Did you Find Any Public Web Service Description over the World Wide Web?

Asif Sohail Abid, Atif Sohail Abid, Muhammad Younus Javed, and Adnan Zahid

***Abstract*—Web service is an implementation of Service Oriented Architecture. It is interoperable, loosely coupled through the components over the internet, and widely used by the consumers. A web service interface is written in WSDL. To understand about the details of a web service consumer uses it description. This description is in structural and non-structural form. The structural description tells the consumer what a service look like, whereas non-structural description helps to understand about the functionalities of the web services. The number of web services in public web service directories are decreasing day by day then the numbers of web services crawled by the search engines. Moreover many public web service that don't have their descriptions and it is also alarming that a large percentage do not contains more then 1K words in their descriptions. To understand a web service there may be other related web pages over the World Wide Web that contains the information about the service. In this paper, the non structural descriptions of a web service are focused. We have tried to search the information about a web service by passing different set of queries to the search engines. The related pages that are collected from the www are ranked, this result is shown to the consumer after filtering at a specific threshold value.**

***Index Terms*—WSDL, web services, public web services, web services descriptions.**

I. Introduction

There is a significant increase seen in web services over the internet world. The Service-oriented architecture (SOA) is a flexible set of design principles used during the phases of systems development and integration in computing. A system based on a SOA will package functionality as a suite of interoperable services that can be used within multiple, separate systems from several business domains. A computing paradigm that is driven by SOA is Service-oriented Computing (SOC) [1].

A web service is an implementation of SOA that possess an interface that supports interoperability between different devices using standard protocol [2]. A web services is for interoperability, and are loosely specified. They are coupled by components present on the internet. They hold a purpose of being used far and wide by suppliers, customers, and business partners [3]. The web services interface is written in machine processable format, it has an interface described in Web Services Description Language (WSDL). To interact with a web service the description of the web service is consider by the consumer want to do so. This interaction is made using Simple Object Access Protocol (SOAP) message. This message is typically transmitted using HTTP, with an XML serialization including other web standards.

A model for describing non semantic Web services is provided by WSDL, it is an XML based language. The WSDL is also used to describe the functions provided by a web service. Web services are described by different component. For example data containers used in message exchange is shown by <type> element; the transmitted information type is shown in <messages>. It normally contains parameters that are passed to the web service. These parts are associated with a type definition. The port type are used in WSDL for describing operations (functions) that can be performed by the Web service, binding defines the communication protocol, and services describes data format for each operation , and it also aggregates multiple related functions respectively. To utilize a web service for a specific need the system consumer will need to go through following tiresome steps [4].

A. Finding a Web Service

If the web service required is not already known to the consumer, then the consumer will go for the Public Web Services Directories PWSD, or he may use some Public web Service PWS discovery approach.

B. Understanding the Service Using Its Description

The descriptions of the web service will tell the consumer about the details of it. WSDL file is plays an important role in understanding it, as it contains the descriptions of the web service. These descriptions can be categorized into two types

1) Structural

It contains declaration, message type, transmission protocol and location of the service, which can help the consumer to understand 'what it looks like' but he cannot know 'what are the functionalities it provides'.

2) Non-Structural

Whereas on the other hand non structural descriptions of services are the text, that are written in some native natural language and embedded in the WSDL. This description tells the consumer that what are functionalities that a specific service contains ad how he can use them.

C. Integration of the Service

When a consumer has discovered a required service, and it's usage and functionalities are cleared then he may integrate the service into the system.

Amazon, Google, and Yahoo are providing a large number

Manuscript received January 18, 2013; revised March 20, 2013.

Asif Sohail Abid and Adnan Zahid were with Faculty of Computing, Riphah International University, Islamabad, Pakistan (e-mail: Asif.sohail@riu.edu.pk, adnanzahid2011@yahoo.com).

Atif Sohail Abid and Muhammad Younus Javed are with College of Electrical and Mechanical Engineering, National University of Sciences and Technology, Rawalpindi, Pakistan (e-mail: atifsohailabid@yahoo.com, myjaved@ceme.edu.pk).

of web services and they are not publishing those services over the public registries or brokers, under the light of their decision to do so [5]. But public service registries cannot be ignored in an entire manner, a central registry will continue to be universally. Moreover the public web services registries do not meet the terms of UDDI specifications. These registries can be used by invoking online by the consumer [6]. Searching the descriptions of the services using some search engine crawlers is also focused by different researchers. To look at the effectiveness of the public service registry available, one should read between the lines to the work carried out by the researchers in recent past. The work carried by Eyhab *et al*. provides fair evidence of services that are available. The numbers of services in public registry are decreasing then services crawled by search engine's crawlers. 53% of the UDDI Business Registry (UBR) registered services are invalid. An due to several reasons public UDDI business registry working as major mechanism over the internet has been shut down permanently.

