



Web-Suche im Umbruch?

Entwicklungstendenzen bei Web-Suchdiensten

Joachim Griesbaum¹, Bernard Bekavac²

¹Universität Konstanz
Informationswissenschaft
Fach D 87
D-78457 Konstanz
{griesbau@inf.uni-konstanz.de}

²HTW Chur
Hochschule für Technik und
Wirtschaft
CH-7004 Chur
{bernard.bekavac@fh-htwchur.ch}

Anmerkung: Der Text spiegelt den aktuellen Stand bei Websuchdiensten von Mitte März 2004. Es ist davon auszugehen, dass einzelne Angaben und Sachverhalte durch die aktuell sehr dynamische Entwicklung innerhalb kurzer Zeit überholt sind. Die aufgezeigten Entwicklungstendenzen und die gezogenen Schlussfolgerungen werden aber voraussichtlich über einen längeren Zeitraum stabil und gültig bleiben.

Zusammenfassung

Angesichts der gegenwärtigen Umbrüche im Suchdienstemarkt versucht dieser Beitrag Entwicklungstendenzen bei Websuchdiensten aufzuzeigen. Ausgehend von einer Darstellung grundlegender Problemfelder werden neue technologische Ansätze in den Themenfeldern Erschließung neuer Wissensbestände und Ausgabe von Suchergebnissen angeführt. Es wird deutlich, dass die Suchdienste zunehmend Inhalte des Deep Web zugänglich machen. Die großen, international bedeutenden Suchdienste integrieren hierzu zunehmend spezielle Wissensbestände und Suchoptionen, insbesondere Produktsuchdienste und geografische Suchdienste. Damit entwickeln sich diese Dienste in einer steigenden Anzahl von Fällen zu den geeignetsten Anlaufstellen zur Befriedigung von Informationsbedürfnissen. Damit diese technologischen Fortschritte, angesichts der geringen Nutzungskompetenz der Anwender, wirksam werden gehen die Suchdienste dazu über, neue Ansätze bei der Ergebnisausgabe umzusetzen, um dem Nutzer die Auswahl der bestgeeigneten Ergebnistypen aus den verschiedenen vorhandenen Indizes abzunehmen. Voraussetzung dafür ist, dass die Suchdienste in der Lage sind die jeweiligen Informationsbedürfnisse der Nutzer zu erkennen. Personalisierung gilt momentan als Schlüsseltechnologie, um dieses Ziel zu



erreichen. Somit deutet sich an, dass sich der Fokus der technologischen Entwicklung langsam wandelt, weg von einer Fixierung auf optimale Rankingverfahren, hin zu einer grundlegenden Orientierung an den subjektiven Bedürfnissen des Nutzers.

1 Fragestellung

Die Bedeutung von Suchdiensten im Web ist kaum zu überschätzen, haben doch Google und andere populäre Suchdienste wie etwa Yahoo, Msn, Lycos, Altavista die Art und Weise vieler Menschen auf Informationen zuzugreifen grundlegend verändert.¹ Als Gatekeeper des Web [Machill & Welp 2003], besitzen sie entscheidenden Einfluss darauf, welche Inhalte des Internet für Nutzer tatsächlich greifbar sind. Mit dem Start der neu entwickelten Yahoo Search Technologie im Februar 2004² werden Verschiebungen auf dem Suchdienstemarkt prognostiziert und ein verstärkter, neu entfachter Wettbewerb zwischen den populären Suchdiensten – insbesondere Yahoo und Google – erwartet.³ Von der dadurch forcierten technologischen Weiterentwicklung werden erhebliche Qualitätssteigerungen bei Web-Suchdiensten erhofft.⁴ Ausgehend von zentralen Problemfeldern und Herausforderungen der Suche im Netz versucht dieser Beitrag angesichts der Umbrüche im Suchdienstemarkt aufzuzeigen, welche Entwicklungspotenziale bei Web-Suchdiensten gegenwärtig vorhanden sind. Die Darstellung beschränkt sich aus Platzgründen auf die Beschreibung aktueller Ansätze in den zentralen Themenfeldern Erschließung neuer Wissensbestände und Ausgabe von Suchergebnissen. Ein Anspruch auf Vollständigkeit kann und soll nicht erhoben, eine einseitige Fokus auf die dominierenden Suchdienste Yahoo und Google allerdings vermieden werden, da neue Suchtechnologien häufig zuerst von Suchdiensten umgesetzt werden, die im Vergleich zu

¹ Als führend bzw. dominant kann gegenwärtig Google bezeichnet werden. URL <http://www.wordspy.com/words/google.asp> (letzter Zugriff 26.02.04). Die „American Dialect Society“ nominierte das Verb „google“ gar als nützlichstes Wort des Jahres 2002 URL <http://www.americandialect.org/woty.html> (letzter Zugriff 26.02.04).

² Welche die Suchtechnologie des Konkurrenten Google auf den Yahoo Portalseiten ablöst. Vgl. Heise News Meldung, Yahoo sucht selbst, vom 18.02.04 URL <http://www.heise.de/newsticker/meldung/44752> (letzter Zugriff 25.02.04).

³ Chris Sherman, Yahoo! Birth of a New Machine. URL <http://www.searchenginewatch.com/searchday/article.php/3314171> (letzter Zugriff 25.02.04).

⁴ Ebd.

international⁵ oder national⁶ dominierenden Anbietern eher ein Nischendasein führen und nur einer relativ geringen Zahl von Nutzern bekannt sind.

