

Google auf dem Weg zum Monopol? Aktuelle Entwicklungen bei Internetsuchmaschinen: Märkte, Technologien und die Rolle der Bibliotheken (Preprint)

erschienen in: *BuB. Forum für Bibliothek und Information*, 55; Jg. 2003 (9) , S. 568 – 576

Internet-Suchmaschinen sind inzwischen eine der wichtigsten Möglichkeiten der Informationsrecherche und –beschaffung im Internet. Sie sind einfach zu bedienen und liefern in Sekundenbruchteilen Ergebnisse aus einem riesigen Datenpool, der mehrere Milliarden Dokumente umfassen kann.

Eine Suchmaschine hat in den letzten Jahren besonders viel Aufmerksamkeit erregt: „Google“. Google ist inzwischen die mit großem Abstand meistgenutzte Suchmaschine und wird von vielen Internet-Nutzern als erste (und häufig auch als einzige) Möglichkeit der Informationsrecherche im Internet benutzt. Die Suchmaschine "Google" ist inzwischen für viele zum Synonym für "Suchen im Internet" geworden.

Doch ist dieses Quasi-Monopol, welches Google inzwischen erreicht hat, auch eine Gefahr? Wenn Google (vermeintlich) immer irgendetwas zu einem Thema findet, wer braucht dann noch andere Suchmaschinen, Verzeichnisse oder gar Datenbanken?

Dass auch Google höchst unvollständig ist, und die riesigen Datenmengen, die in Datenbanken und Bibliothekskatalogen liegen – das sogenannte „Invisible Web“ - , überhaupt nicht abdeckt, ist vielen Nutzern gar nicht bewusst.

Google (und andere Suchmaschinen) werden selbst für die wissenschaftliche Recherche häufig eingesetzt. Dies zeigt eine Umfrage der Universitätsbibliothek Bielefeld [1]. Bei dieser Umfrage gaben 70% der Befragten an, für eine *wissenschaftliche Recherche* häufig allgemeine Internet-Suchmaschinen einzusetzen. Damit waren Suchmaschinen nach dem Bibliothekskatalog (den 88% häufig einsetzten), das am zweithäufigsten genannte Instrument zur wissenschaftlichen Recherche. Nur 44% der Befragten gaben an, Fachdatenbanken zu nutzen, und spezielle wissenschaftliche Suchmaschinen wurden sogar nur von 19% der Befragten genutzt.

Aktuelle Entwicklungen auf dem Suchmaschinen-Sektor

In den letzten Jahren hat sich die Suchmaschinen-Landschaft grundsätzlich geändert. Ein kurzer Rückblick soll diese erstaunliche Entwicklung noch einmal verdeutlichen. Bis Ende der 90er Jahre waren AltaVista und Yahoo noch die meistgenutzten Suchdienste. Eine Statistik für den englischsprachigen Raum verdeutlicht dies [Tabelle 1].

Google, gestartet im Jahr 1998, konnte bis ins Jahr 2000 nur ca. 1% der weltweiten Nutzung aller Suchmaschinen verzeichnen.

Da die Suchmaschine durch ihre sehr einfache Bedienung und erstaunlich hohe Treffergenauigkeit Suchexperten und den Durchschnittsnutzer gleichermaßen überzeugen konnte, stieg seit Anfang 2000 die Nutzung sehr rasch an. Im Jahr 2001 führte Google erstmals die Rangliste der beliebtesten Suchdienste an. Seit dem hat sich der Vorsprung immer weiter vergrößert, stagnierte aber laut den Statistiken von OneStat in den letzten Monaten bei etwa 55% [Tabelle 2].

Da im deutschsprachigen Raum der Suchdienst Yahoo keine so große Rolle spielt, wie im englischsprachigen, ist der Anteil von Google hier noch höher. Der Dienst WebHits [4] wertet

täglich die Zugriffe auf zahlreiche Web-Seiten aus und gibt auch Informationen über die Nutzung von Suchmaschinen bekannt. Anfang Mai 2003 erreicht Google laut WebHits im deutschsprachigen Raum einen Anteil von 66% an der Nutzung aller Suchdienste [Tabelle 3].

Auch bei einem Vergleich der Anzahl der Suchanfragen pro Tag liegt Google vorne [Tabelle 4]. Obwohl diese Zahlen mit Vorsicht zu genießen sind, wird deutlich, dass Google mit Abstand die meist genutzte Suchmaschine ist.

Das Erstaunliche: Google verzichtet vollständig auf Werbebanner auf Web-Seiten oder Anzeigen in Print- und TV-Medien. Einzig und allein die durchweg positiven Kritiken von „Suchmaschinen-Experten“ und die sehr einfache Bedienung halfen der Suchmaschine innerhalb von zwei Jahren vom Insider-Tipp zur klaren Nummer 1 aufzusteigen. Anders als andere Suchmaschinen, konzentrierte sich Google zudem ausschließlich auf den eigentlichen Bereich der Indexierung von Internet-Seiten – auf die Entwicklung von kosten- und personalintensiven "Portalen", wie sie bei anderen Suchmaschinen Ende der 90er Jahre überall aus dem Boden sprossen, wurde bewusst verzichtet.

Die Beliebtheit von Google zeigt auch die Tatsache, dass die Suchmaschine bei den alljährlichen *Search Engine Awards* [6] in 9 von 13 Kategorien den Sieg einfahren konnte. Google wurde u.a. mit 65% der Stimmen zur besten Suchmaschine gekürt, auf Platz 2 landete Alltheweb mit gerade einmal 13% der Stimmen.

Inzwischen hat Google schon so viele Vorschusslorbeeren geerntet, dass Google sogar als Sieger in der Kategorie "Best News Search Engine" gewählt wurden, obwohl hier erst eine Beta-Version angeboten - im Vergleich dazu landeten die schon lang etablierten News-Suchmaschinen von Yahoo, Alltheweb und AltaVista nur auf den Plätzen.

Trend zu Monopolen – Gefahr oder Chance?

Seit einiger Zeit und besonders in den ersten Monaten diesen Jahres ist eine deutliche Konzentration auf dem Suchmaschinen-Sektor festzustellen. Neben Google gibt es eigentlich nur noch zwei ernstzunehmende Konkurrenten: Overture und Yahoo.

