

AN INTELLIGENT RETRIEVAL FRAMEWORK IN SEMANTIC WEB BASED ON AGENTS

Lin Xu and Bo Peng

College of Information and Electricity Engineering, CAU, 100083

Abstract: Semantic Web is a research project of the World Wide Web Consortium (W3C), which aims to solve the problem of that the Web data is hard to process automatically because of the lack of semantic. The objective of this project is to implement the intelligent reasoning and processing automatically for the information published in the Web. Agent, as an intelligent object, is particularly suitable for lots of applications which are implemented in the Semantic Web. Based on the research of WWW and Intelligent Agent, this paper combined with some technologies, such as Information Search, Knowledge Representation, Modeling of Ontology, then proposed and implemented a Multi-Agent System Framework on the basis of Ontology semantic information retrieval, which includes four parts: information collecting, storing, querying and semantic reasoning.

Key words: Semantic Web; Intelligent Agent; Ontology; Intelligent Retrieval; Multi-Agents system

1. INTRODUCTION

The Web was designed as an information space, with the goal that it should be useful not only for human-human communication, but also that machines would be able to participate and help. One of the major obstacles to this has been the fact that most information on the Web is lack of semantic, that the notion of the Semantic Web has been introduced. "The Semantic Web is an extension of the current web in which information is given well-defined meaning, better enabling computers and people to work in

cooperation.", according to Tim Berners-Lee, the founder of the Semantic Web.

At XML 2000 Tim Berners-Lee, Director of the Wide Web Consortium, outlined his vision for the Semantic Web and explained the layered architecture that developed in the next ten years (Figure1).

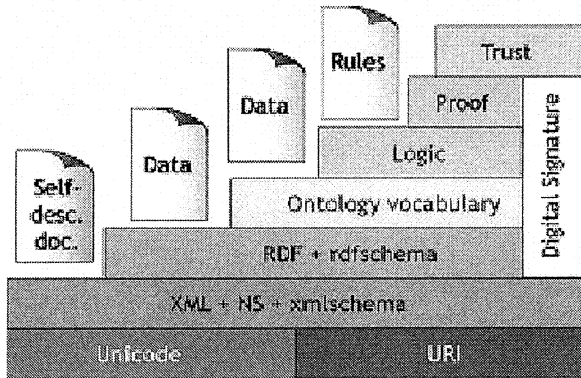


Figure1. The architecture of the Semantic Web

(1)The bottom of the Semantic Web is made up of Unicode and URIs(Uniform Resource Identifiers). Unicode is a character encoding system, like ASCII, designed to help developers who want to create software applications that work in any language in the world. URIs stand for Uniform Resource Identifiers, URIs are short strings that identify resources in the web, such as documents, images, downloadable files, services and so on. The Semantic Web can involve all resources and different languages with the help of Unicode and URI.

(2)The second layer of the Semantic Web is XML+NS (Name Space) +XML Schema. XML is a standard format for serializing data using tags. The NS (Name Space) of an element, is the scope within which it is valid. The XML+NS (Name Space) +XML Schema layer keeps the principles of SGML(Standard for General Markup Language) in place but its spec is thin enough to wave.

(3)The third layer consists of RDF and RDF Schema. RDF(Resource Description Framework) provides a way of describing resources via metadata. It restricts the description of resources to triplets. RDF Schema is a schema defines the terms that will be used in the RDF statements and gives specific meanings to them. RDF and RDF Schema layer is weak in some special area, so the Ontology layer is needed.

(4)The fourth layer is Ontology, the Ontology is designed to standardize means that can be used on the Web.

(5)Logic is the fifth layer, it is an inference engine that could explain how it followed the semantic links that it did.

(6)The next step in the architecture of the Semantic Web is trust and proof. Very little is written about them, which is a shame since it will become very important in the future.

The RDF/RDF Schema layer and the Ontology layer are the core of the architecture. The description information of Web exists in the form of metadata, it must be more intelligent if the metadata is collected and stored because of the semantic.

From the above descriptions, it is evident that applying the semantic in web retrieving is necessary. Agent, as an intelligent object, is particularly suitable for lots of applications which are implemented in the Semantic Web. So based on the research of WWW and Intelligent agent, this paper combined with some technologies, such as Information Retrieval, Knowledge Representation, Modeling of Ontology, then proposed and implemented a Multi-Agent System Framework on the basis of Ontology semantic information retrieval.

2. THE ARCHITECTURE OF THE INTELLIGENT RETRIEVAL FRAMEWORK

The Intelligent Retrieval Framework described the metadata of web by RDF models, which used for web resource only, to show details of the web. In order to store and query the RDFdata, an appropriate relational database will be ok, also the RDFdata can be treated as XML data. OWL, the Web Ontology Language, is a semantic markup language for publishing and sharing ontologies on the World Wide Web, also it is used in the Intelligent Retrieval Framework to support the knowledge reasoning.

