THEORETICAL FOUNDATIONS FOR GIS

Cristina Fierbinteanu * **Serban Petrescu** **
Catalin Tudose ** **Tudor Metea** ** **Mihai Buneci** ***

* *Titu Maiorescu University*
Calea Vacaresti 187, sect.4, Bucharest, Romania
** *University Politehnica of Bucharest*
Splaiul Independentei 313, Bucharest, Romania
*** *National Institute for Economic Research*
Calea 13 Septembrie 13, Bucharest, Romania

Abstract: *One of the main problems in the field of geographical information systems resides in their history itself: they have been developed from digital mapping systems by practitioners who used ad-hoc, readily available, easiest to use and understand but not necessarily most suitable models and pieces of software from a variety of fields. The most serious problem is that the lack of conceptual modeling, integrating the different data types involved, makes the behaviour of a GIS difficult to predict in the real world. In this paper we present current attempts to lay theoretical foundations for geographical information systems.*

Keywords: GIS, ontology, spatiotemporal objects, active database systems

1. INTRODUCTION

A spatiotemporal object has three kinds of attributes: spatial, thematic and temporal. The storage and retrieval of spatial information that changes over time continues to be an unresolved issue with modern GIS software, although much research effort has been invested in this area recently. As a matter of fact, GIS came to stand for Geographical Information Science, a recognition of the fact that geographical information has special characteristics and its study constitutes a field of scientific investigation in its own rights. Most of the recent research into spatiotemporal data modeling focuses though on the handling of discrete changes in spatial entities. Less attention is given to continuous changes, which are nevertheless essential in some of the most useful GIS applications, like, for example, flood risk analysis.

2. IMPORTANCE OF THIS RESEARCH

Geographical information systems (hereafter referred to as GIS) evolved rapidly during the past 20 years as a natural development of digital mapping systems. A digital mapping system usually accommodates the following activities:

- terrain data acquisition and its represantation by an uniform digital format (not depending on the source being a topographical survey, the scanning of already existing maps, aerial images, etc.)
- grouping of certain physical entities - e.g. elevations, rivers, lakes - into virtual "objects" (points, polylines, polygons) owning "properties" such as geographical coordinates, altitude, lengths, surfaces, neighborhoods, overlappings, labels, etc.
- storing the above mentioned virtual objects and their properties into a database

- retrieving from the database the objects of our interest and representing them in an adequate format (printed, on the screen, etc.)

Using the methodology described above it is possible to develop complex maps, allowing for the “overlaying” of different “themes” (e.g. relief, administrative regions, pedological map) into successive layers (if the database contains the necessary information). Although its roots are in digital mapping, GIS became a complex instrument of modeling, analysis and management of spatial resources, integrating the benefits of different factors of progress, as follows:

- the experience accumulated in other fields of information technology, such as databases, digital image processing, computer aided design, object oriented programming
- the spectacular growth of computational power available, along with the sharp fall of computer prices
- the availability of new methods for primary data acquisition (GPS, high resolution satellite images, etc.), many of these following the declassification of some pieces of military technology.

In spite of the remarkable progress, many of the unresolved requirements and problems are still as actual as they were 20 years ago:

(1) the complexity of the clients’ requirements (there is no general solution to accomodate all the applications)
(2) data acquisition, which may represent a very significant percentage of a GIS implementation cost
(3) databases have to accomodate huge amounts of data
(4) performance (GIS complexity growth implies very large amounts of data being simultaneously accessed by several users, which could easily ”suffocate” an uninspired design)
(5) unlike banking transactions or airplane ticket reservation, GIS database updates are ”long” transactions, inducing consistency and concurrent database access problems
(6) consistency maintenance among the different layers of a map
(7) time tracking of objects from the database. A town is growing up, and the 1990 map is not accurate in 2007, even if some items (a lake for example) are unchanged
(8) SQL-like languages are inefficient for GIS typical querying (e.g. ”select a point within a polygon”)
(9) large polygonal areas problem, for example Norway’s coastal area laced with fiords or an archipelago with thousands of isles, generating an enormous number of points or extremely complicated polylines
(10) interoperability with other information systems (GIS or non-GIS)
(11) complex object definition within the database (vectorial information, raster data, numerical or multimedia data)
(12) the capacity to handle vectorial data, as well as raster data, and to combine it correctly (e.g. overlaying a picture taken by satellite on a ”classical” map)

