Automatic Discovery and Composition of OGC Web Services

Aluna: Eveline Russo Sacramento
Orientadora: Vânia Maria Ponte Vidal
Co-orientador: Marco Antônio Casanova

Doutorado em Ciência da Computação
Início: Março de 2005
Término (previsto): Julho de 2009
Etapas concluídas: Exame de qualificação

Departamento de Computação – Universidade Federal do Ceará (UFC)
Fortaleza, CE – Brasil
Departamento de Informática – Pontifícia Universidade Católica (PUC)
Rio de Janeiro, RJ - Brasil

eveline@lia.ufc.br, vvidal@lia.ufc.br, casanova@inf.puc-rio.br

Abstract: The Semantic Web is based on the idea of annotating Web content with machine-interpretable metadata such that computers are able to process this content on a semantic level. These semantic annotations make it possible a higher degree of automation and allow that more precise results can be achieved in tasks like Web services discovery and composition.

This thesis proposes an approach that uses semantic annotations for the automatic discovery and composition of Geospatial Web Service. The thesis first discusses how to combine semantics provided by metadata catalogues with semantics provided by ontologies to produce catalogue services capable of efficiently searching suitable Web services. Then, it addresses how to compose more complex Web services from basic OGC Web services in an automatic or semi-automatic way. The thesis focuses on the composition of data services and portrayal services, and adopts existing SWS technology (WSMO/WSML/WSMX).

The first contribution of the thesis is to integrate SWS technology into existing OGC Web catalogues for the purpose of enhancing the functionality and improving the hit-rate of the discovery service. The central contribution is that this thesis goes beyond simple query mediation, covering semantic service composition. It uses the expressiveness of WSML, coupled with the power of a WSML reasoner, to determine the best set of services to be composed to fulfill specific user’s goals.

Keywords and phrases: Geographic Information Systems; Interoperability and Information Integration; Data Modeling and Data Semantics; Semantic Web Services; Spatial Databases.
1. Introduction

A Web service is a software system designed to support interoperable machine-to-machine interaction over a network. It has an interface described in WSDL and interacts with other components using SOAP messages. The main operations supported by Web services are publication, discovery, selection, invocation and composition of services. In the spatial domain, we count on Geospatial Web Services (GWS) that provide access to geospatial data and expose basic processing functionality using Web service technology. There are many types of GWS, such as Web mapping services, spatial analysis services, metadata services, and visualization services. They were developed independently of the W3C protocol architecture for Web services.

The geospatial community has already developed a set of specifications through standard-setting bodies, including the ISO Technical Committee 211 (ISO/TC 211), the Open Geospatial Consortium (OGC) [19], among others. The OGC comprises more than 350 companies, government agencies, and universities with the purpose of creating and promoting the development of technologies that facilitate interoperability between systems that process spatial and location data. The OGC working groups publish their results as syntactic interface specifications and standards for data interchange. Two important OGC initiatives are the Geography Markup Language (GML) [21] and the Web Feature Service (WFS) [22]. Metadata standards for the description of geodata exist as well as catalogue services to search them [3]. OGC also proposed an architecture called the OpenGIS Services Framework (OSF) [5] that identified services, interfaces, and exchange protocols that the applications may use [20].

However, challenges remain in supporting the crucial tasks of discovery and retrieval of information sources that meet the user’s needs [24]. The proposed model for GWS composition also lacks some important features [11]. It does not allow the integration of Web services into higher abstraction levels or frameworks. No descriptive language exists that allows combining any number of Web services into a single or complex service chain. Even using a simple service-service interaction, no rules or execution constraints can be defined. The OGC discussion does not address semantic aspects at all. Outside the geospatial domain, these issues have been addressed by a number of consortia, software vendors, or initiatives.

The idea of the Semantic Web is to annotate software being offered via Web service interfaces with machine interpretable descriptions describing what the software does and how it does it. By developing ontologies that capture service semantics, one will be able to publish services in a way which is both semantically rich and yet machine interpretable. Semantic Web Services (SWS) apply such Semantic Web technology in the Web Services area. In this way, a higher degree of automation and more precise results can be achieved in many tasks. For example, it allows more sophisticated service discovery with a precision which is higher than that currently possible with UDDI or WSDL. To support service discovery, one may resort, for example, to taxonomies that structure service capabilities and domain ontologies that provide a suitable vocabulary covering a specific application area. When composing complex Web services from basic Web services, automated planning algorithms from AI can be employed [27].

