Reasoning and Realization Based on Ontology Model and Jena

Zhang Wei [#1,*1], Duan Liguo [#2], Chen Junjie [#3]

[#] *Computer and Software College,Taiyuan university of Technology*
No.79. West Yingze Street, Taiyuan, China
[1]zhangwhhx@163.com
[2]tyutdlg@163.com
[3]chenjj@tyut.edu.cn
[*] *Information center, Shanxi Medical College for Continuing Education*
No.22 Shuangta East Street, Taiyuan, China

***Abstract*—As the core of answer extraction of question answering system of restricted domain, reasoning plays an important role in the system. In question answering system of restricted domain, we use ontology model. How to reason with ontology model depicted with OWL using the protégé and Jena API are discussed, then, the reasoning application of Taiyuan hospital information retrieval system is also discussed in detail.**

***Keywords*—Jena; description logic; reasoning machine;**

I. Introduction

As the core of answer extraction of question answering system of restricted domain, reasoning plays an important role in the system. With the vast application of OWL language, many reasoning technologies based on the OWL language appear, which will increasingly cause the rapid development of ontology application. There are abundant and indirect-stated knowledge in ontology, however, knowledge discovery is a high-cost activity, it is because the process of reasoning is extremely complicated. If knowledge discovery is done when the information is searched, it will make the efficiency of query very low. Commonly, we use the method that the reasoning machine is used to mine the knowledge of the ontology in advance, then, we add the result of reasoning to the prior knowledge base to improve the search result.

II. Correlation Technique

A. Ontology, Ontology Language and Development Tool

Ontology is a concept in philosophy that describes the thing's essence. In the domain of Artificial Intelligence (AI), ontology is an engineering artefact. It contains a glossary set, which describes some conditions and an axiom set. So the ontology is usually a lexicon and a description of terms [1]. In 1991, Neches proposed that an ontology is to define the basic terms and relations comprising vocabularies of a topic as well as the rules for combining terms and relations to define extensions to the vocabularies [2].With the help of the ontology, we can eliminate the conflict of different understandings of the concepts and engender the same cognition. A typical ontology contains five parts: classes, relations, functions, objects and axioms [3].

The OWL, i.e. Web Ontology Language is designed for use by applications that need to process the content of information instead of just presenting information to humans. Ontology languages, such as RDF,DAML+OIL,OWL[4] are based on XML standard, XML is a technical means of the realization of the metadata (the data of describing data), XML realizes formatting from the bottom of data and documents. RDF (Resource Description Framework) is built on XML standard to describe metadata and the relation between metadata, it defines a simple data model and describes the resource and the relations among resources by property and value. Therefore, RDF model can be regarded as an entity relation diagram. To describe resource in detail, a more expressive language is required. Only with automatic reasoning function can the description mechanics be applied in automated processing. It is based on the above considerations, ontology languages such as DAML+OIL and OWL are pushed forward to develop. These ontology languages are created to build on the basis of description logic research and introduce Web features, among the languages, OWL is a Web ontology language created by W3C working group in 2001 and has been become the recommended standard of W3C since February 10, 2004[5]. Its maximal feature is related description logic, which means reasoning machine of description logic can reason OWL of ontology. OWL is future standard of Web ontology language, it can construct a complicated conceptual relation network by describing concepts, conceptual properties and mutual relations. OWL has greater machine interpretability of Web content than that supported by XML, RDF, and RDF Schema (RDF-S) by providing additional vocabulary along with a formal semantics.

Protégé is an ontology editor developed by Stanford University and is an open source software. It was designed excellently and it has many plug-ins. Its interface is graphic, and it has many tags to support the edit of Classes, Properties, Individuals, Forms respectively.

Jena [6] is a Java API developed by HP laboratory, and it is an API of RDF and OWL, and it also provides a reasoning function of OWL and RDF. As far as reasoning, Jena itself provides a reasoning engine based on rules. It implements

978-1-4244-6439-5/10/$26.00 ©2010 IEEE

reasoning with OWL by rules, but it can't support all reasoning function of OWL-DL.

