Research on Semantic Information Retrieval Based on Ontology

Zeng Dan
Library of Wuhan University of Technology, Wuhan, P.R. China, 430070
(E-mail: zengd@whut.edu.cn)

Abstract The purpose of this paper is to resolve the problem of the accuracy on traditional information retrieval, which brings ontology-based semantic information retrieval. The anthor wlilizes the method of establishing the domain semantic model with ontology technology, the membership of concept added to the process of semantic modeling, and to provide semantic annotation to facilitate computer calculation processing. The model considers two ways to solve the traditional search engine users search results returned relevance is not high, that is, the user query whether is "true expression" and search engines whether can reveal the nature of index web documents.
Key words Ontology; Semantic retrieval; Information retrieval

1 Introduction

Traditional information retrieval technology mainly is based on the keyword syntax matching. Advantage of this technique is simple and quick and easy. However, this is based on the simple matching syntax level, Lack of knowledge expression, processing and understanding, leading to unsatisfactory search quality. Ontology is due to good ability of concept hierarchy and support the logical inference, and expressed semantic by the relationship between the concepts [1], semantic information retrieval has become an important topic.

Abroad (especially Europe and the United States) is a leading research on the ontology, launched a series of ontology development methods (such as IDEF5 method, framework laws, enterprise modeling method, Methodology, cyclic access method, Uschold method, etc.), development tools (Ontolingua, OntoSaurus, WebOnto, Prot6962000, OntoEdit, etc.) and relevant standards. Ontology in Information Retrieval applications in foreign countries, well-known projects include (onto)-Agent (An Ontology Based WWW Broker to Select Ontologies), Ontobroker (Ontology Based Access to Distributed and Semi-Structured Information) and SKC (The Scalable Knowledge Composition Project) other. Ontology research in China started is relatively late, relatively little research. The main research contents include product information modeling, virtual enterprise modeling, and common sense knowledge base. Such as the large-scale knowledge system of Institute of Computing Technology Chinese Academy of Sciences, the common sense Knowledge of Chinese Academy of Sciences Institute of Mathematics, Artificial Intelligence Institute Zhejiang University on product information integration based on ontology research. Among them, the more influential study is the usefulness of common sense knowledge led by Lu Ruqian Chinese Academy of Sciences Academy of Mathematics [2].

2 The Information Retrieval Based on Semantic

To understand the rich logic semantic and reasoning search in the network information resources, retrieval systems must have:

(1) Some knowledge to express objects concept and their inter-relationship between the logical semantic;

(2) A certain vocabulary system to describe the object classes and their relationships, then to establishment the corresponding metadata elements;

(3) A certain assignment mechanism to establish metadata element and description relationship of the corresponding resources (or resource fragment);

(4) Some of the markup language and syntax to mark metadata elements and their the assignment relationship;

(5) The search reasoning mechanism to search, verification and reasoning use of knowledge systems and markup languages.

If all of these mechanisms to deal with computer-understandable way, and set up the system in the popularity of network resources, can achieve the network environment based on semantic retrieval and reasoning, which is the goal of Semantic Web.

In fact Web-founder Tim Berners Lee in 1998 proposed the concept of Semantic Web , with the W3C in February 2001 launched Semantic Web Activity , semantic retrieval becomes into the network research and development mainstream in network environment.

According to W3C, Semantic Web research goal is to develop a series of computer-readable expression and processing of semantic information in language and technology to support extensive and effective automatic reasoning in network environment. At this stage, it will focus in support on the network information resources and their semantic content and semantic representation, support to semantic-based agent system data analysis, understanding and processing automatically, support for inter-agent systems semantic-based knowledge exchange, support different applications and systems based on semantics of data exchange, transformation and re-use automatically. To achieve this goal, Semantic Web requires a multi-technology framework (Figure 1).

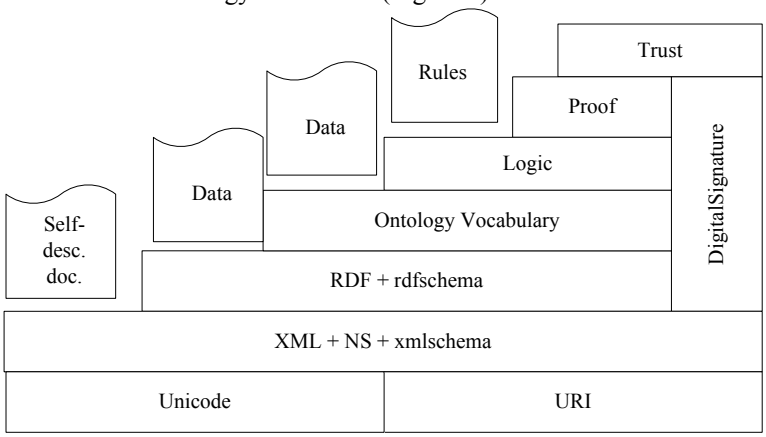

Figure 1 A Multi-Technology Framework

Shown in Figure 1, Semantic Web is a set of network information storage, organization, and that all aspects of security authentication complete system, involving XML, ontology, digital signature techniques and methods, it contributes to the semantic level based on network information organization and retrieval, is the development direction of WWW.

