

Suchmaschine, ERM & Co.: Ressourcenmanagement im Backend des Bibliothekars

Tobias Höhnnow

This document appeared in

Bernhard Mittermaier (Eds.):

eLibrary - den Wandel gestalten

5. Konferenz der Zentralbibliothek

Proceedings of the WissKom 2010: 5. Konferenz der Zentralbibliothek, 08.-11. November 2010, Jülich

Schriften des Forschungszentrums Jülich / Reihe Bibliothek/Library, Vol. 20

Zentralbibliothek (ZB)

Forschungszentrum Jülich GmbH, Zentralbibliothek, Verlag, 2010

ISBN: 978-3-89336-668-2

Suchmaschine, ERM & Co.:
Ressourcenmanagement im Backend des Bibliothekars

Tobias Höhnow

Zusammenfassung

Die Bibliothek des Wissenschaftsparks Albert Einstein zu Potsdam hat zusammen mit dem Kooperativen Bibliotheksverbund Berlin-Brandenburg (KOBV) eine eigene Suchmaschine (ALBERT) auf Basis von Lucene + Solr in Betrieb. So wird eine „Google-like“-Suche mit schneller und übersichtlicher Präsentation der Ergebnisse ohne Verwendung komplexer Suchanfragen und weiterer Kenntnis zugrunde liegender Quellen angeboten. Im Gegenzug sind zur Verwaltung immer umfangreicherer Kollektionen Bibliothekare auf Anwendungen wie Electronic Resource Management-Systeme (ERM) angewiesen. Auch bieten Web 2.0-Funktionalitäten neue Möglichkeiten, wie etwa die Aggregation von Inhaltsverzeichnissen lizensierter Zeitschriften als RSS-Feed. Auf diese Trends hat die Bibliothek auf Grundlage quelloffener Software mit selbst entwickelten Anwendungen reagiert. Der Aufsatz gibt einen Überblick über diese Anwendungen ausgehend vom Frontend des Benutzers (ALBERT) hin zum Backend-Bereich des Bibliothekars. Dabei werden die Funktionalitäten der einzelnen Systeme verdeutlicht, auf die Kriterien und Anforderungen des Update-Prozesses eingegangen sowie die Vor- und Nachteile thematisiert.

Abstract

The Library of the Wissenschaftspark Albert Einstein in Potsdam is operating together with the Kooperativer Bibliotheksverbund Berlin-Brandenburg (KOBV) an own search engine (ALBERT) on basis of Lucene + Solr. Thus is provided a Google-like search possibility comprising a quick and clearly representation of the results without applying complex queries and knowledge on underlying sources. In return librarians relies on systems for Electronic Resource Management (ERM) to administrate collections that are getting more comprehensive. Additionally Web 2.0 functionalities bears new opportunities like the aggregation of table of contents from licensed journals as RSS feed. To respond this trend the library has developed self-made applications on the basis of Open-source software. The article surveys these applications from the user frontend (ALBERT) to the librarians backend. Hereby are clarified the systems functionalities, criterias and requirements of the update procedures as well as the pros and cons.

Einleitung

Die Bibliothek des Wissenschaftsparks Albert Einstein fungiert als gemeinsame Bibliothek der Forschungseinrichtungen Deutsches GeoForschungsZentrum (GFZ)¹, des Potsdam-Instituts für Klimafolgenforschung (PIK)² und der Forschungsstelle Potsdam des Alfred-Wegener-Instituts für Polar- und Meeresforschung (AWI)³. Angesiedelt auf dem traditionsreichen Campus auf dem Telegrafenberg in Potsdam versorgt die Bibliothek den größten Standort außeruniversitärer Forschung (zwei Zentren der Helmholtz-Gemeinschaft und ein Institut der Leibniz-Gemeinschaft) im Land Brandenburg mit elektronischen und gedruckten Medien sowie Informationen, die für die aktuelle Forschung benötigt werden. Die Bibliothek ist Ansprechpartner zu allen Fragen des wissenschaftlichen Publizierens.

Suchmaschine statt Bibliothekskatalog

Die Anforderungen und Erwartungen der Nutzer an Bibliothekskataloge und Recherchesysteme haben sich durch die Verbreitung von Suchmaschinen grundlegend geändert. Erwartet wird eine „Google-like“-Suche mit möglichst schneller und übersichtlicher Präsentation der Ergebnisse ohne Verwendung komplexer Suchanfragen und weiterer Kenntnis zugrunde liegender Quellen. Kurz: es soll schnell und frei sein und stets Ergebnisse liefern, so wie Google es vormacht (Soules 2010). Neben der schlichten Bevorzugung des direkten Links vom Nachweis zum Volltext erwarten Wissenschaftler und Wissenschaftlerinnen einen One-Stop-Shop möglichst sämtlicher zur Verfügung stehender Literatur. Dies schließt heutzutage den Zugang zu vielfältigen Quellen ein und geht weit über den traditionellen „Besitz-Bestand“ einer Bibliothek hinaus. Weiterhin bieten Dienste des als Web 2.0 bzw. Bibliothek 2.0 bekannten Schlagworts neue Möglichkeiten der Bereitstellung und Darstellung von Inhalten sowie des Vernetzens von Informationen (Herb 2007).

