# THE SEMANTIC RETRIEVAL OF SPATIAL DATA SERVICE BASED ON ONTOLOGY IN SIG

Sheng-Tao Sun [a,b,]*, Ding-Sheng Liu [a], Guo-Qing Li [a], Wen-Yang Yu [a]

[a] Center of Earth Observation and Digital Earth, Chinese Academy of Sciences, Beijing, China 100094
[b] Graduate University, Chinese Academy of Sciences, Beijing, China 100080

**Commission IV, WG IV/1**

**KEY WORDS:** Spatial Information Grid, Spatial data service query, Ontology, Semantic heterogeneity

**ABSTRACT:**

The research of SIG (Spatial Information Grid) mainly solves the problem of how to connect different computing resources, so that users can use all the resources in the Grid transparently and seamlessly. In SIG, spatial data service is described in some kinds of specifications, which use different meta-information of each kind of services. This kind of standardization cannot resolve the problem of semantic heterogeneity, which may limit user to obtain the required resources. This paper tries to solve two kinds of semantic heterogeneities (name heterogeneity and structure heterogeneity) in spatial data service retrieval based on ontology, and also, based on the hierarchical subsumption relationship among concept in ontology, the query words can be extended and more resource can be matched and found for user. These applications of ontology in spatial data resource retrieval can help to improve the capability of keyword matching, and find more related resources.

## 1. INTRODUCTION

The purpose of Grid is to integrate and share all the resources (including data, computing equipments and software) in the network transparently and seamlessly. Within the earth observation field, a lot of efforts have been made to explore the application of grid for sharing of the spatial data and computing resources. Up to now, these efforts have achieved great success.

Spatial Information Grid (SIG) was developed by Center of Earth Observation & Digital Earth (CEODE) and it is a spatial information infrastructure, which has the ability to provide services on demands. It aims at sharing, integrating organizing, and collaborating enormous distributed spatial resources. And also it can connect, manage, access, and integrate various spatial data and computing resources to implement spatial information applications and services (K. T. He, 2005).

In SIG, spatial data service are described in some kinds of specifications, such as WSDL, WCS, WFS, WMS, WPS, etc., which use different meta-information of each kind of services. This kind of classification description method of spatial data service describes service in the view of resource usage and function, and can help to realize the standardization of resource description. However, this kind of standardization cannot resolve the problem of semantic heterogeneity, which may limit users to obtain the required resources. Then we should find a way to solve this heterogeneity of semantic problem.

Ontology in semantic web plays an important role in extracting and formalizing semantics. Ontology consists of logical axioms that convey the meaning of terms within a community. The logical axioms represent hierarchies of concepts and the relations among concepts. Explicit and formal definition of semantics of terms guided researchers to apply formal ontologies to semantic heterogeneity as a potential solution.

The main problems of semantic heterogeneity in spatial data service query mainly include four types:

(1) Name heterogeneity. The same entity and phenolmenon has different names in various application situations (one word with multiple equivalents of the same meaning). This kind of heterogeneity may limit users to use the exact name in certain situation and get all matching resources.

(2) Concept heterogeneity. The same concept and noun has different meanings in diverse contexts (one word with several equivalents of different meanings). This kind of heterogeneity may cause users to get many resources in different scenes but not all of them can satisfy users' requirement.

(3) Data type heterogeneity. The same property value of one entity can be described in several data types (such as string, integer, float and so on). This kind of heterogeneity may meet mismatch of property value in distinct data type.

(4) Structure heterogeneity. Different resources of the same category can be described under diverse meta-information structures and description schemes (different numbers and meanings of description fields). This kind of heterogeneity may cause resources mismatch under different description schemes.

These semantic heterogeneities above can't be solved through using traditional lexical analysis and string match, which make

---

* Corresponding author. stsun@ceode.ac.cn. No.9 Dengzhuang South Road, Haidian District, Beijing, China 100094

many challenges in the improvement of spatial data service query.

