TELKOMNIKA Indonesian Journal of Electrical Engineering
Vol.12, No.1, January 2014, pp. 167~ 178
DOI: http://dx.doi.org/10.11591/telkomnika.v12i1.3889

Research on Data Integration of the Semantic Web Based on Ontology Learning Technology

Hongsheng Xu, Ruiling Zhang*
College of Information Technology, Luoyang Normal University
Henan LuoYang, 471022, China
*Corresponding author,e-mail: zhangruilingls@163.com

Abstract

Goal of ontology learning is to use machine learning and statistical techniques, by means of automatic or semi-automatic, and obtain the expected from the existing data resources. Web data integration based on ontology is web data mapping information in the source to the process of ontology concepts. The goal of semantic Web is to provide a computer for semantic Internet information can be understood, thus the computer to identify the information, and the automatic interpretation, exchange and processing. The paper presents the research on data integration of the semantic web based on ontology learning technology. Theory and experiments show that compared with the method of concept lattice construction algorithm has certain superiority.

Kata kunci: *Ontology Learning, Data Integration, Concept Lattice*

1. Introduction

The main purpose of the semantic web is to extend the current WWW, makes the information network are of semantic, is the computer can understand and handle, facilitate the interaction and cooperation between human and computer. So the study on how to represent the information is to understand and process in the form of a computer, which is semantic Berners-Le [1]. Hierarchical semantic network is presented, which is based on XML and RDF/RDFS, and on this basis to construct ontology and reasoning rules, to complete based on semantic knowledge representation and reasoning, which can be understood by computer and processing.

Now bearing the Web page important information data are mostly through the backstage script generated from a database or direct extraction, and displayed on a web page is relatively fixed format, which Web form is to use a format more extensive. Because of the structure of the Web table to determine the relative comparison and it is thus easier to construct a class of general wrapper to extract the Web table information.

The expression of HTML in information, ignoring the deep structure and meaning of information, they try to label webpage through domain semantic information and solve the problem. Key to realize the work is to construct, fusion, using semantic information, through the annotation processing, at the conceptual level and the site of realizing the integration of these annotation design method, the new method they proposed a Generative Semantics information.

A frame consists of a number of known as the groove structure; each slot can be based on actual needs is divided into a plurality of side. A slot for a description of the object's attributes; one side is used to describe a corresponding attribute, every aspect and constraints are given with the. The framework consists of 85 grooves, respectively describes the 34 attributes of teachers. Frame representation is a strong adaptability, summarizing, structuring, reasoning, and can put the declarative knowledge and procedural knowledge, combining the knowledge representation method. The main shortcoming of frame representation method is not a good representation of process knowledge, usually with other methods in combination.

In numerous practical applications in data integration and unpredictable user may query the content, so the intermediate layer body should try to contain the local ontology of knowledge, it is necessary to keep consistent with the local ontology, and fully consider the application of information integration needs. We conclude that in the process of constructing a middle layer of ontology should be (complete principle) of concepts and relations in the local

Received June 17, 2013; Revised August 14, 2013; Accepted September 13, 2013

ontology if it meets application requirements, this concept should appear in the intermediate layer ontology. (FL principle) of different local ontology concept appeared in the middle layer in the ontology if there is relationship between, the relationship should be reflected in the intermediate layer ontology.

The explicit input of traditional personalized recommendation method of main source of data is to the user. The current user in order to obtain the recommendation must be evaluated on a number of projects, to express their preferences, including explicit navigation, keyword and item attributes and purchase history. Obviously, the availability of the side effects of the methods require the user to make the relevant evaluation caused will reduce the whole system. Different from the traditional recommendation method is based on Web mining, personalized e-commerce recommendation data input is hidden input, it will not affect the normal user browsing behavior.

Ontology can be used as a universal semantic model Web information and application in Web information extraction. Web the process of information extraction based on ontology is a Web data mapping information in the source to the process of ontology concept. In the extraction process of how to automatically identify the concept of ontology in information sources are the main problems of Web information extraction based on ontology the need to consider. RDF is a standard W3C based on the XML recommendation. RDF proposed a simple model used to represent any data type, the data type by between node and node labeled arcs. Table 45 shows the nodes to resources on the Web, arc used to represent these resources attribute. Therefore, this data model can easily describe objects and the relationship between them. W3C recommended in RDF standard to solve semantic limitation of XML. DAM is a language created the first stage of project DARPA. DAML extends RDF, increased T more complex class, attribute definition. The paper presents the research on data integration of the semantic web based on ontology learning technology.

2. Research of Ontology Learning Technology in Semantic Web

Constructing domain ontology is usually done by experts in the field, the only way to ensure the completeness and accuracy of the ontology semantics. However, in the ontology building process, the computer needs a precise definition of formal knowledge; also need the knowledge engineer's participation. In addition, a domain ontology perfect is an iterative process of evolution in the construction.

