Revista Digital Universitaria
1 de mayo 2011 • Volumen 12 Número 5 • ISSN: 1067-6079

ARTICULO

SPATIAL ANALYSIS BY MEANS OF A DISTRIBUTED GEOGRAPHICAL INFORMATION SYSTEM

Miguel Torres, Marco Moreno-Ibarra, Rolando Menchaca-Méndez, Rolando Quintero y Giovanni Guzmán
Centro de Estudios de Cómputo
Instituto Politécnico Nacional

1 de mayo 2011 • Volumen 12 Número 5 • ISSN: 1067-6079

Spatial Analysis by means of a distributed geographical information system

Abstract

To improve the spatial retrieving by means of Geographical Information Systems (GIS), a distributed GIS-Application has been designed. This application is mainly focused on making spatial analysis, throughout Spatial Analyzer Module (SAM). SAM contains several spatial and mathematical models that have been designed and implemented for detecting landslides and flooding areas. Users can make an analysis to retrieve the spatial data from different sites using a specification based on Extensible Markup Language (XML). Geospatial data are stored in a geographic database into the Enterprise GIS. Thus, all spatial analysis processes are executed in the server application to improve the performance of the análisis and information retrieving. The GIS-Application has been implemented into ArcMap module of ArcGIS system.

Keywords*:* GIS, XML, landslides, flooding areas, distributed environment, spatial database.

Análisis espacial por medio de un sistema de información geográfica distribuido

Resumen

Con la idea de mejorar la recuperación espacial, por medio de un Sistema de Información Geográfica (SIG), una aplicación SIG distribuida ha sido diseñada. Esta aplicación se enfoca principalmente en realizar operaciones de análisis espacial, a través de una herramienta denominada denominada Módulo de Análisis Espacial (SAM). Este módulo contiene varios modelos espaciales y matemáticos que fueron diseñados e implementados para detectar áreas con riesgo de deslave e inundación. Los usuarios pueden llevar a cabo un análisis para recuperar los datos espaciales de diferentes sitios, utilizando una especificación basada en el lenguaje XML. Los datos geoespaciales están almacenados en una base de datos geográfica dentro de un SIG empresarial. Asimismo, todos los procesos de análisis espacial se ejecutan en el servidor de aplicación para mejorar el rendimiento del análisis y recuperación de información. La aplicación SIG ha sido implementada en el módulo ArcMap del sistema ArcGIS.

Palabras clave*:* SIG, XML, deslaves, inundaciones, ambiente distribuido, base de datos espacial

Introduction

Nowadays, Geographical Information Systems (GIS) are powerful and useful tools as means of information, visualization and research or as decision making applications [1]. Recently, intelligent spatial analysis is the main need presented in the Geocomputation trends. Spatial data have an important role in this situation; many times, the information is extended at different places. The problem is greater, because the spatial data present different formats and specifications such as scale, projection, spatial reference, representation type, thematic, DBMS type, and date. For these reasons, the heterogeneity of the spatial data complicates the spatial analysis.

We propose a distributed schema based on XML to develop a standard spatial specification, which is used to recover spatial data. This schema allows exchanging, transferring, and storing geographical information into a geographical depository.

Proposals and projects have been designed to retrieve the spatial data. In the majority of the cases, the recovered data represent the information in a raster format (jpeg, gif) [2]. This information is difficult to use to make spatial analysis, because its intrinsic characteristics are not explored. Important projects have been designed to represent geographical information in the web (Web-Mapping), but these applications are closed to handle spatial data [3].

In this paper, we propose a mechanism to make spatial analysis by means of SAM. SAM can recover geographical objects using a distributed environment. SAM is a tool designed to simulate natural phenomena with special methods focused on detecting landslides and flooding areas. XML is used to develop an encoding specification based on qualitative and quantitative properties of the spatial data. The main goal of this application is to find the solutions for making spatial analysis and recovering spatial data in a distributed environment. This is a great challenge in the new trends of Geocomputation field.

