

# Information

## WISSENSCHAFT & PRAXIS

<http://kvk.uni-karlsruhe.de>

Die erste Metasuchmaschine für Bücher  
**Karlsruher Virtueller Katalog KVK**  
Online seit 1996



Über 1,5 Millionen Suchanfragen im Monat.

Mit einem Klick in  
**54 Bibliotheks-  
und Buchhandels-  
katalogen aus  
20 Ländern** suchen!



Zahlreiche virtuelle Regional- und Fachkataloge nutzen die KVK-Technik:

- [artlibraries.net](http://artlibraries.net)
- [Open Access Katalog OASE](http://Open Access Katalog OASE)
- [Schweizer Virtueller Katalog CHVK](http://Schweizer Virtueller Katalog CHVK)
- [Utah's Catalog](http://Utah's Catalog)
- [Virtuelle Deutsche Landesbibliographie](http://Virtuelle Deutsche Landesbibliographie)

Ein Service der  
24-Stunden-Bibliothek



Universität Karlsruhe (TH) | Universitätsbibliothek  
Forschungsuniversität • gegründet 1825

Schwerpunkt

Suchmaschinen

Suchverhalten im Web

Usability von Webportalen  
und Suchmaschinen

Visualisierungskomponenten

Suche in Firmen-Websites

Google Scholar

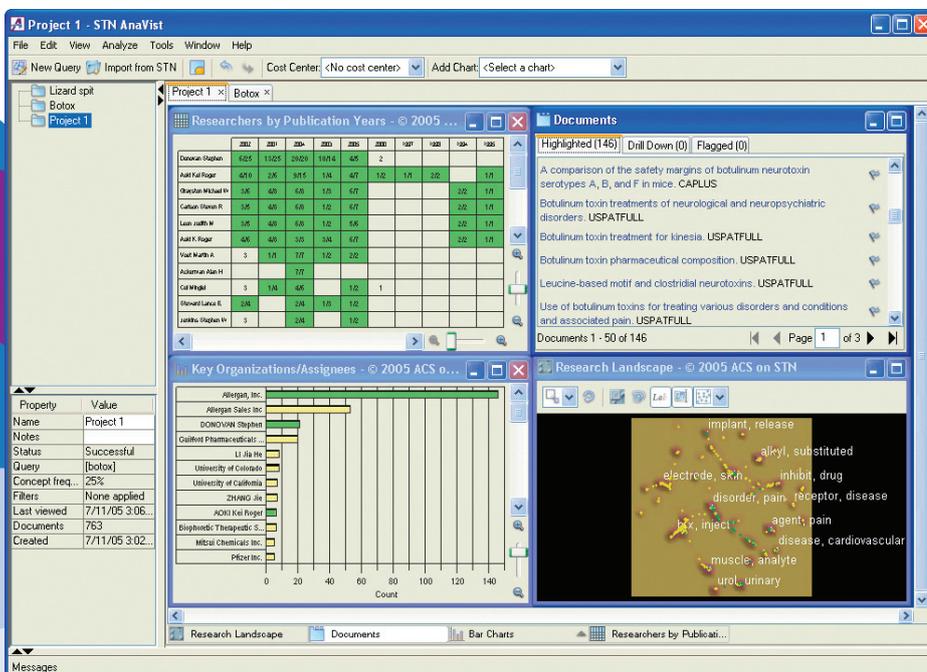
Historische Rezensionen

PsychSpider

Wissenschaftssuchmaschine Base

Open Access für Schwellenländer

# STN<sup>®</sup> ANAVIST<sup>™</sup>



dynamische, interaktive  
Visualisierung von  
Rechercheergebnissen

STN<sup>®</sup> AnaVist<sup>™</sup> – Die interaktive Analyse- und Visualisierungssoftware lässt Trends in Forschung und Patentanmeldung erkennen.

## Nur STN AnaVist bietet

innovative Funktionalitäten, einschließlich der intellektuellen Standardisierung von Feldinhalten aus Patent- und anderen Dokumenten, kombiniert mit der Möglichkeit, die Beziehungen zwischen den Daten in einer interaktiven Arbeitsumgebung darzustellen.

## Mit STN AnaVist können Sie

- die Patent- und Forschungslandschaft analysieren
- Märkte und Wettbewerber beobachten
- neue Anwendungsfelder für existierende Technologien erkennen
- neue Geschäftsfelder identifizieren
- Managementstrategien effektiver planen

## STN AnaVist unterstützt Geschäftsentscheidungen

STN AnaVist bietet neue Möglichkeiten der gezielten Informationsgewinnung zur Unterstützung wichtiger Geschäftsentscheidungen.

FIZ Karlsruhe  
STN Europe  
Telefon: +49 7247 808-555  
E-Mail: [helpdesk@fiz-karlsruhe.de](mailto:helpdesk@fiz-karlsruhe.de)  
[www.stn-international.de](http://www.stn-international.de)

CAS  
STN North America  
Telefon: +800 753-4227  
E-mail: [help@cas.org](mailto:help@cas.org)  
[www.cas.org/supp.html](http://www.cas.org/supp.html)

Japan Association for International  
Chemical Information (JAICI)  
STN Japan  
Telefon: +81 3 5978-3601 (Technical Service)  
+81 3 5978-3621 (Customer Service)  
[www.jaici.or.jp](http://www.jaici.or.jp)

# Editorial

Seit dem ersten Themenheft der IWP über Suchmaschinen sind mittlerweile etwas mehr als zwei Jahre vergangen. Seitdem haben sich nicht nur die Web-Suchmaschinen weiter rasant entwickelt, sondern es wurden auch viele Spezialanwendungen geboren und die den Suchmaschinen zugrunde liegende Technologie wurde auf andere Anwendungsfelder übertragen. So ist es nicht verwunderlich, dass auf den Call for Papers für dieses Themenheft verstärkt Vorschläge für Beiträge über speziellere Fragestellungen sowie Vorschläge für Praxisberichte eingereicht wurden.

Aus der Fülle der Einreichungen musste der Herausgeber eine Auswahl treffen, die sowohl den thematischen Schwerpunkten als auch der Balance aus Aufsätzen aus der Wissenschaft und aus der Praxis gerecht wird. Diese Auswahl war bei der hohen Qualität vieler der vorgeschlagenen Beiträge nicht leicht. Die Qualität wurde im Peer-Review-Verfahren durch jeweils zwei Reviewer gesichert. An dieser Stelle sei allen Gutachtern herzlich für Ihre Mühe gedankt. Sie haben wertvolle Anregungen gegeben, die zu einer weiteren Verbesserung der Beiträge geführt haben.

Ich hoffe, dass die nun vorliegende Zusammenstellung einen interessanten Einblick in das Forschungsfeld Suchmaschinen sowohl aus dem informationswissenschaftlichen als auch aus dem informationspraktischen Blickwinkel bietet.

In der Diskussion um Suchmaschinen kommt oft das eigentlich Naheliegende zu kurz, nämlich das Suchverhalten der Nutzer. Nadine Höchstötter von der Universität Karlsruhe hat sich in dieser Hinsicht verdient gemacht und gibt in dieser Ausgabe einen Überblick über ihre Forschungen der vergangenen Jahre. Dabei kommen nicht nur Fragen wie die Länge und Komplexität der Suchanfragen zur Sprache, sondern auch Trends in Reaktion auf aktuelle Ereignisse.

Ein weiteres Thema, das im Suchkontext oft ein wenig stiefmütterlich behandelt wird, ist die Usability der Suchinterfaces. Franz Aschoff und Eva Rausch von Trautenberg (Hochschule Darmstadt in Kooperation mit T-Online) wiederholen hier eine Untersuchung aus dem Jahr 2005 (erschieden im

ersten Suchmaschinen-Themenheft der IWP) und fügen einige weitere interessante Fragestellungen hinzu. Durch den Vergleich mit der Vorgängeruntersuchung wird es möglich zu erkennen, welche Standards sich bei den Suchinterfaces etablieren und welche weniger erfolgversprechend sind. In der Interaktion zwischen Nutzer und Suchmaschine treten neben den klassischen Interfaces zunehmend auch stärker grafisch orientierte Benutzeroberflächen. Bernard Bekavac und Kollegen (Hochschule Chur) zeigen die unterschiedlichen Visualisierungsmöglichkeiten, die zurzeit bei Suchmaschinen verwendet werden, auf und vergleichen diese.

Dass Suchmaschinentechnologie zunehmend auch für enger begrenzte Bereiche eingesetzt wird, zeigt sich im Bereich der Suche auf einzelnen Websites. Eine solche Suchfunktion ist mittlerweile Standard zumindest bei größeren Websites, wird aber oft von den Sitebetreibern nicht in ihrer Bedeutung für den Nutzer erkannt. Martin Braschler und Kollegen (Hochschule Winterthur, Hochschule Chur in Kooperation mit Unternehmen) untersuchen die Qualität der Site-Suche bei Unternehmen in der Schweiz. Dabei werden die Bereiche Suchindex, Vergleich von Anfragen und Dokumenten, Benutzerinteraktion und Suchergebnisse untersucht.

Neben der Suche im Unternehmenskontext haben in den letzten Jahren auch Spezialsuchmaschinen, die teils von den großen Web-Suchmaschinen angeboten werden, an Bedeutung gewonnen. Die bekannteste davon ist sicher Google Scholar, eine Suchmaschine für wissenschaftliche Literatur. Dirk Lewandowski (Hochschule für Angewandte Wissenschaften Hamburg) geht der Frage nach, wie gut sich diese Suchmaschine für die Recherche nach informationswissenschaftlicher Literatur eignet. Es zeigt sich, dass Google Scholar die deutschsprachige Literatur bei weitem nicht vollständig erfasst. Allerdings kann diese Suchmaschine eine Hilfe beim schnellen Zugang zu Volltexten sein.

Der zweite Block des Themenhefts widmet sich interessanten Projekten aus der Praxis. Daniel Burckhardt von der Humboldt-Universität Berlin stellt das Projekt „Histori-

sche Rezensionen online“ vor. Dabei wird mit Hilfe von Suchmaschinentechnologie ein Index von Besprechungen geschaffen und durchsuchbar gemacht. Der Datenbestand von bereits über 35.000 Rezensionen konnte so ohne manuelle Katalogisierung oder Zusammenstellung aufgebaut werden. Anhand des Projekts wird aufgezeigt, wie mit moderner Technologie trotz knapper Ressourcen ein anspruchsvolles Angebot geschaffen werden kann.

Schon seit einigen Jahren existiert die Psychologie-Suchmaschinen PsychSpider. Erich Weichselgartner und Christiane Baier (Zentrum für Psychologische Information und Dokumentation Trier) zeichnen die Entwicklung nach und berichten über aktuelle Entwicklungen bei dieser Suchmaschine. PsychSpider kann als Referenzprojekt für die Entwicklung eigener, fachspezifischer Suchmaschinen dienen.

Eine nicht auf ein einzelnes Fach beschränkte, sondern von einem Bibliotheksstandpunkt aus umfassende Wissenschaftsuchmaschine wird seit einiger Zeit an der Universitätsbibliothek Bielefeld unter dem Namen BASE entwickelt. Dirk Pieper und Sebastian Wolf zeigen hier die Entwicklung der letzten Jahre und die Hinwendung zu Open-Access-Inhalten aus Repositorien und deren Kombination sowohl mit freien Web-Inhalten als auch mit Inhalten aus dem Bibliothekskatalog.

Die Beiträge des Themenhefts geben also insgesamt einen spannenden Einblick sowohl in die Forschung und daraus abgeleitete Handlungsempfehlungen als auch in konkrete Projekte, die sich Suchmaschinentechnologie bereits erfolgreich zu nutzen machen. Es ist zu wünschen, dass viele weitere Firmen und Institutionen die Potentiale, die sich durch den Einsatz von Suchmaschinentechnologie bieten, nutzen. Die vorliegenden Aufsätze mögen ihnen Anregungen für die Umsetzung geben.



Dirk Lewandowski  
dirk.lewandowski@bui.haw-hamburg.de  
www.durchdenken.de/lewandowski

# Inhalt

3/2007

## SCHWERPUNKT SUCHMASCHINEN

Gastherausgeber: Dirk Lewandowski

### 129 EDITORIAL

von Dirk Lewandowski

### 132 LESERBRIEF

Das „elektronische Zeitalter der Chemieinformation ist eingeläutet worden“ – und sonst nichts? (Gerd Bauer)

## SUCHMASCHINEN

135 Nadine Höchstötter  
Suchverhalten im Web – Erhebung, Analyse und Möglichkeiten

141 Franz Aschoff und Eva Rausch von Traubenberg  
Usability von Webportalen und Web-Suchmaschinen im Vergleich

149 Bernard Bekavac, Josef Herget, Sonja Hierl und Sonja Öttl  
Visualisierungskomponenten bei webbasierten Suchmaschinen: Methoden, Kriterien und ein Marktüberblick

159 Martin Braschler, Josef Herget, Joachim Pfister, Peter Schäuble, Markus Steinbach und Jürg Stuker  
Kommunikation im Internet: Die Suchfunktion im Praxistest

165 Dirk Lewandowski  
Nachweis deutschsprachiger bibliotheks- und informationswissenschaftlicher Aufsätze in Google Scholar

169 Daniel Burckhardt  
Historische Rezensionen online: Eine thematische Suchmaschine von Clio-online

173 Erich Weichselgartner und Christiane Baier  
Sechs Jahre PsychSpider: Aus der Praxis des Betriebs einer Psychologie-Suchmaschine für freie Web-Inhalte

179 Dirk Pieper und Sebastian Wolf  
BASE – Eine Suchmaschine für OAI-Quellen und wissenschaftliche Webseiten

### 132 TAGUNGSBERICHT

Achim Oßwald

Bangalore Commitment: Workshop on Electronic Publishing and Open Access

### 185 PERSONALIE

In Memoriam Dr. Udo Schützsack (Wolfrudolf Laux)

REZENSIONEN 186

LITERATURAUSLESE 188

MEDIAINFO 189

IMPRESSUM 191

TERMINKALENDER 192

## Call for Papers: WissKom2007

### 4. ZB-Konferenz „WissKom2007: Wissenschaftskommunikation der Zukunft“

Die Zentralbibliothek des Forschungszentrums Jülich veranstaltet in der Zeit vom 6. bis 8. November 2007 die Konferenz „WissKom2007: Wissenschaftskommunikation der Zukunft“.

Die Wissenschaftskommunikation hat sich seit der weitgehenden Digitalisierung der Informationsversorgung erheblich verändert. Dies betrifft den wissenschaftlichen Erkenntnisprozess genauso wie die Informationsversorgung durch Bibliotheken. Ansätze wie eScience, Gridcomputing, die Entwicklung des WEB 2.0, aber auch der Einsatz bibliometrischer Wissenschaftsindikatoren und die Open Access Bewegung bedeuten eine strukturelle Neuausrichtung der Wissenschaftskommunikation. All das thematisiert die Jülicher Konferenz WissKom2007- sie ist bereits die vierte Konferenz der Zentralbibliothek zu Themen im Spannungsfeld von Wissenschaft und Information.

Das Programmkomitee lädt Experten aus den oben genannten Bereichen ein, Abstracts zu den zentralen Themen der Konferenz einzureichen:

- eScience
- Wissenschaftsindikatoren
- Web 2.0
- Primärdaten-Management (Open Access)

Abstracts (200 bis 300 Wörter) sind mit der Angabe der Kontaktdaten (Name, Institution, Adresse etc.) bis zum 30. April 2007 an folgende E-Mail Adresse zu richten:

zb-konferenz@fz-juelich.de  
Forschungszentrum Jülich GmbH,  
Zentralbibliothek  
Programmkomitee  
Dr. Rafael Ball  
52425 Jülich  
Telefon: (0 24 61) 61 30 13  
www.wisskom2007.de

Bitte teilen Sie uns mit, ob Ihr Beitrag bevorzugt als Vortrag oder als Poster berücksichtigt werden soll. Über die Annahme der Beiträge entscheidet das Programmkomitee bis zum 25. Mai 2007.

Aktuelle Informationen zur Konferenz: [www.wisskom2007.de](http://www.wisskom2007.de)  
Edith Salz

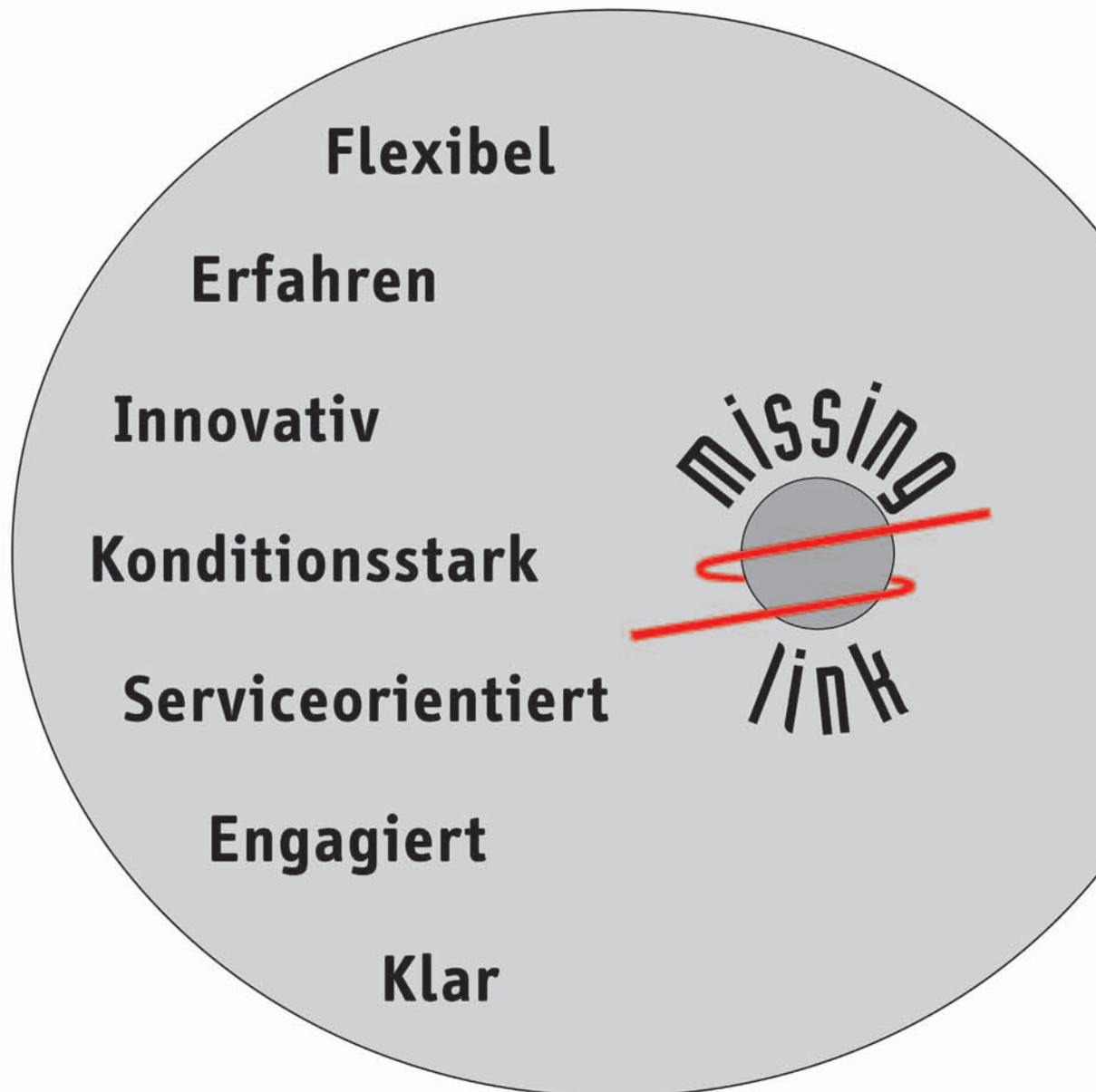
Forschungszentrum Jülich GmbH, Zentralbibliothek

**Missing Link**

Westerstrasse 114-116 | 28199 Bremen

Tel.: [0421] 504348 | Fax: [0421] 504316

# Erwerbungspartner, mit denen Sie rechnen können



[www.missing-link.de](http://www.missing-link.de)

[info@missing-link.de](mailto:info@missing-link.de)

**Internationale**

**Versandbuchhandlung**

# Das „elektronische Zeitalter der Chemieinformation ist eingeläutet worden“ – und sonst nichts?

## Bemerkungen zum Artikel von Gregor Fels in IWP 58(2007) 51-56

Wenn man die historische Entwicklung der Chemieinformation zwischen „gestern und morgen“ liest und vorwiegend nur eine Aufzählung „elektronischer Medien“ findet, ist man erstaunt darüber, dass die großen Leistungen der Chemie in den letzten hundert Jahren ohne Computer überhaupt möglich geworden sind.

Angesichts der Vorliebe des Autors für elektronische Medien und für Recherchen mit Hilfe von http und www fällt der Teil „Chemieinformation gestern“ entsprechend lückenhaft aus. Bedauerlicherweise werden auch wichtige Informationen einfach weggelassen.

Während das amerikanische Referatorgan „Chemical Abstracts Service (CAS)“ vorwiegend im Zusammenhang mit der mühseligen Benutzung von meterlangen Reihen von CAS-Bänden erwähnt wird, erfolgt keine Bemerkung über das deutsche Chemische Zentralblatt (CZ), das über hundert Jahre lang weltweit das einzige Referatblatt der gesamten Chemie war.

CAS und auch das russische „Referativnyj Žurnal Chimija“ wurde nach dem Vorbild des CZ aufgebaut, das nach dem Zweiten Weltkrieg als gesamtdeutsches Unternehmen noch weiter existierte.

Auch über das System SPRESI<sup>1)</sup> (Speicherung und Recherche struktur-chemischer Informationen), das noch heute benutzt wird, ist Stillschweigen gewahrt worden. Mit dem einleitenden Satz zum Unterkapitel „Informationsflut“: „Internet und Web stellen heute die Infrastruktur für die Chemieinformation dar“ soll wohl zum Ausdruck gebracht werden, dass man damit das Problem der Informationsflut in den Griff bekommen hat.

Angesichts von inzwischen auf 1,4 Millionen chemierelevanten Dokumenten angewachsenen Informationsflut im Jahr 2004 wird eine überraschend einfache Lösung vorgeschlagen: „Es sei notwendig, diese wichtigen Fachinformationen zu strukturieren und zu verarbeiten.“

Diese Tätigkeiten hat der Mensch, ohne oder mit Computerunterstützung, schon ausgeübt als der „Informationsberg“ noch lange nicht so hoch war. Was ist also neu bei dem Vorschlag des Autors? Bei aller Begeisterung für „elektronische Medien“ sollte doch nicht übersehen werden, dass nicht der Computer, sondern der Mensch das eigentliche informationsverarbeitende System ist!

Auch „intelligente“ Computersysteme müssen sich an Denk- und Wahrnehmungsstrukturen des Menschen orientieren. Ob der Einsatz computergestützter

Verfahren effektiv ist davon abhängig, inwieweit es gelingt Probleme zu lösen, die mit der Art der Wissensordnung, der ständigen Änderung von Wissen und dem Problemlösungsverhalten des Menschen zusammenhängen.

Seit vielen Jahren liegen diesbezügliche Veröffentlichungen mit Methoden und Anwendungsbeispielen vor.

Zur Erinnerung stichwortartig einige Literaturhinweise auf Probleme, die auch im Vorfeld von computergestützten Diensten beachtet werden sollten.

### Problem 1: Der multidisziplinäre Zustand von Wissen<sup>2)</sup>

Alle Wissensgebiete sind heute mehr oder weniger multidisziplinär geworden. Das trifft besonders auch für die Chemie zu. Da der Fortschritt der Wissenschaft in den Grenzgebieten liegt, ist der Problemlöser gezwungen, auch Aussagen aus sehr entfernt liegenden Gebieten zu entnehmen. Eine entsprechende Ordnung ist mit den herkömmlichen hierarchisch aufgebauten Klassifikationssystemen nicht zu erreichen. Wie die Erfahrung zeigt, kann die Anwendung eines disziplinunabhängigen Ordnungsprinzips auf der Grundlage von Kategorien nützlich sein.

### Problem 2: Visualisierungsverfahren müssen besser den Funktionen des menschlichen Gedächtnisses angepasst werden<sup>3)</sup>

In Anlehnung an die Erfahrung aus der Chemie, wo mit der Darstellung der Strukturformel einer chemischen Verbindung wesentlich mehr Informationen erkennbar werden, als mit dem Namen wird mit einer strukturbildartigen Darstellung von Wissen eine Einordnung in das Arbeitsgedächtnis des Menschen erleichtert. Diese Methode befördert den „assoziativen Zugriff“ im Gedächtnis und erleichtert die Umstrukturierung vorhandener Gedächtnisinhalte.

Mit dieser Art von Visualisierung können auch Wissenslücken und besonders dicht besetzte Teile des Wissensvorrates leichter erkannt werden.

### Problem 3: Darstellung und Auswertung von Wissensänderungen<sup>4)</sup>

Bedeutung und Struktur des Wissens sind einem ständigen Prozeß der Änderung unterworfen. Da auch Änderungen des Wissens wichtige Informationen vermitteln, sollten sie auch auswertbar gemacht werden.

Wissensänderungen werden am zweckmäßigsten über Änderungen von Wissensstrukturbildern veranschaulicht. In der Chemie ist die Darstellung dynamischer Vorgänge mit Hilfe von Strukturfor-

melbildern in chemischen Reaktionsgleichungen schon lange eine bewährte Methode.

Die graphische Darstellung und Auswertung von Wissensänderungen bei dem oft schwer durchschaubaren Wandel des Begriffsfeldes eines multidisziplinären Wissensgebietes ist ein Problem, das durch Anwendung einer strukturbildartigen Wissensdarstellung auf der Grundlage neuerer Ergebnisse der Gedächtnisforschung und der kognitiven Psychologie einer Lösung näher gebracht werden.

### Problem 4: Unterstützung der schöpferischen Tätigkeit bei der Problemlösung<sup>5)</sup>

Durch eine strukturbildartige Darstellung des vagen und oft widersprüchlichen Ausgangswissens wird das assoziative Denken unterstützt, um lösungsfördernde Begriffskombinationen zu erkennen und um weiterführende Fragen formulieren zu können. Auch die Multi-Aspektivität des Problemwissens wird sichtbar gemacht und damit auch bessere Voraussetzungen für eine wirkungsvolle Problemlösungsstrategie geschaffen.

### Literatur

1) Ševjakowa, L.A.; Scharnow, H.G.; Schildmann, G.: Der Aufbau eines Speichers chem. Verbindungen, – die Realisierung einer neuen Etappe des integrierten Chemieinformationssystem der UdSSR u. der DDR. In: Informatik 24 (1977) 1, S.17-19

2) Bauer, G.: Graphische Darstellung interdisziplinärer Wissensstrukturen. In: Fortschritte in der Wissensorganisation Bd.9 (2006), 15-17

Bauer, G.: Die vielseitigen Anwendungsmöglichkeiten des Kategorienprinzips bei der Wissensorganisation. In: Fortschritte in der Wissensorganisation Bd.9 (2006), 22-23

Fugmann, R.: Die Nützlichkeit von semantischen Kategorien auf dem Gebiet der Informationsbereitstellung. In: Fortschritte in der Wissensorganisation Bd.9 (2006), 34-34

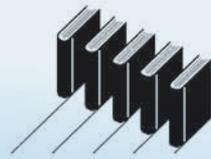
3) Bauer, G.: Visualisierung durch „geordnete“ Strukturbilder des Wissens. In: Fortschritte der Wissensorganisation Bd.5 (1998), 269-278

Schildmann, G.: Unterstützung des Forschungsprozesses durch Visualisierung von Wissensstrukturen. In: Fortschritte in der Wissensorganisation Bd.5 (1998), 279-284

4) Bauer, G.: Strukturbildartige Repräsentation von Wissensänderungen. In: Fortschritte in der Wissensorganisation Bd.3 (1995), 175-187

5) Bauer, G.: Promoting creative processes by a thesaurus-like representation of knowledge structures. In: Tools for Knowledge Organization and the Human Interface Vol.2 (1991), Frankfurt (M), 62-69

Dipl.-Chemiker  
Dr.rer.nat Dr.sc.phil. Gerd Bauer  
Schleswig



**DABIS.eu**

Gesellschaft für Datenbank-Informationssysteme mbH

*Ihr Partner für Archiv-,  
Bibliotheks- und DokumentationsSysteme*

## **BIS-C 2000**

**Archiv- und  
Bibliotheks-  
Informationssystem**

**DABIS.eu - alle Aufgaben - ein Team**

**Synergien: Qualität und Kompetenz  
Software: Innovation und Optimierung  
Web - SSL - Warenkorb und Benutzeraccount  
Lokalsystem zu Aleph-Verbänden**

**Software - State of the art - Open Source**

<b>Leistung</b>	<b>Sicherheit</b>
<b>Standards</b>	<b>Offenheit</b>
<b>Stabilität</b>	<b>Verlässlichkeit</b>
<b>Generierung</b>	<b>Adaptierung</b>
<b>Service</b>	<b>Erfahrenheit</b>
<b>Outsourcing</b>	<b>Support</b>
<b>Dienstleistungen</b>	<b>Zufriedenheit</b>
<b>GUI - Web - Wap - XML - Z 39.50</b>	

### **Archiv**

**singleUser  
Lokalsystem  
multiDatenbank  
multiProcessing  
skalierbar  
Unicode  
Normdaten  
multiMedia**

### **Bibliothek**

**multiUser  
Verbund  
multiServer  
multiThreading  
stufenlos  
multiLingual  
redundanzfrei  
Integration**

## **DABIS.com**

**Heiligenstädter Straße 213  
1190 - Wien, Austria  
Tel.: +43-1-318 9 777-10  
Fax: +43-1-318 9 777-15  
eMail: office@dabis.com  
http://www.dabis.com**

## **DABIS.de**

**Herrgasse 24  
79294 - Sölden/Freiburg, Germany  
Tel.: +49-761-40983-21  
Fax: +49-761-40983-29  
eMail: office@dabis.de  
http://www.dabis.de**

## Die Verwaltung des Deutschen Bundestages sucht für das Referat Pressedokumentation in Berlin einen/eine



### Dipl.-Dokumentar/-in (FH) oder Dipl.-Informationswirt/-in (FH).

#### Die Pressedokumentation des Deutschen Bundestages

Das Referat Pressedokumentation ist die zentrale Auskunftsstelle des Deutschen Bundestages für den Bereich Presseinformation und die größte Informationsstelle ihrer Art innerhalb des Bereiches der obersten Bundesbehörden. Die maßgeblichen Informationsprodukte der Pressedokumentation sind eine täglich erstellte Presse-Frühinformation sowie ein sachlich und geographisch umfassendes Inhouse-Pressearchiv, das durch ein eigenes Lektorat verbal erschlossen wird und im Intranet des Deutschen Bundestages verfügbar ist. Der Auskunftsbereich bearbeitet Anfragen nach Presseinformation aus internen und externen Quellen.

#### Ihre Aufgabe

- Lektur der Tages- und Wochenpresse mit Auswahl und Bewertung relevanter Beiträge
- inhaltliche Erschließung (Indexierung) der Quellen nach den spezifischen Indexierungsregeln und Dokumentations-sprachen für das elektronische Archiv der Pressedokumentation
- Erstellung der elektronischen Pressemappe
- Mitwirkung bei der IT-Koordination des Referates (insbes. bei der fachseitigen Betreuung des laufenden Produktionsbetriebes)

#### Unser Anforderungsprofil

- Fachhochschulausbildung der Studienrichtungen Dokumentationswesen, Mediendokumentation, Informationswirtschaft
- fundierte Kenntnisse der Indexierungs- und Klassifikationstheorie
- Erfahrung in der Inhaltsanalyse und Indexierung von Presseinformationen
- sehr gute Allgemeinbildung
- fundiertes Hintergrundwissen in den Bereichen Politik, Wirtschaft, Gesellschaft, Kultur; Gespür für das aktuelle politische Geschehen wird vorausgesetzt
- spezialisierte Kompetenz in mindestens einem der relevanten Sachgebiete
- sehr gute Kenntnisse der Presselandschaft des In- und Auslands
- sehr gute IT-Anwendungskennntnisse, Erfahrung im Umgang mit den neuesten Entwicklungen im Bereich der IT-gestützten Information und Dokumentation, Erfahrung mit Dokumentenmanagement
- gute englische und französische Sprachkenntnisse
- Fähigkeit zu Konzentration und Arbeit unter Zeitdruck; Belastbarkeit und Bereitschaft zur Ableistung von Früh- und Wochenenddienst
- Kooperationsbereitschaft und Teamfähigkeit

Bitte geben Sie auch außerberuflich erworbene Erfahrungen, Fähigkeiten und Tätigkeiten an.

Die Bezahlung erfolgt nach der Bundesbesoldungsordnung (Bes.-Gr. A 12 BBesO) bzw. dem Tarifvertrag für den öffentlichen Dienst (Entgeltgruppe 11 TVöD). Der Dienstposten ist teilzeitgeeignet.

Die Bundestagsverwaltung gewährleistet die berufliche Gleichstellung von Frauen und Männern. Die Verwaltung des Deutschen Bundestages sieht sich der Gleichstellung von schwerbehinderten und nicht schwerbehinderten Beschäftigten in besonderer Weise verpflichtet und begrüßt deshalb Bewerbungen schwerbehinderter Menschen. Bei gleicher Eignung werden Bewerbungen schwerbehinderter Menschen bevorzugt berücksichtigt.

**Ihre Bewerbungsunterlagen schicken Sie bitte unter Angabe der Kennziffer ID 4/15 bis zum 11. Mai 2007 an:  
Deutscher Bundestag – Verwaltung –  
Personalreferat ZV 1 – Platz der Republik – 11011 Berlin**

# Suchverhalten im Web – Erhebung, Analyse und Möglichkeiten

Nadine Höchstötter, Karlsruhe

Suchmaschinen stellen eine wichtige Möglichkeit dar, um neue Informationen oder Webseiten im Internet zu finden. Das Online-Marketing wurde dadurch verstärkt auf das Suchmaschinen-Marketing ausgerichtet, bei dem vor allem darauf geachtet wird, dass man in Suchmaschinen gut sichtbar ist. Bei der effizienten Planung von Marketing-Aktionen ist es deshalb wichtig, nicht nur profunde Kenntnisse über die Funktionsweise von Suchmaschinen zu haben, sondern auch das Suchverhalten von Suchmaschinennutzern zu untersuchen und die Ergebnisse geeignet umzusetzen. In diesem Beitrag wird auf die Erforschung des Suchverhaltens im Web und auf die Interaktion von Menschen mit Suchmaschinen ab 1997 eingegangen. In erster Linie werden verschiedene Erhebungsmethoden erklärt und Analysen vorgestellt. Es wird auch diskutiert, wie die Ergebnisse umgesetzt werden können, um die Auffindbarkeit von Webseiten zu verbessern.

## Searching behaviour of search engine users – investigation, analysis, opportunities

Search engines are one of the most important possibilities to find new information or web pages. Hence, online marketing activities focus more and more on search engine marketing strategies. For the efficient planning of marketing activities, it is important not only to know about ranking algorithms but also to analyze online searching behaviour of search engine users and to implement those results appropriately. In the following article we will depict the research of web searching behaviour and the interaction of users with search engines since 1997. Primarily, we will describe different observation methods, we will introduce analyses and statistics, and we will discuss how these results can be used to improve the locating of web pages.

## 1 Einleitung

Im Januar 2005 nutzten 13 Millionen Deutsche Suchmaschinen. Im Schnitt werden Suchmaschinen von jedem Nutzer 17 Mal pro Monat genutzt. Google hält seine Quasi-Monopolstellung in Deutschland mit 87 Prozent aller gestellten Suchanfragen. Darauf folgen Yahoo mit 3,3 Prozent und MSN mit 2,4 Prozent<sup>1</sup>. Deutschland war zudem im Jahr 2004 der zweitgrößte Markt für Suchmaschinen-Werbung. Es wurden ungefähr 165 Millionen Euro für Anzeigen in Suchmaschinen und Programme für Suchmaschinen-Marketing ausgegeben. Für das Jahr 2010 werden von Forrester 400 Millionen Euro prognostiziert (Lake 2005). Das zeigt, dass Suchmaschinen eine Dienstleistung darstellen, die sowohl im Alltag als auch im Berufsleben eingesetzt wird, um Informationen oder neue Webseiten beziehungsweise Produkte zu finden. Damit entwickelte sich das Online-Marketing verstärkt zu einem Suchmaschinen-Marketing, bei dem es vor allem darum geht, die obersten Plätze in den Ergebnislisten einzunehmen, wenn eine Suchanfrage passend zu einer Seite eingegeben wird. Um Informationen oder Webseiten tatsächlich effizient im Web anzubieten, reicht jedoch ein reines Suchmaschinen-Marketing nicht. Anbieter von Web-Inhalten sollten sich auch überlegen, wie sie anhand des Suchverhaltens von Internetnutzern ihre Online-Sichtbarkeit verbessern können, die nicht nur durch eingehende Links und Buchung von Anzeigen in Suchmaschinen gefördert werden kann, sondern auch durch eine geeignete Wortwahl auf den Webseiten und eine Ausrichtung an der Bedürfnisstruktur der Suchenden (Schmidt-Mänz/Gaul (2005), Schmidt-Mänz/Koch (2006)).

Der vorliegende Beitrag geht im nächsten Kapitel auf die gängigsten Methoden der Datenerhebung zur Untersuchung des Suchverhaltens ein und legt deren Vor- und Nachteile dar. Im dritten Kapitel werden die Auswertungsmöglichkeiten, die gerade bei der Erhebung von Suchanfragen in Frage kommen, erklärt und an-

hand der Literatur der letzten Jahre zusammengefasst. Im letzten Kapitel wird darauf eingegangen, wie die Kenntnis des Verhaltens dafür verwendet werden kann, Suchmaschinen und Webseiten zu verbessern.

## 2 Erhebungsmethoden

Im folgenden Kapitel werden Methoden vorgestellt, mit denen das Suchverhalten und die Interaktion von Menschen mit Suchmaschinen erhoben werden können. Es wird dabei jeweils die Vorgehensweise kurz erklärt und anschließend die Vor- und Nachteile genannt.

### 2.1 Logfile-Analysen

Die häufigste Methode, die eingesetzt wird, ist die Analyse von Logdateien, die bei Suchmaschinen gespeichert werden. Eine Logdatei beinhaltet das elektronische Protokoll der Interaktionen aller Suchmaschinennutzer mit einer Suchmaschine, welches automatisch erstellt wurde. Informationen, die dabei gespeichert werden, sind beispielsweise die IP-Adresse, die Suchanfrage des Nutzers und die Zeit, zu der die Suchanfrage einging (Jansen 2006). Informationen wie die Angabe, ob Operatoren oder andere Einschränkungen wie die Bildersuche genutzt wurden, sind ebenfalls verfügbar. In vielen Fällen werden auch Cookie-Angaben gespeichert, die wie die IP-Adresse Rückschlüsse auf den Rechner geben, von dem aus die Suche gestartet wurde. Außerdem sind noch die Position der aufgerufenen Dokumente innerhalb der Ergebnisseiten und deren URL verfügbar (siehe Tabelle 1 auf der nächsten Seite). Eine der bekanntesten und ersten Auswertungen einer solchen Logdatei wurde von Silverstein et al. (1999) durchgeführt. Diese Logdatei wurde 1997 von der Suchmaschine Altavista zur Verfügung gestellt.

<sup>1</sup> [www.webhits.de/deutsch/index.shtml?deutsch/webstats.html](http://www.webhits.de/deutsch/index.shtml?deutsch/webstats.html) [1.2.2007]

Tabelle 1: Übersicht über die gängigsten Einträge einer Logdatei (In Anlehnung an Silverstein et al. (1999) und Jansen (2006))

Information in der Logdatei	Beschreibung
Identifizierung des Nutzers	IP-Adresse des Computers. Oft werden diese als Hash-Zahlen ausgegeben, damit Rückschlüsse auf einzelne Nutzer, aber nicht auf die IP möglich sind.
Zeitstempel (timestamp)	Gibt an, zu welcher Uhrzeit eine Suchanfrage abgesetzt wurde. Der Zeitstempel wird seit dem 1. Januar 1970 in Millisekunden gemessen.
Cookie	Gibt ebenfalls Auskunft über den Nutzer, der eine Suchanfrage abgesetzt hat; es handelt sich hierbei ebenfalls um eine eindeutige Kennung. Wenn ein Nutzer keine Cookies zulässt, bleibt dieses Feld leer.
Suchterme	Einzelne Wörter, Zahlen und Begriffe, genau so, wie sie vom Nutzer eingegeben wurden.
Ergebnisfenster	Die Seitennummer des Fensters mit den Suchergebnissen, die von einer Suchmaschine nach der gestarteten Suche zurückgegeben wird. Auf der Seite sind auch Anzeigentexte sichtbar, die passend zu einem Suchterm in der Suchanfrage ausgegeben werden.
Zusätzliche Information	Angabe von speziellen Suchbereichen oder sonstigen Einschränkungen, die von einem Nutzer vorgenommen wurden.
Browser	Angabe des Browsers, der genutzt wird.

Der Vorteil bei der Auswertung solcher Logdateien ist die verfügbare Menge an Informationen zu einem Nutzer. Durch die IP-Adresse oder die Cookie-Information kann der Suchprozess von der ersten bis zur letzten Anfrage ausgewertet werden. Ein Vorteil bei der Auswertung von solchen Massendaten ist auch, dass man einen Eindruck von deskriptiven Statistiken von Suchanfragen bekommt. Die Verteilung von Suchanfragen und deren Längen kann beispielsweise erhoben werden. Ebenso kann berechnet werden, wie groß der Anteil von Operatoren in der Menge aller erhobenen Suchanfragen ist. Damit sind zahlreiche Möglichkeiten gegeben, den Suchprozess zu beobachten. Bei Spink et al. (2000) wird beispielsweise auf die Veränderung einer Suchanfrage während des Suchprozesses eingegangen. Ein Problem bei der Betrachtung von Suchsitzungen (*sessions*) ist allerdings, dass nicht genau festgestellt werden kann, wann diese beendet ist. Ein Nutzer ruft einen Link in der Ergebnisliste auf und „verlässt“ dadurch die Suchmaschine. Zur Definition, wann genau eine Suchsitzung eines Nutzers beendet ist, werden deshalb verschiedene Möglichkeiten gegeben. Die erste Suchanfrage eines Nutzers eröffnet dabei immer die Sitzung, fraglich ist hingegen, wann diese Sitzung als beendet gilt. Eine Möglichkeit ist die Einführung einer zeitlichen Begrenzung. Ruft der Nutzer einen Link innerhalb der Suche auf und kehrt von dieser aufgerufenen Seite innerhalb von beispielsweise zehn Minuten nicht zurück,

dann gilt die Sitzung als beendet. Die nächste Suchanfrage des Nutzers gilt dann als die erste Anfrage einer neuen Suchsitzung.

Ein weiterer Vorteil bei der Betrachtung von Logdateien ist auch, dass sich die Nutzer bei dieser Erhebungsmethode unbeeobachtet fühlen, da sie nicht wissen, dass ihre Daten erhoben werden. Die Analyse von Logdateien von Suchmaschinen ist damit als Feldexperiment zu sehen. Die Ergebnisse sind objektiver und realistischer beziehungsweise als repräsentativer einzuschätzen, als dies bei anderen Methoden der Fall ist.

Nachteilig bei der reinen Auswertung ist das fehlende Wissen über die Sozio-Demographie des Nutzers. Außer einer Identität in Form einer Nummer sind keine Informationen wie Alter, Geschlecht, Familienstand, Gehalt oder Angaben zur Bildung bekannt. Ein weiterer Nachteil ist die gefährdete Privatsphäre von Nutzern, wie das Beispiel der Logdatei der Suchmaschine AOL zeigt, die 2006 zum allgemeinen Download bereitgestellt wurde. Die Logdatei enthielt ungefähr 20 Millionen Suchanfragen von rund 500.000 Nutzern über den Zeitraum März bis Mai 2006<sup>2</sup>. Durch die Suchanfragen zu einem Nutzer, der durch eine gesonderte Identitätsnummer klar nachverfolgbar ist, sind Rückschlüsse auf dessen Gewohnheiten und zum Teil auch auf seinen Namen oder den von Bekannten möglich. Oft werden von Personen, die ihre eigene Website besitzen, ihre eigenen Namen eingegeben, um nachzuprüfen, wie ihre Seite gegenüber anderen Suchergebnissen auf den Ergebnisseiten abschneidet. Das Gleiche gilt bei der Suche nach Geschäftspartnern oder Bekannten, die gesucht werden.

Problematisch ist auch die Bereitstellung solcher Dateien, weil Suchmaschinen den Zugriff auf ihre Logdateien erlauben müssen. Dies wird wiederum von nahezu allen Suchmaschinen kategorisch abgelehnt. Im Hinblick auf das Beispiel der Veröffentlichung der AOL-Logdateien ist dies durchaus verständlich. Es werden meist, wenn überhaupt, auch zu Forschungszwecken nur sehr kleine Datensätze von einem Tag verfügbar gemacht. Beispiele zu der Analyse von Logdateien und ein Überblick solcher Ergebnisse sind in Jansen/Spink (2006) und Spink/Jansen (2004) zu finden.

### 2.2 Liveticker

Logdateien von Suchmaschinen stehen in den meisten Fällen nicht zur Verfügung. Es ist jedoch durch die Beobachtung sogenannter Liveticker möglich, auf eine ähnlich umfangreiche Datenbasis zuzugreifen. Manche Suchmaschinen wie Lycos oder Fireball stellen derartige Liveticker<sup>3</sup> zur Verfügung. Die Webpage wird beispielsweise alle zehn Sekunden aktualisiert und zeigt die gerade eingegangenen letzten zwanzig Suchanfragen. Mit Hilfe dafür konstruierter Programme ist es möglich, diese Ticker abzufragen. Um eine vollständige Liste mit allen Suchanfragen zu erhalten, muss der Liveticker öfter als im Zehntel Sekundenabstand abgefragt werden, denn innerhalb von zehn Sekunden gehen in der Regel mehr Suchanfragen ein als im Liveticker angezeigt werden. Die automatische Abfrage des Livetickers erfolgt deshalb ungefähr alle halbe Sekunde. Dadurch sind aber einige Suchanfragen doppelt in der Datenbasis. Die gespeicherten Listen (immer ein Stapel von zwanzig Suchanfragen in unserem Beispiel) werden jeweils miteinander verglichen und so kombiniert, dass alle redundanten Suchanfragen entfernt werden und eine zeitlich geordnete Reihe von Suchanfragen entsteht. So ist es möglich bei Stoßzeiten alle Suchanfragen zu erhalten. In ruhigeren Zeiten, in denen weniger Suchanfragen an die Suchmaschine geschickt werden, werden sehr viele redundante Listen erzeugt. Insgesamt lassen sich auf diese Weise jedoch nahezu vollständige Datenbestände zur Auswertung erzeugen. Von eigens dafür vorgesehenen Suchanfragen (100.000 fingierte eindeutige Suchanfragen) waren ungefähr 99,5 Prozent in der Erhebung enthalten. Anschließend wurden diese wieder gelöscht, um die Ergebnisse nicht zu verfälschen (Schmidt-Mänz 2007).

Vorteil bei dieser Erhebungsmethode ist, dass man nicht von der Bereitstellung von Logdateien durch Suchmaschinen abhängig ist. Zudem kann eine solche Erhebung über einen sehr langen Zeitraum fortgeführt werden. Durch die Beobachtung verschiedener Liveticker ist der Vergleich von verschiedenen Suchmaschinen und deren Nutzung möglich. Vorteilhaft

2 [www.infoworld.com/article/06/08/07/HNaol\\_searchdata\\_1.html](http://www.infoworld.com/article/06/08/07/HNaol_searchdata_1.html) [1.2.2007]

3 [www.lycos.de/suche/livesuche.html](http://www.lycos.de/suche/livesuche.html) [1.2.2007]

ist auch, dass die Anonymität der Nutzer auf jeden Fall gewahrt bleibt und die erhobenen Daten eine objektive Sicht auf die Nutzungsweise geben, da sich die Nutzer unbeobachtet fühlen.

Nachteilig sind die fehlenden Informationen zu der genauen Interaktion der Nutzer während einer Suchsitzung. Es können also keinerlei Auswertungen auf Basis verschiedener Nutzer oder Suchsitzungen gemacht werden. Die einzige Information, die bei Livetickern gegebenenfalls angezeigt wird, ist die Nutzung von speziellen Suchbereichen. Zudem wird bei dem Aufruf der nächsten Ergebnisseite die Suchanfrage erneut angezeigt, wodurch die Anzahl der gestellten Suchanfragen verfälscht werden kann. Es ist auch nicht möglich, nachzuvollziehen, inwieweit Suchanfragen automatisiert eingegeben wurden.

Ergebnisse aus der Erhebung von Livetickern sind bei Schmidt-Mänz (2007) und Zien et al. (2000) beschrieben.

### 2.3 Umfragen

Eine weitere Möglichkeit ist die Durchführung einer Umfrage. Bei einer Umfrage zum Thema Suchmaschinen bietet es sich an, diese online durchzuführen; ein Vorteil sind die geringeren Kosten und die elektronische Verfügbarkeit der Daten. Schwierig ist die Wahl der Fragen, mit deren Hilfe das Verhalten von Personen bei der Suche abgefragt werden soll. Beispiele solcher Umfragen sind in Schmidt-Mänz (2007, S. 29-59) ausführlich zusammengefasst. Bei der Durchführung einer Online-Umfrage ist darauf zu achten, dass mit leichten Fragen begonnen wird, um dann systematisch das Abstraktionsniveau zu erhöhen. So kann am Anfang mit der Frage nach der Häufigkeit der Internetnutzung und der Nutzung von Suchmaschinen begonnen werden, um später Fragen nach der Nutzung von Operatoren oder speziellen Suchbereichen zu stellen. Es wird davon abgeraten, gleich zu Beginn nach demographischen Angaben zu fragen. Der Teilnehmer sollte zuerst den Sinn und Zweck der Umfrage und die gestellten Fragen kennen, bevor er um persönliche Angaben gebeten wird. Eine weitere Erleichterung ist die Aufteilung der Umfrage in mehrere Seiten, um der Ermüdung von Teilnehmern durch Scrollen vorzubeugen. Eine weitere Hilfe ist die Darstellung des Fortschritts, indem Angaben zur Länge der Umfrage gemacht werden (zum Beispiel „Seite 2 von 5“).

Der große Vorteil von Umfragen besteht darin, dass Angaben zur Person erfasst werden können. Es ist dadurch möglich, verschiedene Gruppen einander gegenüberzustellen. Hölscher/Strube (2000) vergleichen so Anfänger mit Experten, die durch die Dauer der Internetnutzung bestimmt wurden, was über die Angabe des Jahres des Internetbeginns be-

stimmt werden konnte. Bei Schmidt-Mänz (2007) werden weitere Gruppen wie Männern mit Frauen oder Angestellte mit Studenten verglichen. Es können auch Fragen nach der Funktionsweise von Suchmaschinen gestellt werden, um herauszufinden, inwiefern Suchmaschinennutzer eine Vorstellung darüber haben, wie Suchmaschinen Suchanfragen bearbeiten oder ob es unterschiedliche Suchmaschinen-Nutzertypen gibt (Körber 2000). Ein weiterer Vorteil ist die Formulierung so genannter „offener“ Fragen, wodurch Störfaktoren in Suchmaschinen oder Verbesserungsvorschläge erhoben werden können, wenn dies gewünscht ist.

Der Nachteil von Umfragen ist, dass sich die Teilnehmer der Befragung bewusst sind und sich deswegen bei „kritischen“ Fragen in ein besseres Licht rücken wollen. Die Häufigkeit der Nutzung von Operatoren oder der Phrasensuche wird beispielsweise überdurchschnittlich oft angegeben werden. Bei den persönlichen Angaben ist ebenfalls mit zum Teil falschen oder gar keinen Antworten zu rechnen. Eine Möglichkeit ist die Einführung von Kontrollfragen beziehungsweise die Durchführung einer Konsistenzprüfung (Jahr 1950 als Interneteinstieg nicht möglich). Es zeigte sich, dass bei einer Umfrage eine Anhebung des Antwortniveaus stattfindet, aber bei dem Vergleich von verschiedenen Gruppen durchaus wieder Unterschiede deutlich werden (Schillewaert et al. 1998; Schmidt-Mänz 2007, S. 96).

Eine weitere Problematik ist die Repräsentativität der Umfrageergebnisse. Da nicht von „dem“ Internetnutzer ausgegangen werden kann, muss auf eine breite Teilnehmerschaft geachtet werden. Eine weitere Möglichkeit ist die Befragung einer festen Zielgruppe, wie die Nutzer einer bestimmten Suchmaschine (Spink et al. 1999). Eine ausführliche Diskussion zur Repräsentativität und Beispiele zu Umfragen sind bei Schmidt-Mänz (2007, S. 64) zu finden.

### 2.4 Laborexperimente

Seltener als die Analyse von Logdateien und Umfragen werden Laborexperimente durchgeführt (Körber 2000; Machill et al. 2003). Weitere Beispiele solcher Laborexperimente sind bei Schmidt-Mänz (2007, S. 30) aufgeführt. Bei Laborexperimenten werden die Probanden bei der Nutzung von Suchmaschinen in einem dafür eingerichteten Computerlabor beobachtet. Es werden hierfür den Teilnehmern Suchaufgaben vorgelegt, die dann unter Beobachtung gelöst werden müssen (zum Beispiel „Wie heißt der fünfte Präsident der USA?“). Dabei sind verschiedene Fragestellungen möglich, es kann nach allgemeinen und ganz speziellen Themen gesucht werden. Damit lassen sich wiederum Rückschlüsse auf die Nutzung von

Operatoren und speziellen Suchfunktionen bei unterschiedlich schweren Aufgabenstellungen ziehen.

Ein Nachteil sind die Kosten der Durchführung eines solchen Experiments. Die Einrichtung und Bereitstellung eines Labors ist dabei kein unerheblicher Faktor. Weiterhin ist nachteilig, dass sich im Gegensatz zu Umfragen Teilnehmer nicht nur beobachtet fühlen, sondern den Beobachter direkt neben sich sitzen haben. Unter diesen Bedingungen wird sich kaum ein Proband wie bei der alltäglichen Suche verhalten. Es kann auch passieren, dass der Interviewer, der die Beobachtung durchführt, den Probanden aktiv beeinflusst. Zudem sind die Suchanfragen oft sehr konstruiert und entsprechen nicht unbedingt echten Informationsbedürfnissen, die oft spontan aufkommen und nach denen im Internet recherchiert wird. Ein weiterer Nachteil ist, dass solche Experimente sehr zeitaufwendig sind und deswegen auch nur wenige Probanden für die Teilnahme gewonnen werden können (meist zehn bis zwanzig Personen oder weniger). Daraus entsteht neben der möglichen Verminderung der Objektivität das weitere Problem der mangelnden Repräsentativität der Ergebnisse.

Ein Vorteil hingegen ist, dass der Beobachter nach einzelnen Schritten des Probanden bei der Suche beziehungsweise der Interaktion mit Suchmaschinen und nach deren Sinn und Zweck fragen kann (zum Beispiel „Wieso werden an einer bestimmten Stelle Operatoren eingesetzt?“). Ein weiterer Vorteil ist die Kenntnis der Probanden und deren soziodemographischen Merkmale. Diese werden oft durch einen ergänzenden Fragebogen erhoben.

### 2.5 Sonstige

Eine Kombination der verschiedenen Methoden ist durchaus denkbar. So gibt es beispielsweise Online-Panels, bei denen Internetnutzer bei verschiedenen Online-Aktivitäten beobachtet werden, worunter auch die Interaktion mit Suchmaschinen fällt. Der Vorteil besteht darin, dass detaillierte Angaben über Nutzer verfügbar sind, aber der Nutzer nicht unter der gleichen Beobachtung steht, wie dies bei Laborexperimenten der Fall ist. Auch die Analyse von Logdateien kann um einen Fragebogen erweitert werden, indem jeder n-te Nutzer einer bestimmten Suchmaschine diesen präsentiert bekommt, wodurch neben den Logdateien auch Hintergrundinformationen zu den Nutzern erhoben werden (Spink et al. 1999). Bei Laborexperimenten können auch Blickaufzeichnungsverfahren eingesetzt werden, die zusätzlich zeigen, wohin Nutzer tatsächlich schauen, wenn Ergebnisseiten von Suchmaschinen aufgerufen werden. Dies ist eine Erweiterung zu der Analyse, welche Ergebnisse aufgerufen

werden. So kann beurteilt werden, ob tatsächlich vor dem ersten Aufruf eines Links in den Ergebnisseiten die anderen Angaben in der Liste evaluiert wurden. Es ist ebenfalls interessant, inwiefern Links, die zu Werbezwecken eingesetzt werden (beispielsweise Adwords), betrachtet werden (Hotchkiss et al. 2005).

### 2.6 Zusammenfassung

Als die objektivste und aufschlussreichste Methode ist die Auswertung von Logdateien anzusehen. Hier sind die umfassendsten Analysen von Suchanfragen möglich. Zudem kann diese Erhebungsmethode durch eine Online-Umfrage erweitert werden, um auch soziodemographische Informationen zu den beobachteten Nutzern zu erheben. Durch den Charakter eines Feldexperiments fühlen sich die Nutzer unbeobachtet. Die Kosten und der Zeitaufwand der Erhebung sind dabei gering, weil die Logdateien automatisch angelegt werden und in elektronischer Form zur weiteren Analyse zur Verfügung stehen. Bei der Speicherung der Logdateien über einen langen Zeitraum wird lediglich der Speicherplatz wichtig, was auch einen Hauptfaktor der Kosten ausmacht. Bei der Analyse ist unbedingt auf die Wahrung der Privatsphäre der Suchmaschinennutzer zu achten.

Für tiefere Analysen, bei der die Vorgehensweise bei der Suche erforscht werden soll, eignen sich eher Laborexperimente und Umfragen, weil sich die Teilnehmer dabei aktiv zu ihrem Suchverhalten äußern. Logdateien geben nur das Resultat der Suche, also das letztendliche Verhalten wieder, während Laborexperimente und Umfragen das Verhalten während der Suche beziehungsweise Meinungen und Einstellungen von Suchmaschinennutzern aufgreifen können. Diese Erhebungsmethoden eignen sich besser, die Psychologie von Suchmaschinennutzern bei ihrer Recherche zu untersuchen.

