

Entwicklungstrends im Web Information Retrieval: Neue Potentiale für die Webrecherche durch Personalisierung & Web 2.0-Technologien?

Joachim Griesbaum
Informationswissenschaft
Universität Konstanz
D-78457 Konstanz

Abstract: In diesem Beitrag werden Entwicklungstrends des Web Information Retrieval, die derzeit vor allem in den Bereichen Personalisierung und beim Einsatz von Web 2.0-Technologien zu beobachten sind, dargestellt und Potentiale und Problemfelder derartiger neuer Ansätze für die Webrecherche diskutiert. Hierzu werden zunächst aus Nutzersicht Suchoptionen in Blogsuch- und Social-Bookmarking-Diensten aufgezeigt. Diese stellen, u.a. durch die Möglichkeit individuelle Profil- oder Alertdienste zu definieren, dem Nutzer mächtige Hilfsmittel zum Aufbau und Erhalt von Expertise zur Verfügung. Wichtige Problemfelder dieser Dienste, zuvorderst das der Qualität nutzergenerierter Inhalte, insbesondere die Spamproblematik, werden angerissen. Im zweiten Teil des Artikels wird ein Überblick über die aktuellen Entwicklungen bei den populären Websuchdiensten Google, Yahoo und MSN mit dem Fokus auf Personalisierung und Web 2.0-Technologien gegeben. Während Microsoft durch „Suchmakros“ und Google durch „Google Co-op“ derzeit einen Schwerpunkt auf nutzerspezifisierbare Suchdienste legen, bietet Yahoo neben dem Experimentierfeld „Yahoo Pipes“ mit „Mein Web“ u.a. derzeit schon Social-Bookmarks an. Noch ist weitgehend unklar, inwieweit diese neuen Dienste und Suchoptionen dauerhaft von den Nutzern angenommen werden, zumal ihr Gebrauch meist mit einer Registrierung verbunden ist, d.h. eine Anmeldung erfordert. Neben erweiterten Recherchemöglichkeiten und einer Verbesserung der Suchergebnisse sind demzufolge auch vermehrt „Lock-in-Effekte“ zu erwarten. Abschließend wird ein gegenwärtig in der Startphase befindliches, vom Verfasser initiiertes Projekt vorgestellt, das auf der Grundlage der vorgestellten partizipativen Konzepte den Aufbau einer offenen, nicht-kommerziellen Community im weiteren Gebiet der Informationswissenschaft zum Themenfeld Web Information Retrieval anvisiert.

1. Einleitung

Zum Jahreswechsel 2006/2007 kündigte Wikipedia-Gründer Jimmy Wales die Entwicklung einer neuen Suchmaschine „Wikia Search“ an.¹ „Wikia Search“ fußt auf dem Wiki-Gedanken und verfolgt den Ansatz, durch menschliche Urteile über die Qualität von maschinellen Suchergebnissen ein besseres Ranking zu erreichen, und will damit die Suche im Web revolutionieren. Ein sehr ambitioniertes Vorhaben und zugleich das gegenwärtig wohl prominenteste Projekt, das dahin zielt, Konzepte und Technologien des sogenannten Web 2.0² gewinnbringend für das Web Information Retrieval zu nutzen. Derartige und ähnliche Ideen – über die Partizipation der Nutzer die Suche im Web zu verbessern – werden derzeit von einer Vielzahl sogenannter Web 2.0-Dienste und in zunehmenden Maße auch von den populären Suchdiensten aufgegriffen und schrittweise umgesetzt.

Aus dieser Perspektive wird nachfolgend ein Überblick über den gegenwärtigen Stand des Web Information Retrieval gegeben. Ziel ist dabei nicht, nur den *State of the Art* der Web-Suche deskriptiv nachzuzeichnen, sondern vor allem aus Nutzersicht Potenziale aufzuzeigen und Problemfelder zu thematisieren. Ein wichtiger Punkt, denn zweifellos stellen Web-Suchdienste inzwischen einen unverzichtbaren Bestandteil der Lebenswirklichkeit vieler Menschen dar. Die Recherche im Internet ist zentraler Bestandteil der informationellen Absicherung eines großen Teils der Bevölkerung und wird zunehmend wichtiger für eine Vielzahl persönlicher Entscheidungen.³ Folgt man [Eberspächer & Holtel 2007], S.V verändern Suchdienste wie Google, Yahoo, MSN die Art und Weise, wie wir mit Information und Wissen umgehen. Ob und inwieweit dies informationskompetent⁴ gelingt, ist zum einen stark daran gebunden, über welche Kenntnisse zu Suchdiensten wir verfügen und zum

¹ URL search.wikia.com/wiki/Search_Wikia (letzter Zugriff 13.06.2007). Vgl. auch das Interview mit Jimmy Wales „Eine offene Gesellschaft braucht offene Institutionen“ in der Süddeutschen Zeitung vom 19.02.2007 URL www.sueddeutsche.de/computer/artikel/719/102617/ (letzter Zugriff 13.06.2007).

² „Web 2.0“ ist ein unscharfer Begriff. In einer ersten Näherung können damit neue Technologien und Dienste des Internet verstanden werden, die mit einer veränderten Wahrnehmung und Nutzung des Web einhergehen und damit verbunden sind, dass in zunehmendem Maße die Benutzer (gemeinsam) Inhalte erstellen und bearbeiten; vgl. [O'Reilly 2005]. Im Kontext dieses Textes bezieht sich der Begriff im Kern darauf, dass Suchdienste zunehmend kollektive Verhaltensmuster berücksichtigen bzw. auf kooperativ generiertem Wissen aufsetzen.

³ So spielen Suchmaschinen etwa bei Kaufentscheidungen eine tragende Rolle. Vgl. u.a. Pressemitteilung vom 21.03.2006, comScore Study Confirms the Importance of Search in Influencing Offline Buying, URL www.comscore.com/press/release.asp?press=796 (letzter Zugriff 13.06.2007), Pressemitteilung vom 14.06.2005, Suchmaschinen-Marketing: Exklusiv-Studie belegt die Notwendigkeit von guten Suchmaschinen-Rankings – auch für etablierte Marken, URL www.eprofessional.de/html/pressemitteilungen/pm-14-06-2005_w3b.html (letzter Zugriff 13.06.2007).

⁴Zum Begriff Informationskompetenz vgl. The Association of College and Research Libraries (ACRL). Information Literacy Competency Standards for Higher Education. URL www.ala.org/ala/acrl/acrlstandards/informationliteracycompetency.htm (letzter Zugriff 13.06.2007).

anderen auch davon abhängig, inwieweit wir bereit sind, unsere darauf aufsetzenden Recherchefähigkeiten auch tatsächlich anzuwenden. Insofern sei an dieser Stelle der Appell an den Leser gerichtet, die nachfolgende Darstellung auch auf den eigenen Recherchealltag zu beziehen und die vorgestellten Dienste und Suchoptionen selbst einmal spielerisch auszuprobieren.

2. Blogsuch- und Social-Bookmarkdienste

2.1 Blogs & Blogsuchdienste

Blogs sind im Kern Webseiten, auf denen regelmäßig – oft täglich – neue Einträge bereit gestellt werden – vgl. für eine ausführlichere Erläuterung [Blood 2002].⁵ Im Unterschied zu anderen Publikationstechnologien ist es mit Weblogs auch für Nutzer, welche nur über geringe Webdesign-Kenntnisse verfügen, sehr einfach, Inhalte im Web zu veröffentlichen. In der Mehrzahl der Blogs publizieren Privatpersonen persönliche Erfahrungen und Erlebnisse. Neben diesen „private Blogs“ werden aber auch zunehmend sogenannte „corporate Blogs“ von Unternehmen bzw. sonstigen Organisationen als Kommunikationsinstrument genutzt [Neuberger et al. 2007]. Ein Beispiel hierfür stellt etwa der offizielle Google Blog dar.⁶ Es kann an dieser Stelle keine vertiefende Diskussion zu sozio-demografischen Merkmalen, gesellschaftlichen Auswirkungen bzw. zentralen kommunikationstheoretischen Aspekten, etwa der Glaubwürdigkeit von Blogeinträgen, geführt werden. Festzuhalten ist dennoch, dass die sogenannte Blogosphäre eine tragende Rolle auf dem Weg zum partizipativen Web 2.0 einnimmt und den in Blogs generierten Inhalten insbesondere eine hohe Aktualität zugeschrieben wird.⁷ Aus technischer Perspektive weisen Blogs im Vergleich zu Standard-HTML-Seiten folgende Eigenschaften auf:

- Einträge werden umgekehrt chronologisch sortiert.
- Jeder Beitrag verfügt über einen permanenten individuellen Link (Permalink).
- Blogs verfügen über verschiedene Technologien, welche die Rückverlinkung zu anderen Blogs erleichtern (Trackback/Pingback).⁸
- Bloginhalte werden in Form strukturierter XML-Formate (RSS-Feeds) bereitgestellt.

⁵ Ergänzend u.a. auch die Einträge zu Weblog in der Wikipedia de.wikipedia.org/wiki/Weblog, en.wikipedia.org/wiki/Weblog.

⁶ URL googleblog.blogspot.com/ (letzter Zugriff 15.06.2007).

⁷ Aus journalistischer Perspektive wird Weblogs insbesondere in Krisensituationen publizistische Relevanz zugeschrieben. Vgl. [Neuberger et al. 2007].

⁸ Vereinfacht ausgedrückt unterstützen diese Technologien die bidirektionale Verlinkung von Blogs in folgender Form: Wenn auf Blog A ein Link zu Blog B gesetzt wird, kann dies auf Blog B in Form eines Beitrags angezeigt werden. Vgl. URL sw-guide.de/webdienste-blogging/was-sind-trackbacks/ (letzter Zugriff 15.06.2007).

Aus der Sicht eines Informationssuchenden sind Dienste, welche die Inhalte der Blogosphäre durchsuchbar machen, zunächst vor allem für Recherchen zu aktuellen Themen relevant. Beispiele für Blogsuchdienste sind u.a. Blogsuche.de, Technorati.com und Daypop.com. Google stellt mit blogsearch.google.com seit September 2005 ebenfalls einen speziellen Blogsuchdienst bereit.

Im Vergleich zur Standardsuche bei Google und Co. weist die Suche in Blogsuchdiensten erweiterte Optionen auf, die sich zunächst primär aus der Kombination bzw. der Nutzung der oben erwähnten technischen Eigenschaften von Blogs ergeben. So bietet die Google-Blogsuche zusätzlich eine Sortierung nach Datum und ermöglicht es, Datumsbereiche mit deutlich höherer Genauigkeit zu suchen als die Google Standardsuchmaschine. Des Weiteren besteht die Option, über RSS-beliebige Suchergebnisseiten zu abonnieren bzw. weiterzuverarbeiten. Blogsuchdienste bieten also nicht nur den Zugriff auf einen Teilbestand des Web, dem eine besonders hohe Aktualität zugeschrieben wird, sondern erlauben strukturiertere Anfragen als Standardsuchmaschinen. Folgende Abbildung zeigt die zeitbasierten Anfrageoptionen der Google- Blogsuche in der Übersicht.