Die Bibliothek des Wissenschaftsparks Albert Einstein hat zusammen mit dem Kooperativen Bibliotheksverbund Berlin-Brandenburg (KOBV) seit 2007 eine eigene Suchmaschine ALBERT⁴ (All Library Books, journals and Electronic Records Telegrafenberg) in Betrieb, die für das wissenschaftliche Personal den zentralen Einstiegspunkt in seine Suche bietet und den klassischen Bibliothekskatalog ersetzt. Damit wurde eine einfache und benutzerfreundliche Suchoberfläche zum schnellen

¹ <http://www.gfz-potsdam.de/>

² <http://www.pik-potsdam.de/>

³ <http://www.awi.de/de/institut/standorte/potsdam/>

⁴ <http://waesearch.kobv.de/lse/>

Auffinden fachwissenschaftlich relevanter Daten und Informationen, sowohl in bibliothekseigenen Beständen als auch in zusätzlichen Quellen, geschaffen. Neben den klassischen Katalogdaten werden über verschiedene Suchräume Artikel, Bibliographien, Datenpublikationen u.v.m. suchbar und verfügbar gemacht, um weitere für die Informationsvermittlung auf dem Campus relevante und spezifische Inhalte zu erschließen (Bertelmann, Herm et al. 2007). ALBERT basiert auf der quelloffenen Suchmaschinen-Technologie Lucene, die aus dem Apache-Projekt stammt⁵. Die Software ist ebenso komplett in Java geschrieben wie der als Erweiterung funktionierende Open-Source-Suchserver Solr⁶. Mit Apache Solr sind Features wie facetierte Navigation, Treffer-Highlighting und verteilte Suche möglich. Diese Möglichkeiten wurden in die im Herbst 2010 gestartete, stark überarbeitete und substantiell verbesserte Version 2.0 der Suchmaschine integriert (Abb. 1).

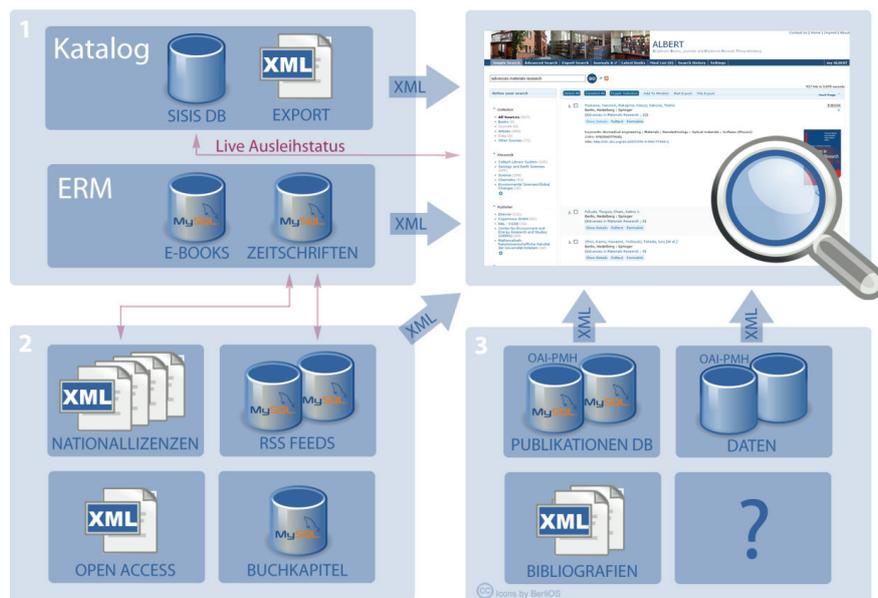


ABB. 1: INTEGRATION VERSCHIEDENER QUELLEN IN ALBERT: (1) KLASSISCHE KATALOGDATEN, (2) ARTIKEL, (3) ANDERE QUELLEN.

Die eingesetzten Technologien gestatten eine sehr performante Suche, so dass ausgeführte Abfragen stets über die gesamten Indices ausgeführt werden. Die Darstellung der Ergebnisse der in ALBERT zugrunde liegenden rund 6 Mio. Datensätze

⁵ <http://lucene.apache.org/>

⁶ <http://lucene.apache.org/solr/>

werden so erst im Nachhinein in einzelne Suchräume (Collections) differenziert. Im Detail betrachtet, umfassen die „Collections“ folgende Suchräume:

(1) Der Suchraum *Books* entspricht dem klassischen Bibliothekskatalog und zeigt alle Ergebnisse einer Suche aus dem Monographien-Bestand (ca. 80.000 Einträge) an. Ergänzt wird dies durch verfügbare elektronische Bücher aus dem Angebot der Nationallizenzen⁷ der DFG (Deutsche Forschungsgemeinschaft) sowie anderer spezifischer Quellen, die auf der Ebene der Suchmaschine integriert werden und nicht zuerst in den klassischen Bibliothekskatalog aufgenommen werden müssen.

