

Chapter 16

Predicting Similarity of Web Services Using WordNet

Aparna Konduri
University of Akron, USA

Chien-Chung Chan
University of Akron, USA

ABSTRACT

As vast numbers of web services have been developed over a broad range of functionalities, it becomes a challenging task to find relevant or similar web services using web services registry such as UDDI. Current UDDI search uses keywords from web service and company information in its registry to retrieve web services. This method cannot fully capture user's needs and may miss out on potential matches. Underlying functionality and semantics of web services need to be considered. This chapter introduces a methodology for predicting similarity of web services by integrating hierarchical clustering, nearest neighbor classification, and algorithms for natural language processing using WordNet. It can be used to facilitate the development of intelligent applications for retrieving web services with imprecise or vague requests. The authors explore semantics of web services using WSDL operation names and parameter names along with WordNet. They compute semantic interface similarity of web services and use this data to generate clusters. Then, they represent each cluster by a set of characteristic operations to predict similarity of new web services using nearest neighbor approach. The empirical result is promising.

INTRODUCTION

Web Services offer a promise for integrating business applications within or outside an organization. They are based on Service Oriented Architecture (SOA) (Barry, 2003) that provides loose coupling between software components via standard interfaces. Web Services expose their interfaces using

Web Service Description Language (WSDL) (Christensen, Curbera, Meredith, & Weerawarana, 2001). WSDL is an XML based language and hence platform independent. A typical WSDL file provides information such as web service description, operations that are offered by a web service, input and output parameters for each web service operation. Web Service providers use a central repository called UDDI (Universal Description, Discovery and Integration) (Clement et al., 2004) to advertise and

DOI: 10.4018/978-1-61520-757-2.ch016

publish their services. Web Service consumers use UDDI to discover services that suit their requirements and to obtain the service metadata needed to consume those services. Users that want to use a web service will utilize this metadata to query the web service using SOAP (Simple Object Access Protocol) (Mitra, 2003). SOAP is a network protocol for exchanging XML messages or data. Since SOAP is based on HTTP/HTTP-S, it can very likely get through network firewalls. The advantages of XML and SOAP give web services their maximum strength.

With web applications and portals getting complex and rich in functionality, many users are interested in finding similar web services. Users might want to compose two operations from different web services to obtain complex functionality. Also, users might be interested in looking at operations that take similar inputs and produce similar outputs. Let us say, web service *A* has an operation *GetCityNameByZip* that returns city name by zip code, web service *B* has an operation *GetWeatherByCityName* that returns weather by city name and web service *C* has an operation *GetGeographicalLocationBasedOnZip* that returns city name, longitude, latitude and altitude of a location by zip code. Operations from web services *A* and *B* are related, i.e., output from one operation can be used as an input to another. So, these operations can be composed to obtain weather by city name. Operations from web services *A* and *C* are similar. They take similar inputs. Outputs are also similar, i.e., output of operation from web service *C* is fine grained when compared to output of operation from web service *A*.

As more and more web services are developed, it is a challenge to find the right or relevant web services quickly and efficiently. Currently, UDDI supports keyword match just based on web service data entries in its registry. This might potentially miss out on some valid matches. For example, searching UDDI with keywords like zip code may not retrieve web service with postal code information.

Semantics of a web service in terms of the requirements and capabilities of a web service can be really helpful for efficient retrieval of web services. WSDL does not have support for semantic specifications. A lot of researches are done on annotating web services through special markup languages to attach semantics to a web service. Akkiraju et al. proposed WSDL-S to annotate web services (Akkiraju et al., 2005). Cardoso and Sheth used DAML-S (Cardoso & Sheth, 2003). DAML-S was used as annotations to compose multiple web services by (Ankolekar, Burstein, Hobbs, Lassila, Martin, McIlraith, Narayanan, Paolucci, Payne, Sycara, & Zeng, 2001). The OWL-S introduced by (Martin et al., 2004) was used by Ganjisaffar et al. as annotations to compute similarity between web services (Ganjisaffar, Abolhassani, Neshati, & Jamali, 2006), and a semantics-based composition-oriented discovery of web services based on OWL-S was introduced in (Brogi, Corfini, & Popescu, 2008). The main challenges of ontology-based approach are that manual annotation is a time consuming task and how to deal with discrepancies between ontologies.

Another line of research has been done to automate the process of discovery or composition of relevant web services in response to a given query based on the contents of WSDL files. Normally, heuristics are used to infer the functionality or semantics of a web service based on its description, operations along with parameters that these operations take. Wang and Stroulia have used vector space model enhanced by WordNet (Fellbaum, 1998) to perform semantic structure matching for assessing web service similarity (Wang & Stroulia, 2003a, 2003b). Dong et al. built a web search engine called Woogle based on agglomerative clustering of WSDL descriptions, operations and parameters (Dong et al., 2004). Wu and Wu provided a suite of similarity measures to assess the web service similarity (Wu & Wu, 2005). Kil et al. proposed a flexible network model for matching web services (Kil, Oh, & Lee, 2006). A kernel based web services

14 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/chapter/predicting-similarity-web-services-using/42368

Related Content

A Single Pass Algorithm for Discovering Significant Intervals in Time-Series Data

Sagar Savla and Sharma Chakravarthy (2007). *International Journal of Data Warehousing and Mining* (pp. 28-44).

www.irma-international.org/article/single-pass-algorithm-discovering-significant/1788

HYBRIDJOIN for Near-Real-Time Data Warehousing

M. Asif Naeem, Gillian Dobbie and Gerald Weber (2011). *International Journal of Data Warehousing and Mining* (pp. 21-42).

www.irma-international.org/article/hybridjoin-near-real-time-data/58636

An Outlier Detection Algorithm Based on Probability Density Clustering

Wei Wang, Yongjian Ren, Renjie Zhou and Jilin Zhang (2023). *International Journal of Data Warehousing and Mining* (pp. 1-20).

www.irma-international.org/article/an-outlier-detection-algorithm-based-on-probability-density-clustering/333901

The Importance of Data Within Contemporary CRM

Diana Luck (2009). *Data Mining Applications for Empowering Knowledge Societies* (pp. 96-109).

www.irma-international.org/chapter/importance-data-within-contemporary-crm/7548

Data

(2023). *Principles and Theories of Data Mining With RapidMiner* (pp. 35-49).

www.irma-international.org/chapter/data/323367