The rest of the paper is organized as follows. In Section 2 we present the architecture of the GIS-application and describe the functionality of the system modules. Some obtained results are shown in Section 3. Section 4 exposes our conclusion related to the work.

GIS-Application architecture

The developed GIS-application presents client-server architecture. This tool contains the following modules: Enterprise GIS, Communication Module, Spatial Analyzer Module, Spatial Database and XML Administration Module. Figure 1 shows the architecture of the tool.

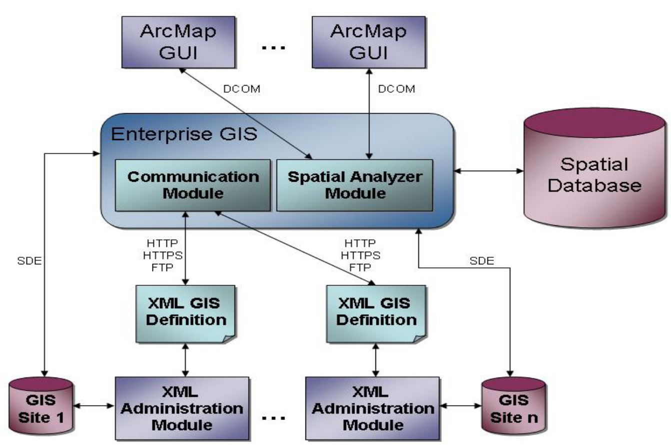

Figure 1. GIS-Application Architecture

The functional mechanism of the GIS-application is the following: ArcMap users need to make a request. This request is sent by DCOM technology to the communication module into the Enterprise GIS. This module processes the request and sends the parameters via HTTP or HTTPS

1 de mayo 2011 • Volumen 12 Número 5 • ISSN: 1067-6079

protocols.

In the remote GIS site, the XML administration module verifies the initial XML definition and queries the local XML definition to locate and compare the qualitative and quantitative characteristics of the spatial data such as scale, projection, spatial reference, representation type, thematic, DBMS type and attribute data. If the XML definition matches with the local XML definition, the spatial data will be recovered and sent by means of the Spatial Database Engine (SDE) mechanism. The geographical objects are stored in the spatial database, in which they will be analyzed by SAM.

SAM is focused on detecting landslides and flooding areas. In the following subsections, we will describe the most important issues on the GIS-application such as spatial database, communication module and spatial analyzer module.

Spatial database into GIS-Application

Our GIS-application contains a geographical database, which has been designed and implemented by Geodatabase. It is a storage mechanism provided by ArcInfo, which is focused on generating independent geographical depositories [4].

The spatial database is a special environment that models the characteristics of the geographical phenomena. Geodatabase provides a topological model that is integrated by class of elements. This model is similar to the spatial coverage. The Geodatabase model is supported by an object-oriented relational database. On the other hand, it is considered as a hybrid between object-oriented and relational techniques. Using this technology, users can access to the spatial and descriptive attributes from different sources by means of the SDE [5].

This hybrid mechanism defines an open interface to database system and allows handling geographical information in an intrinsic way. In this case, the behavior of the spatial objects is defined by the system. The entities are represented as spatial objects with properties, behaviors and relationships between them. The spatial database has been designed using the ArcInfo system. All these components are involved in the analysis to determine landslides and flooding areas. The spatial data that match with the XML specification are recovered and stored in the spatial database.

The model of this depository has been developed using Unified Modeling Language (UML), which generates COM objects. These objects implement the behavior and the schema, in which the objects are created, stored, and handled into the spatial database. This schema establishes the relationship between objects and the spatial database. On the other hand, it defines the spatial relationships between objects in different classes depending on primitives of the representation (lines, points and polygons). Figure 2 shows the elements that compose the spatial database.

