ICIC Express Letters
Part B: Applications
Volume **8**, Number **3**, March **2017**

ICIC International ©2017 ISSN 2185-2766
pp. 655–663

MFSMA: MULTI-FEATURE SIMILARITY MEASURING ALGORITHM FOR SEMANTIC ANNOTATION OF WSDL DOCUMENTS

Wei Lu[1], Yong Yang[1,*], Weiwei Xing[1], Xiaoping Che[1]
Liqiang Wang[2] and Yuanyuan Cai[1]

[1]School of Software Engineering
Beijing Jiaotong University
No. 3, Shangyuancun, Haidian District, Beijing 100044, P. R. China
{ luwei; wwxing; xpche; 12112089 }@bjtu.edu.cn; *Corresponding author: 12112088@bjtu.edu.cn

[2]Department of Computer Science
University of Central Florida
Orlando, FL 32816, USA
lwang@cs.ucf.edu

Received August 2016; accepted November 2016

Abstract. *Semantic annotation of WSDL (Web Services Description Language) document is an efficient, convenient, and practical method to implement Semantic Web Services. Semantic similarity is the backbone of semantic annotation. There are some limitations of previous semantic similarity measuring approaches. Most of them focused on measuring semantic similarity between concepts in a specific domain ontology. However, terms used in Web Services often are from multiple domains with different knowledge sources, which makes traditional approaches not applicable. In addition, previous works provide low discrimination due to incomplete utilizing of the knowledge resources. To address these, we propose MFSMA (multi-feature similarity measuring algorithm) that consists of two parts as structural similarity and lexical similarity to measure semantic similarity. Our method combines three common used approaches (Edge-based, Feature-based, and Information Content-based) with mapping them to three proposed features (depth, width, and density) in structural representation. Finally, we implement a comparison experiment, and results show that our approach provides better discrimination among different experiment sets. Theoretically, proposed approach can be applied in semantic annotation of any type user defined Web Services description documents.*

Keywords: Semantic similarity, Semantic annotation, Web Services, Semantic Web Services, Ontology

1. **Introduction.** Web Services provide a standardized way to achieve inter-operability between heterogeneous software systems [1]. However, lack of semantics makes it complex to accurately discover and compose Web Services [2]. Semantic Web Services is proposed for adding semantics into traditional Web Services [3, 4, 5]. One implementation of Semantic Web Services is semantic annotation of traditional Web Services by tagging a term in a WSDL document with a concept in a domain ontology, because, ontologies (that formally represent knowledge as a set of concepts and relations between concepts) can provide a definitive and exhaustive classification of entities in all spheres of being. Semantic similarity, which reflects how closely associated concept pairs are, is the backbone of semantic annotation. According to the result of semantic similarity measuring one can select the most appropriate ontology concept to annotate target term in WSDL document.

Previous semantic similarity measuring approaches have limitations when terms are described by different languages (such as, XML, RDF, OIL, or OWL). Furthermore, those approaches are inflexible because most of them are proposed to measure semantic similarity in a single ontology. In addition, the results are not obvious discrimination of different concept pairs that represent similar objects. The reason is that single feature of the object information is considered while the discrimination is hidden in the not considered features.

In this paper, we propose a novel semantic annotation approach with implementing a multi-feature similarity measuring algorithm named MFSMA to compute semantic similarity between concept pairs in WSDL and ontology (written in OWL) documents. We base on an assumption that semantic similarity between terms is heavily influenced by the related terms in the same knowledge source (description document).

Contributions of the proposed algorithm are summarized as follows:

- A new approach is proposed to support calculating semantic similarity between concepts (or terms) in documents that are written in different description languages;
- Relationship that is reflected by internal features (depth, width, and density) among nodes in the transformed tree structure is fully utilized;
- The discrimination is amplified that provides better basis of annotation.

The remainder of this paper is organized as follows. Section 2 reviews some previous work. Section 3 describes the proposed approach. Section 4 gives the experiment results. Sections 5 presents conclusion and future work.

2. **Related Work.** In most of previous work, semantic similarity is measured between concept pairs in single ontology. These work can be roughly classified as: Edge-based [6, 7, 8, 9], Feature-based [10, 11, 12] and Information Content-based [13, 14, 15, 16, 17].

