Acquisition and Management of Semantic Web Service Descriptions

Maria Maleshkova

Institute of Applied Informatics and Formal Description Methods (AIFB),
University of Karlsruhe (TH), Germany
maleshkova@aifb.uni-karlsruhe.de

Abstract. The increasing importance and use of Web services have resulted in a number of efforts targeted at automating Web service discovery and composition based on semantic descriptions of their properties. However, the progress in the automation of Web service discovery is still held back by the fact that the description of Web services in terms of semantic metadata is still mainly manually. This Ph.D. thesis addresses this problem by developing an approach for the acquisition and management of semantic Web service descriptions in order to facilitate efficient service discovery and composition. Specifically, this involves the collection of information about a Web service, the acquisition of semantic descriptions based on the collected information, and the structured storage of the generated semantic descriptions.

1 Introduction

A clearly developing trend in the recent years is towards exposing the functionalities provided by existing software components in the form of services and facilitating in this way software reuse and added value through service composition. This development is strongly supported by the Web, which enables ubiquitous access via a set of Web standards and protocols to software components residing on different platforms. As a result, Web services are seen increasingly as the basic construct for the development of rapid, low-cost and easy-to-compose distributed applications in heterogeneous environments.

When it comes to finding appropriate services and composing distributed applications, current technologies still require a large amount of human interaction. This results in the restricted number of use cases for service integration and the limited scalability of solutions involving manual activities in the process of service discovery and compositions. In order to address these problems, the idea of supplementing Web services with a semantic description of their functionality, that could facilitate their discovery and integration, has been developed. However, the information, on which these semantic descriptions are based, is rarely discussed. In addition, there is still no established method for acquiring semantic Web service descriptions, which not only describe the functionality but also specify functional and non-functional properties in a way that supports automatic

matchmaking processes. Therefore, there is the strong need of developing an automatic approach, which enables the acquisition and management of semantic Web Service descriptions with the goal of supporting efficient service discovery and composition.

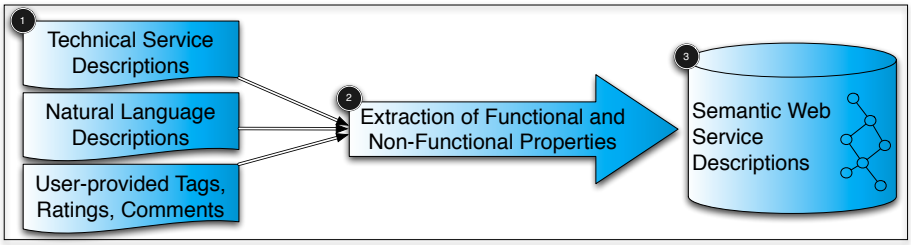

Fig. 1. Acquisition and Management of Semantic Web Service Descriptions

This Ph.D. thesis targets to explore the problem of semi-automatically acquiring and publishing semantic descriptions for Web services and aims to use Semantic Web techniques to develop a methodology to enhance the situation. Figure 1 shows the three main challenges connected with the acquisition and management of semantic Web Service descriptions.

1. **How can information about Web services be collected?** To answer this question, the relevant sources of information have to be identified, including not only technical and text descriptions, but also information inspired by Web 2.0 such as user tags, ratings and comments. After this, automatic mechanisms for data collection from the Web have to be developed.
2. **How can semantic Web service descriptions be automatically acquired?** This question is in the focus of the Ph.D. thesis and is addressed by exploring existing methods for metadata acquisition and identifying their advantages and disadvantages. The main goal is to develop a method, which semi-automatically acquires metadata by extracting functional and non-functional service properties in order to facilitate discovery and composition.
3. **How can semantic Web service descriptions be stored?** The acquired semantic descriptions have to be stored in a way that facilitates the manual as well as the automatic matchmaking processes. It is important to employ the metadata's syntactical or semantical interconnections in order to enable indexing and efficient service search.

2 Related Work

There is already some work done related to the automatic creation of semantic Web service descriptions. In particular, there are two main areas of research: the acquisition of a suitable Web service domain ontology and the actual process of

annotating Web services. Sabou et al. [1] present two ontology building process in the context of two concrete research projects, revealing some of the major aspects that make Web service ontology building difficult.

Focusing on the Web service annotation task, Patil et al. [2] apply graph similarity techniques to select a relevant domain ontology for a given WSDL file from a collection of ontologies. Hess and Kushmerick [3] employ Naive Bayes and SVM machine learning methods to classify WSDL files in manually defined task hierarchies. However, none of the developed approaches focuses on facilitating Web service discovery by specifying functional and non-functionally properties and in the same time taking into consideration temporal conditions on effects, trust and access control policies.

There are a number of state of the art Web service repositories including UDDI, Bindingpoint, .NET XML Web Services Repertory, WebserviceX.NET, Web Service List, Xmethods and SalCental. An overview of these repositories presented in [4] shows that in contrast to traditional software libraries, Web service repositories rely on little metadata to support service discovery, mainly because of the difficulty of automatically deriving metadata describing Web service collections. In order to overcome the lack of metadata, there are a number of approaches which aim to enhance existing Web service repositories, in particular UDDI, with complex semantic markup [5]. Still, the existing repositories do not target the structured storage of semantic Web service descriptions in order to facilitate the effective service discovery but rather use semantic information as an extension to stored descriptions.

