Computers & Geosciences 35 (2009) 783–790

Contents lists available at ScienceDirect

Computers & Geosciences

journal homepage: www.elsevier.com/locate/cageo

A taxonomy of geospatial services for global service discovery and interoperability

Yuqi Bai, Liping Di *, Yaxing Wei

Center for Spatial Information Science and Systems, George Mason University, 6301 Ivy Lane, Suite 620 Greenbelt, MD 20770, USA

ARTICLE INFO

Article history:
Received 7 May 2007
Received in revised form
1 December 2007
Accepted 31 December 2007

Keywords:
Geospatial web service
Classification
Classification scheme
Classification node
Taxonomy
Knowledge representation

ABSTRACT

Geospatial service taxonomies represent the knowledge about the characteristics of geospatial services from the enterprise, computational, information, engineering, infrastructure, or technology viewpoints. This paper presents a lightweight taxonomy of geospatial services with the aim of promoting the global sharing of and interoperability among geospatial service instances. This taxonomy focuses on the knowledge connected with service interoperability. As a hierarchical taxonomy, it consists of six layers: service category, service type, version, profile, binding and uniform resource name (URN), from the root down to the leaves. Each layer is composed of classification nodes, with each node identifying one classification concept. Each concept, with a concrete semantic meaning, can be used to classify service instances. The application of this classification scheme to the Global Earth Observation System of Systems (GEOSS) Component and Service registry is also introduced. The results of this study may lead to the further development of service taxonomy to thoroughly capture the knowledge about geospatial services. The lessons learned may be useful to others representing and manipulating geoscientific knowledge.

1. Introduction

As the Word Wide Web becomes the dominant computing platform, geospatial services, as software programs, develop rapidly. The Open Geospatial Consortium (OGC) has been leading the development of standards for geospatial and location-based services in the last 14 years. So far 378 compliant or implementing software products and/or service instances have been registered with the OGC.[1] They cover a broad range of service types: Catalog/Registry services, Data Access services, Portrayal and Display services, Data Transformation services, and Location-based services.

The increasing availability of geospatial services brings new challenges in the areas of service discovery and interoperability. Besides the descriptions of service characteristics, interface characteristics, and association with data sets/content for each service instance, there should be a classification of geospatial service. Service providers and consumers can reach a mutual understanding of what service instances can do and how they work, if they reference the same classification scheme.

From the geoscientific knowledge perspective, it is also very desirable to capture and represent the classification of geospatial services. A geospatial service classification scheme/taxonomy enables annotation, discovery, integration, and use of geospatial services within the cyberinfrastructure for the Earth and Space geosciences. As a matter of fact, "classification" is a well-known knowledge representation problem, along with "spreading activation" and "subsumption."[2] The linkable between classification and knowledge representation is that the process of

* Corresponding author. Tel.: +1 301 982 0795; fax: +1 301 345 5492.
E-mail addresses: ldi@gmu.edu, ybai1@gmu.edu (L. Di).

[1] All Registered OGC products. http://www.opengeospatial.org/resource/products.

[2] Knowledge representation—Wikipedia. http://en.wikipedia.org/wiki/Knowledge_representation.

0098-3004/$ - see front matter © 2008 Elsevier Ltd. All rights reserved.
doi:10.1016/j.cageo.2007.12.018

classification is always based on predefined classification schemes/taxonomies, and the later have properties that enable the representation of entities and relationships in structures that reflect knowledge of the domain being classified (Kwasnik, 1999).

In the last decade, the construction, management, and application of taxonomy have been applied in many research fields: enterprise information management (Abrol et al., 2005), Web search (Pahlevi and Kitagawa, 2002), software design (Kienle and Muller, 2007; Kovacic, 2005), software agents (Kerschberg et al., 2001; Huang et al., 2000), Web security (Bazaz et al., 2006; Abbas et al., 2006), fault analysis (Hayes, 2003), and data grid (Srikumar et al., 2006). In the geospatial science field, Grimshaw proposed a taxonomy for geographical information systems based on the general taxonomy of information systems (Grimshaw, 1996). It uses a three-dimensional framework, decision–technology–strategy, to provide a classification system that reflects both the static and dynamic characteristics of geographical information systems. Compared with geographical information systems, geospatial services usually have concrete functionalities, specific communication protocol definitions, and expected usage scenarios. These distinct differences have prevented this taxonomy from being properly utilized in classifying geospatial services.

