ONTOLOGY-BASED METADATA FOR COMPUTER TERMINOLOGY

Khin Kyawt Kyawt Khaing, Moe MoeHlaing
Computer University (Magway)
khinkyawtkyawtkhaing.k4@gmail.com

ABSTRACT

Nowadays,Information Technology is rapidly improving to provide possibilities for integration among applications using various databases. It involves retrieving and integrating information from multiple data sources. The method how to easily retrieve the information and knowledge are needed. Therefore, information retrieval system should provide advanced searching facilities to retrieve information. To organize large information repositories and access these repositories efficiently, metadata can be used. Closely related to metadata is ontology. Ontology is the key component of the Semantic Web. Ontology is used not only in semantic web but also in databases and applications that need to share domain information (specific area of knowledge). Ontology comprises as the concept, properties and restrictions on properties. This paper intends take the facilities of ontology in the building of a computer terminology.This system accepts the query as input and retrieves the relevant definition, image, etc. This system is implemented by using Java API for RDF (Resource Description Framework) called Jena package.

1. Introduction

Information and knowledge are essential for human. The method how to easily retrieve the information and knowledge are needed. Information retrieval system is a device interposed between a potential user of information and the information collection itself.

Information Retrieval (IR) is concerned with the process involved in the representation, storage, searching and finding of information which is relevant to a requirement for information desire a human user.

The capability of representing with knowledge base on text is important. This paper explains how to represent this knowledge by using ontology concept. The main focus of this paper is to study how to create domain ontology. First, this system will capture the concept of computer terms and then set the related information into the domain (ontology-based metadata database) by using RDF (Resource description Framework), a frame-based language. RDF is proposed by W3C consortium to use in semantic web. Finally, this paper accepts the query as input and the user can retrieve the relevant definition, image, antonym etc.

To develop ontology-based metadata, Resource Description Framework (RDF) schema can be used. RDF is a recent W3C recommendation designed to standardize the definition and use of metadata descriptions of web-based resources. RDF is equally well suited to representing data. RDF uses XML as its serialization syntax.

2. Search Strategies with Support of Ontology

To get easier and more effective information from databases, a variety of searching functionalities can be used. These functionalities include content-based visual search, hybrid search and ontology-based search. This paper proposed the ontology-based search functionality.

2.1 Content-based Visual Search

In content-based visual search, the user will be able to provide an image, video or 3D model as input query to the system. Then, the system will retrieve the related results.

2.2 Hybrid Search

Hybrid search is the combination of ontology-based search and content-based visual search.

2.3 Ontology-based Search

The ontology-based search will give the opportunity to the users to take advantage of the ontological data structure and look for specific information. The search can be conducted using two different methods. With the first method, predefined concepts will be available as links in the web interface (such as search base on location, date, etc.). A tree-like interface is the method and it gives an illustrative example of the structure of the underlying knowledge. By using this method the user can select a concept to start the search process. Using the second search method, the user has the option to type in keywords in a text field. The ontology is queried and the objects that were found to contain the

keywords in their metadata are displayed in the result set.

3. Related Work

In recent year, several IR systems have been developed using XML case representation. An XML based markup language is introduced as a standard way of representing cases[5]. The main focus of semantic web languages such as RDF is to provide more descriptive information. RDF graph model is distant from its XML tree syntax; therefore relationships can be more easily modeled [3]. If ontology developers want to modify ontology, they have to understand RDF syntax and its model [2, 4].In this system user can get only information of the concepts and complicated attributes specified in predefined ontology. To get more complete and efficient retrieval system, a set of API classes for retrieving the relations between concepts should specified in the ontology. And then these classes are needed to build in Jena tool.

4. Ontology

Ontologies are metadata schema, providing a controlled vocabulary of concepts, each with an explicitly defined and machine processable semantics. By defining shared and common domain theories, ontologies help both people and machines to communicate concisely supporting the exchange of semantics and not only syntax. Moreover ontology is a formal model of the kinds of concepts and objects that appear in the real world together with the relationships between them. There are roughly four kinds of ontologies: document ontologies, metadata ontologies, domain ontologies and service ontologies. In this approach system, it consists of domain ontology.