(2) *Journals* versammelt alle Ergebnisse zu einer Suche des Zeitschriften-Bestands, in der elektronische Zugänge sowie Printbestände eingeschlossen sind. Die mehr als 10.000 Titel umfassen neben eigenen Subskriptionen auch Zugänge über Konsortialverträge und Zeitschriften-Archive der DFG-Nationallizenzen sowie offen zugängliche Zeitschriften (s. a. Abschnitt „ERM“). Das thematische Standardset beinhaltet alle zugänglichen Kernzeitschriften der Bereiche Geowissenschaften und Geographie (~3.600 Titel) und kann auf weitere beliebige Fachgebiete erweitert werden. Mit den Features *Show Latest Articles* bzw. *Show Archive* können zudem frisch freigeschaltete im Volltext verfügbare Artikel in ALBERT angezeigt und als RSS-Feed dauerhaft abonniert werden (s. a. Abschnitt „RSS“) bzw. das aus den DFG-Nationallizenzen integrierte Archiv einer Zeitschrift angezeigt werden.

(3) Mit *Articles* existiert ein Suchraum mit relevanten wissenschaftlichen Artikeln zur Ergänzung der Suchmöglichkeiten in Fachdatendanken (~5,4 Mio Einträge). Die unter dem Label *Paper Current* versammelten Artikel sind die neuesten aus relevanten Zeitschriften (>700 Titel) und beziehen vor allem jene Aufsätze mit ein, die auf ihren Verlagsservern bereits zugänglich, aber noch nicht als Printausgabe erschienen und entsprechend nicht in ISI Web of Science⁸ oder anderen Fachdatenbanken zu finden sind - letztlich also zeitgemäßer ein Current Contents-Dienst (~200.000 Einträge). Diese Nachweise sollen so die Indexierungslücke zwischen Fachdatenbanken sowie Artikeln mit Preprint-Status wie 'online first', 'article in press' oder 'ahead of print' auf den Verlagsservern schließen und werden nach etwa einem Jahr aus dem entsprechenden Suchmaschinenindex entfernt.

⁷ <http://www.nationallizenzen.de/>

⁸ <http://isiknowledge.com/>

Das Label *Paper OA* (Open Access) enthält weltweit frei zugängliche relevante Artikel, die ihren Ursprung in den von der virtuellen Fachbibliothek GEO-LEO⁹ geharvesteten geowissenschaftlichen Open Access-Repositories haben (ca. 30.000 Dokumente). Hier werden die Vorteile von als Open Access deklarierten frei verfügbaren wissenschaftlichen Artikeln genutzt und neben den Metadaten auch die Volltexte indexiert.

Das generell als *Paper* bezeichnete Label beinhaltet einerseits Aufsätze aus relevanten elektronisch lizenzierten Buchreihen, wie den Special Publications der Geological Society of London, und andererseits die zur Verfügung stehenden Metadaten der Zeitschriften-Archive der DFG-Nationallizenzen. Die Bibliothek des Wissenschaftsparks Albert Einstein hat 44 vor allem naturwissenschaftliche Archiv-Pakete lizenziert und die für 11 Produkte zur Verfügung stehenden Metadaten in den Suchmaschinenindex integriert (~5,4 Mio. Einträge). Im Falle des Vorhandenseins weiterer Metadaten-Pakete sollen diese zukünftig in den Suchmaschinenindex einbezogen werden.

Die im Rahmen der DFG-Nationallizenzen bereitgestellten Metadaten im MAB2-Format werden mit Hilfe des von der DNB (Deutsche Nationalbibliothek) angebotenen Conversion Tools *MabToMabxml* in das Format MabXML umgewandelt. Anschließend wurden die teils sehr umfangreichen XML-Dokumente geparkt und in kleinere zusammengehörige Segmente unterteilt. So entstand für jeden Zeitschriften-Titel ein eigenes XML-File, das anschließend anhand der ISSN mit dem Themengebiet der EZB (Elektronische Zeitschriftenbibliothek)¹⁰ verknüpft wurde. Hierbei wurde auf den Möglichkeiten des flexiblen Systems zum Electronic Resource Management (ERM) aufgebaut (s. a. Abschnitt „ERM“). Dieser neue Weg der Anreicherung und qualitativ verbesserten Aufbereitung der Metadaten der DFG-Nationallizenzen, ermöglicht durch das aktive Ab- und Zuschalten des Themengebiets in ALBERT, hilft Redundanzen zu vermeiden. So können nur jene Zeitschriften-Archive durchsuchbar gemacht werden, die tatsächlich das Interesse des Nutzers finden. Hier findet folglich eine erste Personalisierung des Sucheinstiegs statt.

