



Internet-Recherchen – Was bessere Suchergebnisse verhindert

Dr. Torsten Beyer

Dr. Beyer Internet-Beratung

Bis in die frühen 1990er Jahre war es noch möglich, für alle wichtigen Internet-Domains ein manuell gepflegtes, hierarchisches Verzeichnis zu erstellen. Anfangs hatten in der Regel nur Bibliotheken, Forschungs- und Regierungseinrichtungen eine eigene Internetseite. Das erste Geschäftsmodell von Yahoo basierte übrigens auf einem solchen Verzeichnis in Verbindung mit aktuellen Nachrichten, Werbung und weiteren Dienstleistungen.

Mit der zunehmenden Kommerzialisierung des Internets, Mitte der Neunziger Jahre, explodierte die Zahl der verfügbaren Seiten und Inhalte, so dass diese Verzeichnisse nur noch als grober Einstieg dienten und das Zeitalter der Suchmaschinen begann. Es ist mit Firmen wie Altavista, Lycos, Excite oder Fireball verbunden, die es so heute alle nicht mehr gibt. Ihr großes Problem war der immer kostspieligere Aufbau eines eigenen aktuellen Suchindex und das Eliminieren von minderwertigen, doppelten und veralteten Inhalten in den Trefferlisten. Letztlich war das Scheitern an dieser Aufgabe auch das Ende der Unternehmen.

Larry Page und Sergey Brin sorgten mit einem neuartigen Algorithmus für das Sortieren von Suchergebnissen für eine substantielle Verbesserung der Qualität der Trefferlisten. Sie gründeten im September 1998 unter Mithilfe von Eric Schmidt das Unternehmen Google, dessen Marktanteil in den führenden Industrienationen außer China (Baidu) und Russland (Yandex) heute überall jenseits der 90 Prozent liegt. Einziger noch verbliebener globaler Konkurrent mit einem

Marktanteil von ca. 3 Prozent ist die Suchmaschine „Bing“ von Microsoft, die früher die Namen „Live Search“ bzw. „MSN“ trug. Nicht von ungefähr steht das Verb „googeln“ schon seit 2004 als Synonym für „im Internet suchen, recherchieren“ im Duden.

Was macht den Erfolg von Google aus?

Die zentrale Neuerung von Google war das Einbeziehen der Verlinkung einer Seite als Qualitätsfaktor, analog zu wissenschaftlichen Publikationen in Fachzeitschriften. Diese sind umso wichtiger, je öfter sie in anderen Publikationen zitiert werden. Darüber hinaus wird auch die Autorität der verweisenden Quelle in die Bewertung eines Links einbezogen.

Altavista und Co legten mehr Wert auf den Textinhalt einer Seite und waren so relativ einfach zu manipulieren. Dafür reichte schon das häufige Wiederholen des gewünschten Suchbegriffs auf einer Seite, im Extremfall mit weißer Schrift auf weißem Hintergrund.

Natürlich ist auch die Verlinkung einer Seite durch sogenannte „Linkfarmen“ manipulierbar, aber die Google-Algorithmen erkennen solche Aktivitäten heute recht gut und bestrafen Seiten, die dort übermäßig gelistet sind, durch schlechtere Platzierungen bis zum Ausschluss aus dem Suchindex.

Die genauen Kriterien, wie sich eine Trefferliste bei der Suche in Google ergibt, sind so geheim wie die Zusammensetzung eines teuren Parfüms und das wichtigste „Kapital“ von Google. Erschwerend kommt hinzu, dass sich die Rezeptur ständig verändert. Das sind dann die bei Webmastern und Suchma-

schinenoptimierern gefürchteten Updates, die Namen wie „Panda“, „Penguin“, „mobile friendly“ etc. tragen und die Rankings manchmal ziemlich durcheinanderbringen können. Google führt solche Updates und Änderungen am Algorithmus inzwischen fast täglich durch, ohne dass wir als Anwender viel davon merken. Ziel jeder Änderung sind besserer Suchergebnisse und dadurch ein immer weiterer Ausbau des eigenen Marktanteils im Suchmaschinenmarkt. Dafür betreibt Google weltweit zahlreiche Rechenzentren mit zigtausend Servern, was das Unternehmen sicher viele Millionen Euro pro Jahr kostet. Genaue Zahlen sind nicht bekannt und lassen sich daher nur schätzen.

