Web Services Roadmap: The Semantic Web Perspective

Peep Küngas
Norwegian University of Science and Technology
Department of Computer and Information Science
Trondheim, Norway
peep@idi.ntnu.no

Mihhail Matskin
Royal Institute of Technology
IMIT/LECS
Kista, Sweden
misha@imit.kth.se

Abstract

Recently the field of Web services has gained focus both in industry and academia. While industry has been mostly interested in standardisation and promotion of the technology, academia has been looking for ways to fit the technology into other frameworks, such as the Semantic Web. Anyway, despite of the increased academic and commercial interest to Web services, there are currently only few case studies available about Web services in the Semantic Web context. Moreover, according to authors' knowledge, there is no publicly available study analysing which data is currently mostly provided/required by Web services.

In this paper we target these shortcomings by providing a case study of semantically annotated commercial and governmental Web services. We analyse interaction and potential synergy between commercial and governmental Web services. Also the role ontologies for semantic integration of Web services is analysed. Moreover, we identify the most common data exploited by current Web services.

1 Introduction

Recently the field of Web services has gained focus both in industry and academia. While industry has been mostly interested in standardisation and promotion of the technology, academia has been looking for ways to fit the technology into other frameworks, such as the Semantic Web. Introduction of the Semantic Web provides enormous opportunities for utilising Web services in knowledge-based systems.

Web services' research in the Semantic Web context varies from automated annotation [4, 5, 3, 12, 10] and ontologies [1] to automated Web service composition [8, 7, 14, 15, 16]. While automated annotation intends to provide methods to extract semantics from existing Web services, research related to ontologies and Web services has focused to the modelling of Web services. The latter has led to such initiatives as WSMO, SWSO, OWL-S and WSDL-S.

Moreover, a lot of research is devoted to the integration of Web services into agent and P2P systems. Naturally many articles consider semantics in P2P and agent systems. Thus there is a lot of activity around Semantic Web services and loosely-coupled systems.

Anyway, despite of the increased academic and commercial interest to Web services, there are currently only few case studies available about Web services in the Semantic Web context. Moreover, according to authors' knowledge, there is no publicly available study analysing which data is currently mostly provided/required by Web services and which Web service domains are available.

In this paper we target these shortcomings by providing a case study of semantically annotated Web services. Both commercial and governmental Web services are considered. We analyse interaction and potential synergy between commercial and governmental Web services. Also the role of ontologies for semantic integration of Web services is analysed. Moreover, we identify the most common data exploited by current Web services.

Our case study is based on analysis of 493 commercial and 96 governmental Web service operations. First, we annotated these operations semantically. Then simple ontologies were built to extend usage of these operations within the Semantic Web and to support automated Web service composition. It turned out that while ontologies enhance the usage of the commercial Web services, they have no significant impact on the governmental Web services. However, ontologies facilitate automation of semantic integration of commercial Web services with governmental ones.

The rest of the paper is organised as follows. In Section

2 we explain which Web services we selected for analysis. Section 3 analyses the constructed Web services roadmap. Section 4 describes the challenges, which occurred during this research. Finally, Section 5 reviews related work while Section 6 presents conclusions and elaborates future work.

Table 1. Annotation overview.

Domain	Operations	Concepts	Ontology size
X-Road	96	595	128
Commercial	493	578	189
Merged	589	1149	317

2 Annotating Web services

The roots of the work presented in this paper are strongly related to our previous and current work on automated Web service composition [11, 6]. After developing a system for automated Web service composition, we wanted to evaluate its performance and applicability in a "real-world" configuration. While performing the evaluation we became interested in the current Web services roadmap. More specifically, we wanted to identify which kind of Web services are available and which are the most common inputs and outputs of current Web services. A special focus was set to analysis of potential interactions between commercial and governmental Web services.

In order to analyse available Web services, we first annotated Web service operations under consideration. By annotating we mean a process of giving logical names to inputs and outputs of Web service operations. These logical names refer to particular concepts in an ontology and represent the semantics of data, which is exploited by Web services. In this paper, when counting the number of relations in developed ontologies, we state explicitly only the number of relations, which represent subclass/superclass relations. Relations, which represent links between Web service operations and data, are not counted.

Although there are some tools available, which allow semi-automatic annotation of Web services, our annotations were constructed manually. The reason is that the current annotation tools are not mature enough for general usage. Additionally we had to develop new ontologies for describing data structures and relations between their fields. This task also needs proper tools, which were not available at the moment we started this thread of research.

