

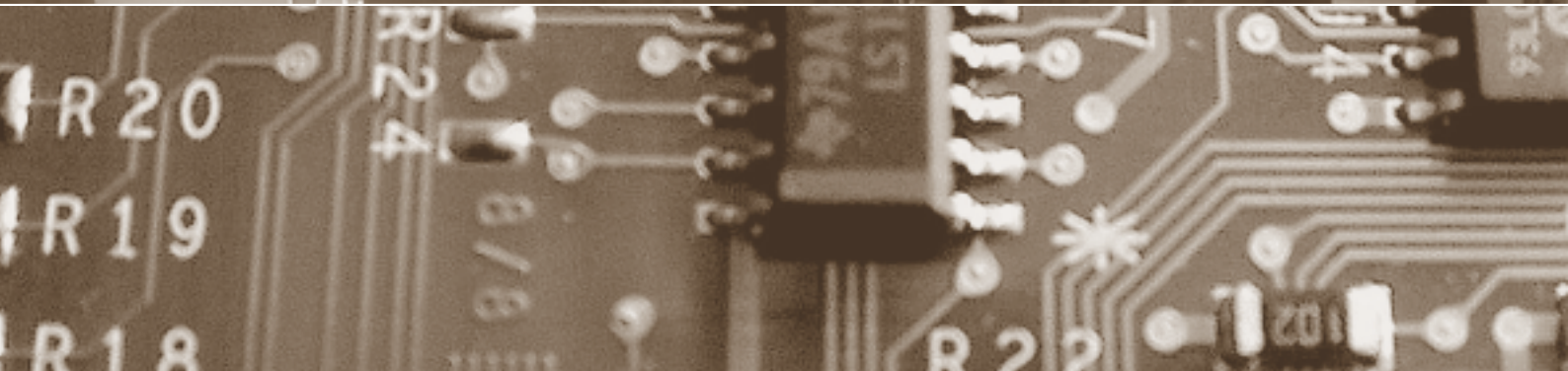
Schwerpunkt:

Suchen – Vergessen

fokus: Suchmethoden im Netz: heute – morgen

fokus: Wer sucht, der findet (nicht immer): Internetzensur

report: BWIS-II-Reform: kritische Bemerkungen



Herausgegeben von
Bruno Baeriswyl
Beat Rudin
Bernhard M. Hämmerli
Rainer J. Schweizer
Günter Karjoth

fokus

Schwerpunkt:

Suchen – Vergessen

auftakt

Geschürte Ängste – gleichgültige

Menschen

von Gerhart R. Baum

Seite 101

Vergessen – nicht vergessen –

nicht finden

von Beat Rudin

Seite 104

Suchmethoden im Netz: heute – morgen

von Felix-Robinson Aschoff

und Abraham Bernstein

Seite 106

Daten «löschen» oder wirksam
entfernen?

von Thomas Burris

Seite 110

Wer sucht, der findet (nicht immer)

von Daniel Zinn

Seite 114

Best Practices for Social Networks

von Giles Hogben

Seite 120

zwischenakt

Weit vom (Daten-)Geschütz ...

von Frank U. Frey

Seite 123

Die Entwicklung von Suchtechnologien für das World Wide Web gehört heute zu den zentralen Herausforderungen der Informatik. Eine Alternative zu den heutigen Algorithmen-basierten Suchmaschinen stellen hierbei Social-Search-Ansätze dar. Das Semantic Web beinhaltet schliesslich die Vision, komplexe natürlichsprachige Anfragen beantworten zu können.

**Suchmethoden
im Netz: heute –
morgen**

Nicht alles, was im Internet steht, ist gut. Es kann zutiefst menschenunwürdig und rechtswidrig sei (wie Kinderpornografie) oder auch bloss einem Regime nicht gefallen. Aus diesem Grund gibt es rund um den Erdball Internetzensur – zum Teil mit Unterstützung der Betreiber von Internet-suchmaschinen.

