Establishment of Spatio-Temporal Models of Geographic Entities Behavior

Research Team

Marinos Kavouras, *Professor, National Technical University of Athens, Greece*

Eleni Tomai *(Research Student), National Technical University of Athens, Greece*

George Panopoulos *(Research Student), National Technical University of Athens, Greece*

Introduction

The rationale behind the current research program was the development and generation of an ontology of geographic objects taking into consideration their behavior in time. The stages of the research comprised of: 1) the resolution of theoretical questions and the study of the characteristics/ attributes of geographic entities, 2) the development of a methodology for the generation of a spatio-temporal geographic ontology, and finally 3) the developments of a model for the space-time behavior of geographic entities

First step of the research

The first step comprises of resolving several theoretical issues that are of major importance in the spatio-temporal ontological research. These are: a) the use of theoretical tools for the development of a formal geospatial ontology, b) the definition of characteristics and attributes for the geographic entities and c) the investigation of the presence of vagueness in the geographic realm and the implications it imposes on the behavior of the entities in question.

A) Theoretical tools for geographic representation. In literature [1], three basic theoretical tools are recognized for the development of a formal ontological theory of spatial representation: mereology, location and topology.

B) Characteristics and attributes of the geographic entities. Those attributes that define semantically a geographic entity are called essential properties or rigid properties. The term attribute in this case is also used with its philosophical meaning (e.g. the attribute of something being a mountain), and with the aspect of a characteristic.

C) Vagueness in the geographic realm. Many geographic concepts are considered to be affected by soritical statements [4] and therefore present vagueness. Apart from the philosophical issues presented about vagueness, we confront two pragmatic issues when trying to establish a geographic ontology; regions that exhibit vague boundaries, and vague categories of geographic entities.

Second step of the research

The second step was dedicated to the development of a methodology for the generation of a spatio-temporal ontology.

The first action was to look into existent categorization schema of geographic information to find out which are their characteristics, their qualities and their drawbacks. An experiment was conducted for the semantic comparison of these schemas and for the visualization of the results through a spatialization method.

The schemas that were used for the experiment are:
1.   CORINE LC [2]
2.   GDDD- (Geographical Data Description Directory - GDDD MEGRIN) [5]
3.   WordNet [11]

The purpose of the present research was to identify semantic information from definitions and to enrich the representation of categories with semantic properties and relations, such as those reported by [6]), in order to disambiguate geographic categories. The ability to represent and visualize the degree of semantic similarity with concept mapping tools [9], [10] greatly facilitated the entire process.

This formalized semantic information was further used to disambiguate similar categories by explicitly and objectively identifying similarities and heterogeneities between them

Ontology	Category
CORINE LC	Peat bog, Water course, Water body, Salt marsh, Saline, Intertidal flat, Coastal lagoon, Estuary, Sea and ocean, Inland marsh
MEGRIN	Bog, Canal, Lake/ pond, Salt marsh, Salt pan, Watercourse
WordNet	Body of water, Bog, Canal, Lake, Pond, Salt pan, Watercourse, Marsh, Estuary, Sea, Ocean, Lagoon

Table 1. The categories used for the experiment

In order to determine the similarity between two categories, we take into account the values of the properties/relations they possess. If the values of a given semantic property or relation coincide then the two category types are similar in terms of that property/relation. If the values of a property/relation are distinct, then similarity between the two categories is equal to zero.

The similarity measure S between two categories a, b, is set by the ratio model (based on Tversky's similarity measure):

$$S(a,b) = \frac{C}{A + B + C} \tag{1}$$

where, C is the number of properties/relations which categories a and b share, but also exhibit common values for, A is the number of properties/relations of category a but not of b, and B is the number of properties of category b but not of a. As it can be understood, the ratio is bounded between 0 and 1, the former denoting complete dissimilarity, the latter, coincidence of entities.

In order to visualize the different ontologies we use Multi-Dimensional Scaling (MDS) [7]. The method uses a similarity/dissimilarity matrix to project the data into the projection space, which in our case, is a two dimensional space. MDS is a dimensionality reduction method that represents multi-dimensional data sets by using a stress function; therefore, distances among data reflect the corresponding (dis) similarities. The value of the stress function is an indicator of the goodness-of-fit of the result. The higher its value the more the distortion imposed on the visualization of the entities; therefore, distances are greater than the corresponding dissimilarities. The output is a scatter plot of the data where similar entities are close in the representation space while dissimilar ones are far away. The visualization result is shown in Figure 1.