In the Edge-based approaches, an ontology is modeled as a directed graph in which semantic distance represents the semantic similarity between two concepts in the ontology. The semantic distance is the number of the least links (also called shortest path) separating the concept pairs. In the feature-based approaches, overlapping and non-overlapping feature sets of one concept are utilized when measuring semantic similarity. Information Content-based approaches complement the above two methods by considering the quantification of semantic information (Information Content, IC) that concepts have in common. Single ontology corpora maybe lead to textual ambiguity and data sparseness which makes the result inaccurate.

With the widespread adoption of the Semantic Web paradigms, many ontologies have been developed in the past few decades for various purposes and domains [11, 18, 19, 20, 21, 22, 23]. Edge-based extension approach connects two ontologies by a bridge, and then, uses extension Edge-based measure to compute semantic similarity [22]. It classifies ontologies into one primary ontology and secondary ontologies. The limitation is that a primary ontology must be selected first, and the authors assume that the primary ontology selected will always provide better result which is not the truth. Feature-based extension approaches rely on matching synsets and term description sets, such as [11]. The term description sets are words extracted by parsing term definitions. This kind of methods usually considers ontology logical knowledge and losses taxonomical information (e.g., meronym, glosses, related concepts, etc.). Information Content-based extension approaches rely on information theory that utilized notion of mutual information. These extension methods can obtain more accurate results; however, they are relying heavily on well-defined ontologies.

Semantic similarity measuring in different knowledge sources concentrates on items from user-defined documents [26, 27, 28, 29].

Patil et al. proposed a framework for semi-automatically marking up Web Services description with ontologies called MWSAF in [26]. They used a combination of lexical and structural similarity measures as Equation (1).

$$MS = \frac{w_1 * ElemMatch + w_2 * SchemaMatch}{w_1 + w_2} \tag{1}$$

where *ElemMatch* and *SchemaMatch* represent lexical similarity and structural similarity of two concepts, respectively. Especially, *SchemaMatch* considered the similarity of sub-concepts and the ratio of matched sub-concepts. However, the organization structure of elements was not fully utilized both in WSDL and OWL document that would decrease the accuracy of the semantic similarity measurement.

The authors proposed a lexicon-based alignment semantic annotation approach in [27]. They generated a synonym of a concept according to WordNet. Then, a 2D matrix that holds the synonyms of the word for each sense in one dimension, and derivation hierarchies of the senses in other dimension was obtained by the synonyms. In the lexicon-based alignment, they perform matching over level-sense synsets by using name equality between all elements in the generated synonyms. A table, in which each cell is a tetrad containing name equality concept pairs and their levels, will be obtained. At last, the semantic similarity of synonyms is calculated by Equation (2).

$$md\left(c_{a_{sense_i}}, c_{b_{sense_j}}\right) = \left(\frac{2 * d_{nl}}{d_{sl^1} + d_{sl^2}}\right)^2 \tag{2}$$

where d_{nl} denotes the derivation order of common node, and d_{sl^1} and d_{sl^2} denote the derivation order of the first and second sense leaves, respectively.

In [29,30], the authors proposed a semi-automatic WSDL Web Services description documents. Firstly, they classified WSDL services description (which is broken down into XSD data types, interfaces, operations and messages) to its corresponding domain. And then, similarity between a WSDL element and the concepts of the selected domain ontology will be computed to identify which ontology concept to annotate the WSDL element. The semantic similarity measure did not detail in the paper.

The limitations of all the above approaches are not fully utilizing information in both WSDL document and domain ontology. For example, [26] does not consider the importance of each concept in both WSDL and ontology, and, the authors do not consider the sub-concepts of compared concept pairs in [27].

3. **Proposed Approach.** Due to different representations of WSDL and OWL, direct semantic matching between items in WSDL and OWL documents is difficult [30]. A good solution is expressing knowledge in original documents as trees like [26, 31]. The first step of our approach is mapping items in WSDL and OWL documents to intermediate tree structures according to rules similar as [26, 31].