**Wer sucht,
der findet
(nicht immer)**

«Social Networks» erleben einen regelrechten Boom und können neben gewünschten Effekten auch unerwünschte Nebenwirkungen zeitigen. Die ENISA (European Network and Information Security Agency) hat ein Papier erarbeitet, worin einerseits die Bedrohungen dargestellt und andererseits Empfehlungen für Betreiber und Nutzer aufgestellt werden.

**Best Practices for
Social Networks**

impresum

digma: Zeitschrift für Datenrecht und Informationssicherheit, ISSN: 1424-9944, Website: www.digma.info

Herausgeber: Dr. iur. Bruno Baeriswyl, Dr. iur. Beat Rudin, Prof. Dr. Bernhard M. Hämmerli, Prof. Dr. iur. Rainer, J. Schweizer, Dr. Günter Karjoth

Redaktion: Dr. iur. Bruno Baeriswyl und Dr. iur. Beat Rudin

Rubrikenredaktor: Dr. iur. Amédéo Wermelinger

Zustelladresse: Redaktion digma, c/o Stiftung für Datenschutz und Informationssicherheit, Kirschgartenstrasse 7, CH-4010 Basel
Tel. +41 (0)61 270 17 70, redaktion@digma.info

Erscheinungsplan: jeweils im März, Juni, September und Dezember

Abonnementspreise: Jahresabo Schweiz: CHF 158.00, Jahresabo Ausland: Euro 112.00 (inkl. Versandkosten), Einzelheft: CHF 42.00

Anzeigenmarketing: Publicitas Publimag AG, Mürtchenstrasse 39, Postfach, CH-8010 Zürich
Tel. +41 (0)44 250 31 31, Fax +41 (0)44 250 31 32, www.publimag.ch, service.zh@publimag.ch

Herstellung: Schulthess Druck AG, Arbenzstrasse 20, Postfach, CH-8034 Zürich

Verlag und Abonnementsverwaltung: Schulthess Juristische Medien AG, Zwingliplatz 2, Postfach, CH-8022 Zürich
Tel. +41 (0)44 200 29 99, Fax +41 (0)44 200 29 98, www.schulthess.com, zs.verlag@schulthess.com

BWIS-II-Reform: kritische Bemerkungen Mit der beabsichtigten Reform des Bundesgesetzes über Massnahmen zur Wahrung der inneren Sicherheit (BWIS) sollen dem Staatsschutz weitreichende Eingriffsbefugnisse zugestanden werden. Der Beitrag setzt sich kritisch mit den heiklen Aspekten des Entwurfs auseinander.

OpenID: Single-Sign-On-Spezifikation OpenID ist ein dezentral angelegtes Single-Sign-On-System, bei welchem der Benutzer über seine Identitäten und über die Weitergabe seiner Identitäts-Attribute entscheidet, wobei die Unabhängigkeit von Identity Providern gewährleistet wird.

Schmerzvolle Erfahrung Die Einwilligung in Persönlichkeitsverletzungen gehört zur Ausübung der höchstpersönlichen Rechte, welche der urteilsfähigen minderjährigen Person zustehen, wie das Bundesgericht im Fall der schmerzhaften Behandlung einer 13-jährigen festgehalten hat. Was heisst das z. B. bei Facebook-Eintragungen, die sich bei einer Stellenbewerbung negativ auswirken können?

Videoüberwachung im Lichte der Verfassung Die Datenschutzkommission des Fürstentums Liechtenstein hat entschieden, dass eine flächendeckende Videoüberwachung in der Fussgängerzone der Hauptstadt Vaduz rund um die Uhr zur Vermeidung von Vandalenakten und Straftaten unverhältnismässig sei.

report

GESETZGEBUNG

BWIS-II-Reform: kritische Bemerkungen

von Lucien Müller, Nina Widmer und Rainer J. Schweizer

Seite 124

IT-SICHERHEITSAUSBILDUNG

Über den Nutzen offensiver Lehre von Martin Mink

Seite 130

TECHNIK

von Robert Ott

Seite 134

RECHTSPRECHUNG

Verwaltungsrat lebenslang auf dem Web?

