspezifische Suchen oder dem Export in Literaturdatenbanken) benötigt.

**Wichtigkeit  
implementierter  
Zitationsanalyse**

Die Extraktion von Zitationen erfolgt im Normalfall automatisiert, d. h. die Dokumente werden maschinell durchsucht, wobei anhand eines Algorithmus Textsegmente von anderen unterschieden werden und als Zitation, Autorenname, Erscheinungsjahr etc. klassifiziert werden, um sie anschließend entsprechend weiter verarbeiten zu können. Dieser Vorgang gestaltet sich insbesondere durch die breite und oftmals nicht standardisierte Datenbasis von Google Scholar als schwierig. Dass die Algorithmen hier schnell an ihre Grenzen geraten, räumt auch der Konzern selbst ein und bittet gleichzeitig um Unterstützung von Menschenhand:

**automatisierte  
Zitationsextraktion**

„Eine automatische Filterung von Informationen aus unterschiedlichen Gebieten ist nicht immer ganz einfach und es können sich leicht Fehler einschleichen. Senden Sie uns bitte eine Suchanfrage, mit der Ihr Artikel gefunden werden kann, und die korrekte Beschreibung, die angezeigt werden soll.“ (Google Scholar 2010a)

**trifft schnell auf  
Grenzen**

Jacsó (2009; 2008) hat sich besonders intensiv mit dieser Problematik auseinandergesetzt und die Fehler und ihre Ursachen detailliert herausgearbeitet. Auf dieser Basis sollen an dieser Stelle einige Beispiele zur Illustration herangezogen werden:

**Beispiele für typische  
Fehler**

- Namen werden von Google Scholar häufig falsch aus dem Web extrahiert, was etwa an Namen wie „Password“, „P Login“, „Subscribe“ oder „Ltd“ erkennbar wird, die besonders in der Anfangsphase der Suchmaschine bei entsprechenden Autorensuchen zu bis zu sechsstelligen Ergebnissen führt(en).
- Umgekehrt werden die tatsächlichen AutorInnen häufig nicht als solche erkannt.
- Google Scholar differenziert bei der Zuordnung von AutorInnen nicht zwischen tatsächlich von diesen geschriebenen Texten, und solchen, die diese nur zitieren. Da Zitate also mitgezählt werden, „verfügen“ AutorInnen häufig fälschlicherweise über hunderte Publikationen.
- Verschiedene Versionen oder Publikationsorte werden teilweise nicht richtig erkannt, so dass sie nicht (wie vorgesehen) als andere Versionen des Textes unterhalb des Ergebnisses auftauchen und registriert werden.

**keine effektive Nutzung  
bereits vorhandener  
Metadaten**

Zum einen resultieren diese Probleme aus der schwierigen Aufgabe der automatischen Identifikation und Extraktion der benötigten Daten. Zum anderen wird jedoch auch kritisiert, dass kooperierende Verlage zwar korrekte Metadaten mitlieferten, Google Scholar diese jedoch offensichtlich nicht effektiv nütze (Jacsó 2009).

**Google Scholar als  
ergänzendes Instrument**

Andererseits ließe sich positiv erwähnen, dass Google Scholar gegenüber dem weitläufig genutzten *SCI (Science Citation Index)* nicht nur den Vorteil bietet, kostenfrei zu sein, sondern zudem auch Publikationen berücksichtigt, die aufgrund der selektiven Auswahlkriterien im SCI nicht erscheinen. So könnte man die Zitationsanalyse Google Scholars auch als ergänzende Alternative wahrnehmen (Noruzi 2005; Harzing/Wal 2008). Vor allem für unabhängige WissenschaftlerInnen und Institutionen, die sich die teuren Gebühren kostenpflichtiger Zitationsanalysedienste nicht leisten können, dürfte Google Scholar interessant sein – alleine schon aus Mangel an bezahlbaren Alternativen. Immerhin eignet sich die „zitiert durch“-Funktion, um schnell und bequem weitere Quellen ausfindig zu machen. Hierfür spielt es keine herausragende Rolle, ob etwa die Namen der AutorInnen richtig erfasst wurden, da diese sowieso in den jeweiligen Veröffentlichungen nachvollzogen werden können. Wenn in solchen Fällen beispielsweise die AutorInnen „A Registered, P Login, P Options, SD Access“ angezeigt werden, dürfte dies für Stirnrüzeln oder Schmunzeln sorgen, beeinträchtigt jedoch nicht die eigentliche Funktionsweise, die dennoch über den entsprechenden Hyperlink gegeben ist.

**als Impact-Indikator  
ungeeignet**

Allerdings fallen die Mängel der Zitationsanalyse, ungleich stärker ins Gewicht, wenn sie als Indikator für den Impact von WissenschaftlerInnen oder Publikationen erhalten soll. Hier sind sie gravierend, da sie in einer quantitativen Interpretation nicht nachvollzogen werden können und die entsprechenden Zahlen schlichtweg nicht reliabel sind. Nun beansprucht Google Scholar auch gar nicht unmittelbar ein Zitationsanalyse-Tool mit einer solchen Funktion zu sein. Dennoch wird es aber als Maßstab für Relevanz genutzt: zum einen, weil davon ausgegangen werden muss, dass viele NutzerInnen die Fehlerhaftigkeit der Daten nicht als solche wahrnehmen. Zum anderen nutzt Google Scholar die implementierte Zitationsanalyse für sein eigenes Ranking der Ergebnisse und trägt damit unmittelbar zur Relevanz von AutorInnen und Publikationen bei. Zudem wird sie bereits von externen Diensten genutzt, die ihre Berechnungen auf Google Scholars Zitationsanalyse stützen.<sup>33</sup>

**dennoch de facto  
relevant ...**

**... und  
aufmerksamkeitssteuernd**

So gilt für die Zitationsanalyse von Google Scholar, was auch oben schon für die Suchfunktion herausgestellt wurde: Die teils vernichtenden Urteile von BibliothekarInnen und SuchmaschinenexpertInnen haben wenig Gewicht, solange Google Scholar von den NutzerInnen dennoch gerne – und häufig eben auch ohne diese Probleme zu reflektieren – angenommen und verwendet wird. Die KritikerInnen mögen richtig liegen, wenn sie Google Scholar als Tool zur Zitationsanalyse ablehnen. In den entsprechenden Fachkreisen dürfte dies auch berücksichtigt werden, allerdings ändert dies nichts an der dennoch gegebenen Relevanz, die der Funktion qua Anwendung fortlaufend zugeschrieben wird. Die Feststellung, dass ein Papier aufgrund fehlerhafter Zitationsanalyse in Google Scholar *zu unrecht* auf- oder abgewertet wird, ändert nichts daran, dass die Suchmaschine de facto die Aufmerksamkeit steuert, die diesem zu- bzw. abhanden kommt – ob zu unrecht oder nicht. Ähnliches gilt auch für den populären SCI, dessen Aussagekraft ebenfalls immer wieder in der Kritik stand und steht, ohne dass dies seine Relevanz ernsthaft einschränkt.

<sup>33</sup> Z. B. [code.google.com/p/citations-gadget](http://code.google.com/p/citations-gadget), [harzing.com/pop.htm](http://harzing.com/pop.htm) (jeweils zuletzt aufgerufen am 12.03.2010).

## 4 Google Books

Im Gegensatz zu den Web-Suchmaschinen Google und Google Scholar durchsucht und organisiert Google Books (auch *Google Book Search*, bzw. im deutschen Sprachraum *Google Bücher* und früher *Google Buchsuche*, ehemals *Google Print*) nicht Online-Dokumente aus dem WWW, sondern beschäftigt sich viel mehr damit, Offline-Dokumente, nämlich gedruckte Bücher, online im Volltext durchsuchbar zu machen. Kurzum: „Google is reaching beyond the Internet and into the real world“ (Vaidhyanathan 2007, S. 1212).<sup>34</sup>

Dies geschieht, indem ihre Inhalte gescannt und digitalisiert werden (4.1). Dazu werden Verfahren der Texterkennung (*OCR, Optical Character Recognition*) eingesetzt, die dafür sorgen, dass auch der Text als solcher erfasst und durchsuchbar wird. Teilweise erhält Google jedoch auch bereits Dateien im Textformat, die den Schritt der Texterkennung überflüssig machen. Das erklärte Ziel von Google Books besteht darin, „(...) gemeinsam mit Verlagen und Bibliotheken einen umfassenden, virtuellen Katalog aller Bücher in allen Sprachen zu erstellen“ (Google Books 2010b).

In Form einer eigenen Webseite samt Domain besteht Google Books seit November 2005, allerdings reichen die Vorarbeiten so weit in die Unternehmensgeschichte zurück, dass der Konzern selbst äußert, man könne „(...) sicherlich sagen, dass das Projekt so alt wie Google selbst ist“ (Google Books 2010a). Demnach gehen die Grundlagen bis auf das Jahr 1996 zurück, als die Gründer Sergey Brin und Larry Page am *Stanford Digital Library Technologies Project* an der Idee einer digitalen Bibliothek arbeiteten. 2002 begann der Konzern dann selbst (unter Federführung von Page) sein eigenes Digitalisierungsprojekt auf die Beine zu stellen. Dazu wurde zunächst eine neue Scanmethode entwickelt, denn bisherige Digitalisierungsprojekte (von denen es eine ganze Reihe gab und gibt) stießen vor allem auf das Problem, dass der Scanvorgang entweder sehr langsam war oder die Bücher beschädigte. 2004 wurde das Projekt *Google Print* ins Leben gerufen, für das in Kooperation mit Verlagen und AutorInnen Bücher digitalisiert wurden bzw. werden (heute auch als *Partnerprogramm* bezeichnet). Zudem startete man auch Partnerschaften mit Bibliotheken (heute *Bibliotheksprogramm*, oder auch *Google Books Library Project* bzw. kurz *Google Library*), in deren Rahmen man mit der Digitalisierung der jeweiligen Bibliotheksbestände begann (4.1). Im Unterschied zu Google Print wurden dabei die RechteinhaberInnen (sofern vorhanden) nicht mit einbezogen, was intensive Kontroversen und Rechtsstreitigkeiten nach sich zog (4.4.2). Auch Google Books übernimmt einige Funktionsmerkmale der Universalsuchmaschine, bietet jedoch auch gänzlich neue Optionen wie etwa die Möglichkeit, Bücher zu rezensieren (4.2/4.3).

Wie die Universalsuchmaschine Google und die Spezialsuchmaschine Google Scholar, beeinflusst auch dieser Dienst das Wissenschaftssystem (4.4). Dies zeigt sich zum einen mit Blick auf die Suche nach wissenschaftlichen Büchern (4.4.1), zum anderen äußert es sich durch Auswirkungen auf AutorInnen und Verlage, vor allem in Form von urheberrechtlichen Kontroversen (4.4.2).

**Google erreicht die Buchwelt**

**Ziel: Digitalisierung aller Bücher**

**Wurzeln in der Frühzeit von Google**

**neue, effektive Scanmethode**

**Kooperationsprogramme mit Verlagen und Bibliotheken**

**Einfluss auf das Wissenschaftssystem**

---

<sup>34</sup> Freilich ist das Internet ebenfalls Teil der „realen Welt“, weshalb diese Dichotomie etwas irreführend sein kann (Lehdonvirta 2008).

## 4.1 Was wird gefunden?

### **Verlage kooperieren über Partnerprogramm**

Wie einleitend bereits erwähnt, speist Google Books seine Inhalte nicht aus dem WWW, sondern aus Kooperationen mit Bibliotheken und Verlagen, in deren Rahmen Bücher (und auch einige Zeitschriften) digitalisiert werden. Dementsprechend ist der von Google Books erfasste Korpus eng an die Angebote dieser Partner geknüpft. Dazu gehören zunächst diverse große *Verlage*, die über das sogenannte *Partnerprogramm* ihre Bücher in Google Books einbringen. Bei der Vorstellung des Vorläufers Google Print im Rahmen der Frankfurter Buchmesse 2004 nannten Brin und Page eine Reihe renommierter Verlage als Teilnehmer.<sup>35</sup> Wie bei den Kooperationen für Google Scholar, veröffentlicht der Konzern auch für dieses Partnerprogramm keine vollständigen Listen der Teilnehmer. Da Google Verlage und AutorInnen nach wie vor zur Partizipation aufruft<sup>36</sup>, ist anzunehmen, dass die Zahl der Kooperationspartner weitaus höher ist.<sup>37</sup>

### **neben Büchern auch Zeitschriften**

Die Verlage lassen dem Konzern digitale Kopien oder gedruckte Exemplare ihrer Bücher zukommen, den dieser dann kostenlos in den Google-Books-Index aufnimmt. Neben Büchern hat Google Books im Dezember 2008 auch damit begonnen, diverse populäre (vornehmlich US-amerikanische) Zeitschriften in sein Programm aufzunehmen. So sind beispielsweise inzwischen alle Ausgaben des Magazins *Life* von 1936 bis 1972 vollständig über Google Books einsehbar.