Warum treibt Google diesen Aufwand?

Haupteinnahmequelle von Google sind bezahlte Werbeanzeigen oberhalb, neben und unterhalb der Trefferlisten, die in einem ausgeklügelten Auktionsverfahren vergeben werden (Google Adwords). Wer eine schlecht optimierte Webseite hat oder in einem hart umkämpften Markt aktiv ist, greift oft zu dieser Möglichkeit, die sich regional, zeitlich und auch hinsichtlich der Kosten sehr gut steuern lässt. Leider braucht man als Unternehmen dazu heute in der Regel einen Dienstleister für SEA (Search Engine Advertising), da die Konfiguration solcher Werbekampagnen inzwischen recht kompliziert ist und sich auch hier ständig die Algorithmen ändern. Es ist wichtig zu wissen, dass ein konstanter Anteil der Nutzer (ca. ein Drittel) Anzeigen grundsätzlich ignoriert und nur die sogenannten „organischen Suchergebnisse“ anklickt.

Warum sind Suchergebnisse oft unbefriedigend?

Man könnte meinen, es läge am Google-Algorithmus, was man bei dem immensen personellen und technischen Aufwand ausschließen kann, oder an der unzureichenden Formulierung von Suchanfragen durch den Nutzer (oft wird nur ein Wort gesucht). In der Tat könnten viele Anfragen bei genauerer Kenntnis der „erweiterten Suche“ präziser formuliert werden und man bekäme so viel kleinere und passendere Trefferlisten. Aber der eigentliche Schwachpunkt ist die unzureichende Suchmaschinenoptimierung vieler Webseiten.

Man kann das Problem mit folgendem Vergleich aus der Analytik sehr gut beschreiben: Ist die Probenvorbereitung unzureichend, dann kann später auch der beste Laborant mit dem besten Analysengerät keine korrekten Messwerte produzieren. Übertragen auf die Internet-Recherche ist der Laborant der Nutzer, das Analysengerät ist der Algorithmus der Suchmaschine und die Probenvorbereitung steht für die Optimierung einer Homepage zur bestmöglichen Indexierung.

Leider passieren bei der Suchmaschinenoptimierung (SEO = Search Engine Optimisation) seit den Anfangszeiten des World Wide Web die schlimmsten und auch von noch so guten Suchalgorithmen nicht mehr zu korrigierenden Fehler, die letztlich zu unbefriedigenden Suchresultaten führen.

Das liegt zum Teil am mangelnden Fachwissen der Agenturen, die Webseiten erstellen oder am Setzen falscher Schwerpunkte. Eine optisch ansprechende Webseite wird nicht besser gefunden als eine relativ schlichte, man denke an Wikipedia, denn der Suchroboter sieht nur den Text und dessen Verlinkung, nicht aber das Layout. Im Gegenteil: aufwändige Layouts können seine Arbeit behindern oder ihn ganz aussperren.

Deshalb sind Suchmaschinenoptimierung und die Erstellung ansprechender Texte, die konkrete Probleme des An-

wenders lösen, wichtiger als das Webdesign. Das hören allerdings Marketingverantwortliche und Geschäftsführende nicht gerne. Denn sie beurteilen eher das Layout und nicht den Quellcode.

Primär muss eine Seite von potentiellen Kunden aber erstmal gefunden werden. Wenn sie dann noch ansprechend aussieht, eine gute Nutzerführung und für den Besucher ansprechende Inhalte aufweist, ist nahezu alles perfekt. Dabei gilt die Maxime: „Der Köder muss dem Fisch schmecken, nicht dem Angler“.

Google veröffentlicht übrigens regelmäßig Richtlinien für Webmaster und Tipps zur Suchmaschinenoptimierung, die eigentlich zur DNA jedes Online-Redakteurs und Webverantwortlichen gehören sollten. Vieles davon spielt sich im nicht sichtbaren Bereich, also im Quellcode der Seite, ab.

