



Context-Aware and Linked Open Data Based Service Discovery

Nasredine Cheniki^{1,2}(✉), Yacine Sam¹, Nizar Messai¹, and Abdelkader Belkhir²

¹ LIFAT - University of Tours, Tours, France

{nasredine.cheniki,yacine.sam,nizar.messai}@univ-tours.fr

² LSI, University of Science and Technology Houari Boumediene,
Bab Ezzouar, Algeria
kaderbelkhir@hotmail.com

Abstract. Context-aware Service Discovery process is the task of identifying relevant services that could satisfy user needs while considering contextual information. Depending on such context information, the discovery process may become a challenging task as it requires to consider different kinds of constraints. In tourism domain, for example, context-aware service discovery process has to take into account functional information (what functionalities does a service offer) and non-functional information (such as user profile and preferences or information about her/his surrounding environment). To deal with context-aware services discovery issue, researchers proposed solutions that rely on semantic Web technologies and particularly ontologies to describe and discover services. However, ontologies are generally not-interconnected and limited because they focus on specific domains. In this paper, we propose a novel discovery approach for context-aware services which relies on Linked Open Data (LOD) instead of ontologies. First, we propose a model to describe both functional and non-functional properties of context-aware services. Then, we enrich the service model instances with semantic data from LOD. Based on enriched descriptions, we propose matchmaking algorithms to satisfy a user's needs in a given context. We performed experiments on touristic services. Results show that using LOD for context-aware service discovery is promising regarding discovery effectiveness.

Keywords: Context-Aware Services · Service discovery
Semantic similarity measures · Linked Open Data

1 Introduction

The continuous improvement in performance of mobile networks over the last few years, in conjunction with the significant evolution of hardware and software of mobile devices (smartphones, tablets, PDAs, etc.) has greatly transformed the way users interact with information systems. A new era of computing, called “ubiquitous computing”, based on these two evolutions, has emerged. Nowadays, a mobile user has an ubiquitous access to multiple resources, in particular, mobile

applications and services. In order to satisfy user needs in terms of functional and non-functional service requirements, discovery process has to be applied. This process performs, at first stage, a functional matchmaking to return services having similar functionalities to those requested by the user. At the second stage, non-functional parameters (such as contextual information) are taken into consideration to rank services depending on user's task or situation [1]. Incorporating context information during discovery process reduces the initial set of services (resulted from functional matchmaking) and eliminates results that do not correspond to the user's needs and hence increase the performance of the services retrieval process.

For more than a decade, Semantic Web technologies, especially domain ontologies were a centric concept in proposed context-aware resources discovery approaches. Researchers have used ontologies to enable semantic matchmaking between user's request and services descriptions.

Despite the interest of using ontologies in context-aware mobile services discovery, they present in our opinion two major weaknesses. First, they are resources limited and their knowledge covers generally restricted domains. Second, they are generally isolated, i.e., their concepts are not interconnected with concepts in other ontologies, which limits their exploitation. For instance, it is not possible to enrich a concept definition in one ontology from an equivalent one defined in another ontology because of missing links. These two limitations implicitly impact any service discovery process that is based on ontologies.

Researchers over the past few years have addressed this issue in two different ways. Either by replacing domain specific ontologies by multi-domains ontologies such as WordNet or by using textual knowledge sources such as Wikipedia to bootstrap service discovery [2]. However, both directions have drawbacks. Although multi-domains ontologies are relatively rich, they remain non-interconnected. Textual knowledge sources also suffer from concepts interconnection problem. In addition, textual sources require exhaustive processing to be exploited (in concepts relevance calculation, for example). We are thus convinced that the use of data from Web of Data, also called LOD (Linked Open Data) [3], will make it possible to overcome the previous limitations. Such data is mainly intended to be processed by machines, and therefore designed and thought to be structured, semantically described, open and linked to each other.

Our contribution in this paper consists in exploiting LOD to perform functional and non-functional matchmaking of contextual services. First, LOD data is used to semantically annotate functional and parts of non-functional services information. Then, semantic similarity measures that involve LOD annotations are developed in order to perform semantic LOD-based matchmaking between user requests and services descriptions.

The remainder of this paper is organized as follows. Section 2 discusses related works in research domain. We present Linked Open Data and LODS similarity measure in Sect. 3. We describe the two phases of our proposed discovery approach in Sect. 4. Evaluation of our approach based on a dataset of touristic services is detailed in Sect. 5. We summarize our work and draw its potential perspectives in Sect. 6.

