

A Study on Knowledge Discovery of Relevant Web Services with Semantic and Syntactic approaches

Mrs. M. Akila Rani ,
Assistant Professor,

Department of Computer Science and Engineering,
N.P.R College of Engineering & Technology,
Natham-624401, Tamil Nadu, India.
Email: akilakamalam@gmail.com

Dr. D. Shanthy,
Professor,

Department of Computer Science and Engineering,
PSNA College of Engineering & Technology,
Dindigul-624622, Tamil Nadu, India.
Email-id: dshan71@gmail.com

ABSTRACT

Web mining is the application of data mining techniques to discover patterns from the Web. Web services defines set of standards like WSDL(Web Service Description Language), SOAP(Simple Object Access Protocol) and UDDI(Universal Description Discovery and Integration) to support service description, discovery and invocation in a uniform interchangeable format between heterogeneous applications. Due to huge number of Web services and short content of WSDL description, the identification of correct Web services becomes a time consuming process and retrieves a vast amount of irrelevant Web services. This emerges the need for the efficient Web service mining framework for Web service discovery. Discovery involves matching, assessment and selection. Various complex relationships may provide incompatibility in delivering and identifying efficient Web services. As a result the web service requester did not attain the exact useful services. A research has emerged to develop method to improve the accuracy of Web service discovery to match the best services. In the discovery of Web services there are two approaches are available namely Semantic based approach and Syntactic based approach. Semantic based approach gives high accuracy than Syntactic approach but it takes high processing time. Syntactic based approach has high flexibility. Thus, this paper presents a survey of semantic based and syntactic based approaches of Web service discovery system and it proposed a novel approach which has better accuracy and good flexibility than existing one. Finally, it compares the existing approaches in web service discovery.

Index Terms: WSDL, SOAP, UDDI, Web service discovery.

1. INTRODUCTION

Web mining is the emerging technology in Web. According to analysis targets, web mining can be divided into three different types, which are Web usage mining, Web content mining and Web structure mining.

Web usage mining is the process of extracting useful information from server logs i.e users history. Web usage mining is the process of finding out what users are looking for on the Internet. The information gathered through Web mining is evaluated (sometimes with the aid of software graphing applications) by using traditional data mining parameters such as clustering and classification, association, and examination of sequential patterns.

Due to the tremendous increase of web services, the search becomes a time consuming process and retrieves a vast amount of irrelevant web services. This motivates the need for the efficient web service mining framework. Finding and invoking the portable composition for web services lead to

challenging activity because of the huge amount of web services availability and short content of WSDL description.

The semantic web service description may have more than one interface relationships among other web services causes complex association. Therefore, web service requester did not attain the exact useful services. Various complex relationships may lead to an incompatible for delivering and identifying efficient web services. These problems can be addressed by the mining framework supported by capability profile specifications based on environment ontology.

A Web Service [1] is a software program identified by an URI, which can be accessed via the internet through its exposed interface. In addition Web services can invoke other Web services.

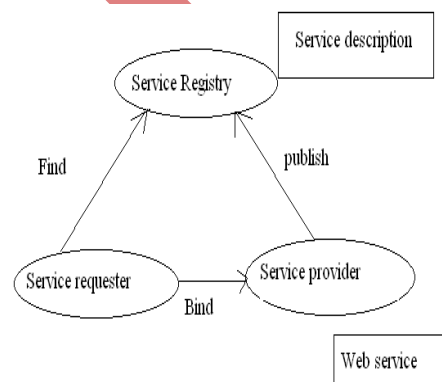


Fig 1: Web Service usage scenario.

The common usage scenario for Web services (Fig 1) can be defined by three phases; *Publish*, *Find*, and *Bind*; and three entities: the service requester, which invokes services[17]; the service provider which responds to requests; and the registry where services can be published or advertised. A service provider *publishes* a description of a service it provides to a service registry. This description (or *advertisement*) includes a profile on the provider of the service (e.g. company name and address); a profile about the service itself (e.g. name, category); and the URL of its service interface definition (i.e. WSDL description). When a developer realizes a need for a new service, he *finds* the desired service either by constructing a query, or browsing the registry. The developer then interprets the meaning of the interface description (typically through the use of meaningful label or variable names, comments, or additional documentation) and *binds* to (i.e. includes a call to invoke) the discovered service within the application they are developing. This application is known as the *service requester*.

At this point, the service requester can automatically invoke the discovered service (provided by the service provider) using Web service communication protocols (i.e. SOAP).