Weitere strukturelle Probleme

Es erfordert enorm viele Ressourcen für die Betreiber von Suchmaschinen, Änderungen an Webseiten frühzeitig zu erkennen, neue Seiten zu indexieren und nicht mehr verfügbare zu löschen, um einen möglichst aktuellen Suchindex zu generieren. Im Prinzip ist der aktuelle Suchindex immer ein Abbild der mehr oder weniger fernen Vergangenheit des Internets, da sich einfach zu viel an jedem Tag ändert.

Ein weiteres Problem ist die Tatsache, dass Seiten, die schon lange unter der gleichen URL existieren, aufgrund ihrer in der Regel besseren Vernetzung mit anderen, einen Vorteil gegenüber neueren Seiten haben. Unternehmen können sich somit durch falsch durchgeführte „Relaunches“ ihre mühsam aufgebaute Reputation ruinieren, wenn dabei die bestehende Seitenstruktur geändert wird, ohne Weiterleitungen einzurichten. Das ist wie ein Umzug ohne Nachsendeantrag. Für Suchmaschinen ist die Seite dann neu, was temporär oder schlimmstenfalls sogar langfristig zum Verlust von Besuchern und Seitenabrufen durch ein schlechteres Ranking führt, weil externe Links ins Leere führen und daher

nicht mehr in die Relevanzbewertung einbezogen werden.

Nicht wenige Firmen verlagern den Zugang zu Literatur, Applikationen oder Preislisten in einen Login-Bereich, um damit Adressen zu sammeln oder Marktbegleiter auszusperrern. Das schadet dann doppelt, weil viele Nutzer von einer Registrierung Abstand nehmen und die Inhalte natürlich nicht über Google gefunden werden können. – Und der Wettbewerb wird andere Wege finden, um an die Informationen zu kommen.

In den sozialen Medien gibt es eine Tendenz, diese zu geschlossenen Netzwerken mit exklusiven Inhalten auszubauen. Die Inhalte werden dann nicht mehr über Google gefunden oder nur, wenn der Autor sie explizit dafür freischaltet. An erster Stelle zu nennen ist hier Facebook, aber auch die bekannten Business-Netzwerke wie XING und LinkedIn verfolgen diese Strategie, damit die Nutzer immer mehr Zeit innerhalb der Netzwerke verbringen und sie so höhere Werbeeinnahmen erzielen können. Leider führt auch das dazu, dass man viele gute Inhalte nicht über Suchmaschinen finden kann.

Das letzte große Problem sind die vielen Sprachen, in denen Webseiten erstellt sind. Im Zuge der Globalisierung ist sicher für jedes international tätige Unternehmen eine englischsprachige Webseite essenziell. Nur was macht man noch: Deutsch? Französisch? Spanisch? Russisch etc.? Leider ist auch Google noch nicht in der Lage, eine Suchanfrage automatisch in alle möglichen Zielsprachen zu übersetzen und diese Treffer dann auch anzuzeigen bzw. vernünftig zu übersetzen. Für die Nutzer ist es daher immer sinnvoll, auch englische Suchbegriffe zu benutzen, um mehr Treffer zu erhalten.

Fazit

Leider ist und bleibt es aus den oben genannten Gründen heute, wie auch schon vor 25 Jahren, unmöglich, über eine einzige Suchmaschine alle verfügbaren Inhalte zu einem Suchbegriff zu finden, die dann auch noch vernünftig nach Relevanz sortiert sind. Wobei das durch

Suchmaschinenoptimierung am einfachsten zu verbessern wäre.

Daneben bleibt die fehlende inhaltliche Qualitätsprüfung von Webseiten ein zentraler Unterschied zu kostenpflichtigen Datenbanken wie den Chemical Abstracts, Patentdatenbanken, wissenschaftlichen Fachzeitschriften oder der guten alten Bibliothek.

Dieses Problem wird sich durch die dezentrale, extrem dynamische inhaltliche Struktur des WWW und die Freiheit für jeden, alles was er möchte zu veröffentlichen und mit anderen zu teilen, wohl nie lösen lassen – auch nicht von einem Algorithmus.