

WISSENSVERNETZUNG CHANCEN NEUER PROZESSE

HANNES DIEDRICH

Wer Informationen sucht, wird im Internet meist irgendwie fündig. Dennoch erweist sich eine zielgerichtete »intelligente« Suche nach konkreten Informationen und ihre Verknüpfung trotz der immer raffinierteren Suchmechanismen von Google & Co. als schwierig. Hoffnung verbreitet derzeit die Idee vom semantischen Web. Das »einzige Bedeutungsnetz von allem und jedem« (WWW-Erfinder Tim Bernes-Lee) wird als universelle Weiterentwicklung des World Wide Web gehandelt. Das Bundesministerium für Bildung und Forschung will deshalb bis 2006 insgesamt 13,7 Millionen Euro in die Förderung dieser Technologie stecken. Bis dahin wollen die Projektpartner eine erste Anwendung der neuen Technologie präsentieren: Das »Smart Web« soll Service-Informationen zu den einzelnen Spielen der Fußballweltmeisterschaft liefern und auch Reiseanfragen beantworten können [@1].

Im semantischen Web wird allen möglichen Dingen – also beispielsweise Texten, Bildern, Musik und Videos – eine Bedeutung zugewiesen. Diese Angaben werden als Metadaten bezeichnet, als Informationen über Informationen. Der Vorteil: Das durch die Verknüpfung dieser Daten entstehende Netz aus Bedeutungen ist für Software »lesbar«, so dass nicht nur die »Dinge« sondern auch ihre Zusammenhänge untereinander von Algorithmen ausgewertet und verstanden werden können. Das semantische Web überführt das WWW von einem »Netz von Dokumenten« in ein »Netz von Informationen«. So bekommen fünfstellige Zahlen z.B. die Zuweisung »Postleitzahl«, wenn sie eine solche sind. Und das Wort »Kohl« würde ggf. in die Kategorie »Gemüse« fallen. Suchmaschinen würden so Informationen »verstehen« und wie das menschliche Gehirn assoziative Querverbindungen zwischen unterschiedlichen Dokumenten herstellen. Entsprechend könnten sie eine Frage wie: »Wer war 2004 deutscher Bundeskanzler?« mit »Gerhard Schröder« beantworten und Links mit weiteren Informationen zur Person des Bundeskanzlers (und nicht einer anderen Person gleichen Namens) angeben. Das Internet würde so zu dem globalen Gehirn, über das schon Visionäre wie H.G. Wells philosophierten, und wäre mehr als nur die Summe seiner Daten.

Drei Techniken, die weit über HTML hinaus gehen, gehören zum semantischen Web: Die HTML-Weiterentwicklung XML (Extensible Markup Language) bildet den syntaktischen Rahmen und definiert die Inhalte von Web-Dokumenten, der URI (Uniform Resource Identifier) bestimmt auf eindeutige Weise den genauen Ort eines Dokuments im Netz, und RDF (Resource Description Framework) führt Bedeutung, Ort und Inhalt zusammen (ausführlich dazu: [2]).

Bisher ist das semantische Web aber nur eine Idee. Eine Reihe von Problemen und Einschränkungen existieren noch, beispielsweise durch Sprachbarrieren: Nicht-englischsprachige Seiten sind entweder nur mit Seiten ihrer Sprache verknüpft oder müssen grundsätzlich in Englisch kategorisiert werden. Doch selbst innerhalb einer Sprache gibt es für viele Dinge unterschiedliche Bezeichnungen. Die Daten werden in ihrer Bedeutung immer in ein mehr oder weniger starres Korsett vorgegebener Kategorien gezwängt werden. Die Feinheiten abzubilden, die die »natürliche Sprache« der Kategorisierung lässt, führt, wenn es nicht sogar unmöglich ist, zu einem riesigen Wust von Meta-Daten.

Vernetzte Vorlieben

Wie ungleich einfacher ist es da, Kategorien nicht zu benennen, sondern gleichartige Dinge unbenannt in »Bedeutungswolken« zusammenzufassen. Der Online-Buchhändler Amazon macht es mit seinen Produktempfehlungen vor: »Kunden, die dieses Buch gekauft haben, haben auch diese Bücher gekauft«. Die bisherigen Einkäufe der Kunden werden hierbei miteinander verglichen und nach größtmöglicher Übereinstimmung gruppiert. Wenn viele Benutzer, die zum Beispiel einen Reiseführer von Berlin gekauft haben, auch einen entsprechenden Stadtplan mit in ihren virtuellen Warenkorb legen, liegt es nahe, dass diese zwei Produkte irgendwie miteinander verwandt sind.

Der Trick bei den so genannten »Gruppenfiltern« ist jedoch, die Zusammenhänge, die Semantik, in die Köpfe der Benutzer auszulagern und indirekt darauf zuzugreifen. Etwas so schwer Greifbares wie »Geschmack« oder »Vorliebe für« lässt sich durch die Auswertung der Bedeutungen, die der Nutzer Objekten zuweist, indem er sie kauft oder eben nicht, und dem Abgleich dieser Daten mit den Daten anderer, auf eine sehr einfache Art beschreiben. Ein Nebeneffekt dieser Methode ist, dass der soziale Druck, etwas nicht gut finden zu dürfen oder umgekehrt etwas gut finden zu müssen, durch die Anonymität des Online-Shoppings nicht zum Tragen kommt. Dadurch wird die Treffgenauigkeit der Empfehlungen nochmals erhöht.