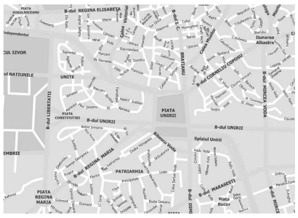

Fig. 1. Standard map (vectorial)

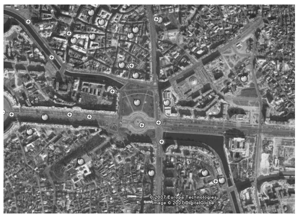

Fig. 2. Space image (raster)

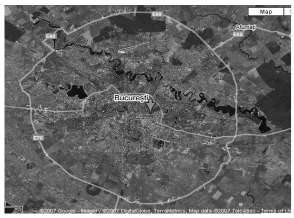

Fig. 3. Hybrid view (vectorial+raster)

(13) the capacity to manipulate map ”objects” using an object oriented programming language, which is close enough to the concept of ”geographical object”
(14) the capacity to handle vectorial data, as well as raster data, and to combine it correctly (e.g. overlaying a picture taken by satellite on a “classical” map)
(15) the capacity to manipulate map “objects” using an object oriented programming language, which is close enough to the concept of “geographical object”

Most of the issues enumerated above are GIS-specific:

- data acquisition (number 2 above) is a difficult process due to multiple, possibly inconsistent data sources (e.g. relief topological survey done at a scale differing from that of the construction plans of a road), requiring detailed verification performed by a human decision-maker

- 2-dimensional representation of a geoid introduces errors, and zooming from 1:100000 to 1:500 requires additional data and algorithms
- the incredibly large amounts of data and the high transfer rates ensured by image capturing technologies require high processing speed and huge storing space
- the continuously growing complexity of the information stored in databases, determined by the ever complicated requirements coming from a growing number of users

One of the main problems in the field of geographical information systems resides in their history itself: they have been developed from digital mapping systems by practitioners who used ad-hoc, readily available, easiest to use and understand (but not necessarily most suitable) models and pieces of software from a variety of fields. For this reason the GIS "database" of thematic (aspatial) attributes uses classical (dBase) files and sometimes totally unsuitable spreadsheet software. The most serious problem is that the lack of conceptual modeling, integrating the different data types involved, makes the behaviour of a GIS difficult to predict in the real world (which means thousands of users simultaneously requiring answers to complex queries from a database continuously being updated).In this paper we present current attempts to lay theoretical foundations for geographical information systems.

3. ONTOLOGY OR ONTOLOGIES?

The term ontology has been given different interpretations in the AI technical literature. Andrew U. Frank (Frank 2006, Frank 2007) introduces a temporal, multi-tier ontology, aiming to integrate different approaches. The four layers proposed are: the physical environment , observations of the environment and activities, the reality of objects (objects and actions, classes and operations), the social reality (legal reality, subjective reality, communication). Frank's conclusion is that, although most efforts to structure ontologies strive for maximum generality (a single set of rules applicable to every case study in the domain), no single ontology proposed by now covers all areas important in a GIS. In our project we assume that an ontology is a collection of concepts (vocabulary) used as building blocks for knowledge-based systems. In knowledge engineering terms, this is a task ontology (Heijst 1997, Vanwelkenhuysen 1996).

4. ACTIVE DATABASES

Active databases are able to monitor and react to specific circumstances of relevance to an application. The description of the active functionality of the database is supported by a knowledge model.

The knowledge model of a rule can have up to three components: an event, a condition, and an action. An event can have various different types of causes: structure operation, behavior invocation, transaction, abstract or user-defined, exception, clock, external[9]. According to the type of the event, the applications of active databases can fall into one of three categories: database system extensions (e.g. implementing integrity constraints using active rules), closed database applications, and open database applications. In the last category, some monitoring devices outside the database record external conditions, producing events that trigger actions in the database. GIS applications fall into the last category (for example in an air traffic control system, a rule could trigger an action in the system when two airplanes are getting closer than a given distance; or in a flood monitoring system, a rule could inform when the level of waters goes above a certain limit).

