Extracting Semantic-preserved Ontology from Relational Database

Songqing Wang, Zhuang Miao[*], Yingpeng Du, Yang Li and Jiabao Wang

College of Command Information Systems, PLA University of Science and Technology, Nanjing, China, 210014

emiao_beyond@163.com*

Keywords: ontology extraction, relational database, semantic preserved.

Abstract. To solve the problems that ontology is lack of semantic information in ontology extraction, the advantages and disadvantages of methods that building ontology from database are analyzed and we find that the maximum semantic ability of ER model is ALUIN. On the base of our research, a method of extracting semantic-preserved ontology is proposed. The method is the base of ontology storage and query. The experiment shows that our algorithm can preserve the semantic meaning and has the low complexity.

Introduction

Data in a relational database is generally geared to the needs of a certain professional field, this characteristic makes the relationship between relational database structure is very stable, the data in the database contains a lot of information, thus drawn from the relational database of the ontology, has a fixed structure and rich semantic information, analysis and operations easily. Now there are more ways to extract the database, but has less than in the semantic aspects, so to keep relational database semantic extraction has the research value.

OWL DL equivalent to the description logic SHION. SHION is passed on the basis of ALC increased the relationship, inverse relationship level concept, enum, and absolute number a few, such as more than ALUIN which constraints, complexity and power of expression than ALUIN. Therefore describing ER in OWL DL pattern extraction is feasible after the formation of the ontology[1].

In the ER model to the OWL DL in the process of transformation, using only the part of the operator of the OWL DL ontology expression[2]. Based on the above model of ER and ALUIN equivalence relation, combined with the design method of ontology[3], and keep the semantic method of these four patterns[4] and ER2WO in-depth analysis, this paper designed a kind of semantic-preserved ontology extraction algorithm ER2OWL_P. The algorithm is based on ER model maximum semantic, the database ER model for extraction of ontology, conform to the largest database semantic ontology, it can improve the efficiency of ontology query.

Semantic-preserved Extraction Algorithm based on ER model

ER model can be abstracted as a five-element model $s = (L_s, isa_s, att_s, rel_s, card_s)$ [5]. Where, L_s is the union set of Entities, attributes, contact, roles and domain, isa_s is binary relation of 'ISA', att_s is the only attribute mapping function of the entities, rel_s and $card_s$ are mapping functions. Reference [6] evaluated the semantic ability of ER model and proved the equivalency of description ALUIN and ER model. Based on the conclusion, we proposed a semantic-preserved ontology extraction algorithm ER2OWL_P. The main steps of the algorithm are as follows.

Step 1: ER five-element model of input variables of inspection, L_s will be set apart from the rest of the four collection, skip to Step 2;

Step 2: transform the elements of L_s. Entities, relations, attributes, entities, domain and character symbols are transformed to corresponding ontology classes. If there is need to transform the concepts, turn Step2, otherwise turn Step3;

Step 3: set the remaining four transformation. The relations in isa_s are transformed. The entities in att_s are created as attribute axioms and class axioms. The relations in rel_s are created as attribute axioms and class axioms. The class and properties in $card_s$ are add with constraints. jump to Step 4.

Step 4: complete the input ER pattern extraction, and output the generated ontology. Algorithm ER2OWL_P pseudo code is shown below.

Input: ER model
Output: OWL ontology
Begin：
For each variable *t* in *S*
Switch(*t*)
Case *L*:
   for each variable *w* in *L*
     if(*w*==*E*), create *class* $\phi(E)$
     if(*w*==*R*), create *class* $\phi(R)$
     if(*w*==*A*), create *property* $\phi(A)$
     if(*w*==*D*), create *datatype* $\phi(D)$
     if(*w*==*U*), create *datatype* $\phi(U)$
Case *ISA*:
   for v each atom *u, v* in *ISA*
     if(*u, v* $\in (\phi(E))$), create *subclassof(u,v)*
Case *Att*:
   for each atom *w, y, z* in *Att*
     if($w \in \phi(E)$ && $y \in \phi(A)$ && $z \in \phi(D)$), create *class(w partial restriction(y allValuesFrom(z) cardinality(1)))*
Case *Rel*:
   for each atom *r, s, w, v* in *Rel*
     if(r&&$w \in \phi(E)$ && $s \in \phi(U)$), create *class(r partial restriction(s allValuesFrom(w) cardinality(1)))*
     if(s&&$w \in \phi(E)$ && $r \in \phi(R)$), create *ObjectProperty(s, domain(r, range(w)))*
     if($v \in \phi(U)$ && $s \in \phi(U)$ && $w \in \phi(E)$ && $r \in \phi(R)$), create *ObjectProperty(v, domain(w, range(r)),inverseof(s))*
Case *Card*:
   for each atom *w, r, s* in *Card*
     if($h = \min cards(w,r,s) \neq 0$), create *class(w partial restriction(s mincardinality(h)))*
     if($h = \max cards(w,r,s) \neq 0$), create *class(w partial restriction(s maxcardinality(h)))*
End

