RESEARCH ARTICLE OPEN ACCESS

Ontology - based Semantic Value Conversion

JieWang[1]
[1](School of Computer Science, Jinan University, Guangzhou 510000, China.)

----------**********----------

Abstract:

In the face of massive Web data, an important task is to integrate relevant data of interest from users in Deep Web into a unified and structured form. In order to facilitate the identification and processing of computers, each data item needs to be annotated. Ontology is the key to semantic tagging. In order to facilitate the sharing and exchange of information, the semantic value of ontology instances must be consistent and compatible. In this paper, a semantic value conversion method based on ontology is proposed. This method is divided into two parts, one is to define the conversion function as the entry point of the semantic value conversion. The second is to use OWL as the ontology language description, and establish an OWL ontology relational database to store the relationship between ontology instances and various semantic values. The semantic value conversion method improves the integrity and consistency of the ontology instance semantic information.

Keywords **—Deep Web; Information extraction; Semantic annotation; OWL ontology; Relational database; Semantic value conversion.**

----------**********----------

I. INTRODUCTION

The amount of data on the web and the speed of its growth make it increasingly difficult for users to get the information they need. Deep Web is validated by dynamic request of Web information, only through the Web page provides the query interface to access Web database, and the results combined with format control after the return in the form of Web pages, its content is stored in the real database [1].For Deep Web data integration, Web information extraction and semantic annotations are two very important modules. Web information extraction is the Web as a source of information, the user is interested in the query result is extracted, due to the performance of data to extract the format is not exactly the same, so you need to the specific target information into available for users to recognize and deal with the unification of the structured storage mode, such as a relational database.As a result of the extracted results are based on information from a number of different sites, each site design and no unified mode, so the structured model also not unified, and some sites have these information domain knowledge, some sites without domain knowledge, people can recognize, but the computer could not identify, so for the extracted information (data item), the result of the need to find a set of semantic vocabulary semantic annotation add semantic information to carry out, not only convenient people recognition, also good for computer processing.The semantic annotation is the instance data of the tag ontology class and is mapped to the ontology class [2]. However, for some reasons, the semantic value of ontology instances can vary. For example, time can be jet-lagged for geographical reasons; Weight can be different from different units of measurement due to cultural background. In order to facilitate the sharing and exchange of information, the semantic

value of ontology instances must be consistent and compatible.

In this paper, on the basis of the Web information extraction technology, combined with the OWL ontology and relational database, in view of the extracted from Deep Web sites and storage items of data in a structured way to construct ontology, a semantic value conversion method based on ontology for semantic annotation. OWL as ontology description language, is adopted to establish the OWL ontology relational database as the basis of semantic value, by constructing a conversion function as the entry of the semantic value, so that the ontology instances that the semantics of the data items is consistent.

II. CONSTRUCT THE ONTOLOGY

Ontology-based semantic value conversion method is used to perfect semantic annotation. Ontology is the key to increase semantic annotation of extracted data, providing lexical and semantics for semantic annotations. Therefore, the construction ontology is the premise of realizing the semantic value conversion method.

A. Ontology and semantic annotation

Ontology is originally a concept of philosophical domain, which is an objective description of real world existence. The concept of ontology has been introduced into the computer field. Stand in the Angle of the Deep Web data integration, this paper think that ontology is the nature of symbol entities referred to objects of a formal semantic description, i.e., according to the unified data items stored in a structured form (instance data) of the semantic information. In general, ontology has a certain universality, is a specific domain knowledge conceptualization, explicit formalized specification, is in the field of consensus, but because there are a number of instances in the field is endless and dynamic change myriad, therefore, only the ontology and specific application, instance to be meaningful. Associating the actual application with the abstract concept of ontology is the work of semantic tagging. Semantic annotation technology correspondingly enriches the instances in ontology.

