On Transformation from The Thesaurus into Domain Ontology

Ping Li, Yong Li

Department of Computer Science and Engineering, Qujing Normal University
Qujing, 655011, China
E-mail: qjncliping@126.com

Abstract: **In order to reduce the workload of constructing domain ontology, this paper proposes a transformation method from thesaurus into domain ontology. We can take building the cereal crops' domain ontology as an example to detail narrative of the process of the transformation and to prove the effectiveness of the ontology construction method.**

Keywords:domain ontology;thesaurus; ontology construction

I. INTRODUCTION

Ontology, a conceptual modeling tool which can be used to describe the information system on the semantic and knowledge, has been widespreadly concerned by researchers at home and abroad as a focus research. Traditional ontology construction process is complex, time-consuming and over-reliance on experts and so on. Automatic or semi-automatic ontology building can reduce the dependency on experts and speed up, reducing the development time and the cost of ontology construction. Now, thesaurus has been used to index and retrieve knowledge in books, information management and information retrieval, and it has formed a scientific organization.

We contain some simply fixed semantic relationships and concepts of classification and grading, which is the same with ontology in some aspects. Seeing thesaurus with the inherent characteristics, it has more clear vocabulary semantic structure than any other, and it is similar to ontology construction based on building knowledge classification and semantic relationship. In order to reduce the complexity of ontology construction and improve building efficiency, many scholars shared reuse of existing thesauri, built on the basis of the ontology, to save time and effort. It also enabled these vocabularies further development in the new environment. The main ideas of conversion from the thesaurus to domain ontology are as follows. First, applying thesaurus relationship among vocabularies as a data source, we can transform vocabulary of the thesauri into ontology classes or concepts, transform the relationship among words of the thesaurus into the relationship among the ontology classes. On the basis of those, the basic elements and relationships in the domain ontology structure can be built. Second, we can use owl language to describe and express ontology. Last, we can generate the domain ontology owl file. This paper introduces how to build cereal crops domain ontology based on agricultural sciences, and use protégé tool to build the domain ontology and use owl to describe the specific content of the domain ontology. What Carried out in the latter part of the body of the query and extended reasoning provides a good foundation[1].

II. DESIGN ABOUT THE TRANSFORMATION METHOD FROM THESAURUS TO THE DOMAIN ONTOLOGY

About the transformation method from thesaurus to the domain ontology, detailed process has been shown in Figure 1.
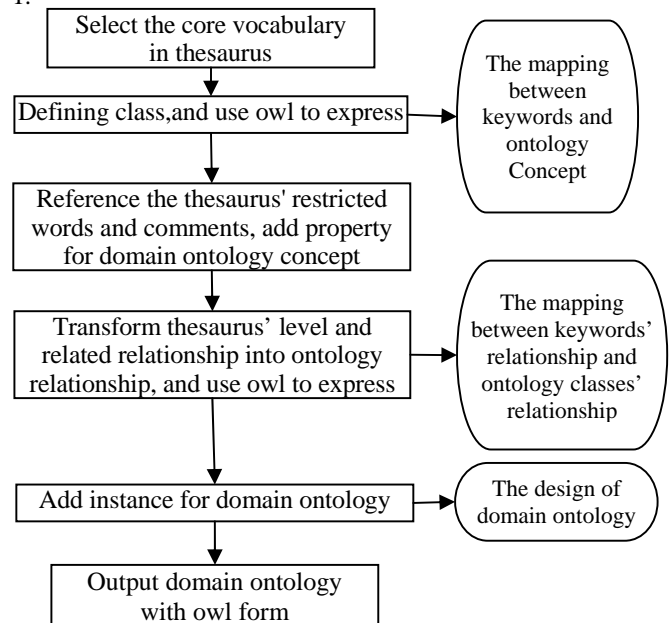

Figure 1. Detailed design module of ontology construction based on thesaurus

A. Rule about the transformation from Thesaurus to the domain ontology

Now, the research of domain ontology creation is still in its infancy, and it does not have a standard method of ontology construction. This paper, learning from the existing methods of building domain ontology and following ontology construction guidelines, proposed a domain ontology construction method based on Thesaurus. The method is logic, have better scalability, and can be applied to construct some domain ontology. At the time of ontology construction, some requirements and rules are needed to deal with the problems that are encountered in the process of transformation and to confine the scope of the ontology. The requirements and rules for constructing domain ontology are follows.

