GeoReferencing the Semantic Web: ontology based markup of geographically referenced information

Kaoru Hiramatsu [1] and Femke Reitsma [2]

[1] NTT Communication Science Laboratories, 2-4, Hikaridai, Seika-cho, Soraku-gun, Kyoto, 619-0237 JAPAN, hiramatu@cslab.kecl.ntt.co.jp

[2] Maryland Information and Network Dynamics Laboratory, University of Maryland, College Park, MD 20742 USA, femke@geog.umd.edu

Abstract: Geographic references that include geographic relations to well known locations are useful for explaining real world information and for knowledge discovery. In order to utilize the spatial characteristics of this information on the web, our research takes advantage of recent developments in the Semantic Web community. To circulate geographically referenced information on the Semantic Web, we have created a geographic ontology for describing them and have implemented two tools, a web service that calculates geographic relations, and a plugin to an open-source GIS that allows us to generate RDF of topological and direction spatial relations among geographic features. This ontology based approach allows us to associate geographically referenced data to any other non-spatial information related to the geographic feature that is expressed on the Semantic Web. These two different approaches to developing geographic references for the Semantic Web have provided us with the opportunity to evaluate the advantages and disadvantages of each, providing us with direction for future work.

1 Introduction

Geographic references are useful for explaining real world information. Such geographic references include geographic relations to well-known locations (e.g. distance to a hotel from an airport and route information which includes street names and directions) and are often used in Web pages of World Wide Web (WWW). In many cases, the geographic references are easy to understand for human users but hard for computers because almost all of them are in plain text form and related knowledge is necessary to solve their semantic ambiguity. If computers could handle such geographic references flexibly, such as hyperlinks in WWW, they would create more expressive geographic information than plotting icons on the map. This is the promise of the Semantic Web, where formal semantics allow us to automatically utilize information in new ways based on their associated ontology. For example, it would be possible to create travel guides by connecting geographic references in Web pages and retrieve information by intuitive geographic relations for human users.

This is particularly relevant given the proliferation of spatial data portals, where finding spatial data is often difficult for humans, and impossible for intelligent computational agents. The Semantic Web supports a future of intelligent agents inferring knowledge and thereby supporting spatial decision making. As evident in spatial decision support (SDSS), spatial decision problems typically combine heterogeneous spatial and non-spatial data, requiring the integration of different types of information [1]. Seamless integration of geographic information with other information based on its semantic content regardless of its representation has also been utilized for GIS (Geographic Information Systems) [2, 3]. Semantic interoperability requires formally defined concepts and terms, as can be expressed in ontologies.

In this paper, we illustrate our approach to geographic references using technologies of the Semantic Web. Our approach consists of two parts, creation of a geographic ontology and its instances, and implementation of two tools that utilize the ontology. This has allowed us to explore some of the advantages and disadvantages of each and provides direction towards future work. The geographic ontology includes typical geographic feature classes and geographic relations (e.g. topological, distance, and direction relationships). It is written in OWL[1] so that we can use it with other geographic ontologies that are based on RDF[2]. The first tool is a web service that enables us to calculate the geographic relations between the geographic instances using their coordinates. They are accessible using SOAP[3] based on service descriptions in WSDL[4] so that we can use them from other applications such as RDF editors, inference engines, and search engines. We have also implemented a plugin for a lightweight open-source GIS (Geographic Information System) that generates RDF describing the spatial relations among geographic features. Both tools are presented on the web and the plugin is available for download from http://www.mindswap.org/2004/geo/geoStuff.shtml .

---

[1] http://www.w3.org/2001/sw/WebOnt/
[2] http://www.w3.org/RDF/
[3] http://www.w3.org/TR/SOAP/
[4] http://www.w3.org/TR/wsdl20/

2 Approach

2.1 Geographic Ontology and Instances

On the Internet, there is much spatial data and many gazetteers. They are carefully constructed and well maintained by specialists. In order to use geographic data on the Semantic Web, we created two geographic ontologies, which are written in OWL (Table 1). The ontology of geographic features was developed to provide appropriate geographic references to test the tools we created (described below), for example, expressing classes such as country, city, and continent. An ontology of spatial relationships was also developed in order to express the topological, direction, and distance relationships between geographic features and relevant mereological relations.

