

Ontology Learning for Semantic Web Services

Auhood Alfaries, David Bell, Mark Lycett

*School of Information Systems, Computing and Mathematics
Brunel University, Uxbridge
United Kingdom*

Email:{a.alfaries |david.bell |mark.lycett @brunel.ac.uk }

Abstract.

Semantic Web Services promise automatic service discovery and composition, relying heavily on domain ontology as a core component. With large Web Service repository, manual ontology development is proving a bottleneck (with associated expense and likely errors) to the realisation of a semantic Web of services. Providing the appropriate tools that assist in and automate ontology development is essential for a dynamic service vision to be realised. As a statement of research-in-progress, this paper proposes combining different ontology learning paradigms in Web Services domain, highlighting the need for further research that accommodates the variation in Web Service descriptive and operational sources. A research agenda is proposed that recognises this variation in artefacts as they are selected, pre-processed and analyzed by ontology learning techniques.

Keywords: Ontology Learning, Web Services, semantic Web Services, Ontology.

1 Introduction

Service Oriented Architecture (SOA) is an emerging architectural approach with the potential to better accommodate organisational change. Web Services are the predominant technological means of delivering on the SOA ideal and there is a clear increase in interest in both the underlying architecture and delivery mechanism (Azoff 2007, Heffner & Peters 2008, Martin 2007b, Tsai et al. 2006, Yu et al. 2008). In practice, however, several barriers exist that militate against the effective use of Web Services – the need for manual intervention in discovery and adoption stands out as a challenge. As recognised by the semantic web community in (Berners-Lee et al. 2001, Shadbolt et al. 2006), Web Services cannot be automatically discovered and composed as the description of those services is not rich enough in its semantics (Martin 2007b).

The development and deployment of Semantic Web Services (SWS) by in the business community is slow however, in good part because their development is an expensive, error prone and labour intensive task. A problem also exists in the fact that

existing ‘stocks’ of Web Services are unlikely decommissioned as they represent significant organisational investments. Ontologies are the general means by which semantics are added into Web Services (Akkiraju et al. 2005, Burstein et al. 2005, Sheth et al. 2006). The challenge therefore is to develop ontologies from existing services and to enable those ontologies to adapt and evolve in line with the domain and any demands made on it (Cuel et al. 2008). Ontology Learning provides an automated means of dealing with these issues, but it is an area that is not well explored.

Given the above, this paper presents early outcomes of research in progress, which develops a research agenda to direct work on Ontology Learning (OL). In achieving this aim, the paper is structured as follows. Section 2 presents an overview of the current status of the development and use of Web Services in industry. Section 3 introduces similar work by discussing OL approaches and, more specifically, OL for semantic Web Services. Section 4 then examines how Ontology Learning Techniques (OLT) can be best used in achieving the full potential of SWS and derives the agenda for research. Section 5 concludes the work.

2 The industry perspective

The need for business systems to be adaptive in the face of organisational change has been met in part by a movement toward Web Services. As background, Figure 1 illustrates key components, roles and operations in a Web Service environment, which are sufficient to allow two parties to share and invoke services remotely given predefined agreement between provider and requester. Currently, service matching requires human intervention to ensure compatibility.

