Ontology-based Method for Geospatial Web Services Discovery

Guiwen Lan[1] Quanyi Huang[2]

[1]School of Geodesy and Geomatics, Wuhan University, Wuhan, 430079, P. R. China
[2]Center for Public Safety Research, Tsinghua University, Beijing, 100084, P. R. China

Abstract

Finding suitable geospatial web services is a crucial task in the open and distributed environment. However, limiting to the current web service technology, keyword-based search in catalogs does nothing to overcome the semantic heterogeneity, leading to find suitable services difficultly or to discover numerously redundant results. Based on study of geospatial web services and the semantic heterogeneity existing, this paper proposes an ontology-based approach for geospatial web services discovery by making use of elements in service capabilities documents. The matching algorithm between request and service description is discussed in detail, and applied to a prototype system for emergency management.

Keywords: Geospatial web service, Ontology, Semantic heterogeneity, Service discovery.

1. Introduction

Geographic information is the key to effective planning and decision-making in a lot of disciplines. The success of planning and decision-making could be interpreted as "getting the right resources to the right place at the right time; to provide right information to right people to make the right decisions at the right level at the time". The emergence of geospatial web services makes it possible for people to find necessary geographic information over the Internet. Over the past years the Open Geospatial Consortium (OGC) and the International Organisation for Standardisation (ISO) have developed geoinformation technology standards that provide the essential basis for syntactic interoperability and cataloguing of geospatial web services.

Composability is often seen as one of the main strengths of OGC's geospatial web services. Therefore OGC's standards are wildly supported by GIS vendors and used in a lot of domains. While in an environment where services are previously unknown, a service that is appropriate for answering a given question from among a large number of available services has to be discovered first. Service discovery, thus, is a crucial task that will become even more important in the context of geospatial web services. However, the use of different vocabulary in the different application domains might lead to semantic heterogeneity problems when only simple keyword-based search is employed to find relevant information in a catalogue. In this paper we proposed an approach for geospatial web services discovery, the remainder of this paper is organized as follow: Section 2 presents the semantic problems caused when a client searches for services; Section 3 gives a brief review on definitions of "ontology", and then introduces some ontologies for geospatial web services; Section 4 introduces an semantic matchmaking approach, which is applied to a prototype system.

2. Heterogeneity of geospatial web services

Different types of heterogeneity have been identified for GIS and geospatial web services in general, that block interoperability between information providers and users/clients, including system heterogeneity, syntax and schematic heterogeneity, semantic heterogeneity. System interoperability refers to the ability to deal with the hardware, operating systems, and communication heterogeneity, such as instruction sets, communication protocols, different file systems, file types and so on. Syntax and schematic interoperability is relevant to data representation, formatting, data models between web services. Semantic interoperability has more to do with the meaning of the data, it assures that contents of data and services are correctly understood when services are requested. E Klien et al. distinguished semantic heterogeneity into two types [3]: (1) Naming heterogeneity, the same real world facts are understood in the same way but are named differently; (2) Cognitive heterogeneity: Because of different perspectives on the same real world facts there may not be a common base of definitions of the underlying facts between two disciplines. Problems can occur if

these semantic differences are concealed because the same term is used for different concepts, while keyword-based search does nothing to overcome those problems. [3]-[9] have presented examples to illustrate problems of semantic heterogeneity.

Syntactical heterogeneity of web services can be overcome mainly using two common web service standards: Web Service Description Language (WSDL) and Simple Object Access Protocol (SOAP). WSDL is used to describe a Web service in terms of its interfaces and SOAP formalizes the XML-based message transportation between Web services [7].

In the GI community, standards from bodies like the OGC provide the basis for syntactic interoperability. OGC and ISO adopted a so-called "publish-find-bind" pattern of web service architectures. Service discovery is handled by a service registry that provides service metadata with details on service types, as defined in ISO 19119 (Services). Service metadata can be queried and managed with catalog services. When semantic heterogeneity is concerned, there is still no answer given by OGC. Today there is no OGC specification that deals with semantics in support of service (and data) discovery, though an attempt has been recently undertaken in the OGC GeoSemantic Web Interoperability Experiment.

