An Ontology-Based Approach for Geographic Information Retrieval on the Web

Mei Kun, Bian Fuling
Research Center of Spatial Information & Digital Engineering, Wuhan University
Hubei, P.R.CHINA
kunjuly@gmail.com

***Abstract*—Finding and accessing suitable geographic information in the open and distributed web environment of current geographic information services (GIServices) is a crucial task. Problems of semantic heterogeneity arise during an exact word matching search for GIServices and spatial database. Moreover, there are only a few services registered in metadata catalogues so that we cannot ignore web contents which also contain geographic information. This paper introduces an architecture of a spatially aware search engine, for semantic interoperability of distributed and heterogeneous GIS on the internet. The proposed architecture is based on mediation and ontologies, which help to providing support for query disambiguation, query term expansion and web resource retrieval. It also gives the structure of the Location Ontology and Service Ontology and shows the detail of semantic query process. Based on a real-world scenario of restaurant guide search, the application of our approach shows that the geographic information retrieval can be efficiently supported.**

Keywords-Semantic interoperability; Ontology; GIS; OWL

I. INTRODUCTION

With the development of the web service technology, a large amount of geographical information is currently being stored and delivered over internet, including various spatial databases, Geographic Information Services (GIServices) and webpages with spatially referenced data. In order to reuse spatial data to avoid duplicating the costs of data modeling, acquisition storage and maintenance, Interoperability is essential for many applications, such as spatial decision support systems that require the integration of spatial data and GIServices.

Most of the problems related to Interoperability are caused by the heterogeneities. Information systems heterogeneity may be considered as structural (schematic), semantic (data), and syntactic (database) heterogeneity (Bishr, 1998). Among them, Semantic heterogeneity of data sources caused serious problems. Since domain experts use the concepts and terminology specific to their respective field of expertise, i.e., their knowledge about the world, to translate the meaning of foreign set of concepts and terms to their own terminology. Software systems usually do not have any knowledge about the world and have to be told explicitly how to translate one term into another [1]. This paper is concerned with semantic web-based information retrieval of geographical information.

A focus of current research efforts is on providing accurate meta-information for both served data and served geo-data processing functionality commonly termed semantic interoperability (Lutz et al., 2003). The use of semantics based on ontologies as a source of meta-information is described by Rüther and Bandholtz (2005). Ontologies are also applied for searching and indexing tasks of the German Environmental Information Network.

II. RELATED WORK

A. *Semantic interoperability and ontology*

Semantic interoperability can only take place with agreements [2]. There are two levels of agreement required for semantic interoperability. The first is the agreement on what the real-world objects and phenomena are and on how they are interrelated and referred to (i.e. the terminology used in an application domain). The second is the agreement on the computer representations of real-world objects and phenomena and the computational operations that can be applied to them. For a full-fledged semantic interoperability among users and systems, semantic agreements must exist at both levels.

Ontology plays a key role in enabling semantic interoperability. Ontology is derived from philosophy and has gained a specific role in branches of computer science recently, such as knowledge representation, database design, information integration, object-oriented analysis, and agent-based system design. The most often cited definition describes an ontology as “an explicit specification of a conceptualization” (Gruber 2002). Ontologies in this modern sense are language-dependent and therefore often characterized as specifications of domain *vocabularies*. Such vocabularies can, for example, contain the terms used in the data and operations of web services.

Ontologies are classified in four groups, according to their dependency on a specific domain or point of view (Guarino, 1997): i) Top-level ontologies describe very general concepts; ii) Domain ontologies describe the vocabulary related to a generic domain; iii) Domain ontologies describe a domain or task; iv) Application ontologies are at the lowest level in inheritance view combines, integrates, and extends all sub-ontologies for the application.

The geographic ontology (geo-ontology) provides a model of the terminology and structure of the geographic space.

Recently, the use of ontology is discussed in GIS building (Devogele et al., 1998; Laurini, 1998), and creation of GIS software components from ontologies (Fonseca and Egenhofer, 1999). Two geographic ontologies CORINE LC (provide consistent localized geographical information on land cover of the member states of the European Community), MEGRIN's GDDD-Geographical Data Description Directory (contains information on available digital geographic information from Europe's National Mapping Agencies), and one lexical ontology－WordNet, which includes geographic categories, become well-known.

B. Ontology representation languages

Several studies have been devoted to ontology representation languages: they range from informal natural languages to formal languages based on predicate logic or graph concepts. Among the formal languages, we consider two such languages, namely, Geographic Mark up Language (GML) and Web ontology Language (OWL). GML is being established by the Open GIS Consortium (OGC) [3] as a standard language for encoding and sharing geographic information. OWL has been passed as a W3C recommendation for defining and instantiating web ontologies [4]. OWL is developed as a vocabulary extension of RDF (the Resource Description Framework).

