ICA Workshop on Generalisation and Multiple representation – 20-21 August 2004 – Leicester

Multi-representation in spatial databases using the MADS conceptual model

Christelle Vangenot

EPFL – Database Laboratory – 1015 Lausanne – Switzerland
christelle.vangenot@epfl.ch

ABSTRACT

This paper proposes a conceptual data model providing full support for multiple representations of the same real world data. Our proposition is built around two complementary approaches: the integrated approach, that leads to the definition of customized database items and the inter-relationship approach, where the representations are linked through inter-representation links. Both approaches use the stamping technique to differentiate among multiple representations of a given phenomenon and to access a particular representation. This proposal has been tested and validated with users, and implemented as a front-end to existing DBMS.

Introduction

Databases are intended to store representations of real world phenomena providing the information needed to support application requirements. Databases are often shared by several applications that perform various processing tasks and hence have different requirements in term of how information is kept, described, organized (in terms of data structures), coded, what constraints, processes, and rules apply, how it is presented... Consequently, the same real world phenomenon may well be represented in a database in different ways for different purposes. This also applies in the spatial domain where for instance, the same river may be seen by some applications as having a linear extent and as having an area extent by some other applications and finally as a collection of objects, each one defining a river segment for another application.

Many factors are driving the elaboration of representations for the same phenomenon to be stored in a geographical database. The intended use of the data is a first factor; we call it the viewpoint. Viewpoint characterizes the user vision of the real world. For instance, a traffic management viewpoint views roads as segments linking different points in space, while a road maintenance viewpoint is likely to be interested in the status and nature of the road pavement. Resolution, both spatial and semantic, is another essential factor: spatial resolution determines the level of detail in capturing geographic features (i.e. a river as a line or an area) and semantic resolution defines the desired level of detail for thematic data (i.e. the value of land use taken among the three values built-up area, cultivated area, and wild area or in a more detailed set of values: industrial, individual habitat, low-cost habitat, rural habitat, commercial zone, park, etc).

1) The underlying MADS data model

Our objective was to propose a conceptual data model with multi-representation facilities. The approach we choose is to propose a set of multi-representation concepts and rules as an extension of an existing spatio-temporal conceptual data model, MADS[1] [12].

MADS is an object+relationship spatio-temporal conceptual data model. In this model, the real world of interest that is to be represented in the database is composed of complex objects, their relationships in between both characterized by properties (attributes and methods) and both possibly participating into generalization hierarchy (is-a links). Spatiality and temporality may be associated at the different structural levels: object, attribute and relationship. The spatiality of an object conveys information about its location and its extent while its temporality describes its lifecycle. For instance in Figure 1, the object type TouristSite has a spatiality (an area). Attributes may have spatial (e.g. the attribute Entrance of the object type Museum in Figure 1) or temporal (e.g. the attribute OpenTime of the object type Museum) domains of values.

A set of predefined spatial and temporal abstract data types, organized in a generalization hierarchy, are used for describing the spatial and temporal extents of data. Attributes may be space- or time-varying (like the attribute Exhibition of Museum), in this way supporting the continuous view of space and time. Relationships are either classical n-ary relationships among individual objects or n-ary associations among sets of objects (multi-association). Relationships may hold one or several specific semantics such as topological and synchronization, that define constraints between spatial (respectively, temporal) objects, or aggregation.

2) Perception Stamps

Multi-representation has been added in MADS as an additional orthogonal dimension to the structural, spatial and temporal ones. To allow users retrieving the desired representations from the set of existing representations, representations have to be distinguishable and denotable. To this extent, we propose to use *perception stamps* (simply hereinafter denoted stamps). Stamps are added on data, whether they are object type instances or attribute values, and on meta-data, object and relationship type definitions or attribute definitions.

Stamps are vectors of values characterizing the context of each perception. In our work we restrict stamps to be pairs of (resolution, viewpoint). Stamps have a twofold semantic: they allow to distinguish the multiple perceptions of the same phenomenon and also to filter access to data during querying.

