Knowledge-Based Systems 56 (2014) 15–25

Contents lists available at ScienceDirect

Knowledge-Based Systems

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

Ontology-based annotation and retrieval of services in the cloud

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

ARTICLE INFO

Article history:
Received 19 July 2012
Received in revised form 2 October 2013
Accepted 6 October 2013
Available online 17 October 2013

Keywords:
Ontology
Semantic web
Semantic annotation
Cloud computing
Knowledge management

ABSTRACT

Cloud computing is a technological paradigm that permits computing services to be offered over the Internet. This new service model is closely related to previous well-known distributed computing initiatives such as Web services and grid computing. In the current socio-economic climate, the affordability of cloud computing has made it one of the most popular recent innovations. This has led to the availability of more and more cloud services, as a consequence of which it is becoming increasingly difficult for service consumers to find and access those cloud services that fulfil their requirements. In this paper, we present a semantically-enhanced platform that will assist in the process of discovering the cloud services that best match user needs. This fully-fledged system encompasses two basic functions: the creation of a repository with the semantic description of cloud services and the search for services that accomplish the required expectations. The cloud service's semantic repository is generated by means of an automatic tool that first annotates the cloud service descriptions with semantic content and then creates a semantic vector for each service. The comprehensive evaluation of the tool in the ICT domain has led to very promising results that outperform state-of-the-art solutions in similarly broad domains.

1. Introduction

The economy has recently taken a downturn, and this has forced many companies to reduce their costs in IT. According to experts, one interesting approach that could be used to deal with this situation might be for these companies to outsource parts of their business through third parties [37]. This would allow companies to free any resources invested in outsourced business processes and to focus their investments on those processes that are unique to the organisation. The future Internet will be based on services, and this new trend will have a significant impact on domains such as e-Science, education and e-Commerce [40]. The Web is consequently evolving from being a mere repository of information to a new platform for business transactions and information interchange. Large organisations are increasingly exposing their business processes through Web service technology for both the large-scale development of software and the sharing of their services within and outside the organisation. New paradigms for software and service engineering, such as Software-as-a-Service (SaaS) and the cloud computing model, are promising sources for the creation of new levels of efficiency through the large-scale sharing of functionality and computing resources.

Software-as-a-Service (SaaS) [2] is a software distribution model which not only provides software but also offers additional services such as maintenance, help and support. The advantage of the use of SaaS is that the software is distributed and hosted on the Internet, thus eliminating the need for users to install the software and any related data on their personal computers. The software and all of the additional data required are stored on an external server, which is provided by the company that offers these services. But SaaS is only one of the different service delivery models that are envisioned by the cloud computing movement. Cloud computing [9] represents a paradigm shift in the delivery architecture of information services. 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. This body defines 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*" [29]. Over the last few years, the growth of the cloud computing industry has led to an exponential increase in the number of services being delivered from the cloud, and one of the direct consequences of this increase has been a considerable rise in the time and effort required to manually find the service that best meets the needs of an individual or an organisation.

* Corresponding author. Tel.: +34 868888522; fax: +34 868884151.
E-mail address: valencia@um.es (R. Valencia-García).

0950-7051/$ - see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.knosys.2013.10.006

Semantic web technologies have simultaneously achieved a certain degree of maturity [41]. They provide a consistent and reliable basis with which to confront the challenges involved in the organisation, manipulation and visualisation of data and knowledge. The possibility of using knowledge-oriented query answering to exploit the benefits of semantics has consequently become a top-class research challenge. Ontologies are the paramount technology of the semantic web. An ontology can be defined as "a formal and explicit specification of a shared conceptualisation" [42]. Ontologies provide a formal, structured knowledge representation, with the advantage that they are reusable and shareable. Ontologies provide a common vocabulary for a domain and define the meaning of the terms and the relations between them with different levels of formality. The knowledge in ontologies is mainly formalised, and five kinds of components are used: classes, relations, attributes, axioms and instances. The definition of ontologies is sometimes diluted in the sense that taxonomies are considered to be full ontologies [42].