## 3 Analyse und Auswertungen von Suchanfragen

Bei der Analyse von tatsächlich gestellten Suchanfragen, wie dies vor allem bei der Analyse von Logdateien und Livetickern von Suchmaschinen der Fall ist, gibt es drei grundsätzliche Auswertungsmöglichkeiten. In den folgenden Abschnitten werden unter Angabe der Literatur, in der die zugehörigen Ergebnisse detailliert zu finden sind, diese vorgestellt und die Ergebnisse zusammengefasst wiedergegeben. Auf Ergebnisse, die aus Umfragen und Laborexperimenten stammen, soll im Folgenden nicht eingegangen werden, da diese oft sehr unterschiedliche Aspekte des Nutzerverhal-

tens betrachten. Hierfür sei auf Schmidt-Mänz (2007, S. 29-59) verwiesen.

### 3.1 Deskriptive Auswertungen

Unter deskriptive Auswertungen fallen die beschreibenden Angaben, mit denen verschiedene Datensätze von Suchanfragen verglichen werden können. Die Größe eines Datensatzes wird durch die Dauer der Erhebung, die Anzahl der Suchsitzungen, der Suchanfragen und -terme angegeben, diese können auch als Bruttozahlen von Suchanfragen oder -termen gesehen werden. Die Bruttozahl gibt somit die Menge aller Suchanfragen an. Eine weitere Angabe ist die Nettoanzahl von Suchanfragen oder -termen, wobei hier nur die eindeutigen Anfragen oder Terme gezählt werden. Darunter ist die Anzahl der Terme zu verstehen, die mindestens einmal gesucht wurden. Bei Schmidt-Mänz/Koch (2006) wurde ein Datensatz analysiert, der 403 Tage umfasste und durch die Beobachtung des Lycos-Livetickers erhoben wurde. Die Anzahl der Suchsitzungen konnte nicht erhoben werden, dies wurde im vorherigen Kapitel erläutert. Insgesamt wurden in diesem Zeitraum 189 Millionen Suchanfragen und 344 Millionen Terme erhoben, wobei davon 29 Millionen Suchanfragen und elf Millionen Terme eindeutig waren. Diese Zahlen sollen dazu dienen, eine Vorstellung von dem Umfang solcher Datensätze zu bekommen.

Beim Vergleich von Datensätzen von Suchanfragen verschiedener Suchmaschinen aus verschiedenen Jahren (Altavista, Alltheweb, Biwe, Excite, Fireball, Lycos, Metager, Metaspinner und Webcrawler) werden immer wieder ähnliche Analysen durchgeführt, die in ihrer Gesamtheit als Benchmarks von Suchanfragen bezeichnet werden können (Jansen/Spink 2006; Spink/Jansen 2004).

Auf die Menge aller Suchanfragen bezogen, sind Suchanfragen im deutschsprachigen Raum mit der Länge von 1,6 bis 1,8 Termen sehr kurz. Sie sind im Schnitt kürzer als englischsprachige Suchanfragen (Jansen/Spink 2006), weil in der englischen Sprache Substantive als Wortzusammensetzung aneinandergereiht werden und nicht wie im Deutschen als ein Wort geschrieben werden (zum Beispiel *christmas tree* vs. *Weihnachtsbaum*). Bezieht man die Länge nur auf die Menge der eindeutigen Suchanfragen, so variiert diese zwischen 2,0 und 2,7 Termen. Dies wird auch durch den sehr hohen Anteil von 48,4 bis 50,1 Prozent an ein-elementigen Suchanfragen deutlich. In Relation zur Menge aller Suchanfragen treten 7,9 Prozent aller Suchanfragen nur einmal auf. Sieht man diese Zahl relativ zu der Nettoanzahl der Terme, dann sind 58,3 Prozent der eindeutigen Suchanfragen auch einzigartig. Bei den Suchtermen ist diese Zahl bedeutend niedriger, 1,3 Prozent der Suchterme tauchen in Relation aller Such-

terme nur einmal auf und 49,1 Prozent der eindeutigen Suchterme sind auch einzigartig (Schmidt-Mänz 2007, S. 131/132).

Operatoren wie AND oder OR werden in den wenigsten Fällen genutzt; häufiger wird die Phrasensuche zur Verfeinerung der Suchergebnisse angewendet („Karl der Große“), um die Suchergebnisse einzuschränken. Insgesamt lag die Nutzungsweise von Operatoren und Phrasensuche unter drei Prozent, der Anteil der Phrasensuche schwankte dabei zwischen 2,1 und 2,5 Prozent. Nur in wenigen Fällen werden natürliche Fragen wie „Wer ist Karl der Große?“ gestellt. Suchanfragen werden als eine natürliche Frage definiert, wenn sie ein Fragewort enthalten. Nur zwischen 0,1 und 0,2 Prozent aller Suchanfragen enthielten solche Fragewörter. Spezielle Suchbereiche wie die Bildersuche werden auch eher selten benutzt, am häufigsten wird mit 65,8 Prozent (Fireball) oder 87,9 Prozent (Metaspinner) die Grundeinstellung der deutschen Suche beibehalten (Schmidt-Mänz 2007).

Als eine weitere deskriptive Auswertung werden oft die Top-10-Terme oder Suchanfragen angegeben. Wie sich zeigte, verändern sich diese selten und nur geringfügig (Schmidt-Mänz/Koch 2006), deswegen wird im nächsten Abschnitt auf die Auswertung von Suchanfragen im zeitlichen Verlauf eingegangen.

### 3.2 Zeitliche Auswertungen

Unter die zeitliche Auswertung von Suchanfragen fallen beispielsweise die durchschnittliche prozentuale Verteilung von Suchanfragen pro Wochentag oder im Ablauf eines Tages. Für diese Angabe werden jedoch größere Datensätze benötigt, die mehrere Tage oder Wochen abdecken. Oft werden solche zeitlichen Auswertungen schon auf der Basis eines Datensatzes durchgeführt, der lediglich einen Tag umfasst, was nicht sinnvoll ist. Das Tageshoch des durchschnittlichen Suchanfragenvolumens liegt zwischen 14.00 und 15.00 Uhr. Ungefähr sieben Prozent aller Suchanfragen werden um diese Tageszeit abgesetzt. Das Tagestief liegt mit gerade einem Prozent in den frühen Morgenstunden zwischen 5.00 und 6.00 Uhr. Montags werden ca. 14 Prozent aller Suchanfragen formuliert, das Wochentief liegt mit zehn Prozent aller Suchanfragen auf den Samstagen (Schmidt-Mänz 2007). Interessant bei der zeitlichen Auswertung ist vor allem die Betrachtung der Nutzung von Suchtermen über die Beobachtungsperiode hinweg. Die Häufigkeit, mit der Suchterme nachgefragt werden, gibt einen Anhaltspunkt, was Suchmaschinennutzer interessiert. Es wurde hier eine generelle Klassifikation von Suchtermen eingeführt (Schmidt-Mänz/Koch 2006, Schmidt-Mänz 2007, S. 148). Ein Suchterm, der in nahezu jedem Zeitintervall eines Erhebungszeitraumes häufig

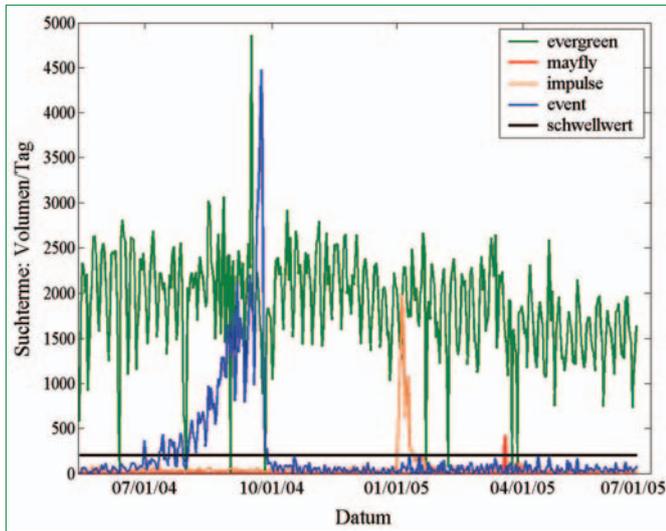


Abbildung 1: Beispielhafte Volumina-Verläufe für Klassen von Suchtermen

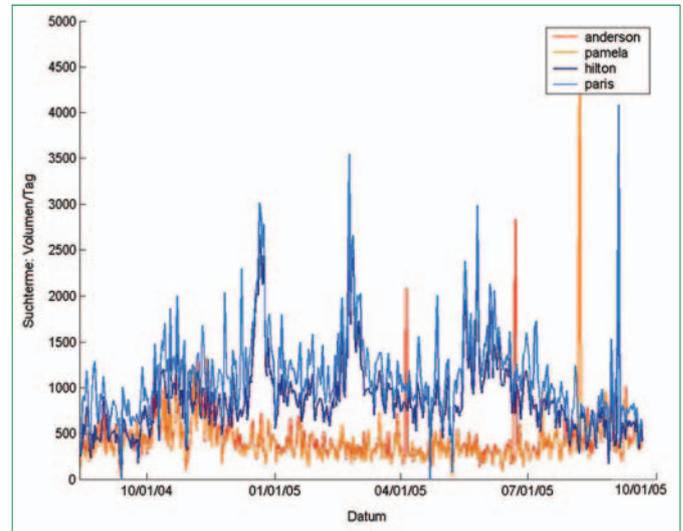


Abbildung 2: Darstellung zweier Cluster von Phrasensuchen

war, wird als Dauerbrenner (engl. *Evergreen*) bezeichnet. Die Zeitintervalle, in welche die Erhebungsperiode zerlegt wird, können beispielsweise als ein Tag oder als vier Stunden beziehungsweise als noch kleinere Taktungen gewählt werden. Ein Term wird dann als häufig eingestuft, wenn sein Auftreten über einem bestimmten Schwellwert (zum Beispiel 100mal pro definiertem Zeitintervall) lag. Es ist für einen Term ausreichend, wenn er nur in 80 bis 90 Prozent aller Zeitintervalle häufig genug war, um als Dauerbrenner bezeichnet zu werden. Unter den Dauerbrennern sind vor allem Terme aus den Begriffswelten der Erotik, der Reise, Multimedia und Musik zu finden. Viele dieser Dauerbrenner stellen auch falsch benutzte Operatoren dar, wie kleingeschriebenes AND oder OR. Füllwörter und Artikel sind ebenfalls unter den Dauerbrennern zu finden. Eine weitere zeitliche Termklasse stellen die Eintagsfliegen (engl. *mayflies*) dar. Sie sind sozusagen das Gegenteil von Dauerbrennern, denn sie sind nur in ein oder zwei Zeitintervallen häufig genug, um in Betracht gezogen zu werden. Bei einem Schwellenwert von 200 Anfragen pro Tag wurden bei dem Lycos-Datensatz 4.837 Eintagsfliegen gezählt, bei einem Schwellenwert von 400 waren es nur noch 2.826. Die Zahlen wären ohne diesen Schwellenwert bedeutend größer. Weitere Termklassen werden durch wiederkehrende Begebenheiten oder Nachrichten beeinflusst. Eine Termklasse ist die der Events, die immer vor dem Eintreten einer angekündigten Begebenheit eintritt (Neujahr, Wahlen, Weihnachten, etc.). Das Charakteristische an dem zeitlichen Verlauf dieser Terme ist, dass sie während einer langen Erhebungsphase nicht häufig auftreten und dass das Volumen zum Ereignis hin ansteigt. Das Maximum des Volumens liegt auf dem Tag des Events. Impulse werden hingegen durch Nachrichten hervorgerufen (zum Beispiel Tsunami, Tod des Papstes). Der charakte-

ristische Kurvenverlauf der Suchanfragen verhält sich gerade entgegengesetzt zu denen der Events. Am Tag des Ereignisses liegt das maximale Volumen an Suchanfragen vor, an den Tagen darauf fällt es schnell wieder ab und bleibt dann wieder unter dem Schwellwert, bis erneut eine Nachricht auftritt, bei der dieser Suchterm wieder wichtig wird. In Abbildung 1 sind die charakteristischen Verläufe der jeweiligen zeitlichen Termklassen anhand der Lycos-Daten aus Schmidt-Mänz/Koch (2006) aufgezeigt.

**3.3 Inhaltliche Auswertungen**

Bei der inhaltlichen Auswertung wird versucht, durch die Klassifikation von Termen in verschiedene Themengebiete (beispielsweise Erotik, Business, Reise) zu zeigen, mit welchen Themen sich Suchende auseinandersetzen. Bei Spink/Gunar (2001) und Spink et al. (2002) wurden solche Klassifikationen eingeführt. Lewandowski (2006) führte ebenfalls eine empirische Untersuchung durch, die sich mit Typen und Themen in Suchanfragen auseinandersetzt. Eine thematische Verlagerung von Suchanfragen von Erotik nach Business wird jedoch durch die Betrachtung der Dauerbrenner bei Schmidt-Mänz (2007) abgelehnt. Sehr kleine Datensätze können durchaus auch thematische Verzerrungen enthalten (Beitzel et al. 2004). Ein weiterer Nachteil bei der inhaltlichen Einteilung von Suchanfragen ist, dass oft nur sehr kleine Teildatensätze genommen werden, um diese wiederum durch eine Person in verschiedene Themengebiete einteilen zu lassen. Automatisierte Verfahren funktionieren ähnlich, indem eine Stichwortliste zu einem Themengebiet erstellt wird, anhand welcher ein Datensatz untersucht werden kann. Enthält eine Suchanfrage mindestens eines der Stichwörter, wird diese dem entsprechenden Themengebiet zugeordnet. In anderen Veröffentlichungen werden Terme anhand der Ähnlichkeiten ihrer Volumina im zeitlichen Verlauf verschiede-

nen Klassen zugeordnet. Hier seien Vlachos et al. (2004) und Schmidt-Mänz (2007, S. 169) genannt. Der Vorteil dieser Methode liegt darin, dass verschiedene Terme nicht unbedingt einem Themenbereich zuzuordnen sind, aber trotzdem in ihrer Kombination dazu verwendet werden, nach einem bestimmten Thema zu suchen. Bei Vlachos et al. (2004) werden zu einem bestimmten Term, der von Interesse ist, zeitlich ähnliche Terme gesucht und diese ausgegeben. Bei Schmidt-Mänz (2007) werden alle Terme, die nicht zu den Klassen der Eintagsfliegen oder Dauerbrenner gehören, verglichen. Die untereinander zeitlich ähnlichsten Terme bilden dann jeweils eine eigene Klasse. Bei einer sehr feinen Einstellung des Algorithmus werden gerade noch die Phrasensuchen ausgegeben. In Abbildung 2 ist ein solcher Verlauf von zwei verschiedenen Clustern dargestellt. Durch die feine Einstellung des Algorithmus wurden nur zwei Cluster im gesamten Datensatz gefunden. Das eine Cluster bestand gerade aus den Termen „pamela“ und „anderson“, während das andere Cluster „paris“ und „hilton“ enthielt. Der Verlauf der Kurven zeigt, wie ähnlich sich diese Terme zueinander in ihren Volumina über die Zeit gesehen sind. Die jeweiligen Terme in den Klassen werden fast ausschließlich zusammen gesucht. Bei einer gröberen Einstellung werden mehr Cluster gefunden, die auch mehrere Terme enthalten (beispielsweise: last, minute, reise). Die Daten stammen aus dem Lycos-Datensatz (Schmidt-Mänz/Koch 2006).

**3.4 Möglichkeiten der Umsetzung der Ergebnisse**

Bei den Möglichkeiten der Umsetzung können verschiedene Bereiche wie Suchmaschinen, Portale oder Webseiten in Betracht gezogen werden. Deskriptive Angaben wie die Verteilung der Suchanfragen und die Nutzung von Operatoren beziehungsweise speziellen Suchbereichen zeigen, dass Suchmaschinen und ihre Nutz-

barkeit für Suchende verbessert werden müssen. Auf Basis der sehr kurzen Suchanfragen ist es verständlich, dass Rankingalgorithmen entwickelt werden müssen, die mit dieser geringen und oft allgemeinen Information umzugehen wissen. Auf der anderen Seite können Suchmaschinennutzer noch nicht mit Suchmaschinen geeignet umgehen, was die Auswertung der Dauerbrenner zeigt, unter denen viele falsch eingesetzte Operatoren und Füllwörter beziehungsweise Artikel zu finden sind, die keine Wirkung auf die Einschränkung einer Suchanfrage haben. Suchmaschinen müssen in ihrer Anwendbarkeit also noch intuitiver aufgesetzt werden. Eine weitere Möglichkeit ergibt sich auch noch durch die Kenntnis von Dauerbrennern. Suchanfragen, die in diese Termklasse fallen, müssen nicht gesondert bearbeitet werden. Für diese Anfragen können Themenlisten erstellt werden, die manuell auf ihre Qualität überprüft werden, da zu diesen Rubriken wiederholt die gleichen Suchanfragen eingehen. Dauerbrenner bieten aber auch Portalen eine Chance, ihre Inhalte zu strukturieren und geeignet darzustellen. Auf Portalen werden ebenfalls Suchanfragen generiert. Die oft durch Informationen überladenen Portale können durch Besucher gar nicht anders bewältigt werden. Die Dauerbrenner-Themengebiete könnten beispielsweise die Navigationsleiste bilden. Dies wird vor allem auch im Bereich von Websites wichtig, die tagesaktuelle Nachrichten oder Informationen anbieten. Oft sind solche Sites auf ähnliche Weise wie Portale strukturiert, es gibt verschiedene Rubriken, anhand der die einzelnen Artikel eingeordnet werden. Durch die Kenntnis von Suchanfragen können Anstiege in den Volumina oder neue Suchterme bemerkt werden, bevor andere Anbieter von Inhalten davon erfahren. Durch diese Kenntnis können Inhalte flexibel am Informationsbedürfnis ausgerichtet werden. Zum einen können Dauerbrenner-Themengebiete als Rubriken dienen, zum anderen kann ein spezielles, aktuelles Thema gesondert dargestellt werden. Aufkommende Trends können besonders kommuniziert werden, indem sie ebenfalls einen wichtigen Platz auf der Hauptseite eingeräumt bekommen.

Anbieter von Websites können durch die Kenntnis von zeitlich ähnlichen Termklassen die Inhalte an den Suchbegriffen ausrichten, die den benutzten Termen auf der Webseite zeitlich ähnlich sind, aber häufiger gesucht werden. Dadurch erhöht sich die Chance, von potentiellen Besuchern gefunden zu werden. Durch die Kenntnis der Begriffswelt der Suchenden sind weitere Optimierungsmöglichkeiten denkbar. Eine weitere Anwendung der Datenbasis von Suchanfragen liegt in der geeigneten Buchung von Anzeigen in Suchmaschinen. Die Datenbasis gibt eine Antwort

darauf, wann und wie lange Anzeigen in Suchmaschinen gebucht werden sollen, um optimale Bietstrategien abzuleiten.

#### 4 Fazit

Insgesamt zeigt die Darstellung der Analyse von Suchanfragen und von Suchmaschinennutzern, dass durch die Erhebung des Suchverhaltens im Web Möglichkeiten der Verbesserung von Suchmaschinen, Portalen und Websites möglich sind, die in vielen Fällen noch nicht (geeignet) umgesetzt wurden. Es sind jedoch noch weitere Untersuchungen vonnöten, um eine profundere Kenntnis der Bedürfnisstruktur zu bekommen. Es können so beispielsweise verschiedene Anordnungen von Ergebnisfenstern untersucht werden, um herauszufinden, an welcher Stelle Anzeigen am besten platziert werden sollten. Eine weitere Möglichkeit ist die Untersuchung verschiedener Eingabefelder in der Suche. In ein Feld werden beispielsweise alle gewünschten Wörter eingegeben, in ein anderes alle unerwünschten. Die Erforschung des Suchverhaltens ist bei weitem noch nicht abgeschlossen. Die Ergebnisse dieses Forschungsgebiets müssen aber mit anderen Forschungsrichtungen wie der Weiterentwicklung von Ranking-Algorithmen oder des Suchmaschinen-Marketings zusammengebracht werden.

#### Literatur

Beitzel, S.; Jensen, C.; Chowdhury, A.; Grossman, D.; Frieder, O. (2004): Hourly Analysis of a Very Large Topically Categorized Web Query Log. In: Proceedings of the 2004 ACM SIGIR Conference on Research and Development in Information Retrieval (ACM-SIGIR), Sheffield, UK, 321-328.

Hölscher, Ch.; Strube, G. (2000): Web Search Behavior of Internet Experts and Newbies. In: Proceedings of the 9th International World Wide Web Conference, 337-346.

Hotchkiss, G.; Alston, St.; Edwards, G. (2005): Eye Tracking Study – An in Depth Look at Interactions with Google Using Eyetracking Methodology. [www.enquiro.com/Eyetracking2-Sample.pdf](http://www.enquiro.com/Eyetracking2-Sample.pdf) [01.02.2007].

Jansen, B. (2006): Search Log Analysis – What it is, What's been done, How to do it. In: Library & Information Science Research (28), 407-432.

Jansen, B.; Spink, A. (2006): How Are we Searching the World Wide Web? A Comparison of Nine Search Engine Transaction Logs. In: Information Processing and Management, 52(1), 248-263.

Körber, S. (2000): Suchmuster erfahrener und unerfahrener Suchmaschinennutzer im deutschsprachigen World Wide Web. Ein Experiment. Magisterarbeit, [www.grin.com/de/preview/27.html](http://www.grin.com/de/preview/27.html) [01.02.2007].

Lake, Ch. (2005): Die Online-Werblandschaft in Europa. DoubleClick, [http://emea.doubleclick.com/uploadpdf/pdf/europe\\_online\\_de\\_05.pdf](http://emea.doubleclick.com/uploadpdf/pdf/europe_online_de_05.pdf) [01.02.2007].

Lewandowski, D. (2006): Query Types and Search Topics of German Web Search Engines, Information Services & Use, 26, 261-269.

Machill, M.; Welp, C. (Hrsg.) (2003): Wegweiser im Netz – Qualität und Nutzung von Suchmaschinen. Gütersloh: Verlag Bertelsmann Stiftung.

Schillewaert, N.; Langerak, F.; Duhamel, T. (1998): Non Probability Sampling for WWW Surveys: A Comparison of Methods. In: Journal of the Market Research Society, 40 (4), 307-322.

Schmidt-Mänz, N. (2007): Untersuchung des Suchverhaltens im Web – Interaktion von Internetnutzern mit Suchmaschinen. Hamburg: Verlag Dr. Kovac.

Schmidt-Mänz, N.; Gaul, W. (2005): Web Mining and Online Visibility. In: Weihs, C.; Gaul, W. (Eds.) (2005): Classification – the Ubiquitous Challenge. Heidelberg: Springer, 418-425.

Schmidt-Mänz, N.; Koch, M. (2006): A General Classification of (Search) Queries and Terms: In: Proceedings of the Third International Conference on Information Technology: New Generations (ITNG'06), Las Vegas, Nevada, USA, 375-381.

Silverstein, C.; Henzinger, M.; Marais, H.; Moricz, M. (1999): Analysis of a Very Large Web Search Engine Query Log. In: ACM SIGIR Forum, 33(1), 6-12.

Spink, A.; Bateman, J.; Jansen, B. (1999): Searching the Web: Survey of Excite Users, Internet Research: Electronic Networking Applications and Policy, 9(2), 117-128.

Spink, A.; Gunar, O. (2001): E-Commerce Web Queries: Excite and Ask Jeeves Study. In: First Monday, Peer-Reviewed Journal on the Internet, 6(7), 4.

Spink, A.; Jansen, B. (2004): Web Search: Public Searching of the Web, Kluwer Academic Publishers, Dordrecht.

Spink, A.; Jansen, B.; Ozmutlu (2000): Use of query reformulation and Relevance Feedback by Excite Users. In: Internet Research: Electronic Networking Applications and Policy, 19(4), 317-328.

Spink, A.; Jansen, B.; Wolfram, D.; Saracevic, T. (2002): From E-Sex to E-Commerce: Web Search Changes. In: IEEE Computer, 35(3), 107-109.

Vlachos, M.; Meek, Ch.; Vagena, Z. (2004): Identifying Similarities, Periodicities and Bursts for Online Search Queries. In: Proceedings of the 2004 ACM SIGMOD International Conference on Management of Data. Paris: ACM Press, 131-142.

Zien, J.; Meyer, J.; Tomlin, J.; Liu, J. (2000): Web Query Characteristics and their Implications on Search Engines. Research Report, RJ 10199 (95073), Almaden Research Center.

Benutzerforschung, Suchmaschine, Marktforschung, Data Mining, Informationsverhalten, Empirische Untersuchung, Methode, Übersichtsbericht

#### DIE AUTORIN

#### Dr. Nadine S. Höchstötter



geb. Schmidt-Mänz arbeitet als Dozentin an der Universität (TH) und der Berufsakademie in Karlsruhe. In ihrer Dissertation befasste sie sich mit dem Suchverhalten in Suchmaschinen. Ihre weitere Forschung ist in der Vorhersage von Suchanfragevolumen und aufkommenden Trends angesiedelt.

Sophienstr. 167, 76185 Karlsruhe  
nsh@topicflux.de  
www.topicflux.de

# Usability von Webportalen und Web-Suchmaschinen im Vergleich

Franz Aschoff und Eva Rausch von Trautenberg, Darmstadt

*Im Rahmen eines berufspraktischen Studienseminars bei der T-Online AG (Darmstadt) wurden Suchmaschinen und Portale mit integrierter Suchfunktion hinsichtlich Funktionalität und Usability untersucht. Ausgehend von der von Fauldrath und Kunisch bereits 2005 durchgeführten Untersuchung wurde erneut geprüft, welche Standards sich bei den Suchdiensteanbietern etabliert haben und welche Trends erkennbar sind. Der bei der zurückliegenden Evaluation verwendete Kriterienkatalog wurde um einige Kriterien erweitert und als Bewertungsschablone der vorliegenden Evaluation eingesetzt. Besonders wichtig für den Nutzwert eines Suchdienstes ist die Relevanz der Treffer. Der bisher weit verbreitete Weg, entsprechende Spezialsuchen anzubieten, verkompliziert die Benutzung der Suchmaschine erheblich, weshalb auch aufgezeigt wird, welche Strategien die Suchdienste zur besseren Antizipation des Nutzer-Informationsbedürfnisses bereitstellen. Die somit erzielte verbesserte Einbindung eigener Dienste und Inhalte stellt einen enormen Gewinn dar, da sowohl Nutzer als auch Anbieter gleichermaßen davon profitieren.*

## **Usability of search engines and portals – an investigation**

*In the context of a traineeship at the T-Online AG (Darmstadt), search engines and portals with integrated search function were examined regarding functionality and usability. Based on an investigation by Fauldrath and Kunisch in 2005, the primary goal was to investigate the existence of standards and trends. This previous evaluation was based on a catalogue of criteria which now had been expanded by some aspects for this reinvestigation.*

*The value of benefit of a search engine is the relevance of its search results. Up to now, the common way to improve these was adding even more special searches for the corresponding search domains. The major disadvantage of this approach is the significantly more complicated handling of the search-interface. As a result of that, this evaluation intends to point out current efforts of anticipating the users' information needs. This leads to a better integration of the website's products and search-services, which results in a great benefit for the user as well as for the search engine provider himself.*

## **1 Einleitung**

Suchmaschinen und Webkataloge sind wichtige Orientierungshilfen im Internet und stellen laut einer Statistik vom Bundesverband Digitale Wirtschaft den zweitgrößten Nutzungsschwerpunkt nach Email im Internet dar (Bundesverband Digitale Wirtschaft 2006).

Wie rasant Suchmaschinen an Bedeutung gewinnen würden, zeichnete sich bereits im Jahr 2000 ab, als laut einer Studie von Jansen und Pooch durchschnittlich jede 28. aufgerufene Webseite die Suchergebnisse einer Suchmaschine war (Jansen/Pooch 2000). Ein Blick auf aktuelle Statistiken zur Beliebtheit der Suchdienste zeigt, dass dieser hart umkämpfte Markt in den letzten Jahren sehr stark an Vielfalt eingebüßt hat und von wenigen Anbietern dominiert wird. Aufgrund unterschiedlicher Erhebungsmethoden sind statistische Angaben zur Marktverteilung nie absolut verlässlich, doch es ist deutlich zu erkennen, dass Google seine marktbeherrschende Position weiter ausbauen konnte. Webstats.de attestierte dem Branchenriesen sogar einen Marktanteil von über 90 Prozent (Heise 2006a). Die Entwicklung neuer Funktionen und Suchdienstleistungen wird daher maßgeblich von Google dominiert. Die nächst größeren Suchmaschinen wie Yahoo und MSN liegen zwar weit abgeschlagen hinter Google, geben sich aber unbeirrt kämpferisch. Um den Nutzer nicht an die Konkurrenz zu verlieren, wird das Service-Angebot ständig erweitert. Auch die Usability der eigenen Dienste wird fortlaufend optimiert, völlig neue Konzepte für eine gute Nutzbarkeit der Webdienste würden jedoch eher dazu beitragen, den User zu irritieren. Internet-Nutzer haben heute eine genaue Vorstellung davon, wie eine Websuche funktionieren muss und welche essenziellen Elemente eine Suchma-

schine zu enthalten hat. Nach Jakob Nielsen erwarten die Nutzer einer Suchmaschine ein Eingabefeld, einen mit „Suche“ beschrifteten Knopf und eine lineare Listendarstellung der Suchergebnisseite (Nielsen 2005). Nennenswerte Änderungen finden nur noch bei den Zusatzfunktionen um diese mittlerweile fest etablierten Standards statt.

Die vorliegende Untersuchung ermittelt, ob sich bestimmte Merkmale der Funktionalität und der Usability als Standard etabliert haben. Unterschieden wird dabei zwischen den klassischen Suchmaschinen und Webportalen, die neben der Suchmaske zusätzlich moderierte Beiträge einbinden. Eine vergleichbare Evaluation wurde bereits im Jahr 2004 von Jens Fauldrath und Arne Kunisch durchgeführt, deren Ergebnisse 2005 in dem Artikel „Kooperative Evaluation der Usability von Suchmaschineninterfaces“ (Fauldrath/Kunisch 2005) vorgestellt wurden. Eine Bewertung der Relevanz und Qualität der Suchergebnisse findet im Rahmen dieser Untersuchung nicht statt. Mit der erneuten Erhebung der Daten wird vielmehr aufgezeigt, welche Veränderungen seit der ersten Erhebung stattgefunden haben und ob bestimmte Entwicklungstrends erkennbar sind.

## **2 Vorgehensweise**

In diesem Test wurden zwei verschiedene Kategorien von Suchdiensten hinsichtlich ihrer Usability und der angebotenen Funktionen analysiert. Neben den klassischen Suchdiensten, deren Schwerpunkt auf der bloßen Suche im Netz liegt, wurden auch Portalanbieter mit integrierter Suchfunktion in die Bewertung einbezogen, bei denen mit Hilfe eines Eingabefeldes im oberen Bereich der Webseiten eine Suchanfrage getätigt werden kann. Hierbei handelt es sich fast immer um eine Web-Suche, bei der mit Hilfe eines

externen Suchmaschinendienstes wie zum Beispiel Google, Yahoo oder MSN nach Inhalten sowohl im Internet als auch auf den eigenen Webseiten gesucht wird.

Die Evaluation wurde von den beiden Autoren durchgeführt, wobei dieselbe Bewertungsmatrix verwendet wurde, die auch Grundlage der Evaluation von Fauldrath und Kunisch war. Verschiedene aktuelle Entwicklungen, wie zum Beispiel die immer mehr Verbreitung findende „Lokale Suche“ machten jedoch eine Erweiterung der Kriterien erforderlich. Eine genauere Beschreibung sämtlicher Kritikpunkte, die den Bewertungsmaßstab dieser Evaluation darstellen, erfolgt in Abschnitt 3.

Die Gewichtung ist in allen Fällen gleich, allerdings können die einzelnen Punkte durch Veränderung eines Gewichtungsfaktors auf- oder abgewertet werden. Hierdurch ist eine erneute Durchführung der Evaluation mit bestimmten, zum Beispiel aus Paneluntersuchungen hergeleiteten Anforderungen möglich, bei denen die Bewertungsfaktor-Gewichtungen speziell angepasst werden. Da sich bei den Portalanbietern die Suchmaske den verfügbaren Platz mit redaktionell gepflegten Inhalten teilen muss, liegt hier hinsichtlich der Usability nur ein Kompromiss im Vergleich zu den klassischen Suchdiensten vor, da aufgrund der höheren Informationsmenge zwangsläufig auch die Übersichtlichkeit beeinträchtigt wird. Diesem Umstand wird in dieser Bewertung Rechnung getragen, indem für den Kriterienpunkt „exponierte Platzierung der Eingabemaske“ eine Abwertung erfolgt. Als dritte Gruppe können die Metasuchmaschinen gesehen werden, die keinen eigenen Index pflegen, sondern die Suchanfragen an eine bestimmte Menge anderer Suchmaschinen weiterreichen. Sie werden in diesem Test jedoch nicht berücksichtigt, da sie immer mehr an Bedeutung verlieren. Besonders Mitte der neunziger Jahre waren Metasuchmaschinen beliebt, da man in ihnen eine gute Lösung sah, eine quantitative Vergrößerung der Informationsbasis zu erzielen. Die Tatsache, dass die Indexgröße der Suchmaschinen inzwischen bei keinem der untersuchten Anbieter mehr auf den Suchergebnissen angezeigt wird, kann als Indiz dafür gesehen werden, dass die bloße Indexgröße heute kein maßgebliches Qualitätsmerkmal für die Suchdienste mehr darstellt.

Ein weiterer wesentlicher Nachteil der Metasuchmaschinen liegt in ihrer stark eingeschränkten Möglichkeit Suchanfragen zu formulieren. Metasuchmaschinen müssen einen kleinsten gemeinsamen Nenner an Suchfunktionen verwenden, die alle eingebundenen Suchdienste verstehen. Nach Marissa Mayer, Produktmanagerin bei Google, ist es vor allem die Geschwindigkeit bei der Ausgabe der Su-

chergebnisseite, die über den Erfolg einer Suchmaschine entscheidet. Und gerade hier liegt der entscheidende Nachteil der Metasuchmaschinen. Diese müssen warten, bis alle eingebundenen Quellen ihre Suchergebnisse ausgeliefert haben, bevor eine Ergebnisseite präsentiert werden kann.

Bei der Bewertung der angebotenen Funktionen wurden bei dieser Evaluation nicht eventuell vorhandene Angaben auf den Betreiberseiten herangezogen, sondern jeder Kriterienpunkt wurde explizit nachgeprüft. Die Untersuchung haben beide Autoren zunächst unabhängig voneinander durchgeführt. Anschließend wurden die Ergebnisse des ersten Evaluationsdurchgangs diskutiert und bei etwaigen Unstimmigkeiten erneut bewertet, bis ein einstimmiges Ergebnis ermitelt werden konnte.

Folgende Suchmaschinen und Portale wurden bewertet:

- Suchmaschinen: Abacho, AlltheWeb, AltaVista, Fireball, Google, HotBot, Seekport.
- Portale: Allesklar, AOL, Excite, Lycos, MSN, T-Online, Web.de, Yahoo!

### 3 Kriterienkatalog

Um darstellen zu können, welche Veränderungen seit der erstmaligen Erhebung von Fauldrath und Kunisch im Jahr 2005 bei den untersuchten Suchmaschinen und Webportalen festgestellt werden können, dient der damals eingesetzte Kriterienkatalog erneut als Grundlage dieser Evaluierung. Allerdings wurde dieser zusätzlich um acht Kritikpunkte erweitert, die in die Kategorien „Zusatzdienste“ und „Ergebnisseiten“ eingegliedert wurden. Der eingesetzte Katalog ist in die Bereiche Suchmaske, Suchfunktion, Zusatzdienste, Ergebnisseiten und Hilfe untergliedert.

Bei den meisten Bewertungspunkten wird abgefragt, ob ein bestimmtes Merkmal vorhanden ist oder nicht. In manchen Fällen werden jedoch mehrere Aspekte zu einem Kritikpunkt zusammengefasst, sodass hier die Angabe eines differenzierten Wertes erforderlich ist, der sich aus der Summe der Einzelkritikpunkte zusammensetzt. Bei der Auswertung der Ergebnisse wird zudem ermittelt, ob das untersuchte Merkmal bei allen Anbietern vorhanden ist und somit einen Standard darstellt oder ob es bei mindestens zwei Dritteln der Anbieter vorhanden ist und somit als Quasistandard gesehen werden kann.

#### 3.1 Suchmaske

Der erste Aufruf einer Suchmaschinen-Webseite zeigt dem Nutzer die Suchmaske, über welche die Sucheingabe vorgenommen werden kann. Um die Aufmerksamkeit des Nutzers nicht übermä-

ßig zu beanspruchen, wird entsprechend der DIN EN ISO 9241-10 ein erwartungskonformes und selbstbeschreibendes Design vorausgesetzt. Schon auf den ersten Blick sollte der Nutzer Antworten bekommen auf die Fragen „Worum geht es hier?“, „Was kann ich hier machen?“, „Was gibt es hier?“ und „Warum sollte ich gerade auf dieser Seite sein – und nicht auf irgendeiner anderen?“ (Krug 2005). Vor diesem Hintergrund wird bewertet, ob das Eingabefeld ausreichend exponiert im oberen Bereich der Webseite platziert ist und ob die Bedienfelder gut erreichbar und selbsterklärend sind. Auch wenn die Suchmaschinen zusätzliche Filterfunktionen direkt über die Startseite anbieten, sollte das „KISS“-Prinzip („keep it small and simple“) auf jeden Fall eingehalten werden. Nur so wirkt die Seite immer noch übersichtlich und nicht zu überladen. Über die Suchmaske sollten die Filterfunktionen für Sprache, Herkunftsland der im Suchergebnis angezeigten Webseiten, Domain und Jugendtauglichkeit der Inhalte deutlich erkennbar sein. Das gleiche gilt auch für die Hilfefunktion, die die Suchmaschine und deren einzelne Funktionen erklärt.

#### 3.2 Suchfunktionen

Ein weiteres Leistungsmerkmal, welches in die Bewertung mit einfließt, ist die Möglichkeit, ein- oder auszuschließende Wörter in den Suchanfragen zu definieren und nach zusammenhängenden Phrasen zu suchen. Diese Funktion wird von den Suchmaschinen auf verschiedene Weise implementiert: Bei einigen Suchdiensten kann dies über die Verwendung der boolescher Operatoren AND, OR oder NOT in den Suchanfragen erfolgen, andere stellen dafür spezielle Eingabefelder in einer „Erweiterten Suche“ bereit. Darüber hinaus wurde bei der Bewertung berücksichtigt, ob eine Anfangs-, Mittel- oder Endtrunkierung möglich ist. Die angebotenen Funktionen wurden mit Hilfe eigener Suchanfragen überprüft, da die Dokumentation der Anbieter in vielen Fällen nur wenig verlässliche Informationen bietet.

Einige Suchmaschinenprovider bieten die Option, in speziellen Bereichen der indexierten Webseiten zu suchen. Üblicherweise können hierfür der Body-Bereich des HTML-Dokuments, der Titel der Webseite und die URL ausgewählt werden. In der Bewertung wird nur generell berücksichtigt, ob eine solche Feldsuche angeboten wird, nicht jedoch deren Umfang. Darüber hinaus wird untersucht, ob es für eine weitere Verfeinerung der Suche möglich ist, regionale oder sprachliche Eingrenzungen vorzunehmen. Auch die Möglichkeit zur Beschränkung auf einen bestimmten Zeitraum, innerhalb dessen die Seiten in den Suchmaschinenindex aufgenommen worden sind, wurde bewertet. Die Crawler der Suchmaschinen



# buchbasel

11. – 13. MAI 2007

BUCH- UND MEDIENMESSE | KLEINVERLAGSMESSE | LITERATURFESTIVAL | JUGENDLITERATURFESTIVAL  
KINDERLITERATURFESTIVAL | COMICS

Freitag und Samstag 09.30 – 19.00 Uhr, Sonntag 10.00 – 17.00 Uhr  
11. – 13. Mai 2007 | Messezentrum Basel

[www.buchbasel.ch](http://www.buchbasel.ch)

*mch*  
messe schweiz

Medienpartner

Badische  Zeitung

Partner

S.V.V.  Schweizer Buchhändler-  
und Verleger-Verband

Basler Zeitung

Beobachter

BILDUNGSCHWEIZ

Bider & Tanner  
Ihre Buchhandlung in Basel

RADIO X 94.5

Schweizer  Buchhandel

Thalia.ch  
ENTDECKE NEUE SEITEN

indexieren aber nicht nur den Inhalt von HTML-Seiten, sondern unter anderem auch von PDF-Dateien, Tabellenkalkulationsdokumenten, Präsentationen und Textverarbeitungsdateien. Daher wurde in dem verwendeten Bewertungskatalog auch die Implementierung von Funktionen aufgenommen, welche die Einschränkung von Suchanfragen auf entsprechende Dateitypen ermöglichen.

Da zu einem Suchbegriff auch mehrere Benennungen existieren können, wurde zusätzlich auch noch bewertet, ob die Suchmaschine automatisch nach Synonymen sucht oder verwandte Wörter zumindest als Suchvorschlag für eine neue Suchanfrage anbietet.

### 3.3 Zusatzdienste

In der Regel beschränken sich die Suchmaschinenanbieter nicht auf die klassischen Suchfunktionen, sondern versuchen mit neuen Zusatzdiensten den Nutzen ihres Services für den User zu verbessern. Bei einem Großteil der Anbieter findet man daher zusätzliche Optionen zur Suche nach Bildern, Nachrichten, Audio- und Videodateien, Beiträgen aus dem Usenet, Produkten oder anderen externen Datensammlungen wie zum Beispiel der Online-Enzyklopädie Wikipedia. Besonders Portalanbieter integrieren häufig die Einträge moderierter Webverzeichnisse in das Suchergebnis. Ob der Anbieter einen eigenen Verzeichnisdienst pflegt oder nur das vorhandene Verzeichnis eines Fremdanbieters einbindet, wurde in dieser Evaluation nicht berücksichtigt. Anstatt lediglich die angebotenen Suchfunktionen und -Räume zu erweitern, bieten einige Suchmaschinen auch Zusatzhilfsmittel an. Besonders Google hat seine Angebote in diesem Bereich ausgebaut, wie zum Beispiel durch die Implementierung von Google Alerts, Google Kalender, Sprachtools und weiteren. Mit einer „Lokalen Suche“, kann beispielsweise nach einem Namen, einer Branche oder Telefonnummer gesucht werden. Mit Hilfe eines zweiten Eingabefeldes für die dazugehörige Ortsbezeichnung wird die Suchanfrage sinnvoll verknüpft. Diese Suchmöglichkeit wird von einer wachsenden Zahl von Suchmaschinen angeboten. Wenn eine „Lokale Suche“ bei den untersuchten Suchmaschinenbetreibern genutzt werden kann, wird dies auch in der Bewertungstabelle vermerkt. Ebenfalls geprüft wird, ob Werkzeuge zum Übersetzen von Webseiten angeboten werden. Altavista war die erste Suchmaschine, die den Übersetzungsdienst Babelfish<sup>1</sup> anbot. Inzwischen werden derartige Werkzeuge, mit denen sich auch komplette Webseiten in eine andere Sprache übersetzen lassen, von immer mehr Suchmaschinenbetreibern adaptiert.

1 <http://babelfish.altavista.com> [20.2.2007]

Auch können Nutzer über eine Frage- und Antwortfunktion Fragen zu einem beliebigen Thema einstellen, welche dann von einer Gruppe registrierter Moderatoren beantwortet werden. Unter der in dieser Evaluation bewerteten Personalisierung ist nicht die individualisierte Einbindung bestimmter Contentmodule in die Suchmaskenseite, wie es zur Zeit beispielsweise von Google angeboten wird, zu verstehen, sondern die Möglichkeit, Einstellungen bei den Suchoptionen und bei der Anzeige der Ergebnisseiten dauerhaft festzulegen. Als letzter Bewertungspunkt dieser Rubrik wird das Vorschlagen alternativer Wörter oder Suchwortkombinationen während der Sucheingabe untersucht.

### 3.4 Ergebnisseiten

Bei den Ergebnisseiten wird überprüft, ob auf den Ergebnisseiten ein KWIC-Index („Keyword In Context“) zu den einzelnen Suchergebnissen eingebunden wird, also der Textausschnitt des gefundenen Dokuments, welcher das gesuchte Wort enthält. Die ausreichende Kennzeichnung von Paid-Links und das optische Hervorheben der Keywords werden ebenso getestet wie das Vorhandensein einer Angabe über den Typ des Dokuments in Fällen, in denen es sich nicht um HTML-Seiten handelt. Unter der getesteten Clusteranalyse versteht man eine Segmentierung innerhalb der Ergebnisse und die thematische Einordnung der Suchraumes mit gleichzeitigem Vorschlag verwandter Themen. Weiterhin wird untersucht, ob eine Rechtschreibkorrektur vorhanden ist, die automatisch falsch geschriebene Eingaben erkennt und ob Angaben über die letzte Aufnahme des Dokuments in dem Suchmaschinenindex gemacht werden.

Suchmaschinen sind bestrebt, dem Nutzer immer möglichst relevante Ergebnisse auf Suchanfragen zu präsentieren. Hier kristallisieren sich verschiedene Lösungsansätze heraus. Viele Suchdienstleister wählen bisher den Weg, den Nutzer über spezielle Tabs oder Buttons eine Vorauswahl treffen zu lassen. So ist bei den drei Suchmaschinen mit den höchsten Marktanteilen, Google, Yahoo und MSN (Sullivan 2006) immer eine spezielle Bildersuche auswählbar, für ortsbezogene Suchanfragen wird die Lokale Suche angeboten und für die Suche nach tagesaktuellen Themen existiert die Nachrichtensuche. Die Problematik der „tab blindness“ zeigt jedoch, dass ein Großteil der Nutzer derartige Vorauswahlmöglichkeiten nicht wahrnimmt. Ein interessanter Ausweg ist daher nach Joachim Griesbaum das „automatische Antizipieren“ des Nutzerverhaltens, bei dem die Suchmaschine eigenständig anhand syntaktischer Kriterien bestimmte Ergebnistypen einblendet (Griesbaum/Bekavac 2004). Bekannte Implementierungen

einer solchen Funktion werden beispielsweise von T-Online in Form des T-Online Produktcontainers angeboten oder bei web.de über die „SmartInfo“-Einblendungen auf den Ergebnisseiten. Ob auch andere Suchdienstleister Vergleichbares anbieten, wird ebenfalls bei der Bewertung berücksichtigt. Abgesehen davon begegnet man oft dem Problem, dass viele Webseiten nur temporär erreichbar sind oder ihr Inhalt sich mit der Zeit grundlegend ändert. Aus diesem Grund speichern einige Suchmaschinen indexierte Webseiten gleichzeitig in einem Cache, sodass diese auch dann noch zur Verfügung stehen, wenn das ursprüngliche Dokument überhaupt nicht mehr an der ursprünglichen Stelle erreichbar ist. Ob solch ein Cache zur Verfügung steht, wird auch in den Bewertungskatalog eingetragen. Bei der Produktlinküberprüfung wurde untersucht, ob auf der Ergebnisseite ein spezieller Link zu zum Suchbegriff passenden Produkten und Dienstleistungen angeboten wird. Als letztes Kriterium der Ergebnisseitenbewertung wird schließlich noch bewertet, ob die Länge der Trefferlisten anpassbar ist und ob eine Vorschaufunktion mit Möglichkeit zum Öffnen der Seite in einem neuen Browserfenster angeboten wird.

### 3.5 Hilfe

Zu jeder Suchmaschine sollte eine Hilfefunktion bereitgestellt werden, die dem Nutzer die verfügbaren Suchfunktionen erklärt, ihm Antworten und Lösungen auf seine Fragen bietet und ihm bei Schwierigkeiten bei der Bedienung der Suchmaschine hilft. Wie genau die Hilfefunktion von den Suchmaschinen und Portalen umgesetzt wird, spielt hier eine untergeordnete Rolle, da diese auf verschiedene Weise sinnvoll implementiert sein kann. Dieser Test bewertet stattdessen generell, ob in der Hilfe eine FAQ oder ein Glossar vorhanden ist und ob die enthaltenen Themen in einem optisch und inhaltlich sinnvoll gegliederten Inhaltsverzeichnis zusammengefasst werden. Abgesehen vom strukturellen Aufbau der Hilfe wird aber auch deren Inhalt bewertet: Werden alle Operatoren erklärt, werden genügend Beispiele angeführt und findet eine Abgrenzung zwischen erweiterter und einfacher Suche statt.

## 4 Erläuterung der Ergebnisse

### 4.1 Suchmaske

Wie in den Abbildungen 1 und 2 dargestellt, werden im Folgenden die Bewertungsergebnisse der Kriterienrubrik „Suchmaske“ dargestellt. Die Ein-Feld Suche hat sich ohne Zweifel als Standard bei allen Suchmaschinenanbietern etabliert, bei den Portalanbietern ist sie als Quasistandard anzusehen, da einzig der Anbieter Allesklar.de eine Ausnahme

macht. Dem Nutzer werden hier schon auf der Einstiegsseite zwei Suchfelder zur Verfügung gestellt, um neben dem eigentlichen Suchbegriff zusätzlich noch eine ortsbezogene Angabe abzufragen. Für Suchanfragen, die einen speziellen lokalen Bezug haben, ist das zwar durchaus sinnvoll, bei Suchanfragen ohne Ortsbezug ergibt so ein sehr spezielles Eingabefeld jedoch keinen Sinn.

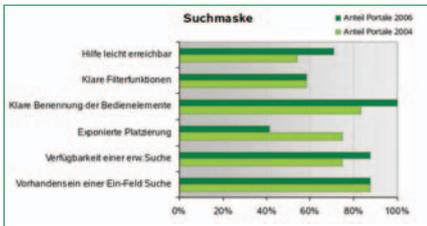


Abbildung 1: Wertevergleich der Kriterienrubrik Suchmaske bei Portalen

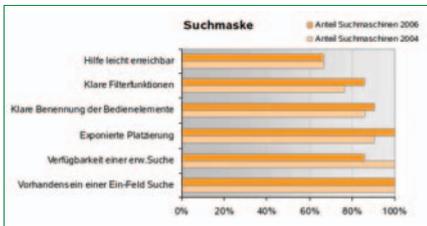


Abbildung 2: Wertevergleich der Kriterienrubrik Suchmaske bei Suchmaschinen

Auffallend ist, dass die Suchmaske bei vielen Portalen nicht ausreichend exponiert platziert ist. Weniger als die Hälfte der Anbieter entsprechen den differenzierten Kriterien der verwendeten Bewertungstabelle. Der Grund hierfür liegt jedoch auf der Hand: Die verschiedenen Portalanbieter müssen dem Problem begegnen, sowohl die Suchmaske als auch die Inhalte so gut wie möglich auf einer Seite zu positionieren, ohne dass sich die Bereiche gegenseitig beeinträchtigen. Dies erklärt auch, warum die exponierte Platzierung bei Suchmaschinen, anders als bei Portalen, ein Standard ist. Hier wird kein Content auf der Startseite eingebunden und das Suchfeld fällt dem Nutzer sofort auf.

Ähnlich wie bei der ersten Erhebung aus dem Jahr 2004 bieten abgesehen von Abacho alle Suchmaschinen eine Erweiterte Suche an.

Generell greifen nur wenige Nutzer auf diese Option zurück: Laut der T-Online-Nutzungsstatistiken machten im August 2006 die Suchanfragen über die Erweiterte Suche in der Gesamtzahl der Suchanfragen gerade mal 0,12 Prozent aus, im Oktober 2006 waren es sogar nur 0,11 Prozent. Nach Machill et al. (2003) kennen 41 Prozent der Suchmaschinenutzer die Möglichkeit Erweiterte Suche nicht einmal. Die klare Benennung der Bedienelemente hat sich zu einem Standard bei den Portalanbietern entwickelt und kann bei den Suchmaschinen als Quasistan-

dard angesehen werden. Gerade bei den Portalen ist eine eindeutige und klare Bezeichnung der Buttons und Felder besonders wichtig, da der Nutzer dem Inhalt schon einen Teil seiner Beachtung zukommen lässt und somit nicht mehr ausschließlich seine Aufmerksamkeit auf die Suche und die vorhandenen Suchkomponenten richten kann. Bei den differenziert bewerteten Filterfunktionen schneiden die Suchmaschinen weit besser ab als die Portale. Ein Familienfilter, mit dessen Hilfe Links zu jugendgefährdenden Inhalten von den Trefferseiten ferngehalten werden, ist abgesehen von Seekport und Abacho bei allen Suchmaschinen implementiert. Ebenfalls differenziert bewertet wird, ob die Hilfefunktion leicht erreichbar ist. Auch hier kann man sowohl bei den Suchmaschinen als auch bei den Portalbetreibern von einem Quasistandard sprechen, wenngleich hier noch erheblicher Verbesserungsbedarf besteht. So ist der Hilfelink bei Abacho beispielsweise in der Fußzeile der Seite platziert, wodurch die Erreichbarkeit nicht als optimal ist. Auch bei Excite ist die Hilfefunktion für den Nutzer äußerst umständlich zu erreichen: Auf den Suchergebnisseiten befindet sich in der Fußzeile der Link „Info/Impressum“, welcher den User auf die gleichnamige Seite führt. Dort befindet sich dann der weiterführende Hilfe-Link, doch die dort angebotene Hilfe bezieht sich lediglich auf sehr allgemeine Fragen zum Portal und nicht auf die Suche. Eine explizite Hilfe zur Suche existiert tatsächlich, allerdings konnte diese nicht auf direktem Wege gefunden werden, sondern erst mit Hilfe der Excite-Suche.

#### 4.2 Suchfunktionen

Unter den Suchfunktionen sind Boolesche Operatoren bei den Suchmaschinen Standard. Auch die Phrasensuche wird mittlerweile von jedem Suchmaschinenanbieter bereitgestellt, selbst wenn in der Hilfefunktion darüber oft kein Wort verloren wird. Auch bei den meisten Portalen ist die Verwendung von Booleschen Operatoren möglich, wenngleich sowohl bei den Portalen als auch bei den Suchmaschinen hierfür oft über eine erweiterte Sucheingabeseite eine spezielle Maske mit mehreren Eingabefeldern für die verschiedenen Suchbegriffe benutzt werden muss. Beispielsweise ermöglicht Allsklar.de die Auswahl des AND- oder OR-Operators ausschließlich über Radiobuttons in der Erweiterten Suche.

Die Feldsuche und Filterfunktion für textbasierte Dateitypen sind als weitere Quasistandards der Suchmaschinen anzusehen. Von den Portalbetreibern bieten immerhin noch 50 Prozent diese Suchfunktion an. Einen Thesaurus zur Suchbegriffserweiterung stellt keine der bewerteten Suchmaschinen dem Nutzer zur Verfügung, von den Portalen unterstützen allerdings immerhin 25 Prozent die-



## 7. DEUTSCHES SINGLE-SOURCE-FORUM

### DITA im Fokus

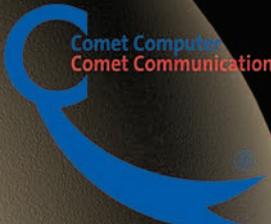
Single-Source-Publishing bedeutet:

- ... Kosten senken
- ... Qualität verbessern
- ... Zeit sparen

**Montag, den 18. Juni 2007  
in München**

Das Branchen-Highlight  
von Comet

Weitere Informationen unter  
089 / 59 98 92 60  
[www.comet.de](http://www.comet.de)



sen Dienst. Möchte ein Nutzer die Suchanfrage nach zeitlichen Kriterien einschränken, wird er bei 57 Prozent der Suchmaschinen zufriedengestellt, kann jedoch nur bei einem Viertel der Portale diese Option zur Suchverfeinerung nutzen.

In dieser Kriterienrubrik fiel ein Ergebnis vollkommen aus dem Rahmen: die Einschränkung der Ergebnisse nach Dateigröße. Weder Suchmaschinenanbieter noch Portalanbieter bieten diese Filterfunktion. Ein Grund hierfür könnte das ansteigende Wachstum der DSL-Anschlüsse in Deutschland sein, wodurch das Herunterladen größerer Datenmengen vergleichsweise unproblematisch und einfach ist. Laut Bitkom ging im Oktober 2006 schon jeder dritte Deutsche über einen Breitbandanschluss ins Internet (Bitkom 2006).

Bei sämtlichen Suchmaschinenanbietern können sprachliche Einstellungen vorgenommen werden. Abgesehen von Abacho, Seekport und Google hat der Nutzer zusätzlich noch die Möglichkeit, gleichzeitig in mehreren Sprachräumen zu suchen. Anders sieht hier das Ergebnis bei den Portalen aus. Allesklar.de und Lycos bieten keinerlei Option, die Suche nach Sprachen einzugrenzen. Anders als bei der sprachlichen Eingrenzung beherrschen alle Portale und Suchmaschinen die regionale Einschränkung der Suchanfragen, bei vielen Anbietern wird dem Nutzer eine umfangreiche Länderauswahl geboten. 71 Prozent der Suchmaschinenanbieter erlauben es, die Recherche nach textbasierten Dateitypen, wie zum Beispiel Word-, Powerpoint-, Excel- oder PDF-Dateien einzuschränken, womit diese Funktion einen weiteren Quasistandard darstellt. Von den Portalen bieten genau die Hälfte diese Funktion an. Eine Übersicht über die Wertevergleiche bei den Portalen und den Suchmaschinen geben die Abb. 3 und 4.

