International Journal of Database Theory and Application
Vol.7, No.3 (2014), pp.73-82
http://dx.doi.org/10.14257/ijdta.2014.7.3.08

An Effective Approach to Fuzzy Ontologies Alignment

Ling Qiu[1] and Yong Liu[2]

[1]School of Computer Science, Sichuan University of Science and Engineering Zigong 643000, China

[2]Artificial Intelligence Key Laboratory of Sichuan Province, Sichuan University of Science and Engineering, Zigong 643000, China

lingqiusichuan@163.com

Abstract

Ontology alignment finds matching elements from different ontologies. A number of ontology alignment techniques have been proposed. But few of them are adaptive to fuzzy ontologies alignment. A key problem is that formulas for computing similarity between elements from precise ontologies can't be directly applied to those elements from fuzzy ontologies. In this paper, we exploit WordNet to compute similarity between two words. And Inspired by formula for computing similarity between two formal concepts from concept lattice, we propose an effective similarity formula for computing similarity between classes from different fuzzy ontologies. Also, we present an algorithm which is global optimal for aligning fuzzy ontologies. A prototype is implemented and the experiment shows that our approach is worthy.

Keywords: *Fuzzy Ontology; Fuzzy Concept Lattice; Ontology Alignment*

1. Introduction

Ontology is explicit specification of conceptualization. It aims to make data sharable and reuse. But in fact, multiply ontologies coexist although they are from same field. Owing to implementation intention and knowledge background of ontology engineers, these ontologies are usually heterogeneous and distributed. In order to fulfill knowledge share and reuse, need for mapping between ontologies becomes stronger. Ontology alignment is the process of bringing ontologies into mutual agreement by the automatic discovery of mappings between related classes. By far, a number of ontology alignment techniques have been proposed [11]. But few of them are adaptive to fuzzy ontologies alignment. A key problem is that formulas for computing similarity between elements from precise ontologies can't be directly applied to those elements from fuzzy ontologies. In this paper, we proposed a new formula for computing similarity between classes and an effective algorithm for fuzzy ontologies alignment.

This paper is organized as follows. Fuzzy ontology and fuzzy concept lattice are introduced in Section 2. Method for computing similarity between classes of fuzzy ontologies is shown in Section 3. The alignment algorithm for two fuzzy ontologies is given in Section 4. A prototype system is introduced in Section 5. The related works are mentioned in Section 6 and conclusion and the next work are arranged in the last section.

ISSN: 2005-4270 IJDTA

2. Related Work

Methods for ontology alignment may be classified into language-based methods, structure-based methods and machine learning methods.

For language-based methods, similarity between classes is evaluated according to class name. Techniques for comparing two of class names includes prefix/suffix comparison, edit distance and n-grams [6]. These techniques are difficult to recognize synonyms classes represented by different names, homonyms classes represented by same names. Also, they are ineffective for names written by complex phrases, sentences and descriptions. In order to improve these techniques, some extra knowledge bases, for example, WordNet, are exploited to extract syntactic information, semantic information, which is introduced into similarity formulas [7]. But the main limitation to language-based methods is that extra thesauri is required, which is usually dependent-language.

Unlike language-based, structure-based methods mainly consider structure information of ontology rather than linguistic information of single class. For example, the number of common, similar children or common parents between classes is used to compute similarity. And then, similarity is extended to others nodes along the graph structure based on the idea that similar nodes entail similar neighbors [8].

In machine learning methods [9], an initial similarity formula is given according to statistical distribution of features about classes such as symbolic, syntactic, semantic, and structural. But parameters are not decided in initial formula. Then, a small set of records about features are selected by users. Some pattern recognition algorithms are applied to initial similarity formula to optimize parameters. The mainly difficult is that initial similarity formal is hard to decide.

