

R. Bertelmann, K. Herm, T. Höhnow, S. Volz

Bibliothekssuchmaschine statt Bibliothekskatalog

Google und Konsorten haben das Suchverhalten unserer Nutzer grundlegend verändert. Erwartet wird eine Suche, die einfach, unkompliziert und übersichtlich sein soll. Längst haben Bibliotheken und Anbieter von Bibliothekssystemen darauf reagiert und die Suchoberflächen entschlackt. Trotzdem sehen viele Bibliothekskataloge nach wie vor wie „Bibliothekskataloge“ aus. Letztlich versuchen viele der Suchmasken immer noch die Vielfalt der erfassten Metadaten und die daraus resultierenden differenzierten Suchmöglichkeiten den Nutzern auf den ersten Blick nahe zu bringen. Das geht, was zahlreiche Studien belegen, häufig an den Bedürfnissen der Nutzer vorbei: Diese wünschen sich einen einfachen und schnellen Zugriff, auf die für sie relevante Information.

Bibliothekskataloge sind längst nicht mehr nur Bestandsverzeichnisse, sondern Zugangssysteme zur Vielfalt der von der Bibliothek vermittelten Informationen. Auch hier bieten Systemhäuser inzwischen Lösungen an, bei denen, im Sinn einer verteilten Suche, weitere Quellen mit einbezogen werden können.

Im Folgenden soll der Lösungsweg vorgestellt werden, den die Bibliothek des Wissenschaftsparks Albert Einstein in Zusammenarbeit mit dem Kooperativen Bibliotheksverbund Berlin-Brandenburg (KOBV) eingeschlagen hat, um mit diesen beiden veränderten Grundvoraussetzungen für ihr Serviceangebot umzugehen.

Die Bibliothek des Wissenschaftsparks Albert Einstein - eine gemeinsame Bibliothek des GeoForschungsZentrums Potsdam, der Forschungsstelle Potsdam des Alfred Wegener Instituts für Polar- und Meeresforschung (zwei Helmholtz-Zentren) und des Potsdam-Instituts für Klimafolgenforschung (ein Leibniz-Institut) - ist eine Spezialbibliothek mit dem thematischen Schwerpunkt ‚Geowissenschaften‘ auf dem größten Campus der außeruniversitären Forschung in Brandenburg, dem ‚Wissenschaftspark Albert Einstein‘ auf dem Telegrafenberg in Potsdam.

Der KOBV setzt bereits seit 2005 Suchmaschinentechnologie in verschiedenen Entwicklungsprojekten erfolgreich ein. Zusammen mit der Bibliothek des Wissenschaftsparks Albert Einstein wurde nun der Prototyp einer „Bibliothekssuchmaschine“ auf Basis erprobter Open-Source- Technologien aus dem Java-Umfeld (wie Tomcat¹, Jakarta-Commons², Log4J³ usw.) als web-basierte Anwendung realisiert, deren Suchmaschinenkern auf der ebenfalls als freie Open-Source Java-Variante erhältlichen Search-Engine-Library Lucene⁴ basiert. Die erste Version der Bibliothekssuchmaschine läuft seit Ende Oktober im Echtbetrieb.

Ziel des Pilotprojektes war die Konzeptionierung, Spezifikation und Implementierung einer neuen, benutzerfreundlichen Suchoberfläche zum schnellen Auffinden fachwissenschaftlich relevanter Daten und Informationen, sowohl in bibliothekseigenen Beständen als auch in zusätzlichen Quellen. Vor dem spezifischen Hintergrund der Spezialbibliothek werden dabei nicht nur Kataloginhalte google-like findbar gemacht, sondern der Suchraum „Katalog“ um weitere für die Informationsvermittlung auf dem Campus relevante und spezifische