In the International Organization for Standardization (ISO) Geographic information-Services (ISO, 2005) standard, a three-level geographic services taxonomy is proposed, where six classes of services have been defined: *Geographic human interaction services*, *Geographic model/information management services*, *Geographic workflow/task management services*, *Geographic processing services*, *Geographic communication services*, *Geographic system management services*. For each of these classes, subclasses, and sub-subclasses are numerated. This taxonomy is of value in describing under which categories service instances could be classified. However, these classes tend to be more general. They cannot be used to describe detailed interoperability-related services characteristics. Consider, for example, the geospatial catalog service. In this taxonomy, it is simply identified as "*Geographic model/information management services—Catalogue service*". But the fact is that geospatial catalog service may strongly affect the standards profiles followed, the underlying information models maintained, and the communication binding protocols supported. Without this information, promoting service discovery and interoperability would be very time-consuming. Users have to manually check each geospatial catalog service instance for these characteristics. Another issue would be service version management. Many differences may exist between different versions of the same type of service. Some parameters may be deprecated, while new mandatory ones have been proposed. This is a big issue when considering service interoperability, but it is not covered in this ISO taxonomy.

US National Aeronautics and Space Administration (NASA)'s Global Change Master Directory (GCMD) project enables users to locate and obtain access to Earth science data sets and services relevant to the global change and Earth science research. GCMD system maintains a four level hierarchy of science keywords[3]: Category-Topic-Term-Variable. For its services keyword, GCMD proposes one category, "*EARTH SCIENCE SERVICES*", and seven topics, "*Data Analysis And Visualization*", "*Data Management/data Handling*", "*Education/outreach*", "*Environmental Advisories*", "*Hazards Management*", "*Metadata Handling*", "*Models*", and "*Reference And Information Services*". By enabling the same keywords being used to describe distinct service instances, GCMD provides an uniform annotation framework for geospatial services. But similar with the taxonomy defined in the ISO 19119:2003 standard, GCMD's keywords are too general to be referenced in service instances level to meet the requirements of service interoperability.

In the OpenGIS Web services Architecture report (Lieberman, 2003), OWS service taxonomy is briefly mentioned to group services that are semantically similar in familiar categories, so as to facilitate browsing and discovery according to already understood (human) or pre-programmed (machine) functionality. Actually this extensible classification scheme is extracted from ISO 19119, Clause 7. In the OpenGIS Web services Architecture Description report (Whiteside, 2005), geospatial services are classified into three tiers: from the top to the bottom, *Application services*, *Processing services*, and *Information Management services*, with service in high tier functionally uses the services in lower tiers. Each tier of services includes multiple specific types of services, many of which are tailored to geographic data and services. Some of the specific services included in each tier are discussed. For example: *Geographic data discovery services* is enumerated as one type of *Application service*. These OGC efforts are a good start. They need to be enhanced to represent necessary knowledge around geospatial services for the purpose of service discovery and interoperability.

To the authors' knowledge, no geospatial service taxonomy that is suitable for global service discovery and interoperability has been described in the literature. The purpose of this paper is to present through a case study how taxonomy of geospatial services could be organized to capture and represent the services' discovery and interoperability characteristics and how it could be used.

The rest of this paper is organized as follows. The case study scenario is introduced followed by an analysis of the service characteristics that are related to Service Discovery and Interoperability. The next section focuses on the design and implementation in detail. The advantages and limitations of the proposed taxonomy are discussed leading to conclusions.

2. Scenario

The Global Earth Observation System of Systems (GEOSS) is designed to consider, analyze, and integrate

[3] NASA/Global Change Master Directory (GCMD) Earth Science Keywords, Version 5.3.8. http://gcmd.nasa.gov/Resources/valids/keyword_list.html.

isolated earth observation systems that have been maintained by involved nations. In 2007, the GEOSS architecture task AR-07-01 initialized the Interoperability Process Pilot Project, where *Component and Service Registry*, *Standards and Interoperability Registry, Clearinghouse* and *Web Portal* are designed and prototyped to promote the discovery of geospatial resources and the interoperability among diverse geospatial services.

As shown in Fig. 1, geospatial services are first registered in the Component and Service Registry by GEOSS Contributors, then discovered through Clearinghouse, and finally invoked by GEO Portal. In this distributed computing environment, it is very desirable to have service providers declare enough information regarding the service behaviors when they register their service offerings. This information can be used by service consumers to facilitate the discovery of available services, to perform quick evaluation of the fitness of particular service for their specific decision or assessment, and to promote interoperability among available services through dynamic integration.