In the paper, protégé is used to create and edit ontology and Jena development kit is used to reason with created ontology model in application program according to the actual needs.

B. Description Logic(DL)

Description logic is developed on the basis of propositional logic and first-order predicate logic, it is an important basis of reasoning of OWL.

The architecture of a standard description logic system is composed of four parts: 1)construction set to represent concepts and relations; 2)TBox(Terminology Box), i.e., an axiom set to describe domain structure, which includes the definition of concepts and axioms; 3)ABox(Asserted Box) ,i.e., an axiom set to describe concrete individual fact, which includes concept assert and relation assert; 4)reasoning system based on TBox and ABox. A knowledge base based on DL is K=TBox +ABox, simply recorded KB(T,A).

Description logic is a knowledge representation language used to describe concept and concept level relation. OWL DL of OWL recommended by W3C is an ontology language based on description logic.

III. ONTOLOGY REASONING APPLICATION EXAMPLE

The example uses Protégé that is a ontology modelling tool to create ontology model. The ontology describes six primitive concepts(primitive classes): people, man, woman, father, mother and parent, a property: hasChild, three individuals: George, Anna and John, a defined class: WhoHasChild. We use OWLViz(a plugin of Protégé 3.4.1) to show result ontology model as Fig. 1:

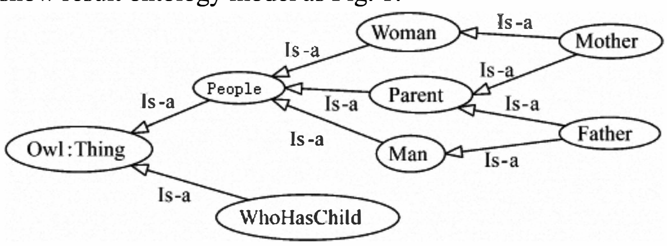

Fig. 1. Visual display of ontology model

Man, Woman and Parent are a child class of People class. Man class and Woman class have no intersection, therefore, they are mutually exclusive relation(declared by Disjoint). There is at least a hasChild relation in Parent class(minimum cardinality, i.e. minCardinality is one),it shows that People whose children are more than one is Parent, Father class is common child class of Man class and Parent class, similarly, Mother class is common child class of Woman class and Parent class. Therefore, we define that People that both is Man and is Parent is Father, in a similar way, we define that People that both is Woman and is Parent is Mother. Then three individuals: John, Anna and George are declared, they are individuals of People class. At the same time, we define that the child of Anna is George (related by property hasChild), John is Parent. Finally we define Defined class WhoHasChild(declared by equivalentClass) to reason and its necessary and sufficient condition is that WhoHasChild ≡hasChild.People≥1. According to given case, the example describes and concludes which People has child, namely, the individuals divided in WhoHasChild class is the result we need. Fig. 2 is a class structure figure available by running the command ,i.e., Compute inferred types of protégé.

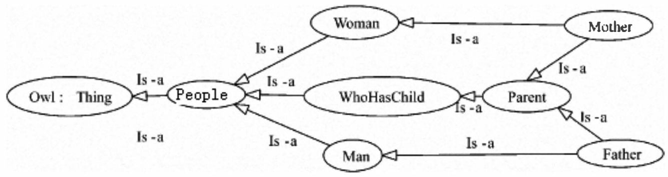

Fig. 2. Asserted hierarchy of protégé

It can be seen that the obtainment of the knowledge of ontology includes two aspects:

