

Protokoll 6. Sitzung Fachausschuss Planung und Entwicklung am 13.10.2010

Vortragsraum der Württembergischen Landesbibliothek, 10:30 Uhr bis 15:00 Uhr

Teilnehmer:

Stefan Abele (WLB),
Klaus Peter Aiple (UB Hohenheim),
Volker Conradt (BSZ),
Marianne Dörr (UB Tübingen, Vorsitz),
Hans-Joachim Fuchs (UB Tübingen),
Guido Hölting (KIZ Ulm),
Jochen Lienhard (UB Freiburg, bis zur Mittagspause),
Christof Mainberger (BSZ, Protokoll),
Leonard Maylein (UB Heidelberg),
Bettina Sabelus (HSB Reutlingen),
Matthias Schulze (UB Stuttgart),
Hannah Ullrich (UB Freiburg, bis zur Mittagspause)

Frau Dörr begrüßt die Anwesenden, verabschiedet bei allgemeiner Zustimmung das Protokoll der letzten Sitzung und bittet Herr Lienhard und Frau Ullrich um die Präsentation der Fortschritte an der Freiburger IPS Installation.

Präsentation des Freiburger Entwicklungsstands von IPS

Herr Lienhard rekapituliert zunächst die Ausgangslage: Metasuchportal seit 2006, Erweiterung um Suchmaschinentechnologie seit letztem Sommer. Die Motivation zum Ausbau der Anwendung durch Suchmaschinentechnologie lag besonders im Drill Down (Facettierung), in der Antwortgeschwindigkeit und in den Rankingmöglichkeiten (insbesondere bei Integration von Abstracts und Volltexten).

Aktuell werden diverse Datenquellen (Biosis, Nationallizenzen) per Z39.50 eingebunden und der OPAC mittels Solr/Lucene indexiert. Dazu wurden mit dem BSZ geeignete Verfahren zur Datenlieferung mittels Marc21 aus dem SWB entwickelt und zahlreiche Fragen geklärt, von der passenden Zeichencodierung, über Kategorien insbesondere im Lokalsatz und im Zusammenhang von Bandreihen und der Einbettung von Normdaten bis zum Handling großer Datenbestände in Zeitscheiben

Eine Indexierung von ca. 2,4 Mio. Titeln erfordert 2,5 Stunden, die Datenaufbereitung im Vorfeld ca. drei Stunden.

Der OPAC-Index wird über eine XML-Schnittstelle in IPS eingebunden. Die Suche in Fachdatenbanken wird über die föderierte Suche in IPS realisiert. Dabei werden verschiedene Fachpakete angeboten. Über eine einheitliche Oberflächengestaltung soll eine homogene Nutzererfahrung in den getrennten Bereichen ermöglicht werden.

Offen ist bislang, wie weitere Indexe angesprochen werden können; hierfür kommen insbesondere Ebsco Discovery Service und die Suchkiste des GBV in Betracht. Dazu müssen für unterschiedlich indexierte Bestände Drill-Down und Ranking koordiniert und Restriktionen durch Lizenzen beachtet werden. Der Ebsco-Index kann finanziert werden; seine Einbindung ist noch unklar.

Diskussion

Herr Maylein schlägt vor, statt dem Einkauf ganzer Indexe eher die Rohdaten zu beschaffen, auch wenn das Einsammeln mit Schwierigkeiten verbunden ist. Der Vorteil liegt in der Autonomie, die Daten nach eigenen Ansprüchen aufzubereiten.

Alternativ können die eigenen Datenbestände (OPAC, etc.) auch bei Anbietern von Discovery-

Services (Serial Solution, Ebsco) indexiert werden, wobei man dann deren Formate akzeptiert.

