Semantic Web Service Discovery and Integration using Service Search Crawler

Viriya Kaewmarin, Ngamnij Arch-int, Somjit Arch-int

Department of Computer Science, Faculty of Science
Khon Kaen University, Khon Kaen, THAILAND
4950201605@stdmail.kku.ac.th, ngamnij@kku.ac.th, somjit@kku.ac.th

Abstract

Currently, Web services have been identified as the salient technology in providing a flexible solution for interoperating the heterogeneous application systems. Due to the dramatic increasing number of available Web services, how to discover the relevant services dispersed over a number of service providers is a big challenge. Although, the current service discovery mechanism can be achieved through the public UDDI, the UDDI structure is not powerful enough for computer-interpretation to enable automatic Web service discovery. This research proposes the layered-architecture of semantic service discovery system by incorporating search crawler as a core component for discovering services resided on the provider websites. The search crawler can operate in multi-threaded environment to enhance the capability of discovering a number of distributed Web services simultaneously, as well as in various UDDI registries. In order to enrich service discovery in a semantic manner, the Web services descriptions returned from the search crawlers are thus modeled into a machine-processable representation language such as OWL-S. The system conformed to the proposed architecture will provide the flexibility and extensibility to accomplish complex Web service requests that meet user-specified functional requirements.

1. Introduction

Nowadays, software development has stepped toward the service orientation. The Service Oriented Architecture (SOA) [1] has been widely used as to describe IT infrastructure which allows different applications to exchange data with one another as they participate in business processes. The technology which currently supports the SOA is a Web service [2]. On account of the fact that the features of the Web service technology has been developed to be an open standard which allows the different application programs to work corporately via the internet system using SOAP (Simple Object Access Protocol) [3] – the use of XML as a universal standard in exchanging data among different task systems. To achieve the Web service discovery, the Web service providers must describe their services with the standard Web Services Description Language (WSDL) [4] and advertise through the Universal Description Discovery and Integration (UDDI) [5]. However, the current service discovery mechanism through UDDI is not powerful enough for computer-interpretation to enable automatic Web service discovery and invocation. Moreover, there are numerous UDDI registries available on the Web. The Web service providers can register with either one UDDI or multiple UDDI’s, or may not register with any UDDI’s because they just post the Web services on their own websites. Therefore, the service requesters have to search for the services in each UDDI or in the service providers’ websites. As a consequence, the services discovery process is often time consuming and may receive irrelevant results. In order to cope with this problem, this research proposes the semantic service discovery and integration system by using the service search crawler as a mechanism to collect the Web service description dispersed over a number of heterogeneous sources, as well as in the various UDDI registries. The collected Web service information was categorized and indexed to the service database of the system to facilitate service discovery and invocation. The search crawler can operate in multi-threaded environment to enhance the capability of discovering a number of distributed Web services simultaneously. Additionally, to enrich the service discovery in a semantic manner, the Web services information returned from the search crawlers are thus modeled and transformed into semantic Web services conforming to the machine-processable representation language such as OWL-S [6].

This research paper is comprised of five parts. The related literatures are revealed in the first part and then the design of semantic services discovery system using search crawler and other components are described in the second part. The next part concerns with the results
of the study leading to the real practice of the service discovery system. The final part consists of the conclusion of the study and the further study.

2. Literature Review

Many studies have been conducted in an attempt to construct the Web service discovery system to search for the Web services. For example, the study of Web Service Search Engine has introduced the architecture and web service discovery system in the form of searching the Web services from the existing UDDIs. The system retrieves using the index and can search for the Web services by identifying keywords. Another example belongs to Semantic Web Service Discovery in the OWL-S IDE. [7] This method uses some methods to solve the problems that cannot be resolved by the UDDI in order to support the Web service description of the semantic service discovery system and other conditions. It matches the document describing Web services with the standard form of UDDI and keeps them in the knowledge database using a logical search.

Although many researches focus on the UDDI development which supports the increasing numbers of Web services, it is still difficult for the service requesters to discover the relevant services. The main obstacle is that the UDDI’s service description does not provide a well-defined semantic for service representation and sufficient expressive power for solving the semantic service discrepancies, occurring with the disagreement on the meaning, interpretation, or intended use of the same or related service information. Additionally, if the service providers do not register their Web services through the UDDI system, it is still difficult for service requester to discover the services.

3. The Reference Architecture

This research has introduced the architecture of semantic service discovery and integration system using service search crawler as depicted in Figure 1. The architecture consists of four main layers, namely, client layer, mediator layer, search layer and resource layer. Each layer can be described in details as follows.

![Architecture of the semantic service discovery and integration system using search crawler](image)

