

Ontology Based Knowledge Management Systems and Its Application to Semantic Digital Library Initiative

Ujjal Marjit

Arup Sarkar

Abstract

The use of digital library in this era advances the speed of information access. In this regard the demand for the digitization of the paper based information into the digital format is evolving. Most of the time this requirements are fulfilled by developing digital library which are capable to hold and store a huge amount of digital data either separately or in archive. The present digital library system is facing the problem of incomplete/partial automation, resource sharing operation integration, interoperability due to heterogeneous data, information and knowledge for information discovery. Adding semantics to the digital library can alleviate the same. So, a proper Knowledge Management System based on Ontology is a key requirement. In this paper a conceptual framework is proposed for better Knowledge Management for the digital library using Ontologies. In addition to that this paper discusses about the necessity of constructing ontology based knowledge management in digital libraries using semantics.

Keywords: Semantic Digital Library, Knowledge Management, Ontology

1. Introduction

Nature of information is changing every day as well as their use. The method to store information, data using papers is old now. Preservation of data in digitized format is a preferred choice today. People are interested in this method not because it is the current trend and heard more fashionable, but it offers several advantages over the traditional one. It is easy to preserve a paper full of data in digital format. It can be stored for an infinite time, no further care to the digital object is required as it is required in case of the books, papers etc. Infinite number of copies can be produced from a single digital object by just a single click of a mouse on the computer screen. They can be accessed from any place in the world through a distributed environment uncountable times within an unbelievable fraction of time by the virtue of high-speed Internet access. So, naturally the importance of digital library is becoming higher than the earlier times. In general a digital library holds more than thousand copies of distinct digital objects. Besides a proper manipulation of these items an efficient searching mechanism is needed for them. In addition to that interoperability issues among the digital objects are very much essential. The aim of this paper is to develop a framework for better discovery of a particular type of digital object as per the user submitted goal based on semantic matchmaking after a proper classification of the documents. In this work ontologies will play a vital role to describe the working domain of digital world as well as each document or digital object as an instance member of that domain. Providentially, the rising ontology-based KMSs can find the content-oriented knowledge that people really want due to the fact that the domain ontology is powerful in knowledge representation and associated inference. Ontologies are meant to provide an understanding of the static domain knowledge that facilitates knowledge retrieval, store, sharing,

and dissemination. For KMSs, ontology can be regarded as the classification of knowledge^[1]. According to Gruber^[2], Ontology is an explicit specification of a conceptualization. That is to say, ontology defines shared vocabulary for facilitating knowledge communication, storing, searching and sharing in knowledge management systems. Semantic Digital Libraries provides the opportunity to expand the usefulness of digital libraries which will probably contain the world's majority of data in the future. Semantics makes unstructured data, particularly, a truly machine usable, machine navigable resource.

This paper is organized as follows: In Section 2, a literature review of the same topics has been discussed here. Section 3 depicts the motivation for this work. A suggested semantic monitoring framework that is ontology based knowledge management is overviewed in Section 4. Section 5 concludes this paper with future research works.

2. Literature Review

Digital library management has passed a long way in the developed world. But it is still a very new and sometimes an undiscovered approach in certain region of the developing world like India. The term "Digital Library" (DL) is used to refer to a range of systems, from digital object and metadata repositories, reference-linking systems, archives, and content management systems to complex systems that integrate advanced digital library services and support for research and practice communities. A DL may offer many technology-enabled functions and services that support users, both as information producers and as information users. Adding semantics to the digital library provides the meaningful information and data that may be accessed without any human interaction. T. Krishna and et.al. Proposed a cost effective web enabled organization document repository framework comprising of Digital Library and related application^[3]. They used open source tools for implementation of a prototype of their model which gives an inexpensive way to introduce Digital Library based document repository system using effective knowledge management. A research work published by Li he and et.al. where they proposed a conceptual model of knowledge management in digital library^[4]. In their model for knowledge management in digital library, they divided it into three levels as "i) knowledge oriented management based on business and process, ii) human oriented management integrated by service and management and iii) ability oriented management with the object of organization innovation."

Another very interesting research work is done by Lifan Guo and et.al^[5]. They introduced an ontology based visualization model to fulfill the purpose of historical knowledge searching by an effective knowledge management with the help of knowledge representation and inference. They have used Apache Lucene^[6] to backup the need of their search models due to its efficient and high performance based text search-engine. In this work they proposed different search models named as – tree navigation, keyword search, ontology relationship search, ontology property search and backtracking search. An in-depth discussion about the necessity of knowledge management in University Digital Library can be found in Linghui Guo's work^[7]. A general framework of knowledge management oriented University digital library is also given here. Abou-Zeid (2003) proposed an ontology engineering process model based on the premise that ontology development is a special application of Nonaka's

knowledge creation model. Nonaka's model views the ontology engineering process as a spiral in which interaction between ontology stakeholders' tacit knowledge and explicit knowledge is continuous and dynamic. Sure^[8] presented a number of semantic web methods and technologies and showed their applicability to practical knowledge management in corporate intranets and in the web. Quan et al^[9] investigated the knowledge management system (Haystack) built on RDF to create, organize, and visualize personal knowledge such as e-mails, documents, tasks, contacts, meetings, and other information. Even though previous studies on ontology-based knowledge management have proven to be active and productive, few studies have as yet discussed an integrated ontology-based knowledge management system focusing not only on searches of web resources and company documents but also on support for identifying employees with particular skills and appropriate company products.