Web personalized recommendation system, usually is the development of visitors to the site navigation mode based on it [2]. But when the personalized system only based on the use of results, and the final recommendation information, related to the concept of useful information may be omitted. To a great extent improve the use of mining and personalized recommendation results webpage using webpage semantics, because it uses more abstract but consistent, methods and machines and people can understand the processing and analysis of data. The basic idea is, will use the data integration based on semantic represented by Ontology terms, proposes to strengthen the semantic navigation patterns and develop recommendation system better.

The ontology should be at the level of knowledge, and has nothing to do with the particular symbol-level encoding. The choice of the representation of the body should not only consider on or realization convenience. Description of the concept should not be dependent on a particular representation of the symbol layer can not rely on a certain language, because the actual system may use different methods of knowledge representation.

Accurate degree of specific life cycle depends on the ontology prototype. In general, the development of ontology to experience the following stages (corresponding to the above process): specification, conceptual, formal, integration, implementation, and finally into the maintenance phase. Knowledge acquisition, evaluation and documentation throughout are the ontology throughout the life cycle. Similar to the TOVE method and it is METHONTOLOGY method with special emphasis on the maintenance of the ontology. The main difference between the two is: METHONTOLOGY stressed that contained in the ontology life cycle maintenance stage, but TOVE stressed that to solve all kinds of problems in ontology maintenance using a variety of techniques, as is shown by equation 1.

$$\sum_i \sum_j (X_{ij} - bi - (\overline{X}_j - b\frac{n+1}{2}))^2 \quad (1)$$

An ontology view P=(C, R, M). Where P is the name of the body view; C set as the concept of R corpse; relationship between concepts in C set, an element in the R is three bytes (R, Cl, C2, Cl, C2 ∈C, R is the relationship between Cl and C2; is a set of first-order predicate the formula set, constraint; expressed in concepts and relations between the mouth is a set of predicate forms of proposition set, expresses the relationship between resin and other body; the point of view of the M is a group of the form of predicate proposition set, express the relationship between the concept and other concepts in C ontology viewpoint in semantic web.

Semi-structured data (XML schema, DTD) and structured data (relational database) is to acquire Ontology (including the concept and the relationship between the tools). For the two types of data sources, which are used to obtain the mapping rules based on the wooden body. In the implementation of the system, the ontology acquisition from XML Schema and part of the DTD is an existing tool based on (HMarfra). HMarfra can realize the mapping from Schema to XML [3]. Then, in the development of Onto Lift, is an intermediate tool XML from DTD to Schema mapping. In this way, will the two tools together, realizes the ontology acquisition from XML Schema and DTD. Access to ontology from relational database part is provided in Java JDBC standard based on the interface, and then according to certain naming conventions will table and attribute names and other information in the database, according to the mapping rules into ontology elements.

If the collection of objects O ⊆O and attribute collection A ⊆A meet O = g (A) and A = f (O) C = (O, A, σ, λ) is called a fuzzy formal context concept. O and A, respectively known as denotation and connotation of the fuzzy concept C, σ and λ is calculated based on the definition of a Fuzzy concept lattice. F is a collection of all fuzzy concept denoted CS (F). CS (F) on the structure of a partial order can be simply defined as follows: If O1 ⊆ O2, (O1, D1) ≤ (O2, D2).Get through this relationship ordered set CS (F) = (CS (F), ≤) is a lattice.

The precondition of ontology sharing is the rational storage of ontology. At present, ontology storage mainly in two ways, one is the relational database, using a formal document. Typical body generally consists of a set of concepts, describe the concept of feature attributes, relations between concepts, the concept and attributes between the constraints to represent static domain knowledge, which in essence the body can use the relational database technology is mature to complete ontology storage, query and other work, as is shown by Figure 1.

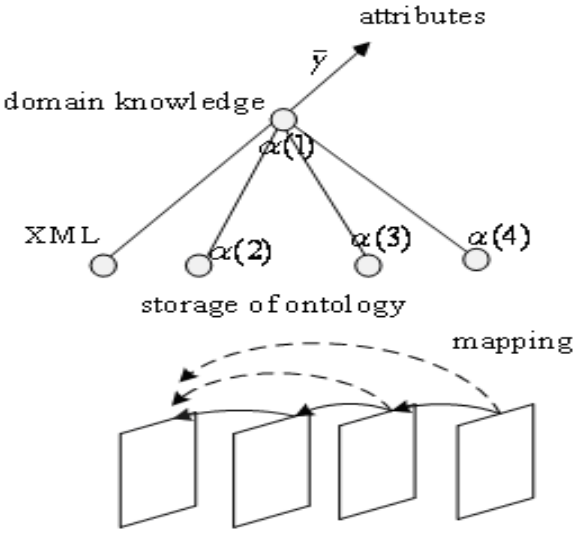

Figure 1. The relational database technology of mapping to complete ontology storage

In general, the ontology mapping requires that the source ontology and ontology representation using the same language. Transformation and translation can solve most of the problems associated with language. A typical language level mismatch including grammar and it is logic expression and language semantics and expressive language. The vast majority of the system in the survey, even if such a compilation work, did not answer the language level of mismatch problem, in the expression of compilation should be how to solve the problem.