Experiment and Discussion

In order to verify the validity of the ER2OWL_P algorithm, we designs a set of experimental verification. Experimental operation tools including Eclipse and MySQL. We used the University database[7] as the experiment database. The experiment of ER2WO ER2OWL_P method, this paper conversion algorithm, database direct extraction algorithm Strova method for validation, the number of classes and properties for extracting ontology were analyzed. Because ER2WO laws of semantic-preserved, this experiment will be based on the ER2WO method, if the class and attribute consistent with ER2WO method, the semantic properties is good, otherwise, the semantic properties. In this paper, the use of PowerDesigner tool to reverse generation of University database, get the University ER model, as shown in Fig. 1.
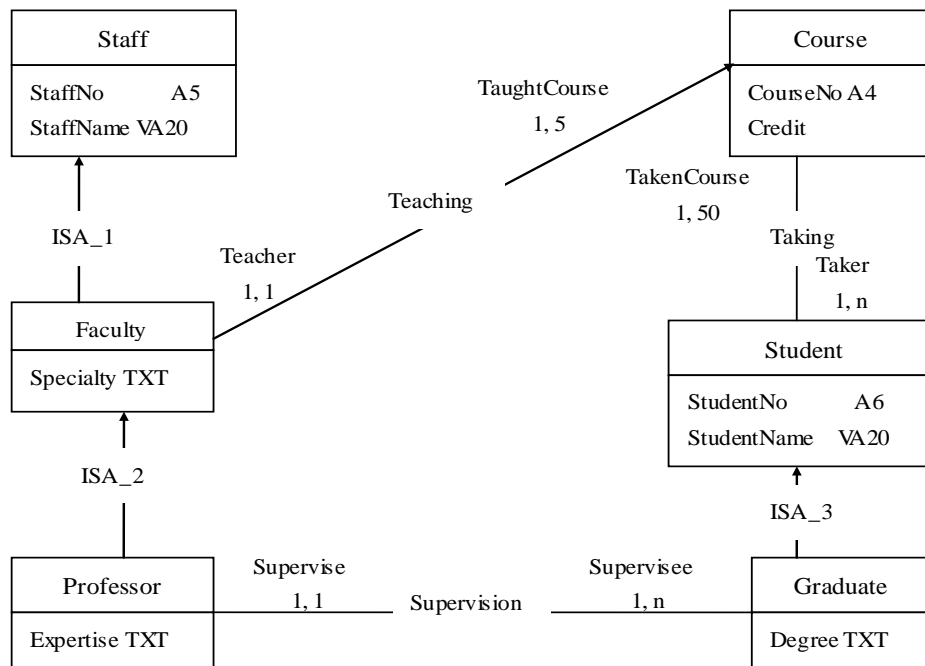
Fig. 1 The ER model of the University database

In the University of ER model, can get the ER model entities in the symbol set, attribute symbol set, role symbol set, contact symbol set, domain symbol set. With three algorithms to extract of University database, get the corresponding ontology, through statistical analysis of the classes of ontology, get the results of three kinds of algorithms. By the transformation of the above, corresponding ontology consists of the following is an identifier with class, as shown in Table 1. Fig.2 shows the class identifier number comparison.

