THEME PAPER

'RESEARCHER' AN OPEN SOURCE SOFTWARE SUIT FOR CONSORTIA

A R D PRASAD

Abstract

The paper describes a suit of Open Source Software suite 'reSearcher' which includes CUFTS (online serials management system), GODOT (an OpenURL compliant system) and dbWIZ (a federated search engine) and CM (citation manager) from Simon Fraser University, Canada. Presents the advantages of an integrated approach where these software are pieced together to provide federated search to a consortium of libraries. The federated search facilitates to query across desperate information resources like e-journals, Z39.50 servers, online databases, DSpace based digital repositories. In fact dbWIZ provides a single stop search engine for a wide variety of information sources. Illustrations are taken from the implementation at Indian Statistical Institute. However, this paper does not cover CM.

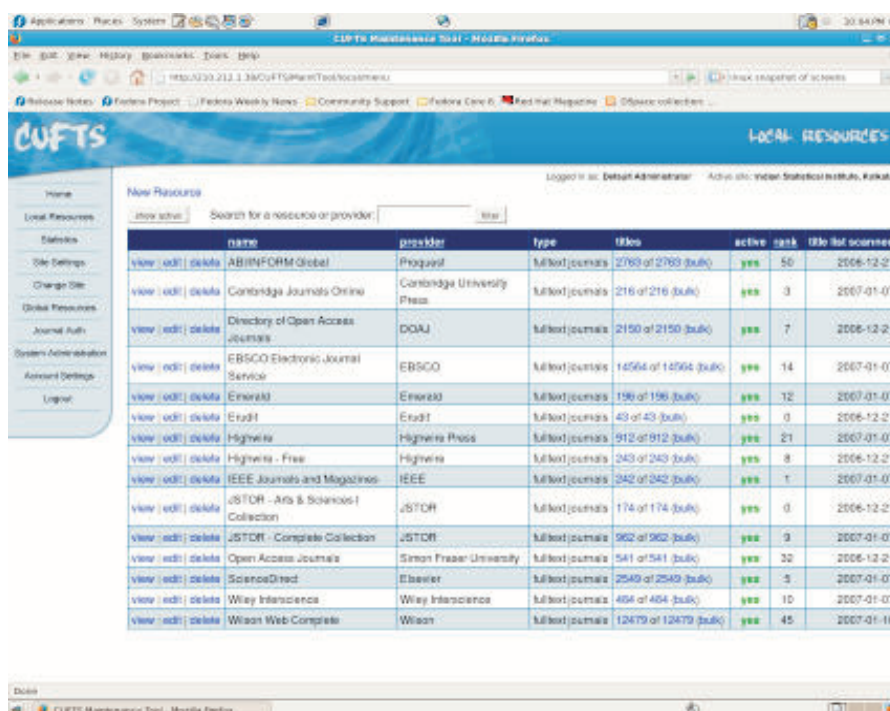**Keywords :** Open Source Software/ CUFTS,/GODOT/ E-resource Management

1. Introduction

The spiraling cost of journals forced both the publishers and libraries to adopt consortia model. In fact the escalating cost is one of the reasons for the emergence of more and more open access journals. Resource sharing, library networking are not new the library professionals. However, the advent of Internet and web technology made if possible to launch online databases, webOPACS, e-journals, digital repositories etc.The new challenge is to provide easy access to information available on the Net in various formats.

Information on the Internet appears in diversified forms like publications in e-journals, digital repositories, databases, MARC records, etc. The web search engines mostly index static web pages (surface web) and provide a search interface. The big question is: How do we get information from deep web? (also called invisible web, hidden web). The information in the deep web includes dynamic pages (generated in response to a query), subscription based information or information which requires authorization, data in database system (e.g. bibliographic data). Though some search engines can index Information in various formats like image, text etc. like jpeg, pdf, word, there should be a better mechanism to access information in databases, webOPACs and digital repositories or institutional repositories.

reSearcher was originally designed to be deployed by library consortia, allowing one installation to support multiple institutions. However, reSearcher may also be installed for single-institution deployment, as presently is the case with Indian Statistical Institute, Kolkata and we may have future plans to include other centres of Indian Statistical Institute at Delhi and Bangalore. The following sections describe the features and advantages of three software in reSearcher suit viz. CUFS, GODOT and dbWIZ

2.   CUFTS

CUFTS is an online serials knowledge base of full-text journal collections, providing libraries with a searchable serials database for their web site, MARC records for their catalogues, direct to article OpenURL link resolving, and electronic resource management (ERM) tools. The software comes with many global resources i.e. Information about various aggrerators and the journal they offer. Using the global resources, each individual library can choose the journals they are subscribing. CUFTS allows authenticated access mechanism for each individual library to maintain the list of journals they are subscribing. However, all the libraries can include open access journals from DOAJ (http://www.doaj.org) and also the Open Access journals of SFU, Hiwire etc.

The above list of aggregaters are taken while installing the software. However, CUFTS allows one add his library list and is highly customizable.