Table 1. Classes and properties in the geographic ontologies

	Sample classes and properties
Geographic feature	Continent, Country, City (Shape) Point, Line, Polygon, MultiPolygon
Geographic relation	(Topological) Equals, Disjoint, Touches (Mereological) isWholeOf, isPartOf (Direction) isNorthOf, isLeftOf, isBehindOf (Space distance) withinMetersOf (Time distance) withinMinutesOf

Following the OpenGIS Simple Features Specification[5] of topological relations based on the Dimensionally Extended 9-Intersection model (DE-9IM) [4, 5], the ontology includes the following eight relations: equals, disjoint, intersects, touches, crosses, within, contains, and overlaps. The ontology allows for quantitative and qualitative distance relations to be established with a primary object, reference object, and a frame of reference [6]. In terms of qualitative expressions of direction, we apply the 8-sector model to express the cardinal directions North, NorthEast East, SouthEast, South, SouthWest, West, NorthWest [7]. We define these direction relations at two levels of constraint, for example, we define A isNorthOf B to be true if the northern most point of A is further north than northern most point of B as the first level of direction relations, and its subproperty A is CompletelyNorthOf B to be true if the southern most point of A is north of the northern most point of B as the second level of spatial direction relations. Therefore, if A isCompletelyNorthOf B, the relation A isNorthOf B is also true. The mereological relations included in the ontology define the whole/part nature of things.

Owing to its RDF basis, the geographic feature classes can also have properties such as labels, administrative codes, statistical data using other ontologies and namespaces (e.g.

[5] www.opengis.org

[8], [9]). This allows us to associate spatial data to any other form of information expressed on the Semantic Web, such as with the OWL sameAs predicate, facilitating inference and reasoning across spatial and non-spatial data.

The geographic instances are translated from the spatial dataset into GML[6] using ESRI's ArcExplorer[7] and then into OWL using an XSLT style sheet. As shown in a sample in Figure 1, we insert a place name and coordinates into the geographic instances because most people use place names to refer to geographic locations, rather than coordinates corresponding to geometric representations of space [10] and the coordinates are suitable for computers to calculate geometric relations between the geographic instances. Using uniquely assigned URIs, we can share them as a part of geographic references on the network.

```
<geo:Country rdf:ID="Tokelau">
  <rdfs:label>Tokelau</rdfs:label>
   <geo:shape>
       <geo:MultiPolygon rdf:nodeID="TokelauShape">
          <geo:xyCoordinates>
                 -171.848052 9.218889, ...
          </geo:xyCoordinates>
          <geo:xyCoordinates>
                 …., ...
          </geo:xyCoordinates>
        </geo:MultiPolygon>
  </geo:shape>
</geo:Country>
```

Fig. 1. Sample instance "Tokelau" (namespaces are omitted)

2.2 Instance creation by GIS plugin

A second tool was developed that allows us to create well formed RDF for expressing topological and direction relations among spatial objects. This tool utilizes the ontologies discussed above and a number of open source APIs in order to markup spatial features. The GeoMarkup tool has been developed as a plug-in of JUMP, which is an extendable lightweight GIS (Geographic Information System) for viewing, editing, analyzing, and processing spatial data, and is accessible via a menu from the JUMP Workbench. Thus far, only the topological and direction relations have been implemented with this tool. Development of distance relations depends on defining appropriate algorithms for varying coordinate systems, which have not yet been fully developed in the ontology.

