

DGI-Forum Wittenberg 2015

Dirk Lewandowski*

Perspektiven eines Open Web Index

DOI 10.1515/iwp-2016-0020

Zusammenfassung: Der Suchmaschinenmarkt wird seit vielen Jahren von nur einer einzigen Suchmaschine, Google, dominiert. Es wurde mittlerweile erkannt, dass diese Situation nicht wünschenswert ist. Wir sprechen nun über mögliche Lösungen. Der Artikel diskutiert unterschiedliche Lösungsansätze und fokussiert dabei auf die Idee einen Offenen Web-Index (OWI), der als öffentliche Infrastruktur verfügbar gemacht werden soll. Die Grundidee ist die Trennung von Datenbestand (Index) und darauf aufsetzenden Diensten, welche in großer Zahl in privater Initiative betrieben werden können. Es geht also darum, die Basis für Vielfalt zu schaffen.

Deskriptoren: Suchmaschine, Index, Konzept, Wettbewerb

The idea of an Open Web Index

Abstract: The European search engine market has been dominated by one single search engine, namely Google, for many years. Most commentators agree that this situation is undesirable, and a discussion has started on ways to solve this problem. This article discusses different approaches, and then focuses on the idea of building an Open Web Index (OWI). This index should be provided in the form of a public infrastructure. The core idea here is to unlink the data underlying a search engine (the index) and the (search) services themselves. Many such services could be built on the basis of one index. Therefore, an Open Web Index would provide the basis for diversity on the search engine market.

Descriptors: Search engine, Index, Project schema, Competition

Les perspectives d'un index du Web ouvert

Résumé: Le marché des moteurs de recherche est dominé depuis de nombreuses années par un seul outil, Google.

*Kontaktperson: Prof. Dr. Dirk Lewandowski, Hochschule für Angewandte Wissenschaften Hamburg, Fakultät für Design, Medien und Information, Department Information, Finkenau 35, 22081 Hamburg, E-Mail: dirk.lewandowski@haw-hamburg.de, www.bui.haw-hamburg.de

Cette situation n'est pas jugée souhaitable et des discussions ont été entamées autour de solutions possibles. L'article examine les différentes approches et se focalise sur l'idée d'un index du Web ouvert (OWI), qui devrait être accessible comme infrastructure publique. L'idée principale est la séparation des données (index) et des services qui s'y greffent et qui peuvent fonctionner en grand nombre sur base d'initiatives privées. Le but est donc de créer une base pour la diversité.

Describeurs: moteur de recherche, index, concept, concurrence

Einleitung

Seit vielen Jahren ist die Situation auf dem Suchmaschinenmarkt fast unverändert: Google hält in Europa einen Marktanteil von mehr als 90 Prozent; Veränderungen sind nicht in Sicht. Auf dem US-amerikanischen Markt konnte sich mit Bing zumindest eine zweite Suchmaschine etablieren. Andere (vermeintliche) Suchmaschinen bestehen vor allem aus Suchinterfaces, die Ergebnisse einer dieser beiden Suchmaschinen anzeigen.

So sehr eine Alternative zu Google zu begrüßen ist, so wird doch schnell klar, dass es im Hinblick auf die Vielfalt der Meinungen oder eben Suchergebnisse bei weitem nicht ausreichend ist, eine oder einige weitere Suchmaschinen zu etablieren. Vielmehr ist nach Wegen zu suchen, wie es gelingen kann, die Basis für eine Vielzahl weiterer, auch spezialisierter, Suchdienste zu schaffen, um weg von nur einer von vielen möglichen Interpretationen der Inhalte des Web hin zu einer Vielzahl von Interpretationen zu kommen, die letztlich den Nutzern dient. Wir befinden uns heute in einer Situation, in der ein Nutzer ernsthaft nur zwischen zwei Interpretationen der wählen kann – und Beschränkungen der Suchmaschinen sorgen dafür, dass die Interpretationsleistung der genutzten Suchmaschine nicht ausgehebelt werden kann: Auch wenn die Suchmaschine häufig angibt, einige Millionen Treffer zu einer Suchanfrage gefunden zu haben, so werden doch immer nur maximal 1.000 Treffer ausgegeben; weitere Ergebnisse lassen sich nicht abrufen. Hier wird das Ranking zum Filter: Während im Ranking Ergebnisse nur in einer bestimmten Reihenfolge dargestellt werden und sich prinzipiell alle Ergebnisse auf-

rufen lassen, sorgt ein Filter dafür, dass nur ein Teil der Ergebnisse überhaupt angesehen werden kann.

In diesem Artikel werden die mit der Dominanz von Google verbundenen Probleme beschrieben und Lösungsansätze diskutiert. Dabei wird insbesondere auf den vom Autor vorgeschlagenen Ansatz der Schaffung eines Offenen Web Index (Open Web Index; OWI) eingegangen.

