



Interoperability in Semantic Digital Library

N.Ramya, Shiva Amrit A

B Tech (3rd Year), Department of Information and Technology,
SRM University, Kattankulathur.

B Tech (3rd Year), Department of Computer Science Engineering,
SRM University, Kattankulathur.

nramya.13@gmail.com, shivaamrit@gmail.com

ABSTRACT— *There is a growing need to digitize and automate the available data in the world and make it available through a digital library. But, the conventional methods of storage and retrieval of data in the form of logical data structure of a database management (DBMS) cannot totally satisfy the requirements for a conceptual definition of data, because it is limited in scope and biased toward the implementation strategy employed by the DBMS. Therefore, the need to define data from a conceptual view has led to the development of semantic data modelling techniques. That is, techniques to define the meaning of data within the context of its interrelationships with other data. The current model of Digital Libraries only allows users to query each digital library system separately. It is therefore desirable to provide a mechanism to foster interoperability between digital libraries. This mechanism would allow users to conduct searches across all systems with just one query. The paper explains issues regarding interoperability of digital libraries using semantic approach utilizing metadata and also proposes a model to achieve ontology based interoperability across digital libraries.*

Keywords— **digital library, interoperability, semantic, ontology, metadata.**

1, INTRODUCTION

Due to the rapid development in Internet technology, traditional libraries are in a stage of transition towards making the library globally accessible on the Internet. Many information storage and retrieval systems were being used for a meaningful retrieval in print-based libraries. In a Web environment, the traditional means and techniques for information storage and retrieval are required to be modified to suite the changing needs. Therefore due to the advent of technology and availability of system development resources there are



many accessible digital libraries found on the Internet. These digital libraries have well researched user interfaces, architectures that allow ease of use and permit various levels of interactivity including searching and browsing. They are aimed to help users to retrieve useful information easily and quickly.

A digital library is a place where various forms of information material such as journals, monograph, visual materials, voice recorder and moving pictures can be retrieve via the Internet. The classification systems for books has been changed into *Ontologies* to represent domain knowledge in machine readable form, and the cataloguing codes have taken shape of *Metadata Schema* for the description web resources.

However, despite the advantages and existence of various digital libraries, the concept of global information sharing is still much less than expected. This is due to difficulty in obtaining materials concurrently through a single portal or a single entry point, in which to obtain information material stored in every digital library, it would require accessing each digital library separately.

Therefore, it is desirable to provide a mechanism to allow interoperability between digital libraries where query searching for information material may be conducted one single time; without the difficulty of going through each digital library. This leads to the motivation of the work performed.

2, DIGITAL LIBRARIES INTEROPERABILITY – ISSUES

Interoperability is the ability of two or more information systems to communicate, execute instructions, share data or otherwise interact, which is fundamental in networked environment. As such interoperability can be achieved through semantic, syntactic and physical layers.

Semantic interoperability is the knowledge-level interoperability with the ability to bridge semantic conflicts arising from differences in implicit meanings, perspectives and assumptions thus creating semantically compatible information environment based on the agreed concepts between different entities.

Syntactic interoperability on the other hand is the application-level interoperability that allows multiple software components such as implementation languages, interfaces, and execution platforms. While physical interoperability is known as the communication-level interoperability, which focuses on infrastructure, network protocol or other applications.

3, ISSUES IN DIGITAL LIBRARY SEMANTIC INTEROPERABILITY

Semantic interoperability entails sharing and reusing metadata in environment characterized by vastly diverse sources of content, styles of content management and approaches to resource description [6].



As such, in Digital Library it is achievable through the availability of metadata standards such as IEEE LOM[1], Dublin Core[2], MICI[3], MARC[4], EAD[5] and TEI[6]. These metadata standards provide the descriptive information for the creation, organisation and administration of metadata for digital collection accessibility and retrievability.

As each digital library has numerous information of collection like books, journals, manuscripts, thesis and articles, each collection is looked thoroughly in terms of its metadata. This is essential as each digital library has its own unique sets of metadata description design to define attributes associated with an information collection.

Metadata is the core of any information retrieval system and so its implications for any digital library are profound: the choice of a metadata scheme underpins any such library's ability to deliver objects in a meaningful way, and greatly affects its long-term ability to maintain and preserve its digital assets.

