Searching Spatial Sense in the Ontological World: Discovering Spatial Objects

Morocho, V and Pérez, L. and Saltor, F.

Research Report LSI-04-30-R

Departament de Llenguatges i Sistemes Informàtics

UNIVERSITAT POLITÈCNICA DE CATALUNYA

Searching the Spatial Sense in the Ontological World: Discovering Spatial Objects

Villie Morocho[1], Lluís Pérez-Vidal[2], and Fèlix Saltor[1]

[1] Departament de LSI-SI, Universitat Politècnica de Catalunya,
Jordi Girona 1-3. E-08034, Barcelona, Spain.
{vmorocho,saltor}@lsi.upc.es
[2] Departament de LSI-IG, Universitat Politècnica de Catalunya,
Av. Diagonal ,647. E-08028, Barcelona, Spain.
lpv@lsi.upc.es

Abstract. The search of semantic understanding in information systems has led us to take advantage from fields like Natural Language processing from the Artificial Intelligence. This is the case of *ontologies* which have been used to solve many problems such as interoperability and integration. This paper presents a point of view over specific-domain ontologies to be used in geospatial activities which are named *spatial ontologies*. An overview of ontologies which could cover this need is presented, focusing on their strengths and weaknesses. In the last part we present the ontology to be used for the semantic integration of spatial schemas in the SIT-SD *(Semantic Integration Tool for Spatial Data)* prototype. This Variable-Level Spatial Ontology is being developed as part of the SIT-SD. In this ontology we have defined the main spatial characteristics to be used in the inference with spatial information.

1 Introduction

The actual trend in the new generation of information systems is toward semantic understanding. In order to face this challenge, various aspects of knowledge and understanding between humans, users-machine and machines have been explored. Knowing that global knowledge is almost impossible to be achieved, it is much rational to divide it in domain-dependent sets of knowledge. Firstly, we present an overview of the development of spatial ontologies. Secondly, we present the construction of one such ontology specially focused on the integration process of spatial database schemas.

Knowledge relative to semantic integrity constraints could be represented inside the ontology. For example: rules that do not allow a building to be intercepted by a street segment; or rules that do not allow a constructed area of a building to be intercepted by another building. Many rules obvious for the human, as above, will be included as additional knowledge for the integrator tool. In Cyc ontology[10], among others, we can find some general knowledge. But in a specific domain, it will be necessary to define its specific knowledge. One way could be by means of relationships. Reed et al.[12] present the *insertion* [1] of geographical entities, FIPS 10-4 ontology, into Cyc. This is the

---

[1] We use the term *insertion* rather than *mapping* to avoid confusion with the terminology used in this paper

simplest form of *ontology mapping* in which the missing terms in the reference Cyc Ontology are easily identified and created. The Reed's work is an insertion of "instances" of some geographic class. In opposite, we propose to increase the semantic knowledge at the type level. For instance, a province should belong only to "one" country. In the section 4 we make a review of some other approaches that have been developed in this direction.

Spatial relationships give more semantic information than relationships from traditional databases. In order to take advantage of this kind of additional knowledge in the geospatial domain, it is necessary to represent it in the ontology in a suitable way. We use in the SIT-SD prototype this kind of information and relationships. This information will be used in the assessment of semantic similarities between classes at the integration level of the federated schema. In this prototype, we have applied some algorithms. Two cases have been covered in previous works [14],[9],[8]. The first case, we applied a word matching between object class and parts of object. The second case, we applied assessment of semantic-neighborhood matching.

The third case, which is the main subject of this paper, appears when we apply an *Variable-Depth Level Spatial Ontology* (VDLSO). VDLSO is being developed as part of the Semantic Integration Tool for Spatial Data (SIT-SD) project [7], [15], where we have defined the main spatial characteristics to be used in the work with spatial objects. In this case, two special subdivisions (from the point of view of spatial features) should be considered: material and immaterial objects. *Material objects*: it is impossible for two such objects to occupy the same space (a downtown area can not share its space with another "different" downtown area). *Immaterial objects*: such objects may occupy the same space (a limit line with a river, or the limit line of a province with a limit line of a country).

The remainder of the paper is organized as follows. Section 2 presents the foundation of spatial ontologies, how a spatial ontology will be used in order to infer with spatial knowledge. Section 3 presents the main aspects to be considered in a developed of a spatial ontology. It show the main difference with another approaches. Section 4 briefly reviews some attempts of including spatial sense in some ontologies. Section 5 presents how we have designed and constructed a spatial ontology to be used in a SIT-SD project. Although the scope of this paper only is the spatial sense. Finally, conclusions and future work are presented in Section 6.

2 Foundation of Spatial Ontologies

There are many definitions and concepts about of an ontology. We agree in that the main purpose of a ontology is to specify *the intended meaning* of a vocabulary, i.e. its underlying conceptualization [4]. Normally, the meaning of the ontology terms, is taken from the consensus of a users community. In our case, we are talking about one particular spatial community. As *spatial data*, we are considering data from applications like Geographic Information Systems, computer-aided design (CAD), robotics, image processing, all of which have at their core spatial objects that must be stored, queried, and displayed. Although in this paper, much of the work will be focused on geospatial knowledge. Moreover, the lines opened in this work could be continued toward applications in very different domains like for instance the microbiology where topological relationships are important too.

The semantic problems appear when the integration of different sources is necessary. *Semantic heterogeneity* have been one of the classical problems in various integration approaches (Multidatabases, Distributed Databases, Federated Databases) and in our days continues to be a challenge to face. Figure 1 presents the main concept of integration with an ontology approach (extracted from Guarino[4]).

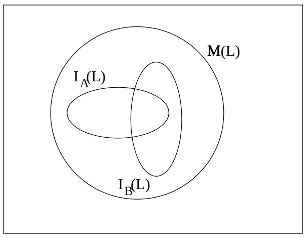

Fig. 1. Two systems A and B using the same language L can communicate only if the set of intended models $I_A(L)$ and $I_B(L)$ associated to their conceptualizations overlap. $M(L)$, set of all models. From Guarino[4]

From the spatial point of view, and specially from the domain of GIS, we should consider some additional aspects in the integration problem. A GIS commonly is an "integration" of many *communities* in a complex system. For instance, a simple touristic map of the city could include data about of streets, transport(subway, subway station, bus route, bus stops), historical buildings, touristic buildings, interest points, gardens and so on. Often the correct design of the map is developed by *layers*[2]. The layers will be: streetLyr, transportLyr, hbuindingLyr, ibuildingLyr, interest_pointLyr, gardenLyr and so on.