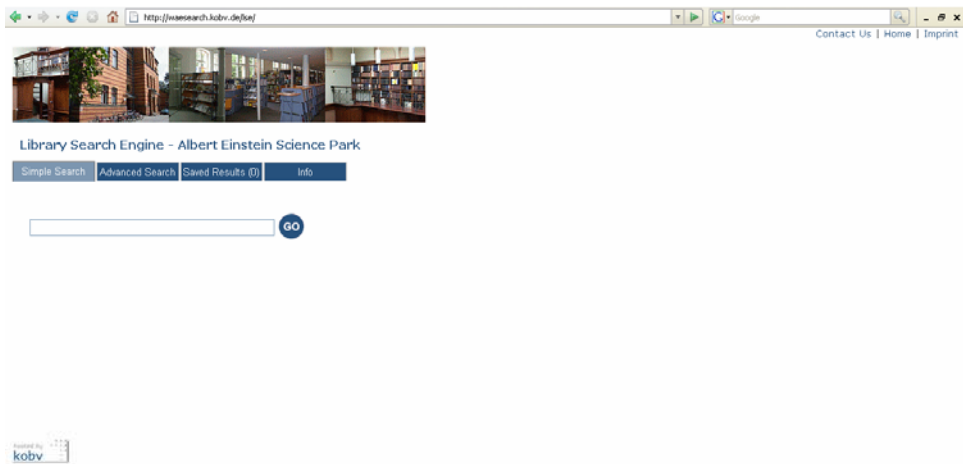
¹ <http://tomcat.apache.org>

² <http://commons.apache.org>

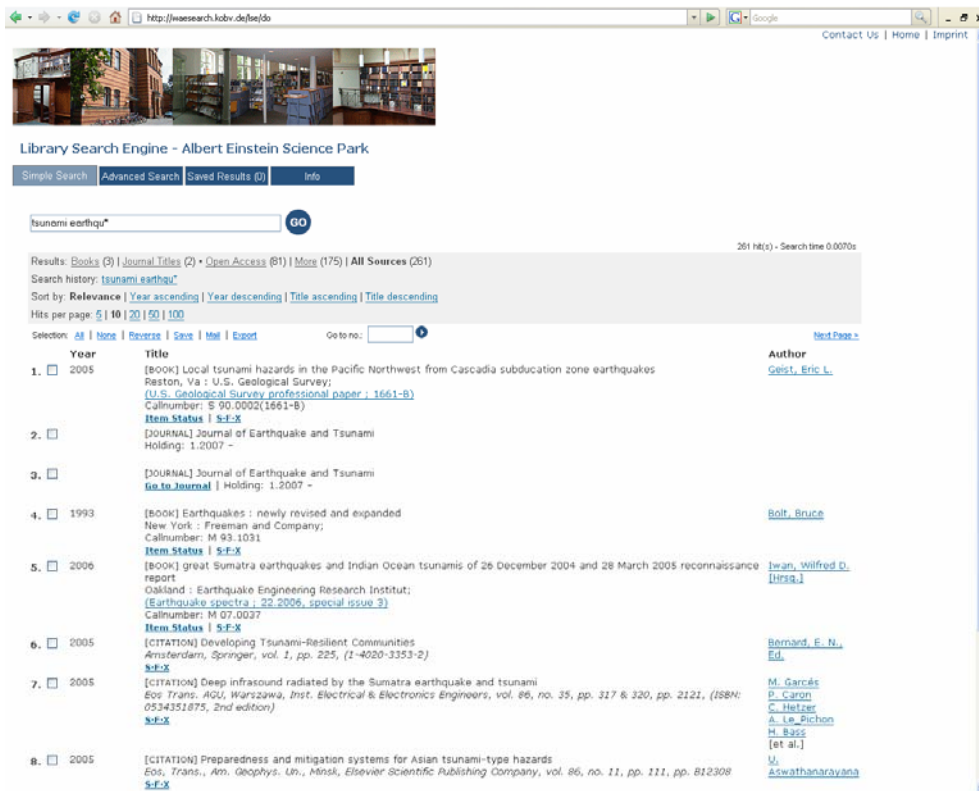
³ <http://logging.apache.org/log4j>

⁴ <http://lucene.apache.org>

fachwissenschaftliche Inhalte als zusätzliche Suchräume erschlossen. Die neue Anwendung dient dem schnellen Ersteinstieg und leitet die Nutzer dann an die jeweiligen Quellen weiter.