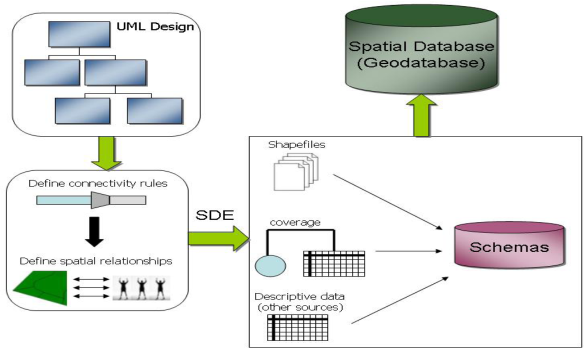

Figure 2. Features that composes the spatial database.

Communication module into GIS-Application

It is usual that the spatial analysis involves a lot of data sources. These data sources are located in different sites that belong to different organizations. For instance, in the case of flooding areas, the analysis involves data of superficial water volume, geology, infrastructure, etc. All these resources are not located in the database of a single organization.

To solve this problem, many efforts have been made to interoperate multi-vendor and distributed GIS. One of the alternatives is GML [6],[7] - a specification developed by the OPC for the transfer and the storage of geographical information based on XML -. However, this OPC recommendation is not yet supported by the majority of commercial GIS. Therefore, it is difficult to find a solution for the problem in short term, because it is necessary that GIS vendors agree in the construction of translators from its data proprietary representations to a standard.The problem to interoperate a multi-vendor and distributed GIS can be divided into three parts: (1) Localization of the required data to perform a spatial analysis; (2) Remote access to the relevant data, and (3) Transformation of different data representations. In this work, we propose a partial but useful solution that treats the first two problems. We have been able to get rid of the first component of the problem, because in many cases the GIS involved in analysis are distributed. To locate the required data and to perform a spatial analysis, we include a module called XML Administration Module. This module is responsible to obtain the geo-information to know if a particular data source is relevant in a particular spatial analysis [8]. The architecture is shown in Figure 1. The obtained geo-information is then codified in a XML document. Table1 shows an example of the contents of the XML document. Part of the information is queried from the GIS and other part is manually loaded by the GIS administrator.

The communication module, located into the enterprise GIS, queries the XML documents of the different sites to display its information to the user [9]. In this way, the users can select the information sources, which are relevant to their analysis. The communication between the enterprise GIS and the XML Administration Modules can be made using different standard protocols such as HTTP, FTP or HTTPS, if a secure transaction is needed.When the data sources are selected, the Communication Module uses the SDE to obtain the spatial data.

1 de mayo 2011 • Volumen 12 Número 5 • ISSN: 1067-6079

```
<?xml version="1.0"?>
<Spatial_Data>
<Description>
<Layer_Name>topo</Layer_Name>
<Theme>Topography</Theme>
<Elab_By>GeoLab-CIC-PN</Elab_By>
<Elab_Date>02/08/1999</Elab_Date>
<Last_Update>12/08/2001</Last_Update>
<Type>Line</Type>
<Topology>Y</Topology>
</Description>
<Geographical_Properties>
<Projection>UTM 14</Projection>
<Datum>NAR_D</Datum>
<Units>METERS</Units>
<Spheroid>GRS1980</Spheroid>
<Boundary>
<Xmin>397041.431</Xmin>
<Xmax>685954.665</Xmax>
<Ymin>2432826.985</Ymin>
<Ymax>3097482.722</Ymax>
</Boundary>
<Scale>1:50000</Scale>
</Geographical_Properties>
<DBMS_Properties>
<Provider>ArcInfo</Provider>
<Table_Name>topo.aat</Table_Name>
<Data_Type>Coverage</Data_Type>
<Attributes>Topo#, TopoD, FNODE#,
                TNODE#, Length, LPOLY#,
                RPOLY#, Height
</Attributes>
<Num_Records>420</Num_Records>
</DBMS_Properties>
</Spatial_Data>
```

Table1. A XML document containing a GIS description

Spatial Analyzer Module into GIS-Application

SAM is a special module, which has been designed to make spatial analysis procedures. The main goal of SAM is to identify zones with high flooding and landslide probabilities. SAM uses vector data to make the analysis. A specific analysis procedure has been designed for each one of the phenomena. The analysis is based on using different spatial data related to the studied phenomenon. In Figure 3 the architecture of SAM is depicted.

