

# Zur Bewertung der Qualität von Suchmaschinen

Dr. Dirk Lewandowski, Heinrich-Heine-Universität Düsseldorf, Informationswissenschaft

erscheint im Tagungsband zur Fachkonferenz „Suchen und Finden im Internet“ des Münchner Kreis (1.2.2006, München).

Nahezu jede Betrachtung des Suchmaschinen-Markts endet in der Forderung nach mehr Transparenz und vor allem nach mehr Konkurrenz auf diesem Markt. Nutzer könnten sich nur umfassend informieren, wenn ihnen ein möglichst großes Quellenspektrum (d.h. möglichst viele hochwertige Suchmaschinen) zur Verfügung stünde, das auch tatsächlich genutzt werden würde. Tatsächlich verlassen sich die Nutzer aber zu einem großen Teil auf Google als alleinige Suchmaschine, seltener werden Yahoo und MSN verwendet, kleinere Anbieter mit eigener Technologie spielen nur eine untergeordnete Rolle [Lewandowski, 2005], S. 21ff.

Während Forderungen nach der Regulierung des Markts oder der Schaffung neuer, „alternativer“ Suchmaschinen leicht geäußert werden, ist noch ziemlich unklar, wie gut – oder wie schlecht – die bisher existierenden Suchmaschinen tatsächlich sind. Wären sie alle schlecht, so ergäbe sich tatsächlich eine Notwendigkeit alternativer Suchmaschinen. Wären sie allerdings alle qualitativ hochwertig, so dürfte es sich wohl nur noch um die Frage nach einer neuen Verteilung der Nutzeranfragen auf die unterschiedlichen Suchmaschinen handeln.

Letztlich geht es um ganz unterschiedliche Fragen, von denen hier drei beispielhaft genannt werden sollen:

- Wie gut sind die ausgegebenen Ergebnisse der unterschiedlichen Suchmaschinen?
- Werden (was als Verdacht von Nutzerseite öfters geäußert wird) kommerzielle Treffer bevorzugt?
- Liefern eigentlich alle Suchmaschinen ähnliche Ergebnisse oder stellt ein Wechsel der Suchmaschine auch immer einen „Perspektivenwechsel“ dar?

Um auf solche Fragen eine Antwort zu geben, müssen die bestehenden Suchmaschinen untersucht werden. Erst danach kann entschieden werden, ob tatsächlich neue, alternative Suchmaschinen vonnöten sind oder ob mit den bestehenden bereits ausreichend Alternativen vorhanden sind, die jedoch von den Nutzern schlicht nicht angenommen werden. Und letztlich kann auf der Basis entsprechender Untersuchungen qualifizierter über die Notwendigkeit nationaler bzw. europäischer Suchmaschinen diskutiert werden.

Aber wie kann nun die Qualität bestehender Suchmaschinen bewertet werden? Es handelt sich um drei Qualitätsebenen, die untersucht werden müssen: Die Qualität der Treffer, die Qualität der Datenbestände (Indizes) sowie die Qualität der Recherche.

## **Trefferqualität**

In der Regel wird die Güte von Suchmaschinen (bzw. von Information-Retrieval-Systemen allgemein) durch Relevanztests bewertet [Lewandowski, 2005], S. 139ff. Die Treffer, die auf Testanfragen hin zurückgegeben werden, werden einzeln bewertet, meist nach einer einfachen Unterscheidung nach „relevant“ und „nicht relevant“. Solche Tests ergaben in der Vergangenheit für die verschiedenen Suchmaschinen zwar unterschiedliche Werte (cf. [Griesbaum, 2004]), in der Tendenz ist jedoch eine Annäherung (zumindest der großen internationalen) Suchmaschinen zu erkennen. Für die aktuelle Suchmaschinenlandschaft mit den zurzeit wichtigsten Anbietern Google, Yahoo und MSN liegt allerdings kein solcher Test vor.