4, METHODOLOGY

Based on analysis findings in Table 1, it was uncovered that the metadata elements identified were stored in every collection in digital library. These data are essential as it is important to identify the use of metadata elements and metadata standards that exist in every Digital Library. This is necessary in order to achieve semantic interoperability based on interoperability methodology and standardization. The methodology requires consistent metadata standards, common use of metadata elements and established mapping technique for interoperability in digital libraries.

Metadata elements are used in every existing digital library. As such, metadata element is an important asset and the rationale for the massive use of this element is because it has a number of benefits in terms of lower costs, greater supply, indexing, higher accessibility and availability.

Based on the details obtained from the Table1, it can be concluded that various standards and elements of metadata are being used in different digital libraries. Common standards being used in the digital libraries among others are Dublin Core, MARC, IEEE LOM, TEI to name a few. Due to the variation of standards and elements used, they result in the hindrance of interoperability in digital libraries as inconsistent metadata elements are utilized to describe the information collection in each digital library.

Therefore, it is difficult to map or find consensus in regular metadata element as each digital library used different types of standards and element. Due to this predicament, semantic interoperability needs consistent metadata elements and standards.



5, CONTEXT-BASED APPROACH

The vision of semantic retrieval can be accomplished with an ontological model of user's interest areas, and the modeling of context of information which has been dealt in the document. This will make the match of user's context with the document context easier at the search stage.

Due to the distributed and heterogeneous nature of the Web, interoperability at semantic level became a great challenge for the system developer. Hence the present study aims at addressing both the challenges using tools and technology available to implement Semantic Web based services. The present work intends to bring semantic retrieval in digital libraries using semantic web technologies for the key tasks like ontology-based annotation

In proposed system a novel approach, namely Transpose-Minify software framework is use to effectively manage the large data sets. The main feature of this model is simplicity, fault tolerance, and scalability. In this model the computing and data resources are co-located, thus minimizing the communication cost and benefiting the service providers.

[12] TABLE 1
 META DATA ELEMETS

Organisation	Collection	Metadata Elements	Retrieve By
Public Higher Institution			
National University of Malaysia	Journal, Thesis, Book, Manuscript, Examination Paper	<i>Keyword, author, title, subject ISBN,</i>	Free for member
Malaya University	Journal, Thesis, Book, Manuscript, Examination Paper	<i>Keyword, title, publisher</i>	Free for member
International Private Higher Institution			
Bond University	Journal, Thesis, Book, Examination Paper	<i>Keyword, title, author, subject</i>	Free for member
Boston University	Journal, Thesis, Book, Examination Paper	<i>Keyword, Title, author, subject</i>	Free for member

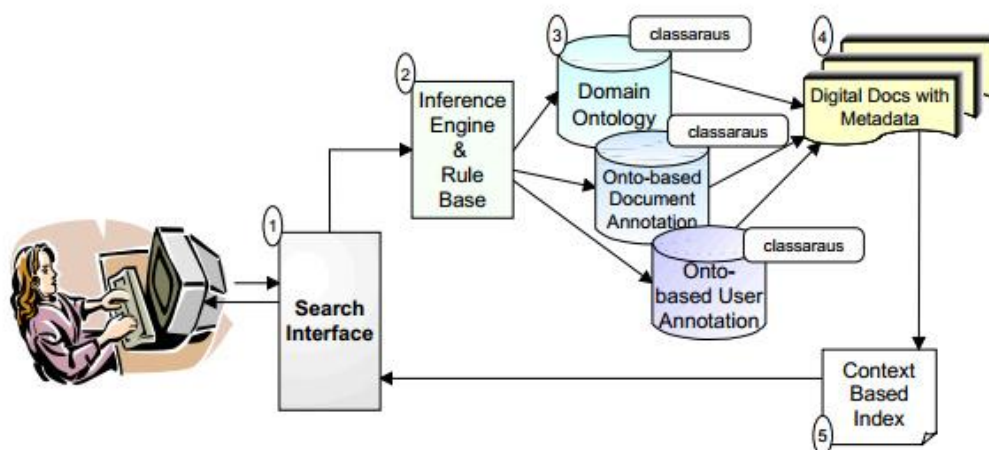


6, CONTEXT-BASED SERVICE MODEL

6.1 Semantic Annotation: In the given model, the ontology has not been only for the domain knowledge modeling, but also for document annotation and the annotation of the registered users of the repository. Context of the information been dealt in the document can be made explicit by document annotation, which can be represented in Web Service Modeling Ontology (WSMO). Similarly the user profiles can be represented as ontology-based annotation.

