

Heterogeneous Information Knowledge Construction Based on Ontology

Jianhou Gan¹, Gang Xie², Yongzheng Yan³, Wanquan Liu^{*4}

¹Key Laboratory of Educational Informatization for Nationalities Ministry of Education,
Yunnan Normal University, Kunming, China

^{1,2}Faculty of Metallurgical and Energy Engineering, Kunming University of Science and Technology,
Yunnan, Kunming, China

³School of Electrical Information and Engineering, North China University of Technology, Beijing, China

⁴Department of Computing, Curtin University, Perth, WA 6102, Australia

*Corresponding author, e-mail: w.liu@curtin.edu.au

Abstract

Describing and representing multi-source and heterogeneous knowledge is a popular research topic in recent years. After investigating knowledge forming process based on multi-source heterogeneous information resources, we present a new approach in which different information resources are put into a mutual RDF(S) data model, and semantic reasoning of RDF(S) is conducted. Moreover, a knowledge base construction framework for multi-source heterogeneous information source with combination of Ontology knowledge model is put forward, and an algorithm of knowledge base construction is also proposed, in which the core issues are knowledge inclusion and updating. Then the time complexity of our algorithm is analyzed. Finally, in order to solve the heterogeneous, and uneven horizontal of geographical distribution of ethnic minority information resources in Yunnan Province, we use the proposed method to construct a domain knowledge base for ethnic minority information resources, and use this model to evaluate the efficiency for the knowledge inclusion algorithm in responding time and indexing responding time for different data resources in our experiments.

Keywords: RDF(S), Ontology, knowledge base construction

Copyright © 2016 Universitas Ahmad Dahlan. All rights reserved.

1. Introduction

With the ubiquitous utilization of modern network technology and digital technology, knowledge based on data analysis from different resources has become an important factor for economic development. Currently, most Internet service is based on a distributed database integration work, and the rapid development of the Internet results in plenty of information and knowledge available in networks. These information and knowledge are usually incomplete located in separate sources, as well as in complex disorder and they cannot directly represent the new knowledge. Thus it is not easy to be understood by common users. As it is quite common to confront with the emergence of multi-sources, heterogeneous and mass information resources in the current Web services, traditional knowledge basis construction method cannot meet the rapid needs of users, thus it is necessary to develop a new knowledge construction method in order to provide rapid new knowledge services. In this paper, we aim to construct the knowledge basis in a different way based on the study of the unified knowledge representation for multi-source heterogeneous information resources with an aim to provide a new model of knowledge service.

The essence of knowledge basis construction is to represent knowledge reasonably and effectively, namely effective knowledge representation [1-5]. There have been plenty of studies on the methods of knowledge representation. Xu and Ye [6] have studied the merits and disadvantages of currently used knowledge representation methods, and claimed that Ontology has great potential in knowledge representation [6]. Fensel, et al., [7] developed the On-to-Knowledge system, a knowledge management system based on Ontology, which aims at solving the limited information sharing problem caused by traditional knowledge basis via using the key-word matching method to search knowledge. In addition, researches on knowledge basis system construction have attracted much attention [8-10], among which a representative

work is KMSphere [8], a knowledge management system developed by Institute of Computing Technology Chinese Academy of Sciences. Cheng [8] stated that the knowledge management system based on Ontology is mainly to solve two issues, the construction of knowledge basis and the evolution of knowledge basis. Researchers have shown the advantages of Ontology on knowledge representation in [8-12]. In this paper, we intend to construct the knowledge basis by adopting Ontology with an aim to improve the restricted information researching in traditional knowledge management system. The traditional knowledge base mainly generate the knowledge base through the simple knowledge discovery. The traditional knowledge base processing is a data processing. Knowledge is searched in the knowledge base according to the needs of the users. Facing numerous and jumbled network knowledge, the traditional knowledge base processing is only limited to the data level, and cannot achieve information interaction.

There are massive information resources in Internet, in which the knowledge formation process is associated with multiple data sources; meanwhile, the systems, grammar, structure and semantics of resources are heterogeneous; hence, these information resources are named as multi-source heterogeneous information resources. In order to construct the multi-source heterogeneous information knowledge basis, the first is to solve the semantic heterogeneity of these information resources [13-18]. Aiming at solving the differences and inter-operable applications, the middleware solution to integrate the heterogeneous multi-source database system is put forward in [19, 20]. Focusing on the accuracy and speed requirements of fusion node, the algorithm of matrix decomposition is proposed [21-24]. This algorithm not only can accelerate the speed of data fusion, but also reach high accuracy when unsymmetrical block matrix happened or some relation matrix lost.

The solution to the integration of heterogeneous data includes XML, SOA, data warehouse, etc.. XML is a general method for structured data representation, and it allows the application procedure to store and transmit the data that can be understood by other application procedure, and separates the format and content of the data from the processing methods. SOA (service-oriented Architecture) is a component model. It links different functional units of the application procedure to a contract by the well-defined interface. Data warehouse technology, as a solution to the integration of heterogeneous databases, not only integrates the data located in different regions through data extraction and transfer tool, the data on different operating systems, different data structures together with certain data patterns, but also ensures the data consistency.

