

Wie Suchmaschinen von Social Software profitieren

Dirk Lewandowski

Hochschule für Angewandte Wissenschaften Hamburg
Fakultät DMI, Department Information
Berliner Tor 5
20249 Hamburg
`dirk.lewandowski@haw-hamburg.de`

Abstract. Dieser Beitrag geht der Frage nach, in welcher Form Web-Suchmaschinen Daten aus Social Software ausnutzen (können), um ihre Ergebnisse zu verbessern. Dabei wird zum einen gezeigt, dass Suchmaschinen bereits in der Vergangenheit implizite, von der Nutzermasse erstellte Daten ausgewertet haben. Zunehmend werden nun explizite Daten verwendet. Dieser Beitrag versucht eine Systematisierung dieser Datenverwendung und stellt die Verbesserungsmöglichkeiten für Suchmaschinen anhand ihrer typischen Komponenten dar.

Keywords. Social Software, Suchmaschinen, Content Acquisition, Indexierung, Qualitätsbewertung, Trefferpräsentation

1 Einleitung

Die Suche im Netz mit Hilfe von Suchmaschinen ist neben E-Mail die meistgenutzte Anwendung im Internet [1]. Pro Monat werden nach Angaben von ComScore allein in Deutschland etwa 5,6 Milliarden Suchanfragen an die allgemeinen Suchmaschinen (d.h. Universalsuchmaschinen) gestellt¹.

Die Recherche über Suchmaschinen ist damit *der* Zugang zu den Informationen im Netz; andere Ansätze (wie Web-Kataloge) haben ihre Bedeutung verloren. Alternative Ansätze wie beispielsweise Social-Bookmarking-Dienste konnten sich als alternativer Zugang zu den Informationen im Netz nicht in der Masse etablieren.

Aktuell wird die Entwicklung der Web-Suche hin zu einer (verstärkten) Sozialen Suche diskutiert. Das Ziel dieses Aufsatzes ist es, zu zeigen, wie Suchmaschinen an verschiedenen Stellen bereits auf die Mitarbeit ihrer Nutzer setzen und welche (impliziten) Nutzerdaten sie sich zunutze machen (können). Die Ausgangsthese ist dabei, dass die Suchmaschinen schon seit langem *implizite* Nutzerbewertungen verwenden, nun jedoch auch vermehrt *explizite* Nutzerbewertungen einsetzen. Letzteres dürfte in Zukunft noch weiter anwachsen.

¹http://comscore.com/Press_Events/Press_Releases/2010/1/Global_Search_Market_Grows_46_Percent_in_2009

Implizite Mitarbeit der Nutzer meint hierbei, dass Daten, die von einer Masse von Nutzern zu einem anderen Zweck als der Zuarbeit zu einer Suchmaschine generiert wurden, für die Verbesserung einer Suchmaschine eingesetzt werden. Ein Beispiel hierbei ist das einfache Setzen eines Links, der später von Suchmaschinen zur Qualitätsbewertung des referenzierten Dokuments eingesetzt wird.

Mit expliziter Mitarbeit ist dagegen gemeint, dass Nutzer Dokumente beschreiben und/oder bewerten in dem Bewusstsein, dass diese Informationen für die Verbesserung und/oder Ergänzung eines Web-Angebots, also beispielsweise einer Suchmaschine eingesetzt werden.

2 Implizite Mitarbeit der Suchmaschinennutzer

In der Diskussion um die Mitwirkung der allgemeinen Nutzerschaft bei der expliziten Verbesserung der Suchmaschinen bzw. ihrer Ergebnisse wird oft vergessen, dass sich Suchmaschinen auch in der Vergangenheit schon die Mitarbeit ihrer Nutzer zunutze gemacht haben, wobei dies aber im wesentlichen implizit geschah, d.h. dem Nutzer war bei seiner Aktivität nicht bewusst, dass diese der Verbesserung der Suchmaschine dient.