6.2 Classaurus: It is a faceted thesaurus used as a vocabulary controlled mechanism for automated permuted indexing in traditional bibliographic databases.

Figure 1
Context Based Approach



7, METADATA INTEROPERABILITY FOR SEMANTIC RETRIEVAL

There is wide variety of metadata schemas available for different kind of digital resources

- Content independent metadata
- Content dependent metadata
 - o Direct content-based metadata
 - o Content descriptive metadata
 - Domain independent metadata
 - Domain specific metadata



This categorization makes it quite clear that to bring semantic interoperability among bibliographic repositories, various types of metadata schemas has to be considered. JeromeDL [4] project has made an effort to bring semantic interoperability among the digital repositories using different bibliographic metadata standards like Dublin Core, BibTeX, MARC21 etc. The WSMO framework is a well define model to bring semantic interoperability among 2 heterogeneous automated retrieval systems by it's the key components like Web Service Modeling Ontology (WSMO) [5], Web Services Modeling Environment (WSMX) [6] including various mediators.

In the model of virtual document delivery (Figure 1), the key tasks are: service discovery, ontology mediation and context matching. The semantic digital library searches for the service registry [7, 8] to locate appropriate repositories. After the service discovery to map the ontology and metadata schema, ontology mediators are used. The ontology mediator converts the mapped ontology to the native syntax. The retrieved document description is used to generate context- based index. This context-based index is then matched with the user query and annotation.

To achieve semantic interoperability for digital libraries, the work performed provides the mapping technique and reuses it in developing semantic interoperability framework. As these elements located in different digital libraries, there rises a need to extract the elements into one central repository.

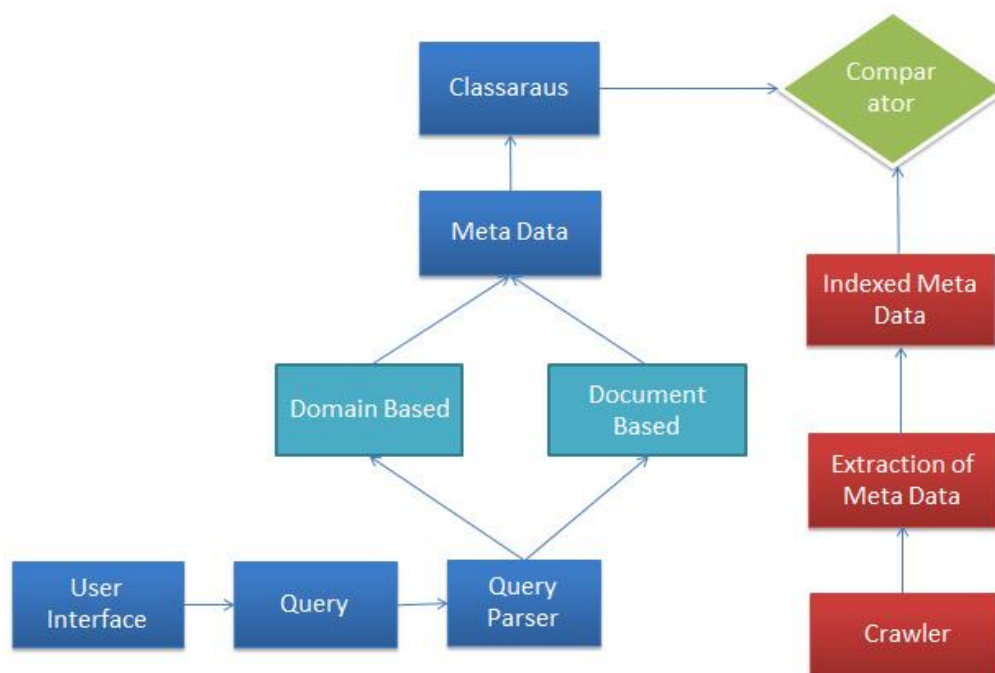
Therefore an extraction process is performed using a tool to extricate the metadata elements into the central repository. It is essential to have a central repository as this would aid in providing a one-stop search facility for all the digital libraries. As such through the use of this facility, semantic interoperability using metadata is made achievable since the metadata extracted can be interpreted and interconvert to locate the collection in the Digital Libraries.