2. EXISTING SYSTEM

Web Services can convert our applications into Web-applications. Web Services are published, found, and used through the Web. Web services are application components. These are self-contained and self-describing. It can be discovered using UDDI. XML(eXtended Markup Language) is the basis for Web services. That is, UDDI is used to list the available services. SOAP is used as transfer protocol for Web Services.

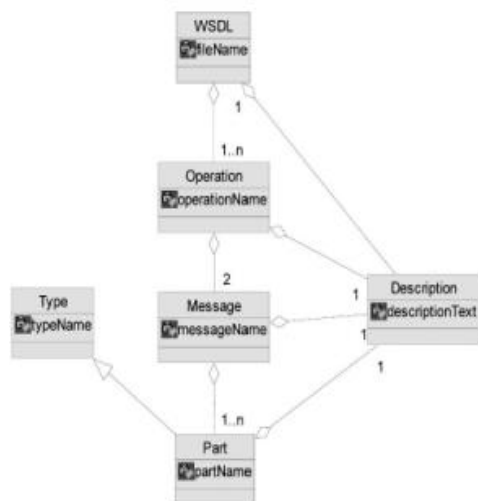


Fig 2: Information available in Web service specification.

Discovering the services which are having “relevant capability” is the process of Knowledge Discovery in Services. Each Web service has an associated XML-based document called WSDL. WSDL file describes Web service functionality and interface information. Fig: 2 explains what are all the information available in Web service specification.

A lot of research has been conducted in the area of web service discovery. The whole set of work can be divided into two categories as Syntactic based and Semantic based approach.

2.1 Syntactic Based Approach

Web services have very brief syntactic descriptions from WSDL files. Semantic descriptions are maintained outside of WSDL documents and referenced to using extensibility elements. The lack of textual information makes keyword-based search models unable to filter irrelevant search results. If a single Web service is unable to satisfy the functionality required by the user, it would be possible force the combination of multiple existing services that can fully satisfy user needs. A web service composition methodology make use of graph based model to find both the similar operations and composable ones. Number of approaches such as Web service flow language [WSFL], business process execution language for Web services [BPEL4WS] [4] and XLANG are available for Web service composition. Most of these techniques support process modeling at the syntactic level and is unable to support reasoning at a conceptual level. This syntactic based approach is very flexible but it gives less accuracy by means of confidence lacking.

2.2 Semantic Based Approach

The semantic matching can be performed by exploiting the semantic representation of concepts and their relations in the Web. The ontology supports semantically enhanced information processing and interoperability. To enable semantic matching of web service specifications the WordNet [11] lexicon may used. The semantic correlations between WSDL concepts can found with the help of the WordNet-based semantic similarity measure. To enrich semantics of web services, ontology based web service standards are introduced (OWL-S)[OWL-S Semantic Markup for Web Services, 2004], Web Services Semantics (WSDL-S) [WSDL-S, W3C, 2006], Web Service Modeling Ontology (WSMO) [J. Bruijn et al., 2006] to enhance a machine processable and understandable semantics. WSDL labels carry poor semantics and service descriptions are lacking. The semantic web service technology enables unambiguous description of web services for automated and machine processable form to provide better service discovery, composition and execution [S.A. McIlraith, T.C. Son, and H. Zeng, 2001]. A Very important issue of current Web service research is the semantic Web service group, such as OWL-S and WSDL-S. A semantic description of Web services can be obtained with the use of OWL-S language. Recent works of Web service discovery have focused on performing semantic matching to enhance the accuracy. However, firstly, constructing ontology as a semantic backbone for large number of Web services is really difficult. Secondly, for manual annotation it requires that the annotator have some skills in ontology engineering. Even though, this approach is time consuming, it gives high accuracy than syntactic approach.

3. PROPOSED SYSTEM

In this research, a new approach called hybrid approach is proposed based on the combination of existing approaches to improve the accuracy with less processing time.

Our proposed work deals with both semantic approach and syntactic approach as combined. Semantic approach is used to categories the similarity between query and available services in the WSDL (Web Service Description Language). WSDL is enriched with synonyms from WordNet leads to better matching precision. To enable semantic matching of Web service specifications, the WordNet[11] lexicon is employed. WordNet entails a lexical database with words organized into synonym sets representing a lexical concept.