We could state that terms used in various layers keep an *univocal sense*, therefore could use the same ontology. However, would it be possible to talk about terms with univocal sense for a whole GIS community or a spatial community? We think that an additional classification is necessary. Some general ontologies have adopted this approach, such as Cyc defines *micro theories*.

All this illustrates that there are many ways of considering the integration problem. The basic case appears when different communities need to communicate, see figure 1. In this case, the communication is possible only if the intended models associated with the conceptualizations do overlap[4]. On the other hand, in a GIS community a common conceptualization over all *themes* may not be possible (see section 3). Or perhaps, it exists only in a very high level of conceptualization. As result we obtain terms which are semantically close for a *theme*, while semantically far for another *theme*.

---

[2] We take the concept of layer from GIS where the layer, commonly, contain similar characteristic objects

For instance, in a design of the same touristic map, for the people designing the `streetLyr`, all plants could be classified as garden. Then, the conceptualization for all of them will be the same, and semantically will be close. Meanwhile, for the people that design the `gardenLyr` the classification will be completely different and with more detail. Obviously, both of them have the same main conceptualization, for both are plants. Then, when could we say that these plants are a garden, and when these plants are a forest, or when these plants are farming? Perhaps adding to the object not only the *word sense*, but also the *spatial sense* (this will be explained in section 3). Let us leave this question in this point in order to complete the example. Normally, each local government cares for this kind of geographic information for each city (it also could be at state level, or country level, and so on). Now, suppose a tourist that need integrated information from city_A and the neighborhood city_B (for example Barcelona and L'Hospitalet). In an utopian and wonderful world, the entire conceptualization of the city_A GIS will correspond to city_B GIS but this is impossible. This paper does not try to solve all problems derived from this integration process. It is the case of situations as, what happend when not correspond the layers from city_A to city_B, and so on.

3 Main aspects to be represented in a spatial ontology

In the representation of geographic information, it is necessary to take into account the main terms defined in this field and how they should be represented in the spatial ontology. (We have reviewed some works for designing our theory such as [13]).

3.1 Theme

In a GIS, the geospatial information corresponding to a particular topic is gathered in a *theme*. It is like defining the context or specific domain of the knowledge. The most common display of a theme is on a map. Maps of topography, railway network, road network, city map and weather are examples of themes over a typical map. Certainly, most ontologies define certain themes in their structure. For example, in Cyc ontology there are micro theories.

3.2 Geospatial Objects

In order to represent entities of the real world, we should abstract up to the conceptual level. A geospatial object corresponds to an entity. A theme is a collection of geospatial objects. Thus, if a geospatial object is formed by two components, the geospatial ontology must be a conjunction of word sense and spatial sense.

- WORD SENSE: Usually a geospatial object is described by a set of descriptive attributes.It should be possible to make inferences with the name of a geospatial object in order to find the semantic of the word. For some years and until our days, most ontologies have been designed to cover enough this sense. Many researchers from the Natural Language processing field have left a valuable legacy that is presently improved to be used in almost all fields of information systems.

SemanticWeb, Semantic Integration, Semantic Understanding and so on. However, the most of these ontologies have few characteristics (or none of them) on spatial sense.

- SPATIAL SENSE: For an object to be considered geospatial, this should have a spatial component. That is to say, if we are talking about a geospatial object, we must be capable of defining *spatial attributes* and *spatial relationships*. Certainly, in this case all entities in our real world could be represented by means of a geospatial object. This representation depends on the degree of transcendence of the object to be represented and, therefore, depends on the the importance of the object for this particular theme. In the beginning, information systems, and more specifically Data Base systems, only have represented the word sense of the entity. In geographical and spatial information, it should search the way for not losing the additional intrinsic semantic information. Therefore, a suitable *spatial ontology* should be capable to express the intended meaning of the terms used by the geospatial community. The ontology can be limited to those structural relationships among terms that are considered relevant for the query. In order to avoid confusion with the terminology used in this paper, we should employ the term *spatial* as a more generic term for objects with spatial component (i.e. such object could be associated with a location relative to the Earth).

  We could conclude from this analysis that, a spatial ontology should cover the following main aspects:
    - Be able to reason or to infer about spatial concepts considering both parts of spatial objects. Although not necessarily with sufficient effectiveness, even if this happens only at the first stage.
    - Takes advantage from engines specialized in spatial reasoning.
    - Links with various databases, thesaurus, and so on, that contain specific information.
    - Supports the query capabilities for spatial information from an ontology. This does not mean queries over the data as such, but to be capable to enrich the query semantically.

  The main semantic information to be represented in a spatial ontology should be the following:
    - TOPOLOGICAL RELATIONSHIPS
      Position Attribute: above, adjacent, below, vertical, horizontal, left, right, near, on. *Spatial Relationship:* inFrontOf, inBackOf, connect, between, distance, cross, through.
    - DIMENSION *Measure:* length, area, angle, and more physical quantity
    - SHAPE Dealing with spatial primitives like line, point, polygon, circle.
    - REFERENCE SYSTEM Latitude and longitude, elevation, altitude.
    - GEOPOLITICAL SUBDIVISION For instance, country contains provinces, country contains states, and so on.

At this point we believe that a spatial ontology that takes advantage of spatial characteristics could aid to solve semantic heterogeneity. For example, if we can infer that the limit of a **park** *normally* is an **street**; or a **subway station** is *always* near, or *often* to over the **subway** but at a different altitude; a **tunnel** is a **railway tunnel** if a

railway crosses through it, or it is a **subway tunnel** if it **overlaps** a **subway**, ori s a **conventional tunnel** if overlaps a **road**. In all of them there are **spatial objects** with topological relationships besides a *frequency* factor which let us assess the similarity among spatial objects.

3.3 The problem of Multi representation and Multi resolution in a spatial ontology

The problem derived from the modeling of our world is the difference in the ways to abstract it. In the modeling process many problems of subjectivity can be included. This fact could yield a wall impossible to cross in the interoperability and integration process. One way to face these problems is, to start the model from a common ontology. Which means every designer should depart from the same knowledge base. Obviously this is an utopia. From the point of view of geospatial objects, there could be additional problems in the representation of *spatial sense*. Multi representation and multi resolution in geographic objects have been challenges faced by many researches [16]. For instance, a city could be represented by a point. But in other abstraction, the same city could be represented by a polygon . Different *shape*, from the *spatial sense* that we are talking in this paper, could be associated to the abstraction of the city. It will depend of the need and the possible use of the information. Therefore, a spatial ontology should be able to deal with *multi-representation* and *multi-resolution* of spatial object. We explain in the section 5 how to face this problem.

4 Reviewing ontologies to be used in spatial inference

There are many attempts to develop ontologies as a tool for dealing with spatial objects. We make a brief analysis of Cyc[5], WordNet[6] and SUMO[11] looking for extensions to add "spatial knowledge" to the ontology.

