RESUMEN DE TESIS DOCTORAL

Representación Ontológica Basada en Descriptores Semánticos Aplicada a Objetos Geográficos
Ontological representation based on semantic descriptors applied to geographic objects

Graduated: Miguel Jesús Torres-Ruiz
Centro de Investigación en Computación del IPN
Av. Juan de Dios Bátiz s/n Esq. Miguel Othón de Mendizábal C.P. 07738 México D.F.
<u>Graduated in November 12, 2007</u>
mtorres@cic.ipn.mx

Advisor: Serguei Levachkine
Centro de Investigación en Computación del IPN
Av. Juan de Dios Bátiz s/n Esq. Miguel Othón de Mendizábal C.P. 07738 México D.F.
sergei@cic.ipn.mx

Abstract

In this dissertation the GEONTO–MET methodology is presented. It is oriented to formalize the geographic domain conceptualization according to the specifications of the Instituto Nacional de Estadística, Geografía e Informática (INEGI). The main goal is to provide semantic descriptions, which represent the properties and relations describing so the behavior of geographic objects and taking into account these features directly from the geographic domain ontology. The thesis also establishes a solid basis for future automatic interpretations of geographic regions oriented to decision making process. In other words, the geographic domain conceptualization can be used to generate new knowledge in an automatic way and make spatial analysis more accurate. Having a semantic description of a geographic domain allows establishing the guidelines of a spatial reasoning. That is, if the behavior of geographic objects involved in an ontology partition is known, then one can understand and interpret the semantics that they represent and explore the implicit inference from the conceptualization. Summing up, this work presents a methodology allowing integrate and share geospatial information. GEONTO –MET provides feasible solutions towards these and other related issues, such as compact data by alternative structures of knowledge representation and avoid the ambiguity of terms, using a domain conceptualization.

Keywords: geographic domain conceptualization, semantic description, geospatial ontology

Resumen

En esta tesis se presenta la metodología GEONTO – MET, orientada a formalizar la conceptualización del dominio geográfico, considerando las especificaciones del Instituto Nacional de Estadística, Geografía e Informática (INEGI). El espíritu de esta metodología es proporcionar un conjunto de descripciones semánticas que reflejen las propiedades y relaciones que describen entre sí el comportamiento de los objetos geográficos, tomando estos elementos directamente de la ontología de dominio geográfico diseñada para este fin. Esta tesis sienta las bases para una futura interpretación automática de regiones geográficas, orientado al soporte de toma de decisiones, haciendo uso de la conceptualización del dominio geográfico; con el objeto de generar nuevo conocimiento en forma automática y realizar un análisis más profundo del entorno geoespacial. El hecho de contar con una descripción semántica de un entorno geográfico, permite establecer los requerimientos necesarios hacia un razonamiento espacial; es decir, el hecho de conocer el comportamiento que presentan los objetos geográficos inmersos en una partición geográfica es de suma importancia para entender e interpretar la semántica que representan y explotar la inferencia implícita que posee la conceptualización. En consecuencia, este trabajo aporta la metodología que permite integrar y compartir información geoespacial. GEONTO – MET proporciona soluciones viables hacia estos tópicos y otros, como el hecho de compactar datos mediante estructuras alternas a modelos tradicionales y evitar la ambigüedad de términos al utilizar una conceptualización del dominio.

Palabras clave: conceptualización de dominio geográfico, descripción semántica, ontología geoespacial

1 Introduction

Ontologies are widely used in Knowledge Engineering and Artificial Intelligence, in applications related to knowledge management, natural language processing, e-commerce, intelligent information integration, information retrieval, database design, and in new emerging fields like the Semantic Web (Gómez, *et al.,* 2004). The primary goal of *ontology* is *sharing* and *reusing* knowledge among people or machines (Tomai and Kavouras, 2004). However, it is difficult to find concrete ontologies within real-world knowledge domain. At this time, there are different and public ontologies for many domains and applications. Users can access them in order to share information, reuse knowledge and integrate data sources for several purposes (Guarino and Welty, 1995). From this point of view, ontology engineering is rapidly becoming a mature discipline which has produced various tools and methodologies for building and managing ontologies (Staab, *et al.,* 2001). However, even with a clearly defined engineering methodology, building a large ontology remains a challenging, time-consuming and error-prone task, since it forces ontology builders to conceptualize their expert knowledge *explicitly* and to *re-organize* it in typical ontological categories such as *concepts*, *properties* and *axioms* (Bateman and Farrar, 2004; Uschold and Grüninger, 1996). Ontology has gained attention among researchers in geospatial information science in recent years. It plays an important role in establishing robust theoretical foundations for geographic information science. In (Mark, *et al.,* 2001) three broad sets of foundational issues are sketched out: (i) conceptual issues concerning what would be required to establish an exhaustive ontology of the geospatial domain, (ii) representational and logical issues relating to the choice of appropriate methods for formalizing ontologies, and (iii) issues of implementation regarding the ways in which ontology ought to influence the design of information systems.

The thesis is focused on the first and second issues to develop the GEONTO-MET approach to conceptualizing the geographic domain. This methodology allows for *integrating* and *sharing* geospatial information as well as for avoiding the *ambiguities* that are sometimes present in geographical data. Many questions can be asked about this issue such as: why is it important to conceptualize the geographic domain? What are the main goals for building approaches oriented to generate geo-ontologies? What are the main benefits in conceptualizing this domain? The answers are directly focused on describing the limitations that *numeric focus* offers in this context (geographic domain). Several implicit characteristics of geospatial information are related to the scale, projection, and reference systems. These have produced problems associated with the semantic heterogeneity, interoperability, data sharing, information integration and imprecise data processing. For instance, if we ask where the Eiffel Tower is, the answer, of a geographic information system is, *"in coordinates 48° 51' 30" N; 2° 17' 40" W"*. However, if we ask the same question to a person, his answers could be *"in Paris"*, *"near Avenue des Champs-Elysées"* or *"on the Rive Gauche"*. The main difference between these answers (*GIS* and *people*) is that the thought of a human being is *conceptual* and the *reasoning* is directly related to the *cognitive* system. Then, numerical methods are not common in our thoughts to find answers about this kind of questions.

Therefore, this work is oriented towards processing geospatial data in a *semantic* focus. This research assumes that when we, as human beings *think spatially*, we relate locations to our cognitive system for reasoning and describing the geographic environment. Features like coordinates, projections, spatial databases or maps are not considered in our mind maps. We think about things that give us a reference associated with different *labels*, in which a location that is well known provides a "*description*" or a reference. This means that approaches to the conceptualization of the geographic domain provide common and shared ontological descriptions designed to establish geospatial specifications or terminology standards to partially solve this interoperability.

In summing up, the *integration*, *sharing* and *interoperability* of geospatial information is the main goal of the GEONTO-MET methodology. Semantic descriptions are built to achieve these purposes. In addition, *Kaab* ontology has been designed using this approach.