B. Ontology description language

Ontology description Language has many kinds, this article selects the OWL (Web Ontology Language) Language description Ontology. OWL is a kind of excellent ontology description language, is based on description logic theory foundation to build language system, it is based on DAML + OIL, is an extension of the RDF, provides a more primitive, more solid theoretical foundation and rich semantic expression mechanism, and uses the object-oriented modeling. OWL is the latest W3C recommendation, and by providing more words to describe attributes and classes, such as disjoint with, cardinality, oneOf, equivalence, and symmetry, it makes it more understandable than XML, RDF, and RDF Schema (RDF - S) in terms of machine-intelligibility of Web content [3].The OWL ontology includes descriptions of classes, attributes, and instances, using Uniform Resource identifiers in OWL to uniquely identify them [4]. The URI corresponds to the ID primary key in the relational data table.

C. The way ontology is stored

There are three main types of ontology storage: plain text storage, specialized management tool storage and relational database storage [5]. Plain text storage is to store the ontology in a text file (such as an OWL file). Special way of management tools storage means there are some software such as OMM support for RDF, OWL, storage management, also provide all kinds of interface, you can use the query language to query the ontology. Relational database storage is to store ontology in a relational database.

Relational database is not the best way to express ontology, ontology semantic information, its internal logic is much more complicated than the relational schema, but database development research history is long, technology is relatively mature, more storage capacity, and support the SQL query, thus query speed is faster than text memory to store a lot, especially when the data volume is very big. As a result, many researchers have proposed the method based

on relational database storage OWL ontology, in order to facilitate the mass data storage and achieve a balance between keeping semantic integrity, OWL ontology relational database can store huge amounts of data, and can keep its maximum semantic integrity [6].

Therefore, this article selects the OWL Ontology Relational Database storage Ontology, literature [7] its definition is given: the OWL Ontology Relational Database (OWL Ontology Relational Database, OORDB) is the concept of the OWL Ontology content according to certain strategy under the condition of without losing its semantic organization to the Relational Database, the use of the existing Relational Database system control and management ability of Ontology.

III. SEMANTIC VALUE CONVERSION METHOD DESIGN

Semantic value conversion method design is divided into two parts: one is to provide a unified conversion function to the user as the entry of the semantic value, the form of the conversion function is unified, but in the form of the input parameters are changeable, it can be according to the different class instances of ontology for semantic value; The second is to design an OWL ontology relational database to store the relationship between ontology instances and various semantic values. The OWL ontology relational database provides the basis for the transformation function, which is the premise for the transformation of ontology semantic value, and also the foothold.

A. Conversion function definition

Theconversion function is the entry that implements the semantic value transformation, defined as: Convert(String className, String instance,String sourceAttriValues, String targetAttriValues). Parameter className is refers to the OWL class, parameter instance refers to the individual instances, the OWL class parameter sourceAttriValues refers to the current instance of the source attribute value, parameter targetAttriValues refers to the target object instance attribute values, separated by a space between multiple attribute values. The conversion function is the source property value of the current instance of the input OWL class, and the output corresponds to the target attribute value. From the perspective of ontology, the OWL class describes the concept of abstract concept is to the individual, so the simplest annotation is to establish the connection between the concept and the individual, also is to find the existence of individuals within the territory, which is then declared it instances of the class. Given the possible overlap of meanings expressed in concepts, an individual can be an instance of multiple concepts at the same time. For example, the data "apple" as an individual, related to ontology, can be labeled as an instance of fruit, a brand, a company or other concept in different applications. Semantic query and other applications often involve the specific attribute values of the instance, so the annotation of a more detailed step needs to give the attribute and attribute value corresponding to the instance. For example, the data "desk" as an instance of the table, corresponding attributes have materials, long, wide, higher, the corresponding attribute value is what material, long breadth height is how much. At the same time, a certain concept can be another attribute of the concept, which USES attribute-of to represent this relationship, and if C is a property of D, call it C = attribute (D). For example, the properties of the table are long and wide, and they are the corresponding properties of the concept with length units, and the corresponding attribute values are inches, meters, inches, etc. When the attribute value changes, the instance changes accordingly..