Both languages have its pros and cons. GML is capable of representing geographic features using a rich built-in vocabulary, based on a well-defined geographic and spatial data model. As OWL is a general purpose language, no specific structures are pre-defined to represent geographic features and they are modeled using user-defined classes. In GML the semantic NT and BT relationships were used to form a hierarchy of objects by linking objects using XLinks. OWL uses the subclass-of axiom to represent the inheritance hierarchy of objects directly. The semantics in the OWL, are well understood, while tracing XLinks via pointers in GML can be obscure [5]. In this paper, we choose OWL as geo-ontology representation language.

We also mention OWL-S [6] as ontology of web services which makes discover, invoke, compose, and monitor web resources in a high degree of automation possible. It has three main parts: the service profile for advertising and discovering services; the process model, which gives a detailed description of a service's operation; and the grounding, which provides details on how to interoperate with a service, via messages.

III. ONTOLOGY-BASED FRAMEWORK FOR GEOGRAPHIC INFORMATION RETRIEVING

In almost all multi-tier component architectures for data retrieving, the general steps that a client usually follows to retrieve information in a distributed environment is to: a) use some search engine that accesses a metadata catalogue of the available resources and establish a link once the relevant information resource is identified; b) query the information resource for the information that is needed; c) retrieve the information from the resource and integrate it with the local database. If multiple information resources are available with no prior knowledge about its content, structure, and semantics, e.g. on the Internet, following the scenario is cumbersome.

A. Architecture

The framework supports query interpretation through ontologies as semantic agreements among users and systems. The architecture is shown in Fig.1 and consists of the following components: wrapper, core search engine and user interface.

The wrapper is used to map the local databases to a common data model and to provide low-level data access functions. Wrapping data access by interfaces has two advantages. First, the requirement that existing database schemes must not be affected is fulfilled, since no database changes are necessary. Second, data of the involved institutions don't want to expose exact locations but would like to restrict access to approximate areas such as bounding boxes depending on defined use and access policies.

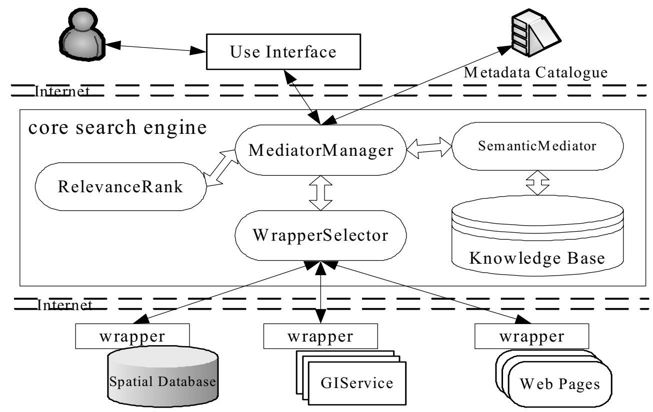

Figure 1. The Architecture of Ontology-based Geographic Information Search Engine

The user interface allows the user to query a subject of interest, a location and a spatial relationship to the location. The subject term or terms may be non-spatial or spatial. Once the semantic mediator (in core search engine) converts query to request using the common model and ontology, it will be echoed on a map or a relevance-ranked list on the user interface where the user is able to confirm the selected of interest. And finally, the results of query are also returned to the user on user interface.

The basic component providing interoperability, including the resolution of semantic heterogeneity, is core search engine. The main part of a search engine is MediatorManager. The basic component is to manage the work of other components. It receives the data produced by other mediators, initiates the search for the appropriate wrapper, receives the processed data from the wrapper and sends the data to corresponding mediator and initiates the system components which resolve semantic conflicts. WrapperSelector is the component of core search engine that provides communication with the wrapper. SemanticMediator participates in resolving semantic conflict. It communicates with MediatorManager and has access to the database. This component uses the Knowledge Base which contains data dictionary, thesaurus, semantic rules, expert knowledge, ontologies and metadata that specify the common model, generates rules on user demand, and store it into the local Knowledge Base. RelevanceRank component takes results retrieved from the search engine and relevance ranks them with respecting to the non-spatial and spatial elements of the query. Various techniques are being explored for combining textual and spatial relevance scores to produce an integrated score. It is also intended to introduce relevance ranking measures that take account of the parent geographical regions of the query.

B. Geographic ontologies for the approach

We concentrate on two kinds of ontologies for the geographic world in this paper. One is a top-level ontology called the Location Ontology which aims at giving the position of spatial-referenced data and GIServices. The other is an application ontology concerned with wrapping GIservices and is called Service Ontology.