Ontology become a hot topic as the key technologies of the Semantic Web ontology modeling the domain knowledge of various disciplines can be achieved not only knowledge sharing and integration, and also offers new prospects for the development of education information. Here are the computer field ontology the OntoAES system (CS-DomainOntology) established the concept of domain ontology set and hierarchy model, and described through the SKOS model of the ontology.

For an ontology viewpoint, the corresponding ontology axioms are actually bound for the ontology concepts and relationships that can be used to detect inconsistencies within the ontology viewpoint. For inconsistencies within the ontology the viewpoint, we have the following definitions.

Definition 1 (the ontology viewpoint inconsistency) for the ontology viewpoint P, if: P. C ∧PR | = Q, P. theta | = Q. We call the P memory in inconsistencies.

Ontology viewpoint inconsistencies between the inspection and the process is also a modification of the ontology the viewpoint of the process, this process is to eliminate the semantic inconsistencies between the ontology body viewpoints. After inconsistency inspection and processing the ontology point of view we construct the intermediate layer ontology.

It is extended to the various relations, He relations such as is-a, has Member, ConsistOf. Rules for subclasses intersect Cl, C2, ..., a division of Cn is C, if I, I (islnstanceOf) Cj∧ I (islnstanceOf) Ck, 1≤ J, K ≤n, J ≠k} report subclass intersection inconsistent. The regular class does not contain individual class Cl, C2, ... , Cn is a division of C, if I, I (islnstanceOf) Cj∧ I (islnstanceOf) C, but I (islnstanceOf) C1 a I (islnstance Of) C2 a ... A I (islnstance Of) Cn, then report subclass does not contain individual, as is shown by equation 2 [4].

$$m_0^3(t) = \left(1 - z^3\right)(1 - c^3(t))\, f^3(t) \tag{2}$$

Relationship between the ontology view P and other body among viewpoints (P (Ω) need middle layer ontology scanned to determine the other ontology view after. In order to quantify relationships is between ontological viewpoints. We introduce the concept of the two body inter-view dependency of. To calculate the dependence of other body point of view on the body, also determined the relationship between the body and the other body viewpoint.

FCA-merge the two bodies merged into a ontology, you need to select the document contains two ontology concept as a formal concept analysis instance, the use of formal concept analysis methods will be merged ontology concepts merged into one concept to form a concept lattice. This concept lattice can be used as the basis of a new ontology construction. Neither FCA-MERGE grammar-based ontology integration, nor the relationship between the concept of the body, and FCA-MERGE method of sample instance selection conditions can be quite harsh.

The development of an ontology (including the concepts and their relationships) tools obtained from the XML and HTML files. It looks like a Web browser. Use it to get the ontology, you need to manually build an initial domain ontology; Then, the user has viewed the process that contains the relevant information in the field of site, the tool will be generated for each site a candidate ontology; Finally, the user the participation of the candidate ontology combined with the initial ontology. Among them, the use of ontology learning methods word frequency statistics and pattern matching (including sub-string matching, content matching, matching dictionary).

Semantic Matching can also be matching operator (device). Developers will match the following definitions: " matching is an operator, two class structure (such as database structure or body) to operate, and generates a mapping between elements of two graph structure, two graph structures correspond with each other in semantics. " So far, Semantic Matching has been developed and tested. Just Semantic Matching rating system is only one type of relationship, i.e. a non formal variable " is-a " relationship.

The modeling process of the ontology body is the continuous cycle of to improve the process of optimize the, so on the evaluation of the ontology model and validation are also

interspersed in the in the process of the entire Ontology developers adopting. The in two major areas of evaluation and validate the carried out to discuss the, with the experts and researchers of the field of carried out, to express partial validation vocabulary used is correct, the feasibility of the relationship between the accuracy as well as the class for the knowledge in the ontology to. And Knowledge Engineering researchers discuss the ontology model feasibility and correctness of the model itself. Consistency, it is completeness, and so on.

The goal of semantic Web is to provide a computer for semantic Internet information can be understood, thus the computer to identify the information, and the automatic interpretation, exchange and processing. Put forward the concept of semantic Web, provides a new idea to solve the problems existing in the current web, as is shown by Figure 2.

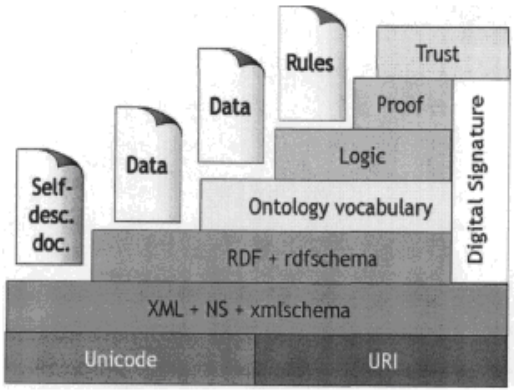

Figure 2. The semantic Web architecture

Ontology provides a rich description of attributes and the types of word, such as: disjoint between classes (disjoint ness), cardinality constraints (cardinality), equivalence (equivalence), symmetry (symmetry), transmission (transitivity), functional (functionality), inverse function (reversibility) and other attributes. Ontology can not only make the resources on the Web with a formal clear, logical reasoning, but also has powerful [5]. Therefore, in the semantic Web, ontology has a very important position; it is base of Web information sharing and exchange at semantic level.

