Re-defining Resource Management using Semantic Technologies

Devika P Madalli

Libraries and library professionals have tried to keep up with technological advances that impact information services. With the increase of resources on the web, librarians are exploring methods of integrating bibliographic resources found when searching or browsing the web. For instance there are currently some tools like citation managers, that librarians can collect some bibliographic data and can integrate it into in-house systems. So far libraries used technology but often those that needed human intervention to interpret. This paper aims discuss various a Semantic Web(SW) technologies that can be exploited for better and efficient management of resources enabling meaning representation so that data can be manipulated 'with meaning and context' by agents. SWTs offer a wide range of possible applications to meaningfully managing information resources. A typical ordering scenario is used to demonstrate the deployment of SW techniques.

Keywords: Metadata, Semantic Web Technologies, Resource Management

1. Introduction

Libraries in a way were key players in the process of democratization of knowledge. Libraries processed and stored data and enabled information flow that enabled exchange of thought and hence enhancement of knowledge. The techniques and methods adopted enroute were traditional but also sophisticated at times. They tried to provide for search or exchange of bibliographic information using various formats and protocols. The present day techniques like online public access catalogues (OPAC), integrated library portals etc. really helped the libraries reach their clientele. Apart from human mediated services, there were some attempts to increase machine to machine communication like use of the popular Z39.50 protocol to map bibliographic data elements.

This paper aims to discuss various a Semantic Web(SW) technologies that can be exploited for better and efficient management of resources. These offer a wide range of possible applications to meaningfully manage information resources. A typical ordering scenario is used here to demonstrate the deployment of SW techniques.

2. What is IRM?

IRM is interpreted as a process, technique and methodology. For the purpose of the present discussion, it is important to define IRM and examine its scope for the proper understanding of the activities involved. IRM is the management (planning, organization, operations and control) of the resources (human and physical) concerned with the systems support (development, enhancement and maintenance) and the servicing (processing, transformation, distribution, storage and retrieval) of information (data, text, voice, image) for an enterprise [1]. Burk and Horton [2] improved this definition by considering IRM as a managerial link that connects corporate information resources with the organization's goals and objectives. In another work Horton [3] offers a more comprehensive definition: IRM is "a managerial discipline which views information as a resource equal to financial, physical, human, and natural resources. IRM addresses the efficient and effective handling of information resources (raw data) and the resulting information assets (knowledge)".

Information resources should be considered as a strategic asset of an organization that should be regularly assessed and audited just as any other tangible asset of an organization such as property and office equipment. Information resources are engines driving the information economy, and in turn is having profound impact on business management, competitive advantage, and productivity [4]. This statement emphasizes the need for some key processes of IRM, which includes information audit, information mapping etc.

3. Libraries and Information Centres and Resource Management Technologies

Tracing the developments that lead to deployment of IRM techniques in libraries, it is observed that the growth is not strictly an evolutionary chain as most of these developments are not mutually exclusive and most of the technologies deployed co-exist. An attempt is made to trace the development of Resource Management Technologies in Library and Information Centres, which can be divided as following sub topics.

3.1 Stand Alone Libraries and Catalogues

In the traditional library, classification and cataloguing are used for better information resource management. It helps in enhancing the discovery, management, accessibility and availability of the holdings of the library. Most of the earlier libraries used to provided the catalogue records only in card formats and a few libraries afford to provide records in machine readable formats (MARC etc) and also as Online Public Access Catalogues (OPACs).

3.2 Library Networks and Union Catalogues

Due to financial pressure resulting from the increasing cost of information resources, libraries in related areas started collaborating with each other to serve their peripheral interest and tried to concentrate on their core acquisition list. The collaboration went beyond just sharing the collections to sharing other resources such as manpower, storage space, technical work, etc. Later library and information center networks emerged when more institutions joined to share their Information Resources. A number of resource sharing networks has been observed at local, regional, national and even at international levels such as OCLC, ILRS, DELENET, INFLIBNET, ENVIS, FOSTIS to name a few. Most of these networks relied on some inventories like Union Catalogues to facilitate the inter library resource discovery and sharing. IRM then was defined in terms of management of information resources in networked environment. Some of the union catalogues were available in both Print and electronic form. The largest print union catalog ever published is the American National Union Catalog Pre-1956 Imprints (NUC), which has been superseded by the creation of large electronic resource sharing networks, such as OCLC's WorldCat.