The early version of Ontology language is similar to XML, which could be simply catheterized into XML. With the development of Ontology languages, most of which are based on XML. RDF is a data model used to describe objects and their relationship. It provides simple language description by using XML [25-28]. RDFS is a simplified Ontology description language, which is regarded as the vocabulary description language of RDF [29-31]. Using the pre-defined terms, RDFS enables users to define the category, property and the relationship between the mentioned two objects. As a description language of Ontology, RDFS can (a) define the resources and their categories, (b) define the properties and describe the relations between objects, and (c) define the relationships among different categories and various properties [32, 33]. Now RDFS is widely used in Semantic Web, Intelligent search engine, data exchange in the semantic layer, automatic link and reference of information, digital library and so on.

In this paper, we first study the information processing based on multi-source heterogeneous resources, by which we describe RDFS for the information, and transforms diverse databases into one mutual RDFS model; then based on the Ontology knowledge model, we present the knowledge basis construction procedure for multi-source and heterogeneous information resources, and also develop an algorithm for knowledge basis construction; finally, we evaluate the applicability of proposed algorithm through a specific knowledge basis construction.

The rest of the paper is organized as follows. Section 2 gives RDFS description for multi-source heterogeneous information resources; Section 3 constructs knowledge basis framework for multi-source heterogeneous information resources; Section 4 puts forward the algorithm for knowledge basis construction with multi-source heterogeneous information resources; Section 5 reports the experimental results; and Section 6 concludes this paper.

2. RDFS Description for Multi-source Heterogeneous Information Resources

2.1. Knowledge Formation Process for Multi-source Heterogeneous Information Resources

In this paper, we are only aiming at specific application sphere and solving the “information silo” issue in sharing and communicating of multi-source and heterogeneous information resources. In terms of the knowledge structure, most of currently used knowledge sources are hard-analyzed unstructured text, semi-structured web page information in complex disorder, or widely used connected databases. In this case, researching on the knowledge formation process for multi-source and heterogeneous information resources will be beneficial if we can develop an efficient method to eliminate the semantic heterogeneity among various data sources; on the other hand, it will be also helpful to transform the unsolved new information resources into a mutual knowledge description. Focusing on different data sources, putting data into mutual RDFS model and describing data with the RDFS are a critical step for knowledge basis construction.

Figure 1 shows the knowledge formation process for multi-source heterogeneous information resources and it can be divided into two steps. Firstly, populating different data sources to an RDF model, and secondly, forming the knowledge model through RDFS semantic reasoning process. Next, we will address these two issues separately.

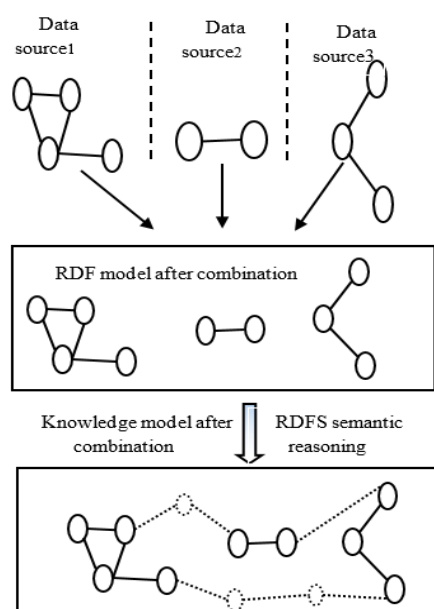


Figure 1. Knowledge formation process of multi-source heterogeneous information resources

2.2. RDFS Description and Semantic Reasoning

2.2.1 RDFS Description

RDFS is a collective word for RDF (Resource Description Framework) and its extension RDF Schema, which is an assertion language using standard vocabulary to represent commands. The basic conception is to use simple statements to represent resources. Each statement contains three parts, namely *Subject*, *Predicate*, and *Object*. In order to adopt RDFS for better describing multi-source heterogeneous information, knowledge is understood to be as a combination of a series of resources. RDFS uses property and property value to describe resources. One RDFS description is defined as following:

Definition 1: *Statement* ::= <subject, predicate, object>

Subject is used to describe RDFS resource, *predicate* is for some specific factor or characteristic of subject, or its relationship to other properties, and *object* is the property value which could also be a *subject*.

As to specific sphere, RDFS offers knowledge description for multi-source heterogeneous information. Based on Definition 1, one has the following steps for RDFS descriptions:

1. Establishing a common vocabulary or set of resources for a specific sphere, among which the common vocabulary or set of resources should be easy to understand and can be used consistently for description;
2. Using the RDF Schema language to establish the common vocabulary in some specific application areas;
3. For the new RDFS vocabulary, one should describe the class, property and resources as a whole, so as to provide a good basis for modeling a specific sphere;
4. Adopting RQL (RDF Query Language) to query for one or more RDF or Schema mode, and returning to the corresponding variable to bind the list [34, 35].

In the RDFS steps mentioned above, it usually uses several construction entities, which include: (1) The main entity is *rdfs:resource*, and two subclasses are *rdfs:class* and *rdfs:property*; (2) The core characteristics includes *rdfs:subClassOf*, *rdfs:subPropertyOf*, *rdfs:type*; (3) The core constraints are *rdfs:range*, *rdfs:domain*, *rdfs:constraintProperty*, *rdfs:constraintResource*.

In the following part of this paper, we give the relevant definitions of RDFS Description and Semantic Reasoning, then a new method for different information resources being put into a mutual RDFS data model is presented. Next, a knowledge basis construction framework is put forward with combination of Ontology knowledge model, and an algorithm of knowledge basis construction is developed.