2 Related Work

Context-aware services discovery is a widely studied notion since the beginning of ubiquitous computing research era. Earlier context-aware discovery systems were based only on user/service location. Location-aware tour guides such as in [4] provide services information according to the user's location. However, in addition to location, context-aware service discovery can benefit from various context information such as device profile, user preferences, and environment parameters in order to satisfy the user's needs. Other works [5–8] tried to include more contextual information such as devices preferences, environment context, user ratings, etc., to perform the services discovery process.

Hence, our approach presented in this article suggests integrating multiple types of context information in order to find services that fit user requirements and improve service discovery effectiveness.

Generally, context-aware services discovery passes through two dependent discovery stages. The first stage concerns functional matchmaking which takes into account required/provided functionalities by user/service. The second stage concerns the non-functional properties, where contextual information are taken into consideration. To perform the two stages discovery, many existing context-aware discovery approaches [9–13] rely on predefined ontologies or (semantic) rules to match services. Ontologies are used at the first discovery stage in order to guaranty a semantic matchmaking that overcomes limitations of the syntactic search. Semantic rules are applied generally on the second stage of matchmaking in order to filter services according to non-functional properties [14–16].

In fact, for the second stage of matchmaking, two directions of matchmaking are applied. The first direction consists of using profile-based similarity matchmaking such as in [17]. This provides automatic matchmaking of services with minimum user's intervention. The second direction consists of using (semantic) rules such as in [15] to provide more flexibility for users to allow them to exhibit their non-functional needs. However, this approach requires more user involvement to specify non-functional requirements.

In our work, we combine rule-based and profile-based discovery. We use rules only to filter services according to user's mandatory requirements. Then, we apply automatic matchmaking based on profile similarity on resulted services.

In addition, due to ontologies limitations stated earlier, we propose, in contrast to studied works, to use LOD resources instead of ontology concepts to describe and discover context-aware services.

3 Linked Open Data and LODS Similarity Measure

Linked Open Data is a set of technologies responsible for the evolution of the traditional Web of documents into a Web of structured and interlinked data. These data are intended for use by machines [3]. Since its appearance in 2007, LOD is constantly evolving into a gigantic source of knowledge covering different domains, freely accessible for everyone. LOD data are generally presented

as resources represented with URIs, where each resource is described by semantically structured information, all represented in form of RDF triples. These triples are stored in RDF-store accessible via dedicated SPARQL endpoints. Since URIs can be used to define unique resources in Web, that makes it easy to link LOD resources with each other. The resulted knowledge database is structured and interconnected that allows information systems to analyze and exploit its resources at a semantic level. An example of LOD datasets is DBpedia [18], the semantic structured mirror of Wikipedia. It transforms every Wikipedia article to a LOD resource described with structured information extracted from the corresponding article.

Based on LOD, a semantic similarity measure called LODS was proposed in [19]. This measure computes the similarity between two concepts linked to corresponding LOD resources by exploiting their features. It uses: (i) a taxonomic structure of ontological resources concepts; (ii) the classification categories of LOD resources; and (iii) the semantic properties used to describe LOD resources. Moreover, LODS takes profit from inter-links between LOD datasets by traversing relationships between resources in related datasets. By doing so, LODS compares resources with more information to reduce the impact of missing information within a single dataset. LODS measure has two parameters ℓ and ℓ' used to limit hierarchy level of extracted categories used in computing similarity. More details about LODS measure is available in [19].

4 Proposed Context-Aware Service Discovery Approach

In this section, we first define a formal model that describes contextual services. Then, we detail the two stages of our discovery approach (functional and non-functional).

4.1 Formal Definition of Contextual Services

In what follows, we formally define contextual services. In particular, we will specify how service annotations are made. Then, we describe how these annotations are exploited to discover services using LODS similarity measure.

A context-aware service s is modeled by a triplet $\langle n_s, \mathcal{F}_s, \mathcal{NF}_s \rangle$, where:

- n_s is the name of the service s .
- The set $\mathcal{F}_s = \{f_1, f_2, \dots, f_i\}$ represents the functionalities of a service s . These functionalities can be annotated by a subset of resources $A_s \in LOD$ where: $\forall f_i \in \mathcal{F}, \exists r_i \in A_s | \{ \langle f_i, \alpha, r_i \rangle \}$ α being an annotation property that attaches a LOD resource to a service functionality.
- The set $\mathcal{NF}_s = \{P_{1,s}, P_{2,s}, \dots, P_{n,s}\}$ represents the non-functional properties of the service. Due to the diversity of non-functional information, they are generally organized in profiles, where each profile $P_{i,s}$ describes a specific context information domain.