Taking into account that service behaviors are further classified by GEOSS as two types: Standard Arrangements and Special Arrangements. Standard Arrangements refer those services that are compliant with public service standards, while Special Arrangements are for those services that are not fully compatible with public service standards, or even no public service standards available to reference. The significant issue is to define a lightweight service taxonomy to capture knowledge around services characteristics, so that geospatial services can be classified according to their service category, particularly what standards are followed.

3. Service characteristics related to Service Discovery and interoperability

As the taxonomy has those properties that enable the representation of entities and relationships in structures that reflect knowledge of the domain being classified, evaluating those service characteristics related to service discovery and interoperability is a good start to reach a well-defined service taxonomy.

The characteristics of geospatial services, as open distributed processing systems, could be understood from the enterprise, computational, information, engineering, infrastructure, and technology viewpoints, respectively, as proposed in the Reference Model for Open Distributed Processing specification (Farooqui et al., 1995). However, from the point of view of service discovery and interoperability, how the service is implemented is less important then what kind of interfaces it supports. That is also the reason why OGC's service specifications are always for the interfaces, rather than the underlying implementations.

The following service interface-centered characteristics are related to service discovery and interoperability.

3.1. Service category

A service category describes the functionality of a geospatial service. Each proposed category covers one role of a geospatial service in the information processing sequence of data–information–knowledge. Each category

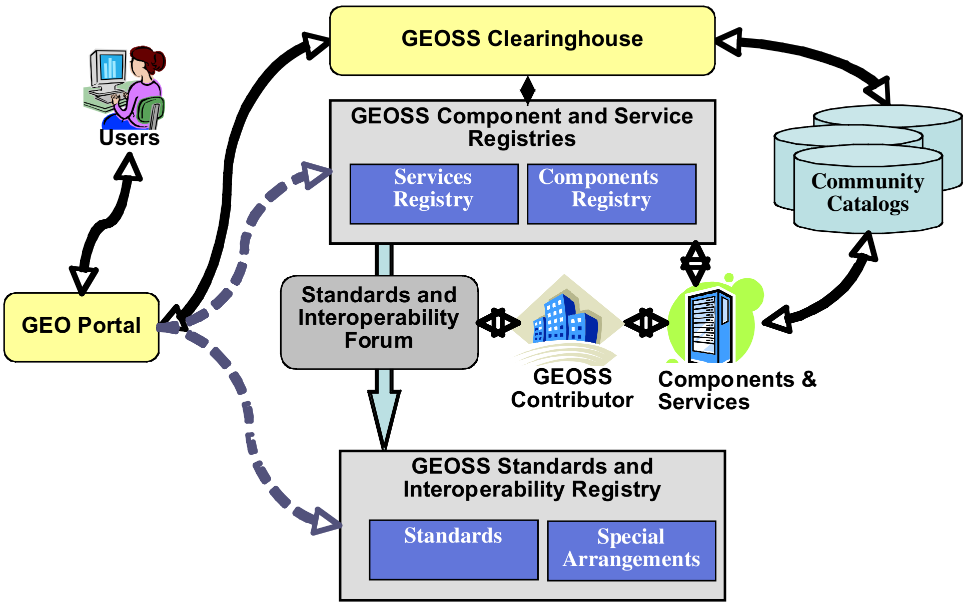

Fig. 1. System architecture of GEOSS interoperability process pilot.

may further have several service types. Some service categories are

- Catalog/Registry services—provide access to a catalog or inventory of data, services, and other community resources,
- Data Access services—provide access to data produced by observations, models, etc., as stored in an archive,
- Portrayal and Display services—provide access to data produced by observations, models, etc., as visual images,
- Data Transformation services—service that applies rules to generate a map view of geospatial data.

3.2. Service interface standard

Each of these standards describes how a service fulfills its functionality through identifying the interfaces it should support. Each interface consists of a named set of operations that characterize the service behaviors. Each operation further specifies a transformation or query that a service may be called to execute, and what the response would be. Examples of service interface standards are

- OGC Catalogue service,
- OpenDAP service,
- OGC Web Feature service,
- OGC Web Coverage service,
- OGC Web Map service.

