

Advanced Profile Similarity to Enhance Semantic Web Services Matching

<http://dx.doi.org/10.3991/ijes.v1i1.2963>

Z. Chouiref¹, A. Belkhir² and A. Hadjali³

¹University Colonel Akli Mohand Oulhadj, Bouira, Algeria

²University of Sciences and Technologies Houari Boumediene, Algiers, Algeria

³Laboratory of Informatique and Automation Systems - ENSMA, Poitiers, France

Abstract—In this paper, we present a fine-grained matching method of the services based on a hybrid similarity measure. We propose a novel encoding of the services descriptions, allowing the match between a request and an advertisement in order to make more efficient publishing and searching process of Web services and reduce the number of comparisons required. By this kind of similarity between concepts of profile, a precise matching method is developed to match the profile of the Web services and user. Searching process in the UDDI registry is done via an algorithm that allows us to extract the search concepts and retrieve the top-k services, thereby further reducing the search engine's response time. The approach is illustrated through some experiments both on real and synthetic data to demonstrate its consistency and effectiveness.

Index Terms—User Profile; Web Service Matching; Service Selection, Profile Similarity.

I. INTRODUCTION

The recent evolution of Internet, driven by the SWS (Semantic Web Services) technology, has extended the role of the Web from a support of information interaction to a middleware for B2B (Business to Business) interactions. Service orientation is a promising paradigm for offering and consuming functionalities within and across organizations in the Web. Indeed, semantic Web services allow a homogeneous use of heterogeneous software components deployed in large networks and in particular the Internet.

The discovery of the SWS is a previous process to their use. It constitutes the process permitting to find the most suitable set of Web services for a user request. It is essentially based on the syntactic research of the WSDL (Web Services Description Language) descriptions (Inputs, Outputs, Preconditions and Effects parameters) using the UDDI's (Universal Description, Discovery and Integration) registers.

However, with the exponential growth of available services, the diversity of users and the conditions under which they access Web services, finding the relevant SWS for particular users is becoming a challenging task. With Web 2.0 applications and particularly e-business and e-commerce applications, Web service discovery is becoming much more important in a Web context. Service computing tries to solve questions based on profiles of users from a contextual informational view of the Web where users have several characteristics such as the client terminal, the client preferences, its location, etc. All these parameters form a particular context of use called the profile.

In addition, the methods available in the UDDI publication do not contain a formal model describing the profile of services' users. Therefore service discovery could not be achieved efficiently without considering their profiles. Indeed, when a user requests a Web service discovery system, (s) he would have services tailored to its context and profile. Currently, the discovery problem can be formalized as a decision problem on the selection of services from a set of alternative services that provide the same functionality but differ in profile parameters. How to describe the information of SWS and matching process of SWS constitutes one of the main focuses of the research in service computing. It is thus necessary to propose a model which provides relevant and adapted results to the user profile. This model allows the representation of information that characterizes both the user and the service. The proposed search method is based on the use of a sophisticated similarity measure, which estimates the correspondence degree between the desired profile and the provided profile.

A user has many and diverse preferences according to the context, his/her interest domain, etc.; the question is how to take advantage of such user's information to improve the discovery process and better satisfy the user request? Suppose a person is looking for a hotel in Italy, current search engines provide a list (which can be huge) of all the hotels, but the question remains which one to choose? To solve this problem, we are leveraging the profile information in a Web services discovery system during the search and selection stages by selecting services corresponding as well as possible to the user request. This selection is based on the use of similarity measures, which estimate the matching degree between the desired profile and the provided profile parameters.

Our main contributions are summarized as follows: (i) A novel model of specification of the profile as well as a setting for the description of Web services are proposed, allowing an efficient match between a service request and a service advertisement, (ii) a Web service discovery architecture including the profile dimension is described, (iii) an hybrid similarity measure for the management of profiles and the rank-ordering of the retrieved services is investigated. In addition, a top-k services building algorithm is provided.

In the remainder of this paper, we first provide, in Section II, a critical overview of existing works related to the SWS discovery issue. Building upon these works, an approach towards efficient SWS discovery and selection is discussed in Section III. Section IV provides an experimental study to show the relevance and suitability

of our proposal, whereas the conclusion and future works are presented in Section V.

II. RELATEDWORK AND ANALYSIS

Users and software entities should be able to discover, invoke, compose, and monitor Web resources offering particular services and having particular properties. They should also be able to do so with a high degree of automation if needed. As the number of services on the Web constantly increases, efficiency and scalability of service discovery techniques become then a critical issue. On the other hand, several approaches have been proposed for enhancing the SWS discovery. Some of these approaches are syntax based [1], while other are semantic based [4][6]. Recently, other approaches, reviewed below, are distinguished by the fact that they rely on QoS parameters [9], context information [10], etc.