3. Ontologies for geospatial web services

In order to solve the semantic heterogeneity and allow machine-automation of data integration, it is desired that the semantics of the data are defined explicitly and represented in a machine process-able way [8], namely "ontology". In this section, we first give a brief review on the definition of "ontology", and then introduce a ontologies for geospatial web service discovery.

3.1. Ontology

The word, ontology, originates from philosophy, and stands for "the study of being existence". Computer science borrows this word from philosophy, which "defines the basic tiers and relations comprising the vocabulary of a topic area as well as the rules for combing terms and relations to define extensions to the vocabulary" [3]. In [10] "ontology" is referred to the science of describing the kinds of entities in the world and how they are related. Although there are a lot of definitions about ontology, two important aspects in defining ontologies can be easily distinguished:

- Explicit: "Explicit" stands for the meaning of the types of concepts that are used in the conceptualization and the constraints of their usage are well-defined.
- Formal: "Formal" refers to the fact that the ontology is defined in an artificial and well-defined language so that it is machine-readable.

Merging the above definitions, E. Klien et al. [3] defined "ontology" as "an explicit formal specification of a shared conceptualization". We adopted the definition from Wei Xu & Sisi Zlatanova, to reckon ontologies as "explicit, formal and shared ontologies".

3.2. Ontologies for geospatial web services discovery

In the GI community, several ontologies have been proposed in recent years. In this paper we introduce three type of ontologies those are most relevant to service discovery, i.e., the global ontology, geospatial domain ontology and geospatial service ontology.

The global ontology is the core-shared vocabulary for all of human consensus reality. It contains bridge concepts between different domains and acts as a common language that all other ontologies must reference. Metadata standards such as Dublin Core Metadata, are generally extended as the basis to define upper level concepts and assertions about these concepts.

Geospatial domain ontology aims at providing the core conceptualization and knowledge structure of geospatial domain. For example, "highway" is a kind of "road", "village" belongs to "residential area". This ontology represents the problem space over which the user will query, including concepts representing the geospatial world from the perspective of the relevant domain and covering the knowledge about the following aspects: (1) spatial-temporal factors, e.g. location, time and unit; (2) physical facts, e.g. physical phenomena, physical properties and physical substances; (3) disciplines, e.g. scientific domains and projects, (4) data collection, e.g. instruments, platforms and sensors. In [5] an ontology of airport is used to exemplify the definition "domain ontology".

Geospatial service ontology enables automatic discovery, invocation and composition of services registered in the catalog maintained by the broker. The design and creation of service ontology can be based on the capabilities document of OGC's web services. In fact, service ontology consists of the following elements: (1) service profile, which describes who provides the service, what the service does, as well as other properties of services, allows the knowledgebase

to infer whether or not a particular service is appropriate to a given problem; (2) service "process model", which states the inputs, outputs, preconditions and effects of a service, allows the knowledgebase to figure out whether or not a service meets the requirements as well as the conditions to invoke the service; service "grounding", which presents the ports, protocols and encoding of invocation, tells knowledgebase how to invoke a service.

To sum up, a client/user initiates the request by combining the geospatial domain ontology with the global ontology. The request is translated to the geospatial service ontology. The geospatial service ontology offers matching services and data sources to match the requests.

4. Ontology-based approach for service discovery

A search engine is generally provided by the broker to search for suitable services for requestors. The responsibility of the matching engine is to match service descriptions with requests. The search engine should be designed to allow the user to describe his requirements by specifying a detailed description so that it can be compared semantically to the descriptions of services. The most relevant service compared to the query needs comes first.