Ontology as the semantic support of semantic web information organization plays a central role. In fact, the original ontology is a philosophical concept, is the philosopher to describe the nature of things. Later scholars borrowed knowledge engineering concept. In the computer field, the current of more authoritative ontology concept is proposed by Studer: the concept of ontology is shared clear formal specification [3]. This definition includes the meaning of 4 layers: conceptual, clear, formal and shared. Ontology aims to acquire, describe, and indicate the knowledge that is in the related fields, providing a common understanding of knowledge, determine recognition vocabulary in the area, and clearly define these terms (terminology) and the interrelationship between vocabularies from formal models of different levels. It is the performance and the describe capacity of ontology on the concept and the relationship between concept that to be the core of the semantic web, has become the key to realize the semantic information retrieval.

Semantic Web provides a new information expression way, so that the information has a semantic, but also makes the data are more well-defined structure. From the perspective of the semantic Web to seek new breakthroughs technology of search engine, to structure information retrieval model based on semantic, to make full use of the semantic information and structured features of data to retrieval, and fundamentally change the traditional search engine retrieval accuracy and recall was not high shortcomings. Ontology is the core concept of the Semantic Web, Information retrieval based on semantic is fundamentally information retrieval based on ontology.

Semantic information retrieval is actually the semantic relations reflected in Ontology applied to the annotation and retrieval of information resources. Specifically, to achieve information retrieval by the analysis and reasoning of relevant documents on the semantic level, and to communicate with the appropriate forms and user-friendly interface[4]. Semantic information retrieval includes ontology document retrieval, case retrieval and semantic retrieval.

Ontology document retrieval is to find a having specific classes and properties ontology document [5]. One of the ideas on how to achieve ontology document retrieval is apply ontology document in the transformation of ordinary search engines, That is, through the RDF document to be dealing with some common search engine indexing and retrieval, while in a sense play the role of semantic information. Another idea is to explore new methods and techniques for ontology searching, such as the site for a particular search engine or based on Jena the search engine. Metadata extracted in the system by search

for the semantic Web document together with the original structure information into a database, support for ontology query of particular class or properties.

The purpose of case retrieval is to find and search on a specified class all instances information in Ontology-based Knowledge Base, It is mainly based on structured query and reasoning and is based on RDF (S), OWL and other underlying knowledge model of the graph traversal and graph model is widely used. By Semantic Web data to complement traditional search results belongs to case retrieval, the Semantic Web resources including official semantic web documents and semantic annotation information. At the traditional search same time, the system will be for the query concept in the RDF Knowledge Base by graph traversal search of all relevant data.

Semantic retrieval is concerned about not only the simple relationship between the properties chain, it is also associated with various complex relationships between concepts. At present, some studies have begun to focus in the retrieval problem of semantic relations, Such as the literature [6-7], etc. of the semantic association search. The main contribution is to analyze the retrieval of semantic relations are facing three major challenges (that is, not enough understanding of the relationship between generic and comprehensive, the lack of relationship rather than to the concept of an object query language and system, search results and scheduling problems) and propose the corresponding solution (to be divided into semantic relations and semantic similarity of two types of association and formal, designed to support the relationship between p-query query context based on user-specified results to be sorted).

3 Retrieval System Development and Implementation

3.1 Create ontology

Ontology constitutes the core of knowledge information systems of in the field. Ontology Support information exchange process from in different aspects, the role of ontology includes: communication between different digital library systems; semantic-based Agent Communication; knowledge-based retrieval; semantic-level understanding of content; uniform format; syntax interoperability; semantic interoperability. Ontology includes concepts, relations, instances and the axioms. Formal ontology be shown in Figure 2.