The predominant problem is the restrictions posed by keyword matching that do not allow retrieval of SWS with similar functionality. These works [4,6] have focused mainly on providing means to describe the functionality of a SWS and to allow a very expressive language for querying services. This development is increasingly significant since it seems to be able to overcome some insufficiencies of the syntactic approaches and tackle some of the UDDI register inadequacies. However none of these papers discusses in depth how a publisher/requester should provide data related to her/his context. We can say that each of the two entities, user and Web service, has its own context [10]. The service context can group the service localization (geographical restriction), the implementation cost, the quality of services (QoS) parameters, etc. The user context can be formed of her/his localization, her/his preferences, etc.

The study of [15] focuses on QoS in discovery systems. The service consumer searches UDDI registry for a specific service through discovery agent which helps to find best quality service from available services which satisfies QoS constraints and preferences of requesters. Context-Awarenessas proposed in [8] performs the necessary changes in the service behavior and/or the data handled in order to adapt the service to the context of the each user. Rong *et al.* [14] suggest with an example that context should be domain oriented or problem oriented in Web services discovery systems. They divide context in two categories as explicit and implicit, with Personal profile oriented context, Usage history oriented context, Process oriented context and other context. Chukmol *et al.* [7] propose the personal opinion on service functionality and quality or invocation cost should also be considered by collaborative tagging-based environment for Web services discovery. The study done in [5] has proposed a novel approach to enhance Web services discovery based on, among others, QoS, customer's preferences and past experiences. The work in [19] presents an alternative approach for supporting users in Web services discovery by implementing the implicit culture approach for recommending Web services to developers based on the history of decisions made by other developers with similar needs.

Let us note that SWS properties include several parameters like the functional (IOPE, functionalities, etc.) and non-functional parameters (QoS, property that identifies the technical standards or protocols for implementing services and categorization). However, the majority of suggested approaches focus only on some parameters like: QoS, localization, etc. The information should have more user centric presentation in the discovery system. Multiple Web services with the same kind of functionality may be available in different contexts; best service among them should be selected. This can be done using profile parameters.

Moreover, few works took into account multiples qualitative and quantitative parameters to help users to find the best service during the discovery process. It is also worth noting that the existing researches in SWS discovery either use information retrieval techniques or semantic-based methods for locating Web services. One of the big problems of Web search is the definition of a correspondence function between the representation of the proposed service and the user request. In order to solve such problem, we propose a new approach to improve the SWS automatic discovery. In particular, we provide a descriptive common form for both service and user profiles and introduce a novel similarity measure to match the user profile and the one of the available service.

III. A PROFILE SIMILARITY-BASED SEMANTIC WEB SERVICES DISCOVERY APPROACH

The service discovery engine should provide rich and machine-processable abstractions that enable to describe service properties and profile as well as the specifications of user needs. Furthermore, discovery requires a matching process that compares the advertisements with the requests to verify whether they describe matching profiles. This is essential to enable the development of reasoning mechanisms to handle the discovery process. In this section, we will describe the service, the user and the profile formalism, and the proposed architecture. Finally, we will show how a new profile-based similarity measure can be used to enhance the discovery process.

A. Formalisation

The discovery process requires a formalism that can be used to encode Web service profile for advertisement and for requests. Before describing the concepts of our model, we first give some formal definitions.

DEFINITION 1. A SEMANTIC WEB SERVICE IS DEFINED AS A QUINTUPLE: $\{N_s, D_s, P_s, O_s, PP_s\}$, WHERE:

- N_s is the name of the service.
- D_s is the functional description of the service.
- P_s is a set of parameters describing the service.
- O_s is the set of operations of the service.
- PP_s is the set of concepts composing the profile of provided service.

Each operation is associated with an input list $I=\{i_1...i_n\}$ and an output list $O=\{o_1...o_n\}$. Services are also described by preconditions and effects. The "IO" (Input, Output) model for service operations is used in our model as a first step towards SWS discovery.

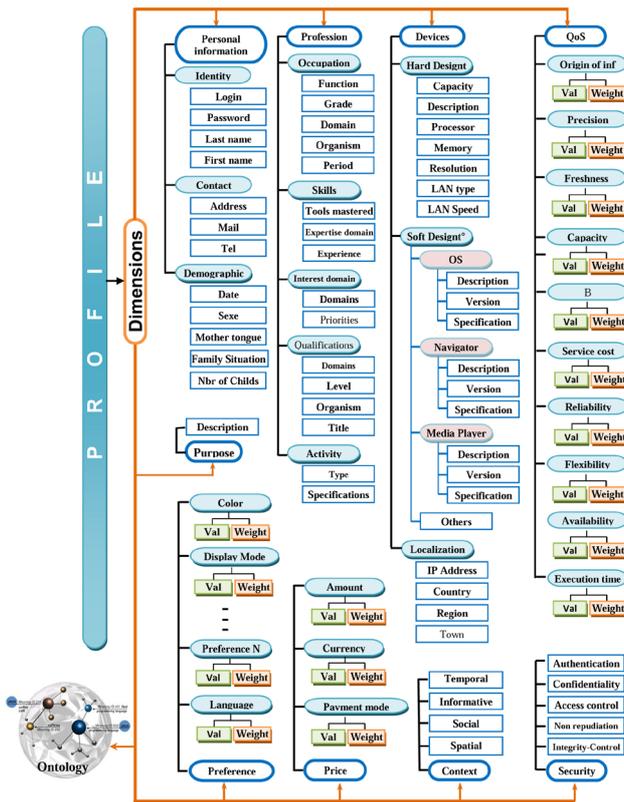


Figure 1. Profile model of the service and the user.

