



**HAL**  
open science

## Multiple Representations of Web Services: Discovery, Clustering and Recommendation

Mustapha Aznag, Mohamed Quafafou, Nicolas Durand, Zahi Jarir

► **To cite this version:**

Mustapha Aznag, Mohamed Quafafou, Nicolas Durand, Zahi Jarir. Multiple Representations of Web Services: Discovery, Clustering and Recommendation. IEEE 9th International Conference on Web Services (ICWS 2011), 2011, Washington DC, USA, Unknown Region. pp.748-749. hal-01465121

**HAL Id: hal-01465121**

**<https://hal.science/hal-01465121>**

Submitted on 18 May 2018

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Multiple representations of web services: discovery, clustering and recommendation

Mustapha AZNAG, Mohamed QUAFAROU, Nicolas DURAND  
*Laboratory of Information and Systems Sciences (LSIS)*  
*Aix-Marseille II University, France*  
*Email: firstName.lastName@univmed.fr*

Zahi JARIR  
*LISI Laboratory, FSSM,*  
*University of Cadi Ayyad, Morocco*  
*Email: zahijarir@ucam.ac.ma*

**Abstract**—This paper analyses several web services representations based on web services descriptions and more generally on the content of WSDL files. We introduce also a new representation called symbolic reputation which is computed from relationships between web services. Different implementation issues are discussed and the results considering real world web services are analysed to determine the usefulness of the introduced representations for three main tasks: web services discovery, clustering and recommendation.

**Keywords**-reputation; representation; WSDL; semantic tagging;

## I. INTRODUCTION

Web services use natural language descriptions (ambiguity of sentences), they are multi-languages and cross-domains. In this context, the representation of web service becomes a major challenge. This paper studies different classes of web services representations and compares them considering the following three important tasks: discovery, clustering and recommendation.

## II. WEB SERVICES REPRESENTATIONS

We have considered the following representations :

- Baseline representation (B) : A web service can be described by a textual description extracted from WSDL document or given by his provider when publishing in the UDDI. The current UDDI registry only allows to search web services by their textual description. We consider this representation as a baseline representation for a web service. Let us remark that the major disadvantage of this representation is that most web services have an poor or an empty description.
- Functions descriptions integration (BFD) : The goal of this representation is to enrich the baseline representation of a service by adding the name and the textual description of all its operations. We apply a textual processing to obtained the resulted representation. Let us remark that such descriptions may be dirty and generally have different sizes. We need more cleaning process to improve such representation.
- WSDL types integration (WBFD) : To complete the second representation, we added all information described by WSDL types. The types are used by messages to transmit informations between web services.

Consequently, WSDL types are good features to describe the functionality of a service and are the most informative element in WSDL document. For this reason, we extract all type names (elements, complex types, simple types, attributes, enumerations) from the database and we apply the textual processing to produce a set of words. Thus, we use the obtained set of words to construct the new representation (WBFD).

- Rules based text tagging of WS descriptions (Rbfd) : The approach consists in defining of extraction rules to identify, extract and annotate relevant multi-word terms from web service descriptions. The processing steps (tokenization, Part-Of-Speech tagging, extraction and output generation) of information extraction have been implemented as modules using the LinguaStream platform<sup>1</sup>. LinguaStream [1] is an integrated experimentation environment targeted to researchers in natural language processing. Let us note that we used TreeTagger<sup>2</sup> for the Part-Of-Speech tagging step. The extracted information is given in a form of XML file.

## III. SYMBOLIC REPUTATION MODEL

### A. Web services dependency graph

Web services dependency graph was used to represent the structure of the web services space, i.e. web services relationships, by generating a network depending on the I/O of each web service. This induced network is then used for web service composition [2]. The general approach consists of searching first for web services operations dependencies using the I/O parameters of their operations, then the web services dependency graph is induced. For seek of simplicity, we consider here that a dependency occurs between two services when they offer at least two dependent operations: a service requires (resp. provides) data from (resp. to) another service. Of course, other dependency definitions can be used without affecting our general symbolic reputation computing approach.

<sup>1</sup><http://www.linguastream.org>

<sup>2</sup><http://www.ims.uni-stuttgart.de/projekte/corplex/TreeTagger/>

Categories	B		BFD		Rbfd		Wbfd		SR	
	P%	R%	P%	R%	P%	R%	P%	R%	P%	R%
Weather	82,61	54,29	85,71	51,43	81,82	51,43	68,97	57,14	68,97	57,14
Address Information	99,09	36,21	95,69	36,88	94,64	17,61	91,60	36,21	96,67	19,27
SMS	68,18	75,00	65,00	65,00	55,56	50,00	48,48	80,00	12,12	80,00
Currency Exchange	69,23	50,00	55,56	55,56	58,33	77,78	16,67	16,67	08,51	22,22
Stock	97,60	85,92	97,56	84,51	94,67	50,00	98,28	80,28	94,12	90,14
Average	83,34	60,28	79,90	58,68	77,00	49,36	64,80	54,06	56,08	53,75

Table I  
WS DISCOVERY FOR EACH REPRESENTATION (PRECISION(P) AND RECALL(R) IN %).

