Future Generation Computer Systems ▮ (▮▮▮▮) ▮▮▮–▮▮▮

Contents lists available at ScienceDirect

Future Generation Computer Systems

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

Creating a semantically-enhanced cloud services environment through ontology evolution

Miguel Ángel Rodríguez-García [a], Rafael Valencia-García [a,*], Francisco García-Sánchez [a], J. Javier Samper-Zapater [b]

[a] Departamento de Informática y Sistemas, Universidad de Murcia, Campus de Espinardo 30100 Murcia, Spain

[b] Departament d'Informàtica, Escola Tècnica Superior d'Enginyeria, Universitat de València. Av. de la Universitat 46100 Burjassot, València, Spain

HIGHLIGHTS

- Cloud resources are automatically annotated with semantic content.
- The semantic annotation approach supports document and ontology evolution.
- Semantic information improves cloud service discovery and selection processes.

ARTICLE INFO

Article history:
Received 27 February 2013
Received in revised form
15 July 2013
Accepted 2 August 2013
Available online xxxx

Keywords:
Semantic annotation
Ontology evolution
Cloud computing
Information extraction
Knowledge management
Ontology
Semantic web

ABSTRACT

Currently, the availability of Web resources has grown enormously to the point that whatever a user needs at a given moment can potentially be found on the Internet. These resources are not limited to data items anymore, functionality delivered through some sort of service architectural model is also offered on the Internet. In the last few years, cloud computing has emerged as one of the most popular computing models to provide services over the Internet. However, as the number of available cloud services increases, the problem of service discovery and selection arises. Experience indicates that semantic technologies can provide the basis for enhanced and more precise search processes. In this paper, we present a platform that makes use of semantic technologies and techniques to facilitate the discovery of cloud resources meeting the users' needs. We propose an architecture that puts together semantic annotation techniques, ontology evolution, term extraction and indexing resources to semantically annotate cloud services, and a semantic search engine that leverages the semantic description of the cloud resources to find them from keyword-based searches. A comprehensive evaluation of the tool in the ICT domain has produced very promising results and is also presented in this article.

1. Introduction

Experts define cloud computing as a "*model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction*" [1]. Additionally, cloud computing can be seen as a technological paradigm shift that permits to offer computing services over the Internet using a pay-per-use model. This later feature of the cloud makes it especially appealing for companies in the current socio-economic climate. Consequently, according to studies by Gartner Inc., one of the most prominent IT analyst firms, cloud computing is the primary source of growth in IT expending. The National Institute of Standards and Technology (NIST) has established a group focused on promoting the effective and secure use of cloud technology within government and industry. They identify three main service models, namely, Software-as-a-Services (SaaS), Platform-as-a-Service (PaaS), and Infrastructure-as-a-Service (IaaS) [1]. The former is the model representing the actual applications that end users leverage to accomplish their objectives. PaaS is the model representing frameworks and common functions on which applications can be built. The latter refers to the services that provide processing, storage, networks, and other fundamental computing resources, on which consumers can deploy arbitrary software. In this manuscript, when we use the term "cloud service" we are referring to services in any of the three categories just mentioned (SaaS, PaaS, and IaaS). As the cloud computing model is gaining momentum, more and more cloud services are becoming available. While service consumers can clearly benefit from such a broad service offer, they are significantly hampered by the time-consuming task of manual service discovery. Indeed, the exponential increase

* Corresponding author. Tel.: +34 686892581; fax: +34 868884151.
E-mail address: valencia@um.es (R. Valencia-García).

0167-739X/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.future.2013.08.003

in the number of services available makes it harder for users to find the ones that meet their needs. So, one of the main concerns in the cloud computing community is that of service search and retrieval [2,3].

Under these circumstances, the authors believe that the success in the application of semantic technologies in other environments to enhance search processes (e.g., semantic Web Services [4]) can be seamlessly applied in the cloud context. Semantic Web technologies [5] are currently achieving a certain degree of maturity. The semantic Web was conceived with the aim of adding semantics to the data published on the Web (i.e., establish the meaning of the data), so that machines are able to process these data in a similar way a human can do. Semantic technologies have been applied with success in several domains such as tourism [6,7], finances [8,9], e-learning [10], leisure [11,12], or bioscience [13,14]. Ontologies are the paramount technology of the semantic Web. An ontology can be defined as "*a formal and explicit specification of a shared conceptualization*" [15]. Ontologies provide a formal, structured knowledge representation, with the advantage of being reusable and shareable. Knowledge in ontologies is mainly formalized using five kinds of components: classes, relations, attributes, axioms and instances. The use of ontologies and semantic technologies can overcome the limitations of traditional search tools [16,17], and they have also proved to be useful in the scope of the mechanisms related to, for instance, Web Browsing [18,19], Service Discovery [4], Recommendation [20], and Information Management [21].

