From Geo-Pragmatics to Derivation Ontologies: new Directions for the GeoSpatial Semantic Web

Frederico Fonseca and **Andrea Rodriguez**

College of Information Sciences and Technology
Pennsylvania State University
University Park, PA 16802, USA
Email: fredfonseca@ist.psu.edu

Center for Web Research
University of Chile and
Department of Computer Science
University of Concepcion, Chile

We currently have a wealth of geographic information on the Web that is available in different forms ranging from images, maps, spatial databases, and tables to simple texts such as informal city guides, description of landscapes, and reports of bird watching activities. One of the most recent initiatives in trying to efficiently index, retrieve, and integrate information on the Web is the Semantic Web (Berners-Lee, Hendler et al. 2001). Berners-Lee's initial example shows the relevance of geographic information: *"At the doctor's office, Lucy instructed her Semantic Web agent through her handheld Web browser. The agent promptly retrieved information about Mom's prescribed treatment from the doctor's agent, looked up several lists of providers, and checked for the ones in-plan for Mom's insurance within a 20-mile radius of her home and with a rating of excellent or very good on trusted rating services."* This vision will be put in practice with the implementation of software agents that will talk with other agents in order to get the tasks done. For the agents to understand each other it is necessary to have ontologies in place that will define the vocabulary for the agents. For instance, in the example, one agent could have used 'close to', 'near', or 'within walking distance' instead of 'within a 20 miles radius', to achieve similar results. Ontologies provide precise definitions and can be linked to different contexts so that these terms have precise meanings that can be handled by the software agents.

In order to capture the richness of geographic information on the Web it is necessary to create representations of the geographic information resources. It is also necessary to link and map more informal resources, such as personal web pages with geographically interesting material, to the highly formalized representation structures that are being built as part of the Semantic Web. The result will be a broader spectrum of data available to users of geographic information.

In addition to the tendency in which ontologies of the Semantic Web are targeted at automated reasoning, the research in geo-ontologies often addresses information sharing and the development of GIS applications. Therefore it is necessary to clarify what we mean by geo-ontologies both in purpose and structure. "What is special about spatial?" (Anselin 1989; Egenhofer 1993) or what is special about geo-ontologies? Representing geographic entities–either constructed features or natural variations on or in the surface of the earth–is a complex task. These entities are not merely located in space; they are tied intrinsically to space. They take from space some of its structural characteristics, such as mereological, topological, and geometrical properties. A geo-ontology is then different from other ontologies because topology and part-whole relations play a major role in the geographic domain. In addition, geographic objects can be connected or contiguous, scattered or separated, closed or open, and they are typically complex and have constituent parts (Smith and Mark 1998). Furthermore, besides using geographic concepts, the Geospatial Semantic Web has other dimensions that involve space and time. Egenhofer (2002) emphasizes that the future of the Semantic Web includes not only a spatial component but also a non-spatial component.

This special issue emphasizes specific aspects of geographic information use on the Internet that basic Semantic Web research has not addressed yet. In particular, we address the three basic dimensions of information on the Geospatial Semantic Web: (1) *Professional*: structured geographic information stored in geographic databases which are organized in web portals or in institutional websites, (2) *Naïve*: unstructured, subjacent, informal geographic information in web pages, and (3) *Scientific*: geographic information within science papers, models, and theories (Fonseca and Sheth 2002).

The special issue starts with Brodaric approaching the scientific aspects of the Geospatial Semantic Web. He proposes an extension of the Pragmatic Web. In doing so, Brodaric's Geo-Pragmatics not only addresses geo-scientific concepts associated with geographical models and the need to represent their contexts, but also tries to cover some of the aspects not yet investigated by the Pragmatic Web. Pragmatics tries to understand how agents, such as scientists are using their scientific concepts in practice. Brodaric's take on Geo-Pragmatics is to look for the origins, effects and uses of the digitally represented geoscientific concepts and theories, so that we understand what they really mean. Brodaric uses an example of a supposedly continuous geological area that has a sharp break in its classification by two different groups of geologists. Geo-Pragmatics gives a framework to understand why and how the two groups reached their final classification so that when the time comes for the classification to be used, it will be done in a meaningful and judicious way.

Delboni et al. go right at the heart of the naïve geospatial web. They combine the work that has been done in GIScience regarding spatial terms such as *near*, *in front of*, *close to*, and *inside* with the most current work on search engines. The result is a framework that enables the use of what they call *positioning expressions* to perform geographic searches, without resorting to geocoded data or gazetteers. The authors also resorted to a bit of geo-pragmatics. They conducted an experiment to see how people really used the expressions in the context of web searches. They found that people use expressions that denote spatial relations both in the description of locations and in geographic queries submitted to a search engine. Thus their work supports the Geospatial Semantic Web by helping create queries that are closer to human language more effective in their results.