1) Try to keep the original naming.
2) Try to keep close to the thesaurus' original structure.

3) As far as possible, try to reflect the information already in the thesaurus to ontology and enrich ontology's architecture and content.

4) Refinement Thesaurus contains conversion rules of the relationship, which enrich the relationship between words.

5) To avoid information overload, the command has nothing to do with the semantic information which is not transformed.

B. The overall design of the domain ontology

First, we need to determine the areas and scope of the ontology. After information gathering, the built cereal crops' domain ontology prototype should be able to have basic elements and characteristics of the ontology, and its concept scope should be closely related to the concept of cereal crops' important concept. Secondly, we need to define the domain ontology's application purpose. The ontology is reqested to be efficiently built, so that the ontology can be really used in Web services, and achieve the intelligent information search. Finally, we need to determine ontology's development tools. Protégé has a good user interface whose plug has better scalability. Therefore, this choice of Protégé develop tools for the construction of the ontology in the field of cereal crops?

C. The detailed design of the domain ontology

1) Determine the set of core concepts in the field of cereal crops

First, identifying the field's most important concepts bases on thesauri and other relevant important concepts, and establish a set of core concepts. Then, referring to the firstly determined concept of cereal crops base on thesauri, and after the collection and analysis of cereal crops' field information, it turns to determine cereal species. Referring to thesauri, cereal crops are divided into rice, wheat, corn, sorghum and other classes, such as sub-concepts of cereal crops. When we transform the determined keywords to the concept of ontology, we can use owl to express it as the following.

```
<owl:Class rdf:about="#cereal crops">
<rdfs:subClassOf>
<owl:Class rdf:about=" rice"/>
</rdfs:subClassOf>
<rdfs:subClassOf>
<owl:Class rdf:about=" wheat"/>
</rdfs:subClassOf>
……
</owl:Class>
```

2) Add property to domain ontology concept

It is not enough to only describe the concepts in the field for a field. It also need to establish the attributes of various ontology elements, because attributes reflect the relationship of ontologies. Meanwhile, we must add attribute and to the concept of the core concept set, and classify attributes, and then, refine the properties of each category. For example, producer, effect, cultivation means, production institution, are set to the cereal crops' property concept. In addition, the property concept has inheritance, and the concept of sub-concepts can inherit properties. For example, cereal crops

have the property of producer, then, its the sub-concepts wheat, corn, rice also inherit the property of the producer. Of that, we can also set characteristics of the sub-concepts. In order to understand adding properties to cereal crops, we can see the E-R diagram in figure 2.
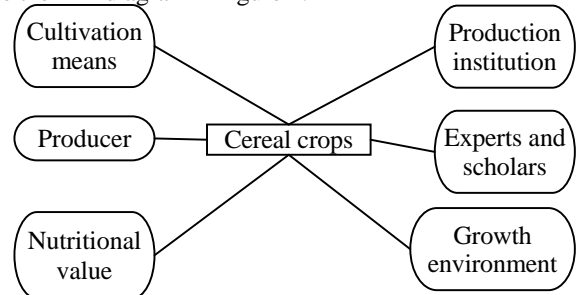

Figure 2. Add cereal crops' attributes

Using owl to express it as follow.

```
<owl:ObjectProperty rdf:about="producer">
</ owl:ObjectProperty >
<owl:ObjectProperty rdf:about="effect">
</ owl:ObjectProperty >
……
```