Social Software

Bei Amazon kann sich der Kunde die Produkte empfehlen lassen, die Menschen mit ähnlichen Vorlieben gekauft haben. Wie wäre es, diese Menschen kennen lernen zu können? Als John Hagel III und Arthur G. Armstrong Ende der 90er Jahre ihr Buch »Net Gain« schrieben, galten virtuelle Gemeinschaften (und Portale) als *die* Einnahmequelle der Zukunft im Internet. Heute, nur wenige Jahre später, spricht kaum jemand mehr von solchen sozialen Gemeinschaften. Wir reden lieber von »Social Software«, auch wenn dieser Begriff einen anderen Aspekt in den Mittelpunkt rückt und die Angelegenheit gleichsam entmenschlicht.

Ein Beispiel für eine »Soziale Software«, die auf demselben Prinzip wie Amazons Buchempfehlungen aufbaut, ist das Angebot von Audio-scrobler [3]. Diese Website ermöglicht mittels eines Zusatzprogramms (Plugin) im Audio-Player, die Titel aller gehörten Musikstücke an einen zentralen Server zu übertragen. Dort werden sie mit den Daten der anderen Benutzer verglichen. Die Ergebnisse schlagen sich in verschiedenen Angeboten nieder: Nutzer können

- andere Nutzer, ihre musikalischen Nachbarn, mit ähnlichem Profil finden,
- über das Internetradio last.fm [4] die Musik aus den Profilen mehr oder weniger (sogar das ist einstellbar) ähnlich interessierter Nutzer hören,
- sich Bands und Songs anhand des bekannten Musikgeschmacks empfehlen lassen.

In und unter diesen nicht ausdrücklich kategorisierten musikalischen Nachbarschaften können sich Gruppen bilden, die wiederum mit Künstlern verbunden werden können und einer Kategorie wie »Artists« (Madonna, Miles Davis), »Genres« (Jazz, German Independent) oder »Locality« (London, Australia) untergeordnet sind.

Jeder kennt jeden (irgendwie)

Wodurch Social Software in der letzten Zeit aber eigentlich bekannt wurde, sind Websites wie Orkut [5], Friendster [6] oder Openbc [7]. Bei diesen Angeboten werden Kontakte, Freundschaften oder Bekanntschaften zwischen den Nutzern registriert. Zwischen zwei Personen, die sich nicht kennen, kann so mittels Pfaden, die über ihnen bekannte Personen und deren Bekannte führen, eine Verbindung konstruiert werden. Dadurch wird ersichtlich, über wie viele Ecken jemand mit jemand anderem verbunden ist – auch oder obwohl sie sich noch nie

gesehen haben. Die Soziologie prägte in den 1960er Jahren den Begriff der »Small World«, der genau diesen hohen Grad abkürzender Wege durch persönliche Beziehungen innerhalb sozialer Vernetzungen bezeichnet (ausführlich dazu: [8]).

Eine Anwendung wäre die Anbahnung eines Kontaktes zu einem gewünschten Geschäftspartner, dem man durch den Pfad über gemeinsame Bekannte ja sozusagen schon vorgestellt ist. Zusätzlich zur Angabe der Verbindungen können weitere (mitunter recht intime) Angaben gemacht werden: Diese sind zwar größtenteils nicht zwingend, werden aber häufig bereitwillig ausgefüllt, weil sie den Status innerhalb der Gemeinschaft erhöhen und dadurch leichter zu neuen Kontakten führen können. Wer welche Daten sehen kann, kann je nach System recht differenziert festgelegt werden. Die Daten liegen dabei auf dem Server des Seitenbetreibers, der diese nach Belieben auswerten und beispielsweise für zielgenaue Werbung nutzbar machen kann.

In diesem Zusammenhang ist Googles Inhaberschaft von Orkut kritisch zu sehen. Denn was will der derzeitige Suchmaschinen-Marktführer mit Daten über Menschen und ihre Beziehungen, dem Bereich, »in dem die neuen Techniken zur Erstellung von Wissensdomänen uns am meisten versprechen« (Datenschleuder 83 2004)? Geht es nur um den Besitz einer möglichst großen Benutzerbasis? Die Verknüpfung der unterschiedlichen Bereiche Websuche, Gmail und Newsgroups? Um den Verkauf statistischer Auswertungen auf Grundlage der gesammelten Daten? Kundenbindung? Auf persönliche Interessen zugeschnittene Werbung? Wie ernst wird der Datenschutz genommen?