Properties inside profiles can be of two types: quantitative and qualitative. Quantitative properties have generally numerical values while qualitative properties have string type values that can be attached to LOD resources. This annotation is modeled as follows: for a property p with set of values $\{v_1, v_2, \dots, v_i\}$, a set of resources $r_i \in LOD$ can be attached to them. Formally, $\exists p = \{v_1, v_2, \dots, v_i\} \in P_{i,s} \mid \langle v_i, \alpha, r_i \rangle, r_i \in LOD$.

To show how to use our model, let's have an hotel service named $n_s =$ "Le Plantagenêt" with only one feature $\mathcal{F}_s = \{Hotel\}$. The annotation set of this service is $A_s = \{dbp : Hotel\}$. It has one DBpedia resource¹ attached to the functionality *Hotel*. This service has a profile that describes service preferences

$\mathcal{NF}_s = \{P_{preferences,s}\}$, where $P_{preferences,s} = \{TotalCapacity = 33; Ranking = 3, ComfortServices = dbr:Internet, SpokenLanguages = \{French, English\}\}$.

4.2 Functional Matchmaking of Services

Based on the above-defined formalism, we define here the first stage of contextual services discovery. This stage of matchmaking concerns the functional part of the service. Let have a user request consists of a set of terms $R = \{t_1, t_2 \dots t_i\}$ where each term can be attached to a subset of annotation resources $A_R \in LOD$, so $\forall t \in R, \exists r_i \in A_R \mid \langle t_i, \alpha, r_i \rangle$.

The similarity between the request R annotated by a set of LOD resources A_R and a service s whose functionalities are annotated by a set of resources A_s is calculated based on LODS measure as follows:

$$Sim_{\mathcal{F}}(R, s) = \frac{\sum_{a \in A_R} \sum_{b \in A_s} LODS^{\ell, \ell'}(a, b)}{|A_R| \times |A_s|} \quad (1)$$

We use a classical aggregation measure [20] that allows the comparison of two objects annotated with semantic concepts. It shall sum the scores obtained from applying LODS measure on each combination of the Cartesian product of the two compared sets. Then, it divides the sum by the number of combinations. The final score is normalized in the interval [0,1].

4.3 Non-functional Matchmaking of Services

At this second matchmaking stage, we first use semantic rules to eliminate services that do not satisfy mandatory user requirements. Then, a profile-based similarity measure is applied to the filtered services in order to compare user and services profiles and determine their degree of relevance.

Given the variety of information that can be found in a contextual environment, each type of information is represented in a separate profile structure. For instance, we can find a profile for user preferences, a second for user device specifications, a third for information about surrounding environment, etc. Profiles

¹ dbp is a prefix that represents the URL <https://dbpedia.org/resource/>.

information can be explicitly provided by the user/service-provider or implicitly leaned by the system.

Hence, non-functional matching is based on a set of user/service profiles. Let consider $P_{1,X}, P_{2,X}, \dots, P_{n,X}$ the set of profiles that describe user/service non-functional properties in a particular context situation where if $X = u$ means that the profile describes user information, and if $X = s$ means that it describes service information. Each profile is described by a set of properties $P_{i,X} = \{x_1, x_2, \dots, x_n\}$. Non-functional similarity between a user u and a service s profiles is calculated as follows:

$$Sim_{NF}(u, s) = \lambda_1 \times f(P_{1,u}, P_{1,s}) + \lambda_2 \times f(P_{2,u}, P_{2,s}) + \dots + \lambda_n \times f(P_{n,u}, P_{n,s}) \quad (2)$$

where λ_i represents the level of profile importance during the matching process. The function f computes the similarity between two profiles. It is based on measure defined in [21]. We extend the latter to take into account the different types of data present in user/service profiles. We integrate our LODS similarity measure into the extended measure in order to evaluate the quantitative information present in profiles.

Let consider $P_{i,u} = \{x_1, x_2, \dots, x_n\}$ and $P_{i,s} = \{y_1, y_2, \dots, y_m\}$ two profiles that describe user and services properties in a particular context situation. The function f measures the degree of similarity between $P_{i,u}$ and $P_{i,s}$ as follows:

$$f(P_{i,u}, P_{i,s}) = \frac{a \sum_{i=1}^{i=N} w_i \times ASim(x_i, y_i)}{a \sum_{i=1}^{i=a} w_i + b \sum_{i=1}^{i=b} w_{a+i} + c \sum_{i=1}^{i=c} w_{a+b+i}} \quad (3)$$

where, \mathbf{a} is the number of properties in common, \mathbf{b} represents the number of properties belonging to X and not to Y , and \mathbf{c} is the number of properties belonging to Y and not to X . Each property has a weight w_i , where $\sum_{i=1}^n w_i = 1$.