3) Add instance

According to the detailed design process of ontology construction cereal crops, and after adding properties to the concept, we should add instance to each concept. For example, we can use wheat, barley, oats, buckwheat as instances of the wheat crops. Look at the E-R diagram in Figure 3
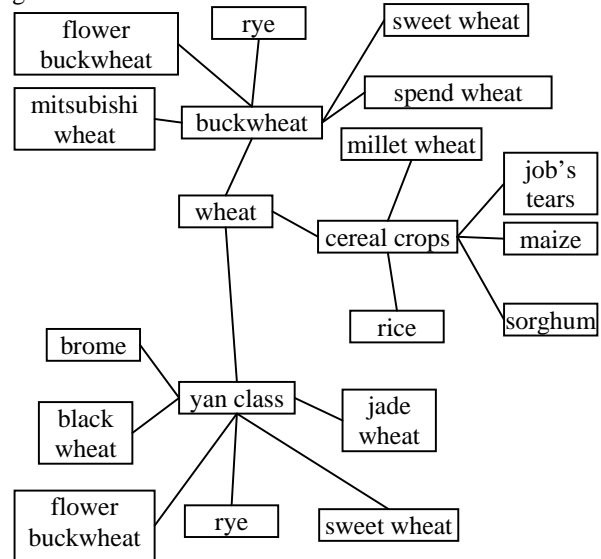

Figure 3. E-R diagram of the cereal crops

In figure 2, the cereal crops' properties include cultivation means, production institutions and so on. And the cereal crops' subclass contains rice, wheat, sorghum, maize, job's tears, and millet. According to the rules of ontology construction, subclass inherits the properties of the parent class. So, all cereal crops' subclass also has the above properties. Further more, the sub-class also has respective

instance. For example, wheat as a subclass of the cereal crops, also contains barley, wheat, buckwheat, and oats. And corresponding to the concept of the ontology, it has specific instance that is wheat's subclass. So, let's take owl as an example to express it as the following.

```
<owl:NamedIndividual rdf:about="barley">
</ owl:NamedIndividual>
<owl:NamedIndividual rdf:about="wheat">
</ owl:NamedIndividual>
...
</ owl:NamedIndividual>
```

4) To determine the relationship among concepts

The MeSH relationship is mainly divided into equivalent relationship, hierarchical relationship, related relationship and so on. Also to can be expressed by Use, Generation, Minute, Race, Ginseng, then respectively abbreviated as Y, D, S, F, Z, C. Among them, Y and D represent equivalent relationship, S and F represent hierarchical relationship, Z represents the first word of the largest tribe in a group of words, C represents related relationship. In the OWL language, using classes, individual and attributes and other among relations' statements to describe the relationships the entities, using OWL to describe the MeSH relationship[2]. Figure 3 shows that, the class of rice, sorghum, corn, coix, millet, wheat have hierarchical relationship with cereal crops, the class of wheat, barley, oats, buckwheat have hierarchical relationship with wheat classes, the class of triangular wheat, flower buckwheat, rye, sweet wheat, flower wheat have equivalent relationship with buckwheat, the class of brome, rye, the bells wheat, hong mai, jade wheat, sulu have equivalent relationship with oats. With owl, the relationship between them is expressed .

a) The owl representation of equivalent relationship:
With two words as equivalentClass in the equivalent relationship's OWL description. The above example E-R diagram is described with " <owl:Restriction> " statement, the class of flowers wheat have equivalent relationship with buckwheat.

```
<owl:Class rdf:ID="# flowers wheat">
<owl:equivalentClass rdf:resource="& buckwheat"/>
<owl:equivalentClass>
<owl:Restriction>
<owl:onProperty rdf:resource="#hasPreferredTerm"/>
<owl:hasValues rdf:resource="# buckwheat"/>
</owl:Restriction>
</owl:equivalentClass>
</owl:Class>
```

The following equivalent relationship is linked from the formal MeSH to the informal MeSH, but also with formal MeSH as OWL classes, then, we got the following semantic definitions.

```
<owl:Class rdf:ID= "oats">
<owl:equivalentClass rdf:resource="& rye"/>
<owl:equivalentClass>
<owl:Restriction>
<owl:onProperty
    rdf:resource="#hasnonPreferredTerm"/>
<owl:hasValues rdf:resource="& rye"/>
```

```
</owl:Restriction>
</owl:equivalentClass>
</owl:Class>
```