4.1 OpenCyc

Started in the mid eighties by D. Lenant (Microelectronics and Computer Consortium). It is, maybe, the most commonly used ontology in our days, with enough commonsense knowledge to support natural language. Due mainly to consistency problems coming from a unique huge knowledge base, it was necessary to separate this in "microtheories". Usually, a microtheory is a knowledge domain. However, when it must deal with several microtheories, it is not clear how Cyc solves the inconsistency. This is the case of GIS integration discussed in section 2. OpenCyc, the public version of Cyc technology, now in release 1.0 should include close to 6,000 concepts and 60,000 assertions. At the present time, it covers 90% of the expectations. It supplies other simple categories by means of links to synset structure of WordNet[6].

Cyc has a graph-like structure where the root is `Thing` from which `Individual -> SpatialThing -> SpatialThing-Localized` is derived, see figure 2 (of course we are taking into account only what is of interest to us). Then Cyc defines as `SpatialThing` "The collection of all things that have a spatial extent or location relative to some other SpatialThing or in some embedding space ...". But when we ask for any spatial thing like for example `street`, it only mentions "street is located inside a city" but without spatial attributes because it is only a sentence. It is also true that it uses common-language terms and locutions. We propose the representation of these spatial relationships and attributes inside the ontology suitable to be used in the inference process.

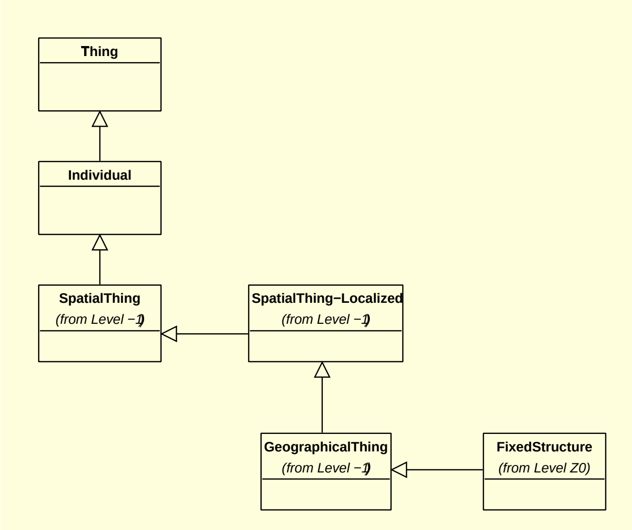

Fig. 2. Basic structure of Cyc from which start our extension

4.2 WordNet

"WordNet is an on-line lexical reference system whose design is inspired by the psycholinguistic theories of human lexical memory". It was developed by the Cognitive Science Laboratory (Princenton University) led by G. Miller [6]. It carries a lexical database with which it is capable to distinguish among the syntactic categories of noun, verb, adjective, and adverb. It could processes natural language although a bigger database will be necessary for a satisfactory result. In its 92 version, it included approximately 95,000 different word forms (51,500 simple words and 44,100 collocations) organized into almost 70,100 word meaning, or sets of synonyms.

WordNet was designed for processing natural language, for which it is a well-suited tool, but this ontology presents lacks on the spatial sense . Thus, WordNet processing is enough to cover the *word sense* needs of our work, but for the *spatial sense* we should look for other solutions.

It is interesting to refer to the EDR Electronic Dictionary too, which was designed to deal with natural japanese language by Yokoi[17]. Cyc and WordNet deal with the english language.

4.3 SUMO an attempt toward Spatial Ontology

Created by the IEEE Standard Upper Ontology (SUO) working group. The Suggested Upper Merged Ontology [11] is an attempt to link categories and relations coming from different top level ontologies. The main focus of this ontology was over Semantic Web area. It appear with important goals and started with the possible spatial sense. There are attempts to deal with spatial information, but the work has stopped (and curiously we can not find news about the ontologies described above in their websites).

SUMO presents two particular ontologies, one for geography and the other for transportation. If we inspect its inner structure, we could find signs of these spatial attempts. But it will be impossible to satisfy the needs presented previously. Therefore, we take advantage of a few characteristics developed here to launch our own vision of *spatial ontology*.

5 Constructing the Spatial Ontology

Reviewing the ontology classification made by Guarino in [4], [3], we collocate our work as a *Variable-Depth Level Ontology* It results of a *high-level* ontology, when we are working with general knowledge, and *low-level* ontology, when we are working with detailed information. This approach is the way more natural of manage geospatial knowledge. Later we will explain our point of view.

In order to formalize our concepts, we abstract the geographical world in some levels. First, the Spatial Object instance (e.g. Barcelona, Diagonal Avenue, Catalunya Square) in which the spatial object is initialized with values. Second, Spatial Object (e.g. city, way, garden) which was defined above (section 2). And third, Spatial Object Class (e.g. city class, street class, green area class) where the spatial objects are grouped by a particular characteristic and behavior.

Furthermore, we define *theme class* as a set of *spatial classes* which depend of a *domain-specific ontology* (e.g. tourism theme class). An instance of tourism theme class could be *tourism of Barcelona*. Thus, the geospatial world will be defined as the set of *theme classes* which depend on a geospatial ontology, or geospatial knowledge base which is the result of a consensus among members of the geospatial community. In figure 3 we show our approach.

In our approach, a *spatial object* could belong to more than one *theme class*. Although, the meaning of the term used to define this spatial object, will depend on the ontology behind it.

For instance, the spatial object **city** could belong to some *theme classes* but the meaning and the representation will be different (remember the problem of multi-resolution and multi-representation discussed in section 3.3). It is clear, that the upper part of the ontology for the **city** term in the tourism theme is irrelevant. And only a few levels of *generalization* are necessary (such as Localized-Thing and Spatial-Thing,

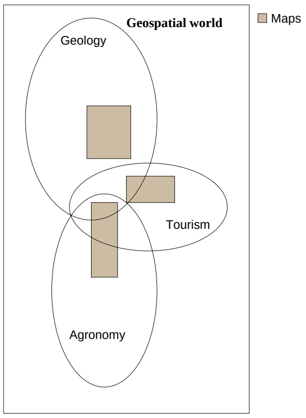

Fig. 3. The geospatial world defined as set of theme classes

see the OpenCyc base structure of USO in figure 2). Conversely, the lower part of the ontology should convey more levels of *specialization* in order to define the knowledge necessary at this level of representation (figure 4). Thus, it will be necessary to prune the ontology [2] . Furthermore, the previous conception of the geometry representation could be different. In the example, at level of tourism theme the city itself is not represented. All the spatial objects in the map, should belong to the same city. On the other hand, the city in a different theme could be represented by a point class from the point of view of the geometry.