### **Kooperation mit großen Bibliotheken**

Wie einleitend bereits skizziert, stellen neben Verlagen auch *Bibliotheken* Bücher für Google Books zum Scannen zur Verfügung. Nach Konzernangaben kooperieren eine Reihe großer Bibliotheken mit Google.<sup>38</sup>

### **Google Books verfügt nicht immer über ganze Sammlungen**

Daraus lässt sich allerdings nicht schließen, dass die jeweiligen Bibliotheksbestände vollständig von Google Books erfasst werden. Zum einen dauert das Digitalisierungsprojekt noch an, zum anderen stellen einige Bibliotheken nur Teile ihres Katalogs zur Verfügung. So hat etwa die Bayerische Staatsbibliothek nur urheberrechtlich nicht mehr geschützte Werke für Google geöffnet, und auch die seltenen Handschriften und Inkunabeln der Bibliothek sind von der Kooperation ausgenommen.<sup>39</sup>

<sup>35</sup> Blackwell, Cambridge University Press, University of Chicago Press, Houghton Mifflin, Hyperion, McGraw-Hill, Oxford University Press, Pearson, Penguin, Perseus, Princeton University Press, Springer, Taylor & Francis, Thomson Delmar und Warner Books ([books.google.at/intl/de/googlebooks/history.html](http://books.google.at/intl/de/googlebooks/history.html); zuletzt aufgerufen am 10.03.2010).

<sup>36</sup> [books.google.com/googlebooks/publishers.html](http://books.google.com/googlebooks/publishers.html) (zuletzt aufgerufen am 10.03.2010).

<sup>37</sup> Im Netz kursieren etwa Angaben, nach denen ca. 30.000 Verlage in 100 Ländern Verträge mit Google abgeschlossen haben (vgl. z. B. [io9.com/5501426](http://io9.com/5501426), zuletzt aufgerufen am 22.05.2010).

<sup>38</sup> Bayerische Staatsbibliothek, Columbia University, Committee on Institutional Cooperation (CIC), Cornell University Library, Harvard University, Universitätsbibliothek Gent, Bibliothek der Keio-Universität, Bibliothèque Municipale de Lyon, Katalanische Nationalbibliothek, New York Public Library, Oxford University, Princeton University, Stanford University, University of California, Universität Complutense Madrid, Universitätsbibliothek Lausanne, University of Michigan, University of Texas in Austin, University of Virginia, University of Wisconsin – Madison ([books.google.com/googlebooks/partners.html](http://books.google.com/googlebooks/partners.html); zuletzt aufgerufen am 20.04.2010).

<sup>39</sup> Weiterführende Informationen zu den einzelnen Kooperationen sind über diese Übersichtsseite erreichbar: [books.google.com/intl/de/googlebooks/partners.html](http://books.google.com/intl/de/googlebooks/partners.html) (zuletzt aufgerufen am 10.03.2010).

Im Gegensatz zum Partnerprogramm, das Verlage weiter zur Zusammenarbeit aufruft, sind derzeit keine zusätzlichen Kooperationen mit Bibliotheken vorgesehen, werden aber auch nicht ausgeschlossen:

„Momentan wollen wir den Bestand weiterer US-amerikanischer Bibliotheken nicht digitalisieren. In der Zukunft werden wir unser Programm jedoch möglicherweise erweitern und bestimmte Sammlungen aus Bibliotheken in den USA und in anderen Ländern aufnehmen.“ (Google Books 2010c)

Dem „work in progress“-Status von Google Books entsprechend, wird es an einigen Stellen noch als Beta-Version deklariert, also als unfertig und vorläufig ausgewiesen (wobei der Zusatz „beta“ nicht mehr unmittelbar unter dem Logo erscheint, wie es früher der Fall war und bei Google Scholar auch derzeit noch der Fall ist). Über den aktuellen Stand des Projektes liegen keine präzisen Informationen vor, daher kann an dieser Stelle keine genaue und aktuelle Angabe über den Umfang der bereits verfügbaren Bücher gemacht werden. Allerdings wurden einige Meldungen über den „Zwischenstand“ des Projekts veröffentlicht:

- Im März 2009 gab Google bekannt, nahezu alle Werke aus dem 19. Jahrhundert aus den Beständen der Bodleian Library der Oxford University digitalisiert zu haben. Diese sind nun über Google Books vollständig und kostenfrei zugänglich. Gleichzeitig betonte man, die Kooperation würde auch in Zukunft weiter fortgeführt.<sup>40</sup>
- Im September 2009 gab Google bekannt, es seien nun über 10.000.000 Bücher über den Service verfügbar.<sup>41</sup>
- Im Februar 2010 verkündete der Konzern, das Digitalisierungsprojekt an der University of Texas sei abgeschlossen.<sup>42</sup>

In jedem Fall scheint die Geschwindigkeit des Digitalisierungsprojektes im Vergleich zu anderen sehr hoch zu sein, was sich entsprechend im Umfang niederschlägt.

Gemäß dem Ziel *alle* Bücher zu erfassen entsprechend, finden sich in Google Books verschiedenste Bücher, ohne besonderem thematischen oder sonstigen Fokus, wie Google selbst erklärt:

„Google Buchsuche findet nahezu alle Bucharten: Belletristik, Sachbücher, Nachschlagewerke, wissenschaftliche Abhandlungen, Lehrbücher, Kinderbücher, naturwissenschaftliche und medizinische Fachbücher, Schulbücher und noch viele mehr. Je mehr Bücher unserer Bibliothekspartner wir hinzufügen, umso größer wird unser Bestand. Sie werden also auch vergriffene, seltene und öffentlich zugängliche, nicht mehr urheberrechtlich geschützte Bücher finden können.“ (Google Books 2010c)

Die bereits erfassten Bücher sind zwar alle im Volltext durchsuchbar, allerdings können sie – abhängig vom jeweiligen urheberrechtlichen Status – teilweise nur beschränkt eingesehen werden (dazu 4.2). Bei Büchern aus dem Partnerprogramm bestimmen die jeweiligen Verlage über den sichtbaren Anteil. Demgegenüber sind urheberrechtlich nicht (mehr) geschützte Bücher komplett abrufbar, so dass gerade auch alte und seltene Werke über den Dienst zugänglich werden. So ist etwa die erste englische Übersetzung von Isaac Newtons „Mathematical principles of natural philosophy“ aus dem Jahr 1729 in Google Books vollständig einsehbar.

**keine zusätzlichen Kooperationen mit Bibliotheken**

**„work in progress“**

**schneller Fortschritt**

**keine besonderen inhaltlichen Schwerpunkte**

**variierender Zugang, je nach Urheberrechtslage**

<sup>40</sup> [booksearch.blogspot.com/2009/03/bodleians-treasures-available-to-all.html](http://booksearch.blogspot.com/2009/03/bodleians-treasures-available-to-all.html) (zuletzt aufgerufen am 07.04.2010).

<sup>41</sup> [googleblog.blogspot.com/2009/10/tale-of-10000000-books.html](http://googleblog.blogspot.com/2009/10/tale-of-10000000-books.html) (zuletzt aufgerufen am 13.03.2010).

<sup>42</sup> [booksearch.blogspot.com/2010/02/sharing-latin-american-works-from.html](http://booksearch.blogspot.com/2010/02/sharing-latin-american-works-from.html) (zuletzt aufgerufen am 10.03.2010).

<b>Suchen statt Lesen</b>	Google selbst betont die Suchfunktion gegenüber dem eigentlichen Lesen: „Prinzipiell möchte Google Buchsuche Ihnen dabei helfen, Bücher zu entdecken, und nicht dabei, sie von Anfang bis Ende zu lesen“ (Google Books 2010c).
<b>lokale bzw. sprachliche Unterschiede</b>	Die Zugänglichkeit der Werke kann in einigen Ländern abweichen. Aufgrund ungeklärter Urheberrechtsfragen sind viele Werke, die über die US-Version des Dienstes frei verfügbar sind, in den lokalen Ablegern nicht zugänglich. Allerdings lassen sich viele davon mittels eines Umwegs über den externen Anbieter <i>archive.org</i> ebenfalls außerhalb der USA aufrufen. <sup>43</sup>  Zudem werden viele Sprachen (noch) nicht korrekt erfasst und sind daher für die Volltextsuche nicht verfügbar, wobei laut Google alle Sprachen für das Partnerprogramm zulässig sind. Zu den derzeit indextierten gehören: Chinesisch (traditionell und vereinfacht), Deutsch, Englisch, Französisch, Italienisch, Japanisch, Koreanisch, Niederländisch, Portugiesisch, Russisch und Spanisch; weitere befinden sich in der Testphase (Google Books 2010d).

## 4.2 Was wird wie angezeigt?

<b>weniger minimalistische Startseite</b>	Im Vergleich zu Google und Google Scholar fällt beim Anblick der Startseite von Google Books unmittelbar auf, dass diese nicht minimalistisch im wesentlichen auf das Eingabefeld zur Suche reduziert ist, sondern weit mehr Informationen beinhaltet (Abbildung 4.2-1). Hier dominiert der visuelle Eindruck, der durch eine Aufreihung verschiedener anklickbarer Buch-Cover gegeben wird, die unter verschiedene Kategorien ( <i>Interessant, Klassiker, Zeitschriften</i> , sowie verschiedene Themenfelder) geordnet sind. Fährt man mit dem Mauszeiger über einen dieser Einbände, erscheinen zusätzliche Informationen, wie Titel und AutorIn (in der Abbildung 4.2-1 ist das Beispiel „Sehnsucht als poetologisches Prinzip bei Joseph von Eichendorff“ erkennbar). Ist man beim Dienst angemeldet, finden sich oberhalb zusätzliche Optionen zu den Funktionen aus <i>Meine Bibliothek</i> (dazu 4.3).
<b>zusätzliche Informationen</b>	
<b>Unterschiede zu anderen Google-Diensten</b>	Zudem werden am Seitenrand unter der Funktion <i>Themen durchsuchen</i> weitere Kategorien verlinkt. Auffällig gegenüber anderen Google-Diensten ist hier auch der Verweis „Google hat eine bahnbrechende Vereinbarung mit Autoren und Verlagen getroffen“ direkt auf der Startseite, bei dem es sich offenbar um eine Reaktion auf die massiven Kontroversen im Zusammenhang mit Urheberrechtsfragen handelt (dazu 4.4.2). Ebenso ist die zentrale Umwerbung der Anmeldefunktion vergleichsweise ungewöhnlich, so wie auch die relativ marginalisiert angeordnete Suchmaske. Insgesamt wird durch das Seitendesign eher zum „Stöbern“ als zur schnellen Stichwortsuche eingeladen.

<sup>43</sup> [archive.org/details/googlebooks](http://archive.org/details/googlebooks) (zuletzt aufgerufen am 15.03.2010).



The screenshot shows the Google Books interface. At the top, there is a search bar with the text 'Google bücher' and a search button labeled 'Nach Büchern suchen'. To the right of the search bar is a link for 'Erweiterte Buchsuche'. Below the search bar, there is a section titled 'Beliebte Bücher durchsuchen'. On the left side of this section, there is a list of links for 'Anmelden' and 'Geschäft und Wirtschaft', 'Kochen', 'Computer', 'Current Events', 'Familie und Beziehungen', 'Spiele', 'Garten', 'Gesundheit und Fitness', 'Haus und Heim', 'Humor', 'Recht', 'Literatur', 'Literaturkritik', 'Literary Criticism & Collections', 'Mathematik', 'Medizin', 'Non-Classifiable', 'Körper, Geist und Seele', 'Darstellende Kunst', 'Haustiere', 'Philosophie', 'Fotografie', 'Poesie', and 'Politikwissenschaft'. The main content area is divided into several sections: 'Interessant' featuring a row of book covers with a tooltip for 'Sehnsucht als poetologisches Prinzip bei Joseph von Eichendorff' by Katja Löhr; 'Klassiker (r)' showing a row of classic book covers; 'Zeitschriften' displaying a row of magazine covers; and 'Mathematik' showing a row of mathematics-related book covers. Navigation arrows are visible on the left and right sides of the book rows.

Abbildung 4.2-1: Startseite von Google Books

Anmerkung: Aufgerufen am 15.03.2010.

Zu den Büchern selbst gibt es im Wesentlichen zwei Ansichten: *Über dieses Buch* (bzw. *Über diese Zeitschrift*) und *Lesen* (teilweise auch *Vorschau*). Wie bereits oben angemerkt, ist die Ansicht *Lesen* je nach Urheberrechtsstatus des jeweiligen Werkes sehr unterschiedlich. Grundsätzlich ist hier zwischen vier Ansichten zu unterscheiden<sup>44</sup>:

**verschiedene  
Ansichten**

- *Vollständige Ansicht*: Das Buch ist komplett einsehbar.
- *Eingeschränkte Vorschau*: Von den Rechteinhabern bestimmte Anteile sind einsehbar.