Similar as [26], SSD (abbreviation of Semantic Similarity Degree) in the proposed approach as Equation (3) consists of two parts: structural and lexicon similarity.

$$SSD\big(W_i, O_j\big) = w_l * S_l\left(W_i, O_j\right) + (1 - w_l) * S_s\big(W_i, O_j\big) \tag{3}$$

where $SSD(W_i, O_j)$ is SSD between term W_i and O_j that are the name of node in WT (tree structure of WSDL documents) and OT (tree structure of OWL files), respectively, and $W_i \in W = \{W_1, W_2, \ldots, W_n\}$ and $O_j \in O = \{O_1, O_2, \ldots, O_m\}$. w_l is a weight.

3.1. **Lexicon similarity measuring.** Lexicon similarity is the measurement of linguistic similarity between WSDL element and ontology concept. We use Levenstein Distance [32] and Abbreviation Expansion to calculate lexicon similarity based on the assumption that the string used for naming elements in WSDL or concepts in OWL ontologies is single word or words connected with special character, i.e., space, capital letter, etc. The lexicon similarity $S_l\left(W_i, O_j\right)$ is calculated as Equation (4):

$$S_l(W_i, O_j) = \begin{cases} LD_{sim}(W_i, O_j), & \text{if } AE_{sim}(W_i, O_j) = 0 \\ AE_{sim}(W_i, O_j), & \text{else} \end{cases} \tag{4}$$

$$LD_{sim}(W_i, O_j) = 1 - \frac{ld(W_i, O_j)}{MaxLength(W_i, O_j)} \tag{5}$$

$LD_{sim}(W_i, O_j)$ expresses the lexicon similarity based on Levenstein Distance. $ld(W_i, O_j)$ denotes the Levenstein Distance between W_i and O_j by adding a virtual root node above both WT and OT, and $MaxLength(W_i, O_j)$ means the larger string length of the two concepts.

$$AE_{sim}(W_i, O_j) = \begin{cases} 0, & \text{if no abbreviation is between } W_i \text{ and } O_j \\ 1, & \text{if abbreviation is between } W_i \text{ and } O_j \end{cases} \tag{6}$$

is the Abbreviation Expansion of W_i and O_j.

3.2. **Structure-level similarity measuring.** $S_s(W_i, O_j)$ denotes the structural similarity between W_i and O_j. $S_s(W_i, O_j)$ is calculated by Equation (7):

$$S_s(W_i, O_j) = \frac{F(W_i, O_j)}{NumOf\left(S_l\left(W_k', O_l'\right)\right)} \tag{7}$$

where $W_k' \in W' = \left\{W_i' \textit{ is the relevant nodes of } W_i\right\}$ and $W' \subset W$, and O_j is the same; $NumOf\left(S_l\left(w_k', O_l'\right)\right)$ is the number of relevant nodes that satisfies $S_l\left(w_k', O_l'\right) > 0$.

$$F\left(W', O'\right) = Max\left\{\sum_{i=1}^{NumOf\left(S_l\left(W_k', O_l'\right)>0\right)} w\left(W_k', O_l'\right) * S_l\left(W_k', O_l'\right)\right\} \tag{8}$$

is a function to select the maximum sum of lexical similarity $S_l\left(W_k', O_l'\right)$ with weight $w\left(W_k', O_l'\right)$ (a weight to reflect the influence of the relevant nodes organization structure in the tree structure). $w\left(W_k', O_l'\right)$ is composed of three parts as Equation (9).

$$w\left(W_k', O_l'\right) = w_d\left(W_k', O_l'\right) * w_w\left(W_k', O_l'\right) * w_\rho\left(W_k', O_l'\right) \tag{9}$$

In Equation (9), $w_d\left(W_k', O_l'\right)$ is a weight which reveals the contribution of node's depth in the tree structure to the structural similarity. Here, depth is the distance between a node and the virtual root node. The core principle is that the larger depth difference is, the smaller $w_d\left(W_k', O_l'\right)$ is. Based on this principle, we proposed Equation (10).

$$w_d\left(W_k', O_l'\right) = f_d\left(f_\Delta(\Delta Dep), f_\Sigma(\Sigma Dep)\right) \tag{10}$$

where ΔDep and ΣDep represent sum and difference between W_k' and O_l', respectively.