3.1 The client layer is the highest layer of the architecture, consisting of the service administrators, service providers, and service requesters. The service administrators communicate with the service search crawler generator of the mediator layer to setting up the time schedule for creating and managing the service search crawlers. They also communicate with the service management engine to manage the service data gathering from the service search crawler and classify services into service categories. The next group is comprised of service providers or Web service developers who need to publish more about their Web services. The service providers can register their services through the Service Registration Engine. Another group comprised of service requesters or Web service developers who need to search for the desired services. The service requesters communicate with the service discovery engine to search for the Web service data stored in the service database. This layer provides location transparency to clients for maximum flexibility. The service requesters can search requests without any prior knowledge of local sources or remote sources to be accessed through Web services. In addition, the service requesters can search for the services by specifying a part of the service name with a simple-condition, or specifying a part of the operation, input and output parameters with complicated-condition identification.
3.2 The mediator layer is designed as the principal layer of the system comprised of the Service Search Crawler Generator (SCG), the Service Management Engine (SME), the Service Database (SDB), the Service Discovery Engine (SDE), and the Service Registration Engine (SRE). The SCG functions are to respond to the system administrators’ command in generating the search crawlers according to the specific time schedule, as well as consolidating the service data obtained from the search crawler and subsequently forwarding it to the SME. The SME functions are composed of analyzing the service data and transforming the service data which is in the form of the WSDL documents into the semantic documents which is in the form of the OWL-S documents, as well as transforming and storing the OWL-S document structure into the Service Database (SDB) for providing further semantic search. In addition, the SME can also respond to the service administrators’ commands in administrating and managing the service data stored in the SDB, namely correcting and improving the service data, and grouping the service data into categories. The next component is the SRE that enables other service providers to register more services and information as described in WSDL format to the system and subsequently forward to the SME. The last component is the SDE whose function is to respond to the service requesters’ commands. The SDE will search for the requested services in SDB according to the conditions the service requesters have specified initially.

3.3 The search layer includes service search crawlers which perform in a multi-threaded environment. The SCG can set up number of the crawler to be generated on this layer. The search crawler will search for Web documents from different websites on the internet and analyze if a Web document provides the Web service links conforming to the patterns of the Web service document. If the Web documents are matched with the patterns, the service search crawler will collect these documents and subsequently forward them to the SCG on the upper layer.

3.4 The resource layer is composed of websites in the internet network. Each website can provide Web documents that describe Web service information, such as the URL of WSDL (e.g., http://api.google.com/GoogleSearch.wsdl), WSDL-S and OWL-S. This layer also includes the UDDI websites which store only the documents describing the Web services.

4. The Service Data Analysis of the Service Management Engine

One of the SME function is to analyze the service data which is in the form of WSDL document and transform it into semantic service data. The algorithm showing the procedure of service data analysis is given in Figure 2 which composes of the following steps.

4.1 Service data classification: This process reads a Web service document and classifies the service description appearing in the document, such as the service name, the operation, the input parameters and output parameters. These service descriptions are prepared before forwarding to the next step OWL-S document transformation module.

4.2 OWL-S document Transformation: This process reads the service descriptions of the WSDL document and transforms them into the OWL-S document for each operation specified in the WSDL document.

4.3 Transforming and storing the OWL-S document structure into the Service Database (SDB): This process transforms the OWL-S document structure into relational structure of service database. To capture the semantic elements of the OWL-S document structure, the service database structure must encompass the service profile, service model, and service grounding structures. A partial structure of the service database is depicted in Figure 3. The semantic service structure enriched to the service database enables the semantic service discovery and invocation in the future.

```
Analyze Service Data Algorithm

While QueryRawdata() != null do {
    url = Query_Rawdata_url();
    Input (url);
    If Query_Service_Profile(url) = 0 then
        Read_ServiceProfile(url);
        Read_SizeOperation(url);
        For (int i=0 ; i<=SizeOperation ; i++)
            getOperationName(i);
        Read_SizeInputParameter (i)
        Read_SizeOutputParameter (i)
        for (int j=0 ; j<=SizeInputParameter ; j++)
            getInputParameterName(j);
            getInputParameterType(j);
        for (int k=0 ; k<=SizeOutputParameter ; k++)
            getOutputParameterName(k);
            getOutputParameterType(k);
        CovertToOWL-S(i);  SaveToFile(i);
        Copy_WSDL_File(); SaveToDatabase();
    Else End; }  Delete_Rawdata_url();
```

Figure 2. Algorithm for Web service document analysis
5. Results of the Study

The study has implemented the service discovery and integration system conforming to the architecture designed in Figure 1. The SCG was implemented with ItSucks API [8] and OWL-S API [9] under Java JDK 1.6 environment. The SDE, SME and SRE were developed by Java Server Page (JSP) under Apache Tomcat Web Server version 6.0. The service database was created by using MySQL database. The SDE provides both normal and advanced search for services, as such the service requestors can specify a part of the service name, description, category, operation, or parameter through the normal search as illustrated in Figure 4. On the other hand, the advanced search permits the service requesters to input more conditions for searching, such as searching from more than one operation or parameter, in order to discover the most relevant services for requesters. Upon submission of a requester query through the SDE, the service result returns the URL endpoint of the Web service, service name, text description, category, operation name, parameter name, parameter type, as well as the WSDL and OWL-S documents. Consequently, the service requesters could select the target address of the desired document for further Web service development and invocation. A portion of search result is shown in Figure 5.
Additionally, in the section of administrating and managing system, the service administrators can issue a command to add, edit, or delete the service data through the SME, as well as update the service profile part as shown in Figure 6. Moreover, the service providers were also allowed to add their Web service data through the SRE in case their service data did not appear in the service database.

6. Conclusion and Further study

The research has introduced the semantic service discovery and integration system using service search crawler as a core component for discovering and integrating services residing on the provider websites. The search crawler can operate in a multi-threaded environment to enhance the capability of discovering a number of distributed Web services simultaneously, as well as in various UDDI registries. The Web services descriptions represented in WSDL format that are returned from the search crawlers are thus analyzed and transformed into a machine-processable representation language such as OWL-S. This allows computers and applications to understand the meaning of the documents describing the Web services and to enable the semantic service discovery and invocation in the future.

References