<sup>44</sup> Siehe detailliert auch unter: [books.google.com/intl/de/googlebooks/screenshots.html](http://books.google.com/intl/de/googlebooks/screenshots.html) (zuletzt aufgerufen am 20.04.2010).

- *Auszugsansicht*: Ergebnisse aus einer Suche in einem derartigen Buch zeigen lediglich sehr kleine Ausschnitte (Abbildung 4.2-2).
- *Keine Vorschau verfügbar*: Die Funktion *Lesen* ist vollständig deaktiviert. Dies kann auf Wunsch der Rechteinhaber erwirkt werden.

### weitere Möglichkeiten

In den ersten beiden Fällen wird beim Klick auf ein Buch dieses direkt in der „Lesen“-Ansicht geöffnet, andernfalls in der „Über“-Ansicht. Sofern die „Lesen“-Funktion verfügbar ist, kann das Buch mittels verschiedener Ansichten (eine oder mehrere Seiten, Vollbild) „durchblättern“ werden, wobei in der Auszugsansicht jeweils einige Seiten fehlen. Wenn vorhanden, kann auch über ein hyperlinkförmiges Inhaltsverzeichnis zu einzelnen Kapiteln gesprungen werden. Zudem ist es möglich, Hyperlinks zu bestimmten Buchseiten zu erhalten, die dann weitergegeben oder auf Webseiten eingebunden werden können. Bei vollständig einsehbaren Büchern können darüber hinaus gewünschte Seitenzahlen direkt per Eingabe angewählt werden und Buchausschnitte markiert und kopiert werden (als Text und verlinkbares Bild). Zudem steht hier die Funktion *Nur Text* zur Verfügung, mit der u. a. Technologien zur Unterstützung von Sehbehinderten kompatibel sein sollen<sup>45</sup>. Außerdem können die Bücher als PDF-Dateien heruntergeladen werden.

**Nietzsche: his philosophy of contradictions and the contradictions of his ...** Von Wolfgang Müller-Lauter

2 Ergebnisse für **Wilhelm Windelband** werden in diesem Buch angezeigt. - Sortieren nach: [Relevanz](#) | [Seiten](#)

**Seite 183 »**

3. ...  
 4. Richard H. Grützmacher, *Nietzsche*, p. 143.  
 5. Alois Richl, *Friedrich Nietzsche*, p. 25.  
 6. **Wilhelm Windelband**, *Lehrbuch der Geschichte der Philosophie*, pp. 577ff.  
 7. Paul Deussen, *Erinnerungen an Friedrich Nietzsche*, p. 100.  
 8. Karl Löwith called attention to this misuse in *Von Hegel zu Nietzsche*, p. 211. In the same passage Löwith gives a series of examples of the obvious and irresolvable contradic-

**Seite 246 »**

...  
 76.  
 Weizsäcker, Carl Friedrich von. *Die Geschichte der Natur*. Göttingen, 1962. Eng.: *The History of Nature*, trans. Fred U. Wieck. Chicago: University of Chicago Press, 1976.  
**Windelband, Wilhelm**. *Lehrbuch der Geschichte der Philosophie*, H. Heimseth, ed. Tübingen, 1957.  
 Wolff, Hans M. *Friedrich Nietzsche. Der Weg zum Nichts*. Bern, 1956.  
 Ziegler, Theobald. *Friedrich Nietzsche*. Berlin, 1900.

Abbildung 4.2-2: Ausschnitt-Ansicht in Google Books  
 Anmerkung: Aufgerufen am 15.03.2010.

### Zusatzinformationen

In der Ansicht *Über dieses Buch* finden sich eine Reihe an Zusatzinformationen zu dem ausgewählten Buch. Neben den üblichen bibliographischen Daten zählen dazu auch Rezensionen (die von NutzerInnen selbst verfasst werden können; dazu 4.3), Referenzen von Webseiten, anderen Büchern, sowie aus Google Scholar. Weiters werden *Andere Ausgaben*, *Ähnliche Bücher* und *Häufige Begriffe und Wortgruppen* angezeigt. Teilweise erscheint auch eine Einbindung von *Google Maps*, in der Orte markiert werden, die in dem jewei-

<sup>45</sup> [googleblog.blogspot.com/2007/07/ever-more-books-to-read.html](http://googleblog.blogspot.com/2007/07/ever-more-books-to-read.html)  
 (zuletzt aufgerufen am 10.03.2010).

ligen Buch genannt werden. Je nach Buch sind mal mehr, mal weniger dieser Features verfügbar und beim Durchsuchen von Zeitschriften finden sich weitere Optionen.

In beiden Ansichten (*Über dieses Buch* und *Lesen*) bleibt eine *Seitenleiste* erhalten, über die Stichwortsuchen innerhalb des Buches durchgeführt werden können. Zudem finden sich hier Links zu Online-Händlern, lokalen Buchhandlungen und Bibliotheken über die das Buch verfügbar ist, sowie Werbeanzeigen und Rezensionen.

### Seitenleiste

Gibt man eine Suchanfrage in die entsprechende Maske im oberen Bereich der Seite ein (*Nach Büchern suchen*), erscheinen die Ergebnisse wie bei Google und Google Scholar in einer hierarchischen Listenform (Abbildung 4.2-3).

The screenshot shows the Google Books search interface. At the top, the search term 'definitionen geschichte' is entered in the search box, with a 'Nach Büchern suchen' button and a link to 'Erweiterte Buchsuche'. Below the search bar, there are navigation options for 'Bücher', 'Listenansicht', and 'Coveransicht'. A list of search results is displayed, each with a book cover thumbnail, the title, author, page count, and a brief description. The results include:

- Mikroskop und Mikroskopie- ein wichtiger Helfer auf vielen Gebieten ...** by Wolfgang Piersig, 2009, 76 Seiten. Description: Im vorliegenden Werk zum Mikroskop wird aufgezeigt, daß diese sehr kleine Gegenstände dem Auge vergrößert darstellen können und es ein Gerät ist, welches ...
- Lehrbuch der Logik auf positivistischer Grundlage mit Berücksichtigung der ... - Seite 592** by Theodor Ziehen, 1919, 866 Seiten. Description: Die 11 Elen treten auf, wenn mehrere Definitionen aneinandergereiht werden, und bestehen darin, ...
- Geschichte des vatikanischen Konzils von seiner ersten Ankündigung bis zu ... - Seite 240** by Theodor Grandera, Konrad Kirch, 1906. Description: ... daß der Papst bei seinen Definitionen in Sachen der Sitten auch über die bürgerlichen und politischen ...
- Geschichte der neuern Philosophie, Band 1-2 - Seite 219** by Kuno Fischer, René Descartes, Baruch Spinoza, 1865. Description: werden bewiesen durch die Sätze, welche vorhergehen : Definitionen, ...
- Grundlagen des Studiums der Geschichte: eine Einf - Seite 2** by Egon Boshof, Kurt Düwell, Hans Kloft, 1973, 329 Seiten. Description: Definition der Geschichte Am Anfang steht eine wichtige terminologische Klärung, ...

On the left side of the page, there are various navigation links such as 'Anmelden', 'Themen durchsuchen', 'Geschäft und Wirtschaft', 'Kochen', 'Computer', 'Current Events', 'Familie und Beziehungen', 'Spiele', 'Garten', 'Gesundheit und Fitness', 'Haus und Heim', 'Humor', 'Recht', and 'Literatursammlungen'.

Abbildung 4.2-3: Beispielsuche in Google Books

Anmerkung: Suche durchgeführt am 15.03.2010.

<b>Ergebnisliste mit Unterschieden zu anderen Google-Diensten</b>	Google Books zeigt in der Ergebnisliste zudem Miniatursichten der jeweiligen Buchcover. Neben den Buchtiteln führen die Ergebnisüberschriften teilweise auch Seitenzahlen auf, die beim Anklicken direkt aufgerufen werden. Weiters finden sich (soweit verfügbar) Informationen zu den AutorInnen, des Erscheinungsjahrs und der Seitenzahl des Buches. Unterhalb eines Textauschnitts wird die Anzeigeform, in der das Werk einsehbar ist, angegeben. Zudem kann hier direkt in die Ansicht <i>Über dieses Buch</i> gesprungen werden und teilweise wird auf weitere Ausgaben verwiesen.
<b>unbekannte Ranking-Kriterien</b>	Nach welchen Kriterien die Rangfolge der Ergebnislisten bestimmt werden, ist unklar. Nunberg kritisiert, diese machten „no sense at all“ und würden häufig angeführt von „inferior or shoddy editions of classic works“ (2009). Auch die Beispielsuche mit dem Begriffspaar „definitionen geschichte“ (Abbildung 4.2-3), erweist sich weder als hilfreich, um Definitionen des Begriffs „Geschichte“ zu finden, noch um etwas über die Geschichte von Definitionen zu erfahren. Stattdessen führt das erste Ergebnis zu einem Werk mit dem Titel „Mikroskop und Mikroskopie – ein wichtiger Helfer auf vielen Gebieten“. Beim Blick auf den Buchumschlag findet sich das Begriffspaar im Untertitel: „Definitionen, Geschichte, Daten, Literatur“.
<b>keine effektive Relevanzbestimmung über Links</b>	Beispiele wie dieses deuten darauf hin, dass Google Books beim Ranking derzeit noch stark von abfrageabhängigen Faktoren bestimmt wird. Diese Annahme erscheint schlüssig, da die Bücher sich (noch) nicht in einem mehr oder weniger dichten Netzwerk aus Hyperlinks befinden, wodurch sich diese Art der Relevanzbestimmung nicht effektiv anwenden lässt. Eine abfrageunabhängige Bewertung, wie sie von der Universalsuchmaschine Google sonst in Form des PageRanks relativ effektiv und erfolgreich vorgenommen wird, ist unter solchen Bedingungen kaum möglich. Auch Nunberg stellt heraus, dass die übliche Google-Architektur bei der Buchsuche an ihre Grenzen gerät und betont die Wichtigkeit von Metadaten, denen jedoch in Google Books kein besonderer Stellenwert zukommt:
<b>unzureichende Berücksichtigung von Metadaten</b>	„It’s clear that Google designed the system without giving much thought to the need for reliable metadata. In fact, Google’s great achievement as a Web search engine was to demonstrate how easy it could be to locate useful information without attending to metadata or resorting to Yahoo-like schemes of classification. But books aren’t simply vehicles for communicating information, and managing a vast library collection requires different skills, approaches, and data than those that enabled Google to dominate Web searching.“ (Nunberg 2009)
<b>Beispiele für mangelhafte Metadaten</b>	Metadaten sind in Google Books oft mangelhaft <sup>46</sup> , da sie offenbar meist automatisch ausgelesen werden, ähnlich wie bei der Zitationsextraktion von Google Scholar (3.4.2). Darunter leidet auch die Funktionsfähigkeit vieler der oben vorgestellten Anzeigemöglichkeiten. So führt Nunberg (2009) eine Reihe kurioser Beispiele an, die dies illustrieren: Demnach erscheinen in Google Books zahlreiche Bücher für das Jahr 1899, die sicherlich nicht in diesem Jahr publiziert worden sind, z. B. Stephen Kings „Christine“ oder Robert Sheltons Biographie über Bob Dylan. Die Dominanz dieses Jahres ist auf einen Metadatenanbieter zurückzuführen, der es standardgemäß verwendet, wenn das Erscheinungsjahr unbekannt ist, wie Google später erklärte (Oder 2009).
<b>auch Kategoriensystem fehlerhaft</b>	Auch das Kategoriensystem beschreibt Nunberg (2009) als fehlerhaft, etwa wenn „Moby Dick“ unter <i>Computer</i> kategorisiert wird, oder „The Cat Lover’s Book of Fascinating Facts“ unter <i>Technology &amp; Engineering</i> . Teilweise geraten die Zuordnungen der Metadaten zum Titel auch gänzlich durcheinander,

---

<sup>46</sup> Darauf weisen neben Nunberg auch weitere BeobachterInnen von Google Books hin (Duguid 2007; Baksik 2006).

so beispielsweise wenn Sigmund Freud and Katherine Jones 1939 das Buch „The Mosaic Navigator: The Essential Guide to the Internet Interface“ veröffentlicht haben sollen. Laut Nunberg handelt es sich dabei nicht lediglich um seltene Ausnahmefehler. Viel mehr treten diese in einem Ausmaß auf, das er als „endemic“ (ebd.) bezeichnet. Google selbst bestreitet dies auch gar nicht und äußerte in Reaktion auf Nunbergs Kritik sogar, Google Books weise nicht hunderte oder tausende derartiger Fehler auf (wie es Nunberg vermutete), sondern diese träten viel mehr millionenweise auf (Oder 2009).

