Geographical Objects Representation by Means of Spatial Ontologies

M. Torres, S. Levachkine, M. Moreno, and R. Quintero

Geoprocessing Laboratory-Centre for Computing Research-National Polytechnic Institute, Mexico City, Mexico
{mtorres, sergei, marcomoreno, quintero}@cic.ipn.mx
http://geo.cic.ipn.mx

Abstract. In this work, we generate spatial ontologies for improving the geographical objects representation in spatial databases. The process to generate spatial ontologies consists of the interaction between a spatial subject domain and a spatial taxonomy. Spatial subject domain is composed by *a priori* spatial knowledge, which is related to the "essential properties" (spatial data semantics) of the geographical objects. On the other hand, spatial taxonomy describes the classification among the spatial data, according to the primitives of the spatial objects representation. Using such method, it is possible to improve the spatial data representation establishing the geo-information conceptualization that results useful for subsequent spatial data processing and interpretation. The main goal of this work is to partially solve the ambiguities and heterogeneity in the spatial data by means of spatial ontologies Thus, this method of spatial data semantic processing can be considered as a knowledge discovery into the spatial databases.

Keywords: Spatial Semantics, Spatial Ontology, Spatial Subject Domain, Spatial Taxonomy, Spatial Databases.

1 Introduction

Nowadays, the spatial databases are very useful and powerful tools to handle, display, and process the geographical information. These databases integrate Geographic Information Systems (GIS), which are composed to store and process spatial data. To solve some ambiguities in the spatial data processing and the interpretation, the geo-information should have good quality from the input to the representation. The "adequate" representation of spatial data is crucial for improving the decision making in different environments [1].

In this paper, we generate spatial ontologies based on the spatial semantics, which can be used to represent geographical objects by means of spatial concepts ("not words"). Such spatial data conceptualization aim to compress the data and facilitate the knowledge discovery into spatial databases.

Up-to-date GIS do not extensively explore the spatial data semantics. To develop a spatial semantic theory is a great challenge in the new trends of Geocomputation field. Thus, the spatial analysis can use alternative methods to represent spatial data: this data representation jointly with the semantic rules - both based on data semantics - can be stored in a knowledge-base to generate new concepts that form the spatial ontologies. These concepts are defined by

the properties and behavior of geographical objects and explored by the human experience. In general, we seek to represent correctly spatial objects for their subsequent processing [2].
Several works related to semantic interoperability have been published. In particular [3] presents an approach to semantic similarity assessment combining two different strategies: feature-matching process and semantic distance computation.
An approach proposed in [4] has been achieved to improve the spatial data handling. As described in this work, LOBSTER system combines the artificial intelligence techniques to provide a query language more flexible and powerful than standard SQL. This system has been developed in Prolog for object-oriented modeling, geomorphology, and query optimization.
In [5] and [6] an ontology-driven GIS as a system integrator has been proposed. In these works, a special model to conceptualize the geographical information and solve problems related to the integration and interoperability in GIS of different types at different levels of detail has been described.
Nowadays, the interoperability in GIS is approached by using the spatial semantics representation. This interoperability is based on the integration of spatial schemas, query languages and sets of semantic rules, which can provide knowledge of data and geographical representation interfaces.
A peculiarity of our proposal is the definition of spatial semantics, which is based on intrinsic properties of the geographical objects. These properties integrate a knowledge-base, which can be represented by concepts that form the spatial ontologies. In other words to generate spatial ontologies, we define a data description that is composed of semantic properties of the spatial data to integrate the spatial subject domain. Moreover, the spatial data are classified by spatial taxonomies. The last component is interacted with the subject domain to generate ontologies. In particular, an improvement of representation of the spatial databases can be achieved by using our approach.
This proposal consists of generating ontology levels, which can be defined by top or down ontologies according to the particularization of the concepts. For this purpose, we consider 2D partition of a map that is composed by a set of primitives of representation. Such maps are stored in spatial databases. However, the approach is focused on describing the semantics of a particular map in two dimensions, which is formed by geographical objects. The approach attempts to find out the spatial semantics throughout spatial ontologies. Using this approach, we seek to enrich the concepts by means of the "interaction" and to improve the spatial data representation.
The rest of the paper is organized as follows. In Section 2 we describe a spatial semantics definition, which is proposed to use in spatial databases. The spatial subject domain and the spatial taxonomy are described in Section 3. Section 4 presents the spatial ontology definition and the interaction approach to generate it. Section 5 contains some preliminary results related to the spatial semantics definition. Our conclusions are sketched out Section 6.

