Combination of Spatial Analysis Methods and Semantics Extraction in a SDI Environment for Agro-ecological Zonation

Eduardo Garea Llano and Rainer Larin Fonseca
Advanced Technologies Application Centre, 7a # 21812, Siboney, Playa, Havana - 12200, Cuba.{egarea, rlarin}@cenatav.co.cu

ABSTRACT

In this paper we present a method for agro-ecological zonation of crops in Cuban mountain conditions in a Spatial Data Infrastructure (SDI) Environment. The proposed method is based on the combination of Geographical Information System (GIS) analysis techniques as map overlay, reclassification, and a proposed algorithm for indexation of the overlaid feature maps previously obtained by decision criteria given by experts in agriculture.

To provide geospatial semantics we extract semantic information from data stored in the geodatabase integrated in the developed GIS application by extracting a geospatial ontology that can serve as an advanced user querying system.

The proposed method was experimented in two scenarios, one for agro-ecological zonation of coffee plantations in Nipe-Sagua-Baracoa mountain region, and other for agro-ecological zonation of forest, fruit and potato plantations in Guaniguanico mountain region. As result a set of maps was obtained, these maps are an important tool integrated to the Cuban SDI that can be used for the development of new plantations on a sustainable base.

1. INTRODUCTION

The use of GIS as a tool for decision making in agricultural applications has had a wide use internationally, many times in form of applications for the agricultural resource inventory and others as analysis tools for the generation of new information.

The combination of raster and vectorial analysis methods has constituted an effective tool for the generation of new information starting from spatial and alphanumeric information integrated to a GIS (Van Westen CJ, 1997; Diez. A, 2002; Alafont L, 2002).

Recently, ontologies have gained increasing interest in the GIS community. They are essential to create and use data standards as well as human computer interfaces and to solve heterogeneity/interoperation problems. The use of ontologies as a middle layer between users and the databases adds a conceptual level over the data and allows the user to query the system on semantic concepts without having any specific information about the database.

In this work we present a method for agro-ecological zonation of crops based in the combination of GIS analysis techniques as map overlay, reclassification, and a method of indexation of the overlaid feature maps combined with an approach to extracting a geospatial ontology from geographical data stored in spatial databases. To provide geospatial semantics we use relations which define geospatial ontology that can serve as a basis for an advanced user querying system in a SDI environment.

This allows to build a mapping between ontological concepts and data and will increase the quality and effectiveness of the spatial analysis in the decision making process.

We present two cases use of the proposed method due to the need of evaluation of the productive potentialities in the mountainous regions of Cuba for different kinds of plantations in consonance with the new plans of sustainable development of these regions.

The remainder of the paper is structured as follows. We first introduce a motivation to clarify in what context of a SDI we refer to combination of spatial analysis methods (section 2). In section 3, methods of spatial analysis are discussed with respect to their role in the zonation process. In section 4, we present a method for binary indexation of the overlaid feature maps. In section 5, we explain our methodology of agro-ecological zonation, in section 6 we explain our idea of semantic extraction, in section 7 we illustrate the idea by conducting a walk-through the zonation process in the mountain regions of Cuba.

2. THE SPATIAL DATA INFRASTRUCTURE

Spatial Data Infrastructure (SDI) is a relatively new term. An SDI is a framework for sharing spatial information and integrating spatial information and knowledge. Through connecting each system together, SDI seeks to solve problems of geographic and spatial nature.

SDI began to be developed in the ninety decade of XX century characterized by its data-centric orientation (Rajabifard. A, 2006). More recently a second generation of SDI has emerged as a process-based model; however, the real demand today is to build effective user-driven SDIs and this requirement will become the key factor to SDI “usability” in the future. The user satisfaction priority is not exclusive of SDI, there is a real trend to reduce the gap between users and providers as a common characteristic in the Information and Communication Technology (ICT) development.

Sustainable development is a goal shared by all countries of the world and constitutes the only chance to survive on Earth. This fact is universally accepted, but often we are not really making decisions thereon. Besides the will of politicians who debate in world summits adopting new actions to tackle the main problems, ICT also should play its role. Inspired by Tim Berners-Lee, inventor of the Web and current director of the World Wide Web Consortium (W3C), a group of individuals from the past decade has established methods for representing not only words, but their meanings, in a format that machines could understand. This new vision was called “The Semantic Web” (Lowe J, 2008). In Web-based GIS context and SDI the Semantic Web could be interpreted as “The Geospatial Semantic Web”.

