Deriving Semantic Description Using Conceptual Schemas Embedded into a Geographic Context

Geoprocessing Laboratory – Centre for Computing Research – IPN, Mexico City, MEXICO

Abstract. Information integration and semantic heterogeneity are not trivial tasks. An integrated view must be able to describe various heterogeneous data sources and its interrelation to obtain shared conceptualizations. In this paper, we propose an approach to make a conceptualization of the real world based on conceptual schemas, which are used to generate a semantic description of the geospatial domain. This description attempts to provide the guidelines to formalize the geographic domain in form of geospatial ontologies according to specific contexts. In this case, we propose that conceptual schemas are built in order to abstract specific and essential *parts* of the geospatial domain and to represent schematically which geospatial entities should be collected and how they must be organized. We perceive that geographic data modeling requires models that are more specific and capable of capturing the *semantics* of geospatial data, offering higher abstraction mechanisms and implementation independence. Therefore, we approach conceptual schemas to describe the contents of the real world abstraction to specify the behavior of the geospatial entities, in which the context plays an important role to guarantee shared and explicit conceptualizations. Our research is mainly oriented to propose an approach related to conceptual issues concerning what would be required to establish ontologies of the geospatial domain. In addition, the work is led to formalize appropriate methods to represent ontologies of the geospatial domain.

1 Introduction

Ontology has gained increased attention among researchers in geographic information science in recent years. Up-to-date, the ontology notion plays an important role in establishing robust theoretical foundations for geographic information science [1]. Under this umbrella, it is possible to unify several interrelated research subfields, each of which deals with different perspectives on geospatial ontologies and their roles in geographic information science. Three broad sets of foundational issues need to be resolved: (1) conceptual issues concerning what would be required to establish an exhaustive ontology of the geospatial domain, (2) representational and logical issues relating to the choice of appropriate methods for formalizing ontologies, and (3) issues of implementation regarding the ways in which ontology ought to influence the design of information systems.

Nowadays, there are diverse institutions that use geospatial data to make a decision in different environments. The use of geographic databases through geographic information systems (GIS) provides tools for managing, analyzing and processing geospatial data. However, information can not be sometimes represented in “adequate” way, since it presents ambiguities that do not allow the appropriate use and analysis. These ambiguities are originated by imprecision of information, heterogeneity and isolation sources. Whereby, it is difficult to develop interoperable applications that allow us to share, integrate and represent geospatial information.

These facts bear with searching solutions oriented to geospatial data representation and integration, semantic heterogeneity and imprecise geographic objects issues. Consequently, commercial GISs do not have tools to extensibly explore the essential properties and relations of geographic objects. Therefore, by means of these applications, it is difficult to explore the semantics of a set of geographic objects.

According to [2] and [3], the ontologies and the knowledge representation are essential for the creation and the use of standards to exchange data, as well as for the design of human computer interaction. Whereby an ontology allows us to solve problems associated to heterogeneity, interoperability, representation, integration and exchange of geospatial data. These problems imply incompatibility between diverse geographic objects, as well as a different spatial conceptualization according to a specific context. For example, we engage with the world day by day in a variety of different ways: we use maps, specialized languages, and scientific instruments; we also engage in making rituals and telling stories; we use information systems, databases, different machines and other software-driven devices of various types. Each of these ways of engaging with the world involves a certain *conceptualization*. What this means is that it involves a system of concepts and categories, which divide up the corresponding universe of discourse into objects, processes and relations in different sorts of ways. Thus, in a religious ritual setting, we might use concepts such as *God*, *salvation* and *sin*; in a scientific setting, we might use concepts such as *micron*, *force* and *nitrous oxide*; in a story-telling we might use concepts such as *magic spell*, *dungeon* and *witch*. These conceptualizations are often tacit, that is, they are often invisible components of our cognitive apparatus, which are not specified or thematized in any systematic way [4].

On the other hand, the conceptualization of geospatial domain is diverse, because the geospatial data used are often imprecise or many subjects have different point of view. Thereby it is important to consider alternative object representations, which are independent of the imprecise nature of the geospatial data [5].

