

Der Offene Web-Index – Fragen und Antworten

1) Was ist überhaupt ein "Web-Index"?

Ein „Index“ ist zunächst einmal ein Instrument (ein Nachschlagewerk, ein geordnetes Verzeichnis) um etwas zu finden. Jeder kennt es aus den letzten Seiten von Sachbüchern. Zu vielen Wörtern aus einem Buch sind dort Seitenzahlen notiert: diejenigen Seiten, auf denen dieses Wort und/oder dieses Thema vorkommt. In dieser Hinsicht wird ein Index auch „Verschlagwortung“ genannt. Im Gegensatz zum Inhaltsverzeichnis eines Buches ist ein Index wesentlich detaillierter, erlaubt den schnellen Zugriff auf gesuchte Inhalte.

Ein Index kann aber viel mehr sein, als nur Wort-bezogen. Dort kann auch notiert sein:

- wo welche Bilder vorkommen,
- wo welche Bilder oder Grafiken zu welchen Themen vorkommen,
- welche Wörter nur in Überschriften vorkommen (also evtl. besonders wichtig sind)
- welche anderen Bücher oder Aufsätze zitiert werden, u.v.a.m.

Der Ersteller des Indexes ist frei darin, was er dort für Informationen erfasst, mit denen das Wiederfinden von Inhalten im Buch erleichtert wird. Ein solcher Index wird aber umso wichtiger, je dicker ein Buch ist.

Das dickste „Buch“, welches die Menschheit je geschrieben hat, ist das World Wide Web (kurz WWW oder Web), aus dem heutzutage der größte Teil des Internets besteht. Ohne einen Web Index wäre im Internet mit seinem gigantisch großen Inhalt kaum etwas wiederzufinden. DARUM ist ein Web Index von zentraler Bedeutung für das Internet – ohne ihn ist es praktisch kaum nutzbar.

2) Was hat ein Web-Index mit Suchmaschinen zu tun?

Suchmaschinen machen mit dem Web Index das, was Nutzer eines Buches mit dem Index tun: nachsehen, zu welchen Suchwörtern (oder Bildern, Themen usw.) welche Webseiten des WWW passen. Prinzipiell könnten dies Menschen machen – aber es würde viel zu lange dauern; darum wird diese Sucharbeit (das Zuordnen von Wörtern zu Webseiten) von Programmen gemacht. Ein solches Programm nennt man eine Such„maschine“. Diese suchen nicht wirklich in dem Moment durch das ganze WWW, in dem User Suchwörter eingeben, das würde nicht schnell genug gehen, sondern nur in ihrem Index.

Jede Suchmaschine braucht einen Index, in dem sie Zuordnungen suchen kann. Eine Ausnahme scheinen Meta-Suchmaschinen, wie metager.de, zu sein. De facto benutzen Meta-Suchmaschinen jedoch die von anderen Suchmaschinenbetreibern erstellten Indexe. Das geht nur mit Erlaubnis dieser Suchmaschinenbetreiber. Die Meta-Suchmaschine ist also vom guten Willen der Index-Eigentümer abhängig.

Weltweit gibt es nur noch vier Betreiber, die einen globalen Index besitzen: Google und Microsoft/Yahoo in den USA, Yandex in Russland und Baidu in China. Die Eigentümer der Indexe stellen diese anderen Suchmaschinen nicht vollständig zur Verfügung, und meist nur gegen Bezahlung. Es gibt keinen uneingeschränkten Zugriff auf alles in den Indexen und der Zugreifende ist vollständig vom Index- Eigentümer abhängig.

3) Was ist das Besondere am Offenen Web-Index/OWI?

Bei einem freien, Offenen Web Index werden Nutzer dieses Indexes den uneingeschränkten Zugriff auf die gespeicherten Informationen haben und werden nicht mehr vom guten Willen der vier globalen Konzerne abhängig sein. Nutzer dieses Indexes werden viele neue Suchmaschinenbetreiber sein, welche die unterschiedlichsten Arten und Formen von Suchmaschinen daraus entwickeln. Diese Suchmaschinen können sehr unterschiedlich zu den bisher bekannten sein – durch andere Bedienung, anderes Design, andere Themengebiete, andere Geschäftsmodelle, und sie werden in Konkurrenz zueinander stehen.