von Amédéo Wermelinger

Seite 138

RECHTSPRECHUNG

von Amédéo Wermelinger

Seite 139

RECHTSPRECHUNG

Videoüberwachung im Lichte der Verfassung

von Philipp Mittelberger

Seite 140

TRANSFER

Zugriffschutz: eine technische Herausforderung

von Roland Portmann

Seite 144

forum

ISSS

Am Puls der Informationssicherheit

von Bernhard Hämmerli

Seite 146

agenda

Seite 147

schlussakt

Nachträgliche zusätzliche «Sicherheit»?

von Bernhard Hämmerli

Seite 148

Cartoon

von Hanspeter Wyss

Suchmethoden im Netz: heute – morgen

Von der herkömmlichen Suchmaschine bis zur Vision einer verständnisvollen Antwort: Potenziale und Begrenzungen



Felix-Robinson Aschoff, Forschungsassistent Information Management Research Group, Institut für Informatik (IFI) der Universität Zürich, Zürich
aschoff@ifi.uzh.ch



Prof. Abraham Bernstein, PhD, Leiter der Forschungsgruppe Dynamic and Distributed Information Systems, Institut für Informatik (IFI) der Universität Zürich, Zürich
bernstein@ifi.uzh.ch

Den enormen Herausforderungen an Web-Suchtechnologien kann langfristig nur durch eine Kombination von technologischen und sozialen Ansätzen begegnet werden.

Das Internet und das darauf aufbauende World Wide Web (WWW) werden heute weithin als eine der wichtigsten technischen Innovationen in der Geschichte gewürdigt. Das WWW stellt hierbei einen vorläufigen Höhepunkt in der Entwicklung menschlicher Kommunikationstechnologien dar. Ausgehend von antiken Feuerzeichen über die Verbreitung des Buchdrucks im ausgehenden Mittelalter bis hin zur Entwicklung der Massenmedien im 20. Jahrhundert erlaubt uns das Internet nun einen Austausch von Informationen in nie gekanntem Ausmass.

Beeindruckend sind hierbei die Wachstumsraten dieses neuen Mediums. Manche Forscher gehen von einer jährlichen Verdopplung der öffentlich zugänglichen Webseiten aus¹. Eine Reihe von Schätzungen weist jedoch darauf hin, dass selbst diese exponentielle Wachstumskurve die tatsächliche Zunahme an Netzinhalten noch unterschätzt¹. Google hat in diesem Jahr angegeben, eine Billion (10^{12}) unterschiedliche Links gefunden zu haben², und die Zahl der weltweiten Internetnutzer wird auf über 1,4 Milliarden geschätzt³.

Eines der grössten Probleme aus technologischer Sicht besteht in der Beherrschung der ungemein grossen, heterogenen und dezentral organisierten Informationsflut, die das Internet täglich bereithält. Die eigentliche Stärke des Internets, nämlich, dass jede Person weitgehend ohne formale Vorgaben beliebige Inhalte publizieren kann, wird aus Sicht der Informationssuche und -findung zur grossen Herausforderung. Absolut gesehen ist die Zahl der prinzipiell verfügbaren Netzinhalte zwar beeindruckend. Aus

Sicht des Nutzers zählt aber letztendlich nicht, ob die gesuchte Information irgendwo im Netz gespeichert ist, sondern nur, ob er auch in der Lage ist, diese zu finden. Suchmaschinen wie beispielsweise Google stellen einen Zugang in dieses Informationslabyrinth dar und haben in den letzten Jahren an wirtschaftlichem sowie auch an gesellschaftspolitischem Einfluss deutlich zugenommen. Die Vorteile und Begrenzungen dieser Technologie sowie alternative Ansätze sollen in diesem Beitrag vorgestellt werden.

