An Introduction to Semantic Information Retrieval in Digital Libraries

Selvam M.

Librarian, Loyola College, Vettavalam, Tiruvannamalai (Dt)

Abstract: *World Wide Web has a huge collection of databases that contain unstructured information which is retrieved by the humans based on the keywords. Due to the enormous growth of information, the web delivers irrelevant information often. There are many information retrieval techniques and algorithms have been used to satisfy the human needs on the web but still it is uncertain. Semantic web is an extension of current web which is designed to overcome this issue by helping the machines to understand the request of the humans and provides the information in a structured manner. This articles explores an information Retrieval through Semantic concepts in Digital Libraries and the World Wide Web which is highly demanded today.*

Keyword: Search Engines, Resource Description Framework, Ontology, Web Ontology Language, Information Retrieval, Digital Libraries

1. Introduction:

Information Retrieval is a method of storing, organizing, and offering access to information sources. It is a common process of accessing information which can be applied to any type of research field. Information Retrieval offers unstructured information which leads distrust on web as well as information repositories. Knowledge based Retrieval tools are used to help the users to find their choice of information but still it needs more advancements in retrieval process. Semantic Web technologies have been integrated with digital libraries to meet the needs of Digital Libraries users. It teaches the computer the relationships between digital objects, their meaning, and how users would like to interact with them. The search and retrieval functions would benefit from the ability to search semantically along with the traditional search methods the library currently employs.

2. Studies on Information Retrieval and Digital Libraries

There are many papers have been published on Modern information retrieval techniques in digital libraries.

Walson [1] described that the information retrieval in a library system based on Electronic and Non Electronic forms by the user. He used Historical survey method to bring the originality of this paper and concluded that both the forms have helped the researches and academicians on different way in a successful manner.

Onwuchekwa [2] stated that the information retrieval methods and its significance in Library system. He also added Information seeking behaviour of the user during Information Retrieval Process. This paper suggested that the Effectiveness of information Retrieval methods had to be evaluated periodically.

Jones [3] reviewed the history of Information Retrieval research and demonstrated the statistical techniques for retrieval. He also explored the relationship between information retrieval (IR) and related research and digital libraries.

This paper aims to explore the fundamentals of semantic web Technology with information retrieval concepts in Digital Libraries.

3. Definition: A Semantic Information Retrieval System

An information retrieval system is designed to obtain the documents or unstructured information required by the user[1] from a large collection of data within a database.

A Semantic Information Retrieval System is a well-structured and the information is organized in a meaningful way such that the machines could be understood the users query (request).
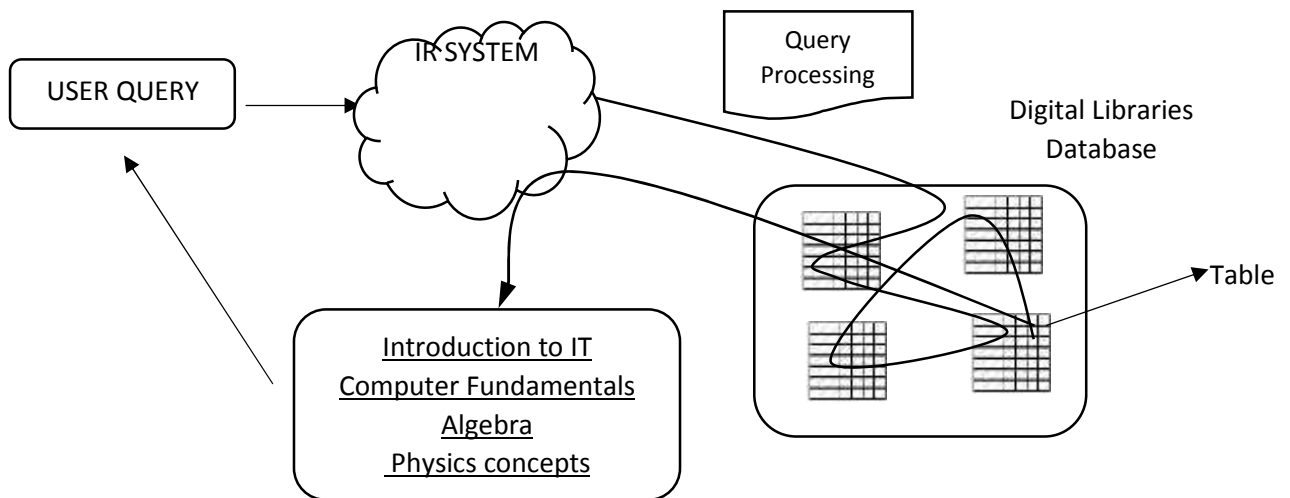
Figure 1 - A Simple Architecture of IR in Digital Libraries