The atomic measure $ASim$ is calculated as follows:

$$ASim(x_i, y_j) = \begin{cases} sim_{num}(x_i, y_i), & \text{if } (x_i \wedge y_i : \text{numerical}) \\ sim_{intrv}(x_i, y_i), & \text{if } (x_i \vee y_i : \text{Interval}) \\ sim_{LODS}(x_i, y_i, \mu), & \text{if } (x_i \wedge y_i : \text{LOD}) \\ sim_{set}(x_i, y_i, \mu), & \text{if } (x_i \wedge y_i : \text{Set}) \\ sim_{str}(x_i, y_i), & \text{if } (x_i \wedge y_i : \text{String}) \end{cases} \quad (4)$$

So, depending on the type of properties, $ASim$ calls a dedicated function to measure the similarity between properties values in two compared profiles:

- sim_{num} is used to process numeric values: $sim_{num}(x_i, y_i) = \frac{\min(x_i, x_i)}{\max(x_i, y_i)}$
- sim_{intrv} is used to compare two values of a property, one of which is an interval. If one of the values is a numeric value and the other one is an interval, the measure returns 1 if the numeric value is included in the interval, it returns 0 otherwise. If both values are intervals, the percentage of intersection between these two intervals is returned.

- If the values of a property are attached to a list of LOD resources, the similarity calculation is based on LODS measure (Eq. 5). Two measures are applied depending on the value of the parameter μ :
 - If the existence of a single resource from the list of the first profile in the list of the second profile (case *AtLeast*) is sufficient, the measure returns the maximum value obtained from applying LODS on all possible resources combinations of the two lists.
 - Otherwise, if we look to satisfy all resources of first profile compared to those in second profile (case *Max*), then the measure returns the average of the cumulative results obtained by applying LODS on all possible combinations of resources of the two lists.

$$sim_{LODS}(x_i, y_i, \mu) = \begin{cases} 1, & x_i = y_i \\ \frac{1}{|x_i|} \sum_{a \in x_i} \max_{b \in y_i} LODS^{\ell, \ell'}(a, b) & if(\mu = Max) \\ \max_{a \in x_i, b \in y_i} LODS^{\ell, \ell'}(a, b) & if(\mu = AtLeast) \end{cases} \tag{5}$$

- The measure sim_{set} (Eq. 6) compares properties that holds set of values. Tversky [22] model is applied on compared sets, if we want to satisfy all list elements of the first profile compared to list elements of the second profile (case *Max*). Otherwise, if at least one element is to be satisfying (case *AtLeast*), the measure returns 1 as soon as there exists an element of the first profile list in the second profile list.

$$sim_{set}(x_i, y_i, \mu) = \begin{cases} Tversky(x_i, y_i) & if(\mu = Max) \\ 1, & if(\mu = AtLeast \wedge \exists a \in x_i, \exists b \in y_i, a = b) \\ 0, & Else \end{cases} \tag{6}$$

- The sim_{str} (Eq. 6) compares string type properties. There are several measurements in literature, we opt for the standard Levenshtein measure.

$$sim_{str}(x_i, y_i) = 1 - Levenshtein(x_i, y_i) \tag{7}$$

5 Experimental Validation

5.1 Contextual Services Benchmark

To validate our services discovery approach, we use touristic services provided by the governmental touristic agency of the Indre et Loire Department (Agence Départementale de Tourisme d’Indre-et-Loire - ADT37) in Centre Val de Loire

region, France) via Tourinsoft platform². These services are important to evaluate our approach since they are described with functional and non-functional (contextual) properties. We obtained 1632 services classified in categories (touristic offers) presented in Table 1.

Table 1. Selected touristic offers

| |
|--|
| Touristic offers |
| Hotels and restaurants |
| Historic Sites and Monuments |
| Parks and Gardens |
| Museums |
| Interpretation centers |
| Collective accommodations |
| Chambres d'hôtes |
| Furnished of tourism |
| Leisure and social service activities facilities |
| Cultural and sporting activities |

We evaluate our touristic services discovery approach in two stages according to the two phases of the discovery process, functional and non-functional. Our strategy consists in involving real users in evaluating the obtained results from both phases of the discovery process. To do this, we recruited students and teachers from the University of Tours in France. It is worthy to mention here that in both evaluation phases, the parameters of the *LODS* measure were set at $\ell = 2$ et $\ell' = 1$ (as suggested in original *LODS* paper).