2 Problemfelder von Web-Suchdiensten

Die zentrale Aufgabe von Web-Suchdiensten besteht darin, die Inhalte des Internet zu erschließen. Angesichts der enorm umfangreichen und stetig wachsenden Datenbestände des Internets eine gigantische Herausforderung. Eine Studie der Berkeley University of California schätzt die Größe des aus statischen Webseiten bestehenden Surface Web auf 167 Terrabytes und die Größe der datenbankbasierten, erst auf konkrete Anfrage hin generierten, Wissensbestände des Deep Web⁷ auf 91,850 Terrabytes [Lyman & Hal 2003]. Den populären Web-Suchdiensten gelingt es zwar große Teilbestände des Surface Web zu erfassen, die Bestände des Deep Web werden allerdings nur zu einem geringen Teil erschlossen. Somit ermöglichen die Suchdienste einerseits Zugang zu enormen Datenmengen – Google z.B. zu mehreren Milliarden Webseiten⁸ – zugleich bieten sie aber nur in sehr beschränktem Maße Zugriff auf das im Netz vorhandene Wissen. Damit ergibt sich häufig die paradoxe Situation, dass der Nutzer einerseits in einem Meer potenziell relevanter Treffern zu ertrinken droht und ihm zugleich umfangreiche Wissensbestände vorenthalten bleiben.

Zu diesem Abdeckungsproblem treten qualitative Problemfelder hinzu. Für den Nutzer ist es nicht nur schwierig bis unmöglich alle potenziell relevanten Dokumente zu finden, auch die Qualität der gefundenen Treffer bleibt in mehrfacher Hinsicht zweifelhaft. Begreift man Relevanz als Grad der Übereinstimmung zwischen Dokumentinhalt und Informationsbedürfnis [Robertson 1981] stellt sich insbesondere im Web neben der Frage der Einschlägigkeit im Sinne inhaltlicher Übereinstimmung auch die Frage der

⁵ Google, Yahoo, MSN, AOL, AskJeeves. Vgl. die Zahlen für Januar 2004 für US Surfer von Nielsen Netratings, One in three Americans use a search engine. URL http://www.nielsen-netratings.com/pr/pr_040223_us.pdf (letzter Zugriff 10.03.04).

⁶ T-Online, Google, MSN, Web.de, Yahoo, freenet, AOL, Lycos. Vgl. die Zahlen für September 2003 für Deutschland von Nielsen NetRatings, in Danny Sullivan, Nielsen NetRatings European Search Engine Ratings. URL <http://searchenginewatch.com/reports/article.php/2156441> (letzter Zugriff 06.03.04).

⁷ Unter dem Deep Web werden alle Wissensbestände verstanden, auf welche die Roboterprogramme der Suchmaschinen auf Grund von Zugangsbeschränkungen durch die Anbieter oder technischen Restriktionen nicht problemlos zugreifen können [Bergman 2000].

⁸ Google durchsucht nach eigenen Angaben im März 2004 über 4 Milliarden Webseiten.

Validität im Sinne der Glaubwürdigkeit der gefundenen Informationen. Denn im Internet gibt es häufig keine oder nur eine rudimentäre redaktionelle Qualitätssicherung – Jeder kann alles veröffentlichen. Eine kürzlich veröffentlichte Studie zum Themenbereich Gesundheit belegt die Gefahr, dass veröffentlichte Informationen falsch, ungenau, veraltet oder unvollständig sind.⁹

Brisant ist die angeführte Abdeckungs- und Qualitätsproblematik insbesondere deshalb, weil den typischen Nutzern der populären Suchdienste die genannten Mängel kaum bekannt bzw. nur in geringem Maße bewusst sind. In verschiedenen Untersuchungen wird immer wieder aufgezeigt, dass nicht nur die Nutzungskompetenz der meisten Suchdienstennutzer gering ist, sondern darüber hinaus gehend die Zusammenstellung und Sortierung der Suchtreffer insgesamt von den meisten Nutzern nicht nachvollzogen werden kann [Machill & Welp 2003], S.207. Beispielsweise ist einem Großteil der Nutzer nicht bekannt, dass Suchmaschinenergebnisse z.T. aus kommerziellen Einträgen bestehen [Machill et al. 2002]. Das Abdeckungs- und Qualitätsproblem wird also durch ein Kompetenzproblem der Nutzer enorm verschärft. Häufig führt dies zu der Situation, dass die bekannten populären Suchdienste auch dann als primäre Recherchemittel verwendet werden, wenn systematischere und effektivere Suchwerkzeuge zur Verfügung stehen [Klatt et al. 2001].

Die auf Grund der technologischen Weiterentwicklungen erhoffte Qualitätssteigerung kann damit qualitativ nicht ausschließlich an eher technischen Kriterien, wie Abdeckungsgrad und Trefferqualität, festgemacht werden, sondern ist letztlich aus einer Perspektive zu bewerten, die den Nutzungskontext und die Auswirkungen auf die informationelle Absicherung der Nutzer mit einbezieht. Aufgrund der geringen Bereitschaft der Suchdienstennutzer sich selbstständig ein höheres Maß an Kompetenz, z.B. durch Inanspruchnahme von Suchhilfen, anzueignen [Machill & Welp 2003], S.440, ist es dabei von entscheidender Bedeutung auf welche Art und Weise sich technologische Weiterentwicklungen tatsächlich in typischen Suchkontexten auswirken, bzw. wie nutzerfreundlich sie umgesetzt werden.

⁹ Study Find Problems With Access To Credible Health Information Online
URAC & Consumer WebWatch Pressemitteilung 12.02.04. URL
http://www.urac.org/news_release.asp?navid=news&pagename=news_releases&id=1049
(letzter Zugriff 10.03.04)

3 Erschließung neuer Wissensbestände

Seit einigen Jahren versucht eine neue Generation von Metasuchdiensten gezielten Zugriff auf Wissensbestände des Deep Web zu ermöglichen. Beispielsweise liefert die Suchanfrage „jaguar“ bei der Metasuchmaschine ez2Find.com¹⁰ neben Surface-Web-Treffern verschiedener abgefragter Suchdienste eine Liste von Deep-Web-Datenbanken und ermöglicht die direkte Anfrageweiterleitung.