Table 2 Class identifier in the ontology

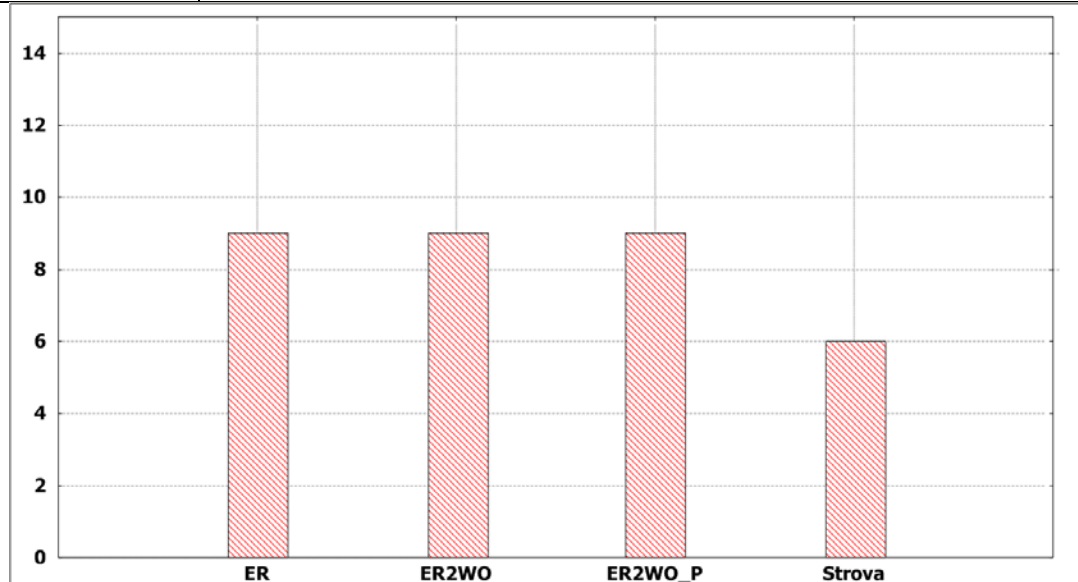| Algorithm	Class identifier
ER2WO	owl:Staff, owl:Faculty, owl:Professor, owl:Student, owl:Graduate, owl:Course, owl:Teaching, owl:Taking, owl:Supervision
ER2OWL_P	owl:Staff, owl:Faculty, owl:Professor, owl:Student, owl:Graduate, owl:Course, owl:Teaching, owl:Taking, owl:Supervision
Strova	Staff, Faculty, Professor, Student, Graduate, Course

Fig. 2 Class identifier number comparison

From Fig. 2 we can see that in the algorithm Strova, class identification and object attribute identification number are less than ER2WO and ER2OWL_P. This is because the database extraction of ontology, the table structure is very strict, the structure of relationships but there is no corresponding table without extraction, which led to the missing during extraction, and the connection between the ER model in relation to the entity. The rest of the two algorithms are to maintain the relationship. Therefore, the algorithm ER2OWL_P is semantic-preserved, and has lower than algorithm ER2WO logic complexity.

Conclusion

This paper discusses the current relational database extraction method were analyzed and compared, keep semantic properties of every method, has carried on the comparative analysis, sums up the semantics of the various models maintain performance. A semantic-preserved ontology extraction algorithm is presented, which has lower complexity of logic.Experiment results show the effectiveness of the algorithm.

References

[1] F. Bobillo, U. Straccia, Fuzzy ontology representation using OWL 2, International Journal of Approximate Reasoning, 2011, vol. 52, pp. 1073-1094.

[2] L. Bellatreche, Y. Ait-Ameur, Chedlia Chakroun, A design methodology of ontology based database applications, Logic Journal of IGPL, 2011, vol. 19, pp. 648-665.

[3] A. Escobar, Notes on the Science of Ontology. England: Literary Licensing, 2009.

[4] S. Sacchi, What do we mean by 'Preserving information'? Towards a formal ontology for digital preservation. England: Holy Cross, 2010.

[5] G Zhang,S Jia , Ontology-based knowledge extraction for relational database schema, in the 2th International Symposium on Electronic Commerce and Security, 2009, pp. 585-589.

[6] YP Du, Z Miao, Y Zhang, W Xu, Q Zhang, Evaluation of semantic ability of er model. Advances in Intelligent and Soft Computing. Springer Berlin Heidelberg, 2014:713-720.

[7] University database. http://cse.seu.edu.cn/people/ysdong/graduates/~zmxu/ER2WO/.