3. Motivation

Use of web is increasing everyday with the need of high speed information access. E-Commerce based business process is also booming. In these circumstances a centralized or distributed environment is required where an efficient framework will be present to handle all the necessary data. So business authorities feel the need for a web based information archival system to store the huge amount of their business data. Besides this Universities and different Institutions also required an advance library system to store their huge amount of books and Research Thesis. At this stage digital library is considered as an important solution where documents or information get stored in the digitized format in centralized or distributive manner. Some researchers also introduce other frameworks for better information management. Besides this our study shows that ontology based knowledge management could be useful for better information search and retrieval. In case of semantic web, ontologies are used to describe a particular domain and all the existing resources there, which are beneficial to find out accurate information as per the user's need. The same concept is also applicable for the existing digital library. In this work the main aim is to develop a framework to make the transition from digital library to semantic digital library. Knowledge discovery is a major issue in our present digital library system. Semantic digital libraries offer expanded facilities for knowledge discovery, data mining of semi-structured text, and mechanisms for linking and searching related concepts.

4. Framework of Ontology-Based Knowledge Management System (OKMS)

The main objective of the proposed architecture is as follows: i) Storage of digital object within a Web Server to prepare the digital library. ii) Classification of documents (digital objects) as per type, content and the time of creation. iii) Preparation of domain ontology to describe the working domain with semantics. iv) Making the digital library more machine and user friendly. v) Development of ontologies to describe the each digital object semantically. vi) Better information discovery (in the format of digital object) as per the user's requirements.

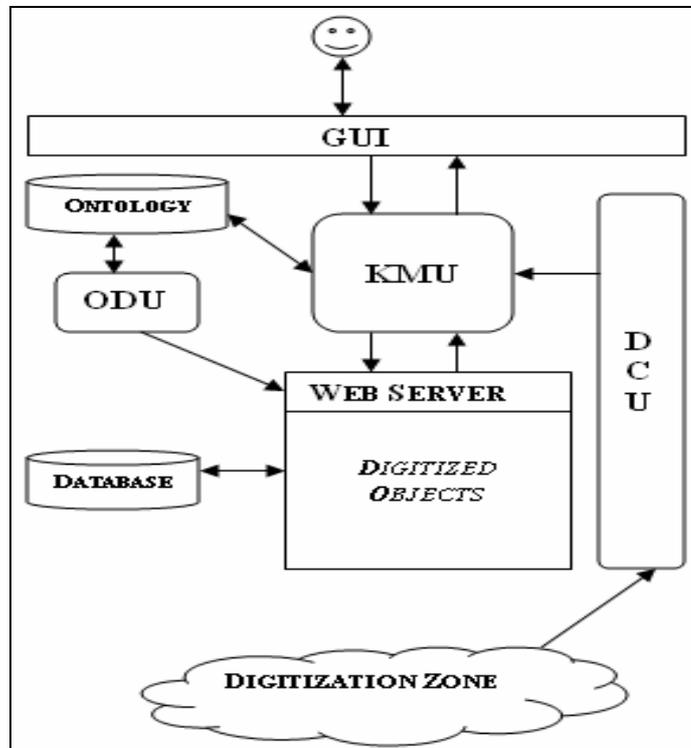


Figure1: Conceptual model of proposed framework for ontology based Semantic Digital Library

Figure-1 describes the proposed ontology based knowledge management unit for digital library. The different units of the proposed architecture are as follows:

End User: In the discussed environment the end user could be a human, a machine client, or an software application or agent searching for certain information or documents.

GUI: In this work GUI could be a simple html based interactive graphical interface or it could be an agent based unit for interactive services for end user as well as it could be useful to represent the user submitted request as a ontological goal description using semantics. The necessary user authentication and their request validation also could be performed here. But a separate module for security measures is always preferable.

Ontology Development Unit (ODU): This unit will be used to develop all the necessary ontological description of the domain and the digital documents. Though in this paper a specific mechanism or algorithm to develop the Ontologies automatically, is not included. All the developed Ontologies will be stored into the repository dedicated for Ontologies, could be identified in the figure1 as a cylinder labeled as Ontology.