We set $f_\Delta(\Delta Dep)$ and $f_\Sigma(\Sigma Dep)$ to be monotonically decreasing and monotonically increasing as Formula (11) and Formula (12), respectively:

$$f_\Delta(\Delta Dep) = e^{-\alpha * |\Delta Dep|} \tag{11}$$

$$f_\Sigma(\Sigma Dep) = 1 - e^{-\beta * \Sigma Dep} \tag{12}$$

where $w_d\left(W_k', O_l'\right)$ is considered to be governed by ΔDep and ΣDep, as Equation (13):

$$w_d\left(W_k', O_l'\right) = \gamma * e^{-\alpha * |\Delta Dep|} * \left(1 - e^{-\beta * \Sigma Dep}\right) \tag{13}$$

where γ is an adjustment factor to control the value of $w_d\left(W_k', O_l'\right)$. Experimental values are $\gamma = 1$, $\alpha = 0.3$, and $\beta = 1$.

$w_w\left(W_k', O_l'\right)$ also is a weight which reveals the contribution of node's width similar as $w_d\left(W_k', O_l'\right)$ defined as Equation (14). Width is the number of sibling nodes. It is noted that the value of $w_w\left(W_k', O_l'\right)$ mainly depends on the sibling nodes of W_k' and O_l'.

$$w_w\left(W_k', O_l'\right) = \frac{Max\left\{\sum_{i=1}^{NumOf\left(S_l\left(sib_f^{W_k'}, sib_g^{O_l'}\right)>0\right)} S_l\left(sib_f^{W_k'}, sib_g^{O_l'}\right)\right\}}{(1+\alpha) * NumOf\left(S_l\left(sib_f^{W_k'}, sib_g^{O_l'}\right) > 0\right) + \alpha * |m - n|} \tag{14}$$

where $\alpha = 0.5$, $sib^{W'_k} = \left\{ sib_m^{W'_k} \textit{ is the sibling node of } sib^{W'_k} \right\}$ and $sib^{O'_l}$ are the same as $sib^{W'_k}$, m and n are the size of $sib^{W'_k}$ and $sib^{O'_l}$, respectively, and $NumOf\left(S_l\left(sib_f^{W'_k}, sib_g^{O'_l}\right) > 0\right)$ is the number of sibling node pairs that have a lexical similarity bigger than 0.

$w_\rho\left(W'_k, O'_l\right)$ also is a weight similar as $w_d\left(W'_k, O'_l\right)$ and $w_w\left(W'_k, O'_l\right)$, which is associated with the node's density (occurrence probability of a node) defined as Equation (15).

$$w_\rho\left(W'_k, O'_l\right) = \begin{cases} 1, & \text{if } \rho\left(W'_k\right) = \rho\left(O'_l\right); \\ f_\rho\left(\rho\left(W'_k\right), \rho\left(O'_l\right)\right), & \text{if } \rho\left(W'_k\right) \neq \rho\left(O'_l\right); \end{cases} \tag{15}$$

and

$$f_\rho\left(\rho\left(W'_k\right), \rho\left(O'_l\right)\right) = \log_{\frac{1}{N_w}}^{\rho\left(W'_k\right)} * \log_{\frac{1}{N_o}}^{\rho\left(O'_l\right)} \tag{16}$$

where N_w is the total number of the sub-nodes in WT, and N_o is the total number of the sub-nodes in OT. $\rho\left(W'_k\right)$ and $\rho\left(O'_l\right)$ are the density of W'_k and O'_l, respectively.

3.3. **Formal description of the proposed algorithm.** Algorithm 1 is formal description of our approach to calculate semantic similarity between concept pairs $\left(W'_i, O'_j\right)$.

4. **Experiments Evaluation.** Most of WSDL documents and OWL files used in our experiments are from Meteors project in University of Georgia[1]. Table 1 in Appendix describes the notations that will be used in the following sections and their descriptions.