Our research is mainly oriented to propose an approach related to conceptual issues concerning what would be required to establish ontologies of the geospatial domain.

In this paper, we propose an approach to make a conceptualization of the real world based on conceptual schemas, which are used to generate a semantic description of the geospatial domain. This description can provide the guidelines to formalize this domain in form of geospatial ontologies according to specific contexts.

The rest of the paper is organized as follows. Section 2 describes some related works. Section 3 describes the proposed approach to conceptualize the geospatial domain. Section 4 shows a case study to build a semantic description based on conceptual schemas. Our conclusions and future works are outlined in Section 5.

2 Related works

Some works related to ontologies and semantics in geospatial information science to be mentioned are as follows. Guarino [6] coined the term "ontology-driven information systems" and provided a broad discussion on their place in the computer and information science. Gruber, one of the pioneers of the use of ontological methods in information science, defines an ontology as "a specification of a conceptualization" in [7].

Smith *et al.* [4] reported the results of a series of experiments designed to establish how non-expert subjects conceptualize geospatial phenomena. Subjects were asked to give examples of geographical categories in response to a series of differently phrased elicitations. The results yield an ontology of geographical categories – a catalogue of the prime geospatial concepts and categories shared in common by human subjects independently of their exposure to scientific geography.

Bishr *et al.* [8] argued that information modeling requires to be controlled to allow successful sharing of information. Also, they suggest that any coherent information model need to be based on accepted ontological foundation to guarantee unambiguous interpretation. In addition, their work attempts to show that ontology based information modeling provides more cognitive foundation for information systems models and therefore minimizes the problem of semantic heterogeneity.

Smith *et al.* [2] designed an ontology of geographic kinds to yield a better understanding of the structure of the geographic world, and to support the development of GIS that are conceptually sound. This work first demonstrated that geographical objects and kinds are not just larger versions of the everyday objects and kinds previously studied in cognitive science.

Fonseca *et al.* [9] proposed a framework to link the formal representation of semantics (i.e., ontologies) to conceptual schemas describing information stored in databases. The main result is a formal framework that explains the mapping between a spatial ontology and a geographic conceptual schema. The mapping of ontologies to conceptual schemas is made using three different levels of abstraction: formal, domain and application levels. At the formal level, highly abstract concepts are used to express the schema and the ontologies. At the domain level, the schema is regarded as an instance of a generic data model. At the application level, authors focus on the particular case of geographic applications. Additionally, they discuss the influence of ontologies in both the traditional and the geographic systems methodologies, with an emphasis on the conceptual design phase.

According to this works and in particular with [9], it is important to distinguish that our research is concentrated to use conceptual schemas to describe the semantic contents of the real world abstraction to specify the behavior of geospatial entities, in which the context plays an important role to guarantee shared and explicit conceptualizations. We will aim to propose issues and methods concerning what would be required to establish and to represent ontologies of the geospatial domain.

3 Geospatial domain conceptualization

This section gives the guidelines to build conceptual schemas to conceptualize the geospatial domain. Thus conceptual schemas are used to generate a semantic description, which can provide the framework to formalize the geospatial domain, according to specific contexts. In this section, we point out the most important components involved in our approach such as conceptual schema and context.

3.1 How to design conceptual schemas for geospatial domain

In the modeling approach, the modeler is required to capture a user's view of the real world in a formal conceptual model. Such an approach forces the modeler to mentally map concepts acquired from the real world to instances of abstractions available in his paradigm choice. On the other hand, the consolidation of concepts and knowledge represented by a conceptual schema can be useful in the initial steps of ontology construction. To adequately represent the geographic world, we must have computer representations capable not only of capturing descriptive attributes about its concepts, but also of describing the relations and properties of these concepts.

We propose conceptual schemas to describe the contents of the real world abstraction in order to specify the behavior of the geospatial entities. In this case, conceptual schemas certainly correspond to a level of knowledge formalization. In this case, conceptual schemas are built to abstract specific parts of the geospatial domain and to represent schematically which geographic entities should be collected and how they must be organized. We perceive that geographic data modeling requires models that are more specific and capable of capturing the *semantics* of geospatial data offering higher abstraction mechanisms and implementation independence.

