# Markov Model Mediator Mechanism vs Object-Oriented Data Model As a Data Warehouse Schema

Mei-Ling Shyu

School of Electrical and
Computer Engineering
Purdue University
West Lafayette, IN 47907

Shu-Ching Chen    Naphtali Rishe

High-Performance Database Research Center
School of Computer Science
Florida International University
Miami, FL 33199

## Abstract

A data warehouse is a conceptual repository that collects information from one or more data sources and provides a schema that allows this repository to be queried by users of the warehouse. The difficulties in designing a schema for a data warehouse whose member databases are multimedia databases are mainly caused by the highly diverse nature of multimedia data, a very wide range of multimedia requirements, and the heterogeneous schemas of the member databases. In this paper, two possible models for the schema of a data warehouse composed of multimedia database systems – the Markov Model Mediator (MMM) mechanism and an object-oriented data model – are compared based on some design issues. The comparative analysis shows that the proposed MMM mechanism provides a better schema for a data warehouse than an object-oriented data model.

## 1 Introduction

Data warehousing employs database technologies for storing and maintaining data. In [4], a data warehouse is defined to be a subject-oriented, integrated, time-varying, non-volatile collection of data. A data warehouse does not create value by itself; value comes from the data in the warehouse [8]. In other words, data warehouses are built in the interests of business decision support and contain a summary of historical data from a number of operational databases [5]. Many queries over data warehouses require summary data that can be supported by the conceptual schemas (or views) of the data warehouses. To speed up query processing, a proper conceptual schema should be designed for a data warehouse since the performance of a query depends on how the schema is implemented.

A data warehouse consists of one or more data sources with heterogeneous schemas. A uniform access mechanism to heterogeneous information stored in databases with different data models and access methods should be provided to fulfill users' requests for readable and useful information. Furthermore, the schema of a data warehouse for a multimedia information system should have the ability to model a variety of multimedia data in terms of their structures, behaviors, and functions.

In our previous studies [6, 7], we proposed a *Markov Model Mediator (MMM)* mechanism that can serve as the schema of a data warehouse. Another possible model for a data warehouse schema is the object-oriented data model. In this paper, we perform a comparative analysis between the MMM mechanism and the object-oriented data model based on some issues relevant to the design of a data warehouse. The comparative analysis shows that our proposed MMM mechanism provides more functionality than an object-oriented data model when it is used to model the schema of a data warehouse for a multimedia information system.

This paper is organized as follows. In the next section, the proposed MMM mechanism is briefly introduced. Section 3 presents the comparative analysis between the MMM mechanism and the object-oriented data model. Conclusions are presented in Section 4.

# 2 The Markov Model Mediator (MMM) Mechanism

The proposed *Markov Model Mediator (MMM)* mechanism adopts the *Markov Model* framework and the *mediator* concept. [9] defines a mediator to be a program that collects information from one or more sources, processes and combines it, and exports the resulting information. A Markov model is a well-researched mathematical construct that consists of a number of states connected by transitions. The states represent the alternatives of the stochastic process and the transitions contain probabilistic and other data used to determine the state that should be selected next. All transitions $S_i \rightarrow S_j$ such that $Pr(S_j \mid S_i) > 0$ are said to be allowed, the rest are prohibited.

There are two types of MMMs. Each database is modeled as a local MMM and each data warehouse is modeled as an integrated MMM. The compact notation $\lambda=(\mathcal{S}, \mathcal{F}, \mathcal{A}, \mathcal{B}, \Pi, \Psi)$, where $\mathcal{S}$ is a set of media objects called states, $\mathcal{F}$ is a set of attributes/features, $\mathcal{A}$ denotes the state transition probability distribution, $\mathcal{B}$ is the observation symbol probability distribution, $\Pi$ is the initial state probability distribution, and $\Psi$ is a set of augmented transition networks (ATNs), is adopted for the MMM mechanism. The formulations of $\mathcal{A}, \mathcal{B}$, and $\Pi$ for an MMM and the construction of the data warehouses for a network of databases are shown in [6, 7]. The augmented transition network (ATN) is a semantic model used to model multimedia presentations, multimedia database searching, and multimedia browsing. For the details of ATNs, please see [1, 2].

# 3 Comparative Analysis

The following subsections discuss some of the capabilities inherent in the MMM mechanism and the object-oriented data models when they are used to model the schema of a data warehouse in a multimedia information system.