2 Spatial Semantics Definition

To define spatial semantics, we use the essential characteristics that involve the spatial data. Our definition is based on providing a set of rules. This set is composed of relationships, properties, functions and behaviors, which define the characteristics of the geographical objects [2]. Fig. 1 shows that the spatial semantics definition starts with a spatial description of the geographical objects. This description consists of topological, logical, attribute and spatial distribution properties, which define conceptually spatial semantic rules. We propose the spatial semantics definition in [2], where 2D spatial objects system is defined by a set of primitives of representation, which form a partition of a map. This partition is composed of basic primitives of representation such as point, line and polygon objects.

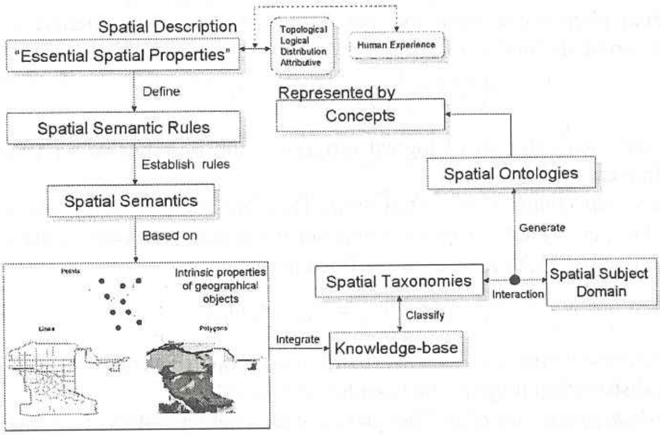

Fig. 1. General schema of the Spatial Semantic Analysis

Let P be the spatial partition and U_i the universe of geographical objects, which consists of a set of representation primitives (lines, points or polygons). P is the set of partitions of the representation primitives that can exist in that partition P and presented in the same partition, as shown in (1):

$$P = U_i \cup \{Rp_l \vee Rp_p \vee Rp_a\} \text{ i=1,…,n,} \tag{1}$$

where: Rp_l is the representation primitive "lines", Rp_p is the representation primitive "points", Rp_a is the representation primitive "polygons" and i represents the thematic number that involves the spatial partition.

In all cases, these sets are associated with the geographical objects, which are denoted by f_G. We also denote: O_T that represents a set of operations related to the topological properties such as: meet, contain, cover, overlap, etc. O_T represents a set of operations based on the logical constraints, for example: the behavior of the land is considered to build a road in a specific zone, O_D represents a set of functions that belongs to the spatial distribution properties, and O_A represents a set of the characteristics associated to the geographical objects such as: elevation, names, area, direction, etc. The properties considered for this definition can be described as follows:

A. Topological properties (P_T). These compose the spatial object layers. They represent the topologic and geometric shapes and the spatial relationships, keeping the consistency as well as the congruency of the geographical structures. Therefore P_T properties can be defined as is shown in (2):

$$P_T = \{\forall f_G \mid \exists p_t \in T, p_t \subset O_T\}, \tag{2}$$

where: T represents the set of topological properties and p_t indicates a particular topological property that can be considered.

B. Logical properties (P_L). They provide rules that define the behavior, relationships and combined properties of the spatial structures. These characteristics are focused on the description of the phenomena, and can be used to simulate and model the geographical phenomena behavior, based on the human experience. Therefore P_L properties can be defined as (3):

$$P_L = \{\forall f_G \mid \exists p_l \in L, p_l \subset O_L\}, \tag{3}$$

where: L represents the set of logical properties and p_l indicates a particular logical property that can be considered.

C. Spatial distribution properties (P_D). They are used to locate any geographical object in the space. These properties consider the spatial reference of the geographical systems. Therefore P_D properties are defined in (4):

$$P_D = \{\forall f_G \mid \exists p_d \in D, p_d \subset O_D\}, \tag{4}$$

where: D represents the set of spatial distribution properties and p_d indicates a particular spatial distribution property that can be considered.