Figure 1. Resulting clusters showing the heterogeneities among same category_types (terms) in different ontologies

The second action was based on the results of the previous research stage to develop a methodology for generating robust geographic ontologies.

On the issue of what should be contained in an ontology, by way of fully conveying the intended conceptualization, Maedche and Staab [8] advocate that an ontology should comprise the following: a) The Lexicon, b) Concepts, c) Relations, and d) Axioms (figure 2). Adopting this premise, we attempted to investigate the role of these components when building a geographic ontology.

Figure 2 Interrelations among elements of a domain ontology

A geographic ontology should cover the variety of geographic concepts in the world. Domain ontologies are often context-driven, that is, their concepts share the conceptualization relevant to the context tackled by the ontology. For instance, a land cover ontology would not contain geographic concepts such as country or communication networks.

A) Concepts. A right selection of concepts in most domain (or task) ontologies is not an issue, since scientific knowledge about the given domain recognizes concepts of interest. In [3], it is argued that a hermeneutic analysis should guide communication between experts in the design of information system ontologies. In the geographic domain, however, the debate is still open of what experts reckon to be a geographic concept due to the diversity of the domain. Thus, the first step in building a geographic ontology is to select the concepts to be included.

B) Relations. By the term "relation" we refer to the following:

- Relations as semantic relations. hese, reflect the relations among concepts at the lexical semantic level. Examples of this kind of relations are hypernym/hyponym, meronym/ holonym. These are extensively met in terminological ontologies such as WordNet. Hypernym accounts for the "is-a"/"kind-of" relation. Subsequently, it can provide the superordinate concept of the initial one due to its subsuming mechanism. At the very opposite, the hyponym accounts for the same relation providing the subordinate concepts of the initial concept. Both relations offer a tool for building hierarchies of concepts (taxonomies). Both meronym and holonym reflect the notion of parthood. Meronym expresses parthood between the initial concept and its parts, and it is depicted by the "has-part" lexical pattern. On the other hand, holonym expresses parthood between the initial concept and the whole that it is part of, and it is expressed by the "is-part-of" lexical pattern. Both relations offer a tool for building partonomies.

- Relations as semantic properties. These, refer to the "properties" of the concepts in the ontology and the "values" they can take. Properties are very difficult to attribute to concepts.

- Relations among relations. These, intend to build a taxonomy of relations themselves. Relations among semantic properties and semantic relations will assist in fully determining and explicitly establishing the semantics of the domain in question.

C) Axioms in a geographic ontology. An exhaustive domain ontology, should contain axioms that consider: the structuring of the concepts and relations, their meanings and constraints with regard to the domain itself, and the laws that make definitions of concepts and relations consistent and complete.
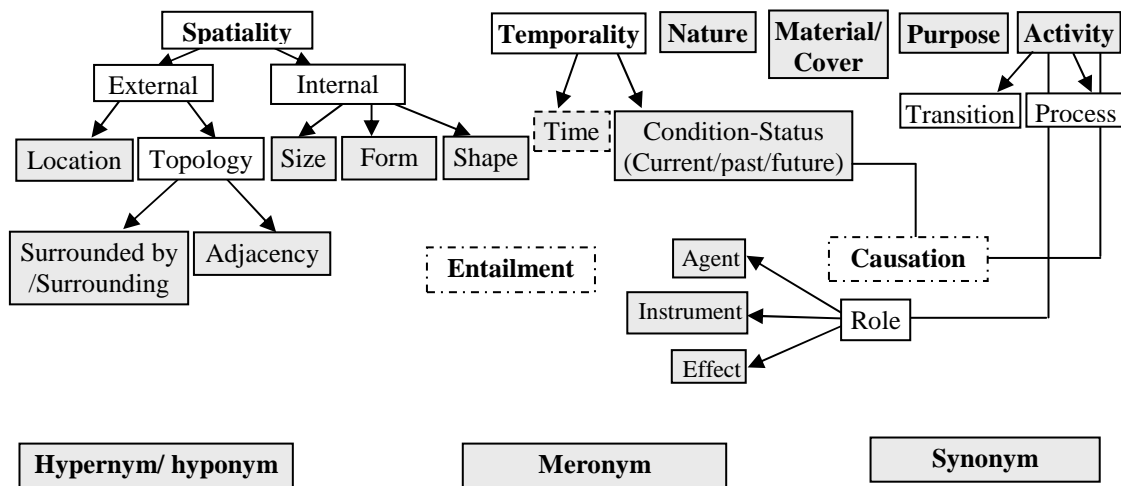
Figure 3. Relations among relations in a geographic ontology

