INTERNATIONAL JOURNAL
FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Harnessing Ontologies and Semantic Web Tools for Enhancing Digital Library Retrieval Systems: A Potential Solution for Scholarly Digital World

Nilesh Shewale[1], Dr. J. Shivarama[2]
[1]Research Scholar, [2]Assistant Professor, Tata Institute of Social Science

Abstract: *The evolution of ICT has brought many changes in our day to day activities and also have brought unprecedented and dramatic changes in library profession and practices. Present rapid development in communication system and recent innovation in technology is witnessed as changing emphasis in the role of information and management. The importance of archiving and finding relevant archived information has been recognized for thousands of years. The advent of computers made it possible to store large amounts of information in digital library and also supports retrieval of information from such large store of information. However, in this process a major issue that emerged was 'relevance' of the information to the information needs of the end user. The purpose of the study mainly focuses on understanding the need* and potential applications *of Semantic Web Technologies / Ontologies in enhancing information retrieving capacity of Digital Libraries* DL's*. To satisfy the purpose of the study, the study was carried out on Digital Repositories available globally. The responses of Administrator / Librarians of Digital libraries used to analyze overall features, problems of DL's and the need of ontologies in enhancing the information retrieval capacity of Digital Library. The study revealed the need for enhancing the digital library features such as browsing, organizing and retrieval efficiency of Digital Library. It was also observed that integration of* domain specific ontologies *may help to retrieve the information more precisely. The study was limited to 100 digital repositories available globally. The study mainly focused on retrieval techniques used in Digital Library Software and the use of domain-specific ontologies which can be applied for enhancing retrieval efficiency. The finding has the implication for Librarians and Administrator of Digital Library information retrieval system. Based on the need and application of ontologies administrators/librarians can think of implementing ontology-based information retrieval system in Digital libraries.*
Keywords: *Digital Libraries, Information Retrieval, Ontology, Semantic Web, Information Discovery*

I. INTRODUCTION

The term "Digital Libraries" has a variety of potential meanings, ranging from a digitized collection of material that one might find in a traditional library to the collection of all digital information along with the services that make the information useful to all possible users. Digital Libraries have a role in preservation/archiving information. Digital library presents opportunities for long-term preservation of resources. Libraries have always been a source of organized knowledge for various application areas, teaching, and research, decision-making. Digital Library not only delivers high-quality knowledge but also provide the wide range of search and browsing services. Digital Library is playing an important role in creating, indexing and retrieving the various forms of digital resources. Digital Libraries are catering to information services interacting with online databases and other Digital resources. The end users approach to meet information needs is also rapidly getting transformed. The Digital libraries significant feature is the efficiency of retrieving the information by browsing the catalogue in an electronic medium.

The new technologies have facilitated the transformation of data into digital formats. Bruce, states

"Information plays a significant role in our daily professional and personal lives and we are constantly challenged to take charge of the information that we need for work, fun and everyday decisions and tasks." (Bruce, 2005)

Scholarly Digital World whereby the emergence of the internet and computer networking have provided the means to information which can be stored, retrieved, duplicated and disseminated in very fast manner. The Scholarly Digital Libraries without retrieval efficiency is no good according to the user perspective. Most of the Digital libraries for its usual resource discovery tools provide are classic "browse" and two different search techniques "simple search" and "advanced search" facilities. The Digital libraries evolved to offer resources in domain specific, document type specific and user level/type specific environments. Some digital libraries contain content from diverse and distributed user communities. On the other part, digital libraries can be complex systems which process hundreds of parallel requests in diverse scenarios (A.R.D. Prasad, 2008). The user interface to use in digital library

should be more dynamic than the static interfaces where users will be allowed to choose the required terminologies and attributes according to their requirements, and it should, as much as possible, hide all the barriers such as the need for learning specific commands, retrieval software/search engine features and so on, and thus should facilitate to enhance users reading and learning habits in a digital library environment (G.G. Chowdhury, 2000). The users make frequent use of the Internet to search for specific academic information, and user's searches for specific academic resources are the main trend in the field of Library & Information Science (Xiangxing Shen, 2008).

The field of Information Retrieval (IR) came into existence in the 1950s out of above necessity. Information retrieval specifically deals with retrieval of information from structured and unstructured data and provide relevant results for the user query. Data stored in the most text databases are semi-structured data in a way that they are neither completely unstructured nor completely structured (Gerard Salton, 1983). The modern web search engines are basically known for implementation of IR System. The rapid spread of full-text searching and the advent of the Internet have contributed to the origin of different types of digital information retrieval systems for instance online database, web portals, digital libraries and so forth (Sanjeev K. Sunny, 2018). The information retrieval process starts when the user enters a query. The queries are the natural state of information needs (i.e. keywords use in a search engine).When the search engine searches for the query, it collectively gathers the document related to search string with a distinct degree of relevancy. User queries are matched with information objects stored in the database. Depending on the application, the objects may be retrieved, for example, text documents, images or videos. Several IR systems are being used on an everyday basis by a wide variety of users

The search features of Digital Libraries like Keyword & Phrase Search, Boolean Logic, Proximity Search, Truncation Search, Range Search and thesaurus based retrieval searching enhances the retrieval. To make the Digital Library Information Retrieval system more meaningful semantic web technologies may be useful. While evaluating the retrieval effectiveness of Digital Libraries, all the studies report the positive effects of a thesaurus in retrieval performance; however, it is noteworthy that to enhance information retrieval, they all necessarily used some other tools/techniques in addition to the thesaurus. Though thesaurus based Search can improve Semantic Information Retrieval System, and use of domain-specific ontology will further improve. The idea is to develop technologies that make the information more meaningful for the machine to process, so as to support more effective retrieval of information (Sanjeev K. Sunny, 2018).The evaluation of different retrieval feature of eight Indian Digital Libraries they found that the search facilities used in digital libraries are providing navigational support sometimes which are very comprehensive. However, the simple search and advanced search sometimes lead to a search result with more recall ratio (Anup Kumar Das, 2007). Most of the search engines use keyword-based search for mining the relevant information from the machine that "obstinately" returns zero hits in response to a query that does not match the character strings in its database files (Saeed Rezaei Sharifabadi, 2006). The Digital Library plays a critical role in extracting information from the database due to lack of semantics. Greenstone version 3 allows for the creation of metadata based digital library collection with the support and services of semantics. The process of configuring an ontology-enabled Greenstone version 3 collection is more complex, requires more effort, and is resource intensive and does not cost free (Hinze, 2008). The domain ontology describes a proper amount of subject knowledge with the definitions of main concepts, properties, relationships and representative instances that have been used to build the catalogue descriptions and to process the user queries. The DjDL supports both keywords as well as ontology-based search (Nisheva-Pavlova, 2011).