Die folgenden Abschnitte sind wie folgt gegliedert: Zunächst wird die Bedeutung der Suchmaschinen beschrieben und auf die Situation auf dem Suchmaschinenmarkt eingegangen. Aus dieser Situation heraus ergibt sich „das Problem Google“, welches im Folgenden in seinen unterschiedlichen Bereichen beschrieben wird. Anschließend werden Lösungsvorschläge zur Schaffung von Alternativen zu Google vorgestellt und diskutiert. Dann wird auf die Initiative für einen „Open Web Index (OWI)“ eingegangen, in der sich Wissenschaftler und Suchmaschinenpraktiker aus unterschiedlichen Bereichen gesammelt haben, um einen Offenen Web-Index als Infrastrukturmaßnahme voranzubringen. Der Artikel schließt mit einem Fazit.

Die Bedeutung der Suchmaschinen und der Suchmaschinenmarkt

Suchmaschinen sind *der* Zugang zu den Informationen im Web; andere Ansätze zur Erschließung dieser Inhalte haben hinsichtlich der Nutzung nur eine nachgeordnete Bedeutung. Suchmaschinen sind noch vor E-Mail der beliebteste Dienst des Internet (Frees, 2015). Dem gegenüber steht die seit Jahren unveränderte Tatsache, dass im Grunde nur eine einzige Suchmaschine, nämlich Google, genutzt wird. Sie erreicht in Deutschland wie auch in den meisten anderen europäischen Ländern einen Marktanteil von über 90 Prozent. Dazu kommen Marktanteile, die durch Partnerschaften auf dem Suchmaschinenmarkt gewonnen werden: Im sog. Partnerindex-Modell greift ein Suchportal auf die Ergebnisse einer anderen Suchmaschine zurück. Dieses Modell ist für beide Seiten attraktiv: Der gebenden Suchmaschine entstehen durch die zusätzlichen Transaktionen kaum Kosten, weil der Hauptteil der Kosten in Aufbau und Pflege der Suchmaschine liegt, nicht in der Verarbeitung der Suchanfragen. Die nehmende Suchmaschine spart die Kosten für den Betrieb einer eigenen Suchmaschine. Die Kooperationspartner teilen sich die Erlöse, die mittels der zusätzlich zu den Suchergebnissen ausgespielten Werbetreffern erzielt werden. Der Überblick der auf dem deutschen Suchmaschinenmarkt bestehenden Partnerschaften (Abb. 1) zeigt, dass nur Google und Bing als gebende Suchmaschinen eine Bedeutung haben, wäh-

rend andere Suchmaschinen keinen eigenen Datenbestand anbieten, sondern auf den Datenbestand einer dieser beiden Suchmaschinen zurückgreifen.¹ Das Partnerindex-Modell hat so zu einer Ausdünnung des Suchmaschinenmarkts beigetragen.

Dieser Blick auf den Suchmaschinenmarkt zeigt, dass es nur vermeintlich eine Vielfalt gibt, die sich tatsächlich auf nur zwei Suchmaschinen reduziert. Weltweit existieren noch zwei weitere bedeutende Indexe, nämlich die von Yandex (Russland) und Baidu (China). Weitere Indexe von neueren Suchmaschinen wie Duck Duck Go erreichen bei weitem nicht die für den Betrieb einer „echten“ weltweiten Suchmaschine notwendige Größe. Betrachtet man in dieser Weise die Entwicklung des Suchmaschinenmarkts nicht in Bezug auf die Dienste (Suchmaschinen), sondern in Bezug auf die Indexe, so findet sich eher eine rückläufige Tendenz. Seit 2009 wurden mehrere Indexe eingestellt; zu nennen sind hier Yahoo, Ask.com und Cuil.

Neben den genannten Suchmaschinen/Indexen werden immer wieder die Initiativen CommonCrawl und Internet Archive als Alternativen genannt. CommonCrawl stellt einen (kleineren) Index zur allgemeinen Nutzung zur Verfügung. Dieser Index ist jedoch statisch und eignet sich damit nicht zum Aufbau von Suchmaschinen, sondern nur für Analysezwecke. Der Index des Internet Archive speist vor allem die Wayback Machine, mit der sich alte Versionen von Dokumenten im Web aufrufen lassen. Es gibt keine Schnittstelle, um auf diesem Index andere Dienste aufzubauen. Dazu kommt, dass man auch hier nicht von einem vollständigen Index des Web – soweit dies möglich ist – sprechen kann.

Das „Problem Google“

Als das „Problem Google“ soll im Folgenden das Bündel an Problemen, das sich aus der Dominanz nur einer einzigen Suchmaschine ergibt, beschrieben werden. Dabei geht es weniger um Google als Unternehmen als um eine Situation, die sich ergibt, wenn ein einziger Anbieter den Zugang zu den Inhalten des Web bestimmt. Allerdings soll auch nicht verleugnet werden, dass diese Probleme durch Googles praktisches Handeln verstärkt werden; auf diesen Sachverhalt wird an geeigneter Stelle hingewiesen. Die hier beschriebenen Probleme beziehen sich auf die Qualität der Suchergebnisse; weitere Probleme wie beispielsweise die

¹ Die Metasuchmaschine Metager bildet hier eine Ausnahme: Sie greift u. a. auch auf die originären Datenbestände von Yandex und Exalead zurück.