Herkömmliche Suchmaschinen

Die Funktionsweise einer Suchmaschine lässt sich stark vereinfacht anhand folgender Schritte aufzeigen:

- Das Web wird kontinuierlich nach neuen Inhalten durchsucht. Dieser Prozess ist heute bei den meisten Suchmaschinen automatisiert und wird durch sogenannte «Web Crawler» durchgeführt.
- Die gefundenen Seiten und Dokumente werden indiziert, d.h. es wird von den Suchmaschinenbetreibern eine Datenstruktur erzeugt, die ein schnelles Durchsuchen der extrem grossen Linksammlungen ermöglicht.
- Ein Webnutzer formuliert ein Informationsbedürfnis, zumeist indem er ein oder mehrere Suchbegriffe in das Suchmaschinen-Interface eingibt.
- Die Suchmaschine greift auf den Index zu, um anhand der Suchbegriffe eine Resultatmenge an Internet-Links zusammenzustellen.
- Ein Ranking-Algorithmus berechnet für jede Seite einen Relevanz-Wert, und die Seiten werden dem Nutzer in absteigender Reihenfolge dieser Werte präsentiert.

Diese schematische Darstellung lässt erkennen, dass es neben vielen anderen Faktoren drei wesentliche Aspekte gibt, die die Chancen des Suchenden, eine bestimmte Webseite zu finden, beeinflussen:

- Welcher Anteil des gesamten Webs kann von einer Suchmaschine gefunden und indiziert werden?
- Wie gut kann der Nutzer sein Informationsbedürfnis in der Abfragesprache formulieren?

■ Wie gut arbeitet der Ranking-Algorithmus, der die Reihenfolge der gefundenen Webseiten festlegt?

Es ist schwierig, verlässliche Schätzungen darüber zu finden, welcher Anteil des Webs durch aktuelle Suchmaschinen erfasst wird. KOBAYASHI/TAKEDA⁴ diskutieren Studien, bei denen die Abdeckung des Gesamtwebs durch die damals (2000) aktuellen Suchmaschinen um die 20% liegt. Eine zusätzliche Schwierigkeit bei der Betrachtung stellt das sog. Deep Web dar. Gemeint sind hiermit vor allem Datenbanken, die Webseiten dynamisch aufgrund von Anfragen erzeugen und somit von Suchmaschinen nur schwer erfasst werden können. Es wird davon ausgegangen, dass das Deep Web noch einmal 500-mal mehr Daten enthält als das «Surface Web» und nur zu ca. 30–40% von Suchmaschinen abgedeckt wird⁵.

Als Abfragesprache zur Formulierung des Informationsbedürfnisses verlangen fast alle Suchmaschinen von den Nutzern die Eingabe von Stichworten. Studien haben jedoch gezeigt, dass viele Nutzer die Möglichkeiten dieser Abfrageart nur sehr eingeschränkt oder in Bezug auf logische Operatoren oft sogar falsch nutzen⁶. Es gelingt den Nutzern also häufig nicht, ihr Informationsbedürfnis adäquat zu formulieren. Eigene Forschungen⁷ weisen darauf hin, dass in manchen Situationen die Verwendung von natürlicher Sprache besser geeignet ist. Dieser Ansatz wird z. B. von der Suchmaschine Powerset (www.powerset.com) verfolgt.

Der angesprochene Ranking-Algorithmus ist für Web-Suchmaschinen sehr relevant, da jede Anfrage meist mehrere tausend Treffer zurückliefert. Ein Nutzer wird also meistens nur solche Links aufrufen, die auf den ersten Seiten der Ergebnisliste präsentiert werden. Für dieses Ranking greifen Suchmaschinen auf eine ganze Reihe von Informationen zurück. Dies schliesst z. B. die Häufigkeit und die Position des Suchbegriffs im entsprechenden Dokument ein sowie die Aktualität der Webseite oder den geografischen Aufenthaltsort der suchenden Person. Ein weiterer wichtiger Anhaltspunkt für diese Berechnung ist die Verlinkungsstruktur der Seite. Hierbei wird angenommen, dass Seiten, auf die oft von anderen Seiten verlinkt wird, eine grössere Relevanz und Zuverlässigkeit haben (vergleichbar mit dem Zitationsindex für wissenschaftliche Artikel). Die Umsetzung dieser Idee mit dem Google-Page-Rank-Algorithmus⁸ vor zehn Jahren hat entscheidend zum Wettbewerbsvorteil dieser Suchmaschine beigetragen.