## 3.1 Schema Independence Support

In object-oriented data models, user requests are method calls. Ideally, the application views would have a full set of generic methods that are implemented in terms of methods of objects across multiple database schemas. While these generic methods are needed, they are difficult to support because resources are often not described using the same object formalism. Hence, the semantics and implementation techniques for inter-model mappings are inevitably needed if an object-oriented data model is used to model the schema of a warehouse. Inter-model mappings involve schema translation or schema transformation, which introduce overhead to the warehouse. In addition to data, methods also need to be integrated. For example, the definition of methods on a warehouse schema should involve the reuse of behavior implemented in the member databases. Pre-existing method definitions at the member database level can be seen as sub-routines of methods at the warehouse level. Issues include dealing with equivalences and discrepancies between the methods defined in different member databases.

On the other hand, the proposed integrated MMMs support schema independence. The only information required from each member database is its media objects and their attributes/features. A local MMM is constructed for each member database and the semantic relationships among the media objects within the database are captured in that local MMM. An integrated MMM is then constructed from the local MMMs of the member databases. As long as each data model of a member database provides the defining classes (tables) and their corresponding sets of features (attributes), the semantic relationships can be maintained in an integrated MMM without mapping between schemas. In other words, schema autonomy can be preserved as much as possible and schema independence can be supported by using the MMM mechanism.

In order to construct the schema of a data warehouse and extract information from the warehouse, there is a need to conceptually merge databases. Here, the object-oriented data model and the MMM mechanism are compared when each of them is selected to model the schema of the warehouse, as shown in Figure 1. In Figure 1(a), the schema of the relational database needs to be translated into an object-oriented schema if an object-oriented data model is chosen as the schema of the warehouse. As depicted by the big dashed ellipse, it takes three steps to translate a relational schema into an object-oriented schema. The three steps are as follows [3]:

1. *Step 1:* relation ⇒ class
2. *Step 2:* foreign key ⇒ association attribute
3. *Step 3:* "is-a" relationship ⇒ inheritance

For details of the translation steps, please see [3]. After the schema translation process, the relational database can be viewed via an object-oriented schema. Since the other member database in this example is an object-oriented database, there is no need for further schema translation. Both the translated object-oriented schema and the object-oriented

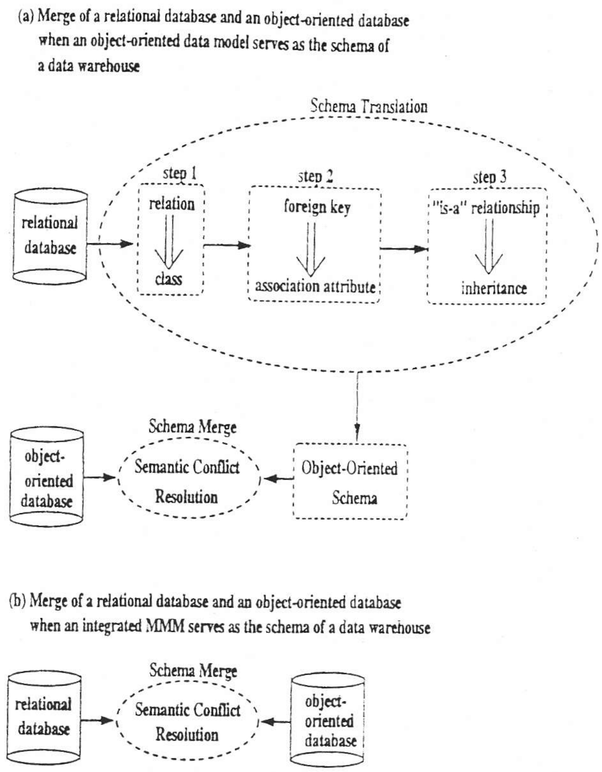


Figure 1: Conceptual merge of a relational database and an object-oriented database if (a) an object-oriented data model or (b) the MMM mechanism serves as the schema of the data warehouse.

member database are the inputs to the merge process, as shown in the small dashed ellipse in Figure 1(a). The merge process requires the resolution of semantic conflicts that occur among the member databases in order to extract information from the data warehouse.

On the other hand, the proposed MMM mechanism supports schema independence. In other words, there is no need to execute the schema translation process for the member databases. As shown in Figure 1(b), the relational database and the object-oriented database enter the merge process without undergoing any schema translation processes. The MMM mechanism still needs to resolve any semantic conflicts among the member databases in the merge process.