The description defines oats as formal MeSH, also links it to informal MeSH of rye, it takes into account the possibility of connection between natural language and open environment. After the above owl, we can regard equivalent relationship as a bridge in the retrieval process, and retrieve the relevant information in different fields, and expand the knowledge of ontology. But, Equivalence class has only simply associated with two classes, there is no description of a specific semantic relationship among words. For example, two Use-Generational relationships' MeSH with each other, is similar to the function and purpose, or related in content and form, or two completely antagonistic equivalent that can not be determined. Reference [3], the equivalent relationship can be refined and adjusted to several semantic relationships, as shown in the following table I.

TABLE I.   THE EQUIVALENT RELATIONSHIP'S THINNING AND CATEGORY

Before refine	After refine
Use（Y）/ Generational（D）	Synonymy relationship
	Near-Synonymy relationship
	Antonymy relationship
	Determinative relationship
	Generic relationship

b) The owl representation of hierarchical relationship:
The hierarchical relationship is hyponymy of a kind of vocabularies, the tag <rdfs:subClassOf> in the OWL language that could specify directly hyponymy. For instance, according to the diagram 3, we can know that there are some classes have hierarchical relationship with the wheat classes, including: wheat, barley, oats, buckwheat;

```
<owl:Class rdf:about="# wheat classes" >
<rdfs:subClassOf>
<owl:Class rdf:about="#wheat" >
</rdfs:subClassOf>
<rdfs:subClassOf>
<owl:Class rdf:about="# barley" >
</rdfs:subClassOf>
......
</owl:Class>
```

A simple hierarchical relationship can be defined in the above code, while you can use many kinds of methods to define all children of a class, the conventional method is to define the union set of all subclasses in a extension class. The definition above also could be defined as follows:

```
<owl:Class rdf:ID="# wheat classes" >
<rdfs:subClassOf rdf:resource="# cereal  crops" >
<owl:unionOf rdf:parseType="Collection">
<owl:Class rdf:about="# wheat" >
<owl:Class rdf:about="# barley" >
<owl:Class rdf:about="# buckwheat" >
<owl:Class rdf:about="# oats">
</owl:unionOf>
</owl:Class>
```

The wheat classes is a subclass of cereal crops, which is defined in the above code, although the wheat originally includes four subclasses: wheat, barley, oats, and buckwheat.

Among seeing reference[3] , the hierarchical relationship between subject terms could be refined into three kinds of relationships, as following table II:

TABLE II.   REFINEMENT AND CLASSIFICATION OF THE HIERARCHICAL RELATIONSHIP

Before refine	After refine
Blong to（S）/Branch（F）	Generic relationship
	Whole_Part relationship
	Instance relationship

c) The owl representation of related relationship:

We can use reference(C) to express the relevant relationships between the subject terms, for it is a reciprocal relationship, the notion is that   "A related B, B related A". In the following example, there is the relevant relationship between "oats" and "nutritional value", so the two words can be connected through the attribute called RelativeTerm (related words).

```
<owl:Class rdf:ID="oats" >
<owl:Restriction>
<owl:onProperty rdf:resource="#hasRelativeTerm"/>
<owl:someValueFrom    rdf:resource="# nutritional
value"/ >
</owl:Restriction>
</owl:Class>
```

The relevant relationship is an important means that can be used to reveal all kinds of contact between subject terms, and expand the search range, and find the related data. The range involved is wide, and complex, conversion process in the ontology construction, and it can be expanded according to the requirement of users in the specific application of environment (such as domain expert), mainly refined into twenty kinds of relationships[13], as the following table III:

TABLE III.   REFINEMENT AND CLASSIFICATION OF THE RELATED RELATIONSHIP

Before refine	After refine
Reference(C)	Object or Action/Property relationship
	Product/Raw material relationship
	Object or Action/Measurement relationship
	Object/Origins relationship
	Discipline or Object relationship
	Vocation or Practitioner relationship
	……

After refining the equivalent relationship of the subject terms, the hierarchical relationship of the subject terms, and the relevant relationship of the subject terms, we can refine and classify the attributes and relationship of concepts. Coarse granularity ontology is refined into fine particle ontology, which makes the concept of ontology constructed and the relationship between concepts more accurate.

