INTERNATIONAL JOURNAL OF ADVANCE RESEARCH, IDEAS AND INNOVATIONS IN TECHNOLOGY

ISSN: 2454-132X
Impact factor: 4.295
(Volume 4, Issue 5)
Available online at: www.ijariit.com

Ontology based digital library search system for enhanced information retrieval in engineering domain

Nilesh Shewale
shewale.nilesh2008@gmail.com
Tata Institute of Social Sciences, Mumbai, Maharashtra

Dr. J. Shivarama
j.shivarama@tiss.edu
Tata Institute of Social Sciences, Mumbai, Maharashtra

ABSTRACT

Recent years have seen an exponential increase in the amount of available information in both print medium and electronic medium. The acceptance of electronic/digital medium has increased dramatically in recent years, which had led to the demand for more organized and accessible information in the digital medium. Hence, Digital libraries (DL's) emerged as the digital counterpart of the traditional library system. Digital library environments had negated the traditional limits on representation and distribution of information, by facilitating global access, round the clock service, classification, and organization of the available information, by adding more information retrieval elements. Another challenge emerged with the advancement of technology, which had increased applications of new forms of information, such as multimedia files, scientific data, unstructured data, and semi-structured data, heterogeneous structured data, which is being generated, stored and utilized by digital libraries worldwide. There a large amount of data generated every day for different domains such as medicine, healthcare, engineering, energy, and more. As a consequence, retrieving relevant information related to any domain from heterogeneous knowledge sources had become a challenging task. Ontologies have emerged as a potential solution in this field, as they have the ability to manage information resources efficiently, manage web complexities, and automate bibliofigureic storage and annotation management and more. Ontologies play a significant role in digital libraries by promoting interoperability by defining a common vocabulary for the ease of sharing information in a particular domain. Ontologies also tackle the challenge presented by the availability of heterogeneous information sources and improve the accuracy of information retrieval. This paper discusses, the need and potential applications of domain-specific ontologies and semantic technologies in DL's, which can be used to address the issue of increasing volume of information as well as enhancing the information retrieval capacity of Digital Libraries. The paper also discusses the designing and developing of domain-specific ontologies for Digital Library and suggest a model of Ontology-based Digital Library search system which retrieves exact results for queries and eliminates irrelevant results by having refined queries.

Keywords—*Digital library, Ontology, Information retrieval, Semantic web technology, Engineering*

1. INTRODUCTION

Digital libraries (DLs) are fundamentally augmentations of physical libraries in digital form, some of the famous DLs are National Science Digital Library (NSDL) and the Internet Public Library. They are global and are available in multiple languages, acting as storehouses of information, facts, audio, and images, easily available to people across the globe facilitating users and learners (Baeza-Yates, Ribeiro-Neto, Fox, & Sornil, 1999).

1.1 Present scenario of digital libraries (DLs) and the features of Digital Library Software's

In this digital era, digital libraries (DLs) have become a vital feature in modern librarianship, which is regarded as a positive upcoming trend for libraries adopted worldwide by library professionals. High acceptance of this trend can be attributed to OSS adaptation, increasing digitization, the attraction of digital preservation and open access availability (Gonçalves, Moreira, Fox, & Watson, 2007). Baeza-Yates et al. (1999) had defined DLs as a set-up having an open architecture and supporting a wide range of dispersed and diverse digital information. Rasmussen and Edie (2004) had defined DL as an organization providing structured resources for selection that offers access to the intellectual digital property and other digital works for interpretation, distribution, integrity preservation over time in a manner that makes them available in economical and timely fashion.

Some desirable characteristics of Digital Library Software's are as follows:

- Secure Authentication and Authorization: Users and contributors have limited access to items, it include mechanisms for secure user identification and secure resource control. DL must also support multisite access, allow multiple level user authentication processes for individual users, collection groups and institutional units (Gonçalves et al., 2007).

- Support for various file formats: It is a desirable feature for DL to be able to support various file formats such as PDF, Word, HTML and more. This has vital ramifications for digital preservation owing to the complications of a document consisting several files as well as are cross-linked. It should also preserve links for videos, images, animations, audios and much more, as long as they are active (Gonçalves et al., 2007).
- Metadata harvesting: Metadata is the raw material for generating indexes and can be developed offline and uploaded on the web later or can be generated and edited online. DLs must be able to retrieve metadata from various sources and index or link them for further applications and services. Moreover, a simple form of metadata must be provided to authentic users but administrative must be taken not to reveal technical metadata elements publicly (Baeza-Yates et al., 1999).
- Support Import & Export: Import & Export are vital tools for digital libraries, for distinguishing items, collections, and communities in a simple directory format and allowing content movement between the library and other systems (Baeza-Yates et al., 1999).
- Flexible browsing facilities: DLs should provide flexible browsing, where the user can use the name of the author, date, title and such elements for browsing. It will require a broad range of browsing facility and searching interfaces to cater for flexible browsing (Tramboo, Shafi, & Gul, 2012).
- Multilingual and Multi-gigabyte support: DL software must allow processing of multiple languages to cater to the multilingual resources available and should have the capacity to manage millions of documents and up to several GBs (Tramboo et al., 2012).