8, PROPOSED MODEL

The interface in Figure 2 is an interactive console for the query entry. The query is then parsed into domain and textual detail based on the ontology. The metadata is extracted from this representation. The claussaurus is then used to derive further information from the metadata. This set is then fed to the comparator. Meanwhile the crawlers would have taken a reference of the dspaces and made indexes of the metadata. The user query which was parsed and refined using the claussaurus is used as a base to search the repository now to yield a result.



Figure 2
PROPOSED MODEL



CONCLUSION AND FUTURE WORK

Interoperability in a heterogeneous environment is a broad connotation encompassing syntactic, as well as semantic interoperability. Interoperability at semantic level is a challenging task. In context-sensitive query processing over heterogeneous information resources requires the matching of concepts. Vocabularies, semantic relationships and mappings are information objects themselves, their life cycle: creation, acquisition, collection, modeling, identification, integration, mediation, search, use, maintenance and preservation etc. is of primary importance and a necessary prerequisite to improved semantic interoperability [9]. Steps are to be taken in all future researches in this regard.



REFERENCES

- [1] C. Patra, "Developing a Digital Library on ceramics," *Emerald Group Publishing Limited*, vol. 24, pp. 453-469, 2006.
- [2] A. Shiri, "Digital library research:current developments and trends," *Emerald Group Publishing Limited*, vol. 52, pp. 198-202, 2003.
- [3] J. Salim, Y. Yahya, N. R. M. Rashid, S. Othman, and M. Mohd, "Diglib: Gateway to Digital Libraries and Searchable Databases," *The International Journal of Learning*, vol. 14, pp. 7-15, 2008.
- [4] S. A. Baird, "Government Role in Developing an Interoperability Ecosystem," in *ACM International Conference Proceeding, Macao*.
- [5] W. E. Moen, "Mapping the Interoperability Landscape for networked information retrieval," in *International Conference on Digital Libraries*, Roanoke, Virginia, United State, 2001, pp. 50-51.
- [6] E. Duval, E. Forte, K. Caedinaels, B. Verhoeven, R. V. Durm, k. Hendrikx, M. W. Forte, N. Ebel, M. Macowicz, K. Warkentyne, and F. Haenni, "The Aiadne Knowledge Pool System," *Journal of Communication of the ACM*, vol. 44, pp. 72-78, 2001.
- [7] A. Leroy, "J2EE and .NET: Interoperability with Webservices " in *Scientific Engineering of Distributed Java Applications*. vol. 2952: Springer Berlin / Heidelberg, 2004, p. 155.
- [8] M. Hatala, G. Richards, T. Eap, and J. Willms, "The Interoperability of learning object repositories and services: standards, implimentations and lessons learned," in *13th International World Wide Web Conference*, New York, USA, 2004, pp. 19-27.
- [9] L. Davis, D. Flagg, R. Gamble, and C. Karatas, "Classifying Interoperability Conflicts," in *COTS-Based Software Systems*. vol. 2580: Springer Berlin / Heidelberg, 2003, pp. 62-71.
- [10] H. Salah, B. Mahmoud, and B. Nacer, "An Architecture for the interoperability of workflow models," in *Proceedings of the first international workshop on interoperability of heterogeneous information system*, 2005, pp. 31-38.
- [11] J. Park and S. Ram, "Information Systems Interoperability: what lies beneath?," *ACM Transactions on Information Systems (TOIS)*, vol. 22, pp. 595-632, 2004.
- [12] S. Ram and V. Ramesh, "Schema Integration: Past, Present, and Future," in *Management of Heterogeneous and Autonomous Database Systems*, A. K. Elmagarmid, Marek Rusinkiewicz, and A. Sheth, Eds.: Morgan Kaufmann, 1999, pp. 119-130.



BIOGRAPHY



N. Ramya, is currently perusing her 3rd year in B Tech - Information and Technology in SRM University.



Shiva Amrit A is currently persuing his 3rd Year in BTech – Computer Science and Engineering in SRM University. He is involved in a project on Securing adhoc on Demand Distance Vector protocol funded by Defence Research and Development Organization(DRDO) at SRM University.