http://waeseach.kobv.de



Der erste Such-Einstieg erfolgt über eine einzeilige Maske (Simple Search). Mehrere eingegebene Suchterme werden standardmäßig mit dem Operator ‚und‘ verknüpft. In der ‚Advanced Search‘ können hiervon abweichende differenziertere Suchanfragen formuliert werden. Durchsucht werden bei ‚Simple Search‘ alle indextierten Angaben, bei der Advanced-Search die jeweils gewählten Bereiche. Insgesamt wurden für die Anwendung die vollständigen Metadaten inklusive Verlagsangaben, Abstracts und weitere Angaben, sowie in den Fällen, in denen aus den Metadaten auf ein offen zugängliches PDF-Dokument referenziert wird, auch dessen vollständiger Inhalt indextiert.

Die eingesetzten Technologien gestatten es, eine sehr performante Anwendung und Suche zu realisieren. Aufgrund der so erzielten Schnelligkeit der vom KOBV implementierten Anwendung werden Suchen immer über den gesamten Inhalt ausgeführt und erst in der Darstellung der Ergebnisse differenziert. In der Ergebnisnavigation können die Ergebnisse gesammelt (Link „All Sources“) oder nach unterschiedlichen Datenquellen getrennt angezeigt werden (weitere Links).

So werden durch den Link „Book“ die Ergebnisse aus dem klassischen Bibliothekskatalog präsentiert. Dieser Teil des Suchindexes wird über regelmäßige Exporte aus dem Bibliothekssystem Sisis-Sunrise und einer Allegro-Datenbank, die Katalogdaten einer Teilbibliothek enthält, gespeist. Jede Titelanzeige enthält einen Link „Item-Status“, mit dem ins Bibliothekssystem auf den jeweiligen Datensatz verwiesen wird. Hier können dann die üblichen Funktionen des elektronischen Benutzer-Kontos (Vormerkung, Kontenübersicht und dergleichen) genutzt werden.

Der Ergebnislink „Journal Titles“ erschließt die Suchtreffer im Zeitschriftenbestand des Wissenschaftsparks (Print und elektronisch; inklusive der Nachweise der älteren Jahrgänge vieler Zeitschriften, die über das Programm 'Nationallizenzen' der DFG zugänglich sind). Insgesamt etwa 5000 Zeitschriften und Schriftenreihen, die von der Bibliothek im Hintergrund mit einer Allegro-Datenbank und der EZB verwaltet und daraus die Daten für die Bibliothekssuchmaschine gewinnt. Angaben zu vorhandenen Printjahrgängen und zu Lizenzzeiträumen ergänzen die Links zu den vorhandenen elektronischen Zugängen. Doch die Suchraumerweiterung weist weit über den klassischen Bibliothekskatalog hinaus und bietet weitere Quellen zur Suche an:

- Current Contents: hier werden die aktuellen Inhaltsverzeichnisse von zur Zeit 260 Zeitschriften ausgewertet, soweit deren Inhalte über RSS-Feeds zugänglich sind. Diese Aufsatzinformation wächst monatlich um etwa 6000 aktuelle Einträge und schließt die Lücke zwischen Erscheinen der Zeitschriften und der Verzeichnung der Aufsätze in den relevanten Fachdatenbanken. Entsprechend soll hier über eine „Moving Wall“ nur die aktuelle Information für ein Jahr vorgehalten werden. Umso mehr, da über die RSS-Feeds, vom bibliothekarischen Standpunkt aus gesehen, meist nur sehr rudimentäre Metadaten zur Verfügung stehen. Entscheidend ist jedoch, dass in den XML-Daten der Feeds immer eine direkte Verlinkung zum lizenzierten Volltext des Aufsatzes zur Verfügung steht.