4. Need of Semantic web:

The internet, WWW has a biggest database with a huge collection of information covering all the disciplines in a weakly structured form (text, audio and video). Search engines today will pull up results based on keywords and phrases because that is what it is told to do. So it leads to the following limitations to be considered;

- ❖ Searching information.
- ❖ Extracting information.
- ❖ Maintaining information.
- ❖ Uncovering information.
- ❖ Viewing information.

If the Search engines are beginning to understand the semantics of the pages it can overcome the above limitations by,

- ❖ Information will be organized in conceptual spaces according to its meaning.

---

[1] Onwuchekwa, Edeama O and Jegede, Olumakinde Richard Information Retrieval Methods in Libraries and Information Centers. An International Multidisciplinary Journal, Ethiopia. Vol. 5 (6), Serial No. 23, November, 2011, Pp. 108-120.

- ❖ Automated tools will support maintenance by checking for inconsistencies and extracting new knowledge.
- ❖ Keyword-based search will be replaced by query answering: requested knowledge will be retrieved, extracted, and presented in a human-friendly way.
- ❖ Query answering over several documents will be supported.

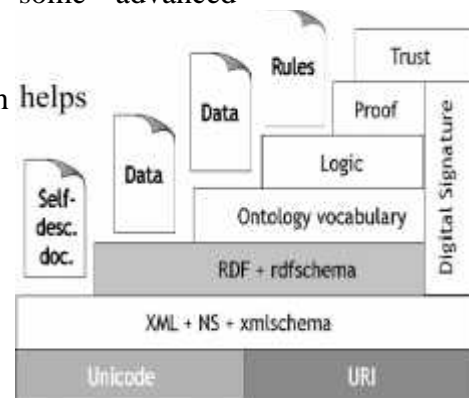## 5. Components of Semantic Web:

Semantic web can't work by all itself; it needs society scale applications such as semantic web agents, processors of semantic data and some advanced collaborative applications.

URI expanded as **Uniform resource Identifier** which helps to identify the web resources. It classified into two forms: URL (Uniform Resource Locator) which is used to identify an Internet resources location and URN (Uniform Resource Name) which is used to identify an internet resources with implying its location.

Figure 2 – Components of Semantic Web

URL:http://www.selvam.org/iis.review1.htm
URN: www.selvam.org/iis.review1.htm#one
URI: http://www.selvam.org/iis.review1.htm.html#one

XML (eXtensible Markup Language) facilitates exchange of data and allows users to add their own tags. It contains structured information of a document. It is a basic language for semantic web.

RDF & RDF SCHEMA (Resource Description Framework) is a data format which is used to describe the web resources such as Title, Author, content and copyright information. It is an XML based framework. It also helps to describe the relationship between the resources and its properties in a machine understandable format. It has three objects:

    Resources: A resource may be an entire Web page; a part of a Web page;
    Properties: A property is a specific aspect, characteristic, attribute, or relation used to describe a resource.
    Statements: A specific resource together with a named property plus the value of that property for that resource is a RDF statement.

OWL: Web ontology Language represents the knowledge based ontology language which is supported by W3 consortium for semantic web. It works based on Description logics. OWL has three sub languages; OWL lite, OWL DL and OWL full.

SPARQL: The Simple Protocol and RDF Query Language (SPARQL) standard is a query language for RDF. It provides a declarative interface for interacting with an RDF database.

