

# Service-Finder: Web Scale Semantic Discovery

Kalina Bontcheva<sup>1</sup>, Emanuele Della Valle<sup>2</sup>, Michael Erdmann<sup>3</sup>,  
Holger Lausen<sup>4</sup>, Nathalie Steinmetz<sup>4</sup>, Andrea Turati<sup>2</sup> \*

<sup>1</sup> University of Sheffield, 211 Portobello Street, Sheffield S1 4DP, England  
K.Bontcheva@dcs.shef.ac.uk

<sup>2</sup> CEFRIEL, Via Fucini 2, 20133 Milan, Italy  
*firstname.lastname@cefriel.it*

<sup>3</sup> ontoprise GmbH, An der Raumfabrik 29, 76227 Karlsruhe, Germany  
erdmann@ontoprise.de

<sup>4</sup> seekda OG, Museumstraße 21, A-6020 Innsbruck, Austria  
*firstname.lastname@seekda.com*

**Abstract.** The Web is moving from a collection of static documents to a set of Web Services. Search engines provide fast and easy access to existing Web pages, however up till now there is no comprehensive solution providing a similar easy and scalable access to Web Services. The European research project Service-Finder aims at developing a platform for service discovery where service related information from heterogeneous sources is automatically integrated in a coherent semantic model to allow effective discovery.

## 1 Introduction

The aim of Service-Finder is to provide an efficient access to publicly available services. (Web) Service technology can be used to expose functionalities in many domains. However creating a Web Service is only the very beginning of the complete service life cycle. Potential users must be made aware of its existence and features.

Initially UDDI [2] was proposed as solution to publish and search services. Although the standard was supported by major software vendors (IBM, Microsoft, SAP) it has not prevailed in the domain of publicly available Web Services. Today the process of publishing is mainly done by providing access to the interface description and providing a couple of Web pages explaining the particularities of a service. This information is then found by the crawlers of search engines such as Google or Yahoo. In addition to using standard search engines, a couple of portals specialized on providing access to public services are available. However most of them cover only a small portion of the actually available services and suffer from the problem of missing or outdated information [1].

The aim of the Service-Finder project is to design and implement an architecture that significantly improves the current situation by:

- Employing automated methods to gather Web Services and related resources.
- Leveraging semi-automatic means to create semantic service descriptions from information available on the Web.
- Facilitating community feedback to build up and improve semantic annotations automatically.
- Describing the aggregated information in semantic models in order to allow reasoning over them.

---

\* Author names are in alphabetical order

## 2 Semantic Discovery in Service-Finder

Within the Service-Finder we will use the benefits of employing a semantic model for discovery, not focusing only on reasoning aspects, but mainly on the creation and maintenance of accurate description at Web scale. As first step Web Services and their related information are identified by a crawler. Using natural language text processing and information retrieval techniques a basic semantic meta model for each service is created. The information is presented in a Web 2.0 style portal allowing the community to improve and correct the annotations. The explicit and implicit user feedback gathered will be used to automatically improve the automated service annotator.

The *Service Crawler* starts with the premise that every service is described with technical interface specification (WSDL). Using specific metrics we ensure that only meaningful services are included in the crawling process (specifically tutorials), while ill formatted and non-reachable services are excluded. Moreover, service descriptions will exist in different versions that will have to be correlated. Not all services available are described using the WSDL technology, especially in the area of mashups and consumer-oriented services where providers often utilize the REST paradigm. REST services do not have a standardized interface description, which implies that we need to develop methods to identify such services.

The *Automatic Service Annotator* will use several ontologies to enhance the data obtained by the crawler. First it will associate every service with a concept of a service-category ontology. In addition it will use elements from a generic Web Service ontology that captures properties such as "hasProvider", "knownSince" and similar. Moreover for selected vertical domains we will use domain-specific ontologies. Our approach is to allow automatic creation of service descriptions, as well as to support users in creating and correcting descriptions in a community driven process. One of the innovations of Service-Finder is that we can benefit from user feedback to verify the automatically produced results, solve conflicts and contradictions this way and thus obtain higher quality metadata in a semi-automatical but scalable fashion.

Within the Service-Finder interface we are able to track user choices and to gather large scale data on users' behavior and services' usage. The *Cluster Engine* analyzes this data from the Service-Finder portal, in addition to services' details which it gets from the *Conceptual Indexer and Matchmaker* component and builds both user and service clusters. The Conceptual Indexer and Matchmaker in turn needs information about clusters in order to customize the searches executed through the Service-Finder portal.

## 3 Conclusion

Using these methods, the Service-Finder project will apply semantic Web Service discovery at Web scale. Previous approaches suffer of the lack of existing semantic annotations (respectively the high costs of obtaining them). Through crawling and leveraging community feedback this bottleneck can be overcome for the case of publicly available services.

## References

1. Daniel Bachlechner, Katharina Siorpaes, Holger Lausen, and Dieter Fensel. Web service discovery a reality check. In *3rd European Semantic Web Conference*, 2006.
2. T. Bellwood, L. Clément, D. Ehnebuske, A. Hately, Maryann Hondo, Y.L. Husband, K. Januszewski, S. Lee, B. McKee, J. Munter, and C. von Riegen. Uddi version 3.0. <http://uddi.org/pubs/uddi-v3.00-published-20020719.htm>, July 2002.