The union of OGC and Semantic Web efforts resulted in an area called Semantic Geospatial Web [10], that will bring significant advances in the meaningful use of spatial information. With the flexible incorporation of geospatial semantics into the Web, geospatial information retrieval will become precise to the level that the results of user queries will be immediately useful, without weeding out irrelevant hits. So, a convenient way to provide universal access to geospatial data sources is to implement
GWS that encapsulate the data sources, because this strategy obviates the need for traditional means of data distribution. Under this new perspective, the Web search and retrieval of geospatial data and services will not only be based on syntactical or lexical information, as do many current approaches, but on the semantic information specified by metadata, gazetteers and geo-ontologies [4].

However, there are situations where a single basic service can not meet the search criteria, thereby requiring the composition or chaining of two or more services to produce a more complex service. Furthermore, if one of the component services is unavailable, a replacement should be found and inserted, so the complex service can still be provided. The product of Web service composition is called an orchestration. Web service composition can be static, when Web services are composed at design time, or dynamic, when Web services are composed at execution time (on-the-fly); the composition can be manual, when the Web service discovery and invocation tasks require human intervention; or automatic, otherwise.

SWS technology provides means for describing services, and also infrastructural capabilities to discover services and to enable interoperation. However, it does not provide the reasoning capabilities required to decide which service is needed, which provider is best, how to negotiate the parameters of a service and what actions to take when using a service. If a service is simple and used in a straightforward way, the reasoning capabilities will also be simple. However, complex reasoning capabilities, such as negotiation or dynamic planning, may be required for industrial strength applications. In these scenarios, SWS technology has to be used in conjunction with other computer science disciplines, such as AI [27], that provides sophisticated knowledge representation approaches, inference machines and planning algorithms.

2. Contributions

This thesis proposes an approach that uses semantic annotations for the automatic discovery and composition of GWS. The thesis first discusses how to combine semantics provided by metadata catalogues with semantics provided by ontologies to produce catalogue services capable of efficient searching for suitable Web services. Then, it addresses how to compose Web services from basic OGC Web services in an automatic or semi-automatic way. The thesis focuses on the composition of data services and portrayal services, and adopts existing SWS technology (WSMO/WSML/WSMX) [8].

The first contribution of the thesis is to integrate SWS technology into existing OGC Web catalogues for the purpose of enhancing the functionality and improving the hit-rate of the discovery service. The central contribution is that this thesis goes beyond simple query mediation, covering semantic service composition. It uses the expressiveness of WSML [9], coupled with the power of a WSML reasoner, to determine the best set of services to be composed to fulfill specific user’s goals.

To help facilitate Web services automated composition, during the planning step, we can resort to a Library of Typical Plans [15]. A plan [1] is a description of how an agent (or class of agents) usually proceeds towards some goal. It consists of either a set of partially ordered operations or plans, or of a set of specialized alternative plans able to achieve the goal. A Library of Typical Plans can be used by plan-recognition algorithms to detect which possible plans the agents may be trying to perform. The recognized plans can then be used in simulations of future events. Another way to obtain a plan is to select a pre-existing plan and adapt it, if necessary. The geospatial services taxonomy offers a high-level geospatial service thematic classification and can also be used to assist in the grouping and search available GWS. Web services composition will be
implemented by a Prolog planner [6, 14, 15]. The resulting plan will be published in the WSMX environment.

One of the major challenges of this thesis lies in identifying and solving differences between OGC and SWS paradigms to account for the specific requirements of expressing geospatial semantics.

3. Present State of the Work

This thesis focuses on data services, that allow access and updates to data held by a geographic data source, and portrayal services, that offer specific support for the visualization of geospatial data. They include Web Feature Service (WFS) and Web Map Service (WMS).

These services will first be specified using a formalism based on a temporal logic (STRIPS) [13]. Then, the obtained axioms will be translated to the clausal format required by the Prolog implementation of a planning algorithm [6, 14, 15].

The next step consists of conducting experiments on the Web, using WSMO/WSML/WSMX, one of the emerging SWS initiatives. This technology will be used because of its strong decoupling of the different components, its high impact on service mediation [2] and the support offered by design tools, such as WSMT and WSMO Studio [27]. WSMO [12, 25] defines four modeling elements (ontologies, goals, Web services and mediators) to describe several aspects of a SWS. The ultimate goal of WSML is to enable the total or partial automation of the following tasks: discovery, selection, composition, mediation, execution, monitoring, etc.

The specifications of the OGC metadata catalogue service interface will be translated from ISO 19119:2005 [17] metadata standard to OWL [26]. Then, the implementation of the services will be specified in WSML, generating WSMO ontologies, and integrated into SWS technology. The second step, called data mediation, uses the WSMX data mediation architecture. It consists of resolving semantic mismatches between WSMO ontologies obtained from goals and Web services, using mappings rules between their concepts. The third step, called process mediation, uses the WSMX process mediation architecture. It consists of reconciling the differences between the processes implemented by OGC services. Figure 1 shows a draft of the proposed architecture.