SAM contains three components: Analysis Block, List of Procedures and List of Resources. 1) Analysis Block is composed of a set of processes to make data analysis. 2) List of Procedures stores the sequence of steps to find risk areas as well as the description of required data. 3) List of Resources contains the spatial and attribute data description that can be queried using the network.

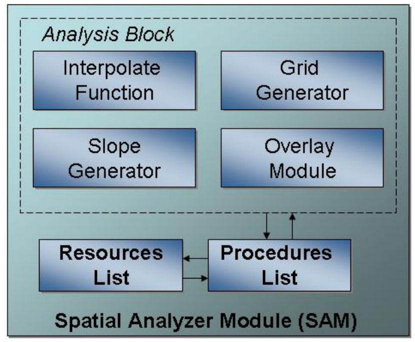

Figure 3. Architecture of Spatial Analyzer Module

Analysis Block. *It contains the functions to detect the risk areas. These functions are the following:*

- **Interpolate Function.** It is used for several tasks. The used method is a minimum curvature spline in two dimensions from a set of points. For computational purposes, the entire space of the output grid is divided into blocks or regions of equal size. They are represented in a rectangular shape. The equation 1 shows the spline function that has been used [10]:

$$S(x,y) = T(x,y) + \sum_{j=1}^{N} \lambda_j R(r_j), \quad (1)$$

where $j = 1, 2, ..., N$
N is the number of points.
λ_j are the coefficients obtained from the system of equations.
R_j is the distance from the point (x, y) to the j^{th} point.

To use this function, it is necessary to provide the set of points and tolerances, which depend on the specific case study.

- **Slope Generator.** The Earth surface presents many irregularities and different topographic characteristics. It is possible to evaluate the behavior of the fluid in any area. In this process the slopes are used to determine the speed of the fluids and to identify accumulation zones. The process is described in details in [11]. Figure 4 shows the process of generation of the slope layer.

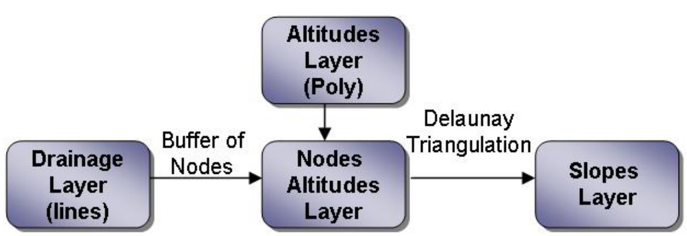

Figure 4. Generation process of the slope layer

- Grid Generator. It is used to process some analyzed data, especially in density map

1 de mayo 2011 • Volumen 12 Número 5 • ISSN: 1067-6079

-
- generation. The vector grids are regular of an m x m magnitude, in which m is the cell size. The cell magnitude in the grid is determined by the studied phenomenon, scale and covered area. Two alternatives can be used to generate the grids. First, specifying the beginning and ending grid coordinates $((x_0, y_0),(x_1, y_1))$ and establishing the number of required divisions for the grid. The second alternative is to specify the beginning coordinate (x_0, y_0), cell size, number of columns and rows in the grid [12]. Figure 5 shows the specifications of the grid.