Overture erwarb im März 2003 kurz hintereinander die Suchmaschine AltaVista und die Suchmaschine Alltheweb, die bisher von der norwegischen Firma FAST (Fast Search & Transfer) betrieben wurde – zusammen für 210 Mio. Dollar (geradezu ein „Schnäppchen“, bezahlte die Holding CMGI im Juni 1999, in der Hoch-Zeit der New-Economy, für die Übernahme der Suchmaschine AltaVista von Compaq noch sage und schreibe 2,3 Mrd. Dollar).

Fast gleichzeitig erwarb Yahoo für 230 Mio. Dollar den Konkurrenten Inktomi. Noch erhält Yahoo übrigens die Suchergebnisse auch von Google – doch diese Partnerschaft wird zum Ende des Jahres 2003 aufgelöst.

Eine Grafik soll die entstandene Konzentration und die Abhängigkeiten der Suchdienste untereinander verdeutlichen [Grafik 1].

Diese Konzentration bietet Vorteile, bürgt aber auch Gefahren. Ein Vorteile ist z.B., dass hier drei finanzstarke Firmen miteinander konkurrieren, was sich durchaus positiv auf die Entwicklung intelligenterer Such- und Indexierungsmethoden auswirken kann – schließlich möchten Yahoo und Overture Google den ersten Rang unter den Suchmaschinen wieder

streitig machen und dies kann nur über zwei Wege führen: Mehr Inhalte und besseres Suchmethoden.

Aber auch die Gefahren sind nicht zu übersehen. Schon heute wird Google als eine Art „eierlegende Wollmilchsau“ betrachtet, die zu jeder Frage eine Antwort weiß. Doch auch Google ist „nur“ eine kommerzielle Firma, die bestrebt ist, Profit zu erwirtschaften. So hat Google bereits mehrmals Internet-Seiten aus dem eigenen Index entfernt, um einem Rechtsstreit zu entgehen [7]. Zudem kann niemand von Google – oder einer anderen Suchmaschine – verlangen, in den Suchmaschinen-Index aufgenommen zu werden. Die Konzentration führt auch dazu, dass es ganz einfach weniger Auswahl bei den Suchmaschinen gibt. Overture hat bereits in einer Presseerklärung [8] angekündigt, dass AltaVista in Zukunft nur noch der Erprobung neuer Arten der Werbung dienen soll während Alltheweb weiterhin für die Erprobung neuer Suchtechnologien benutzt werden soll.

Zu vermerken ist auch, dass die Suchmaschinentechologie der Suchmaschine "Scirus", die sich auf die Suche nach wissenschaftlichen Webseiten spezialisiert hat, derzeit von FAST bereitgestellt wird. Ob und wie sich die Übernahme von Alltheweb durch Overture z.B. auf Scirus auswirkt, bleibt abzuwarten.

Indexierungstechniken

Wie konnte Google in nur 2 Jahren vom absoluten Insider-Tipp zur eindeutigen Nr. 1 in der Suchmaschinen-Landschaft aufsteigen? Google ging einen neuen Weg, bei der Relevanzbewertung der Treffer. Einfach ausgedrückt: Die Rangliste der Treffer hing nicht mehr davon, wie oft und wo ein Wort im Dokument vorkommt, sondern wie viele Links auf ein Dokument verweisen.

Zu Verdeutlichung sollen zuerst die verschiedenen Indexierungstechniken der unterschiedlichen Suchdienste aufgezeigt werden und danach ein Überblick auf die verschiedenen Relevanzbewertungstechniken gegeben werden.

Bei den verschiedenen Indexierungstechniken muss zwischen drei Arten von Suchdiensten unterschieden werden: Internet-Verzeichnisse (z.B. Yahoo), allgemeine Suchmaschinen (z.B. Google) und Metasuchmaschinen (z.B. MetaGer).

Internet-Verzeichnisse

In Internet-Verzeichnisse, seien sie nun allgemein oder auch wissenschaftlich spezialisiert, werden die Dokumente von einem Redaktionsteam (also von Menschenhand) ausgewählt, geprüft und in ein passendes Themengebiet einsortiert. Je nach Verzeichnis und Themengebiet werden dabei unterschiedliche Qualitätsansprüche gestellt. Internet-Verzeichnisse können in Qualität und Größe sehr unterschiedlich sein und umfassen zwischen einigen tausend und einigen Millionen Internet-Seiten. Meist wird nur der Titel einer Internet-Seite und eine sehr kurze Zusammenfassung des Inhalts indexiert und gespeichert, nie der gesamte Volltext einer Seite.

Suchmaschinen

Suchmaschinen indexieren dagegen ganz anders: Automatisch arbeitende Programme ("Robots" oder "Crawler" genannt) durchsuchen das Internet Tag und Nacht nach neuen oder veränderten Seiten und speichern diese in riesigen Datenbanken. Diese „Robots“ folgen den

Links auf Internet-Seiten, ähnlich wie ein menschlicher Nutzer, der alle Links einer Seite anklickt. Als Start-Adressen werden dabei bereits in der Suchmaschinen-Datenbank vorhandene Internet-Adressen benutzt. Pro Tag können mehr als 100 Mio. Internet-Seiten indexiert werden.

Eine Filterung findet, wenn überhaupt, nur durch formale Kriterien statt (sei es nun eine bestimmte Sprache oder bestimmte Inhalte). Es gibt keine Gliederung oder Systematisierung. Alle Dokumente werden in einer großen Datenbank gespeichert, die mehrere Milliarden Seiten umfassen kann. Gespeichert wird in der Regel der komplette Text einer Seite, Wort für Wort, Zeichen für Zeichen. Ausnahme: Viele Suchmaschinen indexieren nur die ersten 100 KB einer Seite. Sehr lange Texte, insbesondere Dokumente im PDF-Format (z.B. elektronisch verfügbarer Dissertationen oder Zeitschriftenartikel) werden daher häufig nicht vollständig indexiert.

Die Zyklen, in denen diese Robots ihre Runde ziehen, variieren zwischen einigen Tagen und mehreren Wochen. Einige Suchmaschinen sind in der Lage, die Besuchsfrequenz der Robots gezielt danach zu steuern, wie häufig die zu indexierende Seite aktualisiert wird. Die meisten Suchmaschinen speichern die gesammelten Dokumente nicht komplett als Text. Zur Suche wird vielmehr eine Indextabelle angelegt, die die Wörter einer Seite in einer Ja - Nein - Struktur enthält. Mit Hilfe dieser Indextechnik lässt sich die Größe der indexierten Dokumente auf ca. 4% reduzieren.