3.3. Standard version

Published specifications usually have several revisions in their lifecycle. All the standards defining bodies, e.g. the International Standard Organization (ISO), the International Electrotechnical Commission (IEC), the International Telecommunication Union (ITU), and the OpenGIS Consortium (OGC), define their own process for creating each distinct version. As the availability, cardinality, and optionality of each service interface, operation, or parameter could be changed in each revision, the version number is a very important characteristic when referencing public interface standards.

3.4. Service binding

A service binding is a concrete protocol and data format specification. It applies to all the operations that service supports. The reason it is defined in the service interface standard is that it enables a service consumer to learn the run-time behavior of the service at design time. This is a key point for promoting service interoperability, as no one prefers to access a service if its run-time behavior is unpredictable. Because the service binding defines concrete protocols and data formats for communication, it is another important characteristic of geospatial services from the Service Discovery and interoperability point of view. Some service binding examples are Simple Object Access Protocol (SOAP), Hypertext Transfer Protocol (HTTP) GET, and HTTP POST.

3.5. Standards profile

A profile consists of an agreed-upon subset and interpretation of a specification. Profiles are usually developed to meet the needs of particular information communities.

4. Proposed geospatial service taxonomy

4.1. Taxonomy structure

The structure of taxonomies varies from parent–child trees to relationship schemes, from simple groups to an alphabetical list. A tree structure of classifications is called a containment hierarchy. At the top of this structure is a single classification, the root node, which applies to all objects. Nodes below this root are more specific classifications that apply to subsets of the total set of classified objects.[4]

As the aforementioned service characteristics naturally match the tree pattern, we defined this geospatial service taxonomy following parent–child structure. Fig. 2 shows the logical multi-layer structure of this taxonomy.

From the top to the bottom, the taxonomy consists of service category, services standard, standard version, service binding, service profile, and Uniform Resource Name (URN) layers. The first five layers are introduced to capture the aforementioned service characteristics related to service discovery and interoperability. Each of these five layers consists of a group of classification nodes in the taxonomy. Nodes in the upper layers and those in the lower layers form a parent–child relationship. In particular, a node in the upper layer may have more than one child and a node in the lower layer always has one and only one parent.

Usually in the parent–child taxonomy, there are two ways to identify each taxonomy concept: by code or by path. “By code” means each taxonomy concept will have an unique or Non-unique code that can be directly referenced outside of the taxonomy. In the “by path” scenario, every taxonomy concept must be referenced by the path beginning from the top concept. In this taxonomy, an URN layer is introduced to hybrid these two scenarios. This layer consists of many identification nodes. Each identifies one classification node appearing in the other five layers. Concepts that are identified by URN nodes can be referenced by code; other concepts can only be referenced by path.

Fig. 3 illustrates parts of the proposed taxonomy content. Four service categories are listed here: data access service, catalog/registry service, portrayal and display service, and data transformation service. For the catalog/registry service category, the OGC catalogue service is listed as an example of a service standard. This classification node, as a parent, has two child nodes: 2.0.1 and 2.0.2. This describes the fact that this service standard has two recognized service versions, with one version number 2.0.1 and the other 2.0.2. Each version has three

[4] Taxonomy–Wikipedia. http://en.wikipedia.org/wiki/Taxonomy.

possible bindings: CORBA, HTTP, and Z39.50. In service profile layer, four available profiles for HTTP binding and three for Z39.50 binding are listed accordingly. Each classification node, no matter in which taxonomy layer it resides, may have an affiliate service URN identification node. Three identification nodes are listed for demonstration purposes.

4.2. Classifying services

Service instances can be classified by referring to one or more of the predefined concepts. How this classification information could be described is an another issue for taxonomy application. The electronic business Registry Information Model (ebRIM) specification defines classification Scheme, Classification Node, and Classification classes to describe the classification of registry objects. The following mechanism is used, following this idea, to formally classify service instances using the taxonomy concepts.

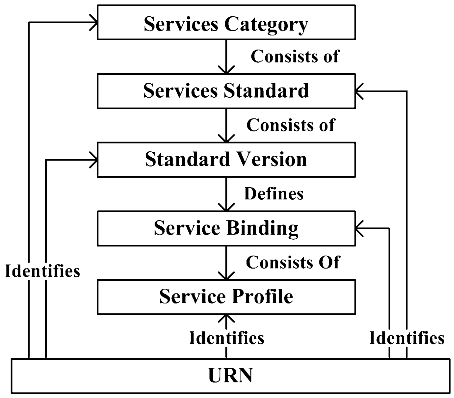

Fig. 2. Multi-layer logical structure of proposed taxonomy.