⁹ <http://www.geo-leo.de/>

¹⁰ <http://rzblx1.uni-regensburg.de/ezeit/>

(4) Der als *Data* bezeichnete Suchraum beinhaltet publizierte wissenschaftliche Forschungsdaten verschiedener Quellen. Zur Zeit gehören in diese Kollektion die Daten der GFZ Scientific Drilling Database (SDDB)¹¹ und der WDC Pangaea¹². Auch aufgrund der zunehmenden Bedeutung wissenschaftlicher Daten wird dieser Bereich in Zukunft ausgebaut werden und schafft eine bessere Verknüpfung zwischen klassischen Publikationen und Datenpublikationen.

(5) Im letzten Suchraum *Other Sources* sind die Publikations-Datenbanken des Deutschen GeoForschungsZentrums (GFZ) und des Potsdam-Instituts für Klimafolgenforschung (PIK) enthalten (~15.000 Dokumente). Diese Datenbanken dienen einerseits dem Publikationsmanagement der Forschungseinrichtungen, fungieren aber zunehmend parallel auch als Open Access-Repositoryen. Die teils offen zugänglichen Volltexte sind folglich komplett in den Suchindex integriert. Darüber hinaus finden sich hier verschiedene Bibliographien aus dem Forschungsumfeld des Telegrafenberges, z. B. des International Lithosphere Program (ILP)¹³, des Deutschen Kontinentalen Reflexionsseismischen Programms (DEKORP)¹⁴ sowie eine Bibliographie zur Seismologie (insgesamt ~ 30.000 Einträge). Die Implementierung weiterer Bibliographien ist durchaus denkbar.

Neben der Bereitstellung kontextbezogener Informationen wie Verfügbarkeit gedruckter Medien, Buchcover, bibliografische Beschreibung und Möglichkeiten der Verfeinerung der Ergebnisse (Refinement), verfügt ALBERT weiterhin über die Alternative permanente Bookmarks zu generieren (Permalink). So können Kollektionen von Kernzeitschriften z. B. einzelner Arbeitsgruppen dauerhaft gebildet werden, die mit stets aktuellen Zugangsdaten versehen sind. Deutlicher wird der Vorteil der in dieser Weise zur Verfügung gestellten neuesten Inhalte dieser Zeitschriften. So können sowohl thematische RSS-Feeds generiert werden wie auch beliebig zu definierende Zusammenstellungen z. B. bestimmter Kernzeitschriften. Der Wissenschaftler oder die Arbeitsgruppe erhält somit stets die neuesten Artikel seines Interesses und ist ohne weiteres Zutun ständig Up-to-date (s. a. Abschnitt „RSS“). Durch die Integration ausschließlich im Volltext verfügbarer Zeitschrifteninhaltsverzeichnisse bzw. deren Inhalte ist die Frustrationsgrenze („Please select a purchase option.“) beim Nutzer hinsichtlich Zugangsbarrieren gering.

¹¹ <http://dc110dmz.gfz-potsdam.de/contenido/lakedb>

¹² <http://www.pangaea.de/>

¹³ <http://scliip.gfz-potsdam.de/>

¹⁴ <http://www-app1.gfz-potsdam.de/www/pb3/dekorp/welcome.html>

Um Nutzern eine unkomplizierte Erweiterung ihrer Anfragen auf bibliografische Datenbanken oder wissenschaftliche Suchmaschinen zu ermöglichen, wird der Linkresolver SFX eingesetzt. SFX dient hier also zur Weiterleitung in ergänzende Suchräume und erfüllt damit jene Funktion, die in anderen Suchsystemen über verteilte Suchen (mit dem Nachteil der oft langen Antwortzeiten) eingebunden ist.

Electronic Resource Management eigener Bauart

Zur Verwaltung immer umfangreicherer Kollektionen sind Bibliothekare im Backend-Bereich auf Anwendungen wie Electronic Resource Management-Systeme (ERM) angewiesen. Auf diese Entwicklung sind Anbieter kommerzieller Bibliothekssysteme mit Bereitstellung eigener Produkte bereits eingegangen (Collins 2008) und (Breeding 2008). Der Entscheidung der Bibliothek, nicht auf solche Systeme zurückzugreifen und eine eigene Applikationen zu entwickeln, liegen mehrere Faktoren zugrunde. Zum einen lag bei Beginn des Projekts noch kein den Ansprüchen und Bedürfnissen der Bibliothek entsprechendes System vor. Zum anderen weisen Eigenentwicklungen eine hohe individuelle Anpassungsfähigkeit an die speziellen Bedürfnisse einer Bibliothek auf, während kommerzielle Systeme wegen ihres an Profit orientierten Zuschnitts meist wenig flexibel sind, um auf lokale Bedürfnisse einzugehen. Normalerweise lassen sich Eigenentwicklungen ins vorhandene Umfeld leichter implementieren, jedoch muss der Aufwand der Systempflege im Vergleich zu einem kommerziellen Produkt natürlich kontinuierlich kritisch hinterfragt werden (Hawthorne and Watson 2007). Hier können aber offene Systeme mit entsprechenden Exportmöglichkeiten bzw. Schnittstellen (z. B. der EZB) Abhilfe schaffen oder zumindest unterstützend wirken. Unter Zuhilfenahme der Funktionalitäten des Web 2.0, existieren mittlerweile bereits Ansätze, die versprechen eine preiswerte wenn nicht gar kostenfreie Verwaltung elektronischer Ressourcen zu gestatten (Murray 2008). Auch sind Teillösungen, die beispielsweise COUNTER-gerechte Nutzungsstatistiken elektronischer Zeitschriften bereit stellen, sicherlich eine Option, jedoch sind auch hier die Anschaffungskosten als problematisch zu bezeichnen (Schüller-Zwierlein 2007). Insgesamt ist die Eigenentwicklung der Bibliothek auch mit Blick auf kostenintensiven Kauf und der jährlichen Wartungs- und Lizenzkosten einer kommerziellen Lösung wirtschaftlich.