Die hohe Suchgeschwindigkeit einer Suchmaschine wird dadurch erreicht, dass bei einer Suche lediglich die Indextabelle durchsucht wird. Ein simples Ja oder Nein führt zu einem Treffer - oder eben nicht.

Hieran wird auch noch ein zentraler Nachteil der heutigen Suchmaschinen deutlich: Die Suche erfolgt meist immer nur nach Zeichenketten. Wortbedeutungen, Pluralformen, Synonyme etc. werden überhaupt nicht berücksichtigt.

Dass es in Zukunft intelligentere Suchtechnologien geben muss, steht außer Frage. Welche dies sind und welche bereits im Einsatz sind, werde ich später noch beschreiben.

Metasuchmaschinen

Metasuchmaschinen machen sich die Vorarbeit, die Suchmaschinen oder Internet-Verzeichnisse leisten, zu nutze. Sie sehen auf den ersten Blick aus wie "normale" Suchmaschinen. Die Suchbegriffe werden in ein Eingabefeld eingegeben und die Suche wird gestartet. Metasuchmaschinen besitzen jedoch keine eigene Datenbank, in denen nach diesen Begriffen gesucht wird, sondern jede Suchanfrage wird an verschiedene andere Suchdienste weitergeleitet (dies können sowohl Suchmaschinen als auch Internet-Verzeichnisse sein, zum Teil sogar andere Metasuchmaschinen). In früheren Zeiten waren Metasuchmaschinen wichtige Dienste für die Informationsrecherche. Doch mit der deutlich verbesserten Qualität der Suchmaschinen, insbesondere im Hinblick auf Relevanzbewertung und Aktualität, spielen Metasuchmaschinen als Rechercheinstrument im Internet nur noch eine eher unbedeutende Rolle.

Relevanzbewertung

Sind die Dokumente von der Suchmaschine indexiert geht es nun daran, die Trefferliste nach einer Suche des Nutzers in irgendeiner Form aufzubereiten. Dies geschieht über die Relevanzbewertung, dem sogenannten Ranking. Es gibt verschiedene Arten des Rankings.

„Klassisches Ranking“

Das "klassische Ranking", wird wie folgt ermittelt: Die Suchmaschine prüft, wie oft das gesuchte Wort im Text vorkommt und wie häufig das Wort im Vergleich zur Textlänge ist. Die Relevanz steigt, wenn das Wort in der URL, im Titel, in Überschriften oder in den sogenannten Meta-Tags vorkommt.

Meta-Tags dienen dazu zusätzliche Informationen in eine Internet-Seite einzubinden. Dies können z.B. Autorennamen, Schlagwörter oder eine Kurzbeschreibung der Seite sein. Durch den zunehmenden Einsatz des sogenannten *Link-Rankings* haben die Meta-Tags allerdings an Bedeutung verloren. Im Internet haben diese Meta-Angaben, anders als in Datenbanken und Bibliothekskatalogen, nie eine große Rolle gespielt. Es gibt auch keine festgelegten Regeln, nach denen diese Meta-Tags ausgefüllt werden müssen. Häufig werden Meta-Tags sogar missbraucht um die Relevanz eines Dokumentes künstlich zu erhöhen, indem z.B. ein Wort mehrmals in ein Meta-Tag eingetragen wird oder in dem Wörter in Meta-Tags eingefügt werden, die mit dem eigentlichen Inhalt der Seite nichts zu tun haben. Suchmaschinen sind in letzter Zeit zunehmend davon abgegangen den Inhalt eines Meta-Tags überhaupt zu berücksichtigen. Einzig und allein die Kurzbeschreibung wird z.B. noch von Google erfasst.

Die Methode des *klassischen Rankings* lässt keine Schlüsse auf die Qualität eines Dokuments zu. Hier ist besonders die Intelligenz des Nutzers gefragt, d.h. er muss nachdenken, wie er Suchbegriffe möglichst so einsetzt und ggf. kombiniert, dass er sinnvolle Treffer erhält. Und selbst dann ist der Erfolg der Suche häufig Zufall. Je größer die Menge der Informationen, die durchsucht werden, desto unwahrscheinlicher ist es, mit dieser Methode eine erfolgreiche Suche durchzuführen.

Ein Grund warum die heutigen Suchmaschinen - allen voran Google - einen anderen Weg beschreiten: das *Ranking nach dem Grad der Verlinkung*. Dies ist auch ein Hauptgrund für den großen Erfolg der Suchmaschine Google.

Ranking nach Verlinkung (Link-Ranking)

Für Google ist bei der Relevanzbewertung nicht vorrangig entscheidend, wie oft und wo ein Wort auf der Seite vorkommt, sondern wie viele Links auf eine Seite führen. Dabei geht man von der (meist berechtigten) Annahme aus, dass ein Link auf eine Seite in gewisser Weise eine Empfehlung für diese ist. Entscheidend für die Ermittlung des Rankings ist auch, von *welcher* Seite der Link kommt. Jede Internet-Seite erhält von Google ein sogenanntes PageRank. Wer in seinem Browser die „Google Toolbar“ installiert hat kann diesen Wert auch selbst für jede Internet-Seite ermitteln [9].

Google benutzte anfangs zur Ermittlung des PageRank einige bekannte Startseiten im Web (z.B. Yahoo.com, aber auch einige Seiten amerikanischer Hochschulen). Diese erhielten in der Regel einen sehr hohen PageRank. Dieser PageRank-Wert wird nun auf die externen Links auf der Seite verteilt und so „vererbt“. Ein Link von einer Seite mit einem hohen PageRank bringt also mehr Punkte, als ein Link, von einer Seite, mit einem niedrigen PageRank.

Zusätzliche Ranking-Punkte gibt es, wenn der Suchbegriff als Linktext gefunden wird. Teilweise beschreibt der Text, der mit dem Link hinterlegt ist in höherem Maße den Inhalt einer Seite, als die dort genannten Informationen im Titel oder in den Meta-Tags.

Die dahinter steckende Technologie ist allerdings noch viel komplizierter und füllt dutzende Seiten mit technischen Beschreibungen [10].