In [7] an algorithm was proposed to search for appropriate e-services for requestors by matching the inputs and outputs of the request against that of the services. The similarity between services is evaluated through the computation of coefficients obtained by comparing input/output parameters that services exchange during their execution and operations that they are able to perform. Following the intuitions from [7], we build an algorithm for matching a query description from a request and a document description of a geospatial service. We first give an overview of OGC' compliant services, and then present a method to compute semantic similarity coefficient of two concepts in the ontology, based on this, the matching algorithm is proposed at last.

4.1. An overview of OGC web services

The standardised service interfaces originating from OGC, including WMS, WFS and WCS, have commonalities in their design. Fig.1 gives us an overview of OGC web services. One such commonality regards the metadata describing the capabilities of a service. Many of the metadata structures are common, based on the ISO 19115 international standard for geographic information metadata, and are provided by the common operation GetCapabilities. The response from a GetCapabilities request is an XML description of the service's information content and supported request parameters, and is therefore both machine- and human-readable. This capabilities document conforms to an XML schema, partly unique for the particular type of service, which allows clients to validate the response. The capabilities document consists of information of the following sections [1]:

- ServiceIdentification: Metadata about this specific server. The schema of this section shall be the same for all OWSs.
- ServiceProvider: Metadata about the organization operating this server. The schema of this section shall be the same for all OWSs.
- OperationsMetadata: Metadata about the operations specified by this service and implemented by this server, including the URLs for operation requests. The basic contents and organization of this section shall be the same for all OWSs, but individual services can add elements and/or change the optionality of optional elements.
- Contents: Metadata about the data served by this server. The schema of this section is specific to each OWS type, as defined by that Implementation Specification. Whenever applicable, this section shall contain a set of dataset descriptions, which should each be based on the MD_DataIdentification class specified in ISO 19115 and used in ISO 19119.

After a service requestor finds suitable service(s), (s)he can invoke the GetCapabilities operation to get detailed information about this service. For example, a WMS service whose base URL is "http://sms.webmap.cn/scripts/openserv.exe?map=/sms_ogc/sms1500.map&", its GetCapabilities operation can be invoked by sending a request "http://sms.webmap.cn/scripts/openserv.exe?map=/sms_ogc/sms1500.map&SERVICE=WMS&REQUEST=GetCapabilities" to get its capabilities document. Guided by the document, subsequent calls to the more specialised operations of the service can be done manually or in an automated manner by the client, using suitable parameters.
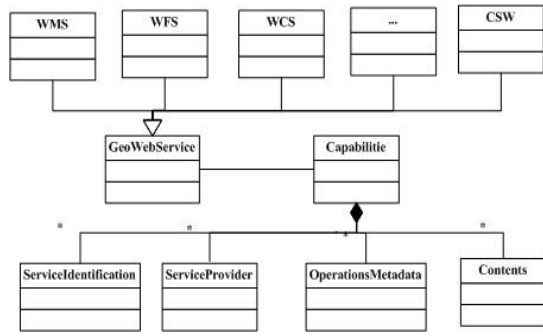
Fig.1 An overview of OGC web services.

4.2. The semantic similarity coefficient of two concepts

Geospatial web service discovery involves the identification of service advertisements/descriptions that match a service request. The matching program breaks down requests and service descriptions into concepts. A match between a request and a document of service description therefore consists of matches of all the concepts provided by the request and the description document. The match of two concepts is done by computing semantic similarity coefficient of them.

Let Ci (i=1,2,…,n) be any concept in the ontology. Further, let D(i) be the domain of Ci, Let s: D(Ci) $\times$ D(Ci) be function on the D(Ci) mapping to R+ (R+ denotes the set of the positive real number, S(C1,C2) denotes the similarity coefficient of C1 and C2. Rules for computing similarity coefficient of two concepts are showed in Table.1.