The runtime strategy of an active database is closely related to the underlying DBMS. With Oracle9i it is possible to monitor changes in the data, using the Java Message Service (JMS) to dequeue messages placed into the Advanced Queuing system by the PL/SQL trigger. The JMS notifies a Java client application (for example a GIS-based monitoring system) that data has changed, so the application can update its display (O'Rourke 2003).

5. METHODOLOGY

It is necessary to research both the evolution of one single object (e.g. the shape of a river, the position of a car), as well as the spatiotemporal relationships among objects (e.g. the overlapping of a river channel and a river meadow, the existence of a car within a given distance from a hospital). The information necessary to keep track of spatial changes of geographical objects over time is complex and therefore special algorithms are required for storing it, querying and displaying the corresponding map. We are planning to investigate the applicability of parallel processing and knowledge engineering (e.g. constraint solving) specific methods in the field of spatiotemporal data modeling and also draw on active database research results.

6. STUDY ON THE INTEGRATION OF HETEROGENEOUS MAPS

One step towards the dynamical integration of heterogeneous maps is the identification of corre-

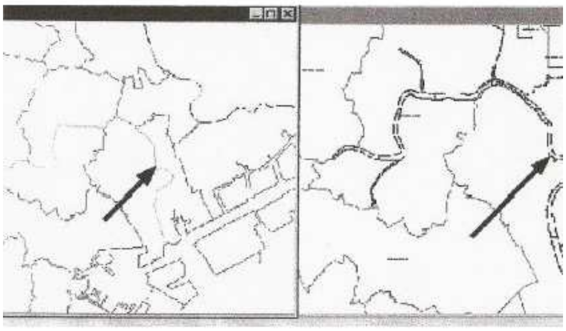

Fig. 4. Related objects in maps of different scales

sponding objects on maps of different scales. This process is difficult because related objects:

- do not always overlap flawlessly, due to the limited precision of the mapping
- may have different graphical representations (Figure 4).

In our previous research we considered a large-scale map and a mid-scale map. We assumed that one of the maps has been in use within the GIS and the other is being newly imported into the system. The problem is to determine, for each entity of the old map, the corresponding entity/entities in the newly acquired map. We presented a model in which the only concept used at the user interface level is the concept of theme. A theme has at its basis entities and layers. We used this model to determine corresponding objects in maps of different scales (Cretu 2005, Fierbinteanu 2001).

7. FIRST STEPS TOWARDS A DATA MODEL FOR SPATIOTEMPORAL THEME MAPS

In (Wang 2003) several spatial operations were defined (Table 1). All these operations require a parameter (or parameters) holding geometric information . The parameter can be a geoentity, a geoentity class or a theme. As geoentity classes and themes are defined as sets of geoentities and sets of geoentity classes respectively, the operations having a theme or a geoentity class as a parameter will return a list of tuples.

A few typical temporal operations were defined too, as well as an extended WHERE clause for SQL statements. An example of using this is shown below.

SELECT block
FROM block, building
WHERE (OVERLAY (block, building) > 2 and building.classification = 'hotel') AND (VALIDTIME (building) and building.classification = "hotel")

8. FURTHER WORK

The aim of our project is to develop a conceptual framework for geographical data modeling that will provide support for efficiently tracking the evolution of spatial attributes of map objects. Interdisciplinary study is required to achieve this objective (cartography, information systems, topology as a branch of mathematics, but also the theory of knowledge as a subdomain of philosophy). At present there exists no such conceptual model for spatial databases. We are looking for a simple but powerful model (similar to that introduced by Codd in 1970, for classical databases). The discovery of such model would have a strong impact on both the academic community and industry.