The first step for the database administrator is to identify the perceptions that are to be supported by the database and to associate a unique stamp to each one of them. This defines the set of stamps that are allowed for use with the database. For instance, if we consider two descriptions made by two different tourist offices describing the same geographical area, the designer has identified two perceptions: one for each tourist office. Moreover, the second tourist office works at a coarser resolution. Thus the database designer will create two stamps:

---

[1] MADS stands for "Modélisation d'Applications à Données Spatiales" (in French).

s= < Viewpoint = "Tourist Office 1", Spatial resolution = 10>

t= < Viewpoint = " Tourist Office 2", Spatial resolution = 50>

Once created, those two stamps may be used to stamp any schema element. Stamping an element defines for which perception the element is relevant. Thus, an object or relationship type relevant for several perceptions bears several stamps. It may also show different attributes depending on the perception. Its attributes may be stamped with a subset of the stamps associated to the type. An element that has a single representation may also bear multiple stamps, meaning that the same representation is shared by the perceptions identified by the stamps.

3) Strategies for multi-representation

Multiple representations of a given phenomenon may basically be organized according to two strategies, both based on stamping, that may be used separately or in combination:

In the first strategy, the *multi-representation strategy*, the idea is to merge the different representations of the same real world phenomenon in a single database element and associate to it the stamps identifying the perceptions for which it is relevant. We call such an element a perception-varying database element.

Every concept of the model may be perception-varying:

– *Object and relationship types* may be perception-varying types. They may bear several stamps and show different sets of attributes according to the considered stamp. Figure 1(a) shows several perception-varying object types: *TouristSite, Museum* and *Monument.* They have two definitions, one for stamp *s* and one for stamp *t*. For instance, *Museum* for stamp *s* includes the attributes *Entrance*, *Exhibition* and *OpenTime* and for stamp *t* the attributes *Entrance* and *Description*.

– Attributes may have *different definitions* i.e. different cardinalities and/or value domains according to the stamp. For instance, the attribute *District* of *TouristSite* has several definitions, one of domain string for stamp *s* and one of domain integer for stamp *t.* This means we could store two values for districts (e.g. "*Côte d'Or*" for stamp *s* and *21* for stamp *t*).

– *Attributes* may contain a value that is function of the stamp. We call them perception-varying attributes. The notation of such an attribute is *f(S)*. For instance, the attribute *Name* of *TouristSite* or the attribute *Geometry* of *TouristSite* are perception-varying attributes. It is thus possible to store several values at different resolutions for the *geometry:* one for stamp *s* and one for stamp *t*.

– *Relationship types*, *their roles and cardinalities* may also be perception-varying: For instance, relationship types may change their semantic according to the perception and be a topological relationship of adjacency in one perception and be a topological relationship of intersection in another one.

– Finally, stamps on *is-a links* provide for perception-varying generalization/specialization hierarchy. For instance in Figure 1(a), as the specialization of museums in either private or public museums is only of interest for stamp *s*, the is-a link is stamped with *s*.