The remainder of the thesis is structured as follows: Section 2 describes the GEONTO-MET approach to conceptualize the geographic domain. Section 3 depicts some results about this dissertation. Conclusions and suggestions for future work are outlined in Section 4.

2 GEONTO-MET approach

GEONTO-MET methodology (Torres, *et al.,* 2005; Torres and Levachkine, 2007; Torres, 2007) is oriented to formalize geographic domain conceptualization according to the specifications from the INEGI. The main goal is to provide semantic descriptions, which represent the properties and relations being described so that the behavior and features of geographic objects are taken into account directly from the geographic domain ontology.

GEONTO-MET is composed of four principal stages: *Analysis* provides an abstract model of the geographic objects involved in this domain. *Synthesis* makes the conceptualization of the geographic domain. A set of application ontologies (in tourist and topographic contexts) and domain ontology called *Kaab-Ontology* is generated by the *Processing* stage. Finally, *Description* produces an alternative representation of geographic objects as well as the integration of them in a semantic description template. The general framework of this methodology is shown.
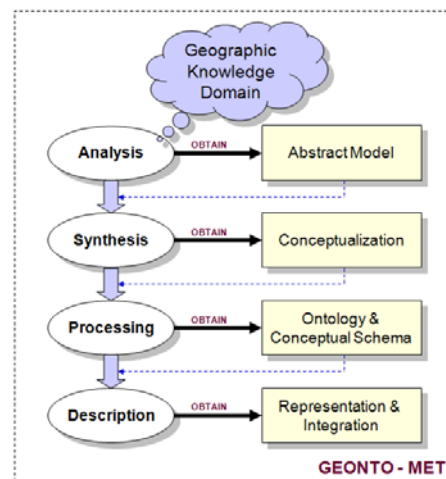

Fig. 1. Framework of GEONTO-MET

According to Figure 1, the conceptual structure proposed for the GEONTO-MET approach initially considers the *geographic knowledge domain*. It is based on several data sources such as the metadata dictionary of INEGI (INEGI, 1999) and the geographic ISO specification (ISO-19115, 2004). These sources provide the general knowledge and the way of how domain experts perceive the abstract world to reflect a simplify view of the geographic environment. In this paper, we are focused on describing the *synthesis* and *processing* stages and the results obtained in the *description* phase.

On the other hand, this approach is based on a set of axiomatic relationships allowing for direct translation of the relations between concepts to the conceptualization. In this way, the *semantic resolution* is improved, that is, the definition of such relationships can be iteratively refined. To achieve this, we use a couple of sets ($A_1 = \{is, has, does\}$ and A_2 specific prepositions related to geospatial context). These sets are necessary and sufficient to define the rest of relationships, involved in the conceptualization of geographic domain.

The essence of this methodology is to reduce the *axiomatic relationships* within conceptualization. One could think that this reduction is a limitation for the richness of expression that conceptualization can implicitly contain. Nevertheless, the universe of possible relations is not *a priori* defined, due to the *"relation"* in a classic sense is not predefined. In fact, the reduction of axiomatic relations has two main advantages: the first, it is possible to define as many *"typical relations"* as needed, because this type of relation is treated as a concept. In other words, *"typical relations"* are part of conceptualization, they are not considered as axioms, and these are defined as *concepts*. The second advantage is that relations have a semantic association to themselves, not only from an axiomatic definition, but also from the conceptualization itself (the context of each relation).

To illustrate the idea, let us consider one axiomatic relation widely used: the relationship *"part_of"*. Such relationship means that one concept is a constituent element of another concept. By using GEONTO-MET, it is possible to create this relationship as a concept *(concept-R)*, by defining the concept *"part"* (in the way that the concepts are defined) and using the axiomatic relationships (*is*, *does*, *has*) to create the *concept-R* equivalent to the relationship *"part_of"*.

For instance, let's consider the following sentence: "**heart** *part_of* **body**"; in this, two concepts (*heart* and *body*) are involved as well as one axiomatic relationship (*part_of*). By using the methodology, the same relationship could be expressed "**heart** *is* part of **body**", in which three concepts (*heart*, *part* and *body*) and two axiomatic relationships: a fundamental one (*is*) and an auxiliary one (*of*) are described. The advantage is that the semantic of the relationship is the same and it is not necessary to previously define the relation *"part_of"*. Then, we only need to define *"part"* as a concept (having the semantic richness of concepts) and use it to define the new relation. On the other hand, GEONTO-MET is composed of a set of elements that are used to make the *geographic domain conceptualization*. They are described in the next subsections.

2.1 Axioms in GEONTO-MET

It minimizes the number of axioms by means of the reduction of the axiomatic relationships. In order to do this, we have defined a small set of axiomatic binary relationships, divided into two subsets. The first, subset A_1 contains three relations that will be called *fundamental relations*, and it is denoted by equation 1.

$$A_1 = \{is, has, does\} \tag{1}$$

The *"is"* relationship means an *existence* or identity, such it is used to characterize the concepts in the conceptualization. It implies the inheritance of properties and abilities. Also, it allows for making a hierarchy of the same concepts. Some names for this relationship could be *"son-of"* or, *"is-a"*. As binary relation, *"is"* has the following properties. Let C be a set of concepts:

- **It is anti-symmetric:** $\forall a, b \in C, a(is)b \wedge b(is)a \Rightarrow a = b$, if concept A *"is"* B and B *"is"* A, necessarily A and B are the same concept. For instance, "lake" *"is"* "waterbody" but "waterbody" *("is")* not lake, because there are other geographic concepts that are also "waterbody".
- **It is reflexive:** $\forall a \in C, a(is)a$, each concept A *"is"* itself. For instance, "Chapultepec Lake" *"is"* "Chapultepec Lake", because a geographic object has its own identity.
- **It is transitive:** $\forall a, b, c \in C, a(is)b \wedge b(is)c \Rightarrow a(is)c$, if A *"is"* B and B *"is"* C then A *"is"* C. For instance, if "road" *"is"* "og_artificial" and "og_artificial" *"is"* "geographic object", then "road" *"is"* "geographic object".

The *"has"* relationship describes *aggregation* or *association*, such it is used to define the properties that build up a concept. The properties of this binary relation are as follows:

- **It is non-symmetric:** $\exists a, b \in C, a(has)b \wedge b(has)a \Rightarrow a = b$, there are concepts A and B so that if A *"has"* B and B *"has"* A, then A and B are the same concept. For instance, "state" *"has"* "counties" but "counties" do not have (*"has"*) "states".
- **It is irreflexive:** $\exists a \in C \ni a \neg(has)a$, there is some concept A such that it does not have itself as property. For instance "urban area" does not have (*"has"*) "urban area".
- **It is transitive:** $\forall a, b, c \in C, a(has)b \wedge b(has)c \Rightarrow a(has)c$, if A *"has"* B and B *"has"* C, then A *"has"* C. For instance, if "country" *"has"* "state" and "state" *"has"* "county", then "country" *"has"* "county".