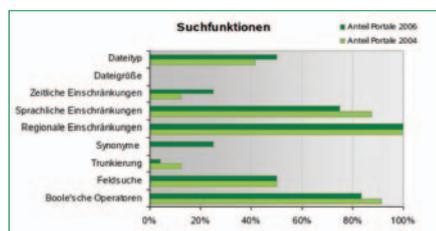


Abbildung 3: Wertevergleich der Kriterienrubrik Suchfunktionen bei Portalen

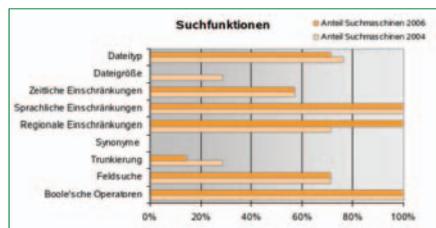


Abbildung 4: Wertevergleich der Kriterienrubrik Suchfunktionen bei Suchmaschinen

### 4.3 Zusatzdienste

In den Abbildungen 5 und 6 werden die Bewertungsergebnisse der Kategorie „Zusatzdienste“ dargestellt. Hier wird sichtbar, dass im Vergleich zu den Evaluationsergebnissen aus dem Jahr 2004 immer mehr Suchmaschinenanbieter eine Bilder- und Videosuche anbieten. Damals boten 71 Prozent der Suchmaschinenanbieter eine Bildersuche und 29 Prozent eine Videosuche an. Außer bei Hotbot ist die Möglichkeit der Bilderrecherche heute bei jeder der bewerteten Suchmaschine möglich und stellt somit einen Quasistandard dar. Ähnlich verhält es sich bei den Portalanbietern. Hier hat sich die Bildersuche und die Möglichkeit nach Produkten zu suchen als ein Quasistandard etabliert. Einzig Allesklar.de bietet diese Zusatzdienste nicht an. Mit bereits 63 Prozent entwickelt sich die Audiosuche bei den Portalen zunehmend und könnte schon bald ein Quasistandard sein. Viele Portalanbieter binden ihr eigenes Angebot bei diesem Zusatzdienst mit ein, wie zum Beispiel Musicload bei T-Online.

Ein Feature, welches immer stärker nachgefragt wird, ist die „Lokale Suche“. Schon 63 Prozent aller untersuchten Portale bieten diese zusätzliche Branchensuche an. Während Allesklar.de und AOL die Suchergebnisse lediglich als Auflistung präsentieren, werden bei T-Online, Web.de, Yahoo und Google die Suchergebnisse in einer Karte eingezeichnet dargestellt, die sich beliebig vergrößern, verkleinern und verschieben lässt. Auch die Nachrichtensuche ist mittlerweile ein wichtiges Feature beider Anbieter geworden. Mit 71 Prozent bei den Suchmaschinen und 88 Prozent bei den Portalen ist dieser Zusatzdienst ein häufig angebotenes Tool und könnte sich bald als Standard etablieren. An dieser Stelle ist eine Abgrenzung des Zusatzdienstes Nachrichtensuche und der Möglichkeit, nach Zeiträumen zu filtern sinnvoll. Die Möglichkeit zur zeitlichen Einschränkung ist bei Suchmaschinen sowie bei Portalen nur noch selten zu finden. Anstelle dieser Funktion tritt immer häufiger die Nachrichtensuche, welche zumindest bezüglich der Aktualität ein Ersatz für den mittlerweile obsoleten Datumsfilter sein kann.

Fast die Hälfte der Suchmaschinenbetreiber ermöglicht dem Nutzer eine vertikale Suche in Datenbanken, über die Hälfte der Portale bietet diesen Zusatzdienst an.

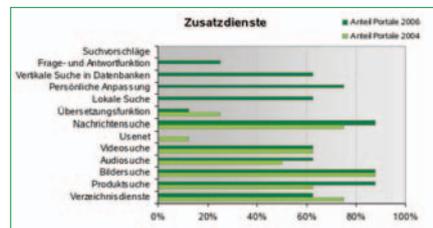


Abbildung 5: Wertevergleich der Kriterienrubrik Zusatzdienste bei Portalen

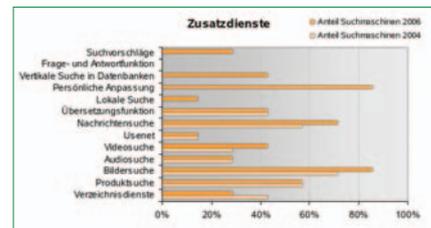


Abbildung 6: Wertevergleich der Kriterienrubrik Zusatzdienste bei Suchmaschinen

Hierbei unterscheiden sich die weiteren Suchmöglichkeiten innerhalb der Suchmaschinen und innerhalb der Portale. So binden die beiden Portale von Web.de und T-Online beispielsweise die Online-Enzyklopädie Wikipedia in ihren Dienst mit ein, ein weiteres Portal bietet unter anderem eine Jobsuche an. Eine Frage- und Antwortfunktion ist momentan bei keiner Suchmaschine vorzufinden, aber bei einem Viertel der Portale. Dabei lässt sich ein Trend hin zur Frage- und Antwortfunktion der Art erkennen, wie sie von Yahoo! Clever und Lycos IQ umgesetzt worden ist und bei denen die Nutzer von einer Community Fragen beantworten lassen können.

Immerhin fast 30 Prozent der Suchmaschinen machen dem User als Eingabehilfe Vorschläge für Suchbegriffe, im Gegensatz zu den Portalanbietern, von denen keiner diesen Zusatzdienst anbietet. Im Vergleich mit den Ergebnissen von Fauldrath und Kunisch fällt das Angebot von Verzeichnisdiensten besonders aus dem Rahmen. Deutlich weniger Suchmaschinen- und Portalbetreiber als es noch im Jahr 2004 stellen einen Katalog zur Verfügung. Damals boten noch 43 Prozent der Suchmaschinen und 75 Prozent der Portale diesen Dienst. Heute sind es nur noch 29 Prozent bei den Suchmaschinen und 63 Prozent bei den Portalen. Selbst Yahoo!, welches als Webverzeichnis startete, hat den „Verzeichnis“-Tab in der Anordnung nach hinten gerückt. Anstelle dessen werden immer häufiger die „Lokale Suchen“ angeboten, die eine komfortablere Suche nach Firmen, Branchen oder Personen ermöglichen.

### 4.4 Ergebnisseiten

Wie in den Abbildungen 7 und 8 erkennbar, sind sowohl für die Suchmaschinen als auch für die Portale die Anzeige des Titels, der URL und einer kurzen Beschreibung des gefundenen Dokuments Standard. Auch die Kennzeichnung von Paid-Links ist bei beiden Suchdienststypen Standard. Hier ist eine deutliche Verbesserung im Vergleich zu den Ergebnissen von 2004 zu erkennen. Damals kennzeichneten nur 63 Prozent der Portale ihre Paid-Links eindeutig, somit konnte man dies damals noch nicht als Quasistandard betrachten. Eine weitere Verbesserung der angebotenen Funktionalitäten zeigt sich bei dem Kriterium „Rechtschreibkor-

rektur“. Mit damals 57 Prozent bei den Suchmaschinen und 25 Prozent bei den Portalen war dieses Feature nur selten vorzufinden. Sowohl die Portalanbieter als auch die Suchmaschinenanbieter haben diese Funktionalität jedoch in den letzten Jahren immer häufiger implementiert. Dadurch schnitten beide Anbietergruppen in der Erhebung mit 71 Prozent und 75 Prozent wesentlich besser ab und so kann heute von einem Quasistandard gesprochen werden. Zusätzlich werden bei allen getesteten Suchmaschinen die gefundenen Schlüsselwörter auf der Ergebnisseite und in den Dokumentbeschreibungen optisch hervorgehoben und auch 88 Prozent der Portalanbieter verwenden das Keywordhighlighting.

Die Bereitstellung von älteren Versionen oder nicht mehr vorhandenen Webseiten mit Hilfe eines Caches findet bei 14 Pro-

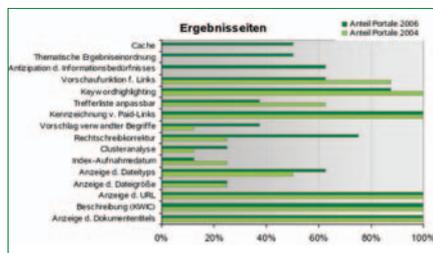


Abbildung 7: Wertevergleich der Kriterienrubrik Ergebnisseiten bei Portalen

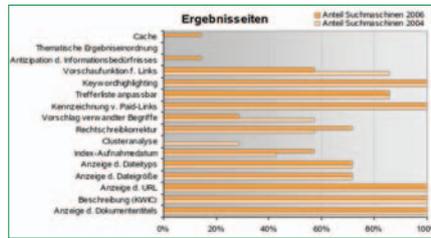


Abbildung 8: Wertevergleich der Kriterienrubrik Ergebnisseiten bei Suchmaschinen

zent der Suchmaschinen und 50 Prozent der Portalanbieter statt. Fast alle Suchmaschinen weisen in den Ergebnislisten auf den Dateityp und die Größe der Datei hin, wodurch zwei weitere Quasistandards erkennbar sind. Dieses Feature ist bei den Portalanbietern seltener anzutreffen. Nur ein Viertel der Betreiber führt die Dateigröße auf.

Es ist jedoch auch zu beobachten, dass einige Merkmale von immer weniger Suchmaschinen angeboten werden. Deutlich wird dies beispielsweise an der Möglichkeit, die verlinkten Seiten der Suchergebnisliste in einem neuen Fenster anzeigen zu lassen. Während dieses Feature vor zwei Jahren noch ein Quasistandard war, sind es mittlerweile nur noch 57 Prozent der Suchmaschinenanbieter und 63 Prozent der Portale, die diese Option zur Verfügung stellen. Ein naheliegender Grund ist die zunehmende Leistungsfähigkeit der verwendeten

Webbrowser, die eine solche Funktion überflüssig machen. Der Webbrowser Firefox öffnet beispielsweise einen Link in einem neuen Tab, wenn dieser mit der mittleren Maustaste angeklickt wird. Seit der Veröffentlichung der Version 1.0 am 4. November 2004 nimmt dessen Verbreitung rasant zu, wodurch der bisherige Marktführer MS-Internet Explorer immer mehr Marktanteile einbüßt (Heise 2006b).

Auch eine Clusteranalyse, welche eine Segmentierung innerhalb der Ergebnisse durchführt, ist recht selten anzutreffen, ebenso wie eine thematische Einordnung des Suchraums mit gleichzeitigem Vorschlag ähnlicher Themen. Lediglich 38 Prozent der Portalbetreiber bieten dem Nutzer einen Vorschlag verwandter Begriffe passend zur Suchanfrage an. Bei den Suchmaschinenanbietern sind es mit 29 Prozent sogar noch weniger, die dieses Feature zur Eingrenzung des Suchterms anbieten. Die Einbindung sowohl eigener Produkte und Dienstleistungen als auch solcher von Fremdanbietern, wie zum Beispiel Wetterinformationen, Fahrplanauskünfte, Stadtpläne, Aktienkurse und ähnlichem in Form einer kurzen Zusammenfassung im oberen Bereich der Suchergebnisseite, hat sich bereits bei 63 Prozent der Portalanbieter durchgesetzt. Bekannte Implementierungen dieser Art sind zum Beispiel die One-Box von

## Weitkämpfer Medien Consulting

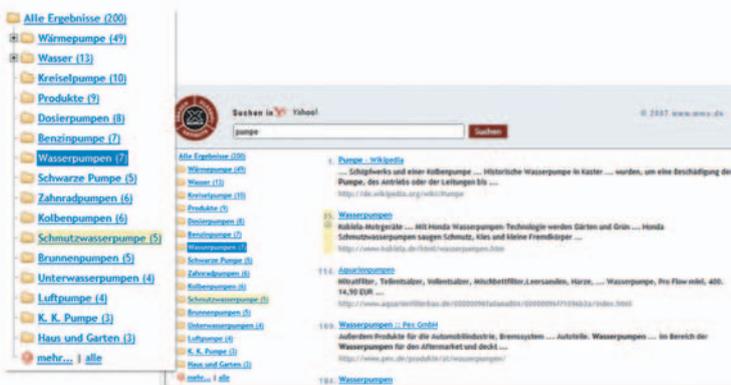
## X-SEARCH Clustering Engine



### Die neue X-SEARCH Clustering Engine

Sie betreiben ein Internet oder Intranet Angebot? Und Ihre Nutzer erhalten oft ewig lange Trefferlisten?

X-SEARCH Clustering Engine organisiert die Suchergebnisse Ihrer Search-Engine zu **thematisch passenden Kategorien**. Dynamisch und in Echtzeit. In Millisekunden.



X-SEARCH Clustering Engine bietet Ihnen auf einen Blick **alle Suchergebnisse perfekt gruppiert**. Sie brauchen keine Taxonomien und kein Klassifizierungssystem.

X-SEARCH Clustering Engine kombiniert intelligente semantische Algorithmen mit einer **kompletten Linguistik Engine** zu dieser einzigartigen Technologie. Für präziseste Cluster. Für viele Sprachen.

X-SEARCH Clustering Engine kann dank HTTP Schnittstelle mit nahezu jeder Search-Engine »on top« kombiniert werden und erfordert **keine Änderung Ihrer Architektur**.

**Zu schön um wahr zu sein? Hört sich so an. Ist es aber nicht.** X-SEARCH Clustering Engine ist das Ergebnis aus 10 Jahren Forschung und Entwicklung. Gerne überzeugen wir Sie mit **einem Clustering Test Ihrer Daten**. Oder fordern Sie **zunächst weitere Informationen an**:

Weitkämpfer Medien Consulting  
Gesellschaft für Elektronische Medien mbH

Reindlweg 3  
D-82418 Seehausen am Staffelsee

Telefon: +49 (0) 8841 488 24 - 0  
Telefax: +49 (0) 8841 488 24 - 88

Email: wmc@wmc.de  
Homepage: www.wmc.de

Discover The Power of Knowledge

Google, der TPC (T-Online Produktcontainer) von T-Online oder InfoExpress von Yahoo. Bei den Suchmaschinen hingegen ist Google der einzige Anbieter, der diese Art der Antizipation des Informationsbedürfnisses einsetzt. Es bleibt weiterhin abzuwarten, welche Portale und Suchmaschinen diesen Dienst in nächster Zeit integrieren werden.

**4.5 Hilfe**

Die Abbildungen 9 und 10 zeigen, dass über die Hälfte der Suchmaschinenanbieter dem Nutzer eine FAQ-Seite sowie ein Inhaltsverzeichnis bieten. Die Portalbetreiber schneiden bei diesen Kriterienpunkten generell schlechter ab. Allgemein wird hier bei beiden Suchdiensttypen ein Erweiterungsbedarf deutlich, denn von einer gut gepflegten und strukturierten Hilfe profitieren sowohl erfahrene als auch unerfahrene Nutzer. Erstaunlicherweise wird von keiner Suchmaschine und nur von sehr wenigen Portalbetreibern ein Glossar in die Hilfe eingebunden. Allerdings sind 81 Prozent der Suchmaschinen und immerhin 54 Prozent der Portale hinsichtlich der angebotenen Informationen im Hilfebereich völlig zufriedenstellend.

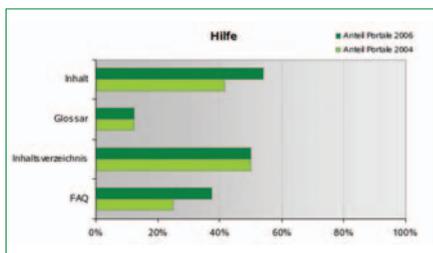


Abbildung 9: Wertevergleich der Kriterienrubrik Hilfe bei Portalen

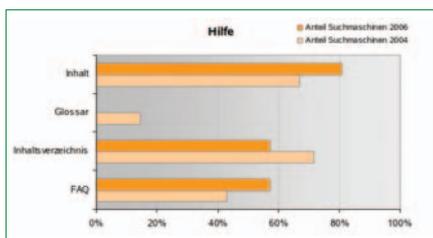


Abbildung 10: Wertevergleich der Kriterienrubrik Hilfe bei Suchmaschinen

**5 Fazit**

Die erneut durchgeführten Usabilitytests zeigen, dass mittlerweile viele Portalanbieter eine umfangreiche Lokale Suche anbieten, häufig mit einer zusätzlichen Kartendarstellung. Für ortsbezogene Suchanfragen bietet diese Kartendarstellung dem Nutzer gegenüber den klassischen Such- und Verzeichnisdiensten einen deutlichen Mehrwert, weil die räumliche Distanz der gefundenen Objekte sehr anschaulich dargestellt wird.

Gemeinsamkeiten der beiden Suchdiensttypen lassen sich vor allem bei den Ergebnisseiten finden, bei denen immer der Titel der gefundenen Webseite, ein Textausschnitt mit den enthaltenen Schlüsselwörtern und die URL angezeigt werden und die gesponserten Links hinreichend gekennzeichnet sind. Auch die regionale Einschränkung der Suchergebnisse ist bei allen Suchdienstbetreibern möglich. Durch diese einheitliche Strukturierung ist ein gewisses Maß an Erwartungskonformität gegeben, die den kognitiven Aufwand des Nutzers bei der Verarbeitung der Suchergebnisseiten reduziert.

Sowohl bei den Suchmaschinen als auch bei den Portalanbietern sind die Webseiten gut für einen iterativen Suchprozess ausgelegt, bei dem nach einer ersten, sehr generellen Suchanfrage zusätzlich weitere, präziser formulierte Suchanfragen gestellt werden können, um somit bessere Ergebnisse zu bekommen. Abgesehen von Abacho wird dies bei allen Suchmaschinen dadurch erreicht, dass auf den Suchergebnisseiten der Bereich zur Eingabe von weiteren Suchanfragen wesentlich kleiner dargestellt wird. Der Wiedererkennungswert bleibt aber bewahrt, weil die Eingabemaske zwar kleiner ist, die einzelnen Funktionen aber die gleiche Anordnung wie auf der Startseite haben. Bei den Portalanbietern wird andererseits der Nutzer bei seiner Suche unterstützt, indem auf den Suchergebnisseiten die zuvor auf der Startseite angebotenen redaktionellen Beiträge, die mit der Suchanfrage nur selten etwas zu tun haben, weitgehend ausgeblendet werden. Auch die Menge der grafischen Werbemittel ist auf den Suchergebnisseiten in der Regel stark reduziert.

Erfreulicherweise ist der Trend zu immer mehr Spezialsuchen erkennbar abgeflaut, stattdessen sind die Suchdienstbetreiber stärker bemüht, die vorhandenen Spezialsuchen durch eine vorausgehende Antizipation des Nutzer-Informationsbedürfnisses besser in die reguläre Websuche einzubinden. Diese Entwicklung kommt aber nicht nur dem Nutzer zugute, für den sich idealerweise die Relevanz der Suchergebnisse erhöht, sondern auch den Suchmaschinen und Portalen selbst, weil sie auf diese Weise die Erreichbarkeit der eigenen Produkte und Dienstleistungen erhöhen können.

**Literatur**

Bitkom (2006): Jeder dritte Haushalt hat einen Breitband-Anschluss. [www.bitkom.org/de/presse/8477\\_42098.aspx](http://www.bitkom.org/de/presse/8477_42098.aspx) [18.02.2007].  
 Bundesverband Digitale Wirtschaft (2006): Online-Nutzung in Deutschland. [www.bvdw.org/fileadmin/downloads/marktzahlen/basispraesentationen/bvdw\\_basisppt\\_online-nutzung\\_20061113.pdf](http://www.bvdw.org/fileadmin/downloads/marktzahlen/basispraesentationen/bvdw_basisppt_online-nutzung_20061113.pdf) [18.02.2007].  
 Griesbaum, J.; Bekavac, B. (2004): Web-Suche im Umbruch? Entwicklungstendenzen bei Web-Suchdiensten. [www.inf.uni-konstanz.de/~gries](http://www.inf.uni-konstanz.de/~gries)

[bau/files/Entwicklungstendenzen\\_bei\\_Websuchdiensten\\_volltext.pdf](http://bau/files/Entwicklungstendenzen_bei_Websuchdiensten_volltext.pdf) [18.02.1007].

Heise Online (2006a): Google in Deutschland über 90 Prozent. [www.heise.de/newsticker/meldung/78315](http://www.heise.de/newsticker/meldung/78315) [18.02.2007].

Heise Online (2006b): Mozilla Firefox gewinnt wieder Marktanteile. [www.heise.de/newsticker/meldung/75311](http://www.heise.de/newsticker/meldung/75311) [18.02.2007].

Jansen, B.J.; Pooch, U. (2000): Web user studies: A review and framework for future work. In: Journal of the American Society of Information and Science Technology. 52(3), 235-246.

Krug, S. (2005): Don't Make Me Think! A Common Sense Approach to Web Usability. Berkeley, Calif.: New Riders Press,

Machill, M., Neuburger, C., Schweiger, W., & Wirth, W. (2003). Wegweiser im Netz: Qualität und Nutzung von Suchmaschinen. In: M. Machill & C. Welp (Hrsg.): Wegweiser im Netz. Gütersloh: Bertelsmann Stiftung.

Nielsen, J (2005): Mental Models For Search Are Getting Firmer. [www.useit.com/alertbox/20050509.html](http://www.useit.com/alertbox/20050509.html) [18.02.2007].

Sullivan, D. (2006): Nielsen NetRatings Search Engine Ratings. <http://searchenginewatch.com/showPage.html?page=2156451> [18.02.2006].

**Benutzerfreundlichkeit, Empirische Untersuchung, Suchmaschine, Portal, Vergleich, Bewertung, Usability**

**DIE AUTOREN**

**Franz Aschoff**



ist Student im Fachbereich Informations- und Wissensmanagement an der Hochschule Darmstadt im Schwerpunkt Wirtschaft. Seine besonderen Interessen sind

Content-Management Systeme und Suchmaschinenmarketing.

Bunsenstr. 14  
 64293 Darmstadt  
[franz@aschoff.name](mailto:franz@aschoff.name)

**Eva Rausch von Traubenberg**



ist Studentin im Fachbereich Informations- und Wissensmanagement an der Hochschule Darmstadt im Schwerpunkt Medien. Ihre besonderen Interessen sind Suchma-

schienenmarketing/-optimierung

Marktplatz 14  
 64283 Darmstadt  
[rausch-von-traubenberg@gmx.de](mailto:rausch-von-traubenberg@gmx.de)

# Visualisierungskomponenten bei webbasierten Suchmaschinen: Methoden, Kriterien und ein Marktüberblick

Bernard Bekavac, Josef Herget, Sonja Hierl und Sonja Öttl, Chur (Schweiz)

Bei webbasierten Suchmaschinen werden zunehmend auch Systeme mit Visualisierungskomponenten für die Ergebnisrepräsentation angeboten. Die Ansätze der Visualisierungen unterscheiden sich hierbei in der Zielsetzung und Ausführung deutlich voneinander. Der folgende Beitrag beschreibt die verwendeten Visualisierungsmethoden, systematisiert diese anhand einer Klassifikation, stellt die führenden frei zugänglichen Systeme vor und vergleicht diese anhand der Kriterien aus der Systematisierung. Die typischen Problemfelder werden identifiziert und die wichtigsten Gemeinsamkeiten und Unterschiede der untersuchten Systeme herausgearbeitet. Die Vorstellung zweier innovativer Visualisierungskonzepte im Bereich der Relationenvisualisierung innerhalb von Treffermengen und der Visualisierung von Relationen bei der Suche nach Musik schließen den Beitrag ab.

## Visualization components in web based search engines – methods, criteria, market overview

Web based search engines using visualization components for the representations of results are becoming more frequent. The approaches for the visualizations diverge however significantly with respect to their objectives and execution. In this paper the applied methods of visualization are described and classified systematically. A market overview on the leading and freely available systems is given followed by a comparison and identification of typical challenges that exist in this context. Finally, innovative new concepts for the visualization of relations in search engine results and the visualization of relations in specific application areas are being introduced.

## 1 Informationsvisualisierung – ein Konzept zur Effizienzsteigerung der Informationsarbeit?

„Ein Bild sagt mehr als tausend Worte“ – dieses Sprichwort zeigt die Bedeutung der visuellen Wahrnehmung und deren Vorzüge gegenüber Texten, Zahlen oder Tabellen. Bilder haben gegenüber textuellen Repräsentationen den Vorteil, dass sie präattentiv wahrgenommen werden, womit die kognitive Belastung durch den Einsatz visueller Darstellungen vermindert und der Einblick in Zusammenhänge erleichtert werden kann.

Unter Visualisierung versteht man den Prozess „Daten, Strukturen und Zusammenhänge grafisch darzustellen, um eine effizientere Analyse und Kommunikation zu erreichen“ (Schumann/Müller 2000, 1). Card definiert Visualisierung als „use of computer supported, interactive, visual representation of data to amplify cognition“ (Card et al. 1999, 6). Neben einer besseren allgemeinen Wahrnehmbarkeit von darzustellenden Inhalten gilt es zudem inhärente, verborgene Zusammenhänge zu offenbaren, um diese ebenso als Basis für Informationsverarbeitungsprozesse nutzen zu können. Als Maßstab zur Bewertung von Visualisierungen dienen hierbei hauptsächlich die Kriterien der Effektivität, Visualisierungen müssen leicht verständlich sein und Expressivität, sie dürfen keine Fakten implizieren, die nicht aus den zu Grunde liegenden Daten hervorgehen.

Suchdienste im Internet stellen heutzutage unverzichtbare Hilfsmittel zur Informationsbeschaffung dar. Bei klassischen Suchmaschinen resultieren Versuche, sich Informationen zu einer bestimmten Problemstellung zu beschaffen jedoch nicht selten mit Trefferzahlen im Hunderttausenderbereich. Ein intellektuelles Sichten und Auswerten der Ergebnisse auf Nutzerseite ist bereits bei wesentlich kleineren Ergebnismengen nicht mehr umsetzbar. Umso wichtiger sind folglich mehrwertschaffende Formen der Ergeb-

nisrepräsentation, die für Nutzer einen Mehrwert leisten im Sinne einer Interpretationsunterstützung. Solche Ergebnisdarstellungen ermöglichen es Nutzern beispielsweise durch neue visuelle Ansätze die Treffer zu selektieren, die bei ihrem konkreten Informationsbedürfnis die höchste Relevanz aufweisen. Die visuelle Unterstützung im Information Retrieval ist ein Weg, der derzeit von diversen kommerziellen Anbietern gegangen wird; Visualisierungen werden folglich vermehrt im Rahmen der Informationssuche eingesetzt mit dem Ziel, einen intuitiven, benutzerfreundlichen Zugang zu Daten zu erlauben.

In letzter Zeit sind vermehrt auch frei zugängliche Systeme für die Websuche auf dem Markt positioniert worden, die klassische Funktionen von Suchmaschinen abdecken und gleichzeitig eine visuelle Ergebnisrepräsentation anbieten.

Der folgende Überblick bietet zunächst eine Einführung mit Kriterien, nach denen Visualisierungen in der Ergebnisrepräsentation von Suchmaschinen strukturiert und kategorisiert werden können (Kapitel 2). Anschließend wird ein Marktüberblick gegeben über frei zugängliche Websuchmaschinen mit integrierter Visualisierungskomponente (Kapitel 3). Zum Abschluss werden zwei aktuelle innovative Ansätze zur Visualisierung vorgestellt (Kapitel 4).

## 2 Visualisierungsmethodiken

Im Nachfolgenden wird eine Systematik der Visualisierungsmethodiken aufgestellt, die aktuell bei Treffermengen-Visualisierungen im Internet angewendet werden. Dies erlaubt im darauf folgenden Kapitel eine bessere Einordnung und einen Vergleich der betrachteten Systeme bezüglich ihrer Anwendungsgebiete.

### 2.1 Suche im eigenen Index oder Metasuche

Zunächst ist es wichtig zu unterscheiden, ob die Suchmaschinen mit Visualisierungskomponenten auf einem eigenen,

respektive einzelnen Index bezüglich der zu visualisierenden Treffermenge basieren oder ob es sich hierbei um eine Meta-suche handelt, die den Index verschiedener anderer Suchmaschinen befragt und dann darauf basierend eine Art Übersicht über die vereinte Gesamttreffermenge gibt. Die Unterscheidung ist vor allem deshalb wichtig, weil bei einem eigenen Index neben den Treffern selbst auch konkrete Werte bezüglich des Ranking sowie anderweitige Informationen wie zum Beispiel statistische Werte bezüglich der Suchbegriffe oder Verlinkungsstruktur innerhalb der Treffermenge vorliegen, die dann in die Visualisierung einfließen können.

Anders ist das bei Metasuchdiensten, diese können allenfalls auf Grund von Rankingpositionen innerhalb der Treffermengen der befragten Suchdienste auf die Relevanz schließen. Da jedoch in der Regel mindestens zwei bzw. mehrere Treffermengen vorliegen, ist ein einheitliches Ranking der Gesamttreffermenge kaum bzw. nur mit sehr einfachen Mitteln zu bewerkstelligen. Dies zeigt sich auch in den im nächsten Kapitel vorgestellten Metasuchdiensten, die im Prinzip nur eine Übersicht der Treffermenge visualisieren und nur geringe Relevanzinformationen aufweisen.

In diesem Zusammenhang ist es natürlich wichtig zu wissen, ob Ranking-Algorithmen hinterlegt sind und ob das Ranking selbst aus der Visualisierung ersichtlich ist.

## 2.2 Kriterien für die Visualisierung

Neben den verwendeten Metaphern zur Darstellung von Treffermengen bzw. Ergebnisdokumenten, wie zum Beispiel Dokumentenlandkarten oder die Verwendung spezieller Icons, werden vor allem auch diverse Interaktionstechniken wie Zooming und Filtering angewandt. Mann (2002) beschreibt Visualisierungsmöglichkeiten nach den Kriterien der verwendeten Metaphern (zum Beispiel Landkarten), Techniken (zum Beispiel Zoom), Komponenten (zum Beispiel Visualisierung der Suchanfrage oder Visualisierung von Dokumenteigenschaften) und Systeme (zum Beispiel TileBars). Wild (2004, 104) unterscheidet unter dem Kriterium „Einsatzort“, an welchen Stellen die Visualisierung in einem Retrievalprozess angewandt wird (Suchformulierung, Ergebnismenge etc.). Im vorliegenden Beitrag betrachten wir ausschließlich die Ergebnismengenvisualisierung, daher ist die entsprechende weitere Unterteilung in set level, site level und document level relevant:

- **Set level:** das Suchergebnis als Ganzes wird visualisiert. Es geht hier darum, einen Überblick über die Ergebnismenge zu erhalten, um „Trends“, „Hot Spots“ und „Clusters“ zu erkennen. Typische Beispiele für den Einsatzort

„Set level“ sind Landschafts- und Landkartenmetaphern.

- **Site level:** das Site level bezeichnet eine Ebene zwischen Dokument und Dokumentenraum. Visualisierungen auf dieser Ebene bezwecken, die Organisationsstruktur einer Website oder eines Pfades, dem ein Benutzer im Web gefolgt ist, aufzuzeigen.
- **Document level:** Visualisierungen auf dieser Ebene zeigen Inhalt, Struktur und Relevanz von Einzeldokumenten. Ein typisches Beispiel für den Einsatzort „Document level“ sind so genannte Thumbnails.

Die betrachteten Systeme sind meist im Bereich des „set level“ angesiedelt, bei einem System kann von einem einfachen „site level“ gesprochen werden. Keines der Systeme bietet eine Visualisierung im Bereich des „document level“ an.

Auf der Ebene der Darstellung von Dokumentenrelationen unterscheidet Wild (2004, 105) sehr differenziert, wie die Dokumentenrepräsentanten und die Relationen zwischen den Repräsentanten in der graphischen Darstellung strukturiert sind. Dabei wird zwischen Makro- und Mikrolevel unterschieden:

- **Makrolevel:** Auf dieser Ebene werden die Darstellungsformen mit Hilfe von vier Attributen beschrieben. Die Visualisierung einer Ergebnismenge ist entweder „hierarchisch“, „netzwerkartig“, „linear“ oder „cluster-artig“.
- **Der Mikrolevel** beschäftigt sich mit dem Verhältnis zweier Einzelelemente bzw. Dokumente. Die zum Mikrolevel gehörenden Attribute sind „nominal“, „ordinal“ oder „quantitativ“. Ein nominales Verhältnis zweier Einzelelemente bedeutet, dass eine Relation zwischen den Elementen vorhanden ist. Weitere Aussagen über die Natur des Verhältnisses können nicht gemacht werden. Ordinale Verhältnisse ermöglichen eine intuitive, aber nicht genau messbare Aussage beispielsweise über die Relevanzverhältnisse (Treffer A ist relevanter als Treffer B). Ein als „quantitativ“ charakterisiertes Verhältnis ermöglicht das Auslesen konkreter Werte (zum Beispiel Treffer A ist 25 Prozent relevanter als Treffer B).

Auf Makroebene kommen bei den betrachteten Systemen alle vier Ausprägungen vor. Auf Mikroebene bieten nur wenige Systeme eine direkte Relation zwischen zwei Trefferdokumenten. Meist ist nur eine nominale Unterscheidung möglich, da sich einzelne Trefferdokumente in einem Cluster befinden und in einem Fall liegt eine Art ordinale Unterscheidung vor, da man zwischen den einzelnen Treffern einer Metasuchmaschine die Rankingposition bei den jeweiligen Quell-Suchmaschinen direkt vergleichen kann. Bei der graphischen Darstellung der Dokumentstellvertreter werden die opti-

schen Merkmale und die Dimension der visuellen Aussage betrachtet. Die graphische Beschaffenheit der Dokumentrepräsentanten kann dabei abstrakt, ikonisch oder realistisch sein, wobei die Zuordnung auch mehrfach vorkommen kann, zum Beispiel wird bei Kartoo ein Trefferdokument in der Übersicht ikonisch dargestellt und bei Mouse-Over dann eine realistische Darstellung in Form eines verkleinerten Screen-Shots (Thumb-Shot) eingeblendet. Wenn beispielsweise dokumentformatabhängige (zum Beispiel HTML, pdf) Icons verwendet werden, so besitzt der Repräsentant eine Dimension. Wenn zudem noch die Größe eines dargestellten Icons Auskunft gibt über die anhand eines Ranking ermittelte Relevanz, so zählt dies als eine weitere Dimension. Deuten zusätzlich bspw. verschiedene Farbabstufungen noch die Suchmaschinenherkunft an, dann wäre die Dimensionalität auf drei festzulegen.

Die maximal darstellbare Menge an Dokumenten sowie eine evtl. gruppierte Darstellung von Teilbereichen (zum Beispiel als Cluster) wird von Wild (2004, 106) unter dem Begriff Skalierbarkeit betrachtet. Die visualisierbare Treffermenge wird durch das Unterkriterium Volumen mit den Attributen small-, medium-, large- und flexible-scale gekennzeichnet. Die Attribute beziehen sich auf die Dokumentenmengengrößen:

- **Small-scale:** wenige Dokumente (bis zu 20)
- **Medium-scale:** viele Dokumente (20-500)
- **Large-scale:** sehr viele Dokumente (mehr als 500)
- **Flexible-scale:** variable Anzahl von Dokumenten.

Der Verdichtungsgrad bei der Darstellung von Teilbereichen einer Treffermenge wird durch die Attribute „aggregiert“, „explizit“ und „detailliert“ unterschieden:

- **Aggregiert:** Mehrere Dokumente werden zu einer Gruppe zusammengefasst. Jeder Gruppe wird ein Dokumentrepräsentant zugeordnet.
- **Explizit:** Jedem Dokument der Treffermenge ist genau ein Dokumentrepräsentant zugeordnet.
- **Detailliert:** Die visuelle Repräsentation erstreckt sich bis auf die Inhalte des einzelnen Dokuments.

Die Interaktivität einer Visualisierung bezeichnet alle Formen der möglichen Veränderung einer Visualisierung durch den Benutzer. Wild (2004, 106) umschreibt mehrere Interaktionsformen, wobei man von interaktiver Vollständigkeit spricht, wenn eine Visualisierungsmethode alle aufgeführten Formen zur Verfügung stellt:

- **Overview:** Mit Hilfe dieser Funktion können die Merkmale identifiziert wer-

den, durch die ein Objekt erkennbar ist. Die absolute oder relative Lage eines Objekts kann festgestellt werden. Die Funktion dient der Überblicksbildung.

- **Zoom:** Vergrößerung und Verkleinerung der Ansicht.
- **Filter:** Die Möglichkeit für den Nutzer irrelevante Daten aus der Darstellung zu entfernen.
- **Details-on-demand:** Zu einzelnen Objekten sollen Zusatzinformationen abrufbar sein, zum Beispiel indem man mit der Maus auf das Objekt zeigt.
- **Relate:** Informationen zu Beziehungen zwischen Objekten (zum Beispiel: gehören Treffer zusammen, kann man sie gruppieren, und ähnliches).
- **History:** Die Funktion soll den iterativen Prozess der Suche auch in der Visualisierung von Treffermengen unterstützen. Typischerweise gehören dazu Funktionen wie „undo“ und „redo“.
- **Extract:** Die Ergebnisse einer Suchanfrage sollen sich extrahieren lassen (E-Mail, drucken, speichern).
- **Configure:** Programmeinstellungen ändern (zum Beispiel Unterscheidung der Quell-Indexe bei Metasuche).
- **Pan:** Veränderung bzw. Verschiebung des zu betrachtenden Ausschnitts.

### 3 Marktübersicht

Im Folgenden wird ein Überblick über die bekanntesten frei zugänglichen Web-suchmaschinen mit Visualisierungskomponenten gegeben, die anhand der in Kapitel 2 erläuterten Kriterien klassifiziert und systematisiert werden.

Die Systeme unterscheiden sich grundsätzlich durch ihre Zielsetzung, es können folgende drei grobe Zielkategorien identifiziert werden, anhand derer der nachfolgende Marktüberblick gegliedert wird:

1. **Vergleichende Metavisualisierungen:** Ziel der Systeme dieser Kategorie ist es darzustellen, welche Treffer bei einer Metasuche von den verschiedenen zugrundeliegenden Suchmaschinen gleiche oder unterschiedliche Treffer ermitteln und welche Schnittmengen vorliegen.
2. **Clustervisualisierungen** haben zum Ziel, die Suchanfrage des Nutzers entsprechend verschiedener Themenkomplexe zu beantworten.
3. **Suchmaschinen mit Relationenvisualisierung** bezwecken eine Visualisierung der Zusammenhänge zwischen den Treffern des Ergebnisraums.

#### 3.1 Vergleichende Metavisualisierung

Die erste Grobkategorie der vergleichenden Metavisualisierung umfasst zwei Suchanwendungen, bei denen die Ergebnismengen von zwei bis vier Suchmaschinen anhand einer Visualisierung dargestellt werden.

Hierdurch lässt sich ermitteln, wie stark sich Indizes und auch das Ranking bei den verschiedenen angebotenen Suchmaschinen unterscheiden und welche Treffer bei einer Suchanfrage als Schnittmengen von mehreren Suchmaschinen ermittelt wurden.

##### 3.1.1 Dogpile

*Dogpile* wurde im Jahr 1996 als eigenständige Metasuchmaschine auf den Markt gebracht und gilt heute als einer der führenden Anbieter in der Metasuche mit Auswertung der Ergebnisse von führenden Suchmaschinen wie Google, Yahoo! Search, MSN, Ask.com, About, MIVA, LookSmart und weiteren.

In einer kooperativen Forschungsstudie, die 2005 von Dogpile, der University of Pittsburgh und der Pennsylvania State University durchgeführt wurde, stellten die Forschungspartner fest, dass lediglich ein überraschend kleiner Anteil von 1-11 Prozent der Suchergebnisse bei den vier Keyplayern Google, Yahoo! Search, MSN und Ask.com (1,1 Prozent) bzw. bei zwei dieser Suchmaschinen (11,4 Prozent) übereinstimmen (Spink et al. 2005). Zur Verdeutlichung dieser geringen Schnittmenge wurde 2005 das Dogpile Search Comparism Tool entwickelt, das die Schnittmengen anhand einer Visualisierung darstellt.

Dieses System ermöglicht jeweils den Vergleich von drei der vier oben genannten Suchmaschinen, wobei die Ergebnisse jeder Suchmaschine anhand einer Kreismenge dargestellt werden mit Treffern auf Documentlevel, durch Kugeln repräsentiert.

Die Schnittmengen der drei Kreise lassen einen Rückschluss zu auf gemeinsame Treffer der entsprechenden Suchmaschi-

nen. Anhand einer Farbcodierung wird dargestellt, welche der dargestellten Treffer einzelner Suchmaschinen durch die Dogpile-Metasuche als Topergebnis ermittelt wurden.

Mittels einer Mouse-over-Funktion werden zu jeder Kugel Hintergrundinformationen, wie URL oder ein kurzer Teaser dargestellt (Abbildung 1).

##### 3.1.2 Langreiter

Die Metaanalyse-Suchmaschine „Google vs. Yahoo“ wird seit 2004 von Christian Langreiter als Experiment betrieben und ermöglicht den Vergleich der Treffermengen von Yahoo und Google bzw. von Google.com und Google.cn.

Aus der Visualisierung wird ersichtlich, welche Treffer bei beiden Suchdiensten ermittelt und auf welche unterschiedlichen Positionen sie jeweils gerankt werden.

Die Ergebnismenge wird durch Kreise, die auf zwei parallel verlaufenden Geraden (eine Gerade repräsentiert jeweils die Treffer einer Suchmaschine) dargestellt, wobei jeder Kreis einem Treffer entspricht. Die Kreisposition (von links nach rechts gerechnet) innerhalb der Geraden entspricht dem Ranking des Treffers bei der entsprechenden Suchmaschine. Wurde ein Treffer bei beiden Suchmaschinen ermittelt, werden die beiden zugehörigen Kreise blau ausgefüllt dargestellt und mit einer dünnen Linie verbunden. Hierdurch wird auf einen Blick ersichtlich, bei welcher Suchmaschine der Treffer weiter nach vorne oder hinten gerankt wurde und ob ein Treffer bei der jeweils anderen Suchmaschine überhaupt ermittelt wurde (Abbildung 2).

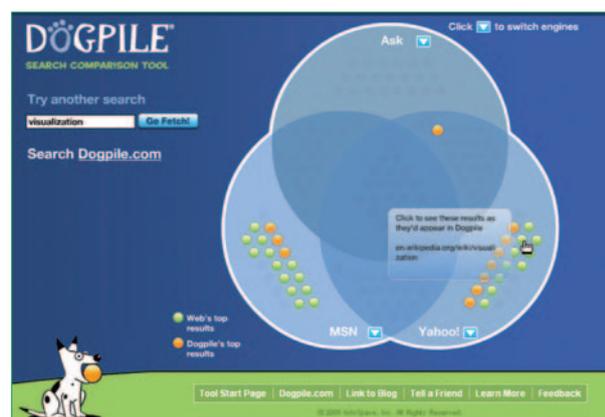


Abbildung 1: Schnittmengendarstellung bei Dogpile zur Suchanfrage „Visualization“

#### DOGPILE Hintergrund

<b>URL:</b>	<a href="http://missingpieces.dogpile.com/">http://missingpieces.dogpile.com/</a>
<b>Unternehmen:</b>	InfoSpace, Inc.
<b>Produkte:</b>	Websuche
<b>Funktionen:</b>	Metasuchmaschine
<b>Visualisierung:</b>	Schnittmengen-Darstellung mit Kugeln als Repräsentation einzelner Suchergebnisse

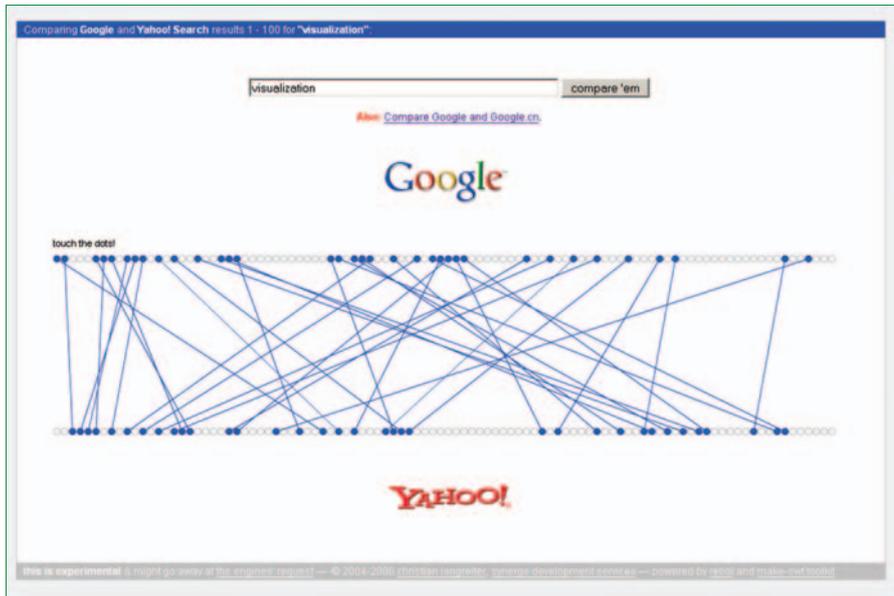


Abbildung 2: Vergleichsdarstellung bei Langreiter zur Suchanfrage „Visualization“

**LANGREITER Hintergrund**

**URL:** www.langreiter.com/exec/yahoo-vs-google.html  
**Unternehmen:** Synerge Development Services  
**Produkte:** Websuche  
**Funktionen:** Metasuchvergleich  
**Visualisierung:** Vernetzte Darstellung auf Dokumentlevel

**3.2 Clustervisualisierung**

Suchmaschinen mit Clustervisualisierung unterteilen die ermittelte Ergebnismenge in grobe Kategorien und erleichtern dem Nutzer somit die Selektion der für ihn relevanten Treffer. Die Cluster werden anhand verschiedener Ansätze visualisiert, wobei Gestaltungsmittel wie beispielsweise räumliche Nähe, Geschlossenheit oder der Einsatz einer Farbcodierung zum Tragen kommen.<sup>1</sup>

**3.2.1 Grokker**

Grokker wurde ursprünglich als kostenpflichtige Desktop-Suchmaschine für unternehmensweite Suche entwickelt, ist aber bereits seit 2005 als Referenz ihrer Leistungsfähigkeit als Websuchmaschine frei verfügbar. Aufgrund einer Kooperation mit Yahoo basieren die Ergebnisse auf dem Yahoo-Index, wobei sich dieser um die Indizes von Wikipedia und Amazon-Beständen erweitern lässt.

Es handelt sich somit um eine Art Metasuchmaschine, die eine freie Indexwahl zulässt.

Neben der klassischen Listendarstellung, wie sie von gängigen Suchmaschinen bekannt ist, setzt Grokker als strukturierendes Kriterium ein Clustering ein, das bis auf Dokumentenebene geht. Die einzelnen Cluster werden auf einer Clus-

ter Map in Form von Kreisen (Setlevel, Cluster), Kugeln (Setlevel, Untercluster) oder Rechtecken mit gefalteter Ecke (Documentlevel) dargestellt, die anhand von Zooming- und Mouse-Over-Funktionen detailliert betrachtet werden können.

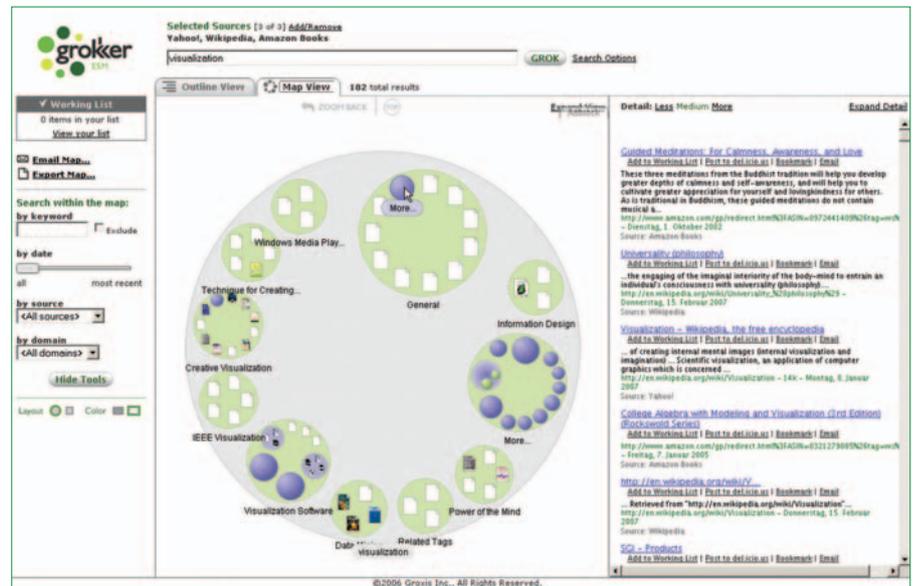


Abbildung 3: Clusterdarstellung bei Grokker zur Suchanfrage „Visualization“

**GROKKER Hintergrund**

**URL:** http://live.grokker.com/  
**Unternehmen:** Groxis, USA  
**Produkte:** Desktop, Enterprise und Web-Grokker  
**Funktionen:** Volltextsuche  
**Visualisierung:** Cluster-Map der Ergebnisse mit Zoom-Funktion (Map View), alternative Darstellung: in Listenform in Foldern organisiert (Outline View)

Folgende Gestaltungsmittel werden bei der Visualisierung eingesetzt:

- Farbe zur Abgrenzung zwischen den Clusterebenen
- Räumliche Größe der Cluster-Kugeln und Dokument-Blätter zur Repräsentation der Dokumentenmenge und -relevanz
- Tagging für die Bezeichnung der Themenkategorie der Cluster

Folgende Zusatzfunktionen bietet Grokker außerdem:

**Workinglist:** Diese ermöglicht das Speichern von Suchergebnissen auf einer personalisierten Arbeitsliste sowie den Export der Ergebniskarte.

**Filterfunktion:** Innerhalb der Ergebnismenge kann eine Einschränkung nach zeitlichen Aspekten, nach Domain oder Quelle ebenso vorgenommen werden, wie die Suche und das Ausschließen von Schlagworten innerhalb der Ergebnismenge.

**3.2.2 Mooter**

Mooter ist als eine der ersten klassischen Clustersuchmaschinen seit 2001 auf dem Markt.

Die Ergebnisse einer Suchanfrage bei Mooter werden zunächst anhand einer spinnenförmigen Clusterdarstellung in die wichtigsten sieben Kategorien aufgeteilt, jede Kategorie durch eine Kugel dargestellt, die durch Linien in der Mitte verbunden sind. Ein weiterführender Link ermöglicht das Laden einer nächsten Cluster Page mit den jeweils folgenden

1 Eine weitere Suchmaschine in diesem Bereich ist Searchmapr (www.searchmapr.com/, [02.02.2007]), die mit einer Treemap-Abbildung arbeitet. Sie wird in diesem Beitrag nicht berücksichtigt, da die Stabilität stark eingeschränkt ist und Suchanfragen nur unzuverlässig ausgeführt werden.

sieben Kategorien. Bei Klick auf einen der Cluster wird eine Neusortierung der Ergebnisse vorgenommen und unterhalb der Visualisierung werden die Treffer anhand einer listenbasierten Darstellung angezeigt.

Über das Icon „next Clusters“ wird eine neue Spinnendarstellung geladen mit sieben weiteren Kategorien.

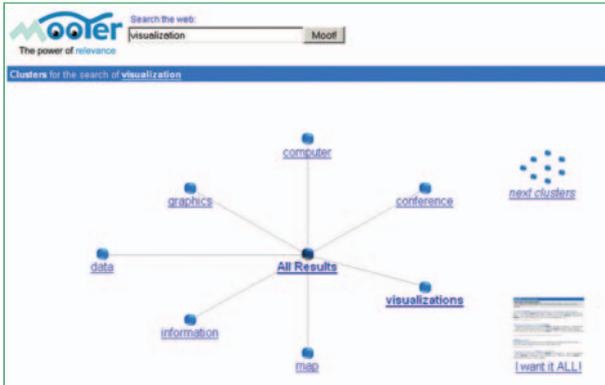


Abbildung 4: Clusterdarstellung bei Mooter zur Suchanfrage „Visualization“

#### MOOTER Hintergrund

**URL:** [www.mooter.com/](http://www.mooter.com/)  
**Unternehmen:** Mooter  
**Produkte:** Websearch  
**Funktionen:** Volltextsuche, Clustering  
**Visualisierung:** Cluster Page der Ergebnisse

#### 3.2.3 Ujiko

Ujiko ist eine Web-Suchmaschine, die Clustering über eine Farbcodierung ermöglicht, die durch den Nutzer gesteuert werden kann. Daneben ist eine Personalisierung durch diverse Einstellungen möglich.

Die Suchmaschine wurde von Kartoo entwickelt und 2005 auf dem Markt positioniert. Neben der Visualisierung in einer so genannten Circular Map integriert sie ein Anreizsystem, das bei gesteigertem Gebrauch der Suchmaschine durch einen Nutzer Zusatzfunktionen freischaltet (für je zehn Klicks in der Suchmaschine wird eine neue Funktion aktiviert). Weiterhin können Suchergebnisse durch den Nutzer thematisch umgruppiert, einer Ordnerstruktur zugewiesen, bewertet oder aus der Ergebnismenge gefiltert werden. Durch eine Auswahl verschiedener „Skins“ (aus dem Englischen für Verkleidung) kann die Darstellung der Suchmaschine optisch verändert werden.

Die Suchergebnisse werden anhand einer Circular Map sternförmig dargestellt und mittels Farbcodierung lassen sich Begriffskluster hervorheben.

Analog zu Google werden die Treffer auf verschiedene Seiten aufgeteilt, wobei Ujiko auf einer Ebene bis zu zwölf Ergebnisse darstellt.

Ujiko bietet diverse Zusatzfunktionen wie beispielsweise

- Filtering nach Schlagworten in den Treffern
- Filtering nach selbst definierten Begriffen
- Ordnerstrukturen
- Bewertung von Suchergebnissen
- Zusätzliche Suchfunktionen bei Kartoo und Yahoo

#### 3.2.4 meX-Search

meX-Search ist eine experimentelle, nicht kommerzielle Metasuchmaschine, die von Karsten Knorr 2004 im Rahmen einer Diplomarbeit erstellt wurde und die Ergebnisse von Yahoo in thematischen Clustern auf Basis der Carrot<sup>2</sup> Clustering Engine<sup>2</sup> sowie des Lingo-Algorithmus<sup>3</sup> zusammenfasst, um sie als sogenannte meX maps zu visualisieren. Die meX maps ordnen in einer Grotter-ähnlichen Darstellung kreisförmige Cluster um einen menschlichen Avatar an. Die Größe der einzelnen, mit Labels versehenen Kreise repräsentiert die Anzahl

auf eines der Clusterlabel, so wird der jeweilige Terminus als Suchbegriff übernommen und eine neue Anfrage gestartet. Die unterschiedliche Sättigung der Hintergrundfarbe der einzelnen Kreise steht dagegen in keinem Bezug zur Ergebnispräsentation. Als irrelevant betrachtete Cluster lassen sich zudem per Mausklick schließen (s. Abb. 6).

#### 3.2.5 Quintura

Quintura ist eine im November 2005 in den Markt eingeführte Websuchmaschine, die wahlweise Daten von Google, Yahoo, MSN Search oder Technorati auswertet. Die eingegebenen Suchtermini werden in Kombination mit verwandten Begriffen als Begriffswolken visualisiert, wobei die Schriftgröße die Anzahl der dahinterstehenden Treffer repräsentiert. Per Mouse-Over-Effekt werden zusätzliche, für den jeweiligen Terminus relevante Tags eingeblendet. Der User kann durch Anklicken weiterer Begriffe seine Suchanfrage verfeinern oder sie umgekehrt durch das Abwählen einzelner Termini erweitern.



Abbildung 5: farbcodierte Clusterdarstellung bei Ujiko zur „Visualization“

#### UJIKO Hintergrund

**URL:** [www.ujiko.com/](http://www.ujiko.com/)  
**Unternehmen:** Kartoo  
**Produkte:** Websearch  
**Funktionen:** Volltextsuche, Clustering, Filtering, Historyfunktion, Ordnerfunktion, Mailfunktion, Speicherfunktion  
**Visualisierung:** Circular Map, Farbcodierung

der dahinter verborgenen Links, die nur auszugswise als weiße Kästchen in den jeweiligen Clustern dargestellt werden. Per Mouse-Over-Effekt erhält man Titel, Url und einen kurzen Textausschnitt der verknüpften Seite. Klickt man dagegen

<sup>2</sup> <http://project.carrot2.org> [02.02.2007]

<sup>3</sup> [www.cs.put.poznan.pl/dweiss/site/publications/download/iipwm-osinski-weiss-stefanowski-2004-lingo.pdf](http://www.cs.put.poznan.pl/dweiss/site/publications/download/iipwm-osinski-weiss-stefanowski-2004-lingo.pdf) [02.02.2007]



Abbildung 6: Clusterdarstellung bei meX-Search zur Suchanfrage „Visualization“ mit Video-Suchagent

**meX-Search Hintergrund**

URL: [http://www.mex-search.com/mex\\_full.html](http://www.mex-search.com/mex_full.html)  
 Unternehmen: meX-search.com  
 Produkte: Websearch  
 Funktionen: Metasuche, Clustering auf Textbasis,  
 Visualisierung: meX map, Flächencodierung

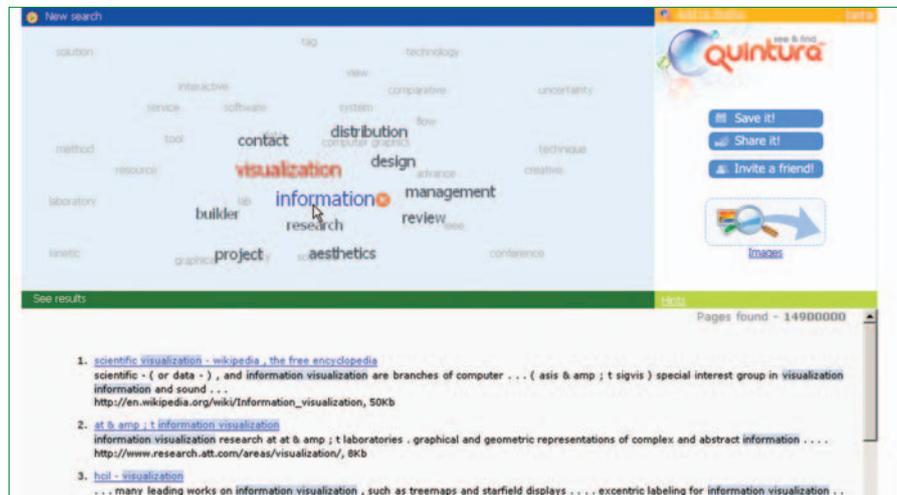


Abbildung 7: Clusterdarstellung bei Quintura zur Suchanfrage „Information Visualization“

**Quintura Hintergrund**

URL: [www.quintura.com](http://www.quintura.com)  
 Unternehmen: Quintura Inc.  
 Produkte: Websearch  
 Funktionen: Clustering auf Textbasis  
 Visualisierung: Begriffswolken, Farbcodierung, Kodierung auf Schriftgröße und Position

**3.3 Erweitere Relationenvisualisierung**  
 Die Relationenvisualisierung bei der Darstellung von Suchmaschinenergebnissen

dient dazu, die inhaltlichen Zusammenhänge zwischen den ermittelten Treffern anzuzeigen.

