ANALYSIS AND CHARACTERISTIC OF INFORMATION RETRIEVAL IN DISTRIBUTED RESOURCES

M. Krishanamurthy

Documentation Research and Training Centre (DRTC)
Indian Statistical Institute, 8th Mile Mysore Road, R.V.College Post,
Bangalore-560059.

Corresponding Author's E-mail: mkrishna_murthy@hotmail.com;Krishna@isibang.ac.in

Abstract— This paper examines the meaning of context in relation to metasearch base query expansion. The global network is the transforming reality of the information environment. Fist of all, it is about distributed information resources and distributing computing. Distributed computing has elicited the trend towards higher level of abstraction and more formalysed ontological support for information systems. This paper discuses the importance of metasearch and metadata on the web a common need for high quality information in the Internet.

1 INTRODUCTION

Today there's a great amount of knowledge and information widely available in multiple formats: printed books and journals, video and audiocassettes, and electronic documents through the Internet. The great challenge to access that knowledge in a fast and efficient way, and then obtain relevant information that could be applied to the solution problems. Including the database research community has stated the following goal '"***The information Utility'*** make it easy for every one to store, organize, access, and analyze the majority of human information Online"(Wecel 2003).

However, there are some problems associated with this vision. The first one is the unstructured nature of the information available in the web, therefore making the retrieval of information from a search in most of the cases with a large amount of entries that are not relevant to our query. At present, the mainstream approach for locating web resources is through key word based search engines relying on information retrieval techniques. Although search engines are certainly useful, they present fundamental drawbacks such as lack of information-space organization. To effectively exploit the web's expanding information sources, the emerging semantic web efforts employ machine understandable abstractions for the representations of resource semantics. In particular, the semantic web promotes the use of ontologies as a tool for reconciling semantic heterogeneity between

web resources (Benetallah *et al*. 2006) .

2. SEARCH FOR INFORMATION

In the Internet, search engines are especially popular. These maintain lists of millions of search term, and associated links to relevant Internet pages, in one ore more databases. Users can find the information they require by using appropriate expressions or else with the help of categorized listing. A search is often based on a simple text and does not support any structured refinements. Meta Search engines often have no database of their own, relying instead on various other search engines to find information and collecting together results. Thus, many suggestions from groups and conferences with an interest in Information Retrieval are concerned with the need to find, and to define, a standard organizational structural for data. Similar activities are also being undertaken individual enterprises. Organizations are making data warehouse available, in which all the important data that accrue with in the organization can be found. RDF and other new description facilities for describing knowledge pieces constitute a newly emerging standard for metadata that is about to turn the world wide web into a machine understandable knowledge base, called semantic web. Thus, it might turn up as a natural choice for a widely useable ontology description language.

3. META SEARCH (ACROSS DATABASE)

The key feature of a library portal is to allow searching across multiple databases without having to repeat a search. This feature is generally referred to as meta search, parallel search, broadcast search or federated search (Fyer 2004). Accessing a resource for the sake of querying it or obtaining results from it requires prior knowledge of the resources, because each resources has its own distinct structures and rules for the interpretation of its data. It is not only the structure of the data at the target end that is required for the delivery of queries and the retrieval of results. The local integration and linking of these resources a high level of customization is required. To address these issues, libraries and information providers have developed mechanism to assist user with database selection, search strategy formulation, and search navigation. One useful mechanism for addressing the problems of selecting from multiple relevant databases is to provide the capability of simultaneously searching multiple selected databases and running individual search result for user examination. A model of this dynamic meta search engine has been demonstrated successfully as the library portal such as “Journal finder” and “Full text Article Locator” at the Grainger Engineering Library of the University of Illinois at Urbana Champaign (Mischo and Schlembach 2001). Based on the approach to federated search, this paper deals a model of a library portal customized for searching across the multiple online resources at the Library of the Indian Statistical Institute Bangalore.