---

[6] http://www.opengis.org/docs/02-023r4.pdf
[7] http://www.esri.com/software/arcexplorer/index.html

The JUMP Workbench with the GeoMarkup output is presented in Figure 2 below. The top left pane displays two geographic features that have been drawn with the editing tools (top right pane) and are selected for spatial markup. The central bottom pane displays three automatically generated URIs for the three selected features, the sets of spatial relations that may be displayed, and the output RDF of the topological, direction, and complete direction relations. The details of these relations follow.

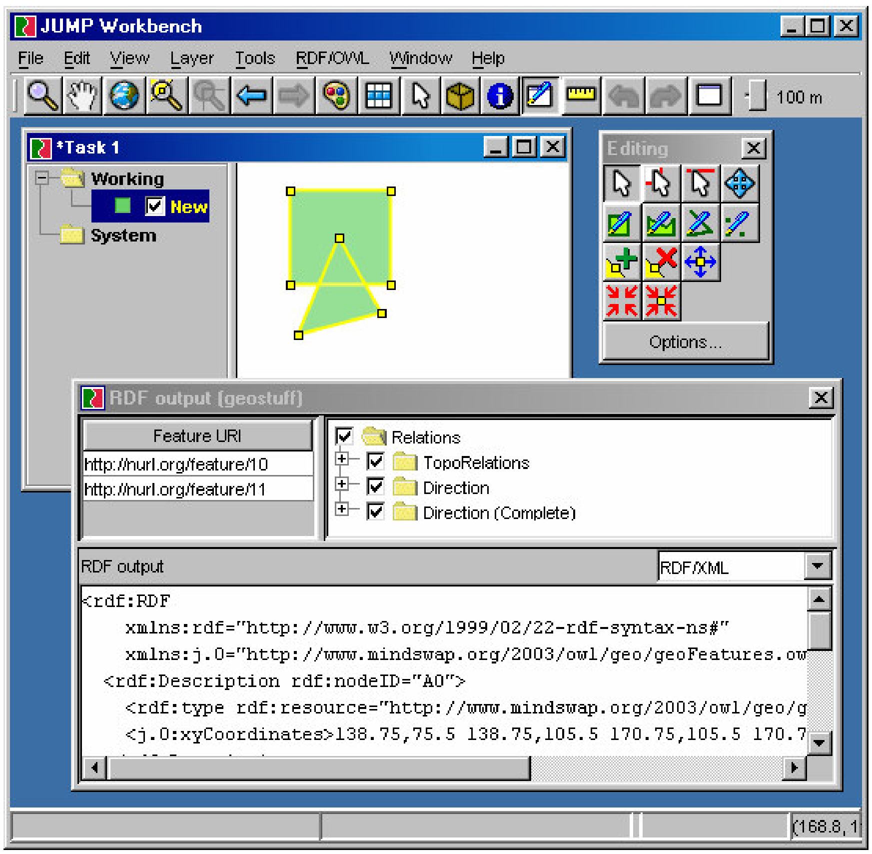

Fig. 2. GeoMarkup Tool

The topological and direction relations applied with the GeoMarkup tool are those defined in the ontology, these relations may be limited to only those selected with the checkbox and can be output in different forms, that is, in full RDF, abbreviated RDF, N-Triple, and N3, as defined by Jena. The output for the topological relations displayed in

Figure 2 above is expressed in full RDF and displayed in Table 2 below. The direction relations can be expressed in the same manner, but are not done so here.