- to obtain the relations of classes and the relations of properties of TBox, more specifically, to obtain the subclass set of the class(namely including inference), the subproperty set of the property, the feature of the property, the domain and range of the property etc. It is close to the concept query, therefore, its inference reliability is higher. What corresponds with this is to divide Parent class under WhoHasChild class by reasoning in the example (because the property of Parent, namely, hasChild≥1 fits in with the necessary and sufficient condition of WhoHasChild class) and it shows that concepts belonging to Parent class is bound to belong to the concept of WhoHasChild class, in the mean time, WhoHasChild class is made the child class of People.
- to obtain the information of ABox, more specifically, to obtain all individuals of a class, properties of the individual, it is close to individual retrieval and information browsing ,therefore, the velocity and efficiency of reasoning is required highly, the integrity of the obtained information become the secondary requirement. By contrast, the individual of WhoHasChild class after reasoning is Anna in the example (individual Anna is People and simultaneously there exists the relation of hasChild, then, Anna is Parent, Parent is also the child class of WhoHasChild, therefore, Anna is the individual of WhoHasChild). That's individual category and its aim is that a described individual is classified as the class which is the most specific and can greatly reflect its characteristics. We can find Anna and John is People having a child.

IV. APPLICATION OF ONTOLOGY REASONING IN TAIYUAN HOSPITAL INFORMATION RETRIEVAL SYSTEM

A. Retrieval System Framework Based on Ontology

Retrieval system framework based on ontology is shown as Fig.3.

The system workflow is as follows:

1) The user submits his query requirement by the user interface.

2) The instruction converter gets and parses the query requirement, then, accesses the knowledge base based on ontology.

3) The knowledge base executes the query statements and returns the query results to the instruction converter.

4) The instruction converter receives the query results and shows the user results according to display mode.

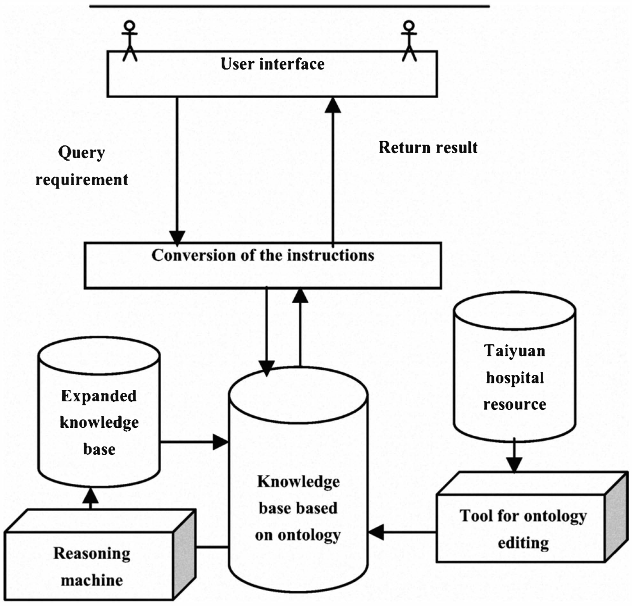

Fig. 3. Retrieval system framework based on ontology

From Fig.3, we can find an important work is to create the knowledge base before the all system is used. The knowledge base is increasingly updated and supplied. The reasoning results are served as supplemental knowledge base to add to prior knowledge base, therefore, the knowledge base got by the way can greatly improve the efficiency and precision of query.

B. The Creation of Taiyuan Hospital Ontology

In the paper, we do some research on hospital domain. After understanding domain ontology concepts and their mutual relations, we use Protégé 3.4.1(an open-source ontology building tool) developed by Stanford University to build manually an experimental hospital domain ontology (HDO). The HDO can provide semantic representation framework of text information and a conceptual hierarchy structure [7]. The part of HDO is described in Fig.4.

C. Expansion of Ontology Knowledge Base and Reasoning Implementation

In our system, we mine the knowledge of ontology in advance and use Jena to complete the excavation work of ontology knowledge. We create three inference rules:

Rule1: (?x belongs to ?y),(?y subclass of ?z)->(?x belongs to ?z)

Rule2: (?x has ?y),(?y same as ?z)->(?x has ?z)

Rule3: (?x subclass of ?y),(?y ward place at ?z)->(?x ward place at ?z)

Rule1 (semantic entailment) shows: if drug effect x belongs to a drug y, and y belongs to z, then drug effect x belongs to z.