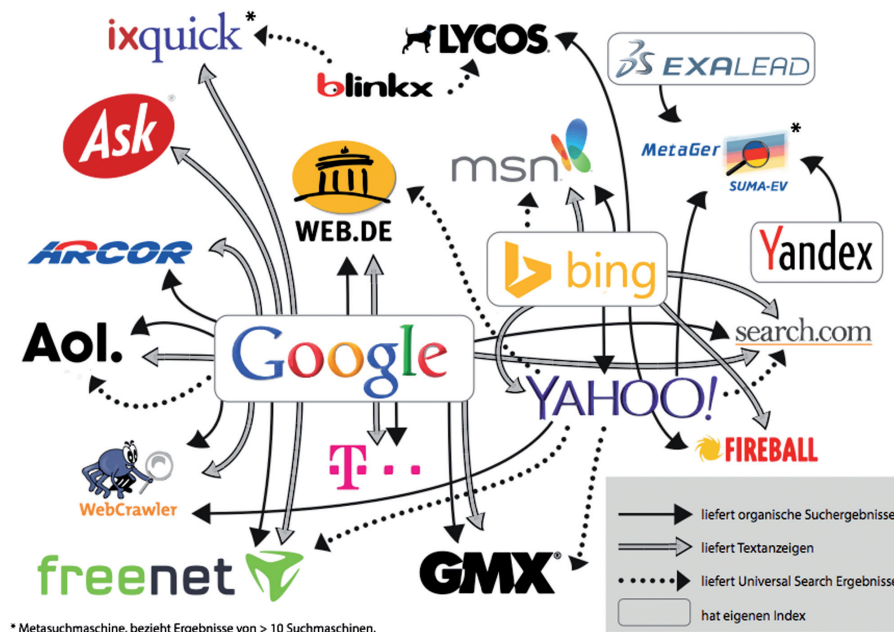


Abbildung 1: Partnerschaften auf dem deutschen Suchmaschinenmarkt (Lewandowski, 2015b, S. 158).

Dominanz bei der Online-Werbung bleiben hier ausgeklammert.

Das „Problem Google“ lässt sich in drei Bereiche unterteilen:

Jede Suchmaschine bietet ihre eigene, und nur eine, Interpretation der Inhalte des Web.

Nur für einen Teil der Suchanfragen lassen sich eindeutig richtige oder falsche Ergebnisse unterscheiden. Nach Andrei Broder (2002) lassen sich Suchanfragen im Web nach den dahinter stehenden Intentionen in drei Typen einteilen: informationsorientiert, navigationsorientiert und transaktionsorientiert. Das Ziel informationsorientierter Suchanfragen sind Informationen zu einem Thema und damit in der Regel mehrere Dokumente. Ein Nutzer möchte sich ein Bild von einem Thema machen und entscheidet selbst, wann seine Suche beendet ist. Im Gegensatz dazu hat eine navigationsorientierte Suchanfrage das Ziel, genau eine entweder bereits bekannte oder zumindest vermutete Website (wieder) zu finden. In diesem Fall gibt es genau ein richtiges Ergebnis. Mit einer transaktionsorientierten Suchanfrage möchte ein Nutzer eine Website finden, auf der dann die Transaktion (bspw. Online-Banking, der Kauf eines Produkts oder das Spielen eines Spiels im Browser) durchgeführt werden kann.

Mit dem Modell von Broder lässt sich auch die Zufriedenheit der Nutzer mit den Ergebnissen der von ihnen bevorzugten Suchmaschine erklären: Navigationsorientierte Anfragen machen einen erheblichen Anteil an der Gesamtmenge der Suchanfragen aus und ihr Ergebnis lässt sich klar nach richtig und falsch unterscheiden. Ein Nutzer

wird bei jeder navigationsorientierten Suchanfrage, die richtig beantwortet wird (die gewünschte Website steht an erster Stelle im Ranking), ein Erfolgserlebnis haben und die korrekte Beantwortung der Anfrage der benutzten Suchmaschine zuschreiben. Dies gilt ebenso für triviale informationsorientierte Suchanfragen, bei denen das Ziel nur ein einziges Dokument ist, um sich oberflächlich zu informieren oder eine Faktenaussage zu prüfen. Es ist anzunehmen, dass sich aus der häufigen positiven Beurteilung des Sucherfolgs in eindeutigen bzw. einfachen Fällen eine Übertragung auf andere, auch schwierigere Fälle von informationsorientierten Suchanfragen ergibt (Lewandowski, 2014b), d. h. Nutzer bewerten die Ergebnisse ihrer bevorzugten Suchmaschine positiv, auch wenn andere Suchmaschinen evtl. relevantere Ergebnisse (oder andere, jedoch ebenso relevante Ergebnisse) liefern würden.

