ABSTRACT of PhD THESIS

Semantic Representation of Raster Spatial Data

Representación Semántica de Datos Espaciales Raster

Rolando Quintero Téllez
Graduated on November 08, 2007
Centro de Investigación en Computación, Instituto Politécnico Nacional, México D.F., México
rquintero@ipn.mx
Advisor: Serguei Levachkine
Centro de Investigación en Computación, IPN, México D.F., México
sergei@cic.ipn.mx

Abstract. When people think spatially, they do not usually consider geographic coordinates nor projections. Facing questions having a spatial sense, people do not answer with maps or coordinates, but use some references whose spatial location is "well known". For instance, the answer of a conventional geographic information system to the question "Where is the CIC?" would be "in coordinates 19.50314°N, 99.14759°W". In contrast, a person would answer "in Zacatenco" or "near to Eje Central". The semantic processing attempts to enrich an abstraction level similar to the one that people use commonly. This processing, applied to spatial data, does not depend on scales, resolutions, projections or others that are fundamental in conventional systems. We assume that the first step for making semantic processing is the semantic description of "raw" spatial data. Such description is the identification of the objects contained in data and the location of such objects within a conceptual framework, where they get a meaning. In this work, we present a methodology for making this semantic description using as a case study the digital elevation models. The methodology is build up of three stages: conceptualization, to define the conceptual framework of the description; synthesis, to process "raw" spatial data and to obtain the spatial objects contained in data; and description, to generate the representation of results from the synthesis according to the conceptual framework.

Keywords: semantic, knowledge, representation, ontology, raster spatial data.

Resumen. Cuando las personas pensamos espacialmente, no lo hacemos teniendo en mente cosas como coordenadas o proyecciones; ante preguntas que tienen una connotación espacial, las personas no contestamos con mapas ni con coordenadas; contestamos con referencias a objetos cuya ubicación es "bien conocida". Por ejemplo, ante la pregunta "¿Dónde está el CIC?", la respuesta de un sistema de información geográfica convencional sería "en 19.50314°N, 99.14759°O", pero una persona nos diría "en Zacatenco" o "cerca del Eje Central". El procesamiento semántico trata de alcanzar un nivel de abstracción parecido al que las personas utilizamos. Este tipo de procesamiento, aplicado a datos espaciales, no depende de escalas, resoluciones, proyecciones ni de ninguna de esas cosas que son importantes en los sistemas convencionales. Consideramos que el primer paso para realizar procesamiento semántico es la descripción semántica de los datos espaciales "crudos", dicha descripción consiste en la identificación de los objetos contenidos en los datos y la ubicación de esos objetos en un marco conceptual dentro del cual obtienen un significado. En este trabajo se presenta una metodología para realizar esta descripción semántica; utilizando como caso de estudio los modelos digitales de elevación. La metodología consta de tres etapas: la conceptualización, en la que se define el marco conceptual para la descripción; la síntesis, en la cual se procesan los datos espaciales "crudos" y se obtienen los objetos espaciales que contienen; y la descripción, en la que se realiza la representación de los resultados de la síntesis en términos del marco conceptual.

Palabras clave: semántica, representación, conocimiento, ontologías, datos espaciales raster.

1 Introduction

In this work we present a methodology for describing semantically spatial objects contained within a Raster Spatial Data Set (RSDS), particularly within Digital Elevation Models (DEM). We attempt to make a description based on the knowledge that people have about spatial things, like things we can see on a landscape. We propose a methodology based on three stages: conceptualization, synthesis and description.

On conceptualization stage, we attempt to capture the knowledge about the domain of problem. In other words, we must find and define the concepts used while people talk or think about landforms. The synthesis stage is the numeric one; we have many algorithms for extracting features from the RSDS. The stage is made in the way images are commonly processed, having pre-processing, processing and post-processing phases. As result of this stage we have parts of the RSDS, called “extracts”, that we consider to be an instance of a concept. In description stage we determine what an “extract” is, and build its semantic representation.

