http://ansinet.com/itj

ITJ

ISSN 1812-5638

INFORMATION TECHNOLOGY JOURNAL

ANSInet
Asian Network for Scientific Information
308 Lasani Town, Sargodha Road, Faisalabad - Pakistan

Information Technology Journal 12 (4): 857-861, 2013
ISSN 1812-5638 / DOI: 10.3923/itj.2013.857.861

Web Service Matching Based on Natural Semantic Annotation

Xia Jiang and Yingming Li
Laiwu Vocational and Technical College, Laiwu 271100, China

Abstract: To quickly and accurately find a needed Web service in service composition, a novel service matching method is proposed based on the nature semantic annotation in this paper. The natural semantic description is added in WSDL and the expanded WSDL is called as a WSDL-NS (WSDL with Natural Semantics). The computing method of semantic similarity is improved and the matching services are found by the service similarity. Finally, a simulation example is given to illustrate the correctness and effectiveness of the proposed method.

Key words: Web service, natural semantics, matching algorithm, similarity, Word Net

INTRODUCTION

Semantic similarity is very important for the natural language processing and retrieval and has been widely used in many domains, such as text block, image retrieval, automatic hyperlinks. The semantic similarity of natural words is usually obtained from the similarity of their concepts (Li *et al*., 2003). Many Web service discovery methods based on semantics are feasible and effective (Ankolekar *et al*., 2002; Wang and Stroulia, 2003), especially there is the semantic annotation method of Web services. From the point of view of practical application, the semantic information is enhanced in the services with semantics annotation (Oliva *et al*., 2011).

WSDL(Web Services Description Language) accords with the W3C standard and is used to describe the Web service interface information (Daniela *et al*., 2005). A WSDL file contains the following elements: Type, Message, Part, Operation, Port Type/Interface, Binding, Port/Endpoint and Service. WSDL has been widely used, but it is lack of semantic information. To deal with this problem, a WSDL-NS (WSDL with Natural Semantics), an expanded WSDL, is proposed based on the natural semantic annotation in this paper. Also, a computing method of semantic similarity is improved and the matching services are found by computing the service similarity. Finally, a simulation example is shown to illustrate the correctness and effectiveness of the proposed method.

A NEW NATURAL SEMANTIC SIMILARITY

Word Net (McHale, 1998) is a set of English vocabulary and a network is structured by it according to the semantics of words. A lot of semantic work is researched by taking it as a data base. The data in Word Net may be used as an objective reflection of a natural language, since the formation and evolution of a natural language are a long process.

The words in Word Net are organized as a tree. The semantic similarity between two concepts is affected from two major factors. One factor is the minimum distance between the concepts, represented by l. The other one is the depth of the minimum public concept, represented by h. Obviously, the minimum distance between the concepts is equal to a sum of their distance to the minimum public concept, respectively. Reference (Li *et al*., 2003) uses formula (1) to calculate the similarity between two concepts and it is optimal when $\alpha = 0.2$ and $\beta = 0.6$:

$$s(\omega 1, \omega 2) = e^{(-\alpha l)} \bullet (e^{\beta h}-e^{(-\beta h)})/(e^{\beta h}+e^{(-\beta h)}) \quad (1)$$

The similarity of some word pairs is shown in Table 1.

However, the similarity of concepts from the minimum distance needs to be improved. For example, for concepts C_1 and C_2, if the distances from them to the minimum public concept are l_1 and l_2, respectively, then

Table 1: Similarity calculated by formula (1)

Word pair	h	l	S $\alpha = 0.2$ $\beta = 0.6$
Noble metal-element	3	2	0.634
Water-earth	3	2	0.634
Crabapple- edible fruit	6	2	0.669
Apple-pear	6	2	0.669
Edible fruit-Baldwin	6	3	0.547
Apple-watermelon	6	3	0.547
Honeydew- edible fruit	6	4	0.448
Melon-Baldwin	6	4	0.448
Bilberry- watermelon	6	4	0.448
Child-terror	7	2	0.670
Buster-scamp	7	2	0.670

Corresponding Author: Xia Jiang, Laiwu Vocational and Technical College, Laiwu 271100, China