Streng genommen ist ja schon der Aufbau der Suchmaschinen-Datenbank ohne den Nutzer nicht zu denken. Die verteilt vorliegenden Dokumente des WWW werden von unterschiedlichsten Nutzern mit den unterschiedlichsten Intentionen erstellt. Deutlicher noch wird die Mitwirkung der Nutzer allerdings bei den Verfahren der Popularitätsmessung, worunter hier sowohl die linktopologischen als auch die Nutzungsstatistischen Verfahren verstanden werden.

Linktopologische Rankingverfahren machen sich die Verlinkungsstruktur des Web zunutze und bewerten Dokumente anhand ihrer Linkpopularität. Das Kollektiv der Nutzer (hier: derjenigen, die Websites selbst erstellen, bzw. genauer: diejenigen, die Links setzen) gibt durch das Setzen der Links Stimmen für bestimmte Dokumente ab. Die Suchmaschinen gewichten diese Links nach ihrer Herkunft und versuchen so, über ein gewichtetes Popularitätsmodell Aussagen über die Qualität bestimmter Dokumente zu erlangen (vgl. [2], Kapitel 8).

Nutzungsstatistische Verfahren hingegen beziehen alle aktiven Nutzer des Web ein, ob diese nun selbst aktiv handeln oder aber nur Inhalte konsumieren. Hier gilt das Anklicken von Dokumenten, kombiniert mit der Verweildauer auf diesen Dokumenten, als Indikator für ihre Qualität (vgl. [2], Kapitel 7). Jeder Klick bedeutet also eine Zuarbeit für die Suchmaschine, die diesen Klick erfassen kann. Gängige Methoden sind Toolbars, die die Nutzeraktivität protokollieren; unabhängig davon, ob der Nutzer sich nun auf dem Angebot der Suchmaschine befindet oder eine von der Suchmaschine unabhängige Seite besucht.

3 Explizite Mitarbeit der Suchmaschinennutzer

Während die implizite Mitarbeit den meisten Nutzern nicht bewusst ist und auch nicht in erster Linie der Verbesserung der Suchergebnisse dient, ist bei der expliziten Mitarbeit zwischen der innerhalb der Suchmaschinen selbst und der in Social Software, die wiederum von den Suchmaschinen ausgenutzt wird, zu unterscheiden. Ein Nutzer beispielsweise, der ein Produkt auf einer Shoppingplattform bewertet, tut dies explizit, dürfte sich aber kaum bewusst sein, dass ebendiese Bewertung wiederum zur Bewertung durch eine Suchmaschine eingesetzt werden kann.

Die explizite Mitarbeit im Angebot einer Suchmaschine ist deutlich bei Google vertreten; über die Dienste SearchWiki und SideWiki wurde in der Presse ausführlich berichtet. In diesen Diensten hat jeder (angemeldete) Nutzer uneingeschränkt die Möglichkeit, Suchergebnisse zu kommentieren und zu bewerten.

SearchWiki wird automatisch innerhalb jeder Trefferliste angeboten, sofern ein Nutzer in einen Google-Dienst eingeloggt ist. Zu jedem Eintrag in den Trefferlisten besteht nun die Möglichkeit, den Treffer an den Beginn der Trefferliste zu stellen, ihn aus der persönlichen Trefferliste zu löschen und den Treffer zu kommentieren. Kommentare sind für alle anderen Nutzer sichtbar; die veränderte Reihung der Trefferliste aber nur für den betreffenden Nutzer selbst. Allerdings bleibt diese Reihung auch bei zukünftigen Suchen nach den gleichen Suchbegriffen bestehen.

SideWiki erweitert SearchWiki in der Hinsicht, dass Nutzer der Google-Toolbar diesen Dienst für jede Seite im Web hinzuschalten können. Sie bekommen dann die Kommentare der anderen Nutzer direkt mit dem aufgerufenen Dokument zusammen angezeigt. Welche Implikationen sich daraus ergeben, verdeutlicht das Beispiel der Homepage von Microsoft, welche durch (vor allem negative) Nutzerkommentare ergänzt wird (s. Abb. 1).

