

*Google Print, Amazon, Yahoo und OCA*

# Die Search Engines erobern die Buchwelt

**Während in der Berichterstattung über die Buchmesse viel auf Google Print und dessen neues deutschsprachiges Suchinterface hingewiesen wurde, hat sich Yahoo (im Rahmen einer Kooperation) daran gemacht, komplette Bücher einzuscannen und für die Suche verfügbar zu machen. Nimmt man Amazon ("Search Inside a Book") hinzu, so laufen demnach gleich drei Scan-Projekte.**

Nach Google (mit Google Print) steigt Yahoo in die Digitalisierung gedruckter Bücher ein. Anders als bei Google, dessen Angebot sich sowohl auf gemeinfreie Werke als auch auf urheberrechtlich geschütztes Material richtet, will Yahoo nur Werke digitalisieren, die bereits der Allgemeinheit gehören. Auch in der Konstruktion des Projekts findet sich ein wesentlicher Unterschied: Yahoo will die Arbeit nicht allein in Angriff nehmen, sondern kooperiert im Rahmen der neu gegründeten "Open Content Alliance" (OCA; [www.opencontentalliance.org](http://www.opencontentalliance.org)) mit dem Internet Archive, Hewlett Packard Labs und Adobe Software. Die Inhalte sollen von verschiedenen amerikanischen und europäischen Universitäten kommen. In der Vergangenheit sah sich Google Print der Kritik ausgesetzt, weil sich das Projekt auf Bestände amerikanischer Bibliotheken beschränkte. Die OCA fordert explizit Verlage, Bibliotheken und Universitäten auf der ganzen Welt auf, sich an dem Projekt zu beteiligen. Das ist sicher auch eine Antwort auf den jüngst von Jean-Noël Jeanneney, dem Präsidenten der französischen Nationalbibliothek, wiederholten Vorwurf, die Digitalisierung durch Google beschränke sich allein auf die angloamerikanische Bestände und damit deren Weltsicht.

Mit der Beschränkung auf frei zugängliche Werke verfolgt die OCA eine Verfügbarmachung des kompletten Inhalts des jeweiligen Buchs in einer Datei. Während man bei Google jede Seite einzeln aufrufen muss, werden die von OCA zu Verfügung gestellten Bücher jeweils als eine PDF-Datei verfügbar sein. Damit

lassen sich sie bequem herunterladen, ausdrucken und weiterverarbeiten. Warum Google dies bei gemeinfreien Werken nicht ebenso hält, ist unklar. Weiterhin wird durch diese Art der Präsentation der Inhalte gewährleistet, dass der Zugriff auf die Werke für alle - auch für andere Suchmaschinen - möglich ist. Die PDF-Dateien sollen beim Internet Archive abgelegt werden, wo sie dann von den Suchmaschinen indexiert werden können. Suchmaschinen wie Yahoo und MSN erfassen übrigens auch längere PDF-Dateien problemlos, während Google die Indexierung nach etwa 100 KB abbricht.

## Open Content Alliance

### Die Unterschiede zu Google Print

Nachdem die Open Content Alliance (OCA) als ein "permanent archive of multilingual digitalised text and multimedia content" gestartet ist - Gründungsmitglieder sind Yahoo!, das Internet Archive, HP Labs, Adobe, die University of California, die University of Toronto, das European Archive, das britische National Archive und O'Reilly - was unterscheidet es von Google Print?

Auch OCA will weltweit Bücher digitalisieren, aber anders als bei Google soll das nur geschehen, wenn die Verleger ausdrücklich die Erlaubnis dafür erteilt haben. Entsprechend geringere Schwierigkeiten wird OCA von den Verlegern bekommen, aber auch entsprechend kleiner mag das Projekt ausfallen. OCA will auch Audio und Video verfügbar machen, während sich Google Print derzeit auf Texte beschränkt. Google Print wird über Google recherchierbar gemacht, die OCA-Inhalte über alle Suchmaschinen. Google Print ist offenkommerziell, während OCA als ein weitgehend gemeinnütziges Projekt auftritt. Aber wenn OCA gemeinnützig wäre, was machen dann Yahoo! und O'Reilly darin, und was sind ihre Interessen?