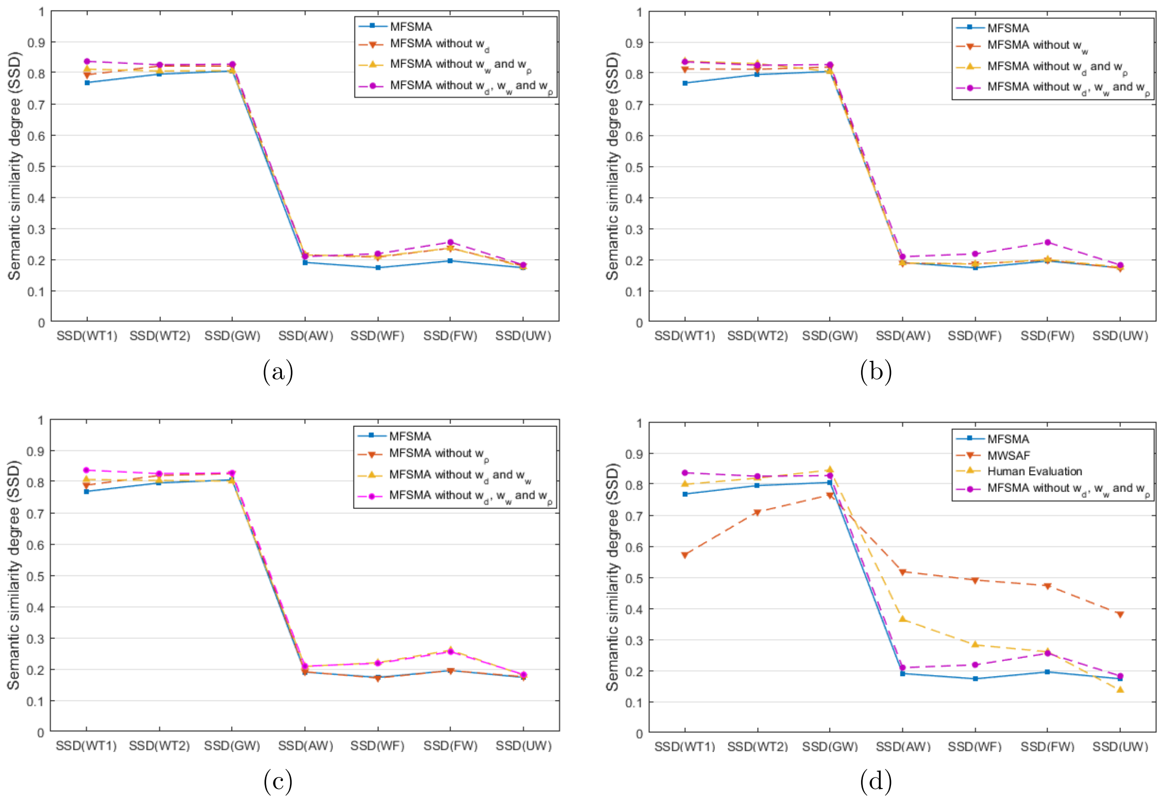

FIGURE 1. Experiment results

[1]Most of the WSDL documents and the OWL file come from Meteors project at *http://lsdis.cs.uga.edu/projects/meteors/downloads/*. *"WT1.wsdl"* and *"WT2.wsdl"* documents are modified by adding and deleting 1 element from *"Global-Weather.wsdl"*, respectively, to change the structure of tree structure (actually, depth, width, and density of node in the structure will be changed).

Algorithm 1 Calculating semantic similarity degree between the concept pairs $\left(W_i^{'}, O_j^{'}\right)$.

Require: $W^{'} = \left\{W_1^{'}, W_2^{'}, \ldots, W_p^{'}\right\}$, $O^{'} = \left\{O_1^{'}, O_2^{'}, \ldots, O_q^{'}\right\}$, $S_s[p][q]$ /*p and q are the size of $W^{'}$ and $O^{'}$ respectively, and each term in the matrix S_s is a structure type with 3 member variables W_{name}, O_{name}, and $S_s(W_i, O_j)$. $W_i^{'}$ and $O_j^{'}$ are quaternary tuple both containing 4 member variables: name, depth, width, and density*/
Ensure: $S_s(W_i, O_j)$ /*the result of structural similarity between W_i and O_j*/