Auch Nutzerkommentare aus fremden Diensten können von den Suchmaschinen ausgewertet werden. Hier ist zu unterscheiden zwischen einer direkten Einbindung und der Aggregation von Beschreibungen oder Bewertungen aus fremden Diensten.

In einer direkten Einbindung werden beispielsweise Produktbewertungen aus einem Vergleichsportal direkt in eine Trefferbeschreibung integriert (Abb. 2). Die Suchmaschine macht sich hier eine bereits auf der ausgewerteten Webseite vorhandene Bewertung bzw. einer dort vorhandenen Aggregation von Bewertungen zunutze.

Im Gegensatz dazu steht die Aggregation von Beschreibungen und/oder Bewertungen durch die Suchmaschine selbst. So können analog dem genannten Beispiel bei dieser Form der Integration die Bewertungen aus unterschiedlichen Portalen zusammengeführt und neu gewichtet werden. Die Suchmaschine bietet hier einen Mehrwert einerseits durch die Aggregation selbst, andererseits durch die Gewichtung selbst, die als besonders vertrauenswürdig eingeschätzte Websites in der Gewichtung bevorzugen kann. Eine solche gewichtende Aggregation wurde bereits bei der US-Version der Suchmaschine Bing realisiert.

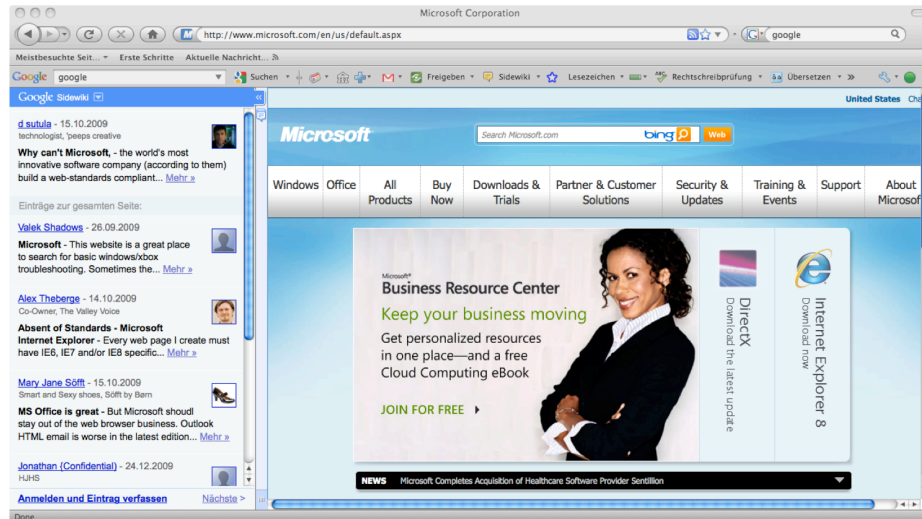


Abb. 1: Nutzerkommentare als Ergänzung der Webseitenpräsentation bei Googles SideWiki

[iRobot Roomba Discovery Robots and robot kit reviews - CNET Reviews](#) -
 ★★★★★ Bewertung von Aimee Baldridge - 19. Jul 2004 - Preisspanne: \$259.00
 CNET's comprehensive iRobot **Roomba** Discovery coverage includes unbiased reviews, exclusive video footage and Robots and robot kit buying guides.
[reviews.cnet.com/...roomba.../4505-3510_7-30975237.html](#) - [Im Cache](#) - [Ähnlich](#)

Abb. 2: Integration von Produktbewertungen in eine Trefferbeschreibung

Während also bei der direkten Einbindung der Bewertungen für den Nutzer offensichtlich ist, woher die Daten kommen (und damit eingeschränkt auch, wer die jeweiligen Bewertungen verfasst hat), kann die Aggregation fremder Bewertungen durch die Suchmaschine selbst zwar im Idealfall zu einer verlässlicheren Bewertung führen, bringt jedoch auch einen – weiteren – Verlust an Transparenz mit sich.