A large number of generic keyword-based search engines, such as Google, Yahoo! Search, and Bing have appeared in order to facilitate the search for and discovery of cloud services. These search engines crawl the entire World Wide Web and create indexes of the pages visited, thus enabling users to find almost everything on the Web, including cloud services [17]. However, these search engines have certain drawbacks resulting from the syntactic nature of their search processes, which hampers the discovery of those services that meet the required criteria. The use of ontologies and semantic technologies can overcome the limitations of traditional search tools [28], and they have also proved to be useful in the scope of mechanisms related to, for instance, Information Retrieval [45], Information Integration [47], Service Discovery [14], Question Answering [46], Recommendation [13], and information management [7]. Ontologies have also been applied to numerous domains such as business management [27], eGoverment [1] and e-Learning [3,38]. More concretely, the use of semantic technologies and metadata improves traditional search engine optimisation [49]. The downside of applying semantics in search processes is that Web content should be annotated, and manual annotation is a time consuming and expensive process [6]. What is more, manual annotations are usually prone to errors resulting from a number of factors such as user familiarity, personal motivation or the complexity of ontologies, which may negatively affect the annotation procedure [36].

The aim of the work presented here is to deal with these two major issues: (1) the semantic annotation of the features of cloud services and (2) the discovery of cloud services that meet users' needs. The first problem is addressed by proposing an automatic semantic annotation platform that can assists users in the process of adding metadata to Web content including cloud services descriptions. The second issue is also addressed here by building a semantic search engine that leverages the cloud service-related semantic annotations in order to improve the precision and recall of the search results.

The automatic semantic annotation platform proposed here is based on the approaches presented in the literature in the field of ontology learning and population. The main goals of these approaches are (i) to reduce the amount of mistakes caused by the manual extraction of semantic information and (ii) to increase the amount of knowledge gathered. However, current semantic annotation methods have some limitations at different levels. Some tools do not provide support for multiple ontologies, and others are not prepared to deal with ontology or document evolution. The approach presented here is able to manage multiple ontologies and provides support for both ontology evolution and document evolution.

The annotation platform gathers semantic content from cloud service descriptions, and the semantic search engine takes advantage of this content to assist users in finding those services that comply with their requirements and expectations. This breakthrough provides a different way of looking for services in the cloud, and is based on natural language processing. The idea is that users should write a brief description of their requirements in natural language. The proposed natural language processing-based framework then translates this description into ontology elements which are matched against the semantic cloud service descriptions. The whole process results in the proposed tool retrieving the list of services that best suit the users' needs.

To sum up, in this paper we present a fully-fledged platform based on semantic technologies that assists in the annotation of cloud service descriptions and enables an enhanced search and retrieval of cloud services. The remainder of the paper is organised as follows. Section 2 provides a description of the most important semantic annotation methods. The components that comprise the proposed platform and its overall architecture are described in Section 3. Section 4 shows a use case scenario in the information and communications technologies (ICT) domain, and a validation and a comparison with the current semantic annotation systems are also provided. Finally, our conclusions and future work are presented in Section 5.

2. Related work

Leech [26] defined annotation as "*the practise of adding interpretative linguistic, information to a corpus*". More concretely, annotation or tagging is a process that permits the mapping of concepts, relationships, comments, or descriptions to a document or to a fragment in a text. In general, annotations can be seen as metadata that are associated with a particular text fragment in a document or another piece of information. Semantic annotations help to bridge the ambiguity of the natural language and their computational representation in a formal language through ontologies. This process consists of inserting tags associated to text fragments in a document that represent ontological elements (concepts, relationships, attributes and instances), thus allowing the creation of documents that can be processed not only by humans but also by automated agents [23]. Although several systems for ontology-based annotation have been proposed over the last decade, there is no standard approach for semantic annotation [44]. The following table (see Table 1) shows some of the best known semantic annotation systems. The classification is based on that presented in [44].

The parameters selected for their representation in the table are the following: 'standard format', 'ontology support', 'support of heterogeneous document format', 'document evolution', 'annotation storage model', 'automation' and 'application domain'. These parameters are explained in detail as follows, and the differences between the proposed annotation method and the current state-of-the-art mechanisms are also highlighted.

2.1. Standard formats

The World Wide Web Consortium[1] (W3C) provides a number of techniques with which to describe and define formats to build ontologies. These include RDF[2] and RDF Schemas,[3] Simple Knowledge Organization System (SKOS), Web Ontology Language (OWL), and the Rule Interchange Format (RIF).

As can be seen in Table 1, RDF and OWL are the most extended formats in the scope of semantic annotation systems. Most of the

[1] www.w3.org.
[2] www.w3.org/RDF.
[3] www.w3.org/TR/rdf-schema.