$$C_1 = C_2 \Rightarrow s(C_1,C_2) = 1;$$
$$s(C_1,C_2) = 0 \Leftrightarrow (C_1 \cap C_2) \subseteq \perp, \perp: \text{Nothing};$$
$$s(C_1,C_2) \neq s(C_1,C_2), \text{ if } C_1 \neq C_2;$$
$$rel(C_1,C_2) = L - PL(C_1,C_2) + k * NCD(C_1,C_2)$$
$$ns(C_1,C_2) = \frac{rel(C_1 - C_2)}{L}$$
$$ds(C_1,C_2) = 1 - \frac{|C_1 - C_2|}{|C_1|}$$
$$s(C_1,C_2) = ns(C_1,C_2) \times w + ds(C_1,C_2) \times (1 - w)$$

Table 1: Rules for computing similarity coefficient

The similarity coefficient S(C1,C2) is calculated by the name similarity ns(C1,C2) and description similarity coefficient (ds(C1,C2)) [5]. They are required to be in the range [0,1]. "rel(C1,C2)" denotes semantic relatedness between two lexically expressed concepts, L and k are constants and PL(C1,C2)

denotes the length of the shortest path between two concepts,NCD(C1,C2) denotes "number of changed directions" of this path. If no such path exists, rel(C1,C2) is zero. "C1-C2" is the difference between two concept descriptions, $|\bullet|$ is the size of a concept descriptions. In [6] the computation of the size of a concept had been discussed in detail.

4.3. The matching algorithm

As mentioned in section 3.2, geospatial service ontology is employed to support service discovery. Most instances of this ontology can be provided by the capabilities document, as discussed in section 4.1. We build a prototype (see in Section 4.4) as a broker that maintains a set of registries (catalog) recording descriptions of geospatial services. A registry in the catalog consists of elements about a service including ServiceIdentification, ServiceProvider, Operations_ Metadata, Contents. The matching process then can be translated to the computation of semantic similarity between the request and the capabilities of geospatial web services. We will first present the main loop in which a request is matched against all the records of service capabilities; then we will discuss the rules for matching each registry (service capabilities) with the request, namely how the degree of match is computed.

The main control loop of the matching algorithm is shown in Table 2. Requests are matched against all the descriptions of services stored in the set of the registries. Whenever description of a service matches against the request, it is recorded and scored to find the matches with the highest degree.

```
match(request) {
recordMatch= empty list
foreach registry in registry_set
 {
    if match(request, registry) then
    recordMatch.append(request, registry)
 }
return sort(recordMatch)}
```
Table 2: Main Control Loop

A match between a request and a record of service description consists of the match of concepts of the abovementioned 4 elements that are picked out from the request. The algorithm to compute degree of match is depicted in Table 3. In Table 3, "reqs" stands for concepts about an element in the request, such as ServiceProvider, "serviceregs" stands for the corresponding element picked out in the registry of service descriptions. The matching of the abovementioned 4 elements is computed following the

same algorithm respectively. The whole degree of match is computed as follow:

$$TotalDegree = SOP \times w_1 + SC \times w_2 + SP \times w_3 + SID \times w_4$$

$$, \sum_{i=1}^{4} w_i = 1, 0 \le w_i \le 1.$$

A match is recognized if "TotalDegree" is higher than a set value (threshold) and the results are sorted according to "TotalDegree". SID denotes the matching degree of ServiceIdentification; SOP, OperationsMetadata; SP, ServiceProvider; SC, Contents. "w_i" denotes the corresponding weight to compute TotalDegree. If there aren't concepts of a certain element presented in the request, the corresponding weight is set to 0. We adopt a strategy that $w_1 \le w_2 \le w_3 \le w_4$ if $w_i \ne 0$, based on the assumption that: (1) if the request contains the ServiceIdentification, the requestor know exact information about the service; (2) if a requester presents some concepts about the ServiceProvider, he is likely to be interested in services provided by a certain provider; (3) generally contents provided by the services is of much concern, and often presented in the requests, however if any of the previous 2 element is also presented, the weight of "Contents" is moderately reduced; (4) generally the requestor pays more attention on "Contents" than "Operation"; (5) a requestor always tries to provide as much as possible knowledge about the wanted service he has previously. Of course, a broker should provide a friendly user-interface to make it easy for the requestors to express their preference.

```
DegreeofMatch(reqs, serviceregs) {
Total= reqs.size
Sum = 0
foreach R in reqs  do {
find S in serviceregs such that
degreeMatch=max(S(R,S))
if (degreeMatch< threshold) {
Total=Total – 1 }
   Else {
     Sum =  Sum + degreeMatch }
return Sum*Total/reqcontents.size }
```

