Engineering Applications of Artificial Intelligence 46 (2015) 154–165

Contents lists available at ScienceDirect

Engineering Applications of Artificial Intelligence

journal homepage: www.elsevier.com/locate/engappai

Semantic-based approach to improve the description and the discovery of Linguistic Web Services

Nabil Baklouti [a,*], Bilel Gargouri [b], Mohamed Jmaiel [a]

[a] ReDCAD Laboratory, University of Sfax, National School of Engineers of Sfax, B.P. 1173, 3038 Sfax, Tunisia
[b] Miracl Laboratory, University of Sfax, Pole Technologique de Sfax: Route de Tunis Km 10 B.P. 242, 3021 Sfax, Tunisia

ARTICLE INFO

Article history:
Received 23 March 2015
Received in revised form
15 July 2015
Accepted 8 September 2015
Available online 6 October 2015

Keywords:
Semantic knowledge
Linguistic information systems
Linguistic Web Services
Description
Discovery

ABSTRACT

This paper deals with the description and the discovery of Web Services in the linguistic information system. The previous studies have proved the lack of semantics within the descriptions of the linguistic Web Services (LingWS for short) which negatively affects the quality of their discovery. In order to overcome this problem, we propose to integrate the nonfunctional linguistic properties and their relationships inside the advertisements and to semantically annotate the several elements of the LingWS descriptions. Thus, we suggest an extension of OWL-S that promotes such a proposal. Moreover, we ensure the semantic annotation through a linguistic domain ontology that we have developed using the ISO standards of linguistic data categories (ISO 12620). This proposal is consolidated with the implementation of an editor called OWL-LingS (stands for OWL for Linguistic Services). As for the discovery task, we have extended the OWLS-MX hybrid matchmaker by integrating the nonfunctional linguistic properties within its matching algorithm. The obtained matchmaker, called OWL-LingS-MX, promotes the creation of the user requests according to the extended description of LingWS. The performance of OWL-LingS-MX matchmaker is shown through a comparison with three famous OWL-S matchmakers on the same computer. These systems are iSeM, OWLSM, and OWLS-MX.

1. Introduction

Nowadays, the Web services have become one of the most relevant research topics in the software engineering field. As the number of Web Services increases, the issue of selecting a desired service(s) becomes a challenging research topic. Initially, Web services were described using WSDL (Web Service Description Language). However, the lack of semantics in WSDL prevents an automatic discovery of Web services (Papazoglou et al., 2007). In order to enhance the description of the Web services, several languages and approaches, such as OWL-S (W3C, 2004), WSMO (W3C, 2004), and SAWSDL (W3C, 2007), have been proposed using semantic models (e.g., ontology).

The Linguistic Web Services (LingWS for short) are a kind of Web Services related to the linguistic information system (e.g., Part of speech Tagger, Tokenizer, Morphological analyzer). Such services are used to compose other LingWS corresponding to well-known applications such as Text Summarization, Machine Translation, and Information Retrieval (Bramantoro, March 2011).

Considering the richness of the linguistic knowledge, researchers in the Natural Language Processing (NLP for short) field have proposed many attempts to improve the description of LingWS using semantic approaches (Klein and Potter, 2004) or semantic wrappers (Ishida, 2006). Nevertheless, they did not offer the possibility to represent all the linguistic features. In fact, the LingWS's are characterized by several features called nonfunctional linguistic properties which are already discussed in previous works (Baklouti et al., 2012a, 2012b). The LingWS description should cover all these properties and their interrelationships like the treatment type (analysis and/or generation), the used formalism (e.g., contextual grammar, unification grammar), the processing level (e.g., morphological, syntactic, semantic), and so on. Unfortunately, the existing semantic approaches are unable to represent this kind of properties and their relations.

As far as the LingWS discovery is concerned, Bramantoro and Ishida (2011) proposed a new technique to measure the semantic similarity between LingWS descriptions through specified concepts in an ontology. They used a domain ontology previously proposed by Hayashi (2007). However, their proposal considers only the LingWS Inputs/Outputs (I/O for short) data type which is not the unique aspect that fully characterizes a LingWS (Hayashi and Narawa, 2012). Concerning the semantic matchmakers, there are many developed tools which ensure the matching of semantic Web services but there are no tools for discovering LingWS.

* Corresponding author at: ReDCAD Laboratory, University of Sfax, National School of Engineers of Sfax, B.P. 1173, 3038 Sfax, Tunisia.
E-mail addresses: nabil.baklouti@fsegs.rnu.tn (N. Baklouti), bilel.gargouri@fsegs.rnu.tn (B. Gargouri), mohamed.jmaiel@enis.rnu.tn (M. Jmaiel).

http://dx.doi.org/10.1016/j.engappai.2015.09.005
0952-1976/

Therefore, it is worth suggesting an appropriate matchmaker that mainly considers the nonfunctional linguistic properties.

In this paper, we aimed to enhance the LingWS description by integrating the nonfunctional linguistic properties and their relationships while annotating both functional and nonfunctional properties using a linguistic domain ontology.