3.   Resource Comparison

With this CUFTS tool a librarian can compare two or more aggregaters to find how many journals are common among the aggregaters. For example, CUFTS configured for Indian Statistical Institute has a list of 14,564 journals from EBSCO Electronic Journal Service and 12479 from Wilson Web Complete. When the 'Resource Comparison' tool is used to find how many journals both the aggregaters are offering, the system showed 4027 duplicates i.e. The list of journals that are commonly offered by both aggregaters.   The output will be useful in making decisions like who should be preferred or how to bargain with the aggregaters.

CUFTS Electronic Resource Managment (ERM): CUFTS offers essential electronic resource management services, allowing a library to maintain the information about its electronic collections, including licensing terms, renewal dates, contacts, etc. CUFTS ERM can send a librarian renewal notification, reminding him of approaching deadlines. ERM allows the following information to be added to facilitate serials management.

- Costs
- Renewal dates
- Licensing terms
- Contacts
- Reports

The ERM may not have all the bell and whistles of a serials management system of many library management software, as one of the main objectives of CUFTS is to make the other components of reSearcher use the CUFTS database information to provide openURL and federated search facility.

4.   Browse & Search Journal titles:

For the end-user CUFTS provides browse and search facility to know whether a particular journal is accessible. For the convenience of the end-users librarians can classify the journals under various subject categories making it intuitive to his clientèle. This is achieved by using one of the tools of CUFTS i.e CJDB (CUFTS Journal Data Base) settings.

In brief CUFTS provides the following facilities:

- E-Resource management
- A – Z serials database
- Full-text link resolving
- Journal searching
- Collection comparison reports
- MARC records

5.   GODOT

Godot is yet another software in reSearcher suit, which serves as a link resolver. GODOT stands for Generalized Online Documents, Ordering, and Texts. GODOT facilitates direct links to full text collections, using the CUFTS knowledge base. It can also display holdings in your catalogue or other organizations catalogues in the consortium. GODOT, it is claimed (we have yet to use this feature of GODOT), can be integrated with Interlibrary Loan systems, including the open source OpenILL (Open Inter Library Loan) and Library Management Systems to provide direct or mediated interlibrary loan requests by your users. The advantages of GODOT are better realized while using the dbWIZ federated search, where GODOT displays the information about search results and in which library in the consortium a particular publication is available, in case your library does not have access (not subscribed) to the desired journals.

Though one can use GODOT interface to find out whether a publication is available in your library or in another library, it is used more as a support to dbWIZ. It can also be used while you are using aggregaters like EBSCO who support openURL. In fact, GODOT works more in the background and it is the task of the librarian to configure it correctly. The end user may not appreciate GODOT directly, except its manifestation while using federated search provided by dbWIZ. But again openURL is very powerful tools which allows end users to navigate from one publication to another publication. In essence, GODOT ...

- Places a link in online databases
- Works with CUFTS for full-text linking
- Searches local and remote catalogues
- Interoperates with many ILS and ILL systems
- is OpenURL compliant
- allows Google scholar
- allows crossRef and DOI (Digital Object Identifiers)

6.   GODOT and ILL

Much of GODOT's functionality and its configuration options are related to submitting Document Delivery requests based on the results of the Z39.50 broadcast search for holdings. A library's interlibrary loan agreements with other libraries can be configured as Direct, Mediated, Information Only or Not Allowed. GODOT also allows ranking different libraries so that preferred or geographically nearer library's holdings are displayed first on GODOT's main screen.

Yet another feature of GODOT is that it can facilitate automatic requesting where GODOT chooses the library to which the request should be sent, instead of the end user choosing a library.

GODOT can be configured in such way that an end user can authenticate himself against the library's Library Management Software (LMS) before placing a Document Delivery request. It is also possible to place restrictions on who can place some types of requests. In addition, the end user data can be retrieved from the LMS so that the user does not have to enter his information.

7.   dbWIZ

dbWiz allows federated search (also known as metasearch, broadcast search, parallel search) across multiple databases, web sites, catalogues, and other online resources from a single interface, and presents the results as an integrated list. In way, dbWIZ provides a single stop search facility and saves the hassle of the end user going to every resource to search for desired information. dbWiz uses Z39.50, XML gateways, and HTML parsing to search and retrieve resources. It allows librarians to customize the resources to be searched and even these resources can be grouped under various subject categorigs depending on the users interests. It has both simple and advanced search facilities. The search results are merged and can be sorted by title, author, date, resource etc. However, it should be noted that dbWIZ may give difference number of search results for the same query even when performed with a gap of few minutes. The reason being that it performs federated search in real time (like Z3950 clients and unlike Internet search engines). The variation in the number of search results can be because of elapsed time in accessing a resource or downtime of the machine hosting the resource. In principle, there is no limit on the number of resources one can search, however, performance will