2.1. The Cuban SDI

The implementation of the Cuban SDI began in 2001. Its first stage was directed to institutional invigoration with creation of the SDI National Commission (Delgado T, 2007).

The projection of the Cuban SDI development was projected until 2010. Some of its main objectives are guided to share geographical information between different Cuban institutions in a cooperative form to support the economic, social, and environmental decision making.

The Cuban SDI is a geospatial service based on usage policies, access and distribution of geographical information according to the national needs.

The Cuban Geographic Portal (GeoPortal) facilitates the access to the geospatial data offered by national providers. It is hosted in a data center provided with powerful servers and its main services are: The information retrieval of the existent spatial information by means of metadata catalogs; access to the geospatial data available in maps, images, and geographical object servers; generation of custom applications, according to the specific user context that allow to make online analysis associating specific thematic data to the geographical objects served in the SDI; available common services in the GeoPortal as a result of actions that facilitate the interrelation of the served spatial data and information provided by the Government to the society which includes territorial information services and geospatial services to the citizen; corporative services that work from the institutional intranets using the SDI geospatial data that allow an improvement in the management decision making of these institutions.

2.2. Tocororo.

It is an interface access node to the global database scheme of the Cuban SDI environment (Capote J.L, 2007).The user interaction with this global scheme is carried out by means of queries to the metadata located in Tocororo's databases. The inputs to the databases provided by Tocororo are grouped in categories created by the system manager. Users can explore every category or make searches in Tocororo. In a first version, Tocororo presented a keyword based scheme of searching. A new version of Tocororo is being created based on the principles of the Semantic Web, it will include an ontological modelling of information stored in a global server scheme and information recovered from database global scheme using Natural Language with criteria based on traditional and geospatial data type using geospatial Semantics Web.

Our work was developed in this context using the tools integrated to the Tocororo interface and the analysis techniques that will be discussed in the next section.

3. SPATIAL ANALYSIS METHODS COMBINED IN THE AGRO-ECOLOGICAL ZONATION

Spatial relations among geographical objects or entities are often as important as the entities themselves (Klien E, 2005). In a new SDI generation geospatial domain ontologies, taxonomic and non-taxonomic relations can be used to define concepts of the real world and to differentiate among them. At this level, spatial relations have an important role for defining and identifying spatial concepts and generating new information. The spatial relations may be expressed through spatial analytical methods.

Berry (Berry. J, 1987) offered a classification of analytical operations in the GIS environment that is very useful for the goal of this work. He differentiated four kinds of actions:

- Reclassification of map thematic values.
- Map overlay.
- Distance and connectivity or computation of shortest distance between poits.
- Map filtering

For the above mentioned methods we will combine reclassification, overlay, and map filtering operations, the rest of methods are mainly operations related with proximity and distance analysis.

3.1. Map Reclassification

Bosque (Bosque .J, 1992) classifies and describes analysis local operation with only one map or map reclassification by local analysis:

- Re-labeling of Initial categories (Figure 1a).

- Addition of the initial categories in a smaller number of values.
- Creation of intervals in a continuous variable.
- Operations with a constant to obtain new categories (Figure 1b) in function of: 1) the previous categories, 2) a constant, 3) a mathematical or logical function (e.g. add a quantity to all pixel values of a raster map).

Trigonometric operations: A new map is generated in which the pixel value is obtained applying a trigonometric operation (sine, cosine, tangent, etc.) to the thematic value of the same base element on the source map.

To these operations can be added; the multivariate pixel to pixel classification that is applied for supervised and no supervised image classification. In practice they are related with thematic map production by conversion of continuous tones of an image by means of a process that categorizes the pixels in land cover types or themes.

These processes are performed by statistical pattern recognition techniques derived from Statistical decision theory, Bayes decision rule, K-nearest neighbour rule, McQueen, ISODATA, K means, etc (Jain A.K, 2000) or use of Artificial Neural Network techniques with Back propagation, LVQ (Learning Vector Quantization) models for the supervised variant and the Kohonen's self-organizing map and the ART2 models for the unsupervised variant, among others.