Representing the geometry in the ontology Why is it necessary to represent the geometry in the abstract model? We believe that *if we want to take advantage from the spatial characteristics of the objects then the geometry should be represented*. In the past, some attempts for representing the geometry in the abstract model have been done. We take into account the work of Borges et al.[1] where primitives with geometry are defined inside the model. They use UML notation but add a stereotype in order to use a small figure in the corner to say what is the geometry of the object. See the figure 5 to get an idea of this.

Therefore in our ontology, geometry should also be represented. But, it should define certain levels of abstraction in order to enable *multi-representation* and *multi-resolution*. Thus, the ontology will represent two dimensions: level of abstraction and level of generalization/specialization. Both of them should be linked. It is like to*have a view of the ontology which depends on the level of abstraction*. In figure 6 we show this approach. It is similar to the use of a GIS graphical tool. When the scale produces a zoom in, the abstract detail will be high, and then more *specialization* in the ontology is necessary (therefore, a process of adding low level). In contrast, when the scale produces a zoom out, the abstract detail will be low, then more *generalization* is necessary (therefore, a process of *pruning* sheets and to add high level on the ontology).

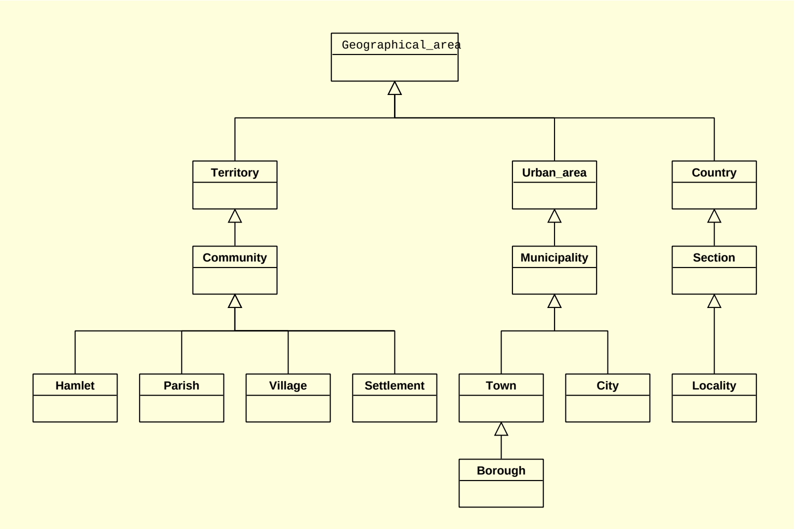

Fig. 4. Down part of the ontology for **city**

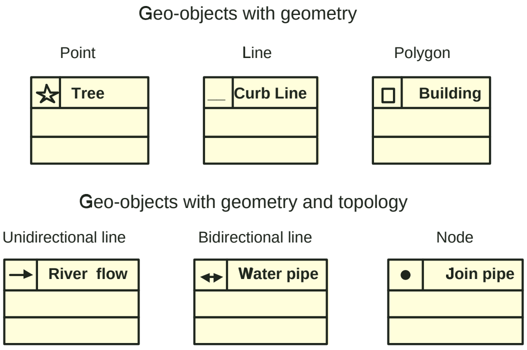

Fig. 5. Primitives from the OMT-G model where the geometry is represented at abstract level

Note also, that **every spatial object should be considered like a multi-classification object. Because every spatial object should belong at least to both spatial sense and word sense**.

Facing the Multi-representation and Multi-resolution To conclude, why is it important to define the level of abstraction? Because it should apply the topological rules in concordance with the level of abstraction. In figure 6 we show a sample of abstraction levels in our ontology. In the figure, it is possible to distinguish the specialization for **building class** (bottom part in the figure). Of course, with this vision, when the abstract level is high, new relationships with another level of knowledge should appear. One example is the case where the abstract level is a city level. Then, the knowledge level in the ontology should be at the buildings level and its specialization. But when it is at the country level, the building specialization should be *pruned* and the relationships with another high level of knowledge should appear. Thus, a dynamic view of the knowledge, that is a *variable-depth ontology*, will be necessary. In our case, we solve in part this dynamic need by including in the ontology a *domain-level*. In other words, it is a new domain of knowledge. We have translated the *domain-levels* to OpenCyc like a micro theory. We have used a graphic representation of an ontology by means of Unified Modeling Language (UML). We have chosen this graphic representation in collaboration with the ODISSEA project (http://www.lsi.upc.es/recerca/esp/recerca-esp.html) in order to translate from OpenCyc to XML Metadata Interchange(XMI).

In this graphical representation we have different packages for representing the abstraction levels. In our approach, the way of pruning the ontology is defining levels previously. Thus, when we need a particular level, we will take into account only this domain-level (microtheory in OpenCyc, and Package in UML representation, see figure 6). In conclusion, we have defined our own ontology extending a microtheory in Cyc with new spatial characteristics and taking into account a standard for geographic information as SDTS.

Conceptual Generalization In order to represent geometry inside of the ontology, we define the basic geometry of the class as the start point, and the possible geometry transformations of the class. For example, in the case of a building class, the initial type for the geometry will be polygon. The possible transformations will be from polygon class to point class (see table 1).

Therefore, it is necessary to record the variation according to geometric shape in order to know the possible representation depending on the shape and on the scale. This phenomena could occur in two representation variations: according to *geometric shape* and according to *scale*. In figure 7 we show the possible variations of the representation of a building, and a city.

In order to make inferences with spatial objects, we could define rules depending on our spatial objects. In this case we show an example using spatial objects from a touristic map city. In the theme we have five “basic” spatial objects which will be discovered. The following are the necessary rules:

“If we could store the knowledge related to behavior among spatial classes represented in the ontology, then the entire knowledge will aid to discover spatial ob-

