Enriching Semantically Web Service Descriptions

Maricela Bravo[1], José Rodríguez[2], and Alejandro Reyes[1]

[1]Systems Department, Autonomous Metropolitan University
Azcapotzalco, DF, CP 02200 - Mexico
{mcbc,jaro}@correo.azc.uam.mx
[2]Computing Department, CINVESTAV-IPN
Gustavo A. Madero, DF, CP 07300 - Mexico
rodriguez@cs.cinvestav.mx

Abstract. Service Oriented Computing (SOC) has incrementally been adopted as the preferred programming paradigm for the development, integration and interoperation of large and complex information systems. However, despite its increasing popularity, the SOC has not achieved its full potential yet. This is mainly due to the lack of supporting tools to enrich and represent semantically Web service descriptions. This paper describes a solution approach for the automatic representation of Web service descriptions and their further semantic enrichment between operation names based on the calculation of four semantic similarity measures. The enrichment approach is accurate because the final decision is done through a voting scheme, in the case of inconsistent results, these are not asserted into the ontology. Experimentation shows that although few similarity relationships are found and asserted, they represent an important step towards the automatic discovery of information that was previously unknown.

Keywords: Public Web Service Descriptions, Ontology Representation, Semantic Web Services.

1 Introduction

Service Oriented Computing (SOC) has been adopted as the preferred programming paradigm for the development of complex information systems. This trend has led to the emergence of service repositories, service frameworks and many supporting technologies which offer facilities for searching, discovering, selecting and invoking Web service operations. However, the SOC has not achieved its full potentiality, mainly because search and invocation of Web service operations still lacks of the level of automation and that facilitates that any service requestor exploits any public available Web service.

Whenever a service requestor searches for a service in public repositories, he obtains a list of services that syntactically match the keywords provided, then he has to check one by one in order to identify which of these services satisfies his *functional* requirements. This is not an easy task as the majority of available Web services are described in WSDL 1.0 1.1, or WSDL 2.0. WSDL offers syntactical information

R. Meersman et al. (Eds.): OTM 2014, LNCS 8841, pp. 776–783, 2014.

regarding the service address, name, operations, input and output messages and all the required information to invoke the service. However, WSDL lacks of functional information regarding the use of operations and parameter values. To advance on a feasible solution, it is necessary to build repositories of semantically enriched web services; but these repositories must reuse existing services. The work reported in this paper represents a step towards this end.

Semantic Web Services, introduced by McIlraith et al. in [1], relay on the incorporation of ontologies to enhance service descriptions. According with T. Gruber [2] "An ontology is an explicit specification of a conceptualization". An ontology also defines formally the relationships that exist between terms and a set of axioms which detail and restrict the concepts. Inspired by the concept of Semantic Web Services, in this paper we introduce a solution for the automatic representation of Web service descriptions as ontological models and their further semantic enrichment based on semantic similarity measures. The type of semantic relationships that are discovered between ***operation names*** are "*isSimilarTo*" and "*isDifferentTo*".

Continuing with the work reported in [3], in this work *semantic enrichment of Web services* is the improvement of Web service descriptions by means of an ontological representation. This process consists of three general phases: 1) *Web Service Ontology Generation*, this phase consists of parsing Web service descriptions and applying a predefined ontology template to automatically produce its corresponding ontology model; 2) *Discovery of Semantic Relationships*, which consists of calculating similarity measures between operation names and for each set of results calculate the upper and lower thresholds, which are then used to identify semantic relationships: *isSimilarTo*, *isDifferentTo* and *unDefined*. To decide on semantic relationships a voting schema is used. An important design goal for this phase was to build a module capable of incorporating ***any set of similarity measures***. 3) *Instantiate new Semantic Relationships*, which consists of asserting semantic relationships between service operations into the ontology.

The rest of the paper is organized as follows: in Section 2, related work is presented; in Section 3, the discovery of semantic relationships between service operations is detailed; in Section 4, the ontological model for the representation of Web services is presented; in Section 5, the experimental setup is described; in Section 6, experimental results are evaluated; and finally conclusions are presented in Section 7.