3. Fuzzy Ontology and Fuzzy Concept Lattice

3.1. Fuzzy Ontology

Definition 1. A **fuzzy ontology** O_F is defined as a 5-tuple [10]: $O_F=\{C,P_F,R_F,A_F,I\}$. where:

- C is the set of classes, which represents various entities in some domain being modeled. We assume that classes are named by one or more natural language terms and are normally referenced within the ontology by a unique identifier.
- P_F is a set of properties. Any $p_F \in P_F$ is defined as a 3-tuple: $p_F=\{c,v,q\}$. c is a class. v is value of p_F. q is a qualifier. An instance is "*This watch is very expensive.*". *price* is explained as property p_F. *watch*, *expensive* and *very* is explained as c, v and q.
- R_F is the set of relationships between classes. Any $r_F \in R_F$ is defined as a 4-tuple: $r_F=\{c_1,c_2,t,f\}$. Both c_1 and c_2 are two classes. t is type of relationship between c_1 and c_2. And t is one of *synonym of*, *kind of*, *part of*, *instance of*, *property of*. But in this paper, only *kind of* relationship is considered. f is a membership degree [0, 1] of t.
- A_F is a set of fuzzy axioms, usually formalized into some logic language. These axioms specify additional constraints on the ontology and can be used in ontology consistency checking and for inferring new knowledge from the ontology through some inference mechanism.
- I is set of all instances.

It is an important research point how to get a fuzzy ontology. Some approaches, like FFCA [1, 2], UML [3], *etc*, have been proposed. But compared to others, Fuzzy Formal Concept

Analysis is more popular owing to high similarity in structure between fuzzy concept lattice and fuzzy ontology.

3.2. Fuzzy Concept Lattice

Definition 2. A **fuzzy formal context** *FFT* is defined as a 3-tuple: $FFT=(B,A,P)$, where

- B is a set of formal objects.
- A is a set of formal attributes.
- $P \subseteq B \times A$. For any $b \in B$ and $a \in A$, $(b,a) \in P$ holds iff object b has attribute a with $\mu(b,a)$, a membership degree [0,1].

Definition 3. Given a fuzzy formal context $FFT=(B,A,P)$ and $X \subseteq B, Y \subseteq A$. $X'=\{a \in A | \mu(b,a) \geq \gamma \ \forall b \in X\}$ is said to be the **common attributes** of X. $Y'=\{b \in B | \mu(b,a) \geq \gamma \ \forall a \in Y\}$ is said to be the **common objects** of Y. γ is a threshold.

Definition 4. Given a fuzzy formal context $FFT=(B,A,P)$. A 2-tuple $ffc=(X,Y)$ is said to be a **fuzzy formal concept** of *FFT* iff $Y=X'$ and $X=Y'$ hold.

For example, consider a fuzzy formal context called *Chinese Cities* where B={*Beijing*, *Shanghai*, *Chongqi*, *Chengdu*, *Kunming*}, A={*Circumstance*, *Economic*, *Transportation*, *Sustainability*, *Science*} and R is showed in table 1. We assume the threshold γ is 0.5. A fuzzy formal concept is, for instance, the pair ({*Shanghai*, *Chengdu*, *Kunming*},{*Sustainability* (0.6)}).

In Table 1 and Figure 1, we abbreviate *Circumstance*, *Economic*, *Transportation*, *Sustainability* and *Science* to *Cir*, *Eco*, *Tra*, *Sus* and *Sci* respectively. Also, we abbreviate *Beijing*, *Shanghai*, *Capital*, *Chongqi*, *Chengdu* and *Kunming* to *Bj*, *Sh*, *Cq*, *Cd* and *Km* respectively.

Definition 5. Given two fuzzy formal concepts (X_1,Y_1) and (X_2,Y_2) of a fuzzy formal context (B,A,P), (X_1,Y_1) is said to be **superconcept** of (X_2,Y_2) and (X_2,Y_2) is said to be **subconcept** of (X_1,Y_1) if $X_2 \subseteq X_1$ or $Y_1 \subseteq Y_2$ holds. The relationship is represented as $(X_2,Y_2) \leq (X_1,Y_1)$.

Definition 6. Given a fuzzy formal context (B,A,P), consider the set of all fuzzy formal concepts of this context, indicated as $\wp(B,A,P)$. Then $(\wp(B,A,P),\leq)$ is a complete lattice called **fuzzy concept lattice**.