Wenn es nun bei informationsorientierten Suchanfragen keine klare Unterscheidung zwischen richtig und falsch gibt, so ergibt sich ein Interpretationsspielraum. Diese Interpretation erfolgt algorithmisch, d. h. die Bewertung der Dokumente erfolgt maschinell auf der Basis von menschlichen Annahmen, die in die verwendeten Algorithmen eingehen. Jede Suchmaschine bietet damit eine eigene algorithmische Interpretation der Web-Inhalte.

Jede Suchmaschine, die selbst Inhalte oder Spezialsuchmaschinen anbietet, gerät in einen Interessenskonflikt, wenn diese Inhalte in der allgemeinen Suche angezeigt werden.

Suchmaschinenbetreiber haben bereits vor vielen Jahren damit begonnen, Spezialsuchmaschinen für bestimmte

Themen und bestimmte Bereiche des Web aufzubauen (sog. vertikale Suchen). Da Nutzer allerdings ihre Suchanfragen weiterhin vor allem in der allgemeinen Websuche eingaben und sich vor dem Abschicken einer Suchanfrage kaum Gedanken über die Auswahl eines geeigneten Suchraums machten, führten die Suchmaschinen Ergebnisse aus unterschiedlichen vertikalen Suchen im Rahmen der sog. Universal Search auf einer Suchergebnisseite zusammen (vgl. Quirnbach, 2009). Ergebnisse aus diesen vertikalen Suchen werden anders als die regulären Ergebnisse präsentiert; oft wird direkt ein Ergebnis großflächig und grafisch angezeigt (Bsp. Google Maps) oder eine Auswahl von Ergebnissen mit einem Vorschaubild präsentiert (Bsp. Google News). In dieser Präsentation bevorzugt die Web-suchmaschine Ergebnisse seiner eigenen vertikalen Suchmaschinen, auch wenn es für alle Bereiche auch alternative Angebote gibt. Diese werden jedoch in die organischen Suchergebnisse verbannt, wo sie weniger auffällig präsentiert werden. Mit dem Aufbau eigener vertikaler Kollektionen hat sich die Rolle der Suchmaschinen als reinen Vermittlern zwischen Nutzern und Dokumenten gewandelt. Zum Problem wird die Bevorzugung eigener Angebote, wenn Nutzer nicht nur vor allem in der allgemeinen Websuche suchen, sondern auch nur eine einzige Suchmaschine verwenden. Dann werden Konkurrenten benachteiligt, weil es für sie aufgrund von Entscheidungen des Suchmaschinenbetreibers auch bei besserer Qualität ihrer Produkte nicht möglich ist, gleich oder gar besser als die suchmaschineneigenen Angebote auf den Suchergebnisseiten gelistet zu werden.

Suchmaschinenoptimierung hat einen Einfluss auf die Suchergebnisse.

Suchmaschinenoptimierung bezeichnet Maßnahmen, die bestimmten Dokumenten bzw. Websites zu einer höheren Sichtbarkeit in den Suchmaschinen verhelfen sollen. Es geht also darum, Dokumente so zu gestalten, dass sie konform zu den Rankingalgorithmen der Suchmaschinen sind. Je nach Grad der Suchmaschinenoptimierung kann man von einer Hilfestellung für die Suchmaschinen, aber auch von einer Manipulation der Suchergebnisse sprechen (Lewandowski, 2015b, S. 161 ff.). Zwar lassen sich mit Suchmaschinenoptimierung die Ergebnisse aller Suchmaschinen beeinflussen, ein besonderes Problem ergibt sich aber gerade durch die Fokussierung auf eine einzige Suchmaschine, d. h. wenn es durch Suchmaschinenoptimierung gelingt, bestimmten Inhalten eine besonders hohe Sichtbarkeit zu verschaffen, dann gelingt dies praktisch für alle Suchmaschinennutzer. Gäbe es eine größere Vielfalt – auch hinsichtlich der Nutzung – auf dem Suchmaschinenmarkt, wären die Effekte der Suchmaschinenoptimierung weit geringer.

Suchmaschinenoptimierung ist mittlerweile eine etablierte Praxis im kommerziellen Bereich. Allerdings sollte nicht vergessen werden, dass Suchmaschinenoptimierung auch für informative Inhalte erfolgt. Vorangegangen sind hier die Zeitungsverlage, die ihre Online-Nachrichtenangebote optimiert haben. Zunehmend werden aber auch Inhalte zum Zweck der PR oder sogar im akademischen Bereich optimiert. Im Zusammenhang mit der Verwendung von nur einer Suchmaschine durch die Nutzer ergibt sich hier eine erhebliche Manipulationsmöglichkeit. Ein aktuelles Beispiel ist die Untersuchung des Effekts von Suchmaschinenrankings auf das Verhalten bei politischen Wahlen. Hier wird Google eine erhebliche Einflussmöglichkeit zugesprochen, die sich analog auch auf externe, entsprechend optimierte Inhalte übertragen lässt (Epstein & Robertson, 2015).

