Mit Peer-to-Peer auf Websuche

Internet ohne Google? Unvorstellbar. Aber die beliebteste Suchmaschine hat Tücken, sie liefert nämlich schwammige Ergebnisse. Jetzt möchte eine Handvoll Forscher des deutschen Max-Planck Instituts die Suche revolutionieren. Neue Geschäftsmodelle für Verlage und die Musikindustrie kommen damit in Sicht.

Andrea Salzburger

Surfen ist Googeln, ein unzertrennliches Paar der täglichen Internet-Routine. Wer heutzutage durchs Netz navigiert und dabei interessante Seiten aufstöbern will, kommt ohne die Websuchmaschinen Google, Yahoo & Co kaum aus.

Der Selbstversuch beweist: Man tippt das gesuchte Wort ein. Knapp zwei Zehntelsekunden reichen aus, um etwa 172.000 Einträge für den Begriff „Bananenrepublik" zu finden. Noch schneller geht es mit „Berlin-Marathon": Hier sind es gleich 2.970.000. Bloß, was will man mit 2 Mio. Einträgen, nach einem Stichwort zusammengeschaufelt?

Insgesamt halten mehrere Tausend Google-Server sogar 8 Mrd. Dokumente in petto. Ihr Standing ist beeindruckend: Weltweit 49,2 % aller Anfragen laufen über die 1998 ins Netz gestellte, werbefinanzierte Suchmaschine. Im deutschsprachigen Raum kratzt die Suchmaschine sogar an der 90-Prozent-Marke. Selbst große Konkurrenten wie Yahoo, Microsoft oder Lycos nehmen sich dagegen kleinwüchsig aus.

Suche für Schlaue

Was überhaupt auf den Google-Suchindex kommt, folge bestimmten Kriterien, erklärt Spezialist Matthias Bender vom Saarbrücker Max-Planck-Institut für Informatik. Dieses indexbasierte 08/15-System hält Bender „für furchtbar dumm". Mit speziellen oder komplizierten Anfragen stoße man sehr schnell an die Grenzen. Dennoch ist das bunte „Doppel-O" bislang das populärste Such-Werkzeug. Ob ein passendes Fundstück dabei ist, ist allerdings Glückssache.

„Das geht besser", weiss Bender. Bei der Websuche bedient er sich nämlich lieber eines Peer-to-Peer (P2P)-Netzwerks: Der promovierte Informatiker hat das Prinzip der Tauschbörsen - etwa Napster, Gnutella oder BitTorrent - auf die Web-Suche übertragen. Dazu baute er mit vier Kollegen ein Programm namens „Minerva", das einen Info-Marktplatz für Gleichgesinnte eröffnet. Was sich die Nutzer im Netzwerk teilen sind schlicht Internet-Links. Nicht irgendwelche, sondern „Links, die auf gute Webseiten zeigen."

Denn das Suchnetz macht sich das Wissen der Teilnehmer zunutze, indem es bewertete, bezeichnete oder auch oft angesurfte Seiten zu einem bestimmten Thema nach vorne sortiert. Für Bender der entscheidende Quantensprung: „Erst so wird eine thematische Suche möglich."

Schreiben über Informatik

Das „Informatikerschloss" Dagstuhl liegt an der Quelle der wissenschaftlichen Forschung über Informatik. Zum erstmals heuer dort abgehaltenen Wissenschafts-Journalismus-Seminar vergab MONITOR mittels einer Ausschreibung ein Stipendium an die Tiroler Journalistin Mag. Andrea Salzburger. Im Folgenden schildert sie ihre Eindrücke und beschreibt ein interessantes Peer-to-Peer-Projekt.