2.2.2. RDF(S) Semantic Reasoning

In the knowledge formation process for multi-source heterogeneous information resources, RDFS description is merely used to standardize the RDF/XML serialization statements. However, only the formal representation of grammar can be conducted for machine implementations, which cannot avoid the ambiguity of RDF language comprehension. In the programming implementation, judgment on the true of RDF (S) statement is required to implement RDFS automated reasoning. Therefore, this paper offers several important conceptions, as the theoretical basis of RDFS reasoning; in the meantime, owing to the fact that RDFS as ontological description language, these theoretical developments are not only the inevitable evidence for automated reasoning, but also suitable for the Ontology reasoning in knowledge basis construction.

Definition 2: *Brief explanation on RDF graph.* As for the V in RDF graph denoted as $I = \langle IR, IP, IEXT, IS, IL \rangle$, where, IR denotes the non-empty resources set and is called the domain of explanation I , IP is the property resources set, $IEXT$ is the resources set whose property is mapped into resources-resources collection, that is $IEXT: IP \rightarrow 2^{IR \times IR}$, IS is a mapping from URI into resources or property, that is $IS: URI \rightarrow IR \cup IP$, IL represents a mapping from the typed literals of V into resources, that is $IL: \text{typed literals} \rightarrow IR$.

From Definition 2, RDF semantic language should firstly map V in the RDF graph into the elements in domain IR , then the elements are mapped as the binary relation in domain IR . Thus, under the simple interpretation of a given RDF graph, it can be clearly defined by the RDF graph assignment which is as a true value judgment method of RDF graph.

Definition 3: *Assigning RDF graph.* One RDF graph will give the following assigning rules:

- (1) If E is a non-type argument, then $I(E) = E$;
- (2) If E is a type argument, then $I(E) = IL(E)$;
- (3) If E is URI, then $I(E) = IS(E)$;
- (4) If E is a non-empty triple node $\langle s, p, o \rangle$, then $I(E) = \text{True}$, and only when, $s \in V, P \in V, o \in V, I(P) \in IP$ and $\langle I(s), I(o) \rangle \in IEXT(I(P))$
- (5) If E is an RDF graph, then $I(E) = \text{False}$, and only when E contains certain triple node E^* which makes $I(E^*) = \text{False}$, otherwise $I(E) = \text{True}$.

In RDF semantic reasoning, RDF graph assigning is a judgment method on defining true value. On the basis of this, RDF semantic specification is introduced to derive a concept for

conducting the true statement over other true statements, consequently it will lead to RDF implication concept and RDFS implication concept.

Definition 4: *Simple implication rules.* Given S and E are both RDF graph, if every simple explanation satisfies that S can meet E , then it is called S simple implicating E ; in other words, if each model of RDF graph S is E 's model, then S simply implicates E .

In practice, defining implication rules is mainly used to achieve machine reasoning. The basic conception is that if an RDF graph contains some forms of triple node $\langle s, p, o \rangle$, after applying implication rules, the triple node can be added into the RDF graph. RDF semantics specification [21] mainly defines three implication rules, namely the simple implication rules, RDF implication rules and RDFS implication rules.

Definition 5: *RDF implication rule.* Given S and E are two RDF graphs, if one graph which applies the simple implication rule of S or RDF implication rule simply implicates E , then it is called S RDF-implicating E .

Definition 6: *RDFS implication rule.* Given S and E are two RDS graphs, if one graph which applied simple implication rule of S , RDF implication rule, and RDFS implication rule simply implicates E , then it is called S RDFS-implicating E .

In other words, S RDFS-implicating E if and only if RDFS closure of S simply implicates E , equivalent to the concept of formal grammar RDFS deduction. It guarantees the validity of the RDFS logical reasoning for semantic concept, and provides a theoretical foundation on knowledge representation reasoning.

3. Knowledge Basis Construction Framework for Multi-source Heterogeneous Information Resources

3.1. Ontology Knowledge Basis Model

The difference between new knowledge basis framework and traditional DBMS (Database Management System) is that traditional DBMS is unable to represent and deal with rule-based knowledge, but new knowledge basis has a uniform symbol and structural model, which is a rational collection of description knowledge and procedural knowledge in a specific sphere; on the other hand, Ontology, as a knowledge representation method, can effectively represent the concepts of structure and the relations between concepts. So it can better achieve the "shared conceptualization" [36-41]. In certain applications, we combine the two in a reasonable way, and then provide a new knowledge basis model, intending to provide required data and normative instructions for the knowledge structure, and build a theoretical basis for building a knowledge basis framework.

Definition 7: *Ontology representation of knowledge basis model.* $\langle \text{Knowledge Model} \rangle :: = \langle \text{Domain Knowledge} \rangle \langle \text{Reason Knowledge} \rangle \langle \text{Task Knowledge} \rangle$, in which Domain Knowledge represents field knowledge used for a detailed description of a particular field of knowledge type; Reason Knowledge represents reasoning and or methodology knowledge, describing the reasoning methods or steps of general knowledge in specific areas, such as matching, a generator, an inference engine, and other basic constructions; Task Knowledge represents the task knowledge, which describes the target knowledge of the system to be achieved in stages, including the sub-tasks of decomposition in the reasoning process and target knowledge in reasoning.