By combining both approaches, the new novel approach can give advantages of both semantic and syntactic approaches. Thus, our proposed approach can give a simple and highly flexible environment. Even thought, this proposed approach may lead time consuming one, its accuracy will compromises every other factors. Thus it may produce high accuracy by fixing some threshold for both Semantic and Syntactic approaches.

4. RELATED WORK

In past years, a lot of research has been conducted in the field of Web service discovery. That works can be generally categorized into 2 types: 1) Semantic based approach, 2) Syntactic based approach. Semantic approaches generally support the integration of web services by exploiting the semantic description of their functionality using ontological approaches. While accurate search and discovery require semantic approaches with technologies such as RDF[16], OWL-S[14] and WSDL-S, currently these technologies are not

widely used in practice. In addition, searching repositories in real time using semantic approaches becomes increasingly more time-consuming as repository sizes increase and the number of reference ontologies increases and becomes more complex.

Conversely, syntactic projects tend to concentrate on string manipulation to correlate services. Rocco et al., (2005) [17] uses string manipulation software to equate web service messages while Pu et al., 2006 [15] uses an eXtensible Markup Language (XML) type-oriented rule based approach.

The specification of a Web service is expressed in WSDL (Christensen et al., 2001), which specifies only the syntax of messages that enter or leave a computer program. In which order messages have to be exchanged between services must be described separately in a flow specification. There are many Web services flow specification languages like BPEL4WS (Curbera et al., 2002) [23] and WSCI (Arkin et al., 2002) [22]. The composition of the flow (i.e., plan) is still manually obtained. Semantic annotations have been widely discussed in the Semantic Web community (Berners-Lee et al., 2001) where preconditions and effects of services are explicitly declared in the Resource Description Format (RDF) (RDF, 1999) using terms from pre-agreed ontologies.

Dong et al., (2004) [5] introduce a web service search engine named Woogle. The approach in this work uses syntactical methods to generate queries in a web service data repository. Consequently, it leads to better search engine result but it does not provide support for semantic web services.

Nowlan et al., (2006) [13] suggested a new approach called naming tendency with syntactic approach. It captures the tendencies of software designers/developers. Naming tendencies are extracted from the more descriptive part name strings of WSDL which are having high frequency. This work combines the nature of message naming with standard string manipulation approach which is one of the syntactic approach. The main drawback of this work is more manual work is needed to identify naming tendency. WSDL provide only syntactic level description of their functionalities.

Jiangang Ma et al., (2009) proposed a Probabilistic Semantic Approach for Discovering Web Services. This work makes use of category matching, thus smaller size of services available for similarity comparison but it is not cost effective.

Kungas et al., (2009) [9] proposed semi automated discovery, selection, composition and management of Web Services. This work is based on syntactic analysis and it uses XML Schema and WSDL interfaces. This approach is cost effective but more care should be taken for naming rule.

5. COMPARISON OF EXISTING APPROACHES

In Web service discovery there are two approaches are available namely Semantic based approach and Syntactic based approach. Each approach has its own advantages and disadvantages.

Ontology based approach and Concept based approach are comes under Semantic based approach as shown in Fig 3..

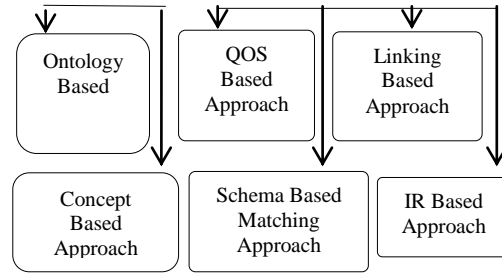
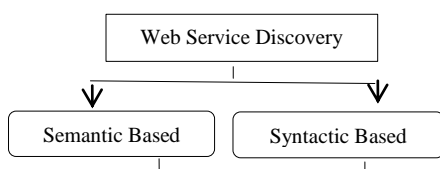


Fig 3: Hierarchy of Web Service Discovery System

QOS based approach, Linking based approach, Schema based matching approach and IR (Information Retrieval)-based approach are comes under Syntactic based approach.

Semantic based approach is very simple approach and it gives good accuracy but it also has some disadvantages like less flexibility and time consumption which leads performance degradation.

Even though the syntactic based approach is complex and gives less accuracy than semantic based approach, it is very flexible approach. There is no performance degradation like semantic approach and it has less number of comparisons in web service discovery, than semantic based approach. All these information are summarized in Table 1.