As shown in Fig. 4, the Classification class is imported from the ebRIM information model. This class is associated with the Geospatial service class and the Taxonomy Concept class. One or more Classification instances classify a Geospatial service instance by referencing defined Taxonomy concepts. The referenced Taxonomy concept could be either Classification Node, or Identification Node. The classification, along with the service instance, is maintained in the service registry. By referencing the same taxonomy, the service instances can be classified uniformly.

4.3. Application of proposed taxonomy in GEOSS Service Registry

GEOSS Component and Service Registry is built up on top of SUN Service Registry system. The SUN Service Registry is a reference implementation of the Java API for XML Registries (JAXR) specification. This specification provides an uniform and standard Java API for accessing different kinds of XML Registries. The backbone of this

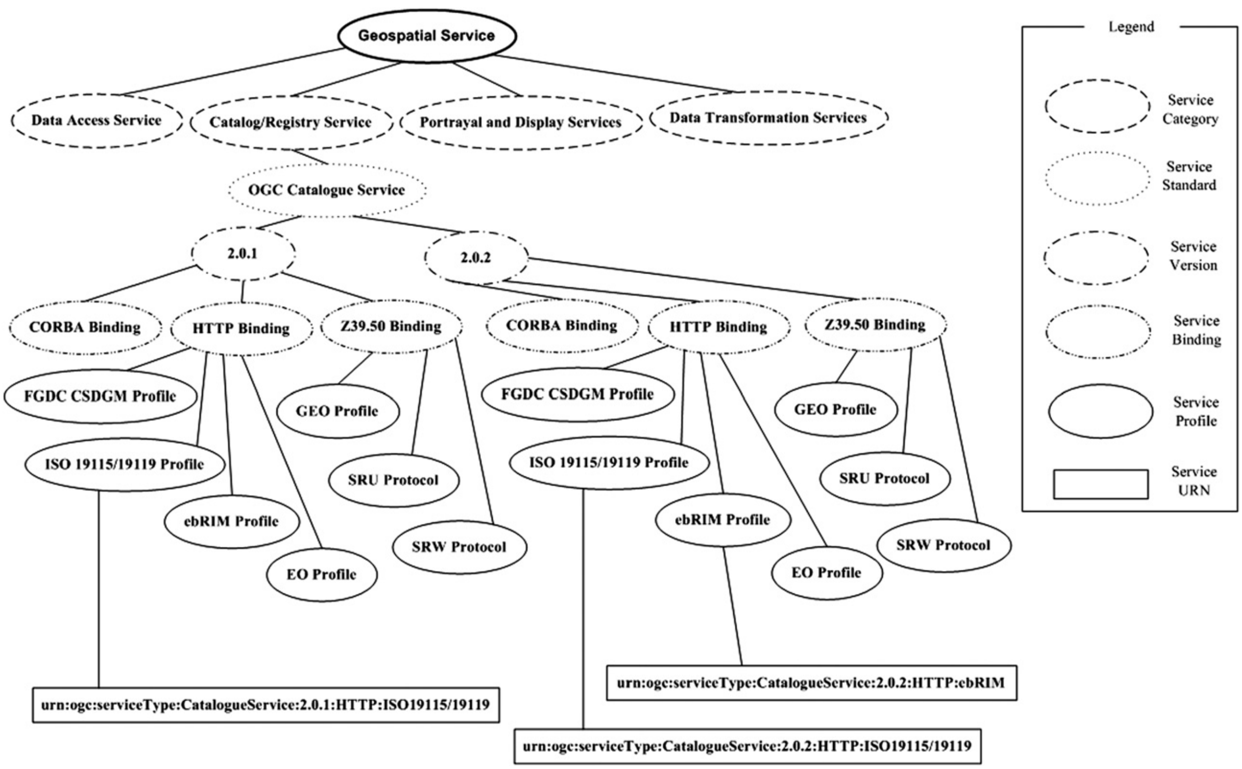

Fig. 3. Partial proposed taxonomy.

