An Evaluation of Geo-Ontology Representation Languages for Supporting Web Retrieval of Geographical Information

P. Smart, A.I. Abdelmoty and C.B. Jones
School of Computer Science, Cardiff University, Cardiff, Wales, UK
P.Smart,A.I.Abdelmoty,C.B.Jones @cs.cardiff.ac.uk

Extended Abstract

The internet is the single largest information resource in the world. It is, however, not being used to is full potential. Currently most the information is written using syntactical machine readable languages such as HTML. These languages are limited in that they are only intended for human consumption. To fully unlock the potential of such a vast resource of information, we need to make the information not only machine readable but machine-understandable. In order to gain machine understanding we need semantic languages which are able to define meaning to the information being stored. Agents (human or machine) could then use this information in variety of different ways.

A large amount of geographical information is currently being stored and delivered over the internet. Internet providers such as the Ordnance Survey are realizing the potential and are currently offering their data in GML format. Geographic digital libraries, such as the ADL, are being established. There is, however, the need to realize the potential of semantically enriching the geographic information to provide more automated and intelligent ways of managing and retrieving the data over the web.

In February 2001 the semantic web initiative was launched by W3C (W3C, 2001) for the semantic representation of data on the Web. Tim Berners Lee, the creator of the World Wide Web, is quoted as saying 'The Semantic Web is an extension of the current web in which information is given well-defined meaning, better enabling computers and people to work in cooperation'. In order for the semantic web to function, computers need access to structured information and inference rules. Key to this is the use of ontologies. An ontology is a specification of a conceptualization, which provides the structured vocabulary and semantics which can be used in the markup of web resources to provide machine understanding.

The SPIRIT (Spatially Aware Information Retrieval on the Internet) project (C.B. Jones *et al.* 2003) is a project funded as part of the Semantic Web European initiative which aims to provide an intelligent, web based, geographical search engine. At the heart of the SPIRIT system is a geographical ontology which provides support for the different components of the system. The ontology plays an important role in supporting query disambiguation and query term expansion of the required query, relevance ranking of the retrieved search results, the creation of the spatial indexes to support the search and the annotation of web resources; web documents and geographic data sets. Figure 1 show the geo-ontology as proposed in the SPIRIT project.

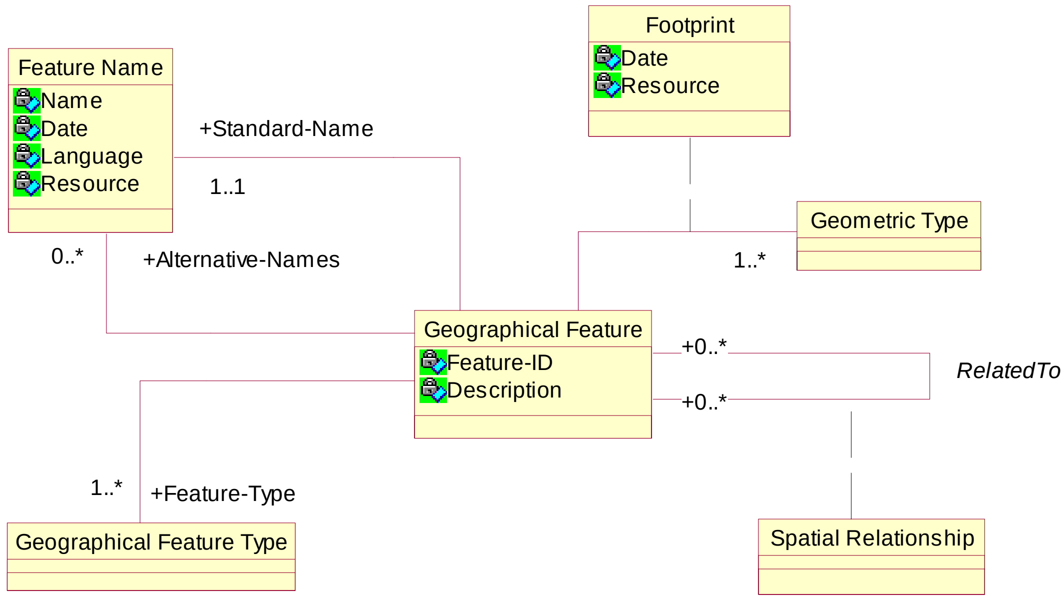

Figure 1. Base Schema of Geographical Feature Ontology (C.B. Jones et al, 2003).