for $i = 1 \rightarrow p$ **do** /*for each element in $W^{'}$*/
  for $j = 1 \rightarrow q$ **do**/*for each element in $O^{'}$*/
    $w_d\left(W_i^{'}, O_l^{'}\right) \leftarrow \gamma * e^{-\alpha * \left|Dep\left(W_i^{'}\right) - Dep\left(O_l^{'}\right)\right|} * \left(1 - e^{-\beta * \left(Dep(W_i^{'}) + Dep\left(O_j^{'}\right)\right)}\right)$; /*weight value of depth*/
    $w_w\left(W_i^{'}, O_l^{'}\right) \leftarrow \frac{Max\left\{\sum_{i=1}^{NumOf\left(S_l\left(Sib_f^{W_k^{'}}, Sib_g^{O_l^{'}}\right) > 0\right)} S_l\left(Sib_f^{W_k^{'}}, Sib_g^{O_l^{'}}\right)\right\}}{(1+\alpha) * NumOf\left(S_l\left(Sib_f^{W_k^{'}}, Sib_g^{O_l^{'}}\right) > 0\right) + \alpha * |m-n|}$; /*weight value of width*/
    if $\rho\left(W_k^{'}\right) == \rho\left(O_l^{'}\right)$ **then**
      $w_\rho\left(W_i^{'}, O_l^{'}\right) \leftarrow 1$;
    else
      $w_\rho\left(W_i^{'}, O_l^{'}\right) \leftarrow \log_{\frac{1}{N_w}}^{\rho\left(W_i^{'}\right)} * \log_{\frac{1}{N_o}}^{\rho\left(O_j^{'}\right)}$; /*weight value of density*/
    end if
    $w\left(W_i^{'}, O_j^{'}\right) \leftarrow w_d\left(W_i^{'}, O_j^{'}\right) * w_w\left(W_i^{'}, O_j^{'}\right) * w_\rho\left(W_i^{'}, O_j^{'}\right)$; /*combination weight value of $w\left(W_i^{'}, O_j^{'}\right)$*/
    $S_s[i][j].W_{name} \leftarrow W_i^{'}$;
    $S_s[i][j].O_{name} \leftarrow O_j^{'}$;
    $S_s[i][j].S_s\left(W_i^{'}, O_j^{'}\right) \leftarrow S_l\left(W_i^{'}, O_j^{'}\right) * w\left(W_i^{'}, O_j^{'}\right)$; /*structure similarity for concept pairs $\left(W_i^{'}, O_j^{'}\right)$*/
  end for
end for
$F\left(W_i^{'}, O_j^{'}\right) \leftarrow FindMaxOf(S_s[i][j])$; /*select the maximum value of $S_s[i][j]$ from all the calculated values*/
$S_s\left(W_i^{'}, O_j^{'}\right) \leftarrow \frac{2 * F\left(W_i^{'}, O_j^{'}\right)}{p+q}$; /*the final structure similarity of concept pair $\left(W_i^{'}, O_j^{'}\right)$*/

From Figures 1(a), 1(b) and 1(c), we can find that the internal features indeed influence the result of semantic similarity matching. In Figure 1(d), we compare the result of MWSAF [26], human evaluation result and our approach (includes 8 cases that are explained in the legend). From Figure 1(d), we can find that our approach provides better discrimination that human evaluation and MWSAF.

5. Conclusion.

In this paper, we propose a new semantic similarity degree matching algorithm called MFSMA which aims to provide better discrimination of semantic similarity matching for annotation decision making. Internal features including "depth", "width", and "density" are taken into account in our approach. Experimental results shows the proposed approach can provide better decision to determine which one of the OWL concepts can be used to annotate corresponding WSDL element, because MFSMA obtains semantic similarity degree with high discrimination with a width value range.

Acknowledgement. This work was supported in part by the National Natural Science Foundation of China (No. 61100143, No. 61272353, No. 61370128, No. 61428201, No. 61502028), Program for New Century Excellent Talents in University (NCET-13-0659), Beijing Higher Education Young Elite Teacher Project (YETP0583). Open Research Project (No. 2016303) funded by Key Laboratory of Geological Information Technology, Ministry of Land and Resources, Beijing, China.

