

Service-Finder Annual Report



<http://www.service-finder.eu>

The Web is moving from a collection of static documents to one of Web Services. Search engines provide fast and easy access to existing Web pages, however up till now no comprehensive solution exists which provides a similar easy and scalable access to Web Services. The Service-Finder project has developed a platform for service discovery where service related information from heterogeneous sources is automatically integrated in a coherent semantic model to allow effective discovery.

To this end, the Service-Finder project created a public Web portal capable of:

- Employing automated methods to gather WSDLs and all related resources such as wikis, blogs or any webpage in which useful information are given;
- Leveraging semi-automatic means to create semantic service descriptions of a Web Service out of the information gathered on the Web;
- Index the semantic aggregated information to enable fast searches and allow reasoning and matchmaking over them;
- Providing a Web 2.0 portal to support users in searching and browsing for Web Services, and facilitating community feedbacks to improve semantic annotations;
- Giving recommendations to users by tracking their behaviour in using the Web 2.0 portal.

Summary of Activities

In the second and last project year, partners designed and developed the Beta and final version of the Service-Finder portal. They started from the results of last year when the Alpha prototype went public and evaluated the quantitative and qualitative aspects of each component and the whole portal. This procedure provided feedbacks to refine requirements and the design of the whole portal. Moreover, during 2009 partners updated the Service-Finder Portal by enhancing the current features (both for quality, quantity and performance) and by adding new advanced functionalities (such as: detecting additional related information, usage of user feedbacks to improve automatic annotation and recommendations, programmatic APIs).

The Beta version of the Service-Finder portal is live since October 2009 at <http://demo.service-finder.eu>. Web users are able to test its functionalities and provide their contributions following a collaborative approach typical of Web 2.0 sites. Even if the project is approaching its end, partners will continue to keep the portal up and running until 2012.

Service Crawling

At the heart of the Service-Finder Web Service search engine is the Service-Crawler. Continuing our results from the first project year, we continuously improved our approach for focused crawling and for the detection of WSDL service descriptions and related resources. We also continuously monitored, evaluated and improved our approach in terms of throughput and focused crawl performance. During the second year of the project we improved our WSDL processing in terms of WSDL deduplication and the related document

processing by adding a term vector similarity tool and a document deduplication tool. We added a new feature to the Service Crawler that enables us to crawl UDDI repositories. Moreover we investigated the crawling for REST services: we developed an approach based on manual feature detection and provide an initial evaluation of the approach.

Automatic Annotation

The Automatic Annotator processes the data from the Service Crawler and produces semantic analyses for the Conceptual Indexer and Matcher. We adapted natural language processing and information extraction techniques (using the GATE platform and bespoke software for Service-Finder) to obtain relevant information from the textual documents (HTML and PDF) relating to the web services and to compare them with the relevant pre-processed WSDL files. The Automatic Annotator processes each provider's set of documents independently, translates the extracted information into RDF/XML or F-logic, and forwards it to the Conceptual Indexer and Matcher, along with compressed archives of the plain text content of interesting documents for keyword indexing. We evaluated the results of the alpha release against a sample of services manually categorized through the portal and used the manual annotations as training data for a machine-learning categorization tool, which gave much better results and which we have integrated into the beta release. Users' annotations (through the portal's Web 2.0 features) can be incorporated into future training and evaluation data.

Conceptual Indexing and Matchmaker

One key principle to realize the goal of effective Web Service discovery is the actual matchmaking process, i.e. retrieving good services for a given user request, related to crawled services, automatic annotations and also feedback collected from actual users. The Conceptual Indexing and Matchmaking component (CIM) takes care for storing all relevant information, which consists of formalized knowledge as well as textual information gathered from web documents related to a web service.

Our evaluations have shown that the combined approach of searching for web services based on formal *and* textual input provides better precision of the search results than textual queries alone. We have developed a ranking algorithm which ranks matching services according to their availability, their response time, past feedback from users, availability of documentation and other factors. This algorithm ensures that useful, matching services will appear at the top of the list.

The CIM is based on the OntoBroker reasoner that internally uses F-Logic to represent the formal knowledge. F-logic allows us to enrich the basic data with ontologically derived facts and, thus, improve the knowledge retrieval and matchmaking steps. With the other Service-Finder components CIM communicates via RDF, i.e. it receives RDF triples from the Automatic Annotator and provides RDF-Graphs as query results for searches triggered by the Service-Finder Portal.

Service-Finder Interface

The main point of access for the final users to the results of the project is the Service-Finder Interface, a Web portal through which it is possible to search and browse for the services crawled, annotated and indexed by the other components. The beta release of the Service-Finder Interface, accessible on the Web at <http://demo.service-finder.eu>, provides all the functionalities to: find services; navigate through them and their providers to learn about their features and characteristics; contribute in a Web 2.0 fashion to the services' description by adding tags, categorizations, ratings, comments and a wiki-like free-text description; finally

invoke the services to test their operations. A machine-to-machine access to Service-Finder data is also possible by means of RDFa annotations embedded in the portal pages.