Table 1: Comparison of existing approaches

FEATURES	SEMANTIC APPROACH	SYNTACTIC APPROACH
Accuracy	High	Low
Simplicity	High	Less
Flexibility	Less	High
Time consumption	More	Less
Number of comparisons	More	Less
Processing time	High	Low

Based on this information we proposed a new approach called hybrid approach which is the collaboration of existing two approaches. This hybrid approach may have the advantages of both approaches. Here, performance degradation may compromise by its accuracy.

6. CONCLUSIONS

In this paper we have tried to provide the huge spectrum of work investigated by researchers globally in the field of Discovering the correct services either semantically or syntactically and highlight the advantages and disadvantages of each system and comparison of these approaches. In this research, we have also introduced a new approach called hybrid approach and finally we conclude that the proposed approach may lead better accuracy than the existing one.

REFERENCES

[1] Amazon Web services, www.amazon.com /gp/aws/landing.html, 2011.

- [2] M.B. Blake and M.F. Nowlan, "Predicting Service Mashup Candidates Using Enhanced Syntactical Message Management," Proc. Int'l Conf. Services Computing, July 2008.
- [3] M.B. Blake, K.C. Tsui, and A. Wombacher, "The EEE-05 Challenge: A New Web Service Discovery and Composition Competition," Proc. IEEE Int'l Conf. E-Technology, E-Commerce, and E-Services, Mar. 2005.
- [4] BPEL4WS, <http://www.ibm.com/developerworks/library/specification/ws-bpel/>, 2011.
- [5] X. Dong, A.Y. Halevy, J. Madhavan, E. Nemes, and J. Zhang, "Similarity Search for Web Services," Proc. Int'l Conf. Very Large Data Bases (VLDB), 2004.
- [6] J. Han and M. Kamber, Data Mining: Concepts and Techniques. Morgan Kaufmann, 2001.
- [7] A. Jhingran, "Enterprise Information Mashups: Integrating Information, Simply," Proc. 32nd Int'l Conf. Very Large Data Bases, 2006.
- [8] J. Koehler and B. Srivastava, "Web Service Composition: Current Solutions and Open Problems," Proc. Workshop Planning for Web Services in conjunction with 13th Int'l Conf. Automated Planning and Scheduling (ICAPS '03), 2003.
- [9] P. Kungas and M. Dumas, "Cost-Effective Semantic Annotation of XML Schemas and Web Service Interfaces," Proc. IEEE Int'l Conf. Services Computing, pp. 372-379, 2009.
- [10] X. Liu, Y. Hui, W. Sun, and H. Liang, "Towards Service Composition Based on Mashup," Proc. IEEE Congress on Services, pp. 332-339, July 2007.
- [11] G. Miller, "Wordnet: A lexical database for English". Communications of ACM 38(11)(1995)39-41.
- [12] NIST Levenshtein Distance, <http://www.nist.gov/dads/HTML/Levenshtein.html>, 2010.
- [13] M.F. Nowlan, D.R. Kahan, and M.B. Blake, "Using Naming Tendencies to Syntactically Link Web Service Messages," Proc. Data Eng. Issues in E-Commerce and Services (DEECS), pp. 90-99, June 2006.
- [14] OWL-S, <http://www.daml.org/owl-s/>, 2011
- [15] K. Pu, V. Hristidis, and N. Koudas, "A Syntactic Rule Based Approach to Web Service Composition," Proc. Int'l Conf. Data Eng. (ICDE '06), 2006.
- [16] Resource Definition Framework (RDF), <http://www.w3.org/RDF/>, 2010.
- [17] D. Rocco, J. Caverlee, L. Liu, and T. Critchlow, "Domain-Specific Web Service Discovery with Service Class Descriptions," Proc. Third IEEE Int'l Conf. Web Services, pp. 481-488, 2005.
- [18] E. Sirin, J. Hendler, and B. Parsia, "Semi-Automatic Composition of Web Services Using Semantic Descriptions," Proc. Web Services: Modeling, Architecture and Infrastructure Workshop in conjunction with Int'l Conf. Enterprise Information Systems (ICEIS '03), 2002.
- [19] Web Services, <http://www.w3.org/2002/ws/desc/>, 2011.
- [20] WS-Challenge, <http://www.ws-challenge.org/>, 2011.
- [21] XMethods, <http://www.xmethods.com/>, 2011.
- [22] Arkin, A., et al. 2002. Web services choreography interface WSCI. <http://www.w3.org/TR/wsci/>.
- [23] Curbera, F., et al. 2002. Business process execution language for web services. <http://www-106.ibm.com/developerworks/webservices/library/ws-bpel/>.