Warum brauchen wir mehr als eine Suchmaschine?

Abgesehen von den bisher genannten, eher grundsätzlichen Überlegungen gibt es auch pragmatische Gründe, die für mehr als eine Suchmaschine sprechen bzw. die es angezeigt erscheinen lassen, sich für eine Vielfalt auf dem Suchmaschinenmarkt stark zu machen. Dies sind insbesondere die folgenden sieben Gründe (vgl. (Lewandowski, 2015b, S. 179 ff.):

1. Einholen einer „zweiten Meinung“: Ein Nutzer kann durch die Nutzung einer zweiten (zusätzlichen) Suchmaschine zu ein und derselben Suchanfrage eine andere Sicht auf die Inhalte des Web erhalten.
2. Mehr bzw. zusätzliche Ergebnisse: Dadurch, dass kein Suchmaschinenindex ein vollständiges Abbild des Web bietet, können sich mit einer Suchmaschine Dokumente finden lassen, die in von einer anderen nicht gefunden werden können. Zwar mag ein Nutzer nur in seltenen Fällen vor dem Problem stehen, gar nichts zu seiner Suchanfrage zu finden, jedoch kann es durchaus sinnvoll sein, mehr Ergebnisse zu der selben Suchanfrage aufzurufen.
3. Andere Ergebnisse: Die Ergebnisse verschiedener Suchmaschinen überschneiden sich auf den vorderen Trefferpositionen stark (vgl. (Spink, Jansen, Blakely, & Koshman, 2006). Dies bedeutet, dass schon das einfache Eingeben der selben Suchanfrage bei einer anderen Suchmaschine zu anderen Ergebnissen führt, auch wenn es natürlich auch Überschneidungen gibt.
4. Bessere Ergebnisse: In Retrievaleffektivitätstests schneiden Suchmaschinen zwar unterschiedlich ab (auch wenn diese Unterschiede in der Regel geringer

sind als erwartet, s. (Lewandowski, 2015a), dies bedeutet jedoch nicht, dass eine Suchmaschine bei jeder Suchanfrage den anderen überlegen wäre. Vielmehr zeigt sich, dass auch eine schlechter abschneidende Suchmaschine zu manchen Anfragen bessere Ergebnisse liefert.

5. Andere Trefferpräsentation: Selbst Suchmaschinen, die auf einen Partnerindex zurückgreifen (wie etwa Yahoo), können durch eine eigene Zusammenstellung und Präsentation der Suchergebnisse zu sinnvollen Alternativen werden. Dies verstärkt sich insbesondere bei neueren Darstellungsformen wie direkten Antworten, Universal Search und Knowledge Graph.
6. Andere Benutzerführung: Eng verbunden mit der Trefferpräsentation ist die Benutzerführung. Die Gestaltung des Suchinterface kann eine Suchmaschine zu einer lohnenden Alternative machen.
7. Andere Suchmöglichkeiten: Suchmaschinen sind hinsichtlich ihrer Recherchemöglichkeiten beschränkt. Insbesondere die eingeschränkten (und teils nicht korrekt funktionierenden) Suchfunktionen von Google wurden immer wieder bemängelt.

Es soll nicht verleugnet werden, dass die genannten Gründe jeweils nur auf einen kleinen Teil der Suchanfragen zutreffen. Allerdings mindert dies nicht ihre Bedeutung, wenn man Suchmaschinen als ernsthafte Recherchewerkzeuge betrachtet.

Lösungsvorschläge

Wie nun kann man nach dem Gesagten dafür sorgen, dass alternative Suchmaschinen etabliert werden und damit Vielfalt auf dem Suchmaschinenmarkt hergestellt wird? Der Schlüssel liegt m.E. darin, attraktive Angebote zu schaffen, die von den Nutzern angenommen werden. Wie solche Angebote der allgemeinen Websuche, aber auch der Spezialsuche, genau aussehen, muss den Anbietern und den Nutzern überlassen werden. Im Folgenden werden die Strategien zur Etablierung weiterer Anbieter auf dem Suchmaschinenmarkt erörtert.

Zunächst einmal gibt es die Hoffnung, dass sich der Markt selbst regulieren wird. Hierbei wird davon ausgegangen, dass die aktuelle (und bereits seit vielen Jahren) bestehende Situation durch den Eintritt neuer Anbieter verändert werden kann. Auf der Basis dieser Annahme stellt sich dann nur noch die Frage, wie schwierig es für einen neuen Anbieter ist, sich auf dem Markt zu etablieren und inwieweit er von einem Monopolanbieter daran gehindert wird (vgl. Monopolkommission 2015). Es ist positiv

festzustellen, dass es Microsoft mit erheblichen finanziellen Mitteln gelungen ist, seine Suchmaschine Bing zumindest in den USA zu etablieren. Letztlich bedeutet dies aber nicht mehr, als dass *eine* weitere Suchmaschine zur Verfügung steht; an der fehlenden *Vielfalt* auf dem Suchmaschinenmarkt ändert dies wenig.