Ontology is a formal definition of a body of knowledge. An explicit formal description of concepts (classes) in a domain of discourses, properties of each concept describing various features and attributes of the concept, and restrictions on properties. Ontology together with a set of individual instances of classes constitutes a knowledge base. Ontologies are used by people, databases, and applications that need to share domain information (a domain is merely a specific area of knowledge, such as medicine, petroleum, financial, management, tec.). Classes are the focus of most ontologies. Classes describe concepts in the domain. For example, a class of wines represents all wines. Specific wines are instances of this class. A class can have subclasses that represent concepts that are more specific than the super class. For example, we can divide the class of all wines into red, white, and rose wines.

4.1 Semantic Web

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. The Semantic Web will extend the current web with representation of semantics, allowing not only humans, but also computers to read and understand the contents of web pages [6]. The purpose of ontologies in the Semantic Web is to provide a kind of semantic typing for the distributed all over the Web to facilitate their interrogation by users through search or query engines.

4.2 Designing Ontology

Ontology is similarly with OOP. OOP is emphasis with object and method. Ontology is emphasis with object and description. All things in our environment that have concept and relationships can transform into ontology. For example, the word 'www' is concept, 'has synonym', 'has definition', are attributes (properties) and detail description of 'world wide web', 'internet' are the value of property.

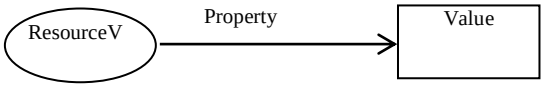


Figure 1. A generic RDF description

- Elliptical nodes represent resources.
- Arc represents resource properties
- Rectangular nodes represent literals

5. Ontology Languages

RDF (Resource Description Framework) is a language for creating a data model for objects (resources) and relations among them. RDF defines a simple model for describing interrelations resources I terms of properties (attributes) and values. RDF is also a generalmethod to decompose knowledgeinto

small pieces, with some rules about the semantics, or meaning which consists of three object types.

- **Resources:** A resource may be an entire Web page; a part of a Web page; a whole collection of pages; or an object that is not directly accessible via the Web; e.g. a printed book. Resource Indicators (URIs).
- **Properties:** A property is a specific aspect, characteristic, attribute or relation used to describe a resource is a RDF statement.
- **Statements:** A specific resource together with a named property plus the value of that property for that resource is a RDF statement.

RDF is developed WWW Consortium. RDF provides a general, flexible method to decompose any knowledge into small pieces, called triples, with some rules about the semantics (meaning) of those pieces.

6. RDF DataQueryLanguage (RDQL)

RDQL is a query language for RDF. RDQL is widely implemented by RDF frameworks. RDQL allows complex queries to be expressed concisely. RDQL's syntax superficially resembles that of SQL, and some of its concepts will be familiar to anyone who has worked with relational database queries. RDQL queries can be executed on the command line against a Jena model using the 'jena.rdfquery' tool.

SELECT ?definition

WHERE

?y <http://somewhere/Dictionary-rdf/3.0# word> "www"

?y<http://somewhere/Dictionary-rdf/3.0# DefinitionOf> ?definition

Result: http://somewhere/"world wide web"

The SELECT part declares the variables to be by the query. The WHERE clause introduces a second variable, concept and defines triples that are matched against the graph. The query finds statements in the graph for which all of the triples in the WHERE clause hold. The USING clause is a convenience, used to declare prefixes for namespaces.

7. Explanation of the System

This section describes the explanation of the system. In this system, it consists of four step processes. The first step of the system is to capture resources form computer domain by using ontology concepts. After using ontology concepts, construct simple computer terminology ontology. And then the system can retrieve information from the ontology-based metadata. Finally, the system will display relevant results for user's need. This proposed system consists of two portions. One is search and another is update portion.

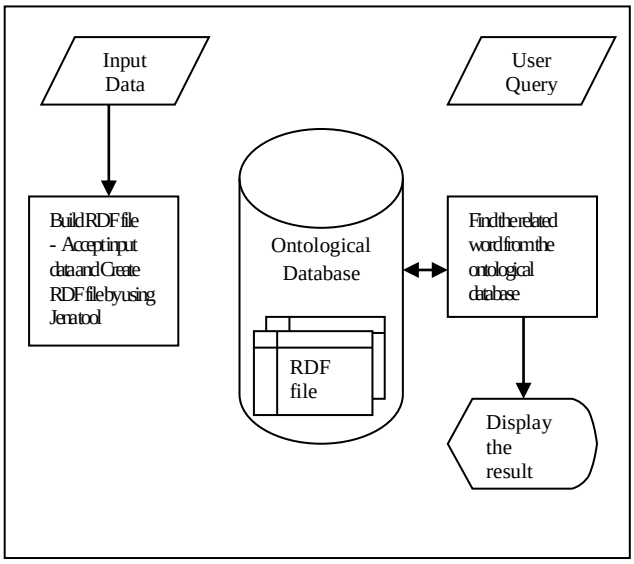
7.1 Searching Relevant Results from Ontology-based Metadata

In searching process, user can type the word in the search text box. This process is searching by software. In the result box, all information is displayed. Retrieving Information from ontology-based metadata has many advantages. In this system, user can get not only query information but also other related information.