Commercial Web services' WSDL documents were retrieved through the list at SalCentral.com in March 2005. From the list of available Web services we annotated most of the available operations, whose semantics was clear. Altogether we annotated 493 commercial Web service operations. Additionally we developed an ontology for commercial Web services, which consists of 189 relations. The overall commercial Web services domain contains 578 concepts.

For governmental Web services we chose the services from X-Road [9] project, which was initiated by Estonian government. X-Road is a middle-tier data exchange layer enabling government databases to communicate with their clients. The system allows officials, as well as legal and natural persons, to search data from national databases over the Internet within the limits of their authority. The system ensures sufficient security for the treatment of inquiries made to databases and responses received.

X-Road project was initiated in 2001 and by March 2005 X-Road had already 41 databases providing services plus 354 institutions and companies using the services. The overall number of available Web service operations was 687. We annotated 96 of them. The domain and the developed ontology consists of 595 concepts and 128 relations. The reason of having a larger ontology for commercial Web services (compared to governmental services) is potentially due to the larger heterogeneity of data in this domain. While governmental Web services are centered around queries about citizens and companies, commercial Web services provide a wider set of Web services.

After merging X-Road and commercial Web service domains, the merged domain consisted of 1149 concepts and 589 operations. 24 concepts were shared between the domains. These concepts represent potential interactions between commercial and governmental services. Table 1 summarises the number of annotated Web service operations, concepts and relations in developed ontologies.

After annotating Web services and building ontologies, we constructed graphs to visualise Web services roadmaps for commercial and X-Road Web services. Due to the lack of space we present here only the roadmap of commercial Web services with ontologies. The roadmap is depicted in Figure 1. Nodes in the graph represent concepts and edges represent potential data flows implemented by Web services. The size of a node shows proportionally its importance in the domain—larger nodes are used by a larger number of Web service operations. Surprisingly the topology of the roadmap has a similar structure as the Web itself [2]. There are tubes, tendrils, disconnected components and strongly connected components.

3 Analysis of the Web services roadmap

In this section we analyse the commercial Web services domain presented in Figure 1 and X-Road Web service operations domain as well as potential interactions between them.