User preference information is more and more widely used in personalized system. The user preference representation is mainly divided into two categories, one is to use the user's personalized information to indicate that the user preference. Two is the model representation directly gives the user preferences, users can express their preferences. We give the user preference ontology model based on semantic information, can give more to the user preference, with stronger expression ability.

SquishQL is mainly used for RDF description database query, as a simple RDF oriented, based on Graph Navigation Language subgraph matching mechanism, SquishQL resembles the SQL syntax. In order to retrieve data, it is using two types of constraints in SquishQL: model and the filter expression. Model performance for the three tuple pattern of <subject, predicate, object>, describes the edges of a graph and a correlation operation. For each element of the three tuple model, allowing a variable or a definite value. And the filter expression can express the values for the three tuple model in the limit variables. SquishQL on the expression of RDF support many functions, because become the basis of a series of RDF query language.

At present, Semantic Matching can work with the classification system of hierarchy, also known as the directed acyclic graph (precursor, DAGs). The body can be converted to a classification system structure, node is in the ontology, is-a relation is the connection, but other relations will be lost in the conversion process [6]. The use of course not cancels algorithm results as the mapping process input values. At present, the work is being done to different types of account of semantic relations.

Of course, ontology can usually be rewritten with a marked connection chart format, but in the process of rewriting some information (such as axioms) will be missed. Concept node, the relationship between the concepts should be a connection between nodes (labeled). Connection is mark (label) the type of relationship. This map and ONION used keyed same. Ongoing relationships (relationships) semantic integration is to the Semantic matching the algorithm. But this work is still in its infancy.

Function is a kind of special relationship, the former n of a 1 element can be uniquely identified in the n elements. The formal definition of F:C1 * C2 * ... The Cn-1→ Cn. For example, Grandmother Of is a function, where Grandmother f (x, y) indicate that the Y is x's grandmother, apparently x can only determine their grandmother y, as is shown by equation 3.

$$x^{(1)}(k+1)=\left(x^{(0)}(1)-\frac{b}{a}\right)e^{-ak}+\frac{b}{a} \tag{3}$$

In the field of ontology building domain ontology construction stage can be accurately described the content of Web site, the domain ontology using specific ontology description language representation. As the representation of ontology language tools, should have the basic functions are as follows: a logical system with strong support, has a strong ability to express knowledge, also has certain reasoning ability, formal, the representation of domain knowledge is formalized, namely machine readable and understandable. Because of using machine understandable representation language ontology representation, can be directly by the computer processing, and interoperability between different systems.

In the learning process, the need for users to have mastered the knowledge and skills to describe, and it is in order to better the positioning of their own learning background, develop more individualized learning process and learning objectives. Learning ability is the comprehensive application of knowledge, skills, experience, ability to external learning resources and tools to solve a problem or complete a task. At present, there have been many studies devoted to learning capability can be applied to the field of education and training, such as IMS to develop the Reusable Definition of Competency or Educational Objective (IMS-RDCEO) standard, which made the description, reference and learning ability to define information exchange model.

Based on prior to build domain ontology and knowledge base, the extraction and storage of semantic information in the indexing process, based on the traditional vector space model of the index increased significantly the semantic information. Based on the semantic indexing provides ontology-driven search and browsing mechanism [7]. A novel method based on graph query mechanism is provided, the user can through the browser, intuitive to draw complex semantic query. These queries can be stored and easily be reused as model. Based on the keyword query, the system will list object semantic association domain object list and query, the user more convenient to search the information of interest.

3. Data Integration Based on Ontology Technology

Domain ontology construction method is mainly the following: comprehensive method, TOVE method, IDEF-5 method, skeleton method METHONTOLOGY method, SENSUS method, KACTUS method, seven steps. The most famous is the seven step method developed by the Stanford University, the method accords with the thinking of human, logic is strong, and has the advantages of convenient operation and extensible. A lot of domain ontology is constructed by seven steps.

The user model is established mainly through user behavior [8]. User behavior is divided into explicit and implicit two. Display behavior provided by registered users active feedback information and authority to provide information, basic personal information, including user preference information, knowledge background, learning ability and performance information etc. The learning resources and the use of model data implicit behavior through user browsing performance, which implied the personal interests and preferences, which can be analyzed and processed by the data mining method, as is shown by equation 4.

$$f(b)-f(a)=\frac{f'(x_3)}{\frac{1}{x_3}}(\ln b-\ln a)=x_3 f'(x_3)\ln\frac{b}{a} \tag{4}$$

For each ontology point of view, according to the application requirements closely associated, we define the concept of relevant factors. We will apply the requirements defined as a triple H = (psi, theta, and phi).Where Ψ represents a concept that may occur in the ontology of the intermediate layer represents the relationship between the conceptual level, Θ, Φ, H is usually determined empirically adopted by the developer according to the rules Ψ and theta constituted.

iPROMPT interactive ontology integration tool. It can be fused elements in the task of ontology integration provide users with help. The role of the iPROMPT identification contradiction lies and potential problems and suggest possible strategies to solve these problems and contradictions.