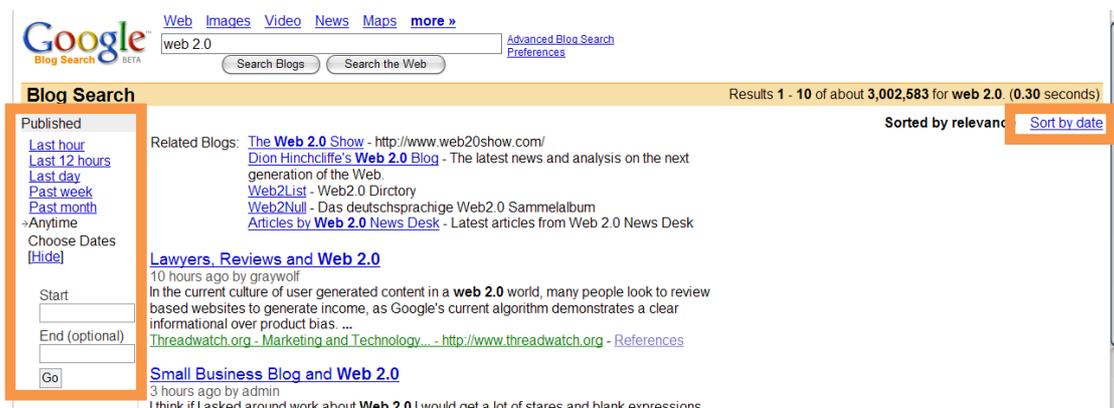


Abbildung 1: Zeitliche Spezifikationen bei Google Blog Search

Technorati.com bietet, neben der Volltextsuche in Blogs, ergänzend auch Zugriff auf nutzergenerierte bzw. von Nutzern bereitgestellte Videos und Bilder. Bei diesen Inhalten beruht die Inhaltserschließung und die Suche nach den Objekten primär auf den, von den Nutzern verfassten, Tags. Tag-basierte Suchoptionen werden nachfolgend anhand von Social-Bookmarkdiensten diskutiert.

2.2. Social-Bookmarkdienste & Tagging-Communities

Social-Bookmarkdienste sind Dienste, die es ihren Nutzer ermöglichen, Bookmarks online – auf dem Server des jeweiligen Anbieters – anzulegen und zu speichern.

Bookmarks können kommentiert, verschlagwortet und anderen Benutzern öffentlich zugänglich gemacht werden. Populäre Social-Bookmarkdienste sind z.B. Del.icio.us (2005 von Yahoo akquiriert), Furl.net, Spurl.net. Im deutschen Sprachraum hat neben Mister-wong.de vor allem auch das auf ein wissenschaftliches Publikum ausgerichtete Bibsonomy eine hohe Bekanntheit erreicht. Die Inhaltserschließung in Social-Bookmarkdiensten ähnelt zunächst der von Webkatalogen – etwa Dmoz.org –, in denen Redakteure Webseiten begutachten und bei Aufnahme in den Katalog einen Titel und eine Beschreibung verfassen. Im Unterschied zu Katalogen werden die Websites nicht in eine Ordnungshierarchie eingebunden, sondern mit Schlagworten – sogenannten Tags – versehen. Die Summe aller Tags aller Nutzer wird auch als Folksonomy⁹ bezeichnet. „Social Tagging“ ist ein Phänomen, das in der Fachwelt breit diskutiert und erforscht wird, für einen Überblick vgl. u.a. [Voß 2007].

Die von den Nutzern vergebenen Terme unterliegen i.d.R. keinerlei terminologischer Kontrolle oder Struktur. Tags werden etwa auch dazu benutzt, um neben inhaltsbezogenen auch formale Aspekte (z.B. Datum) oder gar individuelle emotionale Bezüge („cool“, „lustig“) auszudrücken [Regulski 2007]. Aus klassischer informationswissenschaftlicher Perspektive ist die Verwendung von Polysemien („Jaguar“ als Tier oder Automarke), Synonymen („Auto“ oder „PKW“) oder Flexionen („Suchmaschine“ oder „Suchmaschinen“) und die nicht vorhandene Strukturierung der erfassten Metadaten bzgl. suchergebnisrelevanter Parameter wie Recall und Precision durchaus kritisch einzuschätzen. Dennoch bleibt festzuhalten, dass die Beliebigkeit/Offenheit der Tag-Vergabe nicht nur einen niedrighen Einstieg ermöglicht bzw. den Indexierungsaufwand für die Teilnehmer minimiert – ein wichtiger Erfolgsfaktor –, sondern letztlich einen breiten semantischen Raum aufspannt, in dem die Objekte vielfältig beschrieben und repräsentiert werden können.

Das gemeinsame Indexieren bzw. die daraus entstehende Folksonomy bildet die Grundlage der in Social-Bookmarkdiensten vorhandenen Such- und Browsingfunktionen. Nutzer können über Tagclouds,¹⁰ „Populär“- oder „New“-Listen navigieren, bzw. über eine Stichwortsuche in den erfassten Daten (Titel/Beschreibung/Tags) auf die vorhandenen Bookmarks zugreifen. Die folgende Abbildung zeigt eine Tagcloud in Del.icio.us, die Größe der Tags illustriert deren Verwendungshäufigkeit.

⁹ Folksonomy ist ein Kunstwort, welches aus den Termen „folk“ und „taxonomy“ gebildet wird.

¹⁰ Zweidimensionale Wortwolken, in ihrer Bedeutung für das Browsing annähernd mit den Kategorien in Webkatalogen vergleichbar.



Abbildung 2: Tagcloud in Del.icio.us

Abbildung 2 deutet implizit schon an, dass und wie Social-Bookmarkdienste bzgl. ihrer Suchfunktionalität über Webkataloge hinausgehen. Bei der Ausgabe von Suchergebnissen werden neben einem Termabgleich zwischen Suchanfrage und Bookmarks, Eintragsdatum und Popularität der Einträge berücksichtigt. So können Ergebnislisten in Del.icio.us nach Eintragsdatum (neueste Einträge zuerst) oder Popularität sortiert werden. Popularität bedeutet Anzahl der Nutzer, welche das jeweilige Bookmark gespeichert haben. Folgende Abbildung zeigt zur Veranschaulichung die Ergebnisse in Del.icio.us zur Suchanfrage „suchmaschine seo“.

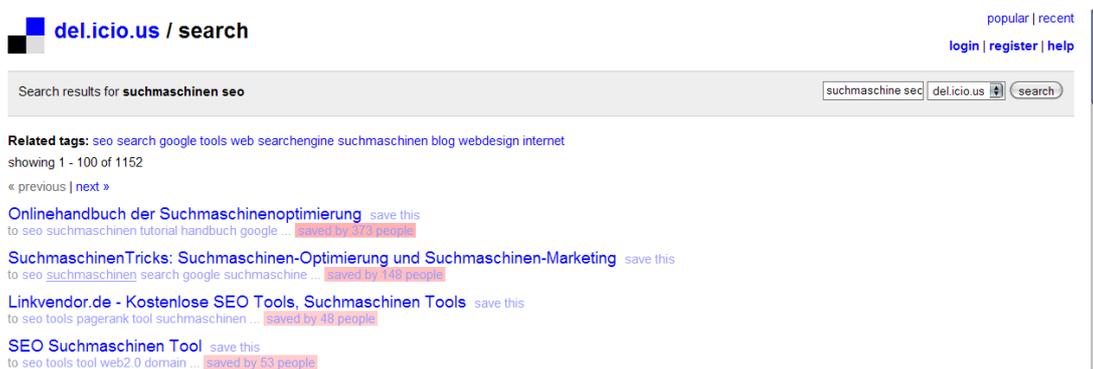


Abbildung 3: Ergebnisliste in Del.icio.us

Links, die von vielen Usern als Favoriten gespeichert werden, werden als qualitativ hochwertig betrachtet. „User Popularity“ wird damit in ähnlicher Weise zu einem suchrelevanten Qualitätsmerkmal wie die Linkpopularität bei roboterbasierten Suchmaschinen.¹¹ Social-Bookmarkdienste ermöglichen es, populäre oder aktuelle Bookmarks aus einer themenspezifischen Perspektive zu recherchieren. Da bei den

¹¹ Vgl. für eine Übersicht zu linktopologischen Rankingverfahren u.a. [Lewandowski 2005].

einzelnen Bookmarkeinträgen die Nutzer angezeigt werden, welche das jeweilige Bookmark gespeichert haben, vernetzen Social-Bookmarkdienste indirekt auch die Mitglieder der Community miteinander. Social-Bookmarkdienste stellen für den Recherchierenden also sowohl einen thematischen als auch einen personenbezogenen Zugang zu Informationsressourcen bereit. In letzterem Sinne können sie auch zum Aufspüren von „Experten“, verwendet bzw. dazu genutzt werden, Subnetzwerke mit Nutzern, die ähnliche Interessen aufweisen, zu initiieren.

Das kooperative Indexieren – „Social Tagging“, „Bookmark Popularity“ und vor allem auch die Vernetzung von Personen und Wissensobjekten – ist gegenwärtig Grundlage und Gegenstand vielfältiger Forschungsansätze im Information Retrieval und Wissensmanagementbereich, u.a.

- zur Verbesserung von Rankingalgorithmen bei roboter- und tagbasierten Websuchdiensten, u.a. [Dala 2007], [Hotho et al. 2006], [Gordon-Murnane 2006],
- als Filter- und Empfehlungssysteme, u.a. [Stoilova et al. 2005], [Niva et al. 2006],
- für das Wissensmanagement in Unternehmen, z.B. [Millen 2005].

Die Idee durch kooperatives Indexieren, nutzergenerierte Metadaten zu generieren, um Wissenserarbeitungs- und Wissensnutzungsprozesse zu befördern, ist dabei zwar nicht neu, erreicht aber mit den Social-Bookmark- und anderen tagging-basierten Communities wie Flickr.com (Fotosharing/Suche) im Web erstmals eine Reichweite und Bedeutung, die solche Konzepte über die Beförderung individuellen und gruppenbezogenen Wissensmanagements hinausheben und damit auch für die Webrecherche nutzbringend werden lassen. Um die Anwendungsmöglichkeiten zu verdeutlichen, werden einige Anfragebeispiele aufgezeigt, die mit Hilfe solcher Dienste beantwortet werden können.

- Welche neuen Webseiten wurden mit „suchmaschinen“ verschlagwortet?
- Welches sind die populärsten Webseiten mit dem Schlagwort „recherchieren“?
- Wer erklärt meine Homepage/Firmenpage zu seinen Favoriten?
- Welche anderen Nutzer haben dieselben Bookmarks wie ich gespeichert?

Diese Beispiele illustrieren, wie sich solche Dienste aus der Sicht von Recherchierenden schon derzeit sinnvoll nutzen lassen, um etwa themenspezifische Expertise aufzubauen und aufrechtzuerhalten, neue Informationsressourcen und/oder auch andere Nutzer mit ähnlichen Interessen zu finden. Des Weiteren bieten auch die Social-Bookmarkdienste RSS-Feeds an, so dass der Nutzer, wie in Blogsuchdiensten,

Alerts definieren, d.h. Suchergebnisse abonnieren bzw. diese in anderen Diensten, die RSS unterstützen, weiterverwenden kann.

Tagging-basierte-Communities stehen gerade am Beginn ihrer Entwicklung und weisen, wie oben anhand der Forschungsansätze angedeutet, vielfältiges Potenzial zur Verbesserung nicht nur, aber vor allem auch des Web Information Retrieval auf. Allerdings ist ebenso festzuhalten, dass derartige Dienste zumindest derzeit für Recherchierende auch erhebliche Defizite aufweisen, die über die angeführten tagging-immanenten Problemfelder hinausreichen. Zunächst ist hier systembedingt die geringe Abdeckung im Vergleich zur roboterbasierten Suchmaschinen zu nennen. Damit sind solche Dienste oftmals nicht für die direkte Beantwortung spezifischer Informationsbedürfnisse geeignet. Ebenso wie Kataloge erfassen sie nur einen geringen Teil des Web und bieten keinen Volltextzugriff. Sie sind also eher für thematisch breit angelegte, i.d.R. themenbezogene Informationsbedürfnisse – z.B. „Migräne“ – geeignet und nicht für spezifischere, vergleichsweise komplexe Suchanfragen – „Migräne Akuttherapie“ – nutzbar. Das Thema „Abdeckung, Vollständigkeit, Recall“ ist sicher differenziert zu betrachten und sollte näher erforscht werden. Für einen ersten Eindruck bzw. Veranschaulichung sind in nachfolgender Tabelle die Trefferzahlen zu verschiedenen Anfragen in den Social-Bookmarkdiensten Del.icio.us und Mister-wong.de sowie der Suchmaschine Google.de und des Webkatalogs Dmoz.org aufgeführt.