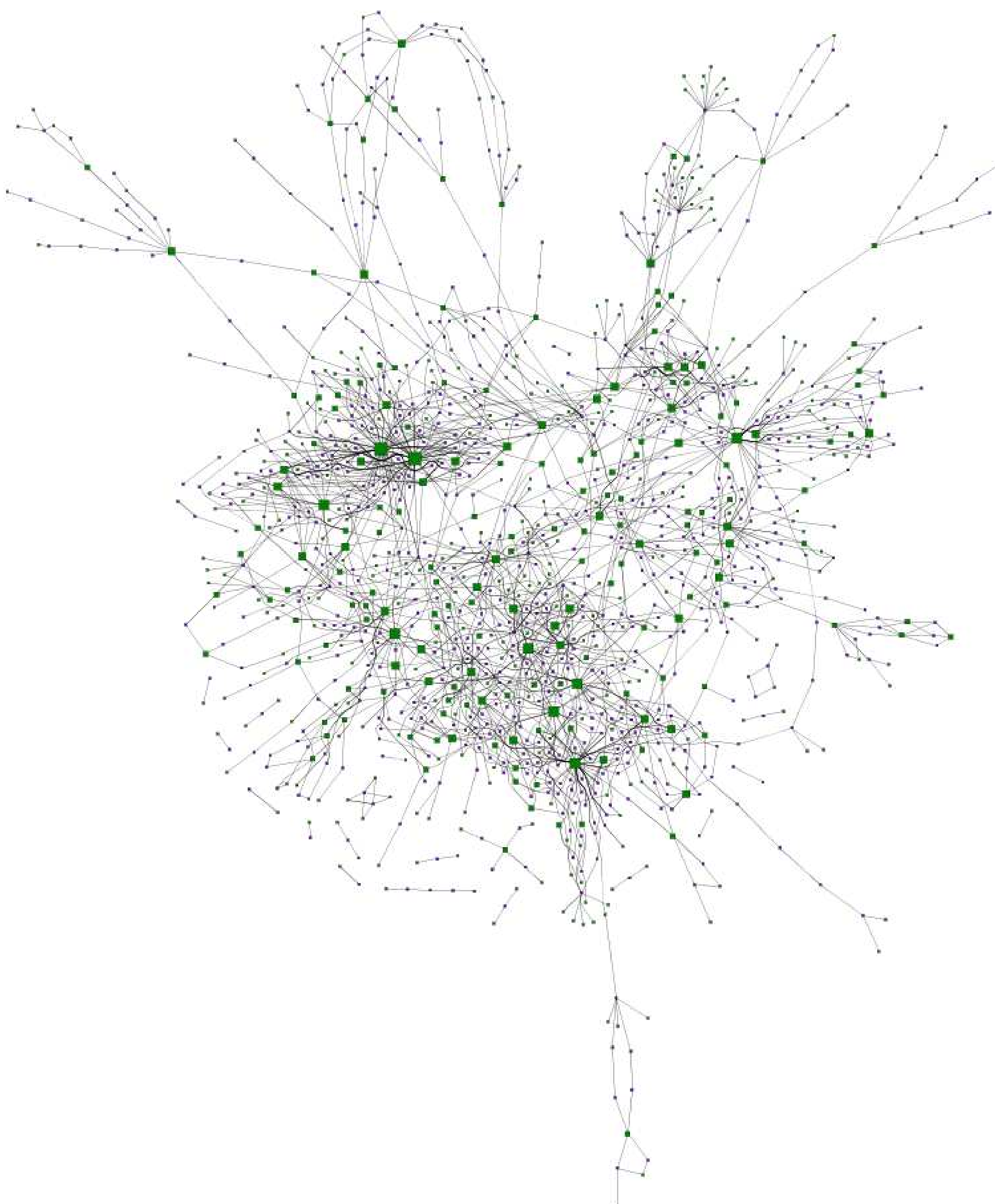

Figure 1. Roadmap of commercial Web services (extended with ontologies).

3.1 The effect of ontologies

The usage of ontologies allows us to reduce the number of isolated Web services in a domain. While ontologies had no effect on X-Road domain itself, they allowed to bind commercial Web service operations to X-Road operations. Moreover, the usage of ontologies allowed to reduce the number of isolated commercial Web service operations from 46 to 22. The ontologies had no effect to X-Road operations because governmental Web services are more homogeneous compared to commercial ones.

3.2 Available Web services

Generally governmental and commercial Web services provide different services. Overlapping areas are related to queries about companies/businesses and contact information for persons and businesses. However, governmental Web services in that area facilitate access to more sensitive information and have therefore limited access.

3.2.1 Available commercial Web services

Most common commercial Web services are currently related to yellow pages services—locating and finding information about businesses, persons and Internet hosts. Also services for measuring distance between physical locations are included in this category. The second category of services deals with communication—instant messaging, sending e-mails, faxes and SMS-es. Third category of services is devoted to financial sector—fetching stock exchange information, market news and performing currency conversions are the key services here.

Naturally, there are services for general Internet search and text translation. Related to this group are services dealing with data conversion, such as mapping documents from one format to another. Some services have been implemented for getting weather forecasts, airport and travelling information. There are also some services for arithmetics, statistics and encryption. Services for direct e-commerce and e-business are still in minority—there are few service for product search, processing credit cards and package tracking. Finally a couple of services have been published for generating graphs, charts and bar codes.

3.2.2 Available governmental Web services

From the overall set of Web services in X-Road we enlist here only Web services, which might be accessible for public usage. They include Web services for checking validity of particular documents, such as driving licenses, passports, diplomas and certificates, which have been assigned by participating institutions. Services for contacting particular persons, companies and other organisations are also important. Finally, there are Web services for checking ownership of real estate properties and vehicles. By making these services public, citizens would become able to check information related to buying/selling real estate and vehicles without intermediaries.

Governmental Web services are characterised by high redundancy. There are many Web services exploiting similar data, both in inputs and outputs. However, a slight variation in inputs and outputs mainly arises from regulations identifying which data should be accessible to specific parties.

3.3 Synergy between commercial and governmental Web services

Given the restrictions associated with accessing governmental Web services, these Web services could be viewed mostly as core services, while commercial Web services could serve as value-added services. In other words, due to the heterogeneity of the commercial Web services domain, these Web services could be applied for customising governmental Web services. Governmental Web services could also be used for accessing particular persons or companies while commercial Web services could provide generic statistical functions for facilitating this process.

Most of the currently available Web services could be used for customising governmental Web services. For example, instant messaging could be exploited as an alternative medium for accessing citizens and officials. Internet search and translation can be applied for fetching more information about companies and persons and facilitating thereby more adequate decision making.

There is still a lot of space for additional Web services. X-Road, for instance, could provide Web services for fetching statistics, which is essential for businesses and decision-making. There is a strong need for commercial Web services, which would take generic person or company data, such as name and birth/foundation date, and return some useful information about them.

There could also exist governmental Web services for delivering information, such as warnings to the citizens travelling in a particular region. This information can be also delivered to travellers by travel agents themselves. Thus there are a lot potential for new Web services combining governmental and commercial Web services.