The main purpose of our work is to try to solve these semantic heterogeneities based on ontology. This paper aims at the study and solution of name heterogeneity and structure heterogeneity based on ontology and improves the search results of spatial data service in SIG. The main contributions of this paper include: (1) Analyzes the name character of concept and spatial entity, and presents Class Meta-Information (CMI) description method in OWL to solve name heterogeneity. (2) Studies the current main spatial data service description specifications and proposes Meta-Info Mapping (MIM) method in spatial data service query to realize the semantic match between retrieval schema and service description specifications. (3) Based on the vocabulary from ontology, the hierarchical subsumption relationship among concepts could be used to extend the query words and help to find more related resources.

## 2. RELATED WORKS

Ontology has been proposed to play a central role in driving all aspects and components of an information system, leading to ontology-driven information systems (N. Guarino, 1998), and in the specific case of GIS, leads to what we call Ontology-Driven Geographic Information Systems (ODGIS). Frederico (1999) introduced a geographic information system architecture based on ontologies and used object-oriented mapping of ontologies, which could provide a great level of interoperability and allows partial integration of information when completeness is impossible.

Max (2002) introduced a new framework of Semantic Geospatial Web, and pointed out that the creation the Semantic Geospatial Web needed the development multiple spatial and terminological ontologies and the processing of geospatial queries against these ontologies. The Semantic Geospatial Web will enable users to retrieve more precisely the data they need, based on the semantics associated with these data.

According to the quick development and application of Web Service, W3C (2004) set a standard OWL-S for Web Service semantic description. OWL-S is an ontology, within the OWL-based framework of the Semantic Web, for describing Semantic Web Services. It will enable users and software agents to automatically discover, invoke, compose, and monitor Web resources offering services, under specified constraints.

Yang An (2004) proposed a service mode of web geography service based on ontology in OWL-S and gave the methods of web service description, discovery and composition. And also, Qiu Tian (2009) presented a matching algorithm for service discovery based on semantic similarity of concepts in OWL-S.

Xiaofeng Zheng (2005) proposed an approach to build up a semantic description and representation for business and services on top of the UDDI (Universal Description Discovery Integration) and WSDL (Web Service Description Language) based service registry. This approach designed a semantic based search engines for Web Service registration and discovery, and provided an enabling solution to make semantic matching of user's queries on the Web Services.

Patrick (2008) presented an extensibility and semantic enablement architecture for web service catalogues, which took the diversity of various standards into account, and used ontology to support different description standards without loosing their specific advantages.

In the semantic application of spatial data service based on Grid, Geren Li (2004) defined a concept of semantic grid and advanced the semantic grid architecture of spatial information systems (SGASIS). In this semantic grid architecture, the ontology transform bridge could transform ontology concept between local ontology and general ontology so as to encapsulate the local GIS and domain application and ensure that all operations are based on semantic.

Lorenzino (2009) discussed an approach of how to semantically coordinate geographic services, which is based on a view of the semantics of web service coordination, implemented by using the Lightweight Coordination Calculus (LCC) language. In this approach (structure preserving semantic matching), service providers share explicit knowledge of the interactions in which their services are engaged and these models of interaction are used operationally as the anchor for describing the semantics of the interaction.

These research works above mainly focused on two aspects: service meta-info semantic description and service discovery semantic process. Moreover, the semantic description is the basis of semantic retrieval, but these works didn't pay much attention to the semantic analysis and match of different service description specifications with retrieval schema, which can integrate various service description specification (such as WSDL, WCS, WFS, WPS, etc.) into a general service semantic description schema and provide uniform service query view.

## 3. THE SOLUTION OF NAME HETEROGENEITY BASED ON ONTOLOGY

Ontology can provide the domain lexical knowledge base of concepts and terms. The relationship between concept and term can be used to get all similarity terms about the keyword, which can help to match more related resources.

Traditional concept relationship description in OWL confuses concept with term, and consider term as concept, which may cause it difficult to solve name heterogeneity of concept and instance. In OWL, concept ought to be the basic unit of ontology as Class (A Class in OWL defines a group of individuals that belong together because they share some properties), while term should rely on concept as semantic Property (Property can be used to state relationships between individuals or from individuals to data value).

To solve this problem, we present Class Meta-Information (CMI) method to describe the relationship between concept and term, which can make it easy to find related terms of the same concept. This method uses the instance with special name (ClassName_0, which can distinguish from other instances) to store complete meta-information of the belonging concept. The basic information stored in class meta-information includes: the corresponding Chinese terms of concept (ChineseNames), the corresponding English terms of concept (EnglishNames), the hierarchical position identification of "Class" (HID), the corresponding names of concept in other knowledge systems (OtherNames).