### 4.3 Einflussmöglichkeiten der NutzerInnen

Wie bei anderen Produkten des Konzerns üblich, werden auch in Google Books Möglichkeiten zur spezifizierten Sucheingeabe geboten, die über die *Erweiterte Buchsuche* erreichbar sind. Neben den üblichen Optionen wie z. B. der Suche nach genauen Wortgruppen oder mittels ausschließender Begriffe, die ebenfalls durch entsprechende Operatoren verwendbar sind, lassen sich hier Suchen auf Bücher mit bestimmten Ansichten, Sprachen und Erscheinungszeiträumen beschränken. Außerdem kann u. a. gezielt nach Titel, AutorIn, Verlag, ISBN/ISSN gesucht werden. Hier gilt wiederum das, was bereits für Google Scholar in diesem Zusammenhang angemerkt wurde (3.3): Viele dieser Funktionen sind aufgrund fehlerhafter Metadaten nur eingeschränkt zuverlässig nutzbar.

Wie sich bereits am Design der Startseite zeigt (4.2), gestaltet sich Google Books im Vergleich zu Google und Google Scholar weniger minimalistisch auf die Sucheingeabe reduziert und bietet stattdessen zusätzliche Möglichkeiten zur Erschließung von Quellen. Neben den oben schon dargelegten Zugängen via themenspezifischer Kategorisierungen usw. (4.2), hebt sich Google Books besonders durch die interaktiven Angebote von den anderen untersuchten Suchmaschinen ab. Über einen kostenlosen Google-Account kann man sich beim Service *Meine Bibliothek* anmelden, wodurch einige partizipative Möglichkeiten eröffnet werden. So können Bücher bewertet und rezensiert werden und über so genannte *Bücherregale* lassen sich individuelle Sammlungen organisieren, die auch öffentlich einsehbar gemacht werden können. Eine solche Interaktivität ist bemerkenswert, da sie sonst weder von Google bzw. Google Scholar und schon gar nicht von herkömmlichen Bibliotheken erlaubt wird (schon aus praktischen Gründen ist eine beliebige individuelle Organisation von Büchern hier kaum denkbar). In diesem Zusammenhang ist auch erneut die Möglichkeit des Verlinkens einzelner Buchpassagen zu nennen, die einen gänzlich neuen Zugang zu Büchern eröffnet. Weiters werden NutzerInnen des Dienstes eingebunden, indem sie in der „Lesen“-Ansicht Probleme melden können, d. h. auf fehlende oder fehlerhafte Seiten aufmerksam machen können.

Für AutorInnen bzw. Verlage besteht die Möglichkeit, Werke aus dem Bibliotheksprogramm auszuschließen (opt-out), über deren Urheberrechte sie verfügen (dazu auch 4.4.2).<sup>47</sup>

Weitere begrenzte Einflussmöglichkeiten wurden durch Hacking ausgelotet. So wurde etwa versucht, das implementierte System zum Digital Rights Management (DRM), das eingeschränkte Inhalte vor Vollzugriffen schützen soll, zu umgehen.<sup>48</sup>

#### **Erweiterte Buchsuche**

#### **zusätzliche Möglichkeiten zur Erschließung von Quellen**

#### **u. a. durch „Meine Bibliothek“**

#### **interaktive Möglichkeiten**

#### **Ausschlussmöglichkeit für RechteinhaberInnen**

#### **Hacking**

<sup>47</sup> [books.google.com/partner/exclusion-signup](http://books.google.com/partner/exclusion-signup) (zuletzt aufgerufen am 13.04.2010).

<sup>48</sup> [kuro5hin.org/story/2005/3/7/95844/59875](http://kuro5hin.org/story/2005/3/7/95844/59875) (zuletzt aufgerufen am 07.04.2010).

## 4.4 Einsatz in der Wissenschaft

Google Books wirkt sich vor allem auf zweierlei Ebenen auf das Wissenschaftssystem aus: Zum einen stellt es eine neue Form der Suche nach wissenschaftlichen Büchern dar (4.4.1), zum anderen fordert der Service durch sein Bibliotheksprogramm AutorInnen und Verlage heraus und wirft grundsätzliche urheberrechtliche Fragen auf (4.4.2).

### 4.4.1 Suche nach wissenschaftlichen Büchern

#### **große Erwartungen**

Die wesentliche Neuerung, die Google Books bei der Suche nach wissenschaftlichen Büchern bietet, liegt auf der Hand: Der Dienst erlaubt eine sekunden-schnelle Volltextsuche in Millionen von Büchern. Googles Ziel, sämtliche Bücher digital verfügbar zu machen, veranlasst einige dazu, Google Books mit der antiken Bibliothek von Alexandria zu vergleichen, die ähnliche Absolutheitsansprüche verfolgte (Kelly 2006). Nunberg (2009) betont der Dienst sei nicht nur auf dem Weg die größte, sondern fast sicher auch die „letzte“ digitale Bibliothek zu werden. Es verwundert daher nicht, dass das Projekt intensiv unter akademischen BibliothekarInnen diskutiert wird, die ja zum Teil gleichzeitig selbst zu den wichtigsten Kooperationspartnern von Google Books gehören (4.1).

#### **macht Bibliotheken nicht obsolet**

Während kaum bestritten werden kann, dass Google Books schon jetzt auf die akademische Welt einwirkt, wird es jedoch in absehbarer Zukunft wohl kaum traditionelle Bibliotheken obsolet machen. Das liegt zunächst schlichtweg daran, dass Google Books eben keine Bibliothek ist. Dadurch dass urheberrechtlich geschützte Werke nicht (bzw. nur eingeschränkt) einsehbar sind, handelt es sich bei dem Dienst keineswegs um einen digitalen Ersatz für analoge Bibliotheken:

„Aufgrund der urheberrechtlichen Schranke ist also auch in Google Book Search der Medienbruch zwischen der gedruckten und der elektronischen Welt der wissenschaftlichen Information nur ansatzweise überwunden; insofern ist Google Book Search weniger eine ‚digitale Bibliothek‘ (im Sinne einer Sammlung online zugänglicher Texte) als vielmehr eine Suchmaschine für Buchinhalte, die bestimmte Stichwörter oder Wortfolgen in einem gedruckten Werk auffindbar macht und zudem einen ersten, ausschnitthaften Einblick in das Werk ermöglicht.“ (Stöber 2006, Absatz 16)

#### **neue Möglichkeiten für historische Quellen ...**

Allerdings ergeben sich schon jetzt umfangreiche Möglichkeiten in Disziplinen, deren Fokus gerade auf die bereits frei verfügbaren historischen Texte gerichtet ist. Neben der Geschichtswissenschaft hat beispielsweise auch die sprachhistorische Forschung ein Interesse an solchen Dokumenten und Google Books wird hier bereits als eine mögliche Quelle genutzt (Gloning 2009). So lässt sich in diesem Bereich durch die Volltext-Verfügbarkeit schnell nachvollziehen, wann und wo bestimmte Wörter verwendet wurden. Es ist anzunehmen, dass einigen alten und schwer zugänglichen Büchern über diesen neuen Weg wieder eine gesteigerte Relevanz zukommt. Zumindest erhalten sie neue Chancen, die sie sonst nicht erhalten hätten, etwa weil sie nicht weiter verlegt werden.

#### **... und durch Verlinkung von Büchern**

Dazu tragen auch die durch Google Books gegebenen Möglichkeiten zur Verlinkung von Büchern und Buchsegmenten bei. Da der Dienst noch jung und technisch nicht ausgereift ist, stecken diese allerdings noch in den Kinderschuhen<sup>49</sup>.

<sup>49</sup> Teilweise wird die Verlinkung zu Google Books aus wissenschaftlichen Artikeln bereits standardmäßig (soweit möglich) hergestellt, z. B.

<http://europeangovernance.livingreviews.org> (zuletzt aufgerufen 03.06.2010).

Einige erwarten sich dadurch jedoch einschneidende Veränderungen. So entwirft Kelly das Bild einer *universal library*, in der die Bücher aus ihrer Isolation gelöst und miteinander vernetzt werden:

„Turning inked letters into electronic dots that can be read on a screen is simply the first essential step in creating this new library. The real magic will come in the second act, as each word in each book is cross-linked, clustered, cited, extracted, indexed, analyzed, annotated, remixed, reassembled and woven deeper into the culture than ever before. In the new world of books, every bit informs another; every page reads all the other pages.“  
(Kelly 2006)

Zwar klingt dies aus heutiger Perspektive noch ein wenig nach Science Fiction, jedoch sind die technischen Grundsteine für derartige Funktionen durch Google Books tatsächlich bereits gelegt. Die Suche nach Büchern beschränkt sich dabei nicht auf die Stichworteingabe in einer Suchmaske, sondern erlaubt pluralisierte und heterogene Zugangsformen, die durch die Funktion individuell und öffentlich nutzbarer „Bücherregale“ schon rudimentär vorhanden sind. Diese erlauben potenziell eine neue Unabhängigkeit von traditionellen Katalogisierungssystemen, die Bücher ihrer Ordnungslogik unterwerfen. Die Volltextsuche per Stichworteingabe hat demgegenüber den Vorteil, nicht an unflexible Kategorisierungen gebunden zu sein, die sich später als untauglich erweisen können (dazu auch: Becker 2009). Stattdessen können hier gänzlich neue Ordnungen geschaffen werden.

Allerdings macht dies Bibliotheken eben nicht obsolet. Unabhängig davon, dass Google Books in seiner derzeitigen Form ohnehin keinen Ersatz darstellt, da viele Titel hierüber gar nicht verfügbar sind, erfüllen Bibliotheken sehr viel mehr Dienstleistungen als die physische Bereitstellung von Büchern. Das wird gerade durch die massiven Probleme automatisch generierter Metadaten deutlich. Es ließe sich sogar argumentieren, dass *aufgrund* der heterogenen Distributionsmöglichkeiten von Wissen, Bibliotheken umso nötiger sind, um unter diesen Bedingung Orientierung und Ordnung zu stiften (Fister 2009; Stöber 2006). Tatsächlich sehen viele (darunter wohl auch die teilnehmenden Bibliotheken) Google Books auch weniger als Konkurrenten, sondern als günstigen Weg, NutzerInnen auch im Online-Zeitalter zu den eigenen Beständen zu führen (Dames 2006). Zweifelsfrei geht langfristig ohnehin kaum ein Weg an der Digitalisierung und Online-Verfügbarmachung von Bibliotheksbeständen vorbei:

„(...) on the demand side, everyone wants and expects to do almost everything over the Internet, at least in the initial stages of search. Even our most curmudgeonly faculty, the ones who love breathing the dust of musty tomes, start their day from home looking at the online catalog and browsing collections that they can get to online. Except for the most arcane materials and users, that which is not available online will simply not be read.“ (Courant 2006)

Google bietet dabei einige nahezu unwiderstehliche Vorteile: Der Konzern verfügt nicht nur über die sehr teure und aufwendige Technologie, die umfangreichen Bestände schnell und schonend zu digitalisieren, er bietet sie auch noch umsonst an. Bei geschätzten 200 Millionen US-Dollar, die alleine für das Bibliotheksprogramm veranschlagt wurden (Baksik 2006, S. 400)<sup>50</sup>, wird klar, warum es für viele attraktiv erscheint. Es ist jedoch gerade diese Attraktivität des Quasi-Monopolisten und die damit einhergehende Alternativlosigkeit, die KritikerInnen befürchten lässt, es könnten dadurch ungewollte kul-

**neue  
Zugangsformen ...**

**... ersetzen jedoch  
Bibliotheken nicht**

**Trend zur Online-  
Verfügbarmachung**

**Vorteile von  
Google Books**

**Befürchtungen**

<sup>50</sup> Vaidhyathan (2007, S. 1216) nennt dagegen einen Wert von „nur“ 170 Millionen US-Dollar. Da weder die genauen Kosten, noch der exakte Umfang des Projekts bekannt sind, handelt es sich bei derartigen Angaben um ungefähre Schätzwerte. In jedem Fall lässt sich wohl urteilen, dass das Projekt sehr teuer ist.

turhegemoniale Prozesse in Gang gesetzt werden (Jeanneney 2006). Damit in Verbindung stehen auch Urheberrechtsfragen, die Gegenstand einer anhaltenden Debatte sind (4.4.2).