Educational Semantic Web, the ontology has a very important position, the understanding of the structure of knowledge ontology and software agents to share information on the network becomes and machine-readable description of structured knowledge, field knowledge reusability [9]. This article describes the-oriented the ontology Language and Ontology of the Semantic Web to build tool, focusing on introduced the OWLDL language and the description logic, the definition of is used in the ontology construction and the formal description of the of OWLDL language on the ontology body members and relationships.

The concepts in ontology to organize according to taxonomic point of view, the formation of classification structure system. Constructing a classification structure should follow some basic principles, such as: meet the basic principles of taxonomy, a clear distinction between the concept and the concept of father, brother, there are many ways to meet the necessary semantic constraint hierarchy definition concept, the most commonly used are of two kinds: the top-down method, the method from the definition of this field began the concept of routine, and then more special if the definition of the concept of. Bottom-up method, the method to define concepts from the most detailed level, namely, the leaves of the hierarchy, and then the concept of combination of this part of the more general concept of higher layers. It is worth noting that, contain a hierarchy model is just a classification relationship (taxonomy relation), as is shown by equation 5.

$$\hat{\sigma}=\sqrt{\frac{\sum_{i=1}^{n} y_i\cdot(y_i-\hat{y}_i)}{n-2}} \tag{5}$$

Ontology learning ontology concept extraction, semantic relations between extraction and classification system of the automatic construction of key technology, through the experiment, the algorithm has been tested, and the ontology evaluation method were discussed. The integration of a variety of machine learning algorithms, this method has a higher accuracy in terms of learning concepts extraction and semantic relations [10]. The general body WordNet and HowNet as the corpus, it can be applied to different fields of specialization. At the same time, through the on-demand access to the web document, the method can generate real-time ontology.

Protégé provides ontology storage to the MySql function. The MySql JDBC driver is copied to the installation directory Protégé, changed the name to driver.jar. Connect the Protégé and database, the establishment of Protégé in Mysql database. Body building has become the basis for semantic knowledge sharing and Interoperability Based on semantic Web, more and more body is gradually being accepted and used. With the semantic Web technology continues to advance, the basic body of knowledge representation will be continuously improved, will achieve the semantic and reasoning based on semantic Web, lay a good foundation for further establishment of trusted Web.

Ontology is engineering. Research and development of the content of the ontology, including two aspects, one is the ontology research and creates a specific field; the second is to

research and develop universal knowledge ontology repository. Ontology representation is transformation and integration. Research is used to represent the ontology knowledge representation system, provides methods and tools for formal, so that the body can easily be shared and reused; provide ontology evaluation and compare different framework, integrated method of conversion between different ontology method and different ontologies; provide interoperability between different ontology means. The main research in a specific domain ontology or general ontology is as the foundation of the application [11]. From the discussion above we can see, the body has become an important tool in knowledge engineering, knowledge acquisition, representation, analysis and application, has the vital significance, as is shown by figure 3.

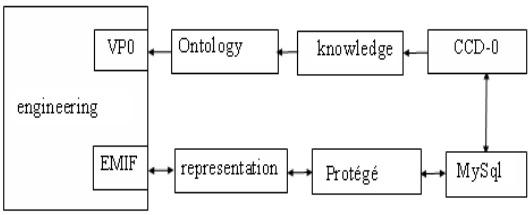

Figure 3. Ontology knowledge representation system

The development of a comprehensive is ontology learning tools. Its main characteristic is that it can support the ontology acquisition from multiple data sources. At present, it can be done from unstructured data (text) and semi-structured data (HTML, dictionary) acquisition of concepts and their relations [12]. For learning ontology from unstructured data, it uses the method of weighted word frequency statistics to obtain concept, used to obtain a classification relationship is based on the concept of hierarchical clustering method, the use of methods based on association rules to obtain the non-taxonomic relation; for the HTML data, it will be pretreated to plain text, and then use the learning method of unstructured data ontology based on Ontology acquisition from; for the dictionary, it uses the learning method based on template. The system can deal with Devin and the English data sources.

The implementation utilizes owl of the class and its hierarchy class label, in accordance with the ontology framework established using Protégé build the class hierarchy, it is a typical tree structure, the root is the default Protégé OWL Plugin owl: Thing movie domain ontology in its core class hierarchy Movie. The property implementation OWL attribute is the association between the object of the class of expression.

The ontology is a conceptual system or basic knowledge system is a higher level of knowledge of the Knowledge Base abstract. In a given field of knowledge, the ontology defines the description and representative of the vocabulary of the domain knowledge, including the definition of the basic concepts of computer identification as well as the relationships between them, and therefore it constitutes the field-oriented knowledge representation system the core, is encoded by a field of knowledge and interdisciplinary knowledge, sharing of knowledge known as possible. This chapter details the computer field of OntoAES ontology and learning resource description ontology.