Suchanfrage*	Del.icio.us	Mister-wong.de	Dmoz.org	Google.de
erdbeertorte	5	3	0	133000
metasuchmaschinen	42	12	1	95100
entwicklung des internets	28	5	1	1050000
rechercheablaufplan	0	0	0	8

Tabelle 1: Trefferzahlen in Social-Bookmarkdiensten, Suchmaschinen und Webkatalogen

*Die Suchanfrage „erdbeertorte“ ist laut Google Zeitgeist die Top Gaining Query für May 2007, URL www.google.com/press/intl-zeitgeist.html#de (letzter Zugriff 19.06.2007). Die Anfragen „metasuchmaschinen“, „entwicklung des internet“, „rechercheablaufplan“ sind Suchanfragen, über die Besucher, den Webserver der Informationswissenschaft Konstanz gefunden und besucht haben, URL www.inf-wiss.uni-konstanz.de.

Der Vergleich in Tabelle 1 ist sicher problembehaftet und nicht verallgemeinerungsfähig – die ausgewählten Suchanfragen zeigen zudem, dass eine hohe Trefferzahl nicht a priori als Qualitätsmerkmal zu sehen ist („metasuchmaschinen“) – illustrieren aber das Abdeckungsproblem („rechercheablaufplan“). Er deutet zugleich an, dass sich mit Bookmarkdiensten ein höherer Recall erreichen lässt als mit Webkatalogen.

2.3 Spamproblematik in Web 2.0-Communities

Ein Problemfeld, welches die grundlegende Verwendungsfähigkeit von Web 2.0-Communities nicht nur aber vor allem auch zu Suchzwecken beeinträchtigt, stellt deren missbräuchliche Nutzung mit dem Ziel der Manipulation der Ergebnisse von Websuchdiensten dar. Dieses Spamproblem reicht weit über den engeren Kontext von Social-Bookmarkdiensten und anderen tagging-basierten Communities hinaus und betrifft generell die Nutzung der neuen sozialen Medien des Web 2.0. Mit der steigenden Verbreitung und Nutzung von Blogs, Social-Bookmarkdiensten und anderen Sozialen Medien wird deren Anwendung bzw. Einbindung in das Suchmaschinenmarketing¹² zunehmend attraktiver.¹³ Der Grund hierfür liegt darin, dass der Erwerb von Reputation und der Aufbau von Linkpopularität in derartigen Communities auch als wichtige Erfolgsfaktoren gesehen werden, um eine hohe Sichtbarkeit in den populären Suchdiensten, allen voran Google, zu erreichen.¹⁴ Hier öffnet der niedrighschwellige Zugang zu und Gebrauch von Web 2.0-Diensten einem möglichen Missbrauch Tür und Tor. Nichts belegt diesen Tatbestand besser, als dass das z.T. automatisierte, Anlegen von Links in Blog-Kommentaren und Wikipedia-Einträgen 2005 dazu geführt hat, dass die populären Websuchdienste ein „Nofollow-Link-Attribut“ kommunizierten, dessen Gebrauch bewirkt, dass derart gekennzeichnete Links bei der Sortierung nicht mehr berücksichtigt werden.¹⁵ Seit 2007 wendet auch die bekannteste Web 2.0-Seite, die Wikipedia, dieses Attribut für externe Links an.¹⁶ Damit lässt sich festhalten, dass die Spamegefahr das Problem bzw. die Frage der Glaubwürdigkeit des „user generated content“ verschärft und die Nützlichkeit von Blogsuch- und Social-Bookmarkdiensten und anderen auf Web 2.0-Inhalten aufsetzenden Suchdiensten teilweise konterkariert.¹⁷ Derzeit ist weitgehend unbekannt und unerforscht, welches Ausmaß dieses Spamproblem einnimmt. Bezogen auf den konkreten Recherchealltag des Lesers ist festzuhalten, dass – auch wenn z.B. von

¹² Unter Suchmaschinenmarketing lassen sich alle Maßnahmen subsumieren, die dazu dienen, in Suchdiensten gute Positionen und hohe Sichtbarkeit zu erreichen, vgl. [Griesbaum 2003].

¹³ Vgl. Sherman, C. (2006), What's the Big Deal With Social Search? URL searchenginewatch.com/showPage.html?page=3623153 (letzter Zugriff 19.06.2007).

¹⁴ Vgl. Newcomb, K. (2007), Why Are Search Marketers Getting Social? URL searchenginewatch.com/showPage.html?page=3625530 (letzter Zugriff 19.06.2007).

¹⁵ „From now on, when Google sees the attribute (rel="nofollow") on hyperlinks, those links won't get any credit when we rank websites in our search results. This isn't a negative vote for the site where the comment was posted; it's just a way to make sure that spammers get no benefit from abusing public areas like blog comments, trackbacks, and referrer lists.“ URL googleblog.blogspot.com/2005/01/preventing-comment-spam.html (letzter Zugriff 19.06.2007).

¹⁶ URL

en.wikipedia.org/wiki/Wikipedia_talk:WikiProject_Spam/2007_Archive_Jan#External_links_in_articles_are_now_.22nofollow.22_per_Jimbo_Wales (letzter Zugriff 19.06.2007).

¹⁷ Während Blogsuchdienste, die Spamproblematik mit den Suchmaschinen teilen, wird in einer Gegenüberstellung von editorial betreuten Katalogen und Social-Bookmarkdiensten deutlich, dass erstere aufgrund der immanenten redaktionellen Qualitätssicherung, tendenziell qualitativ hochwertigere Indizes/Einträge/Inhalte aufweisen.

Social-Bookmarkdiensten Qualitätssicherungsverfahren zur Spamkontrolle genutzt werden¹⁸ – die Ergebnisse/Inhalte dieser Dienste kritisch interpretiert bzw. hinterfragt werden sollten.

2.4 Zwischenergebnis

In einem ersten Ergebnis lässt sich festhalten: Blogsuch- und tagging-basierte Communities, hier vertiefend behandelt anhand von Social-Bookmarkdiensten, bergen hohes Innovationspotenzial für die Websuche. Blogsuchdienste bieten vergleichsweise hochgradig strukturierbaren Zugriff auf aktuelle Inhalte des Web. Social-Bookmarkdienste ermöglichen, ähnlich Webkatalogen, Zugriff auf von Menschen indexierte Ressourcen und eröffnen durch die indirekte Vernetzung der Community-Teilnehmer völlig neuartige personale Suchoptionen, die sich konzeptuell als Expertensuche kennzeichnen lassen. Aus der Sicht von Recherchierenden schaffen diese Suchwerkzeuge damit völlig neue Möglichkeiten, auf die Ressourcen des Internet zuzugreifen. Allerdings zeigt die dargestellte Spamproblematik auf, dass bei den in diesen Communities bereitgestellten Ressourcen – im Unterschied zu großen renommierten Webkatalogen – nicht a priori von einer hohen Qualität ausgegangen werden kann. Zwar zeigen verschiedene Untersuchungen, etwa [Giles 2005], dass „user generated content“ qualitativ nicht schlechter sein muss als Information, die in kontrollierten professionellen, bevorzugt wissenschaftlich-technischen und wirtschaftlichen, Bereichen produziert wird. Dennoch gilt: die Offenheit von Web 2.0-Communities sorgt für Anreize und Möglichkeiten, gezielt Spamtechniken zu manipulativen Zwecken anzuwenden. Aus diesem Grund sollten Recherchierende Suchergebnisse dieser Dienste ebenso kritisch betrachten wie Suchergebnisse roboterbasierter Suchdienste.

Aus technischer Perspektive eröffnet der Einsatz der RSS-Technologie Möglichkeiten zur direkten Weiterverwendung von Suchergebnissen bzw. gestattet es, jede Suchanfrage auch als Interessenprofil zu definieren, d.h. Ergebnislisten zu abonnieren. Dadurch sind Alert-Dienste wesentlich einfacher und vielfältiger zu spezifizieren als etwa über Google-Alerts.¹⁹ Damit wird traditionelles „Ad-Hoc-Retrieval“ mit „Routing-Task-Retrieval“ [Ferber 2003], S.94, bei dem zu bestehenden „alten“ Anfragen im Zeitablauf neue relevante Dokumente herausgefiltert werden, miteinander verschmolzen. Aus technischer Perspektive sind damit diese Dienste ideal dazu geeignet, langfristige Informationsbedürfnisse, wie etwa einen themenspe-

¹⁸ So prüft beispielsweise Mister-Wong.de importierte Bookmarks auf Spam. Vgl. auch Mister-Wong Anti-Spam Regeln URL www.mister-wong.de/blog/mister-wong-anti-spam-regeln/ (letzter Zugriff 20.06.2007).

¹⁹ www.google.com/alerts.

zifischen Expertiseaufbau oder auch die strategische Markt- und Wettbewerbsbeobachtung („Competitive Intelligence“), zu unterstützen. Aus der Sicht von Recherchierenden wäre es wünschenswert, wenn auch die populären Suchdienste Google, Yahoo und MSN derartige Technologien zur Überwachung und Weiterverwendung von Suchergebnissen anbieten würden.

3. Entwicklungstrends bei Google, Yahoo und MSN

Entwicklungstrends bei den populären Websuchdiensten lassen sich derzeit grob in drei Bereiche gliedern, vgl. hierzu auch [Horowitz 2007], [Broder 2006], [Griesbaum & Bekavac 2004], [Lewandowski 2004].

1. Erschließung neuer Wissensbestände.
2. Präsentation von Suchergebnissen.
3. Personalisierung & Einsatz von Web 2.0-Konzepten & Technologien.

Der Schwerpunkt der folgenden Ausführungen bezieht sich auf den dritten Bereich, ohne dass damit Weiterentwicklungen in den anderen Bereichen ignoriert werden sollen. So ist etwa die in 2.1 dargestellte Google-Blogsuche ein Beispiel für den Aufbau spezieller Indizes. Weitere Beispiele, die hier nicht behandelt werden, aber für Recherchierende relevant sind, sind etwa die Google-Buchsuche²⁰ oder auch „Yahoo Subscriptions“.²¹ Im zweiten Bereich ist derzeit vor allem die Google „Universal Search“²² anzuführen, die ein erhebliches Echo, vor allem in der amerikanischen Suchmaschinenmarketing-Industrie, hervorruft.²³ Es handelt sich dabei um ein Redesign der Suchergebnisseiten von Google. Neben Webseiten werden nun vermehrt Bilder, Videos, News und weitere Ergebnisse aus der Vielzahl der vertikalen Google-Suchdienste eingebunden. Des Weiteren wurde die Navigation modifiziert, d.h. eine „Kontextnavigation“ über den Ergebnissen eingefügt, eine neue „universal navigation bar“ im Kopf der Ergebnisseite bereitgestellt und unterhalb der Suchtreffer ein Bereich „Searches related to:“ ergänzt, in welchem „ähnliche“ Suchanfragen zur Anfragenverfeinerung aufgeführt werden. Zur Veranschaulichung ist nachfolgend die

²⁰ books.google.de.

²¹ search.yahoo.com/subscriptions?ei=UTF-8.

²² googleblog.blogspot.com/2007/05/universal-search-best-answer-is-still.html (letzter Zugriff 20.06.2007)

²³ Vgl. u.a. Newcomb, K. (2007), Google Goes Universal, Adds Navigation, URL blog.searchenginewatch.com/blog/070516-164804 (letzter Zugriff 20.06.2007), Stelter, P. (2007), Distilling Universal Search, URL searchenginewatch.com/showPage.html?page=3626082 (letzter Zugriff 20.06.2007). Obwohl erst im Mai 2007 „freigeschaltet“, lieferte die Google Suchanfrage „Optimizing Content for Universal Search“ am 20.06.2007 bereits über eine Million Treffer.

Ergebnisseite zu der im offiziellen Google Blog vorgeschlagenen Suchanfrage „darth vader“ dargestellt.