Es können aber auch ganz andere Anwendungen als Suchmaschinen entstehen, diese „Datenschätze“, dieses „Gold“ des Internets, mit dem Offenen Web Index „auszugraben“ und der Menschheit zur Verfügung zu stellen. So können beispielsweise durch Auswertung von Linkstrukturen und deren Clusterbildungen neue Einsichten auf inhaltliche Zusammenhänge entstehen. Grundsätzliche Erkenntnisprobleme der Menschheit können in neuem Licht erscheinen, beispielsweise Fragen wie „In welchen Kategorien spielt sich menschliches Denken ab?“, aber auch ganz konkrete, praktische Fragen wie die Variation in der Preisgestaltung für ein gleiches Produkt in Märkten weltweit. Eine völlig neue Vielfalt unterschiedlichster Suchmaschinen und völlig neuer Werkzeuge wird entstehen.

Voraussetzung dazu ist der uneingeschränkte Zugriff auf den Index, unabhängig von kommerziellen Interessen seiner Betreiber. Kein Betreiber wird mehr die Möglichkeit haben, missliebiges Wissen zu unterdrücken oder Konkurrenz auszuschalten. Die nur teilweise, oder gänzliche, Nichtindexierung einer Website ist jedem der vier Index-Betreiber bisher freigestellt, was viele Inhalte des WWW Menschen praktisch unsichtbar macht.

Der Offene Web Index ist nach der Erfindung der Druckerpresse, des Rundfunks, des Computers, des World Wide Web und der Suchmaschine, der nächste Schritt in der Evolution menschlichen Wissens. Er erlaubt den FREIEN Zugang und Umgang mit Wissen.

4) Warum ist es so schwierig, einen solchen Web-Index zu erstellen?

Es ist eine große, aber lösbare Aufgabe, den Offenen Web-Index zu erstellen und aktuell zu halten. Aber in Europa ist kaum eine Organisation oder Firma oder selbst ein Staat willens oder in der Lage, dazu alleine die finanziellen Mittel aufzubringen. Das liegt an der kaum vorstellbaren Größe des Web. Um seine Inhalte in einem Index zu erfassen, werden Zehntausende von Rechnern benötigt. Dazu gehört eine technische und administrative Infrastruktur, die von sehr vielen Menschen unterschiedlicher, aber hoher Qualifikation betrieben werden muss. Kostenschätzungen dieses Aufwandes beginnen bei einigen Hundert Millionen bis hin zu einigen Milliarden Euro.

Als Schlagwort zur Beschreibung dieses gemeinschaftlichen Projektes ist „IT-Airbus“ von manchen genannt worden. Das Projekt ist dem europäischen Airbus vergleichbar, wird aber nicht einen großen Konzern hervorbringen, sondern viele kommerzielle und nicht-kommerzielle Projekte aller Größen und Ausrichtungen.

Ein Alleingang eines Staates oder eines Konzerns ist auch nicht wünschenswert, denn die Vielfalt aller europäischen Sprachen und allen Wissens soll im Offenen Web Index gespiegelt sein.

Daher ist ein solches Projekt, welches Europa unabhängig von den bisher ausschließlich nicht-europäischen Index-Eigentümern macht, am besten in der Staatengemeinschaft der EU verankert.

5) Was kann ein Open-Web-Index/OWI mehr als Suchmaschinen füttern?

Ein Web-Index bringt die unstrukturierten, heterogenen Daten des Internet in eine strukturierte Form. Nur damit ist es möglich, Internetdienste aufzusetzen, die diese Daten für die verschiedensten Zwecke aufbereiten. Einige Beispiele (neben Suchmaschinen) dazu:

- Kartendienste können Orte mit Daten aus dem OWI versorgen. Konkret: Daten zu Städtenamen, Hotels, Restaurants, Geschäfte, Sehenswürdigkeiten können auf Klick des Nutzers aus dem OWI in die Karten übernommen, die Informationen dazu aus dem OWI extrahiert, und in die Karten integriert werden. Dazu in wählbarer Tiefe weitere Hintergrunddaten, z.B. die Historie von Städten, Sehenswürdigkeiten; Preise von Hotels, Restaurants, Geschäften usw.
- Preis-Vergleichs- und Verbrauchenportale können ihre Daten direkt aus dem OWI beziehen; damit wesentlich einfacher, vollständiger und transparenter aufgesetzt

werden.

- Linkstrukturen können durch den OWI analysiert werden: Wer verlinkt wen, wo bilden sich Linkcluster, welche Bedeutungen haben diese. Diese Informationen können dann wiederum genutzt werden, um darauf aufsetzende neue Werkzeuge zur Informationsgewinnung zu schaffen; beispielsweise durch Generierung von Kategorienschema.
- Aus den aus dem OWI exportierten Daten lassen sich Trends jeglicher Art ablesen; welche Themen/Kategorien interessieren momentan die Nutzer? Wo sind z.B. politische Veränderungen erkenn- und vorhersagbar? Google-Flu-Trends macht das deutlich: https://en.wikipedia.org/wiki/Google_Flu_Trends

Genau so wie Google bisher/öffentlich Grippewellen erkennt, wird Google auch Werkzeuge zur Erkennung politischer Trends haben. Solches Wissen beinhaltet auch der OWI. Darauf aufsetzende Werkzeuge können es extrahieren und öffentlich machen.

