

Augmented Systems in Ontology Based Intelligent Information Retrieval

Ravindra Pratap Singh, Poonam Yadav

Abstract: In order to solve the problem of information overload, current information retrieval systems need to be improved. Intelligence should be embedded to search systems to manage effectively search, retrieval, Filtering and presenting relevant information. This can be done by ontology based information retrieval techniques. Systems are evaluated by using their information retrieval features. Ontologies are introduced to describe the semantic information of knowledge leases in order to meet requirement of utmost sharing and reuse of the domain knowledge. The paper provides an overview of ontology-based information retrieval techniques and software tools currently available as prototypes or commercial products.

Keywords: Ontologies, Information retrieval, Semantic indexing, Semantic mapping.



1. ONTOLOGY AND INFORMATION RETRIEVAL

JING-YAN et al. described that characteristics of an ideal information retrieval systems are that searching is fleet and result is accurate. Traditional methods of information retrieval can be divided into three essential kinds: text retrieval, data retrieval and topic retrieval as described in [1].

Text retrieval system: The user's information retrieval demand is represented as keywords which are compared with each word of whole text in text retrieval system. Its searching method is based on word-frequency analysis, and typical system is Google. The excellences of this kind of information retrieval system are includes a large amount of output results and super automatic search control. Its shortcomings are that return information is accumulated and reality is too low.

Data retrieval system aims at the structural information system, retrieval demand and information data all stands by certain formats, and is provided with the decided structure. It permits appointed field search, and various kinds of commercial service database all adopt this searching mode. The quality of data retrieval depends on code organization, and may cost a lot of money. The output result is relative accurate, but can not generate all expectancy information.

Topic retrieval system collects information by manpower manner or semi-automatic manner. It scans accessing document by dedicated method, and append to some beforehand determinate topic. User can visit downwards gradually from one layer to next layer, so finally can get retrieval result. The representative system is Yahoo.

The excellence of topic retrieval system is that user can easily master searching process. Ontology is the formalized description of share conceptual model to the information resource, so any conceptual object C can be defined as:

$$C = \{D, W, R\}$$

Where, D indicates such domain knowledge, W is a status set of correlative thing in applied domain, R is a set of concept relation in domain space $\{D, W\}$.

It has two effects that ontology is used in information retrieval.

(1) Automatic analysis to the domain attribute of document.

The information retrieval system looks for keywords according to a certain means in searched information document; these keywords can be used in document classification which is dependent on ontology knowledge. So information retrieval system can sort documents to some possible domains combining with primary content of information document, then filtrates off irrelevant domains, gets final result of correlative domains which the document is classed to.

Moreover, it can be make keyword matching based on the approximate semantic network of ontology, automatically class correlative documents to a certain domain. In this way, information retrieval system searches some selective documents which are correlative to user searching demand, and do no longer scan all information documents. So searching time and cost can be compressed greatly and user's semantic relativity judgment may be expedited.

(2) Intelligent formulation and visualization to user's searching demand.

For the searching keywords given by user, we can effectively estimate feasible domain depending on ontology

knowledge. Then information retrieval system enumerate consistent concept and definition of the domain to user, user can make judgment and selection. On the one hand, it can help user confirm their knowledge demand, user can more distinctly visualize information demand which is not clearly presented or little does user think. On the other hand, it can let information retrieval system ascertain the position of searching keyword in ontology, so help machine to understand user's searching intention, and provide more accurate and more relevant information.

Because of ontology defines essential kernel concept and their semantic relation, it can be used as the tools of resource description and knowledge presentation.

2. INTELLIGENT AGENT TECHNOLOGY

At present, Agent widely has applied in the fields of information retrieval and service, electronic commerce, distributed computing and so on. Yi Xiao et al. described Agent is a concept developed in the field of the artificial intelligence (AI), is a hot spot in AI the recent years. Generally believed that, Agent is the intelligent entity which can run independently and continuously under the environment of isomerism computation, and has the properties as follows as described in [2].

- (1) Autonomy - Agent is able to run continuously without the disturbance from outside or nobody, and can control its own behavior and internal condition.
- (2) Adaptability - Agent can sense the external environment, and carries on the self- adjustment according to the change of external environment, therefore has the ability of adapting user's work way and the characters.
- (3) Learning - Agent has the ability of adjusting and revising own behavior by utilizing the information of the external environment.
- (4) Cooperation - Agents can communicate each other by some kind of Agent communication language (for example KQML or the FIPA-ACL language), and mutually cooperate to complete a big task.
- (5) Mobility - Agents are able to move in the network. These properties of Agent may be represented by the basic model of Agent system in Fig.1.