Die Vorteile dieses Rankings liegen auf der Hand. Auch mit einfachsten Suchanfragen, die nur aus einem Wort bestehen, werden die vermeintlich relevantesten Seiten (eigentlich sind es die Seiten, auf die die meisten Links führen) an erster Stelle präsentiert. Dies sind durchaus auch häufig qualitativ hochwertige Angebote. Da die meisten Suchanfragen in Suchmaschinen nur aus ein oder zwei Wörtern bestehen, konnte Google als „Erfinder“ dieser Ranking-Technologie in kürzester Zeit große Erfolge feiern. Auch wenn andere Suchmaschinen, wie z.B. Alltheweb, inzwischen ein ähnlich gutes Ranking anbieten können, ist Google die klare Nr. 1 – denn gerade im Internet sind die Nutzer in dieser Hinsicht konservativ und bleiben bei einem bewährten Produkt.

Das Link-Ranking hat jedoch auch Nachteile. Ein großes Problem für die Suchmaschinen ist das sogenannte „Spamming“, das massenhafte Überfluten von Suchmaschinen durch falsche oder irreführende Informationen.

Google versucht dem Problem auf folgende Weise Herr zu werden: Verweise vom eigenen Host werden ignoriert, und eine große Zahl von Seiten, die ihrerseits nur den Zweck haben, Links zu erzeugen (sogenannte „Link-Farmen“), helfen ebenfalls wenig, da diese Seiten meist ein sehr niedriges PageRank haben.

Dennoch gelingt es verschiedenen Anbietern immer wieder das Link-Ranking zu missbrauchen und so ein künstlich erhöhtes Ranking bei bestimmten Suchabfragen zu erhalten.

Auswirkung auf das Suchverhalten

Das Link-Ranking hat sich auf das Suchverhalten der Nutzer ausgewirkt. Beim *klassischen Ranking* war die Intelligenz des Nutzers gefragt, er musste sich mit der Suchsyntax auskennen, wissen was ein Boolescher Operator oder ein Trunkierungszeichen ist und wie man diese richtig einsetzt. Die Qualität des Suchergebnisses hing im Wesentlichen von der Qualität der Suchabfrage ab.

Dies hat sich mit dem zunehmenden Einsatz des *Link-Ranking* geändert. Auch mit einfachsten Suchanfragen werden zumindest unter den Top 10 relevante Dokumente angezeigt. Die meisten Nutzer schauen sich in einer Trefferliste nur die erste und maximal noch die zweite Seite einer Trefferliste an (also maximal die ersten 20 Treffer). Ein Problem hierbei ist: Dem Nutzer wird suggeriert, dass er auch mit einfachsten Ein-Wort-Suchabfragen immer relevante Ergebnisse finden kann. Und sollte man einmal mit einer Ein-Wort-Suchabfrage nicht weiterkommen, genügt es meist ein zweites Wort hinzuzufügen. Bei einfachen Recherchen nach konkreten Informationen kommt man hierbei auch meist zum Erfolg, ganz anders sieht es bei wirklich intensiven wissenschaftlichen Recherchen aus.

Es gibt keine detaillierten Auswertungen der Suchanfragen, doch es ist zu vermuten, dass die Qualität der Suchabfragen durch Google eher abgenommen als zugenommen hat. Ein-Wort-Suchabfragen sind die Regel. Phrasensuche wird kaum eingesetzt. Die Trunkierung, in Datenbanken und Bibliothekskatalogen eine der wichtigsten Suchmöglichkeiten, ist in Google und anderen Suchmaschinen erst gar nicht möglich und daher auch kaum einem Nutzer bekannt. Die erweiterte Suchmaske, mit der auch in Google eine Suchanfrage z.T. hervorragend eingegrenzt werden kann, wird kaum genutzt. Schon eine eigene Studie von AltaVista aus dem Jahr 1999 ergab, dass nur wenige Nutzer die erweiterte Suchmaske benutzen.

Bezahltes Ranking

Nicht immer sind die Treffer, die unter den Top 10 einer Trefferliste zu finden sind, auch die relevantesten. Eine Besonderheit stellt hierbei das bezahlte Ranking in Suchmaschinen dar. Eine häufig gestellte Frage: Wie verdienen Suchmaschinen überhaupt Geld? Dies geschieht häufig dadurch, in dem Werbung in die Trefferliste eingebunden wird. Man unterscheidet drei verschiedene Werbeformen.

- „Keyword-abhängige“ Werbung
- „sponsored results“
- bezahltes Ranking

„Keyword-abhängige“ Werbung wird sehr häufig in Suchmaschinen eingesetzt. Eine Firma oder auch eine Privatperson kann bei einer Suchmaschine eine Anzeige schalten, die nur nach Eingabe bestimmter Suchbegriffe eingeblendet werden soll. Die Preise für eine solche Anzeige hängen in der Regel vom gesuchten Wort ab. So kostet eine Anzeige, die bei der Suche nach „webdesign“ erscheinen soll, mehr als eine Anzeige, die bei der Suche nach „bibliothek“ auftauchen soll, weil ganz einfach der Begriff „webdesign“ häufiger gesucht wird als der Begriff „bibliothek“. Auch Werbebanner werden z.T. abhängig von der Suchanfrage eingeblendet, z.B. in der Suchmaschine AltaVista.

Neben dieser Werbeform gibt es die sogenannten „sponsored results“. Sie tauchen in der Regel ganz am Anfang einer Trefferliste auf und sind – mehr oder weniger deutlich – als „Werbung“ gekennzeichnet.

Google hat auch hier eine Vorreiter-Rolle eingenommen, in dem es bezahlte Links deutlich als „Anzeige“ kennzeichnet, während in anderen Suchmaschinen nur mehr oder weniger versteckt ein Hinweis auf einen bezahlten Link angebracht ist. Bei AltaVista findet man z.B. je nach Suchbegriff bis zu 10 diese „sponsored results“, die am Anfang der Trefferliste zu finden sind.

Das *bezahlte Ranking* ist die undurchsichtigste Form der Werbung. Einfach ausgedrückt: Das Ranking (in der Regel die Top 10) richtet sich danach, wer am meisten bezahlt um bei einem bestimmten Suchbegriff ganz vorne zu landen – unabhängig von der tatsächlichen Qualität und dem Inhalt der Seite. Suchmaschinen-Betreiber Overture betreibt mit dieser Strategie recht erfolgreich seine Suchmaschine.