2 Related Work

In this section related work concerning semantic Web services, service directories and similarity measures is described. McIlarith et al. [1] described an approach to markup Web services to enable automatic Web service discovery, execution, composition and interoperation. Sycara et al. [4] presented one of the first semantic languages and infrastructure to mark up Web services: DAML-S. An ontology mapping solution was presented by Pathak et al. [5] to support the translation of service ontologies and user ontologies facilitating service discovery and matchmaking using non functional characteristics. Klush, Fries and Sycara [6] presented OWLS-MX a OWL-S service matchmaker which incorporates reasoning on logically defined preconditions and effects. The main limitation with OWL-S is that the majority of public available service descriptions are in WSDL language, few OWL-S public service descriptions exist. Gomadam et al. [7] introduced the notion of semantic template to capture the requirements of a service requestor using SAWSDL and model references. Authors also describe an automatic approach for Web service composition addressing the problems of process heterogeneities and data heterogeneities. Du, Song and Munro [8] described a method for transforming existing Web service descriptions into an enhanced semantic Web service framework which incorporates composition relationships between services. Their composition relationship definition links a service output with a different service input through a similarity measure. However, this output-input relationship rather defines a data type compatibility than a functional compatibility. Elgazzar, Hassan and Martin [9] presented an approach to improve Web service discovery by clustering Web services into functionally similar groups. The main limitation of their approach is that they do not provide any semantic representation of clusters or mechanisms to infer and reason about their results.

OWL-S is an ontology-based service description language [15], which supplies service providers with a set of constructs for describing the properties and capabilities of their Web services. An OWL-S Service presents a service Profile, is described by a Service Model (or Process Model); and supports a Service Grounding. The Semantic Annotation for WSDL (SAWSDL) specification [16] defines a set of mechanisms to add annotations to WSDL documents, such annotations reference ontologies. SAWSDL is helpful for the discovery and invocation of Web services. The Web Service Modeling Ontology (WSMO) is a complete ontological model [17] that describes: Ontologies, Web Services, Goals, and Mediators. The WSMO incorporates the Web Service Modeling Language (WSML), a language for the specification of Semantic Web services.

Related works rely on the incorporation of ontologies as a mechanism to achieve semantic interoperability. The main drawbacks of these related works is that users need to provide their ontology concepts and create manually mappings between ontologies. None of reported works have presented a fully automated enrichment approach using public available Web services described with different WSDL versions.

3 Discovering Semantic Relationships

The objective of this phase is to find similarities or differences between operation names using semantic measures and establish new semantic relationships between individuals into the ontology.

Calculate Semantic Similarities between all Operations. The operation names are short texts from one to seven words. These names are written in several formats and contain no relevant information in some cases. So, in order to get the similarity between operations a pre-processing phase is required. Preprocessing phase involves obtaining lexical units that are part of the operation name. First, text normalization is performed in order to transform operations names into a single canonical form, for example: *getFlightPrice*. Then, lexical units are obtained from operation names, for example: [*get*][*Flight*][*Price*]. Finally, the processing also includes a lexical discrimination of several words that do not contain important meaning in the operations, which area: *http, for, return, result, soap*. These words are filtered out prior to calculate semantic similarity between operations.

The level of similarity between operations pairs is performed by calculating the average of semantic similarity measures between all words filtered. Four measures were used for this process: the Wu and Palmer measure [10] that calculates semantic similarity by considering the depths of the two synsets in the WordNet taxonomies, along with the depth of the lowest common subsumer; the Lin measure [11], which is a universal definition of similarity in terms of information theory that is not directly stated as in earlier definitions, rather, it is derived from a set of assumptions; the path measure [12] relies on the length of the shortest path between two synsets for their measure of similarity limiting to IS-A links and scale the path length by the overall depth of the taxonomy; the Lesk measure [13] proposed that the relatedness of two words is proportional to the extent of overlaps of their dictionary definitions.