Document Classification Unit (DCU): According to our view the classification of digital objects or documents is required. It could be very beneficial to narrow down the range of the objects to search by identifying the nature of the request submitted by the end users. At present only three main categories is prescribed. First one is the category by type or format of the documents like PDF, MS Word, ODF file, FLV, JPEG etc. Second categorization is according to their contents i.e. verifying that what kind of information a particular object holds confidential, governmental, private or public etc. The third categorization will be done by using their original time of creation, their historic value and importance. It obvious that there could be more and more categories and sub-categories are possible if further assessment is done.

Knowledge Management Unit (KMU): This is another and most important part or unit of this architecture. It would be used to handle all incoming search request through the GUI to find out particular information or a document holding the expected information from the digital library. This unit will also perform the matchmaking operation between user submitted requests (Onlogical description of the request) and the ontological description of the existing digital objects within the digital library. If no match found, it will be notified to the end users. Otherwise matched documents will be handover to the end users.

Web Server: This is the place where all the digitized objects will be kept after their proper classification. An external database could be used to develop the all the necessary records securely and persistently.

Digitization Zone: It is the place where all the digitized objects come from. Naturally it can not be a single point or place. That's why it shown as cloud. It could be online station where paper based documents are getting scanned and then submitted to the DCU. It could be a simple home user who is uploading already digitized object from his personal collection to increase the strength of the digital library, or it could be an academician who wants to store some copies of important research papers into the digital library for his students. So in a single sentence, this architecture could be useful by handling ontology based knowledge management to discover the pre-reserved knowledge from a digital library in the form of digitized documents or objects.

5. Conclusion & Future Research Directions

In this work a conceptual framework is proposed to introduce the ontology based knowledge management in the existing digital libraries which will help to transform any standard digital library into the Semantic Digital Library. A dedicated knowledge management unit is proposed here for the efficient and effective knowledge management within the digital library. It will also handle the user requests for digital documents or information in digital format. But the main job of this module is to perform search operation to find out and retrieve the accurate information as per the user's requirement based on semantic matching between user request and ontological description of working digital libraries and their contents. A document classification technique is also proposed here to narrow down the range of the searching digital objects.

No implemented prototype is introduced here since this work is still at conceptual and experimental level. Another major drawback of this framework is no security mechanism is introduced here. Though some researchers suggest the use of open source security tool to take the security measure.

But an implemented and enhanced security mechanism within the framework is always preferable. The following future works and research challenges are foreseen such as extension of ontology towards covering additional knowledge elements of the domain, providing fresh metadata, semantic matching and fuzzy approaches.

Our future work will consider the security requirement for the framework. At this stage no proper ontology development methodology is described. The modules also need an enhance design with proper user interface for handling all the maintenance related issue with the administrative privilege.

References

1. N. Guarino, Understanding, Building, and Sing Ontologies: A Commentary to Using Explicit Ontologies in KBS Development, *International Journal of Human and Computer Studies*, 46:293-310, 1997.
2. GRUBER, T" Toward Principles for the Design of Ontologies used for Knowledge Sharing" *International Journal of Human and Computer Studies*, 46:293-310, 1997.
3. KRISHNA Tanay, THAKUR Rakesh Kumar, KUMAR Deepak. (2006). Cost effective Document Repository Management. 1st International Conference on Digital Information Management, 2006 IEEE. 6 Dec. 2006. (pp. 344-350).
4. HE Li, XIN Guan, CHANGAN Shao. (2008). The Research on Knowledge Management in Digital Library. *IEEE International Symposium on IT in Medicine and Education*, 2008. ITME 2008. 12-14 Dec. 2008. (pp. 354-359).
5. GUO Lifan, ZHOU Yigang. (2009). Ontology Based Digital library Search Model Research. *IEEE International Conference on Granular Computing*, 2009, GRC '09. 17-19 Aug. 2009. (pp.184-187).
6. The Apache Software Foundation. Apache Lucene. Available at <http://lucene.apache.org/> (Accessed on 5/11/2010)
7. GUO Linghui. (2010). On Construction of Knowledge Management in University Digital Libraries. 3rd IEEE Explore. pp525-529.
8. ABOU-ZEID, E. S. (2003). What can Ontologists Learn from Knowledge Management? *Journal of Computer Information Systems*, 43(3), 109-117.
9. QUAN, D., HUYNH, D., & KARGER, D. R. (2003). Haystack: A Platform for Authoring End User Semantic Web Applications. *Proceeding of The International Semantic Web Conference* (pp.738-753)

About Authors

Mr. Ujjal Marjit, System-in-Charge, C.I.R.M, University of Kalyani.

Mr. Arup Sarkar, Research Scholar, Department of Computer Science & Engineering, University of Kalyani.