DEFINITION 2. A REQUEST IS DEFINED AS A QUINTUPLE: $\{N_R, I_R, O_R, PR_S, \square\}$, WHERE:

- N_S is the name of the request.
- I_r is a set of inputs parameters of the request.
- O_r is a set of outputs parameters of the request.
- PR_s is the set of concepts composing the profile of required service.
- \square ($0 < \square \leq 1$) is the interest score defined by the user.

DEFINITION3. EACH PARAMETER I IN THE LIST OF INPUTS AND EACH PARAMETER O IN THE LIST OF OUTPUTS IS DEFINED AS A QUADRUPLE $\{N_T, V_T, T, CO\}$, WHERE:

- N_T is the name of the attribute.
- V_T is the value of the attribute.
- T is the data type of the attribute value.
- CO is the concept of ontology which is connected to the attribute to improve its quality¹.

DEFINITION 4. PROFILE P IS MODELED THROUGH A SET OF CONCEPTS $\{C_1, \dots, C_N\}$ CALLED PROFILE CONCEPTS C_i , WHERE:

Each concept C_i is characterized by an infinitely countable set of values and weight ($\in [0, 1]$) associated to each parameter, i.e. $C_i = (\text{value}, \text{interest score})$. A service request or advertisement can be represented by the set of intervals, qualitative or quantitative values associated to

¹Some approaches use the notion of ontology to determine search context and user's interests [18][19].

its profile concepts. These are used to limit the set of the returned services that match the profile values specific to the user.

The information of the profile can be static (personal data...), evolutionary (preferences...) and temporary (localization...). These pieces of information must be captured to make correspondence between demands and offers of services, on the syntactic and semantic level. This improves the relevance of answers during a discovery session. The proposed model contains mainly several dimensions (Figure 1) that involve most information characterizing a profile. This general structure takes the form of a tree that includes a set of hierarchically organized concepts. The structure thus defined is flexible in the sense that different features can be spread through the tree structure of the description proposed. It allows modeling the user's profile soliciting the service as well as the semantic Web service's profile offered.

B. An Illustrative Example

To illustrate our proposal, let us take a tourist's case that likes to travel during vacations in Italy, (s)he tries to book a hotel{five stars} that possesses a {restaurant}, knowing that (s)he is a {smoker} and has a {car}. His/Her query can be defined as follows:

$Qu = \{Hotel, NumberStars, Region\}$.

1) Step 1. Simple search.

$Att = \{Name\ Of\ Service = Hotel \wedge Inputs = (NumberStars, Region)\}$, can be used for the case of a simple research of services. For instance, the results returned by this method are the hotels given in Table I, that correspond to the desired number of stars, and the hotels nearest of user's location which describes her/his current position.

2) Step 2. Profile-based advanced search.

The results returned by simple search don't really satisfy the user's needs (in particular, the needs related to the parking and a smoking zone). To improve the satisfaction of the results, one needs to use elements of service's profile (Hotel.Facilities) described in Table II, and user's profile which can be defined as follows:

User Profile: $Pu = \{Food(0.2), Smoke(0.2), Parking(0.6)\}$.

Result. Both hotels are five stars hotels and close to the user, although the price of the first hotel is better than the second, but the second hotel is more suitable to the user's profile.

C. Weighted Characteristics

It is possible that some characteristics are more important than others. Our similarity function uses different weights for each characteristic. It is obvious that the determination of the weights play an important role in the final result. Several methods have been proposed: the objective methods [16][17] for which the requestor is not solicited and methods for which values of weights resulting directly from the requestor's opinions.

A simple method is used in this paper that assumes weights are given by the user. In our example, on each user profile's attribute, is added a number between 0 and 1 to express the relative importance of this attribute with respect to the other. Thus the value 0.2 on the first two att-

TABLE I.
ATTRIBUTES OF THE FIRST AND THE SECOND HOTELS

Attributes	Value
Hotel.idH	182667
Hotel.Name	LittleItalyHotelNukuAlofa.
Hotel.Description	This hotel is located in center city at 2891 Vuna Road Kolomotua close to the hospital.
Hotel.Stars	5
Hotel.Price	152
Hotel.propertyType	0
Hotel.FreeRooms	⊥(not defined)
Hotel.Facilities	Facilities.IdF
Hotel.idH	365295
Hotel.Name	LittleItalyHotelRestaurant
Hotel.Description	This hotel is located at 50 km from center city and thesea at VunaRoad,Tongatapu,Tonga.
Hotel.Stars	5
Hotel.Price	153
Hotel.propertyType	0
Hotel.FreeRooms	22
Hotel.Facilities	Facilities.IdF