1.2 Present scenario of digital libraries for Information retrieval – Properties, Challenges, and types of information retrieval systems in DLs

Information retrieval (IR) is the related to acquiring, organizing and attaining knowledge from the information. With the propagation of IR systems as well as online resources had revolutionized the concept of a library and led to the emergence of digital libraries. Information retrieval is a crucial element for the success or failure of DLs, it provides high efficiency and ease of use to diverse users. Information retrieval systems functions in diverse ways of facilitating users in their information searches (Raftopoulou et al. 2008.). There are two highly relevant properties of Information retrieval for DLs namely, word stemming and ranked retrieval. Word Stemming, is the property that involves truncating linked words with a root word so that none of the relevant information is overlooked. This property decreases the vocabulary significantly but increases total recall in DLs. Although, this property also leads to the retrieval of irrelevant information by decreasing precision (Jones, 2006). Ranked Retrieval, is another property of information retrieval system relevant for DLs, which increases the result values by changing the position or order of the results for displaying most relevant documents on the top of the retrieved list. This property is particularly important for DLs, as it promptly retrieves crucial results and increases the efficiency of the software (Jones, 2006).

With the enormous and continuous increase in the number of online information resources as well as demand for access, information retrieval field had been extended to multiple applications such as answering questions, tracking, subject detection, multimedia retrieval, data structuring, text mining and much more (Schatz, 1997). With this increase in applications, the challenges associated with have also increased, following are some of the significant challenges for information retrieval system in DLs:

- Global Information Access: Ideally, IR systems are required to understand the information required through effective interaction with the automated system that has access to structured and unstructured data in multiple languages across the web, which is a significant challenge for the designers (Jones, 2006).
- Contextual Retrieval: It is a huge challenge for the IR systems to use search technologies as well as knowledge related to user context and search query and in one framework and provide appropriate and relevant results for any user's information requirements (Lohilahti & Kai, 2009).
- Multi-lingual Complexities: It is a significant challenge for information retrieval as it is very difficult to determine the context of the required information in every search query expressed in various languages. It is very difficult to select appropriate search keywords for obtaining the required information due to ambiguities that are present in different languages (Schatz, 1997).
- Irrelevant Information Retrieval: Relevance is highly subjective property, any retrieved document can be more relevant to one user and less to another having a similar query. Thus, designing a retrieval system is always influenced by intricacies linked with relevance as well as the evaluation of the system becomes difficult on the basis of relevant documents missed by it (Schatz, 1997).

Depending on the range of searches queries and complicated retrieval algorithms there are various types of information retrieval systems in DLs, namely, Keyword Search, Boolean Search, Vector Space Model, Latent Semantic Analysis, Probabilistic Models and Language Model, which are discussed here in brief.

- **Keyword Searches:** Many of the information retrieval techniques are based on keyword searches in one way or other, it mostly analyses the collected documents for the query word and is a quite an effective technique for information retrieval. Word mismatch problem and multilingual issues are often faced with this type of IR (Allan et al., 2002).
- **Boolean Search:** This type of IR uses three basic logical operators: OR, AND, and NOT, where OR logic is used for determining similarity terms and concepts; AND for documents with multiple searches and NOT for excluding terms from search (Allan et al., 2002).
- **Vector Space Model:** This Model has three stages, namely, file indexing, term weighting, and similarity ranking. With file indexing, irrelevant words are removed by using stop list, which allows representation of significant words. The second step is weighing for obtaining both precision and relevant recall using term frequency, inverse document frequency, and length normalization. Last is the similarity ranking applied by comparison function between query and document vector (Allan et al., 2002).
- **Latent Semantic Analysis:** It is also known as latent semantic indexing (LSI), which is a statistical and mathematical technique used for deducing contextual information from within sentences. It is an automated process, which uses a document matrix showing the frequency of each word, Singular Value Decomposition (SVD) and weighting functions (Allan et al., 2002).
- **Probabilistic Model:** It is a model of retrieval estimate providing the likelihood of document is similar to the search query. It ranks document on the basis of their probability that a document will be relevant to a given query (Allan et al., 2002).

- **Language Model:** It is a natural language model applying algorithms that combine statistical and semantic information. This model index phrases instead of words and improves the precision of search (Allan et al., 2002).

1.3 Need of semantic or context-based retrieval system with special reference to digital libraries

Most of the Information Retrieval (IR) technology is based on the simple keyword search, which is limited in capabilities to acquire and use concepts for relevant and timely retrieval. With the growth of online information, DLs are faced with numerous challenges associated with accuracy, relevance, capturing data and such (Limbu, Connor, Pears, & Macdonell, 2014). In the following section, the need for context/semantic-based information retrieval is presented.