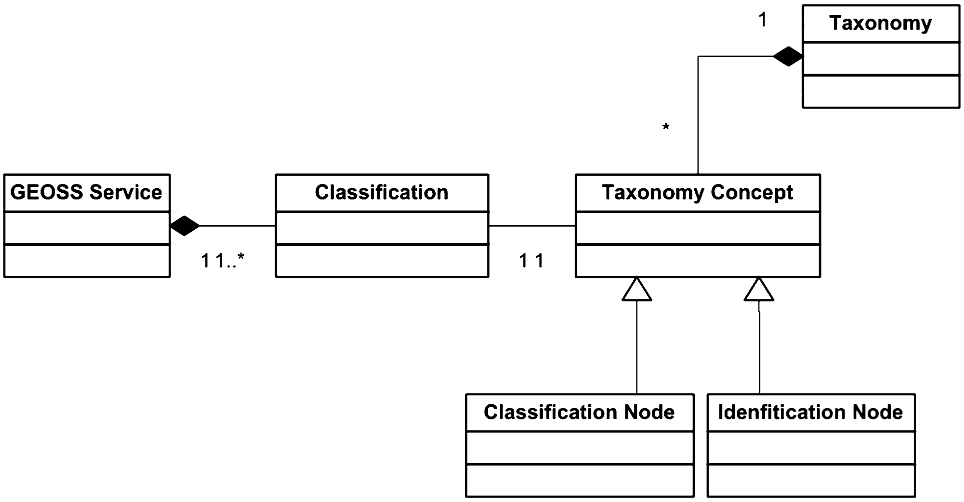

Fig. 4. Service classification mechanism.

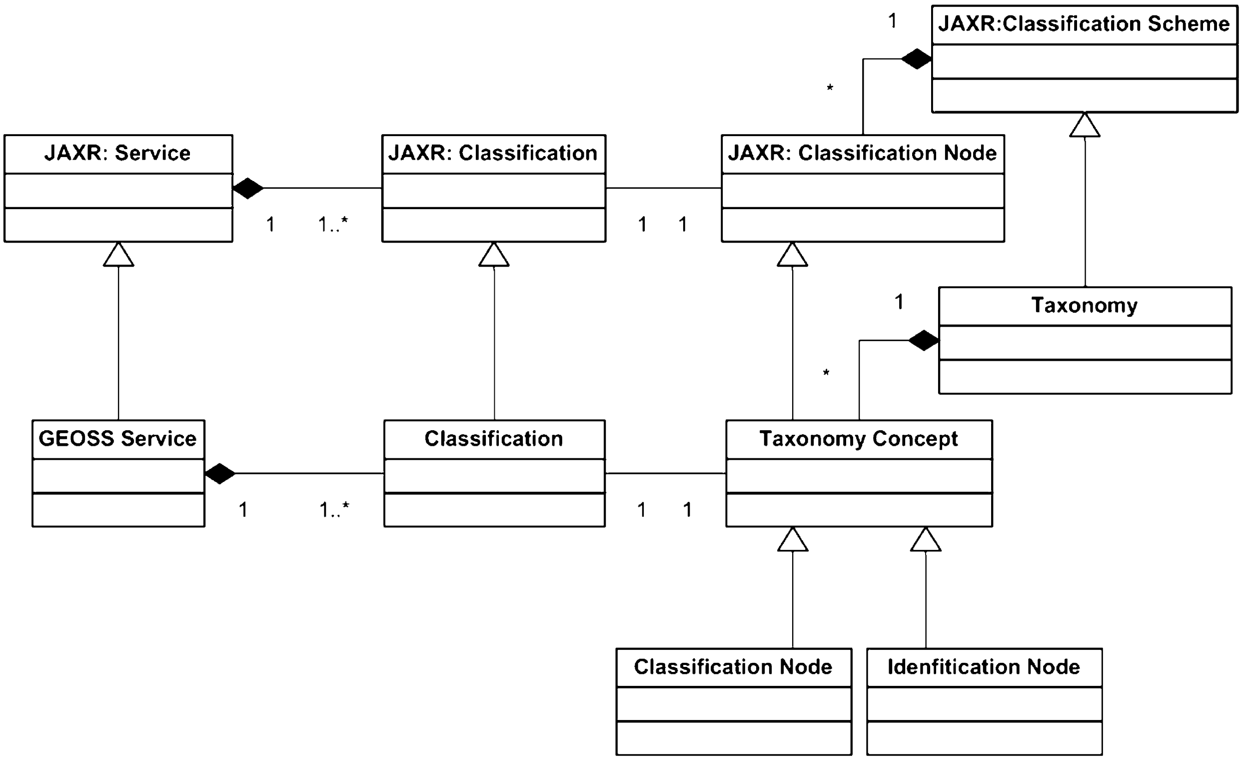

Fig. 5. Taxonomy application UML diagram.

specification is an information model for an xml registry that combines ebXML and the Universal Description Discovery and Integration (UDDI).