Table 3: Algorithm to compute matching degree between request and service description

4.4. A prototype system

The semantic matchmaking approach has been applied to a prototype system. As in recent years people have found that it is necessary for emergency managers to share geographic information. The prototype system is designed as a broker (catalog service) for lots of information sources, including OGC compliant Web Services such as WMS, WFS, which are provided by the emergency management agencies. When a provider publishes his service to the system, metadata about the service is collected. Moreover, the system will invoke the GetCapabilites of the service and fetch the capabilities document. After parsing this document, all the metadata are stored in a database. The search engine of the system employs the proposed approach to match clients' request and metadata of services. The visualization of OGC's compliant web services has been done by employing OpenLayers library (a javascript library, see in http://www.openlayers.org/).

5. Conclusions

Geospatial web services attract more and more interest within the GI community due to their increasing popularity in main computer science. Therefore to find suitable information from those services has been a hot topic in recent years. In this paper we first give a brief review on the semantic problems of geospatial web services, and then ontologies relevant to service discovery are presented. Service description provided in service capabilities document is used to match against request from clients.

The match algorithm for service discovery is discussed in this paper, and applied in a prototype system for emergency management application. To sum up, our matching algorithm provides a way for discovery, selection and interoperation of geospatial web services, which is a crucial feature in the open and distributed environment. However worth-mentioned is that although there are a lot of literatures dealing with semantic interoperability of geospatial web services, it is still a long way to go to put it into practice. Our future research and development will focus on how to design and create ontologies for emergency management domain, and integrate it with the ontologies proposed in the section 3, then enhance performance of the search engine.

References

[1] A. Whiteside, editor. *OpenGIS® Web Services Common Specification (OGC 05-008)*, Open Geospatial Consortium Inc., pp.10-32,2005.

[2] D. Bianchini, V. De Antonellis, B. Pernici, P. Plebani, Ontology-based methodology for e-service discovery, *Informatiom systems*, 31:361-380, Elsevier Ltd.,2006.

[3] E. Klien, M. Lutz, W. Kuhn, Ontology-based discovery of geographic information services: An application in disaster management, *Computers,*

environment and urban systems, Elsevier Science, 30:102-123, 2006.

[4] S. Gao, Omer F. Rana, Nick J. Avis, D. Chen, Ontology-based semantic matchmaking approach, *Advances in Engineering Software*, Elsevier Ltd., 38:59-67,2007.

[5] D. Kolas, M. Dean, J. Hebeler, Geospatial semantic Web: architecture of ontologies, *Aerospace Conference (2006 IEEE)*, Big Sky,Montana, 2006.

[6] N. Karam, S. Benbernou, M. Hacid, M. Schneider, Semantic matching of Natural Language Web Queries, *Lecture Notes in Computer Science*, Springer Berlin/Heidelberg, pp.416-429, 2004.

[7] M., Paolucci, T.Kawamura, T.R. Payne, K. Sycara,. Semantic matching of web service capabilities. *In 1st international semantic web conference (ISWC2002)*,Sardinia, Italy, 2002.

[8] P. Yue, L. Di, W. Yang, G. Yu, P. Zhao, Semantic-based automatic composition of geospatial web services chains, *Computers & Geosciences*, , Elsevier Ltd., 33: 649-665, 2006.

[9] W. Xu, S. Zlatanova, Ontologies for disaster management, *Lecture notes in geoinfomation and cartography, Geomatics solutions for disaster management*, Springe-Verlag, Berlin Hdeidelberg, pp. 185-199, 2007.

[10] http://www.w3.org/TR/2004/REC-owl-guide-20040210/.