Available online at www.sciencedirect.com
SCIENCE DIRECT®
Web Semantics: Science, Services and Agents on the World Wide Web 3 (2005) 340–365
JOURNAL OF Web Semantics
www.elsevier.com/locate/websem

# Learning domain ontologies for semantic Web service descriptions

Marta Sabou [a,*], Chris Wroe [b], Carole Goble [b], Heiner Stuckenschmidt [a]

[a] *Department of Artificial Intelligence, Vrije Universiteit Amsterdam, De Boelelaan 1081a, 1081 HV Amsterdam, The Netherlands*
[b] *Department of Computer Science, University of Manchester, UK*

**Abstract**

High quality domain ontologies are essential for successful employment of semantic Web services. However, their acquisition is difficult and costly, thus hampering the development of this field. In this paper we report on the first stage of research that aims to develop (semi-)automatic ontology learning tools in the context of Web services that can support domain experts in the ontology building task. The goal of this first stage was to get a better understanding of the problem at hand and to determine which techniques might be feasible to use. To this end, we developed a framework for (semi-)automatic ontology learning from textual sources attached to Web services. The framework exploits the fact that these sources are expressed in a specific *sublanguage*, making them amenable to automatic analysis. We implement two methods in this framework, which differ in the complexity of the employed linguistic analysis. We evaluate the methods in two different domains, verifying the quality of the extracted ontologies against high quality hand-built ontologies of these domains.

Our evaluation lead to a set of valuable conclusions on which further work can be based. First, it appears that our method, while tailored for the Web services context, might be applicable across different domains. Second, we concluded that deeper linguistic analysis is likely to lead to better results. Finally, the evaluation metrics indicate that good results can be achieved using only relatively simple, off the shelf techniques. Indeed, the novelty of our work is *not* in the used natural language processing methods but rather in the way they are put together in a generic framework specialized for the context of Web services.


*Keywords:* Ontology learning; Semantic Web services; Ontology learning evaluation

In the last few years the Web encountered two revolutionary changes which aim to transform it from a static document collection in an intelligent and dynamic data integration environment. First, the Web service technology allowed uniform access via Web standards to software components residing on various platforms and written in different programming languages. As a result, software components providing a variety of functionalities (ranging from currency conversion to flight booking or book buying) are now accessible via a set of Web standards and protocols. Naturally, the real value of Web services is in their composition which allows creating new and complex functionalities from the existing services. The second

* Corresponding author. Tel.: +31 20 598 7752; fax: +31 20 598 7653.
*E-mail address:* marta@cs.vu.nl (M. Sabou).

1570-8268/$ – see front matter © 2005 Elsevier B.V. All rights reserved.
doi:10.1016/j.websem.2005.09.008

novel Web technology, the Semantic Web, developed techniques for augmenting existing Web data with logics based formal descriptions of their meaning. This semantic markup is machine processable and therefore facilitates access and integration of the vast amount of Web data.

A major limitation of the Web services technology is that finding and composing services still requires manual effort. This becomes a serious burden with the increasing number of Web services. To address this problem, semantic Web researchers advanced the idea of augmenting Web services with a semantic description of their functionality that could facilitate their discovery and integration. More precisely, Web services are described in terms of concepts provided by a *domain ontology*. These concepts denote entities in the domain of the Web service (e.g., *Food*, *Hotel*) as well as functionalities that can be performed by services in the given domain (e.g., *OrderFood*, *BookHotel*). To ensure high quality reasoning on these semantic Web service descriptions, it is essential that they rely on the use of quality domain ontologies, i.e., ontologies that have a broad coverage of their domain's terminology. This would allow many (if not all) services to use the same or a small number of different ontologies thus reducing the need of mappings at reasoning time. Note that in practice different types of ontologies are used ranging from catalogs of domain concepts to formal domain models.

Despite their importance, few domain ontologies for Web service descriptions exist and building them is a challenging task. One of the problematic aspects is that for building a high quality domain ontology one ideally needs to inspect a large number of Web services in that domain. Since many domains witnessed a rapid increase in the number of available Web services to several hundreds (1000+ in bioinformatics), tools that support ontology curators to build a Web service domain ontology from these *large and dynamic data sets* become crucial.

Our work addresses the problem of (semi-) automatically learning Web service domain ontologies. We report on the first stage of this work in which we aim to get a better understanding of the ontology learning task in the context of Web services and to identify potentially feasible technologies that could be used. Early in our work we learned that the context of Web services raises several issues that constrain the development of an ontology learning solution. We designed a framework for performing ontology learning in the context of Web services which addresses these issues in two ways. First, it exploits the particularities of Web service documentations to extract information used for ontology building. In particular, the sublanguage characteristics of these texts lead to the identification of a set of heuristics. These heuristics are implemented as pattern based extraction rules defined on top of linguistic information. Second, the learned ontologies are suited for Web service descriptions as they contain both static and procedural knowledge.

We implemented two learning methods that follow the basic principles of the framework but use different linguistic knowledge. The first method uses basic Part-of-Speech (POS) information and was developed and tested in the context of the WonderWeb project[1] [52]. The second method uses deeper dependency parsing techniques to acquire linguistic knowledge. It was designed and tested on data sets provided by the *my*Grid project[2] [55]. In this paper we present both methods and compare them by applying and evaluating them in the context of both projects.

This paper is structured as follows. We first present some introductory notions about semantic Web service technology, concluding the important role that Web service domain ontologies play as well as some requirements that they should fulfill (Section 1). Then, we analyze why it is difficult to build such domain ontologies. We do this by describing the process of building domain ontologies in the context of the two research projects that served as case studies for developing and evaluating our framework (Section 2). We conclude Section 2 with an overview of the issues that constrain the development of an ontology learning solution in the Web services context. Then, we present an ontology learning framework that deals with these constraints and the two concrete implementations of this framework in Section 3. Implementation details and some considerations about the usability of the extraction tools are provided in Section 4. In Section 5 we present an overview of existing ontology learning evaluation practices and show how they were adapted for our work. In this section we also detail our experimental results. We

[1] http://wonderweb.semanticweb.org/.
[2] http://www.mygrid.org.uk/.