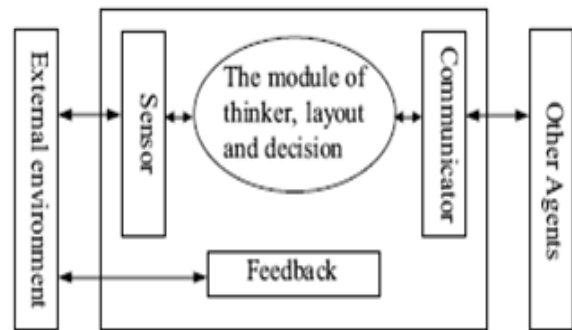


Fig.1 [2]: The basic model of Agent system

The basic model of Agent system generally, there are three most basic properties of Agent such as the autonomy, learning and the cooperation. A basic Agent system should has one of these properties, a advanced Agent system must have the autonomy and one of other two items, but an ideal Agent system must realize all of these properties and introduce other characteristics (for example mobility according to the actual demand, etc).

3. CG INTELLIGENT INFORMATION RETRIEVAL SYSTEM ARCHITECTURE

Users input keywords which they want to search, then retrieval system return the matching document to users. Because of synonyms and polysemy, it's very difficult to understand user exact needs by keyword. Often the initial keywords do not get the results they want, the mildly relevant or irrelevant documents were also retrieved. Pan Ying et al described that CG introduces how to use the computer to generate, process and display graphics. CG is very broad in its content. It is difficult for learners to understand the whole system of disciplines as described in [3].

Inquiry system is B/S model, Java program which invokes Jena API functions to realize the complex Semantic inquiries. Jena is a Java framework for building Semantic Web applications. It provides a programmatic environment for RDF (Resource Description Framework), RDFS (RDF Schema) and OWL, SPARQL (Simple Protocol and RDF Query Language), and includes a rule-based inference engine. Jena can be used to create and manipulate RDF graphs. The system mainly includes: the user retrieval interface, the retrieval application, Jena Ontology analytical tools, and the CG ontology. The System architecture is shown in fig.2.

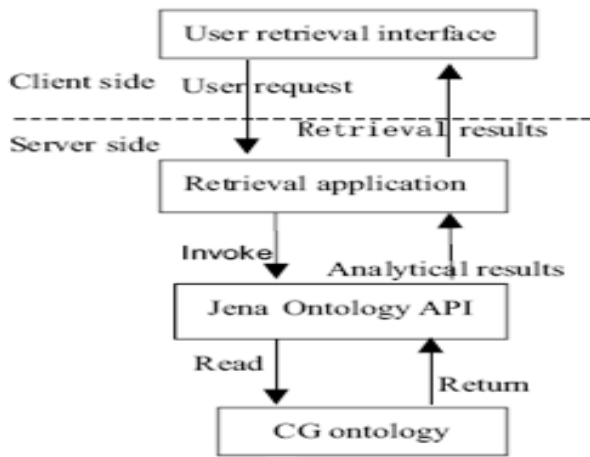


Fig 2[3]: CG Information Retrieval System architecture

On the client side, users input queries in the retrieval interface. On the server side, retrieval application accepts the request of users, and then invokes Jena API to analysis CG ontology and Forms search results, the final results are returned to the users at last.

4. INFORMATION RETRIEVAL AGENT SYSTEM BASED ON KEYWORDS DISTRIBUTION

Jae-Woo LEE described that for efficient information retrieval, it is important that keywords are defined very well as appropriate terms about all documents in the Internet or distributed computing systems. And there are three kinds of algorithms, retrieval algorithms, filtering algorithm and indexing algorithms. Retrieval algorithms are extracting important or meaningful information from database, knowledgebase or documents. Those are classified to sequential scanning and searching indexed text files. Filtering algorithms are simplifying result of text or information for efficient information retrieval. Indexing algorithms are constructing data structure for searching information exactly. Indexing is one of the most important data for efficient information retrieval. For automated indexing there are various algorithms, stemming, stop list, thesaurus, etc as discussed in [4].

Stemming is the process of matching morphological term variants for using general terms. Stemming is for reducing the size if index files and improving information retrieval system's performance.

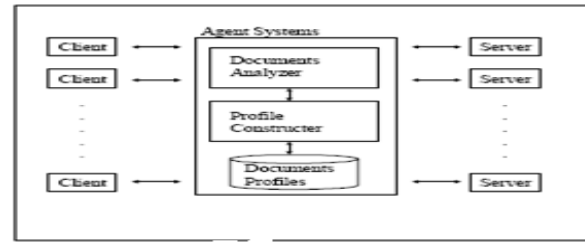


Fig. 3[4] A model for information retrieval agent system

The model is composed of two agents systems, the one is about analyzing documents of server systems and the other is constructing profiles for documents browsing. As the analyzing documents agent, for key paragraph extraction we use meaningful term's frequency and the key word distribution characteristics in a document, and those terms are selected by using stemming, filtering stop-lists, synonym for search meaningful terms in a document. The agent receives a web client's information retrieval request and extracts key paragraph with frequency and distribution using the keywords of the client. And constructing profiles agent constructs profile of the documents with the keywords, key paragraph, and address of the document browsing. And then many clients can search many documents or knowledge easily using the profile for information retrieval and browse the document.