TABLE II.
PROFILES OF THE FIRST AND THE SECOND HOTELS RESPECTIVELY

Attributes	Value
Facilities.idF	1
Facilities.RoomService	Room Concierge
Facilities.Food	Dining Area/Restaurant
Facilities.Transport	Shuttle Service to/from Airport
Facilities.Touring	Tour and Information Desk
Facilities.Smoke	Smoking Zone
Facilities.Money	ForeignCurrencyExchange
Facilities.Parking	⊥ (not defined)
Facilities.Communication	Fax Services
Facilities.Internet	WiFi Internet
Facilities.idF	2
Facilities.RoomService	Room Concierge
Facilities.Food	Dining Area/Restaurant
Facilities.Transport	⊥ (not defined)
Facilities.Smoke	Room for Smoking
Facilities.Business	CenterforBusinessActivities
Facilities.Parking	Parking (Free)
Facilities.Kids	Children's Activity
Facilities.Internet	ClubBroadband Internet Access

ributes means that these conditions are of the same importance. The last attribute expresses the fact that the user has a stronger preference for hotels with a parking.

D. Towards a new discovery architecture

Our approach is supported by the architecture depicted in Figure 2. It is capable of integrating the profile into the process of semantic Web services discovery. It is composed of basic elements of SOA (Service Oriented Architecture)[3]: the service provider, the service requester and the service registry UDDI, to which we add an "Interoperability Module" and a "Discovery and Selection Engine".

1) *Interoperability module*: In order to achieve interoperability, our architecture is endowed with module which consists of the following:

a) *Administrator of the profile* :its tasks are: initialization, modification, search and consultation of the profile.

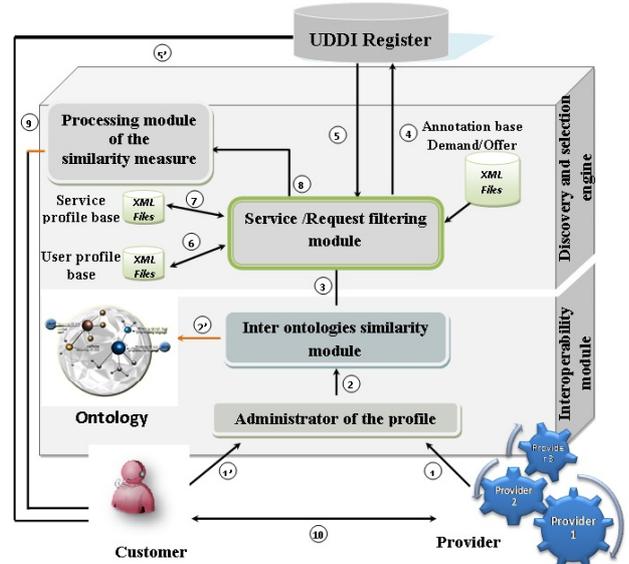


Figure 2. Proposed discovery architecture.

b) *Inter ontology similarity module*: It assures a semantically annotation of each parameter of the profile of the service and the request by means of an associated ontology O.

c) *A profile database*: to record the profile information.

2) *Discovery and Selection Engine*: The result of the query can be send directly to the customer through a simple search as in SOA, or via an advanced search through the following components:

a) *Service/Request profile filtering module*: The result of the request (5) is intercepted by this module that generates XML's files including profile descriptions of the preselected service as well as those of the request from files stocked respectively in the basis of the profile service and the basis of the profile user (6), (7) (see Figure 2).

b) *Similarity measure treatment module*: It aims at comparing different parameters of the profile related to Web services and the users (8) using a proposed similarity measure, and sent the best services to the customer (9).

E. A novel similarity measure

In this section, we first present some of the most-known semantic similarity measures that are used to compute the similarity between two concepts, and then we introduce our similarity measure.

1) *Overview of similarity measures*

Matching operation is recognized to be in a key of many applications such as ontology engineering, information integration (schema, catalogue, data, etc.), Web service composition, Web query answering, peer-to-peer information sharing [17]. The quality of matching results depends -highly- on the similarity measure that is used to estimate the resemblance between concepts of the matched entities. Several similarity measures were

proposed in various applications fields like [12] (intrusion detection), [13] (textual data)...etc.

a) *Path-based measures*

• **Shortest path length metric:** Rada et al. [20] defined the distance between concepts in an *is-a* semantic net as the length of the shortest path separating them.

$$d(c1, c2) = \text{minimum number of edges separating } c1 \text{ and } c2$$

Furthermore, Authors gave an example of a formula that transforms distance into similarity:

$$S(c_1, c_2) = [1 - d(c_1, c_2)]$$

• **Castano, Ferrara, Montanelli, and Raccas metric:**

Castano et al. [21] propose a similarity measure. It is given by the following formula:

$$sim_{ter}(t1, t2) = \begin{cases} \max_{i=1, k} \{W_{t1 \rightarrow_i^n t2}\} & \text{if } k > 1 \\ 0 & \text{otherwise} \end{cases}$$

where: k is the number of paths between $t1$ and $t2$ in Thesaurus Th; $t1 \rightarrow_i^n t2$ denotes the i^{th} path of length n ; $W_{t1 \rightarrow_i^n t2} = W_{1_{tr}} \times \dots \times W_{n_{tr}}$ is the weight associated with the i^{th} path, where $W_{j_{tr}}$ is the weight associated with the j th relationship in the path.

b) *Depth-relative measures*

As path-based measures have many drawbacks, many authors have presented metrics that capture more efficiency the semantic similarity between concepts.