3.4 Most common semantic data

In this section we analyse, which kind of data is most commonly exploited in the currently available Web services. We consider X-Road and commercial Web service domains and the domain resulting from merging these two domains.

3.4.1 X-Road domain

From 595 concepts in X-Road domain, 25 concepts were involved in 10 or more operations. The most important concepts were national identification code, first name and last name. Other concepts include different dates (birth date, event occurrance date, etc.), business registry code, address, e-mail address, street name, postal code, country name and code, company name and various messages.

3.4.2 Commercial Web services domain

From 578 concepts in commercial services domain, 40 concepts were involved in 10 or more operations. The most important ones were user name and password followed by ZIP code, date, license key, city name, postal code and e-mail address. User name, password and license key are currently most important mechanisms for controlling access to commercial Web services. While governmental Web services handle authentication at system level, commercial Web services limit access by expecting users to provide user names, passwords, license keys, access/account codes, etc.

Additional important concepts were country name, currency, stock symbol, message content (for instance messaging, SMS and e-mails), search string, name, IP address, URL and words to be translated. Other concepts include weather, (US) state name, interest rate type, location, country codes, date, distance, postal address and delivery tracking numbers.

3.4.3 Merged domains

From 1149 concepts in the merged services domain, 64 concepts were involved in 10 or more operations. Most important of them were user name, password, national identification code, first name, last name, ZIP code, date, e-mail address, postal code, license key, city name, country name and message content.

24 of 1149 concepts were overlapping between X-Road and commercial services with ontologies. These concepts represent the data, which can be currently directly passed between commercial and governmental Web services. Furthermore, these concepts represent the current interactions between these two domains.

The overlapping concepts are birth date, city name, company name, country code/name, county, currency, date, day, e-mail address, fax number, first name, last name, gender, IP address, message content, mobile phone number, month, name, phone number, postal code, street name, year and a general string. In general, these concepts are mostly related to everyday communication.

4 Challenges

While annotating Web services and building ontologies we encountered a number of challenges, which either limited our efforts or made in some cases annotation even impossible. A very important factor is the usage of a wide variety of languages in WSDL files. Although most of the WSDL files were documented in English and the same language was used for naming inputs and outputs, many services contain data in other languages as well. This makes the extraction of semantics from WSDL files difficult.

Moreover, there is often too little or even misleading information available about Web services and data fields. For instance data field name *country* could be a country name in a particular language or any available country code. There is a general bias not to document data fields in commercial Web services. While X-Road services had mostly data fields commented, only one percent of all available commercial Web services had comments for data fields. Anyway, the situation for service and operation documentation in commercial Web services was much better. In particular, 127 of 404 available Web services and 2443 of 3764 operations were documented.

Data with the same meaning is encapsulated in different data types. For instance, an address may be represented with a string or a data type containing fields for each element of an address. Furthermore, sometimes an address contains a country name while in other cases it represents only a street name and house/appartment number.

In summary, the main challenges are as follows:

- different languages
- lack of documentation in WSDL documents
- different data structures with the same meaning
- dynamically changing WSDL documents
- availability of WSDL documents

Therefore, in order to facilitate annotation and further usage of annotated Web services, developers should document WSDL documents in a more detailed manner. Additionally, it would be desirable to have annotations in multiple languages within the same WSDL document, or at least a tag indicating which language is used for documentation and naming data fields. There could also exist a sort of versioning system for WSDL documents such that existing annotations could be modified on-the-fly.

5 Related work

Sabou et al [13] present a case study of using DAML-S ontology for annotating semantically Web services. They identify difficulties of writing DAML-S services. Since programmers are assumed to have knowledge about WSDL, SOAP and DAML, the language is too comprehensive and knowledge-demanding for human users.

Automated Web service annotation has been considered by Heß et al [4, 5], Burstein [3], Sabou [12] and Patil et al [10]. Semantically annotated Web services have been exploited by several researchers in automated Web service composition [8, 7, 14, 15, 16]. Research on Semantic Web services representation has led to initiatives like WSMO, SWSO, OWL-S and WSDL-S.

6 Conclusions and future work

In this paper we extracted and analysed most important semantic data objects in the currently available Web services. We also provided an overview of the currently available Web services. Both commercial and governmental Web services were analysed. Representative set of governmental Web services was selected from X-Road [9] project.

We also analysed potential interactions between governmental and commercial Web services. Based on the analysis we proposed some new Web services, which could be beneficial for businesses and governmental institutions.