1.1 Previous works

Many of existing works are guided from a numeric point of view and a big number of them are focused on the flow analysis and extraction of drainage lines (Ackermann 1993; Hodgson 1995; Etzelmüller and Sulebak 2000). Also, we have explored some other areas related to landform analysis and processing; we find that geo-morphometry has been deeply studied by researchers, but almost always with a numeric approach (Weibel and DeLotto 1988; Etzelmüller and Sulebak 2000; Sulebak, Tallaksen et al. 2000); only some works used “categories” or “classes” for making analysis. We have studied approaches about conceptualizations and ontologies, where we have found philosophical works that lies with existence and reality. Those works serve as basis for the knowledge representation (Smith and Mark 2001; Smith and Mark 2003). Similarly, we have studied works about ontologies, in practical and philosophical terms; as a way for modeling and understanding reality.

2 Methodology

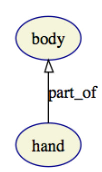
We propose to make a semantic representation of spatial data, which is composed of three stages: conceptualization, synthesis and description.

2.1 Conceptualization

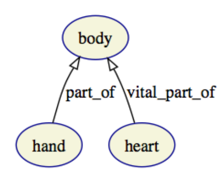
The conceptualization stage has two parts, represented by means of ontologies: high-level and domain. Below the methodology for conceptualizing the geographic domain will be outlined (Quintero 2007, Torres 2007). The main part of the methodology is the idea of minimizing the number of axiomatic relations used to define concepts and relationships. We propose a set with only three fundamental axiomatic relations (is, has and does) and a set of auxiliary relations (prepositions). It could seem that the reduction of axiomatic relations restricts the expressiveness of the conceptualization. Nevertheless, we state that the number of relations to make conceptualization could be bigger than in schemas where all relations are pre-defined. This brings with it two advantages: first, it can be defined as many “classic relations” as needed and, second, relations have a semantic associated to them, because they are defined by means of other concepts or relations. For instance, let us consider a widely used relation: “part_of”. Consider the following statement: “heart part_of body” (Figure 1a), in which there are two concepts (heart and body) and one axiomatic relation (part_of). Now, if we consider that “heart vital_part_of body” (Figure 1b), it is necessary to include “vital_part_of” in the set of axioms. Using the proposed approach, the first statement would be “heart is part of body” (Figure 1c), where we involved three concepts (heart, body and part) and two axiomatic relations (is and of). In order to add the second statement, with our approach, we have only to define the concept “vital_part” (that could be inherited from the concept “part” and the concept “vital”) and use it in the form described previously (Figure 1d).

Therefore, we propose that for conceptualizing the geographic domain, it is necessary to use and/or define the following elements: axioms, relations, concepts, properties, abilities and constraints. In (Quintero 2007, Torres 2007) are fully defined these elements.

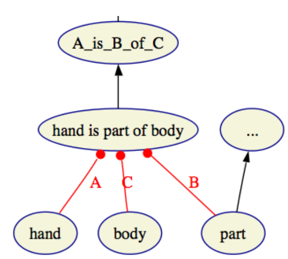
a



b



c



ISSN 1405-5546

d

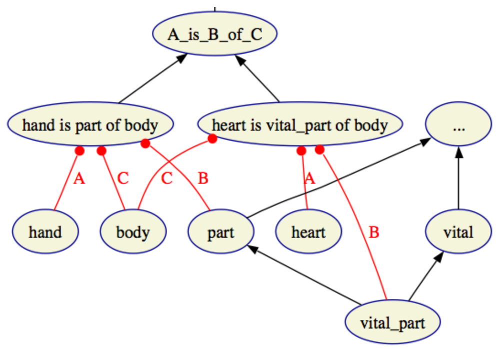


Fig. 1. Comparison between conceptualizations

2.2 Synthesis

In synthesis stage, we attempt to decompose RSDS in "extracts" according to the conceptualization. For obtaining "extracts", we use a strategy similar to the one used by people for clustering objects: applying different criteria to different clustering level. We will apply different clustering criteria to each set of objects. In each level we can use different algorithms or criteria. As result of the conceptualization of criteria and algorithms, we obtain the ontology of application.

The conceptualization of the application produces the ontology with all possible results of applying the extraction algorithms (the results of algorithms, not the algorithms themselves). As well as concepts in the domain ontology, as part of conceptualization of application we must set up the relationships of existence between extracts (unnamed concepts) in the application ontology and concepts in the domain ontology.