Calculate Mean and Standard Deviation. We calculate the arithmetic mean and standard deviation using the similarities of all operation pairs. Arithmetic mean is obtained from Equation 1, while standard deviation is defined from Equation 2.

$$\bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i \tag{1}$$

$$\sigma = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(x_i - \bar{x})^2} \tag{2}$$

Suppose we have a data set, $x_1, \ldots, x_n$ then
N is the number of n values of our data set
x_i is a data contained in our data set

We use the arithmetic mean and standard deviation to obtain the upper and lower thresholds. The upper threshold is used to identify those operation pairs that have a positive similarity according with the semantic measure applied. The lower threshold is used in the identification of operation pairs that are definitively different in accordance with the semantic measure applied.

Set the Upper and Lower Thresholds and Discover Relationships. Two thresholds are defined from the arithmetic mean and standard deviation in order to determine the limits that represent the corresponding semantic relationships. The upper threshold defines the limit for the similarity relationship *Operation*$_i$ *isSimilarTo Operation*$_j$.

$$T_{upper}(sim) = \bar{x} + \sigma \tag{3}$$

where $\bar{x}$ is the arithmetic mean and σ is the standard deviation.

The lower threshold defines the limit for the difference relationship between operations $Operation_i$ *isDifferentTo* $Operation_j$.

$$T_{lower}(sim) = \bar{x} + \sigma \tag{4}$$

where $\bar{x}$ is the arithmetic mean and σ is the standard deviation.

For each similarity between operations that are over the upper threshold a semantic relationship *isSimilarTo* is defined. Also, for each similarity under the lower threshold a semantic relationship *isDifferentTo* is defined.

For those operation pairs whose calculations resulted under the upper threshold and over the lower threshold no semantic relationship is established, because there is no numerical certainty to establish the similarity or difference.

4 Ontological Model

The general model was designed to represent the following service description implementations: WSDL 1.0, WSDL 1.1 and WSDL 2.0. It consists of: a *Service class*, which represents a WSDL service description; *Endpoint class*, which specifies a unique network address that the service consumer uses to invoke the methods of the service; *Binding class*, which specifies the SOAP binding style and transport; *Interface class*, the *porType* of a Web service defines the operations that can be invoked, and the input and output messages that are used to execute the operation, in the ontological model depicted in Figure 1, instead of a *porType* class, an *Interface* class was created to make this model compatible with the WSDL 2.0 specification; *Operation class*, represents the methods offered by the service interface, the WSDL specification defines an input message and output message for each operation; *Parameter class*, the *Parameter* class represents the super class of the *ParameterInput* class and the *ParameterOutput* class.

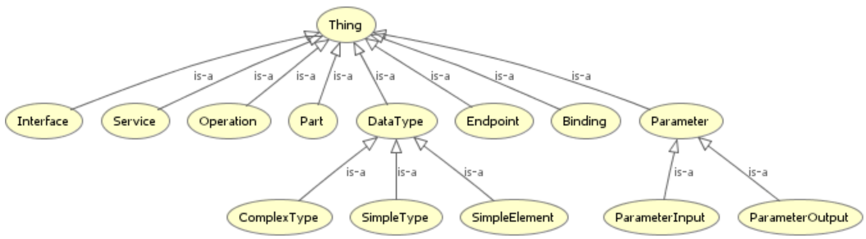

Fig. 1. General ontological model for the representation of Web services

5 Experimentation

An experiment was executed using the test collection OWLS-TC3, which contains 1080 Web service descriptions. A subset of 43 Web services was selected using as a criteria their file names - file name starting with "Country" - this selection was done arbitrarily.

1) *Web Service Ontology Generation*. SDWS was executed following a sequence of predefined steps: selection of any set of Web services; uploading the set of selected services, parsing the services according to their representation language; populating a new ontology based on the respective template; and finally downloading the new produced ontology.