### B. Random walks models for reputation computing

A random walk is a sequence of nodes in graph constructed by the following process: select a random starting node in the graph, then move to select a neighbour of this node at random and so forth. The random walk analysis have been applied to different fields as a fundamental model for random processes in time . More particularly, a formal model was proposed in [3] to allow the computation of the reputation of a web page using the web hyperlink structure. We have adapted this model to compute the symbolic reputation of web services.

We consider the dependency graph constructed in Section III-A. The probability of visiting a service  $s$  for topic  $t$  at step  $n$  of the walk, denoted  $R^n(s, t)$ , is defined as follows:

$$R^n(s, t) = (1 - d) \sum_{q \rightarrow s} \frac{R^{n-1}(s, t)}{O(s)} + \begin{cases} \frac{d}{N_t}, & \text{if } t \in \Omega(s) \\ 0, & \text{otherwise} \end{cases} \quad (1)$$

The previous probability is the basis of the algorithm that compute the symbolic reputation. In fact, suppose that with the probability  $d$  the random walker jumps into a service uniformly chosen at random from the set of services that contains the term  $t$ . In this context, the probability that a random walker visits a service  $s$  in a random jump is  $\frac{d}{N_t}$  ( $N_t$  denotes the total number of services on the UDDI that contains the term  $t$ ) if the vectorial representation of service  $s$  contains the term  $t$  and it is zero otherwise. The probability that the walker visits the service  $s$  at step  $n$  after visiting one among its parent service  $ss$  is  $\frac{1-d}{O(s)} R^{n-1}(q, t)$  where  $R^{n-1}(q, t)$  denotes the probability that the walker visits the service  $ss$  for topic  $t$  at step  $n - 1$  and  $O(s)$  denotes the number of services that require data from  $s$  (parents of  $s$  in the dependency graph).

## IV. EXPERIMENTAL RESULTS

We have considered different web service sources like WebservicesX.net<sup>3</sup>, xMethods.net<sup>4</sup> and seekda.com<sup>5</sup>. Our general Meta-UDDI architecture allows to use of different UDDI which are used through specific wrappers. We have collected around 8,500 services. Unfortunately, these services are not classified into categories. This is a real

<sup>3</sup><http://www.webservicesx.net/ws/default.aspx>

<sup>4</sup><http://www.xmethods.net/ve2/index.po>

<sup>5</sup><http://www.seekda.com>

problem when we want to evaluate the usefulness of our representations using precision and recall measures. For this reason, we have considered a new web services source, i.e. service-finder.eu<sup>6</sup>, which classifies its services using an ontology. The number of services which are classified into a category by service-finder is only reduced to only 1,647 web services. However, 465 services among them are not available: they are discarded. Moreover, the collected web services are multi-languages. We have automated the recognition of the language and keep only those using English language. Finally, only 993 web services are used. Let us denote 'Collected WS' the number of service collected for each category, 'Available WS' the number of available services and 'Used WS' the number of services used for each category after the process of recognition of the language. See Table I for details the results obtained using each representations and differents categories of web services.

## V. CONCLUSION

This paper revisits the traditional representations of web services using their textual descriptions and/or the elements of the WSDL structure as types, attributes, etc. We have proposed two new web services representations. The first one is based on semantic tagging of web services descriptions to keep only the most significant parts. Whereas, the second representation, i.e., symbolic reputation, is more contextual as it considers not only the service itself, but its neighbours.

## REFERENCES

- [1] Bilhaut, F. and Widlcher, A. *LinguaStream: An Integrated Environment for Computational Linguistics Experimentation*, the European Chapter of the Association of Computational Linguistics, 2006, Trento, Italy
- [2] Abrehet Mohammed Omer, Alexander Schill, *Dependency Based Automatic Service Composition Using Directed Graph*, Next Generation Web Services Practices, International Conference on, pp. 76-81, 2009 Fifth International Conference on Next Generation Web Services Practices, 2009.
- [3] Davood Rafiei Alberto O. Mendelzon *What is this Page Known for? Computing Web Page Reputations*, 9th International World Wide Web Conference, Amsterdam, 2000.

<sup>6</sup><http://www.service-finder.eu/>