[www.opencontentalliance.org](http://www.opencontentalliance.org)  
und [www.archive.org](http://www.archive.org)

Die OCA will grundsätzlich auch geschützte Werke von Verlagen annehmen und arbeitet zurzeit in Kooperation mit Verlagen an einem Geschäftsmodell. Als erster Verlag hat sich O'Reilly bereit erklärt, Bücher einzustellen - allerdings momentan ohne jede Form der Vergütung. Google wirbt um die Verlage mit zwei Argumenten: Einerseits lassen sich durch das Verfügbarmachen von Werken die Verkäufe erhöhen. Andererseits werden auf den Seiten, die die Bücher präsentieren, Anzeigen gezeigt. Die Erlöse aus diesen Anzeigen werden zwischen Google und dem Verlag geteilt. Bei eingescannten Bibliotheksbüchern wird auf Anzeigen verzichtet.

Ein wesentlicher Unterschied zwischen Google und OCA besteht darin, dass die Open Content Alliance geschützte Werke nur auf Initiative der Verlage aufnimmt, während Google grundsätzlich alles ohne vorherige Absprache scannt, die geschützten Werke aber nur in Ausschnitten zugänglich macht. Gerade diese Herangehensweise ist es, die Google den Ärger mit Verlegern und Autoren gebracht hat (Verbände beider Interessengruppen haben Google in den USA verklagt).

Der dritte Anbieter von gescannten Büchern im Volltext, Amazon, konzentriert sich auf aktuelle Bücher, die im Handel erhältlich sind. Im Rahmen des "Search Inside a Book"-Programms werden auch hier Titel auf ähnliche Weise wie bei Google Print verfügbar gemacht: Die Titelseite, der Klappentext, das Inhaltsverzeichnis und das Register lassen sich von jedem Nutzer vollständig einsehen. Weiterhin kann im Buch gesucht werden. Wird ein Suchergebnis aufgerufen, mag man ähnlich wie auch bei Google Print zwei Seiten vorwärts und rückwärts blättern. Um "Search Inside" zu nutzen, muss sich der Nutzer mit seinem Amazon-Passwort einloggen. Auch die Amazon-eigene Suchmaschine A9.com bietet neben der Websuche eine gleichzeitige Suche nach Büchern, die auch die eingescannten Werke berücksichtigt.

Wie sind die drei Angebote im Vergleich zu sehen? Bei Amazon ist die Ziel-

setzung des Projekts am Klarsten: Der Verkauf der gedruckten Bücher soll gefördert werden. Die Open Content Alliance hat sicher die ehrenhafteste Zielsetzung, nämlich die gescannten Werke allgemein zugänglich zu machen. Yahoo ist hier nur ein Partner innerhalb des Projekts und stellt die Suchtechnologie zur Verfügung. Eine weitergehende Initiative seitens Yahoo (etwa die Finanzierung des kostspieligen Scannens) ist nicht zu erwarten.

Die Zielsetzung von Google Print bleibt unklar: Es geht sicher nicht allein darum, "die Informationen der Welt zugänglich zu machen", wie das Unternehmen behauptet. Neben dem Anzeigengeschäft ist langfristig eher an einer weitergehenden Zusammenarbeit mit den Verlagen zu denken: Wenn die Bücher schon einmal elektronisch vorliegen, dann könnte man sie gleich in dieser Form verkaufen. Google partizipierte in diesem Fall am Gewinn; die gemeinfreien Werke wären eher dafür da, das Angebot möglichst vollständig zu machen. Allerdings hat sich Google bei den Verlagen bisher keine

Freunde gemacht. Es dürfte schwer werden, zu einer guten Grundlage der Zusammenarbeit zu finden.

Über den Umfang der bisher bei den Projekten eingescannten Bücher lässt sich noch nichts sagen. Größenvergleiche insbesondere zwischen Amazon und Google stehen aus. Möchte man nach Buchinhalten suchen, empfiehlt es sich, alle Angebote auszuprobieren. Der Zugriff auf die Suche ist über folgende Seiten möglich:

● Unter [print.google.de](http://print.google.de) hat Google vor kurzem ein deutschsprachiges Suchinterface freigeschaltet. Neben der einfachen Volltextsuche kann in der erweiterten Suche nach Titel, Autor, Verlag und ISBN gesucht werden.

● Bei Amazon lässt sich "Search Inside" unter [www.amazon.de/exec/obidos/tg/browse/-/14224751](http://www.amazon.de/exec/obidos/tg/browse/-/14224751) aufrufen. Bei der Anzeige der regulären Suchergebnisse kann man alternativ mit einem Klick auf die "Search Inside"-Ergebnisse umschalten. Eine elegante Methode, die Suche im Web mit der in Büchern zu kombi-

nieren, bietet A9.com (hier muss man auf der Startseite "Books" mit ankreuzen).