Das „Invisible Web“

Wenden wir uns noch einmal dem Indexieren von Dokumenten zu. Die bisher vorgestellten Suchdienste decken weitgehend das sogenannte „Visible Web“ ab, das für Suchmaschinen sichtbare Netz. Suchmaschinen indexieren in der Regel nur statische, textbasierte Internet-Seiten. Auch das weitverbreitete PDF-Format, in dem häufig Zeitschriftenartikel in elektronischer Form gespeichert sind, wird von vielen Suchmaschinen indexiert. Hier sei aber noch mal die „Grenze“ von ca. 100 KB erwähnt, die häufig dazu führt, dass diese Dokumente nicht vollständig indexiert werden –insbesondere die Literaturverzeichnisse, die sich in der Regel am Ende eines Dokuments befinden und häufig viele relevante Informationen enthalten, werden deshalb oft gar nicht indexiert.

Überhaupt nicht indexiert werden Nicht-Text-Informationen (Grafiken, Video, „Flash“-Filme), dynamisch generierte Seiten und natürlich passwort-geschützte Inhalte.

Nach einer Studie von Brightplanet [11], ist die Datenmenge, die im „Invisible Web“ zu finden ist, mehr als 400 mal so groß wie die des sichtbaren Webs. Suchmaschinen decken also nur einen kleinen Teil aller verfügbaren Informationen ab.

Inhalte aus Datenbanken und Bibliothekskatalogen machen einen großen Teil des „Invisible Web“ aus.

Warum sind diese Daten für Suchmaschinen eigentlich „Invisible“? Dazu ein Beispiel: Führt ein Nutzer eine Suche in einem Bibliothekskatalog durch, wird das Ergebnis zwar auch als normale Internet-Seite präsentiert – doch diese wird dynamisch generiert und ist nur für diesen Nutzer sichtbar. Sie wird erst sichtbar, nachdem ein Nutzer eine Suchanfrage gestellt hat. Erkennen kann man solch dynamisch generierten Seiten in der Regel an sehr langen URLs oder an einem oder mehreren Fragezeichen innerhalb der URL.

Die Robots, die für die Suchmaschinen die Seiten sammeln, können sich nur von Link zu Link weiterbewegen. Sie können keine Suchmasken ausfüllen und keine Passwörter eingeben. Deshalb bleibt für eine Suchmaschine eine solche Seite unsichtbar.

Mit verschiedenen Tricks lässt sich das „Invisible Web“ zumindest zum Teil für Suchmaschinen sichtbar machen. So kann man theoretisch von einer statistischen Internet-Seite aus einen Link z.B. auf eine Ergebnisseite aus einem Bibliothekskatalog legen. Klickt ein Nutzer auf den Link, wird im Hintergrund automatisch eine Suche im Katalog durchgeführt und das Ergebnis präsentiert. Da man in diesem Fall keine Suchmaske ausfüllen muss, können Suchmaschinen diese Ergebnisseite dann indexieren – sie wird sozusagen auch für Suchmaschinen sichtbar.

Diese Methode ist natürlich nicht möglich, wenn für eine Suche z.B. ein Login erforderlich ist oder für jede Suchabfrage eine sogenannte Session-ID vergeben wird. Diese Session-ID's sind in der Regel nur wenige Minuten gültig, dann ist eine solche Suchanfrage auch per Direktlink nicht mehr zu erreichen. Diese Seiten bleiben für Internet-Suchmaschinen unsichtbar.

Teaser-Informationen: Mehrwert oder Datenmüll?

Über die oben gezeigte Methode der *Direkt-Verlinkung* von Suchergebnissen können Inhalte aus Datenbanken und Bibliothekskatalogen, die eigentlich dem „Invisible Web“ zuzurechnen sind, für Suchmaschinen sichtbar gemacht werden.

Das Hochschulbibliothekszentrum Nordrhein-Westfalen in Köln ist dabei im Projekt „Virtuelles Bücherregal NRW“ die einzelnen Datensätze der eigenen Verbunddatenbank ins HTML-Format umzuwandeln, um sie über eine Suchmaschine suchbar zu machen. Anfang Mai waren ca. 13 Mio. der 20 Mio. Datensätze bereits umgewandelt [12]. Zudem sind die Datensätze untereinander verlinkt, wodurch eine Indexierung durch eine Suchmaschine ja erst möglich ist. Von Datensatz Nr. 1 gibt es einen Link zu Datensatz Nr. 2 usw. bis zum letzten Datensatz.

Auf diese Weise sind bisher über 175.000 Datensätze in den Google-Index gelangt [13], aber die Zahl kann durchaus noch steigen, denn eine Begrenzung gibt es für Google offensichtlich nicht.

So können nun auch Datensätze aus der HBZ-Verbunddatenbank über Internet-Suchmaschinen gefunden werden und sind damit auch Nutzern zugänglich, die bisher nie in Bibliothekskatalogen recherchiert haben.

Dieses Verfahren kann jedoch auch zu einer Verfälschung des Rankings missbraucht werden. Dies soll am Beispiel der „Königstraum-Datenbank“ des Verlags für die Deutsche Wirtschaft

AG gezeigt werden [14]. Zu jedem Stichwort, welches in der Datenbank zu finden ist, wurde automatisch eine eigene (statische) Internet-Seite generiert. Diese Seiten enthalten einen Link auf die Datenbank des Verlags. Klickt man auf den Link wird eine Liste der verfügbaren Dokumente zu diesem Stichwort gezeigt. Klickt man nun auf eines der Dokumente erhält man nur die Information, dass dieses Dokument kostenpflichtig ist. Die automatisch generierten Internet-Seiten sollen also nur einen Anreiz schaffen, in die Datenbank einzusteigen und ggf. Kunde zu werden, um sich die Dokumente ansehen zu können. Daher spricht man auch von Informations-„Teasern“ oder „Teaser“-Dokumenten.

Die Teaser-Dokumente selbst sind alle miteinander verlinkt. Von jedem Teaser-Dokument führen bis zu 300 versteckte Links auf andere Teaser-Dokumente. Dadurch soll ein hohes Link-Ranking erreicht werden. Durch geschickte Platzierung der Suchbegriffe im Dokument wird zusätzlich eine hohe Relevanz suggeriert. Ein erfolgreiches Konzept. Gibt man z.B. den Begriff „Zettelflut“ bei Google ein, wird das Teaser-Dokument aus der Königstraum-Datenbank an erster Stelle gelistet [Grafik 2].

Die Suchmaschine Alltheweb hat aus der Königstraum-Datenbank bereits fast 450.000 Dokumente indexiert [15].