**3.3.1 Kartoo**

Die Metasuchmaschine *Kartoo* ist eines der Produkte des gleichnamigen französischen Unternehmens, das 2001 erfolgreich auf dem Markt gestartet ist und mittlerweile in verschiedenen Sprachversionen zur Verfügung steht. Es handelt sich wiederum um eine Metasuchmaschine, die die Treffer in einer Trefferkarte visualisiert.

Treffer werden nach Stichworten spezifiziert und unterteilt und als Höhenzüge dargestellt, Dokumente werden durch Blätter symbolisiert, die in ihrer Größe in Abhängigkeit von der Relevanz variieren können. Je mehr gemeinsame Themen zwei Treffer haben, desto näher rücken sie zusammen. Die Themenstichworte werden dazu verwendet, Beziehungen zwischen den Treffern anzuzeigen.

Die Besonderheit von Kartoo ist das topographische Interface, das sich einer Landkartenmetapher bedient. Kartoo bezeichnet sich selbst als „World leader in cartographic visualization“, kann mit beliebiger Indexierungssoftware zusammen eingesetzt werden und verarbeitet sowohl strukturierte als auch unstrukturierte Daten. Das System ist in verschiedenen Versionen verfügbar: als Suchfunktion für die Web-Site (SiteBox), für das Intranet (Suchsystem für bis zu zehn Quellen parallel), für Datenbanken und das Dokumentenmanagement.

**3.3.2 Webbrain**

Im Kern handelt es sich bei *Webbrain* um ein System zur Organisation von Informationen; es erlaubt, diese zu vernetzen, Beziehungen aufzuzeigen, über die Grenzen einzelner Informationssysteme hinweg: „TheBrain illustrates how information is related, provides a visual context for documents and data, and offers a framework for collaboration.“ Die Struktur ist dabei nicht hierarchisch, sondern assoziativ, so dass es fast keine Einschränkungen hinsichtlich der Verknüpfung gibt. Das entscheidende Charakteristikum ist dabei, dass die Assoziationen selbst angelegt werden müssen – die Produkte von The Brain erlauben eine visuelle Organisation, erstellen diese aber nicht automatisch. The Brain spricht von „Thoughts“; das kann potentiell alles sein: Dokumente, Web-Seiten, verschiedene Informationstypen. Diese werden miteinander verbunden in einer Weise, die der Denkweise des Menschen entsprechen. The Brain verspricht eine neue Arbeitsoberfläche zu sein.

The Brain ist in verschiedenen Versionen verfügbar:

1. Personal Brain: Integrierter Informationszugriff auf dem eigenen Computer: Informationen aus dem Internet und allen persönlichen Anwendungen werden in einer graphischen Karte (Netzwerk) miteinander verbunden.



Abbildung 8: Topographische Kartendarstellung bei Kartoo zur Suchanfrage „Visualization“

### KARTOO Hintergrund

- URL:** /www.kartoo.com/  
**Unternehmen:** KartOO SA, gegründet 2001, Frankreich  
**Produkte:** Kartoo Metasuche, Kartoo Intranet, Kartoo Portal  
**Funktionen:** Indexierung (unterschiedliche Formate möglich), Crawling; Suche (in heterogenen, verteilten Quellen); Visualisierung: Topographische 2D-Ergebniskarten  
**Visualisierung:** Darstellung der Suchergebnisse in einer Karte (verschiedene Datenbestände möglich), die bearbeitet und gespeichert werden kann

- Webbrain: Setzt auf dem Open Directory Project auf und erlaubt sowohl eine Suche in diesem, als auch eine Navigation (Browsing).
- Enterprise Knowledge Platform: Zusammenführung verschiedener Informationsquellen in Unternehmen, Einsatz als Kollaborationswerkzeug mit entsprechenden Funktionalitäten.

Site Brain ist ein Navigationssystem für die eigene Web-Site oder das Intranet (nur einsetzbar in Verbindung mit einem anderen Produkt von Brain Technologies).

Nicht selten werden die Produkte von The Brain auch als Mindmapping-Produkte bezeichnet.

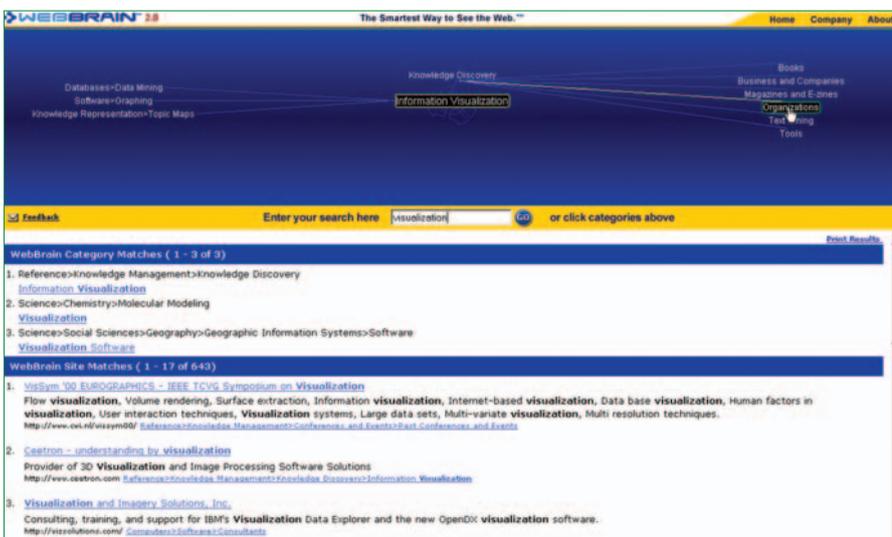


Abbildung 9: Netzwerk-Darstellung bei Webbrain zur Suchanfrage „Visualization“

### WEBBRAIN Hintergrund

- URL:** www.webbrain.com/  
**Unternehmen:** TheBrain Technologies Corporation  
**Produkte:** The Brain (Enterprise Knowledge Platform), Personal Brain, Web Brain, Site Brain  
**Funktionen:** Organisation: Erstellen des Assoziationsnetzes; Navigation im Netz; Suche (im Netz der Thoughts)  
**Visualisierung:** Netzwerk-Darstellung der enthaltenen Konzepte & Beziehungen

## 10 Dinge, mit denen wir uns auskennen

GUI, CMS, Podcast, Web 2.0, AJAX, XHTML, TYPO3, BITV, Load Balancing, IPTV

### 3-point concepts

Agentur für Kommunikation



3-point concepts GmbH  
 Chausseestraße 56  
 10115 Berlin



Tel.: +49-(0)30-28 51 98-0  
 Fax: +49-(0)30-28 51 98-28



info@3-point.de  
 www.3-point.de

### 3.4 Vergleich der Systeme auf Basis der Systematisierung von Visualisierungsmethodiken

Alle beschriebenen Systeme wurden unter Verwendung der in Kapitel 2 aufgestellten Systematik der Visualisierungsmethodiken untersucht. In Tabelle 1 wird das Ergebnis zusammenfassend dargestellt, indem für jedes einzelne System aufgezeigt wird, welche der unterschiedlichen Kriterien erfüllt werden.

Tabelle 1: Einordnung der betrachteten Systeme in die Klassifikation der Visualisierungsmethoden

		Dogpile	Langreiter	Mooter	Grokker	Kartoo	Ujiko	Webbrain	Mex	Quint	
Einsatzort	set level				X	X			X	X	
	document level	X	X	X	X	X	X		X		
site level											
Primärstruktur	Makrolevel	Hierarchisch						X			
		netzwerkartig		X			X				
	Linear						X	X			
	Cluster			X	X	X	X		X	X	
Mikrolevel	Nominal	X	X	X			X	X			
	Ordinal				X	X			X		
	quantitativ										
Skalierbarkeit	Volumen	small-scale									
		medium-scale	X	X		X				X	
		large-scale			X		X				
		flexible-scale						X	X		X
	Verdichtungsgrad	aggregiert				X		X		X	X
explizit		X	X	X	X	X	X	X	X		
detailliert											
Interaktivität	Interaktive	overview								X	
		zoom				X					
	Vollständigkeit	filter				X	X	X			X
		details-on-demand	X	X		X	X	X		X	X
		relate	X	X	X	X	X	X	X		
		history					X	X		X	
		extract				X	X	X			
		configure	X			X	X	X		X	X
Ranking	Rankingalgorithmen hinterlegt	X	X	X	X			X			
	Ranking in Visualisierung dargestellt		X								
	Keine Auskunft möglich		X							X	

Dabei ist relativ klar zu erkennen, dass die bekannteren und in Fachbeiträgen viel diskutierten Systeme Grokker und Kartoo auch die differenziertesten Visualisierungstechniken verwenden. Das vom gleichen Hersteller wie Kartoo stammende System Ujiko weist ebenfalls eine Vielzahl von Kriterien auf, ist jedoch im Allgemeinen nicht so bekannt wie Kartoo selbst (bzw. Grokker), was sich vermutlich auf die spätere Markteinführung zurückführen lässt.

Die untersuchten Systeme lassen sich folglich bezüglich des Funktionsumfangs in zwei Gruppen aufteilen: Kartoo, Grokker und Ujiko einerseits, die restlichen weniger umfangreichen Systeme andererseits. Quintura schneidet, trotz einer Fülle an Interaktionsmöglichkeiten, im Vergleich zu anderen Systemen in der oben angestellten Bewertung eher schlecht ab, was jedoch auf dem Umstand beruht, dass bei Quintura die Ergebnisvisualisierung zur Nutzerführung hinsichtlich der Formulierung von Suchanfragen genutzt wird und nicht ausschließlich als Repräsentation der Treffermenge.

Die Systeme der ersten Gruppe besitzen alle eine Clustervisualisierung, deren Mehrwert gegenüber der text- und listenbasierten Ergebnisrepräsentation herkömmlicher Suchmaschinen zumindest theoretisch relativ einleuchtend ist.

Bei den weniger umfangreichen Systemen ist auffällig, dass sie bei den Kriterien Primärstruktur, Skalierbarkeit, Verdichtungsgrad und Interaktive Vollständigkeit größtenteils die gleichen Attri-

können. Denn obwohl alle Systeme eine Visualisierung der Ergebnismengen vornehmen, unterscheiden sich die Oberflächen massiv. Dem Nutzer erscheinen die Systeme jeweils in unterschiedlicher Aufmachung, er muss sich somit nicht nur von der gewohnten textbasierten Ergebnisrepräsentation auf eine visuelle Darstellung umstellen, sondern für jedes System einzeln den Funktionsumfang und die Interaktion neu erlernen. Hierzu wird er vermutlich nur dann bereit sein, wenn der mehrwertschaffende Aspekt, der durch eine ungewohnte Systemoberfläche mit Visualisierung erzielt wird, plausibel erscheint und sehr bald durch eigene Erfahrung nachvollziehbar wird.

## 4 Ausblick

Wie die Marktübersicht deutlich macht, ist die Entwicklung geeigneter visueller Suchsysteme zur intuitiv nachvollziehbaren Repräsentation von Suchergebnissen ein aktuelles Entwicklungs- und Forschungsfeld diverser Suchdienstanbieter im Internet. Die klassischen Formen der Listendarstellung können aufgrund ihrer Begrenztheit den Ansprüchen des knowledge discovery nicht im notwendigen Ausmaß gerecht werden. Die weiter zunehmende Datenflut verlangt nach verbesserten Formen der Interaktion zwischen dem Informationssuchenden und den Informationssystemen. Die Visualisierungsmethoden sind ein Erfolg versprechender Ansatz. Den Ausblick hierzu sollen zwei weitere Visualisierungssysteme liefern, die sich jedoch auf spezielle Bereiche konzentrieren und dadurch mächtigere Formen von Informationspräsentation ermöglichen. Im ersten Fall handelt es sich um den Dienst LivePlasma, der eine graphische Ergebnisvisualisierung im Bereich Film und Musik anbietet. Im zweiten Fall wird das Forschungsprojekt Visual Relations kurz beschrieben, bei dem versucht wird, die Suchergebnisse von Unternehmensdaten mit diversen Visualisierungsmethoden, u. a. auch mit einer Verknüpfung mit geographischen Informationssystemen (GIS), darzustellen.

### 4.1 LivePlasma

LivePlasma<sup>4</sup> ist ein Dienst, der die Suche nach Gruppen oder Künstlern, Filmtiteln, Schauspielern oder Regisseuren ermöglicht. Die zugehörige graphische Suchergebnisvisualisierung stellt dabei nicht nur eine Art Affinitätsdiagramm dar, das diverse Arten von Relationen bzw. Abhängigkeiten zum gesuchten Begriff visualisiert, sondern ermöglicht sie über interaktive Techniken die Exploration im Bereich der Musik- und Film-Welt.

Im Zentrum der Visualisierung steht ein Graph (Netz), der auf Basis gemeinsamer Kriterien zum Suchbegriff aufgebaut wird. Durch ähnliche Farbgebungen wird

4 <http://liveplasma.com>, Stand: 02.02.2007

zusätzlich die Zugehörigkeit zu „Künstlerfamilien“ verdeutlicht und durch die Größe der Darstellung die Popularität und der Einflussreichtum eines bestimmten Künstlers.

Interaktionstechniken ermöglichen das Zooming innerhalb des Netzwerks sowie über Thumbnails die direkte Verbindung zu einem entsprechenden Produkt in Form einer CD bzw. DVD bei amazon.com.

bearbeitet wird, greift die behandelte Thematik auf. Ziel des Projektes ist der Aufbau eines Information Retrieval-Systems, das heterogene Datenbestände sowie inhärente Beziehungen identifiziert und visualisiert. Die geeignete Visualisierung der zugehörigen „Relationen“ gehört, neben dem Nachweis der Dokumente selbst, zu den zentralen Fragestellungen innerhalb des Projektes. Im Fokus stehen vor allem intuitiv bedienbare Vi-

aus strukturierten bzw. unstrukturierten Unternehmensdatenbeständen mit räumlichen Daten aus Kartenmaterial nicht generell automatisierbar ist.

Soll solch eine GIS-Aufbereitung von Informationen aus Suchergebnismengen brauchbare Mehrwerte erzeugen, dann müssen die zugrunde liegenden Unternehmensdaten zum einen schwerpunktmäßig Informationen mit geographische Bezugspunkten beinhalten und zum anderen sollten diese vornehmlich in strukturierten Datenbeständen, wie zum Beispiel Datenbanken vorliegen, damit eine eindeutige Identifikation möglich ist. Dies ist offensichtlich bei denjenigen Branchen der Fall, bei denen geografische Daten ohnehin eine zentrale Rolle spielen, wie zum Beispiel bei Transport und Logistik, Reiseveranstaltern oder der Immobilienbranche.

Ein weiterer wichtiger Faktor bei der Wahl der Visualisierung ist die zu Grunde liegende Datenstruktur. Während auf eine Suchanfrage hin üblicherweise eine netzartige Datenstruktur zu erwarten ist, bildet die Treffermenge einer Relationensuche eine Hierarchie, ausgehend vom ausgewählten Datensatz. Demgemäß muss die jeweilige Benutzeroberfläche in der Lage sein, zwischen diesen Ansätzen zu wechseln.

Die hohe Varianz der Anzahl der zu visualisierenden Elemente stellt eine weitere Herausforderung dar. Zugriffsrechte auf Dateien, die Qualität der Suchanfrage sowie der Homogenitätsgrad der Daten können starken Einfluss hierauf ausüben, weshalb sich der Einsatz von Verzerrungstechniken wie „Fish Eye“-Views empfiehlt.

Abbildung 12 zeigt den ersten, vereinfachten Entwurf einer Benutzeroberfläche für das Projekt „Visual Relations“. Als Anfrageobjekt wird zunächst ein Datensatz als Ausgangspunkt gewählt, von dem ausgehend die Relationensuche zu verknüpften Datensätzen angestoßen wird. Die Darstellung orientiert sich grundsätzlich an der Metapher einer Zielscheibe, wobei die eingeblendeten Ringe lediglich der Orientierung dienen sollen und in ihrer Anzahl nicht zwingend den Grad der Relationen widerspiegeln. Der Dateityp ist hierbei dem Attribut Farbe zugewiesen. Per Mouse-over werden weitere Details zu den Treffern eingeblendet, weiterhin lässt sich anhand einer Linie nachverfolgen, wie die einzelnen Treffer zueinander in Relation stehen bzw. welche Dokumente unmittelbar miteinander verknüpft sind. Hierbei ist für eine bessere Übersichtlichkeit nicht immer der gesamte Relationengraph abgebildet, sondern lediglich die Relationen vom aktuell betrachteten Treffer ausgehend bis hin zum Anfrageobjekt. Geogra-

5 www.visualrelations.com [2.3.2007]

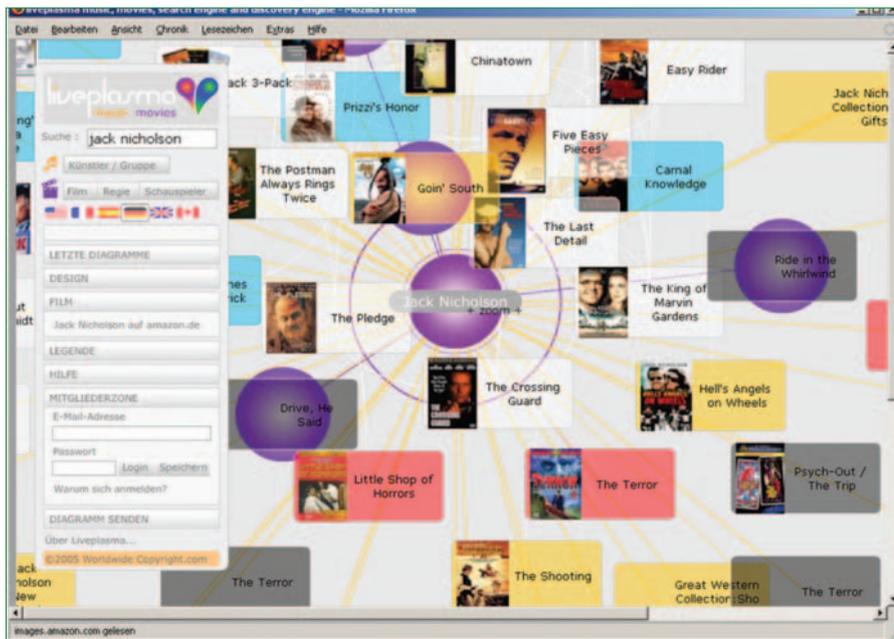


Abbildung 10: Benutzeroberfläche bei Liveplasma bei Anfrage „Jack Nicholson“

Abbildung 10 zeigt die Benutzeroberfläche von LivePlasma. Auf der linken Seite befindet sich die Sucheingabe mit der Schauspielersuchanfrage, hier nach „Jack Nicholson“. Die zugehörige Treffermenge wird im Hintergrund mit Knoten bzw. Kreisen, Verbindungen und Thumbnails angezeigt. Die Farbgebungen visualisieren verschiedene Filmrichtungen (Drama, Komödie, Horror etc.) und Relationen (Schauspieler in, Regisseur von etc.). Die zentrale Interaktionsmethode stellt das Diagramm in der Mitte dar (siehe auch Abb. 11). Ein Klick im Diagrammzentrum ermöglicht das weitere Zoom-in in die Abbildung, so dass nähere Einzelheiten bezüglich der Relationen zu den anderen Objekten ersichtlich werden. Der Klick auf die Zwischenebene ermöglicht das Zoom-out. Der Klick auf die äußerste Ebene oder außerhalb der Kreise verschiebt das Diagramm an die entsprechende Position.

Die direkte Auswahl eines anderen Objektes mittels Mausklick stellt dieses dann in das Zentrum des Interaktionsdiagramms.

#### 4.2 Projekt „Visual Relations“

Das Forschungsprojekt „Visual Relations“<sup>5</sup>, das aktuell an der Hochschule für Technik und Wirtschaft (HTW) Chur im Arbeitsbereich Informationswissenschaft

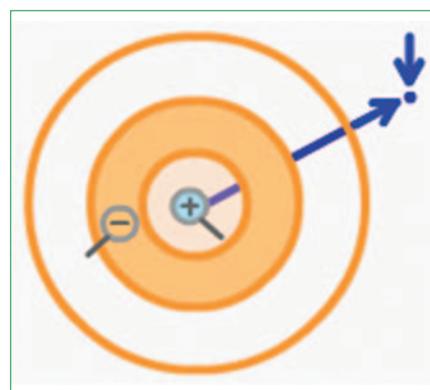


Abbildung 11: Prototyp der Visualisierung der Relationen

sualisierungen, wobei insbesondere der Mehrwert geografischer Visualisierungen sowie darauf basierender Metaphern untersucht werden soll. Darstellungen in Form von Land- und Straßenkarten oder 3D-Betrachtungen von Erdteilen (Geographische Informationssysteme – GIS) sind weit verbreitet und scheinen aufgrund ihrer Vertrautheit ohne hohen zusätzlichen Aufwand erlernbar. Zur Evaluation des Einsatzpotentials dieser GIS-Visualisierungsformen wurden zunächst unterschiedliche, auf realen Unternehmensdaten basierende Testkollektionen erarbeitet und analysiert. Die Analyse zeigte, dass eine Verknüpfung von Sachdaten

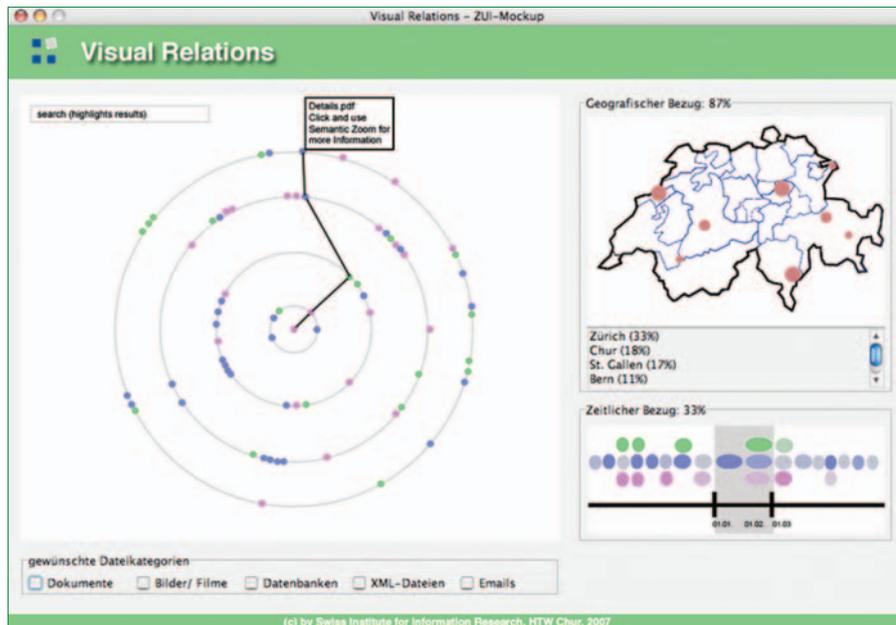


Abbildung 12: Erster Entwurf einer Benutzeroberfläche für das Projekt „Visual Relations“

fisch referenzierte Daten werden zusätzlich rechts oben auf einer Karte abgetragen, wobei die Größe der Kreise die Anzahl der Treffer an einem Ort repräsentiert, so dass Trends leicht auf einen Blick erkennbar sind. Der allfällige Zeitbezug der Daten wird rechts unten mittels einer Fish-Eye-Darstellung visualisiert, die durch einen Zeitregler das Zoomen bis auf einzelne Daten hin ermöglicht. Die einzelnen Bereiche interagieren per Linking und Brushing miteinander.

Um sehr große Datenmengen visualisieren zu können, wurde der Einsatz von Fish-Eye-Techniken im Bereich der Relationenvisualisierung als sinnvoll erachtet. Ab einem gewissen Grad an Relationen sollen diese allerdings, um eine Unübersichtlichkeit zu vermeiden, nicht mehr direkt präsentiert werden. Pfeile entsprechender Größe oder geeignete Landmarken dienen dem Nutzer an dieser Stelle als Wegweiser, in welche Richtung er sich mittels Panning weiterbewegen kann und welche Treffermenge in dieser Richtung zu erwarten ist.

### Danksagung

Das Projekt „Visual Relations“ wird durch die schweizerische Kommission für Technologie und Innovation (KTI) gefördert und mitfinanziert. Die KTI ist die Förderagentur für Innovation des Bundes. Die Autoren danken der KTI für die finanzielle Unterstützung.

### Literatur

Bauer, L.; Herget, J.; Hierl, S. (2007): Conceptual challenges for the evaluation of digital repositories with multiple access options. A case study. Conference ICSD, Bangalore February 2007 (to be published)

Card, S. K.; Mackinlay, J.D.; Shneiderman, B. (1999): Readings in information visualization: Using vision to think. San Francisco: Morgan Kaufmann Publishers, 1999. ISBN 1-55860-533-9

Hierl, S. (2006): Konzeption eines Evaluationsframeworks für Information Retrieval Systeme mit integrierter Visualisierungskomponente. Master Thesis des Fachbereichs Wirtschaftsinformatik an der Hochschule Liechtenstein: Liechtenstein, 2006

Mann, T.M. (2002): Visualization of search results from the World Wide Web, Dissertation Universität Konstanz: Konstanz, 2002

Schumann H.; Müller, W. (2000): Visualisierung: Grundlagen und allgemeine Methoden. Berlin: Springer, 2000

Spink, A.; Jansen, J. et al. (2005): Different Engines, Different Results. Web Searchers Not Always Finding What They're Looking for Online. A Research Study by Dogpile.com, 2005. <http://comparsearchengines.dogpile.com/OverlapAnalysis.pdf>

Wild, F. (2005): Visuelle Verfahren im Information Retrieval. In: Information. Wissenschaft und Praxis 56 (2005) 1

Wild, F. (2004): Visuelle Verfahren im Information Retrieval: intelligentes Sehen als Grundlage der Informationserschließung. Magisterarbeit in der Philosophischen Fakultät IV (Informationswissenschaft) der Universität Regensburg: Regensburg, 2004

### Referenzierte Systeme

Dogpile	<a href="http://missingpieces.dogpile.com/">http://missingpieces.dogpile.com/</a>
Grokker	<a href="http://live.grokker.com/">http://live.grokker.com/</a>
Kartoo	<a href="http://www.kartoo.com/">www.kartoo.com/</a>
Langreiter	<a href="http://www.langreiter.com/exec/yahoo-vs-google.html">www.langreiter.com/exec/yahoo-vs-google.html</a>
Liveplasma	<a href="http://liveplasma.com">http://liveplasma.com</a>
meX-search	<a href="http://www.mex-search.com/mex_full.html">www.mex-search.com/mex_full.html</a>
Mooter	<a href="http://www.mooter.com/">www.mooter.com/</a>
Quintura	<a href="http://www.quintura.com/">www.quintura.com/</a>
Searchmapr	<a href="http://www.searchmapr.com/">www.searchmapr.com/</a>
Ujiko	<a href="http://www.ujiko.com/">www.ujiko.com/</a>
Visual Relations	<a href="http://www.visualrelations.com">www.visualrelations.com</a>
Webbrain	<a href="http://www.webbrain.com/">www.webbrain.com/</a>

Graphische Darstellung, Rechercheergebnis, Marktübersicht, Software, Suchmaschine

### DIE AUTOREN

#### Prof. Dr. Bernard Bekavac



Bereich Information Engineering, Studiengang Informationswissenschaft, Hochschule für Technik und Wirtschaft Chur. Aktuelle Forschungsthemen: Informationsvisualisierung, Information Retrieval, Port

taltechnologien

[bernard.bekavac@fh-htwchur.ch](mailto:bernard.bekavac@fh-htwchur.ch)

#### Prof. Dr. Josef Herget



Professor für Informationswissenschaft und Leiter des Arbeitsbereiches Informationswissenschaft an der Hochschule für Technik und Wirtschaft Chur. Visiting Professor der Donau-Universität Krems.

Forschungsschwerpunkte: Informationsmanagement, Informationssysteme, Wissensmanagement

[josef.herget@fh-htwchur.ch](mailto:josef.herget@fh-htwchur.ch)

#### Sonja Hierl, MSc BIS, Dipl. Informationsspezialistin (FH)



Projektleiterin am Swiss Institute for Information Research der Hochschule für Technik und Wirtschaft Chur, derzeit Promotion an der Universität Regensburg. Forschungsschwerpunkte: Neue Konzepte im

Information Retrieval, Evaluation visueller Systeme

[sonja.hierl@fh-htwchur.ch](mailto:sonja.hierl@fh-htwchur.ch)

#### Sonja Öttl, M.A.



Projektmitarbeiterin am Swiss Institute for Information Research der Hochschule für Technik und Wirtschaft Chur, derzeit Masterstudium für Information Engineering an der Universität Konstanz. For

schungsschwerpunkte: Visualisierung und Mensch-Maschine-Interaktion

[sonja.oetli@fh-htwchur.ch](mailto:sonja.oetli@fh-htwchur.ch)

HTW Chur  
Informationswissenschaft  
Ringstrasse, CH - 7000 Chur

# Kommunikation im Internet: Die Suchfunktion im Praxistest

Martin Braschler, Josef Herget, Joachim Pfister, Peter Schäuble, Markus Steinbach und Jürg Stuker, Schweiz

*Die Ausgangslage für Suchfunktionen auf Unternehmens-Websites stellt sich grundlegend anders dar als bei Internet-Suchdiensten. Nicht die möglichst breite Erschließung von Informationen vieler verschiedener Websites, sondern der erfolgreiche Dialog mit den Nutzern steht im Vordergrund. Das Gelingen dieser Kommunikation wird in diesem Beitrag untersucht. Von 56 Schweizer Unternehmen und Behörden wurde jeweils die Suchfunktion der Website untersucht. Dabei gruppierten sich die insgesamt 74 Einzeltests in die vier Hauptkriterien: Suchindex, Vergleich von Anfragen und Dokumenten, Benutzerinteraktion und Suchergebnisse. Die Resultate zeigen großes Potential für Verbesserungen. Dabei wurden kein Ranking der Websites und keine Auszeichnung im Rahmen der Studie beabsichtigt. Vielmehr wurde auf Basis der Ergebnisse eine Einteilung in gute, mittlere und schlechte Websites für die einzelnen Kriterien durchgeführt. Die Analyse zeigt, dass keine Website in sämtlichen getesteten Kriterien sehr gut abschneidet. Die Studie schließt mit verschiedenen Empfehlungen, die dazu führen können, die Qualität der Suchfunktion wesentlich zu verbessern. Hierbei wird unter anderem empfohlen, der Aktualität und Vollständigkeit des Suchindex besondere Beachtung zu schenken, Metadaten der Dokumentenbasis zu pflegen, bewährte Standards für die Benutzerinteraktion zu beachten und Trefferlisten sowie potenzielle Suchanfragen auch auf eine mögliche Nutzersicht hin zu optimieren.*

## **Evaluation of site search as a customer communication tool for corporate websites**

*Website search differs from search engines on the internet – it is not the main goal to comprise as much information from different websites as possible, but to satisfy user needs on corporate websites. This contribution deals with this task: How do search functions satisfy customer needs, does the website search contribute to a successful dialog with the customer? In this study 56 corporate websites from Switzerland and their search function have been analysed. This has been fulfilled by 74 tests addressing the following four main criteria: search index, matching of queries and documents, user interaction and search results. The results of this study indicate significant potential for improvement. It was not aimed to provide a ranking of the websites or to grant awards, however, a classification in good, average and bad websites in respect of fulfilling the different criteria was executed. No one website was able to deliver superior results in all criteria tested. The study concludes with several recommendations on how to improve the quality of the search function, for example to emphasize the completeness and up-to-dateness of the search index, to care about the meta data of the document basis, to respect standards in human-computer-interaction and to optimize the result output according to the user needs.*

## **1 Websites als Kommunikationsinstrument**

Suchdienste stellen heute im Internet das zentrale Werkzeug zur Informationsbeschaffung dar. Diese horizontalen Internet-Suchdienste wie Google und Yahoo eignen sich für die generelle Suche innerhalb des gesamten Internets.

Auch die Kommunikation zwischen Kunden und Organisationen aller Art erfolgt zunehmend über das Internet. Websites von Unternehmen entwickeln sich zu Kommunikationsdrehscheiben und sind häufig die erste Anlaufstelle, um Informationen über Unternehmen, Angebote, Produkte oder Ansprechpartner zu suchen – und hoffentlich zu finden. Orientierungshilfe liefert vorrangig die Navigation über die Verlinkungen innerhalb der Websites, zumeist werden zusätzlich auch eine Sitemap und eine integrierte Suchfunktion angeboten. Doch wie gut ist diese vertikale Suche geeignet, dem Kunden auch die gewünschte Information zu liefern?

Aus der Benutzerforschung ist schon seit langem das Phänomen bekannt, dass der Informationssuchende häufig nicht einmal in der Lage ist, sein Informationsproblem exakt zu beschreiben bzw. sein Anliegen genau zu formulieren. Dieses Phänomen wird als „anomalous state of knowledge“ bezeichnet. Problemverstärkend stellt sich zudem folgende Frage: Wissen wir denn überhaupt, was auf eine Suchanfrage hin dem Informationssuchenden als relevantes Ergebnis präsentiert wird? Oder wird dies einem zufälligen – uns zumeist auch unbekanntem – Algorithmus der Suchmaschine überlassen? Der Erfolg der Kommunikation, also ob der Kunde die richtige Information tatsächlich erhält, entzieht sich weitestgehend der Kontrolle des Unternehmens!

Die Betreiber von immer umfassender und komplexer werdenden Websites müssen sich schlussendlich fragen lassen, ob sie zum einen überhaupt wissen, was die Interessenten suchen und zum anderen, ob das, was die Suchfunktion als Resultat ermittelt auch das ist, was wirklich kommuniziert werden soll. Die Qualität der Beantwortung der Suchanfrage, darum geht es schließlich, bleibt also häufig dem Zufall überlassen – eine unbefriedigende Situation sowohl für Kunden als auch für Anbieter. Anders als bei persönlichen oder schriftlichen Kommunikationssituationen, bei denen wir selbstverständlich großen Wert auf eine konsistente und angemessene Beantwortung von Anfragen legen, scheint dies für die Kommunikation im Internet nicht zu gelten. Dort ist letztendlich der Suchende für den Erfolg seiner Suchfrage selbst verantwortlich.

Die diesem Beitrag zu Grund liegende Studie „Evaluation der Suchfunktion von Schweizer Unternehmens-Websites“ (Braschler et al. 2006) versucht nun folgende Fragen zu klären: Wie gut gelingt es den Nutzern mit der Site-Suche vorhandene Informationen zu ermitteln? Wo gibt es Problembereiche, die eine erfolgreiche Kommunikation stören bzw. verhindern?

Die Bedeutung der Suchfunktion für eine Website wird beispielsweise in Untersuchungen von Forrester Research herausgestellt (Cremers, 2006). Besonders wenn nach spezifischen Informationen auf unbekanntem Websites gesucht wird, verlassen sich die Nutzer mehr auf die Suchfunktion der Website, als dass

sie sich selbst auf den Sites die gesuchten Informationen navigierend erarbeiten (Abbildung 1). Zudem zeigen die Ergebnisse der Forrester-Umfrage (Cremers, 2006), dass durchschnittlich 13 Prozent der Nutzer zu einer anderen Site wechseln, wenn sie ihr Informationsbedürfnis nicht sofort gestillt sehen.

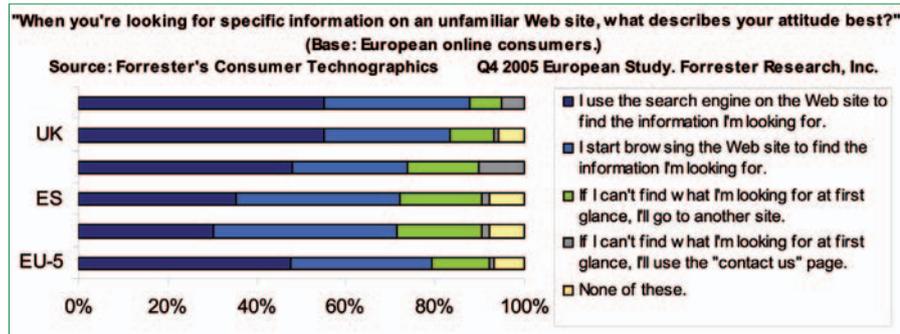


Abbildung 1: Welche Zugriffsmöglichkeit bevorzugen Nutzer auf unbekanntem Websites?

Die Suchfunktion auf einer Website ist aus Nutzersicht folglich der zentrale Anknüpfungspunkt zur Ermittlung von Informationen aus einem Web-Angebot – besonders auch angesichts der stetig wachsenden Informationsmenge (Lyman/Varian 2003). Die Anbieter von Informationsangeboten im Web müssen dieser Entwicklung Rechnung tragen, indem sie leistungsfähige und den Erwartungen der Nutzer entsprechende Such-Tools einsetzen.

Zur Diagnose der Leistungsfähigkeit der Suchfunktion wurde in enger Zusammenarbeit der Projektpartner Eurospider, Hochschule für Technik und Wirtschaft (HTW) Chur, namics und der Zürcher Hochschule Winterthur ein Evaluations-Raster erarbeitet. Die vorliegende Studie fasst die Ergebnisse der ersten Evaluationsrunde zusammen und gibt einen generalisierten Überblick, wo die Stärken und Schwächen der getesteten Site-Suchen liegen. Sie soll dazu anregen, den Stellenwert der Suchfunktion auf einer Website neu zu überdenken und die Suche als ein Mittel zur effektiven Kommunikation zwischen Anbieter und Nutzer wahrzunehmen.

**1.1 Ausgangspunkt und Fokus der Studie**

Die Anforderungen an die Suchfunktion von Unternehmens-Websites sind grundlegend anders als die Anforderungen an Internet-Suchdienste. Während letztere einen „breiten“ Informationszugang zu einer sehr großen Anzahl Websites, unabhängig der einzelnen Inhalte, anbieten (oft als „horizontale Suche“ bezeichnet), bewegen sich die in dieser Studie betrachteten, individuellen Unternehmens-Websites in einem klar abgegrenzten Gebiet, in welchem gezielt Information zur Verfügung gestellt wird (auch „vertikale Suche“ genannt). Während also Internet-Suchdienste zwischen unzähligen Anbie-

tern gleichartiger oder widersprüchlicher Information gewichten müssen, liegt die Herausforderung für Site-Suchen anders: die Nutzer sind möglichst ohne Umweg auf die gewünschte, meist sehr spezifische Information zu führen. Die Nutzer werden auch hier zunehmend anspruchsvoller, sie erwarten beispielsweise die

Möglichkeiten, die Stichwortsuche mit kontrolliertem Navigieren zu kombinieren (Alvarado et al. 2003).

Innerhalb der Studie wird die Suchfunktion einer Unternehmens-Website („Site-Suche“) als Retrieval-Applikation verstanden, die aus den Komponenten der Suchapplikation, der Konfiguration und dem bereitgestellten Informationsangebot aufgebaut ist. Die Studie stellt dabei gezielt die Bedürfnisse der Nutzer, nicht einzelne technische Funktionen, in den Vordergrund und betrachtet dabei losgelöst von

Tabelle 1: Schematischer Aufbau des Evaluations-Rasters

Suchindex	Vollständigkeit des Suchindex
	Aktualität des Suchindex
	Anfrage- und Dokumentenererschließung
Anfragen- und Dokumentenvergleich	Anfrageausführung
	Ausdrucksstärke der Abfragesprache
	Metadatenqualität
Benutzerinteraktion	Präsentation Trefferliste
	Nutzerführung
	Performance
Suchergebnisse	Navigationsanfragen
	Informationsanfragen
	Faktenanfragen
	Allgemeine Fragen an die Testpersonen

Tabelle 2: Aufbau eines Subkriteriums

Anfragen- und Dokumentenvergleich	Anfrageausführung
	Tests:
	1) Auswertung von sehr langen Anfragen Ergebnis-Codierung: 0 = nicht gefunden, 1 = gefunden
	2) Können Phrasen mit Stoppwörtern gesucht werden Ergebnis-Codierung: 0 = Abbruch, 1 = nicht gefunden, 2 = gefunden
	3) Ist die Suche gegenüber Flexionen robust? Ergebnis-Codierung: 0 = nicht gefunden, 1 = gefunden

der eigentlichen Implementierung die Suchfunktion.

Mit der Studie wird kein Produktvergleich angestrebt, da die zur Realisierung der Suchfunktion verwendete Software nur mittelbar am gesamten Suchprozess beteiligt ist und in Verbund mit dem eigentlichen Informationsangebot sowie der Aufbereitung für die Unternehmens-Website nur einen Faktor für eine gute Site-Suche darstellt. Tatsächlich ist davon auszugehen, dass etliche der untersuchten Websites auf der gleichen Softwarebasis realisiert sind.

**2 Testanordnung und Durchführung**

Das verwendete Evaluationsraster ist in vier Cluster gegliedert, mit deren Hilfe die wesentlichen Dimensionen einer Site-suche abgedeckt werden. Diese Hauptkriterien werden jeweils durch mehrere Subkriterien und diese weiterhin durch weitere Aspekte tiefer differenziert, so dass insgesamt 47 Kriterien gemessen werden können (Tabelle 1). Diese Kriterien werden durch insgesamt 74 Einzeltests evaluiert (Tabelle 2). Das Evaluationsraster besteht überwiegend aus manuellen sowie einigen automatisiert durchgeführten Tests. Für jeden einzelnen Test wurde ein Testprotokoll formuliert, das Punkt für Punkt abgearbeitet werden musste.

Zur Illustration ein Beispiel: Die Aufgabenstellung an die Testpersonen für den dritten Test in der Rubrik „Anfragen- und Dokumentenvergleich“ lautet im Testprotokoll: „Aus bestehenden Dokumenten eine Textpassage mit ca. zehn Wörtern in das Anfragefeld kopieren (dabei bitte keine Phrasensuche verwenden) und überprüfen, ob die Dokumente gefunden wurden.“ Die Bewertung erfolgte mittels vorgegebener Antwortmöglichkeiten (Ergebnis-Codierung, vgl. Tabelle 2), deren Werte zur Ergebnisberechnung herangezogen wurden.

Das vollständige Evaluationsraster ist in einem Pre-Test zunächst im Hinblick auf seine Durchführbarkeit und die Reproduzierbarkeit der Tests bzw. der Testanweisungen überprüft worden und letztere sind gegebenenfalls präzisiert worden. So konnte der möglicherweise verzerrende

subjektive Eindruck der Testpersonen bestmöglich eliminiert werden. Nach dem Pre-Test erfolgte die eigentliche Evaluation der insgesamt 56 getesteten Websites. Der Gesamtaufwand für die Evaluation einer Site-Suche

lag bei vier bis sechs Stunden. Die Tests wurden „von außen“, also über das öffentliche Internet durchgeführt, die Betreiber der getesteten Sites waren folglich über den Testzeitpunkt nicht informiert. So konnten Anpassungen der Website-Betreiber an den getesteten Systemen ausgeschlossen werden. Zur Auswertung wurden die codierten Ergebnisse zunächst normalisiert. Der Wert der Subkriterien errechnete sich aus der gewichteten Summe der zugehörigen Testergebnisse. Die Hauptkriterien wurden wiederum aus den gewichteten Ergebnissen der Subkriterien berechnet. Die Beurteilung der Relevanz einer Information hängt von vielen Faktoren ab: dem Vor-/Hintergrundwissen, der Reihenfolge des Auffindens, dem sich wandelnden Informationsbedürfnis, den persönlichen Präferenzen, der Vollständigkeit der Antwort. Dies erschwert eine absolute Bewertung von Suchresultaten im wissenschaftlich objektiven Sinn. Zur Bewertung der Effektivität von Retrievalsystemen hat sich das sogenannte Cranfield-Paradigma durchgesetzt (Cleverdon, 1967). Dabei wird großer Wert auf den Vergleich zwischen verschiedenen Systemen gelegt. Die Erkenntnisse aus der Evaluation sind allgemein umso aussagekräftiger, je mehr Systeme in die Evaluation einbezogen worden sind (Voorhees, 2002). Diese Studie versucht diesem Umstand durch die Evaluierung von 56 verschiedenen Websites Rechnung zu tragen. Dies, gekoppelt mit einer großen Anzahl Einzeltests zu den einzelnen untersuchten Kriterien, erlaubt es, zuverlässige Aussagen, insbesondere hinsichtlich der typischen Performance von Site-Suchen, zu machen.

### 3 Ergebnisse der Studie

#### 3.1 Die Hauptkriterien im Überblick

Zunächst werden die Resultate in den vier Hauptkriterien (Suchindex, Vergleich von Anfragen und Dokumenten, Benutzerinteraktion, Suchergebnisse) dargestellt, um in den folgenden Kapiteln die einzelnen Kriterien genauer zu beschreiben (siehe Abbildung 2).

Es fällt bei Betrachtung des Boxplots (siehe Abbildung 1 und Tabelle 3) auf, dass die Kennwerte der vier Hauptkriterien deutlich verschieden sind. Während absolute Leistungswerte grundsätzlich vorsichtig einzuordnen sind, da sie maßgeblich durch die Ausgestaltung der ein-

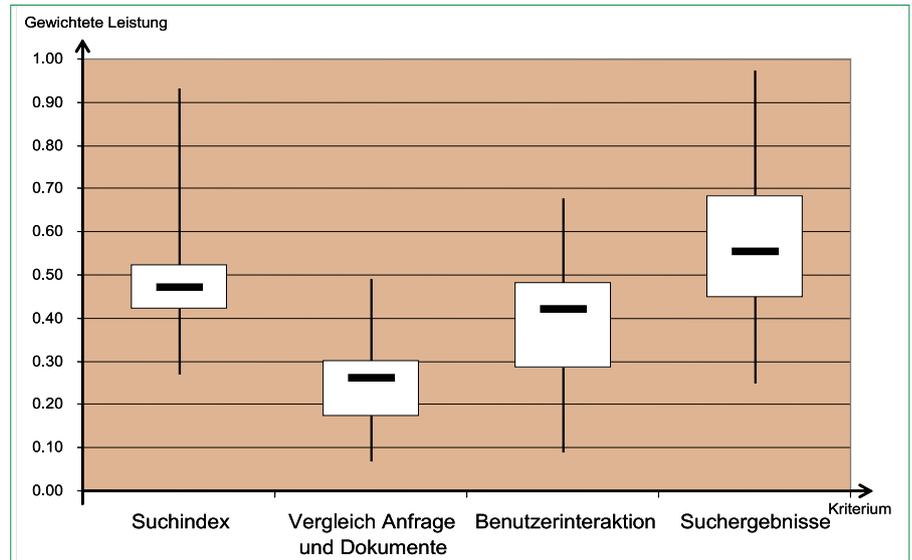


Abbildung 2: Boxplot der vier Hauptkriterien

zelnen Einzeltests beeinflusst sind, ist trotzdem augenfällig, dass die Leistungswerte im Bereich „Vergleich von Anfragen und Dokumenten“ durchgängig gegenüber den anderen Kriterien abfallen. Die mittlere Leistung ist bei 0.26 angesiedelt, während sie für die anderen drei Kriterien deutlich die Marke von 0.4 übertrifft. Die Streuung ist dabei am zweitgeringsten, das bedeutet, dass die Leistungswerte der großen Mehrheit der Site-Suchen in diesem Bereich als ungenügend betrachtet werden muss. Nicht zuletzt ist auffällig, dass auch die Maximalleistung, d.h. die beste getestete Unternehmens-Website, deutlich unter der theoretisch möglichen Bestleistung zurückbleibt. Dabei erreichen nur gerade drei weitere Site-Suchen mindestens 80% der Leistung dieses „Klassenbesten“. Es ist weiterhin festzuhalten, dass die vier Bestleistungen in den vier Hauptkriterien von vier unterschiedlichen Websites erreicht wurden. Das zeigt, dass ein großer Abstand nicht nur zur theoretischen Bestleistung besteht, sondern auch zu einer real heute erreichbaren Leistung (unter der Voraussetzung, dass man die Hauptkriterien als weitgehend unabhängig betrachtet. In der Realität sind natürlich allfällige Zielkonflikte zwischen den einzelnen Kriterien nicht völlig auszuschließen). Düster sieht es am anderen Ende der Skala aus: Mehr als die Hälfte der Websites konnte sich für kein Hauptkriterium in der Spitzengruppe etablieren. Hier ist davon auszugehen, dass die betreffenden Informationsanbieter große Probleme haben werden, mit der vorlie-

genden Site-Suche die Kommunikation mit den Nutzern erfolgreich zu gestalten. Die Testpersonen hatten im Rahmen der Evaluation ihre persönlichen Eindrücke gesondert als „allgemeiner Eindruck“ festzuhalten (Note zwischen 1 [„ungenügend“] und 10 [„exzellent“]). Die durchschnittliche Bewertung über alle Websites liegt bei 4.84 (Median 5) und damit analog zu den einzelnen Hauptkriterien weit vom theoretischen Maximum entfernt. Das Erlebnis der Testpersonen bei der Nutzung der Site-Suchen deckt sich also durchaus mit der objektiven Auswertung der Einzeltests.

Die Testpersonen machten des Weiteren rege von der Möglichkeit Gebrauch, einen kurzen Kommentar zum Umgang mit der Suchfunktion zu geben. Notiert wurden vor allem Probleme, welche aufgetreten waren. So vermissten die Testpersonen immer wieder „griffige“ Einschränkungsmöglichkeiten, die es erlauben würden, die Suche geeignet zu lenken.

#### 3.2 Suchindex

Das Kriterium Suchindex beurteilt die Qualität der Datengrundlage einer Site-Suche. Der Index, d.h. die technisch für den Zugriff optimierte Datengrundlage, beeinflusst Möglichkeiten, definiert aber auch Einschränkungen sowie qualitative Aspekte der darauf agierenden Suchapplikation. Die Qualität des Index wurde aufgrund folgender Subkriterien beurteilt: Vollständigkeit, Aktualität und Anfrage- und Dokumentenerschließung.

Bei der *Vollständigkeit des Suchindexes* wurde geprüft, ob die Erschließung der geprüften Website durch die Suchfunktion möglichst vollständig ist. Als Referenzsysteme zur Beurteilung von Stichproben dienen öffentliche Suchdienste wie Google und Yahoo. Des Weiteren wurden – insofern auf der Site vorhanden – lange, respektive komplexe URLs und binäre Dokumente in der Form von Adobe PDF-Dateien auf deren Indexsta-

Tabelle 3: Statistische Zahlenwerte der vier Hauptkriterien (gewichtete Leistung)

Statistiken	Suchindex	Vergleich Anfragen/Dokumente	Benutzer-Interaktion	Suchergebnisse
Maximum	0.93	0.49	0.68	0.97
3. Quartil	0.52	0.30	0.48	0.68
Median	0.47	0.26	0.42	0.55
1. Quartil	0.42	0.17	0.28	0.45
Minimum	0.27	0.07	0.09	0.25

tus getestet. Der Median von 0.44 ist insofern bemerkenswert, da die Anforderung alle öffentlich zugänglichen Seiten des Webangebotes über die eigene Suchfunktion zugänglich zu machen, von jedem Anbieter angestrebt wird. In diesem Sinne wird also eine Minimalanforderung nur ungenügend erfüllt. Zudem zeigt die maximale Leistung von 1.00, dass die Aufgabe lösbar ist. Schlussendlich sind alle benötigten Elemente im Wirkungsbereich des Betreibers. Ein aufgetretenes Problem ist der Umgang mit PDF-Dateien. Dabei ist vor allem der Umstand von Bedeutung, dass es viele verschiedene Typen und Konfigurationen von PDF gibt, welche sich bei der Extraktion von Text, so wie von der Suchfunktion benötigt, unterschiedlich verhalten. Der auffälligste Wert ist die *Aktualität des Suchindex*, bei welchem das Gros aller Probanden den Test nicht bestanden. Anforderung hierbei war es, dass bei inhaltlichen Änderungen auf der Website, der Index und somit die Suchfunktion diese sofort berücksichtigen. Zwei typische Schwächen waren einerseits neu angelegte oder geänderte Seiten, beispielsweise eine aktuelle Medienmitteilung, die mittels der Suchfunktion nicht aufgefunden werden konnte. Andererseits der Fall, dass in der Trefferliste Links zu Ziel Dokumenten aufgeführt wurden, die nicht mehr existierten. Bei einem Klick auf den Link wurde der User dann mit einer Fehlermeldung konfrontiert. Grund für diese Fehlerklasse ist eine mangelhafte Integration der Suchtechnologie mit dem Pflegeprozess der Inhalte. Der Autor oder spätestens das Content Management System sind über die Änderung einer Seite informiert – weshalb aber nicht die Suchfunktion?

Die *Anfrage- und Dokumentenerschließung* erreicht gute Werte. Sehr gut bestanden die getesteten Websites die Anforderung diakritische Zeichen wie ä oder à mittels Normalisierung korrekt zu behandeln.

### 3.3 Anfragen und Dokumente vergleichen

Ein Suchresultat wird berechnet, indem die Anfrage mit den erschlossenen Dokumenten verglichen wird. Es geht dabei nicht primär um die Suchqualität, sondern darum, ob die Auswertung einer Benutzeranfrage technisch sauber realisiert wurde, was natürlich auch einen Einfluss auf die Suchqualität hat. Das Vergleichen von Anfragen und Dokumenten wurde aufgrund folgender Subkriterien beurteilt: Auswertung der Anfrage, Ausdrucksstärke der Abfragesprache und Metadatenqualität.

Die Site-Suchen, die bei der *Anfrageausführung* schlecht abgeschnitten haben, sind eher Datenbank- als Retrieval-Applikationen. Ihnen fehlt die erforderliche Robustheit, um mit beliebigen unstrukturierten

Abfragen fertig zu werden. Wenn man beispielsweise eine Textpassage in einem Dokument in die Zwischenablage kopiert und anschließend in das Suchfeld einfügt, dann sollte das Dokument gefunden werden. Ein leeres Resultat ist nicht akzeptabel. Das Dokument sollte auch gefunden werden, wenn die Textpassage aus einer alten Version stammt, in der leicht andere Formulierungen (z.B. Plural anstatt Singular) verwendet werden.

Bei der *Ausdrucksstärke der Abfragesprache* fällt die große Streuung auf (Werte zwischen 0.00 und 0.92). Die schlechtesten Systeme unterstützen keine Operatoren, so dass in besonderen Fällen das Informationsbedürfnis nicht adäquat ausgedrückt werden kann. Beispiele: Tel AND NOT „Tel Aviv“.

Die Beurteilung der *Metadaten* fiel ernüchternd aus und zeigt deutliche Mängel in der Produktion der Online-Inhalte, die in vielen Fällen wohl noch nicht adäquaten Prozessen unterliegen. Erstaunlich ist zudem weiterhin, dass in einigen Fällen nicht einmal ein korrektes Datum angegeben wird. Untersuchungen von Dumais et al. (2003) haben gezeigt, dass das Datum beim Wiederfinden von Dokumenten, die man schon einmal gesehen hat, eine wichtige Rolle spielt, vor allem deshalb, weil man das Erscheinen des gesuchten Dokumentes zeitlich einordnen kann. In vielen Fällen fehlt die Angabe, zu welcher Teilkollektion ein gefundenes Dokument gehört, was dann im Suchprozess das explizite Ein- und Ausschließen von Teilkollektionen erschwert.

### 3.4 Benutzerinteraktion

Das Kriterium Benutzerinteraktion beurteilt den Teil der Suchapplikation, mit welcher der Nutzer direkt arbeitet resp. kommuniziert. Die Qualität der Benutzerinteraktion wurde aufgrund folgender Subkriterien beurteilt: Präsentation Trefferliste, Nutzerführung und Performance. Der Median für das gesamte Kriterium Benutzerinteraktion mit einem Wert von 0.42 sowie einem Maximum von 0.68 sind überraschend tief. Das ist insbesondere deswegen erstaunlich, da dieser Aspekt der Suchapplikation – da sichtbar und bedienbar – einfach, verständlich und testbar ist. Zudem lassen sich die meisten Anforderungen an die Benutzerinteraktion mit horizontalen Internet-Suchdiensten, die auf derselben Kollektion arbeiten, vergleichen.

Die Nützlichkeit der Trefferliste hängt stark vom Vorhandensein bestimmter Informationen und Funktionen sowie von deren Darstellung ab. Diese Anforderungen wurden im Subkriterium *Präsentation der Trefferliste* geprüft. Die Nutzeranforderungen für eine öffentliche Site-Suche sind gut erforscht. Durch die intensive Nutzung von öffentlichen Suchdiensten haben Nutzer zudem Darstellungen sowie gewisse Funktionen erlernt. Davon

abzuweichen ist bei ähnlichen gelagerten Anforderungen nicht zu empfehlen und löste bei den Testpersonen Irritation aus. Die Varianz über alle Tests ist dennoch sehr groß. Grund dafür sind anspruchsvoller umzusetzende Anforderungen wie beispielsweise die Qualität von Trefferzitate oder die Hervorhebung des Suchbegriffes im Treffer. Aber auch grundlegende Aspekte (vgl. Kleinfeldt&Baphna 2000 wie die Anzeige der URL oder des Pfades des Treffers, die Anzeige des Dokumentdatums oder der Größe des Dokumentes wurden in einigen Fällen vernachlässigt). So wie beim Hauptkriterium Suchindex war auch hier die Behandlung von binären Dokumentformaten, insbesondere von Adobe PDF, problembehaftet.

*Nutzerführung* ist eine sehr wichtige Anforderung, um Nutzer bei der Suche zu unterstützen. Mit einem Median von 0.2 ist dieser Wert, mit Ausnahme der Aktualität des Suchindex, der schlechteste Wert eines Subkriteriums in der ganzen Evaluation. Dies, obschon es bei der Suche um die Nutzer geht und sich viele grundlegende Probleme der Informationssuche wie beispielsweise Subjektivität von Relevanz oder fehlendes Kontextwissen bei Suchbeginn exakt in diesem Feld bewegen. Bereits an der Anforderung, eine Suche mit der Eingabe-Taste auszulösen, scheiterten zwei der Websites. Bezüglich der Editierbarkeit des Suchbegriffs auf der Trefferliste schon acht. Anspruchsvollere Funktionen wie Korrekturvorschläge für Tippfehler oder eine aktive Unterstützung durch Erweiterung des Suchbegriffs sind bei fast allen Websites abwesend.

Die *Performance* ist in fast allen Fällen gut, wie der Median von 1 und die schiefe Verteilung zu Gunsten der Maximalbewertung zeigt. Im Rahmen des Tests zeigte sich klar, dass die Performance ein klarer Unzufriedenheitsfaktor ist. Ist die Antwortzeit gut, wird kaum gelobt. Ist dies aber ungenügend, so reagierten die Testpersonen stark irritiert bis hin zu Testabbrüchen. Bei den als langsam bewerteten Suchfunktionen ließ sich außerhalb des Testrasters gut beobachten, dass die Antwortzeit einen Zusammenhang mit der Größe der Resultatsmenge hat.