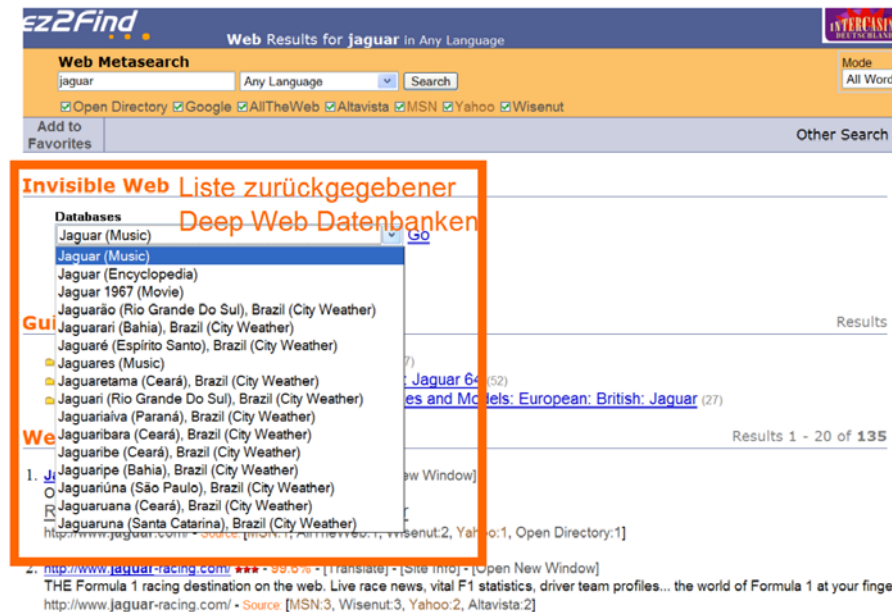


Abbildung 1: Zugriff auf Deep-Web-Datenbanken bei ez2Find.com

Spezielle Deep-Web-Verzeichnisse katalogisieren ähnlich Webverzeichnissen Deep-Web-Ressourcen. Completeplanet.com verzeichnet beispielweise rund 100 000 Deep Web Sites und Suchdienste.¹¹ Incywincy¹² durchsucht Websites, die im Open Directory Project (ODP) erfasst sind und liefert als Ergebnisse ODP-Kategorien und Einträge zurück. Zusätzlich werden die in OPD vorhandenen Seiten mit Hilfe eines Roboters indexiert. Enthalten solche Sites eine Suchfunktion, dann wird die Anfrage an diese Suchfunktion weitergeleitet.

Weiter gehend versuchen Suchdienste wie Turbo10 und Surf wax.com direkten Zugriff auf die Bestände des Deep Web zu ermöglichen. Dazu bieten sie ihren Nutzern die Option sich aus einer Vielzahl vorgegebener Quellen

¹⁰ URL <http://www.ez2www.com/> (letzter zugriff 12.03.04)

¹¹ URL <http://www.completeplanet.com/> (letzter Zugriff 07.03.2004).

¹² URL <http://www.incywincy.com/> (letzter Zugriff 07.03.2004).

(vornehmlich Deep-Web-Datenbanken) individuelle Kollektionen zusammen zu stellen und diese anschließend gezielt zu durchsuchen.

Dabei ermöglicht Turbo10.com, nach eigenen Angaben, im März 2004 Zugriff auf über 2000 Deep Web Datenbanken.¹³

Neben diesen Deep Web Gateways versuchen auch die populären Suchdienste mit dem Wachstum des Surface Web Schritt zu halten¹⁴ und gezielt zusätzliche Wissensbestände recherchierbar zu machen.

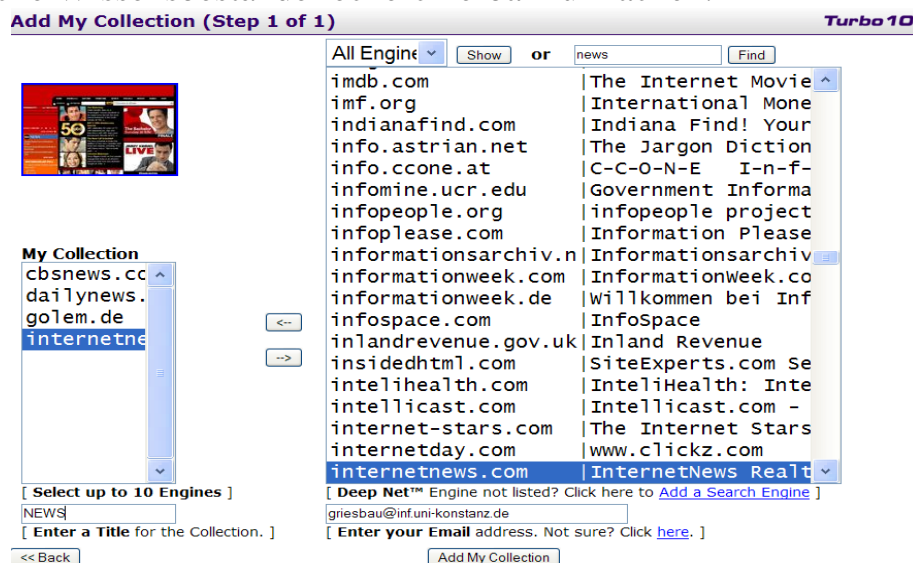


Abbildung 2: Zusammenstellen einer Deep-Web-Kollektion bei Turbo10.com

Google erschließt seit Jahren kontinuierlich neue Datenbestände. Im Unterschied zu den oben genannten Ansätzen der Deep Web Gateways wird aber kein allgemeiner Ansatz gewählt, der darauf zielt ein möglichst breites Spektrum von Deep-Web-Wissensbeständen durchsuchbar zu machen, vielmehr werden schrittweise neue spezialisierte, themenspezifische bzw. dokumenttypspezifische Suchdienste angeboten. Beispiele hierfür sind etwa Google Groups, Google News, Google Catalogs.¹⁵ „Book Searches“ ist ein aktuelles Projekt, mit dem Google das Ziel verfolgt, Inhalte populärer Bücher zu erschließen, indem Exzerpte aus ihnen indexiert werden.¹⁶

¹³ URL <http://turbo10.com/> (letzter Zugriff 07.03.2004).