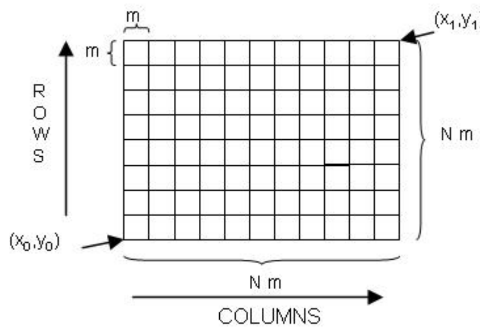

Figure 5. Generation process of the grid

- Overlay Functions. This module has been designed to make topological overlays, which can be used to identify the risk areas. A set of operations has been defined, and it is applied to the spatial analysis. This is made to establish the conditions and to combine different information layers using logical operators. These functions combine spatial and attribute data. The implemented operations for topological overlay in this application are: intersection, union and identity, which are represented by $\cap$, $\cup$ and I respectively.

List of Procedures. It stores the set of procedures for each one of the analysis processes. It has a description of the required data type and the restrictions. However, the users can change the selection criteria. This mechanism provides a list of functions as an alternative for the analysis, in which the parameters can be modified. The detection tasks are the following:

- Detection Mechanism of Flooding Areas. To identify the flooding risk areas, it is necessary to build an infiltration spatial layer [13]. With this layer, we can define the permeability in the study area. Therefore, the infiltration layer is designed using the geology. Later, these spatial data are intersected with the drainage and fracture density layers. Finally, the infiltration is obtained as described in [12]. The process is shown in Figure 6.

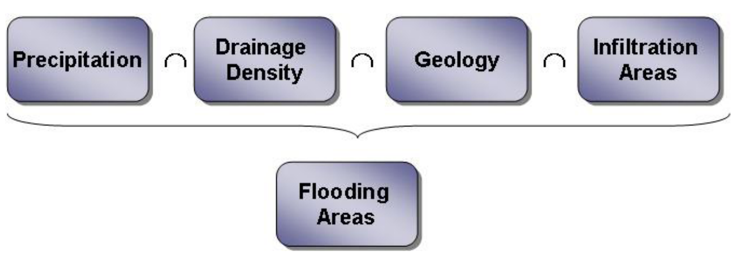

Figure 6. Detection mechanism of flooding areas

- Detection Mechanism of Landslide Areas. To identify the areas with the landslide risk, it is necessary to build a slope layer [12]. The geology, vegetation, and land use layers play an important role in this analysis. Thus, the soil composition determines the affectation when the climatologic conditions are changed. Additionally, the layers of basins, precipitation, and drainage density can be used to find possible landslides originated by precipitations. Figure 7 shows the process to detect landslides areas.

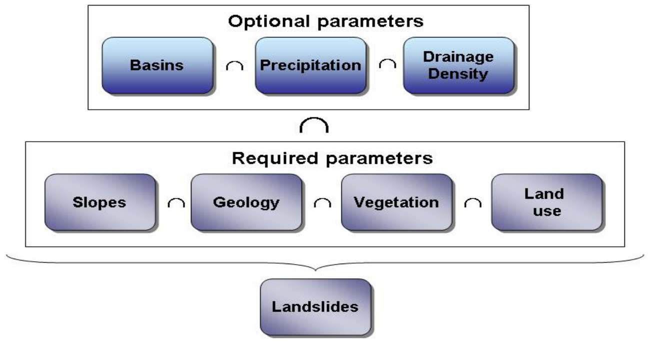

Figure 7. Detection mechanism of landslide areas

List of Resources. This mechanism stores the description of the remote and distributed GIS sites. This description is used to recover the spatial data from the spatial database. Thus, the users are able to select the cartographic digital products list. This list provides a general view of the qualitative and quantitative characteristics of the spatial data.

Results

Using the GIS-application, we can detect potential risk areas, which could be affected by flooding or landslides. The designed methodology has been applied to the Tamaulipas State, Mexico. Some results are shown in this section. Figure 8 shows the following spatial layers recovered from distributed GIS sites: drainage density, infrastructure, and population layers.