1) Location Ontology: In most GIS applications, the assumption is that people may want to find information about something that relates to somewhere. Therefore, the location aspect is of utmost importance and the use of ontologies for this purpose has been long-praised. However, the top-level location ontology, shown as Figure 2, is not a complete geo-ontology, but pragmatically provides basic concept which may be refined in domain ontologies [7].

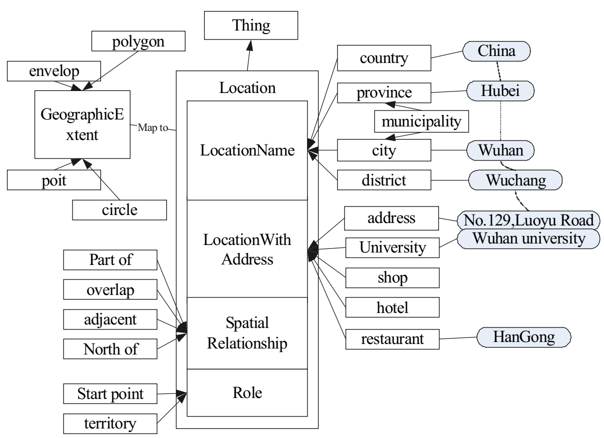

Figure 2. A Fragment of the Location Ontology

The most common way to refer a location is to use LocationName, which may be qualified by defining spatial relationships (such as part-of, overlap or adjacency). The location ontology are encodes geographical terminology and the semantic relationships between geographical terms.

LocationWithAddress is a particular way to represent LocationName. Based on a zip code and a building index or a corresponding GIService, coordinates can be derived for addresses. When objects are instantiated in the ontology, the containment relationship of a given location instance to these instances can be inferred from these coordinates or by the names used in the address.

After location ontology is instantiated by a LocationName and SpatialRelationship, it can be mapped to Geographic-Extents, for example, a hotel instance is mapped to a polygon instance having some coordinates, which enables the inference of containment. GeographicExtent consists of sub-concepts like Point, Circle, envelop, and polygon, described by using some coordinate system. In addition, different spatial relationships like containment and overlapping may be defined on these, implemented by axioms which may use the functions of GISevise.

Furthermore, we add a multi-valued Role property, which can be changed according to the domain context, to constitute a location ontology. The instance of a location ontology can be represented in different geographic extents when it plays different roles. For example, if a city plays the role of an origin of a flight course, it is represented as a point. But if the city plays the role of a territory, it may be represented as a polygon.

2) Service Ontology: The service ontology is based on OWL-S. The OWL-S ServiceProfile has a properties serviceName, which needs not be unique, so different services with the same name may be registered for different locations and textDescription. Moreover, it has properties serviceProduct (to classify services based on products they are concerned with), and also some of our extensions like subject (to be matched with user interests together with serviceProduct) and URL.

When looking at public UDDI business (or metadata catalogue) registries today, there are only a few services registered. Thus, our scenarios cannot be covered by using only web services, and we decided to also treat web content as a kind of service. Normal content does not have properties like input, precondition and effect, but most other properties from an extended OWL-S profile, like name, description, date, provider and coveredRegion. Inversely, content may be seen as a constant service (or better a service without parameter), since it needs not be constant (a Web page, for instance, may be dynamic). A content page which then is a client of a service may even have parameters, e.g., the zip code input field of a weather page, which can be submitted using the html post/get instructions. Of course, this is not a Web service technology, but may be seen as a service.

C. Semantic query process

1) Formulating and interpreting queries: The first step in the procedure is to specify the geographic extent of the query, using location ontologies to search in metadata catalogue, which is required in any query involving location. The next step is to decompose the query into sub-queries. Once a list of ordered sub-queries is ready, each sub-query is checked against the Semantic Mediator, using the domain ontology, to detect semantic conflict between the receiver and the provider. If there is conflict, the mediator will identify the type of semantic conflict and will solve it by reformulating the query request. Then, the mediator decides where the requested data are located, and resends the subqueries to adequate information sources. Next, Semantic Mediator will help to match the candidate operations for resolving the sub-query. In cases where two or more operations are matched to one sub-query, the Semantic Mediator is consulted for each sub-query to determine the types (e.g. POINT, LINE, POLYGON) representing instances under their respective roles. Once operations for sub-queries are determined, entities specified in the sub-queries are associated with input and output parameters of the respective operations. In the last step, each of the input parameters from all operations is validated. An input parameter to an operation is considered valid if it does not depend on an external dataset.