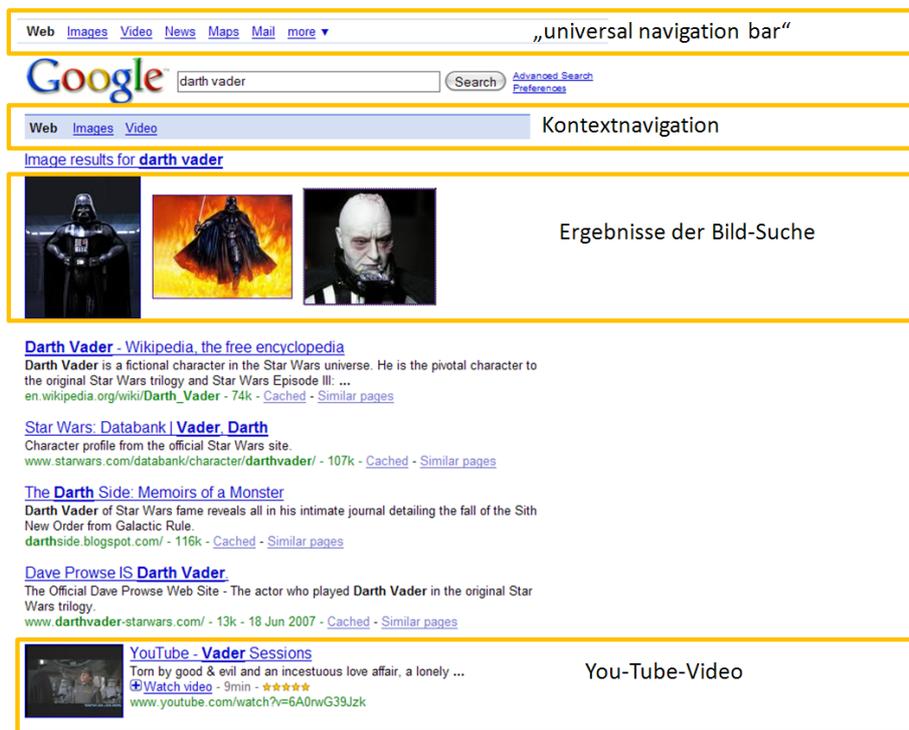


Abbildung 4: Google Universal Search

Damit folgt Google der Idee, in Abhängigkeit von der Art des Informationsbedürfnisses passende Dokumenttypen als Ergebnis auszugeben und dem Nutzer die Auswahl bzw. Anwahl vertikaler Suchdienste soweit als möglich abzunehmen bzw. zu ersparen.²⁴ Aus der Sicht der Recherchierenden ist diese Entwicklung zu begrüßen. Inwieweit diese Veränderung tatsächlich mit einer Verbesserung der Suchergebnisse verbunden ist, ist davon abhängig, in welchem Maße es gelingt, tatsächlich zu den jeweiligen Informationsbedürfnissen passende Dokumenttypen auszugeben und sinnvoll zu sortieren. In dieser Hinsicht ist „Universal Search“ ein wichtiger, aber sicher nicht der letzte Schritt, unterschiedliche Datenquellen und Ergebnisdokumenttypen in Suchergebnisseiten zu integrieren.²⁵ Ein Schritt den, wie nachfolgende Abbildung zeigt, andere Suchmaschinen, wenn auch in abgewandelter Form, ebenfalls schon durchgeführt haben.²⁶

²⁴ Sullivan, D. (2003), Searching With Invisible Tabs, URL searchengine-watch.com/searchday/article.php/3115131 (letzter Zugriff 20.06.2007).

²⁵ Vgl. auch die Aussage von Mayer, M. "While today's releases are big steps in making the world's information more easily accessible, these are just the beginning steps toward the universal search vision. Stay tuned!" URL googleblog.blogspot.com/2007/05/universal-search-best-answer-is-still.html (letzter Zugriff 20.06.2007).

²⁶ Vgl. auch Hewcom, K. (2007), Ask.com Launches Major Updates, URL searchengine-watch.com/showPage.html?page=3626058 (letzter Zugriff 20.06.2007).

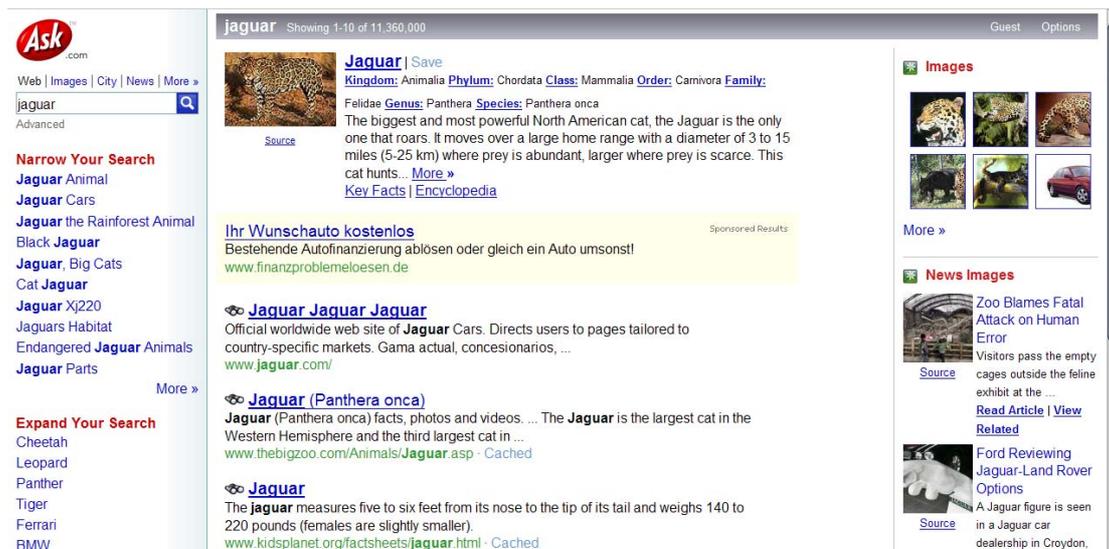


Abbildung 5: "Universal Search" bei Ask.com²⁷

3.1 Personalisierung und Web 2.0-Technologien: „Social Search“

Die folgende Darstellung stellt die Entwicklung von Personalisierungsoptionen und den Einsatz von Web 2.0-Konzepten und -Technologien bei den Suchdiensten MSN, Yahoo und Google in den Mittelpunkt. Beide Bereiche sind nicht klar zu trennen. Personalisierung zielt darauf, individuelle Vorlieben und Bedürfnisse zu erkennen und bei der Ausgabe von Suchergebnissen zu berücksichtigen. So lassen sich etwa Interaktions- und Profildaten dazu verwenden, Informationsbedürfnisse besser nachvollziehen zu können [Griesbaum & Bekavac 2004]. „Social Search“ bedeutet, dass die Interaktionen und Beiträge einer Vielzahl von Nutzern dazu verwandt werden, um die Zusammensetzung und Reihenfolge von Suchergebnissen zu verbessern. In diesem Sinne lässt sich „Social Search“ auch als kollektive Aggregation von Personalisierung verstehen. Auch wenn im Einzelfall die Erfassung umfangreicher individueller Suchprofile und das Tagging in Web 2.0-Communities keinen direkten Bezug zueinander aufweisen wird doch in beiden Fällen Wissen erzeugt, das von Suchdiensten dazu genutzt werden kann, Suchergebnisse zu verbessern. Und genau dieser Aspekt wird als zentraler Schlüssel zur Verbesserung des Web Information Retrieval betrachtet.²⁸ In der Realität existiert eine Vielzahl unterschiedlicher Formen von „Social Search“ – die in 2.2. behandelten Social-Bookmarkdienste stellen dabei eher eine simple Ausprägung dar. Letztlich werden in

²⁷ URL www.ask.com/web?q=jaguar&x=0&y=0&qsrc=2406&o=0&l=dir (letzter Zugriff 20.06.2007).

²⁸ Mills, E. (2006), Internet search gets Web 2.0 style URL news.com.com/2100-1038_3-6108962.html?part=rss&tag=6108962&subj=news (letzter Zugriff 21.06.2007), Hines, M. (2005), Future of search rides on relevance, URL news.com.com/2100-1032_3-5555954.html (letzter Zugriff 21.06.2007).

der Kombination, Zusammenführung bzw. Integration von „Social Search“ mit algorithmenbasierten Suchverfahren, wie sie derzeit in den primär roboterbasierten Suchdiensten angeboten werden, enorme Potenziale gesehen, die Websuche qualitativ auf ein neues Niveau zu heben.²⁹

Entscheidend für den Erfolg derartiger Ansätze ist zum einen das Ausmaß, in dem es den Suchdiensten gelingt, das inhärente Wissen und die Eigenschaften der Nutzer für die Suche explizit zu machen,³⁰ zum anderen die Qualität der erfassten Daten.³¹ Dabei ist es aus der dargestellten theoretischen Perspektive zunächst zweitrangig, ob es sich um die implizite Erfassung von Interaktionsdaten oder z.B. um die explizite Vergabe von Labels/Tags handelt. Im Zweifel gilt die Aussage, über je mehr Daten ein Suchdienst verfügt, desto besser ist er in der Lage, Informationsbedürfnisse nachzuvollziehen und die Bestände der vorhandenen Indizes inhaltlich und qualitativ zu bewerten. Im Folgenden werden in einer Übersicht die wesentlichen Ansätze dargestellt, welche die drei großen globalen Player in diesem Bereich verfolgen.³²

3.2 MSN: Suchmakros

MSN bietet die Möglichkeit, sogenannte Suchmakros zu erstellen. Suchmakros grenzen die Suchergebnisse aus dem MSN-Gesamtindex ein. D.h. aufsetzend auf der MSN-Websuche können Nutzer eigene Suchmaschinen als Untermenge der MSN-Suchmaschine definieren. Im Standardmodus lässt sich eine Liste von Websites festlegen, die das Makro durchsuchen soll. Ein Beispiel hierfür ist etwa ein Suchmakro zu einer Band, das die offizielle Webseite der Band und populäre Fanseiten und Foren durchsucht.³³ Im erweiterten Modus können alle Anfrageparameter der MSN-Suche zur Definition eines Suchmakros verwendet werden. Folgende Abbildung zeigt ein solches erweitertes Makro im Vergleich zur MSN-Standardsuche. Das Makro schränkt die Ergebnismenge auf PDF-Dokumente ein, welche die Phrasen „information retrieval“ und „web 2.0“ enthalten und in deutscher Sprache verfasst sind. Dieses Makro könnte man als „Artikelsuchmaschine für Web 2.0-Retrieval“

²⁹ U.a. Sherman, C. (2006), What's the Big Deal With Social Search? URL searchenginewatch.com/showPage.html?page=3623153 (letzter Zugriff 21.06.2007).

³⁰ Hines, M. (2005), Future of search rides on relevance URL news.com.com/2100-1032_3-5555954.html (letzter Zugriff 21.06.2007).

³¹ Sherman, C. (2005), Yahoo Integrates Personal & Social Search with MyWeb 2.0 URL searchenginewatch.com/showPage.html?page=3516381 (letzter Zugriff 21.06.2007).

³² Aus Platzgründen können kleinere, d.h. weniger bekannte innovative, meist als Web 2.0 Suchmaschinen bezeichnete Dienste – etwa Sproose.com – hier nicht behandelt werden. Der interessierte Leser sei hierzu verwiesen an web-information-retrieval.de/component?option=com_bookmarks/Itemid,61/mode,0/catid,18/navstart,0/search,*/ (letzter Zugriff 21.06.2007).

³³ Vgl. das Suchmakro „tokio_hotel“ unter gallery.live.com/liveItemDetail.aspx?li=658e14ea-7d40-4112-86f3-244f5a285b13&bt=13&pl=4 (letzter Zugriff 21.06.2007).

bezeichnen.³⁴ Die Anfrage „marketing“ liefert bei dieser Artikelsuchmaschine 13 Suchtreffer gegenüber über 160 Millionen Ergebnissen, welche die MSN-Standardsuche für dieselbe Anfrage zurückgibt.