5.2 LOD-Based Services Annotation

Touristic services returned by Tourinsoft, as seen Fig. 1, are described by several properties specifying different functional and non-functional information such as the classification of the service, its location, payment methods, etc. We use classification properties that determine the functionality of the service to describe service functionalities. These properties are attached to *LOD* resources in order to semantically describe the functional aspect of the services. Qualitative non-functional properties (such as *PrestationsEquipements* property) that can be attached to *LOD* resources are also annotated with corresponding resources.

We have performed a manual annotation process in order to guarantee a high level of accuracy, so we focus on measuring discovery effectiveness. However, in a real situation, we reckon that using an annotation tool that performs automatic *LOD*-based mapping is necessary. We choose DBpedia as a primary annotation

² <http://www.tourinsoft.com/>.


```

1 <service>
2 <title type="text">Hostellerie du château de l'Isle</title>
3 <m:properties>
4 <d:SyndicObjectName>Hostellerie_du_château_de_l'Isle
5 </d:SyndicObjectName>
6 <d:GmapLatitude>47.3282285</d:GmapLatitude>
7 <d:GmapLongitude>1.0458362</d:GmapLongitude>
8 <d:ObjectTypeName>Hôtellerie</d:ObjectTypeName>
9 <d:Classements11Prefectoral>3 étoiles</d:Classements11Prefectoral>
10 <d:LanguesParlees>Français#Anglais</d:LanguesParlees>
11 <d:CapaciteTotale>12</d:CapaciteTotale>
12 <d:CodePostal>37150</d:CodePostal>
13 <d:Classification xml:space="preserve"> Hôtels - restaurant
14 </d:Classification>
15 <d:PrestationsEquipements>Boulodrome#Jardin
16     #Jeux pour enfants en extérieur
17     #Jeux pour enfants#Parc#Parking#Parking privé#Restaurant
18     #Salle de réception#Salon#Télécopieur#Terrasse
19 </d:PrestationsEquipements>
20 <d:Web xml:space="preserve"> http://www.chateau-delisle.com</d:Web>
21 <d:ModesPaiements>American Express#Carte bleue
22 #Chèques bancaires et postaux #Chèques de voyage</d:ModesPaiements>
23 </m:properties>
24 </service>

```

Fig. 1. Example of a touristic service description

source of LOD data since it is a multi-domains knowledge source. This allows us to annotate services of different categories.

5.3 Functional Matchmaking Evaluation

We recruited a total of 22 users, 10 of whom initially chose a set of keywords, in French, that they can use in a search scenario for tourist services. The Table 2 shows the 9 most frequent queries submitted by users, the right column provides a corresponding English translation.

The 9 queries selected by users were executed by our algorithm (that implements $Sim_{\mathcal{F}}$ measure) then the result was presented to users for evaluation. For each query, we retrieve the first 10 results that have different annotations (to avoid retrieving services with identical annotations). This makes it possible to evaluate the effectiveness of proposed measure regarding retrieving semantically related services.

After that, the 90 results of all queries were presented in a questionnaire format to be evaluated by 10 users. For each query, we present top 10 retrieved services to users in order to indicate whether the service is relevant to the query or not (binary evaluation). To reassure that there is a high degree of consensus between responses of all participant users, we calculated the degree of agreement

Table 2. A selection of the 9 most frequent queries.

| Q | French keyword | English translation |
|-------|----------------------------------|---------------------------------|
| R_1 | Hôtel | Hotel |
| R_2 | Restaurant | Restaurant |
| R_3 | Hébergements locatifs/collectifs | Rental/collective accommodation |
| R_4 | Chambres d'hôtes | Bed and Breakfast |
| R_5 | Auberges de jeunesse | Youth hostels |
| R_6 | Auberge de campagne | Country Inn |
| R_7 | Gîte | Gîte |
| R_8 | Site et monument historiques | Historic sites and Monuments |
| R_9 | Château | Castel |

using Fleiss's Kappa measure³. The latter returned the score $Kappa = 0.629$ indicating that there is an interesting satisfactory agreement of users' responses.