D. Attribute properties (P_A). They describe the characteristics about a phenomenon. It is necessary to consider *qualitative* and *quantitative* attributes, because these specific attributes support the knowledge of the similarities between the geographical objects. Therefore P_A properties are defined in (5):

$$P_A = \{\forall f_G \mid \exists p_A \in D, p_a \subset O_A\}, \tag{5}$$

where: A represents the set of attribute properties and p_a indicates a particular attribute property, which involves the geographical object.

Therefore, the initial exploration to define *a priori* and intuitive spatial semantics (S_E) is the description of the content of all these properties and relations that are involved in any spatial partition P. See equation (6).

$$S_E = content\left(\bigcup_{i=1}^{n} U_i \cup \{P_T, P_L, P_D, P_A\}\right),^1 \tag{6}$$

All the properties are considered in the object systems of a finite space of geographical objects. Therefore, the content of the set of spatial rules reflect the topological, logical, spatial distribution and attribute properties of the spatial data. In general, this definition looks for improving the spatial data representation for their subsequent processing [2]. However, we consider the *content* of all these properties and relations, the spatial semantics is defined by the *interaction* (between the spatial taxonomy and the spatial subject domain). Moreover, *iterative process* offers the possibility to feedback the approach to generate the spatial ontologies again in the following interaction. Each new defined concept is put into the spatial subject domain. Then the interaction process is repeated-up to no new concepts found. The resulting spatial ontology represents the *semantics* of the spatial object system.

---

1 The concept of "*content*" will be particularized in the following sections in dependence on application.

In sense, the spatial semantics relates the content and the entities of spatial data to the real world. The main purpose is to identify the characteristics of geographical objects, which belong to different spatial databases and adapt them to the resolution level of the geo-information [7].

3 Spatial Subject Domain and Spatial Taxonomy

In our approach, the interaction between the subject domain and the taxonomy generates the "ontologies", according to the spatial semantic rules. These rules are composed of the "essential characteristics" of the geographical objects.

3.1 Spatial Subject Domain

Spatial Subject Domain is defined as a set of "names" that describe the primitives of spatial representation. Thus, we can start with *a priori* knowledge of the geographical objects that appear, e.g. in the map legend. For instance, "blue" lines are united under the concept (name) "river" and "black" lines are united under the concept "fracture", etc. In reverse, the different concepts are united under the same description of the spatial representation that is *"line"*. The interaction between the subject domain and the taxonomy is used to locate concepts into the spatial subject domain that belong to a case of study, and to process these concepts to generate spatial ontologies [2].

In our approach, it is important to count on *a priori* knowledge that is defined by the user depending on the case of study. It is necessary to elaborate a description of the thematic to be analyzed, considering the main elements that compose this theme, such as the data model and the resolution levels of the spatial information. All the characteristics, which are considered in spatial description of the spatial subject domain, should represent relationships between themselves too. Spatial subject domain has to recognize the different semantic levels of *a priori* knowledge that is stored in this domain. Fig. 2 shows the definition of spatial subject domain and its interactions and a brief description for the topographic data is presented in the following schema.