B. Build OWL ontology relational database

The semantic value transformation is based on the source attribute value of the source instance, and the target instance is converted to the target attribute value by querying the ontology database. On the basis of literature [7], the process of constructing ontology semantic value transformation relational database is as follows:

1) Create a class table. This table contains five fields: class URI, describe type, class description value, the conversion formula and standby conversion formula, which kind of URI

for the table's primary key [8], standby conversion formula is the inverse operation of transformation formula. Define a Boolean global variable, useInverseFormula, to determine which formula is invoked to perform the property value conversion, and when the value is false, the transformation formula is called; Instead, call the standby transformation formula. The words described in the OWL language include subC1assOf, equivalentClass, disjointWith, unionOf, one of, complementOf, and intersectionOf, which extend the language's ability to query and reason.

2) Create a property table. The property table is used to indicate the relationship between instances. Specific instances of a group, may exist various relationships between concepts, and domain ontology model is given of the concept of domain, specifically whether there is what kind of relationship, with the aid of marked points out. This table contains four fields: attribute URI, attribute types, domain and range, type fields shows that the attribute is an object or data type attribute, attribute the URI for the table's primary key.

3) Create a property value relational table. There are four fields in this table: attribute URI, target attribute value, source attribute value, and relationship value. The attribute URI is the foreign key, referencing the primary key of the property table. The compound primary key of this table is the property URI, target attribute value, and source attribute value. Target attribute values and the source attribute values corresponding conversion formula or standby conversion formula of target instance attribute values and attribute value, relationship value corresponding to the conversion formula and standby conversion formula of relationship value. property .In this case, the target instance and the source instance are the instances of the domain of the same property in the property table.

4) Create an instance table. Define the following fields for the table: instance URI, class URI, attribute URI, and attribute value. The example constraints in OWL include sameAs, differentFrom, and a11Different.

This database is only used for semantic value conversion, and the specific method of building a complete OWL ontology relational database is found in the literature [7].

Each affiliation must include, at the very least, the name of the company and the name of the country where the author is based (e.g. Causal Productions Pty Ltd, Australia).

Email address is compulsory for the corresponding author.

C. Semantic value conversion process

1) Call the convert function, enter the class name, the instance, the source property value, and the target attribute value. If the value of the source attribute and the target attribute value are more than one, then the multiple values are separated according to the space to form multiple corresponding single source attribute values and single target attribute values. Instead, the source attribute value itself is a single source attribute value, and the target attribute value itself as a single attribute value.

2) According to the single target attribute value and the value of the individual source, the corresponding relationship value is found in the property value relation table. And the global variable, useInverseFormula, is given to the global variable, and the value of the single target attribute value and the value of the single source attribute are false when the corresponding record is in the property value relational table. When the input of a single target attribute values corresponding to the attribute value relationship a record of a single source attribute values in the table, and enter a single source value corresponding to the records of a single target attribute value, the value is true.

3) The connection of the property value relational table and the property table is used to query the range of attributes for the input class name and the input attribute value.

4) In the class table, the range and useInverseFormula are obtained to query the conversion formula or the alternate conversion formula in the class table.

5) Substitute the source instance and the relational value into the formula and get the target instance.

6) Change the source instance to the target instance in the instance table, and the individual source attribute value is changed to a single target attribute value.

7) If there are multiple single attribute values, repeat the above steps until all single source attribute values are converted to a single target attribute value and get the final target instance.

IV. EXAMPLE OF SEMANTIC VALUECONVERSION

A. Example of semantic value conversion

Information extraction is carried out on multiple B2C websites, and the semi-structured data of web form is saved as a structured form of data item alignment. Then, the corresponding ontology is constructed, and the structured data is marked as an instance of the ontology class with semantic annotations, and the mapping is made with the ontology class. Obviously, every data records in a B2C website contains "the price" data item, therefore, when constructing ontology, price as an attribute of commodity type, the corresponding attribute value is how much is the price. Everyday, we often use a digital said the price, in this particular situation, we will default to the semantic understanding of this number is consistent, but once out of the specific situation, only a number used to represent the value of the price is not enough. If, the price of the instance iPhone7 attribute value is 7000, due to the instance data from a number of different B2C website, can't use the default is 7000 yuan RMB in daily to understand the property value, because we don't know it refers to the RMB or hk dollars or other currencies, is yuan or thousand yuan or other units. Prices should, therefore, both as the property of commodity, and also should be described as a class and have their own attributes: currency and unit, that is to say, and commodity prices is the attribute of ontology - relationship, expressed as price = attribute (commodity).