3 Contributions

A number of Web service tasks can be automated by using semantic descriptions. In particular, service offers and requests can be matched automatically. However, semantic descriptions still have to be manually generated. The major goal of this Ph.D. thesis is to develop an semi-automatic method for the acquisition of semantic descriptions of Web services, based on metadata and tags extracted from the Web. The resulting semantic descriptions will be stored in an online semantic Web service descriptions repository, which can be used for both manual and automatic service discovery. The acquisition of semantic descriptions facilitates matchmaking approaches and can reduce the required human interaction to a minimum. Especially in use cases, where Web services are used as building blocks in distributed applications or where Web services provide functionality integrated in business processes, the automated deriving of semantic description plays a key role.

There are three main contributions, which are to be archived. First, the automatic collection of information describing a Web service is facilitated. Second, semantic descriptions of Web services are semi-automatically acquired from the collected information, in the form of semantic annotations. Finally, Web service semantic annotations are stored in an online repository, which provides semantic

indexing to allow for efficient queries and support matchmaking processes. Each of these contributions is described in detail in the following sections.

Web Service Crawlers are developed to support the automatic collection of information about Web services. Web service descriptions are usually given in a standardized form such as WSDL. Unfortunately, WSDL provides mainly technical description of services and this information is insufficient for automatic service discovery because it presumes increased human interaction. Therefore, it is necessary to automatically collect additional information. A Web service crawler can effectively be used to perform this task. In particular, this involves the identification of relevant sources and types of Web service information. Sources include natural language text descriptions, for example, documentation on provider Web-sites, and technical Web service descriptions such as WSDL, or WSMO variants. In addition, sources of information inspired by Web 2.0 such as user tags, ratings, and community inputs are also considered. The definition of the Web service crawlers is based on semantic attributes for service description, based on information needed for service discovery and composition. This involves mainly the specification of functional and non-functional Web service properties.

Web Service Semantic Annotations comprise the acquired semantic Web services descriptions, which are based on the data collected by the Web service crawlers. The main focus of this Ph.D. thesis is on developing a mechanism for the acquisition of semantic descriptions from the collected data, based on a determined formalism for the specification of semantic information. This mechanism is developed by using existing techniques for data mining and ontology learning and thereby, automatically constructing service annotations. In addition, the formalism, chosen as the basis for the semantic annotation, annotates functional and non-functional properties of Web services in a unifying way. The approach described in [6] will be used as a basis where semantic, temporal and security constraints are described by using a combination of the π-calculus and description logics and will be adjusted based on requirements resulting from the specification of the developed mechanism for acquisition of semantic descriptions.

Semantic Indexing is necessary in order to facilitate efficient queries on top of the semantic annotations. The structured storage of the acquired semantic annotations is targeted to improve performance of automatic matchmaking processes and speed up manual search. During semantic indexing, the collection of Web service semantic annotations will be analyzed and organized in a way that allows the efficient answering of expressive queries.

4 Evaluation

The planned evaluation consists of three main steps. First, the developed mechanism for data collection is evaluated by determining a fixed domain of source data and assessing how much data was correctly retrieved. This evaluation can easily be performed by using the well-established recall/precision metrics. Second, semantic annotations evaluation assesses the performance of extracting Web

service properties relevant for matchmaking from the corpus of collected data. Naturally, the quality of semantic annotation extraction has a direct influence on the quality of the service discovery and composition, which are based on these semantic annotations. This evaluation can be done by comparing the Web service functional and non-functional properties present in the collected data to the ones present in the acquired semantic descriptions. Third, the chosen structure for storage of Web service semantic annotations is evaluated by comparing the performance of matchmaking algorithms ran on the repository to the performance of these matchmaking algorithms ran on the same semantic annotations but in unstructured list collection. This evaluation will point out what improvements the developed repository structure brings in means of search performance. Finally, the best evaluation for the developed approach would be the making public of the semantic annotations repository so that other researchers can use it to test their discovery and composition algorithms.

5 Work Plan

The work plan is divided into three stages, each stage marking one year of the Ph.D. studies. **Initialization** State of the art report on formalisms for Web service semantic descriptions, in the context of service discovery (M6). State of the art report on data-mining techniques and ontology learning (M6). State of the art report on possible structures for semantic annotations storage (M6). Identified sources of Web service information (M12). Identified formalism for semantic annotations (M12). **Development** Definition of semantic attributes for the Web service crawlers. First Web service crawlers prototype (M18). Identified structure for annotations storage (M18). Identified data-mining and ontology learning approach (M24). First prototype implementation of acquisition of semantic annotations (M24). Evaluation of the first prototypes (M24). **Refinement** First prototype of semantic annotations repository (M30). Refined methodology based on the first prototypes (M30). Refined implementation (M36). Evaluation (M36).

References

1. Sabou, M., Wroe, C., Goble, C., Stuckenschmidt, H.: Learning Domain Ontologies for Semantic Web Service Descriptions. Journal of Web Semantics 3(4) (2005)
2. Patil, P., Oundhakar, S., Sheth, A., Verma, K.: METEOR-S Web service Annotation Framework. In Proceedings of the 13th World Wide Web Conference (2004)
3. Hess, A., N. Kushmerick, N.: Machine Learning for Annotating Semantic Web Services. In AAAI Spring Symposium on Semantic Web Services (March 2004)
4. Sabou M., Pan, J.: Toward Improving Web Service Repositories through Semantic Web Techniques. Workshop on Semantic Web Enabled Software Engineering (SWESE) at ISWC (November 2005)
5. Akkiraju, R., Goodwin, R., Doshi, P., Roeder, S.: A Method for Semantically Enhancing the Service Discovery Capabilities of UDDI. IIWeb 87-92 (2003)
6. Agarwal, S., Studer, R.: Automatic Matchmaking of Web Services. International Conference on Web Services (ICWS'06). IEEE Computer Society (September 2006)