Eine andere Idee beruht auf der Annahme, dass das Problem gar nicht in dem Vorhandensein nur einer Suchmaschine läge, sondern in der (un)fairen Ergebnisanzeige. Die Lösung ist dann, Google zu einer fairen Ergebnisanzeige zu verpflichten, indem alle Dokumente nach den selben Regeln behandelt werden und damit jedes Dokument zumindest potentiell die selbe Chance hat, als Ergebnis zu einer Suchanfrage angezeigt zu werden. Problematisch an dieser Forderung ist jedoch die Überprüfbarkeit ihrer Einhaltung und auf der theoretischen Ebene die Frage nach der Eindeutigkeit von Ergebnisssets, d. h. ob es überhaupt ausreichend ist, nur eine Art der Ergebnisreihung zu haben, oder ob nicht gerade die Vielfalt von unterschiedlichen Ergebnisreihungen durch unterschiedliche Suchmaschinen das Ziel sein sollte.

Eine weitere Forderung, die wiederum von der Notwendigkeit einer fairen Ergebnisanzeige ausgeht, zielt auf die Zerschlagung von Google durch die Trennung von allgemeiner Websuchmaschine und den anderen angebotenen Diensten. Dies würde insbesondere dem Problem der Bevorzugung eigener Angebote begegnen. Es erscheint jedoch mehr als fraglich, ob sich dadurch tatsächlich die Marktsituation verändern würde.

Ein immer wiederkehrender Vorschlag ist der Aufbau einer alternativen Suchmaschine. Mehrfach wurde gefordert, mit öffentlicher Hilfe eine Alternative zu Google (oder allgemeiner: zu den großen kommerziellen Suchmaschinen) aufzubauen. Zuletzt forderten Hege & Flecken (2014) den Aufbau einer öffentlich-rechtlichen Suchmaschine und stellten deren Aufgabe in den Kontext des öffentlich-rechtlichen Rundfunks. In vielen Fällen wurde mit der Notwendigkeit einer europäischen Alternative zu Google argumentiert; dies vor allem im ursprünglichen Vorschlag des Aufbaus der deutsch-französischen Suchmaschine Quaero. Zu dem gemeinsamen Projekt kam es allerdings nie, weil sich die deutsche und die französische Seite nicht über einen gemeinsamen Weg verständigen konnten. Stattdessen wurde in den beiden Ländern jeweils ein einzelnes Projekt (Quaero bzw. in Deutschland Theseus) verfolgt; beide Projekte hatten dann allerdings nicht mehr den Aufbau einer alternativen Suchmaschine zum Ziel.

Aus grundlegenden Überlegungen ist der Aufbau einer alternativen Suchmaschine abzulehnen: Zuerst einmal würde eine weitere Suchmaschine den Suchmaschinen-

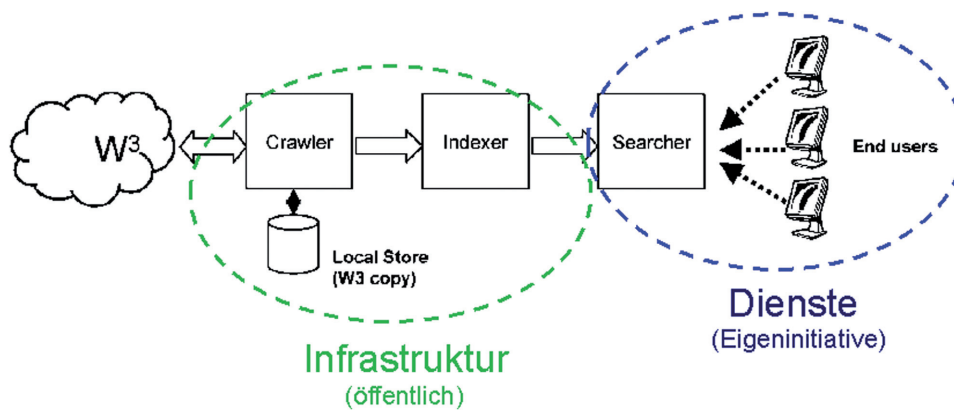


Abbildung 2: Entkoppelung von Infrastruktur und Diensten (auf Basis des grundlegenden Aufbaus von Suchmaschinen aus Risvik & Michelsen, 2002).

markt nicht grundlegend verändern, auch wenn sie nennenswerte Marktanteile erreichen würde. Es ergibt nur einen geringen Unterschied, ob die Auswahl zwischen zwei oder drei (oder einer Handvoll) Suchmaschinen besteht. Des Weiteren kann auch eine hochwertige Suchmaschine an Gründen scheitern, die der Suche nicht inhärent sind, beispielsweise an Fragen des Marketings. Ein solches Scheitern würde allerdings bedeuten, dass die dafür nötigen enormen Investitionen verloren wären.