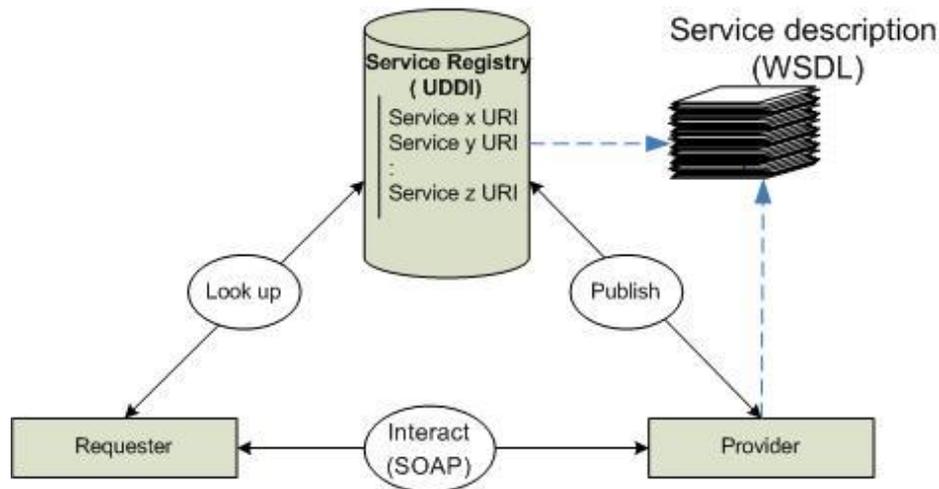


Figure 1: Web Service Architecture

Delivering semantics into Web Services is achieved through annotating Web Service description to a suitable ontology (Sheth et al. 2006) – this is the basis of SWS. Ontology provides a ‘shared conceptualisation’ specifying the semantics of business data, processes and services. Ontology also allows logic-based reasoning by machines – a necessary step in automating the process of service discovery and composition. Currently, domain ontologies are developed manually through the collaboration of highly skilled domain experts and ontology engineers. Ontology building is therefore an expensive and time consuming task that lacks the appropriate automated support tools (Buitelaar & Cimiano 2008). A number of SWS approaches have been proposed including OWL-S (OWL-based Web Service ontology) and SAWSDL (Semantic Annotation for WSDL) and good overviews of the current state-of-the-art on semantic Web Services are provided by (Martin 2007a, Martin 2007b). In all the proposed approaches, ontology development is a challenge that provides a considerable barrier to adopting SWS and, consequently, preventing Web Services from reaching their full potential (Gedda 2008).

3 Ontology Learning

In broad terms, Ontology Learning is grounded in a combination of Ontology Learning Techniques. Most of these techniques are drawn from well-established disciplines such as Machine Learning (ML), Natural Language Processing (NLP) and statistics (Buitelaar & Cimiano 2008, Gomez-Perez & Manzano-Macho 2004). Unsurprisingly, there is often a significant overlap between these three techniques in

practice. For example, statistical techniques are combined with machine learning and classified as such in some literature (Cimiano 2007). Linguistic-based methods are commonly applied with statistical approaches to calculate the relevance of concept to the given domain, these methods include techniques based on linguistic patterns, pattern-based extraction, methods that measures the semantic relativeness between terms within a domain, for other NLP methods refer to (Cimiano 2007, Gomez-Perez & Manzano-Macho 2004, Zhou 2007). In some approaches a combination of all three types are applied. Text-To-Onto (Maedche & Volz 2001) and OntoLearn (Navigli & Velardi 2004), for example, use statistical techniques applied with machine learning algorithms. Other approaches combine linguistic analyses methods and machine learning algorithms like OntoLt (Buitelaar et al. 2004) and ASIUM (Gacitua et al. 2008).

Most comparative surveys compare text-based approaches, however, and there is little work found on comparing learning from unstructured sources to that from structured sources. Web Service sources resemble a specific OL domain in which an OL approach needs to be tailored to cater for the specific nature of these sources. This tailoring involves applying a combination of techniques, covering a pre-processing step to produce syntactically analysed data, followed by the application of an efficient combination of ML and statistical techniques that are applicable in the Web Service domain.

Most work on OL for SWS is related to Web Service matching as exemplified in (Guo et al. 2007). The relationships between WSDL elements are captured and transformed into ontological concepts and relationships, typically using simple pattern detection. Though WSDL documents provide important application level service description they alone are not sufficient for OL as (a) they provide technical description only and (b), in many cases, Web Services use XSD files to provide data type definitions. The need to include other Web Service resources in the OL process is therefore an important one (exemplifying that OL from semi-structured sources should be further addressed).

In that light, the approach introduced by (Sabou et al. 2005) applies NLP to textual description of Web Services and so learns Web Service ontologies from textual

descriptions attached to implementation files (i.e., Javadoc). Noun phrases and service functionality are learnt from verbs by applying a preprocessing pipeline on textual description of Web Services. Linguistic techniques are then applied to extract syntactic patterns and apply dependency parsing. The limitation of this work is that it is confined to Javadoc files, which are not common means of description in Web Services (Guo et al. 2007). The focus on extracting concepts and service functionality from textual description only, ignoring the structural aspect of the Javadoc file, can be improved and extended by considering other Web service sources such as structured sources as in WSDL and XSD documents.

4 Moving forward

Given the aim of automatically learning ontologies from Web Services, the review illustrates two main points. First, there is a need to clarify and address the demands on OL in light of the mix of (semi) structured elements that typically accompany Web Services. Second, there is a need to investigate the appropriate mix(es) of OL techniques in meeting those demands. Both points are illustrated in Figure 2 – highlighting a need to identify techniques for effectively combining a range of Web Service software artefacts with appropriate OL methods.