Table 2: Similarity calculated by formula (2)

Word pair	h	l	l_1	l_2	S $\alpha = 0.2$ $\beta = 0.6$	S' $\alpha = 0.2$ $\beta = 0.6$ $\gamma = 1$	S' $\alpha = 0.2$ $\beta = 0.6$ $\gamma = 0.5$	S' $\alpha = 0.2$ $\beta = 0.6$ $\gamma = 0$
Noble metal-element	3	2	0	2	0.634	0.574	0.603	0.634
Water-earth	3	2	1	1	0.634	0.701	0.667	0.634
Crabapple-edible fruit	6	2	0	2	0.669	0.605	0.636	0.669
Apple-pear	6	2	1	1	0.669	0.739	0.703	0.669
Edible fruit- Baldwin	6	3	0	3	0.547	0.495	0.521	0.547
Apple-watermelon	6	3	1	2	0.547	0.592	0.569	0.547
Honeydew-edible fruit	6	4	0	4	0.448	0.405	0.426	0.448
Melon-Baldwin	6	4	1	3	0.448	0.471	0.460	0.448
Bilberry-watermelon	6	4	2	2	0.448	0.495	0.471	0.448
Child-terror	7	2	0	2	0.670	0.606	0.637	0.670
Buster-scamp	7	2	1	1	0.670	0.740	0.704	0.670

their minimum distance is l_1+l_2. Although l_1 and l_2 can change, if l_1+l_2 is not changed, the similarity calculated by the method is equal. With the increase of the concept depth, the similarity is not changed linearly. Thus the formula (1) is not reflected from this kind of change. Therefore, this method is expanded as follows. The change of l_1 and l_2 is taken into consideration and a new formula (2) is obtained:

$$s'(\omega 1, \omega 2) = e^{(-\alpha l)} \cdot (e^{\beta h}-e^{(-\beta h)})/(e^{\beta h}+e^{(-\beta h)}) \quad (2)$$

where, $\ell = \gamma \cdot ((4l_1 \cdot l_2/(l_1+l_2)^2)-0.5)+l$, $0\le\gamma\le 1$, $l_1\ne 0$ and $l_2\ne 0$.

The distances from each word pair to the minimum public concept are shown in Table 2 and the new similarity calculated by formula (2) is given. γ is a discrete parameter. The formulae (1) and (2) are equivalent if $\gamma = 0$.

NATURAL SEMANTIC ANNOTATION OF WEB SERVICES

There are various kind of methods to add semantics to Web services. Some semantic service representation languages (Klusch *et al*., 2009) are proposed, such as OWL-S, WSMO and METEOR. However, these languages are represented based on some kind of semantic description framework of Web services (Tocha *et al*., 2011). They all more or less abandon intrinsic semantic information of natural language. However, natural semantic annotation of Web services is represented based on natural language and does not require users to learn some semantic description framework specially. Thus it can promote the application of semantic Web.

An annotated Web service can be represented as follows:

$$S = (L,s)$$

where, S is the Web service with natural semantic annotation. l is the annotation information of natural language in Web service. s is the static Web service without semanteme.

"verb + noun" mode is used to add natural semantics in static Web services, such as "search+ticket", "book+hotel". Therefore, natural language annotation information can be described formally as follows:

$$L = (n,v)$$

where, n represents noun and v represents verb. The similarity of any two annotation information $L_1(n_1,v_1)$ and L_2 (n_2,v_2) can be represented in the following:

$$sim = sim(L_1, L_2) = (sim(n), sim(v))$$

sim (n) = sim (n_1, n_2) represents the similarity between n_1 and n_2. sim (v) = sim (v_1, v_2) represents the similarity between v_1 and v_2.

when, n_1 and v_1 are the father concepts of n_2 and v_2, respectively, L_1 (n_1, v_1) is the father annotation of L_2 (n_2, v_2) and L_2 (n_1, v_1) is the son annotation of L_1 (n_2, v_2). For two similarities of annotation information, sim_1 = (sim (n_1), sim (v_1)) and sim_2 = (sim (n_2), sim (v_2)), $sim_1 = sim_2$ if and only if sim (n_1) = sim (n_2) and sim (v_1) = sim (v_2).