The search engines have problems due to ambiguity in the words so the ontology or context based ontology finds the words which are syntactically different but semantically are similar. Based on the similar studies conducted by many of the researchers it is observed that ontologies and semantic tools help in retrieving the precise information from digital libraries (Pahal, N., Gulati, P., & Gupta, P. 2010).

A. Challenges while Searching in Web and Digital Libraries:
1) Focused Search Engines –Evolution of domain-specific search engines those who have adopted a faceted searching technique which often lacks domain specific search and the recall is greater than precision.
2) Organization of the Web – Basic HTML cannot preserve the context and syntactical information of the term which plays a major role at the time of Information retrieval on the Internet.
3) Lack of Global Standard and Proven Frameworks - There are only a few standard and schemes for Knowledge representation and Domain mapping such as OWL, KOS, Thesaurus, Taxonomies and so on in the web. RDF Schema a language for writing ontology which provides a standard platform for representing knowledge semantically using ontology languages.
4) Lack of Availability of Formal Domain-Specific Ontology – Domain-Specific Ontology for various fields are not available and creating them itself is a challenge.

II. NEED FOR THE STUDY

Search engines and search techniques have fallen short of user expectations as they do not give context-based retrieval. Library and Information Science professionals are bound by partial technological understanding and they frequently need the assistance of IT specialist to deliver services of Digital Libraries. Implementing customized Semantic Web-Ontology based solutions in Digital Library for improving the efficiency of Retrieval by helping Library and Information Science professionals to a great extent. Semantic web technologies in digital libraries allow new features such as interoperability, flexibility, retrieval, browsing and sharing of knowledge. The researchers say that the simple, standard and metadata based digital libraries which use metadata structures such as Dublin Core or MARC lack in expressing particular information need of users. Where information retrieval can be more powerful by use of domain specific ontologies and semantic technologies.

III. OBJECTIVE OF THE STUDY

A. To find out difficulties faced by administrators/librarians while accessing the knowledge in Digital Libraries.
B. To identify the potential applications of semantic web technology tools for discovering information in Digital Libraries

IV. SCOPE OF THE STUDY

To study the existing problems and solutions in Digital Library retrieval, following steps were adopted to study the need for ontology-based retrieval in Digital Library. To satisfy the objective the study of search/retrieval features was carried out on 100 Digital Repositories available globally. Study of user behaviour while searching will help for setting parameters to provide better technique while searching in DL's. Therefore with the assumption of the potential applications of semantic web technologies and domain specific ontologies with DL's will help the user in extracting precise information from DL's.

V. RESEARCH METHODOLOGY

This research study had examine the trends in the deployment semantic web technologies in Digital Library. To access the need for implementing the semantic web technologies following method of survey was designed. Questionnaire as a tool was used for a survey research specifically online survey to capture data in the faster way of collecting data from the respondents. The researcher designed the questionnaire where all possible aspect of searching in digital repositories and use of ontology in the digital library is incorporated. Which accessed the need for semantic technology in Digital Repositories. The questionnaire was designed in online form and was distributed through email to the Librarians /Administrators. The questionnaire was designed for Librarians/Administrator and they are asked questions pertaining to the use of Digital Library, Degree of usage of search engines for retrieving information from digital repositories, Knowledge about Semantic web technologies and use of semantic technologies/ Ontologies in Digital library and the notion of Integrating Digital Libraries with ontologies to enhance the precise information retrieval in Digital Repositories.

VI. ANALYSIS AND INTERPRETATION

The data collected from administrator / Librarians are analyzed which represents the degree of use of Digital Library, Features of Digital Library, efficiency of search features, Knowledge of ontologies and semantic web technologies and the notion of integration of ontologies with Digital Libraries.

A total of 100 structured questionnaires was distributed to the administrator/librarians at the global level. Out of 100, the maximum 77% responses were received. The online form of questionnaire was used to get real-time results for quick and easy analysis. The detail finding about the use of Digital Library features and use of semantic technology in DL's is presented below:

Fig 1 represents, the total number responses received out of which 54 % users have responded for DSpace as mostly used software as compare to others. Dspace is also one of the software which is mostly used all over the world as per the statistics provided by Registry of Open access repository (ROAR) and Open Directory of Open Access Repositories (OpenDOAR).
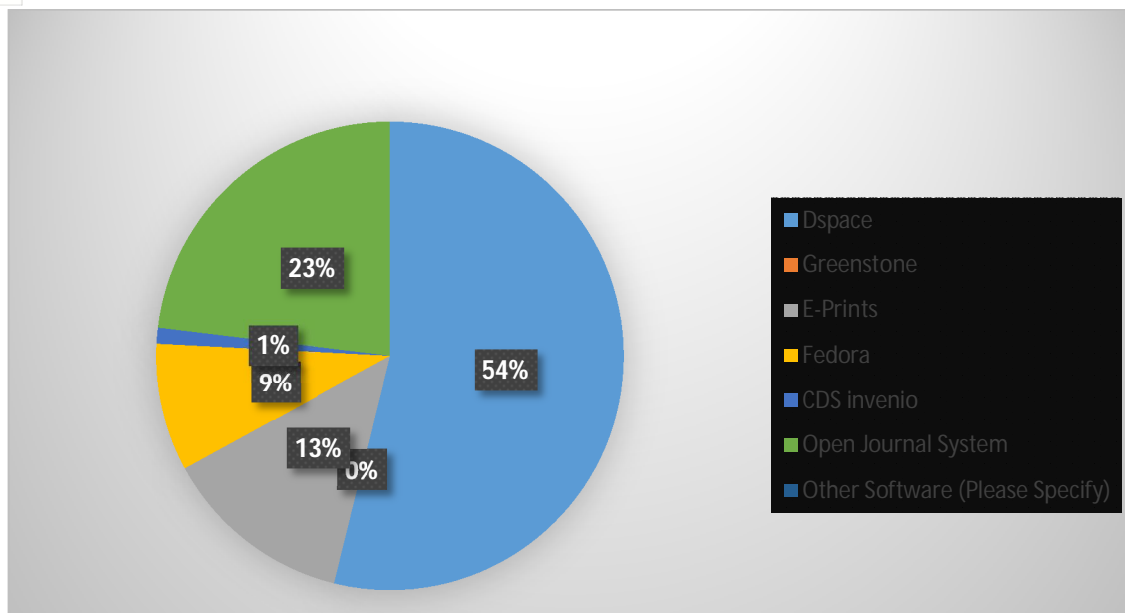
Fig.1 Digital library software used in Organization

Table 1 and Fig 2 explains how long administrator/librarian are using the Digital library software, the majority of respondents have responded i.e. above 5-6 years 57.33% 3-4 years 30.67 %, 1-2 year 10.67% and less than one year 1.33%. Therefore, it is observed that most of the users are well known about digital library software and using the software for a long time.