CONCLUSION

There are some significant attempt to merge information retrieval and ontological models. The integration of agent technology and ontology is a significant impact on the effective use of the web services. Machine learning Techniques will be used for automatically extracting semantic information and text categorization. The evaluation methodology based on using standard test collections assumes that the relevance can be approximated by typical similarity of the query to document and hence user is not required to make relevance assessment about retrieved documents. The three main challenges in utilizing ontologies in information retrieval are

1. To make sure that irrelevant concept will not be kept, and that relevant concepts will not be discarded. Descriptions of documents in information retrieval are supposed to reflect the content of documents and establish the foundation for the retrieval of information when requested by users.
2. To map the information in documents and queries into the ontologies.
3. To improve retrieval by using knowledge about retaliations

between concepts in the ontologies.

REFERENCES

- [1] JING-YAN WANG, ZHEN ZHU, "Frame work of multi-agent information retrieval system based on ontology and its application" 978-1-4244-2096-4/08/\$25.00 ©2008 IEEE
- [2] Yi Xiao, "Intelligent Information Retrieval Model Based on Multi-Agents" 1-4244-1312-5/07/\$25.00 © 2007 IEEE
- [3] Pan Ying, Wang Tianjiang, Jiang Xueling, "Building Intelligent Information Retrieval System Based on Ontology" 1-4244-1135-1/07/\$25.00 ©2007 IEEE.
- [4] Jae-Woo LEE, "A Model for Information Retrieval Agent System Based on Keywords Distribution" 0-7695-2777-9/07 \$20.00 © 2007IEEE
- [5] Yan HU, Yanyan XUAN, "Research on Web Information Extraction Based on XML" 978-0-7695-3334-6/08 \$25.00 © 2008 IEEE
- [6] Qian Zhu, Xianyi Cheng, "The Opportunities and Challenges of Information Extraction" 978-0-7695-3505-0/08 \$25.00 © 2008 IEEE
- [7] Jae-Woo LEE, "A Model for Information Retrieval Agent System Based on Keywords Distribution" 0-7695-2777-9/07 \$20.00 © 2007IEEE
- [8] David H.C. Du, "Intelligent Storage for Information Retrieval" 0-7695-2452-4/05 \$20.00 © 2005IEEE
- [9] JIANG Xin-Hua¹, LIU Yong-Min, "A New Artificial Intelligent Information Retrieval Methods" 978-0-7695-3559-3/09 \$25.00 © 2009 IEEE
- [10] R. Kalavathy*, R.M. Suresh†, R. Akhila‡, "Kdd and data mining"
- [11] DRIDI~, "Ontology-Based Information Retrieval: Overview and New Proposition"
- [12] Panjce Panov, Sajso Djzeroski, Larisa N. Soldatova, "OntoDM: An Ontology of Data Mining" 978-0-7695-3503-6/08 \$25.00 © 2008 IEEE
- [13]Chen-Yu Lee' Von-Wun SOO~-, "Ontology=based Information Retrieval and Extraction" 0-7803-8932-8/05/\$20.00 © 2005 IEEE
- [14] Bill Kules, Jack Kustanowitz and Ben Shneiderman, "Categorizing Web Search Results into Meaningful and Stable Categories Using Fast-Feature Techniques" ACM 1-59593-354-9/06/0006...\$5.00.
- [15] D Evangelin Geetha G Krishna Naidu T V Suresh Kumar K Rajani Kanth, "Simulative Performance Evaluation of Information Retrieval Systems" 978-0-7695-3136-6/08 \$25.00 © 2008 IEEE
- [16] A. Del Bimbo, "A Perspective View on Visual Information Retrieval Systems"

ABOUT THE AUTHOR:

Dr. R. P. Singh received B.Tech. and Ph.D. degrees in electrical engineering from K.N.I.T; Sultanpur (U.P.) and Institute of Technology, Banaras Hindu University, Varanasi (U.P.), India, in 1988 and 1993, respectively.

Dr. Singh is a person of multidimensional personality with 23 years of experience in the field of teaching and research including 4 years as Director/ Principal. Dr. Singh is a member of various academic societies of national and international repute. He has to his credit 40 research papers and authored four books. These books are well received by the students &

academicians alike. In 1994-1995 he was the recipient of "The President of India" as Director/ Principal award for research contributions. Currently he is working as Director/ Principal in Bimla Devi Educational Society's

Group of Institutions (Integrated Campus)

JB Knowledge Park, Faridabad and his email id is ravindra10765@gmail.com .

Poonam Yadav obtained B.Tech in Computer Engg.&Science from Kurukshetra University Kurukshetra, India and M.Tech in Information Technology from Guru Govind Singh Indraprastha University in 2002 and 2007 respectively. She is currently



working as Assistant Professor in D.A.V College of Engg. & Technology, Kanina (Mohindergarh). Presently she is pursuing her Ph.D. research at NIMS University, Jaipur. Her research interests include Information Retrieval; Web based retrieval and Semantic Web etc. Yadav is a life time member of Indian Society for Technical Education and her email id is poonam.ir@gmail.com.



Mrs.