Table 2. RDF ouput of GeoMarkup Tool

```
<rdf:RDF
   xmlns:j.0="http://www.mindswap.org/2003/owl/geo/geoRelations.owl#"
   xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
   xmlns:j.1="http://www.mindswap.org/2003/owl/geo/geoFeatures.owl#" >
 <rdf:Description rdf:nodeID="A0">
   <rdf:type rdf:resource = "http://www.mindswap.org/2003/owl/geo/geoFeatures.owl#Polygon"/>
   <j.1:xyCoordinates>138.75,75.5 …</j.1:xyCoordinates>
 </rdf:Description>
 <rdf:Description rdf:about="http://nurl.org/feature/10">
   <j.0:intersects rdf:resource="http://nurl.org/feature/11"/>
   <j.0:crosses rdf:resource="http://nurl.org/feature/11"/>
   <j.1:shape rdf:nodeID="A0"/>
   <j.0:overlaps rdf:resource="http://nurl.org/feature/11"/>
 </rdf:Description>
 <rdf:Description rdf:nodeID="A1">
   <j.1:xyCoordinates>154.25,90.5 ...</j.1:xyCoordinates>
   <rdf:type rdf:resource="http://www.mindswap.org/2003/owl/geo/geoFeatures.owl#Polygon"/>
 </rdf:Description>
 <rdf:Description rdf:about="http://nurl.org/feature/11">
   <j.0:intersects rdf:resource="http://nurl.org/feature/10"/>
   <j.1:shape rdf:nodeID="A1"/>
   <j.0:overlaps rdf:resource="http://nurl.org/feature/10"/>
   <j.0:crosses rdf:resource="http://nurl.org/feature/10"/>
 </rdf:Description>
</rdf:RDF>
```

2.3 Web Services for Geographic References and Experimental Client

To circulate the geographic references on the Semantic Web, we prepared two Web services, an instance repository and a relation calculator, using Jena[8], Apache/AXIS[9], and the JTS Topology Suite[10]. The instance repository stores indices of URIs, names, and coordinates for enabling us to retrieve the geographic instances by geographic regions and names. The geographic instance set for this experiment is small enough to be handled by one Web service on one PC but full datasets would be so large that we have to classify them under sources, classes, and regions to manage them distributedly. There exist many

[8] http://www.hpl.hp.com/semweb/jena2.htm
[9] http://ws.apache.org/axis/
[10] http://www.vividsolutions.com/

strategies for such management but further discussion of them is beyond the scope of this paper.

The relation calculator enables us to test geographic relations based on spatial coordinates in the geographic instances. The relation calculator also enables us to find unrealized relations between geographic instances based on coordinates. As the first phase of our research, we have prepared a relation calculator that tests topological relations between two geographic instances. The other relations in the geographic ontology will be implemented in next phase.

Using the geographic ontology, its instances, and the Web services, we have implemented a prototype client which has a GUI (Figure 3.) It works with the Web services and helps us create geographic references by checking geographic relations interactively as follows.

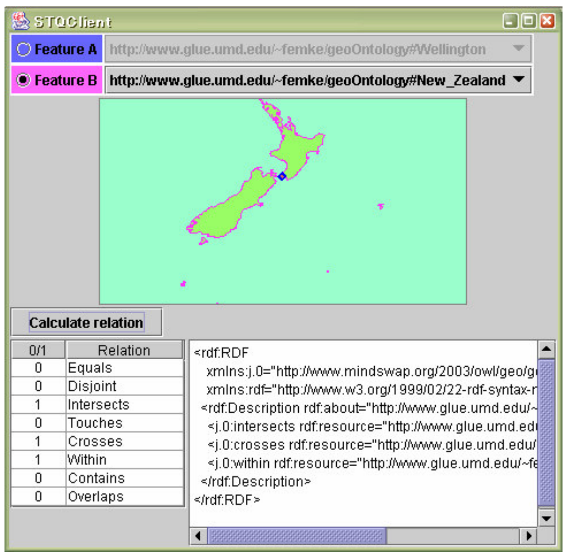

Fig. 3. A prototype client of the Web Services

First the client downloads all geographic instances that are stored in the instance repository, as discussed above in Section 2.1. The downloaded instances are inserted into pulldown menus at the top of the client and drawn as polygons on a map view. Next, a user selects geographic feature A and B by using the pulldown menus or clicking polygons on the map view and presses the button “calculate relation.” Then the client sends coordinates of the selected features to the relation calculator to test the geographic relations between A and B. The results are returned to the client and it displays true or false for each tested relation in the table and generates RDF data in the text pane at the bottom of the client. In this example, the user selected Wellington (the capital of New

Zealand) as A and New Zealand as B. The result is A intersects, crosses, and is within B as shown in the client. Additionally, the result in the text pane can be inserted into RDF descriptions of real world information by cut and paste.