When the relationship inconsistency occurs, if no child relationship constraint (i.e., r r1 child relationship, or r1 is r child relationship) between r and r1, that relationship (r, c1, c2) and (r1, C1, c2) may conflict (r, c1, c2) and (r1, c1, c2) need to be configured by the ontology through the intermediate layer is determined by: if a conflict is determined, the (r, c1, c2) and ((RL} Cl, c2) should not appear in the ontology of the intermediate layer, then the relationship (r, C1, C2)), and (r1, c1, c2) of the respective main body viewpoint will remove them; if determined not to conflict, (r, c1, c2) and (r1, c1, c2) from the ontology body of the intermediate layer structure. The decision in the ontology of the intermediate layer and it is only a relationship that does not appear in its ontology in which the viewpoint is deleted.

The ontology can be used to describe the knowledge structure of internal problems, so in the use of most get their ideas as the knowledge base to use. In view of ontology knowledge

description ability, so use the idea to express user interest, from the composition properties of user interest to understand and express information. Using the relationship formed by the object with the user interest ontology library, structural model, and this model can help the interest from the library for users find more useful, as is shown by equation 6.

$$\sum_{i}\sum_{j}(X_{ij}-bi-(\overline{X}_{j}-b\frac{n+1}{2}))^{2} \tag{6}$$

For a finite path v0l1v1l2 queries in Q ... Lnvn (V0 Q root), to quickly find the satisfied WEC mapping in the ontology tree X, in our study we have introduced a similar to the inter-relevant successive tree index structure, and said that this index structure for XSIRST. Assumption for the ontology tree X, the edge label set L= {l1, l2, ... X, ln}, XSIRST is a forest, each tree in the forest with Li (L≤ l ≤n) root, X into the edge labels for Li nodes as the leaves, the leaves contain the node's leaf order interval.

The main content of ontology check is to be represented as an ontology entities or find another body and the corresponding entity or. At the local level based on methods of calibration is called local method (Local Methods), also known as the local classification method (Local Grouping of Objects). That is to say, only the ratio of an element and other elements, rather than in the global scope of the body of work. Once you can get used to determine similarity or local method of similarity, and then used to estimate the “check”, which includes some more comprehensive method.

Class structure is using the concept of the idea to create [13]. In the concept hierarchy structure, has the concept of inheritance between all components from the upper to the lower paths, each path has the characteristics of complete conceptual basis. So that heavy interests degree and the inheritance of any two concepts by deductive method.

Use the name WRBT to match the dialect mode is the key steps of positioning of the concept of ontology. Assuming the Schema definition of domain ontology we use DAML+OIL language to describe, is pattern matching column names can be used four byte header ((n, D, P, s), where n represents a data type attribute in Schema, use the tag dimly: Data type Property in DAML+OIL. D said N belonging to the class, using the tag dimly: Class in DAML+OIL. P is a dialect of N mode.

According to the abstract is definition of specific application domain knowledge modeling. Usually, the application of ontology is a mixture of concepts, these concepts from the domain ontology and ontology, extension and application of body also includes specific methods and specific tasks. Representation type body, mainly describes the concept of formal behind knowledge, but not committed to any particular field, it provides substantial ontology representation of neutral, provides a representation framework.

4. Research on Data Integration of Semantic Web based on Ontology Learning Technology

Method for learning text based ontology, and it is analyzes the term acquisition, classification, relationship to get the concept of key technologies such as ontology based on learning from text. Through the experiment, it is showed that, with seed concept from plain text to extract concepts, concept classification, and provide the basis for ontology development.

Maintenance of the ontology model includes scalability and portability of the original ontology. The continued expansion of the rapid development of information and knowledge, knowledge areas continue to expand and improve, more new knowledge is introduced and learning, knowledge of the original structural system will also be modified, so the original ontology model need to constantly add the new concept of the word, but also need to adjust and improve the existing class structure and the relational model. This requires the extraction and classification of the classes in the ontology model, as well as the class hierarchy relatively high demand, the ability to easily add and modify, as is shown by Figure 4.

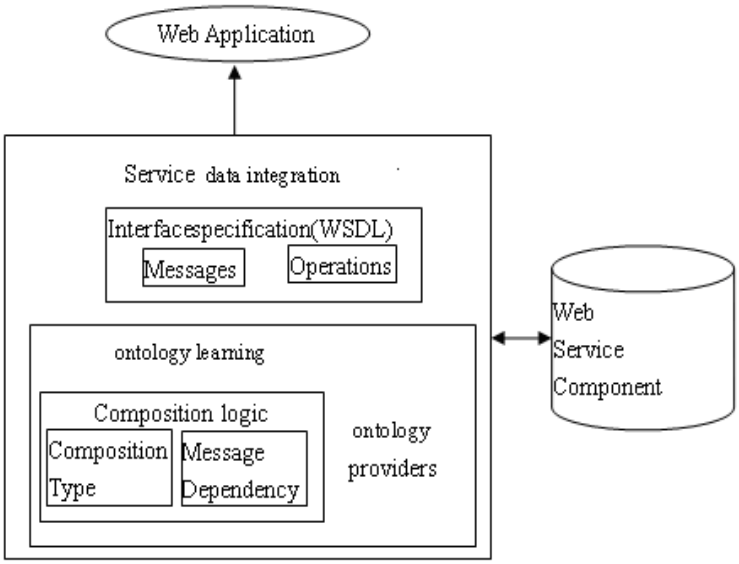

Figure 4. Data integration of the semantic web based on ontology learning diagram