3.3 Web and the OPACs

When the Web became popular, it made IRM relatively cost effective and also complex due to the abundance of information available in disparate forms. Most libraries, big or small, except in a few sectors, now have some kind of access to the Internet. Most of the Library Management Systems including Open Source tools, provide webOPAC functionality, which exposed library and information centre holdings beyond the borders of Networks and to 'netwide' access.

3.4 Digital Libraries and Metadata

By late 1990s Digital Libraries started emerging as solution to seamless Information Resource Transfer as the libraries started large scale digitization projects and made it available in web using Digital Library tools such as E-prints, Greenstone, DSpace etc. In Digital Environment 'Metadata' sort of substituted 'catalogues' with a slight expansion/change of the scope of the word.

The syntax and semantics of how information is presented in a resource has been recorded in the metadata. An information resource has essentially three features, 'content, context and structure', which metadata stores. Metadata aids the identification, discovery, administration, and tracking of use of a particular resource. Metadata encapsulates the structure and organisation of objects, and also enables interoperability between different systems.

The well cited 'Willard Model' [5] of IRM identifies five key elements viz Identification, Ownership, Cost and Value, Development, Exploitation. A review of relation between Metadata and Resource Management, inspired by Williard Model is given below:

- **Identification:** It enhances resource discovery and recording essential features in an inventory. To identify the resource by means of a string or number conforming to a formal identification system.[6] An identifier embodies the information required to distinguish what is being identified from all other things within its scope of identification.

Eg: Metadata stores the ISBN (International Standard Book Number) . A different ISBN is allocated to each manifestation of a title (Paperback , Hardback, CD – ROM etc.). So, a user with a specific ISBN can find the resource efficiently.

➲ **Ownership:** Ownership establishes the responsibility for maintenance of resources Eg: If a Digital Library is holding a resource, the library is also responsible for exposing information regarding applicable rights and permissions to resources.

➲ **Authenticity and Quality:** The record of provenance of a resource is an evidence of its authenticity and quality. Eg: The work flow of a digital library, evaluation of the resources to be included is an important step. There could be some automatic heuristic based approach for determining the source/ host of a web resource but often enough there is not enough clue. Therefore this step may have to be manually applied using some world accepted checklists of evaluation of e-resources to establish authenticity and quality. Workflows often include peer review process

➲ **Versioning:** The record of lifecycle of a resources; when it was created, modified, used and disposed of, is necessary for resource management. This documents the processes that have taken place during the life cycle of an information resource and handles the transactions.[7] Versioning is an important thing in a collaborative environment, tracking of different versions of a resource helps to keep one updated of the changes and hence facilitates better management. The version history of a resource may also act as an important part of publishing a resource , the latest and completed items can only be published. For Eg. In Wikepedia, one can access the older versions of the same article.

➲ **Preservation of Resources:** Archiving of information resources is necessary to ensure long-term accessibility of the preserved information resources. The technologies used and procedures followed should be recorded for future use. For instance: "Digitisation" can be used as effective technology for preservation of resources which further leads to better IRM. It is also necessary to record information such as the technology used for digitisation, the format in which the resources are stored. This information is helpful to retrieve the data from obsolete software, migration to a new system etc.

