

Jedermanns Suchmaschine

Neuartige Software könnte in Zukunft Internet-Suchmaschinen wie Google und Yahoo Konkurrenz machen. Systeme wie „Yacy“ nutzen das Prinzip der unkontrollierbaren Tauschbörsen für Musik und Videos. Die Idee dahinter: Verteilte Datenbanken unter allen Teilnehmern enthalten einen stetig wachsenden Index über die Seiten des Internets – unabhängig von Konzernen.



Kopf hinter dem Yacy-Netzwerk: Michael Christen aus Frankfurt. Er hofft, dass sich neue Nutzer und Entwickler anschließen. Bisher entwickelte der Informatiker Yacy zum Großteil allein.

Die bisher nur von wenigen Computereeks getestete kostenlose und von keinem Konzern gesteuerte Suchmaschine „Yacy“ wird soeben erst geboren. Die Idee basiert auf dem Open-Source-Gedanken und stammt aus Deutschland. Open Source heißt: Jeder kann daran mitarbeiten und hat Zugriff auf den Programmcode. Erste Tests und Beurteilungen von Experten lassen darauf schließen, dass das System möglicherweise das Zeug dazu hat, sich zu einem Kandidaten für eine Alternative zu Google & Co. zu entwickeln.

Das Prinzip ähnelt dem der bekanntesten Internet-Suchmaschine Google. Google beherbergt seine Informationen auf mehreren hundert Servern. Ein Teil der Maschinen klappert eigenständig in Sekunden-schnelle Hunderttausende von Internet-Seiten ab, folgt den darauf enthaltenen Links und verzeichnet die vorgefundenen Daten in einem riesigen Inhaltsverzeichnis. Der Anwender bekommt beim Aufruf von www.google.de von einem der Google-Server immer die gleiche, denkbar einfach zu bedienende Suchmaske ausgeliefert.

Die gerade entstehende neue Suchmaschine Yacy lagert ebenfalls auf mehreren Servern – allerdings nicht auf denen einer Firma, sondern auf den PCs der Anwender. Wer sich die kostenlose Software von <http://freshmeat.net/projects/yacyproxy> installiert, macht seinen Computer für die Dauer seines Surfens im Internet zum Server. Die Software protokolliert auf Wunsch fast jede Seite mit, die der Nutzer aufruft – mit Ausnahme jener, für die ein Kennwort nötig ist. Indexiert werden also ausschließlich öffentlich zugängliche Daten. Wer möchte, kann auch einen eigenen „Crawler“ auf den Weg schicken. Der klappert dann von einer beliebig einstellbaren Internet-Seite aus alle Links ab, durchsucht die darin vorgefundenen Informationen in einer einstellbaren Tiefe und verschlagwortet sie. Das entstehende Inhaltsverzeichnis (der Index) ist am Privat-PC durchsuchbar – so wird nach der Eingabe der besonderen Adresse <http://localhost:8080> eine Suchmaschine auf dem eigenen PC aufgerufen.

Das wäre an sich noch nichts Besonderes, und die Suchmaske hat im derzeitigen Entwicklungsstadium auch noch keinen herausragenden Komfort. Eine weltweite unabhängige, von keinem Konzern kontrollierte Suchmaschine dürfte aber dadurch entstehen, dass die Nutzer ihre privaten Indizes anonym freigeben und unter-

einander austauschen können. Und das erlaubt Yacy schon in der jetzt verfügbaren Vorversion 0.3 filigran.

Mit den richtigen Einstellungen bekommt man an seinem privaten PC über die persönliche Suchmaschine auch Zugriff auf die Indizes anderer Internet-Surfer. Man kann auf Grund besonderer Mechanismen nicht erkennen, wer im einzelnen welche Seiten besucht hat – wohl aber, was die Seiten gezeigt haben. Umgekehrt erhalten andere Surfer Zugriff auf die Indizes selbst besuchter Seiten. Und mit Hilfe einer Sperre werden nur die Seiten verteilt, die keine persönlichen Informationen enthalten, wie etwa Nutzerkennungen und Passwörter.

Der Austausch dieser Informationen geschieht im Hintergrund direkt zwischen den Yacy-Teilnehmern. Es gibt keinen zentralen Server. Lediglich beim ersten Einstieg in das Yacy-Netz werden ein paar altbekannte zentrale Server kontaktiert, unter anderem die der Entwickler. Diese übermitteln zunächst nur die Adressen anderer gerade aktiver Yacy-Surfer. Und beim nächsten Aufruf werden nicht mehr die Server der Entwickler befragt, sondern die bekannten benachbarten PCs anderer Surfer von der letzten Sitzung.

Nach und nach entsteht so ein mächtiger dezentraler Datenverbund von Informationen über die besuchten Webseiten aller Yacy-Teilnehmer. Je mehr Leute daran teilnehmen, desto besser wird das Inhaltsverzeichnis.