There are various possible representation languages for representing the geographic ontology proposed by SPIRIT. In this paper, two languages, namely, DAML+OIL (Ian Horrocks, 2002), a description logic-based language and GML (OGC Recommendation Paper, 2001), an XML-based language are critically reviewed and compared as to their suitability for representing the geo-ontology. The SPIRIT ontology shall be taken as a case study for the evaluation of the languages and examples of both languages shall be given to support the arguments presented.

This paper provides an overview of related work in the area of geo-ontology development as well as an overview of both GML and DAML+OIL. The potentials and limitation of both languages are examined for representing and supporting the geo-ontology. The set of required features of a geo-ontology representation language are proposed.

In order to fully explore the capabilities of DAML+OIL and other ontology languages, a specification of requirement needs to be drawn up, various prototype ontologies need to be developed and tested and the ability of each language to meet the specification needs to be measured.

A summary of the necessary and desirable features for a geo-ontology representation language are as follows. The language should be capable of the basic functionality.

1. Representation of real world geographic concepts.
2. Representation of data properties for each geographical concept.
3. Representing relationships between concepts.
4. Representing specialization/generalization concept hierarchies.
5. Representing of simple composition hierarchies, i.e., concepts made up of sets of other concepts.

It is desirable for the language to provide the following functionality.

6. Representing constraints on data properties of concepts.
7. Representing constraints on relationships.

8. Expressing integrity rules over individuals
9. Expressing integrity rules between individuals belonging to different concepts.
10. Representation of advanced composition hierarchies, i.e. a class of all houses which are within 10 miles of a motorway.

DAML+OIL proposes to be a very rich and expressive language for the conceptualization of real world phenomena. This study is about how to utilize the seemingly powerful features of the language to represent the geographic domain. A characteristic feature of the geographic domain is the potentially very large data sets and related information to be represented. One area of particular interest is the explicit versus implicit maintenance of spatial data. The potential size of a fully populated DAML+OIL geo-ontology is huge. An important question arises as to which information should be explicitly stored. The tradeoff is between computational complexity and storage costs. The question also relates to the ability of the language to automatically derive the implicit information by utilizing spatial reasoning techniques. The question relates also to the scalability of the language and its ability to represent and reason with potentially thousands of individuals.

At a simplistic level DAML+OIL provides, via its well grounded semantics, a powerful way of capturing real world geographical concepts. Definition of user-defined terminologies is possible to allow designers to create ontologies which suits their own individual needs. Concepts can be captured via classes, inheritance mechanisms for representing generalization and specialization hierarchies, binary relations to provide links between classes or properties, cardinality constraints to limit relations, and class constructors to allow for the creation of defined classes. There is also support for transitive and symmetric properties, which are useful when specifying simple spatial relationships, such as adjacency. However, a number of issues arose on using the language for representing the SPIRIT ontology which impeded the languages ability to conform to all of the specifications laid out. Two of the main issues are summarized below.

One of the main concerns is that DAML+OIL doesn't support its own data types. Instead, it uses XML data types which don't allow for logical comparison between two data values. This makes it difficult to perform any spatial reasoning tasks which works on co-ordinate data, i.e. to test the existence of a ‘contains’ relationship between two MBRs.

Another point of concern is tool support. Tools are needed to effectively encode domain experts’ knowledge. Currently tools such as Oiled (Sean Bechhofer *et al.,*. 2001) and protégé 2000 (W. Grosso *et al.,* 2000), allow for the creation of DAML+OIL ontologies, and via links to the FACT (Ian R. Horrocks, 1998), and RACER (Volker Haarslev *et al.,* 1999) reasoners respectively, the consistency of the ontologies can be checked. However, the development of advanced querying tools is still ongoing for use with DAML+OIL. At present most of the querying languages available work on RDF(S) triples (RQL (G. Karvounarakis *et al.,* 2002) being one such querying language). Therefore the full extent of the DAML+OIL language is not catered for, and in particular the support for inference mechanisms for advanced querying of the knowledge bases.