D) The Lexicon. The contents of the lexicon in a domain ontology are the descriptions of concepts in Natural Language, the documentation of the SRs/ SPs, and their values for each concept and finally the list of axioms. What is more, the lexicon should include data not related to the ontology itself, but additional information about the SRs/SPs. This includes the lexical and syntactic patterns, in the definitions of the concepts that make them identifiable, by NLP techniques. In addition, the lexicon should include information on how the ontology should be implemented, that is, information on the ontology language, on the algebra of the axioms, etc.

Third step of the research

The last step of the program was to establish a model for the behavior of geographic entities through space and time. We base our model in previous attempts of modeling time. Our model comprises of the following:

1. We define the entity as a abstract depiction of the phenomena of real world
2. The course of entity is described by the various versions of their objects in the system.
3. The model contains standardized rules for each entity, which determine the type or the percentage of change
4. We used the object oriented approach for the determination of structures of data.
5. The objects share the descriptive attributes of entity that are separated in three distinguishable descriptive regions: the Thematic domain, in which belong the thematic characteristics of object, the Spatial domain, where is described the geometric characteristics and the location of each object and the Temporal domain, which portrays the time information of the entity.
6. The time which is recorded in these "imprints" is "Valid Time", that is to say, the real time in which the change takes place.
7. The spatial characteristics of each entity change through time and the notion of moving objects is used to describe them.

Conclusions

The research attempted a systematic approach to designing geographic ontologies, which address the issue of completeness. This approach applies NLP techniques for determining the proper-for-inclusion concepts in the ontology, and for assessing semantic relations and semantic properties of these concepts. In addition, relations among relations are an integral part of a geographic ontology; these second-order relations deal with the complexity of the semantics of the particular domain. As for designing a comprehensive geographic ontology, the focus should also be on ascertaining axioms, which secure the semantics of the domain. Also we developed a model for describing change in time of as geographic entity, using the object oriented approach and the notion of moving objects.

References

1. **Casati R., Smith B. and Varzi A., "Ontological Tools for Geographic Information",** *Formal Ontology in Information Systems*, Amsterdam: IOS Press, **1998.**

2. European Environmental Agency: **CORINE Land Cover Methodology and Nomenclature, 2000**. http://reports.eea.eu.int/COR0-part2/en.

3. **Fonseca F., M. Egenhofer, Agouris P. and Camara G., "Using Ontologies for Integrated Geographic Information Systems"**, *Transactions in GIS* **Vol. 6-3: 231-257, 2002.**

4. **Fisher P., Sorites Paradox and Vague Geographies,** *Fuzzy Sets and Systems,* **113: 7-18, 2000.**

5. **Geographical Data Description Directory (GDDD): The European Dataset Catalogue**, http://www.eurogeographics.org/gddd/lists/features.htm, **1994**.

6. **Kokla M., Kavouras M., "Extracting Latent Semantic Relations from Definitions to Disambiguate Geographic Ontologies",** *GIScience 2002 Abstracts*, **87-90, 2002.**

7. **Kruskal J. B., Wish M., "Multidimensional Scaling",** *Sage University Paper series on Quantitative Applications in the Social Sciences*, number 07-011. Sage Publications, Newbury Park, CA, **pp.96, 1978.**

8. **Maedche A., Staab S., "Ontology Learning for the Semantic Web",** *IEEE Intelligent Systems,* **16-2: 72-79,2001**

9. **Skupin, A., "A Cartographic Approach to Visualizing Conference Abstracts",** *IEEE Computer Graphics and Applications*. 22, 50 – 58. 2002

10. **Tomai E. Kavouras M., "Sharpening" Vagueness: Identifying, Measuring, and Portraying its Impact on Geographic Categorizations"**, *GIScience 2002 Abstracts*, **189 -192, 2002.**

11. **WORDNET 1.7.1** - a Lexical Database for the English Language, Cognitive Science Laboratory, Princeton University. http://www.cogsci.princeton.edu/~wn/