Herr Mainberger hält es zur Beurteilung der Alternativen für grundlegend zu klären, welche Datenquellen die Bibliotheken insgesamt zur Recherche anbieten wollen, welcher Teil dieser Liste (zu welchem Preis) bereits im Sortiment der Discovery Service Anbieter vorhanden ist, wieviel eine beauftragte Indexierung jener Quellen, die dort nicht indexiert sind, bei den Anbietern jeweils kostet und ob dies nach den Lizenzen für die Datenquellen möglich ist. Schließlich muss auch der Personalaufwand für die Normalisierung von Datenquellen beachtet werden, wenn man dies in eigener Regie tut.

Herr Maylein wirft ein, dass die Konfiguration der Indexierung einer Quelle zunächst nur anfangs anfällt und der Aufwand sich reduziert, wenn man sich am elementaren Bedarf der Nutzer orientiert. Herr Aiple betont, dass beim Vergleich der Service Anbieter zu beachten ist, dass sie die einzelnen Metadatensätze in unterschiedlichen Umfang indexiert anbieten.

Herr Fuchs stellt zwei Vorteile der Service-Anbieter heraus: Sie können als Aggregatoren bei den Datenproduzenten bessere Preise erzielen und integrieren Durchgriffsmöglichkeiten auf die Volltexte in ihre Angebote (Link-Resolver-Funktionalität).

Herr Lienhard berichtet, dass die Abdeckung des Bedarfs der UB Freiburg im Ebsco-Index noch nicht befriedigend ist; auch Frau Sabelus bestätigt dies: ein Vergleich war für die HSB Reutlingen ernüchternd und eine Teststellung wird momentan nicht zufriedenstellend betreut. Der Bestand der HSB Reutlingen wurde von Ebsco auf Basis eines Verbundabzugs integriert.

Ebsco hat in seinem Angebot die Erweiterung um eine föderierte Suche integriert. Serial Solution sammelt und normalisiert für Summon Daten (ca. 500 Mio. Artikeldaten) und bietet sie unter einer Oberfläche an. Beide bieten die Indexierung von lokalen Quellen im zentralen Index.

Wie Herr Mainberger und Herr Fuchs ausführen, läuft das parallele Absuchen verschiedener Suchmaschinenindexe auf eine föderierte Suche hinaus, solange diese nicht Ranking und Drill-down über den Austausch von Wortstatistiken koordinieren bzw. sowieso zu einem gemeinsamen Index gemergt werden.

Herr Fuchs stellt fest, dass der SWB bereits einen größeren Suchraum darstellt als der IPS-Index an der UB Freiburg, der momentan lediglich deren OPAC umfasst.

Auf Nachfrage von Frau Dörr erläutert Herr Mainberger die Problematik der Lokaldaten und der Verfügbarkeit. Hierbei geht es um die Frage der Integration des Services unter einer Oberfläche, also ob der Ausleihstatus erst auf der Einzeltrefferanzeige bzw. über einen weiteren Klick ins Ausleihsystem ersichtbar ist oder bereits in der Kurztrefferanzeige des Portals angezeigt wird. Die zweite Variante kann über eine regelmäßige Übertragung der Änderungen der Ausleihdaten ins Portalsystem erreicht werden (Primo) oder jeweils über Anfragen pro Treffer auf der Kurztrefferanzeige im Ausleihsystem. Beides ist technisch, jeweils mit Vor- und Nachteilen, performant realisierbar.

Kurzbericht über bisherige Touchpoint-Erfahrungen und Informationen zu Exlibris Primo

Herr Mainberger berichtet von der TouchPoint-Teststellung, die das BSZ seit Juni durchführt. Die Nachfolge-Software für Elektra von OCLC weist zwar eine Reihe von Fortschritten auf: Die Softwarearchitektur ist besser gegliedert, insbesondere ist die Oberfläche von der Programmlogik entkoppelt und damit robuster gegenüber Systemupdates. Allerdings gibt es noch Defizite in der Ausgestaltung von Mandanten (Sichten) und der Einbindung von Http-Quellen. Diese ist per Z2Web vorgesehen, was eine Neuimplementierung der Anbindungen in Perl erforderlich macht. Dies würde für die mittlerweile mehreren hundert in die Portale des BSZ eingebundenen Datenquellen einen immensen Aufwand bedeuten. Eine Einführung des Systems ist daher derzeit schwer vorstellbar.