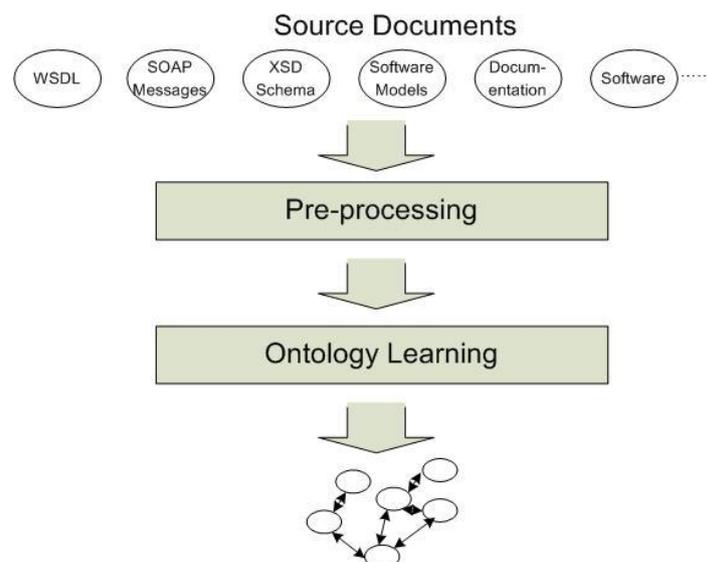


Figure 2: Ontology Learning from Web Service Source Artefacts

The choice of ontology learning strategy, whether it is bottom-up or top down, can be identified based on the data sources and domain (Zhou 2007). Web Service sources are diverse in a number of areas – containing both structured and unstructured data and generating both static and dynamic sources. WSDL and XSD files are examples of static data sources. WSDL files providing a usable source of service interface information, including inputs, output and basic service functionality. SOAP messages, dynamically generated by Web Services and client applications in use, contain instances of server requests issued by clients and instances of service responses issued by service providers. Messages are created when a service is invoked and are an example of a dynamic source. Extending work by Guo et al. in (Guo et al. 2007) to include XSD schema and SOAP messages may offer a number of interesting opportunities – revealing additional concepts and relations through more complex transformation rules. For example, WSDL structures may be transformed into ontological relationships, elements are analysed so that the message : parts relationship is transformed into has property. Applying similar, but more extensive, transformation rules to XSD and SOAP may result in more effective methods. Possible opportunities include using: (1) domain specific rules, (2) advanced source document pre-processing heuristics and (3) source document bootstrapping approaches. WSDL files alone are limited to typically providing a technical description of the underlying service.

Support for variation in Web Service style may also be appropriate. When interpreting document style Web Services, a major part of the service description is found within the referenced XSD schema (Curbera et al. 2002). Interpreting the underlying schema in unison with other Web Service artefacts would result in a considerable increase in the number of identified concepts (when compared to interpreting WSDL in isolation). Moving beyond service description and exploring dynamic SOAP analysis allows executing services to be interpreted and opens further avenues for ontology learning. Service invocation and messaging, via SOAP messages, provides related instance data for each service description. It is this instance data that may provide opportunities for revealing additional relations, axioms and patterns (Daga et al. 2002).

Current OL approaches are in the most part general and need to be specialised to cater for both the technology of the Web Service domain and the business domain in which these services operate. Identifying efficient learning techniques that are applicable in the Web Service domain is a challenging task. Learning techniques from different paradigms need to be combined and tested on varied sources in order to identify effective multidisciplinary techniques for ontology learning.

Drawing the discussion together, a number of research questions exist as follows. For Web Service source documents:

- What are the specific benefits gained from processing differing source document types?
- What ontological elements (e.g.: concepts, relationships) can be extracted, refined or justified by particular Web Service source documents?
- How can source documents (e.g.: WSDL, SOAP) be combined for even greater effectiveness?
- What are the relative merits of differing structural types in the source documents (e.g. from unstructured narrative documentation to semi-structured WSDL and beyond)?
- How can static and dynamic source documents be utilised and combined?

For pre-processing requirements:

- How can linguistic techniques (e.g.: Tokenisation, stemming) typically used on narrative sources, be applied to Web Service source documents?
- What specific syntactic or semantic pre-processing is appropriate for each Web Service source document?

For OL techniques:

- What are the most effective ML techniques (supervised versus unsupervised) for each source document or group of sources?
- What is the most effective OL approach (combining and comparing current approaches) for each source document or group of sources?
- How can techniques from ML and Statistical analysis be combined to benefit from both the structured and unstructured sources?
- How can OL techniques cater for the nuances of dynamic data sources (e.g.: SOAP messages)?

- What steps are required (i.e. tailored framework for Web Service domain) to realise OL in a Web Service environment?