The proposed conceptual schemas are composed of two types of concepts (C): *terminal* (C_T) and *non-terminal* (C_N). The first ones are concepts that do not use other concepts to define their meaning (they are defined by "simple values"). The meaning of non-terminal concepts is conceived by other concepts, which can be terminal or non-terminal concepts (see Eqn. 1).

$$C = C_N \bigcup C_T \tag{1}$$

Each concept has a set of *aspects.* Aspects are characteristics that describe the properties, relations and instances that involve the geospatial objects. From-now-on, we shall use the term "relation" to denote unary relations/properties as Berendt *et al.* [10]. From this point of view, all aspects of a terminal concept are simple, e.g. the type of all aspects that belong to the set of primitive types (punctual, linear and areal objects) is denoted by (T_P), as shown in Eqn. 2.

$$\begin{aligned} T_P &= \{number, character, string, enumeration, struct\}, \\ A &= \{a_i \mid type(a_i) \in T_P\}, \end{aligned} \tag{2}$$

where T_P is the set of primitive types; A is the set of aspects.

Then, the set of *terminal concepts* is defined by Eqn. 3.

$$C_T = \{c(a_1, a_2, ..., a_n) \ni a_i \in A,\ i = 1,..,n\} \quad \textbf{(3)}$$

In the same way, the *non-terminal concepts* have at least one aspect that does not belong to T_P. It is denoted by Eqn. 4.

$$C_N = \{c(a_1, a_2, ..., a_n) \ni \exists a_i \notin A\}, \text{ where } c \text{ is a concept.} \quad \textbf{(4)}$$

Finally, the set of relations R is defined by the pairs that are associated to Γ and Φ, in which Γ and Φ are non-reflexive, non-symmetric, and transitive relations (Eqn. 5).

$$R = R_\Gamma \cup R_\Phi = \{(a,b) \mid a\Gamma b,\ a \in C_N,\ b \in C\} \cup \{(a,b) \mid a\Phi b,\ a \in C_N,\ b \in C\} \quad \textbf{(5)}$$

According to definitions presented above, it is necessary to express the semantics that can provide a conceptual schema by means of a description D. Therefore, we consider the concepts C embedded into the conceptual schemas through geospatial objects, which are represented by primitive types as well as the set of relations R involved among geospatial objects (see Eqn. 6)

$$D = \langle C, R \rangle \quad \textbf{(6)}$$

Fig. 1 depicts a conceptual schema, which has been designed for the geospatial domain. Thus, this schema is adaptive for any context. In other words, it attempts to reflect the main features involved in this domain. For instance, if we have topographic, tourism, or geologic contexts, it is possible to describe the entities, characteristics and relations embedded between geographic objects. The main features involved into geospatial domain have been abstracted of the real world in order to obtain a conceptualization. This conceptualization provides us explicit vocabulary that represents the ontological commitment of the cognitive and intuitive perception of the subjects.