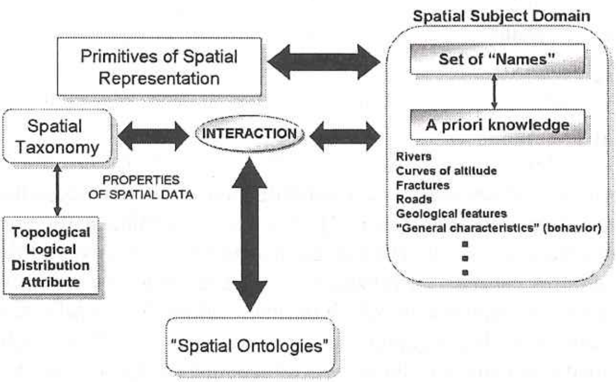

Fig. 2. Interaction of the Spatial Subject Domain

```
<Spatial Data Description>
 <Layer_Name:Topo>
 <Thematic:Topography>
 <Type:Line>
 <Geographical_Properties>
  <Projection:UTM 14>
  <Datum:NAR_D>
  <Units:METERS>
  <Spheroid:GRS1980>
  <Scale:1:50000>
 <Geographical_Features>
  <Type of airports>
   <Values domain:
          International
          National
          Local>
  <Population>
   <Values domain:
          State:
                  Low inhabitants
                  Medium inhabitants
                  Large inhabitants
          City:
                  Low inhabitants
                  Medium inhabitants
                  Large inhabitants
          County:
                  Low inhabitants
                  Medium inhabitants
                  Large inhabitants>
  <Type of roads>
   <Values domain:
          Pavement:
                  One rail
                  Two rails
                  Three rails
                  Four rails
                  More than four rails
          Unpavement
          Breach
          Path>

  • • •
```

3.2 Spatial Taxonomy

We define a *spatial taxonomy* as a classification method to describe every spatial entity (primitive of representation) [2]. The use of spatial taxonomies is served to classify and organize *a priori* knowledge in a hierarchical way. In this classification, the spatial entities can be generalized according to some spatial semantic rules. We consider a subset of spatial data, which is generated by the spatial taxonomy for every spatial thematic layer that is stored into the spatial database. By using this schema, we define the spatial semantics of the geo-information. This method allows processing the geographical information according to the semantics of the geographical objects. Fig. 3 depicts the spatial taxonomy and its interactions.

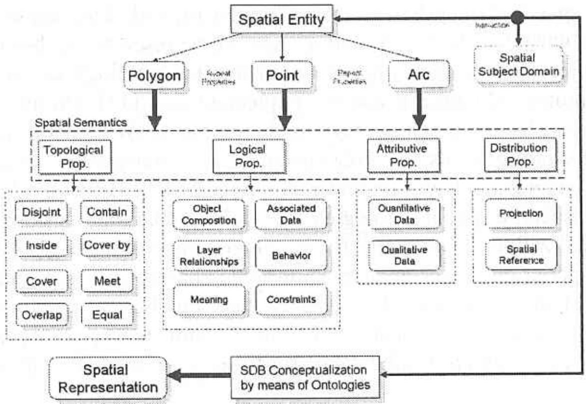

Fig. 3. Spatial Taxonomy and its interaction with the Spatial Subject Domain

To generate the spatial taxonomy, we establish spatial semantic rules. They are defined by the properties of the spatial data (Section 2). The acquired knowledge represents other characteristics of the spatial data, which are used to partially automate the spatial representation. This representation sometimes calls for the human intervention to solve problems related to the *ambiguity* between different contexts.

4 Spatial Ontology Generation

Most widely accepted common conceptualization of the geographical data is based on the description of geographical objects and fields [7]. The spatial semantics definition (Section 2) is proposed and aimed to correctly represent spatial data in an alternative and universal way to generate spatial ontologies.

We consider a spatial ontology as an explicit, shared and structured specification of conceptualization of geographical objects. The spatial ontologies represent definitions, functions, attributes and relationships of geographical objects by means of "concepts" [2].

Besides that, ontologies can be considered as "languages", which use a specific vocabulary to describe entities, classes, properties and functions related to a certain view of the geographical world [8].

Our approach is designed to solve ambiguities that can exist with single characteristics of the geographical objects, due to the spatial ontology is defined by concepts (not by words) of the geographical objects. Spatial ontologies can be classified by levels according to their dependence on a specific task or point of view.

These levels are generated for a specific spatial ontology (*top-ontology*) and it can be particularized to define a particular ontology (*down-ontology*). There are also different levels of information detail. Low-level ontologies correspond to very detailed information and high-level ontologies belong to more general information. The generation of more detailed ontologies should be based on the high-level ontologies, such that each new ontology level incorporates the knowledge presented in the higher level. These new ontologies are more detailed, because they refine general descriptions of the level from which they have been generated [9, 10]. The use of explicit spatial ontologies contributes to better correct spatial representation, because every geographical object description is based on an implicit ontology to avoid *conflicts* between the ontological concepts and the implementation [11]. On the other hand, spatial ontologies play an essential role in the conceptualization of spatial databases, allowing the establishments of correspondences and interrelations among the different domains of geographical objects and relations [12, 13]. By using this approach, we can generate specific spatial ontologies after defining the top-ontology to particularize the conceptualization in other specific ontologies (down-ontologies). The use of ontologies in spatial databases enables knowledge sharing and information-integration. The proposed approach provides dynamic and flexible information exchange and allows partial integration of spatial data when completeness is impossible. This can help the next generation of spatial databases to solve *semantic ambiguities.*