For example, the class meta-information of concept "Biosphere" includes: ChineseNames (生物圈,生命层,生物界,生物) and EnglishNames (biosphere, living, living thing). And also, individuals of concept can use this CMI to store different names of certain instance, such as instance "Beijing" of concept "City" has ChineseNames (北京,首都,京) and EnglishNames (Beijing, Peking, BJ).

The corresponding terms under different circumstances could present in CMI of one concept, and the same terms in different CMI of concepts could present the semantic relationship between these concepts. This kind of relation can well present the semantic relationship between concept and term, which can enhance the semantic description capability of OWL and solve name heterogeneity of concept and instance.

## 4. THE SOLUTION OF STRUCTURE HETEROGENEITY BASED ON ONTOLOGY

In the field of spatial data service, there exist several kinds of data resources mainly include: data resource based on OWS (OGC Web Service, such as WCS, WFS, WMS, and WPS), and data service based on basic Web Service. Web service uses WSDL file to describe service meta-info, and OWS framework uses OGC (Open Geospatial Consortium) Capability XML (eXtensible Markup Language) file to describe service meta-info. Both of them adopt XML format, but they have different meta-info structures, which may limit the interoperation and cross search among these services.

When users retrieval these services through query interface, they often face uniform resource search view and the main search conditions include: service name, service type, fee, provider, linkage, etc. Therefore, we need to build up the mapping between this query schema and various service description standards, and solve the structure heterogeneity among different meta-info specifications.

First, we describe these kinds of service in ontology, shown as Figure 1 and Figure 2, which represent the meta-info structures as hierarchical semantic properties.
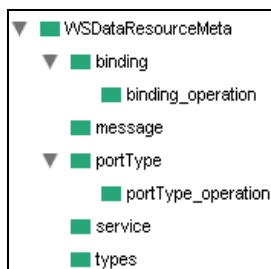


Figure 1. The hierarchical semantic properties representation of WSDL

Then we make reasoning rules for the mapping and transform between these hierarchical semantic properties and general query schema of resource search view shown as Figure 3.

We can get meta-info directly from service description metafile, but different Meta fields map to diverse query schema fields.

That is to say, the query condition needs to be matched with one or multiply Meta fields of different services. These reasoning rules can setup the transform among these fields, and realize service meta-info automatically analysis and extraction from various service specification to general query schema.
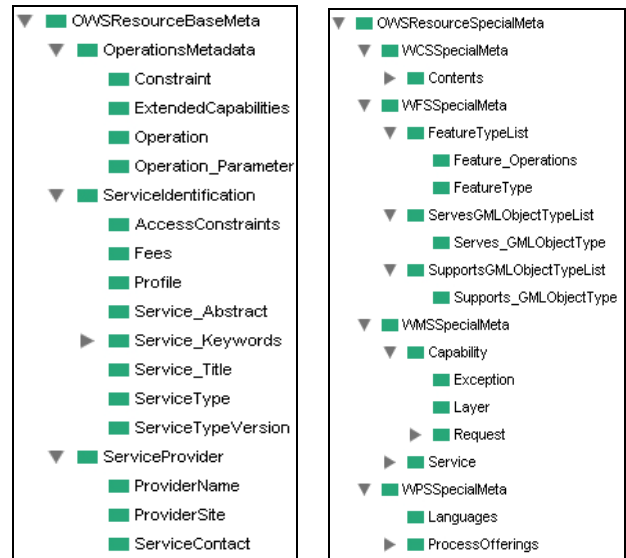


Figure 2. The hierarchical semantic properties representations of OWS (include WCS, WFS, WMS, and WPS)
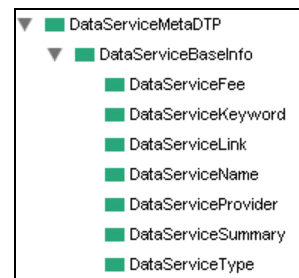


Figure 3. The hierarchical semantic properties representation of general query schema
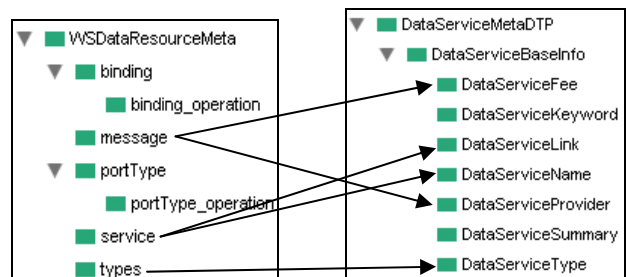


Figure 4. The mapping relations of semantic properties between WSDL and general query schema

The mapping relations of semantic properties between WSDL and general query schema can be shown as Figure 4.