Table-1: Use of Digital Library software in the organization

Year	Percentage	Number of Respondents Responded
Less than 1 year	1.33%	1
1-2 year	10.67%	8
3-4 year	30.67%	23
5-6 year & above	57.33%	43
Total	100%	75
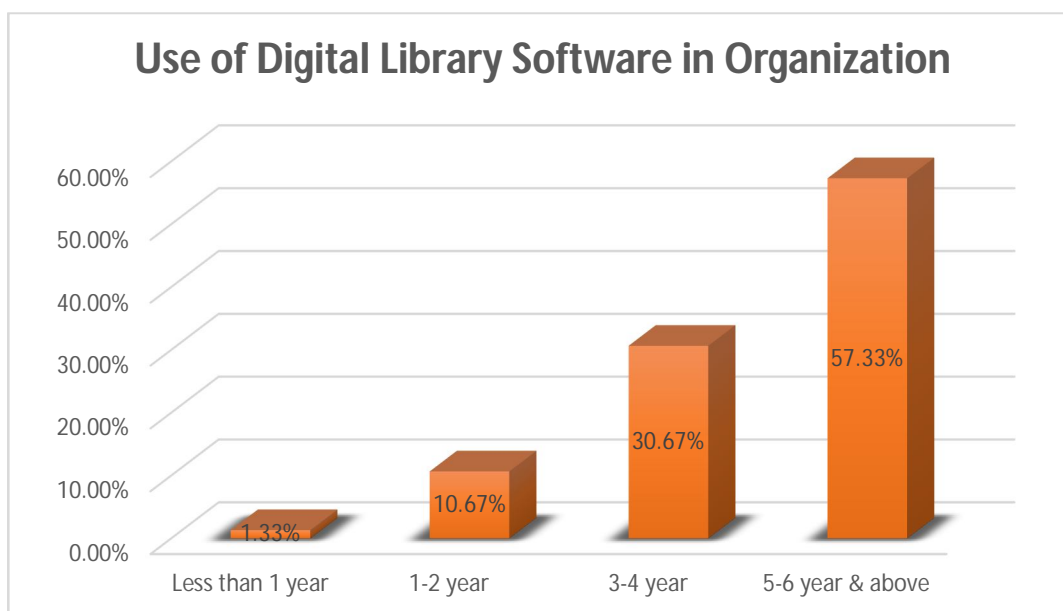
Fig.2 Use of Digital Library Software in Organization

Fig 3 presents the data about the role played in handling the repository where a maximum of the respondents have replied i.e. 36 % as repository administrator, 32% repository manager, 17 % as metadata editor and 16 % as a repository developer. At the global level, it is observed that the role played by the administrator for handling as repository administrator and Repository Manager.
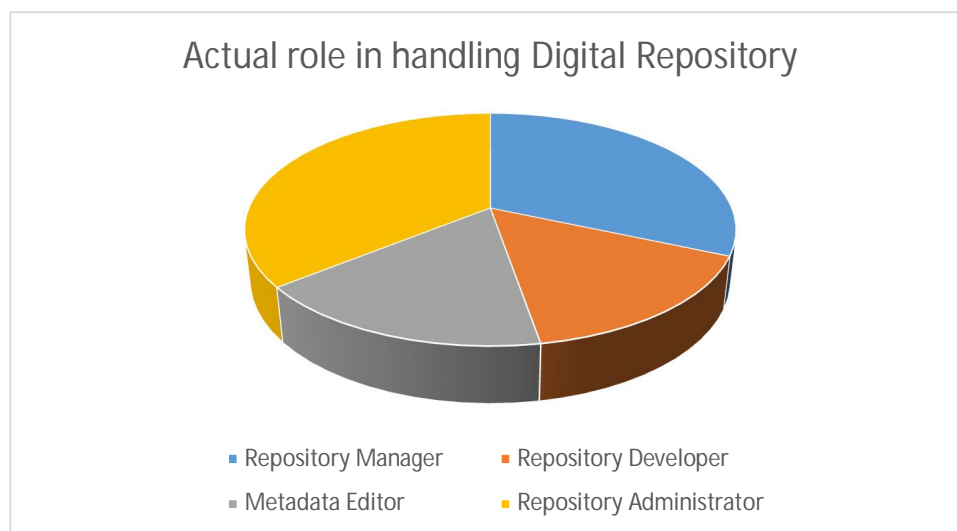

Fig.3 Actual role in handling Digital Repository

Table 2 and Fig 4 represents the important feature which digital library software should have. Most of the respondents have responded the features which many of the digital libraries packages should have, for instance, adding metadata, creating communities and sub-communities, PURL, DOI, Handle Number and so forth but it was also observed that most of the respondents have indicated for the support of context-based retrieval, Application of ontologies, Knowledge organization standards are in the list of important features of Digital libraries which are unfortunately not seen in any of the Digital Library software by default.

Table-2 Important Features for Digital Libraries

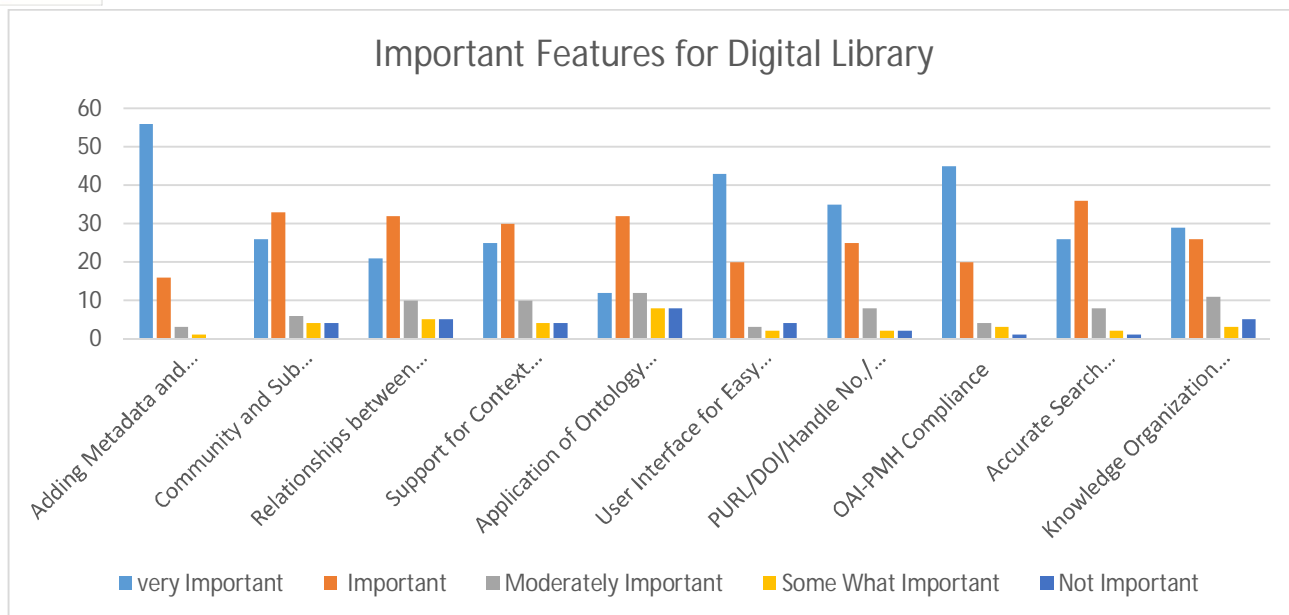| Important Features for Digital Libraries	very Important	Important	Moderately Important	Some What Important	Not Important	Total Respondents **Responded**
Adding Metadata and Metadata Standards	56	16	3	1	0	76
Community and Sub Communities	26	33	6	4	4	73
Relationships between objects or Context Based Retrieval	21	32	10	5	5	73
Support for Context Based Digital Library searching	25	30	10	4	4	73
Application of Ontology for Search and Browse of Items	12	32	12	8	8	72
User Interface for Easy search and Browse the contextualized Documents	43	20	3	2	4	72
PURL/DOI/Handle No./ Unique Id of Items	35	25	8	2	2	72
OAI-PMH Compliance	45	20	4	3	1	73
Accurate Search Mechanism such as Simple, Advance, Boolean Method for Document Retrieval	26	36	8	2	1	73
Knowledge Organization Standards & knowledge Representation	29	26	11	3	5	74