4. Semantic Web and Semantic Metadata

Conventional metadata lacks facility to achieve 'context based' resource management. It captures the structure but not really in a semantic way that should actually point to content and its context. The conceptual structure of the content is essential to retrieve the contextual information. Though the classic application of metadata has been for information search and retrieval, there are several other functions of metadata as enlisted below:

1. Resource description
2. Interoperability
3. Enabling Access
4. Provenance and authenticity
5. Preservation
6. Context Correlation
7. Enabling reuse

Apart from these uses, new schemas also are used for describing structures, personalization, results from monitoring and tracking, processes and interactions that can enrich the learning environment in various ways.[8]

4.1 Semantic Technologies

The growth and availability of data and the use of it in decision-making and planning is growing exponentially. Systems should be able to take decisions by understanding the data and in a way interpreting the pieces of data. Semantic technologies [9] include software standards and methodologies that are aimed at providing more meaningful information, which can be processed by systems. With semantic technologies, adding, changing and implementing new relationships or interconnecting programs in a different way can be just as simple as changing the external model that these programs share. For better search of bibliographic metadata this should be available and understandable to the machines. This gives rise to high need of provision of bibliographic data in semantic web formats.[10]

Semantic technologies include tools for [11]:

- **Autorecognition of topics and concepts** : The tools to represent the meanings and knowledge separately from the data, content, and behaviour of the information resource. In a given question, semantic technologies can directly search topics, concepts, associations across a number of sources.
- **Automatic extraction of Information and meaning** : The tools to extract the information automatically , and analyse it semantically by taking the context of keywords and produce appropriate results
- **Categorization** : The tools to categorize and connect data with contextual information to make it easier to organize and search.

In order to deliver the metadata in a format suitable for the semantic web, some issues will have to be solved. [10] such as, Uniform Resource Identifiers (URI's), Ontologies, Vocabularies and Query Interfaces. The available semantic technologies as discussed below, can be used for resolving these issues.

- URI
- XML and XML Schema
- RDF and RDF Schema
- Ontologies
- SPARQL

Use of these technologies are discussed here with respect to following typical library use scenario:
Library Ordering System: A semantically enriched system that helps libraries in acquisition of resources. The system will store all the details about vendors, online shops, user profiles and semantic relations between concepts in given domains .
End User Interface: Puts a query or requisition in the system about the required documents. The system processes it and forwards the order.

4.1.1 URI's

URI is the identifier for a specific resource on the web. It is now well known that URIs are one of the cornerstones of the semantic web as in semantic web 'whatever is being described and is having an existence must be having one URI'. There are two forms of URI, URL (Uniform Resource Locator) and URN (Uniform resource Name). URL is the most popular URI which helps one to visit a webpage, ie. to locate the page. The URN deals with the name of the resource, for example, the DOI (Digital Object Identifier) of a resource is a URN. For example, the D-Lib Magazine URL "http://www.dlib.org/", which leads to the magazine website. The URN of the magazine is "doi:10.1045/dlib.magazine" and the articles of each issue will have their own URN, the article "Measuring Citation Advantages of Open Accessibility, by Samson C. Soong et. al." published in the "November/December 2009" issue has the URN "doi:10.1045/november2009-soong".

In case of libraries, just as every web resource is having one URI, it is partly true in the case of library holdings.[10] Libraries can assign their own identifiers for better resource management. However, for better adaptability on semantic web the URI scheme should be generic and a good dereferencing mechanism delivering a proper representation of resource is a must.

In respect to the above scenario each vendor, resource, user, etc. will be identified by a URI, and each object in the collection of vendor will have their own URI . For example, if the system has "amazon" in the vendor list it will have the URI "http://www.amazon.com", and suppose the book one wants to purchase is "You Are Not a Gadget: A Manifesto", it will have a URI, something like, "http://www.amazon.com/You-Are-Not-Gadget-Manifesto". Similarly, every vendor or resource will have their own URI and system fetches information using these URIs . The system will process the acquired information with other factors, for example, it will compare the price, margin, time to deliver of the item, mode of payment etc. and will place the order to the appropriate vendor to deliver the item based on specifications stored for such transactions.

