Search Algorithms for Discovery of Web Services

Janette Hicks, Madhusudhan Govindaraju, Weiyi Meng
Department of Computer Science,
State University of New York (SUNY) at Binghamton, NY
Email:(jhicks, mgovinda, meng)@binghamton.edu

Abstract

Web services are designed to standardize interactions between heterogeneous applications using Internet technologies. Within the framework of Internet search technologies, Web services provide structured channels to access search engines and web-accessible databases. Our work involves research in methods to discover Web Service Description Language (WSDL) documents, which provide interface formats, expected data-types, supported protocols and precise service endpoints. This project extends current discovery research through use of the Google Web service, UDDI category searching, and private registry querying with preliminary experiments resulting in a very high percentage of success. The goal is to find WSDL documents for a given domain name, parse the desired service document to obtain invocation formats, and automatically invoke the Web service. Contributions of this research will support enhancements of HTML-dependent search tools by providing access to data inaccessible through surface HTML interfaces.

Key Words: Web Service, UDDI, Search Algorithms

1. Introduction

The goal of this project is to enhance discovery of Web services through design of automated discovery algorithms to provide a foundation for utilization of Web services in enhancing meta-search engine capability. Web services have the potential to significantly enhance several aspects of meta-search engines. A known problem is the difficulty of incorporating search results from HTML search engine pages. Parsing and extracting the output (i.e. search result records) accurately is complex due to the lack of semantic tags in HTML documents. In contrast, description of the service and details of invocation and response are well defined with Web services. Service specifications are defined in the WSDL document for a service via XML-Schema based type system. Paralleling discussions on tapping the "hidden web" [1], Web services have functionality to provide access to information not currently available on the "surface" web, the web represented by static HTML pages. While search engines are also more homogeneous in grouping data, Web services allow groupings of functionality that is heterogeneous in nature. For example, a meta Web service for travel-related Web services grouped into an overall plan including airfare, hotel, and rental car. Web services allow for expansion of existing search engine technology and access to greater depths of the Internet offered information while providing data in a highly structured format.

This discovery research targets any kind of Web service that currently exists and is accessible from the Internet. Google and Amazon represent the broad range of search services for which Web service interfaces are currently available. The project includes search of any type of Web service, since *discovery* techniques should be applicable to future Web services developed for search engines.

In researching Web service discovery standards we found that other than those based on a central registry service; no standard is currently widely implemented. Proposed standards include Microsoft's Web Services Dynamic Discovery (WS-Discovery) [2], which is a dynamic discovery specification that defines a multicast-based ad-hoc protocol to find Web services on a network. An older Microsoft standard, Discovery of Web Services (DISCO) [3], is a static standard using *.disco* documents for a domain to identify the WSDL documents. The WS-Inspection Language (WSIL) [4] is a similar lightweight discovery standard promoted by Microsoft and IBM that merges their earlier efforts: IBM's Advertisement and Discovery of Services (ADS) and Microsoft's DISCO. WSIL

documents contain locations of WSDL documents and pointers to other WSIL documents to create hierarchical lookups for services [5].

Another source of Web services is independent Web service "brokers" that feature their own registries of Web services, such as Xmethods at www.xmethods.com. The main source of discovery proposed in current research is the use of a universal public registry, such as the public Universal Description, Discovery, and Integration (UDDI) [6] Universal Business Registry (UBR) jointly hosted by Microsoft, IBM, SAP, and NTT Communications.

The research question addressed in our work is how to determine if a given domain or web site provides Web services interface(s) and, if so, derive the WSDL document location(s). The search includes current discovery methods and new algorithms we have developed for this purpose. Once the desired WSDL document is found, services and operations are parsed and presented to the user who decides on options for dynamic invocation of the service. The project design and implementation are set up to start with a given domain name and to measure, using developed methods, how well a specific WSDL document is successfully found for this domain across all Internet resources. Current standards only cover discovery of WSDL documents in a central registry or propose standards that are not universally implemented, resulting in a lack of coverage of all existing WSDL documents.

This work expands on the topic of Web service discovery by trying alternative methods to both current centralized and de-centralized standards and by proposing a consolidation of discovery methods. In this "Combined WSDL Discovery" approach, we have implemented several mechanisms for finding WSDL documents: 1) utilizing search engines to find web pages with WSDL documents; 2) using search engines to find distributed lists of WSDL documents (WS-Inspection or WSIL documents); 3) using web crawling programs to discover WSDL documents; 4) querying UDDI registry structures using the common business name search and also using a category-type search; and 5) running queries against a popular commercial registries.

An initial pool of 40 WSDL documents was used in developing these discovery techniques. Final experiments using a test base of 80 WSDL documents produced results with very good success rates in finding those documents using these methods. In developing techniques to achieve this success, numerous experiments were conducted, with minor variations in search terms used, types of queries, and sources.

2. References

[1] Wang, Jiying, "Information Discovery, Extraction and Integration for the Hidden Web", *Proceedings of the VLDB PhD Workshop*, 2003.

[2] Microsoft Web Services Dynamic Discovery (WS-Discovery) Specification, April 2005.

[3] Skonnard, Aaron., "Publishing and Discovering Web Services with Disco and UDDI", *MSDN Magazine*, Feb 2002.

[4] IBM, Microsoft, Web Services Inspection Language (WS-Inspection 1.0), Nov. 2002.

[5] Brittenham, Peter, "An Overview of the Web Services Inspection Language", *IBM Developerworks*, June 1, 2001.

[6] UDDI Version 3.0.2, Feb 2005. Available from World Wide Web: (http://uddi.org/pubs/)