Stärker in Richtung einer Vielfalt auf dem Suchmaschinenmarkt geht der Vorschlag, Google zur Öffnung seines Index zu zwingen. Die Idee hier ist, Konkurrenten den Zugriff auf den Datenbestand der Google-Suchmaschine zu gewähren, damit diese eigene Suchdienste aufbauen können. Zugrunde liegt die Erkenntnis, dass eine wesentliche Hürde für den Aufbau einer neuen Suchmaschine gerade im Aufbau und der Pflege des Index besteht. Heutige Partnermodelle (s. o.) erlauben nur einen beschränkten Zugriff auf *Top-Ergebnisse* der Suchmaschinen (im Gegensatz zu einem vollständigen Zugriff auf den Index) und deren Anzeige (im Gegensatz zu einer umfassenden Weiterverarbeitung).

Schließlich gibt es noch den Vorschlag des Aufbaus eines Web-Index als Infrastrukturmaßnahme (Lewandowski, 2014a). Die Kernidee hierbei ist die Entkoppelung von Index und Diensten. Es geht also darum, eine von allen nutzbare Datenbasis zu schaffen, auf der dann eine Vielzahl unterschiedlicher Dienste aufgebaut werden können. Während die Suche im Web und in Teilen des Web eine Kernanwendung eines solchen Index ist, ist sie keineswegs auf Suchanwendungen beschränkt. Vielmehr können alle Dienste, die in irgendeiner Weise auf Webdaten aufbauen, von einem solchen Index profitieren, also bspw. auch Dienste zur Analyse der Struktur des Web.

Die Entkoppelung von Index und Diensten ist in Abbildung 2 anhand des grundlegenden Aufbaus von Such-

maschinen (Risvik & Michelsen, 2002) dargestellt. Eine Suchmaschine besteht aus den Komponenten *Crawler*, *Local Store*, *Indexer* und *Searcher*. Der Crawler erfasst die Dokumente des Web mittels Verfolgen von Links und speichert die gefundenen Dokumente im Local Store ab, der im Idealfall eine vollständige und aktuelle Kopie des Web ist. Der Indexer bereitet die Dokumente so auf, dass auf der einen Seite aussagekräftige Repräsentationen der Dokumente entstehen (vgl. (Lewandowski, 2015b), S. 48 ff.), auf der anderen Seite die Dokumente effizient durchsuchbar werden. Der Searcher schließlich sorgt für den Abgleich zwischen Suchanfragen und Dokumenten und bringt die gefundenen Ergebnisse in eine Reihenfolge (Ranking).

Die Entkoppelung bedeutet nun, die Bereitstellung des Index (also das Crawling und eine grundlegende Aufbereitung der Dokumente) als öffentliche Aufgabe zu betrachten, analog zu anderen Infrastrukturaufgaben des Staates wie den Bau eines umfassenden Straßennetzes. Die auf diesem Index aufbauenden Dienste werden dann in Eigeninitiative betrieben, gleich ob von Unternehmen, öffentlichen Einrichtungen oder Hobbyentwicklern. Die Betreiber der Dienste erhalten Zugang zur Infrastruktur und sind weder darauf angewiesen, selbst eine Infrastruktur aufzubauen noch sich auf den Zugang zu anderen privatwirtschaftlichen Infrastrukturen zu verlassen.

Die Initiative „Open Web Index (OWI)“

Zur Verbreitung der Idee des Open Web Index hat sich die „Open Web Index Initiative“ gebildet. Diese besteht aus Experten aus Wissenschaft und Industrie. Von Beginn an war klar, dass sich die Idee des Open Web Index nur durchsetzen lässt, wenn auch die Industrie beteiligt ist. Die Schaf-

fung eines offenen Web-Index erfordert erhebliche finanzielle Mittel, so dass die Sinnhaftigkeit des Vorhabens nicht nur durch die Wissenschaft und öffentliche Einrichtungen dargestellt werden kann, sondern auch durch die Industrie in den geschäftlichen Potentialen von Diensten, die auf einem solchen Index aufbauen. Die Initiative besteht seit 2014; Sprecher sind Wolfgang Sander-Beuermann (SuMa e. V.) und Dirk Lewandowski (HAW Hamburg). Das Ziel der Initiative ist es zunächst, Aufmerksamkeit für die Idee des Open Web Index zu schaffen und die Überlegenheit der Idee gegenüber anderen Ansätzen darzustellen. Allerdings arbeitet die Gruppe auch an Lösungen für grundsätzliche technische Fragen, die sich durch den Aufbau und die Bereitstellung eines solchen Index ergeben. Während viele der aktuellen Arbeiten eher im Hintergrund ablaufen, finden sich bereits veröffentlichte Informationen auf der Website www.openwebindex.eu. Dort wird auch über den weiteren Fortschritt der Arbeiten berichtet.