Diese Trends können zum einen anhand der exportierten Daten gewonnen werden, zum anderen aber auch aus den inhaltlichen Veränderungen der Webseiten selber: Zu welchem Thema entstehen gerade viele neue Seiten? Wie verändern sich Seiteninhalte über die Zeit?

- Aus den Strukturen und Inhalten des Index können KI-Programme ihre Daten gewinnen und so die Basis für eine Open-Source-KI schaffen.
- Das wichtigste Merkmal des OWI ist, dass er eine offene Basis für gänzlich neue Internetdienste schafft - auch für solche, die wir uns derzeit noch nicht vorstellen können.

Mit dem OWI können Geschäftsmodelle auf dem Wissen der Welt aufbauen. Sie sind dann unabhängig von fremden Gatekeepern. Viele Innovationen sind nur auf der Basis eines OWI möglich; ohne OWI gibt es keine gleichberechtigte Teilhabe an (Innovations-)Chancen. Der OWI schafft damit den notwendigen Nährboden zur Entwicklung von Internet-Startups.

6) Wie kann ein OWI organisatorisch gestaltet werden?

Erste Überlegungen dazu gehen von einem Stiftungsmodell aus, vergleichbar der deutschen Stiftung Warentest. Diese Stiftung muss unabhängig und staatsfern sein.

7) Wie ist die Finanzierung des OWI?

Für die Startphase wird eine öffentliche Finanzierung aus Mitteln der EU benötigt. Je mehr der OWI dann von kommerziellen Betreibern genutzt wird, desto mehr ist ein Return on Investment zu erwarten. Eine sich selbst tragende Wirtschaftlichkeit dürfte nach ersten Schätzungen nicht vor Ablauf von ca. 5 Jahren zu erwarten sein.

8) Wie wird sichergestellt, dass mit dem OWI kein staatliches Monopol aufgebaut wird?

Die Organisations- und Rechtsform des OWI muss von Beginn an unabhängig und staatsfern konzipiert sein. Da der OWI selber keinen Zugang für Endnutzer bietet, ist es ausgeschlossen, dass er eine der jetzigen Situation vergleichbare Monopolstellung hervorbringt.

9) Wie ist der OWI für Endnutzer zugänglich - und wer darf den OWI nutzen?

Der OWI ist nicht für Endnutzer zugänglich, sondern nur für Firmen oder Organisationen, die darauf aufsetzende Dienste anbieten oder entwickeln. Er folgt einem B2B-Geschäftsmodell. Nutzen kann ihn JEDE Firma oder Organisation, welche die AGBs unterschreibt.

10) Soll der OWI ein "europäisches Google" werden?

NEIN! Der OWI stellt eine Dateninfrastruktur bereit, die andere Firmen oder Organisationen nutzen können, um darauf neue Dienste aufzubauen.

11) Wie stellt der OWI Datenschutz und Privacy sicher?

Jeder Kunde des OWI muss dessen AGBs unterschreiben. Kunden des OWI sind NICHT Endnutzer, sondern Firmen oder Organisationen, welche auf dem OWI aufsetzende Dienste anbieten oder entwickeln. Der OWI hat ein B2B-Geschäftsmodell. Die AGBs des OWI definieren Datenschutz und Privacy - daran MUSS sich der Kunde halten. Tut er es nicht, wird sein Zugang gesperrt.

Falls auch große Internet-Dienste (wie z.B. Google) einmal Kunde des OWI sein wollten, sind sie genauso diesen Datenschutz-Regeln unterworfen.

12) Eine Technische Frage:

Ein Kunde des OWI stellt eine Abfrage an den OWI-Index, der Millionen von Ergebnissen an den Kunden zurückliefern würde - wie behandelt der OWI dieses Problem?

Die Datensätze des OWI-Index erhalten ein Pre-Ranking. Dieses Pre-Ranking kann u.a. auch vom Kunden des OWI gesteuert werden. Der Kunde erhält auf solche Abfragen mit Millionen möglicher Antworten, im Normalfall beispielsweise nur die ersten 1000 Antworten.

Er KANN auch ALLE Antworten bekommen, ABER jeder Kunde hat zunächst ein begrenztes Freikontingent an Antworten. Ab einem zu definierenden Limit wird die Abfrage kostenpflichtig und u.U. auch kostspielig für den Kunden.