MULTI-REPRESENTATION: A NEW PARADIGM FOR DATABASES?
Invited Talk - Extended Abstract[1]

Stefano Spaccapietra[1], Christelle Vangenot[1], Christine Parent[2]
[1]Database Laboratory EPFL-IC-LBD   *[2]HEC-Inforge University of Lausanne*

Key words: multi-representation, spatial databases, temporal databases, database modeling.

While the real world is supposedly unique, the perception we have of it is extremely personal. For instance, a traffic manager is likely to consider roads as segments linking different points in space, while a road maintenance service is likely to be interested in the status and nature of the road pavement. Perception is also time varying, as the same person looks differently at the world at different instants. In order to reconcile diverging requirements from users sharing the same database, research in databases has always looked for ways to support multiple representations of the same data.

Currently, a number of different techniques support some facet of multi-representation to some extent, e.g. views, semantic modeling constructs (namely, is-a links and roles), versions, timestamping, multi-scale hierarchies in geographical databases, and profiles in current Web services.

The view mechanism is the most widespread technique, available in commercial DBMSs. The mechanism relies on the hypothesis that all information requirements about each object type have been integrated[2] into a global all-embracing representation, which is stored as the basic

---

[2] View integration has always been a fundamental step in the process of designing a new database.

representation. Views may then be defined to derive ad-hoc representations from the basic representations already defined in the database. Views are virtual constructs that are linked to the database schema, but remain external to it.

Semantic modeling is an approach that aims at describing as accurately as possible the real world and its semantics. Whenever multiple representations are needed, its aim is to inscribe them as coexisting representations in the database schema. It has, up to now, focused on supporting multiple classifications. It relies on the concept of generalization/specialization hierarchies, where each Is-a link defines two object types as alternative representations for the same real world objects. Is-a links come with associated consistency constraints between the two representations, known as population inclusion and property inheritance. Generalization hierarchies are commonly supported by object-oriented DBMSs, and by some object-relational DBMSs. However, only static hierarchies are supported, i.e. there is no provision for an object to move from one class to another. Moreover, an object can only be instantiated in one class, the most specific one it belongs to. Such constraints basically restrict the data modeling expressive power to classification refinement, making the approach not fully suitable to support the complexity of classification schemes that may be superimposed to the real world.

Consequently, research in object-oriented and semantic modeling has proposed an additional concept, the role concept, as a way to provide more flexibility than is-a links, supporting multi-instantiation and more dynamic classification schemes. The concept has been investigated by many authors, suggesting different properties and rules, but is not supported by commercial DBMSs.

Versioning is another technique, mainly used to support evolution of representations over time, including both schema evolution and data evolution.

Meanwhile, evolution of data values at the instance level has been extensively addresses by research in temporal databases.

Moving from time to space, users of geographical databases have been looking for so-called multi-scale databases, holding multiple representations of the geometric features of spatial objects at different resolution levels. While current GIS (Geographic Information Systems) offer quite limited facilities to support multiple resolution, GIS research has developed a variety of hierarchical models to cope with multiscale data.

Finally, personalization in web services (based on profiling techniques) is yet another facet of multi-representation, that includes presentation aspects related to visualization in user interfaces.

In summary, different techniques address different facets of multi-representation, each one using specific concepts and mechanisms. However, full support remains to be achieved. Moreover, there is no system that offers all of these techniques, and it is even questionable whether such a system is desirable. It would most likely be too complex for users to grasp the heterogeneity of possibilities and understand when to use which facility.

This contribution reports on the results of the MurMur project, in which the authors are involved[3], a 3-years research and development effort that aims at specifying and implementing a generic functionality to provide fundamental support for all facets of multiple representations. The project specifically articulates on the viewpoint and resolution facets, but the proposed solutions do not depend on any specific facet nor on any specific underlying data model or application. New concepts have been introduced as an extension to the MADS data model, a conceptual spatio-temporal data model developed by EPFL and University of Lausanne, but they could similarly be added to any other data model with similar expressive power.