Knowledge ontology representation, ontology, ontology, ordinary top element (core) ontology, domain ontology, ontology, task ontology, domain ontology, a task ontology and application ontology, this classification is proposed by Guarino method on the expansion and refinement, but there is a cross between these 10 kinds of ontology, hierarchical classification not clear [14]. It can also be based on ontology according to the number and types of conceptual structure, according to the classification of the body different aspects of characterization and description of the body of the real world, as is shown by equation 7.

$$f'(\xi)=\frac{1}{\xi}=\frac{\ln a-\ln b}{a-b} \tag{7}$$

An ontological viewpoint in the ontology of " form " as a directed labeled graphs form (ontology), in order to be more conducive to knowledge expression form of transformation for semantic reasoning, ontology, a viewpoint of " viewpoint constraint " really express the constraint in the concepts and relationships, form as a collection of first-order predicate formula. The relationship between " point of view " of an ontology perspective is composed of two parts: one part is the body view and other body among viewpoints, expresses the relationship between the body view; a part of the body is the relationship between the concept and other body view viewpoint, composed of a series of related assertions, form for the predicate form of the proposition.

C-OWL is OWL language on the semantic and syntactic extensions, and its purpose is to context the ontology (contextual ontology). From this point of view, this term contextual ontology means to keep the ontology the content locally, by explicit mapping (such as bridging rules bridge rules) can be mapped to the content of the other ontology. And OWL input mechanism is opposite in OWL input mechanism, a set of local (main body) model is based on a unique sharing model to complete globalization.

RDF is an expression of the language of Web resources, and its main purpose is to express the metadata of Web resources. RDF provides universal expression information architecture to facilitate the interaction between applications and build common RDF parser. RDF graph data model expression can also be used to convert XML syntax RDF graph. Applied to the RDF XML syntax called a RDF / XNIL.RDF / XML design simplifies multi-statement of the same resources. Defined by the W3C RDF Vocabulary Description Language RDF Schema is to complete vocabulary defined, as is shown by equation 8.

$$u = \int_0^x x dx + \int_1^y (\frac{x^2}{2} + \frac{1}{y}) dy = \frac{x^2}{2} y + \ln y \quad (8)$$

Identification of the concept of information based on the text data, including the concept connotation (Natural Language) and extension (instance) extraction. Extraction of ontology knowledge of higher level from the text and other data, such as the rules and axioms, evaluation of the automatic extension of ontology learning and Ontology (Evaluation of ontology learning and population) for ontology learning more complete reference and evaluation means, beyond the lexical overlap and classification evaluation [15]. For example, ontology learning evaluation method, task oriented, ask context, machine learning, text contains.

The paper presents the research on data integration of the semantic web based on ontology learning technology. The prototype system to integrate the information in some designated sites on the Web page (the structure of these Web pages usually are more stable), and the body as a collection of some classes and relations. System uses C++Builder7.0, and VS2010.Net programming tools mixed development, back-end database using SQL SERVER2008. Compare of data integration of the semantic web based on ontology learning technology with FCA, as is shown by Figure 5.

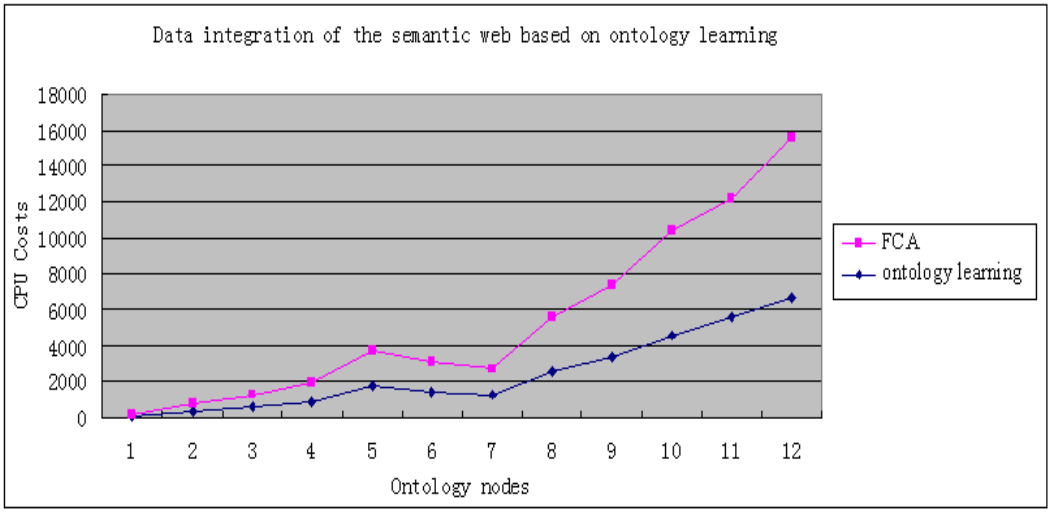

Figure 5. Compare of data integration of the semantic web based on ontology learning technology with FCA diagram