Um eine Migration zu ermöglichen, müssten entweder die Elektra-Anbindungen nachnutzbar sein oder bereits eine relevante Menge von Adapter vorliegen, was bei IPS der Fall ist, ebenso wie evtl.

bei MetaLib.

Herr Mainberger informiert über Gespräche mit der Firma ExLibris, in denen ein Einsatz von MetaLib mit Primo-Oberfläche im Rahmen der derzeitigen Wartungspreise von Elektra skizziert wurde. Zu klären wäre, inwiefern Primo-Central hierbei genutzt werden kann und wie sich das Angebot in einen preislichen Rahmen einbettet, nach dem einzelne Bibliotheken auch Primo nutzen. Erforderlich wäre eine Preisanfrage, die Universitätsbibliotheken, Landesbibliotheken, Hochschulen sowie kommunale Einrichtungen umfasst.

Diskussion

Frau Dörr sieht für eine einheitliche Entscheidung über die Ablöse von Elektra Zeitdruck, da das Thema in den Bibliotheken aktuell ist und bereits einzelne eigene Wege einschlagen. Ein gemeinsames Vorgehen wäre im Hinblick auf die Wirtschaftlichkeit vorzuziehen. Frau Sabelus ergänzt, dass die Hochschulen und Pädagogischen Hochschulen auf einen Service des BSZ angewiesen sind. Die UB Tübingen und das KIZ Ulm werden zwar den aDIS-OPAC zunächst einführen, sehen dies allerdings lediglich als Übergangslösung zu einer umfassenderen Portal-Lösung, die schwieriger zu entscheiden und bis zu einer zweiten Phase zurückgestellt ist.

Zusätzlich, betont Herr Conradt, steht mit den Lokalen Sichten auf den SWB eine weitere OPAC-Variante zur Verfügung. Herr Fuchs favorisiert solche zentralen Lösungen, die allerdings konsequent bis zum WorldCat gedacht werden sollten.

Herr Maylein erkundigt sich, ob es eine Perspektive gibt, schrittweise Daten zur Recherche in den SWB zu laden und sich bei der Normalisierung kooperativ zu unterstützen. Dazu zieht er zunächst die fünf bis zehn größten Verlage in Betracht. Herr Conradt antwortet, dass es eine strategische Zielrichtung des BSZ darstellt, Daten zu aggregieren, wie es bereits im Bereich Swets-Aufsatzdaten und E-Books Praxis ist. Die Normierung ist aufwendig und die Abstimmung mit den Datenlieferanten nicht zu unterschätzen. Momentan sei allerdings noch eine geringe Nutzung der Swets-Aufsatzdaten zu beklagen. Herr Fuchs erklärt dies mit der getrennten Haltung in einem eigenen Datenbereich, der die Recherche nach den Aufsatzdaten in den lokalen Sichten unkomfortabel macht.

Herr Maylein hat bei der Indexierung verschiedener Fremddaten (Nationallizenzen, E-Books, Munzinger-Daten) gute Erfahrungen gemacht, wobei der Konverter durch nutzerorientierte Abstriche bei der Aufbereitung vereinfacht werden konnte.

Frau Sabelus schlägt vor, die Initiative nicht auf Baden-Württemberg zu begrenzen. Herr Conradt stellt dar, dass im Kontext der Fernleihe zwischen den Verbundsystemen der Austausch ganzer Lucene-Indices verabredet wird. Für sächsische Bibliotheken, die mit vuFind arbeiten wollen, werden Rohdaten zum OPAC, E-Books und Swets im Marc21-Format zur Verfügung gestellt.