¹⁴ Resourceshef, Google Ups Total Page Count, News Meldung vom 15.02.04. URL http://www.resourceshef.com/archives/2004_02_01_resourceshef_archive.html#107702946623981034 (letzter Zugriff 07.03.2004).

¹⁵ URL <http://www.google.com/options/> (letzter Zugriff 10.03.2004).

¹⁶ Chris Sherman, Google Introduces Book Searches. URL <http://searchenginewatch.com/searchday/article.php/3290351> (letzter Zugriff 10.03.2004). Ein Beispiel für die Suchanfrage "A Cup of Tea Signature Vertical Note Cards" findet sich unter <http://www.google.com/search?num=100&hl=en&lr=&ie=UTF-8&oe=UTF->

Ein Schwerpunkt der Bereitstellung neuer Datenbestände liegt insbesondere bei der Erschließung von Produktdatenbanken kommerzieller Anbieter, die sich wiederum auf direkte oder indirekte Weise gewinnbringend vermarkten lassen. Ein Beispiel ist Google's Produktsuchmaschine Froogle.¹⁷ Froogle-Treffer werden mittlerweile bei Produkthanfragen wie „monitor“ auf Google.com als „Product Search“-Ergebnisse eingeblendet.¹⁸

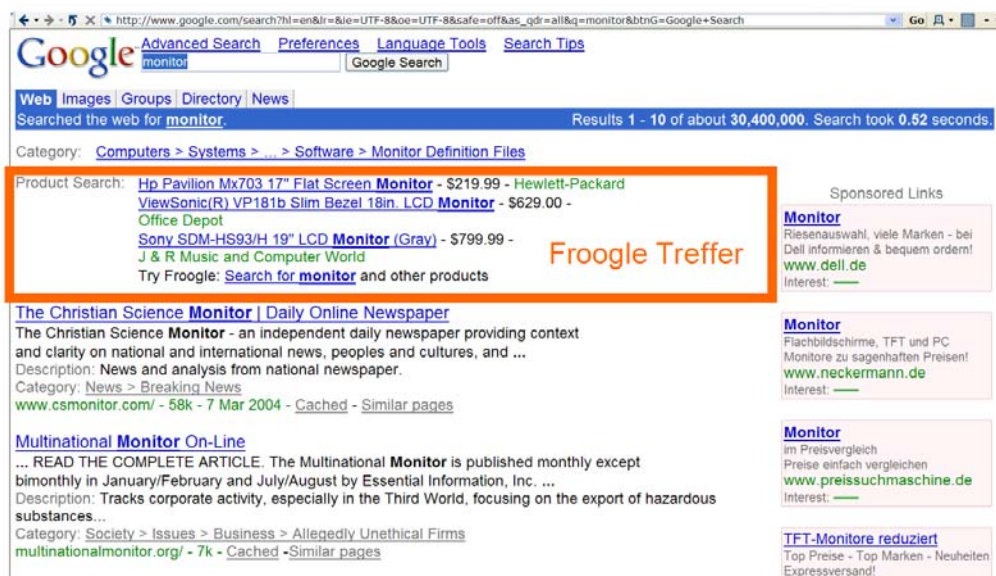


Abbildung 3: Ergebnisse aus der Produktdatenbank von Google.com

Einen weiteren Kernpunkt der Erschließung neuer Wissensbestände bilden Ansätze, die darauf zielen, geografische Suchoptionen verfügbar zu machen. Ziel ist es örtlich spezifiziertes Suchen zu ermöglichen und damit regionen-, bzw. stadt-, stadtteilbezogene oder gar straßengenaue Suchanfragen zu ermöglichen. Ein Beispiel wäre etwa die Suche nach einem Arzt in einer speziellen Stadt. Solche Anfragen in der Art von „san francisco dentist“ sind zwar bei vielen Suchdiensten möglich, aber häufig nicht sehr sinnvoll, da bei typischen Web-Resultaten die notwendigen Ortsangaben nicht strukturiert

8&q=%22A+Cup+of+Tea+Signature+Vertical+Note+Cards%22 (letzter Zugriff 10.03.2004).

¹⁷ URL <http://froogle.google.com/froogle/about.html> (letzter Zugriff 11.03.2004).

¹⁸ URL <http://www.google.com/search?hl=en&ie=UTF-8&oe=UTF-8&q=monitor&btnG=Google+Search> (letzter Zugriff 07.03.2004).

vorhanden sind und deshalb nicht erfasst werden können, so dass i.d.R. eine Vielzahl irrelevanter Treffer zu erwarten ist.¹⁹

Spezielle geografische Suchdienste wie Allesklar.de²⁰ oder Citysearch.com und auch viele Webkataloge erfassen zwar heute schon strukturiert geografische Daten wie Stadtname oder Postleitzahl, respektive Zip-Code zu den vorhandenen Einträgen, insgesamt bieten sie aber nur Zugriff auf einen verschwindend geringen Anteil der Wissensbestände des Netz. Die Entwicklung lokaler Suchdienste, bzw. -optionen ist für Suchdienste sehr attraktiv, da diesem Bereich ein enormes monetarisierbares Werbepotenzial zugesprochen wird.²¹ Deshalb ist es nicht erstaunlich, dass neben spezialisierten geografischen Suchdiensten wie Metrobot.com,²² auch die Major Player Google und Yahoo versuchen derartige Suchoptionen zu integrieren bzw. auszubauen.

Googles Prototyp „Search by Location“ versucht den unstrukturierten Datenbeständen des Surface Web geografische Informationen zu zuordnen und ermöglicht geografische Spezifikationen der Anfragen, so dass der Nutzer explizit festzulegen vermag, auf welche Städte sich seine Anfrage bezieht.²³ Bei der Ergebnispräsentation blendet Google zusätzlich zu den gefundenen Ergebnisseiten eine zoombare Straßenkarte ein, auf der die Adressen der gefundenen Treffer eingezeichnet sind.

¹⁹ Danny Sullivan, Local Search Part 1: New Developments In Local Search & Moves By Overture. URL <http://searchenginewatch.com/searchday/article.php/3091341> (letzter Zugriff 11.03.2004).