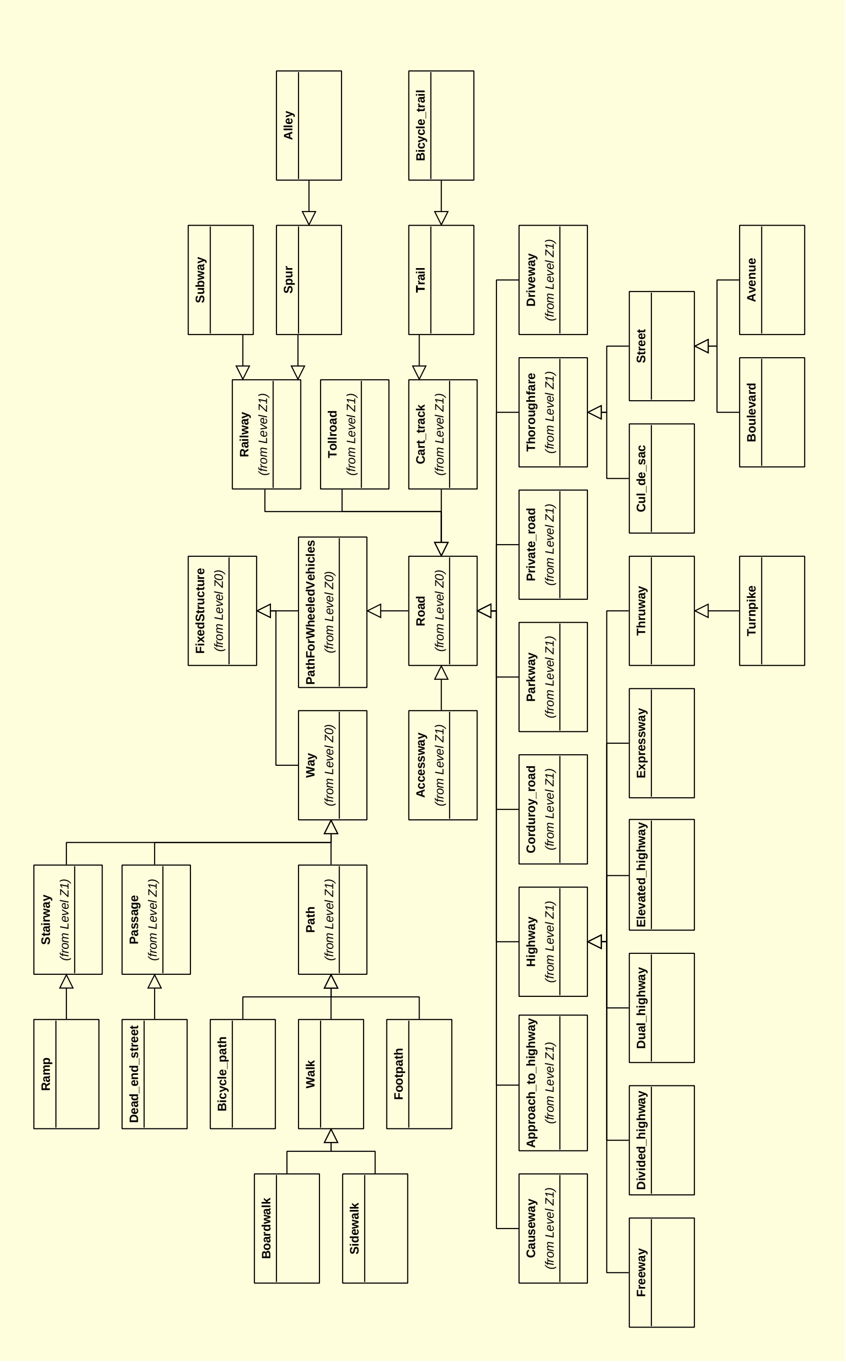

Fig. 6. Ontology for **road** network extending OpenCyc with SDTS Standard

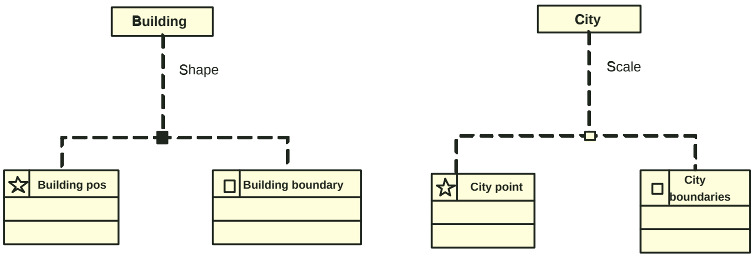

Fig. 7. (1) Variation according to shape (overlapping). (2) Variation according to scale (disjoint)

Table 1. Transformation from polygon class to point class

—Let, ss (spatial sense) be a function that returns the *shape value* of the spatial object. And, $r = 0..n$ (n possibles geometry representations of the spatial object). If the spatial object has not transformations then a unique value of $r = 0$ exists. The function pos returns the localization of the spatial class. Let cg be a function that returns the center of gravity of the polygon.
B^a= spatial object (both, spatial sense and word sense)
B = spatial sense(B^a)
B^w = word sense(B^a)
$ss_r(spatialobject) = shape\_type$
—Then:
$B = B_{t0} = ss_0(Building) = Polygon\_class$
$B_{t1} = ss_1(Building) = Point\_class$
—And, the rule of transformation in this example, will be the following:
$pos(B_{t1}) = pos(cg(B))$

Table 2. Using *wellDef* function in order to valid two spatial objects

---

—Let B^a and B'^a be two objects that belong to a building class and, S^a to a street class, U^a to a subway station class, UL^a to a subway class, BS^a to a bus stop class. These objects are defined into our ontology as parents of the specialization sheets. Then, the rule could be defined using this class, because every element down to these elements must follow the same rule.
—Also, the start geometry is the following:
B, B' belong, to polygon class
S to line class
U to point class
UL to line class
BS to point class
—Then:
$B\ wellDef\ S$ = TRUE $\Leftrightarrow (B^o \cap S^o) = \emptyset$,
$B\ wellDef\ B'$ = TRUE $\Leftrightarrow ((B^o \cap B'^o) = \emptyset)$,
—Let C be a buffer, created at a distance ***dist*** around BS
—Then:
$BS\ wellDef\ S$ = TRUE $\Leftrightarrow BS\ near(dist)\ S \Leftrightarrow S \cap C \neq \emptyset$,
—Let D be a buffer, created at a distance ***dist'*** around U
$U\ wellDef\ UL$ = TRUE $\Leftrightarrow U\ near(dist')\ UL \Leftrightarrow U \cap D \neq \emptyset$

Where *wellDef* is a function which verifies if the spatial object related to another is well-defined.

jects, and therefore spatial classes and schemas". This kind of topology rules could be grouped by the main geometry class to which they belong. In *Point Rules, Line Rules*, and *Polygon Rules*.

5.1 Topological rules in the ontology to discover spatial objects

In real life, the knowledge about topological rules is a big help to discover spatial objects. For example, when somebody takes a touristic map, a "basic reasoning" is developed to discover the streets and blocks: ***a*** *street does not cross a building* (it is the fundamental form of reasoning). It could take advantage from spatial characteristics of the object in order to achieve the same objective by a machine. We could assume the possibility of obtaining the basic structure of an entity. That means, the possibility to make processing of data in order to use the topological rules to discover spatial objects.