**Mangel an  
Nutzungsstudien**

Abseits solcher Grundsatzdiskussionen ist mangels Nutzungsstudien bislang wenig über die Auswirkungen von Google Books auf konkrete Suchpraxen im wissenschaftlichen Kontext bekannt. Die Mängel im Bereich der Metadaten, die zweifelhafte Relevanz der Ergebnisse und die vielfach nicht (oder nur sehr eingeschränkt) möglichen Einblicke in die Bücher, deuten darauf hin, dass die Buchsuchmaschine in naher Zukunft die Suche nach wissenschaftlichen Werken *nicht* revolutionieren wird. Für den akademischen Bedarf ist sie in vielen Fällen schlichtweg nicht praktikabel und wird daher den Gang in die Bibliothek vorerst sicherlich nicht gänzlich unnötig machen. Das schließt nicht aus, dass Google Books in einzelnen Fällen durchaus brauchbar sein kann, so wie oben für historische Werke aufgeführt. Gerade für die frei verfügbaren Bücher ergeben sich neue Verwertungsmöglichkeiten, die offenbar auch schon erfolgreich genutzt werden. So dürften alle Fächer mit einem Interesse an historischen Büchern Google Books zumindest als Ergänzung zu bisherigen Suchpraxen in Betracht ziehen. Ein vollwertiger Ersatz kann Google Books auch hier nicht sein, da viele Bücher nach wie vor nicht im Programm sind und einige wohl auch nie sein werden (so wie etwa die Inkunabeln der Bayerischen Staatsbibliothek, die nicht für das Projekt freigegeben wurden). Hinzu kommt das Problem der häufig mangelhaften Metadaten, die einen vollwertigen und alleinigen Einsatz von Google Books unwahrscheinlich machen.

**keine Revolutionierung  
der Suche**

**praktische  
Anwendungsbereiche  
der Volltextsuche**

Abseits historischer Felder dürfte die Volltextsuche in einigen Fällen auch bei eingeschränkt zugänglichen Büchern hilfreich sein, um sehr spezielle Textpassagen zu lokalisieren. Hat man etwa die Quelle eines Zitats „verlegt“, kann die entsprechende Seitenzahl in dem jeweiligen Buch mit Google Books durch Eingabe genauer Wortfolgen schnell gefunden werden – vorausgesetzt natürlich, dass es hier bereits indiziert ist. Auch die Ursprünge sehr spezieller (Fach-) Begriffe lassen sich unter Umständen durch eine Suche in Google Books ergründen. Aufgrund des – im Gegensatz zur Websuche – noch sehr ineffektiven Rankingsystem und den unzuverlässigen Metadaten erscheint Google Books derzeit für allgemeinere Suchanfragen jedoch eher wenig zweckdienlich.

**Langzeitwirkung kaum  
abschätzbar**

Die Langzeitwirkung des Projekts ist derzeit kaum abschätzbar. Nicht nur weil wenig über die tatsächliche Nutzungspraxis bekannt ist und Google Books noch immer im Aufbau ist. Auch ist die Plattform in hohem Maße abhängig von zukünftigen Rechtsurteilen und der Politik der RechteinhaberInnen.

## 4.4.2 Einfluss auf Verlage und AutorInnen

**unkontroverse ...**

Wie dargelegt (4.4.1), ist die Suche nach wissenschaftlichen Büchern, wie sie von Google Books angeboten wird, nicht unumstritten. Sie berührt vor allem ein Feld, das traditionell von Bibliotheken eingenommen wurde, die eine Infrastruktur für das Suchen und Finden solcher Publikationen boten und bieten. Dennoch scheint die Wirkung von Google Books auf diesen Sektor zunächst keine unüberwindbaren Hürden zwischen dem Unternehmen und Bibliotheken hervorzubringen. Im Gegenteil ist das Bibliotheksprogramm ein Resultat gemeinsamer Kooperation. Während es zwar durchaus kritische Stimmen zu dieser Praxis gibt, ist sie doch zwischen diesen Akteuren eher unumstritten, verfolgen doch beide – zumindest vom Anspruch her – das Ziel der Verfügbarmachung und Organisation von Wissen. Kontrovers wird diese Kooperation jedoch, wenn die Perspektive einer dritten Partei mit einbezogen

**... und kontroverse  
Kooperationen**



wird, nämlich die der RechteinhaberInnen der Werke, d. h. in erster Linie Verlage und AutorInnen. Während viele davon im Rahmen des Partnerprogramms durchaus mit dem Konzern zusammen arbeiten, entstanden durch das Bibliotheksprogramm intensive Streitigkeiten (vgl. Baksik 2006), die bis heute andauern. Die Grundproblematik resultiert dabei daraus, dass sich die RechteinhaberInnen vom Bibliotheksprogramm übergangen fühlten. Während sie im Partnerprogramm selbst darüber entscheiden, welche Ressourcen sie an Google weitergeben, verlaufen die Vereinbarungen zwischen dem Konzern und den beteiligten Bibliotheken bilateral, ohne sie miteinzubeziehen: Die Bibliotheken stellen die Bücher zur Verfügung, Google scannt und digitalisiert sie und überlässt den Bibliotheken eine vollwertige digitale Kopie (wobei einige Bibliotheken die daraus resultierende Kontroverse umgehen, indem sie nur urheberrechtlich nicht mehr geschützte Werke zur Verfügung stellen).

Der Konzern sah in dieser Praxis kein Problem, da nur kleine Ausschnitte aus den betroffenen Werken angezeigt wurden. Damit wähnte man sich im Rahmen einer Regelung, die im amerikanischen Urheberrecht auch unter *fair use* bekannt ist und in der Wissenschaft angewendet wird: Sie regelt, dass unter Einhaltung gewisser Bedingungen, urheberrechtlich geschützte Werke verwendet werden dürfen, ohne die jeweiligen RechteinhaberInnen um Erlaubnis fragen zu müssen, und kommt etwa beim Zitieren zur Anwendung. Es entbrannten jedoch Streitigkeiten darüber, ob Google diese Umstände einhielt, da sie mitunter Auslegungssache sind. So kann z. B. einerseits der allgemeine Nutzen der Verfügbarmachung durch Google betont werden, andererseits lässt sich darin der Zweck der eigenen Bereicherung sehen (vgl. Baksik 2006, S. 402 f.). Weiterhin sorgte die so genannte *Opt-out*-Regelung für Kontroversen, nach der betroffene RechteinhaberInnen sich bei Google Books melden müssen, wenn sie ihre Inhalte aus der Plattform ausschließen wollen. Diese Regelung stieß auf vielfachen Widerstand, u. a. da viele AutorInnen gar nicht in der Lage zu so einem Widerspruch sind, etwa weil sie nichts von dieser Regelung wissen. Gleichzeitig scheint jedoch auch jedes Digitalisierungsprojekt von vornherein zum Scheitern verurteilt, sollte es versuchen, alle Urheberrechtsansprüche eindeutig zu klären. Denn man muss davon ausgehen, dass dies bei einem nicht unbeträchtlichen Teil gar nicht, oder nur sehr schwer möglich ist (Baksik 2006, 406 f.). Während OptimistInnen vor einigen Jahren noch glaubten, Google habe durch seine neue Plattform den „Gordian knot of copydom“ (Kelly 2006) geöffnet, ist 2010 wohl eher das Gegenteil der Fall:

Es kam zu verschiedenen Klagen im In- und Ausland gegen den Konzern, die an dieser Stelle nicht einzeln aufgeführt werden sollen. Bis heute von besonderer Bedeutung ist eine Sammelklage unter Führung der Autorenvereinigung *Author's Guild*, die nach über zwei Jahren Verhandlungen schließlich 2008 in einem Vergleich mündete, dem *Google Book Settlement*<sup>51</sup>, das 2009 neu aufgelegt wurde. Dabei handelt es sich um eine äußerst komplexe Vereinbarung – allein die entsprechende Aufreihung der FAQ (Frequently Asked Questions) umfasst 86 Fragen. Entsprechend schwierig ist sie für juristische LaiInnen zu durchschauen.

Auch lassen sich derzeit in dem hier gegebenen Rahmen keine seriösen Einschätzungen dazu formulieren. Das hängt nicht zuletzt damit zusammen, dass die bisherige Vereinbarung eine weitere Neuauflage erfahren könnte. Das erscheint sogar wahrscheinlich, da sich weiterhin Protest gegen das *Google Book Settlement* regt. So fühlen sich etwa UrheberInnen von visuellen Inhal-

**fair use?**

**Opt-out**

**viele ungeklärte  
Urheberrechtsfragen**

**diverse Klagen gegen  
Google**

**juristischer Vergleich:  
Google Book Settlement**

**Konsequenzen und  
Entwicklungen kaum  
absehbar**

<sup>51</sup> [googlebooksettlement.com](http://googlebooksettlement.com) (zuletzt aufgerufen am 20.04.2010).

ten (darunter z. B. die *American Society of Media Photographers, ASMP*) in ihrem Recht übergangen und klagen ebenfalls gegen den Konzern.<sup>52</sup>

**weiterhin Profiteure  
und „Aussteiger“**

Es sollte auch nicht vergessen werden, dass trotz Klagen von AutorInnen und Verlagen, viele nach wie vor von Google Books profitieren. So wird etwa Paul Manning, Vizepräsident des Buchvertriebs für *Springer*, von Google zitiert: „Die Google Buchsuche ist eines der einfachsten und besten Online-Marketingtools.“<sup>53</sup> Gleichzeitig haben bereits Hunderte AutorInnen von der Opt-out-Regelung Gebrauch gemacht, die offenbar nicht mehr mit Google Books zusammenarbeiten möchten.<sup>54</sup>

**grundsätzliche  
urheberrechtliche  
Fragen**

Es bleibt zu beobachten, wie sich diese Dynamiken weiter entwickeln. Dabei können gerade juristische Details, die in der Komplexität des *Google Books Settlements* bislang „untergehen“, zu großen Veränderungen führen. So ist die Frage nach den langfristigen Auswirkungen von Google Books auch nicht durch eine Deskription der (technischen) Möglichkeiten der Plattform zu klären – wie es hier versucht wurde – sondern wohl eher durch eine juristische Analyse und weitere Beobachtung. Entsprechend scheint der Dienst Google Books selbst auch gar nicht so sehr zu tiefgreifenden Veränderungen der Wissenschaftskommunikation zu führen, sondern viel mehr die grundsätzlichen urheberrechtlichen Fragen dahinter. Wie oben bereits angedeutet, ist daher Kelly (2006) auch in seiner Einschätzung zu widersprechen, Google habe den „Gordian knot of copydom“ geöffnet. Vielmehr scheint hier ein ganzes Netz (inklusive Fallstricke) durch Google und seine KlägerInnen gespannt worden zu sein. Wissenschaftliche AutorInnen dürften dabei in der Regel eher ein Interesse an der freien Zugänglichkeit ihrer Werke (via Google Books) haben. Schließlich erhöht dies die Chance, rezipiert zu werden und die eigene Reputation zu stärken – während etwaige Tantiemen meist ohnehin sehr niedrig ausfallen. So gesehen könnten sich im Kontext der durch Google Books angestoßenen Urheberrechtsfragen auch Konflikte zwischen wissenschaftlichen AutorInnen und Verlagen entwickeln.

---

<sup>52</sup> [asmp.org/articles/press-release-10-04-07.html](http://asmp.org/articles/press-release-10-04-07.html) (zuletzt aufgerufen am 23.04.2010).

<sup>53</sup> [books.google.com/googlebooks/success.html](http://books.google.com/googlebooks/success.html) (zuletzt aufgerufen am 04.03.2010).

<sup>54</sup> [thepublicindex.org/docs/amended\\_settlement/Allen\\_declaration2.pdf](http://thepublicindex.org/docs/amended_settlement/Allen_declaration2.pdf) (zuletzt aufgerufen am 20.04.2010).

## 5 Vorläufige Bewertung und Ausblick

Wie bereits für die bisherigen Steckbriefe (König/Nentwich 2008; König/Nentwich 2009; Herwig et al. 2009) aus dieser Serie, gilt auch hier, dass es sich um vorläufige Betrachtungen eines „moving target“ handelt. So unterliegen sowohl die beschriebenen technischen Einzelheiten als auch die skizzierten sozialen und ökonomischen Aspekte einem fortlaufenden Wandel. Besonders die Funktionsweise der verschiedenen Suchdienste ist äußerst dynamisch, da sich zum einen die durchsuchten Daten ändern (d. h. in erster Linie, dass immer mehr Daten hinzukommen, teilweise jedoch auch aus dem erfassbaren Bereich entfernt werden etc.), zum anderen die ordnenden Algorithmen laufend bearbeitet werden. So kann es z. B. gut sein, dass dieselbe Suchabfrage einen Tag später oder an einem anderen Rechner durchgeführt, zu stark abweichenden Ergebnissen führt. Auch politische und juristische Einflussfaktoren wirken auf die Google-Dienste, was insbesondere bei Google Books beobachtet werden konnte. Neben urheberrechtlichen Fragen stehen hier auch vor allem datenschutzrechtliche Einwände im Fokus, die in diesem Bericht nur am Rande thematisiert wurden. Es ist durchaus denkbar, dass derartige Konflikte massive Änderungen bei den Suchdiensten nach sich ziehen. Schließlich kann sich auch das Nutzungsverhalten ändern, was ebenfalls Auswirkungen auf die Dienste hat.

