

Review

Personal Discovery Fuzzy Ontology and Automatic Construction Retrievals

Been-Chian Chien*; Chih-Hung Hu; Ming-Yi Ju*Department of Computer Science and Information Engineering, National University of Tainan, Taiwan, R. O. C.****Corresponding author****Been-Chian Chien***Department of Computer Science and Information Engineering, National University of Tainan, Taiwan, R. O. C.***Article information****Received:** March 30th, 2024; **Revised:** April 22nd, 2024; **Accepted:** May 21st, 2024; **Published:** June 10th, 2024**Cite this article**

Chien B-C, Hu C-H, Ju M-Y. Personal discovery fuzzy ontology and automatic construction retrievals. 2024; 2(2).

doi: <https://doi.org/10.70705/ppp.ir.2024.v02.i02.pp46-48>**ABSTRACT**

The term “ontology” refers to a common framework for understanding a body of information. One obvious approach to improving search results for users’ requested relevant information is to include ontology into information retrieval. Here, we provide a paradigm that integrates ontologies built automatically from searched material with information retrieval systems. General semantic ontology, domain knowledge, and an automatically created fuzzy concept hierarchy make up the system’s ontology. We used ImageCLEF’s evaluation approach to estimate the system. Results from experiments show that information retrieval performance may be enhanced with the integrated ontology.

Keywords

Information retrieval; Fuzzy ontology; Concept hierarchy.

INTRODUCTION

Since 1940, the issue of data storage and retrieval has garnered more and more attention. We have a lot of data that needs to be saved, and it has to be retrieved quickly and correctly, according to the issue. There has been a lot of effort to create intelligent retrieval systems that are quick. In regards to the digital library study topic, it is certain that several digital libraries suffer from issues related to cataloging and textual searching when it comes to storing. On the other hand, efficient retrieval is still a major challenge. Research into information retrieval systems aims to improve their efficacy. Defining the user’s ideas or goals is the hardest part of looking for relevant articles in a big data set. In most cases, when asked broad, generic questions, an information retrieval system will not be able to provide detailed, helpful replies. Therefore, retrieval of relevant documents is heavily dependent on the user’s cognition and domain knowledge of documents. There have been several recent proposals for study into merging ontology and information retrieval methods for the semantic web [3, 5, 10, 14]. Integrating ontology into information retrieval entails using knowledge structures to sift through data and find what users are looking for, was required. Search results are more likely to be relevant if they address concepts inside the user’s knowledge area. This suggests that the efficacy might be enhanced. The issues in incorporating domain expertise into an information retrieval system are:

1. What is the appropriated information retrieval model?
2. How to generate and construct ontology?
3. How to find the relevant documents by ontology?

This work explores models for information retrieval based on ontologies and talks about how various kinds of ontologies, such as fuzzy ideas generated automatically from stored texts and generic domain knowledge, affect and impact these models.

The structure of this article is as follows: Section 2 reviews the relevant literature. In Section 3, we see the paradigm for information retrieval based on ontologies that has been suggested. Section 4 outlines the processing technique and the retrieval mechanism. The experiments and their results are presented in Section 5. Lastly, Section 6 presents the conclusion.

2. Related work

Information retrieval systems have been the target of several proposals aimed at enhancing their performance. It was in 1983 that G. Salton put out his now-famous vector space model for IR [11]. A document or query is represented in the vector space model as an array of {0, 1} elements. In a query or document, a value of 0 indicates that the associated term does not exist, whereas a value of 1 indicates that it does. You may also substitute the 0s and 1s in the vector with the weights you’ve allocated to the terms. In documents



or queries, the weight typically shows how often or how important a phrase is. One well-liked and effective strategy is the TF-IDF (Term Frequency and Inverse Document Frequency) technique, which relates document weights to term frequencies [12].

Data retrieval systems that rely on predetermined bodies of domain knowledge are known as ontology-based information retrieval systems. Ontology often includes notions, relationships, and definitions. The ontology provides a definition of the vocabulary of a subject area, including its fundamental concepts and relations [7].

There are essentially two branches to the literature on ontology-based information retrieval systems. Using the manually built ontology to improve the performance of an information retrieval system is the first aspect. A large number of researchers have created ontologies as a model for representing domain knowledge in order to build information retrieval systems [3][14]. In order to extract semantic concepts from either an existing ontology or one that has been explicitly constructed, these systems use queries. Finding the appropriate idea is as simple as connecting the query's keywords to the domain ontology. In their online query engine, Rodger et al. used the manually developed domain ontology [10]. By relating words to ontology ideas, the inquiries are deepened.