Fig.4 Important Features for Digital Libraries

Fig 5 shows that due to lack of support of knowledge organization techniques and retrieval tools such as ontology, RDF, SKOS, Thesaurus based searching and organization of Digital Library content are creating problems in managing and monitoring the Digital Libraries. Respondents have also **stated** that due to lack Digital Right management and Intellectual property rights they are facing problems in security, management and monitoring of Digital libraries.
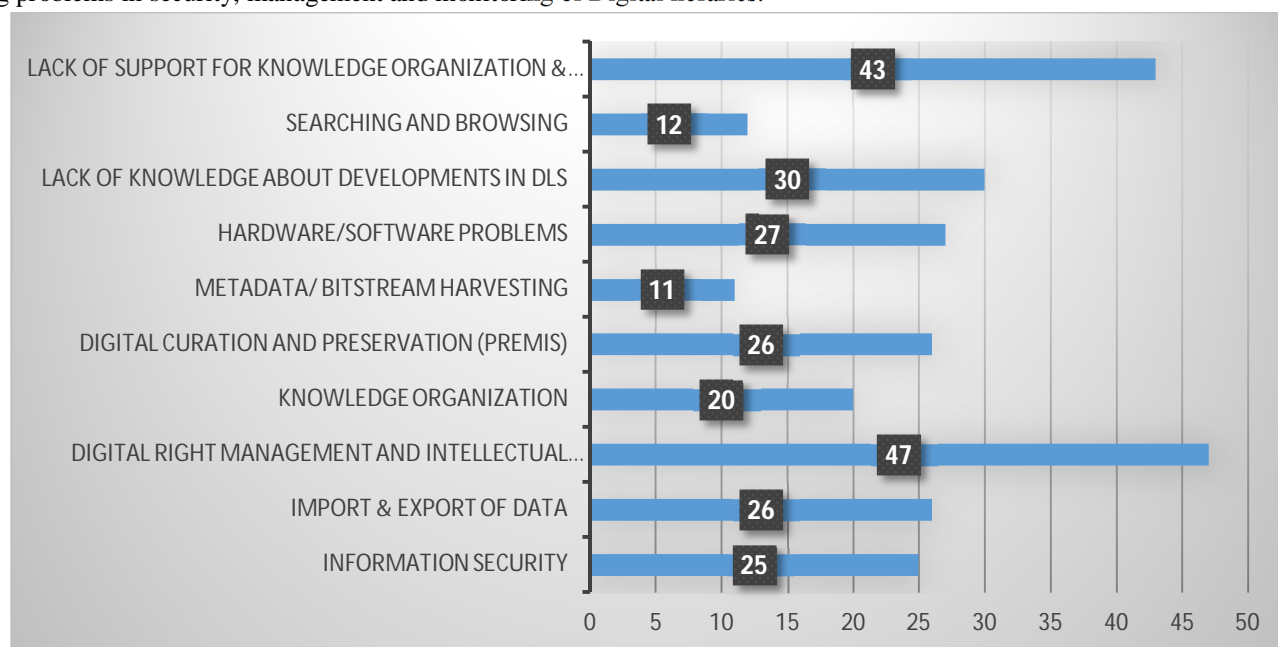

Fig.5 Difficulties in monitoring and managing Digital libraries due to lack of features

Fig 6 indicates that whiles searching in digital libraries most of them have faced problems in finding the precise information, 21% say that they found no match while searching, 18 % observed that they found vocabulary mismatch, 14 % got more irrelevant hits and so forth. It is observed that most of the respondents faced problems while searching in Digital Libraries. The problems such as vocabulary mismatch and irrelevant hits are creating problems while searching in Digital Libraries.
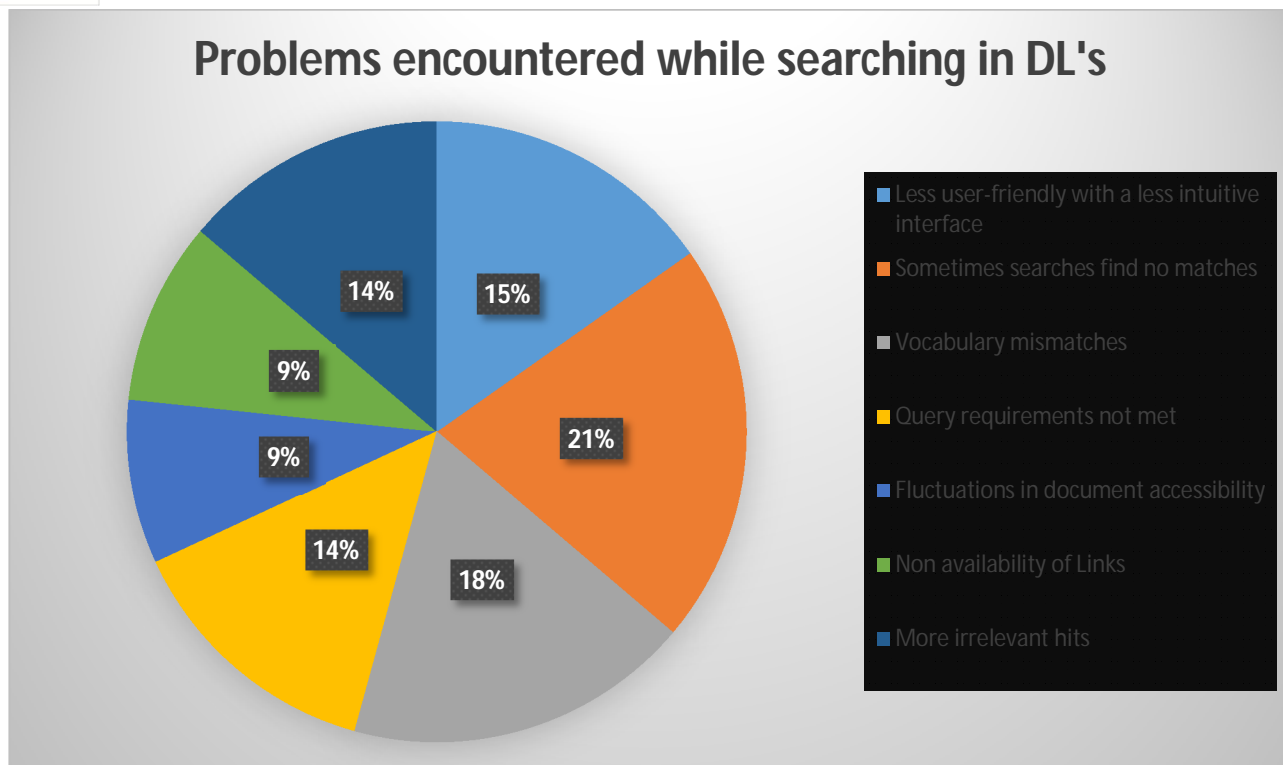
Fig.6 Problems encountered when searching in Digital Library