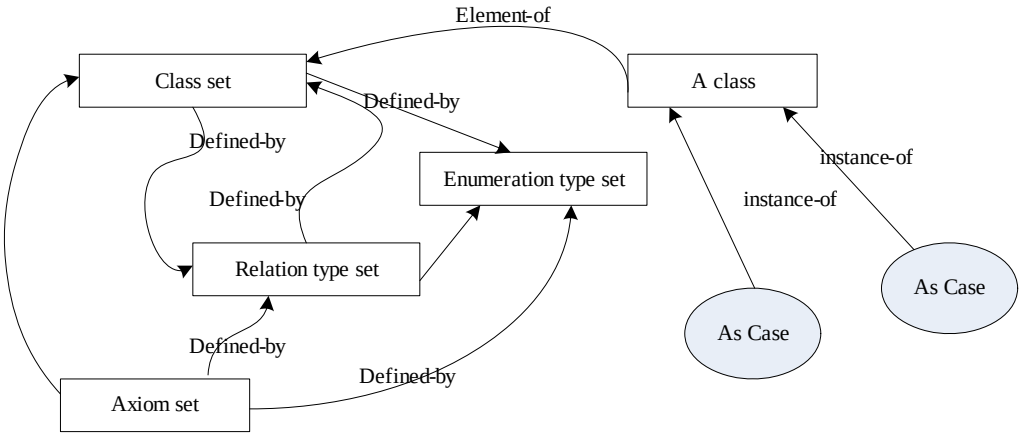

Figure 2 Ontology Conceptual Model

Concept is a group or a class entities or "things" in certain areas. Relation describes the relationship between concepts or between attributes properties. Concept can be defined as hierarchical form of classification system, the various concepts in the classification system linked by category relations. Relations can be described in different aspects characteristics by its property. Instance is the "things" by concept expressed. Ontology and its instance are referred to as Knowledge base. Axiom is used to express a concept or an instance constraint.

Ontology building needs the expert participation in Multidisciplinary. Although ontology engineering tools have been more mature, but the manual construction is still a tedious and difficult task, and ultimately lead to the so-called knowledge acquisition bottleneck. Therefore, need for a simple and reliable extraction method of ontology, that is, an efficient way of ontology structure.

3.2 Information resources semantic annotation

Ontology-based semantic annotation is to use existing ontology, with the XML markup language

for data to be marked, RDF/XML as a data description model, through annotation of digital resources, the data is marked with a clear meaning which can be understood by the machine and ultimately create knowledge base .

James Hendler pointed out: In order to resources contacts with a number of different ontology, mark the prototype should be able to easily create ontology. Users found the selected ontology vocabulary is limited when label page, in order to accurately describe the concept, the prototype should provide a simple ontology editing capabilities, such as to creating new user ontology, inserting a new vocabulary in existing ontology, merging and reusing existing ontology use vocabulary and so on. However, So far, current research prototype ontology-based semantic annotation technology, is generally assumed to be only one body in advance, on the basis of this assumption on how semantic annotation of resources, rather than by applying multiple ontology to be labeled. Obviously, it does not match Web-based semantic annotation environment of digital resources. Semantic annotation process is actually the similarity calculation process, calculation the concept similarity which is extracted from information and domain ontology [8]. First, using multiple domain ontology extracted from the resources which the semantic metadata, the extraction of semantic metadata contains two aspects: named data of entities (classes and attributes) and implied XMLSchema the structural relationship between these entities. Secondly, calculated with the two parts involved with the concept of domain ontology similarity, semantic annotated the digital resources according to the calculated similarity. Specific labeling process is shown in Figure 3. The process of Semantic annotation on the resources is also relevant domain ontology evolution process, with the creation of new domain ontology, domain ontology editor and has been updated to maintain the semantics of synchronization and resources to adapt. The final result is generated heterogeneous knowledge base by annotated heterogeneous resources.

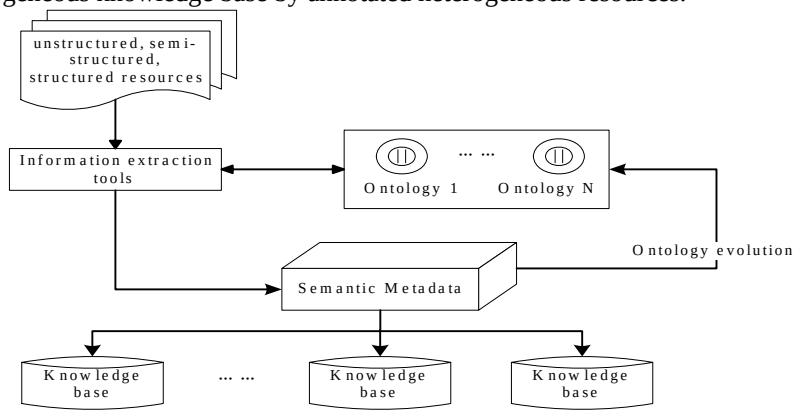

Figure 3 Semantic Tagging Model

4 Information Retrieval Model Semantic Web-Based

By this analysis, a semantic Web-based information retrieval model, as shown in Figure 4.