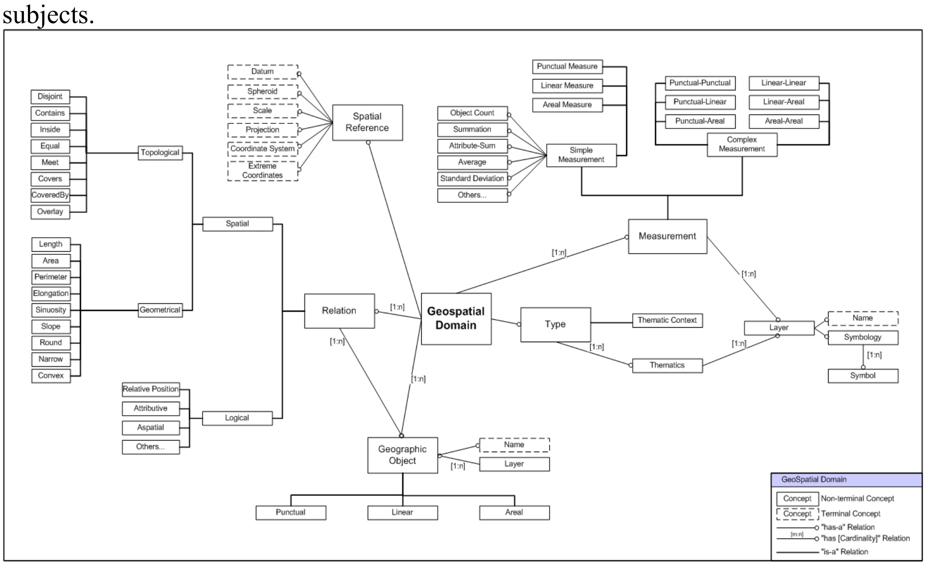

Fig. 1. Conceptual schema proposed to conceptualize the geospatial domain

The conceptual schema shown in Fig. 1 conceptualizes the geospatial domain. This schema represents a raw conceptualization, which contains an optimal number of relations. We are looking for a compact conceptual schema-based abstraction that drives the cognitive process of phenomenon semantic description under specific context. We consider that conceptual schema depicted in Fig. 1, could have more concepts involved in geospatial domain, thereby it can be a certain approximation about the main entities that compose this domain. This domain could be considered as a the first step to collect and organize the concepts contained into the geographic context.

3.2 Context-driven approach to restrict the domain into conceptual schemas

The context term is defined as *"that which surrounds, and gives meaning to something else*[1]*"* or it is the *"discourse that surrounds a language unit and helps to determine its interpretation*[2]*"*. However, to obtain shared conceptualizations and to accomplish with ontological commitments, it is necessary to take into account the *context* term. Also, this term can be used to consistently map different conceptualizations. Due to this, the meaning of a spatial concept may be dependent on a large number of contexts within which the concept is used.

Contexts about a particular use of a spatial concept refer to the knowledge that human uses to constrain the meaning of communication. To reach a common understanding of a vague concept, e.g. ***near***, the system and the user require to share knowledge about the relevant contexts that affect the understanding of the vague concept. Among many potential contexts that may affect how people understand spatial concepts. We focus on three of them: *task*, *spatial contexts*, and *background of the user*. We perceive that context is a key issue in interaction between human and computer, describing the surrounding facts that add meaning.

Context is very useful in geographic information science. For instance, when a user requests geographic information (map) to be displayed by a GIS, it is often because the user is trying to perform a *domain task* that has some information needs. The task becomes an important part of the use *context* for *spatial concepts*. Suppose the same request *"show me a map near Cancun"* may be made by a *subject-A,* who is in a task situation of selecting a clothing store, and by a *subject-B* who is planning vacation. However, *subject-B* is likely to expect a map showing a larger geographic area comparing with *subject-A*. There are evidences that the meaning of spatial concepts, such as *"near"*, is also dependent on the spatial context. Therefore, the relevant spatial context of an object depends on the purpose of the considered geospatial data.

We consider that the context term can be used as a mean to express *exceptions*[3] or *constraints*[4]. This use of context is particularly adapted to a rule-based representation of geospatial knowledge, in which exceptions to the rules contain *context-related* terms. Thus, we present a set of intuitive ideas and preliminary definitions that aim

---

[1] According to the Free On-line Dictionary of Computing.
[2] Definition provided by WordNet.
[3] Example: "remove all buildings except the isolated one".
[4] Example: "the river must be into a valley".

better understanding to the roles that play the context into the conceptualization based on conceptual schema.