Table 1. The Chinese Cities Context

	Cir	***Eco***	***Tra***	***Sus***	***Sci***
Bj	0	0.6	0.9	0.3	0.2
Sh	0.4	0.2	0	0.7	0
Cq	0.6	0.8	0.9	0.3	0.2
Cd	0	0.1	0.2	0.6	0.9
Km	0.2	0.9	0.7	0.8	0

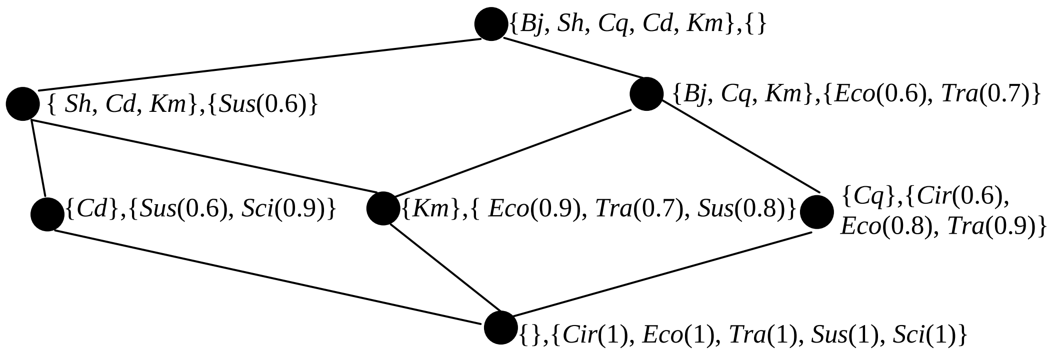

Figure 1. The Fuzzy Formal Concept Lattice from Fuzzy Formal Context

3.3. Mapping between FO and FCL

As it is mentioned in Section 2.1, fuzzy ontology is high similar to formal concept lattice in structure. And fuzzy formal concept analysis is applied to fuzzy ontology construction. So there is an inherent mapping between fuzzy ontology and fuzzy concept lattice. The mapping is as follow.

- Classes are mapped to fuzzy formal concepts.
- Instances are mapped to objects.
- Properties are mapped to attributes. And values of properties are mapped to membership degrees of attributes.
- *synonym of* relationship between classes is mapped to inheritance between fuzzy formal concepts.

For instance, two fuzzy formal concepts ffo_1=({*Sh*, *Cd*, *Km*}, {*Sus*(0.6)}) and ffo_2= ({*Cd*},{*Sus*(0.6), *Sci*(0.9)}) are considered from Figure 1. It is easy to find that ffo_1 is superconcept of ffo_2. ffo_1 and ffo_2 are mapped to two classes cl_1 and cl_2, respectively. *Sus* is a property of cl_1 and *Sus* and *Sci* are two properties of cl_2. There is a *synonym of* relationship between ffo_1 and ffo_2.

4. Similarity Computing between Fuzzy Classes

4.1. Semantic Similarity between Two Words

Formula 1 is used to evaluate similarity *as*(*a*,*b*) between two words *a*, *b* in Wordnet[4]. From formula 1, it is noted that,

- The similarity between two words (*a*,*b*) is the function of their distance and the lowest common subsume *lso*(*a*,*b*).
- If the *lso*(*a*,*b*) is root, *depth*(*lso*(*a*,*b*))=1,*as*(*a*,*b*)>0; if the two words have the same sense, the word *a*, word *b* and *lso*(*a*,*b*) are the same node. *len*(*a*,*b*)=0. *as*(*a*,*b*)=1; otherwise 0<*depth*(*lso*(*a*,*b*))<*deep_max*, 0<*len*(*a*,*b*)< 2**deep_max*, 0<*as*(*a*,*b*)< 1. Thus, the values of *as* (*a*,*b*) are in (0, 1].

$$as(a,b) = \frac{2 * depth(lso(a,b))}{len(a,b) + 2 * depth(lso(a,b))} \qquad \text{Formula (1)}$$