4. RESOURCE DESCRIPTION FORMAT (RDF)

Resource Description Framework (RDF) is an infrastructure that enables the encoding, exchange and reuse of structured metadata. This infrastructure enables metadata interoperability through the design of mechanisms that support common conventions of semantics, syntax, and structure. The Resource Description Framework provides a means for adding semantics to a document without making any assumptions about the structure of the document. It is an XML application customized for adding Meta information to web documents. This project was started in 1995, to filter the unwanted WebPages. It is w3C supported semantic markup scheme. The data model of RDF provides three object types: resources, property types, and statements (Brickley *et al*. 1980).

A resource is an entity that can be referred to by a address at the WWW (URL)

A property defines binary relation between resources and/ values provided by primitive data type definitions in XML.

A statement specifies for a resource a value for property. That is, statements provide the actual characterization of the web documents.

An example is

Author (http://www.isi.bang.ac.in/krishna)=KRISHNA

This states that author of the named web document is Krishna. Values can be structured entities:

Author (http://www.isi.bang.ac.in/Krishna)=X

Name (X) =krishna

Email (X)=Krishna@isibang.ac.in

Where X denotes an actual (i.e., the homepage of krishna) or a virtual URL. In addition, RDF provides bags, sequence, and alternatives to express collection of web resources.

4.1 Characteristics of RDF

RDF is designed to support semantic modularity by creating an infrastructure that supports the combination of distributes attribute registries. This permits communities to declare own vocabularies that may reused, extended or refined t address application or domain specific requirements. RDF can be used in a variety of application areas;

Resource discovery to provide better search engine capabilities

Cataloging for describing the content and content relationships available at a particular web site, page

Knowledge sharing and exchange by intelligent software agents

Content rating

describing collection of pages that represent a single logical document

describing intellectual property rights of webpages

expressing the privacy preferences of a user as well as the privacy policies of a website.

4.2. Formal Data Model

The data model of RDF provides three object types: resources, property types, and statements.

A resource is an entity that can be referred to by a address at the WWW (URL)

A property defines binary relation between resources and/ values provided by primitive data type definitions in XML.

A statement specifies for a resource a value for property. That is, statements provide the actual characterization of the web documents.

An example is

Author (http://www.isi.bang.ac.in/krishna)=Krishna

This states that author of the named web document is Krishna. Values can be structured entities:

Author (http://www.isi.bang.ac.in/krishna)=X

Name (X) =Krishna

Email (X)=krishna@isibang.ac.in

The typical example of this model represented as follows;

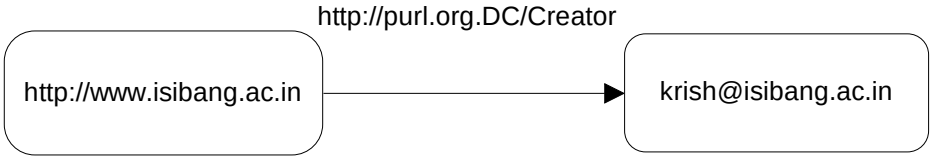

Figure.1 Basic RDF assignment

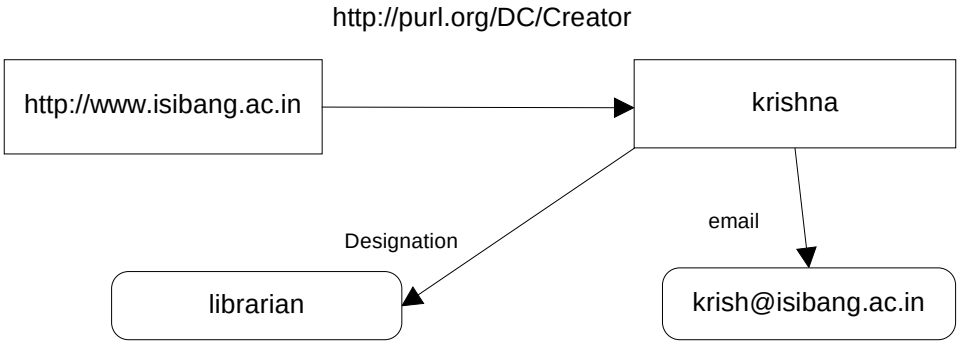

Fig.2 Extended RDF assignment

4.3 RDF Syntax

RDF defines simple, powerful model for describing resources. Syntax representing this model is required to store instances of this model in to machine process able files and to communicate these instances among applications. Creators of RDF decided to use XML syntax which in fact became a standard. The Syntax description presented below.