Based on the conceptual model that will be designed we shall build a library of pre-written analysis and design tools for the WWW. These tools could become the building blocks for applications in a wide range of domains, extending from flood risk analysis and management methodologies to emergency response guides (police/fire/ambulence) and air traffic control systems.

We shall pay special attention to the investigation of GIS application in the domain of virtual learning environments, especially the possibility of using GIS technology to guide navigation in a virtual university campus constructed on the basis of a city metaphor.

REFERENCES

E. Cretu, C. Fierbinteanu, A. Creosteanu, M. Maekawa: *Integrarea dinamica a hartilor eterogene intr-un sistem informatic geografic*, NAV-MAR-EDU 2005, Academia Navala Mircea cel Batran, Constanta, iunie 2005.

Cristina Fierbinteanu, K.Nakayama, M.Maekawa, K.Wang, K.Ishikawa and T.Matsumoto: *Towards the dynamical integration of heterogeneous maps in MapBrowser*,GIS Research UK, GISRUK 2001, University of Glamorgan, April 2001.

Andrew U. Frank: *How to Structure an Ontology, Bremen, Germany*, http:// www.geoinfo.tuwien.ac.at/ presentations/ frank.htm, 2006.

Andrew U. Frank: *Assessing the Quality of Data with a Decision Model*, ISSDQ, Enschede, Netherlands, 2007, ftp:// ftp.geoinfo.tuwien.ac.at/ frank/ enschede_issdq_2007.pdf

G. Van Heijst, A.Th.Schreiber, and B.J. Wielinga: *Using explicit ontologies in KBS development*, International Journal of Human-Computer Studies, vol.46, no.2/3,1997, 183-292.

Norman W. Paton and Oscar Diaz: *Active Database Systems*, ACM Computing Surveys, Vol. 31, No. 1, March 1999, 63-103.

Cameron O'Rourke: *The Active Database: Ways to invoke external processing from the database*, Oracle Magazine, September/October 2003.

J. Vanwelkenhuysen and R. Mizoguchi: *Analysis of representational design decisions to create ontologies and guidelines*, Journal of Japanese Society for Artificial Intelligence, vol.11, no.2, 1996, 136-146.

K.Wang,C.Fierbinteanu and M.Maekawa: *A Data Model for Spatiotemporal Theme Maps: Its Querying and Implementation*, GIS Research UK, GISRUK 2003, City University, London, pp.99-105, aprilie 2003.

Shuo Wang: *Research of Spatiotemporal Data Model for GIS*, PhD Thesis, Department of Information Systems, Graduate School of Information Systems, The University of Electro-Communications, Tokyo, Japan, October 2006.

Operation	Function result
Disjoint(A, B)	true if geometries A and B satisfy the predicate Disjoint
Intersect(A, B)	true if geometries A and B satisfy the predicate Intersect
Touch(A, B)	true if geometries A and B satisfy the predicate Touch
Cross(A, B)	true if geometries A and B satisfy the predicate Cross
Within(A, B)	true if geometries A and B satisfy the predicate Within
Contains(A, B)	true if geometries A and B satisfy the predicate Contains
Overlap(A, B)	true if geometries A and B satisfy the predicate Overlap
Area(A)	area of geometry A
Length(A)	length of geometry A (line or multilines)
Perimeter(A)	perimeter of geometry A (polygon or multipolygons)
Distance(A, B)	shortest distance between two geometries A and B
Shortest-path(A, B)	shortest path between two geometries A and B
Overlay(A, B)	overlaid geometry of geometry A and geometry B
Buffer(A)	buffering geometry of geometry A
ConvexHull(A)	convex hull of geometry A
Intersection(A, B)	geometry representing the intersection of geometry A with B
Union(A, B)	union geometry of geometry A and B
Difference(A, B)	difference geometry of geometry A and B
SymDiff(A, B)	symmetric difference geometry of geometry A and B
Altitude(A)	altitude of geometry A
Inside(A, B)	true if A is three-dimensionally within B
Across(A, B)	true if A cross B with an overpass
Higher(A, B)	true if A is higher than B
Under(A, B)	true if A is under B

Table 1. Spatial operations from (Wang 2003)