

Next-generation library catalogues: review of E-LIB Bremen

MARTIN BLENKLE, RACHEL ELLIS AND ELMAR HAAKE

State and University Library Bremen, Germany

The electronic library of the State and University Library Bremen (E-LIB) is based on search engine technology and offers an integrated catalogue of electronic and print media. All of the library's available media and services can be accessed via a single search interface (based on the idea of a 'one-stop-shop portal') instead of the separate catalogues for different media types still to be found in many libraries. The idea behind this system is to combine all kinds of local holdings and all the external resources available for our users in one single retrieval system and to offer new methods for supporting associative search processes.

There are about 20 million records – nearly 80% of these are electronic full-text documents – in the library-controlled web space of the E-LIB Bremen. The technology behind this system is based on a locally hosted index of heterogeneous, aggregated metadata. The data is organized by the CiXbase¹ system that is not only used and refined in Bremen but also at the ETH Zürich (NEBIS) in Switzerland. The E-LIB Bremen system is both a live and a test system for new services².

Typical advantages of a search engine using a local index are the high response rate, the homogenous display of results³ and a well-balanced result list. In addition, the metadata of search results establish a knowledge base that offers additional options for new forms of analysis that can be processed during the search.

Content

- 3,500,000 books and other local holdings
- 30,000 journals
- 17,000,000 article-level records
3,000,000 of these comprise selected articles from open access repositories
- 200 bibliographic databases
- 8,000 licensed e-books
- 10,000 selected free internet resources

- 30,000 research and expert database records of the University of Bremen.

The metadata of selected document servers in the context of the Open Archive Initiative is imported via OAI-harvester technology. The integration of these open access document metadata in the library catalogue is a new kind of service level, which also serves as a marketing tool to improve the awareness of open access publications. The article-level-based data enables a deep linking functionality to the publisher's full text. The article-level links are included directly in the result presentation. Information about the access conditions are shown for licensed electronic media. The range of data is expanded continuously via integration of further metadata.

Local indexing and interoperability

Various types and formats of metadata are imported and converted to a homogenous XML flat file structure in order to build the local knowledge base. The resulting consistent metadata pool is indexed via search engine technology based on statistical balanced B*-trees algorithms serving as a basis for the search engine. The results are presented in form of XML-structured metadata within the browser. This XML data is converted by stylesheet-based XSL transformation within the client's browser application⁴.

This separation of structure and layout enables nearly complete interoperability of the available metadata and simplifies further processing of this data in other information systems as well as the integration of external services into the E-LIB.

Export functions for citations are provided mainly for different bibliographic database formats such as RIS, Endnote, etc. The 'import into reference database' software is particularly convenient if the data can be delivered to a web-based system like RefWorks or Connotea.

Export features for metadata

- RefWorks, Endnote, RIS, BibTeX
- Social bookmarking: Connotea, del.icio.us, Bibsonomy
- Zotero compliant via COinS
- RSS feature for arbitrary search queries.

This open display architecture has also enabled us to extend our result presentation easily via an integration of web services of other external information providers – thus realizing mashup functions.

Additional information related to the book titles is integrated using an experimental service from Germany's Common Library Network (GBV)⁵ called SeeAlso. For example, we integrate links to titles in Google Book Search, LibraryThing or Wikipedia in the record presentation, if available from the service provider.

Discovery search and recommendation features

The search box is always present at the top of the screen throughout the whole session – a user can

start a new search at any point. The results are presented in a two-column display. The left column displays the results, while the column on the right provides the drill-down options and additional features (see Figure 1).

A search within such a large database often results in a confusing number of results. The search process thus has to be supported by adequate methods of refining the search inputs. In addition to formal and content-driven drill-down functions, you find a number of recommendation tools for the modification of search results. During the search a user has the chance to alter the search strategy in the sense of 'discovery search'. The subject-based search modus then turns into a multilevel iterative navigation within a semantic net that has been built on the entity of all available subject-indexing data of the bibliographic records.

In order to achieve this, all the material of controlled vocabulary of search results is analyzed, statistically evaluated and presented as a tag cloud (see Figure 2). The result of this analysis – the relevance profile of the tag cloud – depends on the chosen term as well as the available title material and its subject index elements.

The screenshot displays a library search interface. At the top, there is a search box containing the text 'Science Philosophy' and a 'Start Search' button. Below the search box are radio buttons for search filters: 'journals only', 'electronic journals only', 'books only', and 'e-books only'. A navigation bar includes links for 'Search', 'Search for authors', 'Folder', 'Virtual book shelf', 'FAQ / Help', and 'Home'. Below the search box, the results are displayed in a two-column format. The left column shows a list of search results, each with a title, author, and a '100%' completion indicator. The right column, titled 'Refine by', provides options to filter results by 'Type of Media' (Books, E-Books, Printed journals, etc.), 'Location of books/journals' (Uni Bremen, TB Bremerhaven, etc.), and 'Users favorites'. It also includes a 'Limit language of titles to:' section and a 'History of publication' bar chart for 'Science Philosophy' showing the number of titles published from 1735 to 2010. At the bottom right, there is a 'Refine your subject:' section with a tag cloud for 'Science Philosophy' including terms like 'Cell & Developmental Biology', 'Life and Medical Sciences', and 'Cognitive science'.