Ontology is an important part of constructing knowledge, and it is the theory analysis, modeling of domain ontology results, a field that is the real world of abstraction for normalized relations between a set of concepts and description, to outline the basic body of knowledge in this field, provides terms as domain knowledge description. Ontology defines a description and a symbol of a field of knowledge of vocabulary; it not only in machine-readable form defines the basic concepts in the field, but also covers the relationship model between these concepts. While an ontology usually defines the concept of a special relationship between the subject and, rather than just some general ordinary concept. The paper presents the research on data integration of the semantic web based on ontology learning technology. The communication is between field and field. The framework provides a unified; ontology reduces in the field of conceptual and terminological confusion.

5. Conclusion

This article uses a variety of ontology learning methods: get the concept, it uses a method based on linguistics; obtain the relationship between concepts based conceptual

clustering methods (including hierarchical clustering and non-hierarchical clustering)and template-based method; an axiom, it uses a template-based approach. The paper presents the research on data integration of the semantic web based on ontology learning technology. In addition, it uses a heuristic method of learning, ontology learning process, while at the same time appear more likely candidate results, it is using some heuristic rules to reduce the hypothesis space, eliminating uncertainty. So, the paper uses a heuristic mixed ontology learning method. At present, it has been possible to obtain the body from the Persian text.

Acknowledgment

This paper is supported by the National Natural Science Funds of China (61272015), and also is supported by the science and technology research major project of Henan province Education Department (13B520155) and Key scientific and technological project of Henan province (102102310058).

References

[1] Berners-Lee T, James Hendler, Ora Lassila. The Semantic Web. *Scientific American*. 2001: 5.
[2] Gruber TR. A Translation Approach to Portable Ontology Specification. *Technical Report of Knowledge System Laboratory (KSL)*. 1993: 92-71.
[3] Sahami M. *Learning classification rules using lattices (Extended Abstract)*. In ECML-95: Proceedings of the Eighth European Conference on Machine Learning, Berlin, Germany: Springer-Verlag. 1995: 343-346.
[4] Yingjie Song, Rong Chen, Yaqing Liu. A Non-Standard Approach for the OWL Ontologies Checking and Reasoning. *Journal of Computers*. 2012; 7(10): 2454-2461.
[5] Fernandes M, Gomez-perez A Juristo N. *Methonlogy: From Ontological Art Towards Ontological Engineering*. In: Proc. of AAAI97 Spring Symposium Series, Workshop on Ontological Engineering, Stanford University. March 24-26th, 1997: 33-40.
[6] Maedche A, Staab S. Ontology Learning for the Semantic Web. *Special Issue on the Semantic Web*, *IEEE Intelligent System*. 2001; 16(2): 72-79.
[7] Da Wei Wang. A New Field Word Segmentation Model Based on Ontology in Digital Library. *IJACT*. 2012; 4(17): 418-425.
[8] Stojanovic L, Stojanovic N. *Volz R Migrating Data-intensive Web Sites into the Semantic Web.* In: Proc. of the 17th ACM symposium on applied computing ACM press. 2002: 1100-1107.
[9] Shamsfard M, Barforoush AA. Learning Ontologies from Natural Language Texts. *Int'1 Journal of Human-Computer Studies*. 2004; 60: 17-63.
[10] Weizhong Xie, Wangyang Hu, Guoyun Cheng. Construction of the public emergencies emergency command system based on ontology. *IJACT*. 2012; 4(18): 664-671.
[11] Huaibin Wang, Haiyun Zhou, Chundong Wang. Virtual Machine-based Intrusion Detection System Framework in Cloud Computing Environment. *Journal of Computers*. 2012; 7(10): 2397-2403.
[12] Maryam Hourali, Gholam Ali Montazer. A New Approach for Automating the Ontology Learning Process Using. *JCIT*. 2011; 6(10): 24-32.
[13] Du Xiao-yong, Li Man. Research on Ontology Learning. *Journal of Software*. 2006; 17(9): 1837-1847.
[14] Liu Dianting, Jia Feifei. Resources Scheduling System for Collaborative Design in Ontology and Multi-agents. *TELKOMNIKA Indonesian Journal of Electrical Engineering*. 2012; 11(3): 1305-1312.
[15] Godin R, Missaoui R, & Alaoui H. Incremental concept formation algorithms based on Galois (concept) lattices. *Computational Intelligence*. 1995; 11(2): 246-267.