2) *Discovery of Semantic Relationships*. After executing SDWS tool the resulting ontology has 43 *Operation* instances. Therefore, the total number of comparison pairs between $\boldsymbol{n}$ operations is given by $\boldsymbol{nc} = (43^2 - 43)/2 = 903$. For each comparison pair four semantic similarities are calculated: Wu-Palmer [10], Lin [11], Path [12] and Lesk [13]. For each set of results, the arithmetic mean and standard deviation are calculated in order to obtain the upper and lower thresholds using formulas (3) and (4). A voting schema is used, which considers three possible results: *isSimilarTo*, when the semantic measure value results over the upper threshold; *isDifferentTo*, when the semantic measure value results under the lower threshold; and *Undefined*, when the semantic measure value results between the upper and lower thresholds. The final decision uses a user-defined majority value. If the number of *isDifferentTo* results is greater than the majority value, then the final meaning is established as *isDifferentTo*. If the number of *isSimilarTo* results is greater than the majority value, then the final meaning is established as *isSimilarTo*. For this experimentation the majority value was established in 2.

3) *Instantiate new Semantic Relationships*. The last phase is to assert the new semantic relationships between individuals into the ontology. For this step we are considering only the *isSimilarTo* and *isDifferentTo* results. Figure 2 shows that the operation get_COMPANY_PROFESSION *isDifferentTo* 6 other operation instances.

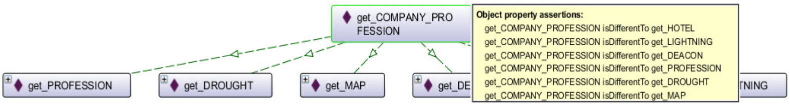

Fig. 2. Some semantic assertions between operations of the *isDifferentTo* relation

6 Evaluation of Results

For evaluation *Precision* and *Recall* measures were calculated [14]. A human (Web service requestor) compared operation pairs and decided based on observation and semantic sound if operation pairs were "*Similar*" or "*Different*". Table 1 shows the *Precision* and *Recall* measures results. From the 903 operation names comparisons, 12 *isSimilarTo* semantic relationships were asserted into the ontology. From these results only 6 were the result of an unanimous vote, the rest of *isSimilarTo* relations are due to the votes of Lin and Lesk measures. From the same set of 903 operation names comparisons, 75 *isDifferentTo* semantic relations were asserted into the ontology. The rest of comparison pairs resulted *unDefined*.

Table 1. Precision and recall results

	Precision	**Recall**	**F-measure**
isSimilarTo	1.000000	0.218182	0.358209
isDifferentTo	0.986667	0.087264	0.160347
unDefined	0.945666	0.862028	0.901912

It is important to note that 4 inconsistencies were found between Lin and Path measures, where the upper threshold of Lin establishes them as *isSimilarTo*, whereas Path lower threshold defines them as *isDifferentTo*. These particular cases occur between the operation names *get_HOTEL* and *get_MAP*. With this particular example it is possible to see that Lin is erroneously giving false positives. Wu-Palmer similarity obtains 100 more *isSimilarTo* results than the rest of measures. This indicates that Wu-Palmer measure may be giving more false positives. One of the important features is the secure establishment of similarity relations in the ontology. In the case of finding inconsistencies the voting scheme results in *unDefined* and therefore these relationships are not asserted into the ontology.

7 Conclusions

We have described a semantic enrichment approach which automatically represents any set of available Web service descriptions as ontologies. We have used SDWS, a tool that facilitates the automatic translation of different service description files into ontological models. SDWS incorporates a set of Web service parsers that allow the automatic extraction of service interface definitions for their semantic representation, the main benefit if this tool is that does not require human intervention, facilitating end users to make use of semantic Web technologies without added complexity.

We have described a semantic relatedness discovery process which calculates four semantic similarities between all operations pairs, then calculates the upper and lower thresholds; and identifies operation pairs that are over and under respective thresholds. To assert new semantic relationships between operations into the ontology, a voting scheme is used assuring that the establishment of semantic relations is sufficiently reliable since it is based on a majority vote. In the case of inconsistent results, these are not asserted into the ontology.