The *"does"* relationship is used to describe an action, such it defines the abilities or operations associated with a concept. Its properties are as follows:

- **It is symmetric:** $\exists a, b \in C, a(does)b \wedge b(does)a \Rightarrow a = b$, there are concepts A and B such that if A *"does"* B and B *"does"* A, then A and B are the same concept.

- **It is irreflexive:** $\exists a \in C \ni a\neg(does)a$, there is some concept A that does not have itself as an ability.
- **It is non-transtive:** $\exists a,b,c \in C \ni a(does)b \wedge b(does)c \Rightarrow \neg a(does)c$, if A *"does"* B and B *"does"* C it does not necessarily imply that A *"does"* C.

The second subset of axiomatic relationships is denoted by A_2. It is composed of prepositions. See the equation 2.

$$A_2 = \begin{Bmatrix} to, before, under, with, against, of, from, in, between, towards, until, \\ for, by, since, on, after, behind, beside, near, through \end{Bmatrix} \tag{2}$$

These relationships are defined as *asymmetric*, *irreflexive* and *non-transitive*, although linguistically some of them do not accomplish such properties[1]. For implementation convenience, we have considered that relationships in A_2 are defined as follows. Let $r \in A_2$ be a relationship, then, i) $\forall a,b \in C, arb \wedge bra \Rightarrow a = b$, ii) $\exists a \in C \ni a\neg ra$ and iii) $\exists a,b,c \in C \ni arb \wedge brc \Rightarrow a\neg rc$.

2.2 Relations in GEONTO-MET
Relations in GEONTO-MET can be classified in two types:
- **Simple:** This type has the form: $a\rho b \in R_S$, where $a,b \in C$ and $\rho \in A_1$.
- **Complex:** It has the form: $a\rho b\pi c \in R_C$, where $a,b,c \in C$ y $\rho \in A_1$ and $\pi \in A_2$.

As we have previously mentioned, *"is"* is a hierarchy relationship that provides the mechanism of *"inheritance"*. By using this relationship, we will be able to create a hierarchy of concepts (existence). The *"has"* relationship provides the capability of *"aggregation"* or *"composition"* of concepts. The *"does"* relationship describes the actions of a concept.

Relationships in A_2 allow for describing and providing *causality* or *intention* to the relationships in A_1. The complex form of relations is used to have a minimum number of axiomatic relationships and give us the semantic to non-axiomatic relationships. R_V describes the set of valid relationships. R_V is the set of all relationships such that comply with restrictions. On the other hand, R_R represents the set of relationships existing in real or concrete conceptualization.

2.3 Concepts in GEONTO-MET
We have defined a concept as a collection of abilities and properties that share a single existence. There are four types of concepts:
- **Relational concepts** (verbs). They are defined as elements denoting an action or operation over other concepts. C_R denotes the set of all concepts-R.
- **Standard concepts** (noun). They are defined as elements belonging to a class. All their abilities and properties are abstract. C_E (concepts-E) denotes the set of this type of concept.
- **Class concepts**. They are concepts-E that allow for making partitions of C_R and C_E. They are denoted by C_L (concepts-L).
- **Instance concepts.** They are concepts whose abilities and properties are concrete.

Also, we can stand that $C = C_R \bigcup C_E \bigcup C_L$, such that $C_E \cap C_R = \varnothing$, $C_E \cap C_L = \varnothing$ and $C_L \cap C_R = \varnothing$, i.e. C_L, C_R and C_E are disjoined sets. An *abstract* concept is a concept-E, which does not have instances.
2.4 Properties in GEONTO-MET

---

[1] For instance, the relationship *"beside"* is a symmetric relation.

Properties are concepts aggregated to one another by means of a relationship of belonging. These define the characteristics of the second concept. A property can be defined as follows: Let $a, b \in C$ be concepts, we say that b is a property of a if $a(has)b \in R_R$. b is a concrete property of a if b is an **instance**. It is called *abstract property* to the one that is not concrete.

For instance, if "mountain" *"has"* "altitude", and "altitude" could be "low", "mid" and "high"; then "altitude" is a property of "mountain" and "low" is a concrete property of "mountain".

Π is the set of properties, Π_C is the set of concrete properties and Π_A is the set of abstract properties. So, in this way we can stand that $\Pi = \Pi_A \bigcup \Pi_C$ and $\Pi_A \bigcap \Pi_C = \varnothing$.

2.5 Abilities in GEONTO-MET
Abilities are concepts that define actions or operations associated with other concepts. As well, they describe how a concept interacts with other concepts. An ability is defined as follows: Let $a, b \in C$ be concepts, so we can say that b is an ability of a if $a(does)b \in R_R$.

b is a concrete ability of a if b is an **instance**. The non-concrete abilities are called *abstract abilities*. H is the set of abilities, H_C is the set of concrete abilities and H_A is the set of abstract abilities. Also, $H = H_A \bigcup H_C$ and $H_A \bigcap H_C = \varnothing$.

2.6 Instances in GEONTO-MET
Instances are a collection of concrete abilities and properties that have a unique existence. In other words, an *instance* is a concept contained within a hierarchy formed by the *"is"* relationship, whose properties and abilities are concrete. Thus, an instance is a concept with no sons within the hierarchy, - it is not a class and all its abilities and properties are instances. I represents the set of all instances in a conceptualization[2].

2.7 Classes in GEONTO-MET
Classes are concepts that have *children concepts* by means of the existence relationship (*"is"*), their sons cannot be an *instance*, and they form a *complete partition*.

In this case, the *"is"* relationship is equivalent to the relationship *"subclass_of"*. C_L denotes the set of classes.

2.8 Constraints in GEONTO-MET
A constraint is described as a *sentence* to avoid inconsistencies in the definition and description of the elements involved in geographic domain conceptualization. A set of constraints defined in GEONTO-MET are sketched out as follows.

i) Constraints for relations in the simple form
Let $a, b \in C$ be concepts, then $a(is)b \in R_V$ if,

- $a(is)b \in R_S \Rightarrow b(is)a \notin R_V$, for avoiding redundant inheritances, because *"is"* relationship is reflexive. This restriction eliminates $\begin{bmatrix} n \\ 2 \end{bmatrix} \Big/ 2$ possible relationships, where $n = card(C)$.

- $a(is)b, b(is)c \in R_S \Rightarrow a(is)c \notin R_V$, for avoiding redundant inheritances, because *"is"* relationship is transitive.

- $a(is)b \in R_S \Rightarrow b \notin I$, instances do not have sons. This restriction eliminates $n_i n$ possible relationships, where $n_i = card(I)$.

---

[2] In a pragmatic form I is a database.

- $a(is)b \in R_S, b \in C_L \Rightarrow a \notin I$, instances cannot inherit from classes. This restriction eliminates $n_l n_l$ possible relationships, where $n_l = card(C_L)$.