• **Measure of Wu and Palmer**[22]

The length of a path between two concepts in a hierarchy is an intuitive measure to calculate similarity. This is a useful and easy to implement. The similarity of two concepts is defined by how closely they are related in the hierarchy.

$$sim(C1, C2) = \frac{2 * depth(C)}{depth_c(C1) + depth_c(C2)}$$

Where, $depth(C)$ is the length of path between C and the root of the hierarchy, $depth_c(Ci)$ is the number of arcs between Ci and root through C .

c) *Measure based on the interpretation of concepts*

D'amato et al.[23] proposed to measure the semantic similarity between two concepts described in logic.

$$S(C, D) = \frac{|(C \cap D)|^I}{|C|^I + |D|^I - |(C \cap D)|^I} *$$

$$\max\left(\frac{|(C \cap D)|^I}{|C|^I}, \frac{|(C \cap D)|^I}{|D|^I}\right)$$

Where $(.)^I$ is a function of interpretation and $|.$ is the cardinal of a set. This measure is interesting because it verifies the semantic properties such as the similarity between two concepts equivalent ($C \equiv D$) is equal to 1, the similarity between two concepts is not null.

2) *Our similarity measure*

A basic similarity measure is useful to define the resemblance rate between objects (profiles) composed of a set of attributes. In general, there are two different approaches for expressing these

attributes: a numeric (as price) and a categorical one (as colour).

In our work, we propose to improve the similarity measure $QSim$ defined in [11] and inspired by Jaccard [2] which operates only on the numerical value properties. We combine syntactic similarity with semantic similarity of each qualitative and quantitative attribute of request and service profile to calculate the overall score in order to recommend to user the best service that matches her/his profile. To evaluate the similarity between two profiles, we have to give a set of definitions:

- Let P be a set of objects' profile (users, documents, Web services, etc.).
- Let n be a set of services. Each service is composed of a set of concepts.
- Let x, y be a profile belonging to P for a given request and service respectively, and $(\omega_1 \dots \omega_n)$ is a set of weights associated to each concept of x and y where $\sum \omega_i = 1$.
- a is the set of common characteristics of x (requested by the client) and y (suggested by the service);
- b is the set of characteristics existing in x and not existing in y (requested by the client but not suggested by the service);
- c is the set of characteristics existing in y and not existing in x (suggested by the service but not requested by the client);

Jaccard Similarity Measure is defined as follows:

$$Jaccard(x, y) = \frac{a}{a + b + c} \quad (1)$$

Quantitative Similarity Measure is defined as follows:

$$QSim(x, y) = \frac{a \sum_{i=1}^{i=n} w_i \times sima(x_i, y_i)}{a \sum_{i=1}^{i=a} w_i + b \sum_{i=1}^{i=b} w_{a+i} + c \sum_{i=1}^{i=c} w_{a+b+i}} \quad (2)$$

Such as $sima$ is the atomic similarity between each characteristic of x and y . It is defined in a set IR^+ which can be modeled as follows (where \perp means not defined):

$$sima(x_i, y_i) = \begin{cases} 1 & \text{if } x_i = y_i \\ 0 & \text{if } x_i = 0 \vee y_i = 0 \vee x_i = \perp \vee y_i = \perp \\ \frac{\min(x_i, y_i)}{\max(x_i, y_i)} & \text{if } x_i \wedge y_i \text{ are quantitative values} \\ 0 & \text{if } x_i \wedge y_i \text{ are qualitative values} \end{cases} \quad (3)$$

Our Similarity Measure checks the properties of $QSim$ defined in [11]. It is formalized as follows:

- Assume that $Qi = (A1 \wedge \dots \wedge An \wedge Pu) / Pu = (C1 \dots C_m)$ be a conjunctive query of functional attributes $(A1 \wedge \dots \wedge An)$ and user profile attributes $(C1 \dots C_m)$.
- Assume that D is a dataset of Web services with qualitative and quantitative concepts $C = (C1 \dots C_m)$ and $D(C_i)$ represents the domain of values of concepts C_i in the dataset D . Each service is composed of a set of functional and profile concept as described in Section III.A.
- Let S_i be the set of returned services after submitting the query using functional attributes,

and R_{s_i} is the relevant services that match user profile for the given query.