The next paper on geographic information searching, by Wiegand & García, makes the bridge between the naïve and the professional Geospatial Semantic Web. First, they emphasize the limitation of the conventional keyword-based approach. They focus on the problem involving semantics in the discovery and selection of appropriate data sources. This is a crucial problem for situations in which mobile clients are involved. Of course, geospatial portals suffer from the same problem, even though their interface is intended to be more interactive. The solution proposed by Wiegand & García is to organize ontologies around data intensive tasks, so that the search for data and dataset-related metadata can be done automatically by taking into consideration semantic contents.

Bilasco et al. also address the issue of adding more semantic content to the way we retrieve professional geographic information. They apply their research to 3-dimensional scenes and propose a geographic digital library, which is used in the description, annotation and indexing of geographic content. The resulting architecture enables the user to query a 3-D library using semantics and advanced topological operations. One of the main contributions of the paper is the definition of a set of semantic profiles for characterization of geometric data according to different users.

Still within the realm of professional geographic information, the next three papers deal with spatial data infrastructures (SDI) and how the Semantic Web can help to make them more effective. SDI are often defined as the set of technologies, policies and institutional arrangements that may support access to spatial data on the Web. In the first of the SDI-related papers, Lutz & Kolas analyze the drawbacks of using Description-Logics alone for the discovery of data sources and the answering of user queries in SDI. They show how rules can be used to overcome them. In their approach, rules are used both to discover and query distributed databases available in SDI as well as to combine the retrieved data to infer new knowledge.

In the next paper, Klien presents the use of semantic annotation of geographic data as a way to improve discovery and retrieval of data available within SDI. Semantic annotation provides support for the transition between concepts in an ontology and feature types in a database schema. Semantic annotations are defined as mappings between database schemas and domain ontologies. Klien suggests exploring geometry and topology in geographic data to automatically derive possible mappings. Based on this method, the process of semantic annotation can be partially automated. The presented strategy is the first step for the specification of a semantic annotation tool for geospatial web services.

In the third paper that deals with SDI, Aditya & Kraak focus on users. The paper criticizes some previous solutions for geospatial data discovery and proposes the development of a search interface using metadata aggregation. In their approach, information summaries of the data available in the SDI can be integrated and presented through the search interface. They also address aspects of visualization of user interaction in the search and discovery of data in SDI.

Dean's derivation ontologies closes the circle of the special issue with a link to Brodaric's Geo-Pragmatics. In their own way, both have an emphasis on the origins of spatial data. Dean suggests that with the focus on the processes that create the spatial patterns seen in spatial databases, derivation ontologies will help our understanding of the nature of the spatial databases they describe. In order to understand Dean's approach we have to keep in mind the two most common uses of ontologies today in information science (for a longer description seeFonseca 2007). When dealing with geographic information systems, ontologies usually represent the domain, which is addressed by the system. For instance, if we have a GIS to manage environmental resources, the most useful ontologies will be the ones that describe ecology principles in general and the ontology that describes the specific region under management. But a different use of ontologies would try to describe the geographic information system itself. Dean proposes his concept of derivation ontologies to describe how a spatial database may have been derived. Derivation will then be a complement to the more common content ontologies. Dean also presents a genetic programming approach to automatically develop derivation ontologies for existing databases.

Thus, the special issue covers the three basic dimensions of the Geospatial Semantic Web, Professional, Naïve, and Scientific. We hope the readers will enjoy the initial directions these papers give to the long road ahead in making geographic information widely and usefully accessible on the Web. We would like to thank Transactions in GIS and especially John Wilson for their support and opportunity to publish these high quality papers. We are also grateful for the excellent and expedite work of the reviewers that made this special issue possible. And finally we thank all the authors that submitted their papers for their excellent research and willingness to work together with us and the reviewers to make this special issue on the Geospatial Semantic Web.

References

Anselin, L. (1989). What is Special About Spatial Data? Alternative Perspectives on Spatial Data Analysis. Santa Barbara, CA, NCGIA.

Berners-Lee, T., J. Hendler, et al. (2001). "The Semantic Web: A New Form of Web Content That Is Meaningful To Computers Will Unleash A Revolution Of New Possibilities." The Scientific American **284**(5): 34-43.

Egenhofer, M. (1993). "What's Special about Spatial?-Database Requirements for Vehicle Navigation in Geographic Space." SIGMOD RECORD **22**(2): 398-402.

Egenhofer, M. J. (2002). Toward the Semantic Geospatial Web. Tthe Tenth ACM International Symposium on Advances in Geographic Information Systems, McLean, Virginia, ACM Press.

Fonseca, F. (2007). "The Double Role of Ontologies in Information Science Research." Journal of the American Society for Information Science and Technology **58**(6): 786-793.

Fonseca, F. and A. Sheth (2002). The Geospatial Semantic Web, UCGIS - University Consortium for Geographic Information Science**:** available at http://www.personal.psu.edu/faculty/f/u/fuf1/publications/e_geosemantic_web.pdf.

Smith, B. and D. Mark (1998). Ontology and Geographic Kinds. International Symposium on Spatial Data Handling, Vancouver, BC, Canada.