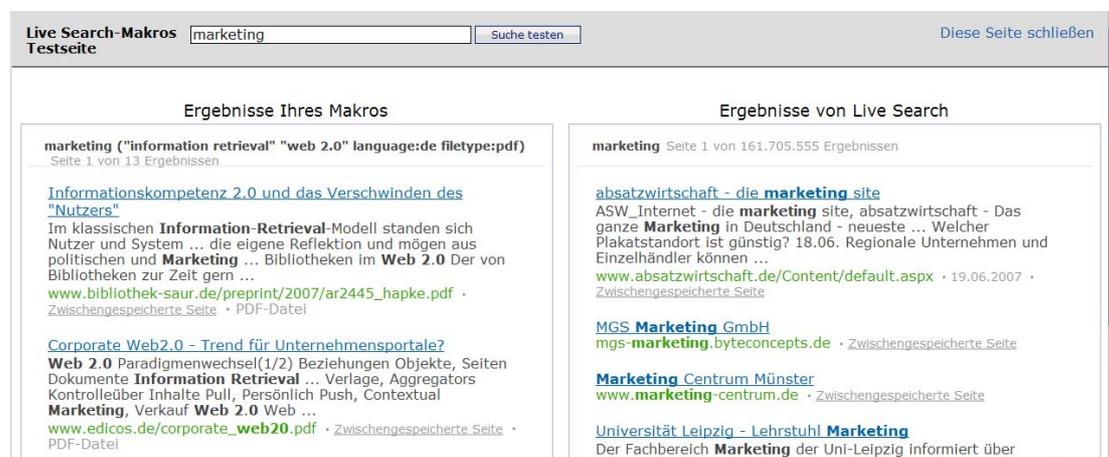


Abbildung 6: MSN-Suchmakros

Suchmakros können mit einem Titel, einer Beschreibung und mit Tags versehen und nach einer redaktionellen Prüfung durch Mitarbeiter von MSN für die Community freigegeben werden. Nach der Freischaltung können andere Community-Mitglieder weitere Tags hinzufügen bzw. eine Beurteilung des Suchmakros verfassen. Gegenwärtig sind auf MSN.de insgesamt 64 „öffentliche“ Suchmakros vorhanden.³⁵

3.3 Yahoo: Clever, MeinWeb, “Search Builder”, Pipes

Yahoo weist in mehrfacher Hinsicht eine starke Community-Ausrichtung auf, denn neben den aufgekauften und derzeit weiterhin weitgehend getrennt vom Yahoo-Portal betriebenen Web 2.0-Diensten, Del.icio.us und Flickr integriert Yahoo mit Yahoo Clever³⁶ und Yahoo MeinWeb³⁷ zwei weitere community-basierte Such- bzw. Antwortdienste direkt in das Yahoo-Portal. Des Weiteren betreibt Yahoo mit dem Yahoo „Search Builder“³⁸ und Pipes³⁹ zwei Dienste, mit denen Nutzer selber Suchdienste zusammenstellen und spezifizieren können.

Yahoo Clever – international als Yahoo Answers bezeichnet – ist ein Frage- und Antwort-Dienst, in dem Community-Mitglieder Fragen stellen können und die

³⁴ Es findet sich unter search.live.com/macros/artikelsuche_web20_ir/artikelsuche_web20_ir/?FORM=MACWLG (letzter Zugriff 26.06.2007).

³⁵ Diese können auch ohne Registrierung und genutzt werden. URL gallery.live.com/results.aspx?c=0&bt=13&pl=4 (letzter Zugriff 21.06.2007).

³⁶ URL.de.answers.yahoo.com/ (letzter Zugriff 22.06.2007).

³⁷ URL.de.myweb2.search.yahoo.com/ (letzter Zugriff 22.06.2007).

³⁸ URL.builder.search.yahoo.com/ (letzter Zugriff 22.06.2007).

³⁹ URL.pipes.yahoo.com/pipes/ (letzter Zugriff 22.06.2007).

gestellten Fragen wiederum durch die Mitglieder der Community beantwortet werden. Gestellte Fragen und vorhandene Antworten werden gespeichert, sind öffentlich zugänglich und werden als Ergebnisse in die Yahoo-Standarwebsuche mit eingebunden. Als Anreizsystem wird ein Punktesystem verwendet, welches Aktivitäten belohnt. Die Qualität der Antworten wird von der Community durch Abstimmungen bewertet und positive Bewertungen dem Punktesystem des Antwortgebers zugeschrieben. Mit steigender Punktezahl erreichen Mitglieder schrittweise weitere Nutzungsprivilegien, wie beispielsweise das Recht, Antworten anderer Community-Mitglieder bewerten zu können. Auf die Frage „Wie viele Mitglieder hat die Yahoo! Clever Community in Zahlen?“ liefert die durch Abstimmung ausgewählte Antwort die Zahl 3702. Ebenso wie in den anderen hier dargestellten Communities werden nicht nur Fragen und Antworten, sondern die dahinterstehenden Personen angezeigt. Yahoo Clever unterstützt auf dieser Basis die Netzwerkbildung.

Bei MeinWeb handelt es sich im Kern zunächst um einen Social-Bookmarkdienst ähnlich Del.icio.us (vgl. 2.2). In Kombination von MeinWeb mit der Yahoo-Websuche verfügt der Nutzer über Suchfunktionalitäten, die über die anderer Social-Bookmarkdienste hinausreichen. Hat man sich in Yahoo als MeinWeb-Nutzer eingewählt und startet die Yahoo-Websuche, so werden sowohl Yahoo Web-Ergebnisse als auch, soweit vorhanden, Treffer in den persönlichen Bookmarks in MeinWeb angezeigt. Dabei wird, sofern die Bookmarks als Volltextkopie gespeichert wurden, auch der Volltext der Sicherungskopien durchsucht – vgl. die folgende Abbildung.

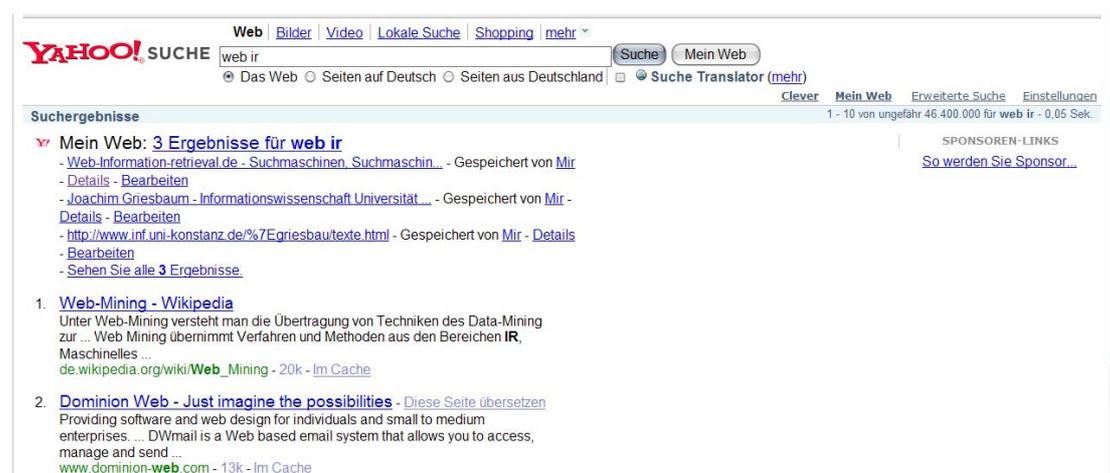


Abbildung 7: MeinWeb-Integration in Yahoo-Websuche

Diese hybride Suche ist derzeit noch sehr rudimentär. Es werden weder vergebene Bookmark-Tags noch die Bookmarks anderer Nutzer bei der Ausgabe von MeinWeb-Ergebnissen in der Yahoo-Websuche berücksichtigt bzw. in den Suchanfragenabgleich miteinbezogen. Des Weiteren steht eine Suchhistorienfunktion „Suchverlauf“ bereit, welche die über die Yahoo-Websuche gesichteten Seiten protokollieren soll.⁴⁰

Der „Search Builder“ ermöglicht es Website-Betreibern eine „eigene“, themenorientierte Suchmaschine auf der Basis der Yahoo-Standardsuchmaschine zu definieren. „Search Builder“ ähnelt bzgl. der Funktionalität den MSN-Suchmakros. Allerdings kann neben der Eingabe von ein- oder auszuschließenden URLs nur eine Termliste ein- oder auszuschließender Suchterme angegeben werden. Eine Definition weiterer Anfrageparameter ist nicht möglich. Die spezifizierte Suchmaschine kann nur auf einer Website des Betreibers eingebunden werden. Eine direkte Weitergabe an die Yahoo-Community ist nicht vorgesehen. Letzteres wird durch einen Drittanbieter wie Rollyo.com ermöglicht. Bei diesem Dienst können die Nutzer in funktional gleicher Weise wie bei Yahoo „Search Builder“ sogenannte Searchrolls spezifizieren, diese mit Tags versehen und den anderen Nutzern zur Verfügung stellen, welche wiederum vorhandene Searchrolls modifizieren und weiterentwickeln können.⁴¹

Mit Pipes⁴² stellt Yahoo einen Dienst zur Verfügung, der es gestattet Daten aus unterschiedlichen Quellen zu aggregiert und miteinander zu kombinieren. Mit Pipes können XML-Datenquellen (RSS) ausgelesen, zusammengefügt, gefiltert, neu sortiert und damit neue (Such- und Alert-)Dienste kreiert werden.⁴³ Unter pipes.yahoo.com/pipes/search?q=informationswissenschaft findet sich eine vom Verfasser in 30 Minuten zusammengestellte Suchmaschine, welche die Webserver verschiedener Lehrstühle der Informationswissenschaft nach PDF-Dateien durchsucht. Folgende Abbildung zeigt den „Quellcode“ dieser durch Yahoo Pipes generierten Spezialsuchmaschine.

⁴⁰ Diese Funktion arbeitete aber am Datum der Untersuchung – 22.06.2007 – nicht.

⁴¹ Vgl. für eine Übersicht vorhandener Searchrolls, URL rollyo.com/explore.html (letzter Zugriff 22.06.2007).

⁴² Vgl. auch www.techmeme.com/070207/p105#a070207p105 (letzter Zugriff 22.06.2007).

⁴³ Vgl. u.a. O'Reilly, T. (2007), Pipes and Filters for the Internet, URL radar.oreilly.com/archives/2007/02/pipes_and_filte.html (letzter Zugriff 22.06.2007).

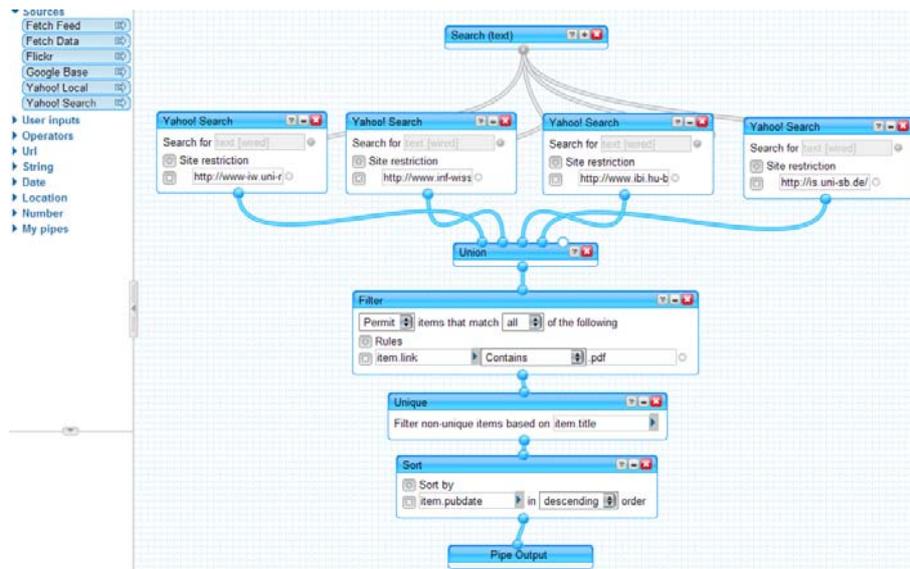


Abbildung 8: Drag and drop „Sourcecode“ bei Yahoo Pipes

Obwohl die Programmierung von Pipes aufgrund der AJAX⁴⁴-basierten Benutzeroberfläche per drag and drop, d.h. verhältnismäßig unkompliziert und intuitiv vonstatten geht, ist gerade die Erstellung komplexerer Pipes doch aufwändig bzw. zeitintensiv und von typischen Suchmaschinennutzer nicht ohne weiteres durchführbar. Insofern sind die Pipes zwar sicherlich ein Meilenstein des Web,⁴⁵ ihr Nutzen für Recherchierende ist aber eher in der Möglichkeit der Bereitstellung spezieller Suchtools durch Dritte zu sehen. Gegenwärtig sind bei Yahoo mehrere tausend Pipes zugreifbar, es handelt sich aber in den meisten Fällen eher um spielerische Anwendungsexperimente, als um ernsthafte Applikation – vgl. auch obenstehendes Beispiel.

