2P-4

Semantic Web Service Discovery Using Weighted Directed Acyclic Graph

Anang Kunaefi[1,2], Takayuki Nagai[1], Hiroshi Nakano[1], Riyanarto Sarno[2]

1) Center for Multimedia and Information Technologies,
Kumamoto University, Japan

2) Software Engineering Laboratory,
Sepuluh Nopember Institute of Technology, Surabaya, Indonesia

Abstract

Semantic Web service is becoming a new paradigm for distributed computing over the Internet. This gives the power to applications to automatically or semi-automatically discover, invoke, compose and monitor web services. The success of semantic web service depends on how we can find web services that satisfy user needs. Many approaches have been proposed for service discovery. Most of them utilizing Ontology to describe web service related information. However, in Ontology based approaches it is still difficult to focus the search on specific attribute of the service. In this research, Web Service Description Language (WSDL) of web service is represented as Weighted Directed Acyclic Graph (WDAG). With this representation, users can dynamically change attribute's weight to refine the search. For example, user can assign a bigger weight to 'input name' attribute than other attributes to relatively focus on the attribute in search.

Keywords: Web Service, Semantic Web Service, Semantic Discovery.

I. INTRODUCTION

Semantic web service can be defined as web service automation for service discovery, selection, composition and execution [3]. To achieve this, at least there are two things we need to do. First, how to represent the service as a knowledge or metadata, and second, how to access and assess the knowledge.

Many studies have been made in this field. Some of them use ontology as a metadata to represent the service. As described in [2], service selection is performed using syntactic similarity in Ontology Web Language for Service (OWLS) representation. Also, because ontology is a method to define relationship among concepts or knowledge, Klusch et al. in [2] evaluates the similarity of two services based on related words in the comparing services.

However, when using ontology, it is difficult to focus the search on specific attribute of the service. Let say, if users only know small information about the service they need, then the discovery process should focus only on that information.

In this research, we use WDAG to accommodate this situation in service discovery directly from the Internet. We show the result of ROC analysis of our method. In the previous research, Sarno et al. in [4] use WDAG for semantic matchmaking of service but limited only for synthesized dataset.

II. METHOD

Unlike ordinary website, web service does not have content. It only has Web Service Document Language (WSDL) as a document contract that contains many information about what the service offers, how to access, and how it works. All of the information is in XML format. To deal with web service environment, we explain our method as follows:

a. Data set preparation

Today, web service are often registered on specific portals (i.e. webservicex.net, xmethods.net) or simply put on the web together with some web pages describing the features of the service.

Therefore, in this research we implemented compact web service crawler to get WSDL documents from the Internet and generate metadata representation of them.

b. Metadata representation

After getting all the WSDLs address from crawling process, the next important task is to parse and mine each WSDL and build WDAG structure from it. For fast query, the graph is converted into graph object stored in database using DB4O (Database for Object).

The purpose of the mining process is to get information related with the service, like service name, operation name, input name, output name and so on.

According to [1] WDAG is arc-labeled, node-labeled and arc-weighted directed acyclic graph. Generated WDAG as a result of WSDL mining process can be seen in Fig. 1. The “Term” part in the structure will be replaced by actual value from the WSDL. The actual value has a vast possibility since there are many kinds of web services in the Internet. There is also possibility that the value is an array of values.

c. Discovery process

The discovery process is performed by computing the similarity between the WDAG structure generated from user input and WDAG structure stored in database. We implemented WDAG similarity algorithm by Jin, J et al. [1] in Java.

III. RESULT

We use Java-based environment for our experiment. We collect 100 web services in 12 categories by crawling webservicex.net and xmethods.net. These categories include weather service, location service, finance service, currency service, validation service and others.