Similarly, the *"has"* relationship contains restrictions in its simple form; Let $a,b \in C$ be concepts, then $a(has)b \in R_V$ if,

- $a(has)b \in R_S \Rightarrow a \notin C_L$, classes do not have properties. This restriction eliminates $n_l n$ possible relationships.
- $a(has)b \in R_S, a \in I \Rightarrow b \notin C_L$, classes cannot be properties of instances. This restriction eliminates $n_l n_l$ possible relationships.
- $a(has)b \in R_S \Rightarrow b \notin I$, instances cannot be properties of non-concrete concepts. This restriction eliminates $n_l n$ possible relationships.

The *"does"* relationship has its restrictions when it is used in its simple form; Let $a,b \in C$ be concepts, then $a(does)b \in R_V$ if,

- $a(does)b \in R_S \Rightarrow a \notin C_L$, classes do not have abilities. This restriction eliminates $n_l n$ possible relationships.
- $a(does)b \in R_S, a \in I \Rightarrow b \notin C_L$, classes cannot be abilities of instances. This restriction eliminates $n_l n_l$ possible relationships.
- $a(does)b \in R_S \Rightarrow b \notin I$, concepts cannot have instances as abilities. This eliminates $n_l n$ possible relationships.

ii) Constraints for relations in the complex form
Let $a,b,c \in C$ be concepts, and $\pi \in P$, then:

- $a(is)b\pi c \notin R_V$ if,
  - $a(is)b\pi c \in R_C, a \in C_E \Rightarrow b \notin I$, instances cannot have sons.
  - $a(is)b\pi c \in R_C, a \in I \Rightarrow b \in C_E$, instances can only be sons of concepts-E.
  - $a(is)b\pi c \in R_C \Rightarrow b \notin I$, *"is"* relation can be used for describing ternary relations, if the concept-R is not an instance.
- $a(has)b\pi c \in R_V$ if,
  - $a(has)b\pi c \in R_C \Rightarrow b \in C_E$, *"has"* relationship can be used for describing ternary relationships, if the middle concept is a concept-E.
  - $a(has)b\pi c \in R_C \Rightarrow a \notin C_L$, classes do not have properties.
  - $a(has)b\pi c \in R_C, \pi \in P \Rightarrow c \notin I$, concepts cannot have instances as properties.
  - $a(has)b\pi c \in R_C, a \in I, \pi \in P \Rightarrow c \notin C_L$, instances do not have classes as properties.
- $a(does)b\pi c \in R_V$ is,
  - $a(does)b\pi c \in R_C, \pi \in P \Rightarrow b \in C_R$, *"does"* relationship can be used for describing ternary relationships if the middle concept is a concept-R.
  - $a(does)b\pi c \in R_C, \pi \in P \Rightarrow a \notin C_L$, classes do not have abilities.
  - $a(does)b\pi c \in R_C, \pi \in P \Rightarrow c \notin I$, concepts cannot have instances as abilities.

2.9 Retrieving geospatial objects through *Kaab* ontology
Ontologies have been implemented in Protégé, applying the GEONTO-MET approach. To carry out the retrieving of geospatial objects, Jena API has been used. It is a structure of Java oriented to build *semantic web* applications. It

provides an integrated programming environment for RDF, RDFS, OWL and SPARQL as well as includes an inference engine based on rules.

This API extracts instances of concepts from *Kaab* ontology, which has been implemented in OWL using GEONTO-MET approach. Therefore, it is necessary to generate a *persistent model* of the conceptualization to establish the link between the components. The instances are retrieved by using SPARQL, which is the query language that offers an access protocol of metadata into the OWL structure. In Figure 2 is depicted the framework to retrieve geospatial objects, by means of *instances* of concepts, which are mapped with geospatial objects stored within a geographic database.

The process starts with the transformation of geographic domain ontology to a persistent model in OWL. The task consists of translating the *abstract entity classes* and *concepts* inherited from these to tabular form, in which SPARQL can access to the elements of the *Kaab-Ontology*. The mapping process is based on a set of predicates to link geographic concepts and instances stored in the OWL persistent model. A searching method by strings is used to retrieve the instances from the *Kaab-Ontology*. This method is composed of *subject*, *predicate* and *objects*. All the strings generated by Jena will be composed of a *triplet* of these elements for semantically represent the state of the conceptualization of a geographic object in a certain context. The instances are retrieved by means of a set of queries expressed in SPARQL. The main purpose of this work is to show the way for accessing the ontology and retrieving the *individuals* of concepts that belong to certain class and accomplish with the *semantic* searching criteria. The advantage for this retrieval mechanism is that the query model is oriented to *data*, whereby the meaning of each defined term in a vocabulary is represented by a language and it is possible to semantically extract *instances* of *concepts* related to the context of the ontology.
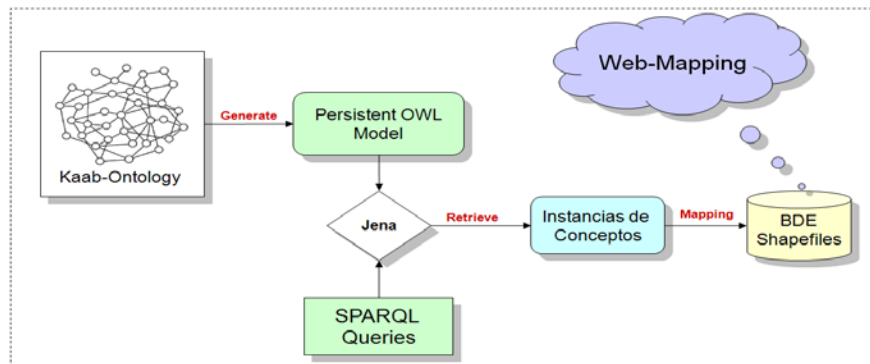

Fig. 2. Framework to retrieve instances and the mapping of geospatial objects

When the instances of concepts are retrieved, they are stored in tables, in which each table has an identifier that is directly mapped to the *shapefiles*[3] for linking the semantic expression with spatial objects. The geographic data that accomplish with these criteria are visualized in a *web-mapping* application.

2.10 Template for a semantic description

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.

---

[3] *Shapefile* is a popular geospatial vector data format for GIS software. It is developed by ESRI as a (mostly) open specification for data interoperability among ESRI and other software products. It spatially describes geometries: points, polylines, and polygons.