In order to illustrate our approach, we will introduce some *axioms*. These axioms should give us the basis and the start point in the algorithm for discovering spatial objects.

Basic assumptions (axioms) These axioms should be used for discovering spatial objects in the domain of touristic city maps tcm, as following:

- Every map that belong to a *theme*, should have at least one `spatial class` common to any other map classified in the same *theme*. Furthermore, such spatial class will be the easiest to discover.

- Based on experience, we have deduced that a tcm should have at least one **spatial object** member of the `Street class`. Therefore, it is possible to find a spatial class member of `Axis class` or `CenterLine class`[3].
- In every tcm, a **spatial object** member of `centerLine class` always crosses the boundaries of the map at least one time.

The process to discover the spatial objects will be divided in the following steps:

- By the first axiom there should be at least one basic spatial class for each theme. And this class is the easiest to recognize.
- It is possible to apply topological rules from this basic class in order to discover the most common classes.
- The number of the most common classes in a spatial system of a certain theme is limited and is known.
- It is possible to continue increasing the knowledge in the ontology, with new topological rules and new spatial objects in order to recognize more specialized classes in a spatial system.

From our example to recognize a touristic city map, the process would be as following:

1. Searching the base axis for streets: Any map with streets should have the base axis. Normally, the distance between two consecutive crossings in the same line should be around 100m. Often, a line is not crossed by any other more than one time.
2. Searching of parcels: Often, a parcel is enclosed by a set of streets.
3. Searching buildings: Normally, a bounding is enclosed by a parcel.

Topological rules between feature classes Everything leads to define rules between spatial objects. Normally these rules should be implemented in GIS tools in order to avoid data inconsistence at the time of input data. The same rules could be part of our ontology. Thus, a process of feedback could be defined to take advantage from the fact that these same rules will be defined in the process of input. These rules will be grouped by the main geometry class which is part of the rule. The tables 3, 4, 5 show the rules ($\star$ = point,$-$=line, $\square$=Polygon geometry type)[4].

In order to avoid confusion with term used in some software products, we take the definition of feature class from OpenGIS and ESRI. *Feature* is the representation of a real-world object on a map. a group of spatial objects which together represent a real-world entity (e.g. a complex feature is a road network). *Feature class* is a collection of geographic features with the same geometry type (such as point, line, of polygon), the same attributes, and the same spatial reference. That is, taking our definition of spatial objects, the geometry representation of the spatial sense. Therefore, feature class allow homogeneous features to be grouped into a single unit. From the point of view of our ontology, the feature class will be the *spatial objects with the same parent*. In our ontology (figure 6), **road** and its children will belong to the same feature class.

---

[3] These two later classes are the base lines when a map is digitalized.

[4] This tables were based in the topology rules from some GIS software products

All rules could be *overload*, that is say, if the spatial object have different shape representation, it could be the same restriction. Thus, if a building (B) is represented by polygon the rule should be, related to parcel (C), $BcoveredC$ (table 5). On the other hand, if it is represented by a point the rule should be from table 3

Thus, the rules for the main spatial objects in a touristic map will be well defined as following:

$B\ wellDef\ S = TRUE \Leftrightarrow (B\ notoverlap\ S)$
$B\ wellDef\ B' = TRUE \Leftrightarrow (B\ notoverlap\ B')$
$BS\ wellDef\ S = TRUE \Leftrightarrow (BS\ covered\ S)$
$U\ wellDef\ UL = TRUE \Leftrightarrow (U\ covered\ UL).$

This knowledge was introduced in the ontology by means of rules in the proprietary language of Cyc. By means of inference could be discover the spatial objects in the our example.

6 Conclusions and future work

In this work we have presented the foundations of spatial ontologies, taking into account the main differences with classical ontologies. In our case, we have faced the multi-resolution and mluti-representing of spatial objects by a division in *ontology levels*, thus as representing this ontology by means UML diagrams. It take advantage of a tool which convert Cyc structure in a XMI file capable to be used by UML tools. We have developed a *variable level ontology* based in Cyc structure capable to use spatial characteristics for inferring. Thus, the ontology generated can work with the *spatial sense* and *word sense*. Such behavior is the main difference with another approaches. This ontology was developed as part of Semantic Integration Tool for Spatial Data (SIT-SD) prototype. A formalization of topological rules also have been presented. By means of this rules the ontology could be enriched with much knowledge. It knowledge is used in order to discover spatial objects by means of their spatial characteristics, in contrast with the word sense of classical ontologies. As future work, it is necessary to adapt a process into geographical tools (such as ArcGIS) where the technical people is introducing topological rules. This rules will feed the ontology in order to obtain a real applicable ontology capable to incorporate enough knowledge for applications. All of which have at their core *spatial knowledge*.

Acknowledgments

Part of this work has been supported by: Politechnic University of Catalonia; FEDER and MCYT under the project TIC2001-2099-C03-01. Special acknowledgments to European Environment Agency and ESRI Spain for their valuable contributions for this research.

References

1. K. A. Borges, C. A. Davis, and A. H. Laender. OMT-G: An object-oriented data model for geographic applications. *GeoInformatica*, 5(3):221–260, Sep 2001.