Dennoch zeichnen sich bei aller Dynamik auch einige Konstanten ab. Es kann kaum daran gezweifelt werden, dass Google mehr als ein kurzzeitiger „Internethype“ ist. Dafür spricht nicht alleine die dominante Positionierung am Suchmaschinen- und Internetwerbemarkt. Kein anderer Anbieter verfügt auch nur annähernd über eine solche breite Datenbasis, die derart effektiv verknüpft, organisiert und verfügbar gemacht wird, wie es bei Google der Fall ist. Dahinter verbirgt sich eine enorm kostenintensive Personal- und Infrastruktur, die alleine schon die mögliche Konkurrenz drastisch einschränkt. Jedoch ist diese Stärke des Konzerns auch das, was KritikerInnen auf den Plan ruft. Zwar sind die Einwände von DatenschützerInnen usw. kaum überhörbar, bislang konnten sie jedoch die Popularität der Dienste unter ihren AnwenderInnen nicht grundsätzlich ins Wanken bringen. Das dürfte zum einen an der schierer Attraktivität der Dienste liegen, zum anderen aber auch an der weitläufigen Unkenntnis über ihre Funktionsweise. Diese und andere Variablen können sich zwar mit der Zeit ändern und so die dominante Stellung des Konzerns verändern, seine stabile und breite Positionierung der letzten Jahre deutet jedoch darauf hin, dass dies nicht kurzfristig geschehen wird.

So wird Google in absehbarer Zeit auch weiterhin auf die Wissenschaft einwirken. Dies ergibt sich aus der Wirkungsmacht des Konzerns in Kombination mit einer funktionalen Überschneidung. Zwar weichen beide in ihrer Zielsetzung zweifellos gravierend voneinander ab (Google strebt als Wirtschaftsunternehmen primär nach Profit, Wissenschaft nach wahrheitsfähigen Aussagen), beide müssen aber ähnliche Aufgaben lösen, die für sie zentrale Bedeutung haben: die Sammlung, Organisation und Verfügbarmachung von Wissen bzw. Informationen. Tatsächlich weist Google selbst akademische Wurzeln auf, was sich teilweise auch in der Funktionsweise der Dienste widerspiegelt. So gründen die Ranking-Methoden im Wesentlichen auf der Idee, Relevanz anhand von Verweisen zu ermitteln, ähnlich wie es bei der wissenschaftlichen Zitationsanalyse geschieht (Mayer 2009). Vaidhyathan charakterisiert die Beziehung zwischen Google und Universitäten entsprechend nicht nur als nah, sondern als „uncomfortably familiar“ (2009, S. 65).

**Ergebnisse aufgrund von Dynamiken vorläufig**

**jedoch auch Konstanten identifizierbar**

**Monopolstellung von Google**

**scheint in absehbarer Zeit stabil zu bleiben**

**entsprechend weiterhin Wirkung auf Wissenschaft**

**durch teilweise überschneidende Aufgabenbereiche**

<b>wurde mit Blick auf die untersuchten Plattformen bestätigt</b>	Dieser Eindruck konnte in diesem Bericht mit Blick auf die Universalsuchmaschine Google, die akademische Suchmaschine Google Scholar und die Büchersuche Google Books bestätigt werden. Alle drei Dienste berühren zunächst insbesondere die Ebene der Suche nach wissenschaftlichen Informationen, für die sie sich einsetzen lassen und eingesetzt werden. Sie bieten dabei neue Wege, die sich vielfach von traditionellen Formen der wissenschaftlichen Informationssuche fundamental unterscheiden. Zum einen ist ihre Datenbasis eine andere, zum anderen folgt die Ergebnisdarstellung nach anderen Prinzipien. Insbesondere BibliothekarInnen kritisieren nicht nur die Qualität der Daten und ihren weitestgehend unbekanntem Umfang, sondern auch die zum Großteil intransparente und fragwürdige Relevanzherstellung durch Rankings. BefürworterInnen betonen dagegen die Kostenfreiheit und die zusätzlichen neuen Möglichkeiten, die durch die Dienste gegeben werden. In jedem Fall ist unstrittig, dass die Dienste bereits intensiv genutzt werden und so auf das Wissenschaftssystem einwirken. Man kann die neuen Wege zu wissenschaftlichen Informationen etwa unter der Annahme demokratisierender Effekte loben oder als Angriff auf wissenschaftliche Standards werten und im Gegenzug die Ranking-basierten Machtstrukturen als „Googlearchy“ (Hindman et al. 2003) kritisieren. Es werden sich empirische Beispiele für jede dieser Thesen finden lassen.
<b>insbesondere relevant für Suche nach wissenschaftlichen Informationen</b>	
<b>intensive Nutzung</b>	
<b>Medienkompetenz notwendig</b>	Einige damit in Verbindung stehende Probleme lassen sich auf der Mikroebene angehen. Zweifellos erfordern die neuen Möglichkeiten und Gefahren, die sich durch die Dienste für die NutzerInnen ergeben, ein besonderes Maß an Medienkompetenz. Entsprechende Kurse können dazu beitragen eben diese zu vermitteln, womit dieses Problem prinzipiell lösbar, oder zumindest abzumildern wäre. In diesem Zusammenhang erscheinen auch weitere Nutzungsstudien vonnöten, die Aufschluss über die Suchpraxen von WissenschaftlerInnen und Studierenden geben. Dabei sollte nicht nur beleuchtet werden, wie sich die Google-Dienste auf ihre NutzerInnen auswirken, sondern auch, wie die User diese qua Anwendung (die von Google schließlich analysiert wird) mitgestalten. Diese Perspektive wird in bisherigen Untersuchungen eher vernachlässigt. Ebenso ist zu fragen, inwiefern durch die Dienste epistemische Veränderungen eingeleitet wurden. Führt etwa die gezielte Stichwortsuche in Verbindung mit einer rekontextualisierten Darstellung in den Ergebnislisten zu einer Vernachlässigung von Zusammenhängen? Wie wirkt sich ständige sekundenschnelle Dauerverfügbarkeit von Wissen aus?
<b>teilweise jedoch Handlungsmöglichkeiten nur begrenzt verfügbar</b>	Für viele der beobachteten Einflüsse auf das Wissenschaftssystem gestalten sich die etwaige Handlungs- und Untersuchungsmöglichkeiten jedoch noch schwieriger – sofern sie überhaupt gegeben sind. Das beginnt schon bei den von Google eingesetzten Algorithmen selbst: Sie sind nicht öffentlich einsehbar und können schon gar nicht von Außenstehenden bearbeitet werden; d. h. sie werden qua Nutzung zwangsläufig akzeptiert. In Verbindung mit der Popularität und der Marktmacht der Dienste führt dies dazu, dass Kritik an ihrer Funktionsweise ins Leere führt. So stellt sich in der derzeitigen Situation nicht die Frage <i>ob</i> die Dienste zweckdienlich für die Wissenschaft sind, sondern vielmehr, <i>wie</i> mit ihnen umgegangen werden sollte, da sie kaum effektiv ignoriert werden können. Verzweifelte Maßnahmen wie „Google-Verbote“ an Hochschulen scheinen kaum eine adäquate Antwort auf dieses Phänomen zu sein. Zum einen, weil sie nicht ausschließen können, dass dennoch „heimlich gegoogelt“ wird, zum anderen, weil die Dienste durchaus auch brauchbare Anwendungsbereiche bieten. Außerdem stellt sich die Frage, ob nicht viele Personen im Hochschulbetrieb längst „vergoogelt“ wurden, um mit Rogers (2009) zu sprechen, d. h. eine Art Gewöhnung an die Einfachheit der Google-
<b>Ignorieren von Google kaum möglich</b>	

Suche eingetreten ist, die sich auf andere Suchdienste auswirkt<sup>55</sup>. Gravierender ist allerdings, dass selbst ein konsequenter Ausschluss der Dienste aus dem Hochschulbetrieb nicht ihre außerwissenschaftliche Nutzung verhindern wird<sup>56</sup>. Alleine schon aufgrund der weitläufigen akademischen Abhängigkeit von öffentlichen Geldern und der Dominanz des Konzerns in einem Kernbereich von Hochschulen – der Vermittlung von (wissenschaftlichen) Informationen – führt kaum ein Weg an Google vorbei. Zudem sind viele wissenschaftliche Verlage und einige Bibliotheken durch die Dienste Google Scholar und Google Books bereits vertraglich an das Unternehmen gebunden.

Das Beispiel Google Books verdeutlicht die tiefe und konfliktreiche Verwobenheit des Wissenschaftssystems mit dem Konzern und lenkt den Blick auf fundamentale wissenschaftspolitische Aspekte. Die Gretchenfrage lautet hier: *Sollten die Dienstleistungen von Google angenommen werden, um so Hochschulkassen von kostenintensiven Aufgaben (wie dem Digitalisieren von Büchern) zu entlasten, oder sollten sie zugunsten wissenschaftlicher Autonomie abgelehnt werden?*

Die intensiven Debatten im Kontext von Google Books zeigen, wie zentral diese Frage ist, selbst wenn sie nicht in dieser „Entweder-Oder-Dichotomie“ beantwortet wird. Sie berührt auch urheberrechtliche und kulturpolitische Kontroversen, die noch längst nicht beigelegt wurden. Sie beschränkt sich auch nicht auf diesen Dienst und betrifft neben den hier beleuchteten Plattformen auch weitere Google-Anwendungen, die sich für wissenschaftliche Zwecke nutzen lassen (Vaidhyanathan 2009). Das Prinzip dahinter ein ähnliches: Google bietet kostenfrei seine umfangreichen Rechen- und Speicherkapazitäten an und erscheint daher auf den ersten Blick äußerst attraktiv. Auf den zweiten Blick ergeben sich jedoch Abhängigkeiten, deren Langzeitwirkungen kaum abschätzbar sind.

Ein simples „Google-Bashing“ wie es in den Debatten immer wieder vernommen werden kann, scheint zur Beantwortung der sich hier aufdrängenden Fragen wenig hilfreich. Andererseits reicht es wohl kaum aus, sich auf die Versprechungen des inoffiziellen Unternehmenscredos „Don’t be evil“ zu verlassen. Selbst wenn man dem guten Willen von Google glauben möchte, kann angesichts der Komplexität der vom Konzern übernommenen Aufgaben (die Sammlung, Organisation und allgemeine Verfügbarmachung des Weltwissens) unmöglich ausgeschlossen werden, dass dabei ungewünschte Nebeneffekte auftreten<sup>57</sup>. Man kann Google kaum anlasten, wenn WissenschaftlerInnen und Studierende leichtfertig mit ihren Suchmaschinen umgehen und Vorwürfe gegen Funktionsmängel können durch den Verweis auf ihren unfertigen Beta-Status abgewehrt werden. Die entsprechenden Effekte treten jedoch unabhängig davon ein, wer nun die Verantwortung trägt und ob ein Projekt offiziell als fertig gilt.

**vertragliche Bindungen**

**wissenschaftspolitische Gretchenfrage**

**Attraktivität durch Kostenfreiheit**

**aber auch Abhängigkeiten**

**weder „Google-Bashing“**

**noch Gutgläubigkeit hilfreich**

**zur Klärung der komplexen Probleme**

<sup>55</sup> Auch Vaidhyanathan beschreibt eine solche „Googlization of students“: „Higher-quality search resources and collections are unlikely to attract students – and will frustrate students who stumble upon them – unless they replicate the reductive simplicity and cleanliness of Google’s interface“ (2009, S. 67). Dies deckt sich auch mit den Beobachtungen von Nygren et al. (2006; vgl. 3.4.1).

<sup>56</sup> Ein ähnlicher Effekt konnte bereits bei der Betrachtung des Verhältnisses von Wissenschaft und Wikipedia betrachtet werden (König/Nentwich 2009).

<sup>57</sup> Ein Beispiel dafür wurde im Mai 2010 bekannt: Während für *Google Streetview* Straßen abgefahren und fotografiert wurden, speicherte man „versehentlich“ über drei Jahre auch umfangreiche Daten aus Funknetzwerken. Ursache war laut Google ein Programmcode aus einem anderen Projekt, der hier unbeabsichtigt eingesetzt worden sei (Lischka 2010).

**globale Ausmaße  
erschwert das  
Abschätzen von Effekten**

Zur Komplexität tragen auch die globalen Ausmaße bei, die sich durch die Online-Verfügbarkeit der Angebote schnell ergeben. Sie erschweren eine Abschätzung der Konsequenzen für alle beteiligten Akteure (d. h. sowohl für Google selbst als auch für seine NutzerInnen, Webmaster, politische EntscheidungsträgerInnen etc.). Gleichzeitig beschränken sich dadurch Handlungsmöglichkeiten auf allen Seiten. Einerseits muss der global agierende Konzern z. B. in vielen Fällen auf nationale Gesetzeslagen Rücksicht nehmen, andererseits stoßen viele national wirksame Gesetze durch die Online-Verfügbarkeit von Google-Diensten buchstäblich an ihre Grenzen. Das gilt auch und gerade für den Wissenschaftsbetrieb, der ebenso global agiert.

