

Web Service Clustering using a Hybrid Term-Similarity Measure with Ontology Learning

Banage T. G. S. Kumara, School of Computer Science and Engineering, University of Aizu, Aizu-Wakamatsu, Japan

Incheon Paik, School of Computer Science and Engineering, University of Aizu, Aizu-Wakamatsu, Japan

Wuhui Chen, School of Computer Science and Engineering, University of Aizu, Aizu-Wakamatsu, Japan

Keun Ho Ryu, School of Electrical and Computer Engineering, Chungbuk National University, Cheongju, South Korea

ABSTRACT

Clustering Web services into functionally similar clusters is a very efficient approach to service discovery. A principal issue for clustering is computing the semantic similarity between services. Current approaches use similarity-distance measurement methods such as keyword, information-retrieval or ontology based methods. These approaches have problems that include discovering semantic characteristics, loss of semantic information and a shortage of high-quality ontologies. In this paper, the authors present a method that first adopts ontology learning to generate ontologies via the hidden semantic patterns existing within complex terms. If calculating similarity using the generated ontology fails, it then applies an information-retrieval-based method. Another important issue is identifying the most suitable cluster representative. This paper proposes an approach to identifying the cluster center by combining service similarity with term frequency-inverse document frequency values of service names. Experimental results show that our term-similarity approach outperforms comparable existing approaches. They also demonstrate the positive effects of our cluster-center identification approach.

Keywords: Information Retrieval, Ontology Learning, Semantic Patterns, Service Similarity, Web Service Clustering

DOI: 10.4018/ijwsr.2014040102

INTRODUCTION

Web services, which share business logic, data and processes through a programmatic interface, represent an important way for businesses to communicate with each other and with clients. The concept of Web services has therefore become a widely applied paradigm in research and industry, with the number of services published on the Internet increasing rapidly over the last few years (Al-Masri & Mahmoud, 2008). With this proliferation of Web services, service discovery is becoming a challenging and time-consuming task. Efficient service discovery is very significant because other major problems in service-oriented computing, such as service selection and composition, are also related to efficient discovery. Reducing the query space by clustering services, thereby avoids many unnecessary similarity calculations in the matching process, and is an efficient approach to increasing the performance of service discovery. Web services can be clustered into functionally similar clusters by considering functional attributes such as *input*, *output*, *precondition* and *effect* (Dasgupta, Bhat et al., 2011). Alternatively, services can be clustered in terms of quality of service by considering their nonfunctional properties, such as cost and reliability (Xia, Chen et al., 2011). Some recent studies have proposed clustering services in terms of social properties (Chen, Paik et al., 2013). Here, we consider functional clustering.

One main issue in clustering is calculating the similarity or affinity between services. Recent studies have proposed several approaches to calculating functional similarity. Simple approaches include checking the one-to-one matching of features such as the *service name* and checking the matching of service signatures such as the *messages* (Elgazzar, Hassan et al., 2010). In some studies, information retrieval (IR) techniques are used (Platzer, Rosenberg et al., 2009). These include similarity-measuring methods such as search-engine-based (SEB) methods (Liu & Wong 2009) and cosine similarity (Chen, Yang et al., 2010; Ma, Zhang et

al., 2008). Some researchers have used logical relationships such as *exact* and *plug-in* (Wagner, Ishikawa et al., 2011) or edge-counting-based techniques (Xie, Chen et al., 2011; Sun, 2010) to increase the semantics in the similarity calculations via ontologies. However, one-to-one matching, structure matching or a vector-space model may not accurately identify the semantic similarity among terms because of the heterogeneity and independence of service sources. These methods consider terms only at the syntactic level, whereas different service providers may use the same term to represent different concepts or may use different terms for the same concept. Furthermore, IR techniques such as cosine similarity usually focus on plain text, whereas Web services contain much more complex structures, often with very little textual description. This means that depending on IR techniques is very problematic. Moreover, there can be a loss of the machine-interpretable semantics found in service descriptions when converting data provided in service descriptions into vectors in IR techniques. In SEB similarity-measuring methods such as normalized Google distance (NGD), there is no guarantee that all the information needed to measure the semantic similarity between a given pair of words is contained in the top-ranking snippets. On the other hand, although ontologies help to improve semantic similarity, defining high-quality ontologies is a major challenge. Several methods have been used to develop ontologies in current approaches, including obtaining assistance from domain expertise, using resources such as WordNet (<http://wordnet.princeton.edu/>, n.d.) and using ontologies already available via the Internet (Xie, Chen et al., 2011). Developing ontology by obtaining assistance from domain expertise is a time-consuming task that requires considerable human effort. In addition, the lack of up-to-date information in a resource might fail to capture the latest concepts and relationships in a domain. Further, the lack of standards for integrating and reusing existing ontologies also hampers ontology-based (OB) semantics matching.

20 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:

www.igi-global.com/article/web-service-clustering-using-a-hybrid-term-similarity-measure-with-ontology-learning/116601

Related Content

Navigation Route based Stable Connected Dominating Set for Vehicular Ad Hoc Networks

Yishun Chen, Weigang Wu and Hui Cao (2015). *International Journal of Web Services Research* (pp. 12-26).

www.irma-international.org/article/navigation-route-based-stable-connected-dominating-set-for-vehicular-ad-hoc-networks/125456/

UML-Based Support for Designing and Validating Web Service Descriptions

Juanjuan. Jiang and Tarja Systa (2006). *International Journal of Web Services Research* (pp. 101-120).

www.irma-international.org/article/uml-based-support-designing-validating/3081/

Improving Recommendation Accuracy and Diversity via Multiple Social Factors and Social Circles

Yong Feng, Heng Li and Zhuo Chen (2014). *International Journal of Web Services Research* (pp. 32-46).

www.irma-international.org/article/improving-recommendation-accuracy-and-diversity-via-multiple-social-factors-and-social-circles/124984/

The Design of QoS Broker Algorithms for QoS-Capable Web Services

Tao Yu and Kwei-Jay Lin (2004). *International Journal of Web Services Research* (pp. 33-50).

www.irma-international.org/article/design-qos-broker-algorithms-qos/3049/

Data Mining Location-Based Social Networks for Geospatial Discovery

Edward Pultar (2012). *Discovery of Geospatial Resources: Methodologies, Technologies, and Emergent Applications* (pp. 204-218).

www.irma-international.org/chapter/data-mining-location-based-social/65115/