4.2. Semantic Similarity between Two Fuzzy Formal Concepts

Definition 7: Consider two fuzzy formal concepts (X_1, Y_1) and (X_2, Y_2) from two different fuzzy concept lattices. Let n, m be the cardinalities of the sets Y_1, Y_2, respectively, *i.e.*, $n=|Y_1|$, $m=|Y_2|$, and suppose that $n\leq m$. **The set** $\tau(Y_1,Y_2)$ **of the candidate sets** of pairs is defined by all possible sets of n pairs of attributes defined as follow: $\tau(Y_1,Y_2)=\{\{<a_1,b_1>\ldots<a_n,b_n>\}|a_h\in Y_1, b_h\in Y_2, \forall h=1,\ldots,n$, and $a_h\neq a_k$, $b_h\neq b_l$, $l\neq h\}$.

Definition 8: Consider a domain ontology**, the concept similarity of two fuzzy formal concepts** (X_1, Y_1) and (X_2, Y_2) from two different fuzzy concept lattices is defined as follow [5]:

$$K\left((X_1,Y_1),(X_2,Y_2)\right)=\frac{1}{\max\left(|Y_1|,|Y_2|\right)}\max_{P\in\tau(Y_1,Y_2)}\left(\sum_{\langle a,b\rangle\in P} f_a * f_b * as(a,b)\right) \qquad \text{Formula (2)}$$

$$Sim\left((X_1,Y_1),(X_2,Y_2)\right)=\frac{|X_1\cap X_2|}{\max\left(|X_1|,|X_2|\right)}*w+K\left((X_1,Y_1),(X_2,Y_2)\right)*(1-w) \qquad \text{Formula (3)}$$

From formulas (2) and (3), it is noted that,

- f_a and f_b are membership degrees of attributes a and b, respectively.
- w is a weight such that $0\leq w\leq 1$, that can be established by the user to enrich the flexibility of the method.
- $as(a, b)$ can be evaluated from formula (1).

Given two fuzzy formal concepts $ffo_1=(\{Bj, Cq, Km\},\{Eco(0.6), Tra(0.7)\})$ and $ffo_2=(\{Cq\},\{Cir(0.6), Eco(0.8), Tra(0.9)\})$. $\tau(\{Eco(0.6), Tra(0.7)\},\{Cir(0.6), Eco(0.8), Tra(0.9)\})$ is evaluated to $\{\{(Eco(0.6), Eco(0.8))\}, \{((Tra(0.7), Tra(0.9))\}\}$. By formula (1), both $as(Tra, Tra)$ and $as(Eco, Eco)$ are evaluated to 1. So $sim(ffo_1, ffo_2)$ is evaluated as follow when w is assigned to 0.5.

$$K(ffo_1, ffo_2)=\frac{1}{3}(0.6*0.8*1+0.7*0.9*1)=0.37 \quad \text{by formula (2)}$$

$$Sim(ffo_1, ffo_2)=\frac{1}{3}*\frac{1}{2}+\frac{1}{3}(0.6*0.8*1+0.7*0.9*1)*\frac{1}{2}=0.35 \quad \text{by formula (3)}$$

5. Alignment Algorithm for Fuzzy Ontologies

Previous algorithms for ontologies alignment focus on only one pair of classes, which just reaches to local optimization. Unlike previous algorithms, we propose a global optimal alignment algorithm *SMDS*. It need be noted that in this paper, only classes mapping are considered.

In order to make algorithm 1 accessible, definitions 9 and 10 related to algorithm 1 are given.

Definition 9. Given two fuzzy ontologies O_{F1} and O_{F2}, D is a **dry subset** of $C_1\times C_2$ iff For any two (a_1,b_1), $(a_2,b_2)\in D$, all of $a_1\neq a_2$, $b_1\neq b_2$ and $D\subseteq C_1\times C_2$ hold, where C_1 and C_2 are two sets of classes in O_{F1} and O_{F2}, respectively. M is said to be a **maximal dry subset** of $C_1\times C_2$ iff for any $S\subseteq C_1\times C_2$, $\sum_{j=1}^{|S|} Sim(s_j), s_j\in S \leq \sum_{i=1}^{|M|} Sim(m_i), m_i\in M$ holds, where both S and M are dry subset of $C_1\times C_2$.