5 Preliminary Results

The following results illustrate the interaction between spatial subject domain and spatial taxonomy, to obtain spatial ontologies by using the spatial semantics definition. Some tests have been made to prove our approach. The system has been partially implemented in *C Builder* to facilitate the spatial ontology generation. In Fig. 4 the concept *"Limit"* of spatial ontology is generated. This spatial ontology presents several "sub-concepts" related to *"Limit"*.

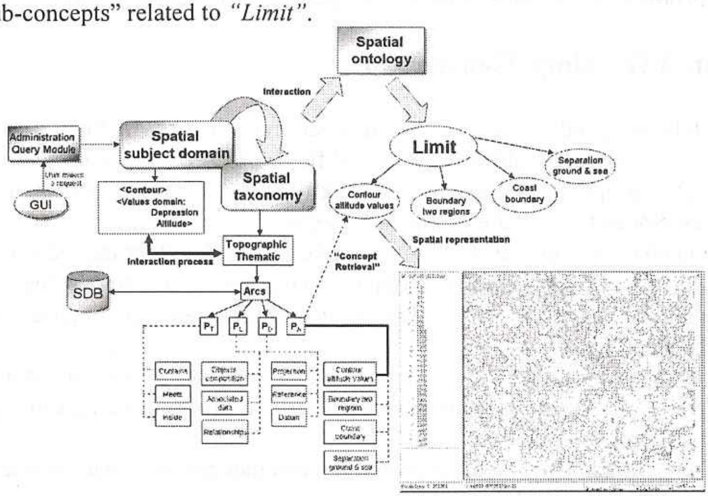

Fig. 4. The concept "Limit" in spatial ontology that represents *contours*

The query functionality is the following: the user performs a request by means of a GIS-application, which sends it to the administration query module for processing the request to obtain the spatial ontology. Inside GIS-application *a priori* knowledge that is stored in the spatial subject domain interacts with the spatial taxonomy, considering in this case, the *"arcs"* as primitive of representation.

Fig. 5 shows the query mechanism to describe the concept "Roads" into the spatial ontology. In this case, the ontology is composed of several "sub-concepts", which are ordered in a hierarchical way. Also, Fig. 5 depicts different levels of concepts, starting with a *top-level* (Roads) and finishing with *down-levels* (One rail, two rails, etc.). When the ontologies provide more levels of concepts, it is possible to particularize these in sub-concepts while the level is less (down-level), the concept is more particular.

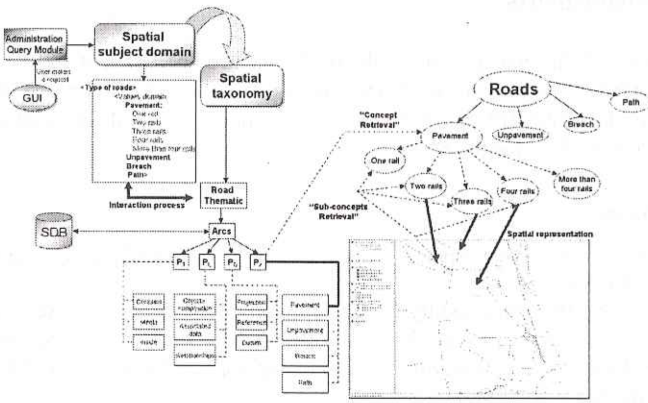

Fig. 5. The concept "Roads" in spatial ontology that represents *pavement roads* with one, two, three and four rails

6 Conclusions

In the present work, the spatial semantics definition to represent geographical objects has been sketched. This definition is developed considering the intrinsic properties of spatial data, which are defined by the basic spatial representation primitives.

By using our method, it is possible to classify geographical objects, generating a *spatial taxonomy*. The proposed definition of spatial subject domain is oriented towards an interaction with spatial taxonomy to conceptualize the spatial databases.