4.1.2 XML and XML Schema

XML (Extensible Markup Language) provides an elemental syntax for content structure within document. It helps to create meaningful tags compared to HTML. XML is created to structure, store, and transport information. The XML Schema is a language for providing and restricting the structure and content of elements contained within XML documents. XML Schemas provide mechanisms to define and describe the structure, content, and to some extent semantics of XML documents. [8, 12]

4.1.3 RDF and RDF Schema

The Resource Description Framework (RDF) [13] is a simple language for expressing data models, which refer to objects ("resources") and their relationships. RDF describes resources with classes, properties, and values. Application-specific classes and properties must be defined using extensions to RDF and one such extension is RDF Schema. RDFS is an extensible knowledge representation language used to create RDF vocabularies.

The use of RDF [14] is not in encoding information about resources, but information about and relations between things in the real world such as people, places, concepts, etc. RDF provides a general, flexible method to decompose any knowledge into small pieces, called triples, with some rules about the semantics of those pieces. The foundation is breaking knowledge down into a labelled, directed graph. Each edge in the graph represents a fact, or a relation between two things. A fact represented this way has three parts: a subject, a predicate (verb), and an object. Any entity which has to be described is known as Resource also known as Subject. It can be a 'webpage' on Internet or a 'person' in a society. Whereas, any characteristic of Resource or its attribute which is used for the description of the same is known as Property, also known as Predicate. For example, a webpage can be recognized by 'Title' or a man can be recognized by his 'Name'. So both are attributes for recognition of resource 'webpage' and 'person' respectively. Similarly, a property must have a value also known as Object. Like, the title of DRTC webpage is 'Documentation Research and Training Centre', name of a person is 'Ranganathan'[15].

If library data which is now in the form of MARC-XML is converted to RDF the querying would become easy. The information about the vendors their stock, the domain of users, their preferences, the budget of the library and other information is form of URIs are stored with their internal relationship. It can provide syntactic interoperability, and the whole description further provides the base layer for building the system for the mentioned scenario.

4.1.4 Ontologies

Ontology is a formal representation of a set of concepts within a domain and the relationships between those concepts. In theory, an ontology [16] is a “formal, explicit specification of a shared conceptualisation”. An ontology [17] provides a shared vocabulary, which can be used to model a domain that is, the type of objects or concepts that exist, and their properties and relations.

A domain-specific ontology models a specific domain, or part of the world. It represents the concepts and their relations apply to a specified domain.

It helps in creating a unified view of data across the applications and resources. for example at present the information is stored in relational databases flat files, object oriented data model. If some change is to be made either one copies the relation to another system or again rewrites the data model. In the above scenario if libraries adapt an ontology-driven approach, it can capture and represent its total holdings about the resources in a language-neutral form and deploy the knowledge in a central repository. With this shared, adapted ontology, the libraries can provide a single, unified interface irrespective of the system used. This would further provide far more intelligent, relevant, and contextually aware interactions than those available with the traditional point-to-point integration approach for information delivery. [15]

4.1.5 SPARQL

Query languages are designed to be applied to data corresponding to a particular data model. RDF [18] is structured data- a graph-based data model, - like tree-based data models. Syntax of SPARQL has been designed to combine information taken from one or more RDF graphs without the need for users at the time of query, to explicitly identify the mechanisms by which the graphs are combined. A standardized query language for RDF data with multiple implementations offers a way to write and to consume the results of queries across this wide range of information. Used with a common protocol, applications can access and combine information from across the Web. It is a protocol and query language for semantic web data sources.

The present day query formats viz, SRU/SRW for Z39.50 protocol or the OAIPMH for metadata harvesting are just built on protocol level they are not sufficient for retrieval in semantic web world. [10] In the above scenario the user queries would be taken from the search bar and then the query is issued in SPARQL syntax.