The corresponding reasoning rules are partly shown as below:

[WSname: (?dr ds:service ?sev), regex (?sev,'<(.*)name="(.*)"><wsdl:(.*)', ?tmp, ?name) -> (?ds ds:DataServiceName ?name)]

[WSLink: (?dr ds:service ?sev), regex (?sev,'(.*)location="(.*)"/>(.*)',?tmp,?link) -> (?ds ds:DataServiceLink ?link)]

The mapping relations of semantic properties between WCS and general query schema can be shown as Figrue 5.
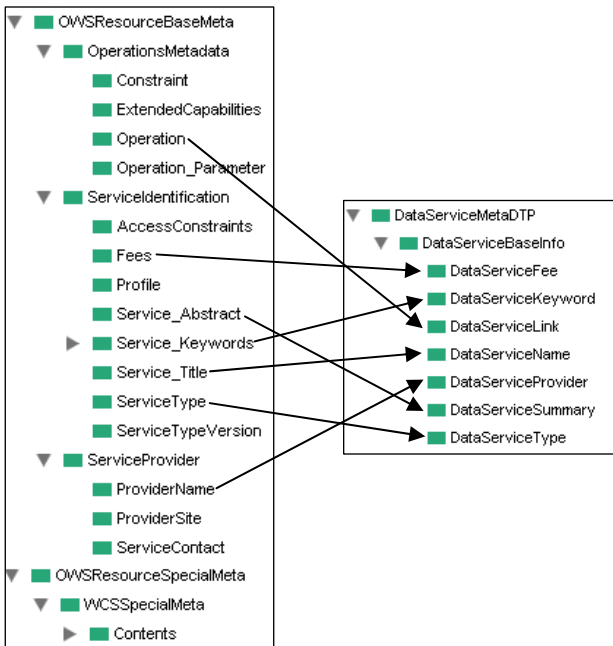


Figure 5. The mapping relations of semantic properties between WCS and general query schema

The corresponding reasoning rules are partly shown as below:

[OWSname: (?dr ds:Service_Title ?title), regex (?title,'<(.*)>(.*)</(.*)>',?tmp,?name) -> (?ds ds:DataServiceName ?name)]

[OWStype: (?dr ds:ServiceType ?st), regex (?st, '<(.*)>(.*)</(.*)>', ?tmp, ?type) -> (?ds ds:DataServiceType ?type)]

[OWSProvName: (?dr ds:ProviderName ?pn), regex(?pn,'<(.*)>(.*)</(.*)>',?tmp,?name) -> (?ds ds:DataServiceProvider ?name)]

[OWSLink: (?dr ds:Operation ?op), regex (?op,'(.*)xlink:href="(.*)"/>(.*)',?tmp,?link) -> (?ds ds:DataServiceLink ?link)]

[OWSfee1: (?dr ds:Fees ?fees), regex (?fees,'<(.*)>(.*)</(.*)>',?tmp,?fee), equal (?fee,'NONE') -> (?ds ds:DataServiceFee 'Free')]

[OWSfee2: (?s1 ds:Fees ?fees), regex (?fees,'<(.*)>(.*)</(.*)>',?tmp,?fee), notEqual (?fee,'NONE') -> (?ds ds:DataServiceFee ?fee)]

The meta-info for service search can be obtained from WSDL file and Capability XML file automatically based on relevant rules, and these rules could provide mapping bridge between two meta-info structures, which can initiatively solve structure heterogeneity of different description specifications. Moreover, these rules are formal representation of domain knowledge and have good maintainability and expansibility.