The work carried out by Lijie Wang et al. shows that 50% of the services they have studied do not have non-structural descriptions and this conclusion is also presented by different other researcher [7], [8]. So if a service is discovered but there is no description found regarding how to use the web service, then the consumer will face a hard time for understanding it or he may be conceived for the functionalities of the service. Moreover, there is just a brief summary of the functionalities for the service available in its non structural part. Consider the following two web services as example.

If we want to use a web service claiming to return digital weather forecast data of National Weather, *http://www.weather.gov/forecasts/xml/SOAP_server/ndfdXMLserver.php,* a consumer will look for its wsdl file for its description to understand, how to use it? And there is a fair description of the web service that can be found on the wsdl file for the service i.e. *http://www.weather.gov/forecasts/xml/DWMLgen/wsdl/ndfdXML.wsdl,* the web service description contains almost 534 words that can help the consumer how to use it, and the consumer is in a position to know functionalities provided by the web service. But if we consider a similar example as Lijie *et al*. has referred in his work. *http://www.webservicex.net/globalweather.asmx,* claiming to provide the weather report of all major cities around the world, looking on the *wsdl* file of the service on the link *http://www.webservicex.net/globalweather.asmx?wsdl,* We will just find this non structural text *"Get all major cities by country name(full / part). Get weather report for all major cities around the world".* This information is still insufficient for a consumer to tell that how he will use the service, either it is free or for commercial use and other similar information.

In this paper we focused our work in finding the descriptions about the web services written in non structural form. We have created an Active-X control and web interface as well for the same problem. It allows the user to input the name of the service then an attempt to extract the information present on different forums, and web pages that refer to explaining the functions of the web services are retrieved. After retrieval of pages the required data is extracted from the received information. As search engine don't provide the search of the WSDL document in particular, we have proposed a scheme for searching WSDL over the internet. This is discussed in detailed in section IV of the paper.

The rest of the paper organization is as follows. The Section II presents the related work to the problem, in this section the WSDL validity and information in WSDL is also discussed in detail. Section III contains motivation of the work. Section IV describes the proposed scheme for finding description of the PWS over the web, Section V contains implementation details, evaluation and results of the proposed work and in last section conclusion and future work is discussed.

II. Related Work

Web services gained the interest of many researchers in recent past. To fulfil the consumer requirement, many researchers dedicated their efforts on web services. A need of web service description language to the programmers is highlighted in Charles Petrie in his work. This need of description language for a programmer is an essence to see that what a software component written a web service can do and how they will use it [9].

Yan li *et al*. in their work on the public web services on the internet studied number, complexity, quality, and function diversity of the services on the internet. Transformation of web of data to web of semantics and services is continuing [10]. A web service mining framework is proposed by George *et al*. service compositions to emerge automatically in bottom up fashion is allowed by this framework.

Pat. P. W. Chan provided an algorithm for building reliable web services. The web services are described by WSDL (web service describing Language) and their interaction with other services are described by Web Service Choreography Interface (WSCI). The performance evaluation, verification of deadlock free web services by Petri-Net modelling are their contribution towards web services [11].

The challenges in discovering web services are address by Eyhab Al-Marsi et. al., a web service relevancy ranking function is presented in their study. This function can be use in finding best available web service in web discovery process [12].

III. Motivation

The researcher work for finding the descriptions of the WSDL and enriching the information about the web service from the web has raised many interesting issues [13]. Releated to our work we are addressing two important issues from them. These issues are the motivation for our work.

A. Validity of WSDL

A web service is described by WSDL file; this file contains structural or non structural description of a web service. These descriptions help the user to understand what a service look like and what a service can do and how to use this service. Lejie *et al*. shown in their study in enriching the descriptions of public web services from the information captured from different web pages that most of the web services do not contain detailed descriptions in their WSDL files, and in most

of the public web services the information given in the WSDL files are too few that it will never help the user to understand what a service can do for him actually. The same results were shown by Eyhab *et al*. in their study. They have focused UDDI business registries that are used for publishing and discovering the public web services and search engines for finding the WSDL. Fig. 1 shows a graph of the results given by Lejie *et al*. and Eyhab *et al*.