Moreover, we presented a spatial data description that is based on a set of spatial semantic rules, obtained from the spatial object analysis. By obtaining this set of rules, the request of the user evaluates and filters the geo-information according to the semantic criteria. We attempt to show an alternative approach to represent spatial data considering their spatial semantic properties, besides that, it can act as a system to integrate geo-information at different levels of detail.

The spatial ontologies represent real world entities using a hierarchical structure, which is composed of "concepts". These concepts can be formed by definitions, functions, spatial properties, rules, relationships that constitute the *spatial semantic analysis*.

In addition, the spatial ontologies catch the *semantics* of geo-data to provide additional information related to the concepts. Also, ontologies can be used to establish agreements about diverse views of the world and consequently carry out the *"meaning"* of the geo-information. Use of ontologies in spatial databases enables knowledge sharing and information-integration. The proposed approach provides dynamic and flexible information exchange and allows partial integration of spatial data when completeness is impossible. This approach can aid to solve *semantic ambiguities* between the available geo-information, because the context of the spatial data can change, according to the case of study or for the representation state by means of concepts of the spatial data.

Acknowledgments

The authors of this paper wish to thank the Centre for Computing Research (CIC-IPN), General Coordination of Postgraduate Study and Research (CGPI), National Polytechnic Institute (IPN) and the Mexican National Council for Science and Technology (CONACYT) for their support.

References

1. Goodchild, M.F.: Geographical Data Modeling. *Computers and Geoscience,* Vol. 18 (1992) 401-408.
2. Torres, M. and Levachkine, S.: Semantics Definition to Represent Spatial Data, In: Levachkine S., Bodanksky E. and Ruas A., (eds.), *e-Proceedings of International Workshop on Semantic Processing of Spatial Data (GEOPRO 2002)*, Mexico City, Mexico (2002).
3. M.A. Rodríguez, M. Egenhofer & R.D. Rugg, Assessing Semantic Similarities Among Geospatial Feature Class Definitions, *Lecture Notes in Computer Science Vol. 1580*, 1999, 189-202.
4. M. Egenhofer & A.U. Frank, LOBSTER: Combining AI and Database Techniques for GIS, *International Journal of Photogrammetric Engineering and Remote Sensing, 56*(6), 1997, 919-926.
5. F. Fonseca & M. Egenhofer, Ontology-Driven Geographic Information Systems, *7th ACM Symposium on Advances in Geographic Information Systems*, Kansas City, MO, 1999, 14-19.
6. Fonseca, F., Egenhofer, M. and Agouris, P.: Using Ontologies for Integrated Geographic Information Systems. *Transactions in GIS*, Vol. 6, No. 3 (2002) 25-40.
7. Fonseca, F., Egenhofer, M. and Agouris, P.: Using Ontologies for Integrated Geographic Information Systems. *Transactions in GIS*, Vol.6, No. 3 (2002) 25-40.
8. Guarino, N.: Formal Ontology, Conceptual Analysis and Knowledge Representation. *International Journal of Human and Computer Studies*, 43, Vol. 5, No. 6 (1999) 625-640.
9. Guarino, N.: Formal Ontology and Information Systems, *in Guarino (Ed.), Formal Ontology in Information Systems*, Editorial: IOS Press (1998) 3-15.

10. Rodríguez, M.A., Egenhofer, M. and Rugg, R.D.: Assessing Semantic Similarities Among Geospatial Feature Class Definitions. *Lecture Notes in Computer Science*, Springer-Verlag, Vol. 1580 (1999) 189-202.
11. Egenhofer, M. and Frank, A.U.: LOBSTER: Combining AI and Database Techniques for GIS. *International Journal of Photogrammetric Engineering and Remote Sensing*, Vol. 56, No.6 (1997) 919-926.
12. Egenhofer, M. and Frank, A.: Naive Geography, in Frank A. and Kuhn W., (Eds.) Spatial Information Theory, A Theoretical Basis for GIS, *Proceedings of the International Conference COSIT '95*, Vol. 988, Lecture Notes in Computer Science, Springer-Verlag, Berlin (1995) 1-15.
13. Gooldchild, M.F., Egenhofer, M., Fegeas, R. and Kottman, C. *Interoperating Geographic Information Systems*, Editorial: Kluwer Academic Publishers (1999).