This work proposed a conceptual schema to describe the contents of the real world abstraction to specify the behavior of geospatial entities. Thus, conceptual schemas certainly correspond to a level of knowledge formalization. They 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 more specific and capable of capturing the *semantics* of geospatial data, offering higher abstraction mechanisms and implementation independence. Our conceptual schema is 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 that can be terminal or non-terminal concepts (see the equation 3).

$$C = C_N \bigcup C_T \tag{3}$$

Each concept has a set of *aspects*. They 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. 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 the equation 4.

$$T_P = \{number, character, string, enumeration, struct\}, \tag{4}$$
$$A = \{a_i \mid type(a_i) \in T_P\},$$

Where T_P is the set of primitive types; *A* is the set of aspects.
Then, the set of *terminal concepts* is defined by the equation 5:

$$C_T = \{c(a_1, a_2, ..., a_n) \ni a_i \in A, \ i = 1, .., n\} \tag{5}$$

The *non-terminal concepts* have at least one aspect that does not belong to T_P. It is denoted by the equation 6:

$$C_N = \{c(a_1, a_2, ..., a_n) \ni \exists a_i \notin A\}, \text{ where } c \text{ is a concept.} \tag{6}$$

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 (See the equation 7).

$$R = R_\Gamma \bigcup R_\Phi = \{(a,b) \mid a\Gamma b, \ a \in C_N, \ b \in C\} \bigcup \{(a,b) \mid a\Phi b, \ a \in C_N, \ b \in C\} \tag{7}$$

According to these definitions, 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 the equation 8).

$$D = \langle C, R \rangle \tag{8}$$

Figure 3 depicts a conceptual schema 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, geologic, or tourism contexts, it is possible to describe the entities, characteristics and relationships embedded between geographic objects, as an inheritance mechanism. This conceptualization provides us explicit vocabulary that represents the ontological commitment of the cognitive and intuitive perception of the subjects.
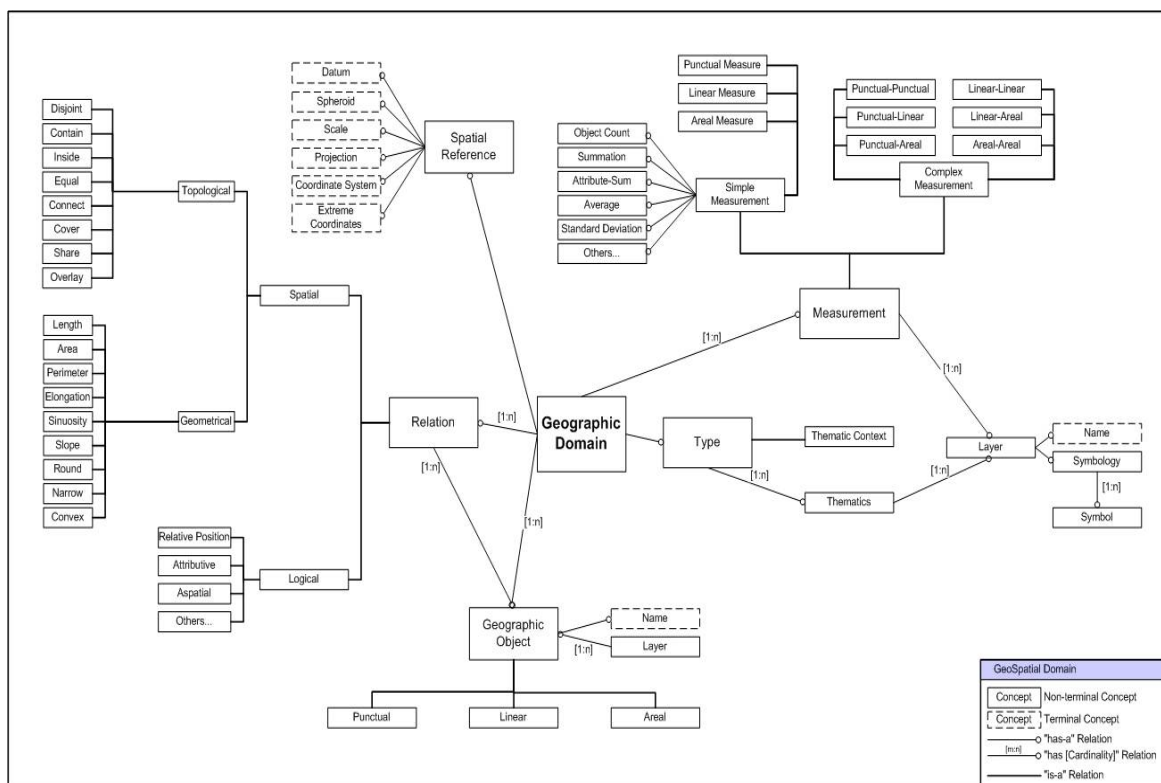
Fig. 3. Conceptual schema used to generate semantic descriptions

As a result from this schema, we obtain a symbolic description that represents a logical organization of C_T and C_N. These concepts are used as a guide to generate the semantic description.

To build the description, a set of tags are considered. The JTS (Java Topology Suite) API is used to populate the C_N and for exchanging the geometric information the WKT strings (Well Known Text) are defined. The template for the semantic description is composed of the following tokens:

- ( )   They represent the object "name".
- < >   They establish the relationship between objects.
- { }   They represent the complete description of each geospatial object.
- **Direction.**   It represents the relative orientation of an object.

Table 1. Structure of the template to define the semantic description

```
{
  [object thematic] (object) <relationship>
  . . .
  [thematic] (object) <relationship>
  direction
};
. . .
{
Other object description
};
```

3 Results

The results are basically oriented to show the functionality of GEONTO-MET, providing some fragments of geographic domain ontology (*Kaab*). The ontology was developed in Protégé 3.2.1. Also, the Jena interface is used to access the ontology and retrieve concepts and relationships for linking the geospatial data. In addition, semantic and ontological representations are depicted.

3.1 *Kaab* ontology
This ontology has been developed by applying the GEONTO-MET methodology. The ontology goal is to aid the *integration* and *sharing* geo-information tasks for avoiding the ambiguity of terms in the development of geographic information systems.

The metadata that are considered in this ontology, were taken from the vector data dictionary provided by INEGI and the ISO 19115 designed by OGC[4], which represent the knowledge domain. In Figure 4, the main partition of *Kaab* ontology is depicted.