- Let P_x, P_y be a profile belonging to P for a given request and Web services respectively, and $(\omega_1, \omega_2, \dots, \omega_n)$ is a set of weights associated to each characteristic of P_x where $\sum \omega_i = 1$.
- We define a threshold λ in order to present only services that have a rate of similarity with the user profile higher than λ .

In our work we focused on the calculation of the similarity between the user profile and the profile of returned services from the simple search (see example above). With these influential variables, we came to the following definition for the similarity measure of profile characteristics:

$$SimPro(P_x, P_y) = \left(\frac{a}{a+b} \right) \times \frac{a \sum_{i=1}^{i=a} w_i \times sima(P_{x_i}, P_{y_i})}{a \sum_{i=1}^{i=a} w_i + b \sum_{i=1}^{i=b} w_{a+i}}$$

(4)

Such as $sima$ is the atomic similarity between each characteristic of x and y . It is defined in a universe U which can be modeled as follows:

$sima: U \times U \rightarrow [0; 1]$

$$sima(x_i, y_i) = \begin{cases} 1 & \text{if } x_i = y_i \\ 0 & \text{if } x_i = 0 \vee y_i = 0 \vee x_i = \perp \vee y_i = \perp \\ \frac{\min(x_i, y_i)}{\max(x_i, y_i)} & \text{if } x_i \wedge y_i \text{ are quantitative values} \\ \text{Value in } [0,1] & \text{if } x_i \wedge y_i \text{ are qualitative values} \end{cases}$$

if $x_i \wedge y_i$ are qualitative value, $sima$ is calculated (in our example) by using the *Jaccard* coefficient (see formula (1)).

F. Algorithm

The proposed algorithm (Algorithm 1) tries to match a query with each advertisement in the repository. It takes a list of query's concept from the user's profile and a list of Web service's profile concepts as input in order to calculate atomic similarity $sima$ for each pair of concepts (of service and request). If the match is not a Fail, it returns the overall score $SimPro$ of each service and appends the advertisement to the result set. Finally the sorted result set is ranked and returned to the client.

IV. EXPERIMENTAL EVALUATION

A. Experimental setup

For testing the proposed approach and models, a preliminary experiment has been set up. The data used in our test are sampled from hotels base (CSV) downloaded from the link address: <http://api.hotelsbase.org/>. Hotelsbase is licensed under a *Creative Commons Attribution-ShareAlike 3.0 Unported License*. Among the files content, the facilities keys file that is used to describe each hotels Web service. These facilities involve in their profile part, a set of elements named category. We run a set of experiments for different parameters values:

- We consider about 100,000 hotels Web services with all properties and facilities. Properties can be: *longitude, latitude, priceMin, priceMax, starsMin,*

ALGORITHM 1. MATCHING ALGORITHM

Input: Request [], services [], tow concepts C1 and C2 /

$C1 \in P_x$ and $C2 \in P_y$.

Output: Atomic Similarity $sima(C1, C2)$, Overall Score $SimPro(P_x, P_y)$

Step 1: Get each service the atomic similarity value of each common attribute.

For S_1 to S_n **do** {For each $service_i \in service$ }
 $a = 0; b = 0;$

For P_{x_1} to P_{x_k} **do** {For each $Profile_j \in request$ }

For P_{y_1} to P_{y_k} **do** {For each $Profile_k \in service$ }

if $(C1, C2)$ quantitative **then**

if $C1=0 \vee C2=0 \vee C1=\perp \vee C2=\perp$ **then**

$sima(C1, C2) = 0; b = b + 1;$

else

if $C1 = C2$ **then**

$sima(C1, C2) = 1; a = a + 1;$

else

if $C1 \neq C2$ **then**

$sima(C1, C2) = \frac{\min(x_i, y_i)}{\max(x_i, y_i)}; a = a + 1;$

if $(C1, C2)$ qualitative **then**

if $C1 = \perp \vee C2 = \perp$ **then**

$sima(C1, C2) = 0;$

$b = b + 1;$

else { $C1$ is the same concept with $C2$ }

if $C1 \equiv C2$ **then**

$sima(C1, C2) = 1; a = a + 1;$

else{the direct path relation $C1-C2$ }

if $C1 \approx C2$ **then**

$sima(C1, C2) = \text{value} \square [0,1]; a = a + 1;$

End For

End For

End For

Step 2:Get each service an overall similarity value.

$$SimPro(P_x, P_y) = \left(\frac{a}{a+b} \right) \times \frac{a \sum_{i=1}^{i=a} w_i \times sima(P_{x_i}, P_{y_i})}{a \sum_{i=1}^{i=a} w_i + b \sum_{i=1}^{i=b} w_{a+i}}$$

Step 3: Return score ranking SC. We retain the list of hotels, if its matching score is greater than a fixed threshold $\square/\square \in [0,1]$ is a user-defined threshold.

starsMax . . . etc. The facilities-key table is segmented in twenty nine category: *Room service, Food, Kids, Sport, Animals, Entertainment, Shopping, Parking, Health, Confort, NewsandTV, Touring . . . etc.* These categories play the role of the services profiles.