Für den Verlag eine zusätzliche und durchaus profitable Möglichkeit Kunden zu gewinnen. Für Suchmaschinen allerdings auch eine Gefahr. Leicht können Suchmaschinen durch solche Teaser-Dokumente überflutet werden, die sogar die Zahl echter Internet-Seiten um ein vielfaches übersteigen könnten.

Noch spielen solche Teaser-Informationen in den Datenbanken der Suchmaschinen eher eine geringe Rolle, doch die Zahl wächst unaufhörlich. An zahlreichen Stellen im Web findet man zudem „Tipps“ wie man sich ein hohes Link-Ranking in Suchmaschinen verschaffen kann – ohne auf „echte“ Links angewiesen zu sein. Wenn diese Verfahren Schule machen, kann es durchaus zu einem Problem für Suchmaschinen damit auch für die Nutzer von Suchmaschinen werden.

Natürlich sind Suchmaschinen nicht wehrlos gegenüber dem Missbrauch des Link-Rankings. In einer Art Wettlauf werden auch bei den Suchmaschinen-Betreibern immer ausgefeiltere Methoden entwickelt, Link-Farmen oder Teaser-Seiten zu entdecken. Google droht auch beim erwiesenen Missbrauch mit der Löschung aller Seiten des Servers aus dem Index. Und da ein Suchmaschine-Betreiber nicht verpflichtet werden kann, eine bestimmte Seite zu indexieren, ist dies rein rechtlich gesehen auch völlig unproblematisch möglich.

Die Auswirkungen auf den Bibliotheksbereich

Wie schon beschrieben hat sich das Quasi-Monopol der Suchmaschine Google und der Einsatz des Link-Rankings auf das Suchverhalten der Nutzer ausgewirkt. Google wird auch in immer stärkeren Maße – gerade von Studierenden – für die wissenschaftliche Recherche eingesetzt. Als einziger Suchdienst eingesetzt, wird es hierbei zwangsläufig zu unzureichenden Ergebnissen kommen. Wenn Bibliothekskataloge und Datenbanken eingesetzt werden, so werden sie häufig ähnlich wie eine Suchmaschine bedient. Ein oder zwei Stichwörter werden in ein Eingabefeld eingegeben, und die Suche wird abgeschickt. Dass man mit dieser Herangehensweise selten zum Erfolg kommt, und nicht einmal annähernd die z.T. sehr viel fortgeschrittenen Suchmöglichkeiten genutzt werden, ist den meisten Nutzer nicht bewusst.

Die Frage ist: Müssen die Nutzer intelligenter werden oder die Kataloge und Datenbanken? Zum einen sollte sicher die Vermittlung professioneller Recherchekennntnisse eine größere Rolle im Studium spielen. Hier bieten Bibliotheken durch Schulungen und Hilfetexte zwar schon häufig eigene Angebote, doch werden sie nur selten von den Studierenden genutzt. Daher ist hier zumindest eine engere Zusammenarbeit mit den Fakultäten in den Universitäten nötig, um z.B. direkt in den Seminaren *Recherchekompetenz* vermitteln zu können.

Doch ist dies nur ein Teil, wie in Zukunft bessere Rechercheergebnisse erreicht werden können. Der andere Weg ist: Auch Bibliothekskataloge und Datenbanken müssen in Zukunft intelligenter werden, die Such-Interfaces müssen einfacher werden und die Qualität der Suchergebnisse auch mit einfachsten Suchanfragen so gut wie möglich werden. Hier sollte man sich Google durchaus als Vorbild nehmen.

Ranking in Bibliothekskatalogen?

Ein Problem, welches bei der Suche in Bibliothekskatalogen auftritt, ist das fehlende oder unzureichende Ranking. Die bisherigen Ranking-Methoden basieren auf rein formale Angaben. So werden die Treffer häufig nach Erscheinungsjahr (die neuesten Treffer zuerst) sortiert. Zwar kann man zumindest die neueren Monographien auf diese Weise an den Anfang einer Trefferliste bringen, aber Zeitschriften erscheinen z.B. dann erst am Ende der Trefferliste, da sie i.d.R. kein Erscheinungsjahr haben.

Dabei könnte man sich durchaus auch eine andere Art der Sortierung vorstellen. Eine Link-Ranking kann nicht stattfinden, denn die einzelnen Treffer aus einem Katalog oder einer Datenbank können in der Regel nicht verlinkt werden.

Doch um die „Wichtigkeit“ eines Buches zu ermitteln, wäre es z.B. möglich zu prüfen, wie oft ein Buch ausgeliehen wurde oder auch in wie vielen Exemplaren ein Buch vorhanden ist. Je häufiger ein Buch ausgeliehen wird oder in je mehr Exemplaren es vorhanden ist, desto höher ist das Ranking.

Ein Weg wäre z.B. auch die dynamische Verlinkung einzelner Dokumente. Der Online-Buchhändler Amazon bietet zu jedem Buch unter der Überschrift „Kunden, die dieses Buch gekauft haben, haben auch folgende Bücher gekauft“ mehrere Links zu vermeintlich passenden Büchern an. Erstaunlicher Weise gelangt man auf diese Weise häufig zu relevanter themenverwandter Literatur, ohne dass einem Titel oder Autor bekannt sind. Ähnliches wäre auch für Bibliothekskataloge möglich („Kunden, die dieses Buch ausgeliehen haben, haben auch folgende Bücher ausgeliehen“).

Natürlich ist hier der Datenschutz von entscheidender Bedeutung: Denn um Informationen dieser Art zu generieren, ist eine Auswertung von Ausleihdaten der Nutzer nötig, die – wenn überhaupt – streng anonymisiert erfolgen darf.

Manche diese Ideen mögen ein wenig utopisch erscheinen und kommen auch aus technischen Gründen häufig nicht in Frage. Daher müssen auch andere Weg berücksichtigt werden, um bessere Trefferlisten zu erreichen.

Linguistische Methoden

Das Link-Ranking ist eine hervorragende Möglichkeit, relevante Dokumente aus einem riesigen, allgemeinen, unstrukturierten Datenpool herauszufischen. Bei einem kleineren, strukturierten Datenpool, ist das Link-Ranking nicht sinnvoll.

Es empfiehlt sich hier eine andere Methode, um relevantere Ergebnisse auch bei einfachen Suchabfragen zu erzielen: Die sogenannten „linguistischen Methoden“. Mithilfe linguistischer Methoden sollen Suchmaschinen „intelligent“ werden, sowohl bei der Indexierung der Dokumente als auch beim Erkennen einer Suchanfrage.