The second part of the task is to automatically build the ontology using the domain corpus. To improve the system's accuracy and coverage, the built ontology is used. A technique for building domain ontologies from Chinese corpora was developed by Wu and Hsu [16]. The linkages and domain keywords make up the domain ontology. The rules for extracting Chinese sentences were also specified by them. After compiling a list of domain keywords, they may use these criteria to determine the relationships between them. With the new indexing formalism in mind, Kang suggested a method that takes document ideas into account [5]. Clusters of words with strong semantic relationships are used to extract these concepts. These ideas were used to derive indexes that were precise in terms of semantics. A document's semantic content may be represented by these indexes.

3. Discussion and conclusion

In order to make information retrieval more successful, we suggested a model that is based on ontologies. The suggested model incorporates a fuzzy taxonomy as its ontology, which is automatically created from the texts. Additionally, one of the ImageCLEF benchmarks was utilized to develop and test the system.

The findings of this study demonstrate that using just a generic semantic ontology in technique TF-IDF₀₂ does not provide satisfactory retrieval results. Experiments Onto₀₁, Onto₀₂, Onto_{M1}, Onto_{M2}, and Onto_{M3} also showed that various ontologies enhance certain themes. The significance of ontology selection in information retrieval is amply shown by the experimental findings. Our results do not show that broad ontologies, such as WordNet, can provide a significant improvement. A proper and appropriate ontology selection from the document database, however, may significantly enhance the retrieval efficacy.

We know from the previous discussion that correct ontology is essential for an information system to get effective retrieval results. Nonetheless, the strategy for choosing appropriate queries into to

the suitable domain knowledge should be considered before the information retrieval task. The researchers are encouraged to develop creative methods for such a problem.

REFERENCES

- [1] The work of B. C. Chien, C. H. Hu, and S. J. Hsu was published in the 2004 IEEE International Conference on Systems, Man & Cybernetics proceedings, which took place in The Hague, Netherlands, from October 10th to the 13th. The paper is located on pages 3128-3133.
- Cross Language Evaluation Forum, accessed online at <http://ir.shef.ac.uk/imageclef2004/>, is referenced as [2].
- the third Improving the retrieval of information from external sources was published in Behavior Research Methods, Instruments, and Computers, Vol. 23, No. 2, 1991, pp. by S. Dumais. spaces 229–236.
- [4] "Semantic similarity based on corpus statistics and lexical taxonomy," paper presented by J. Jiang and D. Conrath at the 1997 International Conference on Research in Computational Linguistics in Taiwan.
- [5] "Exploiting Concept Clusters for Content-based information retrieval," published in 2005 in the Information Sciences - Informatics and Computer Science: An International Journal, was written by B. Y. Kang, D. W. Kim, and S. J. Lee. The article can be found on pages 443-462.
- Pages. 235-312 in G. Miller's 1990 article "WordNet: An on-line lexical database" in the International Journal of Lexicography, volume 3, issue 4.
- [7] "Enabling technology for knowledge sharing," AI Magazine, Vol. 12, No. 3, 1991, pp., an article by R. Neches and colleague number 36–56.
8. The Porter Stemming Algorithm by M. F. Porter, 2006, <http://www.tartarus.org/~martin/PorterStemmer>.
- TT Quan, SC Hui, and TH Cao (1999), "FOGA: A roceedings of the 2004 Symposium on Scholarly Semantic Web, "Fuzzy Ontology Generation Framework for The 2004 September ECML/PKDD Workshop on Knowledge Discovery and Ontologies was held in Pisa, Italy.
- In their 2001 article "A smart web query method for semantic retrieval of web data," published in Data and Knowledge Engineering, H. L. Roger, C. E. H. Chua, and V. C. Storey discuss a web data retrieval system. The article has pages 63–84.
- In 1983, McGraw-Hill published an introduction to modern information retrieval by G. Salton.
- [12] In "Term-weighting approaches in automatic text retrieval,"

G. Salton and C. Buckley discuss term weighting in their 1988 article published in *Information Processing and Management: an International Journal*, volume 24, issue 5, pages 513-523.

[13] University of Tokyo, Department of Computer Science, 2005, Tsujii laboratory, SS Tagger, a part-of-speech tagger for English, <http://www-tsujii.is.s.u-tokyo.ac.jp/~tsuruoka/postagger/>.

In the proceedings of the 16th International Conference on Industrial & Engineering Applications of Artificial Intelligence and Expert Systems, held in Loughborough, UK, June 23-26, 2003, P.

Varga, T. Mészáros, Cs. Dezsényi, and T. P. Dobrowiecki presented “An Ontology-based Information Retrieval System.”.

“Verb Semantics and Lexical Selection” by Z. Wu and M. Palmer was published in 1994 in the Proceedings of the 32nd Annual Meeting of the Association for Computational Linguistics, which was held in Las Cruces, New Mexico.

pp. 1313–1317 in SOAT: A Semi-Automatic Domain Ontology Acquisition Tool From Chinese Corpus, by S. H. Wu and W. L. Hsu, published in 2002 in Taipei, Taiwan, at the 19th International Conference on Computational Linguistics.