

A New Search Engine for Web Portals

Win Nwe Ni, LIBRIS and Eric Bonneau, CSC

Over the past years, the LIBRIS web site has evolved to become a full-featured information hub today. The site offers online users information and services ranging from OPAC searches, online databases and forms and floor plans to the latest news and announcements and events pertaining to LIBRIS.

Although constant efforts have been made to ensure that information within the site is well-organised and easily accessible, the increase in online content makes it more difficult for users to find what they are looking for. New tools need to be put in place to ensure that information is easily accessible.

A new search engine for LIBRIS online content

A search engine would index the LIBRIS web pages and allow users to enter keywords in order to display relevant pages. The new engine would complement the existing OPAC search used to find items within the LIBRIS online catalogue.

Linux based search engine

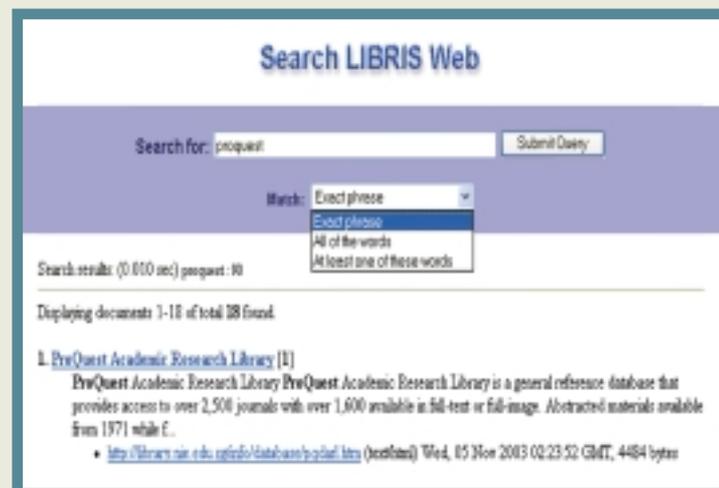
After reviewing a number of search engines on the market, the LIBRIS and CSC team decided to implement an open source solution based on the Linux operating system. The team hoped to limit the cost of setting up a server and therefore a virtual server designed to host the search engine was the preferred choice. For more information on the virtual server, please read the article featured in *ACIS Plug-in Volume 1, Issue No. 4 (October 2002)*.

Search engine features

Open source software, apart from being inexpensive or free of charge, can also provide the functionalities found in more costly commercial packages.

The search engine that was implemented offers three different search methods, and users have the option of choosing the one that most suits their needs. Searches are not case-sensitive and the indexes are rebuilt on demand or based on a defined schedule. The default search method available in the LIBRIS web search engine is the "Exact Phrase". With this method, the search engine will display results that contain at least one occurrence of the exact string of words entered.

First Method – Exact Phrase



Second Method – All of the words

The second method "All of the words" will display results containing at least one occurrence of every single one of the keywords entered.

Third Method – At least one of these words

The last method "At least one of these words" will return results containing at least one of the keywords entered. Not all the keywords need to be present for the result to be displayed.

Ordering the search results

The search engine will select the results based on the search method and then display those results following an algorithm called the weight factor.

For the search engine, a web page is divided into 4 major sections (Description, Keywords, Title and Body). The administrator of the web site decides what sections are the most important and in the case of the LIBRIS web site, more weightage was given to the title of the web documents.

The search results will be presented in order on the screen based on the number of occurrences of the search words and also in which section of the document these search words are found. Search words found in the most important section of a document are given priority.

A search engine as an added feature to your web site

This open source search engine can bring more value to your web site as your users will be able to find information more easily and rapidly. The search engine can also be configured to index selected internal and external web sites.

For more information on the search engine software currently being used on the LIBRIS web site, please visit: <http://search.mnogo.ru/>.