Ontology-Based Fuzzy Retrieval for Digital Library

Tho Thanh Quan[1], Siu Cheung Hui[2], and Tru Hoang Cao[1]

[1] Faculty of Computer Science and Engineering
Hochiminh City University of Technology
Hochiminh City, Vietnam
qttho@cse.hcmut.edu.vn
[2] School of Computer Engineering
Nanyang Technological University
Singapore

Abstract. With the recent advancement of the Semantic Web, researchers are now considering developing ontology-based digital librarires for the sake of efficient information sharing, exchanging and retrieval. In addition, fuzzy queries have been also introduced to help readers to specify their queries more precisely when searching information in digital librarires. In this paper, we first propose an architecture that enables multiple digital libraries to collaborate in the Semantic Web environment. Then we discuss using fuzzy ontology to represent uncertain information in digital libraries and fuzzy queries for retrieving information from fuzzy ontology. An illustrative system is then developed for experiment purpose. Performance of our system is also evaluated and analyzed.

1 Introduction

Digital library is an organized repository of recorded knowledge which can be accessed in a digital and networked environment [1]. With the recent advancement of Semantic Web [2], there is much research considering developing digital library systems on the Semantic Web enviroment [3]. In addition, to improve the accuracy of information retrieval, some digital libraries also adopt fuzzy-based retrieval techniques [5]. In this paper, we propose an ontology-based technique for fuzzy document retrieval on digital libraries. The rest of this paper is organized as follows. Section 2 discusses a proposed Semantic Web-based architecture for exchanging and retrieving information cross multiple digital libraries. Section 3 presents a formal definitions of fuzzy ontology and fuzzy query. Section 4 provides performance evaluation of a experimental system. Finally, Section 5 concludes the paper.

2 Ontology-Based Digital Libraries on the Semantic Web

In this section, we propose an architecture supporting exchanging and retrieving information from multiple ontology-based digital libraries on the Semantic Web environment, given in Figure 1. In our architecture, the scholarly knowledge is not merely stored in a single scholarly database as typically designed in classic digital

D.H.-L. Goh et al. (Eds.): ICADL 2007, LNCS 4822, pp. 95–98, 2007.

libraries. Instead, the scholarly knowledge will be represented using ontological formalism as Scholarly Ontology. It enables scholarly knowledge stored in various Scholarly Ontologies can be shared among multiple digital libraries.

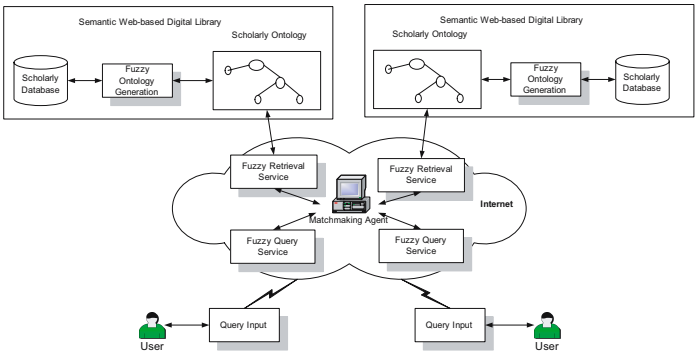

Fig. 1. Ontological-based Digital Libraries on the Semantic Web

3 Fuzzy Ontology and Fuzzy Query

The Semantic Web adopts ontology as standard for knowledge representation. Adopting from [7] , we formally define fuzzy ontology and fuzzy query as follows:

Definition 1 (Fuzzy Ontology). A *fuzzy ontology* F_O consists of 4 elements (C,A^C,R, X), where C represents a set of concepts; A^C represents a collection of attributes sets, one for each concept; $R = (R_T, R_N)$ represents a set of relationships, which consists of 2 elements: R_N is a set of *non-taxonomy relationships* and R_T is a set of *taxonomy* relationships. Each attribute value of an object or relationship instance is associated with a fuzzy *membership value* between [0,1] implying the uncertainty degree of this attribute value or relationship. X is a set of axioms.

Definition 2 (Fuzzy Query). A *fuzzy query* on a fuzzy ontology (C,A^C,R, X) is the fuzzy set $Q_f = \varphi(A')$ where $A' \subseteq A^C$.

To measure the relevance between an ontogical class and a query, we consider both of them as two fuzzy sets. Then we calculate the similarities between the fuzzy sets using fuzzy logic [8].

4 Performance Evaluation

In this section, we introduce an experimental system that makes use of Scholarly Ontology to support fuzzy scholarly retrieval. The Scholarly Ontology is generated from a citation database built on a set of 1400 scientific documents downloaded from the Institute for Scientific Information's (ISI) website[1], as depicted in Figure 2.