The concepts inherited from **OWL:Thing** are considered as *abstract classes*. The *"is"* relationship allows generating essential entities to conceptualize the geographic domain. According to the methodology, it is possible to add a great number of concepts, because the relationships that can be presented between them are directly translated to the conceptualization, avoiding the use of many relationships and enriching the *semantic* of the ontology. A detailed description of each class, as well as the concepts and relationships involved in *Kaab* is stated in (Torres, 2007).
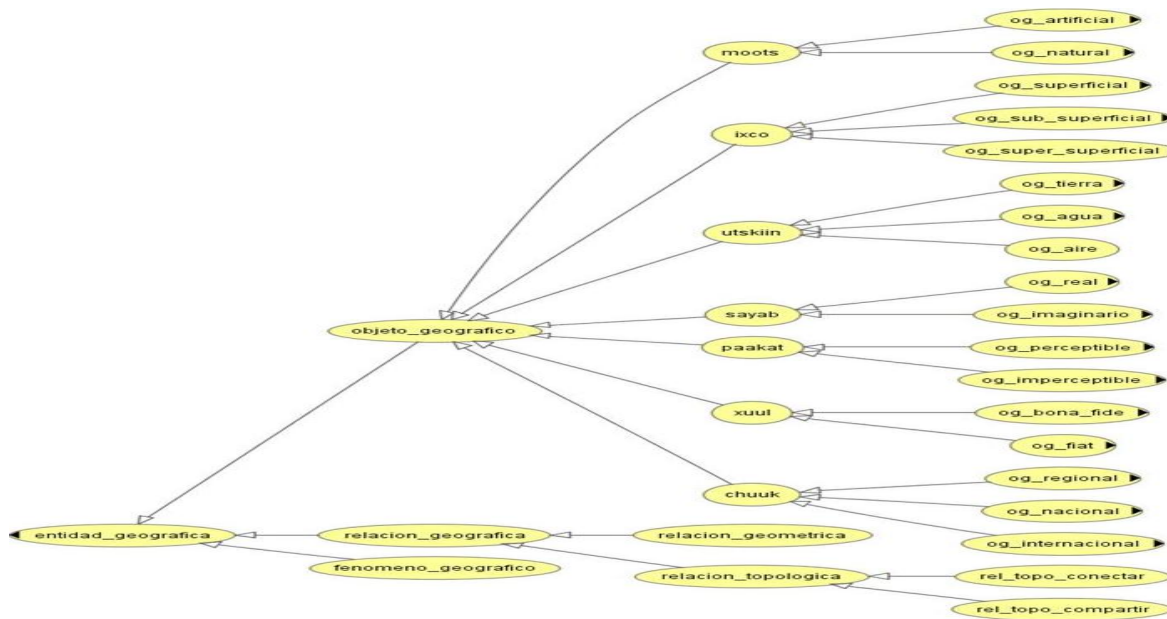

Fig. 4. Main hierarchy of *Kaab*-Ontology

3.2 Ontological descriptions
Ontological descriptions provide the behavior of a geographic data set. This behavior represents the *semantics* of the entities, because the relationships between geospatial objects mainly define the *meaning* and the role that they play in

---

[4] OGC. It is the Open Geospatial Consortium, which provides an standard of metadata for geographic information.

the real world. This fact is oriented to establish mechanisms for *spatial reasoning* and to cognitively understand the way that human beings describe the geographic objects in the world.

Ontological descriptions are performed by means of the axiomatic relationships. In Figure 5 a description of the *"road"* concept with *"is"* and *"has"* relationships is depicted. This description contains the *properties* associated to the concept that are determined by *"has"*. The *"is"* relationship establishes the *root* of the concept to determine its classes within the *Kaab* hierarchy.
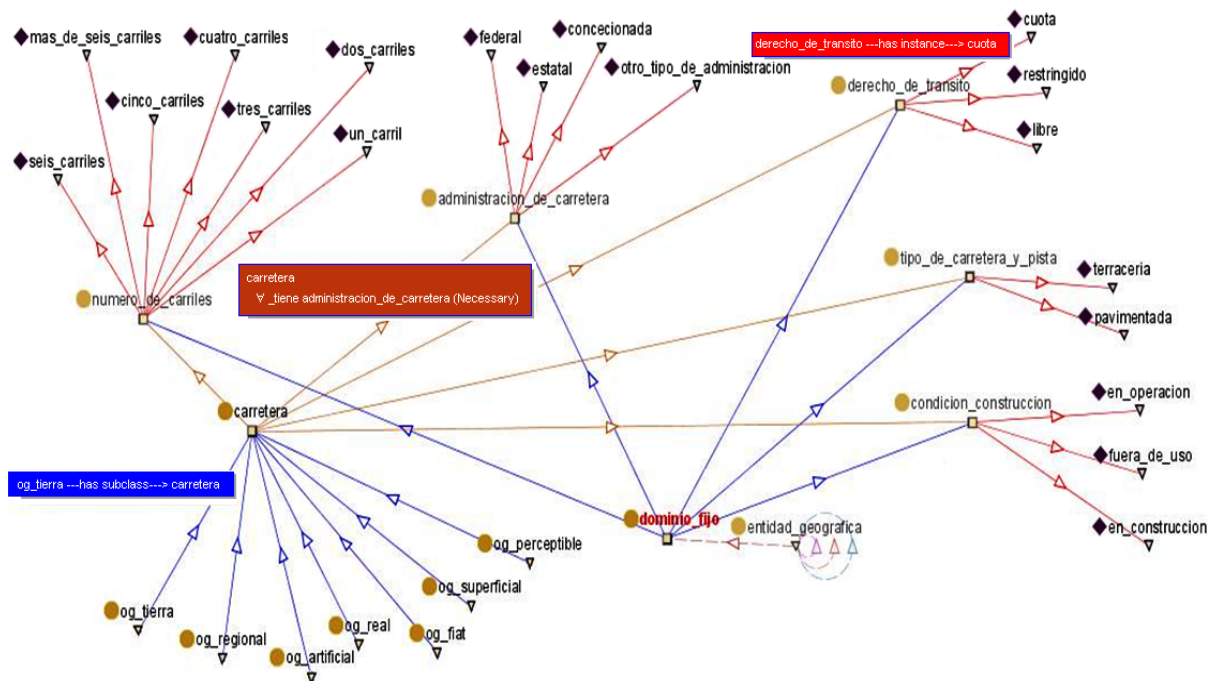

Fig. 5. Ontological description of the concept *"road"*

In Figure 6, the ontological description that represents *abilities* of *instances* is shown. The ability of a concept is defined by the *"does"* relationship. Instances are considered *individuals*[5], because they play the role of typical geographic objects in the real world (as people perceive). They have properties that *characterize* their identity as well as abilities that *describe* the behavior of these instances with others.

---

[5] The *individual* term is a pragmatic idea that is related to the entities of the real world. For example, a *mountain* is the concept; however, if people say *"Everest Mountain",* it is an instance that belongs to the *"mountain"* concept.
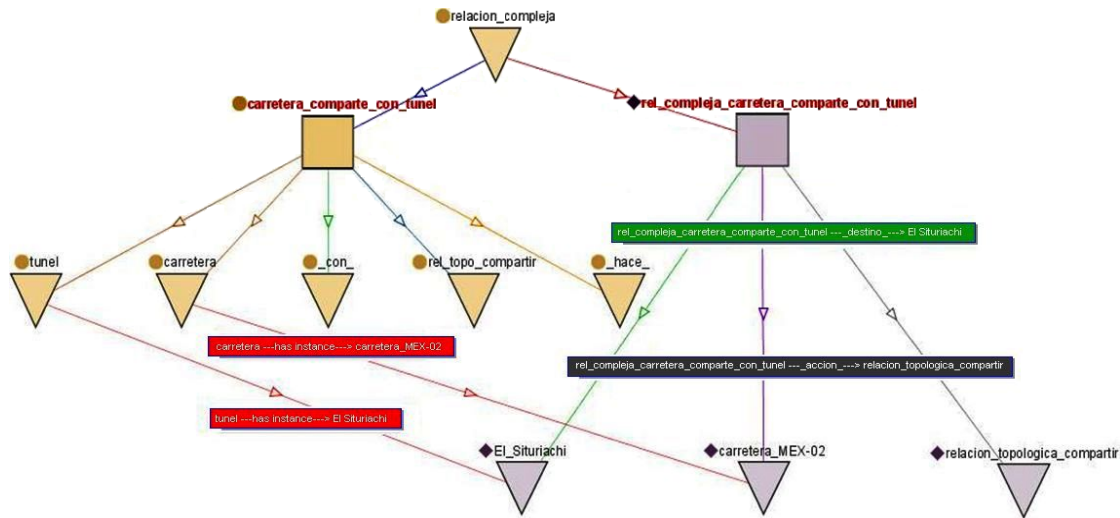
Fig. 6. Ontological description of abilities associated to instances of the "road" concept