Definition 10. For two fuzzy ontologies O_{F1} and O_{F2}, **fuzzy ontology alignment** is represented as a function fol:$(C_1 \times C_2) \rightarrow M$. C_1 and C_2 are two sets of classes in O_{F1} and O_{F2}, respectively. M is a maximal dry subset of $C_1 \times C_2$.

```
Algorithm 1:SMDS(C1, C2, t)
Input: C1, C2, a pair of sets of classes.
       t, a similarity threshold.
Output: M, a dry subset of I.
i←1;
I ←C1×C2;
FOR every (a,b)∈I
   IF Sim(a,b)<t        //Sim(a,b) is used to evaluate similarity between a and b
      Remove (a,b) from I;
M←D        //D is an arbitrary dry subset of I.
FOR every dry subset D' of I
   IF Σ_{j=1}^{|D'|} Sim(d'_j), d'_j ∈ D' > Σ_{i=1}^{|D|} Sim(d_i), d_i ∈ D
      M←D'
RETURN M
```

By definition 10, the task of fuzzy ontologies alignment is to decide function fol. Algorithm 1 is implementation of function fol. It is explained as follow.

Step 1: Decide initial set of matching pairs according to a predefined threshold, which means that pairs whose similarity are smaller than threshold are removed.

Step 2: Find a maximal dry subset M by definition 9.

For example, let C_1 be $\{c_{11},c_{12},c_{13}\}$ and C_2 be $\{c_{21},c_{22}\}$. So $C_1 \times C_2$ can be evaluated to $\{(c_{11},c_{21}), (c_{11},c_{22}), (c_{12},c_{21}), (c_{12},c_{22}), (c_{13},c_{21}), (c_{13},c_{22})\}$. We assume that $Sim(c_{11},c_{21})$=0.9, $Sim(c_{11},c_{22})$=0.85, $Sim(c_{12},c_{21})$=0.3, $Sim(c_{12},c_{22})$=0.7, $Sim(c_{13},c_{21})$=0.85 and $Sim(c_{31},c_{22})$=0.6. And (c_{12},c_{21}) and (c_{31},c_{22}) are removed from $C_1 \times C_2$ because $Sim(c_{12},c_{21})$=0.3and $Sim(c_{31},c_{22})$=0.6 are smaller than 0.7, a predefined threshold. So the revised $C_1 \times C_2$ is $\{(c_{11},c_{21}), (c_{11},c_{22}), (c_{12},c_{22}), (c_{13},c_{21})\}$. All dry subsets of revised $C_1 \times C_2$ are $\{(c_{11},c_{21}), (c_{12},c_{22})\}$ and $\{(c_{11},c_{22}) (c_{13},c_{21})\}$. Because $Sim(c_{11},c_{22})+Sim(c_{13},c_{21})$ is larger than $Sim(c_{11},c_{21})+Sim(c_{12},c_{22})$, $\{(c_{11},c_{22}) (c_{13},c_{21})\}$ is the maximal dry subset.

6. Implementation of Prototype and Experiment

6.1. Implementation of Prototype for our Approach

We implemented a prototype based on the approach described above. The prototype is for OWL Ontology in JAVA. The architecture of prototype is shown in Figure 2. The whole system is composed of six components and WordNet Interface. Application GUI is an interface for users. By Application GUI, users submit two fuzzy OWL ontologies to Ontology Parser. Also Application GUI is responsible for screening the result of ontologies alignment. Ontology Parser is responsible for extracting all classes, properties and their values. Words Similarity Component is implementation of formula (1). And with WordNet Interface, semantic similarity between two words may be evaluated by Words Similarity Component. Concepts Similarity Component is implementation of formula (3) and responsible for evaluating two classes. Ontology aligment Component calls the Similarity Evaluation Component to align ontologies.