Definition 7 can represent the three-level knowledge system of "facts-concept-rule", but the Ontology knowledge model is only capable of better generalizing and abstracting the knowledge representation. The current knowledge basis framework cannot well represent the update process, thus Definition 7 is used as a supplementary. Next, the Ontology management is applied in the construction of the proposed knowledge basis framework.

3.2. Knowledge Basis Framework

Based on the above analysis, we offer a multi-source heterogeneous information resources knowledge basis framework, as shown in Figure 2.

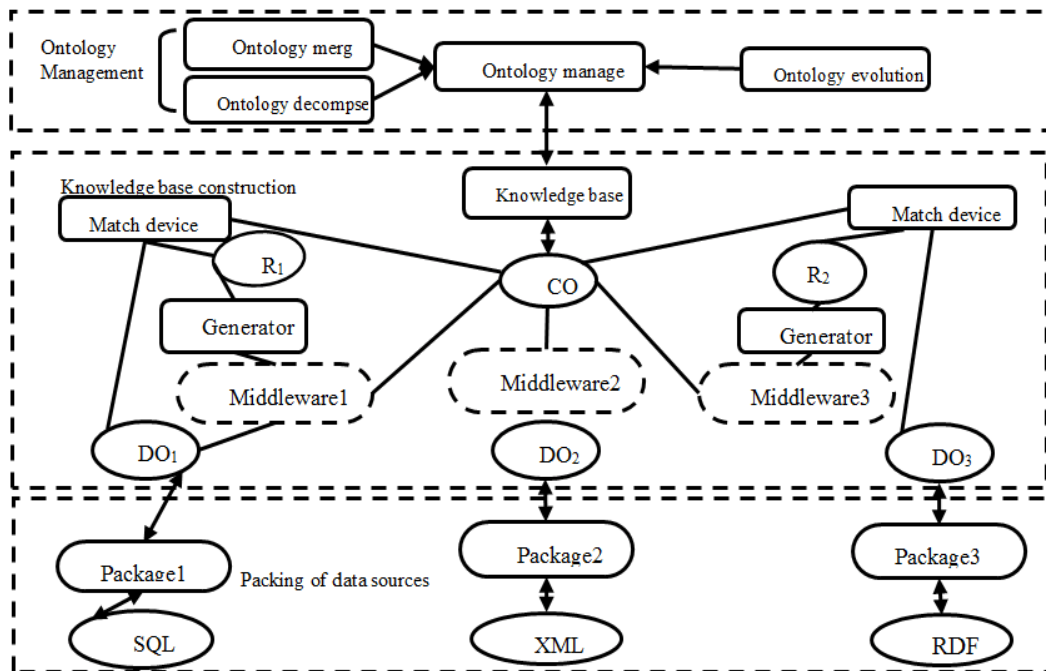


Figure 2. Multi-source heterogeneous information resources knowledge base framework

3.3. Different Data Sources into RDF model

It is required to have different data sources, thus converting different data sources (SQL, XML, RDF) into DO_i through a wrapper i . The converting method for different types of data varies. In application, wrapper 1 is the traditional wrapper (mainly focusing on XML or traditional data which can be structurally described), which one can use the Velocity [42] for converting. Wrapper 2 is a relational database package wrapper, we can use D2RQ, SquirrelRDF, Virtuoso and other tools to convert into RDF graph, and then access through SPARQL. Wrapper 3 is an associated data package wrapper (mainly for the RDF data model in semantic web), which use Pubby [43] (associated data front end) to map URI in supporting Web browser for the packaging process. The wrapper i in the process is based on the RDF (S) description and semantics reasoning.

3.4. Process of Database Construction

One can match DO_i and CO by matching device into the alignment rule R_i (R_i is the generating alignment rule matching device). Given R_i , through intermediate file i , interactive data sources can be established between data (XML, SQL, RDF) and CO. When the data can be described according to the Ontology and then converted into RDF, we can convert the data into the knowledge basis, and provide an interface to outside access, in we can use RDF query language, SPARQL, for knowledge retrieval processing. The whole process is an implementation of ontology knowledge theoretical model.

3.5. Ontology Management

Ontology management is mainly focusing on knowledge updating process, which acts on the entire cycle of knowledge basis construction, including ontology merging, ontology decomposition, and ontology evolution. Jena [44] is convenient tool for merging and decomposing the ontology of RDF data model. The evolution of Ontology requires us to define some relevant rules, as shown in Figure 3.

Through matching, one can establish the relationship between the old O_t and the new O_{t+n} , and then generate the alignment rule R_t . Given R_t , through generators, conversion model can be generated, and then it is converted into I_{t+n} .

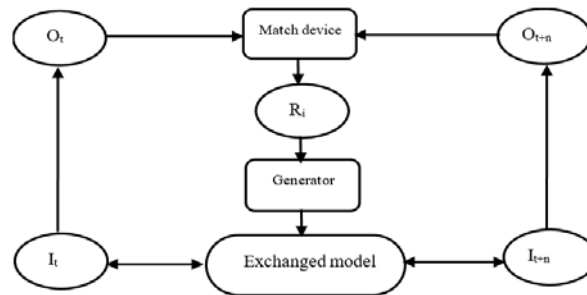


Figure 3. Ontology evolution process

4. Construction Algorithm for Heterogeneous Information Knowledge Base

Based on the knowledge basis framework discussed previously, we can develop a multi-source heterogeneous information resources knowledge basis construction algorithm, which is described in the following steps.