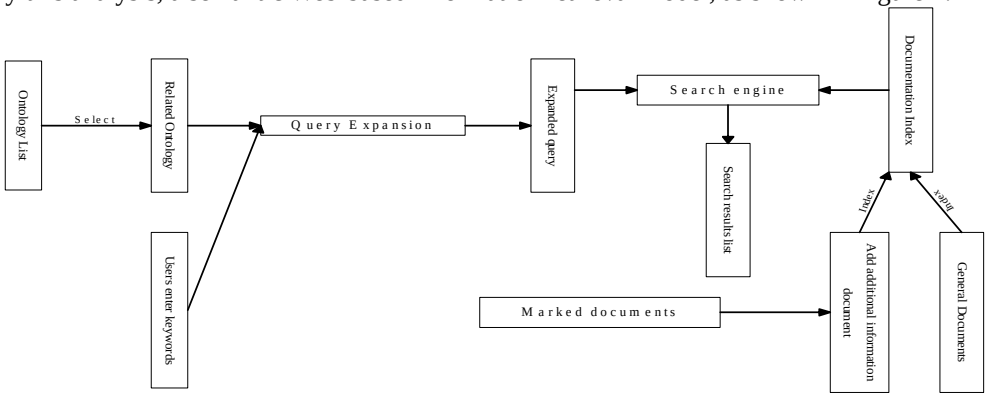

Figure 4 Information Retrieval Model Semantic Web-Based

In this model, First of all, the user selects with queries related to ontology, and then enters keywords. If the user enters the keyword is the concept of domain ontology, with user-selected knowledge in domain ontology on the user input the key words for the text described three kinds of

query expansion, At the same time statistics on the use of frequency expanded each query, The query are sorted according to their use of frequency to facilitate user selection, then search engine query based on user selected search. If the user enters the keyword is not in the concept of domain ontology is not required for query expansion.

5 Conclusions

In order to reduce the response time of user queries can be considered in the domain ontology is established, it is considered after the ontology founding, that is, after the knowledge be contained in domain ontology determined, the concept of the ontology query expansion will expand the results stored in a repository, so that can make online deal be quite simple, that is, just extract the appropriate concept from information base when query expansion. As the light of the current search engines can not index the document with semantic mark, the documents with semantic markup was to add additional information processing, in the process be used to the ontology reasoning mechanism to find out hidden information within the document, hidden information will be processed to the same document, so that search engines can index these additional information, it makes the search engine on the document index to better reflect real content. The model considers two ways to solve the traditional search engine users search results returned relevance is not high, that is, the user query whether is "true expression" and search engines whether can reveal the nature of index web documents.

References

[1] Nicola Guarino, Daniel Oberle, Steffen Staab. What is an Ontology? International Handbook on Information Systems, Handbook on Ontologies

[2] Liu Aijun. Ontology-based Semantic Information Retrieval and Related Technologies[D]. xian:Northwest University, 2008:2-3 (In Chinese)

[3] Studer R., Benjamins V.R Fensel D. Knowledge Engineering Principles and Methods[J]. Dataand Knowledge Engineering, 1998,25 (1):161-197

[4] Ding Shengchun, Gu Defang. The Research of Using Jena in the Realization of Semantic Retrieval Based on Ontology[J]. New Technology of Library and Information Service, 2005,(10) (In Chinese)

[5] Huang Min, Lai Maosheng. Survey of Semantic Search[J]. Library and Information Service, 2008,(6) (In Chinese)

[6] Anyanwu K., Sheth A. ρ -Queries: Enabling Querying for Semantic Associations on the Semantic Web. IW3C2, MTA SZTAKI. Proceedings of the 12th International Conference on WWW. New York: ACM, 2003:690-699

[7] Aleman Meza B., Halaschek C., Arpinar I.B. Context-aware Semantic Association Ranking. Cruz I F, Kashyap V, Decker S,et al, Eds. Proceedings of the 1th international Workshop on Semantic Web and Databases. Co-located with VLDB 2003. Berlin:Humboldt-University, 2003: 33-50

[8] Xu Bin, Li Juanzi, Wang Kehong. Web Service Semantic Annotation[J]. Journal of Tsinghua University(Science And Technology), 2006,(10):1784-1787,1792 (In Chinese)