Naturwissenschaftliche Zeitschriften bilden den Kernbestand der Bibliothek und werden in elektronischer wie gedruckter Form zusammen mit elektronischen Büchern sowie Datenbanken unter Zuhilfenahme eines ERMs administriert und zur weiteren Präsentation in ALBERT aufbereitet. Als Eigenentwurf der Bibliothek basiert das ERM auf der Software-Distribution LAMP (Linux, Apache, MySQL, PHP) und bildet den gesamten Lebenszyklus einer elektronischen Ressource ab (Sadeh and Ellingsen 2005).

Die lizenzierten elektronischen Zeitschriften des Wissenschaftsparks werden zunächst in der EZB verwaltet. Dies hat im Gegensatz zur lokalen Pflege den Vorteil, dass die bibliografischen Angaben und insbesondere die Zugangs-URLs kooperativ gepflegt werden und damit stets aktuell sind. Ein weiterer wichtiger Punkt ist die zentrale Pflege von Konsortiallizenzen sowie DFG-Nationallizenzen in der EZB durch die jeweils verhandelnden Bibliotheken. Zum Update des lokalen ERMs werden wöchentlich jene Zeitschriften und ihre Bezugszeiträume exportiert, die entweder eine Lizenz aufweisen (gelb) oder frei verfügbar sind (grün) und den Themengebieten Geowissenschaften, Geographie oder Umwelt entsprechen (insgesamt ca. 10.000 Titel). Von dort gelangen die exportierten Titel dezidiert in das lokale ERM, wo elektronische Zeitschriften mit gedruckten Zeitschriften vereint und mit zusätzlichen Informationen angereichert werden. Das sind aus dem bibliografischen Bereich Angaben wie vorherige und fortführende Titel, ISI-Abkürzungen, Akronyme, Körperschaften u. Ä. Aus dem individuell für den Wissenschaftspark geltenden Bereich wird der Import mit Angaben wie Rechnungsdaten, Zugriffszahlen, Informationen über Buchbinderlieferungen und vorhandenen gedruckten Bestand verknüpft (Abb. 2). Fachfremde Titel wurden bisher während des Update-Prozesses im ERM deaktiviert, so dass sie dem Benutzer im Frontend (ALBERT) unkenntlich waren. Mittlerweile hat die Möglichkeit des Auswählens der in Frage kommenden Themengebiete diese Deaktivierung abgelöst und es ist nun am Nutzer selbst den entsprechenden Suchraum einzugrenzen. Für viele Zeitschriften existieren vor allem durch die überregionale Literaturversorgung sich einander überschneidende Lizenzzeiträume. Was für den Backend-User noch von Belang ist, interessiert den Nutzer weniger. Er will lediglich den ihm zur Verfügung stehenden Zugangszeitraum der gewünschten Zeitschrift erfahren ohne Kenntnis der Herkunft der Lizenz. Deshalb bildet das ERM in differenzierter Weise

sämtliche Lizenzen einer Zeitschrift ab, aggregiert diese aber für das Frontend zu einer Angabe.

Ein wichtiger Bestandteil des ERM ist die für die Analyse des Bestands notwendige statistische Auswertung der Zugriffszahlen und das Verhältnis zum Subskriptionspreis. Die zur Verfügung stehenden COUNTER-Statistiken werden als CSV-Dateien in die Datenbank eingelesen. Diese Prozedur ist weitestgehend automatisiert und übersichtlich, jedoch könnte an dieser Stelle das bereits etablierte SUSHI-Protokoll zum Datenaustausch unterstützend in das ERM eingebaut werden.