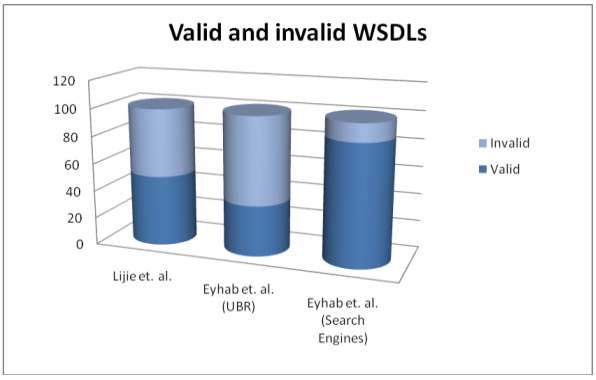

Fig. 1. Valid and invalid WSDLs

According to Lejie *et al*. less then 50 % are valid WSDL, whereas Eyhab *et al*. searched web services from UBR (UDDI Business Registries) that is used for publishing and discovering Web services into registries contains almost 75% invalid WSDS and 87% were validated from the collected WSDL through search engines.

Most of the web services existing today are having invalid or no WSDL files so the developer of the user can find them but he is unable to know about their functionalities and purpose of existence.

B. *Information in the WSDL Files*

The information that is present in valid WSDL is very less. Firstly there are low numbers of valid WSDL available and secondly the information present in these WSDL is unable to explain the functionalities of the web service. The Fig. 2 shows a graph that can be use too understand this statement in light of the work done by different researchers in recent past. It is clearly demonstrated that the information available in almost 72-83 % of valid WSDL is less the 1KB.

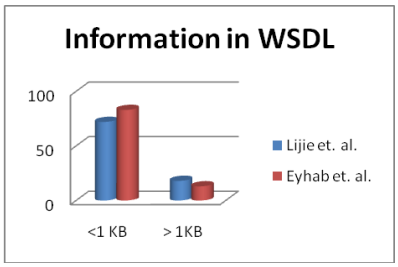

Fig. 2. Amount of Information in WSDL

IV. Proposed Scheme

In this section the proposed scheme is described for collecting the description about the public web service over the web. As information about a Public web service is an essence for the consumer to know how to use it, and what are the functionalities it will provide to the consumer. Moreover these information if not available in the *wsdl* then it will produce a reason for the consumer to search it on the web, and most of the search engine like google, yahoo etc don't provide specific *wsdl* searches. Keeping this in view, we have designed different queries for a entered key word in the interface and passed it to search engines, not only the search engines were considered for searching the description of the public web services, but we have also searched different Public web services directories through different API's available in open source languages i.e. Php. The queries that were taken in consideration contains the exact name, url of the service and synonyms from dictionary if exist. To limit our search we just considered top 15 retrieved results from them. These results are further refined to get a better result. As shown in Fig. 3, there are four major steps involve in the proposed scheme for finding the description of the web service.

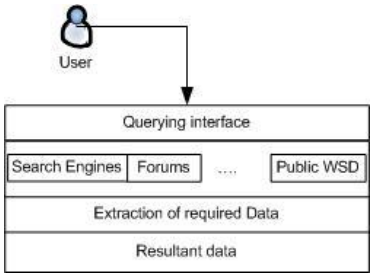

Fig. 3. Proposed Scheme

A. *Querying Interface Layer*

The user query the system using an interface named as "querying interface", this interface accepts the input parameter. The input parameter from the user is passed to underlying layer of search engines, forums and Public Web Services Directories (PWSD). A dictionary is also attached with the interface that will check for any synonym of the word if available, and pass this word to underlying interfaces. But name to most of public web services are given in such a way that they have no synonym.