The main function modules include the collecting agent to collect the information of Web metadata, the storing agent to store the metadata, the reasoning agent to deduce from the semantic correlation on the basis of Ontology and the querying agent to search by the results of the reasoning agent. Besides, it includes a database storing the Web metadata and an Ontology existed in the form of OWL files which supported special field knowledge for the reasoning agent. The architecture of the Intelligent Retrieval Framework is showed as Figure 2.

The flow of the Intelligent Retrieval Framework can be divided into the following four steps:

(1)Collect the information of Web metadata: the collecting agent draws out information from the metadata of the Semantic Web, and the metadata is expected to be stored by the storing agent after filtered by the collecting agent.

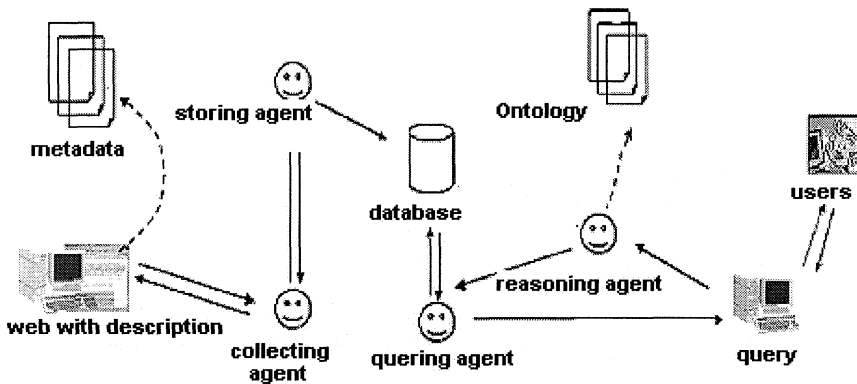


Figure 2. The architecture of the Intelligent Retrieval Framework

(2)Store the metadata: after receiving the metadata from the collecting agent, the storing agent store the metadata into database in an appropriate way.

(3)Intelligent reasoning on the basis of Ontology: when the Ontology is set up, the reasoning agent deduces the key words, or by semantic mapping, or by semantic correlation extending. Then the results that deduced by the reasoning agent would be accepted as a request by the querying agent.

(4)Intelligent querying based on semantic: the querying agent accepts a request from the reasoning agent, then searches the database that stored metadata by semantic. Finally, it returns the results that coincided with users.

3. THE IMPLEMENTATION OF THE INTELLIGENT RETRIEVAL FRAMEWORK

There are several modules in the Intelligent Retrieval Framework, but the core is the four agents: the information collecting agent, the storing agent, the reasoning agent and the querying agent.

3.1 The information collecting agent

The collecting agent collects description information of the Semantic Web metadata, and then hands the metadata to the storing Agent. Generally speaking, the description information of metadata is described in the form of RDF data at present, and the direct way is to insert the RDF data into Web, and often insert into the titles of HTML as XML syntax. Moreover, the RDF data can be stored in the form of files when there is a lot of metadata, and the links of the files would be provided on the Web for other agents or applications.

The collecting agent runs in the platform of JADE, and searches the semantic Web in a given route. It can filter the content of web automatically, draw out the description metadata of the Semantic Web, crawl from one web to another to affirm that the link is effective and delete the link that exist in name only.

3.2 The storing agent

By the communication between agents, the collecting agent passes the web metadata to the storing agent. The storing agent then recovers the metadata in the form of RDF data, after that, it stores the metadata in a proper way.

By and large, there are three methods to store the RDF data nowadays: XML/RDF file layout, RDF database, relational database. The XML/RDF file layout will be ok if there is little data, however, considering augmentability, the mode of querying, efficiency and other factors, it is a better choice to store the RDF data by RDF database or relational database when there is a lot of fact data. Relational database is the most popular application of database at present. However, because it is lack of semantic in relational database, it must be first convert RDF data model into relational model.

Statement is a basic description of a fact, and it is the minimal effective data unit in RDF data model. As a result, the table of storing statement is the most important part in database schema. (Table 1).

Table 1. RDFSatemens table

| Column name | Type | Comments |
|------------------|-----------|--|
| Subject | Id-ref | |
| Predicate | Id-ref | |
| Object | Id-ref | |
| Object_isliteral | Smalllint | flags whether "object" is in literal or resource table |
| Model | Id-ref | |
| isreified | smllint | flags whether this statement is reified |

Besides the RDFSatemens table, other tables are also needed, such as: RDFResources table, RDFLiterals table, Namespace table and LocalName table. The structure of these tables are simple, so omit them here.

RDF is a modeling language based on XML, so the RDF data is an especial XML data in some sense. It is the best way to store the RDF data by Middleware according to the characters of RDF data model. By mapping between XML documents and database, middleware completes the conversion. The typical Middleware product is XML-DBMS, and the Intelligent Retrieval framework in this paper also uses XML-DBMS to implement the conversion between RDF data and Mysql database with three steps: write a mapping document, a filter and java codes.