### 3.2 Database Autonomy Preservation

As mentioned earlier, an object-oriented data model is schema-dependent when it is used to model the schema of a data warehouse, so those member databases whose schemas are not object-oriented must have their schemas translated before the conceptual merge. To allow this, the member databases should provide some detailed information during the schema translation process, such as the functional dependency and inclusion dependence for a relational database. Moreover, the data in a member database will be manipulated in an object-oriented manner regardless of its original schema. Sometimes, data conversion is unavoidable.

On the other hand, database autonomy can be easily maintained by using the MMM mechanism. No member database needs to expose all of its information to its integrated schema; instead, only higher level information such as the media objects and their corresponding set of attributes/features need to be provided. Data in a member database can be manipulated via its own schema without any data conversion.

### 3.3 Multimedia Data Modeling

Most of the current object-oriented data models do not have the necessary functionality to support multimedia data. In current object-oriented data models, relationship semantics do not exist; only simple composition, which describes associations without temporal or spatial structures, exist. Spatial and/or temporal relationships can be expressed by augmenting standard relationships with additional methods and classes in object-oriented data models. The augmented classes and methods manipulate multimedia data within the database as typed objects. However, they generally provide a very limited set of options, so multimedia data cannot be easily queried, reused, or supported by the underlying *multimedia database management system (MDBMS)* without representable spatial and temporal structures. In addition, the current object-oriented data models tend to provide simple class structures rather than a complex multimedia framework.

On the other hand, our proposed MMM mechanism supports the spatial and/or temporal relationships, synchronization, and *Quality of Service (QoS)* controls necessary for multimedia presentations. In the MMM mechanism, the spatial/temporal relationships of the multimedia presentations are maintained by the ATNs associated with the media objects. The input of an ATN is a multimedia input string which can capture the spatial and temporal relationships of multimedia data and provide a means to query the multimedia data [1, 2].

# 4 Conclusions

In this paper, we compared the proposed MMM mechanism to the object-oriented data models for use as the schema of a data warehouse in a multimedia information system. The functionality of such a multimedia data warehouse considered here includes schema independence support, database autonomy preservation, and multimedia data modeling for the heterogeneous multimedia member databases. Our comparative analysis shows that when used to model a data warehouse schema, the proposed MMM mechanism provides more capabilities than the object-oriented data model for the functionality considered. Therefore, the MMM mechanism better serves as a model for a data warehouse schema in a multimedia information system.

## Acknowledgement

This research was supported in part by NASA (under grants NAGW-4080, NAG5-5095, NAS5-97222, and NAG5-6830) and NSF (CDA-9711582, IRI-9409661, HRD-9707076, and ANI-9876409).

## References

[1] S.-C. Chen and R.L. Kashyap, "Temporal and spatial semantic models for multimedia presentations," in *1997 International Symposium on Multimedia Information Processing*, pp. 441-446, Dec. 11-13, 1997.

[2] S.-C. Chen and R.L. Kashyap, "A spatio-temporal semantic model for multimedia presentations and multimedia database systems," accepted for publication in *IEEE Transactions on Knowledge and Data Engineering*, 1999.

[3] J. Fong, "Converting relational to object-oriented databases," *SIGMOD Record*, vol. 26, no. 1, March 1997.

[4] W.H. Inmon, *Building the data warehouse.* Wellesley, MA: QED Technical Publishing Group, 1992.

[5] A. Sen and V.S. Jacob, "Industrial-strength data warehousing," *Communication of the ACM*, Vol. 41, No. 9, September 1998.

[6] M.-L. Shyu, S.-C. Chen, and R. L. Kashyap, "Database Clustering and Data Warehousing," in *1998 ICS Workshop on Software Engineering and Database Systems*, pp. 30-37, Dec. 17-19, 1998.

[7] M.-L. Shyu and S.-C. Chen, "Probabilistic Networks for Data Warehouses and Multimedia Information Systems," Submitted to *IEEE Transactions on Knowledge and Data Engineering.*

[8] H.J. Watson and B.J. Haley, "Managerial considerations," *Communications of the ACM*, pp. 32-37, vol. 41, no. 9, September 1998.

[9] G. Wiederhold, "Mediators in the architecture of future information systems," *IEEE Computer*, pp. 38-49, March 1992.