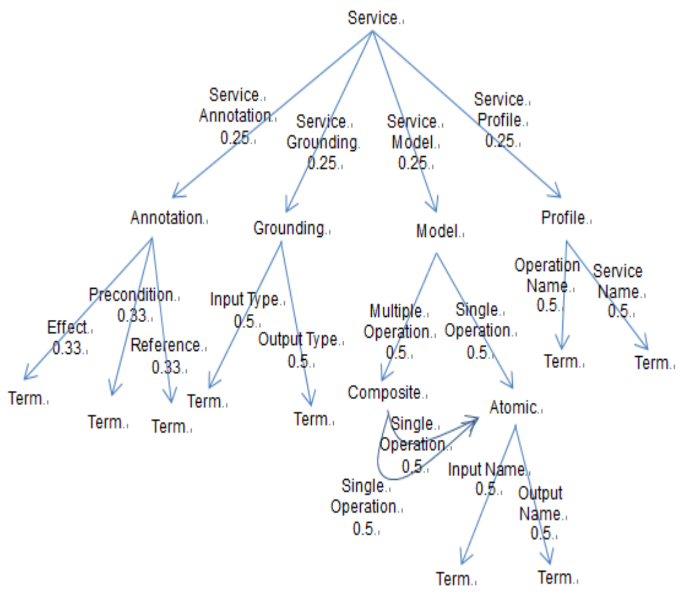

Fig. 1. WDAG structure

To conduct the experiment, we try some scenarios in the query. The scenario includes making query for at least one known attribute and then incrementally adds the number of known attributes.

For example, in one-known-attribute scenario, we define values only for `Service Profile` attribute and its children, `Operation Name` and `Service Name` attributes in the query. Leaving other attributes blank (weight = 0.0), in this case `Service Model`, `Service Grounding`, `Service Annotation` and all of its children. Setting the weight for `Service Profile` to maximum (1.0), give results to the highest similarity score 0.625.

In two-known-attributes scenario, for example, we define values for `Service Grounding` (SG) and `Service Profile` (SP) attributes as follows:

```
SG (weight = 0.5)
  Input Type (weight=0.5),
  Output Type(weight=0.5)
SP (weight = 0.5)
  Operation Name (weight=0.5),
  Service Name(weight=0.5)
```

We can modify the weight according to user needs, as long as the sum of weight of siblings is equal to 1.0.

In three and four known attributes scenario, the highest similarity score are 0.87 and 0.99 respectively, where the highest possible similarity score is equal to 1.0 (exact match).

To evaluate the accuracy of the query result, we use Receiver Operating Characteristics (ROC). Here, we assign a threshold value 0.7, with exception in the case of one-known-attribute scenario, where threshold is decreased to 0.6. All services with similarity score above the threshold will be returned to the user as the result of the query.

As a result, there will be four possible conditions in the query result, that is correct hit (True Positive/TP), correct rejection (True Negative/TN), incorrect hit (False Positive/FP) and incorrect rejection (False Negative/FN). The accuracy is given by the following formula:.

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}$$

By observing scenarios described above for several cases, we get the accuracy of this method is between 72% and 88%. Fig. 2 shows ROC curve for the case of 72% accuracy, and Fig. 3 shows ROC curve for the case of 88% accuracy. More attributes defined in the query lead to higher accuracy.

The execution speed range is between 0.92 seconds and 1.29 seconds per query.

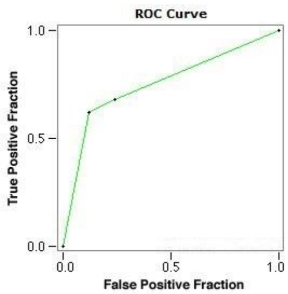

Fig. 2. ROC Curve 1

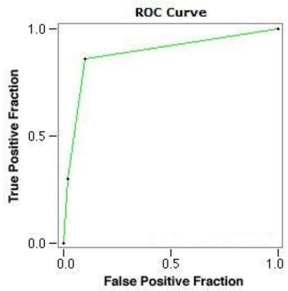

Fig. 3. ROC Curve 2

IV. CONCLUSION AND FUTURE WORK

In this paper, we performed semantic discovery of web service using Weighted Directed Acyclic Graph. The approach is performed using web services available on the Internet. The result shows that this method works well.

Using WDAG structure presented in this paper has also enabled us to automatically execute the web service being searched by users. We believe that using the same structure, we can automatically compose web service. This is what we will do in the future research.

V.REFERENCES

[1] Jin, J.; Boley, H.; Bhavsar, V.C.; Prasad, S.K.; Sarker, B.K.; Yang, L, "*Similarity of Weighted Directed Acyclic Graphs*", Master thesis UNB Computer Science Research Expo, University of New Brunswick, Canada, 2006.

[2] Klusch, Matthias; Fries, Benedikt; Sycara, Katia, *Automated Semantic Web Service Discovery with OWLS-MX*, ACM, 2006.

[3] Fensel, Dieter, *Enabling Semantic Web Services: The Web Service Modelling Ontology*, Springer-Verlag, Berlin, 2007.

[4] Sarno, R; Ghozali, K; Nugroho, B.A.; Hijriani, A, "Semantic Matchmaking Using Weighted Directed Acyclic Graph", In *Proceedings of the third International Seminar on Applied Technology, Science and Arts (APTECS)*, 2011.