Semantic Interoperability*

With Geographic Information Systems (GIS) moving away from monolithic software to distributed network services, the question of how to ensure that these services operate with each other gains in importance. Typically, these services encapsulate a particular functionality exposed through a Web-based interface for communication as proposed by the Open Geospatial Consortium (OGC). The Service Oriented Architecture (SOA) describes the interaction between the service provider, the service broker and the user. The provider registers the service at a registry maintained by the service broker. The broker publishes the service interface, and hence, allows the service to be retrieved and finally executed by the user. The interface describes how to access the service and which data, if any, is exchanged. Besides human users, other services can also request a service via its interface forming service chains. The interface description guarantees syntactic interoperability, i.e., the possibility to exchange data on a technical level. Besides the used internet protocol and access restrictions, this especially includes the used data types. For instance, an interface of a Web-based gazetteer may announce that it requires a place name as input and delivers the place type and spatial footprint as results. The place name and place type would be specified as text strings, while the footprint would be defined according to a particular spatial reference system. What a place type (such as lake or canal) means, i.e., what is required for a particular place to be classified as lake or canal, is not covered by the interface description. This is a serious shortcoming as users accessing the service may assume that the services' conceptualization of the place type fits their own. Consequently, relying only on the names and data types of service parameters is not sufficient. This is especially important for service chains where one service relies on the adequate output of another service without manual control by a human user.

In contrast, to avoid misunderstandings, semantic interoperability requires a shared understanding of the meaning of the exchanged data, such as the place types in the gazetteer

* This is a draft version of Janowicz, K. (2009; forthcoming): Semantic Interoperability. In: Encyclopedia of Geography. B. Warf (Ed.). SAGE Publications.

example. To achieve this, unambiguous specifications of the service functionality, as well as the meaning of the input and output parameters are necessary. These specifications have to be readable to both human users and machines, and also have to be published in a service registry to support discovery. With respect to the gazetteer, the service needs to provide the specifications for the place types. For instance, lakes can be defined as natural bodies of water inside of a landmass. Based on such definitions, users (or other services) can decide whether the requested service fits their requirements. In information science the technology to specify such conceptualization is called an ontology. Ontologies support logical reasoning which is a prerequisite for information retrieval and interpretation. To achieve semantic interoperability between services their ontologies, called application ontologies, are compared to analyze whether the service parameters are based on the same conceptualization. Such analysis, however, requires a common base vocabulary used for the specifications.

An analogy to explain the need for a base vocabulary and the comparison of conceptualizations in application ontologies are Semantic Reference Systems (SRS). Same as spatial reference systems, semantic reference systems define a multidimensional (semantic) coordinate system and operations within such a system. One of these operations is semantic similarity which is defined as the inverse distance between two conceptualizations. It is analog to the distance between coordinates in a spatial reference system. Consequently, while the distance between two places represented in a gazetteer service can be measured based on their spatial footprints, the semantic similarity of the place types can be computed based on their formal specifications. While the question how to achieve semantic interoperability remains an open research challenge, two approaches can be distinguished.

Top-Level Ontology

A top-level ontology contains the commonly agreed upon base vocabulary independent from an application specific perspective, e.g., a particular service. For instance, a top-level ontology distinguishes between abstract and physical objects and defines units of measure. These specifications are then used as building blocks for domain ontologies which specify the vocabulary shared by a particular domain such as hydrology or geography. The domain ontology itself is used by the service provider to develop the application centric ontology necessary for the service interface. In case of the gazetteer service, the provider specifies what is categorized as lake using definitions from the domain ontology, such as the topological inside-relationship or the concept of water body. The mathematical notion of a relationship is defined in the top-level ontology. Based on this top-down approach, the specifications used by particular services can be compared as they rely on the same base vocabulary. This also allows for computer aided search in service repositories and enables users to find services which fit their needs. Semantic similarity is such a technology to support retrieval based on comparing the specifications for overlap and a step towards establishing semantic interoperability.

Folksonomies

While the definition and maintenance of top-level and domain ontologies is up to computer scientists and domain experts, so-called folksonomies are a bottom-up approach to establish a common understanding of the used vocabulary. Folksonomies are a crucial part of the Web 2.0 where users are actively contributing content instead of passively consuming it. By assigning labels (tags), users categorize content on the Web. For instance, in a Web 2.0 gazetteer users might assign tags as place types to mark particular places on a map. Technologies to analyze these tags and hence establish a common understanding based on user provided information are a promising new research direction.

Krzysztof Janowicz, Institute for Geoinformatics, University of Münster

See also Interoperability and Spatial Data Standards, Web Service Architectures for GIS, Semantic

Further Readings

Janowicz, K., Keßler, C., Schwarz, M., Wilkes, M., Panov, I., Espeter, M., and Bäumer, B. (2007). Algorithm, Implementation and Application of the SIM-DL Similarity Server. Second International Conference on GeoSpatial Semantics (GeoS 2007). Mexico City, Mexico,. Springer Lecture Notes in Computer Science,. 4853: 128-145.

Kuhn, W. (2005). Geospatial Semantics: Why, of What, and How? Journal on Data Semantics, Special Issue on Semantic-based Geographical Information Systems. Lecture Notes in Computer Science, 3534: 1-24.

MacKenzie, C. M., Laskey, K., McCabe, F., Brown, P. F. and Metz, R. (2006). OASIS Reference Model for Service Oriented Architectures 1.0 – Committee Specification 1, 2 August 2006.