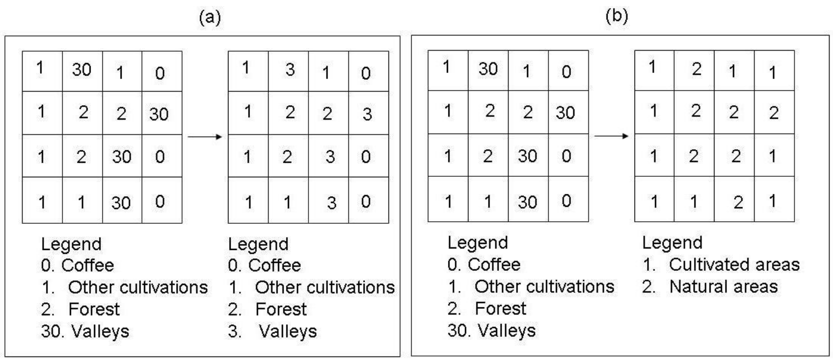

Figure 1: Map reclassification, a) Re-labeling of Initial categories, b) Operations with a constant.

4. METHOD FOR INDEXATION OF THE OVERLAID FEATURE MAPS (MIFM)

The agroecological zonation of crops is a powerful tool for the agriculture sustainable development in the mountain regions due to the high vulnerability of these ecosystems. For that reason, the knowledge of the factors that were considered for the zonation in the produced map is very useful for the final users of the GIS application.

The problem to solve consists in assigning an index to each value in the feature maps that participates in the computation of agroecological areas, the summa of these indexes, in any possible combination, should be result in a new index, from which if will possible to determined which features form each of the categories of the areas. The general scheme of the proposed method is presented in Figure 2.

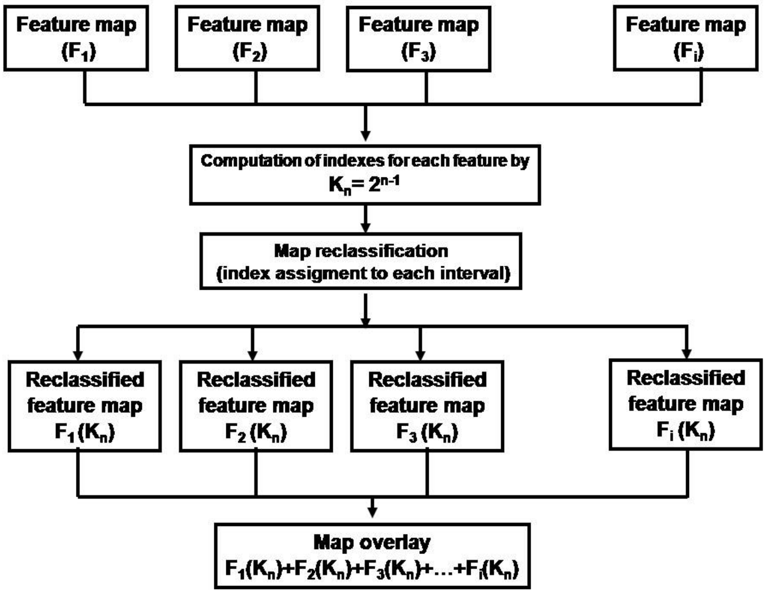

Figure 2: General Scheme of the proposed method.

For the zonation we used various feature maps to describe the categories of each area; the features values are associated to each pixel of a raster map. For a feature, we combined their values. We built a sequence of intervals for a feature taking into account expert's criterion about the characteristics of each spatial distribution of features. We take the first feature value and look for its sequential interval index *n*. We use the following expression to obtain the index for this feature value to achieve that the sum of the coded features by expression 1 will be unique.

$$K_n = 2^{n-1} \qquad (1)$$

where K_n is the index and n is the number of value interval.

The feature indexation is carried out by means of summing of the indexes that compose each interval.

5. METHODOLOGY FOR AGRO-ECOLOGICAL ZONATION

According to the criteria offered by FAO (FAO, 1997), the agro-ecological zonation defines regions based on soil combinations, physiographic and climatic characteristics. The particular parameters used in this definition are centred in the climatic and edafic requirements of crops and in the management systems used for their development.

Each region has a similar combination of limitations and potentialities for land use, and serves as reference point of the recommendations designed to improve the existent situation of land use, increasing the production or limiting the degradation of soil resources.