Some parameters of Web services can also be annotated, such as the pre-condition of invoking services (p), service input (i), service output (o) and service effect (e). The natural language annotation information L_s of Web services can be represented as follows:

$$L_s = (L_f, L_p, L_i, L_o, L_e)$$

where, L_f, L_p, L_i, L_o and L_e are the semantic annotation information of Web service function, precondition of service happen, service input, service output and service effect, respectively.

The semantic similarity of annotated Web services is defined in the following. Given two Web services s_1 and s_2, their description is L_{s1} and L_{s2} respectively. The semantic similarity of their functions, pre-condition of invoking services, service input, service output and service effect are $sim_f = sim(L_{f1}, L_{f2})$, $sim_p = sim(L_{p1}, L_{p2})$, $sim_i = sim(L_{i1}, L_{i2})$, $sim_o = sim(L_{o1}, L_{o2})$ and $sim_e = sim(L_{e1}, L_{e2})$, respectively. Thus, the similarity of the two Web services is defined as follows:

$$sim_{service} = sim(s_1, s_2) = sim(L_{s1}, L_{s2}) = (sim_f, sim_p, sim_i, sim_o, sim_e)$$

If L_{f1}, L_{p1}, L_{i1}, L_{o1} and L_{e1} are the father description of L_{f2}, L_{p2}, L_{i2}, L_{o2} and L_{e2}, respectively, then L_{s1} is the father description of L_{s2}, L_{s2} is the son description of L_{s1}, s_1 is the father service of s_2 and s_2 is the father service of s_1.

Given two similarities, $sim_{service} = (sim_f, sim_p, sim_i, sim_o, sim_e)$ and $sim'_{service} = (sim'_f, sim'_p, sim'_i, sim'_o, sim'_e)$, from the previous definition, have $sim_{service} \geq sim'_{service}$, if and only if $sim_f \geq sim'_f$, $sim_p \geq sim'_p$, $sim_i \geq sim'_I$, $sim_o \geq sim'_o$ and $sim_e \geq sim'_e$.

WSDL-NS AND MATCHING ALGORITHM

WSDL: WSDL plays a very important role in the practical application of Web services. It is always used as an important method of the data interaction between two heterogeneous systems. Its elements are used as a tag in real WSDL files. It mainly consists of the following elements:

- **Type:** Using data definition (string, int) of grammar (such as XML).
- **Message:** Data to transfer, such as input parameters, output parameters. The tag representation of Message elements is shown as follows:

```
<wsdl:message name="getSupport
ProvinceSoapIn">
<wsdl:part name="parameters" element
="tns:getSupportProvince" />
</wsdl:message>
<wsdl:message name="getSupport
ProvinceSoapOut">
<wsdl:part name="parameters" element
="tns:getSupportProvinceResponse" />
</wsdl:message>
```

- **Part:** Message parameters.
- **Operation:** The abstract description of the operations supported by services. The tag representation of Operation elements is given in the following:

```
<wsdl:operation
name="getSupportProvince">
<wsdl:documentation xmlns:wsdl="http://
schemas.xmlsoap.org/wsdl/">
The information of the state, provinces and cities
supported by the Web Service of weather forecast:
</wsdl:documentation>
<wsdl:input message="tns:getSupport Province
SoapIn" />
<wsdl:output message="tns:getSupportProvince
SoapOut" />
</wsdl:operation>
```

- **Port Type/Interface:** An abstraction set of operations supported by one or several endpoints. This name has been changed and it may meet one of them.
- **Binding:** Specific agreement and specification of Specific port type.
- **Port/Endpoint:** The combination of binding and network addresses. This name has been changed and it may meet one of them.
- **Service:** A set of the related endpoints, such as related ports, operations and messages.

WSDL-NS: The format of WSDL files needs to be improved to support natural language annotation when adding semantic information in WSDL files, since the WSDL language itself has no semantic. The language generated in this way is seen as a WSDL-NS language in this study.

To semantically describe Web services, the WSDL language will be extended in the following. A new tag <description> includes an verb description attribute verb and a noun description attribute noun. In this way, the tag <description> can describe the semantic information of Web services.