[1]RDF::=['rdf:RDF>']description*[',rdf:RDF>']
[2]descritpion::= '<rdf:Description'idAboutAttr?'>'
propertyElt* '<rdf:Descriptin>''
[3]idAboutAttr::=idAttr about Attr
[4]aboutAttr::='about='"URI=reference
[5]idAttr::='ID='"URI-reference"" idAttr::=ID='" IDsymbol ""
[6]propertyElt::='<'propName'>'value " '<'propName'>
'<' propName::= resourceAttr'/>
[7]propName::= Qname
[8]value::=description string
[9]resourceAttr::='resource='"'URI-reference'""'
[10]Qname::=[Nsprefix ':']name
[11]URI-reference::=string,interpreted per[URI]
[12]IDsymbol::=(any legal XML name symbol)
[13]name::=(any legal XML name symbol)
[14]Nsprefix::=(any legal XML namespace prefix)
[15]string::= (any XML text, with"<",">", and "&" escaped)

4.4 Expressing RDF in XMl

RDF requires the XML namespace facility to precisely associate each property with the schema that defines property. RDF/XML grammar representation as follows,

<?xml version="1.0"?>
<rdf: RDF
xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#
<rdf: Description about=http://www.isibang.ac.in>
<Creator>Krishna</Creator>
</rdf:Description>
</rdf:RDF>
Similarly the Complex RDF model represented as,
<?xml version="1.0"?>
<rdf:RDF xmlns=http://www.w3.org/1999/02/22-rdf-syntax-ns#>
<rdf:Description about=http://www.isibang.ac.in">

<Creator rdf:resource=”Krishna”/>
<rdf:Description>
<rdf:Description about=”Krishna”>
<email>krish@isibang.ac.in</email>
<Designation>Librarian</Designation>
<rdf:Description>
</rdf:RDF>

4.5. Ontology Mapping

Ontology mapping is the one of the most important issue. On the one hand it is the most difficult and time-consuming task in ontology engineering, and on the other hand it is crucial for reasoning capabilities. In order for ontologies to have the maximum impact, they need to be widely shared. The following ontology can be found at http://library.isibang.ac.in/library~krishna/
Ontology/onto#

<?xml version =’1.0’encoding=’UTF-8’?>
<rdf:RDF
Xmlns: rdf=http://www.w3.org/2008/01/12-rdf-syntax-ns#
Xmlns:rdfs=http://www.w3.org/TR/2008/PR-rdf-schema-20080112#
Xmlns:dc=”http://purl.oclc.org/dc#”>
<rdf:Description about=””>
<dc:Title? Library Portal </dc:Title>
<dc:creator>
<rdf:Bag>
<rdf:li>Krishna</rdf:li>

4.6 Link Analysis

Link analysis is another approach to modeling the context of an information space, which could be a web page and its different types of hyperlinks to databases, homepages and user groups. The content and structure of the information space surrounding documents is an important contextualising factor, which can be used to improve the accuracy of relevance rankings assigned to documents (Lowe 2000). Web pages can be amended by adding multi-destination links to them. The user interests define which links are to be exported for that user and documents content defines which of these links are to be rendered in that document.