- Improving user's search experience: Exponentially increasing online information had presented a requirement of improvement in searching and retrieval method to derive precise information from the massive scale of available content. Enhancing search experience of the users is a major driving factor behind rising demand of semantic/context information retrieval systems, as they provide mechanisms to handle such large amount of data (Mittal et al., 2006).
- Handling of Multimedia Content: Traditional retrieval systems were mostly limited to only text content and documents. The novel mechanisms of semantic/context information retrieval systems enable proper handling of multimedia content as well. It exploits the semantics of user query as well as conceptualize web-content meanings for proper retrieval of multimedia content (Da, Torres, & Falcão, 2006).
- Classification of information sources: Heterogeneous structure of world-wide-web, had led to the availability of multiple sources of information such as text documents, news, e-books, blogs, e-journals, forums and more. In addition to this with the introduction of metadata format for enhancing visualization of content, had created a huge niche for semantic/context based informational systems (Mittal et al., 2006).
- Overcoming limitations of keyword-based retrieval models: The notion behind semantic search is to understand the meaning of the search query and search by meanings of the query rather than the words in the query strings. Traditional methods did not design to recover context of search, which led to the amalgamation of Information Retrieval and the Semantic Web Technologies (Da et al., 2006).
- Representation of Domain knowledge: Elaborated frameworks for semantic/context based IR models allows representation of domain knowledge through concept thesaurus, defining relationship between concepts, query indexing, document indexing and much more (Da et al., 2006).

1.4 Overview of the semantic or context-based retrieval system

Presently, information retrieval domain overcoming various information retrieval challenges by using Semantic Web methods. In recent years, with the explosive increase in Web documents basic keyword-based information retrieval models have been less effective, as different users having similar search query can have diverse goals and so would require different results. This problem is addresses contextual IR that considers users interests and preferences to deliver more precise results (Cioara, Anghel, Salomie, & Dinsoreanu, 2009). In the opinion of Karthikeyan and Dhanapal (2016) there are two main steps involved in context information retrieval, first is semantic text refining that converts available text into machine-readable representation and second is using semantic knowledge inferring by using reasoning algorithms.

Context information retrieval systems' operation is based on the meaning of the search query rather than text matching. As per En, Joemon, and Melucci (2007), context-based IR techniques are considered far better than the content-based systems and also have more meaningful filters to filter out irrelevant information. This method is far more focused on semantic matching rather than syntax matching. The basic procedure involves identifying the context of the search query and identifying suitable results matching the context of the query. For semantically matching queries, content-based similarities are identified, sentiment polarization is identified and then sentimentality is matched for context matching. Primarily, significant terms are identified and their semantic similarity is measured against a search query.

In the opinion Lopes (2009), the context-based system operates on the semantic characteristics of the text and is more real-time efficient. One context-based retrieval technique was proposed by Kalloubi, Nfaoui, and El Beqqali (2016)which was based on figure theory. In this method, the figure was created by linking concepts with named elements. Another context-based technique was suggested by Zakos and Verma (2006), that based the results on the contextual nature of the entities rather than the frequency.

2. PROBLEM STATEMENT

Generally, digital libraries are mainly regarded as repositories of digitally preserved documents, not as the repositories of knowledge. This represents a key challenge for DLs to semantically connect digital content, such that relevant information can be drawn for any search query. Search capabilities in DL are limited to the general search technology that is used by traditional search engines such as keyword search (Baeza-Yates et al., 1999). However, DLs are more complicated as they have multimedia and metadata storage that makes the conventional search capabilities quite limited. Application of domain-specific ontologies in DLs can be used to address this issue as well as tackle the issue of increasing volume of information. In addition to this, domain-based ontologies hold large promises for digital libraries, in the near future. However, there is considerable lack of research that contributes technologically as well as organizationally. Thus, this study examines the application of domain-based ontology in DLs, with focus on engineering domain.

3. OBJECTIVES

The main objective of this research study follows:

- To analyze significant requirement of context based searching for a specific domain in the contemporary environment.
- To study the context-based retrieval system for Digital Repository.
- To Suggest an Ontology-based Digital Library search system model.

Based on the above objectives following research questions were developed and answered as given below:

3.1 What is the need of semantic or context-based retrieval system of a Digital Repository?
As the majority of the Information Retrieval (IR) approaches are based on the simple keyword search method having limited capability to extract and use knowledge from the relevant document in a timely fashion. The explosion of online information and complex structure of world wide web had created numerous challenges for digital Repository systems related to accuracy, speed, relevance, extraction of data to name a few. Semantic or context-based information retrieval system is used in digital repositories firstly for enhancing end user's search experience, as the contemporary user is more drawn towards speed and accuracy during information search. Semantic or context-based information retrieval has the capability to process a large amount of data and extract relevant information from the huge chunks of data. Along with this metadata format used in the information retrieval system based on context or semantic enhances visualization of content, which had created a huge niche for such systems. Secondly, information retrieval in present times is not limited to document or text searches that were mostly employed in previously in the Digital Repository system. Contemporary digital repository system uses mechanisms of semantic or context-based information retrieval systems for enabling proper handling of multimedia content in addition to text content well. This kind of information retrieval system exploits the semantics or context of the user query for searching for relevant information as well as conceptualize retrieval relevant multimedia content. The basic ideas behind semantic or context-based search are understanding the meaning behind the search query provided by the user and searching information by meanings rather than string matching that is used traditionally. Application of semantic web technologies in digital libraries had led to recovering search context and use it for knowledge extraction. Thirdly, the need of context or semantic-based information retrieval system had arisen from the complicated and heterogeneous structure of world-wide-web as well as increasing the availability of various sources of information like, text documents, e-journals, news articles, e-books, blogs, forums and much more had created an explosion of online information, which had arisen the need of information retrieval system that can manage such huge amount of relevant and irrelevant information and extract accurate knowledge in relatively small duration of time. Lastly, the expanded frameworks of semantic or context-based information retrieval models had enabled forming concept thesaurus, defined concept relationships, indexing of query, document indexing and more. Previous studies conducted by Mittal, Krishnan, and Altman (2006), and Limbu et al. (n.d.), had widely covered the use of such information retrieval systems in a digital repository.