**gilt auch für  
entsprechende Studien**

Schließlich sind selbstverständlich ebenso Studien wie die vorliegende von den nur schwer abschätzbaren Folgen der Suchdienste betroffen und müssen entsprechend vorläufig und eingeschränkt vorgenommen werden, wie eingangs schon angesprochen wurde. Angesichts der drängenden Fragen, die sich nicht nur aus wissenschaftlicher Sicht im Zusammenhang mit den Suchdiensten stellen, erscheinen indessen umfangreiche Untersuchungen gleichzeitig dringend nötig. Neben Studien, die auf Einzelaspekte fokussieren (z. B. juristische Analysen, Betrachtungen des Nutzungsverhaltens oder der zugrunde liegenden Daten und ihrer Funktionsweise) sind interdisziplinäre Ansätze der Technikfolgenabschätzung hierzu besonders gut geeignet, da sie der komplexen Wirkungsweise des Gegenstandes am ehesten gerecht werden. Einzelne Disziplinen sind mit der Aufgabe, die sich abzeichnenden Probleme zu identifizieren und ggf. Handlungsoptionen zu erarbeiten, ebenso überfordert, wie die verschiedenen beteiligten Akteure. Der Blick sollte jedoch gerade auf längerfristige Effekte geworfen werden, da evident erscheint, dass durch Google nachhaltige Veränderungen (sowohl im Wissenschaftssystem als auch auf gesamtgesellschaftlicher Ebene) eingeleitet wurden.<sup>58</sup>

**Ansätze der  
Technikfolgen-  
abschätzung besonders  
gut geeignet**

---

<sup>58</sup> So resümiert auch Vaidhyanathan: „The long-term costs and benefits should dominate the conversation. We should not jump at the promises of quick relief or returns” (2009, S. 72).

# Literatur

- Baksik, C., 2006, Fair use or exploitation? The google book search controversy, *Portal-Libraries and the Academy* 6(4), 399-415.
- Becker, K., 2009, Die Macht der Klassifizierung. Abgründe des Wissens an den Klippen der Ordnung, in: Becker, K. und Stalder, F. (Hg.): *Deep Search. Politik des Suchens jenseits von Google*, Innsbruck: Studienverlag, 182-192.
- Becker, N. J., 2003, Google in perspective: understanding and enhancing student search skills, *New Review of Academic Librarianship* 9(1), 84-100.
- Beel, J., Gipp, B. und Wilde, E., 2010, Academic Search Engine Optimization (ASEO): Optimizing Scholarly Literature for Google Scholar and Co., *Journal of Scholarly Publishing* 42(2), 176-190.
- Bergman, M. K., 2001, The Deep Web: Surfacing Hidden Value, *Journal of Electronic Publishing* 7(1 [August]) <doi: 10.3998/3336451.0007.104>.
- Berners-Lee, T., 1998, *Semantic Web Road map*; Letzte Aktualisierung: 14.10.98 <<http://www.w3.org/DesignIssues/Semantic.html>>.
- Brabazon, T., 2006, The Google Effect: Googling, blogging, wikis and the flattening of expertise, *Libri* 56(3), 157-167.
- Bredow, R. v., Dworschak, M., Müller, M. U. und Rosenbach, M., 2010, Das Ende der Privatheit, *Der Spiegel*, 2, 58-69.
- Brin, S. und Page, L., 1998, The Anatomy of a Large-Scale Hypertextual Web Search Engine, *Seventh International World-Wide Web Conference (WWW 1998)*, 14.-18.04.1998, Brisbane <<http://ilpubs.stanford.edu:8090/361/1/1998-8.pdf>>.
- Broder, A., 2002, A taxonomy of web search, *SIGIR Forum* 36(2).
- Brooks, T. A., 2004, The nature of meaning in the age of Google, *Information Research* 9(3) <<http://informationr.net/ir/9-3/paper180.html>>.
- Brophy, J. und Bawden, D., 2005, Is Google enough? Comparison of an internet search engine with academic library resources, *Aslib Proceedings* 57(6), 498-512.
- Calishain, T. und Dornfest, R., 2003, *Google Hacks: 100 Industrial-Strength Tips and Tricks*, Sebastopol: O'Reilly.
- Čas, J., 2005, Privacy in einer Zukunft mit allgegenwärtigen Informationstechnologien – Ein Widerspruch in sich?, in: Nentwich, M. und Peissl, W. (Hg.): *Technikfolgenabschätzung in der österreichischen Praxis. Festschrift für Gunther Tichy*, Wien: Verlag der Österreichischen Akademie der Wissenschaften, 91-112.
- Courant, P. N., 2006, Scholarship and Academic Libraries (and their kin) in the World of Google, *First Monday* 11(8) <<http://firstmonday.org/htbin/cgiwrap/bin/ojs/index.php/fm/article/view/1382/1300>>.
- Dames, K. M., 2006, Library Organizations Should Support Google Book Search, *redorbit.com*, 02.03.06 <[http://www.redorbit.com/news/technology/412990/library\\_organizations\\_should\\_support\\_google\\_book\\_search/](http://www.redorbit.com/news/technology/412990/library_organizations_should_support_google_book_search/)>.

- Degele, N., 2007, Neue Kompetenzen im Internet. Kommunikation abwehren, Information vermeiden, in: Lehmann, K. und Schetsche, M. (Hg.): *Die Google-Gesellschaft. Vom digitalen Wandel des Wissens*, Bielefeld: Transcript Verlag, 63-74.
- Duguid, P., 2007, Inheritance and loss? A brief survey of Google Books, *First Monday* 12(8) <<http://firstmonday.org/htbin/cgiwrap/bin/ojs/index.php/fm/article/viewArticle/1972/1847>>.
- Elmer, G., 2004, *Profiling Machines*, Cambridge, MA: MIT Press.
- Falagas, M. E., Pitsouni, E. I., Malietzis, G. A. und Pappas, G., 2008, Comparison of PubMed, Scopus, Web of Science, and Google Scholar: strengths and weaknesses, *Faseb Journal* 22(2), 338-342.
- Fallows, D., 2005, *Search Engine Users. Internet searchers are confident, satisfied and trusting – but they are also unaware and naïve*, Washington: Pew Internet & American Life Project <[http://www.pewinternet.org/~media/Files/Reports/2005/PIP\\_Searchengine\\_users.pdf](http://www.pewinternet.org/~media/Files/Reports/2005/PIP_Searchengine_users.pdf)>.
- Fister, B., 2009, The Last Library Is Greater than Google. Peer to Peer Review, *Library Journal*, 03.09.09 <<http://www.libraryjournal.com/article/CA6687442.html>>.
- Gloning, T., 2009, Digitale Textcorpora und Sprachforschung. Ältere Koch- und Kräuterbücher, in: Hofmeister, W. und Hofmeister-Winter, A. (Hg.): *Wege zum Text. Überlegungen zur Verfügbarkeit mediävistischer Editionen im 21. Jahrhundert. Grazer Kolloquium 17.–19. September 2008*, Tübingen: Max Niemeyer Verlag, 53-71.
- Google, 2008, Google's Search Engine Optimization Starter Guide, Version 1.1; 13.11.08 <<http://www.google.com/webmasters/docs/search-engine-optimization-starter-guide.pdf>>.
- Google, 2009, *Unsere Philosophie*; [Aufgerufen am: 14.07. 2009] <<http://www.google.at/intl/de/corporate/tenthings.html>>.
- Google Books, 2010a, *Die Geschichte der Google Buchsuche*; [Aufgerufen am: 20.04. 2010] <<http://books.google.at/intl/de/googlebooks/history.html>>.
- Google Books, 2010b, *Google Buchsuche-Bibliotheksprogramm – ein erweiterter Katalog mit den Büchern dieser Welt*; [Aufgerufen am: 20.04. 2010] <<http://books.google.com/googlebooks/library.html>>.
- Google Books, 2010c, *Häufig gestellte Fragen*; [Aufgerufen am: 08.06. 2010] <<http://books.google.com/intl/de/googlebooks/help.html>>.
- Google Books, 2010d, *Wie lauten die Programmrichtlinien?*; [Aufgerufen am: 24.04. 2010] <<http://books.google.com/support/partner/bin/answer.py?answer=18625&topic=322&hl=de&gsessionid=H0FdlaVQoctEDD2gOh2T8w>>.
- Google Scholar, 2010a, *Hilfe*; [Aufgerufen am: 03.04. 2010] <<http://scholar.google.de/intl/de/scholar/help.html>>.
- Google Scholar, 2010b, *Support für wissenschaftliche Verlage*; [Aufgerufen am: 12.03. 2010] <<http://www.google.com/support/scholar/bin/request.py?hl=de>>.
- Google Scholar, 2010c, *Über Google Scholar*; [Aufgerufen am: 02.02. 2010] <<http://scholar.google.de/intl/de/scholar/about.html>>.



- Griffiths, J. R. und Brophy, P., 2005, Student Searching Behavior and the Web: Use of Academic Resources and Google, *Library Trends* 53(4), 539-554.
- Grossman, L., 2006, Google Under the Gun, *TIME.com*, 05.02.06  
<<http://www.time.com/time/magazine/article/0,9171,1156598,00.html>>.
- Haglund, L. und Olsson, P., 2008, The Impact on University Libraries of Changes in Information Behavior Among Academic Researchers: A Multiple Case Study, *The Journal of Academic Librarianship* 34(1), 52-59.
- Halavais, A., 2009, *Search Engine Society*; in Reihe: Digital Media and Society Series, Cambridge: Polity Press.
- Harzing, A.-W. K. und Wal, R. v. d., 2008, Google Scholar as a new source for citation analysis, *Ethics in Science and Environmental Politics* 8(1), 61-73 <doi: 10.3354/ese00076>.
- Head, A. J., 2007, Beyond Google: How do students conduct academic research?, *First Monday* 12(8)  
<[http://firstmonday.org/issues/issue12\\_8/head/index.html](http://firstmonday.org/issues/issue12_8/head/index.html)>.
- Herwig, J., Kittenberger, A., Nentwich, M. und Schirmund, J., 2009, *Microblogging und die Wissenschaft. Das Beispiel Twitter. Steckbrief IV im Rahmen des Projekts Interactive Science*, Wien: Institut für Technikfolgen-Abschätzung  
<<http://epub.oew.ac.at/ita/ita-projektberichte/d2-2a52-4.pdf>>.
- Hindman, M., Tsioutsoulklis, K. und Johnson, J. A., 2003, Googlearchy: How a few heavily-linked sites dominate politics on the web, *Annual Meeting of the Midwest Political Science Association*, 31.03.03, Chicago <doi:10.1.1.7.7520>.
- Jacsó, P., 2005, Google Scholar: the pros and the cons, *Online Information Review* 29(2), 208-214.
- Jacsó, P., 2008, Google Scholar revisited, *Online Information Review* 32(1), 102-114 <doi: 10.1108/14684520810866010>.
- Jacsó, P., 2009, Google Scholar's Ghost Authors, Lost Authors, and Other Problems. Why the popular tool can't be used to analyze the publishing performance and impact of researchers, *Library Journal*, 24.09.2009  
<<http://www.libraryjournal.com/article/CA6698580.html?q=jacso>>.
- Jeanneney, J.-N., 2006, *Google and the Myth of Universal Knowledge: A View from Europe*, Chicago: University of Chicago Press.
- Kamadjeu, R., 2009, Tracking the polio virus down the Congo River: a case study on the use of Google Earth™ in public health planning and mapping, *International Journal of Health Geographics* 8(4)  
<<http://www.ij-healthgeographics.com/content/8/1/4>>.
- Kaumanns, R. und Siegenheim, V., 2008, Von der Suchmaschine zum Werbekonzern, *Media Perspektiven* (1), 25-33  
<[http://www.media-perspektiven.de/uploads/tx\\_mppublications/01-2008\\_Kaumanns.pdf](http://www.media-perspektiven.de/uploads/tx_mppublications/01-2008_Kaumanns.pdf)>.
- Kelly, K., 2006, Scan This Book!, *nytimes.com*, 14.05.06  
<[http://www.nytimes.com/2006/05/14/magazine/14publishing.html?\\_r=2&oref=slogin&pagewanted=all](http://www.nytimes.com/2006/05/14/magazine/14publishing.html?_r=2&oref=slogin&pagewanted=all)>.