class URI	describe type	class description value	conversion formula	standby conversion formula
XX/commodity				
XX/price				
XX/currency			target instance = source instance * relationship value	target instance = source instance / relationship value
XX/scale			target instance = source instance * relationship value	target instance = source instance / relationship value

B. Example of OWL ontology relational database

For the same ontology class instance data, users tend to be more hope they agree to facilitate computer processing semantic environment, section the scenario described above, for example, for multiple class attribute B2C website in prices as a commodity, in order to achieve the price semantic value to establish the OWL ontology relational database, (the ontology database only discuss prices for goods class attribute, actually goods class there are a lot of properties, and different kinds of goods have different attributes, but all goods instance can have this property prices) process is as follows:

1) Create a class table. The price as an attribute of the commodity class varies by currency and unit, so there are four classes in the class: commodity, price, currency and scale. The currency and scale are the properties of the price, but itself as a class description, transformation formula is: the target instance = source instance * relationship value, standby conversion formula is: the target instance = source instance/relationship value. Is given priority to with uri-like keywords, create the corresponding table, as shown in Table 1.(Note:XX/represents a URI network)

TABLE 1
class

2) Create a property table. In the class table, price is the property of commodity, and currency and scale are the property of price class, respectively, as the attribute name of their class name, the domain is the subordinate class, and the range is its own class. The attribute URI is the key word, creating the corresponding property table, as shown in Table 2.

TABLE 2

property

attribute URI	attribute type	domain	range
XX/price	ObjectProperty	commodity	price
XX/currency	ObjectProperty	price	currency
XX/scale	ObjectProperty	price	scale

3) Create a property value relational table. Price has two semantic properties: currency and unit. The currency's range is the world's existing currency, taking RMB, Hong Kong dollar and dollar as an example, its relationship value is the exchange rate between currencies. The unit's range is the decimal counting unit, taking the commonly used yuan and thousand yuan as an example, its relationship value is the advance rate of both. The property value relational table is created with the attribute URI, target attribute value, and source attribute value as the compound primary key, as shown in Table 3.

TABLE 3

property value relationship

attribute URI	target attribute value	source attribute value	relationship value
XX/currency	RMB	dollar	6.8
XX/currency	RMB	HK dollar	0.88
XX/scale	thousand yuan	yuan	0.001

4) Create an instance table. To take a sample of cell phone data from multiple B2C sites and create an instance of an instance of a product class, (considering the length of the case, it's just a couple of data, in fact, the same type of data on multiple B2C sites is a lot more than that) as shown by the following Table 4.

TABLE 4

instance

instance URI	attribute URI	attribute URI	attribute value
XX/iphone7	XX/commodity	XX/price	649
XX/649	XX/price	XX/currency	dollar
XX/649	XX/price	XX/scale	yuan
XX/HUAWEI P10	XX/commodity	XX/price	4388
XX/4388	XX/price	XX/currency	RMB
XX/4388	XX/price	XX/scale	yuan
XX/vertu	XX/commodity	XX/price	68
XX/68	XX/price	XX/currency	HK dollar
XX/68	XX/price	XX/scale	thousand yuan

With the standby conversion formula, the property value relational table can store half of the data records less. Because only when the formula is converted, when a record target attribute value is in tune with the source property value, their relationship values need to be changed to cater for the transformation formula. For example, the target attribute value of the target attribute in table 3 is changed to the value of yuan and source property, and their relationship value should be changed to 1.136, so a new record is needed to store the new relational value. That is, each pair of property values requires two records to store two relational values. Switch and the attribute values after new relationship value is based on the inverse operation of transformation formula, so only need to increase the field in the top class spare transformation formula to store the inverse operation of transformation formula, then each pair of attribute values only need a record to store the value of a relationship. Define a Boolean global variable, useInverseFormula, to determine which formula is invoked to perform the property value conversion, and when the value is false, the transformation formula is called; Instead, call the

standby transformation formula.