Fig 7 presents the degree of contribution made by Librarian / Administrator for enhancing different feature in Digital Libraries. 60 responses for enhancing look and feel of the repository, 46 responses where towards enhancing communities and collection, 38 responses tried enhancing searching and browsing technique of Digital Repository and also received responses for enhancing features such as metadata standards, registry and so forth. A large number of administrators have tried changing the look and feel which will help the user in easy navigation and browsing of information, Metadata standards and Enhancing searching in Digital Libraries.
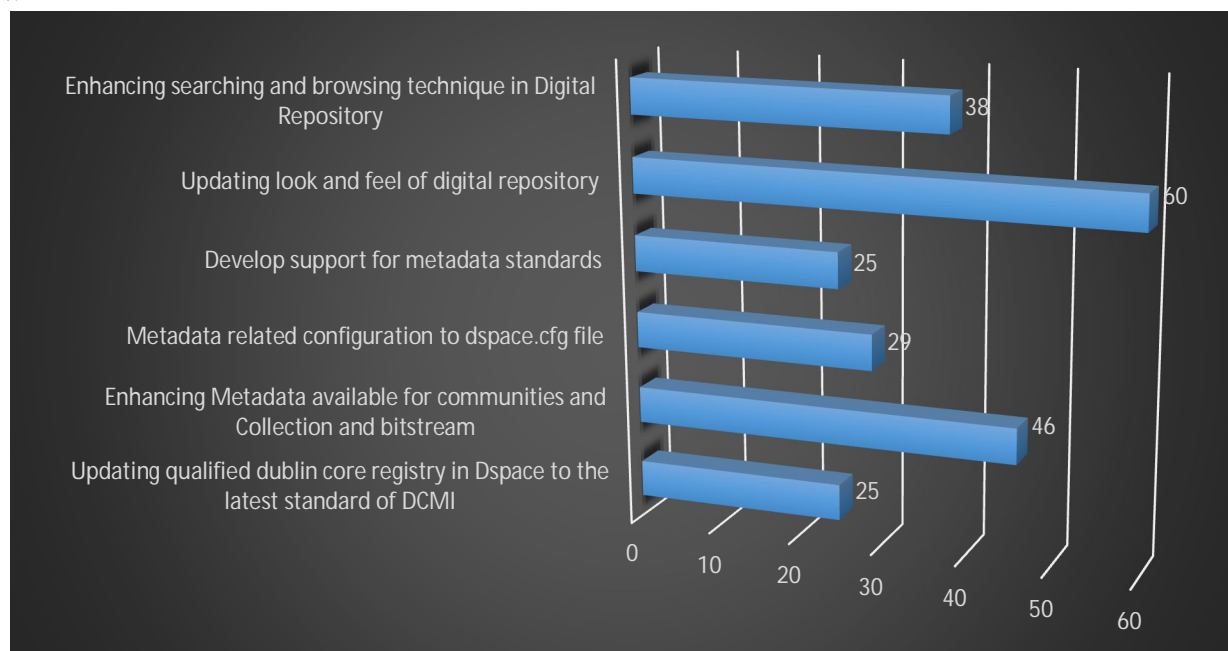

Fig.7 contribution made by Librarian / Administrator for enhancing different feature in Digital Libraries

Fig 8 denotes that, the percent of responses received towards familiarity with the concept of the semantic web where 49 % and 19 % responded that they are familiar and very familiar with the concept. However, 18% also shared their views on being unfamiliar with semantic web concept.
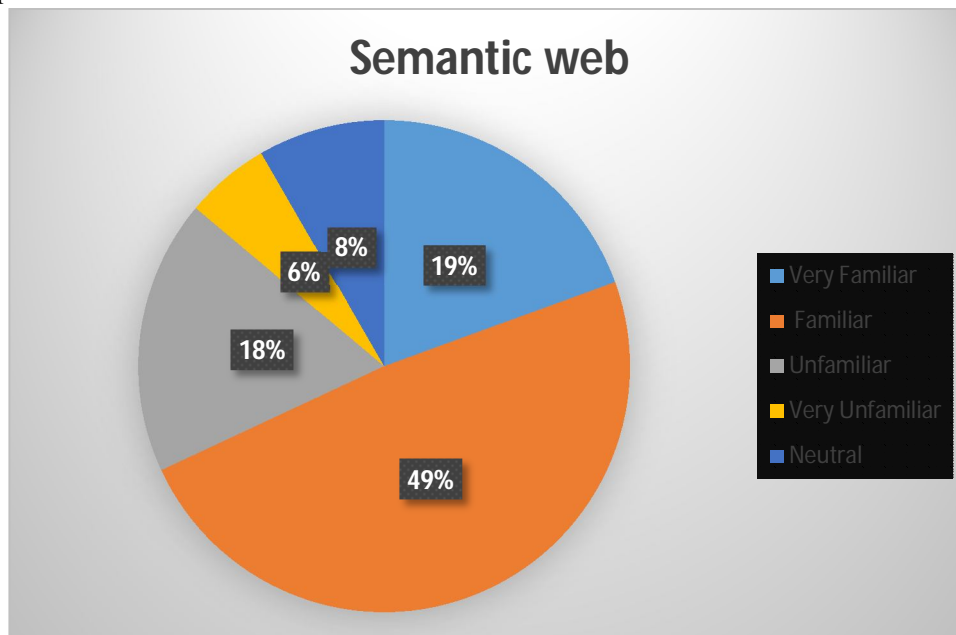