- **Context.** Let a set (of terminal and non-terminal concepts) X, which contains a set of subcontexts Y and X⊆Y. Then, the set of subcontexts composes the universe of the context denoted by $Y \subseteq C_G$, in which C_G is called *geographic context*. X should be a large set ("large" with respect to cardinality $|C_G|$). Thus,
  - A concept C, which can be terminal (C_T) or non-terminal (C_N) concept, belonging to subcontext Y should mentally suggest or bring into our attention Y.
  - C∈Y implies that the name (mention, evocation[5]) of C makes us think of Y. In the real world, Y{occurs, appears, is produced, is achieved, happens, is used} whenever C {occurs, appears, is produced, is achieved, happens, is used}. For example, concept river ∈ context HYDROLOGY.HYDROLOGY is a set, but we wrote here just its name, since it is a named set.
  - Context should be obvious, not hidden. It should be evoked by every C belonging to it.
  - Context is the extension of *concept* to sets (to named sets).
  - A concept may belong to several contexts. For example, river ∈ HYDROLOGY, river ∈ WATER FLOW. A concept (belonging to a context) could be a context, too. For example, MEXICAN HYDROLOGY ∈ HYDROLOGY.
  - Contexts can overlap.

In conclusion, we propose additional intuitive definitions, which give us ideas to describe the context.

- *Problem* or *Objective* (P). It contains initial state and ending state, in other words, the study object (O_i), a result object (O_r) and a set of constraints (K) that involve the problem or objective (see Eqn. 7).

$$P_m = \{O_i, O_f, K_m\}, \quad K \rightarrow K_m \qquad \textbf{(7)}$$

Therefore, we should take into account the context of the problem to obtain a shared conceptualization about the phenomenon of the real world. Then, the context (Ψ) can be denoted by the problems that are defined by itself (see Eqn. 8).

$$\psi = \bigcup_m P_m \qquad \textbf{(8)}$$

In conclusion, semantics is always defined by a specific context and it is given by a collection of geospatial entities, thus an entity inside the semantic space is defined by the context (Ψ).

---

[5] Thinking, depicting, imagining.

4 Case study

In this section, we describe two scenarios, which are focused on showing how to conceptualize the geospatial domain, by means of conceptual schemas in order to obtain a semantic description regarding specific context. The goal is to depict how these scenarios converge in the same semantic description (see Fig. 4). Although their representations are different, they belong to the same context; thereby their semantic description is the same as well as their conceptualization.

- ***Scenario 1: Imagining the real world.*** Suppose that we are looking at a landscape, which depicts several entities such as a forest that has a lake and a river. Moreover, the freeway F25 crosses the highway I37, F25 is used to arrive to Santa Cruz that is the main town of the surroundings (see[6] Fig. 2). So, it is important to make a conceptualization about our observations. In other words, we are making and abstraction process that is used to conceptualize the landscape. Then this kind of conceptualization can be represented in a conceptual schema and restricted by a context. We use the conceptual schema described in Fig. 1 to generate the semantic description.

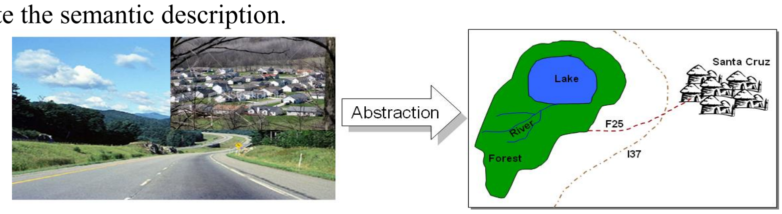

Fig. 2. Scenario 1: Imagining and representing the real world

- ***Scenario 2: Vector map.*** Suppose that we are looking at a map (Fig. 3), it depicts different thematics that consist of different layers, in which each layer contains geographic objects represented by geospatial primitives. The map has *Populations* (POP), *Hydrologic Features* (HYF), *Roads* (ROD) and *Soils* (SOL). Additionally, each thematic and its layers are denoted by a legend. The map is composed of 2 areal objects, 3 linear objects and 1 punctual object.