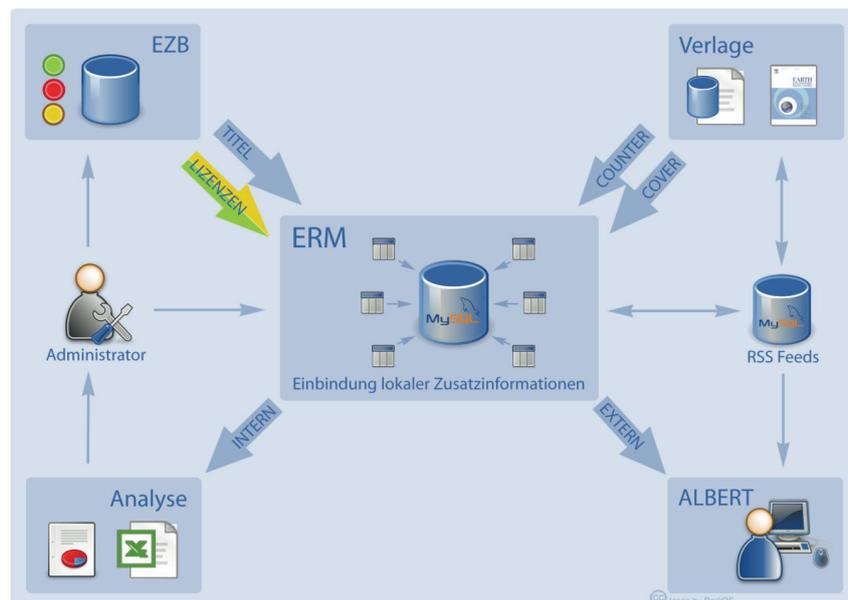


ABB. 2: ERM: AKTUALISIERUNG, ANREICHERUNG UND AUSGABE INS FRONT- BZW. BACKEND

Neben der visualisierten Aufbereitung des jährlichen Subskriptionspreises und dessen Steigerung oder seltener dessen Absinken wird vor allem der Preis pro Volltext-Download ausgerechnet und als Diagramm dargestellt. Als Obergrenze gilt zunächst der aktuelle Preis des Pay-Per-Use dieses Verlags. Allerdings sind die Probleme, die eine isolierte Betrachtung des Download-Preises von Volltexten bringen, bekannt. So wird das Netz von Wissenschaftlern zunehmend als Arbeitsmittel genutzt und es kann angenommen werden, dass populäre oder oft verwendete Artikel gleich mehrfach von ein und demselben Wissenschaftler heruntergeladen werden. Auch eine sehr hohe Nutzung einzelner Zeitschriften im Gegensatz zu recht geringen Zugriffszahlen kann kein alleiniges Maß für Abbestellungen sein. Die mit geringfügigen

Download-Kosten verbundenen Online-Volltexte der Überblickszeitschriften *Nature* und *Science* können somit nicht als Referenzpunkt zu Rate gezogen werden. Neben der Anzahl der zur Verfügung stehenden Aufsätze ist, ähnlich dem Impact Factor, eine hohe Nutzung einer Zeitschrift auch von dessen Zugehörigkeit zu einem Fachgebiet abhängig. Eine exaktere Analyse würde eine Auswertung der Nutzung elektronischer Zeitschriften auf der Ebene einzelner Artikel ermöglichen, was von Verlagen bisher aber sehr selektiv angeboten wird. Mit Ausnahme beispielsweise der seltenen Angabe meistgelesener Artikel eines Fachblatts existieren für viele Zeitschriften nach wie vor gar keine differenzierten Zahlen (Keller 2002).

Unter Beachtung genannter Schwächen von Zugriffsstatistiken werden bei der Bibliothek des Wissenschaftsparks Albert Einstein in die Analyse der Statistik vor allem die Präferenzen der einzelnen Arbeitsgruppen, Sektionen oder Departments einbezogen, ähnlich der Entwicklung, die der Bestandsaufbau hin zur fast ausschließlichen nutzergeleiteten Erwerbung nimmt (Stowell Bracke, Hérubel et al. 2010). Die Tatsache, ohne eine bestimmte Quelle nicht ausreichend über ein zumeist sehr spezielles Fachgebiet im Bilde zu sein, ist Grund genug, auch Zeitschriften mit vermeintlich mittelmäßigen Zugriffszahlen weiterhin im Bestand zu halten. Im Einzelfall wird auch die Bibliothekskommission zu Rate gezogen.

Sämtliche statistische Daten können selbstverständlich zur weiteren Verwendung exportiert werden. Für Einzeltitel gibt es die Möglichkeit, die Daten aller erhobenen Jahre inklusive visualisierter Aufbereitung als PDF-Datenblatt herunterzuladen. Mehrere Titel, beispielsweise eines Jahres und/oder eines Verlages, können als CSV-Datei heruntergeladen und in Excel weiterverarbeitet werden. So ist hier von Vorteil, dass auf die besondere Gegebenheit mehrerer voneinander unabhängiger Institutionen im Wissenschaftspark eingegangen werden kann. Beispielsweise ist etwa die Auflistung und der Export von Zahlen der abonnierten Titel eines speziellen Instituts ohne weiteres möglich. Für besonders häufig abgefragte Konstellationen existieren vorgefertigte Links.