As shown in Fig. 5, this taxonomy itself is logically maintained as a JAXR Classification scheme internally. All the taxonomy concepts are created and maintained as Classification Nodes. In fact, the classification scheme simply describes a registered taxonomy in either the JAXR or the ebXML specification. The taxonomy tree can be defined internally to the registry by instances of Classification Nodes. Each GEOSS service is registered as an instance of a JAXR Service class. The GEOSS service references one or more taxonomy concepts through Classification instances. Each Classification instance links a GEOSS service with a referred taxonomy concept.

With this taxonomy and detailed classification information in place, GEOSS service providers now can simply choose appropriate taxonomy node(s) to be referenced during service registration process. These referred classification information are kept in the underlying database, along with other service metadata information. During service discovery procedure, service consumers can define those taxonomy node(s) that of interest as query criteria, and all the registered services will be shown up, as long as they reference defined taxonomy node(s). By employing this light weight service taxonomy, GEOSS service providers and service consumers can now easily reach a mutual understanding on those registered Services' behavior.

This taxonomy has also improved the service discovery functionality in the way that not only *Exact Match* query criteria can be supported, but *Fuzzy Match* could be fulfilled. For example, for a query with no exact match, those service instances may be suggested with different binding, profile, or version than the defined query parameters. This is achieved through *Query Abstraction* and *Query Substitution*, where similar taxonomy nodes will be used to issue new service discovery queries.

5. Discussion

One challenge faced when organizing a taxonomy is that it is too easy to get too complex and rigid with a taxonomy.[5,6,7] This proposed taxonomy intends to capture and represent geoscientific knowledge only about those geospatial service characteristics that are closely related with how a service could be discovered and reused. Since the design goal of this taxonomy is to promote global service discovery and interoperability, characteristics related to service interfaces are captured, while those related to service implementations are not. Actually, if a hierarchy is weighted down by too many perspectives and disparate rules for grouping and differentiation, it loses some of its power as a clear representation (Kwasnik, 1999).

The application of this taxonomy in the GEOSS Service Registry system shows how service discovery could benefit from this taxonomy. Even though services can also be discovered from pre-registered service metadata without a service taxonomy, a public taxonomy of services not only formally captures the knowledge about one or more aspects of the common characteristics of geospatial services but also provides a sustainable reference architecture for geospatial services development. These functions cannot be performed simply by duplicating several metadata items for geospatial service instances.

This proposed taxonomy follows a tree classification structure. Kwasnik pointed out that the use of trees as knowledge representation has some typical problems: *Rigidity*, *one-way flow of information* and *selective perspective* (Kwasnik, 1999). Our findings are that *Rigidity* can be overcome by having fairly complete knowledge of interest about a domain, before constructing taxonomy. This approach can significantly minimize the future revision of the taxonomy to represent more knowledge. By introducing layered taxonomy base structure, as shown in Fig. 2, this proposed taxonomy does not have the *one-way flow of information* problem. That is because siblings in this taxonomy tree are in fact of the same type of objects. However, the *selective perspective* problem always exists. Actually, the knowledge representation always depends on the context and goal of the representation (Kwasnik, 1999).

The limitation of this service taxonomy exists in that it does not deal with the content of service that is an essential part to reach the full interoperability, e.g. content type, content instance description. So that it cannot support seamless transitions between content discovery and service discovery.

Another critical issue for taxonomy engineering is that whether or not the design reflects the use. The advantage of folksonomy over taxonomy is that it originates directly from the users. Metropolitan Museum of Art's test in 2005 in which volunteers supplied keywords for 30 images of paintings, sculpture, and other artwork shows that there is a huge semantic gap between museums and publics.[8] It could also suggest that subject matter experts, who design the taxonomies, may talk a different language to the one used by the taxonomy users. Our lessons learned from this project show that users-oriented approach, other than experts-centered one would benefit the geospatial service providers and the geospatial service consumers with respect to understanding and utilizing the service taxonomy.

6. Conclusion

This paper discusses geospatial service taxonomy for global service discovery and interoperability. It summarizes the related service characteristics and presented a hierarchical taxonomy, which has been applied in the GEOSS Service Registry system. Our research results may lead to the further development of service taxonomy to thoroughly capture the knowledge around geospatial Web services. Our lessons learnt may be useful to others involved in geoscientific knowledge representation and manipulation.