Our methodology is based on these criteria and we defined an agro-ecological area as a cartographic unit of land resources defined in terms of climate, physiographic and soils and/or land cover. This area has a specific range of limitations and potentialities.

An agro-ecological cell (AEC) is defined as a unique combination of physiographic, soil and climatic characteristics. The AEC is the referenced basic unit used by the FAO methodology for the physic analysis in agro-ecological zonation. In our methodology, we assume these criteria for the cartographic definition of agro-ecological area for some kinds of crops taking into account the special geomorphologic, natural, and social characteristics of mountain regions of Cuba.

The proposed methodology is based also on the specialized criteria of agricultural experts and technical exigencies of the analyzed crops.

We designed this methodology for four kinds of crops: coffee, fruit, forest and potato. Detailed descriptions of these criteria can be seen in Soto´s works (Soto F, 2004). An example of criteria and classification of the agro-ecological areas for coffee crops are presented in Table 1.

	Criteria				
Soil groups	Soil depth (cm)	Height (m)	rainfall (mm)	temp (C^0)	Classification
Alisol, Ferralsol fluvisol Cambisol	>30	>400	>1600	< 16	Optimal
		200-400	1400-1600	16-24	Fairly optimal
			1200-1400	>24	Acceptable

Table 1: Expert's criteria for the coffee zonation.

The proposed general scheme of agro-ecological zonation (Figure 3) combines information of different thematic or feature maps contained in the Cuban SDI and the combination of the previously described methods of analysis.

6. SEMANTICS EXTRACTION

The goal of this first step to the semantics is the use of ontologies and semantic annotation to adapt our proposal to the new version of the Cuban SDI that will be based on the principles of the Semantic Web.

In this section we present the idea of automatic knowledge generation using an architecture proposed by Baglioni (Baglioni, M. 2007) for automatic generation of ontology of agro-ecological zonation.

The architecture begins building an Ontology of Application starting from the database integrated to the GIS, by means of the extraction module.

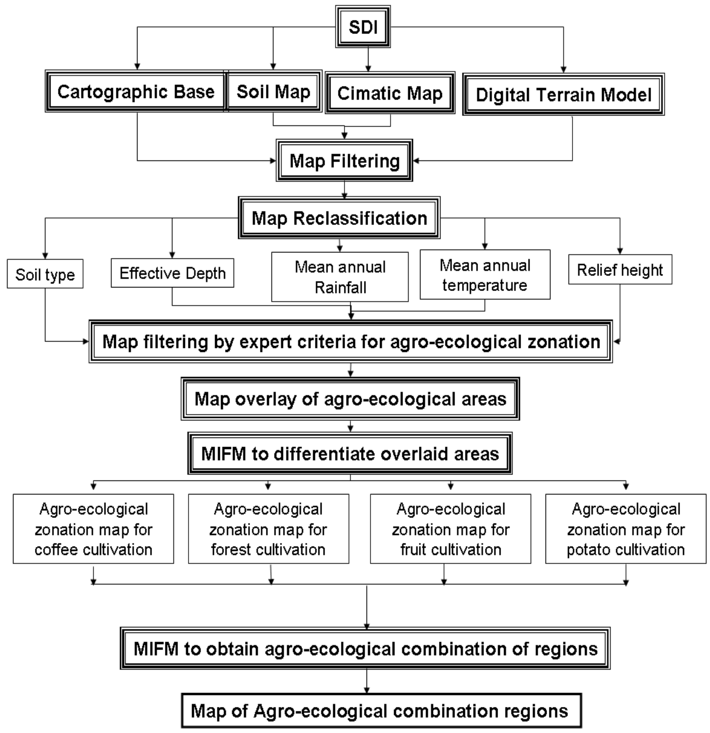

Figure 3: General scheme for agro-ecological zonation.

Since the application ontology is derived from the geodatabase, and strictly depends on the structure of tables and their relations, it is not modeled a-priori.

The objective of the Extraction Rules module is to automatically build the Application Ontology starting from the database schema. There are approaches in the literature that define rules to automatically extract ontologies from relational database such as discussed in (Li M., 2005). Typically, these rules produce concepts and relations from tables depending on the schema of the database, such as the structure of the tables and the features of the primary and foreign keys. When dealing with geodatabase, new rules have been defined (Baglioni, M. 2007) in order to manage direct and indirect location and connect them with the geospatial ontology. This ontology is composed by a group of concepts and the existent relations among these concepts, which are explicitly represented in the structure of the database, and then this ontology is enriched with the Domain Ontology to provide the semantic domain in which we will work. This Enriched Ontology represents the semantic view and the taxonomy of the data, which are mapping to facilitate spatial semantic queries.