Fig.8 Familiarity with Semantic Web Concept

In Fig 9 it was also asked for the familiarity of KOS tools where most of them, have responded the unfamiliarity of KOS tools. The administrator is less aware of the tools such as ontology editors, Ontology visualizer and data mining tools
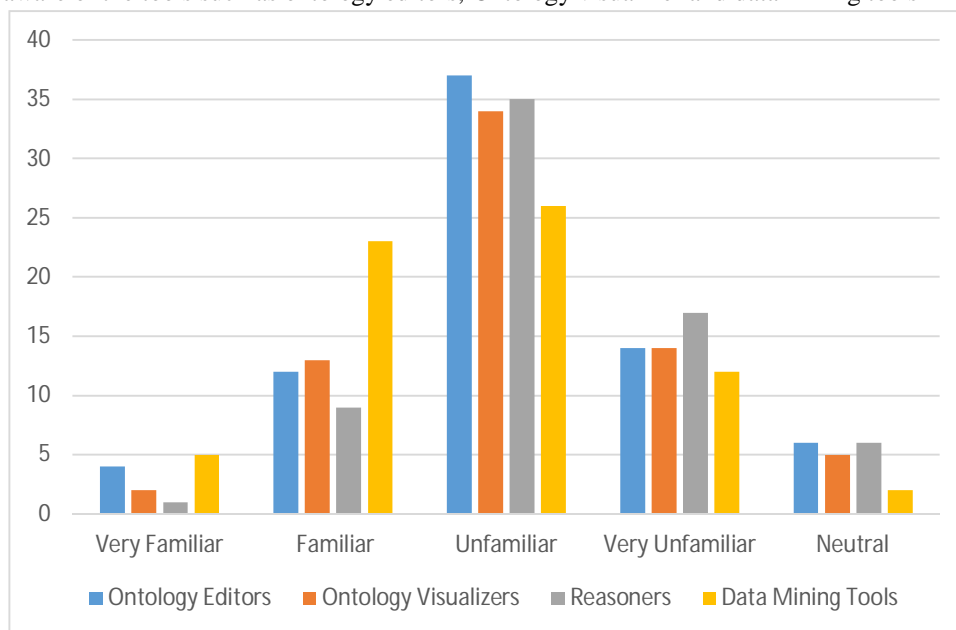

Fig.9 Familiarity with KOS Tools

Fig 10 present the level of integration was made for web 3.0 / Semantic web technology in Digital library software for Organizing, Browsing and Searching of Content. 52% RDF, 13 % SPARQL, 13% Web Ontology Language, 8 % RDFS, 6%N Triples and so forth. It was observed that RDF, Web Ontology Language, SPARQL integration was made for enhancing the features like organizing, browsing and searching in Digital Libraries
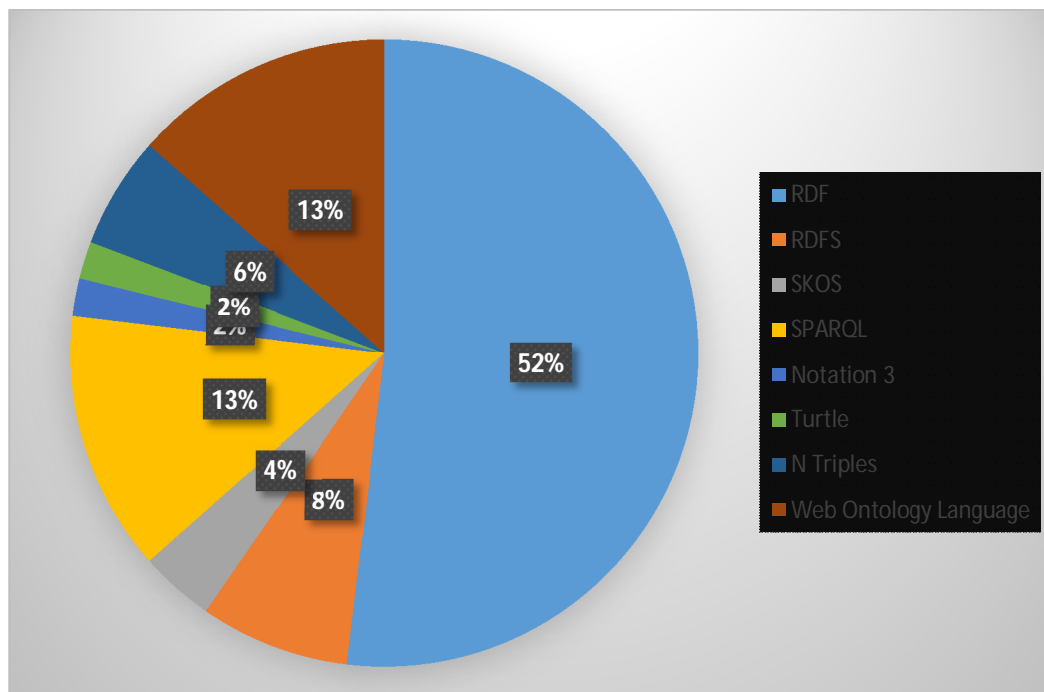
Fig.10 Integrations of Semantic Web Tools in DL's

Fig 11 responses represent that the protégé is the most used software for developing and designing domain-specific ontologies
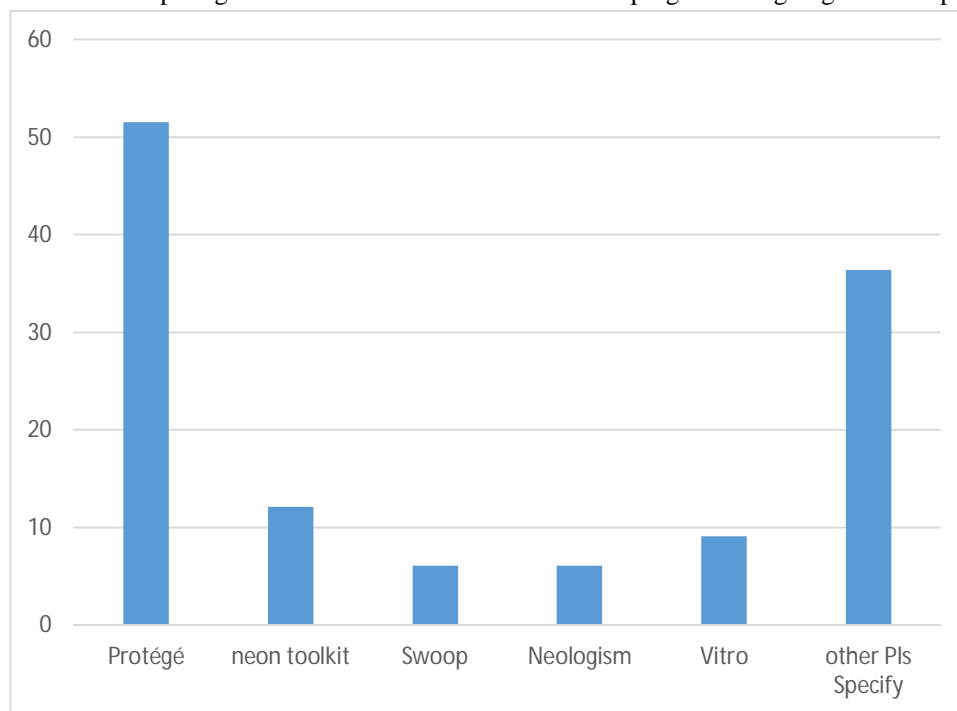