3.3 Semantic description and geospatial data visualization in a web-mapping application

A semantic description of a geospatial region is generated according to the conceptual schema depicted in Figure 3. The XML document is designed in order to share and integrate geospatial information with other geospatial databases. Additionally, a native template has been built to describe the *semantics* of a geospatial region (Figure 7).
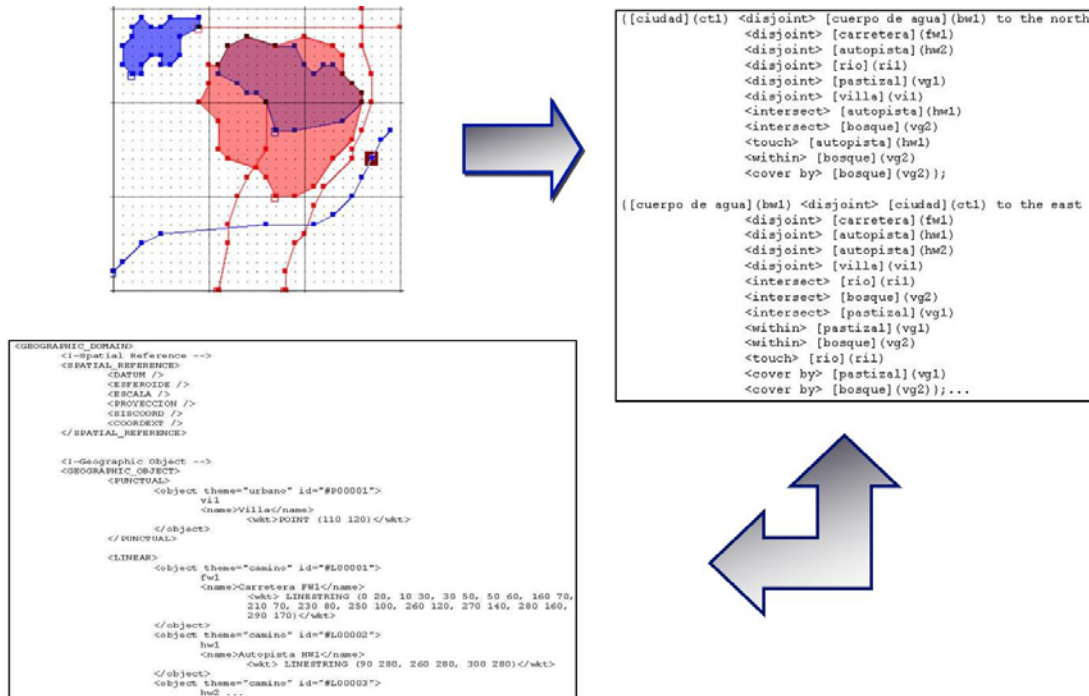

Fig. 7. Semantic description and XML document generated from a geospatial region

A semantic retrieval of concept instances is shown in Figure 8. The query has been implemented in SPARQL and the code accesses *Kaab* ontology, by means of an OWL persistent model to retrieve them. In this query the cost per room in hotels is returned. *"Cost"* is a property of a concept and *"hotels"* are instances. These features belong to the application ontology in a touristic context.
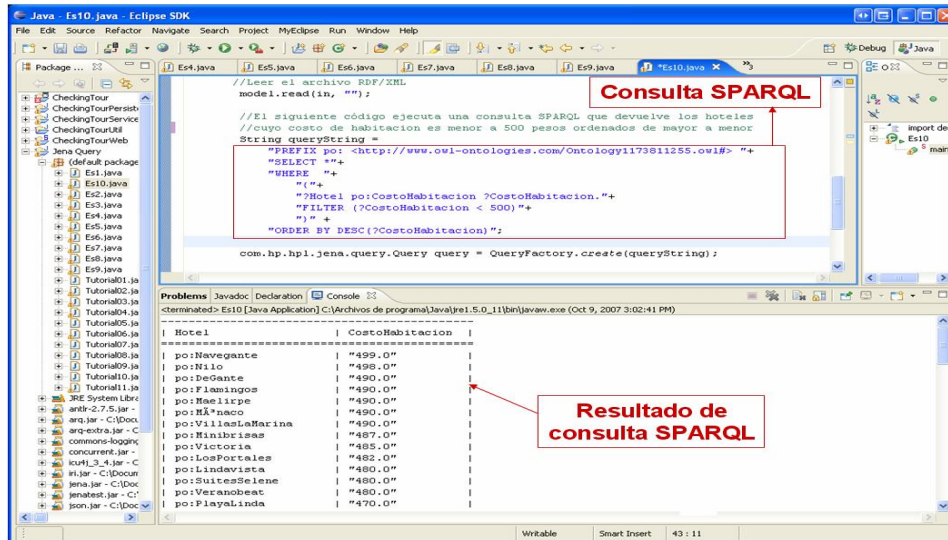