Using *precision@* metric, we calculate the precision value of the first k obtained results. Table 3 illustrates precision results. We notice that our algorithm generally gives good precision values. However, it is remarkable that accuracy decreases when considering more services. This is due to two reasons: (i) the algorithm returns some results that are not relevant for some queries, and/or (ii) for some queries, all relevant results were retrieved before reaching $k = 10$ and the algorithm returns services that have a low similarity value to the query.

Table 3. The *Precision@10* results for the 9 queries

| Queries | Precision | | | | | | | | | |
|---------|-----------|-----|-----|------|------|------|-------|------|------|------|
| | P@1 | P@2 | P@3 | P@4 | P@5 | P@6 | P@7 | P@8 | P@9 | P@10 |
| R_1 | 1.0 | 1.0 | 1.0 | 1.0 | 0.8 | 0.66 | 0.714 | 0.62 | 0.55 | 0.5 |
| R_2 | 1.0 | 1.0 | 1.0 | 0.75 | 0.6 | 0.66 | 0.57 | 0.5 | 0.44 | 0.4 |
| R_3 | 1.0 | 1.0 | 1.0 | 1.0 | 1.0 | 1.0 | 0.85 | 0.75 | 0.66 | 0.7 |
| R_4 | 1.0 | 1.0 | 1.0 | 1.0 | 1.0 | 1.0 | 0.85 | 0.75 | 0.66 | 0.7 |
| R_5 | 1.0 | 1.0 | 1.0 | 1.0 | 0.8 | 0.66 | 0.57 | 0.5 | 0.44 | 0.4 |
| R_6 | 1.0 | 1.0 | 1.0 | 1.0 | 0.8 | 0.6 | 0.71 | 0.75 | 0.66 | 0.6 |
| R_7 | 1.0 | 1.0 | 1.0 | 1.0 | 1.0 | 1.0 | 1.0 | 1.0 | 1.0 | 0.9 |
| R_8 | 1.0 | 1.0 | 1.0 | 1.0 | 1.0 | 1.0 | 1.0 | 1.0 | 1.0 | 1.0 |
| R_9 | 1.0 | 1.0 | 1.0 | 1.0 | 1.0 | 1.0 | 1.0 | 1.0 | 1.0 | 1.0 |
| AVG | 1.0 | 1.0 | 1.0 | 0.97 | 0.88 | 0.85 | 0.80 | 0.76 | 0.71 | 0.68 |

³ https://en.wikipedia.org/wiki/Fleiss'_kappa.

5.4 Non-functional Matchmaking Evaluating

To evaluate the second stage of our discovery approach, we selected 5 users to fulfill their profiles describing non-functional preferences. We asked also users to specify a mandatory constraint that consists in specifying a maximum geographic distance in which services are retrieved. This requirement is met through a semantic rule. We consider that users are located in the same geographical position (we have chosen Tours city center as the reference position). Users respectively chose following mandatory distances 2 km, 2 km, 60 km, 10 km, 10 km.

In order to make the evaluation easy for users, we incorporate only one profile in matchmaking process; it is the profile that describes accommodation preferences. Moreover, we consider only the query R_1 in this evaluation process. We limit also the number of properties involved in the evaluation. Chosen properties are presented in Table 4. Finally, we consider that all properties have the same weight (w_i).

Table 4. Properties considered in non-functional evaluation

| Property | Type | μ | Values examples |
|-------------------|-----------------|---------|---------------------------|
| Spoken language | Set | AtLeast | French, English |
| Payment method | Set | AtLeast | Blue Card, Cash |
| Ranking | Number/Interval | - | 3 stars, [2-4] stars |
| provided services | LOD | Max | dbr:Parking, dbr:Internet |

Context Precision. We first calculate context accuracy that measures the positive impact of involving context information to the service discovery. Context accuracy is defined as the percentage improvement in the result (services that are not relevant to the context are eliminated) over initially retrieved services.

We take the 166 services retrieved by executing R_1 on the first stage of matchmaking. We assume that all relevant services are retrieved (Recall = 100%) and they are all relevant to the query (precision = 100%).

So, the total number of retrieved services $|S| = 166$. After applying the semantic rule that filters services according to their context information, the number is reduced to N services. Context precision (CP) is calculated as follows:

$$CP = \frac{|S| - N}{|S|} \quad (8)$$

The context precision for the 5 users is shown in Fig. 2. Clearly, the precision value is affected by user requirements. User 3, which has eliminated services that have a maximum distance of 60 km, had a low precision because fewer services have been filtered since most of them have a distance lower than the one user required.