For the tag <operation>, one tag <description> is added and used to represent the function semantic description of Web services on the basis of the existed three sub-tags. The detailed description is shown as follows:

```
<wsdl:operation name="getSupport
Province">
<wsdl:documentation xmlns:wsdl="http:
//schemas.xmlsoap.org/wsdl/">
```

```
The information of the state, provinces and cities supported by the Web Service of weather forecast:
</wsdl:documentation>
<wsdl:input message="tns:getSupport Province
SoapIn" />
<wsdl:output message="tns:getSupport
ProvinceSoapOut" />
<wsdl:description verb="get" noun="city
name"></wsdl :description>
</wsdl:operation>
```

To describe the semantic information of preconditions and effects, new tags <preConditionDescription> and <effectDescription> are added in tag <operation>. They describe the semantic information of service preconditions and influence, respectively. We can increase many < description > tags under the two new tags. The detailed description is given as follows:

```
<wsdl:operation name="getSupportProvince">
<wsdl:documentation xmlns:wsdl="http://
schemas.xmlsoap.org/wsdl/">
The information of the state, provinces and cities supported by the Web Service of weather forecast.
</wsdl:documentation>
<wsdl:input message="tns:getSupport Province
SoapIn" />
<wsdl:output message="tns: getSupport
Province SoapOut" />
<wsdl:description verb="get" noun="city
name"></wsdl:description>
<wsdl: preConditionDescription>
<wsdl:description verb="gain"noun="licence"
></wsdl:description>
</wsdl: preConditionDescription>
<wsdl: effectDescription>
<wsdl:descriptionverb="add"noun="number"
></ wsdl:description>
<wsdl:description verb="refresh"noun="record"
></wsdl:description>
</wsdl: effectDescription>
</wsdl:operation>
```

Although, the input/output parameter tags of Web services are defined in tag <operation >, they are a kind of information, i.e., the content contained in tag <message>. So the semantic description of input/output parameters should be put in tag <message>, in other words, a <description> tag is added in tag <message>. The detailed description is as follows:

```
<wsdl:messagename="getSupportProvince
SoapIn">
<wsdl:part name="parameters"element="tns: get
SupportProvince" />
<wsdl:description verb=""noun="country"></
wsdl:description>
</wsdl:message>
<wsdl:message name="getSupportProvince
SoapOut">
<wsdl:part name="parameters"element="tns:get
SupportProvinceResponse" />
<wsdl:description verb=""noun="cityname
"></wsdl:description>
</wsdl:message>
```

WSDL-NS is more strict than WSDL. The name space of WSDL is no longer suitable for WSDL-NS.

Matching algorithm: In the following, a matching algorithm of Web services is given based on WSDL-NS and natural semantics:

Algorithm 1: Matching algorithm of Web services

Input: A target Web service description L, similarity threshold sim = $(sim'_f, sim'_p, sim'_i, sim'_o, sim'_e)$ and a Web service set φ annotated by natural semantics.

Output: A Web service set σ meeting service preconditions.

(1) If φ≠Ø, for any Web service s∈φ, its description *l* is gotten and s is removed from φ; If φ is null, return σ;

(2) If L is the parent description of *l*, s is put into set σ and do (1); else do (3);

(3) If sim (L, *l*) ≥sim, s is put into set σ and do (1).

EXPERIMENT AND ANALYSIS

PanSchema is developed independently by PanSoft company. It is a software development platform supporting a new generation of software development technology. It is built based on MDA (Model Driven Architecture) and used as a system of developing application softwares and generating large size business components.