LOGIC Layer is to describe a formal mathematical logic that reconciles all the different model semantics of the parts (RDF, RDFS, OWL, SPARQL, and RIF).This layer enables to write rules.

PROOF layer executes and evaluate the rules which is written in the logic layer.

TRUST Layer is the top layer of semantic web architecture. The trustworthiness of the content is evaluated.

Digital Signature is an encryption technique defined for lower layers of the stack like Unicode and XML.

6. Integrating Semantic Information Retrieval in Digital Libraries

Information Retrieval and Digital Libraries are used to deliver the digital contents to the users. Digital contents are classified into two categories, Offline Digital Contents and Online Digital Contents.

Offline Contents are collected, stored, classified and retrieved by the user through Information Retrieval System. The contents are categorized and metadata can be created based on the metadata schema (MARC 21, Dublin Core…).

Online Contents are accessible through IR system which the contents are retrieved from webpages, online databases, Online Directories, news groups and open access e journals. Most of the online information retrieval systems have been designed to serve as a commercial basis. It provides access to various remote databases to variety of users.

Figure 3 shows the structure of Semantic information retrieval in Digital Libraries. It's an integration of Semantic Web components and the information retrieval concepts. As information retrieval techniques helps to extract the information on the web whereas the semantic web components allow machines to understand the meaning of digital objects, rather than just the key words used to describe them. This will revolutionize the search and retrieval of digital objects, a key function for digital libraries. Implementing Semantic Technologies in Digital Libraries will provide the right amount of information at right time to the right user.

The main objective of Semantic Web is that it adds logic and meaning to digital objects. Digital objects contain described metadata or information contained within the document that helps to define the information about the document such as the author, publisher and date created.
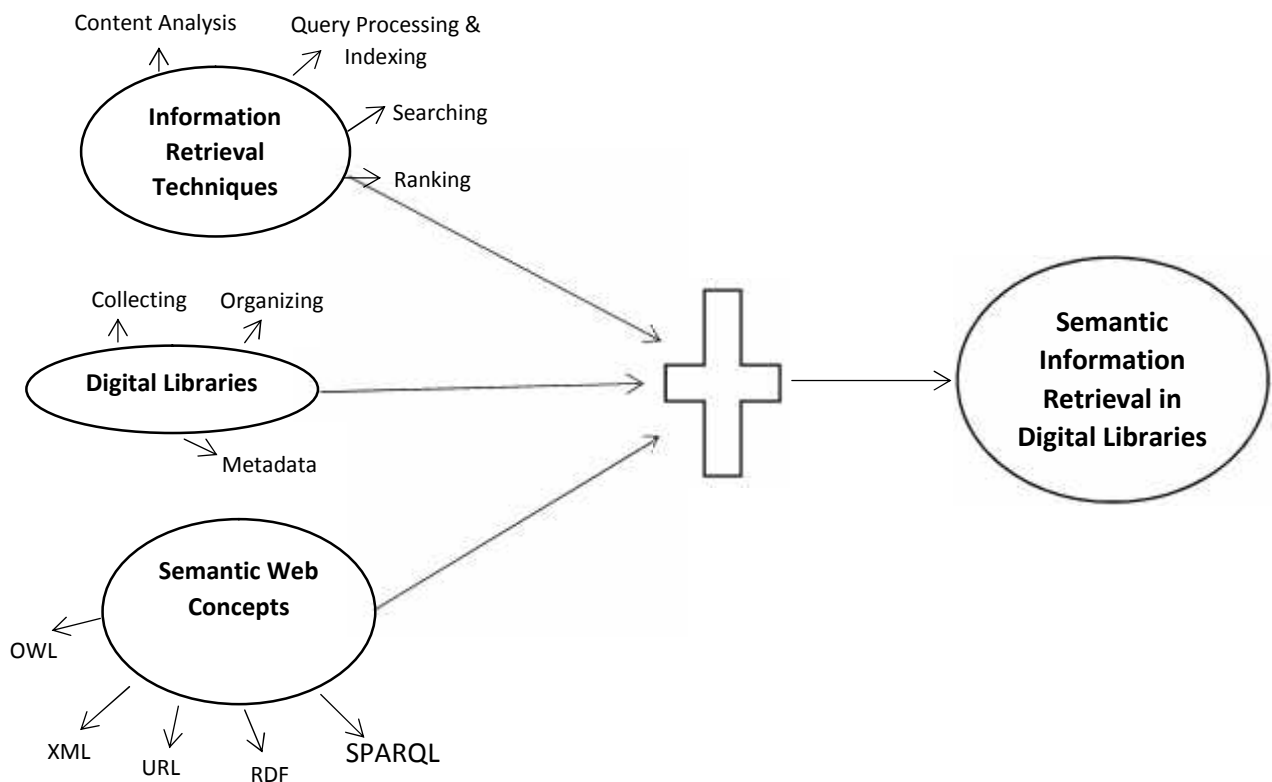
Figure 3 – Semantic Information Retrieval in Digital Libraries