B. *Search Engines, Forums and PWSD Layer*

For the testing purpose we have included google.com, bing.com, yahoo.com, and ask.com in our search engine list on which the queries were passed. We have use different API to retrieved result from them and the required data is extracted from the result. Different websites and forums are also checked for the description about the web service. Here is an algorithm that is used to pass information to yahoo.com; the similar algorithm is used for ask.com, bing.com, and google.com

Algorithm_search(q)
1. Pass the query to search
2. retrieve pages from search
   a. do{
   b. Get the url of the pages
   c. Get the text from the url's
   d. Remove the decorating tags and extract the data using DOM parser
   e. }while(no_of_pages!=0)

Fig. 4. Searching Algorithm

C. *Extraction of Data*

The required information is extracted from the data received from the searches made on web, but this relevancy is

assigned after removing any redundancy of the links from the different searches. As in many cases the search engines result was near or almost same. The extraction of the data is followed by following step

```
Relevency_algo(p)
1. do{
2. Check the page for relevancy
2. Store the relevancy calculated
3. }while(p!=0)
```

Fig. 5. Relevancy Algorithm

D. Resultant Data

The received data from these the pages are now assigned a relevancy number this relevancy is required so as to get. To calculate the relevancy of the segments following formula is considered.

$$R(s,d) = \frac{sim(s,d)}{\sqrt{Dis\tan ce_{d,s} + 0.5}} \quad (1)$$

As Lijie Wang *et al.* [14] describe, d as the relativity and target web service is s, the distance between s and d is represented as Distance d, s, sim(s, d) represents the similarity and 0.5 is use as similarity factor. The pages with a higher relevancy numbers are retrieved from the segment.

V. Evaluation and Results

The proposed system was implemented in php. Apache as a server and php as server scripting language was used. We have used different open source available Php DOM classes.

For the testing purpose, the attempt to search different public web services names to find out the description of these services over web through different search engines was made. The top fifteen results were taken from each search engine and the results were checked for redundancy, and the repeated links were removed in case if any repetition was found. The final resultant set contains pages that are named as candidate pages. These candidate pages were ranked according to their relevancy with the Public web service name using the formula in section IV. The pages with high relevancy from a threshold were selected and the result was taken. After this step the related pages were identified.

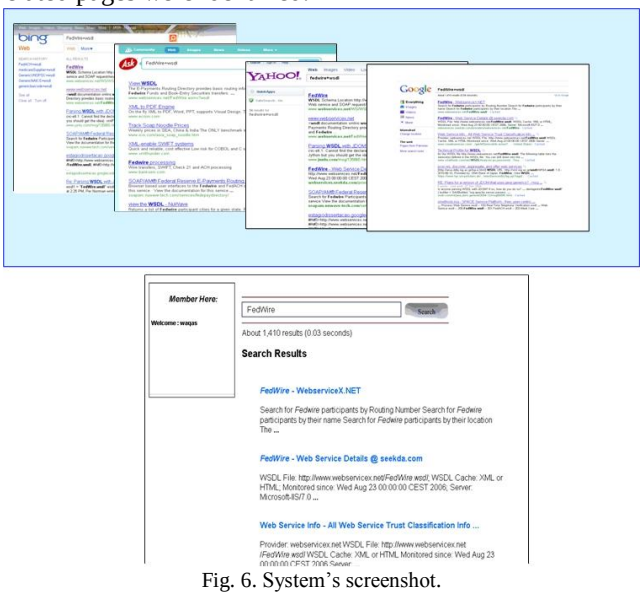

Fig. 6. System's screenshot.

In figure result of the query is shown. We have also shown the response of the same query to google, bing, yahoo and ask.com.

The average result of some queries made to the proposed system is shown in the Table I.

TABLE I: Average Results

	Total candidate pages	Total extracted	related /not related	No's	Percentage
Business (5)	34	14	Related	4	28.6%
			Not Related	10	71.4%
weather(5)	35	12	Related	6	50.0%
			Not Related	6	50.0%
sms(3)	27	16	Related	9	56.3%
			Not Related	7	43.8%
forecast(4)	15	8	Related	6	75.0%
			Not Related	2	25.0%
amazon(4)	18	9	Related	6	66.7%
			Not Related	3	33.3%
utilities(5)	20	11	Related	7	63.6%
			Not Related	4	36.4%
other(4)	15	9	Related	6	66.7%
			Not Related	3	33.3%

To confirm that a given result from the proposed system is related to the Public Web service searched we have manually checked for its relatedness. The graph shows the number of related and non related pages from the candidate pages retrieved as a result for the public web services searched on the proposed system. The average result for related was 58.1 % and 41.9 % were not related to the keyword searched.