Selbst wenn zwei Suchmaschinen in Retrievaltests die gleiche Qualität erreichen würden, müsste es nicht sinnlos sein, in beiden zu recherchieren. Trotz gleicher Qualität könnten die beiden Systeme vollkommen unterschiedliche Treffer liefern. Auch zu dieser Frage, die sich nur durch

die Bestimmung des Überschneidungsgrads zwischen den einzelnen Suchmaschinen beantworten lässt, liegen keine aktuellen Studien vor.

### ***Index-Qualität***

Neben der beschriebenen Qualität der Treffer geht es um die Qualität des Index: Hier sind vor allem die Vollständigkeit, die Aktualität und die gleichmäßige Abdeckung (vor allem auf Sprach- und Länderebene) zu nennen.

Zur Vollständigkeit der Datenbestände der Suchmaschinen liegen nur inzwischen veraltete Untersuchungen vor [Lawrence undGiles, 1998;Lawrence undGiles, 1999], insbesondere fehlen Studien zur tatsächlichen Größe des World Wide Web, anhand derer die Abdeckung durch die Suchmaschinen gemessen werden könnte.

Zur Aktualität der Suchmaschinen-Indizes liegt mit [Lewandowski, et al., 2006] eine aktuelle Untersuchung vor, die zeigen kann, dass keine der momentan wichtigsten Suchmaschinen (Google, Yahoo, MSN) in der Lage ist, die Seiten aus der Testkollektion täglich in ihrem Index zu aktualisieren. Zwischen den Suchmaschinen bestehen große Unterschiede: Während Google mit dem höchsten Anteil der tagesaktuell indexierten Seiten vorne liegt, gelingt es MSN offensichtlich am schnellsten, seinen Gesamtindex aufzufrischen. Weitere Untersuchungen speziell zu Seiten mit niedriger Aktualisierungsfrequenz stehen noch aus.

Eine gleichmäßige Abdeckung des Web innerhalb des Crawls der Suchmaschinen ist vor allem hinsichtlich der Abdeckung unterschiedlicher Länder und Sprachräume von Bedeutung. Dass es den Suchmaschinen nicht gelingt, unterschiedliche Länder auf gleich hohem Niveau abzudecken, können [Vaughan undThelwall, 2004] anhand des Vergleichs der Abdeckung der Länder USA, China, Singapur und Taiwan belegen. Wesentlich mehr US-Sites werden erfasst und diese werden auch deutlich tiefer indexiert als die aus anderen Ländern. Vaughan und Thelwall vermuten dafür Gründe, die in der Struktur des Web und seines bisherigen Wachstums liegen. Für Seiten aus Deutschland bzw. deutschsprachige Dokumente liegen bisher keine ähnlichen Untersuchungen vor, es ist jedoch anzunehmen, dass auch diese Websites nicht in ähnlichem Umfang abgedeckt werden wie die US-Sites.

### ***Recherche-Qualität***

Aber auch die Qualität der Recherche spielt eine Rolle. Zwar wissen wir aus dem Nutzerverhalten, dass der „allgemeine“, „durchschnittliche“ Nutzer seine einfachen Anfragen ohne großen kognitiven Aufwand formuliert [Hölscher, 2002;Machill, et al., 2003]. Für fortgeschrittene Nutzer sind die angebotenen Recherchemöglichkeiten (und ihre Funktionstüchtigkeit) aber durchaus von Bedeutung [Lewandowski, 2004]. Hier zeigen sich große Unterschiede zwischen den Suchmaschinen, allerdings wurden bisher nur einzelne Funktionen untersucht [Lewandowski, 2004].

Alle Nutzergruppen wiederum betreffen Fragen der Benutzerführung. Dabei handelt es sich um ein Thema, das in der Suchmaschinen-Forschung bisher nur wenig beachtet wurde – insbesondere in Hinblick auf die weitere Entwicklung hin zu einer „Personalisierung“ der Suchmaschinen dürfte die Benutzerführung eine zentrale Stellung einnehmen, während die klassische Bewertung der Retrievaleffektivität an Bedeutung verlieren dürfte.