- For the two similarity measures, we performed three different requests through a hotel reservation scenario. Each user profile with the degree of interest for each topic is given as follows:

$P_{U1} = \{Food(0.2); Smoke(0.2); Parking(0.6)\}.$

$P_{U2} = \{Internet(0.2); Food(0.2); Business(0.2); Conference(0.4)\}.$

$P_{U3} = \{Sport(0.2); Health(0.4); Kids(0.3); RoomService(0.1)\}.$

B. Experimental results

The objective of the tests is to show the interest of the profile of the user and the service in the discovery of the best services as well as to confront the two similarity measures (*SimPro* and *QSim*) and to compare the obtained results. Figure 3, Figure 4 and Figure 5 illustrate the similarity degrees obtained between user's three requests and the various services respectively using *SimPro* and *QSim*.

According to the curve depicted in Figure 3, 4 and 5, for *Qsim* we note that the degree of similarity is always weak [0, 0.4] at every time where the number is important, although the service satisfies the user's need well. This is due principally to the fact that *QSim* adds the variable *c* in the denominator $a/a + b$. Mathematically, at every time where *c* increases, the score decreases, what interest us more in the calculation of the similarity is the set of common characteristics (a) of *Px* and *Py* and those that are requested by the client(b). For instance, we have the case of: User 1 with Service 1 (*QSim*= 0.01), this low score is due to the value of *c*=96, on the other hand *a*=2 and *b*=1, so *Simpro*=0.59, the query 1 is satisfied with two attributes.

Now looking at results obtained by our similarity measure *SimPro*, the degree of the similarity is always high if the set of feature (a) suggested by the service and requested by the client is important, no matter the number of attributes provided by the service and non requested by the client, for instance services (1, 2, 3, 4 and 5) correspond to 100 % of the profile of user 2, because the characteristics desired by this user are available in five hotels.

The graphs results show that *QSim* is unable to find some of the relevant services that were directly related to the queries concepts. *SimPro* uses the common information between user profile and service profile to match services based on quantitative and qualitative similarity rather than just quantitative similarity and thus exhibited better answers than *QSim* and *Jaccard* measure.

In short, we think that the influence of a value (and thus a feature) depends on the number of features it was mentioned by the user and published by the provider. The performance of Web service may depend on atomic similarity with weight of each request's characteristic and on the quantity $a/a + b$ that is in fact the average frequency of all mentioned values.

V. CONCLUSION AND OUTLOOK

In this work, we proposed an approach based on profile similarity to enhance the search and recommendation of SWS. In our approach, we applied a new similarity measure for calculating similarity degree between the provided services profile and the desired services profile for better results. The most powerful strength of the proposed approach is the ability to make it possible to systematically measure the similarity between different attributes (either qualitative or quantitative) of user profile and a set of attributes of services profile. As for future work, we first plan to conduct extensive and thorough experiments to study the effectiveness of the



Figure 3. Similarity Results for Query1.



Figure 4. Similarity Results for Query2.

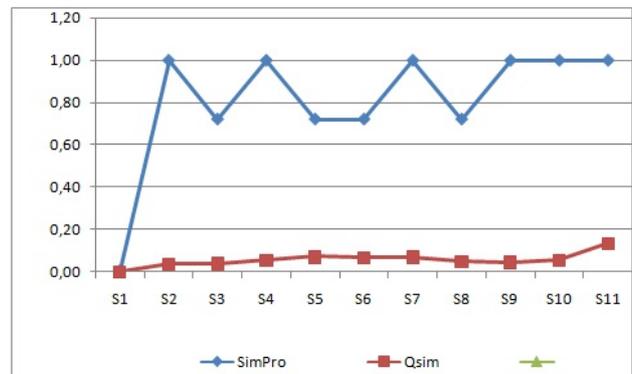


Figure 5. Similarity Results for Query3.

proposed formula for measuring similarity, to analyze the quality of SWS from a user point of view, and to consider the use of ontology for the proposed profile model. Another interesting line of research is to model the parameters of user profile using gradual concepts (such as "cheap cost", "fast service", etc.) which can be represented thanks to fuzzy sets [24], and use the diversity criterion (to return a set of SWS which are as diverse as possible while preserving the satisfaction) in order to improve the quality of the answers.