For example: if we search what drugs antipyretic analgesic and anti-inflammatory drugs are, we can find all subclass of antipyretic analgesic and anti-inflammatory drugs, for example, aspirin, diflunisal.

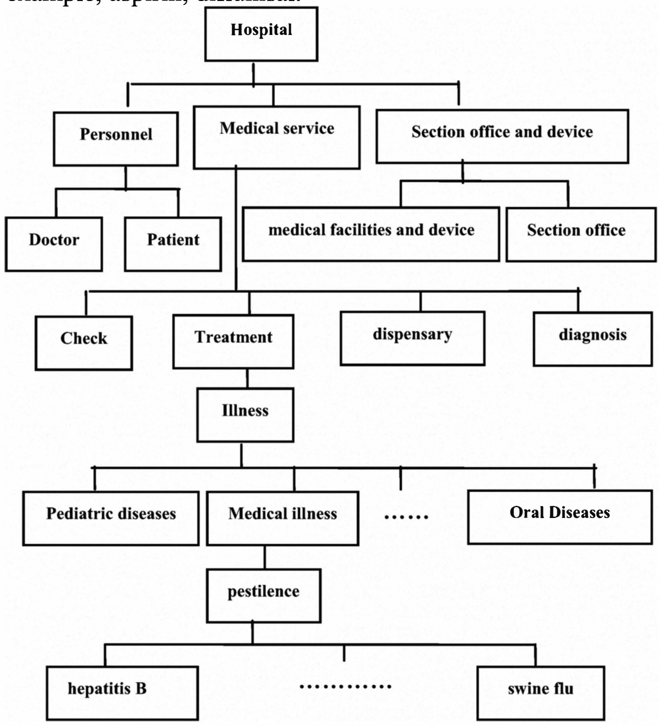

Fig. 4. Ontology hierarchy structure

Rule2 (synonymy expansion) shows: if patient x has illness y, and y is equivalent to z, then patient x has illness z. For example: if we search the patient having poliomyelitis (poliomyelitis , i.e., infantile paralysis), therefore, namely, we search the patient having infantile paralysis.

Rule3 (semantic association) shows: disease x is subclass of disease y, disease y ward is in place z, then, disease x ward is in place z. For example: the communicable disease ward is in third storey, phthisis is a kind of communicable disease, therefore, the phthisis ward is in three storey.

We add three rules to Jena inference engine, then, we perform reasoning based on relevant XML documents, the codes are shown as follows:

```
Model schema=
ModelLoader.1oadModel("file:data/CONCEPT.owl");
//read in the concept file of ontology CONCEPT.owl
Model data=
ModelLoader.1oadModel("file:data/INSTANCE.rdf");
// read in the instance file of ontology INSTANCE.rdf
String rules=
"[Rule 1:(? x belongs to ? y)(? y subclass of ? z)->(? x belongs to ? z)]"
+"[Rule 2:(? x belongs to ? y)(? y same as ? z)-> (? x belongs to ? z)]"]
+"[Rule 3:(? x subclass of ? y)(? y ward place at ? z)-> (? x ward place at ? z)]";
```

```
Reasoner reasoner=
new GenericRuleReasoner(Rule.parseRules(rules));
//add written rules to existing inference rules
```

Three rules are added to Jena inference engine, then, reasoning results are added to the knowledge base, therefore, the knowledge base is expanded.

D. The Contrast between Retrieval Based on Ontology and Retrieval Based on Keywords

The traditional retrieval of databases is based on keywords, yet, the ontology retrieval is based on knowledge, whose recall rate and precision rate are higher than retrieval based on keywords. The examples of specific query contrast are as follows:

1) When we create the knowledge base, antipyretic analgesic and anti-inflammatory drugs includes aspirin, diflunisal etc. When query condition is antipyretic analgesic and anti-inflammatory drugs, the instances with aspirin, diflunisal are also found. Yet, query based on keywords can find only antipyretic analgesic and anti-inflammatory drugs.