Brandheiße Themen wie Robotik, Datenschutz, Computergrafik, Kryptografie, und viele andere werden in Dagstuhl ständig diskutiert. Was läge angesichts dieser geballten Ladung neuester IT-Forschung näher, als junge Fachjournalisten zu einem Schreib-Workshop zusammenzuführen? Nimmt doch so manch bahnbrechendes Forschungsergebnis mangels Übersetzung der komplexen Inhalte erst gar nicht die Hürde der Vermittlung an ein breiteres Publikum. Ein heuer erstmals angebotenes Journalistik-Seminar sollte hier Abhilfe schaffen: Denn die Initiatoren von Dagstuhl forcieren das Kommunikationsbedürfnis zwischen Forschung und Öffentlichkeit. Für den von Winfried Göpfert, Professor für Wissenschaftsjournalismus an der Freien Uni Berlin, und dem Wissenschaftsjournalisten Jörg Göpfert, geleiteten Workshop fanden sich zwischen 9. und 11. Juli also ausnahmsweise Journalisten ein. VertreterInnen aus verschiedenen Wissenschaftsredaktionen - darunter neben dem MONITOR, die FAZ, Financial Times Deutschland, regionale deutsche Tageszeitungen, Internetredaktionen/Portale oder auch Wissenschaftseinrichtungen. Das Spezielle daran: Die intensive Zusammenarbeit mit den Experten vor Ort - angefangen von den Briefings, über die intensive Nachrecherche bis zum Feedback. Während des Artikelschreibens zu einem ausgewählten Themengebiet gaben die Workshop-Leiter Unterstützung. Nachfolgend lesen Sie einen von zwölf im Rahmen des Seminars entstandenen Beiträgen - zur etwas anderen Internet-Suche nach der Peer-to-Peer Logik.

Workshop „Schreiben über Informatik": Journalisten recherchieren direkt an der Quelle der Forschung, die Experten stehen Rede und Antwort. Die Autorin und MONITOR-Stipendiatin Mag. **Andrea Salzburger** steht in der ersten Reihe ganz rechts. (Bild: IBFI Schloss Dagstuhl)

Suchen nach Qualität

Was die P2P-Technik überdies um Meilen weiterbringt, sei „das strukturierte Ausführen von Suchanfragen". Eine optimierte Tabellensuche, im Fachjargon „Hash-Tabelle", hilft dabei. „Jeder Teilnehmer indiziert die lokal besuchten Seiten und erzeugt daraus Statistiken", führt Bender aus. Schließlich würden diese Statistiken, also die Zusammenfassungen, in der Tabelle gespeichert: „Nicht die konkreten URLs", wie er betont. Zusätzlich greift Minerva jedes „Bookmark" (Bezeichnung für Lesezeichen) oder „Blog" (ein Kunstwort aus „Web" und „Logbuch" - für kurze, aktuelle Meldungen) auf, den ein User für seine Lieblingsseiten angelegt hat. Auch „Tags" (Bezeichnungen für Seiten und Files) werden eingelesen, was dann jeweils einem Qualitätsurteil gleichkommt. „Diese Listen tauschen die User unter einander", erklärt Bender. Das beeindruckendste Kriterium heißt damit Treffsicherheit. „Vorausgesetzt, dem Netzwerk gehören möglichst viele Nutzer an!"

Darin liegt allerdings derzeit das Problem: Denn vorerst lässt der große Durchbruch noch auf sich warten. Auch weil P2P fälschlich mit Pirate-to-Pirate gleichgesetzt werde. Dabei beweisen die oftmals illegal benutzten Tauschbörsen, „wie problemlos P2P funktioniert". Um gute Suchergebnisse zu liefern, braucht es aber eine kritische Größe des Netzwerks. Bislang konnten die Forscher allerdings nur Netzwerke mit einigen Hundert Teilnehmern verwirklichen. Das Potenzial ist auf jeden Fall groß: „Das System ist prinzipiell beliebig skalierbar." Dazu weist der Google-Antagonist noch weitere Vorzüge auf: Je größer das Netz, desto stabiler wird es. Zudem ist es hackersicher und gleichberechtigt, von der Grundidee also basisdemokratisch. Damit sei ausgeschlossen, dass China staatskritische Web-Seiten bei Google einfach sperren lässt, meint Bender.