decrease as resources are increased. dbWIZ can search across

- Library catalogues
- Subscription databases
- E-book collections
- Free web-based databases
- Internet search engines
- Institutional repositories
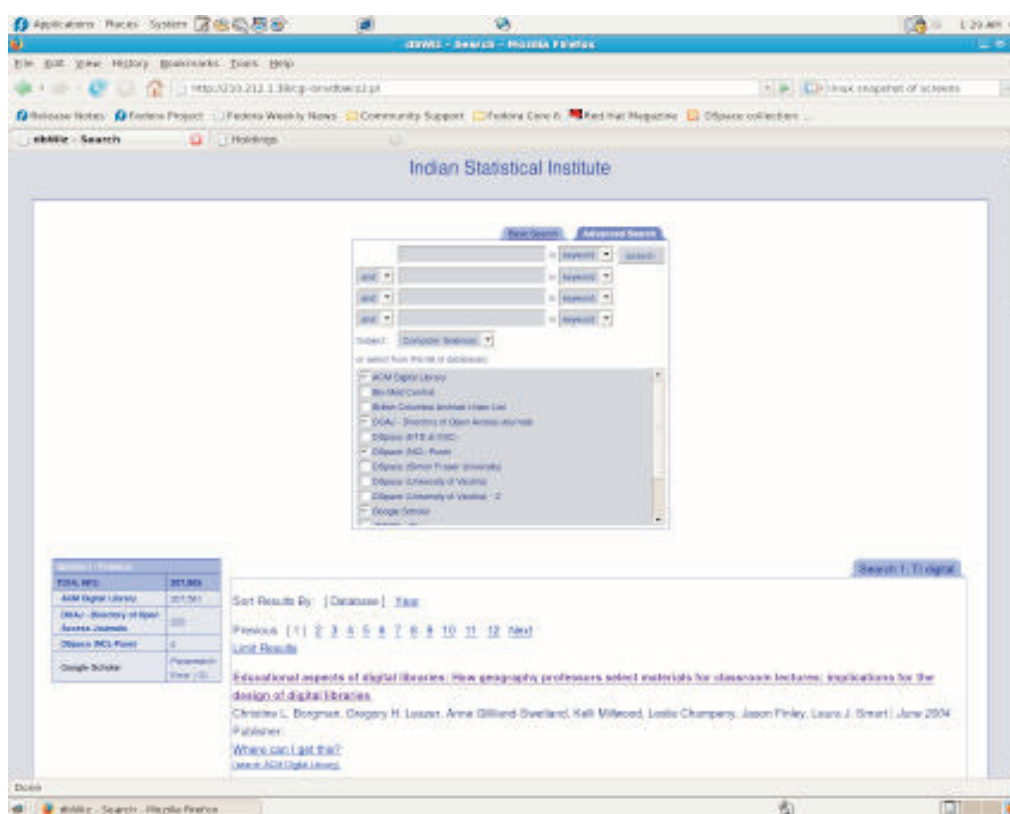
The screen shows the query for the word 'digital', and it is specified that it should appear in title and should be search in ACM Digital Library, DOAJ journals, DSpace repository of National Chemical Labs (NCL, Pune) and Google Scholar. The query produced a total of 207,805 results of which 207,581 are from ACM Digital Library, 222 from DOAJ and 2 from NCL and time got elapsed with Google Scholar. By clicking the title, one access the full text of the article if the user/library has access permissions. Alternatively, the user can click, 'Where can I get this?' button, and the information will be provided by GODOT about availability in other libraries in the consortium.

6.   Conclusion

Though CUFTS, GODOT and dbWIZ can be installed and used independently, the integration of these software provides more powerful environment. The reSearcher suit can be configured on any system with Linux/Solaris/BSD. It requires both PostgreSQL and mySQL and Apache Webserver. It was written in Perl and requires many perl modules to be downloaded from http://search.cpan.org). To make Z39.50 servers accessible, it requires YAZ Z3950 toolkit from http://www.indexdata.dk. The only commercial software that is required is EZProxy, which takes care of authentication requirements to access various subscription based resources for LAN users in an institution.

Though reSearcher suit of software are highly customizable using lots of templates, knowledge of XHTML and Cascading Style Sheets is important. Some background with templating systems (Template Toolkit) and various search and retrieve protocols would also be recommended.

References

1.   http://researcher.sfu.ca

2.   Copeland, L., Long, K., Mundle, T. (1999) The COPPUL virtual library resource sharing software. Library Hi Tech 17 (2), 165-71.

3.   Grogg, J. (2006) Link-Resolver Products. Library Technology Reports 42 (1), 24-30.

4.   Mah, C. and Stranack, K. (2005). dbWiz: Open source federated searching. Library Hi Tech 23 (4), 490-503.

5.   Singer, R. (2006) Helping you buy: Link resolver tools. Computers in Libraries 26 (2), 15-23.

6.   Stranack, K. (2006). CUFTS: An open source alternative for searials managment.. Serials Librarian 51 (2), 29-40.

BIOGRAPHY OF AUTHOR

Dr. A.R.D. Prasad Associate Professor Documentation Research & Training Centre Indian Statistical Institute 8th Mile, Mysore Road Bangalore - 560059.

Email : ard@drtc.isibang.ac.in