²⁰ URL <http://www.allesklar.de/> (letzter Zugriff 11.03.2004).

²¹ Greg Sterling, Local Search Growing, but Small Biz Advertisers Cautious. URL <http://searchenginewatch.com/searchday/article.php/3316741> (letzter Zugriff 11.03.2004).

²² Ein Dienst der als Suchmaschinen Firmenverzeichnisse und Straßenkarten kombiniert. URL <http://www.metrobot.com/def.cfm?j=1> (letzter Zugriff 11.03.2004).

²³ Gegenwärtig ist es möglich U.S.-Städte und Zip-Codes abzufragen. URL http://labs.google.com/help/FAQ_location.html (letzter Zugriff 11.03.2004).

Google Search by Location

google CA 94043
Search terms US address, city & state, or zip

FAQ - Send us your feedback - Discuss

Searched the web for **google** near **CA 94043** Locations 1 - 10 of about **87,900** within **15** miles. Search took **0.76** seconds.

[A] [Google Inc](#)
 [B] [Sgi](#)
 [C] [1310 Shorebird Way, Mountain View, CA](#)
 [D] [350 Ellis St, Mountain View, CA](#)
 [E] [2550 Garcia Ave, Mountain View, CA](#)
 [F] [Alta Vista](#)
 [G] [Advanced Rotorcraft Technology Inc](#)
 [H] [En Pointe Technologies](#)
 [I] [269 Campus Dr, Stanford, CA](#)
 [J] [19420 Homestead Rd, Cupertino, CA](#)

Google Inc [A] - 2 miles NW
 2400 Baysshore Pkwy, Mountain View, CA 94043-1103 - (650) 930-3500
 WSP: TGCA ... Barroso Jeffrey Dean Urs Hölzle **Google** A MENABLE ...
 Happy Birthday, [Google \(TechNews.com\) ...](#) ... Happy Birthday, **Google**, _____ About Filter ...
[and more »](#)

Sgi [B] - 1 miles N
 1600 Amphitheatre Pkwy, Mountain View, CA 94043-1351 - (650) 960-1980
[The Google 2004 Anita Borg ...](#) The **Google** 2004 Anita Borg Scholarships. ...
[SGI - FAQ](#) ... search is powered by The **Google** Search Appliance. ...

Abbildung 4: Google „Search by Location“

Yahoo.com wiederum führt am 10.03.04 mit „SmartView“ entscheidende Verbesserungen beim Spezialsuchdienst Yahoo! Maps ein, der die geografische Suche in Städten mit Hilfe von Stadtplänen gestattet.²⁴ Zusätzlich werden Links zu kulturellen Einrichtungen, Hotels, ÖPNV-Einrichtungen und sonstige lokalen Adressen, soweit in den Yahoo Datenbanken vorhanden, angeboten. Zu den Einträgen werden direkte Links zum Yahoo-Routenplaner und der Web-Suche angeboten. Vorläufig ist der Service auf die U.S. und Kanada beschränkt.

Als erstes Ergebnis lässt sich festhalten: Entgegen früherer Annahmen [Lawrence & Giles 1998] gelingt es den Suchdiensten im Web mit dem Wachstum des Surface Web mitzuhalten. Und nicht nur das, die Web-Suchdienste versuchen in zunehmend Maße Inhalte des Deep Web recherchierbar zu machen. Die eher breiten Ansätze von Deep Web Gateways stehen allerdings noch ganz am Anfang. Angedacht sind spezielle Spider, die versuchen, Inhalte durch dynamische Anfragen an Datenbanken der Deep Web Sites automatisch zu indexieren [Weber 2001]. Spezialisierte Suchdienste verfolgen einen engeren Ansatz und versuchen durch eine Fokussierung auf themen- oder dokumenttypspezifische Teilbereiche spezielle Wissensbestände strukturiert zu erschließen. Die großen, international bekannten Suchdienste, allen voran Google, verfolgen ebenfalls diese Strategie und integrieren zunehmend spezielle Wissensbestände und Suchoptionen. Besonders aktuell sind die Produktsuchdienste und

²⁴ URL <http://www.pandia.com/searchworld/index.html#100304> (letzter Zugriff 11.03.04)

geografische Suchdienste, da diese ein hohes Monetarisierungspotenzial aufweisen. Web-Suchdienste sind derzeit zwar noch sehr weit davon entfernt umfassenden Zugriff auf die offen zugänglichen Wissensbestände des Surface und insbesondere des Deep Web zu ermöglichen, dennoch werden sie bezüglich der Abdeckung stetig besser. Vor allem die Major Player verfügen über immer umfangreichere Web- und Spezialindizes. Damit entwickeln sich die populären Web-Suchdienste in zunehmenden Maße zu den geeignetsten Anlaufstellen zur Befriedigung einer Vielzahl verschiedener z.T. auch sehr spezieller Informationsbedürfnisse.

4 Ausgabe von Suchergebnissen

Suchmaschinenexperte Sullivan karikiert dieses „Zukunftsbild“ illustrativ in nachstehender Grafik und bringt die zu Grunde liegende Problematik in folgendem Satz auf den Punkt: „Doesn't the future of search look great? Whatever type of information you're after, Google and other major search engines will have a tab for it!”²⁵



Abbildung 5:²⁶ Karikatur “How Google will appear in 2005”

Mit der sogenannten „tab blindness“, ist die Tatsache gemeint, dass Suchdienstenutzer in der überwiegenden Zahl der Fälle nicht in der Lage sind, Optionen zur Suchraumeingrenzung wahrzunehmen oder zu verstehen und sie nicht nutzen. Der Mangel an Anwendungskompetenz verhindert eine kompetente Nutzung der Suchdienste: Tabs, Dropdown-Menüs, Radio-

²⁵ Danny Sullivan, Searching With Invisible Tabs. URL

<http://searchenginewatch.com/searchday/article.php/3115131> (letzter Zugriff 11.03.04)

²⁶ Ebd.

