Web Intelligence Implication for Searching Repositories: Federated Search Facilities at LISc

H.C. Nagaratna
JRDTML, IISc, Bangalore, India
hcnc@library.iisc.ernet.in

K.S. Chudamani
JRDTML, IISc, Bangalore, India
ksc@library.iisc.ernet.in

Abstract
Challenge of service coordination in the federated search web is concerned with how to best connect the ultimate service seeker with the ultimate service provider. Web Service matchmaking and resource retrieval are managed in order to provide more power and features in the hands of searchers.

All over the world, open access repositories are being established to archive research and other publications of organizations to make them accessible to users. The tools available are, Dspace, GSDL, Eprints, etc. These enable a user to obtain access to full text articles. Federated searching is enabled by cross matching data and metadata. Z39.50 protocols are used in such searches. Federated search repositories reduce the burden on the search patron by not requiring knowledge of each individual search interface or even knowledge of the existence of the individual data sources being searched.

Introduction
Organizations, professions, individuals information systems are becoming increasingly important in our lives. The process began with personal computers and user software systems. Web is moving from a collection of static documents to a set of web services. The federated search paradigm was created and is evolving in response to the vast number of online databases, repositories and other web resources that now populate what is known as the deep web. Web service technologies are producing a new technology infrastructure in which different web services can be discovered and integrated cooperatively with each other automatically. Additional functionality layers on top of the current web service stack will be implemented.

With the development and maturity of service oriented architectures to support retrieval software agents to federated service transactions organization are using federated services to expose the users functionalities associated with internal systems and retrieval process. Federated service infrastructure activates automatic data integration to enable organization to collaborate and compete in the dynamic global environment. Federated search portals, either commercial or open access, generally search public access bibliographic databases, public access Web-based library catalogues (OPACs), Web-based search engines like Google and/or open-access,.

Federated search portals, either commercial or open access, generally search public access bibliographic databases, public access web-based library catalogues (OPACs), web-based search engines like Google and/or open-access, government-operated or corporate data collection.

Repositories and Search Services
All over the world, open access repositories are being established to archive research and other publications of organizations to make them accessible to users. The tools available are, Dspace, GSDL, Eprints, etc. These enable a user to obtain access to full text articles. Federated searching is enabled by cross matching data and metadata. Z39.50 protocols are used in such searches. One can search by author, title, keyword, etc. Opendoar is one search facility which searches across social science repositories. It uses the Google personalized search interface, which brings in the power of Google search engine. Another type of web-enabled service is union catalogue of journals Nucssi offered by NISCAIR New Delhi. This paper emphasises only on cross searchable repositories.

Federated Searching Activities
According to Peter Jacso (2004) federated searching consists of

1. Transforming a query broadcasting it to a group of disparate databases or other web resources, with the appropriate syntax.
Merging the results collected from the databases,
Presenting them in a succinct and unified format
with minimal duplication, and
Providing a means, performed either automatically
or by the portal user, to sort the merged result set.

Web services in action try to overcome the limitation
of information finding, information extracting,
information representing, information interpreting and
information maintaining. This paper takes a case study
simulating software agents and semantic web services.

Software Agents and Federated Search
Consider the following scenario, regarding e-prints in
IISc. Users go to the web to locate information about
e-books in specific subject. The agent queries the
agents of e-books to find out whether a subject is
available. Semantic Web software agents in the
computer and network will be able to understand the
metadata of all resources in the world including web
pages. The e-repositories guides the federated search
initiatives and research organizations are publishing
most of its research and development.

EPrints is one of the most flexible platforms for building
high quality, high value repositories, recognized as
the easiest and fastest way to set up repositories of
research literature.

In this paper, four important research information
in repositories integration have been dealt with,
namely:

1. E-prints, Open access repository of IISc Research
publication
2. CASSIR Cross Archive Search Services for Indian
Repositories
3. Thesis repositories
4. World repository

These repositories reduce the burden on the search
patron by not requiring knowledge of each individual
search interface or even knowledge of the existence
of the individual data sources being searched.

CASSIR: Cross-Repository Indexing and
Search Service
This service is a part of the ongoing project
“Development of OAI-Based Institutional Research
Repository Services in India”, sponsored by
Department of Scientific & Industrial Research,
Ministry of Science and Technology, Government of
India. This project is being carried out at National
Centre for Science Information (NCSI), Indian Institute
of Science (IISc), Bangalore. The service will harvest
metadata as per the OAI-PMH protocol from the
registered OA repositories in India. It also provides a
web-based search/browse functionality over the
harvested metadata.

Powered by Public Knowledge Project Harvester
Web data designed for metadata describing web
documents is about when and how to present a particular information many attempts have been made to automate and improve the gathering and use of information on the web. Semantic web services automated processing and integration of data and application became easier.

**World Repository**
The world repository is generated using Google based search service. It enables searching across 100 repositories at one go. Google ranking, presentation make the repository most effective.

Its web address is http://library.iisc.ernet.in
User Involvement, Promotion and Awareness
The goal of autonomous software agents is to exhibit communication, cooperation and mobility with their environment. The use of software agents is an integral part of the web and the use of software agents capitalizes on federated search web technologies. Software agent and web technologies include data federation, metadata, data markup and ontologism mechanism for finding and connecting federation search resources, usability issues, information retrieval and organization. Our aim is to bring together researchers who bridge one or more of these technologies in discovering critical information in large and changing data and in enabling the web to empower these technologies.

The organization that knows where to find best information knows it and can utilize it. The quickest wins. Repositories can bring structure to information chaos

1. Our information with machine understandable markup
2. Know what information is authoritative
3. Proof that we can have added trust the information we find
4. Correlates on information with other information that we have
5. Tools to take advantages of repositories only when you bring the information together with web will this information lead to knowledge that enables staff to make well informed decisions

Conclusion
Intelligent search instead of matching keywords or answering queries instead of information retrieval and document exchange between departments via ontology translations, definition of views on documents enables personalisation and contextualisation of information.

References
2. http://etd.ncsi.iisc.ernet.in/
3. http://library.iisc.ernet.in/