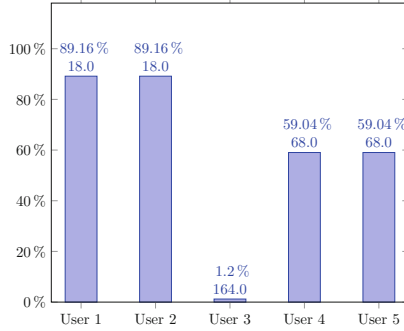


Fig. 2. Context precision for each participating user

Quality of Discovery. The discovery quality of the non-functional service match-making determines whether this matching measure is capable of meeting the user's requirements and preferences. Here we compare scores returned by our measure $Sim_{\mathcal{NF}}$ for each service with the satisfactory scores assigned by users. We also measure the impact of the context on the quality of service discovery by eliminating the second matchmaking phase and ask users to evaluate services returned directly from the first matchmaking phase.

To carry out this evaluation process, the 5 users evaluate the first 5 services returned before and after the applying non-functional matchmaking. For each user, we calculate the average of the first 5 scores attributed by the user to services. Then, we compare results with the average of scores attributed to same services by $Sim_{\mathcal{NF}}$ measure. Figure 3 illustrates the results obtained. It shows that ratings assigned by users are close to those assigned by our measure. This means that our measure is able to classify services according to the degree of user satisfaction. On the other hand, the omission of the context-based matching phase affected negatively the relevance of returned results.

Figure 4 illustrates the value of the standard deviation obtained by applying the Mean Absolute Error (MAE) between the similarity values assigned by users and those assigned by our measure $Sim_{\mathcal{NF}}$. The lower the MAE value, the better the quality of the ranking. Equation 9 shows how MAE is calculated.

$$EAM = \frac{\sum_{s_i \in S} |Sim_{\mathcal{NF}_{s_i}} - Rate_{s_i}|}{|S|} \quad (9)$$

Where $Rate_{s_i}$ is the rating value assigned by the user to estimate the degree of relevance of his profile in regards a service s_i .

The lower MAE values obtained means that scores returned by our measure are closed to those assigned by users. The first three users have evaluation scores close to those returned by our measure, their MAE values are the lowest. The two remaining users have relatively higher MAE values, which means that their

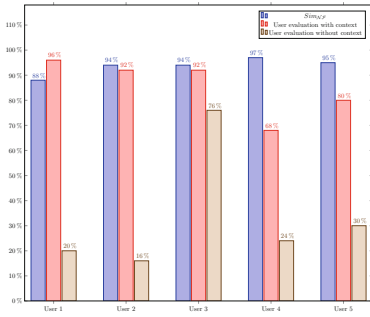


Fig. 3. Evaluation of service discovery with and without use of context information

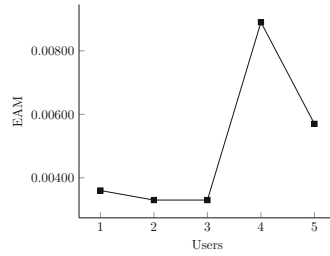


Fig. 4. Resulted MAE for each user

evaluation scores are a little far from those returned by our measure. This difference may reflect a sort of personal bias towards some services or it is due to the ambiguity in some queries.

6 Conclusion

In this paper, we proposed a novel approach to discover contextual services based on Linked Open Data. Our approach exploits LOD to enrich (annotate) functional and non-functional service descriptions and perform semantic match-making. The later was achieved by integrating our LOD-based similarity measure LODS on the discovery process to enable efficient service retrieval by exploiting the semantic aspect of LOD data. Two stages of matchmaking were proposed, the first called functional matchmaking consists of exploiting LOD annotations used to describe service functionalities to return services whose functionalities are semantically related. The second stage called non-functional matchmaking, LODS similarity measure has been integrated into a profile comparison measure to allow automatic matchmaking while taking into account the semantics of qualitative properties linked also to LOD resources.

Experimental Results demonstrate that our discovery approach provides strong matchmaking performance whether in functional or non-functional discovery phases. Hence, it can be used as an effective process to discover and rank relevant services in context-aware systems.

As perspectives, we work to integrate our discovery approach into a tourism recommendation platform that suggests to tourists visit plans and recommended services that are optimized and adapted to their preferences.