The design and implementation of the Service-Finder Interface took into account the revised requirements of the users, the trends in Web 2.0 and interaction design fields and the research results in applying Semantic Web technologies to the development of portals and Web interfaces via the employment of the STAR:chart framework (see also <http://swa.cefriel.it/STAR>). The logs of all users' interactions with the Service-Finder Interface, which reflect their activities to search, browse and annotate services, are the basis for the work of the Cluster Engine as explained in the following section.

Finally a thorough evaluation of the Service-Finder portal, done with the help of the Service-Finder users, revealed both the high level of usability of the interface and the full appreciation of Service-Finder features by its community of users.

Cluster Engine

Cluster Engine is responsible for providing recommendations to users meanwhile they are interacting with the Service-Finder portal.

The literature presents two different approaches for making recommendations: content-based filtering and collaborative filtering. The first one aims at recommending items that are most similar to the ones that the user already appreciated, while collaborative filtering tries to derive similarities between users in order to recommend items that similar users tend to prefer. We chose to base the Cluster Engine over the collaborative filtering approach since it better fits the collaborative typical of Web 2.0 applications and it also foster the serendipity principle, which can help users in finding services.

The Cluster Engine monitors users' behaviour in interacting with the portal to derive users' profiles and, then, compares all users' profiles in order to estimate similarities between users. Exploiting such similarities, it provides users with personalized recommendations based on their implicit preferences in interacting with services.

User Involvement, Promotion and Awareness

The Service-Finder project actively co-operates with other European projects and players in the project area. Two of the consortium partners, seekda and the University of Sheffield are members of STI International (<http://www.sti2.org/>), a collaborative association of interested scientific, industrial and governmental parties that share a common vision in the Semantic Technologies area. Service-Finder co-operates with the NeOn project (<http://www.neon-project.org/web-content/>) in that in the project the NeOn Toolkit is applied for building the Service-Finder ontologies. The Service-Finder and TAO (<http://www.tao-project.eu/>) projects met at ISWC 2008 to share results and demonstrate software. TAO representatives presented an overview of that project as well as some of their software, and Service-Finder presented its intermediate results.

Two of the consortium partners, namely seekda and CEFRIEL are also part of the consortium of the SOA4All project (<http://www.soa4all.eu/>). Service-Finder shares its categorization data with the SOA4All project. Moreover the Service Crawler work within Service-Finder is followed up by the Service location work in SOA4All.

The project targeted and achieved some of the most important dissemination opportunities in the area of Web Services and Semantic Web. This list includes the *International Conference on Service Oriented Computing (ICSOC) 2009* (<http://www.icsoc.org/>), the *European Semantic Technology Conference (ESTC) 2009* (<http://www.estc2009.com/>), the EU Matchmaking Event co-located with the *European Semantic Web Conference (ESWC) 2009*

(<http://www.sti2.org/events/details/39-the-1st-eu-matchmaking-event-creating-opportunities-for-the-future>), the *International Conference on Business Information Systems (BIS) 2009* (http://bis.kie.ae.poznan.pl/13th_bis/), a *Workshop on Web Personalization, Reputation and Recommender Systems at the International Conference on Web Intelligence (WI) and Intelligent Agent Technology (IAT) 2009* (<http://www.wprrs.scitech.qut.edu.au/>) and the *Future Internet Symposium (FIS) 2009*: <http://www.fis2009.org/>). Thanks to these events the innovation aspects of Service-Finder obtained a good visibility.

Further to these international events, the project was present at other national events such as the *Symposium on Advanced Database Systems 2009* (<http://www.sebd.org/>).

Future Research Directions and Exploitation Prospects

The Alpha version of the Service-Finder Portal is live since December 2008 and the Beta version is live since October 2009. Even if the project will terminate in 2009, partners aim to keep the portal up and running until 2012. Future works will include other research and development mainly performed outside the scope of the project:

- for the Service Crawler, we aim to further evaluate our focused crawling approaches, especially the new features integrated during the second year of the project.
- for the Automatic annotator, we will make further refinements in preparation for the final release, and use the technology developed in the project to enhance GATE and in other projects, (such as NeOn);
- for the Conceptual Indexer and Matchmaker, we aim to “feedback” research results, like the RDF interfaces we developed in Service-Finder, into the standard OntoBroker;
- for the Service-Finder Interface, we aim to proceed our research efforts in semantic navigation frameworks and web 2.0 frameworks. We are also investigating exploitation prospects in order to integrate our approach and results into existing Web frameworks;
- for the Cluster Engine, we aim to investigate ways to exploit content-based information in order to overcome some of the limitations of collaborative filtering.

The achievements of the Service-Finder project are exploitable in different scenarios:

- Web 2.0 developers may use Service-Finder for building mash-ups of lightweight services.
- ICT Companies engaged in Application Integration scenarios in large enterprises may actively use Service-Finder to look for “address validation”, “fraud detection”, “credit worthiness”, etc.
- The Service-Finder technologies may be sold as an Appliance capable of providing services search capabilities at very low installation and maintenance costs, due to the automated crawling and annotation features.

Further Information

For further information, please visit <http://www.service-finder.eu> or contact the coordinator of the Service-Finder Project: emanuele.dellavalle@cefriel.it