5. RETRIEVING INFORMATION FROM A DATABASE

Using a library catalog to find the little piece of books/documents can be a good example of database access process and database systems can be applied information filtering

process. Information extraction process is similar to database access for providing information to the user. It is differentiated from the database access process by the nature of the resources from the information is being obtained. In database access process information is obtained from some type of database. While in information extraction the information is less well structured. A web portal technology is a time consuming, frustrating, and a very complex process for end users. Currently, users need to access a large number of heterogeneous and distributed data sources in order to become self-supported. For example, in Indian Statistical Institute Library users need access to several databases and portals of different sites separately, manually filter and organize the search results to get the information about the benefits they are entitled to. Clearly to effectively realize the potential of web based information access, there is a need for facilitating the integrated access to relevant information sources.

5.1 Semantic Query Processing

Traditional searching often produces a set of results which is either too small or too large Semantic searching system can expand the query to reduce the result set to acceptable size, making decisions about documents. The new information portal provides a gateway to the entire spectrum of library journal resources. However, it also allows end-users to choose among them as desired for more efficient search with semantic strategy. Fig. 1 shows the newly designed ISIB Library search aid as a gateway to its journal resources. When the user clicks on the full text link, the metadata elements contained in the link are used to perform a search in the metadata database.

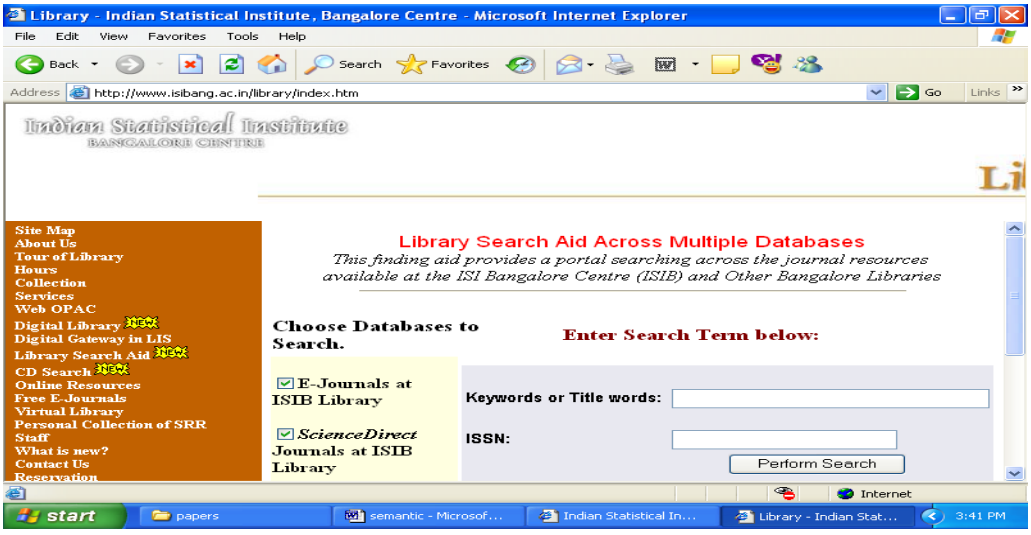

Fig. 1. Web interface: search form

The search term entered, computing, retrieves a number of matches simultaneously from a pre-selected list of information sources, i.e., in this case all are selected: Journals at ISIB Libraries. Users may proceed to choose from here with one single click to locate more details about the journals of interest.