References

1. Dey, A.K.: Understanding and using context. *Pers. Ubiquit. Comput.* **5**(1), 4–7 (2001)
2. Bhargava, P., Lampton, J., Agrawala, A.: Bootstrapped discovery and ranking of relevant services and information in context-aware systems. In: proceedings of the 12th EAI International Conference on Mobile and Ubiquitous Systems: Computing, Networking and Services, ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering), pp. 180–189 (2015)
3. Bizer, C., Heath, T., Berners-Lee, T.: Linked data-the story so far. *Int. J. Semant. Web Inf. Syst. (IJSWIS)* **5**(3), 1–22 (2009)
4. Abowd, G.D., Atkeson, C.G., Hong, J., Long, S., Kooper, R., Pinkerton, M.: Cyberguide: a mobile context-aware tour guide. *Wirel. Netw.* **3**(5), 421–433 (1997)
5. Al-Masri, E., Mahmoud, Q.H.: Device-aware discovery and ranking of mobile services. In: 6th IEEE Consumer Communications and Networking Conference, Institute of Electrical and Electronics Engineers (IEEE), January 2009
6. Al-Masri, E., Mahmoud, Q.H.: MobiEureka: an approach for enhancing the discovery of mobile web services. *Pers. Ubiquit. Comput.* **14**(7), 609–620 (2009)
7. Kim, Y.S., Lee, K.H.: A lightweight framework for mobile web services. *Comput. Sci. Res. Dev.* **24**(4), 199–209 (2009)
8. Elgazzar, K., Hassanein, H.S., Martin, P.: DaaS: cloud-based mobile web service discovery. *Pervasive Mob. Comput.* **13**, 67–84 (2014)
9. Peng, R., Mi, Z., Wang, L.: An OWL-S based adaptive service discovery algorithm for mobile users. In: 4th International Conference on Wireless Communications, Networking and Mobile Computing. IEEE, October 2008
10. Toninelli, A., Corradi, A., Montanari, R.: Semantic-based discovery to support mobile context-aware service access. *Comput. Commun.* **31**(5), 935–949 (2008)
11. Mokhtar, S.B., Preuveneers, D., Georgantas, N., Issarny, V., Berbers, Y.: EASY: efficient semantic service discovery in pervasive computing environments with QoS and context support. *J. Syst. Softw.* **81**(5), 785–808 (2008)
12. Najar, S., Pinheiro, M.K., Souveyet, C., Steffanel, L.A.: Service discovery mechanism for an intentional pervasive information system. In: IEEE 19th International Conference on Web Services. IEEE, June 2012
13. Ruta, M., Scioscia, F., Di Sciascio, E., Piscitelli, G.: Semantic matchmaking for location-aware ubiquitous resource discovery. *Int. J. Adv. Intell. Syst.* **4**(3/4), 113–127 (2012)
14. Raverdy, P.G., Riva, O., de La Chapelle, A., Chibout, R., Issarny, V.: Efficient context-aware service discovery in multi-protocol pervasive environments. In: Mobile Data Management, pp. 3–3. IEEE (2006)
15. Fenza, G., Furno, D., Loia, V.: Hybrid approach for context-aware service discovery in healthcare domain. *J. Comput. Syst. Sci.* **78**(4), 1232–1247 (2012)
16. Chellouche, S.A., Negru, D., Arnaud, J., Batalla, J.M.: Context-aware multimedia services provisioning in future internet using ontology and rules. In: International Conference and Workshop on the Network of the Future (NOF). IEEE, December 2014
17. Bouyakoub, F.M., Belkhir, A.: A service discovery approach based on a quantitative similarity measure for M-tourism platforms. *Commun. Netw.* **4**(03), 227 (2012)
18. Lehmann, J., Isele, R., Jakob, M., Jentzsch, A., Kontokostas, D., Mendes, P.N., Hellmann, S., Morsey, M., van Kleef, P., Auer, S., Bizer, C.: DBpedia - a large-scale, multilingual knowledge base extracted from Wikipedia. *Semant. Web* **6**(2), 167–195 (2014)

19. Cheniki, N., Belkhir, A., Sam, Y., Messai, N.: LODS: a linked open data based similarity measure. In: 25th International Conference on Enabling Technologies: Infrastructure for Collaborative Enterprises (WETICE). IEEE, June 2016
20. Harispe, S., Ranwez, S., Janaqi, S., Montmain, J.: Semantic measures for the comparison of units of language, concepts or entities from text and knowledge base analysis. CoRR abs/1310.1285 (2013)
21. Belkhirat, A., Bouras, A., Belkhir, A.: A new similarity measure for the anomaly intrusion detection. In: 3rd International Conference on Network and System Security. IEEE (2009)
22. Tversky, A.: Features of similarity. *Psychol. Rev.* **84**(4), 327–352 (1977)