## 5. THE EXTENSION OF QUERY WORDS BASED ON ONTOLOGY

Traditional method of spatial data service search is string matching between query condition and resource description meta-info, which is grammatical level match and lacks the full comprehension of query keywords. In semantic level match, ontology can provide complete description of domain vocabulary, such as relationship of equivalence, similarity, subsumption, and other semantic relevancies. These semantic relations among concepts and terms can be used to get deeper understanding of query conditions, and help to find more relevant resources.

This section uses the relationship of subsumption among concepts to extend the query words. The subsumption relation of concepts in ontology can be shown as Figure 6.
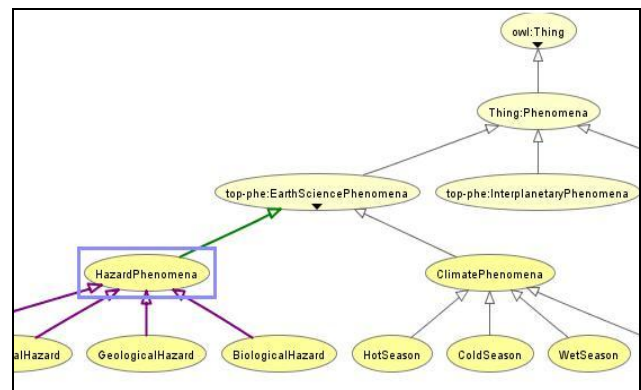


Figure 6. The hierarchical structure of subsumption relation among concepts in ontology

The concepts in ontology are organized as tree structure based on the subsumption relationship. The upper concept is more widely used and abstract, the lower concept is more application-oriented and specific. This tree structure represents subordinate relationship, category, and hierarchical relevance among concepts.

Along this tree structure upwards, we can get more abstract and common concepts and terms, which can help to extend the breath of word search and improve query recall. On the contrary, along this tree structure downwards, we can get more specific and exclusive concepts and terms, which can help to search in certain situation or sub domain for special purpose

and improve query precision. Therefore, the subsumption of concepts in ontology can be useful and helpful to realize efficient, accurate, and exhaustive resource search.

First, we retrieve the keywords in ontology and find the relevant concepts. Then we get the set of parent concepts and set of child concepts based on subsumption relationship in ontology. From CMI mentioned above, we can get all the English terms and Chinese terms of these concepts, and regard them as search keywords. The extension of keywords has more semantic coverage of words and can help to find more related resources which can satisfy user's query requirement.

For example, we set the initial keyword which includes only one word "灾害" in Chinese, and after this semantic words extension, we can get the spatial data service search results shown as Figure 7.

| 检索的关键词为: 灾害 | | | |
|---|---|---|---|
| 关键词对应概念进行匹配结果 | | | |
| 匹配数据服务 | 匹配属性 | 匹配关键词 | 匹配概念 |
| 无 | | | |
| 子概念进行匹配结果 | | | |
| 匹配数据服务 | 匹配属性 | 匹配关键词 | 匹配概念 |
| DS_JDBC_CNIC-DEM | DataServiceKeyword | geological hazard | GeologicalHazard |
| DS_WFS_CEODE-WFS | DataServiceKeyword | 水灾 | HydrologicalHazard |
| DS_WCS_CEODE-MODIS | DataServiceKeyword | flood | HydrologicalHazard |
| 父概念进行检索结果 | | | |
| 匹配数据服务 | 匹配属性 | 匹配关键词 | 匹配概念 |
| DS_WS_CNIC-LANDSAT | DataServiceKeyword | 全球变化 | EarthSciencePhenomena |

Figure 7. The search results of spatial data service

From Figure 7, we can obtain some distinction results of data service query. The first table is the search results of exact matching keyword "灾害" and none service is matched. The second table is the results of matching keywords from sub concepts of "灾害" (such as Fire Hazard, Hydrological Hazard, Geological Hazard, Biological Hazard, Flood, Earthquake, etc. in English, and 火灾, 水灾, 地震, 洪水, etc. in Chinese ) and three data services are matched in different situations of meta-info field and keywords. The third table is the results of matching keywords from concepts of "灾害" (such as Hazard Phenomena, Global Change, etc. in English, and 灾害现象, 全球变化, 全球性灾害, etc. in Chinese) and one more data service is matched.