3.2 Is there any significant requirement of context based searching for a specific domain in the contemporary environment?
A quantitative study was conducted for this research study, the researcher had collected data from engineering educational institutes and analyzed it using regression analysis and the above question was resolved through it. Most of the librarians and administrators who participated in this study revealed that they play the multiple roles as metadata editor, repository manager or repository developer. Moreover, the majority of the librarians had responded that while searching through the digital libraries relevant information is found in the form of text document or pictures. It was also noted from the analysis that the majority of the respondents frequently used digital repositories for accessing the required information. moreover, the data analysis had revealed that digital repositories have multiple purposes uses that includes academic, R&D, searching, browsing, revision, and surfing. The frequency of use and a wide variety of uses for digital repositories had established that a vital role is played by digital repositories for fulfilling the academic requirement of the institutions. The most desirable features of digital repositories were noted to be its capability of indexing items with unique id tags and its efficient search mechanism. Although digital repositories have many desirable features one of the significant challenged faced with them is lack of effective retrieval tools and sustenance for knowledge organization for monitoring and managing content. Majority of the respondents in the survey had also admitted to getting lost while searching for information in digital repositories. Most of the respondents had revealed that while searching or browsing through digital repositories, irrelevant links and non-availability of relevant links is often encountered within the search results. Majority of the respondents were of the opinion that to advance search methods must be included in digital repositories for obtaining more relevant search results in response to the user query. This challenged can be tackled with the use of context-based information retrieval systems in addition to generally used keyword matching retrieval system.

3.3 What are the best practices of Ontology-based Digital Library search system?
Many researchers have studied the application of ontologies in a digital library environment and studied improvement in performance of the digital libraries in terms of retrieval and distribution of knowledge. Shum, Motta, and Domingue (2000) had presented ScholOnto which is a model of ontology-based Digital library server specifically developed for research documents and dissertation thesis. This model was aimed to address the challenge of information browsing, meta-data tracking, performance evaluating and contradictory concepts present in the available literature from a viewpoint of the structure. This model had supported multi-user practice and had also supported distributed user interfaces with the use of a semantic network of formal information. , Cheng et al. (2008) had presented a model for tackling the issue of manual mapping in heterogeneous ontologies, which was usually performed by experts who compared the structures as well as linguistic semantics concepts using various computer applications. This model used domain-based ontologies for mapping diverse document and web pages under similar domain ontology. Liu, Xu, and Li (2011) had presented a model that had used the hierarchy structure of web documents for constructing a domain-specific ontology. The complex structure of the World Wide Wethe b was used here for deriving knowledge from structured and semi-structured web documents and pages, available. Here, Jena was used here to develop, edit, delete and read ontology in the RDF form and SPARQL for addressing queries.

Nisheva-Pavlova and Pavlov (2012) had presented ontology-based search and document retrieval in a digital library for the folk songs (over 1000 digitized Bulgarian folk music songs). Here, search and analysis system was designed for searching music type, lyrics and notes in the available digital resources and domain ontology had aided in sematic oriented searching of the available resources. Another Ontology-based Digital Library search system was presented by Smirnov, Kovalchuk, and Dukhanov (2014), who had integrated virtual modeling and simulation environment in a digital library environment. Here, various semantic ontology models were integrated to develop a workflow design using Virtual Simulation Objects technology. This model was specifically

designed for resolving issues in the public-transportation domain. One such practice was presented by De Carvalho and Tavares (2015), who have used methodological approach for the development of Digital library reference model (DLRM), for improving information organization, sharing, and retrieval. De Carvalho and Tavares (2015) had presented another model domain ontology that had defined real-world semantics using domain-specific languages. With the use of this practice, abstract syntax can be defined for valid expressions as well as also defines the link between abstract syntax and domain-based ontology using real-world semantics.