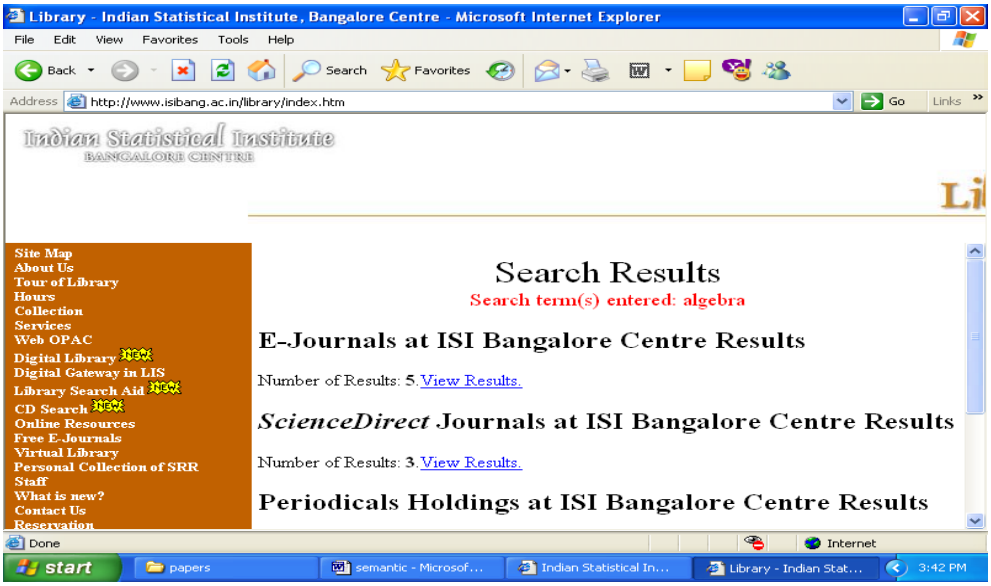

Fig.2. Web Interface: Full display

5.2 Retrieval Mechanism

The advantage of simultaneous searching across multiple sources of library information is evident. For users, the new library portal design presents a gateway to navigate fruitfully from a simple step rather than making a series of search attempts indiscriminately across the previous Web lists, The biggest improvement of this approach is exemplified particularly in searching the database, as it eliminates the need to make multiple searches one year at a time as required previously.

6. CONCLUSION

In this paper it was shown that distributed electronic resources search through Meta search have a clear potential for more efficient. IR; in particular that networking topology plays an important role in retrieval effectiveness. This was investigated within the metasearch engine context, but in fact, IR can benefit from Meta search solutions in a much more

direct and whole way. Ontologies play a substantial role in knowledge based information retrieval and filtering. It provides richer integration and interoperability of data and permits the development of applications that search across diverse communities or merge information form them. In this paper we discussed advantages of introduction of ontology into web-based application and may also bring new uses of the web. Judging from current research discussion, the near future works greater interoperability, share ability and reusability among web application. In all the approaches it should be cover to add new facts (and rules) in simple way.

REFERENCES

1. Berners-Lee,T et al. 200, The Semantic Web, Scientific American, available at: www.sciam.com/article.cfm?articleID=00048144-10D2-1C70
2. Benatallah, B. et al. 2006, Towards semantic-driven, flexible and scalable framework for peering and querying e-catalog communities. Information Systems, 31 p 266-294
3. Bhogal J et al . 2007, A review of ontology based query expansion. Information Processing and Management 43, 866-886
4. Brickley, D., et al. 1998, Resource Description Frame work (RDF) Schema specification. W3C working Draft, available at http://www.w3c.org/TR/WD-rdf-schema
5. Dublin Core metadata initiatives 1999. http://purl.oclc.org/metadata/dublin-core
6. Ding, Y.,2001, Ontology: the enabler for the semantic web, Journal of Information Science, Vol.27 (6),
7. Fyer , D. 2004, Federated search Engines. Online, 28, 16-19.
8. Gordon ,W.,Paynter., and Wittenn,I H. 2001, A Combined phrase and thesaurus browser for large document collections. In 5th European conference, ECDl.(ECDL2001),(Darmstadt)
9. http://dublincore.org
10. http://www.w3.org/RDF/
11. Kifer, M et al.1995, Logical foundations of object-oriented and frame-based languages, Journal of the ACM, 42.
12. Lowe, D., 2000, Improving web search relevance: using navigational structures to provide search context. Available at http://auswb.scu.edu.au/aw2k/papers/lowe/paper.html.
13. Mischo, W.H., 2001, Library Portals, simultaneous search, and full-text linking technologies. Science and Technology Libraries, 20, 133-147.
14. Wecel. , K.2003,Towards an ontological representation of knowledge on the web, Knowledge based information retrieval and filtering from the web (Newyork:Kluwer academic press).
15. W3C Semantic Web. http://www.w3.org/2001/sw/