Experimentation shows that although few similarity relations are found and asserted, they represent an important step towards the automatic discovery of information that was previously unknown and that can be very useful during automatic search, selection and invocation of Web services based on the operation names. As future work, more similarity measurements will be applied using different approaches, such as: syntactic, semantic, structural and pragmatic. These similarity measurements will be extended to more elements of the Web service descriptions, such as input and output parameter names and parameter types, and texts of the document tags.

References

1. McIlraith, S.A., Son, T.C., Zeng, H.: Semantic web services. IEEE Intelligent Systems 16(2), 46–53 (2001)
2. Gruber, T.: A Translation approach to portable ontologies. Knowledge Acquisition 5(2), 199–220 (1993)
3. Bravo, M., Pascual, J., Rodríguez, J.: Semantic Representation of Public Web Service Descriptions. In: Murgante, B., Misra, S., Carlini, M., Torre, C.M., Nguyen, H.-Q., Taniar, D., Apduhan, B.O., Gervasi, O. (eds.) ICCSA 2013, Part V. LNCS, vol. 7975, pp. 636–651. Springer, Heidelberg (2013)
4. Sycara, K., Paolucci, M., Ankolekar, A., Srinivasan, N.: Automated discovery, interaction and composition of semantic Web services. Web Semantics: Science, Services and Agents on the World Wide Web 1(1), 27–46 (2003)
5. Pathak, J., Koul, N., Caragea, D., Honavar, V.G.: A framework for semantic web services discovery. In: Proceedings of the 7th Annual ACM International Workshop on Web Information and Data Management, pp. 45–50. ACM (2005)
6. Klusch, M., Fries, B., Sycara, K.: Automated semantic Web service discovery with OWLS-MX. In: Proceedings of the Fifth International Joint Conference on Autonomous Agents and Multiagent Systems, pp. 915–922. ACM (2006)
7. Gomadam, K., Ranabahu, A., Wu, Z., Sheth, A.P., Miller, J.: A declarative approach using SAWSDL and semantic templates towards process mediation. In: Semantic Web Services Challenge, pp. 101–118. Springer, US (2009)
8. Du, X., Song, W., Munro, M.: A Method for Transforming Existing Web Service Descriptions into an Enhanced Semantic Web Service Framework. In: Information Systems Development, pp. 217–226. Springer, US (2010)
9. Elgazzar, K., Hassan, A.E., Martin, P.: Clustering wsdl documents to bootstrap the discovery of Web services. In: 2010 IEEE International Conference on Web Services (ICWS), pp. 147–154. IEEE (2010)
10. Wu, Z., Palmer, M.: Verbs semantics and lexical selection. In: Proceedings of the 32nd Annual Meeting on Association for Computational Linguistics, pp. 133–138. Association for Computational Linguistics (1994)
11. Lin, D.: An information-theoretic definition of similarity. In: ICML, vol. 98, pp. 296–304 (1998)
12. Leacock, C., Chodorow, M.: Combining local context and WordNet similarity for word sense identification. WordNet: An electronic lexical database 49(2), 265–283 (1998)
13. Banerjee, S., Pedersen, T.: An Adapted Lesk Algorithm for Word Sense Disambiguation Using WordNet. In: Gelbukh, A. (ed.) CICLing 2002. LNCS, vol. 2276, pp. 136–145. Springer, Heidelberg (2002)
14. Baeza-Yates, R., Ribeiro-Neto, B.: Modern information retrieval, vol. 463. ACM Press, New York (1999)
15. Martin, D., Burstein, M., Hobbs, J., Lassila, O., McDermott, D., McIlraith, S., Sycara, K.: OWL-S: Semantic markup for web services. W3C member submission, 22 (April 2004)
16. Roman, D., Keller, U., Lausen, H., de Bruijn, J., Lara, R., Stollberg, M., Fensel, D.: Web service modeling ontology. Applied Ontology 1(1), 77–106 (2005)
17. Kopecky, J., Vitvar, T., Bournez, C., Farrell, J.: SAWSDL: Semantic annotations for wsdl and xml schema. IEEE Internet Computing 11(6), 60–67 (2007)