6.2 Experiments with Prototype

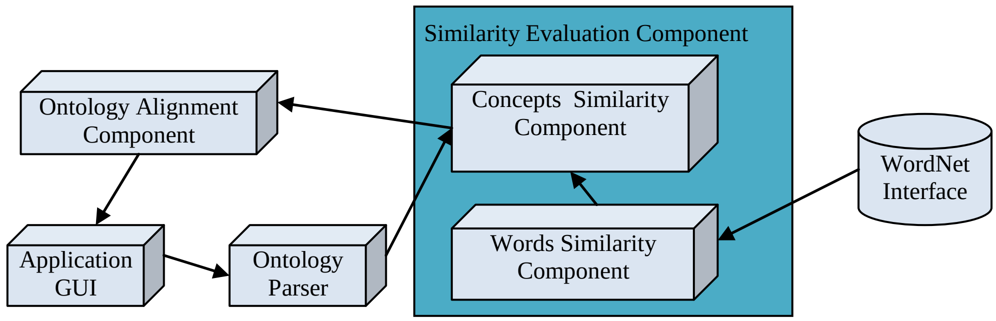

Figure 2. Architecture of Prototype

We have tested the current implementation of the prototype system on two pairs of fuzzy ontologies from Chinese medical domain and sea legal domain. Every pair of fuzzy ontologies are developed by two independent groups. The first Chinese medical ontology CMO_1 ontology contains 78 classes and the second Chinese medical ontology CMO_2 contains 91 classes. The first sea legal ontology SLO_1 contains 51 classes and the second sea legal ontology SLO_2 contains 55 classes.

$$pr = \frac{|M_{right}|}{|M|} \quad \text{formula (4)} \qquad cr = \frac{|M_{right}|}{|R|} \quad \text{formula (5)}$$

Two indexes *pr*, precision ratio and *cr*, recall ratio are introduced to evaluate our approach. From formulas (4) and (5), *M* is the maximal dry subset evaluated by our approach. M_{right} is a subset of *M* and consist of right mapping pairs of classes, which is computed by our approach. *R* is a set consist of all right mapping pairs of classes, which is done by human. In the following, we briefly report experiments performed for two pairs of fuzzy ontology alignments. From Table 2, it need be noted that *Th*, Mr_C, M_C, pr_C, cr_C, Mr_S, M_S, pr_S, cr_S represent threshold, cardinality of M_{right}, cardinality of *M*, *pr* and *cr* about *CMO*, cardinality of M_{right}, cardinality of *M*, *pr* and *cr* about *SLO*. For *CMO*, the *R* is 45. For *SLO*, *R* is 38.

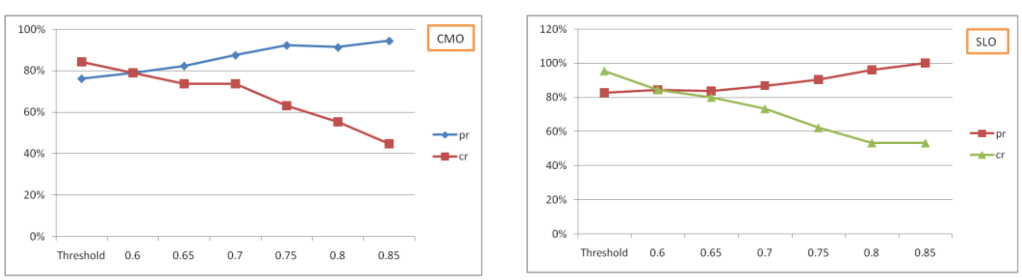

Figure 3. Trend Graphs of Precision Ratio and Recall Ratio about *CMO* and *SLO*

Table 2. The Experiment Result about CMO Alignment and SLO Alignment

Th	Mr_C	M_C	pr_C	cr_c	Mr_S	M_S	pr_S	cr_S
0.6	43	52	83%	96%	32	42	76%	84%
0.65	38	45	84%	84%	30	38	79%	79%
0.7	36	43	84%	80%	28	34	82%	74%
0.75	33	38	87%	73%	28	32	88%	74%
0.8	28	31	90%	62%	24	26	92%	63%
0.85	24	25	96%	53%	21	23	91%	55%
0.9	24	24	100%	53%	17	18	94%	45%

From Figure 3, a uniform thing is that the precision ratio is proportional to threshold and the precision ratio is inversely proportional to threshold. By analyzing our approach, it is inevitable that as threshold becomes larger, more pairs of classes will be removed from M and naturally cardinality of M_{right} will become smaller. But decrease extent of M_{right} is lower than decrease extent of M, which can explain why precision ratio increases. In addition, because R is invariable, recall ratio decrease as threshold becomes larger.