Wissen verteilen

Der Ruf nach einer dezentralen Speicherung solcher Daten ist verständlich. Eine praktische Umsetzung in diesem Bereich ist das Friend-of-a-Friend-Konzept (FOAF) [9]. Mit diesem System können Internet-Nutzer für Mensch und Maschine lesbar, ganz im Sinne des semantischen Webs, Informationen über sich an ihre Homepage (üblicherweise ein Weblog) »heften«. Darin stehen sowohl beliebig umfangreiche Angaben zur eigenen Person als auch Verweise zu den FOAF-Dateien von Bekannten und Freunden. Eine Art »offenes Orkut«, auf das mittels auf dem eigenen Computer installierter Software, wie bei Instant Messaging Diensten, zugegriffen wird, ist genauso denkbar wie die Betrachtung bestimmter Milieus für soziologische Studien. Die Möglichkeiten festzulegen, wer welche Daten sehen kann, sind naturgemäß eingeschränkt. Eine offene, verteilte Lösung ist eben immer auch öffentlich und nicht nur für

Wohlgesonnene einzusehen. Zwar gibt auch der Betreiber eines Weblogs Vorlieben, Neigungen und persönliche Details preis, aber dass diese durch fehlende Semantisierung nicht unmittelbar maschinenlesbar sind, könnte sich in manchem Fall als Vorteil erweisen.

Fazit

Das semantische Web steht noch ganz am Anfang. Problematisch wie utopisch mag die Hoffnung auf eine Verschlagwortung, das »Tagging« des gesamten Webs scheinen. Interessierte Unternehmen könnten sich eventuell mit ihren Partnern auf eindeutige Definitionen einigen und aus kommerziellem Interesse bereit sein, diese zu pflegen. Die Bedeutung konkreter Begriffe aber auch nur innerhalb einer Sprachgemeinschaft festzulegen, dürfte sich teilweise als unmöglich erweisen. Noch frappierender zeigt sich ein grundsätzliches Problem folgender Art: Das Wort »chat« existiert im Englischen (Chat, Plauderei) und Französischen (Katze). Wer nun nach Objekten sucht, denen das Etikett »chat« angeheftet ist, wird sehr diverse Ergebnisse erhalten. Interkulturelle Klassifizierungsunterschiede bilden ein weiteres Hindernis. »Junge« ist sicher eine adäquate Übersetzung des englischen Wortes »boy«. Allerdings bezweifle ich, dass sich die Extensionen dieser Wörter (also die Objekte, die ihre Bedeutung umfasst) entsprechen. Der Einsatz des semantischen Webs in seiner derzeit geplanten Form wird sich weitgehend auf geschlossene Benutzergruppen beschränken – zumal der Aufwand für das gesamte Web vermutlich nicht lohnt.

Neben der einheitlichen Klassifizierung sind urheber- und datenschutzrechtliche Probleme zu bewältigen: Wie soll die Nutzung von Daten durch Vernetzungsmechanismen abgegolten werden? Werden offene Techniken das Web dominieren oder wird man für jeden Datensatz und jeden Verarbeitungsalgorithmus bezahlen müssen? Wie stellen Urheber sicher, dass ihre Daten nicht gegen ihren Willen in ungewollten Zusammenhängen auftauchen? Was ist mit persönlichen Daten? Denn: Das Internet vergisst nichts. Die Semantisierung von Daten wird dazu führen, dass Lebensläufe lückenloser und umfassender als heute mittels der vielen Datenspuren, die jedeR hinterlässt, nachverfolgt werden können. Auf der anderen Seite wird das semantische Web uns (oder denen, die es sich leisten können) eine bisher ungekannt kohärente Sicht auf die Welt verschaffen und durch die automatisierte Verknüpfung von Daten, die bisher nur mit einigem Aufwand verknüpfbar waren, neue Perspektiven eröffnen – und das in Echtzeit. Wie offen und dezentral Social Software in Zukunft sein wird, hängt vor allem von der einfachen Bedienung der

Programme für die Nutzer und der Einigung auf Standards ab. Mit FOAF ist in diesem Bereich schon ein wichtiger erster Schritt getan. Ob Open Source und Idealismus mit proprietären Lösungen und Profitorientierung konkurrieren können, wird sich erweisen. Die zunehmende Verbreitung offener standardorientierter Programme wie Firefox lässt zumindest hoffen.

Literatur

- Datenschleuder 83 (2004): »Orkut - Sehnt sich die Welt danach, sich selbst zu beschreiben?« [WWW-Dokument: <http://ds.ccc.de/083/orkut> - Zugriff: 26.11.2004].
- Mark Frauenfelder (2004): »Das Unvollendete«. In: *Technology Review*, 11.
- Hagel III, John und Arthur G. Armstrong (1999): *Net Gain*. Niedernhausen: Falken Verlag.

Digitale Verweise

- [@1] www.heise.de/newsticker/meldung/print/48662
- [@2] http://de.wikipedia.org/wiki/Semantisches_Web
- [@3] www.audioscrobbler.com
- [@4] www.last.fm
- [@5] www.orkut.com
- [@6] www.friendster.com
- [@7] www.openbc.com
- [@8] http://de.wikipedia.org/wiki/Small_World
- [@9] www.foaf-project.org