Since the ontology should be able to represent the abstraction of the data in the database, it is necessary to explicitly define what types of relations can be present in them to be able to faithfully express their spatial properties. The ontology is defined on the base of two types of concepts ("terminal" and "non terminal") and two types of relations ("has" and "is a") (Torres M, 2005).

Concepts:

- (TC) Terminal Concept: Do not use other concepts to define their meaning (that are defined for "simple value").
- (NC) Non Terminal Concepts: Concepts that use other concepts (terminal or non terminal) in their definitions.

Relations:

- is_a: It defines that an object A can be defined, in a level of bigger abstraction, as an object B, for example:
  A cambisol *is_a* soil type.
- has: It defines that an object A has an object B, for example:
  A soil type *has* a depth.

The domain ontology to be generated should cover the whole area of the semantic knowledge on which the user can carry out queries. This ontology will be applied to the context of agro-ecological zonation, for which the concepts, properties, relations, functions, restrictions and axioms are defined in a "explicit" form using an implementation language that is able to store this knowledge.

7. APPLICATION OF THE PROPOSED METHODOLOGY

7.1 Agro-ecological zonation

The application of the proposed methodology was carried out in several mountainous regions of Cuba; in this work we present the results obtained in Nipe-Sagua Baracoa region situated in the eastern part of the country and in Guaniguanico, situated in the western region (Figure. 4).

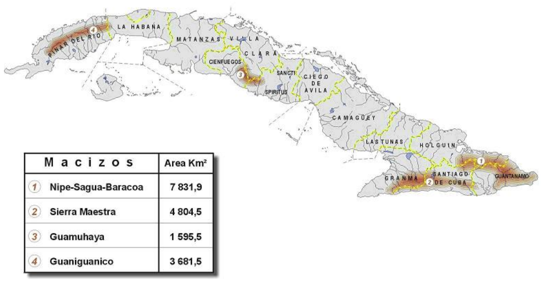

Figure 4: Distribution of mountain regions of Cuba.

The primary spatial and alphanumeric information was taken from Cuban SDI through the "Tocororo" interface access node.

We experimented the agro-ecological zonation for coffee crops in the mountain region of Nipe-Sagua-Baracoa. The agro-ecological zonations for fruits, forest, and potato crops were experimented in the mountain region of Guaniguanico.

Figure 5 shows the distribution of agro-ecological zonation in the mountain region Nipe-Sagua-Baracoa for coffee crops obtained by the proposed methodology.

Figure 6 shows the distribution of agro-ecological zonation for fruit and forest crops in mountain region of Guaniguanico.

Figure 8 shows the result of application of the proposed methodology for determination of the combination of agro-ecological areas.

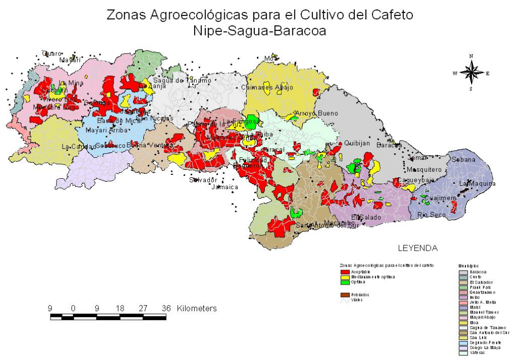

Figure 5: Result of Agro-ecological zonation for Coffee crops.

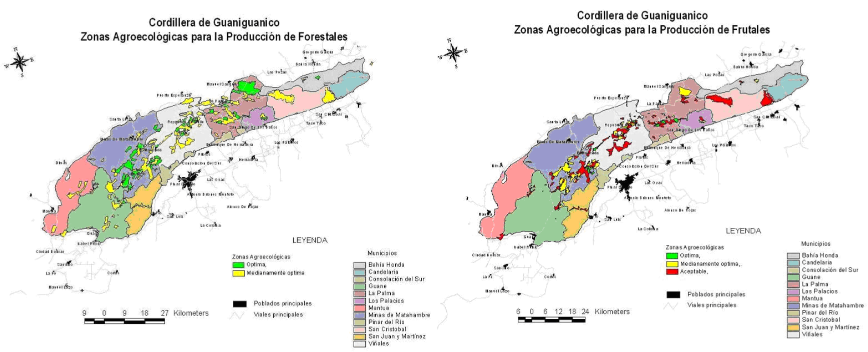

Figure 6: Result of agro-ecological zonation for fruit and forest crops.