Wie bereits argumentiert, sind die Anforderungen an eine Site-Suche nicht dieselben wie an einen horizontalen Internet-Suchdienst. In Bezug auf die Benutzerschnittstelle empfiehlt sich aber eine starke Anlehnung. Gewisse Interaktionsprinzipien werden durch die häufige Nutzung erlernt und die visuelle Leistung sollte dieselben Anforderungen erfüllen. Fast gänzlich abwesend sind leider moderne Arten der Informationerschließung oder spezielle Lösungsansätze entlang der Kommunikationsziele der Suchfunktion.

### 3.5 Suchergebnisse

Broder (2002) unterscheidet in seinem viel beachteten und zitierten Artikel zwischen drei Anfragearten. Bei den *Navigationsanfragen* (navigational queries) geht es darum, eine Website oder eine Homepage zu finden, um von dort aus weiterzuvagieren zu können. Bei *Informationsanfragen* (informational queries) geht es um klassische Informationsrecherchen, bei der Informationen zu einem bestimmten Thema zusammengestellt werden müssen. Bei *Transaktionsanfragen* (transactional queries) soll der Einstiegspunkt für eine Online-Transaktion gefunden werden. Andrei Broder hat diesen Artikel geschrieben, bevor er von Altavista zu IBM wechselte und er hatte Zugang zu den Altavista-Logdateien. Für unsere Zwecke haben wir die Taxonomie der Anfragen modernisiert, indem wir Transaktionsanfragen als eine spezielle Form von Navigationsanfragen betrachten und zusätzlich Faktenanfragen berücksichtigen. Bei *Faktenanfragen* soll ein Faktum gefunden werden, beispielsweise eine Adresse, ein Name oder eine Telefonnummer. Ein einziges vertrauenswürdige Dokument genügt, welches das gesuchte Faktum enthält. Die anderen Dokumente interessieren dann nicht mehr, weil man das Informationsbedürfnis mit diesem einen Dokument befriedigen konnte.

Es kann festgestellt werden, dass bei Navigationsanfragen die die meisten Site-Suchen gute, die restlichen jedoch schlechte bis sehr schlechte Ergebnisse liefern. Bei den Faktenanfragen liefert die Hälfte der Suchapplikationen gute Ergebnisse, ein weiteres Viertel der Applikationen kann als genügend bezeichnet werden. Das letzte Viertel ist klar ungenügend. In der klassischen Information Retrieval Disziplin, nämlich den Informationsanfragen, ist das Ergebnis als schlecht zu werten. Nur Ausreißer sind genügend nahe am theoretischen Maximalwert, die meisten Suchapplikationen sind weit vom Maximalwert entfernt und daher ungenügend bis schlecht.

Die Anfrage nach offenen Stellen (Jobs, Stellen, etc.) ist eine typische Navigationsanfrage. Der Benutzer sucht den Einstiegspunkt, von dem aus er weiter zur Seite mit den aktuellen Stellenangeboten navigieren kann. Die Nutzungsanalysen von vielen Site-Suchen zeigen, dass die Frage nach offenen Stellen zu den am häufigsten gestellten Fragen gehört (persönliche Mitteilung des Suchmaschinen-Herstellers Eurospider basierend auf Kundendaten). Um so mehr erstaunt, dass einige Suchapplikationen kein befriedigendes Resultat liefern.

Dass Faktenanfragen für eine Site-Suche nicht immer trivial zu beantworten sind, zeigt die Testanfrage nach der Anzahl der Mitarbeitenden. Ein Drittel der getesteten Applikationen konnten mit dieser An-

frage nichts anfangen. Der Hauptgrund liegt bei den verschiedenen Formulierungsmöglichkeiten. Gemäß Furnas et al. (1987) besteht eine 80 Prozent Wahrscheinlichkeit, dass der Benutzer der Site-Suche und der Autor einer relevanten Website verschiedene Bezeichnungen benutzen. Eine ideale Suchapplikation sollte einen Benutzer aktiv unterstützen z.B. durch Unterbreiten von alternativen Formulierungen oder Begriffen.

## 4 Empfehlungen

Die Studie konnte den State-of-the-art in Sachen Site-Suche für Unternehmen und Behörden in der Schweiz ermitteln und analysieren. Das Fehlen von Systemen, die in allen Hauptkriterien brillieren, kann als Anreiz aufgefasst werden, eine real existierende Leistung zu erreichen, die deutlich über den Suchfunktionen der heute besten Websites liegt. Das Potential wurde von den einzelnen Site-Suchen in der Untersuchung demonstriert. Die Einzelanalyse der Hauptkriterien lässt einige Ansatzpunkte zur Optimierung identifizieren. Darunter befinden sich insbesondere:

- *Mehr Gewicht auf die Vollständigkeit und Aktualität des Suchindex.* Die Suche ist nur so gut wie die Datenbasis. Viele Websites haben hier Verbesserungsbedarf. Eine mögliche Abhilfe besteht in einer aktiveren Einbindung der Suchfunktion in die Gesamtwebsite (aktive Aktualisierungen des Suchindex statt passives „Crawlen“, d.h. Sammeln der Dokumente durch die Suchapplikation).
- *Bessere Pflege der Metadaten im Hinblick auf Einschränkungen bei komplexeren Suchen.* Eine minimale Qualität (korrektes Datum, korrekte Titel) hilft, dass diesbezügliche Einschränkungen funktionieren. Trefferzitate müssen vollständig und prägnant sein. Wichtige Informationen müssen in Dateiformaten vorliegen, die sauber verarbeitet werden können (die Testpersonen bemängelten häufig Probleme im Umgang mit Adobe PDF-Dokumenten).
- *Orientierung an bewährten Standards für die Benutzerinteraktion.* Die Testpersonen haben fast immer irritiert reagiert, wenn ihnen die Site-Suche intransparent erschien. In Zukunft ist vermehrt zu untersuchen, mit welchen Vorstellungen an die Interaktion die Nutzer sich der Site-Suche nähern. Gegebenenfalls hilft die Orientierung an horizontalen Internet-Suchdiensten.
- *Ausbau der Qualität des Vergleichs zwischen Anfragen und Dokumenten.* Die korrekte Verarbeitung der natürlichen Sprache (Flexionen, Komposita) muss mehr Gewicht erhalten. Gegebenenfalls ist Mehrsprachigkeit zu beachten.

- *Trefferlisten sind nicht nur aus Dokumentensicht, sondern auch aus Nutzersicht aufzubereiten.* Gewisse häufige Informationsbedürfnisse können vorweg genommen werden. Die Suchergebnisse für solche Anfragen können durch vorgängige Optimierung (andere Gewichtung/Stratifikation der Trefferliste, Entitätenerkennung u.a.) verbessert werden. Mit relativ kleinem Aufwand kann so bei einer großen Anzahl von Einzelanfragen eine bessere Leistung erreicht werden.

## 5 Diskussion und Ausblick

Im Nachhinein stellen sich die beiden Fragen „Wurden die richtigen Kriterien evaluiert?“ und zweitens, „Wurden diese Kriterien richtig evaluiert?“. Bei der ersten Frage muss berücksichtigt werden, dass weder ein Sitewettbewerb noch ein Produktwettbewerb mit Ranglisten beachtet war. Dies rechtfertigt, weshalb man auf die Kriterien

- Kosten (einmalige und laufende),
- Sicherheit (Site-Suchen sind im Internet exponiert) und
- Organisation (anspruchsvolle/einfache Wartung)

verzichtete. Die Wahl der verwendeten Kriterien (Suchindex, Vergleich von Anfragen und Dokumenten, Benutzerinteraktionen und Suchergebnisse) basiert auf den Erkenntnissen, welche in den großen Information Retrieval Evaluationsforen TREC (<http://trec.nist.com>), CLEF ([www.clef-campaign.org](http://www.clef-campaign.org)) und NTCIR (<http://research.nii.ac.jp/ntcir/>) gemacht wurden. Die mehr als zehnjährige Tradition dieser Evaluationsforen und die Tatsache, dass die internationale Forschungsgemeinschaft in diesen Foren gut vertreten ist, rechtfertigt die Auswahl. Unter den in dieser Studie verwendeten Kriterien gibt es keine offensichtliche Abhängigkeit, was die Auswahl sinnvoll erscheinen lässt.

Sehr gerne hätten die Verfasser der Studie die Retrieval-Effektivität umfassend ermittelt, wie dies in den oben erwähnten Evaluationsforen gemacht wird. Dabei werden in standardisierten Umgebungen mit speziellen Maßen (Ausbeute, Präzision, P@10, etc.) ermittelt, ob eine Suchapplikation zu den Anfragen die relevanten Informationen *effektiv* findet. Da wir es bei den Site-Suchen nicht mit standardisierten Umgebungen zu tun haben, hätte man die Retrieval-Effektivität mit einem speziellen Experiment ermitteln müssen. Bei 56 Website-Suchen hätte dies jedoch den Rahmen um Größenordnungen gesprengt. Zukünftig ist eine derartige vertiefte Evaluation bei einzelnen Site-Suchen durchaus denkbar.

Die Frage, ob die verschiedenen Kriterien korrekt evaluiert wurden, muss differenziert beantwortet werden. Im Nachhinein

konnten bei vereinzelt Tests Probleme identifiziert werden, welche zu nur teilweise korrekten Resultaten führten. Aufgrund der Erfahrung in den großen Evaluationsforen wurde das Evaluationsraster jedoch so angelegt, dass diese Fehler keinen Einfluss auf die Kernaussagen haben. Die wichtigsten Mechanismen sind die Folgenden.

- Genügend große Anzahl von Tests, was mit 74 erfüllt ist.
- Die Testpersonen müssen die Tests unabhängig vom darunter liegenden Produkt evaluieren. Dies wurde erreicht, indem die Testpersonen keine Beziehung zu den Site-Suchprojekten hatten und die Zuteilung nach dem Zufallsprinzip erfolgte.
- Die Bewertungen erfolgen relativ zu den Bewertungen anderer Site-Suchen. Dies erfordert genügend viele zu testende Site-Suchen, was mit 56 erfüllt ist.

Zusammenfassend kann man die in der Studie gemachten Erkenntnisse als fundiert bezeichnen. Die Kernaussagen werden nicht durch vereinzelte Unschönheiten relativiert, welche aber in zukünftigen Studien trotzdem eliminiert werden sollten. Diese erste umfassende Evaluation der Suchfunktion auf Websites wird weitergeführt. Zum einen stehen die beteiligten Unternehmen für individuelle Tests und Beratungen für Interessierte zur Verfügung. Zum anderen werden im Frühjahr

2007 etwa 100 Unternehmen und Organisationen in Deutschland dem gleichen Test unterzogen.

Als Fazit bleibt festzuhalten, dass die Bedeutung der Suchfunktion als Kommunikationsinstrument noch nicht erkannt wird. Notwendig ist ein Umdenken, die Verantwortung für die Suchfunktion muss künftig von der Kommunikationsabteilung wahrgenommen werden, sie ist keine primär technische Angelegenheit. Die Suche muss sich an der Unternehmens-Website orientieren. Dabei ist die Information geeignet zu strukturieren und anzubieten. Der Site-Suche ist eine hohe Aufmerksamkeit zu widmen, die ihrer Rolle als Kommunikationsinstrument gerecht wird.

### Literatur

Alvarado, C.; Teevan, J.; Ackerman, M.; Karger, D. (2003). Surviving the Information Explosion: How People find their Electronic Information. AI Memo 2003-006. <http://haystack.lcs.mit.edu/papers/alvarado.aim03.pdf>, verifiziert am 28.11.2006

Braschler, M.; Herget, J.; Pfister, J.; Schäuble, P.; Steinbach, M.; Stuker, J. (2006): Evaluation der Suchfunktion von Schweizer Unternehmens-Websites, Churer Schriften zur Informationswissenschaft Schrift 12, Chur, Dezember 2006. [www.informationwissenschaft.ch/index.php?id=89](http://www.informationwissenschaft.ch/index.php?id=89), verifiziert am 15.12.2006

Cleverdon, C. W. (1967). The Cranfield tests on index language devices. Aslib Proceedings, 19, 173-192. Reprinted in (Sparck Jones & Willett 1997)

Cremers, I. et al. (2006). It's Time To Update Site Search Functionality. Best Practices. Forrester Research, Inc.

Dumais, S.; Cutrell, E.; Cadiz, J.J.; Jancke, G.; Sarin, R.; Robbins, D. C. (2003). Stuff I've Seen: A System for Personal Information Retrieval and Re-Use. Proceedings of SIGIR '03, 72-79, 2003. <http://research.microsoft.com/~sdumais/SISCore-SIGIR2003-Final.pdf>, verifiziert am 24.11.2006

Furnas, G.W.; Landauer, T.K.; Gomez, L.M.; Dumais, S.T.: The Vocabulary Problem in Human-System Communication. Commun. ACM 30(11): 964-971 (1987)

Kleinfeldt, S. & Baphna, J. (2000). A Commercial Perspective on Hypertext Search Results. In Proceedings of the Hypertext 2000 & Digital Libraries 2000 Workshop on Information Doors - Where Information Search and Hypertext Link in San Antonio, Texas, USA, 2000. New York: ACM Press

Lyman, P. & Varian, H. R. (2003). How Much Information, 2003. [www.sims.berkeley.edu/how-much-info-2003](http://www.sims.berkeley.edu/how-much-info-2003), verifiziert am 24.11.2006

Schmidt-Mänz, N. & Bomhardt, C. (2005) Wie suchen Onliner im Internet? Absatzwirtschaft, Science Factory 2/2005.

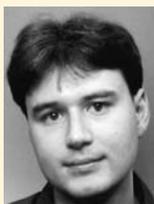
Sparck Jones, K.; Willett, P. (1997). Readings in information retrieval. San Francisco, CA: Morgan Kaufmann

Voorhees, E. M. (2002). The Philosophy of Information Retrieval Evaluation. In C. Peters, M.; Braschler, J.; Gonzalo, M.; Kluck (Eds.): Evaluation of Cross-Language Information Retrieval Systems. Second Workshop of the Cross-Language Evaluation Forum, CLEF 2001. Lecture Notes in Computer Science, Vol. 2406 (pp. 355-370)

**Empirische Untersuchung, Betrieb, Bewertung, Kommunikation, Benutzer, Website, Recherchieren, Schweiz**

## DIE AUTOREN

### Dr. Martin Braschler



Dozent an der Zürcher Hochschule Winterthur  
Zürcher Hochschule Winterthur  
Institut für angewandte Informationstechnologie  
Technikumsstrasse 9, 8400 Winterthur  
brn@zhwin.ch

### Prof. Dr. Josef Herget



Leiter des Arbeitsbereichs Informationswissenschaft an der Hochschule für Technik und Wirtschaft Chur.  
Hochschule für Technik und Wirtschaft HTW Chur  
Swiss Institute for Information Research (SII)  
Ringstrasse/Pulvermühlestr. 57, 7000 Chur  
josef.herget@fh-htwchur.ch

### Joachim Pfister



Wissenschaftlicher Assistent  
Hochschule für Technik und Wirtschaft HTW Chur  
Swiss Institute for Information Research (SII)  
Ringstrasse/Pulvermühlestrasse 57  
7000 Chur  
joachim.pfister@fh-htwchur.ch

### Dr. Peter Schäuble



CEO Eurospider Information Technology AG  
Eurospider Information Technology AG  
Schaffhauserstrasse 18  
8006 Zürich  
schauble@eurospider.com

### Markus Steinbach



Senior Consultant  
namics AG  
Teufenerstrasse 19  
9000 St. Gallen  
markus.steinbach@namics.com

### Jürg Stuker



CEO und Partner der namics AG  
namics AG  
Teufenerstrasse 19  
9000 St. Gallen  
juerg.stuker@namics.com

# Nachweis deutschsprachiger bibliotheks- und informationswissenschaftlicher Aufsätze in Google Scholar

Dirk Lewandowski, Hamburg

## 2 Nachweisformen in Google Scholar

Google Scholar bietet – wenn vorhanden – die gefundenen Artikel im Volltext. Dabei wird unterschieden zwischen einer (kostenpflichtigen) Verlagsversion und einer (kostenlosen) Preprint- oder Autorenversion. Oft finden sich allerdings auch Originalbeiträge als PDF auf den Autorensseiten und damit zum kostenfreien Download. Sind mehrere Versionen eines Artikels vorhanden, zeigt Google Scholar bevorzugt die Verlagsversion, allerdings findet sich dann zu diesem Treffer ein Hinweis auf weitere Versionen, so dass sich unter Umständen auch eine kostenlose Version finden lässt. Diese Funktion ist wohl vor allem für die Auswahl zwischen kostenpflichtiger Originalversion und kostenlosem Preprint gedacht, bündelt aber auch gleiche oder unterschiedliche Preprint-Versionen, die auf verschiedenen Servern liegen. Ist kein Volltext vorhanden, weist Google Scholar den Artikel als „Citation“ nach, das heißt der Artikel selbst ist im System nicht vorhanden, wurde jedoch in einem dem System bekannten Artikel zitiert. Die Literaturangaben aus den bekannten Volltexten werden extrahiert und in Zitationen umgesetzt. Auf diese Weise werden auch Werke erfasst, die über den Abdeckungsbereich von Google Scholar hinausgehen, etwa Monographien und weitere, nicht im Web vorhandene Werke.

Für die vorliegende Untersuchung ist insbesondere interessant, wie groß der Anteil der Nachweise ist. Lässt sich Google Scholar als Ersatz für informationswissenschaftliche Datenbanken wie LISA und INFODATA verwenden? Eine zweite interessante Frage ist die nach dem Anteil der verfügbaren Volltexte: Gelangt man mit Hilfe von Google Scholar problemlos zu den gesuchten Aufsätzen?

## 3 Aufbau der Untersuchung

Für die Untersuchung wurden deutschsprachige, überregionale Zeitschriften aus dem Bereich Bibliothek und Information ausgewählt, also die deutschen Kernzeitschriften des Bereichs. Eine Unterteilung in Bibliotheks- und Informati-

*In dieser Untersuchung werden die Aufsätze der Jahre 2004 bis 2006 aus den acht wichtigsten deutschsprachigen Zeitschriften des Bibliotheks- und Informationswesens auf ihre Verfügbarkeit in Google Scholar hin untersucht. Dabei zeigt sich, dass Google Scholar keinen vollständigen Nachweis der Artikel bieten kann und sich daher nicht als Ersatz für bibliographische Datenbanken eignet. Google Scholar macht einen Teil der Volltexte direkt verfügbar und eignet sich insofern für einen schnellen und problemlosen Zugang zum Volltext, der traditionellen Wege wie die Bibliothek bzw. die Fernleihe umgeht. Für das Fach Bibliotheks- und Informationswissenschaft bietet Google Scholar insgesamt die Chance, seine Inhalte auch über die Fachgrenzen hinaus bekannt zu machen.*

### References of German LIS articles in Google Scholar

*In this study, we ask if Google Scholar is suitable for finding articles in library and information science. Therefore, we check if articles published from 2004-2006 from the eight major German language LIS periodicals can be found in Google Scholar. We find that Google Scholar does not give bibliographic data for all articles and is therefore no substitute for bibliographic databases. But, Google Scholar makes available a certain amount of the articles in full text. Therefore, in some cases it can be a fast and unproblematic way for obtaining full texts, omitting the traditional way of visiting the library or interlibrary loan. For library and information science, Google Scholar could be a good way to make its articles available to a wider audience.*

## 1 Einleitung

Google Scholar ist eine Wissenschaftssuchmaschine, die eine möglichst umfassende Abdeckung aller wissenschaftlichen Literatur erreichen will. Dazu werden sowohl Inhalte aus dem offenen Web (von Institutsseiten oder persönlichen Homepages), Open-Access-Repositories als auch Inhalte von Verlagen, die für Nutzer nur gegen Bezahlung zugänglich sind,

indexiert. Google Scholar öffnet damit Teile des Academic Invisible Web (Lewandowski & Mayr 2006) und führt diese mit Inhalten zusammen, die auch über reguläre Web-Suchmaschinen zugänglich sind. Seit dem Start von Google Scholar wird diese Wissenschaftssuchmaschine stark diskutiert, vor allem auch in Hinblick auf das Verhalten von Bibliotheksbenutzern, die sich zunehmend diesem Dienst zuwenden (Phipps & Maloney 2006). Damit stellt sich die Frage, inwieweit Google Scholar die Recherche in Fachdatenbanken ersetzen kann. Zwar mag man aus der eigenen Erfahrung annehmen, dass Google Scholar keine vergleichbaren Ergebnisse bieten kann, empirische Untersuchungen dazu fehlen jedoch weitgehend.

Die bisherigen Untersuchungen zu Google Scholar konzentrieren sich vor allem auf die Suchmöglichkeiten (Adlington & Benda 2006; Jacsó 2005), den technischen Aufbau (Lewandowski 2005) und die Qualität der gefundenen Ergebnisse (Callicott & Vaughn 2006). Der Frage der Abdeckung gehen Mayr und Walter (Mayr & Walter 2006) auf der Basis von interdisziplinären Zeitschriftenlisten nach. Da allerdings für jede Zeitschrift nur eine Stichprobe von Artikeln erfasst wird, kann hier keine Aussage über die tatsächliche Abdeckung einzelner Zeitschriften gemacht werden. Hier setzt die vorliegende Untersuchung an und untersucht den individuellen Abdeckungsgrad auf Zeitschriftenebene exemplarisch anhand von deutschsprachigen LIS-Zeitschriften.

Die Untersuchung wurde auf Google Scholar beschränkt. Zwar bieten auch andere Anbieter Wissenschaftssuchmaschinen an (vgl. Lewandowski 2006; Lewandowski 2007), vor allem Elsevier mit Scirus („Scirus White Paper“ 2004) und Microsoft mit Windows Live Academic. Diese sind jedoch anders ausgerichtet: Im Fall von Scirus soll das gesamte Academic Web (und nicht nur die wissenschaftliche Literatur) abgedeckt werden, Windows Live Academic (vgl. Söllner 2006) beschränkt sich auf die von den Verlagen angebotenen Aufsätze. Für die hier behandelte Themenstellung erscheint aber gerade das umfangreiche Literaturangebot von Google Scholar interessant. Ein systematischer Vergleich der verschiedenen Wissenschaftssuchmaschinen (Lewandowski 2006) ist jedoch weiterhin ein Desiderat.

onswissenschaft erschien nicht sinnvoll, weil es im deutschsprachigen Raum keine genuin informationswissenschaftlichen Zeitschriften gibt.

Die ausgewählten Zeitschriften sind ABI-Technik, Bibliothek Forschung und Praxis, Bibliotheksdienst, B.I.T.Online, BuB, Information: Wissenschaft und Praxis, Passwort sowie Zeitschrift für Bibliothekswesen und Bibliographie. Die Auswahl orientiert sich weitgehend an den in Stock & Schlögl (2004) ausgewerteten deutschsprachigen Zeitschriften, wobei auf die Auswertung von Pro Libris (regionaler Bezug) und die ISI-Tagungsbände (nur ein Band im Untersuchungszeitraum) verzichtet wurde.

Von jeder Zeitschrift wurden jeweils die Artikel der letzten drei Jahrgänge (2004 bis 2006) ausgewertet. Dabei fand eine Beschränkung auf die Aufsätze statt. Auf Nachrichtenbeiträge oder ähnliches wurde verzichtet. Eine Besonderheit bildet Passwort, wo nur zu einem geringen Teil Aufsätze veröffentlicht werden. Hier wurden nur diejenigen Artikel berücksichtigt, die auch in die Datenbank INFO-DATA nachgewiesen werden.

Jeder Artikel wurde in Google Scholar gesucht und es wurde erfasst, ob der Artikel dort (in Form von bibliographischen Angaben) nachgewiesen wird, ob ein Preprint abrufbar ist und ob die Originalversion (also das Verlags-PDF) entweder kostenlos oder kostenpflichtig abrufbar ist. Beim Artikelnachweis wurde nicht zwischen den verschiedenen Formen, die Google Scholar angibt, unterschieden. Insgesamt wurden 791 Artikel überprüft; die Datenerhebung fand am 3. Januar 2007 statt.

## 4 Ergebnisse

Die Ergebnisse der Untersuchung sind in Tabelle 1 zusammengefasst und werden im Folgenden erläutert.

### 4.1 Nachweise

Bei den Nachweisen ist zuerst festzustellen, dass Google Scholar bei keiner der untersuchten Zeitschriften eine vollständige Abdeckung bieten kann. Am besten schneidet Bibliothek Forschung & Praxis ab (90,7 Prozent der Artikel in Google Scholar nachgewiesen), gefolgt von Bibliotheksdienst (86,6 Prozent), IWP (83,1 Prozent), ZfBB (78,3 Prozent) und ABI-Technik (59,5 Prozent). Interessant ist, dass die Nachweise in Google Scholar einzig beim Bibliotheksdienst aus der Originalquelle kommen, und zwar durch die vom Anbieter auf seiner Website bereitgestellten Inhaltsverzeichnisse. Bei allen anderen Zeitschriften bedient sich Google Scholar aus anderen Quellen, seien es bibliographische Nachweissysteme oder Artikel, die Quellen aus den genannten Zeitschriften zitieren.

Die Artikel aus Passwort, B.I.T.Online und BuB werden nur zu einem geringen Teil (unter einem Drittel aller Artikel) in Google Scholar nachgewiesen. Bei Passwort finden sich keine Inhaltsverzeichnisse online, die Inhaltsverzeichnisse von BIT Online und BuB werden – obwohl online zugänglich – nicht erfasst.

Im Durchschnitt ergibt sich für alle Zeitschriften eine Nachweisquote von 56 Prozent. Dies verdeutlicht, dass Google Scholar (wenigstens zurzeit) die bibliographischen Datenbanken für deutschsprachige LIS-Literatur nicht ersetzen kann. Ein wichtiges Ergebnis ist, dass Google Scholar auch im Fall der Verfügbarkeit aller Nachweise (wie im Fall des Bibliotheksdienstes) die Nachweise nicht vollständig erfasst.

### 4.2 Volltexte

Die zweite Forschungsfrage ist die nach dem Anteil der nachgewiesenen Volltexte. Betrachtet man alle Zeitschriften, liegt dieser im Durchschnitt bei nur 21,4 Prozent. Am besten schneiden hier Bibliotheksdienst (50,2 Prozent aller Artikel im Volltext verfügbar) und Bibliothek Forschung und Praxis (34,7 Prozent) ab. Deutlich abgeschlagen folgt die IWP mit 18,5 Prozent. Dies ist erstaunlich, weil alle Aufsätze dieser Zeitschrift im Portal Informationswissenschaft (vgl. Hauer 2005) verfügbar sind. Allerdings sind die Inhalte des Portals nicht für die allgemeinen Suchmaschinen zugänglich, da diese nur über Links auf die Artikel zugreifen könnten, solche Links (etwa durch Inhaltsverzeichnisse) im Portal aber nicht vorhanden sind.

Sowohl Bibliotheksdienst als auch Bibliothek Forschung und Praxis bieten ihre Volltexte mit einem gewissen Zeitverzug kostenlos an. Diese freigegebenen Inhalte werden von Google Scholar zu einem großen Teil erfasst. Bei den anderen Zeitschriften ist Google Scholar auf andere Zugänge zum Volltext (wie Repositorien und Autoren-Homepages) angewiesen.

### 4.3 Preprints

Es stellt sich die Frage, ob Autoren wenigstens in einem gewissen Maß bereit sind, ihre Volltexte selbst in Form von Preprints zugänglich zu machen, sei es über institutionelle Repositorien oder über ihre eigenen Homepages.

Der Nachweis solcher Preprints in Google Scholar fällt enttäuschend aus. Über alle Zeitschriften hinweg finden sich nur 2,5 Prozent der Artikel als Preprint. Am besten schneiden hier noch die IWP mit 7,3 Prozent der Artikel, die ZfBB (6,7 Prozent) und BIT Online (6,5 Prozent) ab.

In Google Scholar finden sich Treffer aus verschiedenen institutionellen Repositorien, vor allem auch aus E-LIS (Bargmann, Katzmayr & Putz 2005), dem größten Archiv für Preprints im Bereich Bibliotheks- und Informationswissenschaft. Es ist

daher nicht davon auszugehen, dass der geringe Anteil der Preprints auf die generelle Nicht-Erfassung durch Google Scholar zurückzuführen ist. Vielmehr ist zu vermuten, dass Open Access (und damit das Verfügbarmachen von Preprints) bei den Autoren der deutschen LIS-Szene schlicht „noch nicht angekommen“ ist. Ein erstaunliches Ergebnis, wenn man bedenkt, dass gerade in diesem Bereich Open Access stark propagiert wird.

### 4.4 Kostenlose PDF-Dateien

Bei den kostenlosen PDF-Dateien (gemeint sind hier seitenidentische Faksimiles der gedruckten Versionen) liegen aufgrund der vollständigen und kostenlosen Verfügbarmachung älterer Artikel durch die Verlage die beiden Zeitschriften Bibliotheksdienst (mit 48,8 Prozent aller Artikel) und Bibliothek Forschung und Praxis (29,3 Prozent) auf den ersten Plätzen. Allerdings erfasst auch hier Google Scholar bei weitem nicht alle vorhandenen Artikel. Im Fall der IWP (11,3 Prozent der Artikel als kostenloses PDF) ergibt sich wieder der Effekt, dass die PDF-Dateien zwar für alle Artikel vorhanden sind, aber durch das Portal Informationswissenschaft nicht suchmaschinengerecht zugänglich gemacht werden.

### 4.5 Aktivitäten der Verlage

Im Gegensatz zu anderen Fachgebieten finden sich für den deutschsprachigen LIS-Bereich nur wenige Artikel, die kostenpflichtig angeboten werden. Es handelt sich um die noch nicht freigegebenen Aufsätze aus Bibliothek Forschung und Praxis, die nur bei entsprechender Campuslizenz zugänglich sind. Eine direkte Vermarktung von einzelnen Artikeln findet bisher durch keinen Verlag statt.

### 4.6 Zeitliche Veränderungen

Aus den erhobenen Daten lässt sich keine Tendenz zu einer besseren Erschließung neuerer Artikel durch Google Scholar feststellen. Da das System keine zeitlichen Erfassungsgrenzen kennt, also auch alte Artikel erschließen würde, wenn sie später verfügbar gemacht würden, war dies auch nicht zu erwarten. Aber auch beim Anteil der Preprints lässt sich keine positive Tendenz feststellen. Zwar können auch hier Autoren ihre älteren Artikel noch später in Repositorien einstellen, es ist jedoch davon auszugehen, dass sie vorwiegend ihre aktuellen Artikel dort verfügbar machen. Insofern scheinen die Open-Access-Appelle wenig Wirkung auf die Autoren im deutschsprachigen LIS-Bereich zu haben.

## 5 Diskussion

Die Untersuchung zeigt, dass sich Google Scholar nicht als Ersatz für bibliographische Datenbanken im Bereich Biblio-

Tabelle 1: Anteil der in Google Scholar (GS) nachgewiesenen Aufsätze nach Zeitschriften und Art des Nachweises

Alle Zeitschriften	2004	2005	2006	gesamt	Anteil 2004	Anteil 2005	Anteil 2006	Anteil gesamt
Anzahl Artikel	265	268	258	791				
Nachweis in GS	163	157	123	443	61,5	58,6	47,7	56,0
Preprint	6	8	6	20	2,3	3,0	2,3	2,5
PDF kostenlos	70	59	16	145	26,4	22,0	6,2	18,3
PDF kostenpflichtig	0	2	2	4	0,0	0,7	0,8	0,5
Volltext gesamt	76	69	24	169	28,7	25,7	9,3	21,4
ZfBB	2004	2005	2006	gesamt	Anteil 2004	Anteil 2005	Anteil 2006	Anteil 2004-2006
Anzahl Artikel	21	18	21	60				
Nachweis in GS	19	14	14	47	90,5	77,8	78,3	78,3
Preprint	1	3	0	4	4,8	16,7	6,7	6,7
PDF kostenlos	0	0	0	0	0,0	0,0	0,0	0,0
PDF kostenpflichtig	0	0	0	0	0,0	0,0	0,0	0,0
Volltext gesamt	1	3	0	4	4,8	16,7	0,0	6,7
IWP	2004	2005	2006	gesamt	Anteil 2004	Anteil 2005	Anteil 2006	Anteil 2004-2006
Anzahl Artikel	39	40	45	124				
Nachweis in GS	34	37	32	103	87,2	92,5	71,1	83,1
Preprint	3	3	3	9	7,7	7,5	6,7	7,3
PDF kostenlos	3	8	3	14	7,7	20,0	6,7	11,3
PDF kostenpflichtig	0	0	0	0	0,0	0,0	0,0	0,0
Volltext gesamt	6	11	6	23	15,4	27,5	13,3	18,5
ABI-Technik	2004	2005	2006	Gesamt	Anteil 2004	Anteil 2005	Anteil 2006	Anteil 2004-2006
Anzahl Artikel	13	12	12	37				
Nachweis in GS	8	6	8	22	61,5	50,0	66,7	59,5
Preprint	1	0	0	1	7,7	0,0	0,0	2,7
PDF kostenlos	0	0	0	0	0,0	0,0	0,0	0,0
PDF kostenpflichtig	0	0	0	0	0,0	0,0	0,0	0,0
Volltext gesamt	1	0	0	1	7,7	0,0	0,0	2,7
Bibliothek F&P	2004	2005	2006	Gesamt	Anteil 2004	Anteil 2005	Anteil 2006	Anteil 2004-2006
Anzahl Artikel	24	37	14	75				
Nachweis in GS	20	35	13	68	83,3	94,6	92,9	90,7
Preprint	0	0	0	0	0,0	0,0	0,0	0,0
PDF kostenlos	13	9	0	22	54,2	24,3	0,0	29,3
PDF kostenpflichtig	0	2	2	4	0,0	5,4	14,3	5,3
Volltext gesamt	13	11	2	26	54,2	29,7	14,3	34,7
Bibliotheksdienst	2004	2005	2006	gesamt	Anteil 2004	Anteil 2005	Anteil 2006	Anteil 2004-2006
Anzahl Artikel	82	68	67	217				
Nachweis in GS	72	61	55	188	87,8	89,7	82,1	86,6
Preprint	0	0	3	3	0,0	0,0	4,5	1,4
PDF kostenlos	51	42	13	106	62,2	61,8	19,4	48,8
PDF kostenpflichtig	0	0	0	0	0,0	0,0	0,0	0,0
Volltext gesamt	51	42	16	109	62,2	61,8	23,9	50,5
B.I.T.Online	2004	2005	2006	gesamt	Anteil 2004	Anteil 2005	Anteil 2006	Anteil 2004-2006
Anzahl Artikel	10	10	11	31				
Nachweis in GS	4	2	0	6	40,0	20,0	0,0	19,4
Preprint	0	2	0	2	0,0	20,0	0,0	6,5
PDF kostenlos	1	0	0	1	10,0	0,0	0,0	3,2
PDF kostenpflichtig	0	0	0	0	0,0	0,0	0,0	0,0
Volltext gesamt	1	2	0	3	10,0	20,0	0,0	9,7
BuB	2004	2005	2006	gesamt	Anteil 2004	Anteil 2005	Anteil 2006	Anteil 2004-2006
Anzahl Artikel	68	78	85	231				
Nachweis in GS	2	1	1	4	2,9	1,3	1,2	1,7
Preprint	1	0	0	1	1,5	0,0	0,0	0,4
PDF kostenlos	0	0	0	0	0,0	0,0	0,0	0,0
PDF kostenpflichtig	0	0	0	0	0,0	0,0	0,0	0,0
Volltext gesamt	1	0	0	1	1,5	0,0	0,0	0,4
Password	2004	2005	2006	gesamt	Anteil 2004	Anteil 2005	Anteil 2006	Anteil 2004-2006
Anzahl Artikel	8	5	3	16				
Nachweis in GS	4	1	0	5	50,0	20,0	0,0	31,3
Preprint	0	0	0	0	0,0	0,0	0,0	0,0
PDF kostenlos	2	0	0	2	25,0	0,0	0,0	12,5
PDF kostenpflichtig	0	0	0	0	0,0	0,0	0,0	0,0
Volltext gesamt	2	0	0	2	25,0	0,0	0,0	12,5

# Effizienz ist keine Frage der Größe!



## FAUST

Das moderne Datenbank- und Retrievalsystem zur Archivierung, Strukturierung und Erschließung von Massendaten.

### Einzelplatz, Netzwerk, Intranet und Internet

- Flexible Datenstruktur und zahlreiche Musteranwendungen
- Breit einsetzbar in Archiv, Bild- und Medienarchiv, Dokumentation, Bibliothek, Museum



Weitere Infos im Netz: [www.land-software.de](http://www.land-software.de) oder bei LAND Software-Entwicklung, Postfach 1126, 90519 Oberasbach, Fax 0911-695173, [info@land-software.de](mailto:info@land-software.de)

theke- und Informationswissenschaft eignet. Keine der untersuchten Zeitschriften wird vollständig nachgewiesen, auch wenn dies durch die Bereitstellung der vollständigen Inhaltsverzeichnisse wenigstens bei manchen Zeitschriften ohne größere Probleme möglich wäre.

Für den Recherchierenden ist Google Scholar daher vor allem als schneller Weg zu den elektronischen Volltexten von Interesse. Während der Anteil der verfügbaren Volltexte als nicht besonders hoch einzuschätzen ist, kann sich eine Recherche in Google Scholar aber doch lohnen, bevor man sich auf den Weg in die Bibliothek macht oder einen Artikel per Fernleihe bestellt.

Betrachtet man die Bereitstellung der Original-PDF-Dateien durch die Verlage bzw. Fachgesellschaften, so zeigt sich der Weg der „moving wall“, den ZfBB und Bibliotheksdienst gehen, als gute Lösung. Dabei werden die Volltexte nach einer gewissen Zeit kostenlos zur Verfügung gestellt, ohne dass dadurch der Verkauf der aktuellen Zeitschriftenhefte gefährdet wird.

An sich vorbildlich ist der Weg von Information Wissenschaft und Praxis, welche die Original-PDF-Dateien in einem eigenen Portal zur Verfügung stellt. Allerdings können diese Inhalte von Suchmaschinen wie Google Scholar nicht erfasst werden, weil sie nur durch eine spezielle Suche gefunden werden können und

nicht ausreichend verlinkt sind. Hier könnte beispielsweise durch eine Browsingfunktionalität über die Inhaltsverzeichnisse Abhilfe geschaffen werden. Inhaltsverzeichnisse allein bringen jedoch keine Lösung des Problems: Sowohl BuB als auch BIT Online bieten Inhaltsverzeichnisse an, die jedoch von Google Scholar nicht erfasst werden. Allerdings besteht die Möglichkeit, solche Table-of-Contents-Dienste bei Google anzumelden.

Insgesamt könnte die kostenlose Zugänglichmachung von Volltexten (mit oder ohne Zeitverzug) aus dem LIS-Bereich die deutschsprachige Bibliotheks- und Informationswissenschaft stärken. Interdisziplinäre Angebote wie Google Scholar bieten die Möglichkeit, die Erkenntnisse des eigenen Fachs auch Vertretern anderer Fächer zugänglich zu machen. Da Google Scholar ein breit genutztes Angebot ist, sollte es sogar ein strategisches Ziel sein, die eigenen Aufsätze dort verfügbar zu machen, um ihre Sichtbarkeit über die Fachgrenzen hinaus zu erhöhen.

## Literatur

Adlington, J., & Benda, C. (2006). Checking under the hood: Evaluating Google scholar for reference use. In: *Internet Reference Services Quarterly*, 10(3-4), 135-148.

Bargmann, M.; Katzmayr, M; Putz, M.: E-LIS: Open-Access-Archiv für Literatur zum Informations- und Bibliothekswesen. In: *Online-Mitteilungen* 83(2005), 15-24.

Callicott, B., & Vaughn, D. (2006). Google scholar vs. library scholar: Testing the performance of schoogle. In: *Internet Reference Services Quarterly*, 10(3-4), 71-88.

Hauer, M. (2005): Portal Informationswissenschaft: DGI baut Wissenschaftsportal mit AGI und Hochschulen. In: *Information. Wissenschaft und Praxis* 56(2005)2, 71-76.

Jacsó, P. (2005). Google Scholar: The pros and cons. *Online Information Review*, 29(2), 208-214.

Lewandowski, D. (2005). Google Scholar – Aufbau und strategische Ausrichtung des Angebots sowie Auswirkung auf andere Angebote im Be-

reich der wissenschaftlichen Suchmaschinen. URL: [www.durchdenken.de/lewandowski/doc/Expertise\\_Google-Scholar.pdf](http://www.durchdenken.de/lewandowski/doc/Expertise_Google-Scholar.pdf) [10.3.2007].

Lewandowski, D. (2006). Suchmaschinen als Konkurrenten der Bibliothekskataloge: Wie Bibliotheken ihre Angebote durch Suchmaschinentechologie attraktiver und durch Öffnung für die allgemeinen Suchmaschinen populärer machen können. In: *Zeitschrift für Bibliothekswesen und Bibliographie*, 53(2), 71-78.

Lewandowski, D. (2007): Wie können sich Bibliotheken gegenüber Wissenschaftssuchmaschinen positionieren? In: *Wa(h)re Information. Tagungsband des 29. Österreichischen Bibliothekartags in Bregenz 2006* [im Druck]

Lewandowski, D., & Mayr, P. (2006). Exploring the Academic Invisible Web. In: *Library Hi Tech*, 24(4), 529-539.

Mayr, P., & Walter, A.-K. (2006). Abdeckung und Aktualität des Suchdienstes Google Scholar. In: *Information. Wissenschaft und Praxis*, 57(3), 133-140.

Phipps, S.E., & Maloney, K. (2006). Choices in the paradigm shift: Where next for libraries? In: *Internet Reference Services Quarterly*, 10(3-4), 103-115.

Scirus White Paper: How Scirus works. URL: [www.scirus.com/press/pdf/WhitePaper\\_Scirus.pdf](http://www.scirus.com/press/pdf/WhitePaper_Scirus.pdf) [10.3.2005].

Söllner, K.: Google Scholar und Windows Live Academic Search – aktuelle Entwicklungen bei wissenschaftlichen Suchmaschinen. In: *Bibliotheksdienst* 40(2006)7, 828-837.

Stock, W.G., & Schlögl, C. (2004). Impact and Relevance of LIS Journals: A Scientometric Analysis of International and German-Language LIS Journals – Citation Analysis Versus Reader Survey. In: *Journal of the American Society for Information Science and Technology*, 55(13), 1155-1168.

Suchmaschine, Bibliothekswissenschaft, Information und Dokumentation, Zeitschrift, Google Scholar, Fachgebietsabdeckung, Volltext

## DER AUTOR

### Prof. Dr. Dirk Lewandowski



ist Professor für Information Research und Information Retrieval an der Hochschule für Angewandte Wissenschaften Hamburg.

HAW Hamburg  
Fakultät Design, Medien und Information  
Department Information  
Berliner Tor 5, 20099 Hamburg  
[dirk.lewandowski@bui.haw-hamburg.de](mailto:dirk.lewandowski@bui.haw-hamburg.de)  
[www.durchdenken.de/lewandowski](http://www.durchdenken.de/lewandowski)

# Historische Rezensionen online: Eine thematische Suchmaschine von Clio-online

Daniel Burckhardt, Berlin

*Historische Rezensionen online (HRO) ist eine auf geschichtswissenschaftliche Rezensionen fokussierte Suchmaschine. Sie berücksichtigt ausschließlich fachwissenschaftliche Besprechungen, die im Web veröffentlicht wurden und ohne Zugangsbeschränkungen im Volltext zur Verfügung stehen. Eingebunden in das Portal des DFG-geförderten Kooperationsprojektes Clio-online sind über eine einfache Suchmaske gegenwärtig mehr als 35.000 Rezensionen zu über 30.000 Publikationen im Volltext recherchierbar. Statt unter hohem Aufwand Daten zu erstellen oder manuell zu katalogisieren, werden bei HRO bereits vorhandene Texte und bibliografische Daten voll automatisiert gesammelt, verknüpft und in einem zentralen Suchindex aufbereitet. Die Spezialisierung auf eine Textsorte und die Anreicherung der Rezensionstexte mit den in der Verbunddatenbank des GBV bereits erfassten bibliografischen Informationen erlauben zudem zielgerichtete Anfragen nach bestimmten Autoren oder eingeschränkt auf einzelne Themenbereiche, ein Mehrwert für die Nutzerinnen und Nutzer von HRO gegenüber kommerziellen Suchmaschinen.*

## Historical Reviews online – a subject search engine from Clio-online

*Historical Reviews online (HRO) is a search engine focussed on historical reviews. It is limited to academic reviews, which are published online without access restrictions. HRO is integrated into the Internet gateway of Clio-online, a cooperative project for historians supported by the DFG (German Research Foundation). Currently, over 35'000 reviews covering over 30'000 publications can be searched through a simple search field. Avoiding the labour-intensive tasks of content production or manual cataloguing, the central search index of HRO is built by a Web crawler that collects already existing texts and bibliographic data in an automated manner. The specialisation on a single text category and the aggregation of the reviews with the bibliographic information from the GBV Common Library Network enable precise queries for specific authors or limited to certain subject areas. This is a clear surplus for users of HRO compared to commercial search engines.*

## 1 Die Bedeutung thematischer Suchmaschinen

Die zentrale Rolle von Suchmaschinen bei der Informationsrecherche sowie die aktuelle Marktdominanz von Google führten in Frankreich und Deutschland zur Forderung nach öffentlichen Geldern zum Aufbau von Alternativen zu den bestehenden Anbietern.<sup>1</sup> Da die grundlegenden Funktionsabläufe und Algorithmen von Google im akademischen Betrieb entwickelt und erprobt und entsprechend publiziert wurden (Brin/Page 1998; Austin 2006), sind für das Gelingen eines solchen Projekts aber weniger Mittel für Grundlagenforschung sowie zur Implementierung entsprechender Suchalgorithmen als das Know-how sowie ein Geschäftsmodell für den Aufbau und dauerhaften Betrieb einer massiv-parallelen Suchinfrastruktur erforderlich. Dazu gehören Speicherkapazitäten von mehreren

Dutzend Terabyte sowie genügend Rechenleistung zur Bearbeitung von mehreren tausend Suchanfragen pro Sekunde (Langville/Meyer 2006; Brown 2003). Die Umsetzung eines entsprechenden Clusters ohne kommerzielle Interessen ist vermutlich nur dann realisierbar, wenn es gelingt, analog zu verteilten Rechenprogrammen wie Seti@home und Peer-to-Peer-Anwendungen wie BitTorrent die Speicher-, Rechen- und Übertragungskapazitäten auf jeden einzelnen Nutzer einer solchen Suchlösung aufzuteilen (Sietmann 2005).

Ganz anders präsentiert sich die Ausgangslage für spezialisierte Suchlösungen: Die Rechen- und Speicherkapazitäten eines handelsüblichen PC reichen für die Beantwortung einer Suchanfrage pro Sekunde über mehrere Millionen Dokumente. War vor wenigen Jahren noch teure Spezialsoftware von Anbietern wie Autonomy oder Verity erforderlich, sind heute die entsprechenden Algorithmen

und Abläufe gut dokumentiert (Manning/Schütze 1999; Bray 2003) und in Form der Lucene-Programmbibliothek unter der Apache-Lizenz frei verfügbar. Nach Beschränkung auf bestimmte Themen, Anbieter oder Textsorten sowie der Fokussierung auf einen eingeschränkten Nutzerkreis ist der Aufbau eines solchen Angebots somit auch für kleinere Projekte möglich und sinnvoll. Insbesondere wissenschaftliche Bibliotheken könnten bei der Erschließung von hochwertigen, aber bislang nur unzureichend erschlossenen Informationen aus dem „Invisible Web“ eine aktive Rolle übernehmen (Lössau 2004). Entscheidend dabei ist eine geschickte Wahl der zu indizierenden Daten: Die Auswahl sollte einerseits breit genug sein, damit die wichtigsten Bestände aus dem gewählten Bereich abgedeckt werden, andererseits auch eng genug, damit auch bei schlechteren Ranking-Algorithmen als denen von Google die auf den ersten Seiten angezeigten Treffer dem Nutzer relevant erscheinen. Die beiden Messgrößen dafür lauten „Recall“ und „Precision“: *Recall* misst, wie viele der für eine Suchanfrage relevanten Dokumente in der Treffermenge tatsächlich auftauchen, *Precision* erfasst, wie viele der angezeigten Treffer für die ausgeführte Suchanfrage tatsächlich von Bedeutung sind (Bray 2003).

Solange das Semantic Web, die systematische Anreicherung der Inhalte einer Web-Seite mit maschinenlesbaren Auszeichnungen, eine Vision mit ungewissen Realisierungschancen bleibt, gilt dabei:

■ Die wichtigsten Informationen über einen Text stecken im Text selbst. Effizienter als die aufwändige manuelle Erfassung von Metadaten über eine Web-Seite ist deshalb in der Regel die Volltextindizierung des Seiten-Inhalts. Einerseits sind die Nutzer durch die bekannten Suchmaschinen darauf konditioniert, insgesamt seltene, für ihre Fragestellung aber mit hoher Wahrscheinlichkeit im Text enthaltene Begriffe einzugeben, die nicht unbedingt im Titel der Ressource oder in manuell

<sup>1</sup> Siehe Jeanneney (2006) sowie das ausführliche Thesenpapier der Bundestagsfraktion von Bündnis 90/Die Grünen vom März 2005, <http://www.gruene-bundestag.de/cms/publikationen/dokbin/63/63265.pdf> [12.12.2006].

erfassten Schlagworten oder Klassifikatoren auftauchen. Andererseits ist die Berücksichtigung des Volltexts und nicht nur von Metadaten eine wichtige Voraussetzung für ein sinnvolles Ranking der Trefferlisten.

- Statt auf zukünftige maschinenlesbare Zusatzinformationen zu hoffen, lohnt es, nach bereits vorhandenen Informationen zu suchen, die für die Formulierung von Suchanfragen und zur Aufbereitung der Trefferlisten wertvoll sein könnten. So brachten bei der Web-Suche nicht etwa der Einsatz von Methoden der Künstlichen Intelligenz zur Erfassung und Klassifizierung von Seiteninhalten sondern eine systematische statistische Auswertung der durch Verweise zwischen den Seiten implizit kodierten Netzwerktopologie den Durchbruch (Bray 2003). Bei spezialisierten Suchlösungen können auch wesentlich einfachere Auswertungen der Seiteninhalte zum Ziel führen. Für die Umsetzung von HRO, das sich auf Buchrezensionen beschränkt, war die automatisierte Erkennung der besprochenen Titel anhand der in der entsprechenden Web-Seite ausgewiesenen ISBN entscheidend.

## 2 Historische Rezensionen online

H-Soz-u-Kult, einer der Projektpartner von Clio-online<sup>2</sup>, veröffentlicht jährlich rund 1.000 Besprechungen von Neuerscheinungen; über die eigene Web-Site sind mittlerweile über 5.000 Rezensionen einsehbar. Auch die Sehepunkte haben seit dem Start im November 2001 fast 4.000 Publikationen aus den Geschichtswissenschaften rezensiert.<sup>3</sup> Entsprechende Zahlen auch bei den wichtigsten Anbietern im englischen Sprachraum zeigen deutlich, dass die anfänglichen Vorbehalte, die sich sowohl in der fehlenden Bereitschaft zur Überlassung von Rezensionsexemplaren durch gewisse Verlage als auch in der Zurückhaltung potentieller Rezensentinnen und Rezensenten gegenüber der Veröffentlichung in einem ungewohnten Publikationsorgan äußerten, längst überwunden sind (Helmlinger 2003; Hohls 2004). Ähnliche Verschiebungen sind auch in benachbarten geisteswissenschaftlichen Disziplinen wie den Literaturwissenschaften zu beobachten, wo seit 1998 mit IASL online und literaturkritik.de zwei bedeutende deutschsprachige Angebote entstanden sind.

2 [www.clio-online.de/hro](http://www.clio-online.de/hro) [12.12.2006]

3 <http://hsozkult.geschichte.hu-berlin.de/rezensionen/>, [www.sehepunkte.de/](http://www.sehepunkte.de/) [12.12.2006]

4 Internationale Bibliographie der Rezensionen geistes- und sozialwissenschaftlicher Literatur, [www.gbv.de/gsomenu/opendb.php?db=2.14](http://www.gbv.de/gsomenu/opendb.php?db=2.14) [12.12.2006]

5 Als Nebeneffekt dieses Extraktionsvorganges wurden einige Dutzend falsch eingegebener ISBN bei verschiedenen Anbietern erkannt.

Die Idee für den Aufbau von HRO entstand aus der Beobachtung, wonach sich die erschöpfende Suche nach den unter dem Banner der freien Zugänglichkeit angetretenen Besprechungen im Web paradoxerweise fast aufwändiger gestaltete als nach solchen in gedruckten Fachzeitschriften, für die mit der IBR schon länger ein zwar kostenpflichtiges, aber in vielen Universitätsnetzen und wissenschaftlichen Bibliotheken freigeschaltetes Nachweissystem zur Verfügung steht.<sup>4</sup> Deshalb wurden im Projektantrag für die zweite Förderphase von Clio-online bei der DFG Personal- und Sachmittel für den Aufbau eines Nachweissystems für geschichtswissenschaftliche Rezensionen beantragt.

Erste Überlegungen für den kooperativen Aufbau einer entsprechenden Datenbank gingen von einer einheitlichen Metadaten-Auszeichnung der Rezensionen durch die verschiedenen Anbieter im HTML-Header nach Dublin Core aus. Die große Heterogenität der bereits erfassten Angaben, die sich besonders deutlich bei den bibliografischen Angaben der besprochenen Werke zeigt, hätte jedoch eine zeitaufwändige Nachbearbeitung erfordert und die aktive Mithilfe aller Kooperationspartner vorausgesetzt.

Die entscheidende organisatorische Vereinfachung brachte die Einsicht, dass in den Geisteswissenschaften nicht Einzelartikel, sondern fast ausschließlich Monografien und Sammelbände besprochen werden, die über eine eigene ISBN verfügen. Eine Durchsicht der wichtigsten Rezensionsanbieter zeigte, dass die ISBN zumindest bei neueren Besprechungen bei fast allen Angeboten mit aufgelistet wird. Da sowohl Amazon als auch Verbundkataloge von Bibliotheken Web-Services zur Abfrage von bibliografischen Angaben bereitstellen, reicht es aus, auf den Seiten der verschiedenen Anbieter die ISBN sowie – soweit möglich – Angaben zum Rezensenten sowie dem Veröffentlichungsdatum auszulesen. Die besonderen Eigenschaften einer ISBN als zehn- bzw. ab 2007 dreizehnstellige Zahl mit eingebauter Prüfziffer erlauben die automatisierte Extraktion dieser Angaben mit hoher Zuverlässigkeit.<sup>5</sup>

Ein erster „proof of concept“, der Rezensionen von drei Anbietern einsammelte und über Amazon mit den entsprechenden bibliografischen Angabe anreicherte, zeigte schnell die grundsätzliche Machbarkeit dieses Vorgehens. Bei dieser ersten Zusammenführung von rund 10.000 Rezensionen im Volltext zeigte sich auch die hohe Bedeutung eines guten Rankings von größeren Treffermengen für die Nutzerakzeptanz. Die Volltextsuche des MS SQL Server 2000 wurde deshalb durch Lucene, eine Programmbibliothek für Volltextsuchen aus dem Apache-Projekt ersetzt [Gospodnetić/Hatcher 2005, Naber 2005]. Problematisch für ein öffent-

lich gefördertes Projekt waren zudem die Nutzungsbedingungen des Amazon E-Commerce Service. Dieser Dienst ist zwar kostenfrei, verlangt im Gegenzug jedoch bei jeder Anzeige von darüber abgefragten Produktinformationen einen Link auf die entsprechende Detailseite oder die Startseite von Amazon. Dank Unterstützung des GBV konnten für die öffentliche Version der Rezensionsdatenbank stattdessen die wesentlich detaillierteren bibliografischen Daten aus dem PICA-Verbundkatalog eingebunden werden.

Das Interface folgt den üblichen Erwartungen an einen solchen Dienst: eine Textbox zur Eingabe von einem oder mehreren Suchbegriffen, darunter ein „Suchen“-Knopf, der zu einer neuen Seite mit einer nach Relevanz geordneten Resultatliste führt (Nielsen 2005), (siehe Abbildung 1).

Obwohl erweiterte Suchen nur von sehr wenigen Nutzern genutzt werden (Bray 2003), wurde die einfache Suche später durch eine Eingabemaske ergänzt, die gezielte Recherchen nach Autoren, Rezensenten oder nur im Titelfeld erlaubt. Bei jedem Treffer werden Titel, Autor/innen, Ort und Jahr des Buchs, ein kurzes Exzerpt aus dem Rezensionstext, das den Suchbegriff enthält, sowie die URL zum Volltext angezeigt. Ein Klick auf den Titel der Publikation führt zu einer Detailansicht, die nicht nur eine vollständige Übersicht sämtlicher Besprechungen dieses Werks bietet, sondern auch Zusatzangebote wie etwa gescannte Inhaltsverzeichnisse verknüpft und vordefinierte Suchabfragen im KVK, der Clio-Metasuche sowie im GBV zur Bestandsabfrage in lokalen Bibliotheken oder im Buchhandel ermöglicht (siehe Abbildung 2).

## 3 Die technischen Schritte im Einzelnen

Die Implementierung von HRO folgt dem Design-Prinzip einfacher, sequentiell hintereinander ausgeführter Programme, wobei die Zwischenergebnisse im Gegensatz zu üblichen UNIX-Filtern nicht reine Textdateien, sondern in einem SQL-Server abgelegte Datensätze in zwei Tabellen – Review und Publication – sind. Verknüpft werden diese beiden Tabellen über die ISBN, die als Primärschlüssel in der Publication-Tabelle dient.

Aufgrund der ausgefeilten Unterstützung regulärer Ausdrücke, Konvertierungsfunktion für eine Vielzahl von Zeichensatzkodierungen sowie einer Reihe äußerst hilfreicher CPAN-Module wurden sämtliche Programmschritte bis auf die Textindizierung in Perl realisiert.