### ***Einschränkung der Qualitätsbewertung***

Die beschriebenen Faktoren können allerdings nicht die tatsächliche inhaltliche Qualität der Treffer messen, sondern höchstens, in welchem Maß die Nutzer selbst die gefundenen Treffer als hochwertig ansehen. Letztlich lässt sich dieses Problem aber auch nicht lösen, da nicht die (mit Einschränkungen objektiv bestimmbare) Relevanz für die Befriedigung eines Informationsbedürfnisses entscheidend ist, sondern die vom Nutzer empfundene Pertinenz [Lancaster undGale, 2003].

## **Handlungsbedarf**

Die beschriebene Situation zeigt vor allem, dass es bisher an Erkenntnissen über die Qualität der bestehenden Suchmaschinen mangelt. Es scheint einfacher zu sein, Forderungen nach einem Gegenpol auf dem Suchmaschinen-Markt oder nach einer Regulierung der Suchmaschinen zu stellen, als zu erforschen, welche Probleme überhaupt tatsächlich bestehen. Dies ist aber essentiell für das Verständnis vorhandener sowie entstehender Problematiken.

Als Beispiel mag die Frage dienen, ob explizit „deutsche“ Suchmaschinen eigentlich sinnvoll sind und gebraucht werden. Die internationalen Anbieter behaupten, den deutschen Sprachraum ebenso gut abzudecken wie alle anderen Sprachräume auch, aus vergleichenden Untersuchungen zwischen den USA und anderen Ländern können wir aber vermuten, dass dies nicht der Fall ist. Eine Untersuchung dieser Problematik speziell für Deutschland bzw. den deutschen Sprachraum steht allerdings aus.

## **Literatur**

- Griesbaum, Joachim (2004): Evaluation of three German search engines: Altavista.de, Google.de and Lycos.de, Information Research (Band 9), Nr. 4. URL: <http://informationr.net/ir/9-4/paper189.html>
- Hölscher, Christoph (2002): Die Rolle des Wissens im Internet. Gezielt suchen und kompetent auswählen., Klett-Cotta, Stuttgart.
- Lancaster, F.W. und Gale, V. (2003): Pertinence and Relevance, Drake, M.A., Encyclopedia of Library and Information Science Seite 2307-2316, Dekker, New York.
- Lawrence, Steve und Giles, C.L. (1998): Searching the World Wide Web, Science (Band 280), Seite 98-100.
- Lawrence, Steve und Giles, C.L. (1999): Accessibility of Information on the web, Nature (Band 400), Nr. 8, Seite 107-109.
- Lewandowski, Dirk (2004): Abfragesprachen und erweiterte Suchfunktionen von WWW-Suchmaschinen, Information Wissenschaft und Praxis (Band 55), Nr. 2, Seite 97-102.
- Lewandowski, Dirk (2004): Date-restricted queries in web search engines, Online Information Review (Band 28), Nr. 6, Seite 420-427.
- Lewandowski, Dirk (2005): Web Information Retrieval: Technologien zur Informationssuche im Internet, DGI, Frankfurt am Main.
- Lewandowski, Dirk; Wahlig, Henry und Meyer-Bautor, Gunnar (2006): The Freshness of Web search engine databases, Journal of Information Science (Band 32), Nr. 2, Seite 133-150.
- Machill, Marcel; Neuberger, Christoph; Schweiger, Wolfgang und Wirth, Werner (2003): Wegweiser im Netz: Qualität und Nutzung von Suchmaschinen, Machill, Marcel und Welp, Carsten, Wegweiser im Netz, Bertelsmann Stiftung, Gütersloh.
- Vaughan, L. und Thelwall, M. (2004): Search Engine Coverage Bias: Evidence and Possible Causes, Information Processing & Management (Band 40), Nr. 4, Seite 693-707.