7. Conclusion and Future Work

In this paper, we propose an effective approach to fuzzy ontologies alignment. We fully consider the membership degree in formula for computing similarity between classes from two different fuzzy ontologies. Also, we consider global optimization in alignment algorithm for matching classes.

It is easy to see that the quality of alignment severely depends on formula for computing similarity. So in the future work, we shall go on doing experiments in our prototype and improve the formula. Also, we shall introduce restrain conditions to improve alignment algorithm.

Acknowledgements

This work was supported by The Open Project Program of Artificial Intelligence Key Laboratory of Sichuan Province (2013RZJ02).

References

[1] J. Jelsteen, D. Evangelin, J. P. Alice and S. J. J. Nelson, “Ontology Learning Process Using Fuzzy Formal Concept Analysis”, International Journal of Engineering Trends and Technology, vol. 4, no. 2, **(2013)**.
[2] W. Zhou, Z. Liu and Y. Zhao, “Ontology Learning by Clustering Based on Fuzzy Formal Concept Analysis”, Proceedings of Computer Software and Applications Conference, Beijing, China, **(2007)** July 23-27.
[3] F. Zhang, Z. M. Ma, J. Cheng and X. Meng, “Fuzzy semantic web ontology learning from fuzzy UML model”, Proceedings of the 18th ACM conference on Information and knowledge management, Hong Kong, China, **(2009)** November 2-6.
[4] L. Meng, R. Huang and J. Gu, “A Review of Semantic Similarity Measures in WordNet”, International Journal of Hybrid Information Technology, vol. 6, no. 1, **(2013)**.
[5] A. Formica, “Ontology-based concept similarity in Formal Concept Analysis”, Information Sciences, vol. 176, no. 18, **(2006)**.
[6] H. H. Do and E. Rahm, “COMA: A system for flexible combination of schema matching approaches”, Proceedings of the 28th international conference on Very Large Data Bases, Hong Kong, China, **(2002)** August 20-23.
[7] F. Giunchiglia, P. Shvaiko and M. Yatskevich, “S-match: An algorithm and an implementation of semantic matching”, Proceedings of the 1st European Semantic Web Symposium, Heraklion, Greece, **(2004)** May 10-12.
[8] J. Madhavan, P. Bernstein and E. Rahm, “Generic schema matching with Cupid”, Proceedings of the 27th International Conference on Very Large Data Bases, Rome, Italy, **(2001)** September 11-14.

[9] M. Ehrig and S. Staab, “QOM—Quick Ontology Mapping”, Proceedings of the 3rd International Semantic Web Conference, Hiroshima, Hiroshima, Japan, **(2004)** November 7-11.
[10] H. Ghorbel, A. Bahri and R. Bouaziz, “Fuzzy Ontologies Building Method: Fuzzy OntoMethodology”, Proceedings of Fuzzy Information Processing Society, New York, American, **(2010)** July 12-14.
[11] J. Li, J. Tang, Y. Li and Q. Luo, “RiMOM: A Dynamic Multistrategy OntologyAlignment Framework”, IEEE Transactions On Knowledge And Data Engineering, vol. 21, no. 8, **(2009)**.

Authors

Ling Qiu, lecturer, received the master degree from University of Electronic Science and Technology of China in 2010. The main research directions: Computer Application, Artificial Intelligence.

Yong Liu, lecturer, received the bachelor degree from in Sichuan University of Science & Engineering in 2003. The main research directions: Semantic Web, Ontology, Artificial Intelligence.