2. J. Conesa, X. de Palol, and A. Olivé. Building conceptual schemas by refining general ontologies. In *Proceedings of DEXA'03*, pages 639–702. Springer, sep 2003.
3. N. Guarino. Semantic matching: Formal ontological distinctions for information organization, extraction, and integration. In *Information Extraction, International Summer School*, LNCS, pages 139–170. Springer, 1997.
4. N. Guarino. Formal ontology in information systems. In N. Guarino, editor, *Formal Ontology in Information Systems*, pages 3–15, Trento, Italy, Jun 1998. FOIS'98, IOS Press.
5. D. Lenat and R. Guha. *Building Large Knowledge Based Systems: Representation and Inference in the Cyc Project*. Reading, Mass,Addison-Wesley, 1990.
6. G. Miller, R. Beckwith, C. Fellbaum, D. Gross, and K. Miller. Introduction to wordnet: An on-line lexical database. *International Journal of Lexicography*, 3(4):235–244, 1990.
7. V. Morocho, L. Pérez-Vidal, and F. Saltor. Semantic integration on spatial databases: SIT-SD prototype. In *proceedings of VIII Jornadas de Ingeniería del Software y Bases de Datos*, pages 603–612, Alicante, Spain, Nov 2003. ISBN:84-688-3836-5.
8. V. Morocho, F. Saltor, and L. Pérez-Vidal. Ontologies: Solving semantic heterogeneity in federated spatial database system. In *Proceedings of 5th International Conference on Enterprise Information System*, pages 347–352, Angers, France, Apr 2003.
9. V. Morocho, F. Saltor, and L. Pérez-Vidal. Schema integration on federated spatial db across ontologies. In ***P**roceedings of the 5th International Workshop on Engineering Federated Information Systems EFIS*, pages 63–72, Coventry, UK, Jul 2003. IOS Press.
10. OpenCyc. Open cyc. http://www.opencyc.org.
11. A. Pease, I. Niles, and J. Li. The suggested upper merged ontology: a large ontology for semantic web and its applications. In *Working Notes of the AAAI-2002 Workshop on Ontologies and the Semantic Web*, Edmonton, Canada, Jul 2002.
12. S. L. Reed and D. B. Lenat. Mapping ontologies into Cyc. In *AAAI Workshop on Ontologies and the Semantic Web*, pages 1–6. AAAI Press, 2002.
13. P. Rigaux, M. Scholl, and A. Voisard. *Spatial Databases with Application to GIS*. Morgan Kaufmann, 2002.
14. M. A. Rodríguez and M. J. Egenhofer. Determining semantic similarity among entity classes from different ontologies. *IEEE Transactions on Knowledge and Data Engineering*, 15(2):442–456, Mar 2003.
15. SIT-SD. SIT-SD Semantic Integration Tool for Spatial Data. http://www.lsi.upc.es/events/sitsd, Jul 2003.
16. S. Spaccapietra, C. Parent, and C. Vangenot. GIS databases: From multiscale to multirepresentation. In *4th International Symposium, SARA 2000*, volume 1864 of *Lecture Notes in Computer Science*, pages 57–70, Texas, USA, Jul 2000. Springer.
17. T. Yokoi. The edr electronical dictionary. ***C**ommunications of the ACM*, 38(11):42–44, 1995.

Table 3. Topology rules, point related, between feature classes

Let $P, P', L, L' and Y, Y'$ be three pairs of spatial objects belonging to point, line and polygon class respectively.
Define $\sqcap$ as a function to able to return the feature line or point if exist *self* orverlap. The meaning of he symbol in the left part of the table when there are two equal symbols (e.g. $[\star : \star]$) is that the each feature belong to different feature class. Therefore, at the z0 level are different parents. Obviously, different geometry type will lead to different feature class. Thus, be fc a function that return the feature class of the spatial object. For example, from the ontology at figure 6 $fc(Highway) = Road$.
Then:

Point Rules

$[\star : \Box]$	point inside polygon $(P\ covered\ Y) = TRUE \Leftrightarrow (P \cap Y^o = P) \wedge (P \cap \partial Y = \emptyset)$ e.g. Country Capital must be inside each country
$[\star : \Box]$	point be covered by boundary of polygon $(P\ cover\ Y) = TRUE \Leftrightarrow (P \cap Y^o = \emptyset) \wedge (P \cap \partial Y = P)$ - Utility service points might be required to be on the boundary of a parcel
$[\star : -]$	point be covered by endpoint of line $(P\ covered\ endp(L)) = TRUE \Leftrightarrow pos(P \cap L) = pos(start(L)) \vee pos(end(L))$ - Street intersection must be covered by the endpoints of street centerline
$[\star : -]$	point must be covered by line $(P\ covered\ L) = TRUE \Leftrightarrow (P \cap L) = P$ - Bus Stop/Station must fall along Bus Route

Table 4. Topology rules, line related, between feature classes

Line Rules

$[-]$	lines must not overlap $(L\,notoverlap\,L') = TRUE \Leftrightarrow (L \cap L' = \emptyset) \vee ((L \cap L' \neq \emptyset) \wedge dim(L \cap L') = 0)$ - Sidewalk borders normally are lines cannot overlap.
$[-]$	lines must not intersect $(L\,notcross\,L') = TRUE \Leftrightarrow (L \cap L' = \emptyset) \vee ((L \cap L' \neq \emptyset) \wedge (dim(L \cap L') = 0) \wedge (pos(L \cap L') = pos(start(L') \vee end(L'))))$ - Segments could never cross or occupy the same space with other lines.
$[-]$	lines must not have pseudo-nodes $(L\,notpseudonode\,L') = TRUE \Leftrightarrow dim(L \cap L') = 0 \wedge pos(start(L) \vee end(L)) = pos(start(L') \vee end(L'))$ - Segments of a river system might to only have nodes at endpoints of junctions.
$[-]$	lines must not self overlap $(L\,notselfoverlap\,L') = TRUE \Leftrightarrow dim(L \sqcap L) = 0$ - For transportation analysis, street and highway segments of the same object should no overlap themselves.
$[-]$	lines must not self intersect $(L\,notselfcross\,L') = TRUE \Leftrightarrow (L \sqcap L) = \emptyset$ - Contour lines cannot intersect themselves.
$[-]$	lines must not intersect or touch interior $(L\,notcross_{t}ouch\,L') = TRUE \Leftrightarrow (L \cap L' = \emptyset) \vee ((L \cap L' \neq \emptyset) \wedge (pos(start(L) \vee end(L)) = pos(start(L') \vee end(L'))))$ - When the lines should touch at their ends and not intersect of overlap.
$[- : -]$	line must not overlap with another line $(L\,notoverlap\,L') = TRUE \Leftrightarrow (L \cap L' = \emptyset)$, where $fc(L) \neq fc(L')$ - Highway can cross and come close to rivers, but road segments cannot overlap their segments
$[- : -]$	line must be covered by another line $(L\,covered\,L') = TRUE \Leftrightarrow (L \cap L' = L)$, where $fc(L) \neq fc(L')$ - Lines that make up bus routes must be on top of lines in a road network.
$[- : \star]$	line end point must be covered by point $(end(L)\,covered\,P) = TRUE \Leftrightarrow (pos(start(L) \vee end(L)) = pos(P))$ - End points of secondary electric lines must be capped by either a transformer or meter.
$[- : \square]$	line must be covered by boundary of polygon $(L\,covered\,Y) = TRUE \Leftrightarrow (L \cap \partial Y \neq \emptyset)$ - Polylines used for displaying block and lot boundaries must be covered by parcel boundaries.