7.2 Updating Ontology

Proposed of this system consists of updating process function. In this process, the user can insert and update data into database.

8. System Overview



9. System Design

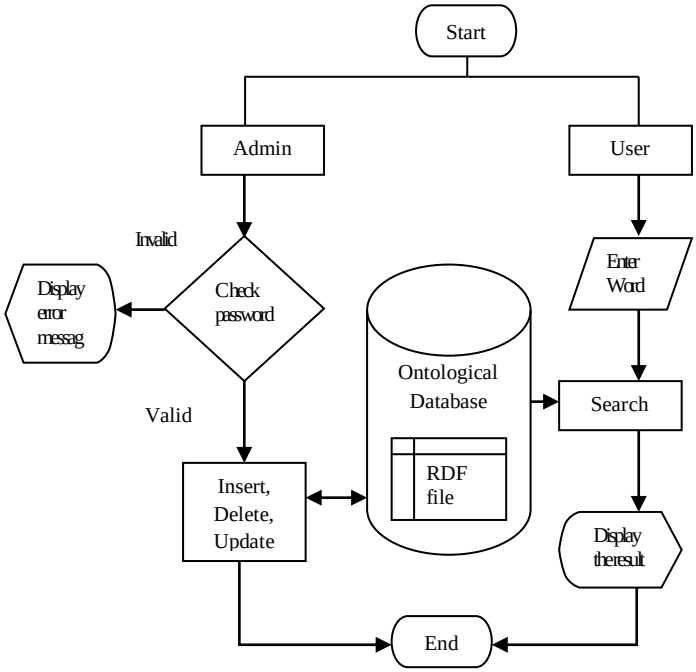


Figure 2. System Flow Diagram

10. CONCLUSION

The main objective of this paper is to study the ontology-based metadata. This paper takes the facilities of ontology in the building of knowledge repository. This paper has described the efficient searching method by using ontology-based metadata in computer terminology domain. It presented how to build computer ontology by using ontology concepts. And also described the use of metadata and what information retrieval can do by applying ontology concepts.The main objective of constructing ontology-based metadata is to query required information from metadata quickly and efficiently. Therefore the result of this system is not only to search quickly and efficiently but also to provide information that seems to fit a user's need can be tailored for a completely different purpose.

10.1 Limitations and Further Extension

The user can query just the words from the proposed ontology-based computer dictionary. So the user read very limited words. And the developers who want to modify this system have to understand and perceive the ontology concept and RDF language.

This system is only computer terms dictionary. To become an informal hierarchical Dictionary like Yahoo!, other specific domain terms dictionaries such as Medical Dictionary and Economic Dictionary need to be effectively developed. There are several requirements to develop RDF for all types of resources.

REFERENCES

[1] O'Reilly, *"RDF and Metadata"*, On-line document http://www.xml.xom/.

[2] "RDQL - A Query Language for RDF", W3C Member Submission, 9 January 2004, On-line document http://www.w3.org/ Submision/ 2004/SUBM-RDQL-20040109.

[3] *"Resource Description Framework (RDF) Concepts and Abstract Syntax"*, W3C Recommendation, 10 February 2004, On-line documenthttp://www.w3.org/TR/rdf-con-cepts.

[4] "Resource Description Framework (RDF) Schema Specification 1.0", W3C Candidate Recommendation, 27 March 2000, On-line document http://www.w3.org/TR/ 2000/CR-rdf-Schema-20000327.

[5] "*RDQL/XML Syntax Specification (Revised);"* W3C Recommendation, 10 February 2004, On-line document, http:// www. w3. org/ TR/2004/REC-rdf-syntax-grammar-20040210.

[6] Resource Description Framework: (RDF) Model and Syntax Specification, w3C, World Wide Web Consortium. **Error! Hyperlink reference not valid.**.

Proceeding of the Third International Conference on Computer Applications 2005, pp 385 to 395.

.