3.4 Suggest a model on Ontology-based Digital Library search system which retrieves exact results for queries and eliminates irrelevant results by having refined queries

For answering this research question, a brief empirical review of existing models of the ontology-based digital library is presented here. Bloehdorn et al. (n.d.) presented a model applying ontology in question answering application of the digital library. This approach provides answers to the questions on the basis of heterogeneous knowledge available in various sources components of distributed ontologies in the digital library. The basic principle is to incorporate structured metadata and unstructured information resource in a unified manner, that can flexibly address queries through structured expression in the natural language. Xiaomei, Zhengdong, and Chongyang (2008) had proposed an ontology-based digital library system's application in e-learning. This model had focused on two main components of information search ontology extraction and query expansion. This model had introduced basic human-like semantic reasoning for knowledge representation which had facilitated more relevant and accurate retrievals. When compared with existing traditional models, this model had shown a 5% improvement in the performance. Martín and León (2010) had applied ontology for expert knowledge management for digital libraries, the proposed model had allowed global access to the information resources, anytime and anywhere, by using artificial intelligence and semantic web technologies. It was noted by the researcher that use of semantic technologies across distributed and heterogeneous structure of digital library had created various opportunities of automation in digital libraries. The specific construction used in this model had improved metadata recording and tracking. Swe (2011) had used ontology in digital libraries for intelligent information retrieval, which was proposed for overcoming the limitations of keyword-searching by in context-based ontology as well as in metadata format. The model proposed was to extract and expand domain concepts from the search query entered by the user. The aim behind the development of this model was to improve the accuracy and relevance of the search result using both applications as well as standard search procedure. The main advantage associated with the use of ontology was that it improved the accuracy of information retrieval in comparison to the conventional search methods used. Popa et al. (2017) had used domain-specific ontologies in medicine, for creating a digital library for improving health literacy. The proposed model consisted of relevant documents from a range of medical domain that was used to provide information to patients regarding their symptoms and conditions. Under this model, information was organized in a structured manner using the unstructured data obtained from various internet articles. In this model, multiple sub-ontologies were used to develop meta-ontologies. After reviewing the existing models, the researcher had identified that all the existing model had improved upon the traditional information retrieval methods by providing more accurate and reliable results. However, it can be noted that none of the available models were able to prune irrelevant search result associated with the search query. In the opinion of the researcher, use of advanced semantic web application in information retrieval, information indexing, and information storage can improve the accuracy and reliability of information retrieval systems as compared to contemporary ones.

3.5 The concept of domain-specific ontology and semantic technologies in DLs

Ontology is a method used for problem-solving, that have reasoning behavior and uses artificial intelligence for sharing and reusing the knowledge. It uses natural language processing, knowledge management, and intelligent information integration, for providing an understanding of any particular domain (Kalibatiene & Vasilecas, 2010). Currently, ontologies are used in multiple areas such as artificial intelligence, information-based systems and more, for organizing information. Ontologies provide an explicit depiction of information that is machine-understandable that increases semantic interoperability. However, developing ontologies is a tedious process but it serves as both user and application interface in DLs. A domain-based ontology or domain-specific ontology, only includes the concepts belong to that particular domain such as engineering, medical science, politics and much more (Wolff, Nelson, & Ebecken, 2014). In the opinion of (Bloehdorn et al. (n.d.), domain ontologies in DL, is a solution that can overcome the problem of irrelevant retrieval to a certain extent. It can be used for organizing bibliofigureic descriptions, representing contents of the document and sharing knowledge between the users. By using domain ontologies in DLs, allows profile transfer, user's browsing pattern to various other digital libraries and databases, so that preferences and browsing behavior of the user can be used by another software application. Domain-based ontologies enhance the performance of the DL by improving content indexing, aids in the development of new information improve representation and retrieval of information in digital environments. In addition to this, automatic contextualized results, integration of information, the automatic or semi-automatic document selection; forming automatic inferences, linguistic and semantic compatibility are some of the potential benefits for the DL (Santos, Campos, & Braga Villela, 2009).

Amongst the major challenges faced while retrieving information from the World Wide Web is the determination of most relevant information on the basis of the search query. Generally, probabilistic calculation and frequency referential are used for obtaining information by search engines. It indicates towards the requirement of more semantic oriented representation of the information on the internet. A major advantage of semantic-based search approach on the Ontologies is that it will provide an outline form searching information objects as well as increase accuracy of the search process and will provide more relevant results. This approach aids in enhancing automation of the process by providing more machine-readable information on the internet (Cioara et al., 2009; Kalibatiene & Vasilecas, 2010).

Semantic technologies are more concentrated on representing knowledge through ontologies, it clearly defines concepts and objects that already exist as well as define the relationship between objects. As ontologies are a proper representation of information in any specific domain, semantic technologies can be used to update and search query by both human user and computer. Sematic technologies in ontology provide a set of rules, that draws knowledge from an organized order of information, providing a semantic meaning (Mittal, Krishnan, & Altman, 2006). Semantic technologies in DLs increase interoperability between different kind of

digital content across the web, facilitates in searching and retrieving information from diverse disseminated resources, enhance the ability to store, index and retrieve text, non-textual as well as multimedia content and aids user in browsing large information repositories. However, building and managing ontologies using semantic technologies is a daunting task, as it requires a large number of resources to enhance information retrieval system. In addition to this balancing relevance and semantics is also a challenge for information retrieval in DL.