Neben der Generierung und Auswertung von statistischen Zahlen sind weiterhin Informationen zu Lieferanten, Buchbinderlieferungen oder Reklamationen vorhanden. Die Verwaltung der Rechnungsdaten erfolgt ebenfalls auf Basis des ERM. Soweit

lieferbar werden Daten zu Rechnungen von der Agentur als CSV-Datei geliefert und können wiederum in die zugrunde liegende Datenbank eingelesen werden. Genauso existiert jedoch auch die Möglichkeit der manuellen Pflege bei Subskriptionen kleinerer Verlage oder individueller Daten. Die Rechnungen selbst werden - sofern erforderlich - digitalisiert und über das Dateisystem verfügbar gemacht. Da die Originalrechnungen zur Weiterbearbeitung sowieso an die zuständige Finanzabteilung weitergegeben werden, wurde so eine vollständig papier- wie ordnerlose Rechnungsverwaltung aufgebaut.

Integration neuester Zeitschrifteninhalte als RSS-Feeds

Neben der Bereitstellung klassischer Katalogdaten können tiefergreifende Informationen beispielsweise auf Articlebene durch Anwendung von Web 2.0-Diensten in das elektronische Bibliotheksumfeld einfließen. So werden die bereits erwähnten *Current Papers* des Suchraums *Articles* in ALBERT angeboten. Zur effektiven Ausnutzung vorhandener Angebote von Verlagen eignet sich die Push-Technologie RSS-Feed ausgezeichnet, um proaktiv relevante Informationen vor allem speziellen Nutzergruppen oder hinsichtlich individueller Fragestellungen zur Verfügung zu stellen (Hart 2007).

Die Bibliothek des Wissenschaftsparks Albert Einstein verwendet zu diesem Zwecke im Backend-Bereich die unter GNU GPL lizenzierte Open Source-Applikation Tiny Tiny RSS¹⁵, die als technische Grundlage eine LAMP-Installation benötigt (wahlweise PostgreSQL statt MySQL als Datenbank) in Verbindung mit einigen PHP- bzw. JavaScript-Bibliotheken. Tiny Tiny RSS wird als webbasierter News Feed Aggregator eingesetzt, der viele hilfreiche Funktionen mitbringt wie Kategorisierung, Labeling, Article Scoring u.v.m. Als nützliches Feature erweist sich insbesondere das Content Filtering, um unerwünschte Inhalte wie Editorials, Acknowledgements oder Author Lists auszuschließen.

Als Alternative zu einem lokal aufgesetzten RSS-Aggregator bietet sich etwa der Service von ticTOCs¹⁶ bzw. dessen Nachfolger Journal TOCs¹⁷ an, das Table of Contents (TOCs) von mehr als 14.000 akademischen Zeitschriften und 500 Verlagen

¹⁵ <http://tt-rss.org>

¹⁶ <http://www.tictocs.ac.uk/>

¹⁷ <http://www.journaltoocs.hw.ac.uk/>

sammelt (Wirth 2009). Der administrative Aufwand dieser frei im Web verfügbaren Anwendung ist logischerweise gering und kann als Vorteil gelten. Hervorgehoben werden muss auch die Möglichkeit, die Dienste von Journal TOCs als Webservice mittels API in eigene Anwendungen oder Webseiten einzubauen zu können. Allerdings verhindert der geringe administrative Aufwand auch die Verwirklichung eigener spezieller Bedürfnisse und führt zu einer Systemabhängigkeit, die eben auch Limitierungen beinhaltet.

Momentan werden mit Tiny Tiny RSS täglich die Inhaltsverzeichnisse sowie vielfach auch die Vorveröffentlichungen von etwa 700 Zeitschriften per RSS-Feed abgeholt und in ALBERT integriert. Bedingung ist die Verfügbarkeit des Volltexts und ein ins Themenspektrum passendes Fachgebiet dieser Titel. Die meisten großen Verlage bieten RSS-Feeds an, wobei als Nachteil die teils sehr unterschiedliche qualitative Ausprägung der zugrunde liegenden Feeds zu bemängeln ist. Auch nachträgliches "Korrekturparsen" oder Workarounds können die Mängel nur partiell beseitigen und sind aufwendig zu generieren. Für Online-Zeitschriften meist kleinerer Verlage oder herausgebender Gesellschaften, die nicht über große Plattformen angeboten werden, existieren leider oftmals keine RSS-Feeds. In Fällen von hohem Stellenwert wurden Individuallösungen gefunden, wie dies zum Beispiel für die elektronische Ressource GeoScienceWorld (GSW) umgesetzt wurde, einem Angebot führender geowissenschaftlicher Organisationen¹⁸. Statt nichtexistenter RSS-Feeds werden hier die angebotenen E-Mail-Alerts genutzt und als Endnote-File exportiert. Unter Umgehung der eigentlichen RSS-Applikation gelangen diese Referenzen direkt in die Datenbank, nachdem ein vorgelagertes Skript für die entsprechende Aufbereitung der Dateien sorgt. Der Aufwand für eine solche Menge thematisch einschlägiger und vielrezitierter Zeitschriften (>30 Titel) kann als geringfügig bezeichnet werden. Der Wert der Integration wiederum ist umso umfassender, da auf diese Weise die neuesten Inhalte von Titeln in die RSS-Datenbank bzw. ins Frontend ALBERT gelangen, die qua Verlagsangebot so dafür nicht geeignet sind. Hier offenbart sich der Vorteil quelloffener Software bzw. von Eigenentwicklungen, da das Anpassen an spezielle Bedürfnisse von Anwendern oder in diesem Falle Bibliotheken überhaupt erst möglich ist. Dem verwendeten RSS-Aggregator konnten überdies leicht weitere Skripte