5. Conclusion

Semantic web and SW technologies have the potential to provide better retrieval, better resource management and better integration. Libraries should also adopt these technologies to give better services to their patrons. With their large collections of semantically valuable bibliographic and authority data, libraries, and especially National Libraries, can make immense contributions to the Semantic Web if their data can be supplied in an interoperable format. The adaptation of the library data formats and practices to the semantic web 'way' would require significant planning, resources and adventure too as rate of change with SW technologies is quite rapid. But still these could be managed in such a way that even drastic changes are transparent to the end users and on the face of it he/she does not notice any change and could still seamlessly access their resources.

References

1. SCHNEYMAN, Arthur H. Organizing information resources. Information Management Review, Summer 1985, pp.35-45.
2. BURK, Cornelius Franklin and HORTON, Forest W. InfoMap: A Complete Guide to Discovering Corporate Information Resources. New Jersey: Prentice Hall, 1988.
3. HORTON, Forest W. Information resources management: harnessing information assets for pro ductivity gains in office, factory and laborator. New Jersey: Prentice Hall, 1985.
4. CAUDLE, Sharon L. Strategic information resource management: fundamental practices. Government Information Quarterly, 1996,13(1), pp. 83-97.
5. WILLARD, Nick. Information resource management. Aslib Information, 1993, 21(5).
6. Guidelines for using resource identifiers in Dublin Core metadata. Available at:http: //dublincore.org/ architecturewiki/Resource Identifier Guidelines (Accessed on 12/12/2009)
7. HAYNES, David. Metadata for Information Management and Retreival.London:Facet Publishing,2004.
8. NILSSON, Mikael. Semantic Web Metadata for e-Learning - Some Architectural Guidelines. Available at: http://www2002.org/CDROM/alternate/744/ (Accessed on 12/12/2009)
9. Semantic Technology Primer. Available at http://www.semantic-conference.com/primer.html (Accessed on 12/12/2009)
10. SVENSSON, Lars G. National Libraries and the Semantic Web: Requirements and Applications. International Conference on Semantic Web and Digital Libraries (ICSD-2007), Bangalore, India, February 2007, edited by ARD Prasad and Devika P. Madalli. Available at: https://drtc.isibang.ac.in/ handle/1849/326 (Accessed on 15/12/2009)

11. Semantic technology. Available at http://en.wikipedia.org/wiki/Semantic_technology (Accessed on 12/12/2009)
12. XML and Web Services In The News - 6 October 2006. Available at:http://www.xml.org/xml/news/archives/archive.10062006.shtml#5 (Accessed on 14/12/2009)
13. BRAY, Tim. RDF and Metadata. O'Reilly xml.com:xml from the inside out Available at http://www.xml.com/pub/a/98/06/rdf.html (Accessed on 14/12/2009)
14. TAUBERER, Joshua. What Is RDF. O'Reilly xml.com:xml from the inside out Available at http://www.xml.com/pub/a/2001/01/24/rdf.html (Accessed on 14/12/2009)
15. PRASAD, A.R.D.and PATEL, Dimple. An Overview of the Semantic Web. 2003. Available at: https://drtc.isibang.ac.in/handle/1849/131 (Accessed on 14/12/2009)
16. GRUBER, Thomas R.. A Translation Approach to Portable Ontology Specifications. Knowledge Acquisition, 5(2):199-220, 1993 . Available at http://tomgruber.org/writing/ontolingua-kaj-1993.htm (Accessed on 14/12/2009)
17. ARVIDSSON, Fredrik and FLYCHT-ERIKSSON, Annika. Ontologies I. Available at: http://www.ida.liu.se/~janma/SemWeb/Slides/ontologies1.pdf (Accessed on 14/12/2009)
18. MELTON, Jim.SQL, XQuery, and SPARQL:What's Wrong With This Picture?.XTech conference 2006: Building Web 2.0 .Available at: http://xtech06.usefulinc.com/schedule/paper/119 (Accessed on 14/12/2009)

About Author

Dr. Devika P Madalli, Assistant Professor, Documentation Research and Training Centre, Indian Statistical Institute, Bangalore.
E-mail: devika@drtc.isibang.ac.in