2) The query based on keywords always ignores synonymous information. For example, our system defines poliomyelitis and infantile paralysis as common concept. When we search poliomyelitis, we can also find infantile paralysis, but, it can not be implemented by the retrieval based on keywords.

3) The knowledge base based on ontology can be expanded by reasoning. For example, the communicable disease ward is in third storey, phthisis is a kind of communicable disease, therefore, the phthisis ward is also in three storey. But it cannot be realized by the retrieval based on keywords.

V. Conclusion

The definition of ontology is the explicit and formal specification of share conceptualization, thus, the use of ontology can bring about the share and reuse of knowledge, therefore, ontology has be used in information retrieval and question answering system increasingly. The paper generally introduces some required knowledge used in Web semantic reasoning, such as description logic, ontology description language OWL, ontology modeling tool protégé and the development kit Jena. On that basis, we use a simple example to illustrate the application of the knowledge in reasoning, then, we discuss in detail the reasoning application in Taiyuan hospital information retrieval system.

As the core of knowledge representation and answer extraction of question answering system of restricted domain, reasoning plays an important role in the system. With the broad application of OWL language in semantic Web domain, there appear many reasoning technology based on OWL. It will increasingly cause the fast development of ontology application of Web. All kinds of reasoning applications will be used to aid to create and apply the ontology of Web. It is necessary to point out when we apply the ontology, besides these basic inference, we can define many additional rules to realize the inference of high level which concentrates on concrete domain and concrete application. The instances of the example are stored in owl file, namely, the instances and the concepts are stored together. For mass instances of the ontology, we can store the instances and concepts separately, for example, the instances are stored in relational database. Jena itself provides permanent storing schema, besides, the Open Source Project The OWL Instance Store also provides the solution of mass data directed towards ABox inference [8].

The further work of the system we still need to do is:

- Input of data information should be more complete.
- To categorize the hospital information retrieval system should be more specialized. For example, the drug classification should be more detailed further.
- To continue to enrich inference rules and make retrieval more intelligent.
- Web service should be added to turn into distributed Taiyuan hospital information retrieval system in the part of knowledge base, so we can realize information share and mutual operation.

Acknowledgment

This research is supported by National Natural Science Foundation of China(No.60970059) and International Scientific and Technological Cooperation Projects of Shanxi Province(No.2009081022). The corresponding author of the paper is Chen Junjie.

References

[1] Zhang Weiming, Xiao Weidong, and Huang Kaige (2002), Semantic Information Model and Application, Beijing: Publishing House of Electronics Industry.(Chinese)

[2] Wang Jin, Chen Enhong, Zhang Zhenya, and WangXufa(2004), An Ontology-Based Cross Language Information Retrieval Model, Journal of Chinese Information Processing,18(3):1-8.(Chinese)

[3] Gruber T.A. Translation Approach to Portable Ontology Specification. Knowledge Acquisition,1993,5(2):199-220.

[4] Asunción Gómez-Pérez, Oscar Corcho. Ontology Specification Languages for the Semantic Web[J].IEEE Intelligent Systems, 2002,17(1):54-60.

[5] Mike Dean, Guus Schreiber. OWL Web ontology language reference[EB/OL]. [2009-01-20].W3C Working Draft, http://www.w3c.org/TR/2003/WD-owl-ref-200330331/,2003.

[6] restol.Jena2.ASemanticWebFramework[EB/OL].[2009-01-20].http://Jena.Sourceforge.net (Accessed Oct.12,2004).

[7] P. Paggio and D. H. Hansen, "Ontology-based Question Analysis in A Multilingual Environment: the MOSES Case Study", Workshop OntoLex 2004, held jointly with LREC 2004, LISBON, Portugal, May 2004.

[8] Instance Store website[EB/OL].[2009-01-20]. http://instancestore.man.ac.uk