Fig.11 Use of Ontology Editor Software for Developing Domain Specific Ontologies

Fig 12 reveals the responses by respondents towards the list of feature is not available in the digital library. It was observed from the responses that, 30 the highest number of response was received for Semantic Search and Ability to link and tag documents, 28 responses for Context-based search, 27 responses for Ontology-Based Browsing, 26 responses for Ontology for Organization of Content and so forth.
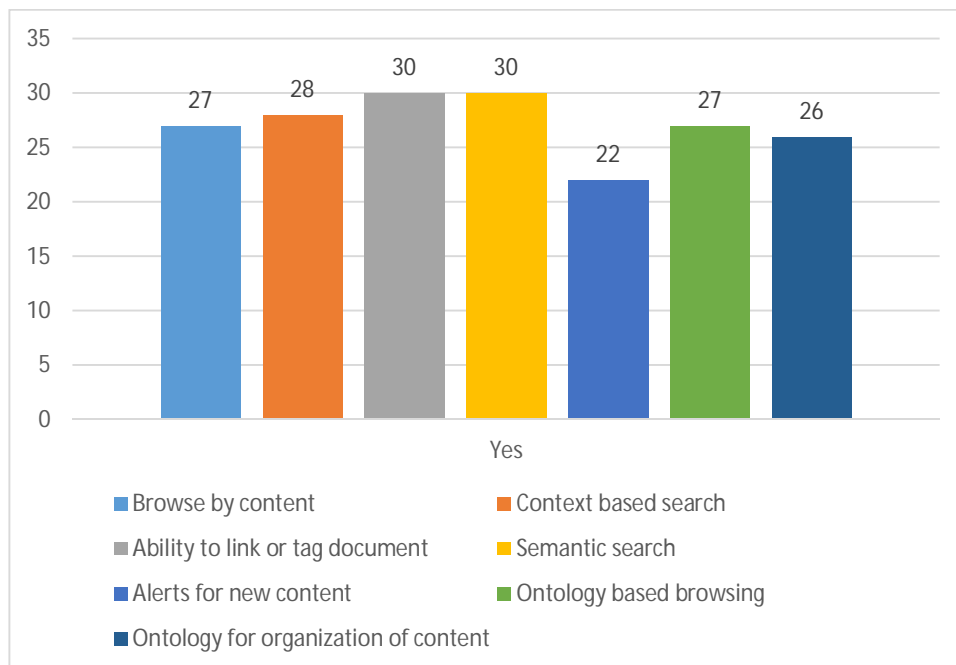
Fig.12 List of the feature is not available in the digital library

It was also asked the respondents about which search technique would help in better retrieval efficiency if integrated into Digital Library Software. Fig 13 represents that, 41% respondents Ontology-Based search, 38% RDF, 16 % SKOS and 5 % opined for others.
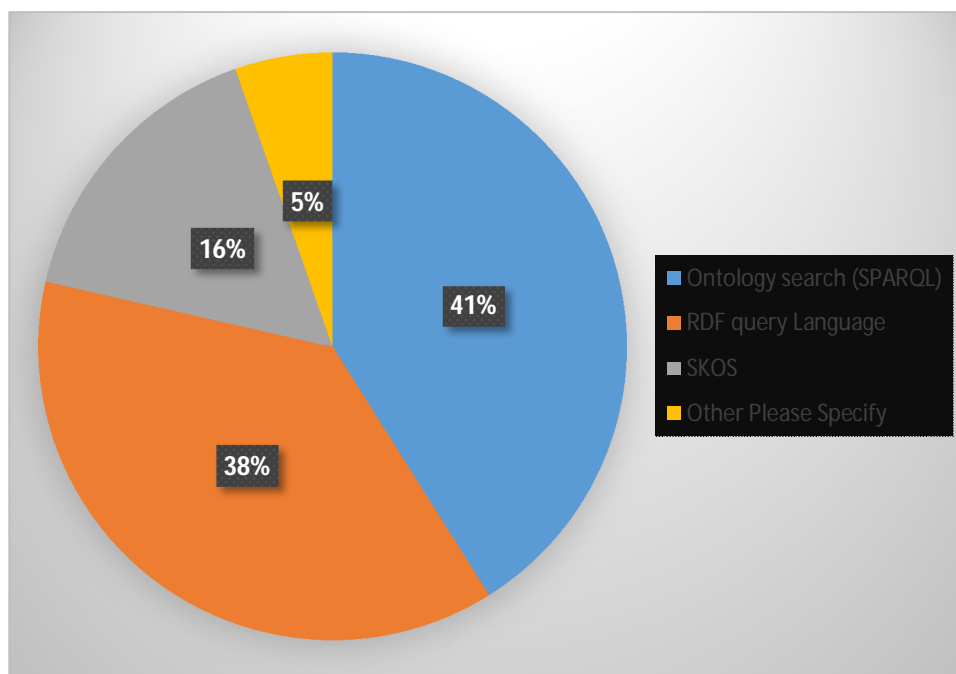