REFERENCES

[1] O. Hatzi, D. Vrakas, M. Nikolaidou, N. Bassiliades, D. Anagnostopoulos and I. Vlahavas, An integrated approach to automated semantic web service composition through planning, *IEEE Trans. Services Computing*, vol.5, no.3, pp.319-332, 2012.
[2] D. Paulraj, S. Swamynathan and M. Madhaiyan, Process model ontology-based matchmaking of semantic web services, *International Journal of Cooperative Information Systems*, vol.20, no.4, pp.357-370, 2011.
[3] T. Berners-Lee, J. Hendler and O. Lassila, The semantic web, *Scientific American*, vol.284, no.5, pp.28-37, 2001.
[4] A. Martinez-Garcia, S. Morris, M. Tscholl, F. Tracy and P. Carmichael, Case-based learning, pedagogical innovation, and semantic web technologies, *IEEE Trans. Learning Technologies*, vol.5, no.2, pp.104-116, 2012.
[5] M. Sabou, C. Wroe, C. Goble and H. Stuckenschmidt, Learning domain ontologies for semantic web service descriptions, *Web Semantics: Science, Services and Agents on the World Wide Web*, vol.3, no.4, pp.340-365, 2005.
[6] R. Rada, H. Mili, E. Bicknell and M. Blettner, Development and application of a metric on semantic nets, *IEEE Trans. Systems, Man, and Cybernetics*, vol.19, no.1, pp.17-30, 1989.
[7] M. H. Margahny and A. Shakour, Fast algorithm for mining association rules, *Journal of Engineering Sciences*, vol.34, no.1, pp.79-87, 2006.
[8] C. Leacock and M. Chodorow, Combining local context and WordNet similarity for word sense identification, *WordNet: An Electronic Lexical Database*, vol.49, no.2, pp.265-283, 1998.
[9] Y. Li, Z. A. Bandar and D. McLean, An approach for measuring semantic similarity between words using multiple information sources, *IEEE Trans. Knowledge and Data Engineering*, vol.15, no.4, pp.871-882, 2003.
[10] A. Tversky, Features of similarity, *Psychological Review*, vol.84, no.4, pp.327-352, 1977.
[11] E. G. Petrakis, G. Varelas, A. Hliaoutakis and P. Raftopoulou, X-similarity: Computing semantic similarity between concepts from different ontologies, *Journal of Digital Information Management*, vol.4, no.4, pp.233-237, 2006.
[12] T. Pedersen, S. V. Pakhomov, S. Patwardhan and C. G. Chute, Measures of semantic similarity and relatedness in the biomedical domain, *Journal of Biomedical Informatics*, vol.40, no.3, pp.288-299, 2007.
[13] P. Resnik, Using information content to evaluate semantic similarity in a taxonomy, *arXiv Preprint cmp-lg/9511007*, 1995.
[14] J. J. Jiang and D. W. Conrath, Semantic similarity based on corpus statistics and lexical taxonomy, *arXiv Preprint cmp-lg/9709008*, 1997.
[15] D. Lin, An information-theoretic definition of similarity, *Internal Conference on Machine Learning*, pp.296-304, 1998.
[16] K. Ishida and T. Ohta, An approach for organizing knowledge according to terminology and representing it visually, *IEEE Trans. Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol.32, no.4, pp.366-373, 2002.
[17] D. Sánchez, M. Batet and D. Isern, Ontology-based information content computation, *Knowledge-Based Systems*, vol.24, no.2, pp.297-303, 2011.
[18] D. Sánchez and M. Batet, A new model to compute the information content of concepts from taxonomic knowledge, *International Journal on Semantic Web and Information Systems*, vol.8, no.2, pp.34-50, 2012.
[19] R. L. Cilibrasi and P. M. Vitanyi, The google similarity distance, *IEEE Trans. Knowledge and Data Engineering*, vol.19, no.3, pp.370-383, 2007.
[20] N. Seco, T. Veale and J. Hayes, An intrinsic information content metric for semantic similarity in WordNet, *European Conference on Artificial Intelligence*, vol.16, pp.1089-1090, 2004.
[21] Z. Zhou, Y. Wang and J. Gu, A new model of information content for semantic similarity in WordNet, *The 2nd International Conference on Future Generation Communication and Networking Symposia*, vol.3, pp.85-89, 2008.