Although the analysed Web services operations and concepts are quite representative, we reviewed only a limited number of Web services compared to what is currently available on the Web. More thorough analysis needs advanced automated methods for Web services' annotation and analysis. Therefore our future research focus is biased towards providing such automated methods and tools.

Acknowledgements

This work was partially supported by the Norwegian Research Foundation in the framework of Information and Communication Technology (IKT-2010) program—the ADIS project. The authors would additionally like to thank anonymous referees for their comments.

References

[1] I. B. Arpinar, B. Aleman-Meza, R. Zhang, and A. Maduko. Ontology-driven Web services composition platform. In *Proceedings of IEEE International Conference on E-Commerce Technology, CEC'04, San Diego, California, USA, July 6–9, 2004*, pages 146–152. IEEE Press, 2004.

[2] A. Broder, R. Kumar, F. Maghoul, P. Raghavan, S. Rajagopalan, and J. W. R. Stata, A. Tomkins. Graph structure in the Web. *Computer Networks*, 3:309–320, 2000.

[3] M. Burstein. Ontology mapping for dynamic service invocation on the Semantic Web. In *AAAI Spring Symposium on Semantic Web Services, Palo Alto, March, 2004*, 2004.

[4] A. Heß, E. Johnston, and N. Kushmerick. Assam: A tool for semi-automatically annotating semantic web services. In *Proceedings of the 3rd International Semantic Web Conference (ISWC 2004)*, Hiroshima, Japan, 2004.

[5] A. Heß and N. Kushmerick. Learning to attach semantic metadata to web services. In D. Fensel, K. Sycara, and J. Mylopoulos, editors, *Proceedings of the 2nd International Semantic Web Conference*, number 2870 in Lecture Notes in Computer Science, pages 258–273, Sanibel Island, Florida, USA, 2003. Springer-Verlag.

[6] P. Küngas, J. Rao, and M. Matskin. Symbolic agent negotiation for Semantic Web service exploitation. In *Proceedings of the Fifth International Conference on Web-Age Information Management, WAIM'2004, Dalian, China, July 15-17, 2004*, volume 3129 of *Lecture Notes in Computer Science*, pages 458–467. Springer-Verlag, 2004.

[7] D. McDermott. Estimated-regression planning for interaction with Web services. In *Proceedings of the 6th International Conference on AI Planning and Scheduling, Toulouse, France, April 23–27, 2002*. AAAI Press, 2002.

[8] S. McIlraith and T. C. Son. Adapting Golog for composition of Semantic Web services. In *Proceedings of the Eighth International Conference on Knowledge Representation and Reasoning (KR2002), Toulouse, France, April 22–25, 2002*, pages 482–493. Morgan Kaufmann, 2002.

[9] I. Odrats, editor. *Information Technology in Public Administration of Estonia, yearbook 2004*. OÜ Piltkiri, 2005.

[10] A. Patil, S. Oundhakar, A. Sheth, and K. Verma. METEOR-S Web service annotation framework. In *Proceedings of the 13th International Conference on World Wide Web (WWW '04), New York, NY, USA, May 17–22, 2004*, pages 553–562. ACM Press, 2004.

[11] J. Rao, P. Küngas, and M. Matskin. Logic-based Web services composition: From service description to process model. In *Proceedings of the Second International Conference on Web Services (ICWS 2004), San Diego, California, USA, July 6–9, 2004*, pages 446–453, 2004.

[12] M. Sabou. From software APIs to Web service ontologies: a semi-automatic extraction method. In *Proceedings of the Third International Semantic Web Conference (ISWC2004), Hiroshima, Japan, November 7–11, 2004*, 2004.

[13] M. Sabou, D. Richards, and S. van Splunter. An experience report using DAML-S. In *Proceedings of the Twelfth International World Wide Web Conference Workshop on E-Services and the Semantic Web*, 2003.

[14] M. Sheshagiri, M. desJardins, and T. Finin. A planner for composing services described in DAML-S. In *Proceedings of the AAMAS Workshop on Web Services and Agent-based Engineering*, 2003.

[15] P. Traverso and M. Pistore. Automated composition of semantic web services into executable processes. In *Proceedings of 3rd International Semantic Web Conference, ISWC 2004, Hiroshima, Japan, November 7–11, 2004*, volume 3298 of *Lecture Notes in Computer Science*, pages 380–394. Springer-Verlag, 2004.

[16] D. Wu, B. Parsia, E. Sirin, J. Hendler, and D. Nau. Automating DAML-S Web services composition using SHOP2. In *Proceedings of the 2nd International Semantic Web Conference, ISWC 2003, Sanibel Island, Florida, USA, October 20–23, 2003*, 2003.