Acknowledgment

The authors would like to thank Douglas Nebert at the USGS for his insightful idea and valuable guidance on the design, implementation and maintenance of the GEOSS Service Registry system.

References

Abbas, A., Saddik, A.E., Miri, A., 2006. A comprehensive approach to designing Internet security taxonomy. In: Proceedings of the

[5] When taxonomy fails. http://www.joiningdots.net/blog/2007/03/when-taxonomy-fails.html.

[6] Why taxonomy fails. http://www.joiningdots.net/blog/2007/04/why-taxonomy-fails.html.

[7] When and why taxonomy fails. http://blog.jackvinson.com/archives/2007/04/11/when_and_why_taxonomy_fails.html.

[8] One Picture, 1,000 Tags. http://www.nytimes.com/2007/03/28/arts/artsspecial/28social.html?ex=1178596800&en=5247f41a41e760e4&ei=5070.

Canadian Conference on Electrical and Computer Engineering. Ottawa, Canada, pp. 1316–1319.
Abrol, M., Doshi, B., Kanihan, J., Kumar, A., Liu, J., Mao, J., 2005. Intelligent taxonomy management tools for enterprise content. In: Proceedings of the 2005 IEEE/WIC/ACM International Conference on Web Intelligence. Compiègne, France, pp. 809–811.
Bazaz, A., Arthur, J.D., Tront, J.G., 2006. Modeling security vulnerabilities: a constraints and assumptions perspective. In: Proceedings of the Second IEEE International Symposium on Dependable, Autonomic and Secure Computing. Indianapolis, USA, pp. 95–102.
Farooqui, K., Logrippo, L., Meer, J.D., 1995. The ISO reference model for open-distributed processing: an introduction. Computer Networks and ISDN Systems 27, 1215–1229.
Grimshaw, D.J., 1996. Towards a taxonomy of geographical information systems. In: Proceedings of the Twenty Ninth Hawaii International Conference on System Sciences. Hawaii, USA, pp. 547–556.
Hayes, J.H., 2003. Building a requirement fault taxonomy: experiences from a NASA verification and validation research project. In: Proceedings of the Fourteenth IEEE International Symposium on Software Reliability Engineering. Denver, USA, pp. 49–59.
Huang, Z., Eliens, A., Ballegooij, V.A., Bra, D.P., 2000. A taxonomy of web agents. In: Proceedings of the Eleventh International Workshop on Database and Expert Systems Applications. London, UK, pp. 765–769.
International Standards Organization, 2005. ISO 19119, Geographic information–Services, p. 75. 〈http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=39890〉.
Kerschberg, L., Kim, W., Scime, A., 2001. A semantic taxonomy-based personalizable meta-search agent. In: Proceedings of the Second International Conference on Web Information Systems Engineering (WISE'01). Kyoto, Japan, pp. 41–50.
Kienle, H.M., Muller, H.A., 2007. A lightweight taxonomy to characterize component-based systems. In: Proceedings of the Sixth International IEEE Conference on Commercial-off-the-Shelf (COTS)-Based Software Systems. Banff, Canada, pp. 192–204.
Kovacic, S.F., 2005. General taxonomy of system approaches for analysis and design. In: Proceedings of the 2005 IEEE International Conference on Systems, Man and Cybernetics. Waikoloa, USA, pp. 2738–2743.
Kwasnik, B.H., 1999. The role of classification in knowledge representation and discovery. Library Trends 48, 22–47.
Lieberman, J., 2003. OpenGIS Web Services Architecture (OGC 03-025, Version 0.3). Open GIS Consortium Inc., USA, p. 58. 〈http://portal.opengeospatial.org/files/?artifact_id=1320〉.
Pahlevi, S.M., Kitagawa, H., 2002. Taxonomy-based adaptive Web search method. In: Proceedings of the 2002 International Conference on Information Technology: Coding and Computing. Las Vegas, USA, pp. 320–325.
Srikumar, V., Rajkumar, B., Kotagiri, R., 2006. A taxonomy of data grids for distributed data sharing, management, and processing. ACM Computing Surveys 38, 1–53.
Whiteside, A., 2005. OpenGIS Web services architecture description (OGC 05-042r2, Version 0.1.0). OpenGIS Consortium Inc., USA, p. 28. 〈http://portal.opengeospatial.org/files/?artifact_id=13140〉.