First we analyze different data sources, and return to the triple node required by the RDF. We then use interfaces and methods provided by Jena for further processing. Firstly, one creates the RDF model using ModelFactory, and then using the *read()* function to read the RDF data, finally applying the iterator StmtIterator to return to the triple node collection.

For a particular field of knowledge, one can implement the knowledge inclusion algorithm and knowledge updating algorithm as described below.

Algorithm 1: Knowledge Inclusion Algorithm

$S ::= \langle \text{subject, predicate, object} \rangle [\langle \text{subject, predicate, object} \rangle]$ let that P is defined as a triple node in S . Data type is the double collection of the key value of Map (key, value). The key mode of P (key, value) is (predicate, {subject, object}), in which the key is the ID;

Input: knowledge included

Output: KB

Step 1: If $S \in KB$, then stop; otherwise go to Step 2;

Step 2: Let $T = S.subClassOf$, and CO be the ontology knowledge base, if $T \notin CO$, then output error and stop; otherwise extract one triple node P from S , turn to Step 3;

Step 3: If $P.predicate = "ID"$, then give key value of ID to $P.object$ and then match to the property collection PS of KB ; if $P.predicate \notin PS$ according to the inherited relation of property in RDF description, include $predicate$; if $P.predicate \in "PS"$, and let $PS = PS \cup \{predicate\}$ then turn to Step 4;

Step 4: Let $S = S - \{P\}$, if $S = \emptyset$ then stop, otherwise extract one triple node P from S , turn to Step 2.

Algorithm1 gives the knowledge inclusion process for multi-source heterogeneous information resources knowledge basis construction, Next we present the Knowledge updating in Algorithm 2.

Algorithm 2: Knowledge Updating Algorithm

Input: Key, ID and triple node needed to be updated;

$PS ::= \{ID\} \langle \langle \text{subject, predicate, object} \rangle \rangle$;

Output: KB

Step 1: If $Key \notin KBID$, it implies that the knowledge basis does not include the knowledge about ID, and stop; otherwise, make $PS = PS - \{ID\}$, extract one triple node P , turn to Step 2;

Step 2: Given $P.predicate$, if $P.predicate \notin KB.predicate$, it implies that the ontology knowledge described is not in the knowledge basis, then search P , judge whether $P.predicate$ is one sub-type of existing resources; if yes, instantiate P ; if no, let $KB.predicate = KB.predicate \cup \{P.predicate\}$ then instantiate it. If $P.predicate \in KB.predicate$, instantiate directly, and then turn to Step 3;

Step 3: Let $PS = PS - \{P\}$, if $PS = \emptyset$, then stop; otherwise extract one triple node P from PS , then turn to Step 2.

From above mentioned two algorithms, the algorithms mainly traverse through triple node. By analyzing the property of triple node, one can achieve expansion or modification of knowledge basis. When the search of triple node ends, these algorithms also stop. For limited number of triple node, the steps of these algorithms are also limited, then the computational time of algorithm is $O(n)$ where n is the number of the triple nodes.

By using the RDF query language SPARQL one can retrieve knowledge from this model. First, input SPARQL as a string, then parse the string to generate an abstract syntax, and then define query rules by algebra operators provided by SPARQL, finally calculate the results in RDF graphs.

The knowledge basis construction algorithm for multi-source heterogeneous information resources includes the above algorithm 1 and the algorithm 2.

5. A Practical Example

Yunnan is an area with rich minority resources. Currently there is not a comprehensive understanding on different material culture among 26 ethnic minorities in Yunnan. The major reason is that the nature of these national resources is heterogeneous, imbalance, and disordered in geographical distribution. By using knowledge basis construction method proposed in this paper, one can construct the multi-source heterogeneous minority information resources knowledge basis, the steps are shown as follows.

We can use Protégé [45] to build the Ontology knowledge basis in ethnic minority domain, and form the RDF graph between domain ontology and application as shown in Figure 4. The upper part of Figure 4 is the ontology section of RDF graph, describing concepts like nation, ideology, customs, etc., and relations like religion property, language property, etc. The lower section of this figure represents the real minority Wa described by Ontology and its related information. Two sections are distinguished through the RDFS.

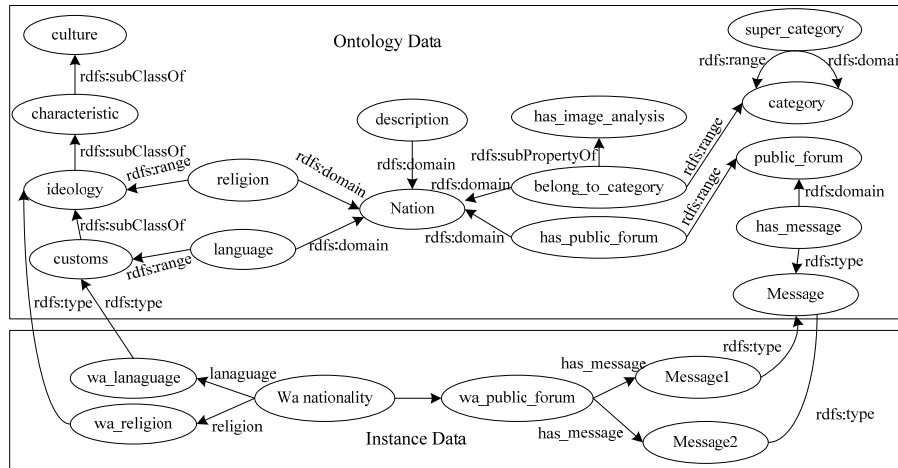


Figure 4. RDF Relation between domain ontology and application example

By using interfaces and methods provided by Jena, one can return the triple node needed for RDF description, and then search through triples in loops. Reasoning rules based on the RDFS are given in Table 1. Next we construct the knowledge basis based on the knowledge including algorithm.