„Noch gibt es ein Henne-Ei-Problem“, sagt Michael Christen, der Kopf hinter Yacy, ein Diplom-Informatiker aus Frankfurt/Main. Mit derzeit noch wenig Teilnehmern ist der gemeinsame Index klein, und bei einem solch kleinen Index gebe es noch wenig passende Treffer bei der Stichwortsuche und daher auch zu geringe Akzeptanz. Ohnehin ist gerade erst die Version 0.3 der Software fertig geworden, eine Version für Auskenner. Die Fassung erlaubt bereits eine zuverlässige Suchen-Funktion über die selbst aufgerufenen Seiten, den Austausch mit den Indizes anderer Internet-Surfer und erstmals auch, einen eigenen „Crawler“ auf den Weg zu schicken.

Wie die Erfahrungen mit Tauschbörsen wie Kazaa eDonkey gezeigt haben, ist solch ein Tauschverbund kaum mehr zu stoppen, wenn er einmal eine bestimmte Schwelle von mehreren hundert Teilnehmern in aller Welt erreicht hat. Irgendwo sind immer gerade genügend Leute online, die einen Teil des Index bereitstellen. So haben die Musik- und die Filmindustrie ihre Mühe, das häufig illegale Treiben in dem dezentralen Netzwerk Kazaa zu stoppen.

Bei den Informationen, die das Yacy-Netzwerk sammelt, geht es dagegen nicht um kommerzielle Medien, sondern um Textinhalte von Internet-Seiten. Ähnlich wie Google & Co. ihre gesammelten Informationen in Datenbanken hinterlegen, agiert das verteilte Yacy-Netz.

Dabei hat Yacy durchaus hehre Motive: die Informationsfreiheit und die Unabhängigkeit

von herausragend bedeutenden Suchmaschinen. Es soll unter anderem zur alternativen Auffindstation für zensierte Inhalte werden. „Die Nutzer von Yacy dürften idealistisch orientiert sein“, sagt Michael Christen. „Dabei geht es um Meinungsfreiheit, Individualität und der Verfügbarkeit von Wissen und Information.“ Ein Beispiel: Eine zentrale Suchmaschine wie Google könnte derzeit durchaus bestimmte Internet-Seiten mit missliebigen Informationen herausfiltern – so wie nach einem Bericht von Heise.de kürzlich bei Google China geschehen. Im Yacy-Netz wäre das nicht möglich: Sobald einer der Teilnehmer die Seite besucht und freigegeben hat, wandern die Informationen in den gemeinsamen Index und werden umgehend auf andere Teilnehmer verteilt.

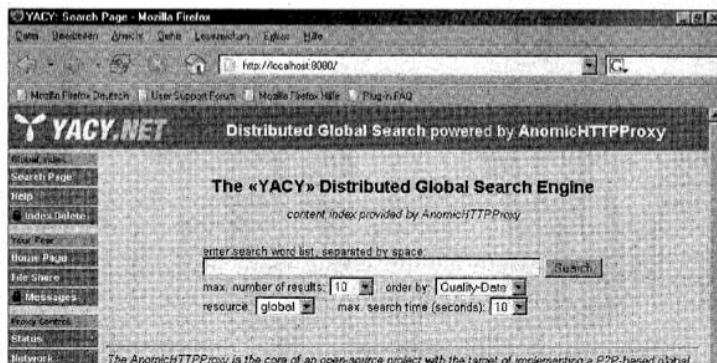
Experten halten solch ein Modell für Suchmaschinen für einen sinnvollen Ansatz. Bereits seit dem Jahr 2000 habe es solche Versuche gegeben, berichtet Wolfgang Sander-Beuermann, Geschäftsführer des SuMa-eV, einem Verein für die Entwicklung alternativer Suchmaschinen. „Das Yacy-Projekt ist am weitesten fortgeschritten“, urteilt der Professor. An der Universität Hannover betreibt er die Suchmaschine Metager.de: eine der letzten Suchmaschinen europaweit, die den US-Konkurrenten Google, Yahoo und MSN nennenswerte Zugriffszahlen abtrotzt.

Dass ein solches verteiltes System für Informationen aus dem Internet konzernunabhängig betrieben werden kann, belegt nach seinen Worten das Usenet. Das war vor allem in den achtziger und neunziger Jahren ein an Universitäten betriebener, dezentraler Nachrichtenaustauschdienst. Allerdings hat es inzwischen an Bedeutung verloren. „Dem Otto-Normal-User erscheint Google immer noch gut genug. Das, was er mit Google nicht findet, sieht er ja auch nicht, und er kann sich dann auch nicht darüber ärgern, dass er es nicht findet“, sagt Sander-Beuermann.

Die wichtigste Eingabemaske zum Recherchieren im Usenet www.dejanews.com, hat 2001 ein amerikanischer Konzern gekauft. Seither können Nutzer zwar weiterhin alternative Programme installieren und so unabhängig von diesem Konzern im Usenet surfen. Komfortabler ist aber die Suchmaske des Konzerns.

Ironie der Geschichte: Dieser amerikanische Konzern war Google. Mit Yacy wäre das wohl nicht möglich.

MARCUS SCHWARZE



Die Suchmaschine Yacy liegt auf der eigenen Festplatte und ist im Internet-Browser unter der Adresse <http://localhost:8080> erreichbar. Ihre Daten bezieht sie aus den eigenen Surfsitzungen und denen anderer Nutzer.