Book Review

Title: Data Models: A Semantic Approach for Database Systems

Author: Sheldon A. Borkin

Publisher: The MIT Press, Cambridge, Massachusetts, 1980.

Since the publication of the ANSI/X3/SPARC Study Group's report on DBMS architecture, a great deal of research into the problem of mapping between data models has been done. This book, which is in effect the author's Ph.D. thesis, explores this problem *vis a vis* mapping between relation and graph data models. Only the definitional aspects of these mappings are presented. No implementation aspects are considered. Detailed algorithms for mapping between graph and relation data models are presented and proven correct. However, because of the complexity of the mapping problem, these algorithms apply only to restricted cases of relation data models.

Chapters 1, 2 and 3 introduce the notions of mapping between data models. The semantic graph and semantic relation data model are introduced because the usual graph and relation data models are mainly syntactic in nature and this is not adequate to describe meaningful mappings. Using semantic data models allows one to interpret what the data models represent in terms of real-world applications. It is shown that the semantic graph and semantic relation data models are related to the more commonly used DBTG-network and Codd's relation data models, respectively. Several levels of equivalence of data models are introduced.

Chapter 4 presents a framework for discussing data models. A data model is described in terms of applications, application models and application states. An application is that portion of the real world of interest to the system user. An application is described in terms of an application model and an application state. An application model specifies the definitional aspects of an application — the schema, consisting of structures and constraints, and the operations. An application state defines the extensional aspects (instances) of an application model (i.e., the data itself). Equivalences of data models are concerned with a data model's ability to express application models, their states, and transitions between application states.

Chapter 5 defines the semantic graph model in detail in terms of the framework presented in Chapter 4. Chapter 6 does the same for the semantic relation data model. Both data models are based on a simple entity-association view of applications. The definition of the two data models concerns how entities and associations are represented in each. The operations defined for the two data models concern the insertion and deletion of semantic units. (Note that retrieval is not treated here as it is not required to show data model equivalence).

Chapter 7 defines several types of equivalence between data models. These deal with instance, schema, and operation equivalence. State equivalence deals with showing that two instances of a database according to different data models represent the same state (i.e., the same data in terms of entities and associations). Application model equivalence deals basically with showing that two schemas represent the same application. Operation equivalence requires that operations can be found in the two data models that have identical effects given equivalent application models and states. Finally, data model equivalence results if all three of the preceding can be shown to hold for two data models.

Chapters 8 and 9 deal with defining a restricted form of the semantic relation data model. The restriction is in terms of additional constraints that need to be added to the semantic relation data model. The constraints are needed because it is shown that there are cases of semantic relation application data models that have no equivalent semantic graph application data model. Most of the two chapters consist of an algorithm for mapping a semantic graph data model to an equivalent restricted semantic relation data model.

Chapter 10 shows that if the algorithms outlined in Chapters 8 and 9 are used to map between semantic graph and semantic relation data models, then the two data models are equivalent.

In his conclusions, the author points to the complexity of the mapping problem as exemplified by the complexity of his algorithms. While the choice of data model significantly affects where the complexity resides, it does not affect the complexity itself since it comes from the application, not from the data model. In the semantic graph data model, the complexity is manifested implicitly in the structure of the graph. In the semantic relation data model, the complexity is explicit in the constraints required.

The approach to mapping taken here is to try to find *direct* mappings between data models. While this is certainly a valid approach, it is not the only one. Another approach would be to find an intermediate *mapping data model.* Such an approach would also reduce the number of mappings required. The reasons for choosing the approach outlined here and its merits over other possible approaches are not discussed.

The data models chosen for the mappings are not ones that normally are used in practice. While it may be true, as the author argues, that semantic data models are required for the mapping problem, the mappings presented here are not of much immediate use in practical situations. In particular, the operation mappings are presented at a very high level. While this makes the operation mapping easier, it ignores the problem of existing query languages and DML operations and mappings between them.

The research outlined in this book does show the way that research needs to go if the mapping problem is to be tackled. Data models and their properties must be formally defined, within some common framework, if the mapping problem is to be tackled systematically. Ad hoc mappings may be necessary in the short run, but formal mappings based on a sound theoretical basis are required in the long run. The research described in this book is a small start in that direction.

Fred H. Lochovsky
Computer Systems Research Group
University of Toronto
Toronto, Canada M5S 1A1