

# RELATION EXTRACTION BASED ON AGROVOC

Qian Wang

*Computer and Information Management Center, Tsinghua University, Beijing, China 100084;  
Address: Computer and Information Management Center, Tsinghua University, Beijing,  
100084, P. R. China, Tel:+86-10-62850228, Email: cauwq@163.com*

Abstract: Relation extraction is an important step in Ontology construction. This paper provides a method to extract relations among conceptions with AGROVOC.

Key words: relation extraction, AGROVOC, thesaurus

## 1. INTRODUCTION

Ontology is a conceptualization of a domain into a human understandable, but machine-readable format consisting of entities, attributes, relationships and axioms (Guarino et al., 1995). How to extract relationships among entities is a difficult task in ontology construction. AGROVOC (FAO, 2007) is a multilingual thesaurus in agriculture which has indexed lots of literatures. We can obtain the relationships from AGROVOC and its text resource.

## 2. RELATION EXTRACTION

Description logic is used to organize the ontology knowledge system. Description logic can be divided into two parts, and one is Tbox which contains the limited aggregate of terminological, the other is Abox which contains the limited aggregate of assertion. In this paper, the concept schema transformed by AGROVOC can be considered as Tbox. We can obtain Abox

from literature database indexed by thesaurus. We can extract the further relations and properties from literature database. For instance, if a property,  $P$ , is tagged as a SymmetricProperty property, we can conclude:

$$P(x,y) \Rightarrow P(y,x)$$

Contrarily, if the rule above is given, the schema can be built as follows:

```
<owl:ObjectProperty rdf:about="P">
  <rdf:type rdf:resource="http://www.w3.org/2002/07/owl#SymmetricProperty"/>
</owl:ObjectProperty>
```

Each literature can be converted into an instance of the tree by distributing the tags (Muhammad et al. 2005). To obtain the basic relationship, we should input those verbs surrounded by agricultural concepts. Some of the key behavioral features of concepts can be extracted by algorithm Morphological variants (Muhammad et al. 2005). Thus the system can obtain the frequently occurring triplets of the form  $\langle s, p, o \rangle$  efficiently. But few property rules defined in OWL (W3C 2004) are identified.

We will give an algorithm for the further Relation Extraction (RE). In this algorithm, we use some functions and sets explained as follows:

- $I$  is a set of instance statements obtained from literature database. Every statement is represented as  $\langle s, p, o \rangle$ .  $s$  is subject,  $p$  is predicate and  $o$  is object;
- $P = \{ \langle s, p, o \rangle \mid \langle s, p, o \rangle \in I \}$ , where  $\langle s, p, o \rangle \in I$ , return all the instances which contain property  $p$ ;
- $R$  is a set of property rules defined in OWL;
- $Support(p_r)$  is the support degree where  $p$  satisfies  $r$ , which has specific formula according to  $r$ ;
- $Confidence(p_r)$  is the confident degree where  $p$  satisfies  $r$ , which also has specific formula according to  $r$ ;

**Algorithms:** Relation Extraction (RE)

**Input:** Literature database indexed by thesaurus;

Threshold  $S$  of  $Support(p_r)$ ;

Threshold  $C$  of  $Confidence(p_r)$

**Output:** Property definition of OWL

1. First, find instances which can be represented with ID composed of literature and indexing order from literature database.
2. Prepare  $I$  from instances database.
3. Prepare  $P$  by classifying  $I$  by property.
4. for each  $p \in P$  {

```
for each  $r \in R$  {  
    compute Support ( $p_r$ ) according to  $r$ ;  
    compute Confidence ( $p_r$ ) according to  $r$ ;  
    If ( Support ( $p_r$ ) <  $S$  || Confidence ( $p_r$ ) <  $C$  ){  
        continue ;  
    }  
     $p_r$  is true;  
}  
output  $p_r$  definition of OWL;  
}
```

### 3. CONCLUSION

In this paper we propose a new method by which we can extract the OWL relationships from literature database indexed by thesaurus. The proposed algorithm is used to construct domain ontology in Agriculture. The extraction process can be completed automatically or semi-automatically by computer so that it depends less upon experts. However, it is not perfect in multiple-level relation mining and other axioms. As for future research, we intend to adjust the extraction arithmetic to support multiple-level relations and mine other axioms.

### REFERENCES

- Guarino N and Giarretta P. *Ontologies and Knowledge Bases: Towards a Terminological Clarification. Toward Very Large Knowledge Bases: Knowledge Building and Knowledge Sharing*, IOS Press, Amsterdam, 1995.
- FAO. *AGROVOC Thesaurus*. [http://www.fao.org/aims/ag\\_intro.htm](http://www.fao.org/aims/ag_intro.htm), 2005.
- Muhammad Abulaish, Lipika Dey: *Biological Ontology Enhancement with Fuzzy Relations: A Text-Mining Framework*. *Web Intelligence 2005*: 379~385
- W3C. *OWL Web Ontology Language Guide*. <http://www.w3.org/TR/2004/REC-owl-guide-20040210/>, 2004.