Buttons oder sonstige weiter gehende Optionen zu Spezifizierung von Suchanfragen jenseits der Eingabe von Suchtermen werden kaum genutzt.²⁷

Das bedeutet, dass die Fortschritte bei der Erschließung neuer Wissensbestände in typischen Suchkontexten faktisch in den wenigsten Fällen wirksam werden. Die Lösung für dieses Problem wird darin gesehen, dass die Suchdienste die Aufgabe der Datenbank- bzw. Indexauswahl übernehmen und Suchanfragen automatisch – für den Nutzer unsichtbar – an die relevantesten (Spezial)Indizes bzw. Ressourcen weiterleiten.²⁸

Erste Ansätze, in Abhängigkeit von der Art des Informationsbedürfnisse passende Dokumenttypen als Ergebnis auszugeben und damit den Informationsbedarf direkt zu befriedigen, sind in Askjeeves „Smart Search“²⁹ und Yahoo Shortcuts³⁰ zu finden. Askjeeves „Smart Search“ zeigt auf die Anfrage „pictures of dna“ an erster Stelle direkt vier Bilder, die DNA Sequenzen präsentieren und erst an zweiter Stelle die typischen Web Results.



Abbildung 6: Askjeeves „Smart Search“-Ergebnisse auf die Anfrage „pictures of dna“

Yahoo liefert bei verschiedenen Arten von Anfragen sogenannte Shortcuts und gibt beispielsweise, sofern vorhanden, auf Anfragen in der Art „map ortname“ an erster Stelle einen Stadtplan als Ergebnis zurück. Bislang funktioniert das „automatische Antizipieren“ des Informationsbedürfnisses durch Verfahren wie „Smart Search“ und „Shortcuts“ nur bei einer sehr

²⁷ Ebd.

²⁸ Ebd.

²⁹ Chris Sherman, Ask Jeeves Serves Up New Answers. URL

<http://searchenginewatch.com/searchday/article.php/3067941> (letzter Zugriff 09.03.04).

³⁰ URL <http://help.yahoo.com/help/us/ysearch/tips/tips-01.html> (letzter Zugriff 08.03.04).

geringen Zahl von Anfragen, in den meisten Fällen ist die Einblendung bestimmter Ergebnistypen an die Eingabe spezieller Parameter oder Operatoren bei der Anfrageformulierung gebunden. Beispiele hierfür sind etwa der Operator „define“ bei Wortdefinitionen³¹ oder exakt spezifizierbare Suchterme wie etwa ISBN-Nummern.

Bislang funktioniert das automatische Antizipieren des Informationsbedürfnisses also nur für einen kleinen Teil, relativ klar typisierbarer, Suchanfragen. Und zwar vornehmlich dann, wenn sich die Art des Informationsbedürfnisses, anhand syntaktischer Kriterien oder durch Schlüsselwörter zuordnen lässt.

Das zweite Ergebnis ist: Für eine qualitative Verbesserung der Suchergebnisse in typischen Suchkontexten ist es entscheidend, inwieweit die Suchdienste in der Lage sind die Informationsbedürfnisse der Nutzer zu verstehen. Jim Lanzone von Ask Jeeves formuliert diesen zentralen Punkt wie folgt: „Understanding the query perfectly would do much more to impact the quality of results for the everyday user (whether novice or expert) than perfecting the results themselves (though obviously we're trying to do both!).“³²

Jeff Weiner von Yahoo nennt das zentrale Ziel: „Ultimately we want to understand the intention of the user, and I think we're going to get closer to that through personalization, ...“³³

Personalisierung meint die Berücksichtigung individueller Vorlieben und Bedürfnisse bei der Ausgabe von Suchergebnissen. Interaktions- und Profildaten von Nutzern werden dazu genutzt individuelle Informationsbedürfnisse besser nachvollziehen zu können. Ziel ist es künftige Suchkontexte und -intentionen der Nutzer zu antizipieren. Nach gescheiterten Ansätzen³⁴ waren personalisierte Suchfunktionen bei Web-Suchdiensten lange Zeit kaum anzutreffen [Khopkar et al. 2003], erst im Januar 2004 startet mit Eureka³⁵ ein Dienst, der das Nutzungsverhalten von Benutzergruppen

³¹ Bei Google und Yahoo. Vgl. ebd. und URL

<http://www.google.com/help/features.html#definitions> (letzter Zugriff 09.03.04).

³² Andy Beal, Ask Jeeves: What's the Future of Search? URL

<http://searchenginewatch.com/searchday/article.php/3316801> (letzter Zugriff 14.03.04).

³³ Chris Sherman, Yahoo! Birth of a New Machine. URL

<http://searchenginewatch.com/searchday/article.php/3314171> (letzter Zugriff 09.03.04).

³⁴ Danny Sullivan, Google May Get Personal. URL

<http://searchenginewatch.com/sereport/article.php/2164251> (letzter Zugriff 14.03.04)

³⁵ URL <http://home.eureka.com/howitworks.htm> (letzter Zugriff 15.03.04).

auswertet. Dabei werden Treffer, die selektiert werden und eine gewisse Verweildauer aufweisen, höher gewichtet. Die Gewichtung beruht auf zwei Faktoren: Zum einem auf dem Verhalten des Nutzers selbst und zum anderen auf dem Verhalten derjenigen, die zum sozialen Netzwerk des Nutzers gehören. Nutzer definieren dabei ihre sozialen Netzwerke selbst, indem sie mit Freunden, Kollegen etc. „Searchgroups“ bilden. In diesen Gruppen lassen sich gemeinsame Interessen erwarten, weshalb die Berücksichtigung des Interaktionsverhaltens der Gruppe zu besseren Ergebnissen führen soll.³⁶

eurekster™ beta

Account: joachim griesbaum
Network: 1 | 2 | 2 Contacts
SearchGroups | Invite Friends | Logout

how eurekster works

How Search Engine rankings work
The best way to find things on the web has changed and improved over time.