The experiment has used a high-performance server for data processing. The system environment is java: jdk1.5.0. The maximum heap threshold is 256M. As we aim at different data sources, ethnic information resource data set is acquired as shown in Table 2.

We have recorded the responding time of executing including algorithm of each data, as shown in Figure 5. From Figure 5, we know that the responding time of including algorithm of different data sources is different, the responding time of Web page is better than other two data sources in the same three tuple number.

By applying SPARQL query language with the proposed knowledge updating algorithm, we can search the triple node, and finally return the relevant search results on RDF graphs. We aim to search and return the "nation" and "culture" as described in RDF data source on "http://ethnic.ynnu.edu.cn/ethnic". Thus we can construct the following knowledge basis model:

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX nation: <http://ethnic.ynnu.edu.cn/ontology/nation#>
SELECT ?X ?Y
FROM <http://ethnic.ynnu.edu.cn/ethnic/ethnic>
WHERE { ?Y nation: culture ?X }
```

Table 1. RDF(S) reasoning rules

Rule Name	Known Condition	Reasoning Results
rdfs1	<s,p,o>(if o is literal)	_n:rdf:type rdfs:Literal
rdfs2	(p rdfs:domain x)&<s,p,o>	s rdfs:type x
rdfs3	(p rdfs:range x)&<s,p,o>	o rdfs:type x
rdfs4	<s,p,o>	s o rdfs:type rdfs:Resource
rdfs5	(p rdfs:subPropertyOf q)&(q rdfs:subPropertyOf r)	p rdfs:subPropertyOf r
rdfs6	p rdfs:type rdfs:Class	p rdfs:subPropertyOf p
rdfs7	<s,p,o>&(p rdfs:subPropertyOf q)	<s,p,o>
rdfs8	s rdfs:type rdfs:Class	s rdfs:subClassOf rdfs:Resource
rdfs9	(s rdfs:type x)&(x rdfs:subClassOf y)	s rdfs:type y
rdfs10	s rdfs:type rdfs:Class	s rdfs:subClassOf s
rdfs11	(x rdfs:subClassOf y)&(y rdfs:subClassOf z)	x rdfs:subClassOf z

Table 2. Descriptive data of experiments

Data Source	File Number	Example Number	Triple Node Number	Database Size (MB)
Relation Database	2	134	2367	1.5
XML	34	767	4783	2.38
Web Page	67	1023	8921	3.7

Being different from the traditional algorithm, using the proposed knowledge including algorithm and the knowledge updating algorithm is better in the following knowledge inference: (1) The traditional algorithm can only be on string matching for the relational databases and XML, which are two data sources. But, using knowledge including algorithm and knowledge updating algorithm, the knowledge unit with semantics will be integrated, and it is easy to understand the semantic relation between the knowledge units to the machine. (2) For RDF data sources, the traditional algorithm can only use the normal form of RDF to directly link the knowledge, and establish the relevance among all types of file storage data, and it lacks semantic information. But, Using the knowledge inclusion sub-algorithm and the knowledge updating sub-algorithm to extend knowledge on the basis of the traditional RDF triples, that is RDF= {Resource, Attribute, Resource Type, Attribute Type}, and the characteristics of the resources and the property is retained, then one can discover new knowledge.

By applying the knowledge updating algorithm, we can search the triple node, and the responding time of different data sources is shown as Figure 6. From Figure 6, we know that the responding time of updating algorithm of different data sources is different, the responding time of Web page is slightly better than other two data sources in the same query. In the single pseudo-distributed environment, we can select the Query SPARQL query in Figure 6 to analyze the response time of RDFS inference model using the traditional algorithm and our algorithm on RDF data sources, via combining with the sub-algorithm knowledge updating.

The results are shown in Figure 7. Through the comparative analysis of the experimental result in Figure 7, we can draw two conclusions: First, with the increasing of the number of the triples data source, the response time of the RDFS inference was significantly increased; Second, the RDFS inference response time to use the improved sub-algorithm knowledge updating is less than the traditional algorithm.