Given that finding the appropriate services manually is time consuming, the addition of semantics to the cloud landscape can help automate the search process [22]. This approach requires the generation of semantic descriptions for all the available cloud services. This, if manually performed, would be again a very time consuming task. Instead, it is possible to apply natural language processing (NLP) techniques to the documents describing the services for adding semantic annotations to them. The computer can then assist the user in discovering the appropriate services by leveraging the semantic information available. In line with this, in this paper we present an ontology evolution-based framework that enables the semantic annotation of the services available in the cloud. The annotations attached to the cloud services are then exploited in order to carry out a semantically-enhanced search over the cloud services space. The ultimate goal is to improve the processes related with the search and retrieval of cloud services that meet the users' needs. The proposed architecture puts together (1) semantic annotation techniques, ontology evolution, term extraction and indexing resources to semantically annotate cloud services, and (2) a semantic search engine that leverages the semantic description of the cloud resources to find them from keyword-based searches. The main contribution of this work is the design and development of a semantic platform for cloud services annotation and retrieval. The platform aims at alleviating the burdens of manually discovering the appropriate cloud services. For this, a semantic annotation mechanism is proposed that advances the state-of-the-art in several ways. Our approach provides support for multiple ontologies, can deal with unstructured documents, supports both document and ontology evolution, and is fully automatic. In all, the main difference between our proposal and other related works is the possibility to automatically annotate cloud services from their natural language descriptions available in several digital formats, and thus assist in enhancing the precision of the service search processes.

The rest of the paper is organized as follows. Some related work is presented in Section 2. The functional architecture of the proposed solution along with its constituent components is described in Section 3. In Section 4, a use case scenario in the information and communications technologies (ICT) domain is shown. Finally, conclusions and future work are put forward in Section 5.

2. Related work

The concept of annotation can be defined as "*the practice of adding interpretative linguistic, information to a corpus*" [23]. More concretely, annotation or tagging is a process that permits to map attributes, comments, or descriptions to a document or to a fragment in a text. In general, annotations can be seen as extra information associated with a particular point in a document or another piece of information. Annotation systems can be classified into three different categories: manual (performed by one or more people), semi-automatic (based on automatic suggestions) [24] or simply automatic (based on computer annotations processes).

Semantic annotations help bridge the ambiguity of natural language and its computational representation in a formal language through ontologies. The process basically consists of inserting tags in a document. These tags represent links between text fragments and ontological elements (attributes, concepts, relationships and instances). As a result of this process, documents are created that can be processed not only by humans but also by automated agents [25]. According to [26], these systems can be classified based on the kind of annotation method used. There are two primary categories, namely, Pattern-based annotation and Machine Learning-based annotation. The former is based on discovery rules, and the latter relies on probabilistic and induction techniques.

Although in the last decade several systems for ontology-based annotation have been proposed, there is not a standard approach for semantic annotation [27]. Table 1 shows some of the most well-known semantic annotation systems as compared against our approach, which is presented in the last row. The parameters selected for their representation in Table 1 are the following: 'standard format', 'ontology support', 'heterogeneous document formats support', 'document evolution', 'automation', and 'ontology evolution'. Next, these parameters are explained in detail.

Standard formats.

Most of the annotation tools developed up to 2006 use the RDF format for the annotations (e.g., Armadillo [28], MnM [34], S-CREAM [36] and CREAM [30]) or RDFa[1] for embedding rich metadata within web documents. The tendency in the last few years, however, has been to utilize OWL ontologies as the annotation format, as described in CERNO [29], EVONTO [31], GoNTogle [32], KIM [33], and Onto-Mat [35]. Finally, there exist other approaches, such as the work presented in [29], that make use of alternative annotation schemas based on text files generated by the TXL programming language [37].

In this work, the second version of OWL, OWL 2 [38], has been used. Its formal model supports a number of important automatic DL inference services, which can be provided by different DL reasoners including HermiT, Pellet2, Fact++ or Racer [39].

Multiple ontologies support.

One property that is often desired in the scope of semantic annotation is that of multiple ontologies support. The tools that offer such a feature can support several ontologies in different domains. The main advantage of semantic annotation systems supporting multiple ontologies is that they can cover different domains. While KIM [33], CREAM [30], GoNTogle [32], or Armadillo [28] support the use of multiple ontologies, CERNO [29], S-CREAM [36], MnM [32], EVONTO [31] or Onto-Mat [35] do not include this feature. Our approach does indeed provide support for multiple ontologies.

Heterogeneous document formats support.

The main approaches for semantic annotation focus on dealing with texts available on the Web, and so the documents being

[1] http://www.w3.org/TR/rdf-primer.

Please cite this article in press as: M.Á. Rodríguez-García, et al., Creating a semantically-enhanced cloud services environment through ontology evolution, Future Generation Computer Systems (2013), http://dx.doi.org/10.1016/j.future.2013.08.003