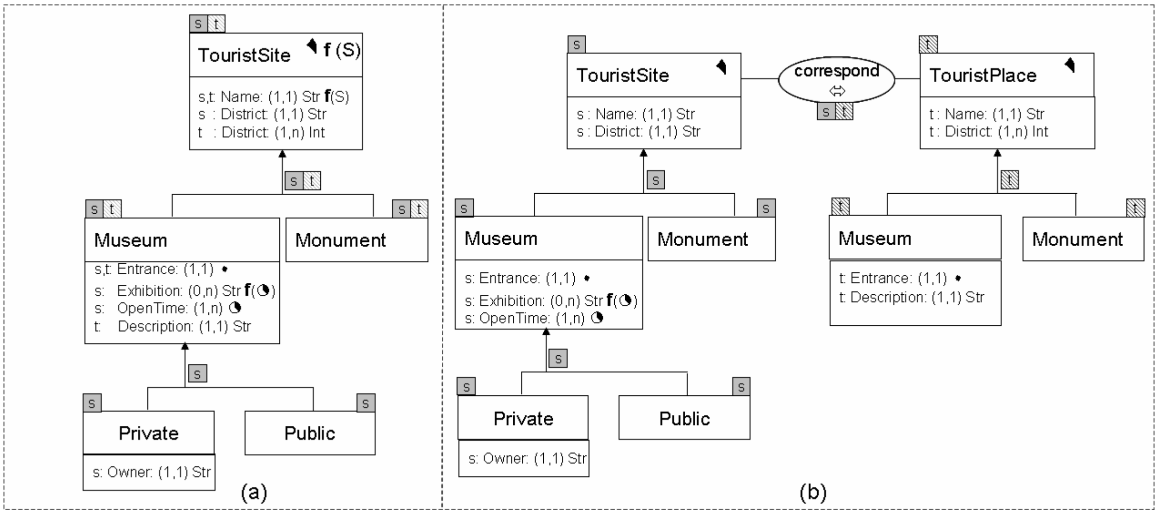

Figure 1: (a) Multi-representation strategy, (b) Inter-representation strategy

The second strategy, the *inter-representation strategy*, is recommended when representational needs for the same phenomena are so diverse that they can hardly be integrated into a common definitional framework. In this case, the different perceptions of the same objects should be related in order to state that they are different perceptions of the same real world object. In order to distinguish inter-representation links from classical ones, they are associated with a specific semantics called *inter-representation semantics*. It may be hold by associations and multi-associations. Some relationships such as is-a links and aggregation relationships have an inherent *inter-representation* semantics that do not need to be explicitly stated. For instance the *correspond* relationship in Figure 1(b) holds the inter-representation semantics (graphically represented by the icon ⇔) which states that the instances of *TouristSite* and *TouristPlace* linked through this relationship are two perceptions of the same real world object.

Manipulation of multi-representation data has also been studied: Data may be queried thanks to an algebra enhanced with multi-representation capabilities. First, the extended algebra allows to select a specific subset of the database according to one or several stamps. Then it is possible to work either on a mono- or on a multi-representation set of data. Stamps are associated to queries to specify which data is to be considered by the query.

4) Results

In conclusion, our proposition is intended to provide concepts to store multiple representations of data in a database according to two facets: resolution and viewpoint. The proposed technique allows for:

– *Modularization of data*: Each stamp delimitates a subset of the database (at the meta-data and data level) and thus allows for object and link filtering according to the viewpoint or the resolution. In the viewpoint dimension, stamps are a way to materialize the universe of discourse of each viewpoint. In the resolution dimension, this capability is particularly relevant as it allows to filter object and relationship types or instances that are no more represented when going from a detailed resolution to a coarser one.

– *Personalization of data*:

- Stamping attributes allows for the definition of customisable data structures according to the viewpoint or the resolution. Especially this allows to filter attributes that are no more relevant when the resolution becomes coarser.
- Stamping values allows for the choice of the one that best fits the objectives of the viewpoint. According to the resolution, users are able to store several values for attributes, one for each resolution. This is particularly interesting for spatial attributes whose value may change according to the resolution but also for some thematic attributes like for instance the land use attribute described in the introduction.

– *Relating the different representations of the same real-world entity*: Stamping provides an easy way to identify which representations stem from a given perception. We must also design the schema to make the DBMS aware of the existence of multiple coexisting representations of the same phenomena. As presented in the last section, we propose two ways to organize this in the database:

- One way is to build a single object type that contains both representations, the knowledge of "which is which" being provided by the stamps of the properties of the type.
- Another way is to define two separate object types, and to link them with a relationship type that holds a specific inter-representation semantics. We propose a complete set of relationships adapted to the different kinds of possible correspondences: is-a links (one to one correspondence with same identifier), traditional relationships (one to one correspondence with different identifier), aggregation relationships (one to many correspondence) and multi-association (many to many correspondence). The two last ones are of particular interest in the resolution dimension as they allow to define hierarchical relationships between objects.

5) State of art and Discussion

Research on multi-representation in spatial databases mostly relates to multi-scale databases, i.e., databases that allow an object to be characterized by several representations of its geometry, each one at a different scale. Explicit storage of multiple geometries is needed as there is no fully automated way to derive geometry at some scale from the geometry at another scale. For this reason view definition can still be used to support multiple points of view, but multi-resolution needs cannot be fully satisfied using such a purely deductive approach.