7.2 Automatic knowledge generation

We automatically generated the application ontology for agro-ecological zonation, **F**igure 8 show the scheme of the obtained ontology.

We carried out the implementation of several tools that allow tasks related to the semantic annotation of spatial data and management of these annotations.

It was also necessary to implement a module that would allow browsing, searching and selecting the desired concept for the annotation as well as creating new concepts. To implement this module use the library JENA[1].

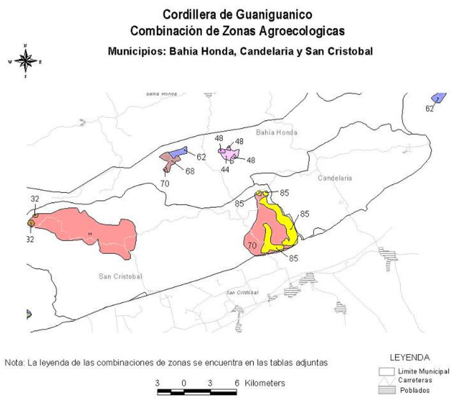

Figure 7: Combination of agro-ecological areas.

To facilitate the process of semantic annotation by the users a visual application was developed. It was implemented in form of plug-in. Using this plug-in user can:

- To carry out semantic annotations on the data shown through the visual interface.
- To carry out searches on the annotations. The user will be able to select a rectangular region of the plane and he will be informed about the data previously annotated in that region.
- To save the annotations in ontology in owl language.
- To open annotations from ontology in owl. language

Then we used the developed tools to assign the ontological concepts to each geographical object in the obtained maps.

We experimented the searching for the optimal areas for coffee crops using automatically annotated data by generated application ontology and using the implemented tools and comparing the results with those obtained using typical searching GIS tools based on a standard (spatial) SQL query language. In Table 2 we show the results of comparison between searches in the implemented application.

These results demonstrate the effectiveness of the proposed approach since the use of the enriched ontology as a middle layer between the query system and the geodatabase, we can abstract from the specific object and refer to the Concept of object. In a query execution phase, the object concept is expanded in the set of its subclasses; therefore, the original query is transformed into a set of spatial SQL-queries, one for each subclass that corresponds to a database table.

1 http://jena.sourceforge.net

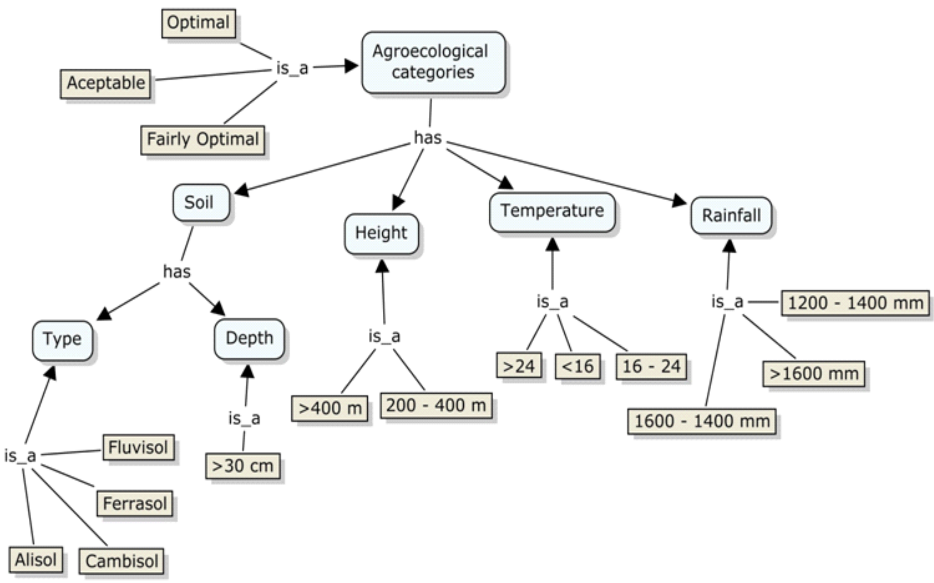

Figure 8: Automatic Generated Application Ontology.

