Latent Semantic Analysis – The Dynamics of Semantics Web Services Discovery

Chen Wu, Vidyasagar Potdar, and Elizabeth Chang

Digital Ecosystems and Business Intelligence Institute
Curtin University of Technology, Perth 6845, WA, Australia
{Chen.Wu,Vidyasagar.Potdar,Elizabeth.Chang}@cbs.curtin.edu.au
http://debii.curtin.edu.au

Abstract. Semantic Web Services (SWS) have currently drawn much momentum in both academia and industry. Most of the solutions and specifications for SWS rely on ontology building, a task needs much human (e.g. domain experts) involvement, and hence cannot scale very well in face of vast amount of web information and myriad of services providers. The recent proliferation of SOA applications exacerbates this issue by allowing loosely-coupled services to dynamically collaborate with each other, each of which might maintain a different set of ontology. This chapter presents the fundamental mechanism of Latent Semantic Analysis (LSA), an extended vector space model for Information Retrieval (IR), and its application in semantic web services discovery, selection, and aggregation for digital ecosystems. First, we explore the nature of current semantic web services within the principle of ubiquity and simplicity. This is followed by a succinct literature overview of current approaches for semantic services/software component (e.g. ontology-based OWL-s) discovery and the motivation for introducing LSA into the user-driven scenarios for service discovery and aggregation. We then direct the readers to the mathematical foundation of LSA – SVD of data matrices for calculating statistics distribution and thus capturing the 'hidden' semantics of web services concepts. Some existing applications of LSA in various research fields are briefly presented, which gives rise to the analysis of the uniqueness (i.e. strength, limitations, parameter settings) of LSA application in semantic web services. We provide a conceptual level solution with a proof-of-concept prototype to address such uniqueness. Finally we propose an LSA-enabled semantic web services architecture fostering service discovery, selection, and aggregation in a digital ecosystem.

1 Introduction

Semantics play an important role in the complete lifecycle of Web services as it is able to help service development, improve service reuse and discovery, significantly facilitate composition of Web services and enable integration of legacy applications as part of automatic business process integration. Unfortunately, current Web Service Description Language (WSDL) standard operates at the syntactic level and lacks the semantic expressivity needed to represent the requirements and capabilities of Web

T.S. Dillon et al. (Eds.): Advances in Web Semantics I, LNCS 4891, pp. 346–373, 2008.

Services. This gap has motivated a lot existing research effort towards the Semantic Web Services (SWS). The fundamental idea underlying current SWS community is that in order to achieve machine-to-machine integration, a markup language must be descriptive enough that a computer can automatically determine its meaning. Following this principle, many semantic annotation markup languages have thus come into view, among them are OWL-S (formerly known as DAML-S) and WSDL-S that have gained great momentum in recent years. The main goal of both OWL-S and WSDL-S is to establish a framework within which service descriptions are made and shared.

The premise of such an ontology-based markup language approach is that every SWS user (be it normal website or end customer) is able to employ a standard ontology, consisting of a set of basic classes and properties, for declaring and describing services. One concern about this descriptive annotation-driven approach is its feasibility: since it would be much more time-consuming to create and publish ontology-annotated (WSDL) content as they would need to be done by domain human experts and powerful editing tools for common users. Other problems might occur when different groups of users and communities want to manage the shared ontology. With this being the case, it would be much less likely for industry companies to adopt these practices as it would only slow down their progress.

In this chapter we carry out SWS research, in particular the service discovery, from another empirical perspective. We believe that one thing distinguishes Web (services) semantics from other forms of semantics is its 'user-centred' commitment towards ubiquity and simplicity, the two most renowned factors leading to the great success of today's Web. By ubiquity, we mean that the underlying technology (such as HTTP and TCP/IP) has to be very robust, lightweight, and non-human intervened to serve for various applications and users. By simplicity, we mean millions of end users can easily access to and personally use the technology without too much expertise in both domain and IT areas. This idea drives us to come up with a novel method to approach the SWS using Latent Semantic Analysis (LSA) technique – the main theme of this chapter. Nevertheless, proposing alternate approach does not mean we completely go against ontology-based approach. On the contrary, we acknowledge that adding semantics in the form of ontology to represent the requirements and capabilities of Web services is essential for achieving unambiguity and machine-interpretability for web services, and hence becomes our long term research objectives. For example, we are currently seeking effective ways that can convert some of our research result – the higher-order association (the very initial stage semantic space) in this chapter – into lightweight ontology in a semi-automatic manner.

The chapter is organised as follows. Section 2 presents OWL-S and WSDL-S, the two widely accepted SWS specifications and characteristics. Section 3 provides in-depth mathematics technique and working mechanism on LSA. Applications of LSA in IR, cognitive, and psychology are introduced in Section 4. This is followed by the rationale to apply LSA in the area of SWS, which has some issues to address. Section 5 proposes the conceptual model of LSA-based SWS. Section 6 then presents semantic search engine prototype based on our conceptual model. Experiment results are reviewed in Section 7, where three LSA and one WSDL parameters are manipulated to gain further understanding of our approach. The chapter concludes in Section 8.