● Die OCA steht noch am Anfang, und es sind noch keine Inhalte verfügbar. Eine Suche nach gemeinfreien Büchern, die auch um die OCA-Texte erweitert werden wird, ist beim Internet Archive möglich (<http://www.archive.org/details/texts>).

Alle drei Initiativen stehen am Anfang, so dass eine abschließende Bewertung, welches Angebot für den Nutzer das beste ist, nicht gegeben werden kann. Sinnvoll und nützlich ist die Verfügbarmachung der Bücher aber auf jeden Fall: Sie erweitert die Web-Suchmaschinen um weitere wichtige Informationen.

Für die Bibliotheken wird es höchste Zeit, sich neue Strategien der Präsentation ihrer Inhalte zu überlegen. Sie werden letztlich nicht damit bestehen können, mit ihren mageren bibliographischen Daten gegen die beschriebenen ambitionierten Projekte anzutreten.

*Dirk Lewandowski,  
[dirk.lewandowski@uni-duesseldorf.de](mailto:dirk.lewandowski@uni-duesseldorf.de)*

## Europäische Kommission

# Gegenprojekt zu Google Print wird konkretisiert

**Die Europäische Kommission hat ihren Plan zum Aufbau einer Europäischen Digitalen Bibliothek konkretisiert.**

Angesichts der 2,5 Milliarden Bücher in den Europäischen Bibliotheken, die Karten und Bilder sowie Audios und Videos nicht mitgerechnet, gesteht die Kommission zunächst ihre begrenzten Möglichkeiten ein, ein Gegenprojekt zu Google Print auf die Beine zu stellen: "Digitising all this material seems not realistic in the light of the costs of the digitisation process and subsequent preservation". Stattdessen werden drei Phasen - Digitalisierung, Online Verfügbarkeit und Digitale Archivierung - eingeführt und ein "Kommunikationsdokument" für die erste Phase verfügbar gemacht. Interessierte Parteien werden um Kommentare gebeten. Diese sollen in einem Papier "Recommendation on digitisation and digital preservation" berücksichtigt werden, das im Juni 2006 ver-

öffentlicht wird. Da die Empfehlungen in diesem Papier auch Einfluss beispielsweise auf die Copyright-Politik der Kommission haben können, sollte ein Interesse etwa der Verleger bestehen, sich zu einem frühen Zeitpunkt an der europaweiten Willensbildung zu beteiligen.

Im Zusammenhang mit dem Aufbau einer Digitalen Bibliothek hat die Kommission 36 Mil-

lionen Euro für Forschung zu fortgeschrittenen Access-Möglichkeiten zu Beständen des kulturellen Erbes sowie 60 Millionen Euro für Entwicklungen verfügbar gemacht, die die Interoperationalität und den mehrsprachigen Zugang zu Beständen nationaler Sammlungen verbessern.

## Yahoo!

### Schritt zum Content Provider

Yahoo! hat neun bekannte Journalisten für die Bereiche "Finanzen" und "Anlagen" für das Schreiben von Kolumnen angeheuert. Zwanzig weitere Journalisten sollen hinzukommen. Der Suchmaschinenbetreiber verzichtet sich davon mehr Traffic.

## Globalspec

### Schritt zum Content Provider (2)

Die Technik-Suchmaschine GlobalSpec hat Inhalte von John Wiley, IEEE und Knovel aufgelegt. Dazu gehören Zeitschriftenbeiträge, Research Paper, technische Spezifikationen und bibliographische Arbeiten.

## AZojomo

### Gutachter im Peer Review-Prozess werden bezahlt

Die Technik-Suchmaschine Azojomo ist die erste Einrichtung, die die Gutachter im Peer Review-Prozess bezahlen will. Wissenschaftler werden aufgefordert, ihre Beiträge einzureichen. Das Geld soll über Anzeigenschaltungen hereinkommen.

## AOL

### Übernahme von Weblogs Inc.

AOL hat Weblogs Inc. übernommen. Der Preis soll zwischen 20 und 25 Millionen Dollar betragen. Das 2003 gegründete Weblogs gibt 85 Blogging Sites heraus und beschäftigt mehr als hundert freie Mitarbeiter, um die Weblogs zu füllen. Einnahmen werden über Anzeigenschaltungen erzielt, darunter allein von Googles AdSense-Programm in Höhe von einer Million Dollar.

## Ask Jeeves

### Bald ask.com

Ask Jeeves soll nach dem Willen seiner neuen Eigentümer bald in ask.com umbenannt werden. Auch das Butler-Maskottchen wird von der Website verschwinden.