C. Example of semantic value conversion

After the database has stored the data, the user can call the data for processing. If the user expects the price attribute of all the commodity instances to be unified in currency as RMB and the unit is processed in the form of yuan, then the first step is to convert the price instance of the instance table to the form of "RMB and yuan" in the form of the semantic value conversion method in order to convert the price instance of the instance table to the "RMB and yuan" in turn. The steps are as follows:

1) Attributes of the current instance iPhone7 as commodity prices instance value of $649, calls the convert function, enter the name for the class: price, instance: 649, source attribute value: $yuan, target attribute value: RMB yuan. The source attribute value and target attribute value have two, and the single source property value is obtained after separation: usd corresponds to RMB; The value of a single target attribute: yuan. Convert the first single attribute value first.

2) According to a single target value of the RMB and a single source attribute value of the dollar, the attribute value relational table query out the corresponding relationship between a value of 6.8. The variable useInverseFormula assigns a value of false.

3) It is the currency class that the value of the property value relation table and the property table are connected to find out that the domain of the domain is the price class and the property is the currency.

4) By using the range currency class and the variable, useInverseFormula, the conversion formula in the class table is found to be the target instance = source instance * relationship value.

5) Convert source instance 649 and relational value 6.8 into the transformation formula: target instance = source instance * relational value, get the target instance 4413.2.

6) In the instance table, the source instance 649 is changed to target instance 4413.2, and the value of the single source attribute value is changed to the target attribute value. Source instance 649 also serves as the price attribute value of the instance iphone 7, so the price attribute value of the instance iphone 7 is also changed to 4413.2.

7) The second single attribute value is processed, where the value of a single source attribute is equal to the value of a single target attribute, so no conversion is required. At this point, all single source attribute values of the current price instance are converted to a single target attribute value, and the final object instance is 4413.2 yuan.

The instance value of the price attribute of the second commodity instance in the instance table, HUAWEI P10, is 4,388 yuan, and the source attribute value is equal to the target attribute value, so no conversion is made. The third instance in the price of goods instance vertu attribute value is 680000 Hong Kong dollars, the same as the above steps, the attribute value is separated into two single attribute value after the first 680000 Hong Kong dollars into RMB 598400, and in the transformation of the second single attribute value, attribute value relational tables, a single target attribute value of input yuan the number of records of the source attribute values, and yuan corresponds to a single source attribute value of input is the record of target attribute values, so the variable useInverseFormula assignment to true. The formula chosen in the following operation is the alternate transform formula: the target instance = source instance/relational value. In the end, the third price case was hk $68,400 converted to rmb598,400.

In the table above is completed for instance all instances of the semantic value conversion price, in the process of practical application, the user can add more relational tables in the attribute value attribute values for semantic value, can also according to the requirements specified in some instance data into different semantic environment.

V. CONCLUSIONS

On the basis of Web information extraction technology, this paper proposes a semantic value transformation method based on ontology for semantic annotation, and introduces the design of semantic value conversion method. In the process of semantic annotation, this method can transform all ontology instances into semantic values in multiple semantic environments according to user requirements, and improve the interoperability of data instances.

REFERENCES

[1] Huang W G, Zhu M, Yin W K, et al. Web Information Automatic Extraction Based on DOM Tree and Visual Feature[J]. Computer Engineering, 2013, 39(10):309-312.

[2] Reeve L, Han H. Survey of semantic annotation platforms[C]// ACM Symposium on Applied Computing. ACM, 2005:1634-1638.

[3] http://www.w3. org/TR/2004/REC-owl-guide-20040210/.

[4] http://www.w3.org/TR/2004/REC-owl-features-20040210/.

[5] Yong L I, Yue-Long L I. On the Method of Storing the OWL Ontology Based on Relational Databases[J]. Computer Engineering & Science, 2008, 30(7):105-107.

[6] Luo J, Chen B. Research on Method About Relational Database Storing OWL Ontology[J]. Computer Engineering, 2010, 36(21):71-72,75.

[7] Wang S H, Zhang X D, Wang Y. An Approach to Building an OWL Ontology Relational Database[J]. Computer Engineering & Science, 2011, 33(12):143-147.

[8] Silberschatz A, Korth H F, Sudarshan S. Database system concepts[J]. 1991, 8(1):1 – 2.