2.3 Description

Description stage consists of representing, according to the conceptualization, the "extracts" obtained on synthesis stage. "Extracts" are described in the application ontology and linked to the domain and geographic ontologies. However, it is not always possible to link an "extract" to best concept on conceptualization. Then, it is necessary to analyze properties of concepts to improve the linking of them to concepts in the domain and geographic ontologies. To determine the best specialization of an "extract", the clustering approach is used. In (Quintero 2007) the specialization algorithm is fully described.

3 Results

The conceptualization was made in three parts: conceptualization of geographic domain, of landforms domain and of application. As result we obtained three ontologies that we called *Kaab*, *Hunxet* and *Wiinkil*, respectively. The *Kaab* ontology has the classes defined to conceptualize the geographic domain, we have used the topographic 1:50000 vector data dictionary for Mexico (INEGI 1996), where are defined more than 70 topographic features which were applied in conceptualization. Detailed description of each class as well as all remaining concepts are presented in (Quintero 2007). The concepts of landforms are conceptualized in the *Hunxeet* ontology, this conceptualization is based on the dictionary of the Spanish Royal Academy of Language. The application ontology, that we call *Wiinkil* is a conceptualization in form of hierarchy of "extracts" obtained from the extraction algorithms.

The developed ontologies must be integrated each other to use and enrich the knowledge described by them. First, in Figure 2 is shown the integration of Kaab ontology with Hunxeet ontology by means of the assignation of the main class in *Hunxeet* (*forma del terreno* – landform) to corresponding classes in *Kaab*. In this way we characterize landforms existentially, according to ontology of geographic domain.

The proposed synthesis algorithm generates three types of extracts: "elev" for elevations, "depr" for depressions and "llan" for plains; by applying it recursively, different combinations of these types are generated. We call signature to a specific combination. The algorithm has four steps: 1) compute the longer plain zone (ZLE), 2) region labeling, 3) segmentation and 4) extraction. In (Quintero 2007) this algorithm is fully detailed.

For testing the methodology, we use a DEM, obtained from USGS web site of Grand Canyon on Colorado State, USA. Figure 3 shows the result of segmentation step. Here, we can see four little images: labeled with "zle" is shown the Largest Plain Zone. Labeled with "elev", "llan" and "depr", we obtained data sets classified under corresponding signature. According to the conceptualization, we must obtain 21 data sets (21 signatures) before starting with extraction step.