Ein Problem des Ranking-Verfahrens besteht jedoch darin, dass schnell zu findende Webseiten inzwischen einen bedeutenden wirtschaftlichen Vorteil versprechen. Das geht so weit, dass Un-

ternehmen mittlerweile grössere Anstrengungen machen, um den Ranking-Algorithmus zu ihren Gunsten zu manipulieren. Dies kann schliesslich dazu führen, dass Suchresultate nicht mehr den Relevanzkriterien der Nutzer unterliegen, sondern verstärkt von der wirtschaftlichen Macht bestimmter Webseitenanbieter abhängen.

Die grosse Stärke der herkömmlichen Suchmaschinen-Technologie liegt im Umstand der Automatisierung und der Skalierbarkeit. Auch

Aus Sicht des Nutzers zählt nicht, ob die gesuchte Information irgendwo im Netz gespeichert ist, sondern nur, ob er auch in der Lage ist, diese zu finden.

wenn der Gesamtinhalt des Netzes nicht annähernd erfasst werden kann, so lassen sich durch die Kombination der oben beschriebenen Verfahren doch enorm grosse Datenbanken zusammensetzen und die Ergebnisse von Suchanfragen vergleichsweise zielgenau präsentieren. Die Stärke der Automatisierung bringt jedoch auch gewisse Begrenzungen mit sich. Automatische Verfahren zum Finden, Indexieren und Präsentieren von Inhalten können gewisse Aufgaben nur sehr begrenzt erfüllen, die Menschen kaum Mühe bereiten. Dies schliesst z. B. das Erkennen und Einordnen von Multimedia-Inhalten wie Fotos, Filme oder Musikstücke ein, die im Netz aufgrund wachsender Bandbreite zunehmend zu finden sind.

Zusammenfassend lässt sich sagen, dass diese Algorithmen-basierten Suchmaschinen zurzeit wohl die mächtigsten Werkzeuge sind, um Inhalte im Netz zu finden. Der umfassende und systematische Eindruck, den die präsentierten Resultatlisten zuweilen auf den Nutzer haben, ist aber eher trügerisch. Metaphorisch ist die Situa-

Kurz & bündig

Die Entwicklung von Suchtechnologien für das World Wide Web gehört heute zu einer der zentralen Herausforderungen der Informatik. Algorithmen-basierte Suchmaschinen stellen zurzeit die mächtigsten Werkzeuge für die Suche nach Webinhalten dar. Sie weisen jedoch auch Schwächen auf, wie die aus Nutzersicht suboptimale Eingabe von Stichworten oder die eingeschränkte Fähigkeit, Multimedia-Inhalte kategorisieren zu können. Eine Alternative stellen hierbei Social-Search-Ansätze dar, bei denen Nutzer Inhalte mit Metadaten versehen können, um sie für andere auffindbar zu machen. Das Semantic Web beinhaltet schliesslich die Vision, komplexe natürlichsprachige Anfragen verstehen und beantworten zu können. Dies setzt allerdings eine bedeutungsbezogene Extraktion von Netzinhalten voraus, wie sie zurzeit noch nicht realisiert ist.

tion eines Suchmaschinen-Nutzers vermutlich mit jemandem zu vergleichen, der mit blossen Augen in den Sternenhimmel blickt: Die grossen und hellen Objekte der unmittelbaren Umgebung werden zwar gesehen, eine weitaus grössere Anzahl an kleineren und weiter entfernten Elementen bleibt im ungeheuer grossen Suchraum jedoch unentdeckt.