Furthermore, we sought to propose an appropriate matching algorithm in order to improve the LingWS discovery. The description and discovery of LingWS would be consolidated by implementing some appropriate tools.

As a result, the integration of the nonfunctional linguistic properties and their relationships inside the description of LingWS helped the lingware system developer to explore the required properties, notably the linguistic ones. Moreover, the evaluation of the proposed matching algorithm using the SME2 environment has improved the discovery of LingWS. The obtained discovery results showed that the proposed approach can be applied for industrial application where the LingWS are numerous and various. In this context, the highly expressive descriptions offering more selecting features ensure the efficiency of the discovery process.

The remaining of this paper is organized as follows: Section 2 shows the nonfunctional linguistic properties. A comparative study between semantic approaches is provided in Section 3. Section 4, however, presents the proposed solution to enhance the description and discovery of LingWS. In Section 5, we depict the proposed solution to integrate and annotate the description elements. Then, Section 6 focuses on the LingWS discovery. Details on the implementation of the description and matching tools are discussed in Section 7. Section 8 presents the comparison experiments and analysis of the results. In Section 9, we briefly comment on some related works before drawing our conclusion in Section 10.

2. Nonfunctional linguistic properties

In general, knowledge items could be considered to describe a lingware system. In the context of LingWS, these knowledge items cover not only the service name and the functional properties but also some nonfunctional properties which are very useful to both describe and discover LingWS. These nonfunctional properties can be classified according to the processing level in the following way:

- *The Lexical Level*: It is characterized by the use of lexical linguistic resources, approaches, formalisms, analysis types, and so on. For example, the developer can choose a LingWS according to its type of lexical analysis which can be either thematic proposing a large conceptual category in which the linguist can navigate to look for a suitable word or structural helping the linguist in the structure choice. In addition, it can be syntagmatic with a statement element regrouped into phrases with an internal structure and a coherent unit.
- *The Morphological Level*: It contains various features such as morphological phenomena, morphological formalisms, and approaches. Among them, we can cite the linguistic approach which segments a text to elementary units that have a linguistic knowledge attached: grammatical category, gender, number, time, person, and so on. The statistical approach analysis starts by splitting sentences into words. Then, a cost is attributed to each bi-gramme according to the calculated apparition frequency in a corpus. Finally, the solution which has the lowest cost is chosen as the best probable. The hybrid approach combines linguistic and statistical criteria. It extracts the relevant terms from both the text statistical analysis and the linguistic filtering of the candidate terms. It produces a sorted list of the most representative terms for a specific domain. The requirement of such an approach can be illustrated as follows: when a developer intends to build a morphological application, using the linguistic approach, he has to take away all LingWS using other kinds of approaches.
- *The Syntactic Level*: Different specificities can characterize a syntactic LingWS such as the syntactic phenomena, analysis, and formalisms. As an example, for the analysis type, we can mention: the Top-Down analysis where the analysis begins from the starting symbol called axiom and tries to rebuild the derivation tree by a prefixed left-right course. The Bottom-Up analysis factorizes the word by picking out or recognizing the right parts of production until finding the axiom. The Deep analysis produces a formal representation of the sentences, under a syntactic tree form. The Surface or Chunking identifies the components limits (i.e., Nominal Group (NG) and Verbal Group (VG)). Finally, the Structural analysis is based on a set of rules to find associations between words in order to construct sentences. We can choose a Linguistic Web service dealing with a particular phenomenon (e.g., Anaphora) using a special formalism (e.g., Tree Adjoint Grammar, Unification Grammar).
- *The Semantic Level*: To develop a semantic application, we can choose some linguistic properties, like the semantic formalisms, phenomena, and resources. The used resource is relevant information. Indeed, if a developer wants to compose an application which needs a Wordnet resource, then he has to eliminate LingWS using a Lexical Markup Framework (LMF[1]) resource for instance.

To sum up, the nonfunctional linguistic properties represent some constraints in the composition task. Indeed, such properties may indicate the incoherence between LingWS as whether to use compatible formalisms or not right from the discovery task. Moreover, the composition can be done according to the personal requirements of the developer, mainly his choice of a particular linguistic resource.

3. Semantic approaches

To deal with the description of the Web Service issues, the software engineering domain provides several approaches that use semantic models (e.g., ontologies) to describe services. Among the well-known approaches, we can mention OWL-S (W3C, 2004), WSMO (W3C, 2004), and SAWSDL (W3C, 2007).

- *OWL-S*: The OWL for Services proposes an ontology of services. It provides three essential types of knowledge about a Web Service: The first is the *Profile* which is used to advertise the service. The service profile elements include preconditions, inputs, outputs, results, and service category. The second is the *Process* which includes inputs, outputs, preconditions, effects, and the behavior of the service (e.g., data and flow control). The third is called *Grounding*, which provides the needed details about the transport protocols.
- *WSMO*: The Web Service Modelling Ontology is an ontology which describes the different aspects of the dynamic composition of Web Services. It describes services using the Web Service Modelling Language[2] (WSML), consisting of four core elements:

[1] http://www.lexicalmarkupframework.com
[2] http://www.wsmo.org/TR/d16/d16.1/v0.21/