4 Ausnutzung von Daten aus Social Software in verschiedenen Suchmaschinen-Komponenten

In diesem Abschnitt soll die Integration von Daten aus Social Software auf die unterschiedlichen Komponenten von Suchmaschinen bezogen werden. Dabei können sich Suchmaschinen in allen ihren Komponenten dieser Daten bedienen, um Inhalte besser auffinden, bewerten und präsentieren zu können.

Als wesentliche Komponenten einer Suchmaschine werden hier die folgenden angesehen:

- Content Acquisition, also das Hinzufügen von Inhalten zum Datenbestand der Suchmaschine. Dieses erfolgt in erster Linie durch das Crawling, kann jedoch durch Daten anderer Herkunft (bspw. Feeds) ergänzt werden.
- Indexierung, also die Aufbereitung (und ggf. Anreicherung) der gefundenen Daten, um sie durchsuchbar zu machen.
- Qualitätsbewertung.
- Trefferpräsentation.

Im Bereich der Content Acquisition können Daten aus Social Software zum Auffinden neuer Links eingesetzt werden, zum Beispiel indem neu auftauchende Links bei Facebook oder Twitter direkt erfasst werden. Die Suchmaschine kann sich hiermit einen (Index-)Aktualitätsvorteil schaffen, wenn über die Social Software Links verbreitet werden, die der Suchmaschine (noch) nicht bekannt sind. Weiterhin können, wenn die Suchmaschinen selbst proprietäre Social Software betreiben, die dort erstellten Daten direkt in den Suchindex integriert werden. Hier verschafft sich die betreffende Suchmaschine nicht nur einen Aktualitätsvorteil, sondern auch einen direkten Vorteil hinsichtlich der Informationen, welche andere Suchmaschinen nicht bieten können.

Im Bereich der Indexierung können Daten aus Social Software beispielsweise zur Beschreibung von nicht-textuellen Inhalten verwendet werden. So verwendet Yahoo für seine Bildersuche tags aus Flickr.

Im Bereich der Qualitätsbewertung kann schon das Vorhandensein einer URL in einem Social-Bookmarking-Dienst oder einer anderen Social Software als Qualitätskriterium verwendet werden. Bedeutender jedoch sind Tag-Häufigkeiten, die Anzahl von Kommentaren zu einem Beitrag oder die Linkhäufigkeit bei Blogbeiträgen. Alle diese Kriterien können zusätzlich auf einen bestimmten Zeitraum bezogen werden.

Im Bereich der Trefferpräsentation schließlich können mit Hilfe der Daten aus Social Software erweiterte Trefferbeschreibungen generiert werden. In erster Linie sind hier Bewertungen zu nennen (wie oben gezeigt), aber auch Kommentare oder von Nutzern erstellte Bilder können hier verwendet werden

5 Fazit

Social Software ist bei den Suchmaschinen „angekommen“. Zwar haben sie auch schon in der Vergangenheit auf die (vor allem implizite) Mitarbeit ihrer Nutzer gesetzt, durch Social Software bietet sich ihnen aber ein noch weit größerer Spielraum. Vor allem die Einbindung expliziter Nutzerbewertungen, ob diese nun bei den Suchmaschinen selbst oder auf der Website eines anderen Anbieters abgegeben wurden, wird momentan stark ausgebaut. Für die Zukunft ist vor allem eine stärkere Aggregation von user-generated content durch die Suchmaschinen zu erwarten, wodurch sich diese weiter von der klassischen Darstellungsform der „zehn blauen Links“ abwenden und sich weiter in Richtung entscheidungsunterstützender Maschinen wandeln werden.

Literatur

1. van Eimeren, B., Frees, B.: Der Internetnutzer 2009 - multimedial und total vernetzt?: Ergebnisse der ARD/ZDF-Onlinestudie 2009. Media Perspektiven (2009) 334-348
2. Lewandowski, D.: Web Information Retrieval: Technologien zur Informationssuche im Internet. DGI, Frankfurt am Main (2005)