3.6 Designing and developing domain-specific ontology for digital libraries

Developing and designing ontology is distinct for every practitioners or developer, some can convert simple lexicons to controlled vocabularies to organized thesauri or convert vocabularies to hierarchies/taxonomies and then convert it to fully developed ontologies having distinct properties. The basic rules for designing domain-specific ontology are complex as there is no set way to model a domain ontology. Consequently, ontology development is essentially regarded as a repetitive process (Kalibatiene & Vasilecas, 2010). In the opinion of Wolff, Nelson, and Ebecken (2014)desirable criteria for designing of domain-based is that it should be open, dynamic, scalable, interoperable, context independent and easily maintained. Hughes, Crichton, and Mattmann (n.d.) had defined three important stages in developing domain-specific ontology namely, capture of concepts and relationships (documentation of key concepts and relationships between concepts and definitions of concepts and relationships), ontology encoding using tools and languages (selecting representation language, writing code and defining basic terms) and implementation (integration, evaluation and phases guidelines). A simple depiction of developing ontology is provided in figure 1 and some of the basic designing steps are as follows (E. De Carvalho & Tavares, 2015; Wolff et al., 2014):

- Determining the specific domain and extent of the particular ontology
- Extracting concepts and selection
- Reusing elements from existing ontologies
- Tally the essential terms for the ontology
- Context building
- Defining object-classes
- Defining class hierarchy and their properties
- Define the features of each class slots
- Creation of instances

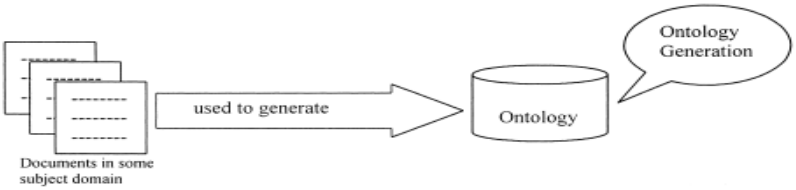

Fig. 1: Error! No text of specified style in document.**: Ontology Regeneration**

Source: (Ding & Foo, 2002)

3.7 Pros and cons of using of domain specific ontology in digital libraries

The major pros of using Domain Specific Ontology in Digital Libraries are as follows:

- Ease of sharing structured information: Domain-based ontologies contain information that can be easily shared amongst the user and software agents, having more common goals. It allows the user to extract information and amass information from various websites, which can be used in question answering and other applications in DLs (Liu, Xu, & Li, 2011).
- Ease of Reuse: Reuse of existing domain knowledge is one of the major driving factors behind the increased application of domain-based ontology. Diverse domain models are based on the same concept of time, so while developing a new domain-specific ontology, already developed detailed domain ontology can be reused. Many of the large domain ontology is the integration of several existing domain ontologies (De Carvalho et al. 2011).
- Interoperability: Domain-based ontology is explicitly committed to a system but supports communication and collaboration between systems that are developed at various websites (Kalibatiene & Vasilecas, 2010).
- Separating domain knowledge from operational knowledge: It is one of the major reasons behind the increasing use of ontologies. Domain-specific ontologies separate components and characteristics and then apply the configuration algorithm (Kalibatiene & Vasilecas, 2010).
- Model-based knowledge acquisition: Domain-specific ontologies have explicit information validation and knowledge verification protocol that aids in acquiring knowledge in a structured manner. By making explicit assumptions in the domain, makes it easier to implement changes when domain knowledge changes. In addition to this, model-based acquisition and analysis aids in reusing and extending ontologies (Wolff et al., 2014).

The major cons of using Domain Specific Ontology in Digital Libraries are as follows:

- Inconsistency and inflexibility: Boundary concepts are used in the development of ad-hoc ontologies that leads to inconsistency and inflexibility in ontologies (Mukhopadhyay, Biswas, & Sinha, 2007).
- Difficult implementation of ontological concepts: The designing of the user interface and code development uses the concept of time, which requires finite coordinates during implementation. This leads to conflicts in geometric computations, topological computation and more (Kalibatiene & Vasilecas, 2010).

4. RESEARCH GAP

After performing the study following gap in the research study were identified by the researcher.

- Existing research studies had not emphasized on the need for formal standards for integrating the ontological application in digital libraries for improved information retrieval.

- Limited studies have been carried out for multimedia content stored in the digital library, and the semantic relation between the multimedia content and text content is not explored completely.
- Existing studies have failed to establish whether or not context-based searching is needed for a specific domain in Digital Libraries.