Altavista, Infoseek	Yahoo, Looksmart	Google, Teoma	 what you and your friends think
keywords and meta tags	what professional editors think	what webmasters think	

search engines and how they work

See how eurekster personalizes search results

- 1 Type in a search term e.g. your name
- 2 Click on a search result that you think is best (this can be on any page of the search results). Stay at that website for at least 1 minute (or we will assume that it wasn't useful for you). Repeat this as often as you like.
- 3 In 3 minutes do the same search again and you will notice that the results you preferred will be at the top of the list of search results (excluding sponsored search results).
- 4 We remember the result you liked so you never have to repeat trawling through a long list of search results again! To get eurekster to personalize your search results, [sign-up here!](#)

Abbildung 7: Eureka

Da Personalisierungsfunktionen erhebliches Potenzial zur Verbesserung der Qualität von Suchdiensten zugeschrieben wird, ist zu erwarten, dass andere Suchdienste ebenfalls versuchen werden eine Qualitätssteigerung durch Personalisierung zu erreichen.³⁷ Insbesondere Yahoo wird hierbei ein hohes Potenzial zugeschrieben, dieser Dienst verfügt über eine große Zahl von Nutzern, die personalisierte Dienste nutzen, bzw. Interaktions- und Profildaten hinterlassen, die sich zur Personalisierung der Suche nutzen lassen. Google scheint im Vergleich dazu strukturell benachteiligt. Das Orkut-Projekt,³⁸ ein sozialer Netzwerkdienst 2004 von einem Google-Mitarbeiter initiiert, sowie die Akquise der Firma Outride, deren technologischer

³⁶ Danny Sullivan, Eureka Launches Personalized Social Search. URL <http://searchenginewatch.com/searchday/article.php/3301481> (letzter Zugriff 15.03.04).

³⁷ Ebd.

³⁸ URL <http://www.orkut.com/> (letzter Zugriff 15.03.04).

Schwerpunkt auf der Personalisierung von Suchergebnissen liegt, im Jahre 2001 werden von einschlägigen Quellen aber als Hinweise interpretiert, dass auch Google künftig versuchen wird die Potenziale von Personalisierung zu nutzen, um die Qualität der Ergebnisse zu erhöhen.³⁹ Ob sich die Potenziale personalisierter Suche aber letztlich materialisieren, bleibt offen, denn fraglich ist, ob Personalisierungsoptionen von den Nutzern akzeptiert werden. Das Gefühl der Nutzer durch personalisierte Suchergebnisse Gefahr zu laufen relevante Ergebnisse zu verpassen, sowie datenschutzrechtliche Bedenken gelten als die primären Hindernisse personalisierter Suchfunktionen.⁴⁰

5 Schlussfolgerung und Ausblick

In Bezug zur eingangs aufgeworfenen Fragestellung zu Entwicklungstendenzen bei Web-Suchdiensten lassen sich folgende Schlussfolgerungen ziehen:

Web-Suchdienste ermöglichen Zugriff auf immer umfangreichere Datenbestände. Deep Web Gateways bieten in zunehmenden Maße Zugriff auf Inhalte des Deep Web. Spezialisierte Suchdienste, etwa geografische Suchdienste eröffnen völlig neue Suchoptionen. Gerade die populären Suchdienste, allen voran Google, integrieren zusätzlich zur „StandardWeb-Suche“ eine Vielzahl spezieller Suchdienste. Damit entwickeln sich diese in immer mehr Fällen zu den geeignetsten Anlaufstellen zur Befriedigung von Informationsbedürfnissen. Das angesprochene Abdeckungsproblem wird dadurch zwar nicht grundlegend gelöst, aber in vielen realen Suchkontexten, bspw. bei der Suche nach Produkten und Dienstleistungen zumindest potenziell entschärft.

Damit sich diese theoretische Qualitätssteigerung für den typischen Nutzer tatsächlich auswirkt, gehen die Suchdienste, angesichts der geringen Anwendungskompetenz der Nutzer, dazu über, neue Ansätze wie „Smart Search“ bei Askjeeves oder „Shortcuts“ bei Yahoo bei der Ergebnisausgabe umzusetzen, um dem Nutzer die Auswahl der bestgeeigneten Ergebnistypen aus den verschiedenen vorhandenen Indizes abzunehmen. Voraussetzung dafür ist, dass die Suchdienste in der Lage sind die jeweiligen Informationsbedürfnisse der Nutzer zu erkennen. Personalisierung gilt

³⁹ Danny Sullivan Google Releases Orkut Social Networking Service. URL <http://searchenginewatch.com/searchday/article.php/3302741> (15.03.04)

⁴⁰ Danny Sullivan, Google May Get Personal. URL <http://searchenginewatch.com/sereport/article.php/2164251> (letzter Zugriff 14.03.04)

momentan als Königsweg, um dies zu erreichen. Ein Bereich, der momentan bei Suchdiensten wenig ausgeprägt ist. Das breite Medienecho auf den Vorreiter Eurekster⁴¹ sowie Aussagen u.a. von Yahoo-Mitarbeitern lassen darauf schließen, dass Personalisierung gegenwärtig als zentrale Entwicklungsperspektive wahrgenommen wird. Wie Sullivan im Januar 2004 schreibt: "Personalized search has long been promised as an important next step for increasing relevancy. Now it comes not from Google or Yahoo but instead from Eurekster.... The potential of using your friends or colleagues is enormous....Eurekster is first out of the block for 2004."⁴² Ob diese Fokussierung tatsächlich zu einem qualitativen Sprung bei Suchdiensten führt, bleibt abzuwarten, zumal unklar ist, inwieweit personalisierte Suchdienste und Suchoptionen von den Nutzern angenommen werden. Erste Ansätze zur Personalisierung sind in der Vergangenheit an der mangelnden Akzeptanz der Nutzer gescheitert.

Befindet sich die Web-Suche also gegenwärtig tatsächlich im Umbruch?