Fig. 8. Instances and value properties associated to the "hotel" concept

In Figure 9 a *web-mapping* application is shown. The instances of concepts retrieved by the SPARQL query are mapped to the geographic objects stored in the *Shapefile*. The goal of this mechanism is to avoid *ambiguities* in the query process of a GIS. A *semantic focus* is proposed to this issue, a conceptualization of a geographic domain is represented by the domain ontology, which provides a readable mechanism to *integrate* and *share* geospatial information.
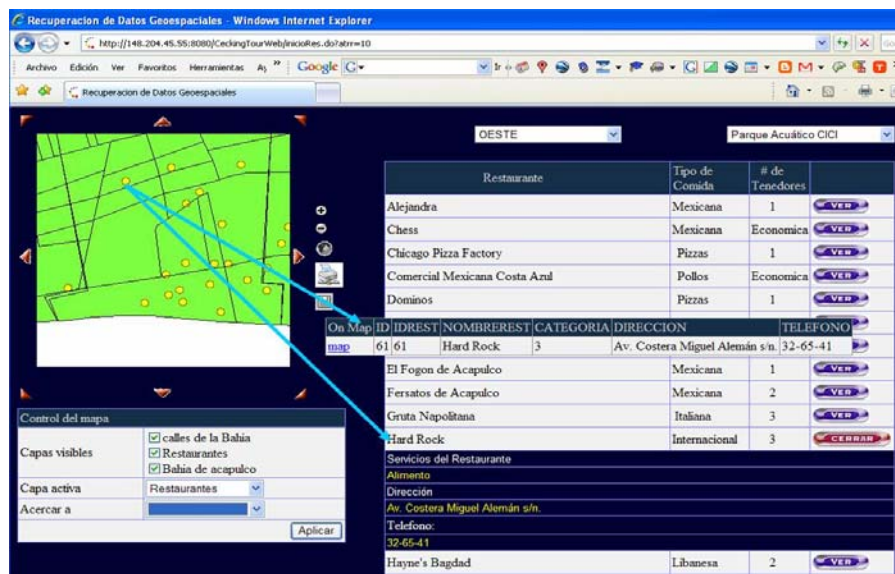

Fig. 9. Web-mapping application that represents instances of concepts

4 Conclusion and future work

In this thesis, a description of GEONTO-MET methodology has been outlined. However, we have only focused on the task of *conceptualization* (Synthesis stage). The advantage of GEONTO-MET is being able to a general framework based on abstract classes to classify and store geographic concepts, defined in a specific context, as well as the explicit relationships. This fact is used to extract the *semantics* that offers the mapping between concepts through their relationships. The *constraints* defined in *Kaab* play an important role to validate the definition of a concept, according to the relationship that corresponds to it. The minimization of axiomatic relationships is focused on *enriching* the conceptualization, directly translating the relationships of the concepts to the ontology. This mechanism offers a higher *semantic granularity* to the conceptualization and it is more flexible for generating ontological descriptions.

The methodology has been designed with a general vision of knowledge representation. It is necessary to do more tests in order to determine its use in other domains. At this time, the geographic domain is totally covered by this approach. Additionally, GEONTO-MET can be compared to the *object-oriented* model; therefore any kind of geographic phenomenon can be ontologically modeled with this approach. We have defined geographic domain ontology as follows: "it is a logical and defined theory in a formal language and represented by an explicit vocabulary, which is based on two sets of axiomatic relationships that are used to reason and understand the entities of the geographic world that can be conceptualized".

As a case study, domain and application ontologies have been designed. They represent the *conceptualization* of each context that defines the semantics of every feature that composes the domain. A mechanism to retrieve geographic concepts based on *Protégé* and *Jena* for accessing the ontologies, by means of SPARQL queries has been also implemented. The retrieved instances of concepts can be visualized in a *web-mapping* application. Templates and descriptions explicitly describe the *properties* and *relationships* of geographic objects represented by concepts, while their *behavior* describes the objects *semantics*.

On the other hand, this approach allows for *integrating* and *sharing* geospatial information. It provides feasible solutions towards these and other related issues. Additionally, by using geographic domain conceptualization it is possible to avoid the *ambiguity* of terms, which represent the *semantics* of the geographic domain in the ontology.

Future works are focused on establishing the guidelines to make *spatial reasoning*. That is, if the behavior of geographic objects involved in an ontology partition is known, then one can understand the *semantics* that they represent and explore the implicit *inference* of the conceptualization. On the other hand, these approaches for processing geospatial data with a *semantic focus*, establish the basis and requirements to produce in short time *iGIS* (intelligent geographic information systems).

References

1. **Bateman, J., Farrar, S.,** "Towards a generic foundation for spatial ontology", *Proceedings of the Third International Conference on Formal Ontology in Information Systems*, Torino, Italy, (2004) 126-138.
2. **Gómez, A., Fernández, M., Corcho, O.,** *Ontological Engineering Second Edition*. Berlin, Heidelberg, New York: Springer-Verlag (2004).
3. **Guarino, N., Welty, C.,** "A Formal Ontology of Properties", *Lecture Notes in Artificial Intelligence,* Vol. 1937, 97-112 (1995).
4. **INEGI,** "Diccionarios de Datos Vectoriales", Instituto Nacional de Estadística, Geografía e Informática (1999).
5. **ISO 19115** (2005), http://standards.iso.org/ittf/PubliclyAvailableStandards/
6. **Mark, D.M., Smith, B., Egenhofer, M., Hirtle, S.,** *"Emerging Research Theme: Ontological Foundations for Geographic Information Science"*. California, United States: University Consortium for Geographic Information Science, Technical Report (2001).
7. **Staab, S., Schnurr, H.P., Studer, R., Sure, Y.,** "Knowledge Processes and Ontologies", *IEEE Intelligent Systems* 16 (1): 26-34 (2001).

8. **Tomai, E., Kavouras, M.,** "From Onto-GeoNoesis to Onto-Genesis. The Design of Geographic Ontologies", *Geoinformatica* 8 (3): 285-302 (2004).
9. **Torres, M.,** *"Representación ontológica basada en descriptores semánticos aplicada a objetos geográficos".* Instituto Politécnico Nacional, Ph. D. Thesis in Spanish (2007).
10. **Torres, M., Levachkine, S.,** "Obtaining semantic descriptions based on conceptual schemas embedded into a geographic context", *Lecture Notes in Geoinformation and Cartography,* Vol. 4977, 209-222, (2007).
11. **Torres, M., Moreno, M., Quintero, R., Fonseca, F.,** "Ontology-driven description of spatial data for their semantic processing", *Lecture Notes in Computer Science*, Vol. 3799, 242-249 (2005).
12. **Uschold, M., Grüninger, M.,** "Ontologies: Principles, Methods and Applications", *Journal of Knowledge Engineering* 11 (2): 93-155 (1996).

Miguel Jesús Torres Ruiz *He is a Professor and Researcher of the Intelligent Processing of Geospatial Information Laboratory in the CIC-IPN. He has published around 90 papers in National and International Journals and Conferences. His research scopes are focused on developing approaches to conceptualize the geographic domain, design of intelligent GIS, ontology-driven information integration and methodologies oriented to the spatial reasoning. Additionally, he is member of the ACM, IEEE, AMIICE, CONACOFI and CONAVEF and he has participated as consultant in several projects for SENASICA-SAPARPA, CENAPRED, and IFE.*

Serguei Levachkine *is a Professor and Researcher of the CIC-IPN and Head of the Laboratory of Intelligent Processing of Geospatial Information. He obtained his Ph. D. degree from the Moscow State University. Additionally, he has published more than 150 papers in National and International Journals and Conferences. His research efforts are mainly in the transition from data to semantics, in a variety of areas such as databases, artificial intelligence, cognitive science, human-computer interaction, information retrieval and integration, intelligent geographic information systems (GIS), and so on.*