D.   Represent of the domain ontology

There are many kinds of ontology languages, protégé to develop in the article, the use of   OWL ontology languages to build cereal crops ontology. In the following, we take cereal crops for example to descript how to express the relation of subject terms. A the word relationship of cereal crops B the word relationship index of cereal crops.

cereal crops D grain crops S crop F wheat classes corn sorghum C nutritional value	grain crops Y cereal crops crop F cereal crops wheat classes S cereal crops corn S cereal crops sorghum S cereal crops nutritional value C cereal crops

We define all relationships of a class as "cereal crops" In the OWL description above. The hypernym is "crop", and the nonPreferredTerm is "the cereal crops ", and the hyponym is "wheat classes", "corn" and "sorghum", and the related word is "nutritional value".

```
<owl:Class rdf:ID=" cereal crops" >
<rdfs:subClassOf rdf:resource="#crop"/ >
<owl:Restriction>
<owl:onProperty
rdf:resource="hasnonPreferredTerm"/>
<owl:someValuesForm>
<owl:Class>
<owl:oneof rdf:parseType="Collection">
<owl:Thing rdf:about="#wheat classes"/ >
<owl:Thing rdf:about="#corn"/ >
<owl:Thing rdf:about="# sorghum"/ >
</ owl:oneof >
</owl:Class>
</owl:someValuesForm>
<owl:Class>
<owl:onProperty rdf:resource="#hasRelativeTerm"/>
<owl:Class rdf:about="# nutritional value"/ >
</owl:Class>
</owl:Restriction>
</owl:Class>
```

III.   CONCLUSION

We propose a method to transformate subject list into domain ontology in this paper, by closing the purpose of ontology application and making full use of the existing concept in subject list, which have saved manpower and material resources and improved efficiency. There has been few people researching the method of developing ontology and how to construct domain ontology at home, these researches belong to start, exploration stage. In the actual transformation work, because the semantic relationship revealed by subject list sometimes is not very accurate, and there are some parts that need domain expert manual intervention after transformating relationship among words in subject list into ontology knowledge relationship. In addition to transformating automatically according to relationship, there are also a lot of manpower and material

resources to that can prove the validity of the knowledge content.

REFERENCES

[1] S.C. Ding, Z Fu, "Research on Semi-automatic Construction of Domain Ontology Based on Space Thesaurus,"Information Studies:Theory & Application, vol. 34(11), pp.113-116, November 2011.

[2] H. Bai, "Some Important Issues for the Ontological Semantic Description of Thesaurus in Chinese Language,"Library Journal, vol .29(11), pp.21-25, November 2010.

[3] L.B. Liu, R.J. Ren, "Research on elaborating the term-term relations of Chinese Thesaurus that build ontology based on thesaurus," The Library Journal of Shandong, vol. (1), pp.73-76, February 2010.

[4] J.Z. Ja, R.J. Wei, "Analysis on the Methods of Converting Thesauri into Ontology," Information Science, vo.l 27(9), pp. 1363-1366,September 2009.

[5] Q.Y. Seng, Q. Liu, "A Thesaurus-based Methodology to Specify the Semantic Meaning of Ontology," Information Science, vol.25(9), pp.1415-1418, September 2007.

[6] X.Y. Du, M. Li, and S. Wang, "A Survey on Ontology Learning Research,"Journal of Software, vol. 17(9),pp.1837-1847 , September 2006.

[7] J.D. Zhang, "Construction and Reasearch of Ontology Based on Thesaurus," Document Information & Knowledge, vol.(4), pp.82-85, July 2006.

[8] A. M. Tang, Q.Zheng,and J. Fan, "Thesaurus-based Approach to Build Domain Ontology," New Technology of Library and Information Service, vol. (4),pp.1-5, April 2005.