Ja, und zwar nicht nur in dem Sinne, dass es Verwerfungen auf dem Markt der Suchdiensteanbieter gibt, sondern auch indem sich der Fokus der technologischen Entwicklung langsam wandelt, weg von einer Fixierung auf optimale Rankingverfahren, wie sie in vor allem in Form von Googles Pagerank die letzten Jahre maßgeblich war, hin zu einer grundlegenden Orientierung an den subjektiven Bedürfnissen des Nutzers. Daraus lässt sich die These ableiten, dass der primäre Fokus technologischer Entwicklungsbemühungen künftig weniger auf der (Weiter)entwicklung universell einsetzbarer, als neutral geltender und deshalb vermeintlich objektiver Sortierverfahren gelegt wird, sondern mehr und mehr auf die subjektiv beste Auswahl an Ergebnissen zielt. Sind solche Verfahren erfolgreich wird damit letztlich das Problem der mangelnden Anwendungskompetenz der Nutzer gemindert. Im Idealfall selektiert das System die passendsten Quellen zur jeweiligen Anfrage und liefert somit automatisch die besten Ergebnisse, ohne dass der Nutzer dazu fortgeschrittene Suchoptionen nutzen muss.

Das grundlegende Qualitätsproblem im Sinne der Validität der ausgelieferten Treffer wird derzeit allerdings nicht systematisch angegangen. Es ist illusorisch zu erwarten, dass Web-Suchdienste „Validitätstests“ bei zurückgelieferten Ergebnisse durchführen (können). Um diesen Mangel zu

⁴¹ URL <http://home.eurekster.com/newsroom.htm> (letzter Zugriff 16.03.04).

⁴² Danny Sullivan, Eurekster Launches Personalized Social Search. URL <http://searchenginewatch.com/searchday/article.php/3301481> (letzter Zugriff 16.03.04).

kompensieren, ist eine möglichst eindeutige Kennzeichnung der Treffer hinsichtlich identifizierbarer qualitativer Merkmale anzustreben. Hilfreich ist dabei nicht nur die oft angemahnte Kennzeichnung kommerzieller Einträge sondern weiter gehend die Einbindung von Zusatzinformationen, die jenseits der typischerweise eingeblendeten Metainformation, die Titel, URL und Beschreibung umfasst, eine genauere Einschätzung der Suchtreffer ermöglichen und damit eine bessere Relevanz- und Validitätseinstufung gestatten. Erste Ansätze hierzu sind bei alternativen Suchdienste wie Alexa.com,⁴³ Zapmeta.com,⁴⁴ Findforward.com⁴⁵ zu finden. Diese Dienste ermöglichen direkten Zugriff auf zusätzliche Informationen, die z.B. Kontaktdaten des Seitenbetreibers, statistische Analysen des Trafficaufkommens und Surfverhaltens umfassen. Zum Teil werden auch Thumbnails eingebunden oder weiter gehend eine Trefferpreview⁴⁶ angezeigt. Diese ermöglicht es, die einzelnen Suchtreffer detailliert zu bewerten, ohne dabei den aktuellen Suchkontext zu verlassen. Findforward.com bietet gar die Option, die einzelnen Treffer mit einem für alle Nutzer sichtbaren Kommentar zu versehen.

Abschließend lässt sich festhalten: Es ist offensichtlich, dass zur kompetenten Nutzung auch bei einer sehr optimistischen Sicht der technologischen Weiterentwicklung von Suchdiensten, die eine erhebliche Qualitätssteigerung der Suchdienste prognostiziert, eine grundlegende Informationskompetenz der Nutzer auch künftig unabdingbar bleibt: Nur so ist er in der Lage die Einflussfaktoren, die das Zustandekommen der Suchergebnisse bestimmen nachvollziehen und die Validität der Suchtreffer beurteilen zu können.

6 Literatur

- Bergman, M. (2000). The Deep Web: Surfacing Hidden Value. White paper. Datum: 2000, <http://www.brightplanet.com/pdf/deepwebwhitepaper.pdf>.
- Khopkar, Y., Spink, A., Giles, C. L., Shah, P. und Debnath, S. (2003). Search engine personalization: An exploratory study. First Monday, 8 Nr.7.
- Klatt, R., Gavriilidis, K., Kleinsimlinghaus, K., Feldmann, M., u.a. (2001). Nutzung elektronischer wissenschaftlicher Information in der Hochschulausbildung : Barrieren und Potenziale der innovativen Mediennutzung im Lernalltag der Hochschulen: Kurzfassung. <http://www.stefi.de/download/kurzfas.pdf>.

⁴³ URL <http://www.alexa.com/> (letzter Zugriff 09.03.04).

⁴⁴ URL <http://www.zapmeta.com/> (letzter Zugriff 09.03.04).

⁴⁵ URL <http://findforward.com/> (letzter Zugriff 08.03.04).

⁴⁶ Ebenso die deutschsprachige Suchmaschine Fireball.de.

- Lawrence, S. & Giles, C. (1998). Searching the World Wide Web. *Science*, 280 Nr.3, 98-100.
- Lyman, P. & Hal, R. V. (2003). How Much Information. Datum: 2003, <http://www.sims.berkeley.edu/how-much-info-2003>.
- Machill, C., Neuberger, C. und Schindler, F. (2002). *Transparenz im Netz. Funktion und Defizite von Internet-Suchmaschinen*. Gütersloh: Bertelsmann Stiftung.
- Machill, M. & Welp, C. (2003). *Wegweiser im Netz. Qualität und Nutzung von Suchmaschinen*. Bertelsmann Stiftung.
- Robertson, S. E. (1981). The methodology of information retrieval experiments. In: *Information Retrieval Experiment*. Jones, K. Sparck (ed.); Butterworth, 9-31.
- Weber, G. (2001). Integration von Datenbanken in Suchmaschinen bei unterschiedlichen Kooperationsgraden. In: *Informatik 2001: Wirtschaft und Wissenschaft in der Network Economy - Visionen und Wirklichkeit*. Tagungsband der GI/OCG-Jahrestagung, 25.-28. September 2001. Bauknecht, K., Brauer, W. und Mück, T. (eds.); Österreichische Computer-Gesellschaft, 345-352.