- König, R., 2009, *Eine Bewegung für die Wahrheit? Gesellschaftliche Wirklichkeitskonstruktion in Wikipedia am Beispiel alternativer Deutungen des 11. September 2001*, Diplomarbeit, Fakultät für Soziologie, Universität Bielefeld.
- König, R. und Nentwich, M., 2008, *Wissenschaft in "Second Life". Steckbrief I im Rahmen des Projekts Interactive Science*, Wien: Institut für Technikfolgen-Abschätzung  
<<http://epub.oeaw.ac.at/ita/ita-projektberichte/d2-2a52-1.pdf>>.
- König, R. und Nentwich, M., 2009, *Wissenschaft in Wikipedia und anderen Wikimedia-Projekten. Steckbrief II im Rahmen des Projekts Interactive Science*, Wien: Institut für Technikfolgen-Abschätzung  
<<http://epub.oeaw.ac.at/ita/ita-projektberichte/d2-2a52-2.pdf>>.
- Kremp, M., 2009, China blockiert Google, *Spiegel Online*, 25.06.09  
<<http://www.spiegel.de/netzwelt/web/0,1518,632473,00.html>>.
- Lehdonvirta, V., 2008, Virtual worlds don't exist, *Breaking the Magic Circle*, Tampere <<http://virtual-economy.org/files/Lehdonvirta-VWDE.pdf>>.
- Lehmann, K. und Schetsche, M. (Hg.), 2007, *Die Google-Gesellschaft. Vom digitalen Wandel des Wissens*, Bielefeld: Transcript Verlag.
- Levy, S., 2010, How Google's Algorithm Rules the Web, *wired.com*, 22.02.10  
<[http://www.wired.com/magazine/2010/02/ff\\_google\\_algorithm](http://www.wired.com/magazine/2010/02/ff_google_algorithm)>.
- Lewandowski, D., 2005, *Web Information Retrieval. Technologien zur Informationssuche im Internet* in Reihe: Informationswissenschaft der DGI, hg. v. Ockenfeld, M., Frankfurt a.M.: Deutsche Gesellschaft für Informationswissenschaft und Informationspraxis e.V.  
<<http://www.durchdenken.de/lewandowski/web-ir/>>.
- Lewandowski, D., 2007, Nachweis deutschsprachiger bibliotheks- und informationswissenschaftlicher Aufsätze in Google Scholar, *Information Wissenschaft und Praxis* 58(3), 165-168.
- Lewandowski, D., 2008, Search engine user behaviour: How can users be guided to quality content?, *Information Services & Use* 28, 261-268  
<doi: 10.3233/ISU-2008-0583>.
- Lewandowski, D. und Mayr, P., 2006, Exploring the academic invisible web, *Library Hi Tech* 24(4), 529-539 (zitiert nach Manuskript)  
<<http://eprints.rclis.org/9156/>>.
- Lischka, K., 2010, Fehler im perfekten System, *Spiegel Online*, 15.05.10  
<<http://www.spiegel.de/netzwelt/netzpolitik/0,1518,694958,00.html>>.
- Maaß, C., Skusa, A., Heß, A. und Pietsch, G., 2008, Der Markt für Internet-Suchmaschinen, in: Lewandowski, D. (Hg.): *Handbuch Internet-Suchmaschinen. Nutzerorientierung in Wissenschaft und Praxis*, Heidelberg: Akademischen Verlagsgesellschaft AKA, 3-17  
<<http://eprints.rclis.org/15514/>>.
- Machill, M., Beiler, M. und Zenker, M., 2007, Suchmaschinenforschung. Überblick und Systematisierung eines interdisziplinären Forschungsfeldes, in: Machill, M. und Beiler, M. (Hg.): *Die Macht der Suchmaschinen – The Power of Search Engines*, Köln: Halem, 07-43.
- Machill, M., Neuberger, C., Schweiger, W. und Wirth, W., 2003, Wegweiser im Netz: Qualität und Nutzung von Suchmaschinen, in: Machill, M. und Welp, C. (Hg.): *Wegweiser im Netz. Qualität und Nutzung von Suchmaschinen*, Gütersloh: Verlag Bertelsmann Stiftung, 13-490.

- Mager, A., 2009, Mediated health: sociotechnical practices of providing and using online health information, *New Media & Society* 11(7), 1123-1142.
- Mager, A., 2010, *Mediated Knowledge: Sociotechnical practices of communicating medical knowledge via the web and their epistemic implications*, Dissertation, Universität Wien.
- Mayer, K., 2009, Zur Soziometrik der Suchmaschinen. Ein historischer Überblick der Methodik, in: Becker, K. und Stalder, F. (Hg.): *Deep Search. Politik des Suchens jenseits von Google*, Innsbruck: Studienverlag, 64-83.
- Mayr, P., 2009, Google Scholar als akademische Suchmaschine, *VÖB-Mitteilungen* 62(2), 18-28.
- Mayr, P. und Walter, A.-K., 2006, *Abdeckung und Aktualität des Suchdienstes Google Scholar*, Frankfurt: DGI <<http://eprints.rclis.org/6208/>>.
- Mayr, P. und Walter, A. K., 2007, An exploratory study of Google Scholar, *Online Information Review* 31(6), 814-830.
- Merton, R. K., 1985, *Entwicklung und Wandel von Forschungsinteressen*, Frankfurt am Main: Suhrkamp.
- Moulton, R. und Carattini, K., 2007, *A quick word about Googlebombs*; Letzte Aktualisierung: 25.01.07 <<http://googlewebmastercentral.blogspot.com/2007/01/quick-word-about-googlebombs.html>>.
- Mullen, L. B. und Hartman, K. A., 2006, Google scholar and the library web site: The early response by ARL libraries, *College & Research Libraries* 67(2), 106-122.
- Nentwich, M., 2003, *Cyberscience: Research in the Age of the Internet*, Wien: Austrian Academy of Sciences Press.
- Nentwich, M., 2009, *Cyberscience 2.0 oder 1.2? Das Web 2.0 und die Wissenschaft, ITA manu:script*, Nr. ITA-09-02 hrsg. v. Institut für Technikfolgen-Abschätzung, Wien <[http://epub.oeaw.ac.at/ita/ita-manuscript/ita\\_09\\_02.pdf](http://epub.oeaw.ac.at/ita/ita-manuscript/ita_09_02.pdf)>.
- Neuhaus, C., Neuhaus, E., Asher, A. und Wrede, C., 2006, The depth and breadth of Google Scholar: An empirical study, *Portal – Libraries and the Academy* 6(2), 127-141.
- Norris, M., 2008, *The citation advantage of open access articles*, Dissertation, Loughborough University <<http://hdl.handle.net/2134/4089>>.
- Norris, M., Oppenheim, C. und Rowland, F., 2008a, Finding open access articles using Google, Google Scholar, OAIster and OpenDOAR, *Online Information Review* 32(6), 709-715.
- Norris, M., Oppenheim, C. und Rowland, F., 2008b, Open access citation rates and developing countries, *ELPUB2008. Open Scholarship: Authority, Community, and Sustainability in the Age of Web 2.0 – 12<sup>th</sup> International Conference on Electronic Publishing*, Toronto <[http://elpub.scix.net/cgi-bin/works/Show?335\\_elpub2008](http://elpub.scix.net/cgi-bin/works/Show?335_elpub2008)>.
- Noruzi, A., 2005, Google Scholar: The new generation of citation indexes, *Libri* 55(4), 170-180.
- Nunberg, G., 2009, Google's Book Search: A Disaster for Scholars, *chronicle.com*, 31.03.09 <<http://chronicle.com/article/Googles-Book-Search-A/48245/>>.

- Nygren, E., Haya, G. und Widmark, W., 2006, *Students experience of Metalib and Google Scholar*, im Auftrag von: BISBAM, Nr. Dnr 63-612-2005.
- OCLC, 2005, *Perceptions of Libraries and Information Resources. A Report to the OCLC Membership*, Dublin (Ohio): OCLC Online Computer Library Center <[http://www.oclc.org/reports/pdfs/Percept\\_all.pdf](http://www.oclc.org/reports/pdfs/Percept_all.pdf)>.
- Oder, N., 2009, Google, „The Last Library,“ and Millions of Metadata Mistakes, *Library Journal*, 03.09.09 <<http://www.libraryjournal.com/article/CA6687562.html?nid=2673&source=title&rid=1407498533>>.
- Page, L., Brin, S., Motwani, R. und Winograd, T., 1998, *The PageRank Citation Ranking: Bringing Order to the Web*: Stanford InfoLab <<http://ilpubs.stanford.edu:8090/422/>>.
- Pomerantz, J., 2006, Google Scholar and 100 percent availability of information, *Information Technology and Libraries* 25(2), 52-56.
- Putnam, R. D., 2000, *Bowling Alone. The Collapse and Revival of American Community*, New York: Simon & Schuster.
- Reischl, G., 2008, *Die Google Falle*, Wien: Ueberreuter.
- Rieger, O. Y., 2009, Search engine use behavior of students and faculty: User perceptions and implications for future research, *First Monday* 14(12) <<http://firstmonday.org/htbin/cgiwrap/bin/ojs/index.php/fm/article/view/2716/2385>>.
- Robinson, M. L. und Wusteman, J., 2007, Putting Google Scholar to the test: a preliminary study, *Program-Electronic Library and Information Systems* 41(1), 71-80.
- Rogers, R., 2009, Zur Frage der Vergoogelung. Hin zu einer unkritisierbaren Maschine?, in: Becker, K. und Stalder, F. (Hg.): *Deep Search. Politik des Suchens jenseits von Google* Innsbruck: Studienverlag, 193-206.
- Schetsche, M., 2006, Die digitale Wissensrevolution – Netzwerkmedien, kultureller Wandel und die neue soziale Wirklichkeit, *zeitenblicke* 5(3) <<http://www.zeitenblicke.de/2006/3/Schetsche>>.
- Schetsche, M., Lehmann, K. und Krug, T., 2007, Die Google-Gesellschaft. Zehn Prinzipien der neuen Wissensordnung, in: Lehmann, K. und Schetsche, M. (Hg.): *Die Google-Gesellschaft. Vom digitalen Wandel des Wissens*, Bielefeld: Transcript Verlag.
- Schießl, M., 2007, Die Anzeigen-Maschine, *Der Spiegel*, Nr. 5, 58-60.
- Schröder, B., 2002, Google filtert, *Telepolis*, 22.07.02 <<http://www.heise.de/tp/r4/artikel/12/12948/1.html>>.
- Sherman, C. und Price, G., 2003, The Invisible Web: Uncovering Sources Search Engines Can't See, *Library Trends* 52(2), 282-298.
- Söllner, K., 2006, Google Scholar und Windows Live Academic Search – aktuelle Entwicklungen bei wissenschaftlichen Suchmaschinen, *Bibliotheksdienst* 40(7), 828-837 <[http://www.zlb.de/aktivaeten/bd\\_neu/heftinhalte2006/Erschliessung0706.pdf](http://www.zlb.de/aktivaeten/bd_neu/heftinhalte2006/Erschliessung0706.pdf)>.
- Spiegel Online, 2006, Google straft BMW ab, 06.02.2006 <<http://www.spiegel.de/netzwelt/web/0,1518,399214,00.html>>.
- Stöber, T., 2006, Der Wandel in der wissenschaftlichen Informationsvermittlung: das Beispiel Google Book Search, *zeitenblicke* 5(3) <[http://www.zeitenblicke.de/tmp/auto1165143180.44/index\\_html](http://www.zeitenblicke.de/tmp/auto1165143180.44/index_html)>.

- Thelwall, M., 2002, Research note: In praise of Google: Finding law journal Web sites, *Online Information Review* 26(4), 271-272.
- Vaidhyanathan, S., 2007, The Googlization of Everything and the Future of Copyright, *University of California Davis Law Review* 40(3), 1207-1232 <[http://lawreview.law.ucdavis.edu/issues/Vol40/Issue3/DavisVol40No3\\_Vaidhyanathan.pdf](http://lawreview.law.ucdavis.edu/issues/Vol40/Issue3/DavisVol40No3_Vaidhyanathan.pdf)>.
- Vaidhyanathan, S., 2009, The Googlization of Universities, in: National Education Association (Hg.): *The NEA 2009 Almanac of Higher Education*, 65-74 <[http://www.nea.org/assets/img/PubAlmanac/ALM\\_09\\_06.pdf](http://www.nea.org/assets/img/PubAlmanac/ALM_09_06.pdf)>.
- Vascellaro, J. E., 2010, Google Tests TV Search Service, *The Wallstreet Journal*, 08.03.2010 <<http://online.wsj.com/article/SB10001424052748704869304575109912574043580.html>>.
- Weingart, P., 2001, *Die Stunde der Wahrheit? Zum Verhältnis der Wissenschaft zu Politik, Wirtschaft und Medien in der Wissensgesellschaft*, 1. Aufl., Weilerswist: Velbrück.
- White, B., 2006, Examining the claims of Google Scholar as a serious information source, *New Zealand Library & Information Management Journal* 50(1), 11-24 <<http://hdl.handle.net/10179/571>>.
- Zimmer, M., 2009, If You Trust Google's Results, You Can Thank... "PigeonRank"?, *michaelzimmer.org*, 07.12.2009 <<http://michaelzimmer.org/2009/12/07/if-you-trust-googles-results-you-can-thank-pigeonrank/>>.