Providing Integrated Access to Multiple Digital Libraries
Issues and Technologies for the INDEST Consortium

JAGDISH ARORA

National Coordinator, INDEST Consortium &
Librarian, IIT Bombay
j_arora@iitb.ac.in

Introduction

The INDEST Consortium subscribes to the following full-text electronic resources and bibliographic databases f00 Government-funded educational institutions:

Full-text Resources: IEL Online, Elsevier's ScienceDirect, Academic's Ideal Library, Springer Verlag's Link, Applied Science and Technology Plus, ABI / INFORM Complete, ACM Digital Library, ASCE Journals and ASME Journals

Bibliographic Resources: COMPENDEX on EI Village, INSPEC on EI Village, SciFinder Scholar, MathSciNet, Web of Science, J-Gate Custom Content for Consortia (JCCC) and J-Gate.

Generally, users at each subscribing institutions interact with each full-text e-resources independently, learn the syntax supported by them and repeat their searches in other resources. Moreover, once a search is done in different digital repositories, users are left with the task of consolidating their result. One of the major objectives of interoperability amongst digital libraries is to provide an integrated search interface that can seamlessly search across multiple digital repositories. This presentation explores the possibility of providing an integrated search interface that can seamlessly search across multiple digital repositories and consolidate search results in a desired sequence.

Present Cross-searching and Linking Options available to INDEST Consortium Members

The INDEST Consortium members at present have the following three options that offer varying degree of cross-searching and linking facilities:

i) The Web Site of INDEST Consortium and Members

A number of member libraries of the INDEST Consortium have developed their website as a gateway to electronic resources, CD ROM and selected qualitative web resources available in public domain. The INDEST consortium also has website with a back-end searchable database for more than 4000 electronic journals and their URLs. The INDEST web site serves as a tool in its own right restricted to the electronic journals subscribed by the Consortium.

ii) Bibliographic Databases

Besides full-text electronic resources, the INDEST Consortium also subscribes to a number of bibliographic databases for all selected institutions. These bibliographic databases include Compendex and INSPEC on EI Village, Scifinder Scholar, MathsciNet and Web of Science. Most bibliographic databases provide extensive linking facilities to the full-text resources and, therefore, offer an opportunity for integrated access and cross- searching of multiple digital repositories. Moreover, Compendex and INSPEC, being on the same platform, i.e. EI Village, also facilitate simultaneous search in both databases, deletion of duplicate records and user-defined customized output. Bibliographic databases, however, have their limitations in terms

time lags in coverage of resources, time-lag in providing links to full-text resources, subject-specific coverage, etc.

iii) JGATE / JCCC

The J-Gate is an Internet gateway and portal service by Informatics (India) Ltd. that offers article level access to 10,000+ electronic journals. The J-Gate Custom Content for Consortium (JCCC) from Informatics India is a virtual library of journal literature created as customized e-journals access gateway and database solution for the INDEST consortium. It acts as one-point access to 4,000+ journals subscribed currently by all the IITs and IISc through the INDEST Consortium or individually. The JCCC offers table of contents browsing, database searching, MyTOC, full-text online links and resource sharing. The resource sharing and document delivery is a unique benefit that JCCC@INDEST offers.

Standards for Implementing Cross Searching of Digital Repositories

A number of standards and technologies have evolved over the year that enables interoperability and cross-searching of digital repositories. XML, DCOM, JAVA, SGML offer interoperability at the mark-up and computer language level. ODBC and JDBC offer interoperability as standard database interfaces. Web browsers serve as a standard client-end application. DOI, PURL, URN, handles and Open URL provides standardized object naming and object identification mechanism enabling location of digital objects. Z39.50 provides mechanism for cross searching of bibliographic databases. MARC and Dublin Core (Z39.85) are standards for creating metadata for documents, non-documents and for web sites respectively.

Technologies and Solution for Cross Searching of Digital Repositories

There are two methodologies to implement cross-searching of multiple heterogeneous digital repositories, i.e. i) a distributed searching approach that provides direct, real-time access to information sources on the web without resorting to crawling or replicating or harvesting metadata; and ii) metadata harvesting approach wherein the bibliographic data is usually harvested on the side of service provider that is used for building advanced services without relying on real-time interactive access to the remote digital repositories.

A web search engine typically represent a technology based on harvesting and keyword searching. Keyword searching works well for the heterogeneous unstructured data sources. However, it fails to exploit the benefits of structured metadata that most full-text resources and bibliographic databases offer. Open Archives Initiatives (OAI) protocols has adopted metadata harvesting approach that allow metadata to be harvested from OAI-complaint databases which can be collected into a single searchable database. The JCCC and J-Gate also used metadata-harvesting approach for cross searching multiple digital repositories.

While hosting and crawling large quantities of data are expensive, difficult and frequently incomplete, distributed searching approach provides direct, real-time access to digital repositories on the web without crawling or replicating data. A number of commercial products and services are now available that offer cross-searching and linking possibilities based on distributed searching. Some of the important ones are:

- VDX / Agora (www.tdgroupcom) (based on e-Lib Project)
- Metalib and SFX (Ex-Libris, based Open URL Technology)
- WebExpress / iPort (OCLIC)
- Digbib (North Rhine Westphalia and Axion)
- MALIBU
- Endeavour (Elsevier Science)

- Science Server (Elsevier Science)
- Webfeat (www.webfeat.com)
- IWA Portal RevCom (ww.iwapps.com)
- Intelliseek

Most commercial products offer hybrid library functionalities creating possibilities for the librarians to provide services integrated with the traditional library services. Such interfaces start a parallel search for bibliographic databases, full-text resources and library catalogues. The search results are collected, collated and presented to the user in a order customized by the users.

The INDEST Consortium does not want to reinvent the wheel. However, we would like to evaluate all technological options available to us and to adopt the products and services best suited to the members of the INDEST Consortium.