¹⁸ <http://www.geoscienceworld.org/>

hinzugefügt werden, um beispielsweise den Export zu ALBERT mit Informationen aus der ERM-Datenbank anzureichern.

Kurzer Ausblick

Abschließend kann festgestellt werden, dass sich die Backend-Ebene zu einem vor allem technisch homogenen Umfeld entwickelt hat und auf dieser Basis weiter ausgebaut werden wird. Die Dokumentenlieferung beispielsweise wird in Zukunft auf dem Fundament einer LAMP-Installation entstehen und als interaktive Web-Applikation etwa von der Ajax-Technologie profitieren

Folgt man den Voraussagen wird die Informationsmenge in Zukunft wohl eher größer als kleiner (Burgelman, Osimo et al. 2010), was sich bei der Weiterentwicklung des Frontends ALBERT im Thema Personalisierung niederschlagen wird. In Betracht kommen neue weiterentwickelte Ansätze des Content Filterings sowie die Verwendung von Visualisierungstools. Verwandte Techniken aus dem Bereich des Text Mining bzw. Data Mining könnten hingezogen werden, um neue verborgene Zusammenhänge aufzudecken. Welche unterstützenden Eigenschaften bei der Kollaboration von Autoren und Wissenschaftlern bibliotheksspezifische Suchmaschinen bieten können, kann ebenfalls Gegenstand der Entwicklungen sein. Für den Endnutzer reizvolle und bequeme Volltext-Recherchen in lizenzierten elektronischen Artikeln oder Büchern, ähnlich Google Books oder Amazons Search Inside!, ohne Anzeige des vollständiges Texts, sind als weitere Optionen denkbar. Auch die Integration von News und/oder Jobs des Wissenschaftsbereichs vor allem hinsichtlich der Aggregation verschiedener Quellen und unmittelbarer Verfügbarkeit (Live-Update) werden zukünftig Themen der weiteren Entwicklung sein.

Literatur

Bertelmann, R., K. Herm, et al. (2007). "Bibliothekssuchmaschine statt Bibliothekskatalog." *Bibliotheksdienst* 41(12): 1302-1306.

Breeding, M. (2008). "Helping You Buy: Electronic Resource Management Systems." *Computers in Libraries* 28(7): 13437.

Burgelman, J.-C., D. Osimo, et al. (2010). "Science 2.0 (change will happen ...)." *First Monday* 15(7).

Collins, M. (2008). "Electronic Resource Management Systems (ERMS) Review." *Serials Review* 34(4): 267-299.

Hart, L. G. (2007). "Library 2.0: RSS Feeds Dynamic Uses for Special Libraries." Special Libraries Association (SLA) 2007 Annual Conference, Denver 3-6 June, 2007, Alexandria, Virginia, USA.

Hawthorne, D. and J. Watson (2007). "Electronic Resource Management Systems: Alternative Solutions." World Library and Information Congress: 73rd IFLA General Conference and Council, 19-23 August 2007, Durban, South Africa.

Herb, U. (2007). "Ohne Web 2.0 keine Bibliothek 2.0." telepolis, 13.09.2010.

Keller, A. (2002). "Elektronische Zeitschriften: Was sagen Nutzungsstatistiken aus?" B.I.T. online 3/2002.

Murray, A. (2008). "Electronic Resource Management 2.0: Using Web 2.0 Technologies As Cost-Effective Alternatives To An Electronic Resource Management System." Journal of Electronic Resources Librarianship 20(3): 156-168.

Sadeh, T. and M. Ellingsen (2005). "Electronic resource management systems: the need and the realization." New Library World 106(1212/1213): 208-218.

Schüller-Zwierlein, A. (2007). "Electronic Resource Management für alle? Automatisierte Contentanalyse mit Excel und VBA." Bibliotheksdienst 41(2): 163-172.

Soules, A. (2010). "New e-sources, new models: reinventing library approaches to providing access." Library Hi Tech News 27(2): 10-14.

Stowell Bracke, M., J.-P. V. M. Hérubel, et al. (2010). "Some Thoughts on Opportunities for Collection Development Librarians." Collection Management 35(3-4): 255-259.

Wirth, A. A. (2009). "ticTOCs: A New Service for "Keeping Current"." Collection Management 34(3): 229-233.