2) Semantic service matching: The service matching is based on OWL-S service profiles [6]. Service (and information) advertisements and demands are described independently by different user groups (i.e., data/service provider, data/service user), not knowing exactly the needs of each other. The most flexible way to match demand against offer is to use semantic technologies, i.e., ontologies and inference. The service selection process is a semantic matching of an implicitly constructed service request profile against the profiles of all known services found in the semantic registry. It evaluates different types of semantic relationships for all dimensions, such as subclass and instance relationships or even relationships defined by arbitrary axioms. User interface help to select only services matching the user's interests.

IV. Implementation

Our application ontology was implemented in OWL to be able to use complex rules for modeling geo-ontologies in the Ontoprise's OntoStudio and to profit from the good inference support of the OntoBroker 4.3 [8] inference engine.

As an example, consider the following query: Identify all restaurants which serve cuisine of type 'Far Eastern' around No.129, Luoyu Road, Wuhan, Hubei, China.

In the first step, The User sends the request to MediatorManager via the User Interface. The search engine analyzes the query request by SemanticMediator. For this, a mount of location, service and processing ontologies are predefined in OntoStudio, which is then stored in Knowledge Base. Using them, the request may be interpreted either directly or semantically. In a semantic way, the range and values of the attributes are semantic categories stemming from an ontology, which then forms background knowledge for interpreting the relational data and provides derived properties. Inversely, the ontology should then be used to generate value selection menus, to guide the process of creating or modifying the user's request.

For example, when using the Location Ontology show in Figure 2 to find the site of "No.129, Luoyu Road", we can generate following query:

```
……
Where GeographicExtant="point"
      and Onto_Related(LocationWithAddress,
          "is a or equal",
          "No.129, Luoyu Ave, Wuhan, Hubei, China.",
          location_ontology)=1
```

Similarily, web service applications can also utilize the ontology to match two different terms semantically. The service advertisements give the description of the restaurant guiding service which provides all kinds of the cuisine. And a request for service should be expressed in the same format of the advertisement. Using the cuisine ontology, the request looking for Far Eastern food may match the service which provide the Chinese, Korean and Japanese food.

In the second step, the Mediator decides which services are useful and where the request data are located, resends queries and returns matched results to user interface.

V. Conclusions

Interoperability of GIS will play an increasingly important role for GIS-supported environmental applications in future. Referring to different levels of interoperability, semantic interoperability represents an issue that causes few unsolved problems. Ontology seems to be an adequate methodology that helps to define a common ground between different information communities. In this paper, using ontology-based approach, GIS users would have an opportunity to access data via the Internet and could simultaneously check the ontology referring to that data. In this way, they could access the data knowing that they can agree on the meaning (the semantics). The increased need for semantic interoperability makes the ontology based approach valuable for GIS application in the distributed environment.

However, the geographic information interoperation based on ontology is not mature, so it is still a high mountain to climb to create wildly used geo-ontology. Our future work can mainly focus on the following ways, like find a more efficient way to define the geo-ontology, especially, establishing relatively mature OWL-S Process semantic model and mapping rules, so that geographic information service based on ontology can really work and finally should be integrated into the Spatial Data Infrastructures.

References

[1] L. Stoimenov, S. Djordjevic-Kajan, "An architecture for interoperable GIS use in local community environment," Computers & Geosciences, 31(2), 2005, pp.211-220.

[2] R. Peachavanish, H.A. Karimi, B. Akinci, "An ontological engineering approach for integrating CAD and GIS in support of infrastructure management," Advanced engineering informatics, 20(1), 2006, pp. 71-88.

[3] Open Geospatial Consortium, in http://www.opengeo-spatial.org/, 2007.

[4] S. Bechhofer, F. van Harmelen, J. Hendler, I. Horrocks, D.L. McGuinness, P.F. Patel-Schneider, L.A. Stein, "Owl web ontology language reference," in http://www.w3.org/TR/owl-ref/, 2007.

[5] A.I. Abdelmoty, P.D. Smart, C.B. Jones, "A Critical Evaluation of Ontology Languages for geographic information retrieval on the Internet," Journal of visual languages and computing, 16(4), 2005, pp. 331-358.

[6] OWL-S Coalition. "OWL-S: Semantic markup for web services," W3C Member Submission, in http://www.w3.org/Submission/OWL-S, Nov. 2004.

[7] N. Weissenberg, R. Gartmann, A. Voisard, "An Ontology-Based Approach to Personalized Situation-Aware Mobile Service Supply," Geoinfomatica, 10(1), 2006, pp. 55-90.

[8] Ontoprise GmbH. "User Guide-OntoBroker," Ontoprise Tutorial Series, in http://www.ontoprise.de/content/index-eng.html, 2007.