From this simple example we can see that the semantic extension of query keywords can find more resources for users, and improve the query results.

## 6. CONCLUSION

The purpose of Grid is to integrate and share all the resources in the network. In SIG, spatial data service are described in some kinds of specifications, such as WSDL, WCS, WFS, WMS, WPS, etc., which use different meta-information of each kind of services. This kind of classification description method of spatial data service describes service in the view of resource usage and function, and can help to realize the standardization of resource description. However, this kind of standardization cannot resolve the problem of semantic heterogeneity, such as name heterogeneity, concept heterogeneity, data type heterogeneity, and structure heterogeneity.

Ontology in semantic web plays an important role to extract and formalize semantics. Explicit and formal definition of semantics of terms can guide researchers to apply formal ontologies as a potential solution to semantic heterogeneity.

We apply ontology in SIG to solve some semantic heterogeneity problems of spatial data resource retrieval, and improve the query result.

(1) The name heterogeneity of query keyword. Ontology can provide the domain lexical knowledge base of concepts and terms. The relationship between concept and term can be used to get all similar terms about the keywords, which can help to match more related resources.

(2) The structural heterogeneity of resource description. Ontology can supply the semantic description of spatial data service, which uses the common description structure of different kinds of services, and gives the mapping between semantic property and meta-information. These mappings can guide the query process to orient several related meta-information which exactly meets the purpose of query, and avoid the retrieval of all meta-information to find the keyword, which can help to find more correct resources.

(3) The term extension of query keyword. Ontology can describe the hierarchical subsumption relationships of concepts and terms. Upward this hierarchical relation, more abstract concepts and terms can be obtained and help to extend the range of query. Downward this hierarchical relation, more specific concepts and terms can be obtained and help to improve the precision of query.

These applications of ontology in spatial data resource retrieval can help to improve the capability of keyword matching, and find more related resources.

## REFERENCES

AN Yang, BIAN Fuling, GUAN Jihong, 2004. Geo Web Services Description and Discovery Based on Ontology. GEOMATICS AND INFORMATION SCIENCE OF WUHAN UNIVERSITY, 29(12), pp.1063-1066.

David Martin, Mark Burstein, Jerry Hobbs, 2004. OWL-S: Semantic Markup for Web Services. http://www.w3.org/ Submission/OWL-S/.

Deren Li, Wei Cui, 2004. Semantic Grid for Spatial Information. GEOMATICS AND INFORMATION SCIENCE OF WUHAN UNIVERSITY, 29(10), pp.843-851.

Frederico T. Fonseca, Max J. Egenhofer, 1999. Ontology-Driven Geographic Informaiton Systems. 7the ACM Symposium on Advances in Geogrpahic Information Systems, Kansas City, MO, pp.1-7.

K. T. He, Y. Tang, W. X. Yu, 2005. Spatial Information Grid and its Application in Geological Survey. International Society of Photogrammetry and Remote Sensing (ISPRS), Workshop on Service and Application of Spatial Data Infrastructure, pp. 283-288.

Lorenzino Vaccari, Pavel Shvaiko, Maurizio Marchese, 2009. A geo-service semantic integration in Spatial Data Infrastructures. International Journal of Spatial Data Infrastructures Research, Vol.4, pp.24-51.

Max J. Egenhofer, 2002. Toward the Semantic Geospatial Web. Proceedings of the 10th ACM international symposium on Advances in geographic information systems, New York, USA, pp.1-4.

N. Guarino, 1998. Formal Ontology and Information Systems, in: Formal Ontology in Information Systems, ed. N. Guarino, IOS Press, Amsterdam, Netherlands, pp. 3-15.

Patrick Maue, 2008. An extensible semantic catalogue for geospatial web services. International Journal of Spatial Data Infrastructures Research, Vol.3, pp.168-191.

Qiu Tian, Li Peng-fei, Lin Pin, 2009. A Web Service Matching Algorithm Based on Semantic Similarity of Concepts. ACTA ELECTRONICA SINICA, 37(2), pp.429-432.

Xiao-feng Zheng, Jing-chang Gao, Zhong-shu Piao, 2005. Semantic Matchmaker Based on Web Serive. Journal of Jilin University (Information Science Edition), 23(5), pp.552-558.