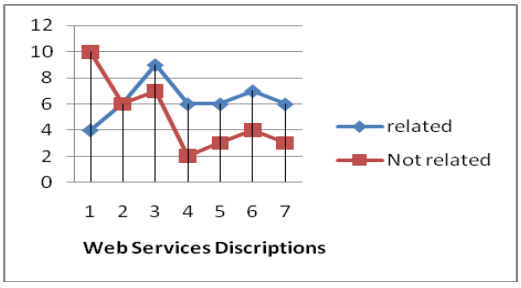

Fig. 7. Related information retrieved.

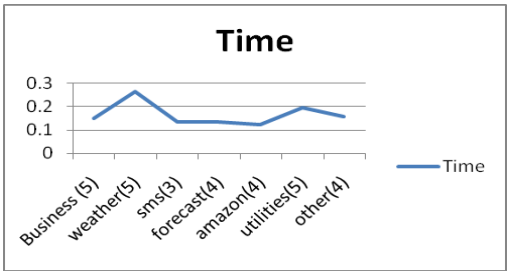

Fig. 8. Average time for quries.

Fig. 8 shows the graph that contains average time for

different queries on the proposed system. The time in the graph is taken in seconds.

VI. Conclusion and Future Work

Keeping in mind the importance of the web service description still web service descriptions is either missing in many of the services or it is too little to understand the functionalities of the web services. The consumer consults different forum, search engines or web sites to get information about public web service. There is a need to improve the search queries that are made.

References

[1] S. Vinoski, "Integration with web services," *IEEE Internet Computing*, vol. 7, pp. 75–77, November 2003.

[2] B. Sleeper and B. Robins. (2001). Defining web service. [Online] Available: http://www.perfectxml.com/Xanalysis/TSG/TSG\DefiningWebServices.pdf.

[3] K. Elgazzar and A. E. Hassan, "Clustering WSDL documents to bootstrap the discovery of web services," in *Proc. IEEE International Conference on Web Services*, 2010.

[4] E. A. Masri and Q. H. Mahmoud, "Investigating web services on the world wide web," *International World Wide Web Conference (WWW 2008)*, pp. 795-804, 2008.

[5] D. Chappell, *Who Cares about UDDI*, New York: Addison Wesley, 2002.

[6] C. Wu and E. Chang, "Searching services on the web: A public Web Services Discovery approach," in *Proc. Third International IEEE Conference on Signal-Image Technology and Internet-Based System*, sitis, pp. 321-328, December 2007.

[7] C. Roberto.Web Services Description Language (WSDL) Version 2.0. [Online]. Available: http://www.w3.org/TR/wsdl20/, 26 June 2007.

[8] A. M. Eyhab and H. M. Qusay, "WSCE: A crawler engine for large scale discovery of web services," in *Proc. International Conference on Web Services, ICWS*, pp. 1104-1111, 2007.

[9] C. Petrie, "Practical web services," *Internet Computing, IEEE*, vol. 13, issue 6, pp. 87-89, November, 2009.

[10] G. Zheng *et al.*, "Service mining on the web," *IEEE Transactions on Services Computing*, vol. 2, no. 1, pp. 65-78, January-March, 2009.

[11] Pat. P. W. Chan and M. R. Lyu, "Dynamic web service composition: A new approach in building reliable web service," *IEEE Computer Society*, AINA, pp. 20-25, 2008.

[12] E. A. Masri and Q. H. Mahmoud, "Discovering the best web service," in *Proc. 16th International Conference on World Wide Web*, pp. 1257-1258, 2007.

[13] L. Yan, L. Yao, Z. Liangjie, L. Ge, X. Bing, and S. Jiasu, "An exploratory study of web services on the Internet," in *Proc. International Conference on Web Services, ICWS*, pp. 380-387, 2007.

[14] L. Wang *et al.*, "Enriching descriptions for public web services using information captured from related web pages on the internet," in *Proc. 5th IEEE International Symposium on Service-Oriented System Engineering (SOSE)*, pp. 141-150, 2010.

Asif Sohail Abid was born in Karachi on 14th July, 1983. He is a PhD Scholar in National University of Sciences and Technology, Islamabad. He has done Master's in Computer Software Engineering in 2012 from the same institute.

He is working as a lecturer in Faculty of Computer, Riphah International University, Islamabad. He has three research publication in the area of Web Engineering.

Mr. Asif is also a member of IEEE Computer Society.