Bereits bei der Indexierung von Volltext-Dokumenten kann mittels linguistischer Methoden folgendes erkannt werden:

- Spracherkennung: Dazu gehört, dass erkannt wird, in welcher Sprache das Dokument geschrieben ist und welcher Zeichensatz verwendet wird
- Entitäten-Erkennung: Wichtige Elemente im Dokument werden automatisch erkannt (z.B. Personennamen, Firmennamen, geographische Bezeichnungen aber auch Abstracts oder Konzeptlisten)
- Textklassifikation: Automatische Textklassifikation anhand von Schlüsselbegriffen, die im Dokument vorkommen. Texte können sowohl formal klassifiziert werden (z.B. kann festgestellt werden, ob es sich um ein Abstract, ein Preprint oder eine Firmenhomepage handelt) als auch fachlich klassifiziert werden (einen ersten – wenn auch nicht sehr erfolgreicher - Versuch, Dokumente automatisch zu Klassifizieren, wurde im Projekt GERHARD [16] unternommen)

Weniger Sinn machen diese linguistischen Methoden natürlich bei rein bibliographischen Datensätzen. Doch auch bei der *Recherche* nach Dokumente können linguistische Methoden helfen, die Suchanfrage automatisch zu verbessern:

- Normalisierung der Schreibung: Bei der Eingabe von Suchbegriffen werden Akzente (z.B. á), Umlaute und bestimmte Schreibformen vereinheitlicht, sodass z.B. bei der Eingabe des Wortes „graphik“ auch das Wort „grafik“ gesucht wird
- Morphologische Normalisierung: Dazu gehören automatische Trunkierung, das sogenannte „Word-Stemming“ (automatische Pluralsuche), die Lemmatisierung (Reduktion des Wortes auf die Grundform) und die Kompositazerlegung (so könnte z.B. bei der Suche nach „Arbeitslosenzahl“ automatisch auch nach „Zahl der Arbeitslosen“ gesucht werden)
- „Phrasing“ und „Anti-Phrasing“: Umwandlung bestimmter Suchbegriffe in Phrasen (gibt ein Nutzer z.B. die Suchbegriffe *new york* ein, wird diese automatisch als Phrase gesucht). Auch der umgekehrte Fall ist möglich, das „Anti-Phrasing“. Gibt ein Nutzer einen ganzen Satz als Phrase ein (z.B. „Wo finde ich Informationen über Bielefeld?“) werden die Schlüsselwörter herausgefiltert (*informationen bielefeld*) und nur nach diesen gesucht

Die norwegische Firma FAST arbeitet bei der Entwicklung dieser linguistischen Methoden mit dem CIS (Centrum für Informations- und Sprachverarbeitung) in München zusammen und hat einige auch schon zum Einsatz gebracht. So werden z.B. in der Suchmaschine Alltheweb Suchbegriffe bei Bedarf automatisch als Phrase gesucht.

Einige dieser linguistischen Methoden werden auch in der wissenschaftlichen Suchmaschine „Scirus“, die ebenfalls von FAST betrieben wird, angewandt. So kann man in Scirus z.B. nach verschiedenen „Informationstypen“ (information types) suchen, u.a. Abstracts, Bücher, Preprints oder Homepages von Firmen und Wissenschaftlern. Diese Informationen wurden durch Entitäten-Erkennung bei der Indexierung automatisch zugeordnet.

Intelligente Suchmaschinen für das „Invisible Web“ – eine Aufgabe für Bibliotheken

Bei der Entwicklung neuer, „intelligenter“ Suchmaschinen – die insbesondere auch die Informationen des „Invisible Web“ abdecken sollen, ist es wünschenswert, dass auch Bibliotheken ihr Know-how einbringen.

Zum Abschluss möchte ich zwei Projekte aus diesem Bereich, die derzeit an verschiedenen Bibliotheken in Deutschland durchgeführt werden, kurz vorstellen. Das Projekt „EASE – European Academic Search Engine“, welches von der Universitätsbibliothek Bielefeld [17] koordiniert wird und das Projekt „Verteilter Zeitschriften Server (VZS)“, welches vom KOBV in Berlin [18] koordiniert wird.

EASE – European Academic Search Engine

Im Projekt "EASE - European Academic Search Engine" wurde ein Konzept erarbeitet, um wissenschaftlich relevante Online-Information in unterschiedlichsten Formaten zukünftig mit intelligenter industrieller Web-Suchmaschinenteknologie zu indexieren und in einem verteilten europäischen Index zusammenzuführen. Eine Vielzahl unterschiedlicher Sichten, d.h. Search & Retrieval-Interfaces, soll dann den Zugriff auf diesen Index ermöglichen. Diese unterschiedlich konfigurierten Such-Interfaces können theoretisch in die jeweils vorhandenen Informationsumgebungen einer Bibliothek oder einer virtuellen Fachbibliothek ohne zusätzliche Hard-/Software integriert werden.

Koordiniert wird das Projekt von der UB Bielefeld in Kooperation mit der norwegischen Firma FAST, der Firma Exorbyte aus Konstanz sowie 13 weiteren Institutionen (Forschern und Forschergruppen aus München, Hannover, Padua und Oxford, Bibliotheken und Bibliotheksverbundzentren). Zur Finanzierung des Projekts wurde ein Antrag im Rahmenprogramm 6 (FP6) der EU gestellt.

VZS – Verteilter Zeitschriften Server

Das Projekt Verteilter Zeitschriftenservers (VZS) setzt sich die Speicherung und das integrierte Angebot lizenzierter digitaler Zeitschriften und freier digitaler Dokumente zum Ziel.

Elemente des „Invisible Web“ und des „Visible Web“ sollen nahtlos miteinander verbunden werden, um Nutzern die Navigation und Suche unter einer einheitlichen Oberfläche zu ermöglichen. Mit dem Vorhaben soll auf diesem Gebiet Kompetenz für offene Navigation und für Suchmaschinen aufgebaut werden, um entsprechende Aktivitäten in Bibliotheken, Bibliotheksverbänden und Konsortien sowie in den Wissenschaften zu initiieren.

Im Rahmen eines Teilprojekts soll ein Entwicklungssystem der Suchmaschinensoftware FAST Data Search an der UB Bielefeld installiert werden.