A disadvantage of the WSMO approach is that it was designed to support W3C Web Services. GWS was not explicitly considered in the development of WSMO. However, the generic approach followed by WSMO enables a seamless interaction between W3C and OGC Web services. Services of both types can be used as components inside the same composition [16], because they do not necessarily need to be WSDL- and SOAP-based in order to be composable within this framework.

Other disadvantage of WSMO is that the process of building the required ontologies is manual, taking a considerable amount of time, and heavily relying on domain experts’ knowledge. It is also hard to capture the user’s goals and formulate them in WSML [18]. Finally, in order to connect GWS, specific adapters need to be implemented.
4. Case Study

Data from FUNCEME (Meteorology and Water Resources Foundation) and SRH (Secretariat of Water Resources) are provided to the public in general through their respective Web sites. These data are generated and maintained in an independent way. FUNCEME, for example, manages information about water quality of some dams located in the metropolitan region of Fortaleza, while SRH manages general information about all dams of the State of Ceará. In this context, a challenge is to integrate and to display map-like views of information stored in these heterogeneous geographic sources in order to integrate information provided by both sources and display a map of the state with all dams.

A solution to this problem, without data replication, is to use mediation architecture and Web Services technology and it is shown in Figure 2. The SWS framework WSMO/WSML/WSMX can be used as a mediator. The FUNCEME database uses an adapter that supports transactions (WFS-T), whereas the SRH database uses an adapter that supports queries (WFS) and a second adapter that supports map visualization (WMS). These Web services are registered with WSMX by FUNCEME and SRH.

A possible goal, submitted to the mediator, would be: “Show the cities with more than 40 thousand habitants in the Ceará State that have dams, with their relevant information”.

Figure 1: The Proposed Architecture.

Figure 2: A Solution to the Case Study.
The following steps abstract a solution to this class of user requests:
1. The client’s request is expressed in an informal language.
2. The request is transformed into a WSML goal and sent to WSMX mediator.
3. The mediator discovers, selects, composes and invokes suitable Web Services.
4. The mediator generates a WMS map by combining the results returned by the Web Services.
5. The mediator sends the WMS map back to the client, which then visualizes it.

This query uses both data sources and the connection between them can be made by city code, a common attribute to these sources. The obtained map plots the State of Ceará, divided by cities, and shows the existing dams to those cities whose population is greater than 40 thousand habitants. Figure 3 shows the WSML Ontology and the WSML Goal. They were manually created at design time by an expert domain.

5. Selected Related Work (Comparative View)

DERI [28] is one of the participant projects of the Semantic Web Services Challenge [23]. The primary requirement of this workshop is to achieve semantic interoperability between a partner using a proprietary system called Moon and a partner using RosettaNet. DERI’s team based its solution on the specifications of SWS technology (WSMO/WSML/WSMX). They enabled interoperability between both partners at syntactic, data and process levels [7]. Syntactic interoperability was achieved by adapters in order to perform XML2WSML (lifting) and WSML2XML (lowering) transformations, because WSMX operates on WSML while RosettaNet and Moon internal systems use XML. Data interoperability was achieved by the WSMX data mediator. Process interoperability was achieved by the WSMX process mediator. In our approach we also adopted SWS technology. DERI’s solution covered mediation and
discovery scenarios, presented by the SWS Challenge. They did not cover composition of Web services as we propose to do.

SWING [24] aims at deploying SWS technology in the geospatial domain. In particular, it addresses two major obstacles that must be overcome for SWS technology to be generally adopted: i) to reduce the complexity of creating semantic descriptions and ii) to increase the number of semantically described services. It will develop methods and tools that can hide the complexity – and automate the creation – of the necessary semantic descriptions. The objective of SWING is to provide an open, easy-to-use SWS framework of suitable ontologies and inference tools for annotation, discovery, composition, and invocation of geospatial web services. We also adopted SWS technology and we propose to cover discovery and composition of Web services. However, SWING’s solution covers only composition of WFS services.

6. Results Evaluation

This thesis proposes an approach that uses semantic annotations for the automatic discovery and composition of GWS and adopts existing SWS technology (WSMO/WSML/WSMX). Most of the current studies focus on discovery and retrieval of GWS [16]. GWS, for example, are mostly isolated, stateless applications [11]. Only recently attention is shifting towards GWS composition. This thesis intends to go beyond simple query mediation, covering semantic Web service discovery and composition. The thesis proposal will be presented next August.

References