5. CONCLUSION

For every research study there are certain inherent limitations, this study aimed to analyze the integration of ontologies in the digital library system for improved information retrieval. This subject area was wide-ranging so the researcher had to select some specific areas as the focus of this study. The methodology selected for the purpose of developing an ontology-based digital library system may not have been the most suitable one as there are various tools and procedures available for developing domain-specific ontologies. Since there is a lack of available ontologies for engineering domain, the presented model can be improved upon by using several other approaches. The development of domain-specific ontologies in the field of engineering integrates is an upcoming field, which opens up a huge field of approaches used for representing relationships between concepts, continuously improvements representation of knowledge, categorizing and indexing domain activities, as well as conceptualizing extracted knowledge for information resources. The future studies can take up a re-examination of other digital library initiatives for understanding the patterns more suitable for application of domain-specific ontology in this field. Furthermore, studies can be conducted for diversifying the application of domain-based ontologies in digital libraries to improve consistency and cohesiveness of the platform. In addition to this, new functionalities, as well as the requirements of the system, can be further evaluated by adding the appropriate description of the ontology framework in digital libraries.

6. REFERENCES

[1] Allan, J., Aslam, J., Belkin, N., Buckley, C., Callan, J., Croft, B., … Zhai, C. (2002). *Challenges in Information Retrieval and Language Modeling Report of a Workshop held at the Center for Intelligent Information Retrieval*.

[2] Baeza-Yates, R., Ribeiro-Neto, B., Fox, E. A., & Sornil, O. (1999). Digital Libraries.

[3] Balaji, P., Sarangi, A. K., & Madalli, D. (2012). Knowledge Organization Systems for Semantic Digital Libraries. *International Conference on Trends in Knowledge and Information Dynamics*, *2*, 10–13.

[4] Baohua, W., Xiaoyan, M., & Fei, G. (n.d.). On the Characteristics of the Digital Library and the Influence to the Work of Reader Service.

[5] Bia, A., & Nieto, A. (n.d.). Information Retrieval in Digital Libraries: efficient catalog searches using tries.

[6] Bloehdorn, S., Cimiano, P., Duke, A., Haase, P., Heizmann, J., Thurlow, I., & Völker, J. (n.d.). Ontology-based Question Answering for Digital Libraries.

[7] Bygstad, B., Ghinea, G., & Klaeboe, G.-T. (2009). ORGANISATIONAL CHALLENGES OF THE SEMANTIC WEB IN DIGITAL LIBRARIES. A NORWEGIAN CASE STUDY.

[8] Castro, L. J. G., Giraldo, O. X., & Castro, A. G. (2010). Using the Annotation Ontology in Semantic Digital Libraries. *ISWC*.

[9] Cheng, C. P., Lau, G. T., Pan, J., Law, K. H., & Jones, A. (2008). Domain-Specific Ontology Mapping by Corpus-Based Semantic Similarity. *NSF CMMI Engineering Research and Innovation Conference, Knoxville, Tennesse*.

[10] Cioara, T., Anghel, I., Salomie, I., & Dinsoreanu, M. (2009). A Context ± based Semantically Enhanced Information Retrieval Model. *IEEE*.

[11] Da, R., Torres, S., & Falcão, A. X. (2006). Content-Based Image Retrieval: Theory and Applications.

[12] De Carvalho, E., & Tavares, G. M. (2015). *Ontologies for Digital libraries*.

[13] De Carvalho, V. A., Paulo, J., Almeida, A., & Guizzardi, G. (2011). Using Reference Domain Ontologies to Define the Real-World Semantics of Domain-Specific Languages, (1).

[14] Ding, Y., & Foo, S. (2002). Ontology research and development. Part 1 -a review of ontology generation. *Journal of Information Science*, *28*(123). https://doi.org/10.1177/016555150202800204

[15] En, B.-L., Joemon, D., & Melucci, J. M. (2007). Context-Based Information Retrieval.

[16] Gonçalves, M. A., Moreira, B. L., Fox, E. A., & Watson, L. T. (2007). What is a good digital library?'' – A quality model for digital libraries. https://doi.org/10.1016/j.ipm.2006.11.010

[17] Janakiraman, A., Ormsby, R., & Subramanian, N. (n.d.). The Role of Information Communication Technology (ICT) in Library and Information Science (LIS) Careers in India.

[18] Jisc. (n.d.). About OpenDAOR - OpenDoar.

[19] Jones, K. S. (2006). Information retrieval and digital libraries: lessons of research.

[20] Kalibatiene, D., & Vasilecas, O. (2010). Ontology-Based Application for Domain Rules Development. *Scientific Papers, University of La*, *756*.

[21] Kalloubi, F., Nfaoui, E. H., & El beqqali, O. (2016). Microblog semantic context retrieval system based on linked open data and graph-based theory. *Expert Systems with Applications*, *53*, 138–148. https://doi.org/10.1016/J.ESWA.2016.01.020

[22] Karthikeyan, N., & Dhanapal, R. (2016). REAL-TIME CONTEXT BASED INFORMATION RETRIEVAL USING SEMANTIC AND FEATURES ON STREAMING CONTENT. *IJRET: International Journal of Research in Engineering and Technology*, *5*(12).

[23] Limbu, D. K., Connor, A. M., Pears, R., & Macdonell, S. G. (n.d.). Improving Web Search Using Contextual Retrieval.