3.4 Google: Suchprotokoll, Co-op

Google stellt mit Co-op und dem Suchprotokoll sowohl personalisierte Suchoptionen als auch „Social Search“-Dienste bereit, die direkt auf der Standardwebsuche aufsetzen. Das Suchprotokoll⁴⁶ bietet eine Vielzahl neuer Personalisierungsfunktionen. Diese beruhen auf der Aufzeichnung des Such- und Surfverhaltens von Google-Nutzern. Voraussetzung ist, dass diese über ein Google-Konto verfügen und über das Web bzw. die Google-Toolbar in Google angemeldet sind. Google speichert– neben dem durch die Toolbar erfassten Surfverhalten – sowohl durchgeführte Suchanfragen

⁴⁴ AJAX ist ein Akronym für Asynchronous JavaScript and XML. Dieses Konzept einer asynchronen Datenübertragung zwischen Server Browser bzw. die hierzu verwendeten Technologien beschleunigen die Interaktion des Nutzers mit der jeweiligen Webapplikation, da jeweils nur Teile der aktuellen HTML-Seite – sukzessiv nach Bedarf vom Server nachgeladen werden müssen.

⁴⁵ O'Reilly, T. (2007), Pipes and Filters for the Internet, URL radar.oreilly.com/archives/2007/02/pipes_and_filtre.html (letzter Zugriff 22.06.2007).

⁴⁶ www.google.de/searchhistory/ (letzter Zugriff 25.06.2007).

als auch angeklickte Suchergebnisse. Folgende Abbildung zeigt das Suchprotokoll-Webinterface.



Abbildung 9: Google Suchprotokoll

Damit ist es möglich, vorherige Suchvorgänge retrospektiv nachzuvollziehen und die gespeicherten Inhalte im Volltext zu durchsuchen.⁴⁷ Des Weiteren bilden die gespeicherten Nutzungsdaten die Grundlage für neue Empfehlungsdienste. Auf der Basis des Suchverhaltens weiterer Nutzer, die vergleichbare oder gleiche Suchanfragen stellen und ein ähnliches Surfmuster aufweisen, werden nach dem Prinzip des kollaborativen Filterns⁴⁸ Interaktionsdaten aggregiert und daraus Empfehlungen abgeleitet.⁴⁹ Diese Empfehlungen können vom Nutzer über ein Tab „recommendations“ in iGoogle oder direkt über die Toolbar abgefragt werden. Folgende Abbildung zeigt die Empfehlungen, die dem Verfasser am 25.06.2007 zugewiesen wurden.

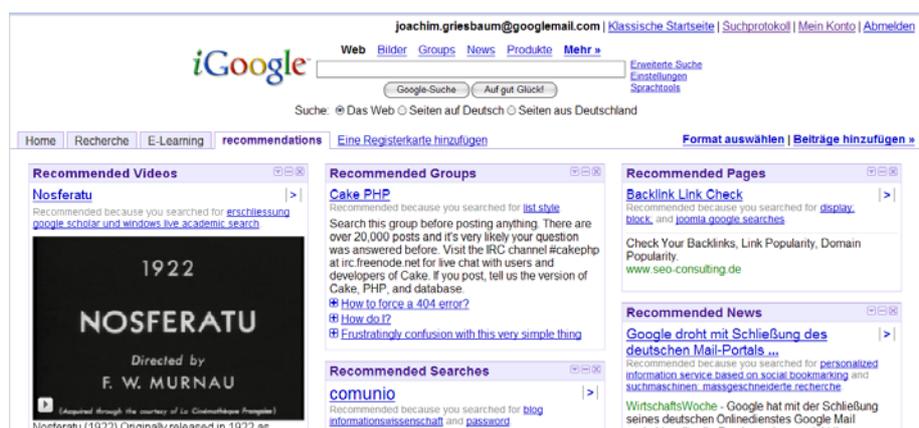


Abbildung 10: Google Empfehlungen

⁴⁷ Für eine detaillierte Darstellung und Diskussion des Google Suchprotokoll vgl. Sullivan, D. (2007), Google Search History expands, becomes web History, URL searchengineland.com/070419-181618.php (letzter Zugriff 25.06.2007).

⁴⁸ Vgl. Zum kollaborativen Filtern [Chakrabarti 2003], S. 115 f..

⁴⁹ Sherman, C. (2007), Google offers „Queryless Search“ & personalized recommendations, URL searchengineland.com/070418-153238.php (letzter Zugriff 25.06.2007).

Google ermöglicht mit dem Suchprotokoll dem Recherchierenden völlig neuartige Suchfunktionen: Zum einen nachlaufenden Zugriff auf bisher durchgeführte Suchprozesse und gesichtete Ergebnisse und zum anderen browsing-orientierte Unterstützung für das Aufspüren weiterer nutzerrelevanter Informationsressourcen. Die Kehrseite bzw. der Preis hierfür ist die Bereitschaft, Google tiefe Einblicke in das persönliche Such- und Surfverhalten zu gewähren. Ob und inwieweit dies aus datenschutzrechtlichen Aspekten problematisch ist und etwa auch zur Strafverfolgung genutzt werden kann, wird derzeit heftig diskutiert.⁵⁰ Auf der einen Seite ist es sicher richtig, Google, als gewinnorientierten marktbeherrschenden Internet-Werbekonzern dies bzgl. besonders kritisch zu beobachten und datenschutzrechtlich bedenkliche Aspekte offenzulegen und zu kritisieren.⁵¹ Andererseits ist die Nutzung der personalisierten Google-Dienste völlig freiwillig. Die Anmeldung und Aktivierung der genannten Dienste erfordert die explizite Zustimmung der Nutzer. D.h. auch wenn die Vermutung zutrifft, dass vermutlich den meisten Google-Nutzern nicht klar ist, dass ihre Interaktionsdaten langfristig aufbewahrt und auch systematisch verwendet werden, kann derzeit nicht davon gesprochen werden, dass Google die Nutzer ohne Warnhinweis zur Preisgabe persönlicher Daten verleitet.

Lässt sich das Google Suchprotokoll als Personalisierung durch implizite Interaktionsdatenaufzeichnung kennzeichnen, ist Google Co-op⁵² Ausdruck dafür, wie Google seit 2006 „Social Search“-Konzepte aufgreift. Google Co-op gliedert sich derzeit in drei Bereiche:

- „Custom Search Engines“ – nutzerspezifizierte Suchdienste,
- „Subscribed Links“ – Auslieferung spezifischer Inhalte zu festgelegten Suchanfragen,
- Topics – Verschlagwortung von Webseiten.

Mit Google „Custom Search Engine“-s (CSEs) bietet Google – vergleichbar den Suchmakros bei MSN bzw. Searchrolls – Möglichkeiten, aufsetzend auf der Google-

⁵⁰ Vgl. u.a. Sullivan, D. (2007), Google Search History expands, becomes web History, URL searchengineland.com/070419-181618.php (letzter Zugriff 25.06.2007), auch Privacy International (2007), A Race to the Bottom - Privacy Ranking of Internet Service Companies, Zitat: „Google's increasing ability to deep-drill into the minutiae of a user's life and lifestyle choices must in our view be coupled with well defined and mature user controls and an equally mature privacy outlook. Neither of these elements has been demonstrated. Rather, we have witnessed an attitude to privacy within Google that at its most blatant is hostile, and at its most benign is ambivalent.“ URL [www.privacyinternational.org/article.shtml?cmd\[347\]=x-347-553961](http://www.privacyinternational.org/article.shtml?cmd[347]=x-347-553961) (letzter Zugriff 25.06.2007).

⁵¹ Vgl. die Kritik der Arbeitsgruppe Artikel 29 der Europäischen Union, URL ec.europa.eu/justice_home/fsj/privacy/news/docs/pr_google_annex_16_05_07_en.pdf (letzter Zugriff 25.06.2007).

⁵² www.google.de/coop/ (letzter Zugriff 25.06.2007), vgl. für eine Übersichtsdarstellung auch www.stonetemple.com/articles/google-coop-overview.shtml (letzter Zugriff 25.06.2007).

Standardsuche, themen- und/oder URL-spezifische Suchdienste zu definieren. CSEs gestatten es festzulegen, welche Webseiten in die jeweilige CSE aufgenommen respektive ausgeschlossen werden sollen bzw. welchen Seiten ein höheres oder niedrigeres Gewicht bei der Sortierung zugewiesen werden soll. Ergänzend können Terme zur thematischen Ausrichtung der jeweiligen CSE definiert werden. Ergebnisse, welche die genannten Terme enthalten, werden bei der Ergebnisausgabe bevorzugt. Folgende Abbildung zeigt eine wiki-zentrierte CSE und deren Ergebnisseite für die Suchanfrage „social search“.

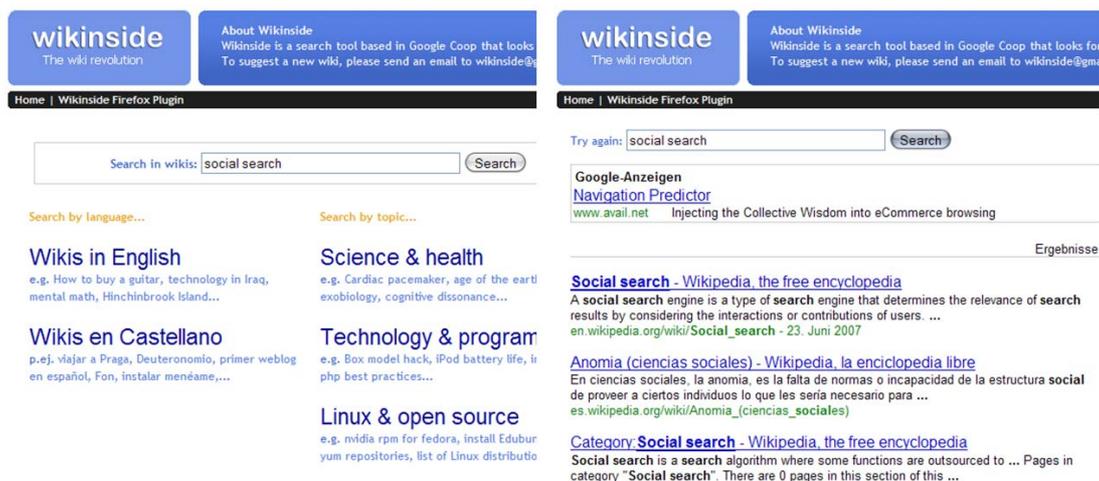


Abbildung 11: Beispiel für eine Google Custom Search Engine (CSE)

Der Autor einer CSE kann über eine Kollaborationsfunktion festlegen, ob andere Nutzer an der CSE mitarbeiten können. Eine CSE kann dies bzgl. für alle Nutzer oder für einen spezifizierten Nutzerkreis freigeschaltet werden. Google stellt für die CSEs eine eigene Homepage zur Verfügung. Zusätzlich können CSEs auch auf anderen Webseiten eingebunden werden. Wird auf den Ergebnisseiten Werbung eingeblendet, so wird der Autor einer CSE an den Einnahmen beteiligt. Nach Angaben eines Google Mitarbeiters existieren derzeit schon über 100000 CSEs.⁵³ Unter www.customsearchguide.com findet sich ein Verzeichnis, in dem CSEs thematisch katalogisiert werden.