A significant part of the research in multi-resolution databases was inspired by the largely hierarchical nature of transitions between scales. Because of the largely hierarchical nature of transitions between scales, multiple representations in multi-scale databases are most often organized into hierarchical data structures, where levels in the hierarchy correspond to increasing detail [5] [17]. Links in the hierarchy provide a path to update propagation. [1] presents different kinds of links (or different types and with different cardinalities) useful to propagate updates. A combination of an aggregation hierarchy, a generalization hierarchy, and a filtering hierarchy forms the map cube model in [16]. Topological structures are added in [10]. The impact of spatial and semantic resolution on data representation from both the modeling and querying points of view is analyzed in [13] and [14], where the authors consider geographical zones that fit into each other. [15] see the database organized as a stratified map space, where each map gathers objects that share the same semantic and spatial granularity. Maps are grouped by map spaces ; i.e., sets of maps showing the same schema at different granularities. Some approaches, [4], are focusing on the definition of granularity. [9] present a deductive knowledge-based system, Geodyssey, that provides a solution for the design and the implementation of multi-scale, multiple temporal, multiple representation spatial databases. The system incorporates a reasoning process to maintain consistency and propagate updates. The query interface can derive simplified representations of objects, if no suitable representation is found. OMT-G is a modeling technique for spatial databases that is based on an extension of the OMT model [3]. Its class diagram describes geo-object classes (i.e. object classes with a geometry like point, line, polygon, node, and arc linking two nodes), spatial relationships (i.e. relationships with a spatial integrity constraint), and geo-fields that provide users with a continuous view of space. OMT-G also supports several geographic representations for a geo-object by allowing designers to attach to any object class several sub-classes describing its alternative geometries. Whenever known, the transformation rules that derive a geometry from another one are described in the transformation diagram. OMT-G also supports one or several alternate graphical legends for each geo-referenced data. An object-relational proposal is formulated in [6], [7].

Compared to the above mentioned research in spatial databases, we mainly depart from it by the fact that we address multi-representation in general; i.e., it is not limited to the geometry of objects, and we do not restrict the approach to hierarchical structures of the set of representations. While hierarchies are very useful to convey specific semantics (e.g., this representation is more detailed than that one), there are complex real world structures that cannot be nicely represented as hierarchies. On the other hand, the hierarchical semantics inherent in spatial resolution may be expressed in our approach through integrity constraints on representations associated with given stamps.

The proposition of [2] represents the major other effort to fully support multi-representation. This solution, VUEL, allows the association of several thematic, graphical and spatial characteristics to the same geographical object. Their concept, the VUEL (View Element) fulfills the requirements specific to MOD ("map on demand") and multiscale SOLAP context.

6) Conclusion and future work

We have specified a solution for the multi-representation and manipulation of spatial data. This work was done as part of the MurMur project [11]. It has lead to the definition of a schema editor, a query builder as well as associated mappings to several DBMS and GIS. The proposed concepts and solution are intended to be easy to understand and close to the way user perceived the real world.

Future work needs to be addressed in the following directions:

Consistency, derivation between representations: In the spatial context, as data from one representation may often result from the derivation of the same data at another resolution, one may expect to state constraints between them. Such constraints may entail for instance that the geometry of roads for stamps *s* should always be spatially included in the geometry of roads for stamps *t*. Moreover, if derivation rules are known, the model may also provide tools to derive the representations instead of storing every representations. This capabilities may lead us to consider the more intricate problem of *update propagation issues* between representations.

Multi-representation in the continuous view of space: Diversity of perception and representation of spatial features also exists in the continuous view of space. In the continuous view, multi-representation may result in the definition of the same field with various spatial resolutions (i.e. in the definition of different spatial locations or of different spatial extents) and in the mapping of the same value but at different semantic resolution to the same field.