Missing Link

Wenige Kompromisse macht der Informatiker auch, wenn es um sein Selbstverständnis als Forscher geht. Nichts weniger als Grundlagenforscher will er sein, dabei ist er aber kein reiner Theoretiker. Also denkt er insgeheim auch über „Geschäftsmodelle" nach. So könnten von der alternativen Websuche in erster Linie solche Netzwerke profitieren, die darauf hinweisen wollen, „dass sie einen bestimmten Buch- oder Plattentitel besitzen, ohne gleich ihr Urheberrecht preiszugeben." Etwa Buchverlage oder die Musikindustrie, genauso wie italienische Museen, die einmal in akribischer Kleinarbeit archäologische Fundstücke digitalisiert haben. „Der Konsument muss ja von den Daten wissen, bevor er sie kauft." Noch bevor das P2P-Suchprinzip als Marketing- oder Vertriebsinstrument entdeckt wird, eignet es sich schon jetzt für geschlossene thematische Gruppen von einigen Hundert Interessierten. Mit Google baue sich ein Informationsmonopol auf, dessen Macht nicht unterschätzt werden dürfe, warnt Bender. Die „allwissende Weltmaschine" profitiert von einem unermesslichen Datenbestand, der bei den Klicks auf Webseiten anfällt, aber vor allem auch von personenbezogenen Daten, die Kunden freiwillig preisgeben. Dabei geben sich die Kalifornier als die wahrhaft Guten aus, getreu dem Firmenmotto: „Don't be evil."

Auch gegen das allfällige Übel des Datenmonopols bietet die P2P-Suche eine Alternative an: Zumindest technisch hat das neue Werkzeug das Zeug dazu, der befürchteten Umklammerung durch Quasi-Monopolisten wie Google etwas entgegenzuhalten. Experimentierfreudige können den Software-Prototyp unter www.minerva-project.org schon jetzt herunterladen. ❑

Das „Informatikerschloss" Dagstuhl

Die Adresse liest sich gut: Octavieallee und Konrad-Zuse-Straße. In diesem Fadenkreuz aus Historie und Ingenieurskunst liegt, idyllisch platziert am Fuße einer mittelalterlichen Burgruine, das historische Schloss Dagstuhl. Seit 1990 beheimatet das fürstliche Gebäude rund 50 Kilometer nördlich von Saarbrücken das „Internationale Begegnungs- und Forschungszentrum für Informatik", kurz IBFI. Verglichen wird das „Informatikerschloss" gerne mit dem US-amerikanischen IT-Eldorado Palo Alto oder Sophia Antipolis im Süden Frankreichs.

Nicht von ungefähr tummelt sich in Dagstuhl die deutsche, europäische und US-amerikanische Informatik-Forscherelite. Im Vorjahr knackte man sogar einen Besucherrekord: 3.053 Gäste nutzten 2006 die spezielle Infrastruktur wie modern ausgestattete Vortrags- und Arbeitsräume, 65 Wohneinheiten oder mit moderner Hard- und Software bestückte Rechnerräume. Für Inspiration sorgen überdies die 50.649 Bände der hauseigenen Fachbibliothek. Ausgeklügelt ist auch die Einladungspolitik: Spitzenforscher werden in den Seminaren mit verheißungsvollem Nachwuchs zusammengebracht. Der Erfolg kann sich sehen lassen: „Hier gibt es bahnbrechende Diskussionen und Durchbrüche, die durch den offenen wissenschaftlichen Austausch sicherlich begünstigt werden. Das bringt neue Forschungsansätze international weiter", ist der wissenschaftliche Leiter Reinhard Wilhelm überzeugt. Im dicken Gästebuch wird die „ausgezeichnete wissenschaftliche Atmosphäre gelobt", die höchste Konzentration auf die Forschungsbeiträge „for more than twelve hours a day" erlaube.

So gingen bereits zahlreiche Publikationen aus den meist einwöchigen oder auch längeren Aufenthalten im „Informatiker-Atelier" hervor. Wobei die Seminarthemen gemäß der Konzeption des IBFI stark interdisziplinär ausgerichtet sind. Meilensteine setzten dabei Themen wie Software-Engineering, Künstliche Intelligenz oder Netzwerke wie Peer-to-Peer-Anwendungen. Dazu kommen Dauerbrenner aus Bioinformatik und Web-Semantik.