5 Conclusions

This research-in-progress paper examines existing techniques and tools available for semi-automatically learning domain ontology from Web Service resources. The purpose of the work is to develop a research agenda that will direct work on ontology learning toward the pragmatic issue of ‘semanticising’ existing Web Services. Current state-of-the art in Web Services indicates that adoption of semantic Web Services is slow and can be encouraged by employing practical tools on top of current standards. Ontology Learning (OL) offers a viable tool for the ontology development lifecycle. A research agenda is proposed that focuses on how OL can address the variation in Web Service software artefacts, the domain in which they operate and the applicability of specific OL approaches.

References

- Akkiraju, R., Farrell, J., Miller, J. et al. (2005) Web Service Semantics-WSDL-S. W3C Member Submission, 7.
- Azoff, M.(2007) Application Development End-User Survey.
- Berners-Lee, T., Hendler, J., Lassila, O. (2001) The Semantic Web - A New Form of Web Content that is Meaningful to Computers Will Unleash a Revolution of New Possibilities. *Sci. Am.*, 284 34-+.
- Buitelaar, P., & Cimiano, P. (eds.) (2008) *Ontology learning and population: Bridging the gap between text and knowledge*. IOS Press, Amsterdam, The Netherlands
- Buitelaar, P., Olejnik, D., Sintek, M. (2004) A Protege Plug-in for Ontology Extraction from Text Based on Linguistic Analysis. *Proceedings of the 1st European Semantic Web Symposium (ESWS)*.
- Burstein, M., Bussler, C., Zaremba, M. et al. (2005) A Semantic Web Services Architecture. *IEEE Internet Comput.*, 9 72-81.
- Cimiano, P. (2007) *Ontology learning and population from text: Algorithms, evaluation and applications*. Springer.
- Cuel, R., Delteil, A., Louis, V. et al. (2008) *The Technology Roadmap of the Semantic Web*. Knowledge Web, Nov.
- Curbera, F., Duftler, M., Khalaf, R. et al. (2002) Unraveling the Web Services Web: An Introduction to SOAP, WSDL, and UDDI. *Internet Computing, IEEE*, 6 86-93.
- Daga, A., Cesare, S. d., Lycett, M. et al. (2005) An Ontological Approach for Recovering Legacy Business Content. *HICSS '05 Proceedings*, 224.1.
- Gacitua, R., Sawyer, P., Rayson, P. (2008) A Flexible Framework to Experiment with Ontology Learning Techniques. *Knowledge-Based Systems*, 21 192-199.
- Gedda, R.(2008) SOA Uptake Still Split A Mid Market Confusion. 19/11/2008.

- Gomez-Perez, A., & Manzano-Macho, D. (2004) An Overview of Methods and Tools for Ontology Learning from Texts. *Knowl. Eng. Rev.*, 19 187-212.
- Guo, H., Ivan, A., Akkiraju, R. et al. (2007) Learning Ontologies to Improve the Quality of Automatic Web Service Matching. *ICWS 2007*, 118-125.
- Heffner, R., & Peters, A.(2008) Topic Overview:Service-Oriented Architecture for CIOs.
- Maedche, A., & Volz, R.(2001) The Ontology Extraction and Maintenance Framework Text-to-Onto.
- Martin, D. (2007) Semantic Web Services, Part 1. *Intelligent Systems, IEEE*, 22 12-17.
- Martin, D. (2007) Semantic Web Services, Part 2. *Intelligent Systems, IEEE*, 22 8-15.
- Navigli, R., & Velardi, P. (2004) Learning Domain Ontologies from Document Warehouses and Dedicated Web Sites. *Computational Linguistics*, 30 151-179.
- Sabou, M., Wroe, C., Goble, C. et al. (2005) Learning Domain Ontologies for Semantic Web Service Descriptions. *Journal of Web Semantics*, 3 340-365
- Shadbolt, N., Berners-Lee, T., Hall, W. (2006) The Semantic Web Revisited. *IEEE INTELLIGENT SYSTEMS*, 96-101.
- Sheth, A., Verma, K., Gomadam, K. (2006) Semantics to Energize the Full Services Spectrum. *Commun ACM*, 49 55-61.
- Tsai, W. T., Malek, M., Chen, Y. et al. (2006) Perspectives on Service-Oriented Computing and Service-Oriented System Engineering. *Proceedings - Second IEEE International Symposium on Service-Oriented System Engineering, SOSE 2006*, 3-8.
- Yu, Q., Liu, X., Bouguettaya, A. et al. (2008) Deploying and Managing Web Services: Issues, Solutions, and Directions. *The VLDB Journal The International Journal on Very Large Data Bases*, 17 537-572.
- Zhou, L. (2007) Ontology Learning : State of the Art and Open Issues. *Information Technology and Management (Bussum)*, 8 241.