In view of the above experimental analysis, the knowledge basis construction method proposed in this paper is effective. It is better than the previous unimproved knowledge basis in terms of RDFS inference time. Namely, by firstly parsing the different data sources and then executing the sub-algorithm of knowledge inclusion and the sub- algorithm Knowledge updating to build the knowledge basis is promising.

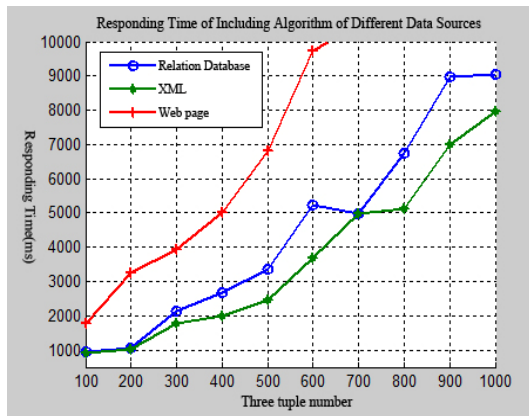


Figure 5. Responding time of including algorithm of different data sources

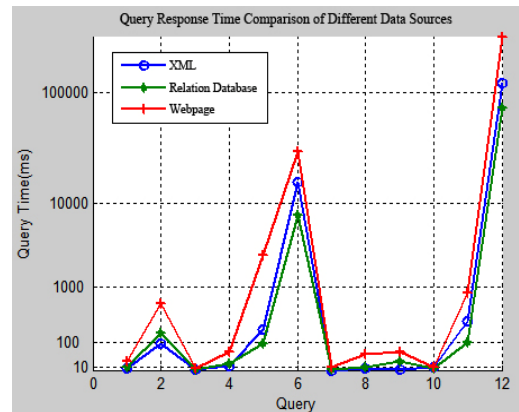


Figure 6. Responding time of updating algorithm of different data sources

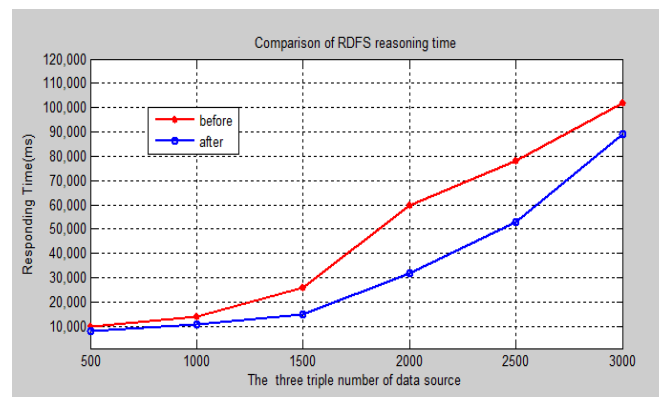


Figure 7. Responding time of RDF inference between the traditional algorithm and our algorithm

6. Conclusion

In knowledge engineering field, domestic and foreign research institutions and scholars have made much effort on knowledge basis construction, but there is not a comparatively complete knowledge system. In this paper, through studying the knowledge formation process of multi-source heterogeneous information resources, we used the RDFS description for semantic reasoning, and analyzed the Ontology knowledge model representation; Then, we provided a framework with knowledge inclusion and updating algorithms for multi-source heterogeneous information knowledge; Finally we demonstrate the effectiveness of the algorithm through a practical application.

Acknowledgements

The research is supported by The research is supported by the National Nature Science Fund Project (61262071), Key Project of Applied Basic Research Program of Yunnan Province (2016FA024).

References

- [1] Schreiber G, Akkermans H, Anjewierden A, De Hoog R, Shadbolt N, Wielinga B. Knowledge Engineering and Management: The Common KADS Methodology. Beijing: China Machine Press. 2003.
- [2] Swain M. Knowledge Representation. Encyclopedia of Systems Biology. New York: Springer. 2013: 1082-1084.
- [3] Jakus G, Milutinović V, Omerović S, Tomažič S. Concepts, Ontologies, and Knowledge Representation. New York: Springer. 2013: 47-62.
- [4] Mendes E. Practitioner's Knowledge Representation-A Pathway to Improve Software Effort Estimation. Berlin, Heidelberg: Springer. 2014.
- [5] G Zhang, Y Li. Multiple Disciplines Product Design Knowledge Representation Strategy Based on Ontology and Semantic Network. *TELKOMNIKA Indonesian Journal of Electrical Engineering*. 2013; 11(10): 6074-6079
- [6] Xu BX, Ye PH. Research on the method of knowledge representation. *Information Science*. 2007; 5.
- [7] Fensel D, Van Harmelen F, Klein, M, Akkermans H. On-To-Knowledge: Ontology-based tools for knowledge management. *Computer*. 2002; 35: 56-59.
- [8] Cheng Y. Research on Ontology-based Uncertain Knowledge Management. Dissertation for PhD Degree. Chinese Academy of Science; 2005.
- [9] Zhenglian S, Jun Y, Yonghua Z, Yongqiang Zh. Ontology-based Equipment Support Knowledge Management Model. *Journal of Equipment Academy*. 2015; 4: 62-66.
- [10] Ning G, Hongjun G. Research on the function of PDM knowledge management system based on the ontology. *Machine design and Manufacturing Engineering*. 2015; 9: 58-62.
- [11] Yang F, Xing Y, Sun H, et al. An Ontology-Based Semantic Similarity Measure Considering Multi-Inheritance in Biomedicine. *Mathematical Problems in Engineering*. 2015; 1: 1-9.
- [12] Cai GJ, Zhao B. An approach for composing services based on environment ontology. *Mathematical Problems in Engineering*. 2013: 1-11.
- [13] Wang JG, Du YY, Su FZ, Zhou CH. Methodology and implementation of spatial data integration based on geo-ontologies. *Geographical Research*. 2009; 28: 696-704.
- [14] Maree M, Belkhatir M. Addressing semantic heterogeneity through multiple knowledge base assisted merging of domain-specific ontologies. *Knowledge-Based Systems*. 2015; 73: 199-211.
- [15] Lourenzutti R, Krohling RA. Ageneralized TOPSIS method for group decision making with heterogeneous information in a dynamic environment. *Information Sciences*. 2016; 330: 1-18.
- [16] Zhang F, Ignatius J, Yajun Zh, Lim Ch P, Ghasemi M. An improved consensus-based group decision making model with heterogeneous information. *Applied Soft Computing*. 2015; 35: 850-863.
- [17] Zhong Zh, Zhiwei Zh, Shuang L. Cross domain boosting for information fusion in heterogeneous sensor-cyber sources. *Signal Processing*. 2015; 4: 1-7.
- [18] Jingjun G, Fanzhi K, Ming Zh, Jianfeng T, Xin L. Research on heterogeneous data integration service in data-intensive application. *Computer Engineering*. 2015; 41: 60-65.