[1] http://www.isinet.com

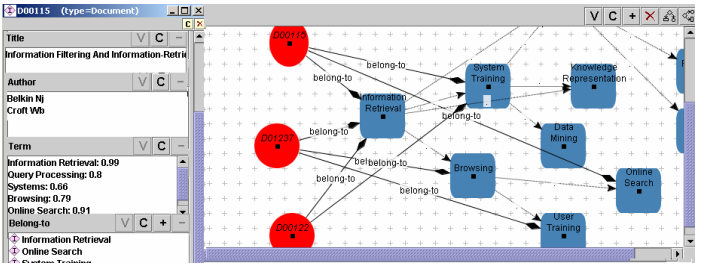

Fig. 2. Browsing documents in the Scholarly Ontology

Fig. 3. Fuzzy retrieval result

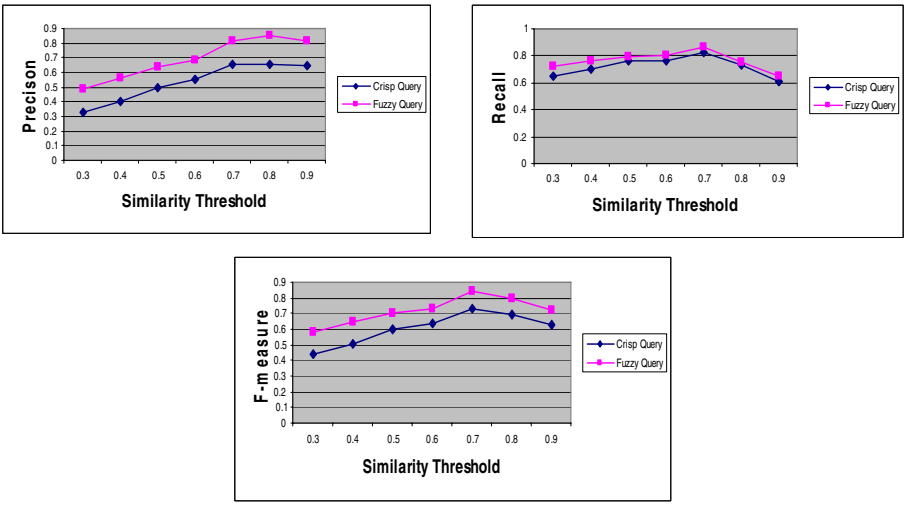

Fig. 4. (a) Performance evaluation based on precision. (b) Performance evaluation based on recall. (c) Performance evaluation based on F-measure.

We have evaluated the performance of fuzzy retrieval in our system. Fuzzy queries are formed based on an interface developed to allow user to specify the fuzzy membership of each query term. In addition, the retrieval resuls are also displayed the similarities between the queries and the retrieved documents as shown in Figure 3.

Figures 4(a), 4(b) and 4(c) give the performance evaluation based on precision, recall and F-measure on both the crisp queries and fuzzy queries. Similar to other keyword-based retrieval systems, crisp queries can obtain quite good performance in terms of recall. However, the precision on crisp queries is not that good. On the other hand, fuzzy queries have achieved better recall, but the improvement is not significant in comparison with crisp queries. However, the precision on fuzzy queries has achieved quite good improvement as compared to crisp queries. Finally, the F-measure on fuzzy queries is also much better than that of crisp queries as well.

5 Conclusion

This paper has proposed two extensions for the current generation of digital librarires. First, we propose to use ontology to represent scholarly information in digital libraries, thus making the libraries enable to share and exchange knowledge in the Semantic Web environment. Second, fuzzy theory is employed to process uncertain scholarly information as the forms of fuzzy ontology and fuzzy queries. A general architecture of digital libraries in the Semantic Web environment has been presented and an experimental system has also been developed to verify our ideas and techniques.

References

[1] Saracevic, T., Dalbello, M.: A Survey of Digital Library Education. Proceedings of the American Society for Information Science and Technology 38, 209–223 (2001)
[2] Berners-Lee T., Hendler J. and Lassila O., The Semantic Web, Scientific American, (2001), Available at: http://www.sciam.com/2001/0501issue/0501berners-lee.html
[3] Naing, M.-M, Lim, E.-P., Chiang, R.H.-L.: Core: A Search and Browsing Tool for Semantic Instances of Web Sites. In: Zhang, Y., Tanaka, K., Yu, J.X., Wang, S., Li, M. (eds.) APWeb 2005. LNCS, vol. 3399, Springer, Heidelberg (2005)
[4] Lim, E.P., Sun, A.: Web Mining - The Ontology Approach. In: Proceedings of The International Advanced Digital Library Conference (IADLC 2005), Nagoya, Japan (August 2005)
[5] IntraText Digital Library, available at: http://www.intratext.com/CERCA/Aiuto.htm
[6] Girill, T.R., Luk, C.H.: Fuzzy Matching as a Retrieval Enabling Technique for Digital Libraries. In: The Digital Revolution: Proceedings of the American Society for Information Science Mid-Year Meeting, San Diego, CA, pp. 139–145. Information Today, Inc. (May 1996)
[7] Quan, T.T., Hui, S.C, Fong, A.C.M., Cao, T.H.: Automatic fuzzy ontology generation for Semantic Web. IEEE Transactions on Knowledge and Data Engineering 18(6), 842–856 (2006)
[8] Zadeh, L.A.: Fuzzy Logic and Approximate Reasoning. Synthese 30, 407–428 (1975)