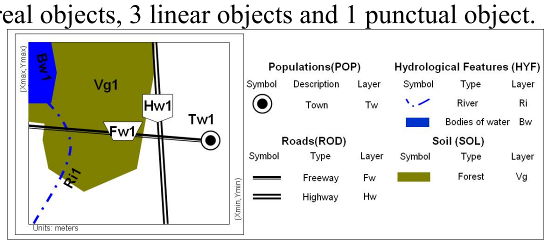

Fig. 3. Thematic map used to illustrate the second scenario

Thereinafter, we use the conceptual schema shown in Fig. 1 to describe both scenarios. According to Fig. 1, in order to obtain a semantic description from conceptual schema, it is necessary to map the geospatial entities into the conceptual schema. Once concepts have been defined into the conceptual schema, we choose the non-

[6] Fig. 3 is only an approximation or an idea of a general landscape described above, we only select some objects to show an illustrative example.

terminal concept to be described (this means to select the aspect to be pointed out). The process continues until we find a terminal concept. When the terminal concept is found, it is necessary to select a pair of geospatial objects, verifying if a relationship between them exists, then a part of description is generated. Terminal concepts are defined by the kind of relation between two objects. In other words, the description starts at the non-terminal concept called "***Geospatial Domain"***. Also, the non-terminal concepts are denoted by means of *rectangles* and the values of the terminal concepts are represented by *ellipses*.

According to the aspect of each non-terminal node, we establish a relation that defines another non-terminal or terminal concept (depending on the objective). This leads to complete the description of geographic objects that compose both scenarios. They converge in the same description according to the context; even though these are represented in different ways (see Fig. 4).

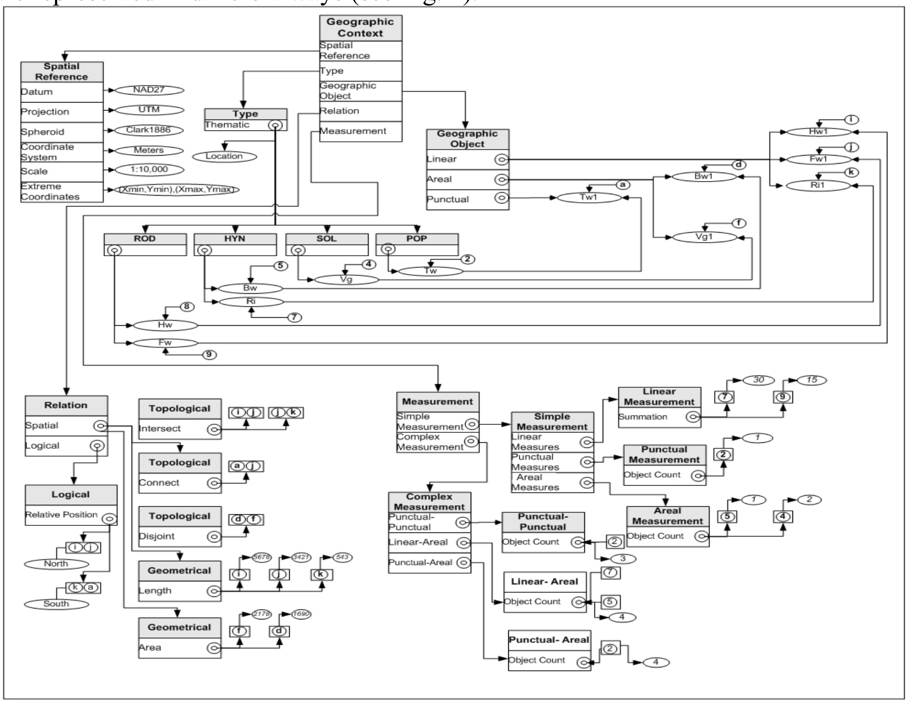

Fig. 4. Semantic description of the scenarios 1 and 2

The method is focused on describing the *semantic content* based on conceptual schemas in a geographic context. However, the description depends on a number of relations, properties and measurements[7] that are considered, whereby it is possible to increase the semantic resolution in the description. The description is made using *tuples* of non-terminal and terminal concepts related among themselves (they are denoted by *Concept* ***relation*** *Concept*). For instance, Fig. 2 and 3 are composed of

[7] A measurement is a procedure for computing values, which are the basis to evaluate characteristics of geospatial phenomena.

several spatial objects. The objects in the layer reflect the relation **"is-a"** (i.e., *HW* **is-a** *Linear Object*). Moreover, the topological relation ***"Intersect"*** is related to *Hw1* and *Fw1*, in which both are linear objects. Thus, in description the ***"Intersect"*** relation is generalized as a spatial relation too.