- [19] Xia YL. Research on Knowledge Integration Algorithm for Multi-source Heterogeneous Information Source. Dissertation of MSci Degree. Yunnan Normal University; 2014.
- [20] Anmei Zh. Research on management and its application of multi-source heterogeneous knowledge about wind power equipment based on ontology. Dissertation of PhD Degree. Hunan University; 2015.
- [21] RDF semantics. W3C recommendation. 2004.
- [22] Kim K, Moon B, Kim HJ. R3F: RDF triple filtering method for efficient SPARQL query processing. *World Wide Web-internet & Web Information Systems*. 2013; 18: 317-357.
- [23] Kim K, Moon B, Kim HJ. RG-index: An RDF graph index for efficient SPARQL query processing. *Expert Systems with Applications*. 2014; 41: 4596-4607.
- [24] Kim J, Park S. Efficient authorization conflict detection using graph labeling in controlling access to RDF. *Computer Standards & Interfaces*. 2015; 38: 51-63.
- [25] Fensel D, Harmelen FV, Horrocks I, McGuinness DL, Patel-Schneider PF. OIL: An Ontology infrastructure for the semantic web. *IEEE Intelligent Systems*. 2001; 16: 38-45.
- [26] Xiaoping Y, Yong T, Zhibo ZH, Zhaoying CH, Yanchong L. Study and Implementation on Semantics-Based Cooperative Temporal XML Index. *Chinese Journal of Computers*. 2014; 37: 1911-1921.
- [27] Fangfang X, Yaoyao L, Jinguang G. Semantic cache replacement strategy for XML algebra-based query optimization. *Wuhan University Journal of Natural Sciences*. 2015; 20: 165-172.
- [28] Qiushi L, Qiuyue W, Shan W. Query understanding for XML keyword search. *Journal of Software*. 2012; 23: 2002-2017.
- [29] Furche T, Linse B, Bry F, et al. *RDF querying: language constructs and evaluation methods compared*. In Proceedings of Reasoning Web. Berlin, Heidelberg. 2006: 1-52.
- [30] Jin-rong X, Jun ZH, Qiu-sheng A. RDF(S) repository privacy protection via inference control. *Journal of Beijing Institute of Technology*. 2014; 23: 495-501.
- [31] Heyang X. Study of Inference Control Algorithm for RDF(S) on Semantic Web. Dissertation of MSci Degree. Beijing University of Posts and Telecommunications; 2015.

- [32] Cai M, Frank M. *RDF peers: A scalable distributed RDF repository based on a structured peer-to-peer network*. Proceedings of the 13th international conference on World Wide Web. New York. 2004: 650-657.
- [33] Runan Zh. Study of transformation method from XML to RDF(s). Dissertation of MSci Degree. Northeastern University; 2012.
- [34] Liu SP, Lin ZS, Mei J, Yue AB. A model-theoretic semantics for XML. *Journal of Software*. 2006; 17: 1089-1097.
- [35] Xiangyue J. Study of query way of fuzzy SPARQL of RDF. Dissertation of MSci Degree. Northeastern University; 2012.
- [36] Magkanaraki A, Karvounarakis G, Anh TT, et al. *Ontology Storage and Querying*. Foundation for Research and Technology Hellas, Institute of Computer Science, Information System Laboratory. Report number: 308. 2002.
- [37] Tiantian H, Ming C. Research of OWL ontology storage and query based on BaseX. *Computer Engineering and Applications*. 2015; 51: 132-136.
- [38] Suihua W, Xiao-dan ZH, Yue W. OWL Ontology storage and query method based on relational database. *Journal of Henan Normal University (Natural Science)*. 2012; 40: 159-163.
- [39] Ying L, Qiming L, Zhipeng ZH, Yongliang D. Ontology-based knowledge modeling for automated construction safety checking. *Safety Science*. 2015; 79: 11-18.
- [40] Slota M, Leite J, Swift T. On updates of hybrid knowledge bases composed of ontologies and rules. *Artificial Intelligence*. 2015; 229: 33-104.
- [41] Sijie ZH, Frank Boukamp, Jochen Teizer. Ontology-based semantic modeling of construction safety knowledge: Towards automated safety planning for job hazard analysis (JHA). *Automation in Construction*. 2015; 52: 29-41.
- [42] Tao Ch, Xuxiang B. Application of Velocity Template Engine Technology in Java Web. *Agriculture Network Information*. 2011; 5: 38-40.
- [43] Pubby. <http://wifo5-03.informatik.uni-mannheim.de/pubby/>.
- [44] Fenghua H, Luming Y. Reasoning of ontology model for typhoon disasters domain based on Jena. *Journal of Computer Applications*. 2013; 33: 771-775.
- [45] Zhe L, Yi L. Building of College English Teaching Resource Ontology based on Protégé. *Journal of Guangxi University for Nationalities*. 2013; 9: 59-62.