[24] Liu, X., Xu, J., & Li, F. (2011). Domain-Specific Ontology Construction from Hierarchy Web Documents. In *2011 Seventh International Conference on Semantics, Knowledge, and Grids* (pp. 160–163). IEEE. https://doi.org/10.1109/SKG.2011.21

[25] Lohilahti, L., & Kai, L. (2009). Context-based multimedia information retrieval. *DTU Library*.

[26] Lopes, C. T. (n.d.). Context Features and their use in Information Retrieval.

[27] Mahesh, G., & Mittal, R. (2008). Digital Libraries in India: A Review. *National Institute of Science Communication and Information Resources, New Delhi, India*.

[28] Martín, A., & León, C. (2010). Expert Knowledge Management based on Ontology in a Digital Library. (PDF Download Available).
[29] Mittal, A., Krishnan, P. V, & Altman, E. (2006). Content Classification and Context-Based Retrieval System for E-Learning. *Educational Technology & Society*, *9*(1), 349–358.
[30] Mukhopadhyay, D., Biswas, A., & Sinha, S. (2007). A New Approach to Design Domain-Specific Ontology-Based Web Crawler. https://doi.org/10.1109/ICIT.2007.20
[31] Mya, T., & Swe, M. (2011). INTELLIGENT INFORMATION RETRIEVAL WITHIN DIGITAL LIBRARY USING DOMAIN ONTOLOGY, 363–373. https://doi.org/10.5121/csit.2011.1232
[32] Nisheva-Pavlova, M., & Pavlov, P. (2012). Ontology-based search and document retrieval in a digital library with folk songs1. *Information Services & Use*, *31*(3–4), 157–166. https://doi.org/10.3233/ISU-2012-0645
[33] Orduna-Malea, E., & Delgado LLpez-CCzar, E. (2017). Academic libraries and open access resources in Latin America. https://doi.org/10.2139/ssrn.2758115
[34] Popa, R., Vasilateanu, A., Goga, N., Doncescu, A., Darminio, P., Raluca-Maria, B., & Catalin, I. (2017). Ontologies applied in medicine: A digital library creation framework for health literacy. *International Conference on E-Health and Bioengineering*.
[35] Raftopoulou, P., Petrakis, E. G. M., Tryfonopoulos, C., & Weikum, G. (n.d.). Information Retrieval and Filtering over Self-organising Digital Libraries.
[36] Rasmussen, E., & Edie. (2004). Information Retrieval Challenges for Digital Libraries. In *Proceedings of the 7th international Conference on Digital Libraries: international collaboration and cross-fertilization* (pp. 95–103). Springer-Verlag. https://doi.org/10.1007/978-3-540-30544-6_10
[37] Santos, N., Campos, F. C. A., & Braga Villela, R. M. M. (2009). Digital Libraries and Ontology. *IGI Globa*.
[38] Schatz, B. R. (1997). Information Retrieval in Digital Libraries: Bringing Search to the Net, *275*.
[39] Shalini R. Urs, A. M. A. (2000). Ontology-based Knowledge Organization Systems in Digital Libraries: A Comparison of Experiments in OWL and KAON Ontologies., 1–17.
[40] Shum, S. B., Motta, E., & Domingue, J. (2000). ScholOnto: An Ontology-Based Digital Library Server for Research Documents and Discourse. *International. Journal on Digital Libraries*, *3*(3).
[41] Smirnov, P. A., Kovalchuk, S. V, & Dukhanov, A. V. (2014). Domain Ontologies Integration for Virtual Modelling and Simulation Environments. *Procedia - Procedia Computer Science*, *29*, 2507–2514. https://doi.org/10.1016/j.procs.2014.05.234
[42] Tella, A. (2016). *Information seeking behavior and challenges in digital libraries*.
[43] Tramboo, S., Shafi, S. M., & Gul, S. (2012). A Study on the Open Source Digital Library Software's: Special Reference to DSpace, EPrints, and Greenstone. *International Journal of Computer Applications*, *59*(16), 975–8887.
[44] Virkus, S. (2010). Characteristics of digital libraries.
[45] Wolff, I., Nelson, G. /, & Ebecken, F. F. (2014). A Simple Strategy to Start Domain Ontology from Scratch. *IJACSA) International Journal of Advanced Computer Science and Applications, Special Issue on Natural Language Processing*.
[46] Xiaomei, X., Zhengdong, N., & Chongyang, S. (2008). An Ontology-Based Digital Library System for E-Learning. In *2008 International Conference on Information Management, Innovation Management and Industrial Engineering* (pp. 344–347). IEEE. https://doi.org/10.1109/ICIII.2008.224
[47] Xu, X., Zhang, F., & Niu, Z. (2008). An Ontology-Based Query System for Digital Libraries. In *2008 IEEE Pacific-Asia Workshop on Computational Intelligence and Industrial Application* (pp. 222–226). IEEE. https://doi.org/10.1109/PACIIA.2008.360
[48] Zakos, J., & Verma, B. (2006). A Novel Context-based Technique for Web Information Retrieval. *World Wide Web*, *9*(4), 485–503. https://doi.org/10.1007/s11280-006-0223-y