Figure 9 shows the areas with flooding and landslide risk. These areas have been recovered from distributed GIS sites. The spatial data are stored in the local spatial database. Therefore, the spatial analysis has been made into SAM providing zones with particular characteristics. The risk areas are concentrated at the center and south of the state. Also, Figure 9 presents classified areas according to level of risk: low risk (blue), middle risk (yellow) and high risk (green). LandSat TM image has been used to complete and improve the spatial analysis.

1 de mayo 2011 • Volumen 12 Número 5 • ISSN: 1067-6079

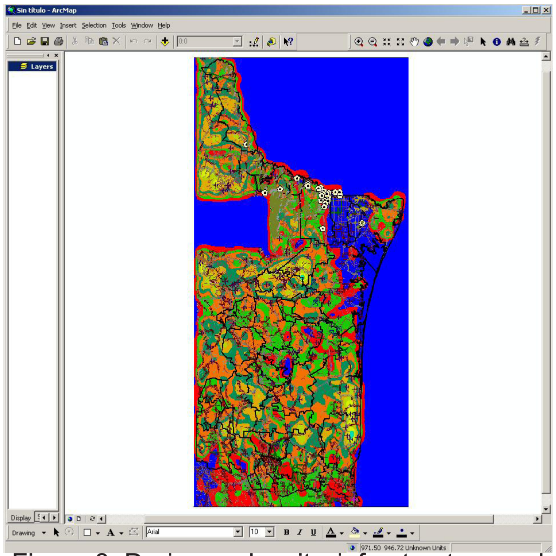

Figure 8. Drainage density, infrastructure and populations layers retrieved from distributed GIS sites

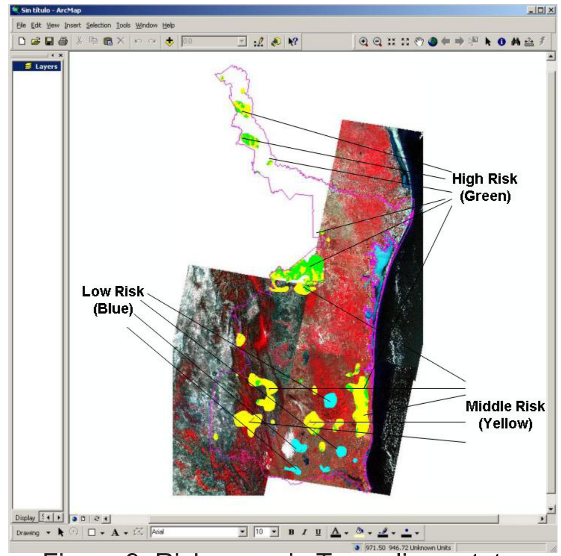

Figure 9. Risk areas in Tamaulipas state, Mexico

Figure 10 presents a zoom operation of the spatial data recovered from different GIS sites. We can see the infrastructure and drainage density around Tampico City.

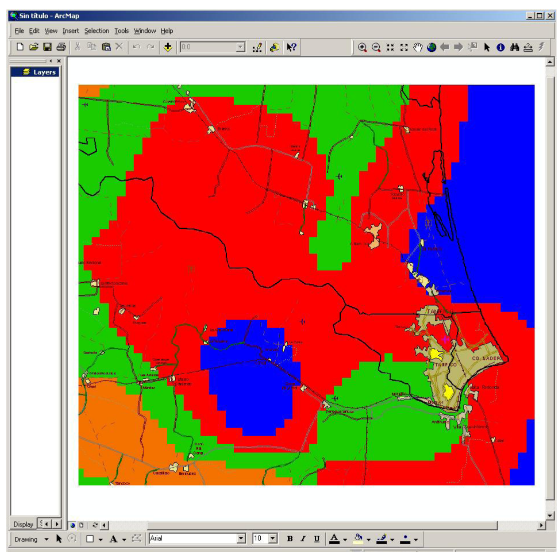

Figure 10. Thematic layers retrieved from a distributed environment

Conclusion

In this work GIS-application works in a distributed environment is described. We propose a XML specification to solve partially the heterogeneity problem of the spatial data. With this application, spatial and attribute data can be recovered from different GIS sites, comparing their qualitative and quantitative properties by means of the XML description.