Figure 1. Two-column display with search box at top of screen



Figure 2. Example of a tag cloud

The main features of the discovery search interface are

- ajax-based suggest function to request and display other user queries without having to reload the web page
- “Did you mean” function and spell-checker based on Levenshtein distance⁶
- tag cloud presentation of related terms or authors via computer linguistic and statistical analysis of results
- recommender for external bibliographic databases based on subject recognition of the query
- presentation of relevant local classification scheme entries offering the display of all books allocated there
- users’ favourites based on ‘URL click activity’
- display of a subject or of an author’s publication history graph based on the publication years of the search results.

Handling of ‘no hits’

The analysis and adequate handling of ‘no hits’ results is an interesting application for a search engine assistant. No or just a few hits are usually either caused by titles actually not to be found in the knowledge base or by spelling mistakes. The E-LIB search engine therefore offers a spell checking and recommendation (“Did you mean”) facility – using the principle of the Levenshtein-Distance algorithm that works with various languages due to the index material available.

In the case of no or just a few hits, the E-LIB system checks the availability of data in the

common library network’s union catalogue via SRU request. In case of a successful response, an order button is offered without switching to another system.

This method can also be used to combine hits from external bibliographic databases with the local holdings – always considering the access conditions for those as well. In the near future the intelligent combination of local indexing and cross-searching will combine the major advantages of both search techniques.

Service-oriented architecture

In addition to the classic search functionalities, one aim of the E-LIB system is to offer new views of the available metadata collection in order to support the data management on the campus in a convenient way. Our students can use the whole variety of E-LIB’s knowledge base directly without diversion within their learning management systems via SRU interface. The results are presented in the ‘look-and-feel’ of the course system. As the search engine offers access to more than 17,000,000 electronic resources with full-text links – among those thousands of e-books – these can easily be used to create electronic course packs.

In addition to the search engine, many of the other E-LIB services are designed in a modular way and can thus be integrated as an independent service into external systems. Such services, e.g. the tag cloud generation and the subject analysis of the request as basis for database recommendations, can be supplied. Provided that the knowledge base has a certain size, is comprehensive and evenly spread on the different subjects, it can be assumed that the services can reach a certain degree of independence from the local holdings and can universally be integrated into other systems.

See E-LIB Bremen at:

<http://suche3.suub.uni-bremen.de/indexEN.html>

References

1. CiXbase is a software based on the OSIRIS Project – a retrieval assistant based on computer linguistic methods. CiXbase:
<http://cixbase.dyndns.org/CiXbase/cixdocs/>
 (Accessed 28 April 2009)
 NEBIS/ETH Zürich:
http://www.nebis.ch/index_e.html (Accessed 28 April 2009)

- CAMbase:
http://cambase.dmz.uni-wh.de/CiXbase/camdb/index_en.html ((Accessed 28 April 2009)
2. Currently 8,000–13,000 daily requests are submitted showing 40,000 different IP numbers per month.
 3. Response times of similar systems based on federated cross-searching techniques often reach a critical stage compared to well-known search engines like Google.
 Rochkind, J, (Meta)search Like Google. The time has come for libraries, too, to negotiate for rights to index full text, *Library Journal*, 2/15/2007 :
<http://www.libraryjournal.com/article/CA6413442.html> (Accessed 28 April 2009)
 4. Browsers that do not offer the facility of a client-based XSL transformation are supplied directly with HTML code that is generated by an Apache/Cocoon framework on the server.
 5. Common Library Network in Germany (GBV):
<http://www.gbv.de/vgm/index?lang=en> (Accessed 28 April 2009)
 'See Also' Service: Voss, J, Getting started with SeeAlso:
http://www.gbv.de/wikis/cls/Getting_started_with_SeeAlso (Accessed 28 April 2009)
 6. Levenshtein-Algorithm:
<http://www.levenshtein.net> (Accessed 28 April 2009)

Reviews of other next-generation library catalogues were published in the previous two issues of *Serials*, as follows:

Serials 21(3), 224-230:

Aquabrowser
Elin
WorldCat Local

Click here to access:

<http://serials.uksg.org/openurl.asp?genre=article&issn=0953-0460&volume=21&issue=3&spage=224>

Click here to access via DOI:

<http://dx.doi.org/10.1629/21224>

Serials 22(1), 68–82:

Encore
Primo
Summa
Summon

Click here to access :

<http://serials.uksg.org/openurl.asp?genre=article&issn=0953-0460&volume=22&issue=1&spage=68>

Click here to access via DOI:

<http://dx.doi.org/10.1629/2268>