Das Web 2.0 und die «soziale» Suche

Im Rahmen der Webforschung wird in den letzten Jahren ein Phänomen diskutiert, das die Nutzungsstrukturen des Internets stark verändert hat. Unter dem Stichwort Web 2.0⁹ wird hierbei eine Entwicklung beschrieben, bei der der Nutzer nicht lediglich Informationen aus dem Internet konsumiert, sondern auch selbst aktiv Inhalte bereitstellt, mit anderen austauscht und Inhalte von anderen durch zusätzliche Information anreichert. Weithin bekannte interaktive Plattformen dieser Art sind das kooperative Internetlexikon Wikipedia oder Austauschplattformen für Multimedia-Inhalte wie Flickr für Fotos oder Youtube für Videos. Der Inhalt dieser Plattformen verändert sich aufgrund der grossen und sehr aktiven Anzahl an Teilnehmern sehr dynamisch, und

den herausgebildet, die als «Social-Search»-Ansätze bezeichnet werden. Sie beruhen darauf, dass Nutzer selbst Netzinhalte wie Fotos oder Videos Metadaten hinzufügen. Diese Informationsanreicherung ermöglicht es wiederum anderen Nutzern, die entsprechenden Inhalte zu finden. Die häufigste Ausprägung dieser Metadaterzeugung ist das «tagging»: Ein oder mehrere kurze Stichworte werden vom Autor dem Informationsobjekt mitgegeben. Das Gesamtgebilde dieser Tags führt daraufhin zu sog. «Folksonomien»¹⁰, also quasi vom «Volk» erzeugten Taxonomien.

Neben Youtube und Flickr gibt es «Social Bookmarking» Sites, die auf dem Social-Search-Prinzip aufbauen. Del.icio.us z.B. erlaubt es Internet-Nutzern, ihre Internet-Bookmarks mit anderen zu teilen und somit von der Sucherfahrung anderer zu profitieren.

Zusätzlich zur menschlichen Fähigkeit, Bilder und Filme kategorisieren zu können, liegt eine Stärke des Ansatzes darin, dass er von grossen Nutzergruppen einfach und intuitiv eingesetzt werden kann. Es werden keine Kenntnisse in Bezug auf formale Abfragesprachen oder Taxonomiesysteme vorausgesetzt. Aus Sicht eines effizienten Web Information Retrievals führt diese Freiheit allerdings auch zu den Begrenzungen des Ansatzes. Die Schwierigkeit liegt darin, dass zwei Personen, die einem Element dieselbe Bezeichnung geben, nicht unbedingt auch dasselbe meinen¹¹. Dieser Umstand verringert die Genauigkeit von Suchresultaten, da ein expliziter gemeinsamer Kontext zwischen den Nutzern oft fehlt (siehe VON AHN¹² für einen interessanten Lösungsansatz des Problems, bei dem zwei Personen spielerisch denselben Begriff für ein gezeigtes Bild finden müssen).

Da bei dieser Suchmethode im Gegensatz zum automatischen Indexieren Aufwand für die

Die Schwierigkeit von «Social-Search»-Ansätzen liegt darin, dass zwei Personen, die einem Element dieselbe Bezeichnung geben, nicht unbedingt auch dasselbe meinen.

nicht auf Text basierende Inhalte sind, wie bereits erwähnt, von herkömmlichen Suchmaschinen nur schwer zu erfassen.

Aufgrund dieser Unzulänglichkeiten der traditionellen Ansätze haben sich neue Suchmetho-