The spatial database has been designed and implemented into Enterprise GIS (ArcGIS). This repository stores all retrievedgeographic objects from several GIS sites. This mechanism provides a topological model that is integrated by class of elements.

SAM is a module designed to make spatial analysis, related to natural phenomena such as landslides and flooding areas. Users can modify the selection criteria to have different scenarios. Using SAM, it is possible to define the importance of the characteristics of the spatial data, which users consider relevant to the analysis. When the communication is established, SAM requires the spatial data by means of List of Resources, therefore the XML specification is compared with the local XML definition to verify the spatial data and recover the information from different GIS sites. This information is sent to the spatial database through the SDE.The mechanisms implemented in the List ofProcedures can be modified and extended to apply in the analysis of different phenomena.

We propose a XML description, which can be used as a standard to solve partially the spatial data integration problem. This technique is an alternative to recover spatial data in a distributed environment, because the XML definition that has been designed to represent a spatial semantics of the geographical objects can be used to find out solutions related to the spatial interoperability.
Acknowledgements

Work partially sponsored by the Mexican National Polytechnic Institute (IPN), by the Mexican National Council for Science and Technology (CONACyT) and by the SIP-IPN under grants 20110875, 20110988, 20110851,20110881 and 20110302.

References

1. M. F. Goodchild (2000) Perspective: Browsing metadata, where do we go from here? *Geo Info Systems10*(5): 30–31.

2. M. Torres, *Herramienta SIG de Escritorio para la Recuperación, Manejo y Análisis de datos espaciales* (CIC-IPN, México, D.F., 2001), (M.S. Thesis in Spanish).

3. M. Worboys, *Innovations in GIS Selected paper from the First National Conference on GIS Research* (Editorial GME, United Kingdom, 1994).

4. A. MacDonald, *Building a Geodatabase* (ESRI Press, United States, 2001).

5. ESRI Inc., *Understanding ArcSDE, Integrating GIS and RDBMS* (ESRI Press, USA, 1999).

6. OGC, Geography Markup Language (GML), Implementation Specification, version 2.1.1, 25 April 2002, http://www.opengis.net/gml/02-009/GML2-11.html

7. R. Sekhar & P.K. Srimani, A Specification of a Spatial Query Language over GML, *Proceedings of the 9th ACM International Symposium on Advances in Geographical Information Systems*, Atlanta, Georgia, USA, 2001, 112-117.

8. T. Arnold-Moore, M. Fuller, A. Kent, R. Sacks-Davis & N. Sharman, Architecture of a Content Management Server for XML Document Applications, *Proceedings of the 1st International*

1 de mayo 2011 • Volumen 12 Número 5 • ISSN: 1067-6079

Conference on Web Information Systems Engineering (WISE 2000), Hong Kong, 2000, 77-88.

9. R. Sacks-Davis, T. Arnold-Moore & J. Zobel, Database Systems for Structured Documents, *IEICE Transactions on Information and Systems, 34-D*(11), 1995, 37-48.

10. L. Mitas & H. Mitasova, General Variational Approach to the Interpolation Problem, *Journal of Computers and Mathematics with Applications, 16*(12), 1988, 983 - 992.

11. M. Moreno, *La Generalización Automática de la Información Geográfica Multiescala* (CIC-IPN, México, D.F., 2001), (M.S. Thesis in Spanish).

12. M. Torres, M. Moreno & S. Levachkine, SIGETAM: Herramienta GIS de Escritorio enfocada a la Detección de Zonas de Riesgo de Deslave e Inundación, *1st International Congress of Informatics and Computing, Guadalajara, México, 2002, 156-168.*

13. A. Linsley, K. Kohler & S. Paulos, *Hidrology for Engineers (Mc Graw Hill, United States, 1998).*