Methodological steps	Time consumed using non annotated data (mlsec)	Time consumed using annotated data (mlsec)
Retrieval of soil groups for each classification category	500	200
Retrieval of height intervals for each classification category	560	224
Retrieval of rainfall intervals for each classification category	530	212
Retrieval of temperature intervals for each classification category	540	216
Retrieval of optimal zones combining all the criteria	700	280

Table 2: Comparison of time consumed by searching.

CONCLUSIONS

In this work a new methodology for agro-ecological zonation in mountain conditions in a SDI environment is presented. The proposed methodology is based on the combination of spatial analysis methods and a new method for indexation of the overlaid feature maps (MIFM). As a result of the application of the proposed methodology in a real scenario, a set of agro-ecological zonation maps for coffee, fruit, forest, and potato crops were obtained for some mountain regions of Cuba.

To provide geospatial semantics we used relations which define geospatial ontology that serves as a basis for an advanced user querying system in a SDI environment. This allowed to build a mapping between ontological concepts and data and increase the quality and effectiveness of the spatial analysis in the decision making process.

BIBLIOGRAPHY

Alafont L., Elaboration of a risk map of environmental combination for the underground waters by means of a GIS, *The Geographical Information Systems in the administration of the geologic risks and the environment*, Madrid 2002 pp 257-268.

Baglioni M, Masserotti M, Renso Ch ,and Spinsanti L., Building Geospatial Ontologies from Geographical Databases.LNCS 4853, Proceedings of the Second International Conference on Geospatial Semantics, 2007, Mexico City.

Bosque J., Geographical Information Systems (in spanish). Rialp.sa, Madrid, 1992 pp 124-138.

Capote J.L, Cruz R, Herrera O. and Delgado T., Automatic development of ontology-based knowledge base using consensus among user profiles. Proceedings of the Second International Conference on Geospatial Semantics, 2007, Mexico City.

Berry. J., A mathematical structure for analysing maps, Environmental management, vol 11, núm 3, 1987, pp 23-29.

Delgado T. and Crompvoets J., Espatial Data Infraestructure in Iberoamerica and Caribbean (in Spanish), IDICT, Havana, Cuba, 2007, 211pp.

Diez. A.: Book "Application of the GIS to the Risk Analysis of fluvial floods". Madrid 2002,.pp 236-251.

FAO, 1997. Agro-ecological zonation, General Guide. Boletín 73. Roma, Italy. 51

Jain A.K., Duin R.P and Mao J., Statistical Pattern Recognition: A Review. IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol 22, No 1, January 2000

Klien E. and Lutz M.. The Role of Spatial Relations in Automating the Semantic Annotation of Geodata A.G. Cohn and D.M. Mark (Eds.): COSIT 2005, LNCS 3693, pp. 133 – 148, 2005. Springer-Verlag Berlin Heidelberg, 2005.

Li M., Du X., Wang S. Learning Ontology from Relational Database. In: Proceedings of the Fourth International Conference on Machine Learning and Cybernetics, Guangzhou (August 2005)

Lowe J.: A Geospatial Semantic Web, available on Internet.

Rajabifard. A. The role of sub-national government and the private sector in future spatial data infrastructures. International Journal of Geographical. Information Science. Vol. 20, No. 7, August 2006, 727–741

Soto F., Agro-ecological zonation for coffee crops in the mountain regions of Nipe-Sagua-Baracoa, Sierra Maestra and Guamuhaya(in Spanish). Researh Report, National Institute of Agricultural Sciences, Havana, Cuba, 2004, 62pp.

Torres M., Quintero Rolando, Moreno M, and Fonseca F. Ontology-Driven Description of Spatial Data for Their Semantic Processing. LNCS 3799, Proceedings of the Second International Conference on Geospatial Semantics, 2005, Mexico City.

Van Westen CJ: Tools for map analysis applied to the selection of a waste disposal site. ILWIS for Windows, Applications Guide, ILWIS Department, ITC, Holland, 1997, pp 28-32.