3 Discussion and Conclusion

In this paper, we illustrated our approach to geographic references based on the Semantic Web. The two approaches to representing geographically referenced information on the web have a number of advantages and disadvantages which provide some guidance for future developments. For the ontologies, although we can express spatial relations, we cannot embark in sophisticated spatial reasoning. OWL is restrictive in expressing spatial relations, where we cannot reason about qualitative spatial relations. Rather, inference of qualitative relations is derived from coordinate based services.

The Web services we have implemented are accessible using SOAP based on service descriptions in WSDL. Therefore, it has the advantage that we can integrate them into other applications. For example, the Web services enable RDF editors to search for related landmarks, check description errors, and apply inference engines to check consistency. They also help search engines to retrieve geographic instances that satisfy geographic relations described in RDF description based on our ontologies. The disadvantages of this approach include scalability issues, where the datasets are very large in OWL format. Storage systems and inference engines that can handle large datasets will be required for practical use.

For the former, we should prepare some services and a database for alignment and mapping between classes and properties that are defined in different ontologies. Using this, we can do inference and reasoning on heterogeneous instances that are based on different ontologies. For the latter, we have to prepare some services that support validation and creation (like geoServices).

Advantages of the GIS plugin include the ability to utilize spatial information in standard GIS formats, such as ESRI's shapefile, for describing RDF datasets based on the geographic ontology. This enables us to import standardized geographic instances into our own descriptions and use them for geographic inference. However, the plugin has the disadvantage of usability, a similar barrier to standard GIS. The plugin is powerful for describing geographic features and relationships but needs some initial setup and operational skill.

Acknowledgements

We would like to acknowledge Jim Hendler and the Mindswap gang for providing a stimulating environment for the development of this work. We also thank Ryusuke Masuoka for many helpful suggestions.

References

[1] Casey, M. J. and M. A. Austin (2002). Semantic Web Methodologies for Spatial Decision Support. International Conference on Decision Making and Decision Support in the Internet Age, DSIage2002, University College Cork, Ireland.

[2] Fonseca, F. and M. Egenhofer (1999). Ontology-Driven Geographic Information Systems. 7th ACM Symposium on Advances in Geographic Information Systems, C. B. Medeiros (eds), Kansas City, MO.

[3] Fonseca, F., M. Egenhofer, P. Agouris and G. Câmara (2002). Using Ontologies for Integrated Geographic Information Systems. Transactions in GIS 6(3): 231-257.

[4] Clementini, E. and P. Di Felice (1995). A Comparison of Methods for Representing Topological Relationships. Information Sciences 3: 149-178.

[5] Egenhofer, M. and J. Herring (1990). A Mathematical Framework for the Definition of Topological Relationships. Fourth International Symposium on Spatial Data Handling, (eds), K. Brassel and H. Kishimoto (eds) Zurich, Switzerland, pp. 803-813.

[6] Hernández, D., E. Clementini and P. Di Felice (1995). Qualitative Distances. Proceedings of Conference on Spatial Information Theory (COSIT'95), A. Frank and W. Kuhn (eds), Semmering, Austria, September 21-23, Lecture Notes in Computer Science 988, Springer.

[7] Hong, J., M. Egenhofer and A. Frank (1995). On the Robustness of Qualitative Distance- and Direction-Reasoning. In Proceedings of Auto-Carto 12, Charlotte, North Carolina, pp. 301-310, (eds).

[8] Cycorp, Inc. OpenCYC Ontology in OWL. http://www.cyc.com/2003/

[9] GeoOnion. http://esw.w3.org/topic/GeoOnion

[10] Jones, C., H. Alani and D. Tudhope. Geographical Information Retrieval with Ontologies of Place. In Proceedings of COSIT'01, pp. 322-335, 2001