Die ersten drei Schritte implementieren über die LWP-Module eine Familie von speziell an die einzelnen Anbieter von Rezensionen oder bibliografischen Angaben angepassten Web-Robots. Diese übernehmen:

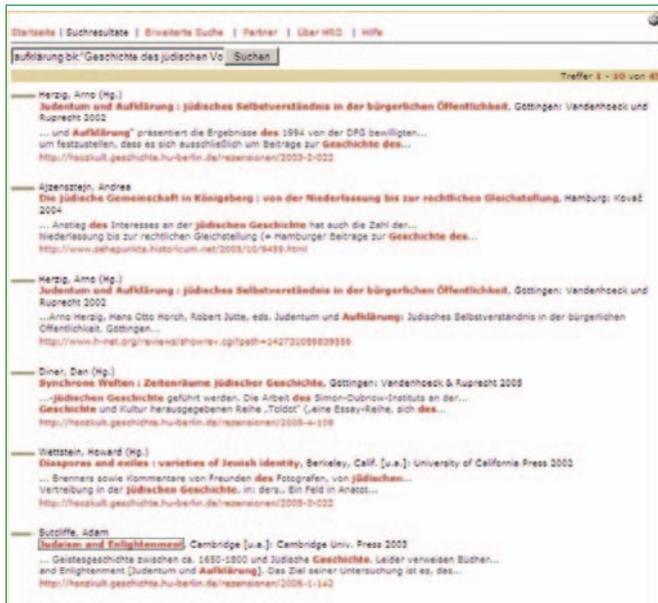


Abbildung 1: Trefferliste zu einer Suchanfrage nach „Aufklärung“ eingeschränkt auf Besprechungen zu Publikationen mit dem Basisklassifikator „15.96 Geschichte des jüdischen Volkes außerhalb des Staates Israel“.



Abbildung 2: Detailansicht mit Verweisen zu zwei Besprechungen sowie Links zu weiteren Suchdiensten.

1. **Einsammeln der URLs der einzelnen Rezensionen:** Aus den meist chronologisch geordneten Archivseiten werden die URLs sämtlicher Rezensionen eines Jahrgangs oder einer Ausgabe extrahiert. Für jede besprochene Publikation wird diese URL, die ISBN und – soweit vorhanden – zusätzliche Metadaten wie etwa der Rezensent oder Ausgabe oder Datum der Besprechung als Datensatz in die Review-Tabelle eingefügt.
2. **Einsammeln der Rezensionen im HTML-Format:** Für jeden im ersten Schritt angelegten Datensatz wird der Seiteninhalt der angegebenen URL abgerufen und im entsprechenden Feld ergänzt.
3. **Abfrage der bibliografischen Informationen anhand der ISBN:** Für sämtliche ISBN in der Review-Tabelle, für die noch kein entsprechender Eintrag in der Publication-Tabelle steht, wird im GBV oder bei Amazon ein bibliografischer Datensatz abgefragt und im MODS-Format<sup>6</sup> in die Publication-Tabelle eingefügt.

Vor der Indizierung erfolgt in einem Zwischenschritt für sämtliche Review-Einträge eine:

4. **Transformation der Rezensionen vom HTML-Format nach Klartext:** Dabei werden nicht zur eigentlichen Rezension gehörige Seitenbereiche wie zum Beispiel eine Navigationsleiste entfernt, bevor über das Modul HTML::FormatText die Konvertierung nach Klartext ohne HTML-Auszeichnungen erfolgt.

Keiner dieser Schritte bietet grundlegende technische Schwierigkeiten. Detail-Arbeit entstand aber durch die Notwendigkeit

6 MODS: Metadata Object Description Schema (XML-Schema für bibliografische Datensätze), [www.loc.gov/standards/mods/](http://www.loc.gov/standards/mods/) [12.12.2006].



OCLC PICA

At the heart of your information

## Volltreffer – was will man mehr



Gönnen Sie Ihren Benutzern dieses Gefühl. Weg von traditionellen Rechercheeinstiegen hin zu modernster Suchtechnologie, wie man sie aus dem Internet kennt.

Die Integration von FAST-Suchtechnologien in die verschiedenen SIS-Produktlinien eröffnet diesen Weg. Modernste Technologie zur Indexierung und Recherche von Katalogdaten und Volltexten verbunden mit effizienter Benutzerführung (Systematik-Filter, Drill Downs) und komfortabler Rechercheunterstützung (linguistische Werkzeuge).

**FAST-Suchtechnologien** der nächsten Generation für **SISIS-SunRise** und **SISIS-Elektra**

OCLC PICA GmbH  
Grünwalder Weg 28g  
82041 Oberhaching  
t +49-(0)89-613 08 300  
f +49-(0)89-613 08 399  
e [info@oclc-pica.org](mailto:info@oclc-pica.org)

[www.oclc-pica.org](http://www.oclc-pica.org)

der Konvertierung der verschiedenen Zeichensätze sowie HTML-Entitäten nach UTF-8<sup>7</sup> und durch die Variabilität der Einzel-Rezensionen bei all jenen Anbietern, die ihre Web-Präsenz nicht dynamisch aus einer Datenbank generieren. Die aufwändigste Prozedur bei der Abfrage bibliografischer Informationen war wiederum die Konvertierung der Daten vom GBV aus dem Pica-Zeichensatz nach UTF-8, die den manuellen Aufbau einer entsprechenden Konversionstabelle erforderte.

Nach diesen Schritten stehen sämtliche Daten bereit, zum:

5. *Aufbau des Volltextindexes*: Durch Aufrufe der Lucene-API wird ein gefelderter Volltextindex über den Rezensionstext sowie die bibliografischen Informationen der entsprechenden Publikation aufgebaut. Da der Index sowohl deutsch- als auch englischsprachige Besprechungen umfasst, wurde auf ein sprachabhängiges „Stemming“ zur Reduktion morphologischer Varianten eines Wortes auf einen gemeinsamen Wortstamm verzichtet.

Das Nutzer-Interface für die Recherche in HRO wurde als Web-Service in das Portal von Clio-online sowie über eine am OAI-Protokoll orientierte Schnittstelle in die Metasuche eingebunden.

#### 4 Akzeptanz und Perspektiven

Seit der Ankündigung von HRO über H-Soz-u-Kult Anfang August 2005 erwies sich die Rezensionsdatenbank als eines der meist genutzten Angebote im Web-Portal von Clio-online. Aus der Auswertung der Logfiles über 32.000 Suchanfragen in den ersten sechs Monaten ergeben sich deutliche Hinweise auf die Interessen der Nutzerinnen und Nutzer. Knapp 12.000 unterschiedliche, fast ausschließlich deutschsprachige Begriffe wurden ins Suchfeld eingegeben. Unter

7 Neben UTF-8 verwenden die eingebundenen Rezensions-Anbieter den west- und osteuropäischen sowie den Windows-Zeichensatz (ISO 8859-1, ISO 8859-2, Windows-1252). Zu den Problemen im Umgang mit verschiedenen Sprachen und Zeichensätzen in Suchanwendungen siehe auch [www.tbray.org/ongoing/When/200x/2003/10/11/SearchI18n](http://www.tbray.org/ongoing/When/200x/2003/10/11/SearchI18n) [12.12.2006].

8 Die häufigsten Begriffe (in Klammern die Zahl der Abfragen) waren: geschichte (373), weltkrieg, deutsche (81), mittelalter, ddr (78), nationalsozialismus (76), mommsen (75), revolution (73), deutschland (71), zeitgeschichte, aly (64), burgund (61), krieg (54), europa (53).

9 [www.oclc.org/research/projects/xisbn/](http://www.oclc.org/research/projects/xisbn/) [12.12.2006]

10 <http://samba.org/ftp/unpacked/junkcode/spamsum/README> [12.12.2006]

11 Für eine vollständige Liste der im GBV benutzten Basisklassifikation siehe [www.gbv.de/vgm/info/mitglieder/02Verbund/01Erschliessung/02Richtlinien/05Basisklassifikation/index](http://www.gbv.de/vgm/info/mitglieder/02Verbund/01Erschliessung/02Richtlinien/05Basisklassifikation/index) [12.12.2006].

den besonders häufig nachgefragten Suchbegriffen finden sich sowohl Epochenbezeichnungen, Themen, geografische Bezeichnungen als auch Autoren- und Personennamen.<sup>8</sup> Eine ähnliche Mischung ergibt sich auch bei einem zufällig gewählten Ausschnitt, der sowohl mehrfach als auch nur ein einziges Mal eingegebene Begriffe umfasst.

Die größte Schwäche des aktuellen Datenmodells mit einem separaten Eintrag in der Review-Tabelle für jede ISBN zeigt sich im Dubletten-Handling: Besonders englischsprachige Publikationen erscheinen häufig zeitgleich in verschiedenen Ausgaben (Hardcover, Paperback) und Verlagen (USA, England) unter entsprechend vielen verschiedenen ISBN. Durch eine passende Abfrage des xISBN-Dienstes<sup>9</sup> sollten die daraus resultierenden Mehrfachanzeigen künftig vermieden werden. Über einen geeigneten „fuzzy-hash“-Algorithmus<sup>10</sup> können auch Mehrfachanzeigen von Sammelrezensionen, die für jeden besprochenen Titel unter einer separaten URL abgelegt werden, unterdrückt werden.

Besonders sinnvoll scheint die Erweiterung des Such-Dialogs um eine geeignete Nutzerschnittstelle für die Abfrage nach den bei vielen Publikationen im GBV mitgeführte Basisklassifikatoren.<sup>11</sup> Diese werden bislang nur über Anfragen der Form `bk:Historiographie` unterstützt. Kombiniert mit geeigneten Sortier- und Export-Optionen für Trefferlisten, könnten so thematische Bibliografien angeboten werden, die sich direkt in Literaturverwaltungen wie EndNote importiert lassen. Die technisch wenig aufwändige Aufbereitung solcher Anfragen in Form eines nach Datum der Besprechung geordneten RSS-Feeds würde auch die Beobachtung von neuen Veröffentlichungen zu einem Themenbereich stark erleichtern.

Durch die weitgehende Automatisierung der Datensammlung und -aufbereitung ist die regelmäßige Aktualisierung von HRO auch nach Auslauf der DFG-Förderung gewährleistet. Zur langfristigen Sicherung des Dienstes sollte mittelfristig die Integration in die technische Infrastruktur eines Bibliotheksverbunds angestrebt werden.

#### Literatur

Austin, D. (2006): How Google Finds Your Needle in the Web's Haystack. In: Feature Column of the American Mathematical Society, [www.ams.org/featurecolumn/archive/pagerank.html](http://www.ams.org/featurecolumn/archive/pagerank.html).

Bray, T. (2003): On Search, the Series. In: [ongoing.tbray.org/ongoing/When/200x/2003/07/30/OnSearchTOC](http://ongoing.tbray.org/ongoing/When/200x/2003/07/30/OnSearchTOC).

Brin, S.; Page, L. (1998): The anatomy of a large scale hypertextual web search engine. In: Proc. 7th WWW, [www.db.stanford.edu/~backrub/google.html](http://www.db.stanford.edu/~backrub/google.html).

Brown, D. J. (2003): A Conversation with Wayne Rosing. In: ACM Queue vol. 1 (6, September), [www.acmqueue.org/modules.php?name=Content&pa=showpage&pid=62](http://www.acmqueue.org/modules.php?name=Content&pa=showpage&pid=62).

Gospodnetić, O.; Hatcher, E. (2005): Lucene in Action. Greenwich, Conn.: Manning.

Helmlinger, P. (2003): Historische Rezensionen im Internet. Entwicklung – Potentiale – Chancen. In: *zeitenblicke* 2 (Nr. 2), [www.zeitenblicke.de/2003/02/helmlinger.htm](http://www.zeitenblicke.de/2003/02/helmlinger.htm).

Hohls, R. (2004): H-Soz-u-Kult. Kommunikation und Fachinformation für die Geschichtswissenschaften. In: *Historical Social Research/Historische Sozialforschung*, 29(1), 212-232, [http://hsr-trans.zhsf.uni-koeln.de/hsrretro/docs/artikel/hsr/hsr2004\\_609.pdf](http://hsr-trans.zhsf.uni-koeln.de/hsrretro/docs/artikel/hsr/hsr2004_609.pdf).

Jeanneney, J.-N. (2006): Googles Herausforderung. Für eine europäische Bibliothek. Berlin: Wagenbach.

Langville, A.; Meyer, C. (2006): Google's pagerank and beyond. The science of search engine rankings. Princeton: Princeton University Press.

Lossau, N. (2004): Search Engine Technology and Digital Libraries. Libraries Need to Discover the Academic Internet. In: *D-Lib Magazine*, 10 (6), [www.dlib.org/dlib/june04/lossau/06lossau.html](http://www.dlib.org/dlib/june04/lossau/06lossau.html).

Manning, D.; Schütze, H. (1999): Foundations of Statistical Natural Language Processing. Cambridge, MA: MIT Press.

Naber, D. (2005): Herr der Suche. Eigene Anwendungen mit Volltextsuche erweitern. In: *c't* 7/2005, 196.

Nielsen, J. (2005). Mental Models For Search Are Getting Firmer. In: *Alertbox* (May 9, 2005), [www.useit.com/alertbox/20050509.html](http://www.useit.com/alertbox/20050509.html).

Sietmann, R. (2005): Wider die Monokultur. P2P-Strategien gegen die Suchmaschinen-Monopolisierung. In: *c't* 16/2005, 52, [www.heise.de/ct/05/16/052/](http://www.heise.de/ct/05/16/052/).

Suchmaschine, Buchbesprechung, Clio-online, Maschinelles Indexierungsverfahren, Bibliothekskatalog, Nutzung, Geschichte, Datenbank

#### DER AUTOR

Dipl. Math. ETH Daniel Burckhardt, M.A.



war Wissenschaftlicher Mitarbeiter an der Philosophischen Fakultät I der Humboldt-Universität zu Berlin. Er war verantwortlich für die technische Konzeption und Realisierung von Historische Rezensionen online.

Institut für Geschichtswissenschaften  
Bereich Historische Fachinformatik  
Humboldt-Universität zu Berlin  
Unter den Linden 6, 10099 Berlin  
[burckhardt@geschichte.hu-berlin.de](mailto:burckhardt@geschichte.hu-berlin.de)

# Sechs Jahre PsychSpider: Aus der Praxis des Betriebs einer Psychologie-Suchmaschine für freie Web-Inhalte

Erich Weichselgartner und Christiane Baier, Trier

Seit Sommer 2000 betreibt das Zentrum für Psychologische Information und Dokumentation (ZPID) den PsychSpider, eine Spezialsuchmaschine für wissenschaftliche Inhalte zur Psychologie. Der PsychSpider ist vorwiegend eine Websuchmaschine, aber auch strukturierte bibliographische Daten werden indiziert. Gegenwärtig bietet der PsychSpider den integrierten Zugang zu ca. 800.000 Dokumenten aus dem Surface Web und dem Deep Web. Die Suchmaschine verfügt über zahlreiche fortschrittliche Merkmale wie Boolesche Operatoren, natürlichsprachige Eingaben, „Query-by-Example“, Relevanzbewertung der Resultate, Strukturierung des Datenbestands in unterschiedliche Kollektionen, im- und exportierbare Indexe, parametrische Suche und automatische Kategorisierung. Die mehrjährige Erfahrung zeigt, dass auch eine Technologie, die eigentlich dazu dienen sollte, möglichst automatisch die relevanten Informationen aus den Myriaden von Dokumenten im Internet herauszufischen, umfängliche menschliche Betreuung braucht. Das liegt an der formalen und inhaltlichen Qualität von Inhalten, aber auch an technologischen Faktoren wie zum Beispiel Content Management Systemen. Die Nutzung des PsychSpider ist über die Jahre kontinuierlich gewachsen und er gehört zu den erfolgreichsten Web-Angeboten des ZPID. Aus strategischer Sicht ist eine nach den Kriterien einer wissenschaftlichen Fachinformationseinrichtung betriebene Suchmaschine unverzichtbar und der weitere Ausbau wird die Attraktivität deutlich erhöhen.

## **PsychSpider: six years experience with a psychology search engine for web content**

Since the summer of 2000, the Institute for Psychology Information (ZPID) maintains the Psychology search engine PsychSpider. The search engine indexes Web content as well as bibliographic (meta-) data. Approx. 800,000 documents from the surface Web and the deep Web are accessible from a single point of entry. The search offers advanced features: Natural language, Boolean and parametric searches, concept querying, ranking of documents by relevance, indexing data into different collections and automatic categorization, to name a few. Several years of experience demonstrate that search engine technology alone is not sufficient to sift through the Web for relevant information but that a sizeable amount of human interaction is required. The reasons are the quality of documents (in terms of adherence to standards and content) and technical innovations like content management systems. The usage of PsychSpider has grown over the years and it belongs to the most successful products of ZPID. From a strategic point of view, a search engine is mandatory for a Psychology information center and planned improvements will enhance user experience.

## 1 Einleitung

„Europa droht im Wettlauf mit Google und Yahoo ein Hightech-Fiasko“ lautet nach dem Ende der deutsch-französischen Kooperation bei Quaero die Schlagzeile von SPIEGEL ONLINE.<sup>1</sup> Damit ist wieder einmal die eminente Bedeutung von Suchmaschinen für die Informationsgesellschaft ins Licht der Öffentlichkeit gerückt, nachdem im Deutschen Bundestag schon vor zwei Jahren die Besorgnis geäußert wurde, allgemeine Suchmaschi-

nen könnten unsachgemäß beeinflussen, welche Informationen im digitalen Zeitalter gefunden werden und welche nicht (Unterausschuss Neue Medien, 17.6.2004). Eher unbemerkt von der breiten Öffentlichkeit bietet das Zentrum für Psychologische Information und Dokumentation (ZPID) schon seit Sommer 2000 unter dem Namen PsychSpider eine disziplinspezifische Suchmaschine für die Psychologie an<sup>2</sup>, die von der Fachcommunity rege genutzt wird (ca. 10.000 Zugriffe pro Monat). Vier Jahre vor dem Börsengang von

Google und der Sorge des Deutschen Bundestags um die Informationsfreiheit hat das ZPID als Fachinformationsanbieter eine Suchmaschine auf den Markt gebracht, welche nach wissenschaftlichen Kriterien ausgewählte nationale und internationale Psychologie-relevante Web-Inhalte indiziert. Mit zunehmender Digitalisierung von wissenschaftlicher Information und deren Verfügbarmachung im World-Wide-Web reicht es für eine Fachinformationseinrichtung nicht mehr aus, nur gedruckte Werke in bibliographischen Datenbanken nachzuweisen. Neben die traditionelle Print-Publikation in begutachteten Zeitschriften ist mit der Verbreitung des Internet eine Fülle neuer wichtiger wissenschaftlicher Kommunikationskanäle getreten. Zudem ist es praktisch, neben den eigentlichen wissenschaftlichen Quellen arrondierende Informationen schnell und einfach abrufen zu können: Institutsanschriften, Anfahrtspläne, E-Mail-Adressen, persönliche Homepages, Förderprogramme, Unterrichtseinheiten, Tagungs- und Kongressinformationen, Stellenangebote usw. Nicht zuletzt ist es ein immer wieder geäußerter Kundenwunsch, von einer einzigen Stelle aus alle relevanten Informationen abrufen zu können. Hier liegt der zusätzliche Vorteil einer Fachinformationseinrichtung, die dafür ihre Daten aus dem „Deep Web“ exklusiv zur Verfügung stellen und nach definierten Qualitätskriterien mit den frei verfügbaren Daten aus dem „Surface Web“ zu einem integrierten Informationsangebot zusammenfügen kann. Umgekehrt ist es für die Produzenten von Information heute enorm wichtig, eine faire Chance in der Trefferanzeige von Suchmaschinen zu erhalten: Das entscheidet über die Visibilität ihrer Werke. Auf den Punkt gebracht kann man sagen: Welche Informationen von einer Suchmaschine gefunden und wie sie präsentiert werden (Ranking), befindet über das Wohl und Weh im Informationszeitalter (Weichselgartner 2006). Eigentlich geht

1 Meldung vom 19.12.2006. [www.spiegel.de/netzwelt/web/0,1518,455558,00.html](http://www.spiegel.de/netzwelt/web/0,1518,455558,00.html) [20.2.2007]

2 [www.zpid.de/PsychSpider.php](http://www.zpid.de/PsychSpider.php) [20.2.2007]

es in der (klassischen) Fachinformation um nichts anderes, als alle zu einer gegebenen Fragestellung gehörigen (wissenschaftlichen) Dokumente nachzuweisen. Wenn im Zuge der Internet-Revolution (kein Massenmedium hat sich je schneller verbreitet, die Nutzer sind gleichzeitig Geber und Nehmer) und der Open-Access-Bewegung immer mehr wissenschaftliche Informationen frei im Internet verfügbar sind, müssen die Fachinformationswerkzeuge dem angepasst werden. Die Mitarbeiter eines Fachinformationszentrums sind erfahren in der Selektion von einschlägig relevanten Werken und können auf Web-Inhalte dieselben Kriterien anwenden wie für Print-Produkte. Allerdings erhalten vom Ressourcen-Einsatz her die gedruckten Werke nach wie vor das deutlich höhere Gewicht, weil es noch keine verbindlichen Qualitätsmaßstäbe für Internet-Publikationen gibt, während im Print-Bereich der Peer Review als Maßstab dient. Die anfängliche Hoffnung, eine Suchmaschine könnte mittels geschickter Steuerung aus den vielen Milliarden Internet-Dokumenten weitgehend automatisch gute und fachlich relevante Inhalte herausfiltern, hat sich (vorerst) zerschlagen. So muss man konstatieren, dass auch eine Suchmaschine fachkundiges menschliches Eingreifen braucht, und zwar mehr als man mitunter leisten kann. Vielleicht deshalb ist der PsychSpider nach wie vor weltweit die einzige Psychologie-Suchmaschine. Selbst die American Psychological Association (APA), ein 150.000 Mitglieder zählender amerikanischer Psychologie-Verband, der gleichzeitig einer der weltweit größten Fachverlage ist, kommt mit seinem PsycCrawler<sup>3</sup> nicht voran. Allerdings scheut die APA nicht nur den hohen Aufwand, sondern auch die Konkurrenz zu den eigenen kostenpflichtigen Datenbankprodukten (Quelle: persönliche Kommunikation). Die ebenfalls in den USA angesiedelte und teilweise werbefinanzierte Suchmaschine Psycsearch<sup>4</sup> kommt für wissenschaftliche Fragestellungen nicht in Betracht. Sie dient kommerziellen Interessen, zum Beispiel der Vermittlung von Therapie- und Beratungsangeboten.

## 2 Produktevaluierung

Im Jahr 2000 war der Ressourcenbedarf noch nicht abschätzbar, aber man wollte potentiellen konkurrierenden Angeboten zuvorkommen und rasch eine gebrauchstüchtige Suchmaschine anbieten. Deshalb hat sich das ZPID für den Zukauf ausgeleitet und am Markt etablierter Suchmaschinentechologie entschieden. Die Wahl fiel unter Berücksichtigung mehrerer Studien, unter anderem einer verglei-

chenden Evaluierung von Alta Vista, Excite, Verity, Open Text Live Link und Netscape Compass Server durch das Fraunhofer-Institut für Software und Systemtechnik (ISST in Berlin) auf Verity Search der Firma Verity Deutschland GmbH (heute Autonomy). Nach einer IDC-Studie war Verity Marktführer und ist es nach eigenen Angaben noch heute (16.000 Kunden weltweit, darunter seit Mitte Dezember 2006 die amerikanische Library of Congress). Folgende Gründe waren im Jahr 2000 maßgeblich: Verfügbarkeit auf Sun-Systemen (Solaris), Einbindbarkeit in Web-Server, Mehrsprachigkeit (Sonderzeichen), schneller Indexaufbau, Indexierung der wichtigsten Dokumentenformate (HTML, ASCII, PDF, DOC), Datenbankschnittstelle, remote Indexierung, komplexe Ausdrücke für Suchanfragen, mehrere Indexdateien, anwendungsbezogene Index-Einschränkung, Index im- und exportierbar und Relevanzbewertung der Resultate. Darüber hinaus offeriert Verity eine umfangreiche Skriptsprache für die Anpassung der Abfrageformulare bzw. der Ergebnisaufbereitung. Die Software bietet die Möglichkeit, den Datenbestand in mehrere sogenannte „Collections“ (Index-Daten) zu strukturieren und nur bestimmte Collections zur Suche zuzulassen. Damit sind anwendungsspezifische Suchen relativ einfach implementierbar. Wichtigste Komponente für die remote Indizierung (also die Indizierung fremder Web-Angebote) ist ein Agent oder Spider, der automatisch nach neuen oder aktualisierten Informationen sucht.

Neben die eher technisch orientierten Auswahlkriterien treten noch administrative, vor allem der Personalbedarf für Wartung und Betrieb. Durch die Verfügbarkeit unter Solaris-Systemen der Firma Sun Microsystems ist für eine wartungsarme Plattform gesorgt, aber auch die Handhabung der Suchmaschine selbst sollte nicht unnötig aufwendig oder komplex sein. So ließ sich in der Tat ein erster Prototyp binnen weniger Tage von einem Nicht-Informatiker implementieren und die Beta-Version eines Basis-Systems bereits nach wenigen Wochen der Öffentlichkeit anbieten.

## 3 Produktinstallation

Lizenziert wurde im Juni 2000 der Verity Search'97 Information Server für Internet/Intranet-Benutzung und 1 CPU. Die Installation verlief rasch und problemlos, lediglich bei der Sicherheit musste nachgebessert werden: Der Server sollte ursprünglich mit Administratorenrechten (Unix-User „root“) laufen und sämtliche Dateien hatten Schreibberechtigung für jedermann (Unix-Nutzergruppe „others“). Nach entsprechenden Korrekturen bzw. dem Anlegen eines dedizierten Nutzer-

Kontos verlief die Einbindung in den Web-Server (Apache) wiederum problemlos. Die meisten Administrationsarbeiten können über die Kommandozeile oder über ein Web-Interface erledigt werden, manche allerdings nur über das Web-Interface. Dazu gehört der „Collection Manager“, mit dem separate Indexe angelegt werden können. Deren Freigabe (oder Sperrung) erfolgt ebenfalls über das HTML-Interface. Das Neu-Anlegen oder Aktualisieren von Indexen funktioniert analog und kann auch zeitgesteuert erfolgen.

## 4 Produktkonfiguration

Für die Einbindung von Suchmasken und Trefferlisten in das Web-Angebot des ZPID wurden vorhandene Muster („templates“) in Anlehnung an das Corporate Design der Einrichtung umgeschrieben. Auch das verlief zügig und reibungslos. Die Suchmaschine unterstützt Boolesche Operatoren, natürlichsprachige Eingaben und sog. „query-by-example“-Abfragen. Auf Grund der Usancen verbreiteter Suchmaschinen, Benutzereingaben standardmäßig mit UND zu verknüpfen, erweist sich die Verity-Voreinstellung ODER als zunehmend problematisch und muss demnächst geändert werden. Die Ausgabe der Treffer erfolgt sehr übersichtlich und kann allen Bedürfnissen angepasst werden. Voreingestellt sind Dokumenttyp, Ranking-Faktor, Titel und drei Zeilen Text des Dokuments. Das Ranking erfolgt nach Relevanz, wobei der Algorithmus die Häufigkeit des Vorkommens von Suchbegriffen im Dokument bzw. deren Platzierung an ausgewiesenen Stellen im Dokument berücksichtigt (URL, Titel, Schlagwortliste, Description-Tag, usw.). Das ZPID zeigt in der aktuellen Trefferliste Ranking-Faktor, Titel, URL und Beschreibung an (Description-Tag, falls vorhanden, sonst Anfang des HTML-Bodies). Am Anfang und am Ende der Trefferliste werden Navigationshilfen sowie die Zahl der gefundenen Treffer und die Gesamtzahl der indizierten Dokumente angezeigt. Verity kann die verwendeten Suchbegriffe in der Anzeige hervorheben („keyword highlighting“), was das Auffinden relevanter Stellen im Dokument enorm erleichtert. Ein weiteres nützliches Merkmal ist die Möglichkeit, nach bestimmten Dokumenttypen suchen zu können, zum Beispiel nur nach PDF.

## 5 Produktaktualisierungen

Wie von einem kommerziellen Software-Hersteller erwartet, werden die Produkte laufend aktualisiert und erneuert. Für die laufenden Aktualisierungen bzw. Fehlerbereinigungen sowie die Inanspruchnahme des Supports sind jährliche War-

3 www.psychcrawler.com [20.2.2007]

4 www.psycsearch.com [20.2.2007]

# Psychologie Information

ZPID

Das Psychologie Webportal

**Wissen  
Wo  
Was  
steht**

... **[www.zpid.de](http://www.zpid.de)**

Zentrum für Psychologische Information und Dokumentation (ZPID)  
Universität Trier • Universitätsring 15 • 54296 Trier  
Fon: +49-(0)651-201-2877 • Fax: +49-(0)651-201-2071  
E-Mail: [info@zpid.de](mailto:info@zpid.de) • Mitglied der Leibniz-Gemeinschaft

tungsgebühren zu entrichten, die einen festen Prozentsatz des Kaufpreises ausmachen. Neue Releases dagegen müssen komplett neu gekauft werden. Dies war im Jahre 2003 nötig, als die Firma das Ende der Unterstützung für Search'97 ankündigte und ein Nachfolgeprodukt präsentierte: Die K2 Enterprise Application Suite mit verbesserter Unterstützung europäischer Sprachen. Zwar gewährt der Hersteller Rabatt für Bildungseinrichtungen (das ZPID ist an der Universität Trier angesiedelt), trotzdem schlagen die Kosten nicht unerheblich zu Buche. Für die langfristige Planung ist es daher wichtig, sich nicht nur über die Wartungskosten, sondern auch über weitere Folgekosten wie Releasewechsel und Consulting im Klaren zu sein. Ebenso können unerwartete Ereignisse die Investition gefährden, wie der Konkurs oder die Übernahme eines Unternehmens. Der Zusammenschluss von Verity und Autonomy im Jahre 2006 hat sich in dieser Hinsicht nicht negativ ausgewirkt, die Produkte werden weiter gepflegt. Erfreulicherweise stehen seit dem letzten Release mittlerweile im Rahmen der K2-Lizenz sogar drei Spider zur Auswahl: Der ursprüngliche *Vspider* (Search'97), der *K2spider* (K2 Enterprise) und *Ultraseek* (von Inktomi im Jahr 2003 übernommen). Abhängig von der Art des Web-Servers und vom Dokumenttyp kann es vorkommen, dass nur einer der drei Spider ein Dokument erfolgreich indizieren kann (alle Spider können den gleichen Index füttern). Konzeptionell ist der *Vspider* für den Kommandozeilen-Betrieb ausgelegt, während die anderen beiden über eine Web-Eingabemaske bedient werden („Dashboard“). Der Trend hin zu graphischen Benutzungsoberflächen auch in der Administration ist wohl ein Zug der Zeit.

## 6 Abgrenzung von anderen Wissenschaftssuchmaschinen

Das ZPID scheint nach wie vor die einzige Einrichtung zu sein, die versucht, eine Suchmaschine mit disziplinspezifischen freien Web-Inhalten aufzubauen. Die Betreiber der multidisziplinären Bielefeld Academic Search Engine (BASE) zum Beispiel konzentrieren sich auf strukturierte Informationen, ebenso wie das HBZ in

Köln als Betreiber des nationalen multidisziplinären Wissenschaftsportals Vascode.

Für BASE<sup>5</sup> werden OAI-Metadaten, lokale Datenbestände der Bibliothek und ausgewählte Webseiten indiziert. Bei den OAI-Metadaten handelt es sich um standardisierte bibliographische Beschreibungen von Open-Access-Publikationen, die in wissenschaftlichen Repository-Servern abgelegt sind (das Akronym OAI steht für Open Archives Initiative). Mit den Datenbeständen der Bibliothek ist der OPAC gemeint, der im Gegensatz zu einer Fachdatenbank fachlich weniger tief erschlossen ist. Ein Blick in das Quellenverzeichnis von BASE<sup>6</sup> zeigt, dass der Schwerpunkt eindeutig auf digitalen Schriftenservern liegt. Als multidisziplinäre Suchmaschine ist BASE weniger vertikal als der disziplinspezifische PsychSpider angelegt.

Vascoda<sup>7</sup> ist ein multidisziplinäres Internet-Portal für wissenschaftliche Information. Es handelt sich um ein integriertes Fachangebot von wissenschaftlichen Bibliotheken und Informationseinrichtungen. Ein Schwerpunkt von Vascoda ist die Verknüpfung der Literatursuche mit der Literaturbeschaffung. In vielen Fällen ist ein direkter Zugriff auf elektronische Zeitschriftenartikel möglich. Falls nicht, kann auf lokale Nachweissysteme und Dokumentenlieferdienste zugegriffen werden. Eine besondere Rolle kommt sowohl beim Nachweis lokaler Nutzungslizenzen von elektronischen Zeitschriften als auch bei der Verknüpfung der Recherche mit dem Zugriff auf den elektronischen Volltext der Elektronischen Zeitschriftenbibliothek Regensburg (EZB) zu (Hutzler, Scheuplein und Schröder 2006). Ebenso wie bei BASE steht für Vascoda die klassische Literatursuche und -beschaffung im Vordergrund. Die Einbindung von weiteren Web-Quellen ist für die Zukunft geplant und soll über eine föderierte Suche realisiert werden. Der PsychSpider wird integrativer Bestandteil der föderierten Suche sein.

Außerhalb des akademischen Umfelds gibt es zwei bekannte multidisziplinäre Wissenschaftssuchmaschinen von kommerziellen Unternehmen: Scirus<sup>8</sup> von Elsevier und Google Scholar<sup>9</sup>, (wobei Google Scholar noch als Beta-Version deklariert wird). Beiden Produkten ist gemein, dass man den bereitstellenden Unternehmen eine Verfolgung von wirtschaftlichen Eigeninteressen unterstellen muss (Elsevier den Verkauf von Verlagsprodukten und Google die Steigerung von Werbeeinnahmen). Bei der Auswahl der Inhalte und der Ausgestaltung der Trefferanzeige (Ranking) ist ein Bias nicht auszuschließen (zur Bestimmung eines Bias siehe zum Beispiel Mowshowitz & Kawaguchi 2002). Auf Grund des disziplinübergreifenden Ansatzes wird häufig eine große Trefferanzahl gefunden, so

dass die Suchergebnisse unübersichtlich sind. Wie auch schon bei den beiden nichtkommerziellen Wissenschaftssuchmaschinen liegt der Fokus auf digitalen Schriftenservern und Bibliothekskatalogen (Google hat diverse Kooperationsabkommen mit großen Bibliotheken geschlossen). Bei beiden werden Informationen aus dem anglo-amerikanischen Sprachraum bevorzugt. Dem gegenüber ist der fachspezifische PsychSpider schlank, transparent und bemüht sich um Vollständigkeit gerade auch in Bezug auf den deutschen Sprachraum.

## 7 Aufbau des Index

Wie werden die Informationen ausgewählt, die der PsychSpider indexieren soll? Wie erfolgt die Abgrenzung zu anderen Disziplinen? Hier kommen die Erfahrungen und Ressourcen einer Fachinformationseinrichtung zum tragen. Bedingt durch den Kontakt zum Fach sind Personen und Institutionen bekannt, die Informationen (potentiell) frei ins Web stellen. Umgekehrt ist der „Community“ das ZPID geläufig und sie kann eine Meldefunktion im Web-Angebot des ZPID nutzen, was allerdings recht selten passiert. Ergiebiger ist der systematische Kontakt zu Autoren, den das ZPID im Rahmen der Erstellung seiner Fachdatenbanken hält, bzw. die Mitarbeit von Redakteuren am Web-Katalog PsychLinker<sup>10</sup>. Suchmaschine und Katalog sind zwei Seiten derselben Münze, nämlich der Psychologie-Ressourcen im Internet. Auf der einen Seite kann man gezielt mit Suchbegriffen operieren, auf der anderen systematisch blättern. Die Suchmaschine PsychSpider, der Web-Katalog PsychLinker und der disziplinspezifische Dokumentenserver PsyDok<sup>11</sup> sind Bestandteil des Projekts „Digitale Psychologie-Information“ und wurden teilweise von der DFG gefördert (Sachbeihilfe an SULB und ZPID). Am Web-Katalog arbeitet ein Stab von ca. 20 freiwilligen externen Redakteuren mit, deren Recherchen auch dem Spider zu gute kommen. Zusammenfassend sind es also die Mitarbeiter des ZPID, unbezahlte Fachredakteure und speziell angestellte studentische Hilfskräfte, die Katalog und Suchmaschine füttern. Der personelle Aufwand ist nicht unerheblich. Ein weiteres aufwendiges Unterfangen ist die Abgrenzung von psychologischen Inhalten auf einem Server, der mehrere Informationsangebote bereit hält (häufig der Fall bei Universitäten). Ein dedizierter Psychologie-Server<sup>12</sup> ist ergiebiger und leichter zu handhaben als ein allgemeiner Universitätsserver mit Bereichen für mehrere Disziplinen.<sup>13</sup> Auch virtuelle Server und Server, bei denen die Inhalte dynamisch erzeugt werden (zum Beispiel mit Hilfe von Content Management Systemen) bereiten Probleme; die Verwendung

5 [www.base-search.net](http://www.base-search.net) [20.2.2007], siehe auch den Beitrag von Pieper/Wolf in diesem Heft.

6 [http://base.ub.uni-bielefeld.de/about\\_sources.html](http://base.ub.uni-bielefeld.de/about_sources.html) [20.2.2007]

7 [www.vascoda.de](http://www.vascoda.de) [20.2.2007]

8 [www.scirus.com](http://www.scirus.com) [20.2.2007]

9 <http://scholar.google.de> [20.2.2007]

10 <http://psychlinker.zpid.de> [20.2.2007]

11 <http://psydok.sulb.uni-saarland.de> [20.2.2007]

12 z.B. [www.psychologie.uni-trier.de](http://www.psychologie.uni-trier.de) [20.2.2007]

13 z.B. [www.uni-muenster.de/Psychologie](http://www.uni-muenster.de/Psychologie), [www.uni-muenster.de/GeschichtePhilosophie](http://www.uni-muenster.de/GeschichtePhilosophie) [20.2.2007]

von interpretierbaren URLs erleichtert die (menschliche) Arbeit ([www.psych.uni-goettingen.de/](http://www.psych.uni-goettingen.de/) ist verständlicher als [www.uni-goettingen.de/de/sh/32971.html](http://www.uni-goettingen.de/de/sh/32971.html)). Generell können Annehmlichkeiten, die Webseiten-Autoren für Menschen einbauen (zum Beispiel Session Cookies, Javascript) der Suchmaschine Probleme bereiten. Der Spider kommt am besten mit barrierefreien Seiten zurecht, das heißt mit Seiten, die mit einem Textbrowser gelesen werden können.

## 8 Probleme mit dem Indexieren

Die Betreiber von Web-Sites adressieren ihr Angebot zuvorderst an menschliche Leser, die einen gängigen Browser benutzen. Letztere sind sehr tolerant gegenüber Verletzungen von Syntaxregeln bzw. dem Einhalten von Web-Standards. Anders verhält es sich mit Spidern (Crawlern). Sie versuchen, so viel Metainformation als möglich aus den vorgefundenen Dokumenten zu extrahieren: Zwischen welchen Metatags steht der eigentliche Inhalt (Text)? Um welche Sprache handelt es sich? Welche ergänzende Information verbirgt sich in XML-Tags? Diese Zusatzinformation, welche über die reine Erfassung des Textes hinaus gehen, sind durchaus wertvoll, erschweren aber mitunter die Indizierung erheblich, weil das „Parsen“ von Dokumenten die regelgerechte Verwendung von Zeichensätzen und Tags voraussetzt. Nicht wohlgeformtes XML zum Beispiel (was im Web recht häufig zu finden ist) führt dazu, dass die — eigentlich vorhandenen — Metainformationen nicht verarbeitet werden und nur der reine Text übrig bleibt. Dies wirkt sich u. a. negativ auf das Ranking solcher Dokumente aus. Mitunter muss beim Indizieren sogar das XML-Parsing deaktiviert werden, weil manche Dokumente sonst überhaupt nicht indexiert werden können. Weitere Probleme resultieren aus Werkzeugen für multimediale und interaktive Inhalte. Die aktuelle Version der Suchmaschinen-Software kann zum Beispiel Information aus eingebettetem Flash oder Java nicht extrahieren. So genannte Flash Intros, die man nicht wegklicken kann (Skip Intro-Link), sind Suchmaschinen-Killer, weil der dahinter liegende Content verborgen bleibt. Die Schwierigkeiten mit dem Portable Document Format (PDF) von Adobe sind auf Grund der neuesten Spider-Releases geringer geworden. Trotzdem kann ein geringer Prozentsatz von PDF-Dokumenten nicht indexiert werden, was an der PDF-Version, dem verwendeten Werkzeug zur Erstellung der PDF-Dokumente oder den gewählten Einstellungen liegen kann (vom Zugriffsschutz ganz abgesehen). Vereinzelt führen auch fehlerhafte Formate zur Nichtindizierung von Dokumenten. Grundsätzlich wird bei jedem Doku-

ment, welches zur Indizierung vorgesehen ist, eine Formaterkennung durchgeführt, um passende Filter auszuwählen. Bei fehlerhaften Formaten ist die Selektion des geeigneten Filters nicht möglich, was zu einem Abbruch des Indizierens führt.

Neben die Schwierigkeiten mit dem Format von Dokumenten treten Schwierigkeiten beim Crawlen. Häufig soll im Wurzelverzeichnis einer Web-Site die Datei robots.txt gemäß dem W3C Robots Exclusion Standard steuern, ob und wie die Site von einem Spider besucht werden darf. Fehlkonfigurationen in der robots.txt, oder auch die explizite Anweisung, nur bestimmte Crawler zuzulassen (meist Googlebot), erschweren den Zugriff auf relevante Webinhalte. Zunehmend finden sich Webangebote, die bei der Auslieferung der Webseiten eine Zeitverzögerung eingebaut haben, um Suchmaschinen auszuschließen. Ein menschlicher Nutzer merkt diese Zeitverzögerung, bekommt mit ein bisschen Geduld aber den gewünschten Inhalt angezeigt. Crawler vermuten bei solchen Zeitverzögerungen technische Probleme und überspringen die Seiten. Nicht alle Web-Angebote, die sich prinzipiell spidern lassen, sind tatsächlich frei zugänglich. Zum Beispiel sind viele Informationen mit einem Zugriffsschutz versehen und bleiben ohne Kenntnis der Login-Daten unerreichbar. Ursachen können kostenpflichtige Angebote oder Vertraulichkeit sein, manchmal aber auch Fehlkonfigurationen. Darüber hinaus muss der Spider meistens bei Angeboten passen, die eine (menschliche) Interaktion voraussetzen.

Ein abschließendes Beispiel soll die Probleme mit dem Spidern verdeutlichen. EdResearch<sup>14</sup> ist eine kostenpflichtige Volltext-Datenbank mit mehreren tausend Fachaufsätzen aus der Erziehungswissenschaft. Im Jahr 2002 waren die Abstracts der Aufsätze kostenfrei über die Web-Site [ericae.net](http://ericae.net) erreichbar; 2003 erfolgte eine Weiterleitung („redirect“) auf die Adresse [edresearch.org](http://edresearch.org). Im Jahr 2004 schließlich wurde ein interaktives Datenbankinterface davor geschaltet, so dass ein Zugriff auf die Abstracts ohne menschliche Interaktion nicht mehr möglich war. Inzwischen ist die Domain [edresearch.org](http://edresearch.org) sogar verkauft und hat mit Erziehungswissenschaft nichts mehr zu tun.

## 9 Technischer Betrieb

Auch im technischen Betrieb zeigt sich, dass die zunehmende Komplexität des Web zu zunehmender Komplexität der Suchmaschine und damit zu höheren Betriebsaufwendungen führt. Das Zusammenspiel zwischen den drei Spidern, dem Datenbanksystem und der Benutzerschnittstelle (den klassischen Komponen-

ten einer Suchmaschine) sowie den Administrationswerkzeugen läuft in einem komplexen System nicht so reibungslos wie in einem einfacher gestrickten. Die Aufwendungen für die Steuerung der einzelnen Werkzeuge, die Auswertung der Logdateien, das Überwachen der Prozesse und das manuelle Eingreifen in den Index steigen kontinuierlich an. Im Zuge dieser Entwicklungen musste die Lizenz aus Performanzgründen um eine zweite CPU erweitert werden, so dass der PsychSpider derzeit auf einer Sun Fire V240 mit zwei UltraSPARC III Prozessoren und 2 GB RAM läuft. Die Größe des Index beträgt ca. 12 GB für 600.000 Dokumente (also ca. 20 KB pro Dokument). Während des Indizierungsvorgangs ist ca. fünfmal so viel temporärer Speicherplatz vonnöten, im Endzustand ist der Index hochgradig optimiert. Nichtsdestotrotz kann der PsychSpider in der Basisfunktionalität nach wie vor mit verhältnismäßig geringem Personaleinsatz betrieben werden.

Da die Indexe transferierbar sind, stellen Sicherungen und Übertragungen auf neue Maschinen kein Problem dar. Allerdings ist die Lizenz an eine bestimmte Maschine gebunden, was im Notfall (zum Beispiel dem Totalausfall einer Maschine) zu Zeitverzögerungen bei der Wiederherstellung des Betriebs führen kann. Die Stabilität der Verity-Systeme ist ausgezeichnet, es gab in all den Jahren keine systembedingten Ausfälle.

## 10 Nutzung

Die Nutzung des PsychSpider war von Beginn an gut und ist über die Jahre langsam aber kontinuierlich gewachsen. Sie liegt derzeit bei ca. 10.000 Suchanfragen im Monat, was für ein deutschsprachiges wissenschaftliches Nischenangebot in Anbetracht der potentiellen Nutzerschaft durchaus bemerkenswert zu sein scheint (Zielgruppen sind die psychologische Forschung, Lehre und Praxis in den deutschsprachigen Ländern). Analysiert man die Nutzung etwas genauer (vgl. Baier & Weiland 2006), so lässt sich auf Grund der Art der Suchanfragen bzw. der Suchterminologie darauf schließen, dass auch viele Laien (Nicht-Psychologen) auf den PsychSpider zugreifen. Wie bei Suchmaschinen üblich gibt die überwiegende Mehrheit nur einen Suchbegriff in die Suchmaske ein, aber auch komplexe Anfragen kommen durchaus mit signifikantem Anteil vor (was unter anderem wiederum für die Nutzung durch Fachleute spricht). Das Verhältnis von *Einfacher Suche* zu *Erweiterter Suche* ist etwa 6 zu 1. Nur in ca. 15 Prozent der Fälle wird die erweiterte Suchmaske mit

14 <http://cunningham.acer.edu.au/dbtw-wpd/sample/edresearch.htm>

den Möglichkeiten der Formulierung einer komplexen Query herangezogen. Auch von der Möglichkeit natürlichsprachiger Anfragen wird kaum Gebrauch gemacht (Beispiel: „Welche Arten von Psychotherapie gibt es?“). Bei der Auswahl der verschiedenen Kollektionen (Indexe) wird die exklusive Suche im Webangebot des ZPID bzw. in der bibliographischen Datenbank PSYNDEX am häufigsten gewählt. Es gibt zehn Kollektionen: ZPID (Web-Angebot der Fachinformationseinrichtung), Veranstaltungen (einschlägige Tagungen und Kongresse), Fachgesellschaften, Fachgruppen, Fachinformationszentren, Institute (universitäre und außeruniversitäre Psychologie-Institute), E-Journals (elektronische Zeitschriften bzw. Dokumentenserver mit frei zugänglichen Volltexten), (Test-) Diagnostik, PSYNDEX Lit & AV (bibliographische Datenbank mit den Segmenten Literatur und AV-Medien), PSYNDEX Tests (bibliographische Datenbank mit dem Segment psychologische Tests) und Web Allgemein (alles andere).

## 11 Weiterentwicklung

Die Vorzüge einer disziplinspezifischen Suchmaschine, die von dokumentarisch ausgebildeten Hochschulabsolventen aus dem Fach betreut wird, sind unübersehbar: Im Unterschied zu den verbreiteten unspezifischen Suchmaschinen ertrinkt der Nutzer nicht in einer Flut von Treffern, sondern sucht per definitionem nur in einem Suchraum, der von relevanten Fachinformationen aufgespannt wird. Die Selektion von Quellen erfolgt nach offengelegten akademischen Standards. Das Ranking-Verfahren ist transparent und wird nicht laufend geändert; Manipulationen durch kommerzielle Einflussnahme oder Zensur finden nicht statt. Dokumente werden zur Gänze indiziert und nicht nur in Teilen (vgl. die 100-KB-Begrenzung für PDF bei Google). Durch Kooperation mit den Autoren von Web-Dokumenten können speziell definierte Meta-Tags gezielt in das Ranking einfließen (zum Beispiel auf der Basis von Dublin-Core). Der PsychSpider wird deshalb als essentielles strategisches Produkt kontinuierlich weiterentwickelt werden.

Inhaltlich wird er im ersten Halbjahr 2007 um den Katalogbestand des Sondersammelgebiets Psychologie der Saarländischen Universitäts- und Landesbibliothek ergänzt. Verhandlungen mit internationalen Content-Anbietern treten auf der Stelle, weil das Hauptaugenmerk kostenpflichtigen Fachinformationen gilt, für die entweder ein Modus des finanziellen Ausgleichs (zum Beispiel Nationallizenz) gefunden oder eine reduzierte Form der Trefferanzeige implementiert werden muss.

Technisch steht die automatische Kategorisierung und Zuordnung von Dokumenten zu verschiedenen Themengebieten auf dem Plan. Autonomy nennt dies „Parametrische Suche“, wobei die Parameter die Begriffe für die Klassen sind. Ein anderer Ansatz ist der „Collaborative Classifier“, der es sehr einfach erlaubt, Kategorien auf Basis von Trainingsdokumenten zu definieren und zu pflegen. Diese Kategorien können dann sowohl für die Anzeige der Dokumente in einer Baumstruktur, als auch für die Definition von relevanten Dokumenten für den Index verwendet werden. Ein drittes Modul für die Fortentwicklung des PsychSpider ist die „Advanced Query Guidance“ – eine automatisch generierte Liste von Begriffen, die im Bezug zur aktuellen Anfrage stehen und so beim „data mining“ helfen.

Ein Wermutstropfen für einen öffentlich geförderten Not-for-Profit Fachinformationsanbieter ist die Lizenzpolitik erfolgreicher internationaler Softwarehersteller. Standard-Lizenzen decken nur Basis-Funktionalitäten ab, und gerade die interessanten fortschrittlichen Module müssen zugekauft werden. Das kann man sich unter Umständen nicht leisten und die Qualität der eigenen Produkte muss darunter leiden.

Durch die Übernahme von Verity durch Autonomy steht eine neue Suchtechnologie im Fokus des Herstellers, die das ZPID nicht lizenziert hat. Die Technologie um den sog. IDOL-Server macht sich einen elementaren Umstand menschlichen Verhaltens zu nutze, nämlich des Lernens aus Erfahrung. Bei IDOL werden die Benutzereingaben analysiert, nach Kennwörtern gruppiert und unter Heranziehung des Bayes-Theorems (Barr & Zehna 1983) über bedingte Wahrscheinlichkeiten werden (modellbasiert) Annahmen über den Zusammenhang zwischen Begriffen und sie enthaltenden Texten berechnet. Mit diesem „Wissen“ ausgestattet kann Autonomy dem Benutzer inhaltlich relevante Dokumente empfehlen. Ein einfaches Beispiel für den Bayes-Ansatz ist die Klassifikation von Spam im E-Mail-Verkehr: Vom Vorhandensein einschlägiger Begriffe in einer E-Mail (Ereignis) wird auf die Eigenschaft „Spam“ (Ursache) geschlossen. Mit Hilfe von Wahrscheinlichkeitsmodellierung nach dem Bayes-Ansatz wird also entweder bewertet, wie gut ein Dokument zu einer Nutzeranfrage passt, oder wie gut es sich in eine bestimmte Kollektion einordnen lässt (Klassifizierung). Ein lernfähiges System ist ein wichtiger Schritt in Richtung Entlastung von aufwendiger menschlicher Intervention hin zu automatischer Verarbeitung. Auch Metadaten verlieren nach diesem Ansatz ihr Gewicht.

## Literatur

Baier, Ch.; Weiland, P. (2006). PsychSpider – Erfahrungen mit dem Betrieb einer spezialisierten Suchmaschine. In: Stempfhuber, M. (Hrsg.): In die Zukunft publizieren. Herausforderungen an das Publizieren und die Informationsversorgung in den Wissenschaften. Bonn: Informationszentrum Sozialwissenschaft.

Barr, D. R.; Zehna, P. W. (1983). Probability: Modeling Uncertainty. Reading, MA: Addison-Wesley.

Hutler, E.; Scheuplein, M.; Schröder, P. (2006). Der schnelle Weg zum Volltext – Einsatz und Nutzung des Verlinkungsdienstes der Elektronischen Zeitschriftenbibliothek. In: Bibliotheksdienst, 3, S. 306-313.

Mowshowitz, A.; Kawaguchi, A. (2002). Assessing bias in search engines. In: Information Processing & Management, 38 (1), S. 141-156.

Weichselgartner, E. (2006). Föderierter Fachsuchverbund. In: c't, 11, S. 13.

Psychologie, Suchmaschine, Systemumgebung, Maschinelles Indexierungsverfahren, Qualität, Arbeitsablauf

## DIE AUTOREN

### PD Dr. Erich Weichselgartner



ist stellvertretender wissenschaftlicher Leiter des Zentrums für Psychologische Information und Dokumentation (ZPID) und Leiter der Bereiche Informationstechnologie und Entwicklung. Zu seinen Forschungsinteressen gehören die elementare menschliche Informationsverarbeitung und die Entwicklung benutzerfreundlicher Informationssysteme. Als habilitierter Psychologe ist er zudem an den Universitäten Regensburg und Trier in der Lehre engagiert.

wga@zpid.de

### Dr. med. Christiane Baier



ist promovierte Medizinerin und Diplom-Informatikerin (FH) am Zentrum für Psychologische Information und Dokumentation (ZPID). Neben der Implementierung und Fortentwicklung des Retrievalsystems für PSYNDEX liegt ihr Aufgabenschwerpunkt in der Betreuung des PsychSpider, der Psychologie-Suchmaschine des ZPID.

Zentrum für Psychologische Information und Dokumentation  
Universität Trier  
54286 Trier

# BASE – Eine Suchmaschine für OAI-Quellen und wissenschaftliche Webseiten

Dirk Pieper und Sebastian Wolf, Bielefeld

*Dieser Aufsatz beschreibt die Entwicklung der Suchmaschine BASE (Bielefeld Academic Search Engine) seit 2005. In dieser Zeit wurde der Index um ein Vielfaches ausgebaut und auch die Nutzungszahlen stiegen deutlich. Der Schwerpunkt liegt auf der Indexierung von Dokumentenservern, die ihre Daten über das „Protocol for Metadata Harvesting“ (OAI-PMH) bereitstellen. Im Gegensatz zu speziellen OAI-Suchmaschine wie OAIster verfügt BASE jedoch über weitergehende Suchmöglichkeiten und indexiert auch wissenschaftliche Webseiten mit Hilfe eines integrierten Web-Crawlers und andere Quellen, wie zum Beispiel den Bibliothekskatalog der Universitätsbibliothek Bielefeld. BASE kommt darüber hinaus als Suchsystem auch in anderen Bereichen, zum Beispiel im Bielefeld eScholarship Repository, als Suchmaschine der Universität Bielefeld und im EU-Projekt DRIVER via Schnittstellen zum BASE-Index zum Einsatz.*

## **BASE - A search engine for OAI sources and scientific web pages**

*This article describes the development of BASE (Bielefeld Academic Search Engine) since 2005. During this time, the BASE index grew considerably and so did the usage frequency. The vast majority of the indexed resources originate from OAI conforming repositories. In contrast to special OAI search engines such as OAIster, BASE provides advanced search options. Scientific web sites and similar resources are indexed, e.g. the library catalogue of Bielefeld University Library. But the area of application is wider than that; it is also used for the Bielefeld eScholarship Repository, it serves as the university's search engine and is in operation via interfaces to the BASE-index in the DRIVER project of the EU.*

## 1 Einleitung

Als das erste Themenheft „Suchmaschinen“ dieser Zeitschrift Anfang 2005 erschien, war BASE<sup>1</sup> ein gutes halbes Jahr unter Produktionsbedingungen im Netz (Summann/Wolf 2005). Das zweite Themenheft bietet nun einen willkommenen Anlass, eine Zwischenbilanz zu ziehen und es zeigt auch, dass Suchmaschinen als Plattform bibliothekarischer Informationsangebote ein zunehmend stärkeres Gewicht bekommen.

Kommerziell betriebene Suchmaschinen wie zum Beispiel Google, Google Scholar, Scirus oder auch Windows Live Search Academic werden von Studierenden gegenüber Bibliotheksangeboten, wie zum Beispiel Bibliothekskatalogen (Lewandowski 2006) und Fachdatenbanken (Urquhart et al. 2005) bevorzugt genutzt. Produkte wie Google Book Search oder Amazon Search Inside führen zu höheren Ansprüchen an Bibliothekskataloge. Die ersten Bibliotheken und Bibliotheksverbände haben darauf reagiert: Bibliographische Angaben werden mit Daten aus Inhaltsverzeichnissen angereichert (Großgarten 2005), bei der Katalogaufbereitung wird zunehmend Suchmaschinentechnologie eingesetzt. Sichtbar wird dies zum Beispiel an (wie BASE) ebenfalls auf FAST<sup>2</sup> basierenden Produkten wie dem Dreiländerkatalog des HBZ<sup>3</sup>, der Entwicklung eines neuen Webkatalogs für Lokalsysteme durch OCLC/PICA<sup>4</sup> oder in ersten Ansätzen zur Integration von Suchmaschinentechnologie in Vascoda<sup>5</sup>.

Darüber hinaus gibt es inzwischen eine Reihe von Projekten im Bibliotheksumfeld, in denen die Open-Source-Software Lucene zum Einsatz kommt<sup>6</sup>. Ziel beim Aufbau der Bielefeld Academic Search Engine (BASE) ist die Realisierung eines suchmaschinenbasierten Informationsangebots, das Zugriff auf verschiedensten über das Internet zugänglichen wissenschaftlichen Content ermöglicht. Die Vorteile beim Einsatz von Suchmaschinentechnologie – einfache, „google-ähnliche“ Nutzung, hohe Performanz, Volltextsuche und hohe Relevanz der Suchergebnisse – sollen verbunden werden mit den aus der Datenbankwelt bekannten Vorteilen der Berücksichtigung bibliographischer Such-

aspekte sowie hoher Datenqualität. Damit wird der an der Universitätsbibliothek Bielefeld seit den neunziger Jahren verfolgte Ansatz der Integration von externem wissenschaftlichen Content und lokaler Datenbankproduktion weiterentwickelt.