5 Conclusion and future works

We propose an approach to make a conceptualization of the real world based on conceptual schemas, which are used to generate a semantic description of the geospatial domain. This description attempts to provide the guidelines to formalize the geographic domain in form of geospatial ontologies according to specific contexts.

On the other hand, we perceive that geographic data modeling requires models that are more specific and capable of capturing the semantics of geospatial data, offering higher abstraction mechanisms and implementation independence.

This approach allows us to process imprecise data and aid to information integration and semantic heterogeneity tasks. Thus, the method is focused on describing the semantic content based on conceptual schemas embedded into geographic context. We have introduced two types of concepts: "terminal" and "non-terminal" as well as two kinds of relations: "*has"* and "*is-a"* to build the conceptual schema. Additionally, we have described a set of intuitive definitions oriented to conceptualize the geospatial domain, referring to conceptual schemas and context.

Therefore, we approach conceptual schemas to describe the contents of the real world abstraction to specify the behavior of the geospatial entities, in which context plays an important role to guarantee shared and explicit conceptualizations.

Future works are mainly oriented to propose conceptual issues related to translate semantic descriptions into geospatial ontologies, as well as what would be required to establish these kinds of ontologies. In addition, our work is led to formalize appropriate methods to represent ontologies of the geospatial domain and to measure semantic contents between geospatial ontologies.

Acknowledgments

The authors of this paper wish to thank the CIC, SIP, IPN and CONACYT for their support.

References

1. Mark, D., Smith, B., Egenhofer, M. and Hirtle, S.: Ontological Foundations for Geographic Information Science, in McMaster, R. and Usery, L. (Eds.) *A Research Agenda for Geographic Information Science*, CRC Press, Boca Raton, FL (2004) 335-350.
2. Smith, B. and Mark, D.: Ontology and Geographic Kinds. *Proceedings of the 8th International Symposium on Spatial Data Handling*, Vancouver, Canada (1998) 308-320.
3. Minsky, M.: *A Framework for Representing Knowledge*, Technical Report, in MIT-AI Laboratory, AIM-306, USA (1974).
4. Smith, B. and Mark, D.M.: Geographical categories: an ontological investigation. *International Journal of Geographic Information Science*. 15(7) (2001) 591-612.

5. Torres, M., Moreno, M., Quintero, R. and Fonseca F.: Ontology-driven description of spatial data for their semantic processing. *Proceedings of the First International Conference on Geospatial Semantics*, Springer-Verlag, 3799, Mexico City, Mexico (2005) 242-249.
6. Guarino, N.: Formal Ontology and Information Systems. *Proceedings of the International Conference on Formal Ontology in Information Systems*, Kluwer Academic Publishers, IOS Press, Trento, Italy (1998) 3-15.
7. Gruber, T.R.: Toward Principles for the Design of Ontologies Used for Knowledge Sharing. *International Journal of Human and Computer Studies*. 43(5/6) (1995) 907-928.
8. Bishr, Y. and Kuhn, W.: Ontology-Based Modelling of Geospatial Information. *Proceedings of 3rd AGILE Conference on Geographic Information Science*, Helsinki, Finland (2000) 24-27.
9. Fonseca, F., Davis, C. and Câmara, G.: Bridging Ontologies and Conceptual Schemas in Geographic Information Integration. *GeoInformatica*. 7(4) (2003) 355-378.
10. Berendt, B., Barkowsky, T., Freksa, C. and Kelter, S. I. E.: Spatial Cognition - An Interdisciplinary Approach to Representing and Processing Spatial Knowledge. *Spatial Representation with Aspect Maps*, Springer-Verlag, (1998) 157-175.