Mit „Subscribed Links“ können Informationsanbieter zu Suchanfragen eigene Inhalte in Google einbinden. „Subscribed Links“ ist damit ein Dienst, der es ermöglicht, in und durch Google die Verbreitung eigener Inhalte zu befördern bzw. spezialisierte Suchergebnisse anzubieten. Damit diese Inhalte für die Google Nutzer sichtbar werden, müssen sie die „Subscribed Links“ der jeweiligen Anbieter abonnieren. Diese

⁵³ Vgl. Stonetemple Consulting “Interview of Google's Rajat Mukherjee”, URL www.stonetemple.com/articles/interview-rajat-mukherjee.shtml (letzter Zugriff 02.07.2007).

können hierzu einen grafischen Hinweis auf Ihrer Website einbinden. Melden sich genügend Abonnenten an, so wird der Anbieter in einem „Subscribed Links“-Verzeichnis bei Google aufgeführt – derzeit weist dieses Verzeichnis 67 Einträge auf.⁵⁴ Folgende Abbildung zeigt ein Anwendungsbeispiel zur Suchanfrage „Bill Clinton“ für einen Nutzer, der die „Subscribed Links“ der Encyclopaedia Britannica abonniert hat. In diesem Fall wird vor den Google-Ergebnissen der für diese Anfrage spezifizierte Eintrag von Britannica.com ausgeliefert.



Abbildung 12: Google Subscribed Links

Google Topics ermöglicht es, Webseiten zu verschlagworten. Topics kann als kontrolliertes Tagging verstanden werden, da Google die Labels, die genutzt werden können, festgelegt hat.⁵⁵ Folgende Abbildung zeigt die Labels der Facette (Unterkategorie) „Health conditions“ in der Kategorie „Health“.

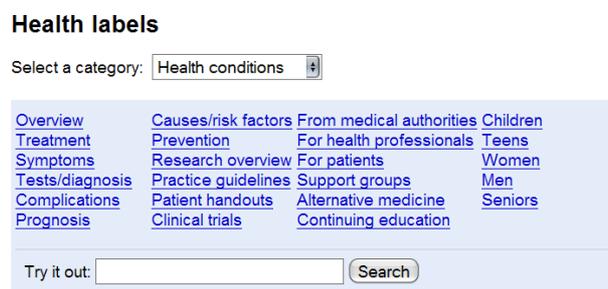


Abbildung 13: Google Topics

⁵⁴ www.google.com/coop/subscribedlinks/directory/Latest (letzter Zugriff 25.06.2007).

⁵⁵ Zumindest derzeit – Stand 25.06.2007 – ist die freie Vergabe von Labels nicht möglich. Die Labels gliedern sich in sechs vorgegebenen Hauptkategorien, die teilweise Unterkategorien beinhalten, auf. Eine Übersicht über die in Google vorhandenen Labels findet sich unter www.google.com/coop/docs/cse/labels.html (letzter Zugriff 26.06.2007).

Zusammenfassend lässt sich festhalten, dass Google mit Co-op eine Vielzahl von Diensten bereitstellt, in denen Nutzer ihr Wissen einbringen und Webseiten auf unterschiedliche Arten inhaltlich und qualitativ kennzeichnen können. Insbesondere durch die Option, über „Subscribed Links“ eigene Inhalte auszuliefern bzw. durch die Möglichkeit, mit CSEs auch direkt Einnahmen erzielen zu können, bietet Google, im Unterschied MSN und Yahoo, auch extrinsische Anreize, Wissen bereit zu stellen. [Weitkämper 2007] bezeichnet Co-op als genialen Schachzug Googles, um das Internet zu kategorisieren.

Aus der Sicht des typischen Recherchierenden ist aber kaum zu erwarten, dass er den Aufwand auf sich nimmt, Webseiten zu verschlagworten. Google stellt zwar ein „Marker Tool“ bereit, das es gestattet, direkt aus dem Browser heraus Webseiten zu annotieren oder in CSEs aufzunehmen, ohne zuvor eine Google Website aufzusuchen, dennoch wird hier die These vertreten, dass sich Google Co-op eher auf Informationsanbieter bezieht bzw. stützt als auf die Vielzahl von Informationssuchenden, die Google täglich als Suchdienst nutzen. Dass Google Co-op aber auch für die Mehrzahl der Recherchierenden eine Verbesserung der Websuche bedeuten kann, zeigt aus quantitativer Perspektive die beeindruckende Zahl von bereits jetzt über 100000 existierenden CSEs. Des Weiteren bringt Google klar zum Ausdruck, dass die in Google Co-op vorhandenen Dienste stets auch dazu dienen, die algorithmenbasierte Google-Standardsuche zu verbessern.⁵⁶ Das Wissen der Community wird also als direkter Input zur Verbesserung algorithmenbasierter Suchverfahren verstanden und genutzt und ist in Googles Kontext nicht unabhängig von diesen oder gar substitutiv zu sehen bzw. findet erst in diesem Anwendungskontext seine Rechtfertigung.

3.5 Zusammenfassung und Ergebnis

In einer zusammenfassenden Perspektive lassen sich aus der Darstellung in Kap. 3 mehrere Ergebnisse ableiten: Zunächst wird ersichtlich, dass alle drei großen Websuchdienste deutliche Anstrengungen im Bereich der Personalisierung und „Social Search“ unternehmen, dabei aber bzgl. der jeweiligen Ausprägung erhebliche Unterschiede bestehen. Allen drei Anbietern ist gemeinsam, dass sie nutzerspezifisierbare Suchdienste bereitstellen. Diese Tatsache könnte als DER derzeitige Entwicklungstrend Trend im Web Information Retrieval bezeichnet werden. Vergleicht man die einzelnen Plattformen, so ist festzuhalten, dass sowohl bzgl. der angebotenen Funktionalität, der angesprochenen Zielgruppe als auch der öffentlichen Wahrnehmung aber erhebliche Unterschiede bestehen. Während etwa Microsofts

⁵⁶ Sterling, G. (2007), iGoogle, Personalized Search And You, URL searchengineland.com/070501-084656.php (letzter Zugriff 26.06.2007).

Suchmakros kaum ein mediales Echo im Web finden,⁵⁷ existiert für die Vielzahl von Google CSEs bereits ein eigenes Metasuchtool. Google und Yahoo unterscheiden sich ebenfalls sehr deutlich. Yahoo setzt mit Clever und MeinWeb und nicht zuletzt anderen zugekauften Community-Websites (Flickr.com, Del.icio.us) derzeit stark auf Community-basierte Dienste, um die Websuche zu verbessern. Bisher sind diese Dienste, sofern überhaupt ins Yahoo-Portal eingebunden, aber nur rudimentär in die Standardwebsuche integriert. Bei Google ist es umgekehrt: Personalisierung (Suchprotokoll) und „Social Search“ (Co-op) sind unmittelbar in die Websuche integriert bzw. erfüllen primär den Zweck, algorithmenbasierte Suchverfahren anzureichern bzw. Input zu deren Verbesserung zu liefern. Google scheint weniger den Aufbau von Communities an sich zu fokussieren, sondern das Ziel zu verfolgen, mit den angebotenen Diensten eine Vielzahl von Hilfswilligen zu akquirieren, welche bereit sind, das Internet zu kategorisieren. Mit der Vielzahl der oben behandelten – seit 2006 neu aufgebauten Diensten – kann Google gegenwärtig eindeutig als Marktführer auch im Bereich der personalisierten und sozialen Suche bezeichnet werden.⁵⁸

Für Recherchierende bieten die genannten Dienste eine Vielzahl neuer Möglichkeiten. Neben einer erwarteten höheren Qualität von Suchergebnissen die etwa durch interessen- oder themenspezifisch anpassbare Suchmaschinen (Pipes, CSEs) denkbar werden, stehen vor allem auch neue Dienste und Formen für das individuelle und kollaborative Wissensmanagement (MeinWeb, Suchprotokoll) bereit.

Hinsichtlich der Bereitschaft, Wissen zu teilen, zeigen die durch Google CSEs gesetzten extrinsischen aber auch die in Yahoo Clever vorhandenen intrinsischen Anreize neue Wege auf, die Nutzer zur Teilnahme an derartigen Diensten zu motivieren. Dass sowohl die finanziellen Belohnungen bei Google CSE als auch die Status-Gratifikationen bei Yahoo Clever auch qualitätssichernde Effekte bewirken, ist zwar zu erhoffen, dennoch ist derzeit weitgehend offen, ob und inwieweit die vorgestellten Dienste manipulationsresistent sind (Spam). Weiterhin ungeklärt bleibt die Frage, inwieweit seitens der Nutzer tatsächlich die Bereitschaft besteht, personalisierte Suchdienste, etwa das Google Suchprotokoll, anzunehmen und auch dauerhaft zu nutzen. Festzuhalten ist, dass Google mittlerweile dazu übergeht, die

⁵⁷ Die Suchanfrage „*microsoft search macros*“ liefert in Phrasenform am 26.06.2007 in Google gerade einmal 216 Treffer, während „*google custom search engines*“ 22 700 Ergebnisse und „*yahoo search builder*“ (beide Anfragen ebenfalls in Phrasenform) 12 900 Treffer liefern.

⁵⁸ Vgl. auch Sterling, G. (2007), iGoogle, Personalized Search And You, URL searchengineland.com/070501-084656.php (letzter Zugriff 26.06.2007), Sullivan, D. (2007) Google Ramps Up Personalized Search, URL searchengineland.com/070202-224617.php (letzter Zugriff 26.06.2007).

personalisierten Dienste zunehmend aktiver zu bewerben.⁵⁹ Aus dieser Perspektive können die „Google Apps für Ihre Domain“⁶⁰ bzw. das Angebot von Google, seinen E-Mail-Dienst Universitäten kostenlos bereitzustellen⁶¹ als auch die kostenlose Toolbar, auch als Hilfsmittel betrachtet werden, um personalisierten Google-Diensten eine breite Nutzerbasis zu erschließen.

Neben den in 3.4 angesprochenen datenschutzrechtlichen Aspekten ergeben sich aus Sicht der Recherchierenden bzw. derjenigen Nutzer, welche personalisierbare Suchoptionen nutzen, „Lock-In-Effekte“ bzw. höhere Wechselkosten zu anderen Suchdiensteanbietern. Führt man sich vor Augen, dass Google den Suchdienstemarkt global dominiert⁶² und z.B. in Deutschland eine fast monopolartige Stellung im Suchdienstebereich⁶³ einnimmt, kann man Personalisierung und „Social Search“ in Bezug auf Google aus einer kritischen Perspektive auch als weiteren Schritt auf dem Weg zur „digitalen Weltherrschaft“ interpretieren.⁶⁴

Im engeren Kontext des Web Information Retrieval stellen sich durch die derzeitigen Entwicklungen eine Vielzahl neuer Forschungsfragen. Beispielweise die zentrale Frage, ob und auf welche Weise und in welchen Kontexten die Soziale Suche und Personalisierung die Qualität der Websuche tatsächlich auf eine neue Stufe heben bzw. grundlegender noch die Fragestellung, wie denn die Effekte derartiger Ansätze methodologisch untersucht, d.h. valide und reliabel geprüft werden können. Die Liste offener und neuer Forschungsfragen ließe sich vielfältig ergänzen und weit fortführen. Greift man in diesem Kontext, den eingangs geschilderten Partizipationsgedanken des Web 2.0 auf, ist es naheliegend auch die Frage zu stellen, wie wir selbst einen Beitrag zur Verbesserung der Suche im Web leisten können.

