The University Consortium for Geographic Information Science

Research Priorities

THE GEOSPATIAL SEMANTIC WEB

THE PRIORITY

Additional research is needed on the Geospatial Semantic Web to provide more support for geographic information than is provided by basic Semantic Web research.

DESCRIPTION OF RESEARCH CHALLENGE

The role of semantics for interoperability and integration of heterogeneous data, including geospatial information, has been long recognized (Sheth 1999; Goodchild et al. 2001). The idea of a *Semantic Web* introduced by Berners-Lee et al. (2001) proposes "a web of data that can be processed directly or indirectly by machines," bringing a higher degree of automation in exploiting data in a meaningful way. Semantics is captured by associating formal descriptions to provide well defined meaning to data and other Web resources so that information processing (retrieval or integration) can be based on meaning instead of on mere keywords. The W3C Semantic Web Activity Working Group (http://www.w3.org/2001/sw/) has been working on a series of standards such as the Extensible Markup Language XML, the Resource Description Framework RDF, the Web Ontology Language OWL.

Ontologies play a critical role in associating meaning with data such that computers can understand enough to meaningfully process data automatically. Compared to syntactic means, a semantic approach leads to high quality and more relevant information for improved decision-making. Equally important is the use of ontologies to achieve shared understanding. Ontologies are also evolving as the basis for improving data usage, achieving semantic interoperability, developing advanced methods for representing and using complex metadata, correlating information, knowledge sharing and discovery. Ultimately, ontologies can be an important tool in expediting the advancement of related sciences, and they can reduce the cost by improving sharing of information and knowledge.

IMPORTANCE OF RESEARCH CHALLENGE

The Geospatial Semantic Web initiative specifically looks for better support for geographic information that the basic Semantic Web research has not addressed. In particular, we see three basic dimensions for geographic information on the semantic web:

Authors:
Frederico Fonseca
School of Information Sciences and Technology
Pennsylvania State University
University Park, PA 16801-3857
E-mail: fredfonseca@ist.psu.edu

Amit Sheth
Large Scale Distributed Information Systems Lab
Department of Computer Science
University of Georgia
Athens, GA 30602
E-mail: amit@cs.uga.edu

University Consortium for GIS
Suzy Jampoler, Director
Carolyn Merry, President

UCGIS
43351 Spinks Ferry Road
Leesburg, Virginia 20176-5631
TEL: (888) 850-8533
FAX: (703) 771-1635
Internet: http://www.ucgis.org

The UCGIS is a non-profit organization of universities and other research institutions dedicated to advancing the understanding of geographic processes and spatial relationships through improved theory, methods, technology, and data.

- *Professional:* Structured geographic information stored in geographic databases which are indexed or described in web pages (Egenhofer 2002).
- *Naïve*: The retrieval of unstructured, subjacent, informal geographic information in web pages.
- *Scientific*: Geographic information science papers, models, and theories.

In order to improve the results of queries looking for information stored in geographic databases it is necessary to support better definition for spatial concepts and terms used across different disciplines and the development multiple spatial and terminological ontologies (Egenhofer 2002).

In the second case we are looking for geographic information in web pages. Queries such as "*I found this interesting web site on the web, where is it located?*" or "*Find other web sites that contain information about places close to this web site or to places mentioned in this web site*" or "*List (or even display) all the location information on the IBM web site, offices, research centers, etc.*."

The third case is similar to what Citeseer (Giles et al. 1998) does today for Computer Science: a specialized search engine types of research and the support for it has been limited.

EMINENT RESEARCH QUESTIONS

Among the challenges we face to make both types of queries feasible are:

- Creation and Management of geo-ontologies: Activities involved in ontology management include designing, developing, storing, registering, discovering, visualizing, maintaining, and querying ontologies. One aspect that makes ontology management particularly challenging is that ontology is based on agreements (and preferably consensus) among domain experts that can be geographically distributed. Ultimately, their survival is based on users' acceptance. This to a good part involves a social and collaborative process. The GIS community can support an initiative in ontology management that can include developing or adapting effective methodologies and tools for ontology management, and applying them to develop domain specific ontologies with broad community acceptance.

- Matching geographic concepts in web pages to geo-ontologies: It is necessary to apply a geospatial characteristic to the interpretation of texts (hermeneutics). Innovative methods are also required to be able to build ontologies from maps, images, and sketches available on the web.

- Ontology integration: In order to provide better results for queries it is necessary to integrate different ontologies not only in the geographic dimension (scientific, professional, naïve) but also on the non-geographic domain. Future research needs to address the necessity of developing and testing the theory of the integration of multidisciplinary ontologies by: (1) performing an empirical study of how different communities categorize the relationship between the different geographic entities; (2) creating relevant geo-ontologies; and (3) designing, prototyping, and assessing computational models to specify, represent, access, and share multiple ontologies of geographic information.

REFERENCES

Berners-Lee, T., Hendler, J., and Lassila, O. (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: 34-43

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

Giles, C.L., K. D. Bollacker, K.D., and Lawrence, S. (1998) CiteSeer: An Automatic Citation Indexing System. In *Proceedings of the Third ACM International Conference on Digital Libraries,* Pittsburgh, Pennsylvania

Goodchild, M.F., Egenhofer, M.J., Fegeas, R., and Kottman, C.A. (eds.) (1999) *Interoperating Geographic Information Systems*. New York, Kluwer

Sheth, A. (1999) Changing Focus on Interoperability in Information Systems: From System, Syntax, Structure to Semantics. In Goodchild, M.F., Egenhofer, M.J., Fegeas, R., and Kottman, C.A. (eds.) *Interoperating Geographic Information Systems*. New York, Kluwer: 5-30