Der Einsatz von Suchmaschinentechnologie ist dabei allerdings kein Selbstzweck. Sie soll in erster Linie dem Nutzer helfen, in den immer größer werdenden Datenmengen das zu finden, was er sucht, zum Beispiel durch den Einsatz von linguistischen Methoden (Finden von ähnlichen Wortformen, multilinguale Suche, Ontologien, usw.), durch verschiedene Verfahren der Relevanzbewertung oder durch die Möglichkeit, nach einer Suche große Treffermengen einzuschränken („Search Refinement“). BASE beinhaltet im Vergleich zu 2004 inzwischen eine Vielzahl unterschiedlicher wissenschaftlich relevanter Datenquellen. Es ist jedoch kein Zufall, dass sich im Projektverlauf der Schwerpunkt auf die Indexierung von Repository-Servern, die ihre Daten nach dem Standard des „Protocol for Metadata Harvesting“ (OAI-PMH) bereitstellen, herausgebildet hat. Zum einen besitzen diese Daten – trotz aller Probleme, die beim Sammeln und Aggregieren von OAI-Daten in der Praxis auftreten können (Pieper/Summann 2006; Summann/Wolf 2006) – gegenüber unstrukturiertem Webcontent eine vergleichsweise hohe Qualität, die es ermöglicht, Suchen nach verschiedenen Aspekten wie Autor, Titel, Schlagwort usw. anzubieten, zum anderen bieten sie im Vergleich zu reinen Nachweisdaten häufig einen Link zum Volltext. Ein weiterer Grund dieser Schwerpunktbildung liegt in der notwendigen Positionierung auf dem Suchma-

1 <http://base.uni-bielefeld.de> [7.2.2007]

2 Fast Search & Transfer: [www.fastsearch.com](http://www.fastsearch.com) [7.2.2007]

3 <http://suchen.hbz-nrw.de/dreilaender> [7.2.2007]

4 [www.oclc.org/dasat/images/4/100764-fast.pdf](http://www.oclc.org/dasat/images/4/100764-fast.pdf) [7.2.2007]

5 <http://lists.vascoda.de/presse.pl?id=5> [7.2.2007]

6 Zum Beispiel die Entwicklung eines neuen Online-Katalogs an der UB Heidelberg, siehe [www.uni-bayreuth.de/SISIS/SSV/SAT/sat\\_20061024\\_files/Lucene.pdf](http://www.uni-bayreuth.de/SISIS/SSV/SAT/sat_20061024_files/Lucene.pdf) [7.2.2007]

schinenmarkt, wobei es auf der Hand liegt, dass schon aufgrund des ungleich geringeren möglichen Einsatzes ökonomischer und technischer Ressourcen bibliothekarische Suchmaschinenangebote mit den oben genannten Produkten nur schwer konkurrieren können. BASE versucht daher mit einem inhaltlichen Profil und mit über den Standard hinausgehenden Suchfunktionen eine Nische zu besetzen. Dazu gehören im Vergleich zu Google unter anderem folgende Merkmale:

- intellektuelle Auswahl wissenschaftlich relevanter (OAI-)Quellen
- hohe Datenqualität, Transparenz der indexierten Quellen über ein Quellenverzeichnis
- Volltextindexierung und Verbindung mit den zugehörigen Metadaten
- differenzierte Anzeige von bibliographischen Daten
- mehr Suchoptionen, Suchhistorie, Sortierung

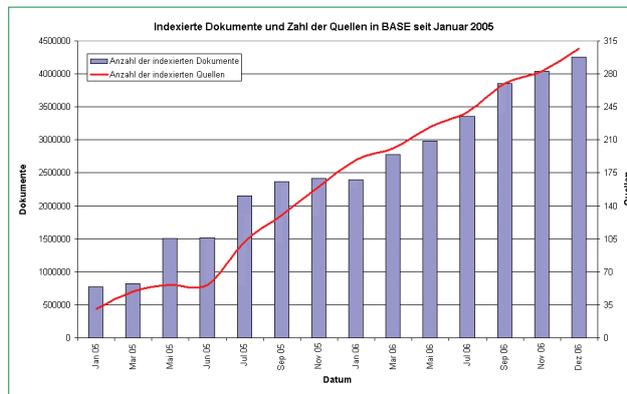
Zum Konzept des Einsatzes von Suchmaschinenteknologie an der Universitätsbibliothek Bielefeld gehört jedoch nicht nur die Positionierung von BASE als suchmaschinenbasierter OAI-Service-Provider. Die FAST-Software ermöglicht es auch, mit Hilfe eines integrierten Web-Crawlers und mit Datenbankverbindungen Dokumente zu indexieren und den Index flexibel in unterschiedlichen Kontexten zu nutzen. Die Nutzung der Daten eines (externen) Suchmaschinenindex in unterschiedlichen Portalen ist in der kommerziellen Suchmaschinenwelt schon seit längerem Praxis (Lewandowski 2005, S. 21 ff.) und sollte auch im bibliothekarischen Umfeld realisiert werden können, was nicht zuletzt auch die weiter unten kurz beschriebene Teilnahme der Universitätsbibliothek Bielefeld am EU-Projekt DRIVER<sup>7</sup> zeigt.

## 2 Entwicklung der Quellen- und Nutzerzahlen

BASE startete im Juni 2004 mit einer Einzelplatz-Installation unter FAST Data Search 3.0. Nachdem in der Zwischenzeit die FAST-Software auf die Version 4.1 aktualisiert wurde, läuft BASE seit Februar 2006 auf einer Server-Farm und weist damit – im Kleinen – die von großen Suchmaschinen bekannte skalierbare Systemarchitektur auf.

Zum Zeitpunkt der Abfassung des Artikels für das erste Themenheft enthielt BASE ca. 800.000 indexierte Dokumente

aus 35 Quellen. Mittlerweile (Stand: Dezember 2006) umfasst der Index 4,3 Millionen Dokumente aus 307 Quellen. Die Quellen werden laufend aktualisiert und die Zahl indexierter Dokumente wächst



und als Konferenzbericht vorliegen kann) und/oder die verschiedenen Server, auf denen ein Artikel abgerufen werden kann, zu einer Gruppe zusammengefasst, die über den Link „Group of X“ angezeigt werden.

Die Einbindung dieser Funktion konnte dank E-Mail-Kommunikation mit Anurag Acharya, dem Chefentwickler von Google Scholar, aktiviert werden. Für das Jahr 2007 wurde von ihm die Möglichkeit in Aussicht gestellt, trefferbezogene Links direkt auf die zitierenden Artikel („Cited by“) setzen zu können.

Neben der trefferbezogenen Integration kann eine Suche in Google Scholar auch direkt von der Suchmaske in der Trefferanzeige gestartet werden („Suche in Google Scholar“), um weitere nicht durch BASE indexierte wissenschaftliche Quellen finden zu können und die beschriebenen Google-Scholar-Funktionen, die von BASE selbst derzeit nicht bereit gestellt werden können, als Add-On zu nutzen. Google Scholar wurde deshalb gewählt, da es sich um das populärste Angebot im wissenschaftlichen Bereich handelt und – beispielsweise im Gegensatz zur allgemeinen wissenschaftlichen Suchmaschine Scirus – eine hohe inhaltlich Übereinstimmung bei den indexierten Quellen besteht. Zudem erscheinen bei Google Scholar – sofern die Bibliothek einen Linkresolver einsetzt und dieser entsprechend konfiguriert ist – trefferbezogene

Links auf die eigenen Bibliotheksbestände („Library links“). Wissenschaftlern und Studierenden, die Google Scholar als primäre Quelle der Informationsbeschaffung nutzen, können auch, ohne dass sie im Bibliothekskatalog recherchieren, auf die eigenen Bibliotheksbestände gelenkt werden. Suchmaschinen sollten daher nicht nur als Konkurrenz zu den eigenen Angeboten betrachtet werden, sondern vielmehr als Möglichkeit, die eigenen Bestände für ein breiteres Publikum sichtbar zu machen.

Als letzter Punkt soll hier noch erwähnt werden, dass auch die barrierefreie Gestaltung der BASE-Webseiten ausgebaut wurde. Mittlerweile erfüllen die Informationsseiten von BASE die Kriterien der Priorität 2, Suchmaske und Trefferlisten erfüllen die Kriterien der Priorität 1 der entsprechenden Richtlinien, die im Bundesgesetzblatt festgehalten sind (Bundesministerium der Justiz 2002).

#### 4 BASE als Suchsystem für verschiedene Anwendungen

Der konsequente Einsatz von Cascading Style Sheets (CSS) zur Trennung von Inhalt und Layout erwies sich als großer Vorteil, nicht nur bezüglich der Integration neuer Funktionen, sondern auch bei der Entwicklung eigener Sichten, wie

zum Beispiel für das Bielefeld eScholarship Repository, in dem BASE als Suchsystem zum Einsatz kommt. Dabei basieren Dokumentenablage, OAI-Schnittstelle und Browsing-Funktionen auf OPUS, die Indexierung der Daten erfolgt jedoch mit FAST und die Suchoberfläche ist ein eigenständiger View der BASE-Oberfläche. Das Bielefeld eScholarship Repository stellt die Forschungsergebnisse der Universität Bielefeld der wissenschaftlichen Community weltweit zur Verfügung. Von der Suchmaske aus ist es möglich, unter anderem nur im Repository oder auch in allen BASE-Quellen zu suchen. Das Suchergebnis wurde an das Design des Bielefeld eScholarship Repositories angepasst. Das Design der verschiedenen Oberflächen wird dabei ausschließlich durch den Einsatz unterschiedlicher CSS-Dateien gesteuert.

Seit 2005 werden mit FAST die Webseiten der Universität Bielefeld und der Universitätsbibliothek indiziert, wobei dieser Teil nicht in BASE einfließt. Hierbei wurden die Stärken des Systems bei der Indexierung von Webquellen deutlich. Durch den Einsatz eines Tools zur Relevanzbewertung, die vom FAST-System bereitgestellt wird, konnte eine wesentliche Verbesserung der Relevanz bei den gefundenen Treffern erreicht werden.

10 <http://europa.eu/eurovoc> [7.2.2007]

SourceOECD 

## Die Online-Bibliothek der OECD



„Mit SourceOECD haben Sie eine ganze Wirtschaftsbibliothek am eigenen PC – ein tolles Arbeitswerkzeug.“

Sebastian Dullien  
Financial Times Deutschland

SourceOECD ist die Online-Bibliothek mit allen seit 1998 erschienenen OECD-Veröffentlichungen.

Mit einem einzigen Abonnement können Sie sich jetzt den Online-Zugriff auf alle OECD-Bücher, Zeitschriften und Datenbanken sichern.

Weitere Informationen finden Sie unter [www.oecd.org/de/publikationen](http://www.oecd.org/de/publikationen)

Gern richten wir Ihnen einen kostenlosen Testzugang ein!

OECD Berlin Centre · Schumannstr. 10 · 10117 Berlin  
Kontakt: Damon Allen · Tel.: (030) 28 88 35-48  
damon.allen@oecd.org · [www.oecd.org/berlin](http://www.oecd.org/berlin)

## 5 EU-Projekt DRIVER

Das DRIVER-Projekt verfolgt das Ziel, in einer ersten Phase eine Testumgebung zur Vernetzung von mehr als 50 in Europa verteilten wissenschaftlichen Dokumentenserver (sog. Repositorien) zu entwickeln. Das Projekt, an dem zehn Partner aus acht Ländern beteiligt sind, wird durch die Abteilung „Forschungsinfrastruktur“ der Europäischen Kommission gefördert (Lossau 2006). Die Universitätsbibliothek Bielefeld ist dabei verantwortlich für die Aggregation und Speicherung von OAI-Metadaten, die Indexierung mit FAST und die Bereitstellung einer HTTP- und SOAP-Schnittstelle. Über diese Schnittstellen wird es für Betreiber des zukünftigen DRIVER-Portals möglich sein, sich flexibel aus dem BASE-Index Quellen für das eigene Informationsangebot zusammenzustellen und in das eigene Portal zu integrieren. DRIVER wird entsprechend der Ausrichtung dann der Teil von BASE nutzen, der die Daten der europäischen Repository-Server repräsentiert.

## 6 Fazit, Ausblick und Perspektiven

Sowohl die steigenden Nutzerzahlen, die zunehmenden Anfragen von Repository-Betreibern zwecks Aufnahme in den BASE-Index als auch die Beteiligung am DRIVER-Projekt lassen das Fazit nach rund drei Jahren BASE positiv ausfallen. Mehr Aufwand als gedacht musste in den Bereich Harvesting und Normalisierung von OAI-Metadaten investiert werden. Beim kontinuierlichen Ausbau des Index durch Aufnahme weiterer OAI-Quellen steht insbesondere auch die weitere Verbesserung und Automatisierung des OAI-Datenflusses vom Harvesting bis hin zur Indexierung mit FAST auf der Agenda.

In der Entwicklung befinden sich derzeit ein Searchplugin, über das man BASE direkt über die Suchmaschinen-Toolbar im Browser durchsuchen kann, und das Browsing über die indexierten Quellen mithilfe der Dewey Decimal Classification (DDC). Alle Dokumente, die mit einer DDC-Nummer versehen sind, können über dieses Browsing ermittelt werden. Derzeit sind dies allerdings erst ca. 25.000 Dokumente und damit nur ein sehr kleiner Teil aller Dokumente im BASE-Index. Allerdings betrifft dies alle von BASE indexierten OPUS-Systeme, sowie andere von DINI zertifizierte Repository-Server, was die Sinnhaftigkeit von Standardisierung über Software oder Zertifizierung verdeutlicht.

2007 wird die lokale Datenbankproduktion (Bibliothekskatalog, Aufsatzdatenbank JADE und weitere) mit FAST eine stärkere Rolle spielen, wobei sich interessante Perspektiven durch die Entwick-

lung des ebenfalls auf FAST basierenden Webkatalogs durch OCLC/PICA hinsichtlich einer möglichen Integration in die BASE-Umgebung ergeben. BASE wird dann für externe Nutzer hauptsächlich die Funktion eines OAI-Service-Providers wahrnehmen, während für Nutzer auf dem Campus Bielefeld zusätzlich die Integration lokaler Quellen angeboten werden kann. Ein weiterer Aspekt, der aufgrund des hohen Aufwands im OAI-Bereich zurückstehen musste, ist die Auseinandersetzung mit den linguistischen Features der FAST-Software. In einem ersten Schritt werden Teile des eurovoc-Thesaurus<sup>10</sup> in die FAST-Indexierungs-Stufen eingebunden, um in Zukunft eine multilinguale Suche in BASE zu ermöglichen.

## Literatur

Bundesministerium der Justiz (Hrsg.) (2002): Verordnung zur Schaffung barrierefreier Informationstechnik nach dem Behindertengleichstellungsgesetz (Barrierefreie Informationstechnik-Verordnung – BITV) : vom 17. Juli 2002. In: Bundesgesetzblatt Teil I 2002 (49), 2654-2660. URL: [www.bgblportal.de/BGBL/bgbl1f/bgbl102s2654.pdf](http://www.bgblportal.de/BGBL/bgbl1f/bgbl102s2654.pdf) [7.2.2007].

Großgarten, A. (2005): Das 180T-Projekt in Köln oder wie verarbeite ich 180.000 Bücher in vier Monaten. In: *Information, Wissenschaft und Praxis* 56 (8), 454-456.

Lewandowski, D. (2005): Web Information Retrieval. *Technologien zur Informationssuche im Internet*. Frankfurt a.M.: DGI.

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

Lossau, N. (2006): DRIVER: Networking European Scientific Repositories. In: *BI.research – Forschungsmagazin der Universität Bielefeld* 2006 (29), 61-65. URL: [www.uni-bielefeld.de/Universitaet/Einrichtungen/Pressestelle/dokumente/BI\\_research/29\\_2006/Forschungsmagazin\\_1\\_06\\_Driver\\_61\\_65.pdf](http://www.uni-bielefeld.de/Universitaet/Einrichtungen/Pressestelle/dokumente/BI_research/29_2006/Forschungsmagazin_1_06_Driver_61_65.pdf) [7.2.2007].

Pieper, D.; Summann, F. (2006): Bielefeld Academic Search Engine (BASE): An end-user oriented institutional repository search service. In: *Library Hi Tech* 24 (4), 614-619.

Summann, F.; Wolf, S. (2005): BASE – Suchmaschinentechnologie für digitale Bibliotheken. In: *Information, Wissenschaft und Praxis* 56 (1), 51-57.

Summann, F.; Wolf S. (2006): Suchmaschinentechnologie und wissenschaftliche Suchumgebung. In: *VÖB Online-Mitteilungen* 2006 (86), 3-8. URL: [www.univie.ac.at/voeb/php/downloads/om86.pdf](http://www.univie.ac.at/voeb/php/downloads/om86.pdf) [7.2.2007].

Urquhart, C.; Thomas, R.; Spink, S.; Fenton, R.; Yeoman, A.; Lonsdale, R.; Armstrong, C.; Banwell, L.; Ray, K.; Coulson, G.; Rowley, J. (2005): Student use of electronic information services in further education. In: *International Journal of Information Management* 25 (4), 347-362.

Suchmaschine, Maschinelles Indexierungsverfahren, Quelle, Bibliothek, UB Bielefeld, BASE, FAST

## DIE AUTOREN

### Dirk Pieper



wurde am 30.3.1967 in Heide geboren und studierte Politikwissenschaft und Volkswirtschaftslehre an den Universitäten Bonn und Hamburg. Danach absolvierte er

die Ausbildung zum Höheren Dienst an wissenschaftlichen Bibliotheken an der FH Köln und ist seit 2004 Erwerbungsdezernent und Koordinator im Bereich Digitale Bibliothek an der UB Bielefeld.

Telefon: (05 21) 106 40 10  
dirk.pieper@uni-bielefeld.de

### Sebastian Wolf



wurde am 26.4.1975 in Göttingen geboren und studierte Wissenschaftliches Bibliothekswesen an der FH Hannover. Seit 1999 ist er in der Universitätsbibliothek

Bielefeld tätig, unter anderem im Bereich Pflege und Weiterentwicklung des Web-Angebots. Er betreut den Suchdienste-Kompass ([www.ub.uni-bielefeld.de/biblio/search](http://www.ub.uni-bielefeld.de/biblio/search)), ein Tutorial zur Nutzung und Funktionsweise von Suchmaschinen. Im BASE-Projekt arbeitet er seit 2002 im Bereich Webdesign, Usability und Funktionalitätstests.

Telefon: (05 21) 106 40 44  
sebastian.wolf@uni-bielefeld.de

Universitätsbibliothek Bielefeld  
Universitätsstraße 25  
33615 Bielefeld

# Bangalore Commitment: Workshop on Electronic Publishing and Open Access

Developing Country Perspective: Bangalore, Indien, 2. bis 3. November 2006

Achim Oßwald, Köln

## Überblick

Am 2./3. November 2006 fand in der süd-indischen IT-Metropole Bangalore ein Workshop<sup>1</sup> statt, zu dem Teilnehmer aus den Ländern Indien, China, Südafrika und Brasilien eingeladen waren. Ziel des Workshops war zu verdeutlichen, welche Vorteile insbesondere Wissenschaftler und wissenschaftliche Einrichtungen, aber auch Informationsfachleute und nicht zuletzt Verleger durch Open Access (OA) haben können. Darüber hinaus wurde die Verabschiedung einer Selbstverpflichtung vorbereitet, die nun als „Bangalore Commitment“ in einer Entwurfsfassung in den einschlägigen OA-Foren diskutiert wird. Hierbei wurde insbesondere auf die sog. *Berliner Erklärung*<sup>2</sup> Bezug genommen. Die Schwerpunktsetzung auf „developing countries“, also sog. Entwicklungsländer bzw. Schwellenländer, bringt eine spezifischer Perspektive in die Umsetzung des Open Access-Konzeptes. Durch den Aufbau gemeinsamer OA-Infrastrukturen und den Erfahrungsaustausch über das OA-Konzept soll die Sichtbarkeit der wissenschaftlichen Forschung und Entwicklung in den jeweiligen Ländern erhöht bzw. ihre Wahrnehmung in anderen Ländern z.T. erst ermöglicht werden. Umgekehrt will man aber auch durch diese Zusammenarbeit die OA-Bewegung weltweit stimulieren.

## Organisation des Workshops

Die Veranstaltung war finanziert und organisiert mit Hilfe der *Indian Academy of Science*, der *MSSRF-Stiftung*<sup>3</sup> sowie des *Open Society Institute* ([www.soros.org/](http://www.soros.org/)) und fand auf dem Campus der führenden nationalen Forschungseinrichtung in Indien, dem *Indian Institute of Science* (IISc; [www.iisc.ernet.in/](http://www.iisc.ernet.in/)) in Bangalore statt. Mehrere leitende Persönlichkeiten der genannten Institute sowie Professoren des IISc nahmen mit Fachbeiträgen

an der Veranstaltung teil. Damit zeigte man auch nach außen die Bewertung und Relevanz des Themas aus der Perspektive der Forschung.

Durch Einladungen an äthiopische, chinesische, südafrikanische, brasilianische und selbstverständlich viele indische Experten<sup>4</sup> versuchten die Initiatoren gerade deren spezielle Perspektiven auf die Entwicklungen des wissenschaftlichen Publizierens und das Thema OA herauszustoichen. Ergänzend hatte der seit langem im OA-Thema aktive indische Hauptorganisator, Subbiah Arunachalam von der MSSRF-Stiftung weitere OA-Expertinnen als argumentative Unterstützung hinzugezogen.

Mit Barbara Kirsop vom *Electronic Publishing Trust for Development*, Alma Swan, Direktorin der Consulting-Firma *key perspectives* ([www.keyperspectives.co.uk/aboutus/index.html](http://www.keyperspectives.co.uk/aboutus/index.html)), die hauptsächlich zum Thema OA für das britische *Joint Information Systems Committee* (JISC, [www.jisc.ac.uk/](http://www.jisc.ac.uk/)) arbeitet, sowie Susan Veldsman von *eIFL.net* (*Electronic Information for Libraries*)<sup>5</sup> waren führende Aktivistinnen der OA-Szene anwesend und brachten ihre Erfahrungen in die Diskussionen ein.

## Wesentliche Inhalte der Tagung

Die Erfahrungsberichte aus Indien, Südafrika und Brasilien über die Etablierung von Institutionellen Repositorien (IR)<sup>6</sup>, d.h. institutionsbezogenen Archiven für wissenschaftliche Veröffentlichungen, waren für alle Teilnehmer motivierende Beispiele, das Thema auch in ihren eigenen Organisationen anzugehen. Vorzeigebispiele waren neben dem der gastgebenden Einrichtung, dem *Indian Institute of Science* (IISc) in Bangalore, auch das eines Bibliothekars eines kleineren indischen Forschungsinstituts, der engagiert berichtete, wie er seinen Institutsleiter erfolgreich von den organisatorischen Vorteilen eines IR für seine Einrichtung

überzeugte. Das IISc hat für die OA-Veröffentlichung seiner Wissenschaftler ein IR (<http://eprints.iisc.ernet.in/>) mit der Bezeichnung und Webadresse *ePrints@iisc* aufgebaut, in dem sich Mitte November 2006 genau 5772 Publikationen von IISc-Wissenschaftlern befanden. *eprints@iisc* ist ein sehr gutes Beispiel für die mit der Einrichtung eines IR verbundenen Aktivitäten wie z.B. der Formulierung von Sammel- bzw. Einreichungsrichtlinien sowie der Bereitstellung von Zugriffsstatistiken u. ä. mehr, mit denen auch interne Kriterien zur Mittelvergabe verknüpft werden können.

- 1 Die Website des Workshops mit weiterführenden Materialien, auf die nachfolgend z.T. verwiesen wird, ist unter <http://scigate.ncsi.iisc.ernet.in/OAworkshop2006/> eingerichtet.
- 2 Die Tagung, bei der die Berliner Erklärung von nahezu allen wichtigen forschungsbezogenen Organisationen in Deutschland verabschiedet wurde, fand vom 20. bis 22. Oktober 2003 in Berlin statt. Folgekonferenzen wurden in Genf, Southampton und Gollm organisiert. Weitere Informationen unter <http://oa.mpg.de/openaccess-berlin/berlindeclaration.html>. In diesem Zusammenhang ist auch [www.soros.org/openaccess/g/index.shtml](http://www.soros.org/openaccess/g/index.shtml) relevant.
- 3 M. S. Swaminathan Research Foundation (MSSRF) ist eine seit 1988 in Indien eingetragene Non-Profit-Stiftung. Ihr Engagement gilt u.a. der nachhaltigen Entwicklung im sozialwirtschaftlichen Bereich, gesunder Ernährung sowie den Themen Information, Ausbildung und Kommunikation (vgl. a. [www.mssrf.org/](http://www.mssrf.org/)).
- 4 Der Berichterstatter konnte im Rahmen eines Forschungsaufenthalts am *Documentation Research and Training Centre* des *Indian Statistical Institute* ([www.drct.isibang.in](http://www.drct.isibang.in)) ebenfalls am Workshop teilnehmen und steht bei Bedarf für Nachfragen gerne zur Verfügung.
- 5 „eIFL.net ist eine unabhängige Stiftung, deren Ziel es ist, elektronische Ressourcen für Bibliotheksbesucher in Entwicklungsländern verfügbar zu machen und die Bibliotheken bei der Zugänglichmachung der Ressourcen zu unterstützen. Der Schwerpunkt des Projekts ist es günstige Abonnements auszuhandeln, indem sich viele Länder zusammenschließen, und dahingehend das Entstehen nationaler Bibliothekskonsortien in Mitgliedsländern zu unterstützen.“ Selbstdarstellung auf der Seite [eifl.net](http://eifl.net).
- 6 Derzeit werden ca. 800 IR gezählt und es kommt ungefähr ein IR pro Tag hinzu.

Für die indischen Teilnehmer wurde seitens des DRTC, dem Documentation Research and Training Centre des Indian Statistical Institute in Bangalore ([www.drct.isibang.in](http://www.drct.isibang.in)) Unterstützung bei der Einrichtung der notwendigen technischen Anwendungen angeboten. Hierfür wird in nahezu allen Anwendungen Open Source Software, also unentgeltlich und lizenzfrei nutzbare Software verwendet, um die Entwicklungsdynamik dieses Softwarebereichs zu nutzen sowie finanzielle Abhängigkeiten zu vermeiden. Ohnehin können gerade in den Entwicklungs- und Schwellenländern Bibliotheken – aus dem Wissenschaftsbereich und erst recht Öffentliche Bibliotheken – sich den Einsatz kommerzieller Software kaum leisten. Als Anwendungssoftware wird vom DRTC bevorzugt auf DSpace (<https://dspace.mit.edu/>) verwiesen, es kommen aber auch anderen Digital Library Programme zum Einsatz.<sup>7</sup>

Im Bereich der Open-Access-Zeitschriften<sup>8</sup>, dem zweiten wesentlichen Aktionsstrang der OA-Bewegung, war der Verleger D.K. Sahu vom Verlag Medknow Publications in Mumbai (früher: Bombay) anwesend. Er wies mit zahlreichen Statistiken und überzeugenden ökonomischen Daten nach, dass es kein Widerspruch sein muss, erfolgreicher Verleger und gleichzeitig OA-Befürworter zu sein.<sup>9,10</sup> Die von ihm verlegten Fachzeitschriften von medizinischen Fachgesellschaften habe nach der OA-Stellung eine deutlich erhöhte Nachfrage sowohl online wie auch bezüglich der Printausgaben erfahren. Die Zahl der eingereichten Beiträge stieg ebenfalls. Besonders beeindruckend sind die weltweiten Zitationsraten, die nicht nur für die aktuellen Jahrgänge anstiegen, sondern auch für retrospektiv OA-gestellte Jahrgänge aus den 90-er Jahren. Dies sollte den Befürwortern eher restriktiver Konzepte retrospektiver Be-

reitstellung von digitalisierten Zeitschriften, wie sie z.B. in Teilen von DigiZeitschriften praktiziert werden, zu denken geben.

Die eingeladenen OA-Spezialisten verwiesen in ihren Beiträgen u.a. auf die folgenden Aspekte, die die Realisierung des OA-Konzeptes gerade für die Entwicklungs- und Schwellenländer wünschenswert sein lassen:

- Die Budget-Probleme der Bibliotheken in diesen Ländern sind wegen der Wechselkursgegebenheiten noch dramatischer als in den Industrieländern. So zeigte z.B. eine WHO-Studie aus dem Jahr 2003, dass in 75 Ländern mit einem Bruttosozialprodukt von weniger als 1000 US\$ pro Einwohner 56 Prozent aller medizinischen Einrichtungen keine Fachzeitschriften abonnieren konnten. Auch in der Bruttosozialprodukt-Gruppe bis 3000 US\$ pro Einwohner waren dies immer noch 34 Prozent aller Einrichtungen.<sup>11</sup>
- Die wissenschaftlichen Ergebnisse können wegen fehlender Rezeptionsmöglichkeiten, z.T. aber auch fehlender Veröffentlichungsmöglichkeiten weder in den Schwellenländern selbst noch weltweit in hinreichendem Maße wahrgenommen werden. Hierzu trägt auch bei, dass die auf dem Index des Institute of Scientific Information (ISI) aufbauende Einstufung von Zeitschriften und Zitierungen Veröffentlichungen aus den Schwellenländern kaum berücksichtigt. Auf diese Weise werden gerade regionalspezifische Forschungsergebnisse unzureichend repräsentiert und nutzbar gemacht. Die wirtschaftliche Entwicklung der Länder wird entsprechend gebremst und wirtschaftliche Abhängigkeiten werden stabilisiert.
- Die Veröffentlichung in den jeweiligen Landessprachen<sup>12</sup> erhöht die Chance, dass Forschungsergebnisse nicht nur weltweit, sondern auch in den jeweiligen Ländern und Sprachregionen selbst wahrgenommen werden. Die chinesischen sowie die indischen Vertreter bei der Konferenz hoben gerade diesen Aspekt heraus.
- Die jeweiligen nationalen Forschungsausgaben werden durch OA-Stellung von Forschungsergebnissen – gerade auch in Form von grauer Literatur – effizienter verwendet, weil die Ergebnisse der Forschung eher wahrgenommen und erneut in Forschung oder volkswirtschaftlich relevantes Handeln umgesetzt werden können. Eher allgemeiner Art ist die berichtete Erfahrung, dass gerade IR sich besonders für die Verbreitung des OA-Konzeptes eignen, da sie die Wahrnehmung einer (Forschungs-)Einrichtung und der Veröffentlichungen ihrer Mitglieder deutlich unterstützen – was sich wiederum in besserer

Ausstattung oder weiteren Projekten niederschlagen kann. Inwieweit die Bereitstellung von Veröffentlichungen parallel zur Einreichung z.B. bei einer Zeitschrift als verpflichtend angelegt sein sollte, war in den Diskussionen umstritten. Die Erfahrungen in Großbritannien zeigen jedoch, dass es das erfolgsversprechendste Konzept darstellt – und schon heute von den Wissenschaftlern bei entsprechender organisationsinterner Politik auch zu ca. 80 Prozent akzeptiert wird. Das Indian Institute of Science sowie weitere Forschungseinrichtungen weltweit sind hierfür gute Beispiele.

## Abschlussklärung: Bangalore Commitment

In inhaltlicher Anknüpfung an das speziell auf den Gesundheitsbereich fokussierende „International Seminar on Open Access for Developing Countries“ im brasilianischen Bahia am 21./22.9.2005, auf dem die „Salvador Declaration on Open Access: The Developing World Perspective“<sup>13</sup> verabschiedet worden war, war von den Veranstaltern für die Abschlussdiskussion das sogenannte *Bangalore Commitment* vorbereitet worden. Im Sinne des OA-Gedankens wurde diese Erklärung jedoch nicht verabschiedet, sondern als Diskussionsvorschlag an die zuständigen OA-Foren weitergeleitet, um unter Einbeziehung entsprechender Kommentare eine möglichst breit unterstützte Erklärung zu haben. Diese sollte dann an die politisch und wissenschaftlich relevanten Organisationen der jeweiligen Länder weitergeleitet werden, um deren weitere Unterstützung zu erlangen und so den Open Access-Gedanken auch in der Praxis weiter voran zu bringen.

Tagung, Indien, Elektronisches Publizieren, Open Access, Archiv, Wissenschaft, Veröffentlichung, Zugriff, Entwicklungsländer

### DER AUTOR

Prof. Dr. Achim Oswald



Fachhochschule Köln  
Institut für Informationswissenschaft  
Claudiusstraße 1  
50678 Köln  
achim.osswald@fh-koeln.de

7 Einen aktuellen Überblick mit Bewertung geben Dion Hoe-Lian Goh et al.: A checklist for evaluating open source digital library software. In: Online Information Review 30(2006) 1, 360-379.  
8 Die jeweils aktuellen Daten und Angaben über OA-Zeitschriften können im Directory of Open Access Journals ([www.openaccess.org](http://www.openaccess.org)) eingesehen werden.  
9 Die Folien seiner Präsentation sind unter [http://scigate.ncsi.iisc.ernet.in/OAworkshop2006/presentations/OA\\_IISc\\_Nov3.pdf](http://scigate.ncsi.iisc.ernet.in/OAworkshop2006/presentations/OA_IISc_Nov3.pdf) einsehbar.  
10 Angesichts der nahezu zeitgleich in Deutschland stattfindenden Diskussionen über OA (vgl. z.B. die Inetbib-Diskussion zur Webseite „Was Verlage sich leisten“ vom November 2006) und die Position des Börsenvereins des Deutschen Buchhandels zum Thema OA entsteht dabei der Eindruck, dass in Deutschland von manchen Verlegern wenig der anderen internationalen Erfahrungen rezipiert wird.  
11 Siehe Folie 5 von Barbara Kirsop unter [www.ncsi.iisc.ernet.in/OAworkshop2006/presentations/BarbaraEPT.ppt](http://www.ncsi.iisc.ernet.in/OAworkshop2006/presentations/BarbaraEPT.ppt).  
12 Allein Indien hat 19 offiziell anerkannte Landessprachen.  
13 Vgl. [www.icml9.org/meetings/openaccess/public/documents/declaration.htm](http://www.icml9.org/meetings/openaccess/public/documents/declaration.htm).

## In Memoriam Dr. Udo Schützsack

Waren es seine markanten grauweißen Haare oder sein vielfältiger Einfluss, der Udo Schützsack zur „grauen Eminenz“ der Agrar-, Biologie- und Lebensmittelinformation und -dokumentation machten? Sicher beides und vieles mehr.

Am 14. Dezember 2006 ist Dr. Udo Schützsack im Alter von 84 Jahren verstorben, nachdem er noch 2002 seinen 80. Geburtstag in gewohnter Frische im Kreise besonders vertrauter Kolleginnen und Kollegen gefeiert hatte.

Udo Schützsack war Botaniker. Sein Studium an der Freien Universität Berlin schloss er 1957 erfolgreich mit einer Dissertation über die „Blürrhythmik der Hamameliaceen“ ab.

Wie manche andere Biologen seiner Generation wand er sich einem damals völlig neuen Arbeitsgebiet zu: dem Dokumentations- und Informationswesen.

Gleich seine erste diesbezügliche Tätigkeit von 1957 bis 1963 an der Bundesforschungsanstalt für Ernährung in Karlsruhe waren für sein weiteres Berufsleben und – trotz vieler verschiedener Arbeitsorte – auch für den Mittelpunkt seines privaten Lebens prägend: Information und Dokumentation auf dem Gebiet der Ernährungswissenschaften.

Über dieses Fachgebiet hinaus wirkte er von 1963 bis 1970 am Institut für Dokumentationswesen in Frankfurt am Main. Unter seinem Leiter Martin Cremer hatte dieses Institut der Max-Planck-Gesellschaft – und damit Udo Schützsack – die Aufgabe der Beratung und Entwicklung für und von Informations- und Dokumentationsprojekten in Forschung und Praxis, eine Aufgabe, die sich später insbesondere im internationalen Rahmen bei der GID (Gesellschaft für Information und Dokumentation) in Frankfurt am Main fortsetzte.

Hier konnte Udo Schützsack etwas entwickeln, was man heute als „Netzwerk“ bezeichnet und vielfältige Kontakte zu Universitätsinstituten, Bundesforschungsanstalten, Industrie und zu den relevanten Behörden und Ministerien des Bundes und der Länder knüpfen.

Zentrales Thema der Arbeit von Udo Schützsack war der Aufbau des „International Food Information Service“ (IFIS) zusammen mit Ernest Mann und anderen internationalen Fachleuten und -instituten. 1971 wurde er „Joint Managing Direktor of IFIS“ und konnte einer – wenn

sie Udo Schützsack nicht kannte überraschten – Fachwelt demonstrierten, dass bei großzügigem und unbürokratischem Einsatz von Finanzmitteln, Fachleuten und moderner Technik ein gewissermaßen globalisiertes und sich weitgehend selbst finanzierendes Informationssystem aufgebaut werden konnte.

Das Interessengebiet von Udo Schützsack reichte jedoch weit über die Ernährungswissenschaft hinaus.

Als Biologe wirkte er insbesondere an den frühen Entwicklungen der Biologiedokumentation in Deutschland mit, sei es bei der Biologiedokumentation am Max-Planck-Institut in Schlitz oder bei den Aktivitäten des Senckenberg-Instituts in Frankfurt am Main.

Die Entwicklung der Agrardokumentation in der Bundesrepublik Deutschland und ihre vielfältige internationale Vernetzung begleitete Udo Schützsack durch Höhen und Tiefen mit unermüdlichem Rat, mit Einfluss, vielfach auch mit konkreter, ja auch finanzieller Hilfe, die er durch seine Tätigkeit beim Institut für Dokumentationswesen und als für internationale Aufgaben zuständiger Mitarbeiter der Gesellschaft für Information und Dokumentation anbieten, vermitteln oder bewirken konnte.

Für die Grundlage jeder Dokumentationsarbeit, die Bereitstellung von relevanten Dokumenten, hier für das landwirtschaftliche Bibliothekswesen, hat Udo Schützsack ebenfalls, z.B. als langjähriges Mitglied des Unterausschusses der Deutschen Forschungsgemeinschaft für die Zentralbibliothek des Landbaus engagiert gewirkt.

Information war für Udo Schützsack in doppelter Weise wichtig: Als Objekt seiner Aktivitäten beim Aufbau von Informations- und Dokumentationssystemen im internationalen Rahmen, aber auch als unabdingbare Voraussetzung seines eigenen Handelns und Wirkens. Udo Schützsack war stets ein intensiver Zuhörer. Mit großer Geduld war er bereit, sich Sorgen und Probleme anzuhören, an Sitzungen teilzunehmen und auf Tagungen besonders die Pausen zu nutzen. Die Fülle des dabei Erfahrenen und Erlebten, die unterschiedlichen Meinungen und Haltungen seiner Gesprächspartner, die Statuspapiere und Sitzungsprotokolle, dies alles war die Basis, auf der er sich seine immer unabhängige, häufig überraschende, fast

immer Wirkung zeigende Meinung bildete.

Bürokratie war Udo Schützsack ebenso verhasst wie fremd, auch wenn er sich ihrer mit Fingerspitzengefühl zu bedienen wusste, wenn es im Verkehr mit Behörden notwendig war. Seine Bereitschaft zum Gespräch, seine Kontaktfreudigkeit ermöglichten ihm Wirkung auch in schwierigen Situationen.

Eine Ursache seines Erfolgs war auch seine Präsenz an den wirklich entscheidenden Stellen, die er wie kaum ein anderer zu orten wusste. Kein Ort war ihm fremd, kein Problem zu schwierig, um ihn von einer Reise dorthin oder an Verhandlungen darüber zu hindern – aber auch kein Termin so wichtig, dass er nicht wegen eines wichtigeren verschoben werden konnte.

Wichtig war Udo Schützsack für seine Aktivitäten und seine Meinungsbildung stets die Unabhängigkeit. Auch wenn die „International Association of Agricultural Librarians and Documentalists“ (IAALD), das Fachinformationssystem Ernährung, Landwirtschaft und Forsten und die Gesellschaft für Bibliothekswesen und Dokumentation des Landbaus (GBDL) niemals auf seinen Rat und seine engagierte Hilfe verzichten mussten, die Übernahme von Funktionen in Gremien und Gesellschaften hat er stets vermieden.

Mit Udo Schützsack verliert die agrarwissenschaftliche Information eine tragende Persönlichkeit. Die, die noch mit ihm zusammenarbeiten durften, werden sich seiner vielfältig-dankbar erinnern. Das Informations- und Dokumentationswesen insgesamt wird sein Ansehen stets in Ehren halten.

*Wolfrudolf Laux, Berlin-Dahlem*

## Das Buchregister. Methodische Grundlagen und praktische Anwendung

Robert Fugmann, Frankfurt am Main:  
Reihe Informationswissenschaft der  
DGI, Bd. 10, 2006. 136 Seiten. Preis 25  
Euro. ISBN-10 3-92547-59-5 bzw.  
ISBN-13 978-3-925474-59-0.



Während im angloamerikanischen Sprachraum mittlerweile von der dritten Welle an Buchpublikationen zum Thema ‚Register‘ die Rede ist, ist man hierzulande ja schon froh, wenn sich überhaupt einmal jemand diesem Gegenstand widmet. So gesehen, ist der schmale Band über Das Buchregister von Robert Fugmann an sich bereits verdienstvoll. Überhaupt ist in den letzten Jahren rund um das Thema Register einiges in Gang gekommen. Im Jahre 2004 hat sich das Deutsche Netzwerk der Indexer (DNI) gegründet. Mit bislang überschaubaren 21 Mitgliedern zwar – aber immerhin.

Im Buch selbst begegnet uns vieles, was wir aus Fugmanns Publikationen der letzten 15 Jahre bereits kennen, angefangen von seiner bisweilen recht eigentümlichen Terminologie (die sich z.B. in Ausdrücken wie Kunstsprache und Natursprache niederschlägt) über seine ebenso hartnäckige wie fragwürdige Verdammung jeglicher automatischer Erschließungsverfahren bis hin zur Vorstellung eines Registers, das für sich in Anspruch nehmen kann, anders als die anderen zu sein – davon weiter unten mehr. In den ersten Kapiteln gibt der Autor zahlreiche Beispiele für die Ausdrucksvielfalt und Mehrdeutigkeit der natürlichen Sprache. Daraus leitet er die Unzulänglichkeit einer Erschließung von Buchinhalten ab, die sich ausschließlich auf das Textvokabular stützt, wie es in Stichwortregistern praktiziert wird. Die Tücken der natürlichen Sprache legen gemäß Fugmann vielmehr nahe, sich bei der Inhaltsbildung von der sprachlichen Form des

Buchvokabulars zu lösen und sich auf dessen Bedeutung zu beziehen. Dies vermag die Vergabe von Schlagwörtern – also von Wörtern, die sich nicht an das Textvokabular binden. Der Autor stellt nachfolgend Merkmale und die Erstellung sog. indikativer Register dar, das sind traditionelle Register mit kurzen, knappen Einträgen. Das Buch mündet schließlich in ein Plädoyer für Register mit ausführlichen Untereinträgen (informative Register), deren Vorzüge und Strukturmerkmale ausgiebig zur Sprache kommen und für die das Register der hier besprochenen Publikation Modell steht.

Schauen wir uns dieses Register nun also näher an: Zunächst fällt auf, dass sich die die Fundstellen nicht wie sonst üblich auf Seitenzahlen beziehen, sondern auf Absatznummern. Mag diese Methode für den Registerersteller auch praktisch sein (da sie ihn unabhängig von den Seitenumbrüchen macht und spätere Einschübe oder Streichungen im Buchtext erleichtert) – die Nutzerfreundlichkeit möchte ich aus eigener Erfahrung bezweifeln. Ich jedenfalls suche aus Gewohnheit auch in Fugmanns Büchern stets zunächst nach Seitenzahlen. Und blättere dann und suche und finde nichts und wundere mich und erinnere mich schließlich an die Absätze und blättere wieder.... Im Sinne des Gebots *save the time of the reader* scheint mir das nicht zu sein. Nebenbei verleiht die Absatznumerierung dem Buch eine recht bürokratisch-juristische und damit wenig lesefreundliche Ausstrahlung, zumal in Kombination mit sehr kleiner und eng gesetzter Schrift und starker Fragmentierung des Textes. Nun aber zum informativen Charakter des Registers, den ihm seine Untereinträge verleihen. Das sieht dann z.B. so aus:

Extraktionsindexieren  
(...)  
trägerisch optimal  
Verführung zu, bei elektronischen  
Volltexten  
Bevorzugung durch Auftraggeber  
aus Kostengründen.

Diese Art von Untereinträgen verstellt den Blick auf das Wesentliche eher, als dass sie ihn schärft, wozu nicht zuletzt auch der leseunfreundliche substantivische Stil beiträgt. Und warum Untereinträge auch dort nötig sein sollen, wo es nur wenige Fundstellen oder sogar nur eine gibt, bleibt das Geheimnis des Autors. Der Leser will doch im Register nicht die Informationen selbst finden, sondern sich mit dessen Hilfe lediglich Zugang dazu verschaffen, ganz im Sinne eines weiteren Gebots, nämlich: *do not duplicate the book in the index.*

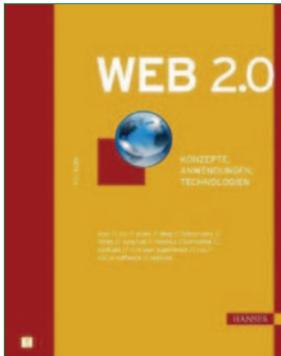
Und worin der zusätzliche Nutzen des ergänzenden systematischen Registers bestehen soll, erschließt sich mir auch nicht. Außerdem: Es ist bisweilen ja schon nicht einfach, einen Verlag davon zu überzeugen, dass ein Buch überhaupt ein Register braucht. Noch schwieriger wird es zu argumentieren, dass dies kein reines Stichwortregister sein sollte, und unmöglich dürfte es schließlich sein, ihm ein informatives Register im Fugmannschen Sinne schmackhaft zu machen. Wenn ich als Autorin dann auch noch auf einem ergänzenden systematischen Register bestehe, wird mir wohl nur noch die Publikation im Selbstverlag übrig bleiben. Und so stehen in diesem Buch überschaubare neunzig Textseiten stolzen 39 Registerseiten gegenüber. Allein zwei Seiten davon nimmt die Bedienungsanleitung ein mit Anweisungen wie: Schlagwörter, die sich im Verhältnis der assoziierten Verwandtschaft zum übergeordneten Schlagwort X befinden, sind unter dem Eintrag <verwandt mit X> gesammelt, deren Verständlichkeit bezweifelt werden darf. Positiv hervorzuheben sind an dem Register seine sehr reichhaltige Verweisstruktur und der Umstand, dass auch der Inhalt von Fußnoten und Abbildungen miteinbezogen wurde. Dennoch: Es hätte den Gebrauchswert der Publikation erheblich gesteigert, hätte sie der Autor mit einem Register versehen, das dem Leser im richtigen (Publikations-)Leben als Vorbild, Beispiel und Modell dienen könnte.

Bei aller Kritik: Das Buch gemahnt eindringlich daran, dass ein gutes Register sowohl den Anforderungen einer Entdeckungsrecherche als auch denen einer Erinnerungsrecherche dienlich sein sollte, um es mit Fugmanns Worten auszudrücken. Heißt: Es soll nicht nur denen zugute kommen, die das Buch gelesen haben, sondern eben gerade auch denjenigen, die es nicht gelesen haben. Nicht nur erinnerte Wortlaute sollen mit ihm wieder auffindbar sein, vielmehr soll das Register auch schnelle Antworten auf die Frage ermöglichen, ob ein Buch zum interessierenden Thema Ergiebiges enthält. Und dass für das zweite Anliegen ein Aufwand zu betreiben ist, der über ein reines Stichwortregister hinausgeht, wird von Fugmann anschaulich und überzeugend dargestellt. Auch seinem Wunsch, dass bei der Rezension von Büchern Vorhandensein und Qualität des Registers (stärkere) Berücksichtigung finden mögen, ist uneingeschränkt zuzustimmen. Dass das von ihm produzierte Register des Guten zu viel tut, steht auf einem anderen Blatt.

Jutta Bertram, Eisenstadt

## Web 2.0. Konzepte, Anwendungen, Technologien

von Tom Alby. München, Wien: Carl Hanser, 2007. 241 Seiten. ISBN 3-446-40931-9. 19,90 Euro.



So ist das mit den Trends. Viele versuchen sich als Trendsetter – und scheitern. Andere suchen nur einen werbewirksamen Titel für eine Veranstaltung, wie das der Verleger Tim O'Reilly 2004 tat, und lösen damit einen Hype aus – wie das neudeutsch heißt. Aber so ganz zufällig war O'Reillys Wortschöpfung „Web 2.0“ nicht. Nach dem Platzen der New-Economy-Blase sollten Fachleute auf einer Tagung diskutieren, wie es nun mit der Internetwirtschaft weiter gehen sollte. In nur kurzer Zeit entwickelte sich das „Web 2.0“ zum Inbegriff einer neuen Bewegung, der nicht nur IT-Enthusiasten angehören. Denn das „join-in-Web“ (das Mitmach-Web), wie das Web 2.0 auch bezeichnet wird, bietet auch IT-Ungeübten die Möglichkeit, im Netz-der-Netz mit eigenen Inhalten, also „user generated content“, präsent zu sein. Den Schlüssel zum World Wide Web erhalten sie durch „Social Software“.

Selbstverständlich geht Tom Alby auch ausführlich auf die Schlüsselbegriffe ein, die sich um das Web 2.0 ranken. Dazu gehören die einer breiten Nutzerschicht bekannten Anwendungen wie Blogs, RSS oder Wikis, wie auch die weniger bekannte Technologien und Entwicklungskonzepte, wie Ajax (Asynchronous JavaScript and XML), Ruby/ Ruby on Rails. In zehn Kapiteln führt Tom Alby in Konzepte, Anwendungen und Geschäftsmodelle des Web 2.0 ein. Neben technischen Aspekten geht er auch auf die soziologischen und psychologischen Aspekte ein, wenn er beispielsweise fragt „Was motiviert die Benutzer, sich in Social Software-Plattformen einzubringen?“ (S. 108) Die Antwort leuchtet ein: „Zunächst einmal ist da die Erwartung der Benutzer, etwas für das Gegebene zurückzubekommen: Wer sein Profil mit vielen Informationen bestückt, der hat auch eine Chance, dass das Profil von anderen angesehen und eventuell Kontakt aufgenommen wird. Wer bei Wikipedia sein Wissen einbringt, der weiß auf der anderen Seite, dass er

von den Beiträgen anderer profitieren kann. Dies kann auch zu einem wirklichen Gemeinschaftsgefühl führen (...) Auch ein gewisser Stolz, Teil dieser Gemeinschaft zu sein, mag dazu gehören.“ (S. 108.f)

Tom Albys „Web 2.0“ zeichnet sich auch durch die vielen Anwendungsbeispielen aus. Ergänzt werden diese durch fünf Interviews mit Praktikern wie Johnny Haesler von Spreeblick.de oder Vasco Sommer-Nunes und Florian Wilken von blog.de. Ein umfangreiches Glossar und ein umfangreiches Stichwortverzeichnis erleichtern die Handhabbarkeit. Getreu dem „Join-in-Web“ gibt es auch ein Blog (<http://web20.buch.alby.de/>) zum Buch, um mit dem Autor in Kontakt zu treten. Ein weiteres Plus ist die leichte Lesbarkeit, die es auch dem Nicht-Informatiker ermöglicht, die technischen Zusammenhänge zu verstehen. Wer mehr über Folksonomy, Ruby on Rails, Ajax oder Long Tail wissen möchte, dem sei Tom Albys „Web 2.0“ sehr empfohlen. Dem IT-Profi wird „Web 2.0“ nicht sehr viel Neues bieten, aber für Agenturen, Designer, Webentwickler, Studierende, BibliothekarInnen bietet dieses engagiert geschriebene Buch viel Nützliches.

Wolfgang Ratzek, Stuttgart

## Medien, Ordnung und Innovation

Dieter Klumpp, Herbert Kubicek, Alexander Rossnagel, Wolfgang Schulz (Hrsg.): Berlin, Heidelberg, New York: Springer, 2006. - 414 Seiten, ISBN 3-540-29157-1 / 978-3-540-29157-2. 59,95 Euro



Die voranschreitende Medienkonvergenz zwingt Medienmanager wie auch Politiker, darüber nachzudenken und Entscheidungen dahingehend zu treffen, ob der bestehende Organisationsrahmen den Herausforderungen gerecht wird, ob eine Reorganisation erforderlich ist oder gar neue Strukturen etabliert werden müssen. Der Konflikt liegt darin, dass einerseits der Gesetzgeber für die Rahmenbe-

dingungen sorgen muss, aber andererseits die Forderung nach mehr Freiheit bei der Umsetzung von sozialen, wirtschaftlichen und technischen Innovationen gefordert wird. In diesem Buch nehmen 41 Experten Stellung zu den drei Teilen „Ordnungsstrategien“, „Regulierungsstrategien“ „Rechtsrahmen und Innovation“. Der erste Teil umfasst neun Beiträge und behandelt eher die technischen Medienaspekte. Zu nennen wären hier beispielsweise „Strategische Aspekte der Frequenzregulierung“ (Matthias Kurth) oder „Der Aufbau von Breitbandinfrastrukturen im internationalen Vergleich“ (Dieter Elixmann; Karl-Heinz Neumann). Der zweite Teil umfasst acht Beiträge und behandelt Themen wie „Broadcasting supervision“ oder „Governance“. Hier geht es unter anderem um „Rundfunkaufsicht: Ordnungsstrukturen im digitalen Zeitalter“ (Klaus Kamps) oder um „Zukunftsmärkte und Kultur: Der Welthandel und die UNESCO-Konvention zum Schutz der kulturellen Vielfalt“ (Verena Metze-Mangold). Der dritte Teil bietet insgesamt 13 Beiträge und stellt in erster Linie juristische Fragestellungen in den Mittelpunkt. Hier wären zu erwähnen „Urheberrecht und Innovation: Basket Two and Beyond“ (Brigitte Zypries), „Urheberrecht und Vertragsfreiheit: Kritische Überlegungen am Beispiel der Verwendungsbeschränkung im IT-Bereich“ (Thomas Hoeren), „Datenschutz bei RFID und Ubiquitous Computing“ (Hansjürgen Garstka) oder „Jugendschutz: Zwischen freiwilliger Selbstblockade und unkontrollierter Selbstkontrolle“ (Karsten Altenhain). Zu den wohl bekanntesten Experten zählen neben Bundesjustizministerin Brigitte Zypries und Hansjürgen Garstka, Vorstandsvorsitzender der Europäischen Akademie für Informationsfreiheit und Datenschutz, und Verena Metze-Mangold, Vizepräsidentin der Deutschen UNESCO-Kommission, auch Jürgen Bulingier, Präsident der Fraunhofer-Gesellschaft, Arnold Picot, Vorsitzender des Münchner Kreises, oder Herbert Kubicek (Professor für Angewandte Informatik, Universität Bremen). Insgesamt liefert „Medien, Ordnung und Innovation“ sehr lesenswerte Beiträge für den organisatorischen Rahmen einer sich dynamisch entwickelnden Informationsgesellschaft. Zu den Beiträgen mit dem höchsten persönlichen Gewinn gehören: Dieter Klumpps „Informationsgesellschaft: Rahmen ohne Bedingungen oder umgekehrt?“, Wolfgang Kleinwächters „Internet Governance: Auf dem Weg zu einem strukturierten Dialog“, Bernd Holznagels und Daniel Krones „Strategien zur Bekämpfung illegaler Inhalte im Internet“ oder Edda Müllers und Michael Bobrowskis „Verbraucherschutz in den Telemedien“.

Wolfgang Ratzek, Stuttgart

Monographien-Zugänge des Monats Februar 2007, zusammengestellt von Annette Bassenge vom Informationszentrum für Informationswissenschaft und -praxis. Die Bücher können unter Angabe der Signatur ausgeliehen werden, sofern es sich nicht um Präsenzbestände handelt. Die Monografien sind gemäß der Kategorien in Infodata angeordnet und innerhalb der Kategorien alphabetisch sortiert.

Fachhochschule Potsdam

Informationszentrum für Informationswissenschaft und -praxis, Friedrich-Ebert-Str. 4, 14406 Potsdam, Tel. (03 31) 580 22 30, Fax (03 31) 580 22 29, bassenge@fh-potsdam.de

## 1. Allgemeines

Ball, R.; Tunger, D.

Bibliometrische Analysen : Daten, Fakten und Methoden. Grundwissen Bibliometrie für Wissenschaftler, Wissenschaftsmanager, Forschungseinrichtungen und Hochschulen

Jülich, DE, 2005. - 79 S.

(Schriften des Forschungszentrums Jülich : Reihe Bibliothek/Library. 12)

ISBN 3-89336-383-1 / ISSN 1433-5557

Bibliometrie / Datenanalyse / Häufigkeitsuntersuchung / Messung / Bewertung  
Sign.: 11 UV1 121

Ockenfeld, M. (Hrsg.); Beling, G. (Red.); Port, P. (Red.); Strohl-Goebel, H. (Red.)

Terminologie der Information und Dokumentation Frankfurt am Main, DE, 2006. - 124 S.

(Reihe Informationswissenschaft der DGI. 9)

ISBN 3-925474-58-7 / ISSN 0940-6662

Information / Dokumentation / Terminologie / Begriff / Fachwort

Sign.: 11 UN3d 116(2)

## 2. Formale Erfassung und inhaltliche Erschließung

Batley, S.

Classification in theory and practice

Oxford, GB: Chandos Publ., 2005. - XI, 180 S.

(Chandos information professional series)

ISBN 1-84334-094-1

Klassifikationsforschung / Klassifikationsmethode / Klassifikationssystem / Dezimale Klassifikation / DK / Facettenklassifikation / Universalklassifikation

Sign.: 11 UN3b 116

## 3. Informationsvermittlung

Brenner-Wilczek, S.; Cepl-Kaufmann, G.; Plassmann, M.

Einführung in die moderne Archivarbeit

Darmstadt, DE: WBG Wiss. Buchges., 2006. - 149 S.

ISBN 3-534-18190-5

Archiv / Archivwesen / Archivgut / Archivierung / Ausbildung Archivwesen / Beruf / Berufsbild / Benutzung / Ordnungssystem / Recherchestrategie / Rechtsfragen / Quellenkunde

Sign.: 11 UA1b 118

Poetzsch, E.