Fazit

Zunächst einmal ist positiv, dass das „Problem Google“ endlich erkannt wurde. Nun wird über mögliche Lösungen diskutiert. Dabei ist m. E. eine konstruktive Lösung einer destruktiven vorzuziehen. Der Aufbau eines offenen Web-Index – und damit die Trennung von öffentlich finanzierter Infrastruktur und privatwirtschaftlich betriebenen Diensten – sorgt für mehr Vielfalt auf dem Suchmaschinenmarkt und darüber hinaus. Frei zugängliches Wissen ist die Quelle von kontinuierlichem Wettbewerb, Innovation und Unternehmertum auf der einen und persönlicher und Meinungsfreiheit auf der anderen Seite. Der Open Web Index schafft dafür die Voraussetzung. Er kann mit anderen Infrastrukturaufgaben des Staates verglichen werden – in Bezug auf die Sammlung und Verfügbarmachung von Wissen ist vor allem der Vergleich mit den Bibliotheken zu ziehen.

Literatur

- Broder, A. (2002). A taxonomy of web search. *ACM Sigir Forum*, 36(2), 3–10.
- Epstein, R., & Robertson, R. E. (2015). The search engine manipulation effect (SEME) and its possible impact on the outcomes of elections. *Proceedings of the National Academy of Sciences*, 2015, 201419828. <http://doi.org/10.1073/pnas.1419828112>.
- Frees, V. B. (2015). Internetnutzung: Frequenz und Vielfalt nehmen in allen Altersgruppen zu. *Media Perspektiven*, (10), 366–377.
- Hege, H., & Flecken, E. (2014). Debattenbeitrag: Gibt es ein öffentliches Interesse an einer alternativen Suchmaschine? In

B. Stark, D. Dörr, & S. Aufenanger (Eds.), *Die Googleisierung der Informationssuche* (pp. 224–244). Berlin: De Gruyter.

- Lewandowski, D. (2014a). Why we need an independent index of the Web. In R. König & M. Rasch (Eds.), *Society of the Query Reader: Reflections on Web Search* (pp. 49–58). Information Retrieval; Digital Libraries, Amsterdam: Institute of Network Culture.
- Lewandowski, D. (2014b). Wie lässt sich die Zufriedenheit der Suchmaschinenutzer mit ihren Suchergebnissen erklären? In H. Krah & R. Müller-Terpitz (Eds.), *Suchmaschinen (Passauer Schriften zur interdisziplinären Medienforschung, Band 4)* (pp. 35–52). Münster: LIT.
- Lewandowski, D. (2015a). Evaluating the retrieval effectiveness of Web search engines using a representative query sample. *Journal of the Association of Information Science and Technology*, 66(9), 1763–1775. <http://doi.org/10.1002/asi.23304>.
- Lewandowski, D. (2015b). *Suchmaschinen verstehen*. Berlin Heidelberg: Springer Vieweg.
- Quirnbach, S. (2009). Universal Search – Kontextuelle Einbindung von unterschiedlicher Quellen und Auswirkungen auf das User Interface. In D. Lewandowski (Ed.), *Handbuch Internet-Suchmaschinen* (pp. 220–248). Heidelberg: Akademische Verlagsgesellschaft Aka GmbH.
- Risvik, K. M., & Michelsen, R. (2002). Search engines and web dynamics. *Computer Networks*, 39(3), 289–302.
- Spink, A., Jansen, B. J., Blakely, C., & Koshman, S. (2006). A study of results overlap and uniqueness among major web search engines. *Information Processing & Management*, 42(5), 1379–1391.



Prof. Dr. Dirk Lewandowski
Hochschule für Angewandte Wissenschaften
Hamburg
Fakultät für Design, Medien und Information
Department Information
Finkenau 35
22081 Hamburg
dirk.lewandowski@haw-hamburg.de
www.bui.haw-hamburg.de

Prof. Dr. Dirk Lewandowski ist Professor für Information Research & Information Retrieval an der Hochschule für Angewandte Wissenschaften Hamburg. Zuvor war er unabhängiger Berater im Themenbereich Suchmaschinen und Information Retrieval sowie Lehrbeauftragter an der Universität Düsseldorf. Seine Forschungsinteressen sind Web Information Retrieval, Qualitätsfaktoren von Suchmaschinen, das Rechercheverhalten der Suchmaschinen-Nutzer sowie die gesellschaftlichen Auswirkungen des Umgangs mit den Web-Suchmaschinen. Zu seinen Veröffentlichungen gehören neben den Büchern „Web Information Retrieval“ und „Handbuch Internet-Suchmaschinen“ zahlreiche Aufsätze, die in deutschen und internationalen Fachpublikationen veröffentlicht wurden. Sein neuestes Werk „Suchmaschinen verstehen“ ist im Februar 2015 erschienen. Dirk Lewandowski ist Editor in-Chief des *Aslib Journal of Information Management* (früher: *Aslib Proceedings*) und Associate Editor der *Online Information Review*.