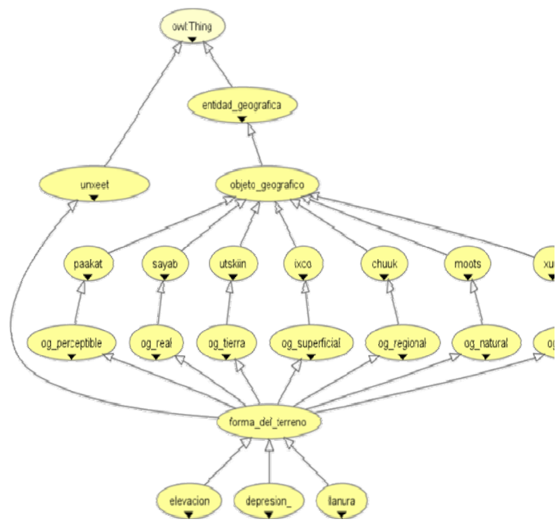


Fig. 2. Fragment of Kaab-Hunxeet integrated ontology

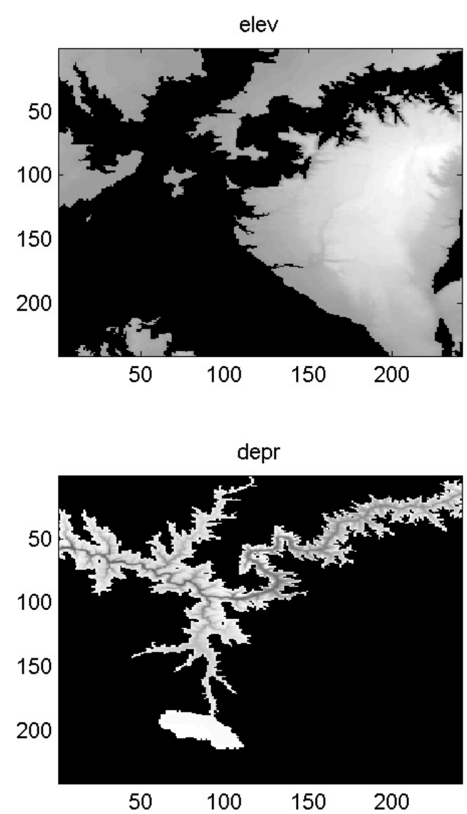


Fig. 3. Results of extraction algorithm

5 Conclusions

In this work, we have described methodology for making semantic descriptions of raster spatial data sets. The conceptualization methodology is the most important part of this research; because we propose to make the conceptualization using only three axiomatic relations, that allows to move the "classic" relations to the conceptualization giving to them semantic richness. As part of case study, we have developed three ontologies: Kaab ontology for the conceptualization of geographic domain, Hunxeet ontology for the conceptualization of landforms domain, and Wiinkil ontology for the conceptualization of our application. Synthesis stage is made in the image processing fashion, with phases of pre-processing, processing and post-processing. Description stage is made by using the conceptualization and by applying some templates for describing spatial knowledge.

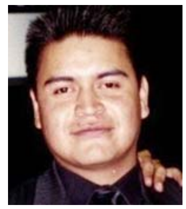
As future work, we consider that it is necessary to analyze and conceptualize geographic relations (topologic and geometric for instance) between concepts identified and described in this work. Also, we want to measure the quality of the description made. We propose the use of building blocks (basic landforms) for building a synthetic model and compare it to the original data set. On the other hand, we propose to make the description by using formal first order logic and comparing the resulting logics, in order to obtain a quality metric.

Acknowledgments

The authors wish to thank to SIP Project: 20110851 and CONACYT project 106692 for their support.

References

1. **Ackermann, F. (1993).** Automatic generation of digital elevation models. *Proceedings of the European Organization for Experimental Photogrammetric Research, Commission B, DTM Accuracy Meeting.* Southampton, United Kingdom. 231-240.
2. **Etzelmüller, B. & Sulebak, J. R. (2000).** Developments in the use of digital elevation models in periglacial geomorphology and glaciology. *Physische Geographie* 41, 35-58.
3. **Hodgson, M. E. (1995).** What cell size does the computed slope / aspect angle represent? *Photogrammetric Engineering and Remote Sensing*, 61(5), 513-517.
4. **Instituto Nacional de Estadística Geografía e Informática (1996).** *Diccionario de datos topográficos 1:50 000 (Vectorial).* Aguascalientes, México: INEGI
5. **Moreno-Ibarra, M. A. (2007).** *Similitud Semántica entre Sistemas de Objetos Geográficos Aplicada a la Generalización de Datos Geo-espaciales.* Tesis de doctorado, Instituto Politécnico Nacional, Centro de Investigación en Computación, México, D.F.
6. **Quintero, R. (2007).** *Representación Semántica de Datos Espaciales Raste.* Tesis de doctorado, Instituto Politécnico Nacional, Centro de Investigación en Computación, México, D.F.
7. **Smith, B. & Mark, D. (2001).** Geographical categories: An ontological investigation. *International Journal of Geographical Information Science*, 15(7), 591-612.
8. **Smith, B. & Mark, D. M. (2003).** Do Mountains Exist? Towards an Ontology of Landforms. *Environment & Planning B, Planning and Design,* 30(3), 411-427.
9. **Sulebak, J. R. & Tallaksen, L. M. (2000).** Estimation of areal soil moisture by use of terrain data. *Geografiska Annaler: Series A, Physical Geography,* 82(1), 89-105.
10. **Torres, M. (2007).** *Representación ontológica basada en descriptores semánticos aplicada a objetos geográficos.* Tesis de doctorado, Instituto Politécnico Nacional, Centro de Investigación en Computación, México, D.F.
11. **Weibel, R. & DeLotto, J. L. (1988).** Automated terrain classification for GIS modelling. *GIS/LIS'88: Proceedings: Accessing the world : third annual International Conference, Exhibits and Workshops,* San Antonio, Texas, USA, 618-627.



Rolando Quintero Telléz

Obtained his Ph. D. degree from the Center for Computer Research of the National Polytechnical Institute of Mexico. He is a Professor and Researcher of the CIC-IPN and the Laboratory of Intelligent Processing of Geospatial Information. Additionally, he has published more than 60 papers in National and International Journals and Conferences. His research efforts are mainly in the semantic processing of spatial data.

Serguei Levachkine

Is a Professor and Researcher of the CIC-IPN and 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