Das HBZ wird im Rahmen dieses Projektes seine Erfahrungen aus dem Projekt „Virtuelles Bücherregal NRW“ einbringen.

Schlussbemerkung

Immer noch gibt es große Defizite bei der Internet-Recherche, sowohl auf Seiten der Nutzer als auch auf Seiten der allgemeinen Internet-Suchmaschinen.

Suchmaschinen sind einfach zu bedienen, bieten aber nicht die ausgefeilten

Suchmöglichkeiten von Bibliothekskatalogen und Datenbanken und sind daher oft

unzureichend für die wissenschaftliche Recherche. Zudem decken sie den Bereich des

„Invisible Web“, in dem ein Großteil wissenschaftlicher Informationen zu finden ist, gar nicht ab.

Unter anderem bieten die beiden genannten Projekte jedoch bereits erfreuliche Ansätze die einfache Bedienbarkeit der Suchmaschinen – die zweifellos ein Hauptgrund für ihre Beliebtheit ist – auch auf die Welt des „Invisible Web“ und die wissenschaftliche Recherche zu portieren, ohne dass Abstriche bei der Qualität gemacht werden müssen. Die weitere Entwicklung auf diesem Gebiet bleibt mit Spannung abzuwarten.

Tabellen:

Tabelle 1

Verteilung der Suchmaschinen-Nutzung laut Statmarket.com [2] von Dezember 1999

Yahoo	56%
AltaVista	11%
Excite	10%
Go	8%

Tabelle 2

Verteilung der Suchmaschinen-Nutzung laut Onestat.com [3] von Januar 2003

Google	54,7%
Yahoo	22%
MSN	9,5 %
AOL	3,7%

Tabelle 3

Verteilung der Suchmaschinen-Nutzung laut WebHits [4] von Mai 2003 für den deutschsprachigen Raum

Google	66,1%
Yahoo	9,4%
MSN	6,2%
T-Online	3,5%
Lycos	2,2%
AOL*	2,1%
Fireball	1,5%
MetaGer	1,4%
Web.de*	1,4%
AltaVista	1,2%

* AOL und Web.de verwenden in ihren Suchmaschinen den Index von Google, angereichert um Einträge aus eigenen Katalogen. Zählt man diese Webseiten, hinzu, gelangt man zusammengerechnet auf einen Anteil von ca. 70% für Google

Tabelle 4

Statistik von SearchEngineWatch [5] - Anzahl der Suchabfragen pro Tag (Stand: Februar 2003)

Google	250 Mio. (inkl. Abfragen über Partner-Sites, u.a. Yahoo und AOL)
Overture	167 Mio. (inkl. Abfragen über Partner-Sites)
Inktomi	80 Mio.
AltaVista	18 Mio.
Alltheweb	12 Mio.

Grafiken:

Grafik 1:

Diese Grafik soll das Netzwerk verdeutlichen, welches inzwischen unter den verschiedenen Internet-Suchmaschinen entstanden ist. Google, Yahoo und Overture „kontrollieren“ den Markt und sind selbst noch untereinander vernetzt.



Grafik 2

Suche in Google nach dem Begriff „Zettelflut“. Ein erster Stelle wird ein Teaser-Dokument aus der Königstraum-Datenbank gefunden.

Das Bild zeigt den Suchmaschinen-Interface von Google. Die Suchbox enthält den Text 'zettelflut'. Die Suchergebnisse sind wie folgt dargestellt:

Zettelflut - Alle Infos auf Koenigstraum.de
Alles zum Thema **Zettelflut** und zu vielen weiteren Themen hält die Koenigstraum.de-Datenbank zum Abruf für Sie bereit. ... **Zettelflut** ...
www.koenigstraum.de/html/zettelflut.htm - 22k - Im Cache - Ähnliche Seiten

Links & Literatur:

- [1] Befragt wurden insgesamt 624 Studierende der Universität Bielefeld im November 2002
- [2] <http://www.statmarket.com/cgi-bin/sm.cgi?sm&feature&stat1221999>
- [3] http://www.onestat.com/html/aboutus_pressbox17.html
- [4] <http://www.webhits.de/deutsch/index.shtml?/deutsch/webstats.html>
- [5] <http://searchenginewatch.com/reports/perday.html>
- [6] Die SearchEngineAwards sind ein Preis, den Internet-User alljährlich vergeben können. Weitere Infos unter <http://searchenginewatch.com/awards/2002-winners.html>
- [7] <http://www.suchfibel.de/news/arc0-2003.htm>
- [8] Zitat aus der Presseerklärung Overtures: „The company [Overture] will use FAST's technology showcase site, AlltheWeb.com, to test and experiment with advanced approaches in search. At the same time, the company will use the AltaVista.com site, which operates on a larger scale, to refine implementations for new products and improve presentation to consumers.” (http://www.corporate-ir.net/ireye/ir_site.zhtml?ticker=OVER&script=410&layout=0&item_id=385605)
- [9] Die Google Toolbar kann unter <http://toolbar.google.com/intl/de/> heruntergeladen werden. Der PageRank-Wert wird in der GoogleToolbar in einen Faktor von 1 bis 10 dargestellt.
- [10] Details zur PageRank-Technologie unter: <http://www.at-web.de/Websitepromotion/pagerank.htm>
- [11] <http://www.brightplanet.com/deepcontent/tutorials/DeepWeb/index.asp>
- [12] Die aktuelle Zahl ist zu finden unter: http://www.hbz-nrw.de/produkte_dienstl/virkiv.html. Mehr Infos unter: <http://kirke.hbz-nrw.de/dcb/index.html>
- [13] Die genaue Zahl kann überprüft werden unter der URL: <http://www.google.de/search?q=inurl:kirke.hbz-nrw.de/dcb/&hl=de&lr=&ie=UTF-8&oe=UTF-8&filter=0>
- [14] Siehe hierzu den Beitrag von Christian Heinisch: Suchmaschinen als Wegweiser zu Inhalten des „Invisible Web“. in Password, April 2003, S. 31-33
- [15] Die genaue Zahl kann überprüft werden unter der URL: http://www.alltheweb.com/search?cat=web&cs=iso-8859-1&q=site%3Akoenigstraum.de&_sb_lang=de+en
- [16] <http://www.gerhard.de>
- [17] <http://www.ub.uni-bielefeld.de>
- [18] Kooperativer Bibliotheksverbund Berlin-Brandenburg (<http://www.kobv.de/>)