[22] H. Al-Mubaid and H. A. Nguyen, Measuring semantic similarity between biomedical concepts within multiple ontologies, *IEEE Trans. Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol.39, no.4, pp.389-398, 2009.
[23] M. Batet, S. Harispe, S. Ranwez, D. Sánchez and V. Ranwez, An information theoretic approach to improve semantic similarity assessments across multiple ontologies, *Information Sciences*, vol.283, pp.197-210, 2014.
[24] A. Rodriguez and M. J. Egenhofer, Determining semantic similarity among entity classes from different ontologies, *IEEE Trans. Knowledge and Data Engineering*, vol.15, no.2, pp.442-456, 2003.
[25] A. Kilgarriff and C. Fellbaum, *WordNet: An Electronic Lexical Database*, 1998.
[26] A. A. Patil, S. A. Oundhakar, A. P. Sheth and K. Verma, METEOR-S web service annotation framework, *Proc. of the 13th International Conference on World Wide Web*, pp.553-562, 2004.
[27] D. Canturk and P. Senkul, Semantic annotation of web services with lexicon-based alignment, *IEEE World Congress on Services*, pp.355-362, 2011.
[28] D. Bouchiha and M. Malki, Semantic annotation of web services, *International Conference on Web and Information Technologies*, pp.60-69, 2012.
[29] D. Bouchiha, M. Malki, D. Djaa, A. Alghamdi and K. Alnafjan, Empirical study for semantic annotation of web services, *International Journal of Networked and Distributed Computing*, vol.2, no.1, pp.35-44, 2014.
[30] M. Klein, D. Fensel, F. Van Harmelen and I. Horrocks, The relation between ontologies and XML schemas, *Electronic Trans. Artificial Intelligence*, pp.128-145, 2001.
[31] B. Xu, J. Li and K. Wang, Web service semantic annotation, *Journal of Tsinghua University*, vol.46, no.10, 2006.
[32] M. Grcar and D. Mladenic, Visual OntoBridge: Semi-automatic semantic annotation software, *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, pp.726-729, 2009.
[33] D. L. Zhang and T. Lv, Concept frequency-based web service annotation, *Journal of Tongji University (Natural Science)*, vol.1, pp.23, 2008.

Appendix. Table 1 illustrates notations used in the experiment and their descriptions, and Table 2 is results of small-scale *"HumanEvaluation"* investigation. 9 persons (3 Ph.D. candidate and 6 graduate students) participated in the evaluation with 8 of them are majoring in software engineering and one majoring in image processing.

Table 1. Notations and their descriptions

Notation	Description
$SSD(WT1)$	Semantic similarity degree between "*WeatherReport*" in *WT1.wsdl* and "WeatherReport" in *WeatherConcept.owl*;
$SSD(WT2)$	Semantic similarity degree between "*WeatherReport*" in *WT2.wsdl* and "WeatherReport" in *WeatherConcept.owl*;
$SSD(GW)$	Semantic similarity degree between "*WeatherReport*" in *GlobalWeather.wsdl* and "WeatherReport" in *WeatherConcept.owl*;
$SSD(AW)$	Semantic similarity degree between "*WeatherSummary*" in *AirportWeather.wsdl* and "WeatherReport" in *WeatherConcept.owl*;
$SSD(WF)$	Semantic similarity degree between "*Weather*" in *WeatherFetcher.wsdl* and "WeatherReport" in *WeatherConcept.owl*;
$SSD(FW)$	Semantic similarity degree between "*Weather*" in *FastWeather.wsdl* and "WeatherReport" in *WeatherConcept.owl*;
$SSD(UW)$	Semantic similarity degree between "*GetWeatherResult*" in *UnisysWeather.wsdl* and "WeatherReport" in *WeatherConcept.owl.*

Table 2. Human evaluation results of concept pairs described in Table 1

SSD(WT1)	**SSD(WT2)**	**SSD(GW)**	**SSD(AW)**	**SSD(WF)**	**SSD(FW)**	**SSD(UW)**
0.80	0.70	0.90	0.60	0.40	0.30	0.10
0.86	0.98	1.00	0.26	0.29	0.42	0.30
0.70	0.60	0.80	0.30	0.30	0.20	0.10
0.72	0.74	0.71	0.06	0.09	0.06	0.00
0.90	0.85	0.80	0.40	0.40	0.30	0.10
0.90	0.90	0.90	0.40	0.20	0.20	0.10
0.95	0.97	0.90	0.50	0.30	0.20	0.05
0.90	0.88	0.80	0.70	0.50	0.50	0.40
0.42	0.77	0.88	0.08	0.03	0.03	0.00