Beyond extending current data modeling practices to a new modeling dimension, the representation dimension, our longer term plan is to assess whether multi-representation is an approach that provides a new insight into the database services that we mentioned above, leading to revisiting the existing solutions and possibly replacing them with new solutions stemming from the multi-representation paradigm. For instance, a question that naturally comes up, once we assume that multi-representation is fully supported, is whether we still need temporal databases, or are they just a specific case of multi-representation. Replacing a collection of heterogeneous techniques with a single and generic modeling dimension opens the way to simpler DMBS implementation and easier understanding of the database approach by application-oriented users. It could lead to a new generation of database systems. But the assessment remains to be done, hence the question mark in the title of this contribution.

From the user perspective, DBMS support for multi-representation provides three very important benefits. The first one is to enable easy development of applications that need to navigate among different representations. An example is current multiscale way-finding applications that provide directions to moving users, in particular car drivers. The second benefit is easier enforcement of inter-representation consistency constraints, thanks to triggers or similar DBMS mechanisms implementing integrity

---

[3] Partners in the MurMur project are Universities of Brussels and Lausanne, EPFL, the Institut Géographique National in Paris, The Cemagref research center in Grenoble and the GIS-provider Star Informatic, Belgium.

constraints and derivation rules. The third benefit is automatic update propagation, an essential feature in consistency enforcement. Automatic update propagation is only possible if the DBMS knows about the multi-representation semantics of the data it handles. We actually experimented our approach in two real case studies. For one of them, provided by the French National Mapping Agency (IGN), the main issue was to turn three existing databases covering the French territory at different scales into an interrelated data store whose consistency could be checked (using, in particular, spatial matching algorithms) and maintained through automatic update propagation. Reducing current high update cost (due to the need to run three independent update processes) is expected to provide benefits that by large will compensate for the investment in multi-representation solutions.

Our work proceeded from this new perspective, aiming first at the definition of fundamental concepts and rules for this new modeling dimension. We analyzed multi-representation requirements and turned the analysis into basic features to be added to a data model. The solution we came up builds on a simple fact: The basic need to support different representations is to be able to identify them unambiguously. To that extent, we use a "representation stamp" (just like in temporal databases timestamps identify which data is valid at a given instant in time) (see figure 1). Throughout our work, we insisted on developing a stamping mechanism that is orthogonal to the other features of the model. This results in an elegant solution, with fewer and cleaner concepts and rules, that should be easily understandable by designers and users. We developed support for two approaches to model multi-representations. In the first one, the needed representations are grouped into a unique object type whose properties may be tailored to be either common to several representations (see, e.g., the Number attribute in figure 1), or relevant for a single representation only (see, e.g., the Dpt. attribute in figure 1). In the second approach, one object type per representation is defined, and these object types are interrelated by one of various mapping relationships (e.g., Is-a, Aggregation, Equivalence) that are available in the data model to express the different semantics of the multi-representation situation.

Figure 1. An illustration of a bi-representation object type

Notations s1 and s2 denote representation stamps. The icons denote spatial abstract data types (line for s1 and area for s2). The semantics is that the Road object type is a representation of roads that is valid for stamps s1 and s2, but has different properties for each stamp (except for the Number property that is fully shared by both representations).

The proposed extensions have been successfully incorporated into the MADS data model, and tested and implemented within the framework of the MurMur project.

This contribution is mainly intended to sketch our vision, present the approach and discuss its main features. More information is available on the website http://lbd.epfl.ch.

REFERENCES

1. C. Vangenot, C.Parent, S. Spaccapietra. *Modeling and Manipulating Multiple Representations of Spatial Data.* SDH'02. Proceedings of the 10th International Symposium on Spatial Data Handling, Ottawa, Canada, July 9-12, 2002.
2. C. Vangenot. *Supporting decision-making with alternative data representations.* Journal of Geographic Information and Decision Analysis. JGIDA Vol. 5, No. 2., 2001.