REFERENCES

- [1] M.B.Blake, M.F.Nowlan, "A Web Service Recommender System Using Enhanced Syntactical Matching" IEEE International Conference in Web Services ICWS, 2007.
- [2] M.laniaZooloqica, "Tables of significant values of Jaccard's index of similarity" Misc. Zool., 22.1: 29-40, 1999.
- [3] D. Booth, H. Haas, F. McCabe, and E. Newcomer. "Web services architecture", Aout (2003). <http://www.w3.org/TR/2003/WD-ws-arch-20030808/>.
- [4] F.Nawaz, K.Qadir, H. Farooq Ahmad, "SEMREG-Pro: A Semantic based Registry for Proactive Web Service Discovery using Publish Subscribe Model", Fourth International Conference on Semantics, Knowledge and Grid, IEEE Xplore, 2008.
- [5] R. Benaboud, R. Maamri, and Z. Sahnoun, "User's Preferences and Experiences based Web Service Discovery using Ontologies," in Proceedings of the Fourth International Conference on Research Challenges in Information Science (RCIS), Nice, France, May.2010, pp. 121-126.
- [6] G.Wen-yue, Q.Hai-cheng, C.Hong, "Semantic Web service discovery algorithm and its application on the intelligent automotive manufacturing system", International Conference on Information Management and Engineering, IEEE Xplore, 2010.
- [7] U. Chukmol, A. Benharkat, and Y. Amghar, "Enhancing Web Service Discovery by using Collaborative Tagging System," IEEE 4th International Conference on Next Generation Web Services Practices, October.2008, pp. 54-59.
- [8] M. Keidl and A.Kemper, "Toward Context/aware Adaptable Web Services," in 13th international World Wide Web Conference, USA, 2004, pp. 55-65.
- [9] T. Rajendran, P. Balasubramanie, "An Optimal Agent-Based Architecture for Dynamic Web Service Discovery with QoS", Second International conference on Computing, Communication and Networking Technologies, IEEEExplore 2010.
- [10] A. K.Dey, G. D. Abowd, "Towards a Better Understanding of Context and Context-Awareness". In : CHI 2000 Workshop on the What, Who, Where, When, and How of Context-Awareness, The Hague, The Netherlands, April 2000.
- [11] A.Belkhirat, A.Belkhir. "A new similarity measure for the profiles management". In 10th Int. Conf. on Computer Modeling and Simulation, 2011.
- [12] A. Belkhirat, A. Bouras and A. Belkhir, "A New Similarity Measure for the Anomaly Intrusion Detection"; Proceedings of the 3rd International Conference on Network and System Security (NSS2009), Queensland, October 2009, pp. 431-436.
- [13] S. Dietze, A. Gugliotta, J. Domingue, "Exploiting Metrics for Similarity-based Semantic Web Service Discovery"; In ICWS, 2009.
- [14] W. Rong and K. Liu, "A Survey of Context Aware Web Service Discovery: From User's Perspective," Fifth IEEE International Symposium on Service Oriented System Engineering, June.2010, pp. 15-22.
- [15] T. Rajendran and P. Balasubramanie, "An Optimal Agent-Based Architecture for Dynamic Web Service Discovery with QoS," Second International Conference on Computing Communication and Networking Technologies, IEEEExplore, July.2010, pp. 1-7.
- [16] M.Zeleny. Multiple Criteria Decision Making. MacGraw Hill, 1982.
- [17] J.L. Fleiss. Statistical methods for rates and proportions. John Wiley, 1981.
- [18] J. Euzenat and P. Shvaiko. Ontology matching. Springer-Verlag, Heidelberg(DE), 2007.
- [19] N. Kokash, A. Birukou, and V. D'Andrea, "Web Service Discovery Based on Past User Experience," in Proceedings of the International'1 Conference on Business Information Systems (BIS), LNCS Volume 4439, 2007, pp. 95-107.
- [20] R. Rada, H. Mili, E. Bicknell, and M. Blettner. Development and application of a metric on semantic nets. Systems, Man and Cybernetics, IEEE Transactions on, 19:17-30, 1989.
- [21] S. Castano, A. Ferrara, S. Montanelli, and G. Racca. "Semantic information interoperability in open networked systems". In Proc. of the Int. Conference on Semantics of a Networked World (ICSNW), in cooperation with ACM SIGMOD 2004, pages215-230, Paris, France, June 2004.
- [22] Z.Wu, M.Palmer, "Verb Semantics and Lexical Selection," Proceedings of the 32nd Annual Meeting of the Associations for Computational Linguistics (ACL-94), LasCruces, New Mexico, pp. 133-138. 1994.
- [23] C. D'amato, N. Fanizzi and F. Esposito, "A dissimilarity measure for ALC concept descriptions", SAC'06: Proceedings of the 2006 ACM symposium on Applied computing, New York, USA, ACM Press, p. 1695-1699,2006. <http://dx.doi.org/10.1145/1141277.1141677>
- [24] L. A. Zadeh, "Fuzzy sets", Information and Control, vol. 8, no. 3, pp.338-353, 1965. [http://dx.doi.org/10.1016/S0019-9958\(65\)90241-X](http://dx.doi.org/10.1016/S0019-9958(65)90241-X)

AUTHORS

Z. Chouiref is with University Colonel Akli Mohand Oulhadj, Bouira, Algeria.

A. Belkhir is with University of Sciences and Technologies Houari Boumediene, Algiers, Algeria.

A. Hadjali is with Laboratory of Informatique and Automation Systems - ENSMA, Poitiers, France.

Submitted 29 June 2013. Published as re-submitted by the authors 23 July 2013.