⁵⁹ Hotchkiss, G. (2007), Just Behave: Google's Marissa Mayer on Personalized Search, URL searchengineland.com/070223-090000.php (letzter Zugriff 26.06.2007).

⁶⁰ URL www.google.com/a/help/intl/de/admins/overview.html (letzter Zugriff 26.06.2007).

⁶¹ Vgl. Heise News vom 13.06.2007, Google für das Leben, URL 193.99.144.85/newsticker/meldung/91071 (letzter Zugriff 26.06.2007).

⁶² Sullivan, D. (2007), Comparing Search Popularity Ratings: Google Climbs & Good News For Live.com, URL searchengineland.com/070323-175711.php (letzter Zugriff 26.06.2007).

⁶³ Vgl. hierzu z.B. die Angaben von www.webhits.de/deutsch/index.shtml?webstats.html (letzter Zugriff 26.06.2007).

⁶⁴ Vgl. Die Welt vom 10.05.2007, Wachablösung für Microsoft, URL www.welt.de/welt_print/article860553/Wachablosung_fuer_Microsoft.html (letzter Zugriff 26.06.2007). Vgl. auch Zitat „It's no longer a joke that the once small, lovable company wants to conquer the world.“ aus Sullivan, D. (2007), Google Search History expands, becomes web History, URL searchengineland.com/070419-181618.php (letzter Zugriff 25.06.2007).

4. Ausblick Web-Information-retrieval.de: Ein Ansatz zur Community-Building im weiteren Gebiet der Informationswissenschaft

Eine Möglichkeit diesen Partizipationsgedanken umzusetzen, besteht in der Bildung einer Community im weiteren Gebiet der Informationswissenschaft (IW), welche sich mit den Themenfeldern Web Information Retrieval, Internetsuchdienste, Suchmaschinenmarketing, Internetrecherche befasst. Die vom Verfasser initiierte Website Web-Information-Retrieval.de soll, so das Ziel, eine offene, nicht-kommerzielle Plattform werden, auf der sich Teilnehmer der IW-Fachwelt einbringen und austauschen können. Folgende Abbildung zeigt einen Ausschnitt der Startseite von Web-Information-Retrieval.de.

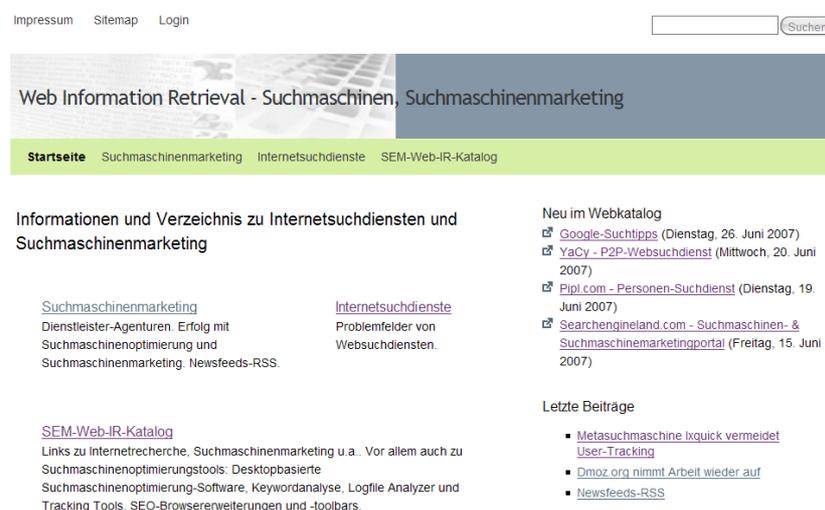


Abbildung 14: Web-Information-Retrieval.de

Kerngedanke ist es, die Kompetenz einer möglichst hohen Zahl von Experten zusammenzuführen und ähnlich Searchengineland.com, Pandia.org und anderen bekannten Websites, die primär den englischsprachigen Raum fokussieren, einen substanziellen Beitrag zur Transparenz im Suchdienste- und Retrievalbereich für den deutschsprachigen Raum zu schaffen, indem z.B. Artikel, Weblinks und Hintergrundinformationen zu den genannten Themenfeldern angeboten werden. Damit stellt die Community Recherchierenden fundiertes Know-how zur Websuche zur Verfügung und leistet somit auch einen Beitrag zur Beförderung von Informationskompetenz. Dabei ist nicht beabsichtigt, in Konkurrenz zu anderen thematisch ähnlichen Plattformen und Websites zu treten,⁶⁵ sondern ergänzend das vielfältige vorhandene Wissen im deutschsprachigen Raum zu bündeln. Beispielsweise können Wissenschaftler die Plattform nutzen, um auf ihre Arbeiten hinzuweisen. De Weiteren

⁶⁵ Beispielsweise recherchenblog.ch, www.durchdenken.de, www.at-web.de, www.b2i.de, blog.hapke.de, weblog.ib.hu-berlin.de.

besteht die Option, z.B. mit Hilfe von Social Bookmarks eine fachbezogene Ressourcensammlung aufzubauen und auch für das persönliche Wissensmanagement zu nutzen.

Web-information-retrieval.de steht am Beginn der Entwicklung und ist sowohl konzeptionell als auch bzgl. der zugrundeliegenden Software und Funktionalität offen. Inhaltlich stehen derzeit einige vom Verfasser erstellte Artikel bereit und insbesondere im Web-Katalog werden wöchentlich mehrere themenspezifische Weblinks eingetragen. Zusätzlich ist eine themenfokussierte Google-CSE eingebunden. Die Website lässt sich also bereits derzeit gewinnbringend nutzen. Alle Leser, ob Forscher oder interessierte Web-Rechercheure, sind eingeladen, sich auch inhaltlich und konzeptuell einzubringen. Zur Kontaktaufnahme genügt eine Email an joachim.griesbaum@web.de.

Literatur

- Blood, R. (2002), *The Weblog Handbook: Practical Advice on Creating and Maintaining Your Blog*, Cambridge, Mass.: Perseus Pub.
- Broder, A. (2006), *The Future of Web Search: From Information Retrieval to Information Supply*, URL www.cs.sjsu.edu/~tylin/ieeesilicon/broder.pdf (letzter Zugriff 20.06.2007).
- Chakrabarti, S. (2003), *Mining the Web. Discovering Knowledge from Hypertext Data*, Amsterdam et al.: Elsevier.
- Dalal, M. (2007), *Personalized Social & Real-Time Collaborative Search*. Poster on WWW 2007, May 8–12, 2007, Banff, Alberta, Canada. URL www2007.org/posters/poster887.pdf (letzter Zugriff 18.06.2007).
- Eberspächer, J.; Holtel, S. (Hg), *Suchen und Finden im Internet*, Springer: Berlin, Heidelberg, New York (2007).
- Ferber, R. (2003), *Information Retrieval. Suchmodelle und Data-Mining-Verfahren für Textsammlungen und das Web*, Heidelberg: dpunkt Verlag.
- Giles, J. (2005), *Internet encyclopaedias go head to head*. In: *Nature* 438, 900-901 URL www.nature.com/nature/journal/v438/n7070/full/438900a.html (letzter Zugriff 20.06.2007).
- Gordon-Murnane, L. (2006), *Social-Bookmarking, Folksonomies, and Web 2.0 Tools*. In: *Searcher*, vol.14, no.6, S. 26-38. URL www.redorbit.com/news/technology/541192/social_bookmarking_folksonomies_and_web_20_tools/index.html?source=r_technology (letzter Zugriff 18.06.2007).
- Griesbaum, J. (2003), *Unbeschränkter Zugang zu Wissen? Leistungsfähigkeit und Grenzen von Suchdiensten im Web. Zwischen informationeller Absicherung und*

- manipulierter Information. In: Ralph Schmidt (Hg.), *Competence in Content, Proceedings 25. Online-Tagung der DGI, Frankfurt M. 2003*, S.37-50.
- Griesbaum, J.; Bekavac, B. (2004), *Web-Suche im Umbruch? Entwicklungstendenzen bei Web-Suchdiensten*. In: Bekavac, Bernard; Herget, Josef; Rittberger, Marc (Hg.): *Information zwischen Kultur und Marktwirtschaft. Proceedings des 9. Internationalen Symposiums für Informationswissenschaft (ISI 2004)*, Chur, 6.-8.Oktober 2004. Konstanz: UVK Verlagsgesellschaft mbH, 2004, S. 283-299.
- Horowitz, B. (2007), *Challenges of the Next Years*. In: Eberspächer, J.; Holtel, S. (Hg): *Suchen und Finden im Internet*, Berlin, Heidelberg: Springer, S. 187-194.
- Hotho, A.; Jäschke, R; Schmitz, C.; Stumme, G. (2006), *Information Retrieval in Folksonomies: Search and Ranking*. In: *Information Retrieval in Folksonomies: Search and Ranking, Volume 4011/2006*, S.411-426. URL www.springerlink.com/content/r8313654k80v7231/fulltext.pdf (letzter Zugriff 18.06.2007).
- Lewandowski, D. (2004), *Technologie-Trends im Bereich der WWW-Suchmaschinen* In: *Information Professional 2011: 26. Online-Tagung der DGI; Frankfurt am Main 2004*, S. 183-195.
- Lewandowski, D. (2005), *Web Information Retrieval. Technologien zur Informationssuche im Internet*. Frankfurt am Main: DGI Schrift. URL www.durchdenken.de/lewandowski/web-ir/ (letzter Zugriff 18.06.2007).
- Millen, D.; Feinberg, J.; Kerr, B. (2005), *Social-Bookmarking in the enterprise*. ACM Queue, vol.3, no.9, Nov. 2005, pp. 28-35.
- Neuberger, C.; Nuernbergk, C.; Rischke, M. (2007), *Weblogs und Journalismus: Konkurrenz, Ergänzung oder Integration?* In: *Media-Perspektiven, 2/2007*, S.96–122. URL www.ard-werbung.de/showfile.phtml/02-2007_neuberger.pdf?foid=19747 (letzter Zugriff 15.06.2007).
- Niwa, S.; Doi, T.; Honiden S. (2006), *Web page recommender system based on folksonomy mining for ITNG '06 submissions*. in: *Proceedings. Third International Conference on Information Technology: New Generation*. IEEE Computer Society, S. 6.
- O'Reilly, T. (2005), *What Is Web 2.0. Design Patterns and Business Models for the Next Generation of Software*. URL www.oreillynet.com/pub/a/oreilly/tim/news/2005/09/30/what-is-web-20.html (letzter Zugriff 13.06.2007).
- Regulski, K. (2007), *Aufwand und Nutzen beim Einsatz von Social-Bookmarking-Services als Nachweisinstrument für wissenschaftliche Forschungsartikel am Beispiel von BibSonomy*, Preprint URL www.bibliothek-saur.de/preprint/2007/ar2460_regulski.pdf (letzter Zugriff 18.06.2007)
- Stoilova, L.; Holloway, T.; Markines, B. (2005), *GiveALink: Mining a Semantic Network of Bookmarks for WebSearch and Recommendation*. In: *LinkKDD '05*, August 21, 2005, Chicago. URL

www.informatics.indiana.edu/fil/Papers/givealink-linkkdd.pdf (letzter Zugriff 18.06.2007).

Voß, J. (2007), Tagging, Folksonomy & Co – Renaissance of Manula Indexing. In: Oßwald, A.; Stempfhuber, M.; Wolff, C. (Hg), Open Innovation. Neue Perspektiven im Kontext von Information und Wissen. Proceedings des 10. Internationalen Symposiums für Informationswissenschaft (ISI 2007), Köln, 30.05-01.06.2007, Konstanz: UVK Verlagsgesellschaft mbH, 2004, S. 243-254.

Weitkämper, N. (2007), Durch die Hintertür das nicht für möglich Gehaltene erreichen: Die Kategorisierung des Internet. In: Password Nr.4, S.14.