Zur Identifizierung der für die Bibliotheken insgesamt interessanten Datenbestände kann man sich, auf Vorschlag von Herrn Hölting, auf jene beschränken, für die die Bibliotheken auch Volltextangebote lizenziert haben. Hinsichtlich der elektronischen Zeitschriften sind diese Informationen in der EZB hinterlegt, aus der man die relevanten Verlage zusammenfassen kann. Eine entsprechende Quelle für Fachdatenbanken stellt DBIS dar. Die Ergebnisse können durch die Experten des Baden-Württemberg-Konsortiums kontrolliert bzw. ergänzt werden.

Frau Dörr fasst die Diskussion in **zwei Modelle** zusammen:

1. Outsourcing Modell (Software as a Service, SaaS):

Das Discovery-System wird an einen Anbieter vergeben. Dazu müssen die Systeme von Ebsco, Serial Solutions und ExLibris evaluiert werden, welche Inhalte sie bieten bzw. integrieren können, welche Funktionalitäten in den bereitgestellten Oberflächen vorhanden sind und ob bzw. welche Konsortialmodelle vorstellbar sind. Hinsichtlich Primo müssten ggf. noch Kosten

für ein Hosting beim KOBV berücksichtigt werden.

2. Modell für eigene, kooperative Indexierung

Hierfür müssen Verlage oder auch Aggregatoren durch das Konsortium nach Metadaten gefragt werden, um eine möglichst breite Abdeckung zu erreichen. Die Verlage dürften zunehmend zugänglich sein, da es überzeugende Beispiele (Wiley) für eine bessere Nutzung gibt, wenn die Daten in einem Index enthalten sind.

Die Daten könnten beim BSZ aggregiert werden. Eine Kooperation mit anderen Verbänden, z.B. dem KOBV, der für Primo bereits Daten aufbereitet, wäre denkbar. Anhand von Swets-Scan und E-Books gibt es hierzu bereits Erfahrung beim BSZ; Aufwände können auch anhand des Mannheimer Vorbilds abgeschätzt werden. Zu beachten ist, dass mit der Aktualisierung der Indexe ein Daueraufwand entsteht.

Herr Maylein schlägt vor, in die Kalkulation noch IPS als Oberfläche aufzunehmen, da er im Aufbau von Oberflächen ein wesentliches Problem erkennt. Herr Mainberger überlegt, ob sich eine solche lose Nebeneinanderstellung von Metasuche und Suchmaschinentechologie wie in IPS nicht auch erschwinglicher z.B. mit Elektra kombiniert mit vuFind erzielen ließe.

Frau Dörr schlägt vor, dem Kuratorium zu empfehlen, eine Evaluation dieser beiden Modelle vornehmen zu lassen. Als Kriterium ist auch der Zeitrahmen der Realisierung relevant.

Herr Conradt stellt fest, dass Elektra bis zu einer Ablösung im Bereich der kommunalen Anwender, der Hochschulbibliotheken sowie des Fernleihportals angeboten wird. Für weitere Bibliotheken stellt auch eine Lokale Sicht auf den Verbund eine vorläufige Option da. Es besteht Zuversicht, dass die lokalen Sichten in einer zukünftigen Version auch mit Drill-Downs ausgestattet werden können. Die Aggregation von Daten, die Bereitstellung als Marc21 und der Austausch von Lucene-Indexen werden vorangetrieben, allerdings muss man hierbei noch Erfahrungen gewinnen.

Er schlägt vor, Tests für Datenaggregation und -Austausch anhand des Lucene-OPACs und der Leipziger vuFind-Installation durchzuführen.

Frau Dörr dankt den Teilnehmern für die Diskussion. Im Kuratorium wird sie die Ergebnisse berichten und die Evaluation der Alternativen empfehlen. Die nächste Sitzung des Fachausschusses soll Ende November stattfinden, der Termin dafür wird noch abgestimmt.