Data interaction in PanSchema involved in many Web services. We apply it to the verification of the proposed methods in this paper. The standard semantic description of each service is given firstly and the Web services are annotated with natural semantic information. Then the requirement Web services are matched based on Algorithm 1. The experiment results are shown in Fig. 1 and 2, where x axis represent sim (n), y axis represents sim (v) and z axis represents the precision ratio in Fig. 1 and the recall ratio in Fig. 2, respectively. The average recall ratio and precision ratio are computed based on different

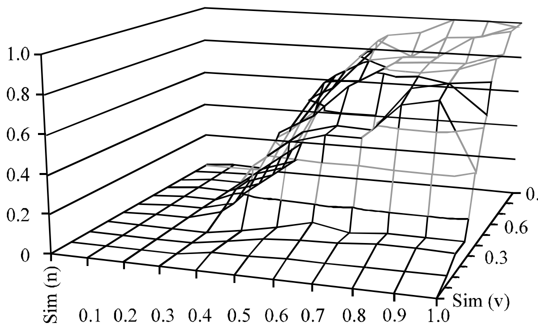

Fig. 1: The surface chart of precision ratio in the experiment

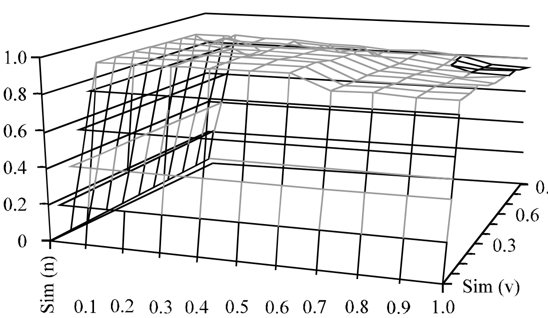

Fig. 2: The surface chart of recall ratio in the experiment

threshold values. We can find that precision ratio is optimized when verb similarity sim (v)≥0.9 and noun similarity sim (n)≥0.8. The recall ratio is optimized when verb similarity sim (v)≤0.7 and noun similarity sim (n)≤0.6.

CONCLUSIONS

The computing method of natural semantic similarity is improved in this paper. By comparing with the traditional methods, the precision of semantic similarity of the proposed method is increased. To add the semantic information in Web services, the Web services with natural semantic information are defined in the paper. WSDL-NS is constructed by means of WSDL and a new discovering algorithm of Web services is proposed based on natural semantic. Thus it lays a foundation for the practical application of Web services. Some simulation experiments are given based on the PanSchema software platform and the recall ratio and precision ratio are analyzed. However, a very high recall ratio is not obtained by this method and the annotated services with errors can not be recognized. Therefore, the future research will focus on the consistency of annotation and services to get the better results.

ACKNOWLEDGMENT

This work is supported by the National Natural Science Foundation of China under grant 61170078; the doctoral program of higher education of the specialized research fund of China under grant 20113718110004.

REFERENCES

Ankolekar, A., M. Burstein, J.R. Hobbs, O. Lassila and D. Martin *et al*., 2002. DAML-S: Web service description for the semantic web. Proceedings of the 1st International Semantic Web Conference Sardinia, June 9-12, 2002, Italy, pp: 348-363.

Daniela, B., D. Calvanese, G.D. Giacomo, M. Lenzerini and M. Mecella, 2005. Automatic service composition based on behavioral descriptions. Int. J. Cooperative Inf. Syst., 14: 333-376.

Klusch, M., G. Fries and K. Sycara, 2009. OWLS-MX: A hybrid semantic web service matchmaker for owl-s services. J. Web Sem., 7: 121-133.

Li, Y., Z.A. Bandar and D. McLean, 2003. An approach for measuring semantic similarity using multiple information sources. IEEE Trans. Knowledge Data Eng., 15: 871-882.

McHale, M.L., 1998. A comparison of wordnet and roget's taxonomy for measuring semantic similarity. Proceedings of the COLING/ACL Workshop on Usage of WordNet in Natural Language Processing Systems, August 16, 1998, Montreal, Canada, pp: 115-120.

Oliva, J., J.I. Serrano, M.D. del Castillo and A. Iglesias, 2011. SyMSS: A syntax-based measure for short-text semantic similarity. Data Knowledge Eng., 70: 390-405.

Tocha, E., I. Reinhartz-Berger and D. Dori, 2011. Humans, semantic services and similarity: A user study of semantic Web services matching and composition. Web Semantics: Sci. Serv. Agents World Wide Web, 9: 16-28.

Wang Y. and E. Stroulia, 2003. Semantic structure matching for assessing web-service similarity. Proceedings of 1st International Conference on Service Oriented Computing, December 15-18, 2003, Trento, Italy, pp: 194-207.