3.3 The reasoning agent

In order to search on the basis of semantic, not only the information of retrieved metadata is needed, but also the reasoning ability. The core of the reasoning agent is intelligent reasoning, it includes semantic mapping and semantic correlation extending in reasoning step.

During the course of semantic mapping, the key words are analyzed. Meanwhile, the querying agent judges the meaning of the key words by Ontology, and then selects the most accurate meaning from that several possible meanings. This method can improve the accuracy of the retrieval results for the key words that have several meanings. Traditional searching tools cannot discriminate different meanings of the same word, so it retrieves only by key words. But in the Retrieval Framework of this paper, every meaning of the key word is defined exactly by the Ontology, and mapping only one URI.

In the course of semantic correlation extending, not only the key words, but also other words that related to the key words are all obtained because of the Ontology. For example, the querying Agent can search one sub concept by subclass relation though this sub concept is not in the list of the key word.

At the end of the reasoning process, the reasoning agent delivers the results to the querying agent in the form of ACLMessage objects, some of the results are gained by semantic mapping, and some of them by semantic correlation extending.

3.4 The querying agent

The querying agent communicates with the reasoning agent, and receives the results from the reasoning agent. After that, the reasoning agent retrieves from the RDF database that storing metadata of the web, and returns the

optimum result to users. Considering the features of RDF models, it is best for users to search by one special RDF searching language.

RDQL is one of the RDF searching languages of Jena, it resembles the “select” language of SQL, and it treats RDF data as triple, at the same time, it ignores Schema or Ontology information. RDQL provides one graph pattern to map data, and its graph pattern is a RDF statement. Users can search the unknown part of the statement according to the restriction of some parts of the statement.

RDQL can use solely, also can be embed in java codes. In this Retrieval Framework, RDQL language is used in the querying agent, and then it was encapsulated as a Query object. After that, the object was called, and users can get searching results in the form of ResultBinding variables, which can be stored and fetched easily by their names. Every model is a Web Unit in the RDF database, and their URL and title can be acquired easily when the model is retrieved. The querying agent returns this message to users, and the whole searching process finishes.

4. CONCLUSION

The semantic Web is the tendency of web in the future. Researches and applications on the Semantic Web will be useful for solving the bottleneck, which restricts the development of web nowadays, and also it can speed up the course of web application. Based on the research of WWW and Intelligent Agent, this paper combined with some technologies, such as Information Retrieval, Knowledge Representation, Modeling of Ontology, then proposed and implemented a Multi-Agent System Framework on the basis of Ontology semantic information retrieval. The purpose is to arise discussion of the Semantic Web and build a nice future for it together.

REFERENCES

- [1]Tim Berners-Lee, J Hendler, O Lassila. The semantic web[J]. New York: Scientific American, 2001,284(5): 34-43.
- [2]Tim Berners-Lee. Semantic Web-XML2000-Slide "Architecture" [EB/OL]. <http://www.w3.org/2000/Talks-xml2k-tbl/Overview.html>.
- [3]T R Gruber. A translation approach to portable ontology specifications. Stanford University Knowledge system laboratory ,Tech Rep: Logic-92-1, 1992, Frank Manola, Eric Miller. RDF Primer. W3C Working Draft 23 January 2003[EB/OL]. <http://www.w3.org/TR/2003/WD-rdf-primer-20030123/>.

- [4]Dan Brickley, R.V. Guha. RDF Vocabulary Description Language 1.0: RDF Schema. W3C Working Draft 23 January 2003[EB/OL]. <http://www.w3.org/TR/2003/WD-rdf-schema-20030123/>.
- [5]Deborah L. McGuinness, Frank van Harmelen. OWL Web Ontology Language Overview. [http://www.w3.org/TR/owl-features/Dave Reynolds.Jena relational database interface – introduction](http://www.w3.org/TR/owl-features/Dave_Reynolds.Jena_relational_database_interface_-_introduction). 12 May 2001[EB/OL]. [http:// www.hpl.hp.com/semweb/doc/RDB/rdb-intro.html](http://www.hpl.hp.com/semweb/doc/RDB/rdb-intro.html).
- [6]Ronald Bourret, XML and Databases[EB/OL], <http://www.rpbourret.com/xml/XMLAndDatabases.htm>.
- [7]Michael K. Smith, Chris Welty, Deborah L. McGuinness. OWL Web Ontology Language Guide. <http://www.w3.org/TR/owl-guide/>
- [8]Miller E. The Semantic Web. Technical Report, Eleventh International World Wide Web Conference (WWW2002), in Honolulu, Hawaii, USA, 2002.
- [9]McBride B. RDF architecture. Technical Report, Eleventh International World Wide Web Conference (WWW2002) in Honolulu, Hawaii, USA, 2002.
- [10]Hendler J. OWL: A Web Ontology Language. Technical Report, Eleventh International World Wide Web Conference (WWW2002) in Honolulu, Hawaii, USA, 2002.