Literatur

- ALPERT, J./HAJAJ, N. (2008), We knew the web was big... The Official Google blog, 25.7.2008, <<http://googleblog.blogspot.com/2008/07/we-knew-web-was-big.html>>.
- BERNERS-LEE, T./HENDLER, J./LASSILA, O. (2001), The semantic web, Scientific American, Jg. 284 (5), 34–43.
- BRIN, S./PAGE, L. (1998), The anatomy of a large-scale hypertextual Web search engine, Computer Networks and ISDN Systems, Jg. 30, 107–117.
- FURNAS, G.W./LANDAUER, T.K./GOMEZ, L.M./DUMAIS, S. T. (1987), The vocabulary problem in human-system communication, Commun. ACM, 30 (11), 964–971.
- HE, B./PATEL, M./ZHANG, Z./CHEN-CHUAN CHANG, K. (2007), Accessing the deep web. Commun. ACM 50(5): 94–101.
- JANSEN, B. J./SPINK, A./SARACEVIC, T. (2000), Real life, real users, and real needs: a study and analysis of user queries on the web. Information Processing & Management, 36(2) 207–227.
- KAUFMANN, E./BERNSTEIN, A. (2007), How Useful Are Natural Language Interfaces to the Semantic Web for Casual End-Users? Proc. of the International Semantic Web Conference (ISWC), 2007.
- KOBAYASHI, M./TAKEDA, K. (2000), Information retrieval on the web, ACM Comput. Surv., 32(2) 144–173.
- O'REILLY, T. (2005), What is Web 2.0., <<http://www.oreillynet.com/pub/a/oreilly/tim/news/2005/09/30/what-is-web-20.html>>.
- SHADBOLT, N./HALL, W./BERNERS-LEE, T. (2006), The semantic web revisited. IEEE Intelligent Systems, 21(3), 96–101.
- VANDER WAL, T. (2005), Folksonomy Definition and Wikipedia. <<http://www.vanderwal.net/random/entrysel.php?blog=1750>>.
- VON AHN, LUIS (2006), Games with a Purpose. IEEE Computer 39(6): 92–94.

Nutzer entsteht, sind die bisherigen Folksonomien in ihrem Umfang mit den Indices grosser Suchmaschinen sicher nicht zu vergleichen. Dieser Umstand wird aber durch die Verteilung des Aufwandes auf die stark steigende Anzahl an Internetnutzern zu einem gewissen Grad kompensiert.

Ausblick auf eine Suche im Semantic Web

Die bereits beschriebene Heterogenität der Formate und Inhalte im ursprünglichen Netz sowie im Web 2.0 führt dazu, dass die gespeicherten Informationen nur sehr eingeschränkt von Maschinen auf ihre «eigentliche Bedeutung» hin analysiert werden können. Bereits relativ einfache Datenstrukturen wie Adressen liegen im Web in vielen unterschiedlichen Formaten vor und machen eine einheitliche Erfassung schwierig. Von der Fähigkeit, gesamte Fragen verstehen und beantworten zu können, wie z. B. «Wann wurde die Schweiz gegründet?», sind die meisten aktuellen Suchmaschinen noch weit entfernt. Diese Art der bedeutungsvollen Suche setzt voraus, dass die Informationen auf einer Webseite so aufgearbeitet werden können, dass die inhaltlichen Bezüge und der Kontext von Maschinen erfasst werden. Dies setzt im Allgemeinen eine Modellierung des gespeicherten Wissens in Form von Ontologien oder von statistischen Informationsmodellen voraus. Diese Ontologien legen explizit fest, was mit welchem Begriff oder Konzept gemeint ist, was die Beziehungen zwischen Konzepten und Worten sind und wie es sich in den semantischen Kontext der Seite einbettet. In der Vision des «Semantic Web»¹³ könnten autonome Software-Agenten auf dieser Grundlage komplexe Suchanfragen beantworten, bei denen inhaltliche sowie räumliche und zeitliche Bedingungen miteinander verknüpft werden: z. B. «Bitte suche mir eine Ärztin in Zürich-Oerlikon, die auf Magenschmerzen spezialisiert ist und Mittwoch Vormittag Sprechstunde hat.»