- E-Books: die Inhaltserschließung von lizenzierten oder gekauften E-Books stellt Bibliotheken vor neue Herausforderungen. Auch hier bietet die Bibliothekssuchmaschine einen Erschließungsansatz: die Metadaten der Aufsätze aus den ‚Special Publications‘ der Geological Society London, eine Serie klassischer Sammelwerke, zu der auf dem Campus der Zugang ab Jahrgang 2004 angeboten, wird, übernommen und sind ebenfalls durchsuchbar (ca. 1600 Einträge).

- Open Access schafft zunehmend den freien und offenen Zugang zu wissenschaftlichen Inhalten. Je mehr Veröffentlichungen auf diese Weise vorliegen, desto stärker stellt sich die Frage, wie diese in die Informationsvermittlung der Bibliothek integriert werden können. Die Virtuelle Fachbibliothek Geowissenschaften Geo-Leo harvestet offen zugängliche Artikel aus geowissenschaftlicher Forschung weltweit und hat der Bibliothek diese Sammlung von ca. 30.000 Texten zur Verfügung gestellt. Diese offen zugänglichen wissenschaftlichen Texte werden ergänzt durch die Publikationen von wissenschaftlichen Primärdaten aus der Scientific Drilling Database des GeoForschungsZentrums.

- Im Weiteren bietet die Bibliothekssuchmaschine Zugang zu fachwissenschaftlich relevanten bibliografischen Inhalten. Dies umfasst die Inhalte der Datenbanken, in denen die wissenschaftlichen Veröffentlichungen der betreuten Institute erfasst sind, sowie eine Reihe von Bibliografien, die im Kontext wissenschaftlicher Projekte an den Instituten entstanden sind. Bibliografische Information also, die allein von ihrer Herkunft her schon thematisch eng mit der Forschung auf dem Telegrafenberg verbunden ist.

- In verschiedenen anderen Entwicklungsprojekten des KOBV wurden neben den Metadaten selbst auch Inhalte von in den Metadaten referenzierten Webangeboten sowie Volltexte indexiert und inhaltlich erschlossen. Wir haben uns im aktuellen Projekt auf die Indexierung referenzierter PDF-Dateien beschränkt, in der Annahme, dass dies meist Volltexte der in den Metadaten beschriebenen Werke sind.

- Um den Nutzern eine Erweiterung ihrer Suchen zu ermöglichen, unter anderem in den auf dem Campus lizenzierten bibliografischen Datenbanken, wird das Linking System SFX eingesetzt. Mit dessen Hilfe werden beispielsweise die Autoren eines Suchresultats übernommen und in einer vorformulierten Suchanfrage für die Fachdatenbank GeoRef angeboten.

Die erste Version der Bibliothekssuchmaschine wird im Wissenschaftspark Albert Einstein seit Oktober 2007 im Echtbetrieb eingesetzt. Mittelfristig ist die Entwicklung einer zweiten Version geplant, die dann beispielsweise Elemente aus Web 2.0 (individualisierbare RSS-Feeds, Anbindung an Social-Bookmarking-Dienste u.a.) und die Möglichkeit, die Bibliothekssuchmaschine via Webservices nahtlos in vorhandene Angebote einzubinden, enthalten soll.

Die Zusammenarbeit in der Entwicklung der Pilotanwendung war für alle Beteiligten sehr produktiv und gelungen. Der KOBV plant, ausgehend vom Piloten eine Default-Applikation zu schaffen und als neuen ASP-Dienst weiteren Bibliotheken anzubieten.

Roland Bertelmann, Tobias Höhnow: Bibliothek des Wissenschaftsparks Albert Einstein
Karin Herm, Sibylle Volz: Kooperativer Bibliotheksverbund Berlin Brandenburg