XML, the basic semantic web language can define structure within a digital object but it does not convey meaning to computers. So, the semantic web process is to add meaning to the XML data found within digital objects. This is achieved through the creation of RDF models and schemas.

RDF models consist of three parts such as subject-predicate-object (often called as triplets). The subject is a resource which can essentially be any digital object. The predicate is a property, such as "has_author" which signifies that the resource has an author. The object can be another property or a class of that property. (Refer 5). The following example explains the concepts of RDF.

Resource (Subject): http://lcv.edu.in,
Property (Predicate): Creator
Value (Object): Selvam
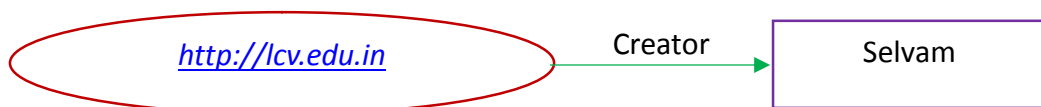
Ontologies are backbone of semantic web and they are usually domain specific. Ontologies also include information about the relationships among terms (concepts).

The combination of XML, RDF, and ontologies makes it possible for computer programs and agents to interpret digital objects based on their meaning and their relationship to other concepts.

There are many digital libraries have implemented these concepts; they are

- Unified Medical Language System (UMLS) developed by the National Library of Medicine
- Europeana (http://www.europeana.eu)
- Friend of a Friend Project or FOAF (http://www.foaf-project.org/)
- JeromeDL (DERI - www.deri.ie/content/**jerome-dl***)*

7. Conclusion:

This article explored the components of semantic information Retrieval and its implementation in digital libraries. Digital libraries should adopt the semantic web technology and facilitate to share their contents to the different user communities of other digital libraries. Digital libraries should provide advanced search facilities and help the users to customize their choice of retrieval. It has to allow the user to create their profiles and share their bookmarks to the social networking sites. The technologies have been introduced in the digital libraries already, now it's a time to the digital librarians to integrate themselves to the technology and bring out their innovative services to the society.

References:

1. Walson, Okede Godbless. Information retrieval in a library based system: electronic and non-electronic. *Comprehensive Research Journal of Education and General Studies* (CRJEGS) Vol. 2 (4) pp. 040 - 043 June, 2014
2. Onwuchekwa, Edeama O. Information Retrieval Methods in Libraries and Information Centers. *African Research Review*. Vol. 5 (6), Serial No. 23, November, 2011.
3. Jones, Karen Sp¨arck. Information retrieval and digital libraries: lessons of research. *Proceedings of the International Workshop on Research Issues in Digital Libraries* (IWRIDL 2006).
4. Moran, Carrie [2010]. The Use of Semantic Web Technologies in Digital Libraries.
5. Europeana. (2010). Europeana. Retrieved from http://www.europeana.eu/
6. Friend of a Friend Project. (2010). The Friend of a Friend (FOAF) project. Retrieved from http://www.foaf-project.org/
7. Hersh, William R. Information retrieval and Digital libraries. Pp 237-275
8. Inkpen, Diana. Information Retrieval on the Internet.