Table 5. Topology rules, line related, between feature classes

Polygon Rules

$[\square]$	polygons must not overlap $(Y\ notoverlap\, Y') = TRUE \Leftrightarrow (Y^o \cap Y'^o \neq \emptyset)$, where $fc(Y) = fc(Y')$ e.g. A voting district map cannot have any overlaps in its coverage
$[\square]$	polygons must not have gaps Let, $superY = Y \cup Y^1 \cup Y^2 \dots Y^n$ where $fc(Y) = fc(Y^1) = fc(Y^2) = \dots = fc(Y^n)$, and $(U) =$ universe set Then: $(notgap(superY) = TRUE) \Leftrightarrow (superY \cap \mathcal{U} = superY)$ - Soil polygons that cannot include gaps nor forms void.
$[\square : \star]$	polygon contains point $(Y\ cointain\ P) = TRUE \Leftrightarrow (Y^o \cap P \neq \emptyset)$ - Parcels must contains at least one address point
$[\square : -]$	polygon boundary must be covered by line $(Y\ covered\, L) = TRUE \Leftrightarrow (\partial Y \cap L = \partial Y)$ - Major road lines form part of outlines for census blocks
$[\square : \square]$	polygon must cover to another polygon $(Y\ cover\ Y') = TRUE \Leftrightarrow (Y \ni Y')$ - Autonomous Communities cover to Provinces.
$[\square : \square]$	Must be covered by $(Y\ covered\, Y') = TRUE \Leftrightarrow (Y \in Y')$ - Provinces must be covered by Autonomous Communities
$[\square : \square]$	polygon must not overlap with another polygon $(Y\ notoverlap\, Y') = TRUE \Leftrightarrow (Y \cap Y') = \emptyset$, where $fc(Y) \neq fc(Y')$ - Lakes and land parcels from two different feature classes must not overlap.
$[\square : \square]$	Must cover each other $(Y\ coincide\ Y')(Y \cap Y' = Y) \wedge (Y \cap Y' = Y')$, where $fc(Y) \neq fc(Y')$ - Vegetation and soils must cover each other.

Where: Objects are indicated by upper-case letters (e.g. A, B), their boundaries are denoted as ∂A, and their interior area as A^o(therefore $A^o = A - \partial A$). The boundary's point object is considered to be always empty (hence the point is equivalent to its interior), and the boundary's line is comprised of its two endpoints. A function called dim, is used to return the dimension of an object, and returns 0 if the object is a point, 1 if it is a line, or 2 if it is a polygon.

Departament de Llenguatges i Sistemes Informàtics
Universitat Politècnica de Catalunya

Research Reports - 2004

- LSI-04-1-R : *Automatic Generation of Polynomial Loop Invariants: Algebraic Foundations*, Rodríguez, E. and Kapur, D.
- LSI-04-2-R : *Comparison of Methods to Predict Ozone Concentration* , Orozco, J.
- LSI-04-3-R : *Towards the definition of a taxonomy for the cots product´s market* , Ayala, Claudia P.
- LSI-04-4-R : *Modelling Coalition Formation over Time for Iterative Coalition Games*, Mérida-Campos, C. and Willmott, S.
- LSI-04-5-R : *Illegal Agents? Creating Wholly Independent Autonomous Entities in Online Worlds*, Willmott, S.
- LSI-04-6-R : *An Analysis Pattern for Electronic Marketplaces*, Queralt, A. and Teniente, E.
- LSI-04-7-R : *Exploring Dopamine-Mediated Reward Processing through the Analysis of EEG-Measured Gamma-Band Brain Oscillations*, Vellido, A. and El-Deredy, W.
- LSI-04-8-R : *Studying Embedded Human EEG Dynamics Using Generative Topographic Mapping*, Vellido, A. and El-Deredy, W. and Lisboa, P.J.G.
- LSI-04-9-R : *Similarity and Dissimilarity Concepts in Machine Learning*, Orozco, J.
- LSI-04-10-R : *A Framework for the Definition of Metrics for Actor-Dependency Models*, Quer, C. and Grau, G. and Franch, X.
- LSI-04-11-R : *QM: A Tool for Building Software Quality Models*, Carvallo, J.P. and Franch, X. and Grau, G. and Quer, C.
- LSI-04-12-R : *COSTUME: A Method for Building Quality Models for Composite COTS-based Software Systems*, Carvallo, J.P. and Franch, X. and Grau, G. and Quer, C.
- LSI-04-13-R : *Enabling Collaboration in Virtual Reality Navigators*, Theoktisto, V. and Fairén, M. and Navazo, I.
- LSI-04-14-R : *DesCOTS: A Software System for Selecting COTS Components*, Carvallo, J.P. and Franch, X. and Grau, G. and Quer, C.
- LSI-04-15-R : *Evaluation and symmetrisation of alignments obtained with the Giza++ software*, Lambert, P. and Castell, N.
- LSI-04-16-R : *A note on the use of topology extensions for provoking instability in communication networks*, Blesa, M.J.
- LSI-04-17-R : *An ISO/IEC-compliant Quality Model for ER Diagrams*, Costal, D. and Franch, X.
- LSI-04-18-R : *A Case Study on Pruning General Ontologies for the Development of Conceptual Schemas* , Conesa, J.
- LSI-04-19-R : *Adding Efficient and Reliable Access Paths to the JCF*, Marco, J. and Franch, X.

- LSI-04-20-R : *Exploiting Simple Corporate Memory in Iterative Coalition Games*, Mérida-Campos, C. and Willmott, S.
- LSI-04-21-R : *On the Semantics of Operation Contracts in Conceptual Modeling* , Queralt, A. and Teniente, E.
- LSI-04-22-R : *Complexity issues on bounded restrictive H-coloring*, Díaz, J. and Serna, M. and Thilikos, D.M.
- LSI-04-23-R : *Chromatic number in random scaled sector graphs*, Díaz, J. and Sanwalani, V. and Serna, M. and Spirakis, P.
- LSI-04-24-R : *Bounds on the bisection width for random d-regular graphs*, Díaz, J. and Serna, M. and Wormald, N.C.
- LSI-04-25-R : *Open Source environment to define constraints in route planning for GIS-T*, Pérez, L. and Silveira, A. da M.
- LSI-04-26-R : *A basic repository of operations for the refinement of general ontologies*, de Palol, X.
- LSI-04-27-R : *Tetrahedral mesh subdivision based on underlying volume data*, Rodríguez, L. and Navazo, I. and Vinacua, A.
- LSI-04-28-R : *The Price of Connectedness in Expansions*, Fomin, F.V. and Fraigniaud, P. and Thilikos, D.M.
- LSI-04-29-R : *Smaller kernels for hitting set problems of constant arity*, Nishimura, N. and Ragde, P. and Thilikos, D.M.
- LSI-04-30-R : *Searching Spatial Sense in the Ontological World: Discovering Spatial Objects*, Morocho, V and Pérez, L. and Saltor, F.

Hardcopies of reports can be ordered from:

Núria Sanchez
Departament de Llenguatges i Sistemes Informàtics
Universitat Politècnica de Catalunya
Campus Nord, Mòdul C6
Jordi Girona Salgado, 1-3
03034 Barcelona,Spain
nurias@lsi.upc.es

See also the Departament WWW pages, http://www.lsi.upc.es/