Fig.13 Search Technique for Enhanced Retrieval Efficiency

The study also revealed when it was asked to respondents that, which digital library software would you like to see the ontology-based search technique. As given above in Fig 1 it was observed that most of the libraries use DSpace as the software, however in Fig 14, 70 plus respondents have responded for Dspace as the software they would like to see the semantic service as a part of.
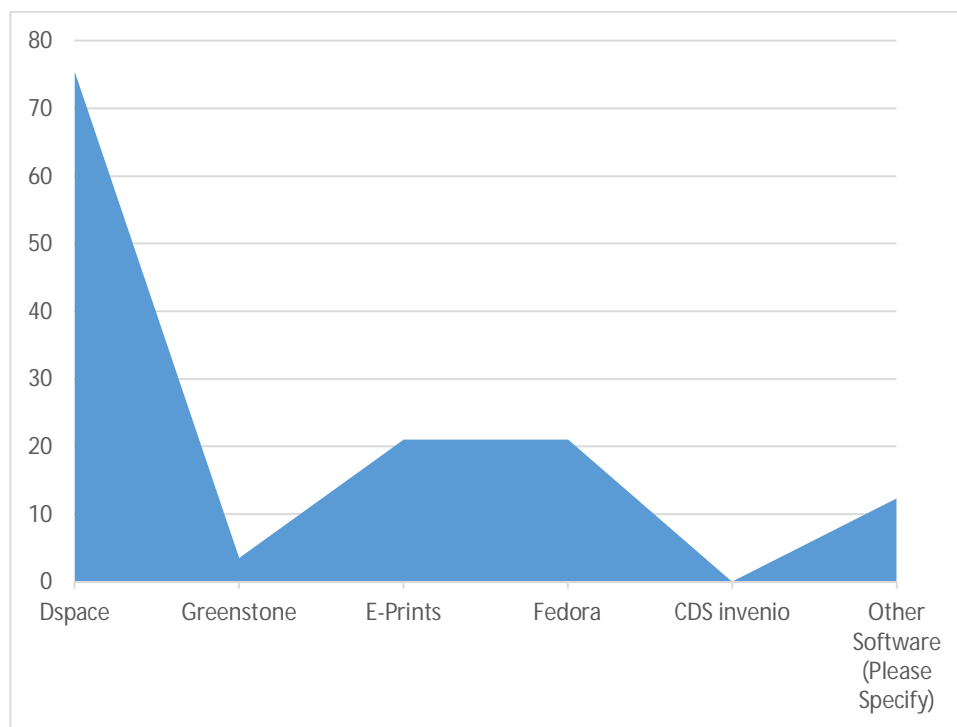
Fig.14 Semantic Service as a part of DL Software

VII.   FINDING & DISCUSSION

The quantitative data analyzed in this study had established that digital data repositories are widely used platform to gather information and this had created a huge niche for searching for relevant domain specific data, which creates a huge opportunity for developing domain-specific ontologies. Ontology is fundamentally a common vocabulary for the scholars or researchers aiming for sharing information in a particular domain.

While searching in digital libraries most of the respondents have faced problems in retrieving the precise information, it was also observed that majority of respondents found no match for queries while searching in DL's. Few faced the problems such as vocabulary mismatch, more irrelevant hits, Fluctuation in document accessibility, less user friendly with less intuitive interface and so forth. They also indicated that vocabulary mismatch and irrelevant hits are creating problems while searching in Digital Libraries. It was also observed that due to lack of support of knowledge organization techniques and retrieval tools such as ontology, RDF, SKOS, Thesaurus based searching and organization it is becoming difficult for managing and monitoring the Digital Libraries by the administrators/Librarians. Based on the study the major problems faced by the administrator/ librarians indicated towards feature of retrieving precise information from Digital Libraries.

Based on the facts observed while analyzing data. The study revealed for the potential need for domain specific ontologies and semantic technologies in Digital Libraries. The knowledge about the integration of ontologies can boost the Organization, Browsing and Search in DL's. Semantic web technologies play important role in developing organizing and retrieving of the digital content. Advance application of indexing and searching can be integrated to improve the search efficiency in Digital Libraries. Features of Digital Libraries can be enhanced by making use of semantic web technologies such as faceted information retrieval, achieving consistency in metadata creation, developing domain-specific ontologies, linked data services, ontology-based searching for enhancing information retrieval. It was also observed that most of the administrators tried to implement the semantic web tools/technology such has RDF, RDFS, SKOS, SPARQL, Notation 3, Turtle, N Triples, Web ontology language. However, few of them were unable to implement the semantic technology due to lack of knowledge or newness with the semantic web concept.

The recommendation/suggestions for the Digital Library administrators/librarians are as given below.

A.   To enhance the effectiveness of Digital Library software the developers can choose and integrate the semantic web technology.
B.   The DL's should display the hierarchies of concepts while searching in the DL's.
C.   The DL's should also provide the recommendation while searching which will help users in finding information precisely.

VIII.   CONCLUSION

Digital Libraries play an important role in every field of subject, but the increase in data and information have the conceptual and vocabulary problems which users face when knowledge discovery /searching digital libraries in the specific domain lack accurate information retrieval system. The present study was designed to access the potential applications of semantic web technologies in DL's for enhancing the Digital Library retrieval effectiveness. And to find out the difficulties faced by the administrators and librarian in using and developing the features in DL's.

Based on the potential need of ontologies/semantic web technologies in the DL's following implication can be made by Digital Library Administrators/ Librarians.

1) The administrator can make a decision of implementing semantic web technologies.
2) Implementation of Semantic technologies can help digital libraries in enhanced searching, browsing and organization of content.
3) Developing and Designing domain-specific ontologies and integration in DL's can help in enhanced retrieval efficiency of DL's.
4) Based on the findings the developers of DL's should improve the features of DL's by implementing semantic tools.

RAFFRENCES

[1]   H. Bruce, "Personal, anticipated information need," Information Research, vol. 10, no. 5, 2005.

[2]   D. P. M. A.R.D. Prasad, "Faceted infrastructure for semantic digital libraries," Library Review, vol. 57, no. 3, pp. 225-234, 2008.

[3]   S. C. G.G. Chowdhury, "An overview of the information retrieval features of twenty digital libraries," Program, vol. 34, no. 4, pp. 341-373, 2000.

[4]   Z. Z. S. H. C. S. Xiangxing Shen, "A review of the major projects constituting the China Academic Digital Library," The Electronic Library, vol. 26, no. 1, pp. 39-54, 2008.

[5]   M. A. Sanjeev K. Sunny, "Evaluating the effectiveness of thesauri in digital information retrieval systems," The Electronic Library, vol. 36, no. 1, pp. 55-70, 2018.

[6]   M. J. M. Gerard Salton, Introduction to modern information retrieval, New York: McGraw-Hill - McGraw-Hill computer science series, 1983 .

[7]   C. D. B. S. Anup Kumar Das, "Information retrieval features in Indian digital libraries: a critical appraisal"," OCLC Systems & Services: International digital library perspectives, vol. 23, no. 1, p. 92 – 104, 2007.

[8]   Saeed Rezaei Sharifabadi, "How digital libraries can support e-learning," The Electronic Library, vol. 24, no. 3, pp. 389-401, 2006.

[9]   A. &. B. G. &. B. D. &. W. I. Hinze, "Greenstone: a platform for semantic digital libraries," 2008.

[10] M. &. P. P. Nisheva-Pavlova, "Ontology-based search and document retrieval in a digital library with folk songs," Information Services and Use, vol. 31, pp. 157-166, 2011.

[11] N. &. M. D. F. Noy, "Ontology Development 101: A Guide to Creating Your First Ontology," Knowledge Systems Laboratory, vol. 32, 2001.

[12] T. R. Gruber, "Toward principles for the design of ontologies used for knowledge sharing," International Journal of Human-Computer Studies - Special issue: the role of formal ontology in the information technology, vol. 43, no. 5-6, pp. 907-928, 1995.

[13] E. a. M. G. McCulloch, "Analysis of equivalance mapping for terminology services," Journal of Information Science, vol. 34, no. 1, pp. 70-92, 2008.

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 ⊙ (24*7 Support on Whatsapp)