Describing the correspondences between the continuous and discrete representation of space. A single database or GIS may contain the description of a phenomenon perceived according to the continuous and the discrete views of space and whose correspondence needs to be recorded. Very few works exists on the issue of describing the correspondence between the continuous and discrete perceptions of the world.

7) References

[1] T. Badard. On the automatic retrieval of updates in geographic databases based on geographic data matching tools. *Proceedings of the 19th International Cartographic Conference, Ottawa'99*, ICA/ACI (Eds.), Ottawa, Canada, p. 47-56, 1999.

[2] Y. Bédard, M.-J. Proulx, S. Larrivée, E. Bernier. Modeling Multiple Representations into Spatial Datawarehouses: A UML-based Approach. *Proceedings of the Joint Workshop on Multi-Scale Representations of Spatial Data, ISPRS WG IV/3*, ICA Com. on Map Generalization, Ottawa, Canada, 2002.

[3] K. Borges, C. Davis, A. Laender. OMT-G: An Object-Oriented Data Model for Geographic Applications. *GeoInformatica*, Vol.5, N°3, p. 221-259, September 2001.

[4] E. Camossi, M. Bertolotto, E. Bertino, G. Guerrini. A Multigranular Spatiotemporal Data Model. *Proc. Of 11th International Symposium on Advances in Geographic Information Systems (ACM GIS 2003),* New Orleans, Louisiana, November 7-8, p. 94-101, 2003.

[5] A. Franck, S. Timpf. Multiple representations for cartographic objects in a multi-scale tree: An intelligent graphical zoom. *Computers & Graphics*, 18(6), 1994.

[6] A. Friis-Christensen, D. Skogan, C.S. Jensen, G. Skagestein, N. Tryfona. Management of Multiply Represented Geographic Entities. *Proceedings of the International Database Engineering & Applications Symposium, IDEAS'02*, Edmonton, Canada, IEEE Computer Society, 2002.

[7] A. Friis-Christensen, C. S. Jensen. Object-Relational Management of Multiply Represented Geographic Entities. *Proceedings of the Fifteenth International Conference on Scientific and Statistical Database Management*, Cambridge, MA, USA, July 9-11, 183-192, 2003.

[8] T. Kilpelainen. Maintenance of topographic data by multiple representations. *Proceedings of the Annual Conference and Exposition of GIS/LIS '98*, Forth Worth, Texas, 1998.

[9] C.B. Jones, D.B. Kidner, L.Q. Luo, G.L. Bundy, J.M. Ware. Database design for a multi-scale spatial information system. *International Journal of Geographical Information Systems*, 10(8): 901-920, 1996.

[10] M. Molenaar. Composite objects and multiscale approaches. *An introduction to the theory of spatial object modeling for GIS*, Taylor & Francis, 1998.

[11] MurMur project: http://lbd.epfl.ch/e/MurMur/

[12] C. Parent, S. Spaccapietra, E. Zimanyi. Spatio-Temporal Conceptual Models: Data Structures + Space + Time, *Proceedings of 7th ACM Symposium on Geographic Information Systems*, GIS'99, Kansas City, Missouri, 1999.

[13] P. Rigaux, M. Scholl. Multi-scale partitions: Applications to spatial and statistical databases, *Proceedings 4th International Symposium on Advances in Spatial Databases, SSD'95*, LNCS 951, Springer-Verlag, 1995, p. 170-183, 1995.

[14] M. Scholl, A. Voisard, J.P. Peloux, L. Raynal, P. Rigaux. Systèmes de Gestion de Bases de Données Géographiques, Spécificités. International Thomson Publishing, 1996.

[15] J. Stell, M. Worboys. Stratified Map Spaces: A formal basis for multi-resolution spatial databases, *Proceedings of the 8th International Symposium on Spatial Data Handling, SDH'98*, 1998.

[16] S. Timpf. *Hierarchical structures in map series*, Ph.D. Thesis, Technical University Vienna, 1998.

[17] S. Zhou, C. B. Jones. Design and Implementation of Multi-Scale Spatial Databases, *Proceedings 7th International Symposium on Spatial and Temporal Databases*, Redondo Beach, CA, July 12-15, 2001.