In comparison with GML, DAML+OIL fares favorably. GML is a modeling language which isn't as semantically rich and expressive as DAML+OIL. It has a pre-defined vocabulary which doesn’t allow for user created concepts, and any operations to be performed on the information would have to be provided by an additional layer of software, which would rely on its own interpretation of the information. However the standardized nature of GML (recommended by W3C), and the very fact that it uses a pre-defined vocabulary is probably the reason it is so popular for geographic markup of web pages. Often there is no need to use such an expressive language to represent straight forward concepts, indeed the current Agent cities Urban ontology (Agent Cities, 2004) written in DAML+OIL only uses the very basic features of the language.

The paper elaborates on the above issues and uses a detailed example for representing the ontology. A critical analysis of which ontology language would be best suited to the rich conceptualization of the geographic domain is given, with specific interest in their ability to reason over space and derive implicit spatial relationships.

Bibliography

W3C, (2001), Semantic Web Introduction, http://www.w3.org/2001/sw

Ian Horrocks, (2002), {DAML}+{OIL}: A Reason-able Web Ontology Language, citeseer.nj.nec.com/579456.html

OGC Recommendation Paper, (2001)**,** Geography Markup Language (GML) 2.0, citeseer.nj.nec.com/422221.html

Sean Bechhofer and Ian Horrocks and Carole Goble and Robert Stevens, (2001), {OilEd}: {A} Reason-able Ontology Editor for the Semantic {Web}, citeseer.nj.nec.com/bechhofer01oiled.html

Ian R. Horrocks, (1998), Using an Expressive Description Logic: {FaCT} or Fiction?, citeseer.nj.nec.com/horrocks98using.html

Agent Cities, (2004), DAML+OIL Urban Ontology, http://iscte.pt/~lhrm/daml/location

C.B. Jones, A.I. Abdelmoty and Gaihua Fu, (2003). "Maintaining Ontologies for Geographical Information Retrieval on the Web", Ontologies, Databases, and Applications of Semantics for Large Scale Information Systems (ODBASE'03). Proceedings of the OTM Confederated Int. Conf., R. Meersam, Z. Tari, Schmidt, D.C. editors, Springer Verlag, LNCS 2888, pp. 934-951.

Volker Haarslev and Ralf Moller, (1999), {RACE} System Description, Description Logics, citeseer.nj.nec.com/article/haarslev99race.html

W. Grosso and H. Eriksson and R. Fergerson and J. Gennari and S. Tu and M. Musen, (2000), Knowledge Modeling at the Millennium -- The Design and Evolution of Protege-2000. Proceedings of the 12 th International Workshop on Knowledge Acquisition, Modeling and Mangement (KAW'99), Banff, Canada, October 1999.

Extensible Markup Language, (XML), http://www.w3.org/XML

G. Karvounarakis and S. Alexaki and V. Christophides and D. Plexousakis and M. Scholl, (2002), {RQL}: A Declarative Query Language for {RDF}, In The 11th Intl. World Wide Web Conference (WWW2002).

Biography of Dr. Alia I Abdelmoty

Dr. A I Abdelmoty is working as a Lecturer of Computer Science in Cardiff University since 2000. Her research interests include qualitative spatial representation and reasoning techniques and applications (intelligent spatial query languages, similarity checking of spatial scenes and maintaining the integrity of large spatial databases), multi-scale spatial databases, spatio-temporal data modelling and manipulation, visual query interfaces to GIS, web-based geo-information retrieval, geo-ontologies and geo-metadata information visualisation.

Biography of Philip D Smart

Philip D Smart has recently completed his degree in Computer Science from Cardiff University, he is currently studying for his Doctorate in the area of Geographic Information Systems, in particular his research efforts are focused at the use of ontologies to provide Geographic Information Retrieval on the Internet.