Information Retrieval : Einführung in Grundlagen und Methoden

Berlin, DE: E. Poetzsch Verl., 2006. - 360 S.

ISBN 3-938945-01-X

Information Retrieval / Information Retrieval System / Datenbank / Dialogdienst / Anbieter / Host / Recherchestrategie / Kommandosprache

Sign.: 11 UN7a 137(5)

## 4. Informationsmanagement

Ratzek, W.

Wenn ich nur wüsste ob meine Botschaft angekommen ist? : Beispiele zur Erfolgsbewertung

Wiesbaden, DE: Dinges u. Frick, 2005. - 140 S.

(B.I.T.online - innovativ. 10)

ISBN 3-934997-12-0 / ISSN 1615-1577

Kommunikation / Kommunikationsprozess / Bewertung / Management / Öffentlichkeitsarbeit / Wissensmanagement

Sign.: 11 UG1 304

## 5. Informationssysteme und Anwendungen

Schwicker, A. C.

Web Site Engineering : Ökonomische Analyse und Entwicklungssystematik für eBusiness-Präsenzen

Stuttgart, DE: Teubner, 2001. - 369 S.

(Teubner-Reihe Wirtschaftsinformatik)

ISBN 3-519-00414-3

Rechnernetz / Gestaltung / Marketing / Kostenbewertung / Electronic Commerce / Informationssystem / Betrieb / Benutzung / Informationstechnologie

Sign.: 11 US5 258

## 6. Digitale Bibliothek

White, A.; Kamal, E. D.

E-metrics for library and information professionals : How to use data for managing and evaluating electronic resource collections

London, GB: Facet Publ., 2006. - XIX, 249 S.

ISBN 1-85604-555-2

Bibliotheksaufbau / Bibliotheksdienst / Elektronischer Dienst / Management / Elektronisches Publizieren

Sign.: 11 UH1 179

## 7. Datenkommunikation / Netze / Dienste

Innovationsforum 2005

Enth.: Günther, S.: Das Web Contact Center. Eine Herausforderung für Bibliotheken; Sieweke, B.: Bibliothecae Quo vadis? Herausforderungen an die Bibliothek von morgen; Wiegemann, S.: Implementierung einer benutzerfreundlichen Oberfläche für mobile Endgeräte am Beispiel eines Bibliotheksinformationssystems

Wiesbaden, DE: Dinges u. Frick, 2005. - 271 S.

(B.I.T.online - innovativ. 11)

ISBN 3-934997-13-9 / ISSN 1615-1577

Bibliothek / Bibliotheksdienst / Bibliotheksorganisation / Benutzer / Benutzerfreundlich / Rechnernetz / Benutzerführung / Gerät / Endnutzer

Sign.: 11 UF2 133

Bouzeghoub, M. (Hrsg.); Goble, C. (Hrsg.); Kashyap, V. (Hrsg.); Spaccapietra, S. (Hrsg.)

Semantics of a networked world : Semantics for Grid databases. Proceedings

Berlin, DE: Springer, 2004. - XIII, 326 S.

(Lecture notes in computer science. 3226)

ISBN 3-540-23609-0 / ISSN 0302-9743

Rechnernetz / Global / Semantisches Netz / Datenbank / Informationsmanagement

Sign.: 11 ZC5 270

## 8. Künstliche Intelligenz

Lamberts, K. (Hrsg.); Goldstone, R. L. (Hrsg.)

Handbook of cognition

London, GB: Sage Publ., 2005. - XX, 455 S.

ISBN 0-7619-7277-3

Künstliche Intelligenz / Kognition / Psychologie / Reasoning / Maschinelles Lernen / Entscheidungshilfe

Sign.: 11 ZB1 716

Wilhelm, O. (Hrsg.); Engle, R. W. (Hrsg.)

Handbook of understanding and measuring intelligence

Thousand Oaks, CA, US: Sage Publ., 2005. - X, 542 S.

ISBN 0-7619-2887-1

Künstliche Intelligenz / Problemlösen / Kognition / Reasoning

Sign.: 11 ZB1 715

## 9. Rechtsfragen

Wayman, J. (Hrsg.); Jain, A. (Hrsg.); Maltoni, D. (Hrsg.); Maio, D. (Hrsg.)

Biometric systems : Technology, design and performance, evaluation

London, GB: Springer, 2005. - XIV, 370 S.

ISBN 1-85233-596-3

Datenschutz / Datensicherung / Mustererkennung / Spracherkennung / Biometrie

Sign.: 11 ZL5 234

## 10. Audiovisuelle DV

Dahm, M.

Grundlagen der Mensch-Computer-Interaktion

München, DE: Pearson Studium, 2006. - 367 S.

ISBN 3-8273-7175-9

Mensch-Maschine-Kommunikation / Benutzerschnittstelle / Benutzerführung / Dialogsystem / Kognition / Computer / Benutzung / Informationsverhalten / Software / Ergonomie

Sign.: 11 ZL4 168

De Bra, P. (Hrsg.); Nejd, W. (Hrsg.)

Adaptive hypermedia and adaptive Web-based systems : Proceedings

Berlin, DE: Springer, 2004. - XIV, 442 S.

(Lecture notes in computer science. 3137)

ISBN 3-540-22895-0 / ISSN 0302-9743

Informationssystem / Hypertext / Elektronischer Dienst / Metadaten / Künstliche Intelligenz

Sign.: 11 ZL2 413

# Informations-Retrieval und Dokumentation

Die komplette Anwendung über das Internet zur Miete! Neue Version (LAMP)

Application Hosting

[http:// www.domestic.de](http://www.domestic.de)



## Journal of Chemical Information and Modeling

New York, Vol. 47, No. 1, January 2007

### ARTICLES

#### CHEMICAL INFORMATION

- Zhang, Qing-You; Aires-de-Sousa, João: Random Forest Prediction of Mutagenicity from Empirical Physicochemical Descriptors, pp.1
- Hristozov, Dimitar; Da Costa, Fernando B.; Gasteiger, Johann: Sesquiterpene Lactones-Based Classification of the Family Asteraceae Using Neural Networks and k-Nearest Neighbors, pp.9
- Dias, Jerry Ray: Strain-Free Total Resonant Sextet Benzoids and Their Antisextet Dualists and Retro-Leap-frogs, pp.20
- Melville, James L.; Riley, Jenna F.; Hirst, Jonathan D.: Similarity by Compression, pp.25
- Mazzatorta, Paolo; Tran, Liên-Anh; Schilter, Benoît; Grigorov, Martin: Integration of Structure-Activity Relationship and Artificial Intelligence Systems To Improve in Silico Prediction of Ames Test Mutagenicity, pp.34
- Vogt, Martin; Godden, Jeffrey W.; Bajorath, Jürgen: Bayesian Interpretation of a Distance Function for Navigating High-Dimensional Descriptor Spaces, pp.39
- Schuffenhauer, Ansgar; Ertl, Peter; Roggo, Silvio; Wetzel, Stefan; Koch, Marcus A.; Waldmann, Herbert: The Scaffold Tree – Visualization of the Scaffold Universe of Hierarchical Scaffold Classification, pp.47
- Batista, José; Bajorath, Jürgen: Chemical Database Mining through Entropy-Based Molecular Similarity Assessment of Randomly Generated Structural Fragment Populations, pp.59
- Agrafiotis, Dimitris K.; Bandyopadhyay, Deepak; Farnum, Michael: Radial Clustering: Visualizing the Aggregate Properties of Hierarchical Clusters, pp.69
- #### COMPUTATIONAL CHEMISTRY
- Sciabola, Simone; Morao, Iñaki; Groot, Marcel de: Pharmacophoric Fingerprint Method (TOPP) for 3D-QSAR-Modeling: Application to CYP2D6 Metabolic Stability, pp.76
- Catana, Cornel; Stouten, Pieter F.W.: Novel, Customizable Scoring Functions, Parameterized Using N-PLS, for Structure-Based Drug Discovery, pp.85
- Eitrich, T.; Kless, A.; Druska, C.; Meyer, W.; Grotendorst, J.: Classification of Highly Unbalanced CYP450 Data of Drugs Using Cost Sensitive Machine Learning Techniques, pp.92
- Takeuchi, Hiroshi: Novel Method for Geometry Optimization of Molecular Clusters: Application to Benzene Clusters, pp.104
- Good, Andrew C.; Hermsmeier, Mark A.: Measuring CAMD Technique Performance. 2. How „Druglike“ Are Drugs? Implications of Random Test Set Selection Exemplified Using Druglikeness Classification Models, pp.110
- Mintz, Christina; Clark, Michael; Acree, William E. Jr.; Abraham, Michael H.: Enthalpy of Solvation Correlations for Gaseous Solutes Dissolved in Water and in 1-Octanol Based on the Abraham Model, pp.115
- #### COMPUTATIONAL BIOCHEMISTRY
- Naim, Marwen; Bhat, Sathesh; Rankin, Kathryn N.; Dennis, Sheldon; Chowdhury, Shafinaz F.; Siddiqi, Imran; Drabik, Piotr; Sulea, Traian; Bayly, Christopher I.; Jakalian, Araz; Purisima, Enrico O.: Solvated Interaction Energy (SIE) for Scoring Protein-Ligand Binding Affinities. 1. Exploring the Parameter Space, pp.122
- Obiol-Pardo, Cristian; Rubio-Martinez, Jaime: Comparative Evaluation of MMPBSA and XSCORE To Compute Binding Free Energy in XIAP-Peptide Complexes, pp.134
- #### PHARMACEUTICAL MODELING
- Salt, David W.; Ajmani, Subhash; Critchton, Ray; Livingstone, David J.: An Improved Approximation to the Estimation of the Critical F Values in Best Subset Regression, pp.143
- Palmer, David S.; O'Boyle, Noel M.; Glen, Robert C.; Mitchell, John B.O.: Random Forest Models To Predict Aqueous Solubility, pp.150
- Yuan, Hua; Wang, Yonyan; Cheng, Yiyu: Local and Global Quantitative Structure-Activity Relationship Modeling and Prediction for the Baseline Toxicity, pp.159
- Zhao, Yuan H.; Abraham, Michael H.; Ibrahim, Adam; Fish, Paul V.; Cole, Susan; Lewis, Mark L.; Groot, Marcel de; Reynolds, Derek P.: Predicting Penetration Across the Blood-Brain Barrier from Simple Descriptors and Fragmentation Schemes, pp.170
- Bordas, B.; Belai, I.; Lopata, A.; Szanto, Z.: Interpretation of Scoring Functions Using 3D Molecular Fields. Mapping the Diacyl-Hydrazine-Binding Pocket of an Insect Ecdysone Receptor, pp.176

Hutter, Michael C.: Separating Drugs from Nondrugs: A Statistical Approach Using Atom Pair Distributions, pp.186

Marcou, Gilles; Rognan, Didier: Optimizing Fragment and Scaffold Docking by Use of Molecular Interaction Fingerprints, pp.195

Hou, Tingjun; Wang, Junmei; Zhang, Wei; Xu, Xiaojie: ADME Evaluation in Drug Discovery. 7. Prediction of Oral Absorption by Correlation and Classification, pp.208

Bruce, Craig L.; Melville, James L.; Pickett, Stephen D.; Hirst, Jonathan: Contemporary QSAR Classifiers Compared, pp.219

Wichmann, Karin; Diedenhofen, Michael; Klamt, Andreas: Prediction of Blood-Brain Partitioning and Human Serum Albumin Binding Based on COSMOS-RS Ü-Moments, pp.228

#### BIOINFORMATICS

Doytchinova, Irini A.; Flower, Darren R.: Predicting Class I Major Histocompatibility Complex (MHC) Binders Using Multivariate Statistics: Comparison of Discriminate Analysis and Multiple Linear Regression, pp.234

Andersson, C.R.; Fryknäs, M.; Rickardson, L.; Larsson, R.; Isaksson, A.; Gustafsson, M.G.: In Vitro Drug Sensitivity-Gene Expression Correlations Involve a Tissue of Origin Dependency, pp.239

#### BOOK REVIEWS

Horn, Anselm H.C.: Introduction to Programming with Fortran: With Coverage of Fortran 90, 95, 2003 and 77. By Ian Chivers and Jane Sleightholme. Springer: London, UK, 2005. XX+592 pp and CD-ROM. ISBN 1-84628-053-5. Softcover, pp.249

Heller, Stephen R.: Adobe After Effects 7.0 – Studio Techniques By Mark Christiansen. Adobe (Peachpit) Press: Berkeley, CA. 2006. 576 pp and DVD, ISBN 0-321-38552-7. Softcover, pp.250

Heller, Stephen R.: Adobe Encore DVD 2.0 – Classroom in a Book By the Adobe Creative Team. Adobe (Peachpit) Press: Berkeley, Ca. 2006. 248 pp and DVD, ISBN 0-321-26795-8. Softcover, pp.250

Heller, Stephen R.: Final Cut Express HD: Digital Video Editing for Everyone (Apple Pro Training) By Diana Weynand. Adobe (Peachpit) Press: Berkeley, Ca. 2006. 512 pp and DVD, ISBN 0-321-33533-3. Softcover, pp.250

Heller, Stephen R.: Adobe Premier Pro 2.0 – Studio Techniques By Jacob Rosenberg. Adobe (Peachpit) Press: Berkeley, Ca. 2006. 624 pp and DVD, ISBN 0-321-38547-0. Softcover, pp.250 S.

## Journal of Documentation

London, Vol. 63, No. 2, 2007

### ARTICLES

Hider, Philip M.: Constructing an index of search goal redefinition through transaction log analysis, pp.175

Hider, Philip M.: Search goal redefinition through user-system interaction, pp.188

Boon, Stuart; Johnston, Bill; Webber, Sheila: A phenomenographic study of English faculty's conceptions of information literacy, pp.204

Harbour, Jenny; Chowdhury, Gobinda G.: Use and outcome of online health information services: a study among Scottish population, pp.229

Borrego, Angel; Urbano, Cristóbal: Analysis of the behaviour of users of a package of electronic journals in the field of chemistry, pp.243

Johnson, Frances C.; Crudge, Sarah E.: Using the repertory grid and laddering technique to determine the user's evaluative model of search engines, pp.259

#### BOOK REVIEW

Hjørland, Birger: Wittgenstein, Language and Information: Back to the Rough Ground!"

Rudzioniene, Jurgita: Managing Academic Support Services in Universities: The Convergence

Urquhart, Christine; Spence, Jaqueline: Document Engineering: Analyzing and Designing Documents for Business Informatics and Web Services

Bawden, David: The Turn: Integration of Information Seeking and Information Retrieval in Context

#### EDITORIAL

Bawden, David: The doomsday of documentation?

## Journal of Information Science

East Grinstead, Vol. 33, No. 1, February 2007

Hornig-Jyh, Paul; Jin-Cheon, Na; Soo-Guan, Christopher Khoo: A hybrid approach to fuzzy name search incorporating language-based and text-based principles, pp.3

Kostoff, Ronald N.; del Río, J. Antonio; Cortés, Héctor D.; Smith, Charles; Smith, Andrew; Wagner, Caroline; Leydesdorff, Loet; Karpis, Georg; Malpohl, Guido; Tshiteya, Rene: Clustering methodologies for identifying country core competencies, pp.21

Mobrand, Kathryn A.; Spyridakis, Jan H.: Explicitness of local navigational links: comprehension, perceptions of use, and browsing behaviour, pp.41

Shachaf, Pnina; Hara, Noriko: Behavioral complexity theory of media selection: a proposed theory for global virtual teams, pp.63

Vegas, Jesús; Crestani, Fabio; de la Fuente, ?.: Context representation for web search results, pp.77

Yang, Chyan; Chen, Liang-Chu: Can organizational knowledge capabilities affect knowledge sharing behaviour?, pp.95

East Grinstead, Vol. 33, No. 2, April 2007

Fry, Jenny; Talja, Sanna: The intellectual and social organization of academic fields and the shaping of digital resources, pp.115

Lin, Hsiu-Fen: Effects of extrinsic and intrinsic motivation on employee knowledge sharing intentions, pp.135

Goss, Cynthia; Lowenstein, Steven; Roberts, Ian; DiGiuseppi, Carolyn: Identifying controlled studies of alcohol-impaired driving prevention: designing an effective search strategy, pp.151

Rowley, Jennifer: The wisdom hierarchy: representations of the DIKW hierarchy, pp.163

Hall, Hazel; Goody, Melanie: KM, culture and compromise: interventions to promote knowledge sharing supported by technology in corporate environment, pp.181

Egghe, L.; Goovaerts, M.: A note on measuring overlap, pp.189

Lourdi, Irene; Papatheodorou, Christos; Nikolaidou, Mara: A multi-layer metadata schema for digital folklore collections, pp.197

Yoon, Soon-Bong; Bock, Gee-Woo; Jang, Seungkwon: An evolutionary stage model of cyberspace: a case of Samsung Economic Research Institute, pp.215

Stuart, David; Thelwall, Mike; Harries, Gareth: UK academic web links and collaboration – an exploratory study, pp.231

## Knowledge Organization

Würzburg, Vol. 33, No. 2, Januar 2007

### ARTICLES

Slavic, Aida: UDC in Subject Gateways: experiment or Opportunity?, pp.67

Gabel, Jeff: Improving Information Retrieval of Subjects Through Citation-Analysis, pp.86

Alexiev, Boyan: Terminology Structure for Learner's Glossaries, pp.96

#### BOOK REVIEW

Sun, Dajin; Carter, Ruth C. (eds.): Education for Library Cataloging: International Perspectives. New York: Hawthorn Information Press, 2006. 492 p. ISBN-13: 978-0-7890-3113-6 (pb), pp.119

Levie, Françoise: L'Homme qui voulait classer le monde, Paul Olet et le mundaneum. Bruxelles: Les Impressions Nouvelles, 2006. 352 p. (80 b&w ill.) ISBN 2-87449-022-9, pp.120

ISKO NEWS, pp.122

Knowledge Organization Literature 33(2006)No.2, pp.125

Personal Author Index 33(2006)No.2, pp.126

Würzburg, Vol. 33, No. 3, März 2007

Information Access For The Global Community: An International Seminar On The Universal Decimal Classification: An International Seminar On The Universal Decimal Classification, pp.137

### ARTICLES

Gnoli, Claudio: Phylogenetic classification, pp.138

Bosch, Mela: Ontologies, Different Reasoning Strategies, Different Logics, Different Kinds of Knowledge Representation: Working Together, pp.153

Zhonghong, Wang; Chaudhry, Abdus Sattar; Khoo, Christopher: Potential and Prospects of Taxonomies for Content Organization, pp.160

Kapterev, Andrei I.: Governing the Professional and Intellectual Potential of a Modern Organization: Sociologic Approach, pp.170

Knowledge Organization Literature 33(2006)No.3, pp.176

Personal Author Index 33 (2006)No.3, pp.184

**Online**

Weston, Vol. 31, No. 2, March/April 2007

**FEATURES**

*Dames, K. Matthew:* Trade Agreements as the New Copyright Law, pp.16  
*Ojala, Marydee:* Cambridge Information Group Buys ProQuest Information and Learning: An Interview with CEO Martin Kahn and President Matt Dunie, pp.21  
*Brody, Roberta:* Everybody's a Critic: Business Watchdogs Online, pp.24  
*Banks, Michael A.:* The Source of It All: William F. von Meister, pp.30  
*DiMattia, Susan S.:* How We Teach (or Should Teach) Online Searching, pp.34  
*Affelt, A.:* Researching the „Perfect Payday“ of Corporate Options Backdating, S.29  
*Padilla, Cynthia:* International Search Engines and Directories: Beyond Tourist Sights and Sites, pp.42

**COLUMNS**

*Notess, Greg R.:* On the Net. Wake Up to Web Page Change Alerts, pp.46  
*Ojala, Marydee:* The Dollar Sign. Searching by and for the Book, pp.49  
*Fichter, Darlene:* Intranet Librarian. The Age of Darwinian Design, pp.52  
*Jascó, Péter:* Péter's Picks & Pans. North Carolina State and Plymouth State Universities' OPACs and Dialog's PsycINFO, pp.55  
*Guenther, Kim:* Web Site Management. Self-Service Best Practices, pp.58  
*Wiley, Deborah Lynne:* Hardcopy. Rethinking Information Work: A Career Guide for Librarians and Other Information Professionals / Outside Innovation: How Your Customers Will Co-Design Your Company's Future / Teaching Web Search Skills: Techniques and Strategies of Top Trainers / Blogging for Dummies, pp.61

**DEPARTMENTS**

*Ojala, Marydee:* HomePage. Dangerous Assumptions, pp.5  
*Sabroski, Suzanne:* Industry News, pp.6  
*Notess, Greg R.:* Search Engine Update. New Search Features, Developments, and Content, pp.63

**B.I.T. online – Bibliothek, Information, Technologie, Bibliotheksmanagement**

Wiesbaden, Jg. 10, H. 1, März 2007

**EDITORIAL**

*Fuhlrott, Rolf:* B.I.T. online geht ins 10. Jahr, S.1

**PERSÖNLICHKEITEN**

*Reinitzer, Sigrid:* Dr. Willi Treichler – Der europäische Bibliotheksvertreter aus der Schweiz und Mitherausgeber von B.I.T. online im Ruhestand, S.3

**ABSTRACTS, S.6**

**FACHBEITRÄGE**

*Ruppelt, Georg:* Arbeit als Motiv utopischer Literatur – Arbeit ist die Religion unserer Zeit, S.11

*Wollschläger, Martin:* Elektronischer Zeitschriftenumlauf – Ein Diskussionsbeitrag zur Gestaltung von Zeitschriftenumläufen, S.19

*Schwartz, Dieter:* Patentstatistische und bibliometrische Analysen sowie deren grafische Umsetzung, S.27

**NACHRICHTENBEITRÄGE**

*Balog, Marion; Obermann, Arnaud; Planck, Stefanie Susanne; Wüst, Volker:* Dienstleistungen der Hauptabteilung Dokumentation und Archive des Südwestrundfunks, S.33

*Fisch, Rainer:* Bücherregale als Teil des Tragwerks. Das „Lipmansystem“ – eine Stahlskelett-Konstruktion der Jahrhundertwende wird saniert, S.37

*Knüfer, Rolf:* Aus dem Ledereinband in die Datenwelt. Alter Realkatalog der Staatsbibliothek zu Berlin wandelt sich in Online-Katalog, S.40

**REPORTAGEN**

*Düro, Michael:* „25 Jahre EU-Recht online“ – Eindrücke von der Feierstunde und ein Bericht vom Workshop, S.44

*Münch, Vera:* E-Books, Bücherbissen, ein hybrides Chemiehrbuch und die Rückbesinnung auf das Kerngeschäft. Bericht von der Kongressmesse Online Information 2006, London, S.47

*Ratzek, Wolfgang:* IFLA Presidential Meeting 2007 im Auswärtigen Amt, S.54

Frau Prof. Dr. Claudia Lux im Gespräch mit Wolfgang Ratzek: „... und ich werde sicher alle in Anspruch nehmen, denn es ist viel zu tun“, S.56

*Ratzek, Wolfgang:* Trends in der eLearning-Branche. Bericht von der Online Educa Berlin 2006, S.62

*Volkman, Bernd:* Modernes Informationssystem im Herzen Berlins. Die Universitätsbibliotheken der Technischen Universität und der Universität der Künste Berlin im Volkswagen-Haus, S.64

*Karimov, Ulugbek Fasilovitsch:* Internationale Konferenz und Ausstellung LIBCOM 2006 – Informationstechnologien, Software und Verlagsprodukte für Bibliotheken, S.66

**FIRMENPORTRÄT**

Fünf Jahre Bibliotheca RFID Library Systems, S.69

**DIGITAL NEWS**

BIB-Control unterstützt bei NSM, S.73

Camm Media-Manager vereinfacht Abruf virtueller Dokumente, S.73

Neue Features unter EBSCOhost, S.74

Neues SFX-Konsortium von ExLibris, S.74

IDM Connector von OCLC PICA, S.75

ProQuest First with Search Platform, S.75

**KURZ NOTIERT**

EBSCO in EContent 100 Liste, S.76

Universität Mannheim führt ALEPH ein, S.76

Automatischer Büchertransport: Gilgen Logistics auf dem Leipziger Kongress, S.76

Dokumentationssystem des Niedersächsischen Landtags, S.76

ProQuest Selected as Finalists for SILA CODiE Awards, S.78

Vergabe des VFI-Förderpreises 2006, S.78

Weltweit größte Wirtschaftsbibliothek noch größer, S.78

Startschuss für DiViBib auf Leipziger Kongress, S.78

Lenk bietet CAD-Planungsprogramm, S.80

Call for Paper – WissKom 2007, S.80

Fortbildungsveranstaltungen des DBV – Technische Grundlagen des Bibliotheksmanagements, S.80

Nationaler Aktionstag für Restaurierung und Digitalisierung, S.81

60 Jahre ekz. Bibliothekservice GmbH, S.81

**NEUE PRODUKTE**

Image Access stellt Großformat-Scanner auf CeBIT vor, S.83

Zeutschel präsentiert neues Scannerkonzept, S.84

Was auch immer Sie verleihen – ADT verleiht Ihnen Sicherheit, S.84

**REZENSIONEN**

*Eich, Ulrike (Rez.):* Management, marketing, and promotion of library services, based on statistics, analyses, and evaluation (IFLA Publication 120/121), S.85

*Eich, Ulrike (Rez.):* Aufbruch als Ziel – BID und „Bibliothek 2007“. Zum Abschluss der sechsjährigen Amtszeit von Georg Ruppelt als Sprecher von BID, S.86

*Schmidmaier, Dieter (Rez.):* Hall, Murray G. und Christina Köstner: Allerlei für die Nationalbibliothek ergattern. Eine österreichische Institution in der Nazi-Zeit, S.88

*Schmidmaier, Dieter (Rez.):* Simon, Theresa: Die Positionierung einer Universitäts- und Hochschulbibliothek in der Wissensgesellschaft (ZfBB SB 91), S.89

*Rohde, Christine (Rez.):* Stadttore zur Medienwelt: Geschichte der Dresdner Bürgerbibliothek, S.90

*Plieniger, Jürgen (Rez.):* Umlauf, Konrad: Medienkunde. 2. Auflage, S.91

*Plieniger, Jürgen (Rez.):* Beger, Gabriele: Urheberrecht für Bibliothekare: Eine Handreichung von A-Z, S.91

**BuB – Forum für Bibliothek und Information**

Reutlingen, Jg. 59, H. 2, Februar 2007

**FOYER**

**WISSENSCHAFTLICHE BIBLIOTHEK**

Boom an der Bayerischen Staatsbibliothek / 2006 eine Million Besucher im Lesesaal, S.84

*Beger, Gabriele:* Urheberrecht wird spezifiziert / Anhörung im Rechtsausschuss des Bundestags, S.85

*Wiepcke, Claudia:* FaMI-Azubis auf den Chefesseln, S.85

*Schubert, Eva:* Standards Einigung auf gemeinsames Datenmodell / Erstes RFID-Anwendertreffen in München, S.86

*Henze, Gudrun:* Aktualisierung der RAK-Musik-Anlage M 9 / Neue Werkverzeichnisse aufgenommen, S.86

**ÖFFENTLICHE BIBLIOTHEK**

Autoren übernehmen Patenschaften für Bibliotheken / Vermittlung von deutscher Literatur in Mittel- und Osteuropa, S.86

EU-Ratspräsidentschaft: Drei Konferenzen für Bibliothekare, S.87

„Experimentarium“ lehrt die Gesetze der Physik, S.87

*Pirsich, Volker:* Impulse für die „Interkulturelle Bibliotheksarbeit“ / Neue DBV-Expertengruppe nimmt Arbeit auf, S.88

Der 200. Lektor betreut die romanischen Sprachen, S.88

*Buhl, Günter:* Revolution in der Stadtbibliothek, S.89

**BILDUNGSPARTNER BIBLIOTHEK**  
 Bildung bestimmt Medieneinsatz / Studie zum Verhalten der 11- bis 17-jährigen, S.89

Computer für Jugendliche wichtiger als Fernsehen / Musikhören ist häufigste Offline-Tätigkeit am Rechner, S.89

Gemeinsam stärker werden / Tipps für Netzwerke und Kooperationen, S.91

Schwarzes Brett: Scharfe Kritik ohne moralischen Zeigefinger, S.91

**TAGUNGEN**

Open Access und digitale Ausleihe / hzb-Fachkonferenz diskutiert über „Die Zukunft der Bibliothek“, S.92

*Götz, Caroline; Kohlmeyer, Ingrid:* „Visionen für unsere Zukunft“ / Ein Symposium der Stadtbücherei Heidelberg, S.93

**NACHRICHTEN, S.94**  
 Neues von IFLA, S.95

*Lutzenberger, Manfred:* Nachruf: Georg Braune – Symptathisch, humorvoll, tolerant, S.96

**TERMINE**

Schöne neue Bibliothekswelt, S.97

**LESESAL**

**DAS INTERVIEW**

„Wir müssen stärker mit der Wirtschaft zusammenarbeiten“ / Der Amerikaner Michael Seadle ist neuer Direktor des Instituts für Bibliotheks- und Informationswissenschaft an der Humboldt-Universität und will es zur „iSchool“ weiterentwickeln, S.102

**PRAXIS**

*Buzinkay, Mark:* Im Reich der Auskunfts-Avatare / Bibliotheksarbeit in der virtuellen Parallelgesellschaft „Second Life“, S.107

*Hörning, Beate:* Vorkoster für den Bestandsaufbau / Historie und Stellenwert bibliothekarischer Buchbesprechungen in Deutschland und den USA, S.111

*Peeters, Regina; Plieninger, Jürgen:* Das „Robinson Crusoe-Syndrom“ – und was dagegen getan wurde / Zehn Jahre „One-Person Librarians“-Kommission des BIB, S.115

**AUSLAND**

*Regneala, Mircea:* Aufbruch nach Zensur und Prestigeverlust / Schwieriger Start rumänischer Bibliotheken in die Europäische Union, S.118

*Hachmann, Ute; Krüger, Susanne:* „Wo Geschichten wahr werden“ / DBV-Expertengruppe für Kinder- und Jugendbibliotheken erkundet Trends in den USA, S.121

**MAGAZIN**

**BLICKPUNKT**

*Plieninger, Jürgen:* Internet Weit verstreut und schwer zu finden / Der schnelle Weg zu bibliothekarischen Fortbildungsangeboten, S.125

**FACHLITERATUR**

*Steinhauer, Eric; Sven Berger, Jürgen Roth, Christopher Scheel:* Informationsfreiheitsgesetz / Matthias Rossi: Ö Informationsfreiheitsgesetz, S.127

*Götz, Martin; Konrad Umlauf:* Medienkunde, S.129

Neue Fachliteratur, S.130

**AUS DEM BERUFSVERBAND**

Aus dem Vorstand: Vereinsausschuss beschließt Überleitung der Fachkommissionen in neuer Struktur – Einladung und Anträge zur BIB-Mitgliederversammlung 2007 in Leipzig – Rückblick auf den DIB-Sommerkurs 2006 in Kempten – „(Rechts-)Sicher handeln“: Einladung zum Sommerkurs 2007 in Stuttgart – Aus den Landesgruppen und Kommissionen: Fachexkursion durch Berlin (Landesgruppe Rheinland-Pfalz und Saarland) – BIB-Workshop „Zeitmanagement“ und Mitgliederversammlung in Zweibrücken (Landesgruppe Rheinland-Pfalz) – Mitgliederversammlung mit „Literarischem Spaziergang“ in Bonn (Landesgruppe Nordrhein-Westfalen) – Studienreise nach Zürich (Landesgruppe Brandenburg) – Weiterbildung für FaMIs an Hochschulen (Kommission zur Information von Fachangestellten und Assistenten) – Service: BIB-Fortbildungen – Mitgliedernachrichten, S.131

**SUMMARY / RÉSUMÉ, S.142**

Reutlingen, Jg. 59, H. 3, März 2007

**FOYER**

**RECHT**

*Kuhlen, Rainer:* Ins Steinzeitalter der Informationsversorgung / Scharfe Kritik an der gemeinsamen Stellungnahme von DBV und Börsenverein in Sachen, S.155

Haager, Michael: Blickpunkt Recht: Wann kommt der biometrische Leseausweis? Überlegungen zum Datenschutz in Bibliotheken, S.158  
**WISSENSCHAFTLICHE BIBLIOTHEK**  
 Hekele, Martin: Herausragende akademische Abschlussarbeiten / VFI-Förderungspreis zum ersten Mal vergeben, S.160  
 60.000 Bände von München nach St. Gallen / Die Kantonsbibliothek erhält einzigartigen Bestand rund um das Buch, S.160  
 Einsatz für kriegsbedingt verlagertes Kulturgut / 20 deutsche Bibliotheken einigen sich auf gemeinsame Strategie, S.161  
**ÖFFENTLICHE BIBLIOTHEK**  
 Scharmann, Ute: Gebührendiskussion zwecklos / Die Stadtbibliothek Wuppertal macht gute Erfahrungen mit dem Kassenautomaten, S.162  
 Born, Sigrid: Leseförderung als zentrale Aufgabe / Eine Fortbildung der Hessischen Fachstelle, S.164  
 Würzburger „Theaterbeutel“ feiert Premiere, S.164  
 Starthilfe für die allerjüngsten „Leser“, S.165  
**AUSBILDUNG**  
 Holste-Flinspach, Karin: Berufserfahrung in Europa sammeln / „Leonardo“ ermöglicht Auslandspraktika für FaMIs, S.166  
 „Du kannst Englisch, mach mal!“ / Wie die FaMI-Auszubildende Juliane Bullan in einer britischen Public Library landete, S.166  
 Lang, Ulrike: Mehrwert Auslandserfahrung / Erfolgreiches Jahr 2006 für Bibliothek & Information International, S.168  
**STANDARDS**  
 Kern, Christian; Pohl, Marianne; Schubert, Eva: Auf der Suche nach internationalen Regeln / Erste Sitzung der ISO zum Datenmodell für RFID in Bibliotheken, S.170  
 Die ISBN hat jetzt 13 Stellen / Bücherflut macht Änderung notwendig, S.172  
**BAU**  
 Erweiterung am Deutschen Platz / Nationalbibliothek errichtet in Leipzig neues Gebäude, S.172  
**KIRCHLICHE BIBLIOTHEK**  
 Röska, Günther: „Entspringt den Fesseln, entflieh den Feinden!“ Zaubersprüche in der Merseburger Domstiftsbibliothek, S.173  
**LESESAAAL**  
**BIBLIOTHEKSKONGRESS 2007**  
 Schleh, Bernd: Von den Besten lernen / Fortbildung in Fülle, Ideen aus dem Gastland Dänemark und ein Fragezeichen beim Motto, S.180  
**VERANSTALTUNGEN DES BERUFSVERBANDES**  
 Information Bibliothek (BIB) beim Leipziger Kongress, S.186

Schleihagen, Barbara: Aktionen und Resolutionen für Informationsfreiheit weltweit / Seit zehn Jahren setzt sich der IFLA/FAIFE-Ausschuss für ungehinderten Wissenszugang und Meinungsfreiheit ein, S.188  
 Thorhauge, Jens: Bereit für die nächste Ära des Erfolgs / Öffentliche Bibliotheken Dänemarks und der nordischen Länder in der Wissensgesellschaft, S.196  
 Hamilton, Stuart: Walking the States to Explore Freedom of Thinking / A British Librarian and Researcher Discovers the USA after „9/11“ on Foot, S.202  
**POLITIK**  
 Barbian, Jan-Pieter: Handy-Poesie, „Hürriyet“ und Hochkultur / „Kulturelle Bildung“: Chancen und Perspektiven für Öffentliche Bibliotheken, S.207  
**PRAXIS**  
 Meinhardt, Haiko: Nutzungsrechte für geschützten digitalen Content / Erfahrungen mit DRM in Bibliotheken aus USA, Dänemark und Deutschland, S.213  
**AUSLAND**  
 Leitner, Gerald: „Österreich liest. Treffpunkt Bibliothek“ eine Erfolgsstory / BÜCHEREIVERBAND schafft effektvolle Lobbyarbeit mit furioser Kampagne, S.217  
 Oechtering, Anne: Eine deutsche Bibliothekarin für Amerikas Elite-Studenten / Vom Arbeitsalltag an der Yale University Library / Ausländische Bewerber haben gute Chancen, S.224  
 Lechtenböcker, Ruby: Ausbauhilfe mit Büchern und Computern in Afghanistan / Schusslöcher, zerrissene Medien, veraltete Kataloge: In Kabuls Bibliotheken gibt es viel zu tun, aber wenig Know-how, S.232  
**MAGAZIN**  
**BLICKPUNKT INTERNET**  
 Plieniger, Jürgen: Internet Sollen Bibliotheken auf Sendung gehen? Podcasts in der Bibliotheksarbeit, S.236  
**FACHLITERATUR**  
 Plieniger, Jürgen: Wolfgang G. Stock: Information Retrieval, S.238  
 Vodosek, Peter: Das Gedächtnis der Stadt: 150 Jahre Wienbibliothek im Rathaus, S.239  
**AUS DEM BERUFSVERBAND**  
 Aus dem Vorstand: Dienstleistungen verbessern – Ressourcen optimal einsetzen: Online-Befragung der BIB-Mitglieder geht bis Ende März – Aus den Landesgruppen: Ergebnisse eines Seminars zum Beschwerdemanagement (Landesgruppe Niedersachsen/Bremen) / Aufruf zur Nennung von Kandidatinnen und Kandidaten für den Landesgruppenvorstand 2007 bis 2010 (Landesgruppe Hessen). – Aus den Kommissionen: Jahresberichte der Fachkommissionen (Teil 1). – Service: BIB-Fortbildungen / Mitgliedernachrichten, S.241  
**SUMMARY / RÉSUMÉ, S.251**  
**EDITORIAL, S.155**

**Wissensmanagement**

Reutlingen, Jg. 9, H. 2, März/April 2007  
**EDITORIAL**  
 Lehnert, Oliver: Es lebe die Technik, S.3  
 News, S.6  
**PRAXIS WISSENSMANAGEMENT**  
 Auer, Thomas: Wissensgesellschaft – Quo vadis?, S.10  
 Porsch, Ronny: Wissensmanagement für Informationsprofis, S.14  
 Meyer, Matthias; Unglert, Thomas; Straus, Marko: Jede Sekunde zählt: Wissensmanagement im Kundensupport, S.16  
**TITELTHEMA**  
 Schieferdecker, Richard; Strothotte, Michael: Wissens-Intranet mit Bordmitteln, S.18  
 Paustian, Klaus; Zehe, Janko: Wissen in Bewegung: Fitness-Check für das Intranet, S.22  
 Dücker, Simon: Enterprise 2.0: Neues Denken statt neue Technologie, S.24  
 Gerick, Thomas: Wissen teilen: Strategische Pharma-Allianzen, S.26  
 Weissenberger-Eibl, Marion A.; Borchers, Carsten: Wissen, verstehen, finden, S.28  
**DOKUMENTATION + KOMMUNIKATION**  
 Voigt, Stefan; Fischer, Martin; Staiger, Mark: Erfahrungen sichern – Projektwissen transferieren, S.32  
 Linde, Frank; Löwen, Natalia: Wissensbilanzen in Hochschulen, S.35  
 Ballod, Matthias: Der Faktor Wissen im Business-Process-Management, S.38  
**HUMAN RESOURCES**  
 Hagmaier, Ardeschyr: Wissenszuwächse messbar machen, S.40  
 Kuntz, Bernhard: Der Lern-Campus für Mitarbeiter, S.42  
 Olafsson, Thor: Personalmanagement im Wandel, S.44  
**TRENDS**  
 Schütt, Peter: Die Rolle der Querdenker, S.46  
 Körper, Nicole: CeBIT 2007: Mehr Effizienz im Büro, S.49  
 Aigle, Christiane: Produktdokumentation – Ein Erfolgsfaktor für Ihr Unternehmen?, S.51  
 Middeke, Uwe: Kundenwissen generieren - aber wie?, S.52  
**WISSENSWERTES**  
 Buchvorstellung: Wissensmanagement im Mittelstand, S.56  
 Sloan, Elka: Gesucht: Wissensmanagement (m/w), S.57

Gegründet von H.-K. Soeken † unter dem Titel Nachrichten für Dokumentation (NfD)  
 Herausgegeben von der Deutschen Gesellschaft für Informationswissenschaft und Informationspraxis e.V. (DGI)  
 Präsidentin: Prof. Dr. Gabriele Beger  
 Hanauer Landstraße 151-153  
 D-60314 Frankfurt am Main  
 Telefon: (0 69) 43 03 13  
 Telefax: (0 69) 4 90 90 96  
 mail@dgi-info.de  
 www.dgi-info.de  
 Mitteilungsblatt des Normenausschusses Bibliotheks- und Dokumentationswesen im DIN Deutsches Institut für Normung e.V., der Fachgruppe Dokumentation im Deutschen Museumsbund und der Arbeitsgemeinschaft der Spezialbibliotheken (ASpB)  
**Redaktionsbeirat**  
 Klaus-Peter Böttger, Mülheim an der Ruhr (Berufsfragen Information und Bibliothek)  
 Dr. Sabine Graumann, München (Informationswirtschaft)  
 Prof. Dr. Hans-Christoph Hobohm, Potsdam (Management von Informationseinrichtungen)  
 Prof. Dr. Rainer Kuhlen, Konstanz (Informationswissenschaft)

Prof. Dr. Dirk Lewandowski, Hamb. (Suchmaschinen, Internet)  
 Prof. Dr. Wolfgang Ratzek, Stuttgart (Informationspraxis)  
 Prof. Dr. Ralph Schmidt, Hamburg (Newcomer Report, Medien)  
**Redaktion**  
 Deutsche Gesellschaft für Informationswissenschaft und Informationspraxis e.V.  
 Marlies Ockenfeld (verantwortlich)  
 Viktoriaplatz 8, 64293 Darmstadt  
 Telefon: (0 61 51) 86 98 12  
 Telefax: (0 61 51) 86 97 85  
 ockenfeld@dgi-info.de  
 Daniel Ockenfeld (Redaktionsassistenten)  
**Gastherausgeber**  
 Prof. Dr. Dirk Lewandowski  
**Verlag**  
 Dinges & Frick GmbH  
 Greifstraße 4  
 65199 Wiesbaden  
 Postfach 1564  
 65005 Wiesbaden  
 Telefon: (06 11) 9 31 09 41  
 Telefax: (06 11) 9 31 09 43  
 Bankverbindung:  
 Wiesbadener Volksbank  
 BLZ 510 900 00, Kto-Nr. 714 22 26  
 Postbank Frankfurt  
 BLZ 500 100 60, Kto.-Nr. 267 204-606

**Objektleitung**  
 Erwin König,  
 e.koenig@dinges-frick.de  
**Anzeigenservice**  
 Ursula Hensel Anzeigenservice  
 Hermann-Schuster-Straße 39  
 65510 Hünstetten-Wallbach  
 Telefon: (0 61 26) 57 08 82  
 Telefax: (0 61 26) 58 16 47  
 ursula.hensel@t-online.de  
 Rocco Mischok  
 Verlag Dinges & Frick GmbH  
 Greifstraße 4  
 65199Wiesbaden  
 Telefon: (06 11) 3 96 99-60  
 Telefax: (06 11) 3 96 99-30  
 r.mischok@dinges-frick.de  
**Gestaltung**  
 Anne Karg-Brandt, Hohenstein  
**Druck**  
 Dinges & Frick GmbH  
 Greifstraße 4  
 65199 Wiesbaden  
 Postfach 2009  
 65010 Wiesbaden  
 Telefon: (06 11) 3 96 99-0  
 Telefax: (06 11) 3 96 99-30  
 Leonardo: (06 11) 93 20 79  
 Twist: (06 11) 9 10 23 78  
 df@dinges-frick.de

**Hinweis**  
 Die Aufsätze stellen ausschließlich die Meinung der Autoren dar. Der Inhalt wurde sorgfältig und nach bestem Wissen erarbeitet. Dennoch kann von Verlag und Redaktion eine Gewährleistung auf Richtigkeit und Vollständigkeit nicht übernommen werden. Die Beiträge und die grafischen Darstellungen unterliegen dem Urheberrecht. Nachdruck und Vervielfältigung jeglicher Art bedürfen der Genehmigung des Verlages und der Autoren.  
**Erscheinungsweise/ Bezugspreise**  
 Sieben Hefte jährlich (Doppelausgabe September/Oktober)  
 Jahresabonnement EUR 179,-  
 Schüler/Studenten EUR 125,-  
 Einzelheft EUR 30,-  
 inkl. Versandkosten/Porto.  
 Das Abonnement gilt für mindestens ein Jahr und kann danach bis sechs Wochen zum Ende des Bezugszeitraums gekündigt werden.  
**Redaktionsschluss für**  
 Heft 4/2007 20. April  
 Heft 5/2007 10. Juni  
 Heft 6-7/2007 10. Juli

14. bis 16. Mai Stuttgart	Frühjahrstagung der Fachgruppe 7 im VdA Einschnitte, Ziele und Durchbrüche in der Dokumentation von Medien	VdA – Verband deutscher Archivarinnen und Archivare e.V., Geschäftsstelle, Wörthstr. 3, 36037 Fulda, Tel.: (06 61) 29 109 72, Fax: (06 61) 29 109 74, info@vda.archiv.net
20. bis 23. Mai Schwechat, Österreich	REAL CORP 007 – 12. internationale Konferenz zu Stadtplanung, Regionalentwicklung und Informationsgesellschaft, und 2. Internationale Wiener Immobilienkonferenz und GeoMultimedia 2007	CEIT ALANOVA, Institut für Stadt, Verkehr, Umwelt und Informationsgesellschaft, concorde technology center schwwechat (concorde.tcs), Am Concorde Park 2, Gebäude F, A-2320 Schwwechat, Österreich, Tel.: +43 1 90360 1241, Fax: +43 1 90360 1299, m.schrenk@ceit.at, www.ceit.at
30. Mai bis 1. Juni Köln	10. Internationale Symposium für Informationswissenschaft Open Innovation – neue Perspektiven im Kontext von Information und Wissen?	Prof. Dr. Achim <i>Oßwald</i> , Fachhochschule Köln, Fakultät für Informations- und Kommunikationswissenschaft, Institut für Informationswissenschaft, Claudiusstraße 1, 50678 Köln, achim.osswald@fh-koeln.de, www.isi2007.de
30. Mai bis 1. Juni Köln	13. Jahrestagung der IuK-Initiative	Informationszentrum Sozialwissenschaften, Lennéstr. 30, 53113 Bonn, Tel.: (0228) 22 81-147, iuk2007@iz-soz.de, www.iuk-initiative.org
11. bis 13. Juli Köln	Summer School – Theorie und Praxis der automatischen Indexierung	Evelin <i>Morgenstern</i> , Initiative Fortbildung für wissenschaftliche Spezialbibliotheken und verwandte Einrichtungen e.V., c/o Zentral- u. Landesbibliothek Berlin, Breite Straße 30/31, 10178 Berlin, Tel.: (030) 90226-668, Fax: (030) 8334478, Mobil: 0160 99530089, morgenstern@initiativefortbildung.de, www.initiativefortbildung.de
14. bis 15. Juni Ilmenau	PATINFO 2007. 29. Kolloquium der TU Ilmenau über Patentinformation und gewerblichen Rechtsschutz. „Gewerbliche Schutzrechte: Rationelle Nutzung ihrer Informations- und Rechtsfunktion in Wirtschaft und Wissenschaft“	Prof. Dr.-Ing.habil. Reinhard <i>Schramm</i> , TU Ilmenau, PATON - Patentzentrum Thüringen, PF 10 05 65, 98684 Ilmenau, Tel.: +49 3677 69 4573, Fax: +49 3677 69 4538, reinhard.schramm@tu-ilmenau.de, www.paton.tu-ilmenau.de/
19. bis 23. August Durban, Südafrika	World Library & Information Congress – 73 <sup>rd</sup> IFLA General Conference & Council Libraries for the future. Progress, Development & Partnership	WLIC Conference Secretariat, Congrex Holland BV, Tel.: +31 20 5040 201, Fax: +31 20 5040 225, wlic2007@congrex.nl, www.congrex.com
20. September Berlin	SuMa-eV Forum 2007: Suchmaschinen – Juristisch, Technisch, Wirtschaftlich, Politisch	Dr. Wolfgang <i>Sander-Beuermann</i> , SuMa-eV, Tel.: (05 11) 762-4383, info@suma-ev.de, www.suma-ev.de
24. bis 26. September Ulm	Jahrestagung der Arbeitsgemeinschaft Medizinisches Bibliothekswesen (AGMB) „Medizinbibliotheken mitten im Zentrum von (E-)Learning, Forschung und Patientenversorgung“.	Christiane <i>Wagner</i> , Zentralbibliothek Medizin, -Informationsvermittlungsstelle, Medizinische Fakultät, Universität Leipzig, Liebigstr. 13, 04103 Leipzig, Tel.: +49 (0)341 97-14013, agmb@agmb.de, www.agmb.de
24. bis 28. September Bremen	INFORMATIK 2007 – Informatik trifft Logistik. 37. Jahrestagung der Gesellschaft für Informatik e.V. (GI)	Prof. Dr. Rainer <i>Koschke</i> , Universität Bremen, Fachbereich 3 – Mathematik und Informatik, 28359 Bremen, Tel.: (04 21) 421-218-9671, Fax: (04 21) 218-4322, info@informatik2007.de, www.informatik2007.de
25. bis 28. September Mannheim	77. Deutscher Archivtag Lebendige Erinnerungskultur für die Zukunft	Thilo <i>Bauer</i> M.A., VdA – Verband deutscher Archivarinnen und Archivare e. V., -Geschäftsstelle-, Wörthstraße 3, 36037 Fulda, Tel.: (06 61) 29 109 72, Fax: (06 61) 29 109 74, info@vda.archiv.net, www.archivtag.de/at2007/
25. bis 28. September Berlin	31. Tagung der Arbeitsgemeinschaft der Spezialbibliotheken (ASpB) Kooperation versus Eigenprofil?	Arbeitsgemeinschaft der Spezialbibliotheken / Sektion 5 im DBV c/o Herder-Institut, Bibliothek, Gisonenweg 5-7, 35037 Marburg, geschaeftsstelle@aspb.de, www.aspb.de
8. bis 10. Oktober Berlin	Herbsttagung der Fachgruppe Dokumentation im Deutschen Museumsbund	Monika <i>Hagedorn-Sauppe</i> , Institut für Museumskunde, In der Halde 1, 14195 Berlin, Tel.: (030) 8301-460, Fax: (030) 8301-504, m.hagedorn@smb.spk-berlin.de, www.museumsbund.de
10. bis 12. Oktober Frankfurt am Main	Online-Tagung 2007 und DGI-Jahrestagung 2007 – Information in Wissenschaft, Bildung und Wirtschaft	DGI-Geschäftsstelle, Hanauer Landstraße 151-153, 60314 Frankfurt am Main, Tel.: (069) 43 03 13, Fax: (069) 4 90 90 96, zentrale@dgi-info.de, www.dgi-info.de
10. bis 14. Oktober Frankfurt am Main	Frankfurter Buchmesse Ehregast Katalanische Kultur	Dr. Juergen <i>Boos</i> , Ausstellungs- und Messe GmbH, Buchmesse Frankfurt, Reineckstraße 3, 60313 Frankfurt am Main, Tel.: (069) 2102-0, Fax: (069) 2102-227/-277, info@book-fair.com
21. bis 24. Oktober Barcelona, Spanien	The International Conference in Trends for Scientific Information Professionals	Infonortics, 15 Market Place, Tetbury, Glos. GL8 8DD, UK, Tel.: + 44 (0)1666 505 772, Fax: +44 (0)1666 505 774, contact@infonortics.com, www.infonortics.com
23. bis 26. Oktober München	SYSTEMS 2007 IT.Media.Communication	Messe München GmbH, Messengelände, 81823 München, Tel.: (089) 94 91 17- 18, Fax: (089) 94 9117-19, info@systems.de, www.systems.de
24. bis 27. Oktober Bad Nauheim	4. Deutsche Competitive Intelligence Tagung	Rainer <i>Michaeli</i> , Deutsches Competitive Intelligence Forum (DCIF), michaeli@dcif.de, www.dcif.de
6. bis 8. November Jülich	WissKom2007 – Wissenschaftskommunikation der Zukunft; 4. Konferenz der Zentralbibliothek im Forschungszentrum Jülich	Edith <i>Salz</i> , Forschungszentrum Jülich, 52425 Jülich, Tel.: (0 24 61) 61-29 07, Fax: (0 24 61) 61-61 03, zb-konferenz@fz-juelich.de, www.wisskom2007.de
7. bis 9. November Berlin	Arbeitssitzung der Arbeitsgruppe Elektronische Medien in der Patentinformation	Michael <i>Lingscheid</i> , Bayer Business Services GmbH, Science & Technology IC-ISS, Olof-Palme-Str. 15, D-51368 Leverkusen / Geb. 4810, Raum 2.17, Tel.: (0214) 30 726, 80, Fax: (0214) 30 414 97, michael.lingscheid@bayerbbs.com
November Bremen	Jahrestagung der IASA-Ländergruppe Deutschland/Deutschschweiz e.V.	Detlef <i>Humbert</i> , Südwestrundfunk Dokumentation und Archive Stuttgart, 70150 Stuttgart, detlef.humbert@swr.de, www.iasa-online.de/
15. bis 16. November Stuttgart	3. Stuttgarter Wissensmanagement-Tage	Oliver <i>Lehnert</i> , Wissensmanagement – Das Magazin für Führungskräfte, Schertlinstr. 21, 86159 Augsburg, Tel.: (0821) 45 03 60 55, Fax: (0821) 45 03 60 10, anmeldung@wissensmanagement-tage.de, www.wissensmanagement-tage.de

# CONTENT and COMMUNICATION

Terminology, Language Resources and Semantic Interoperability  
herausgegeben von **International Network for Terminology** *विश्वविद्यालय*



Band 1

*Hans Friedrich Witschel*  
**Terminologie-Extraktion**  
Möglichkeiten der Kombination  
statistischer und musterbasierter  
Verfahren  
2004. 131 S. - 155 x 225 mm.  
Kartoniert. € 24,00  
ISBN 10: 3-89913-408-7  
ISBN 13: 978-89913-408-7



Band 2

*Bertram, Jutta*  
**Einführung in  
die inhaltliche Erschließung**  
Grundlagen - Methoden - Instrumente  
2005. 315 S. - 155 x 225 mm.  
Kartoniert. € 38,00  
ISBN 10: 3-89913-442-7  
ISBN 13: 978-89913-442-1



Band 3

*G. Budin, C. Laurén, H. Picht,  
N. Pilke, M. Rogers, B. Toft (Ed.)*  
**The Theoretical Foundations  
of Terminology Comparison  
between Eastern Europe and  
Western Countries**  
Proceedings of the Colloquium held on 18 August 2003  
in Surrey, Guildford, UK  
in conjunction with the 14th European Symposium on  
Language for Special Purposes (LSP)  
2006. 196 S. - 155 x 225 mm.  
Kartoniert. € 32,00  
ISBN 10: 3-89913-449-4  
ISBN 13: 978-89913-449-0

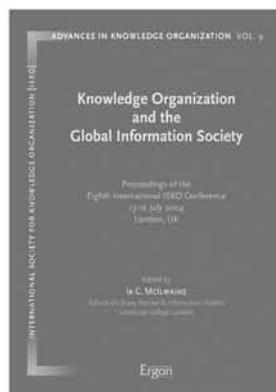
*Please visit our up-to-date catalogue: [www.ergon-verlag.de](http://www.ergon-verlag.de)*

## Advances in Knowledge Organization

ISSN 0938-5495



*Vol. 8* López-Huertas, Maria J. -  
Munoz-Fernandez, Francisco J. (Eds.)  
**Challenges in Knowledge Representation  
and Organization for the 21st Century.  
Integration of Knowledge across Boundaries**  
Proceedings of the Seventh International ISKO  
Conference, 10-13 July 2002, Granada, Spain  
2002. 640 p. - 160 x 230 mm.  
Hardcover. € 49,00  
ISBN 10: 3-89913-247-5  
ISBN 13: 978-3-89913-247-2



*Vol. 9* McIlwaine, Ia C. (Ed.)  
**Knowledge Organization and  
the Global Information Society**  
Proceedings of the Eighth International  
ISKO Conference, 13-16 July 2004,  
London, UK  
2004. 381 p. - 160 x 230 mm.  
Hardcover. € 58,00  
ISBN 10: 3-89913-357-9  
ISBN 13: 978-3-89913-357-8



*Vol. 10* Budin, Gerhard - Swertz,  
Christian - Mitgutsch, Konstantin (Eds.)  
**Knowledge Organization  
for a Global Learning Society**  
Proceedings of the Ninth International ISKO  
Conference, 4-7 July 2006, Vienna, Austria  
2006. 442 p. - 160 x 230 mm.  
Hardcover. € 59,00  
ISBN 10: 3-89913-523-7  
ISBN 13: 978-89913-523-7

**Ergon**

ERGON-Verlag · Grombühlstraße 7 · D-97080 Würzburg Germany  
Phone: ++49 (0)931-280084 · Fax: ++49 (0)931-282872 · e-mail: [service@ergon-verlag.de](mailto:service@ergon-verlag.de)

# Lieber für GENIOS zahlen, als umsonst googeln.

Seriöse Wirtschaftsdaten sollte man sich etwas kosten lassen – es lohnt sich. GENIOS ist der größte Anbieter seriöser, deutschsprachiger Wirtschaftsinformationen. Bei uns können Sie schnell, zuverlässig und unbürokratisch jede Information kaufen – zu Firmen, Branchen oder Personen aus den Bereichen Wirtschaft, Management und Wissenschaft. Wir bieten Ihnen zum Beispiel das Wissen von ● 800 Datenbanken ● 150 Pressearchiven und ● über 400 Fachzeitschriften, außerdem ● 42 Millionen Firmeninformationen ● 10 Millionen Personeninformationen ● 6 Millionen Nachweise aus den Bereichen Wirtschafts- und Sozialwissenschaften sowie ● die Originaldaten des Bundesanzeigers. Und auf Wunsch ● individuelle, maßgeschneiderte Informationslösungen. 25.000 Nutzer profitieren bereits davon.

**Was wollen Sie wissen? [www.genios.de](http://www.genios.de)**



**German Business Information**

GBI-Genios Deutsche Wirtschaftsdatenbank GmbH  
Ein Unternehmen der Frankfurter Allgemeine Zeitung GmbH  
und der Verlagsgruppe Handelsblatt GmbH