Diese netzumspannende Bedeutungsebene ist in dieser Form zurzeit noch nicht realisiert¹⁴. Dies liegt unter anderem daran, dass gemeinsame Standards, wie Ontologien sie darstellen, in der heterogenen Nutzergruppe des Webs nicht einfach durchzusetzen sind. Auf der anderen Seite benötigen Vorgaben, wie sie durch das World Wide Web Consortium (W3C) in Form des Resource Description Frameworks (RDF) oder der Web Ontology Language (OWL) ausgearbeitet wurden, Zeit, sich zu etablieren. Es wird hierbei auch nicht mehr von einer einheitlichen Ontologie ausgegangen, die tatsächlich das gesamte Web abdeckt. Es werden eher lokale Ontologien entstehen, wie dies heute bereits vor allem für wissenschaftliche Anwendungen der Fall ist. Diese

Ontologien können wiederum genutzt werden, um das Bedeutungswissen zwischen verschiedenen Domänen auszutauschen und so nach und nach ein semantisches Web zu realisieren.

Ein zusätzlicher Vorteil der Verwendung von erweiterten Modellen wie Ontologien oder statistischen Sprachmodellen ist, dass sie die Basis für die Verarbeitung von natürlichsprachigen Anfragen sein können. Eigene Studien¹⁵ haben ge-

Zukunft: «Bitte suche mir eine Ärztin in Zürich-Oerlikon, die auf Magenschmerzen spezialisiert ist und Mittwoch Vormittag Sprechstunde hat.»

zeigt, wie sich aufgrund solcher Modelle einfache, aber durchaus mächtige Suchsysteme realisieren lassen.

Insgesamt zeichnet sich eine Verknüpfung der Web-2.0-Ansätze mit Konzepten des Semantic Web ab, was mitunter als «Web 3.0» bezeichnet wird. Die von Nutzer-Communities erzeugten Folksonomien eignen sich beispielsweise gut als Ausgangsvokabular für die Entwicklung von Ontologien, da sie die wesentlichen Begriffe einer Domäne aus Nutzersicht gut repräsentieren. Es scheint sich insgesamt die Einsicht durchzusetzen, dass hybride Ansätze, welche die Stärken technischer Automatisierung mit den Potenzialen sozialer Netzwerke und menschlicher Akteure verknüpfen, für eine zukünftige Suchtechnologie am Erfolg versprechendsten sind. ■

Fussnoten

- ¹ KOBAYASHI, M./TAKEDA, K. (2000).
- ² ALPERT, J./HAJAJ, N. (2008).
- ³ <<http://www.internetworldstats.com/stats.htm>> (letztmals kontrolliert: 18.8.2008).
- ⁴ KOBAYASHI, M./TAKEDA, K. (2000).
- ⁵ HE, B./PATEL, M./ZHANG, Z./CHEN-CHUAN CHANG, K. (2007).
- ⁶ JANSEN, B. J./SPINK, A./SARACEVIC, T. (2000).
- ⁷ KAUFMANN, E./BERNSTEIN, A. (2007).
- ⁸ BRIN, S./PAGE, L. (1998).
- ⁹ O'REILLY, T. (2005).
- ¹⁰ VANDER WAL, T. (2005).
- ¹¹ FURNAS, G.W./LANDAUER, T.K./GOMEZ, L.M./DUMAIS, S. T. (1987).
- ¹² Siehe VON AHN, LUIS (2006).
- ¹³ BERNERS-LEE, T./HENDLER, J./LASSILA, O. (2001).
- ¹⁴ SHADBOLT, N./HALL, W./BERNERS-LEE, T. (2006).
- ¹⁵ KAUFMANN, E./BERNSTEIN, A. (2007).

Meine Bestellung

- 1 Jahresabonnement digma (4 Hefte des laufenden Jahrgangs)
à **CHF 158.00** bzw. bei Zustellung ins Ausland **EUR 123.00** (inkl. Versandkosten)

Name _____ Vorname _____

Firma _____

Strasse _____

PLZ _____ Ort _____ Land _____

Datum _____ Unterschrift _____

Bitte senden Sie Ihre Bestellung an:

Schulthess Juristische Medien AG, Zwingliplatz 2, CH-8022 Zürich

Telefon +41 44 200 29 19

Telefax +41 44 200 29 18

E-Mail: zs.verlag@schulthess.com

Homepage: www.schulthess.com

Schulthess 