

Meta Data Management

Philip A. Bernstein
Microsoft Research
philbe@microsoft.com

Sergey Melnik
University of Leipzig
melnik@db.stanford.edu

By meta data management, we mean techniques for manipulating schemas and schema-like objects (such as interface definitions and web site maps) and mappings between them. Work on meta data problems goes back to at least the early 1970s, when data translation was the hot database research topic, even before relational databases caught on. Many popular research problems in the past five years are primarily meta data problems, such as data warehouse tools (e.g., ETL – to extract, transform and load), data integration, the semantic web, generation of XML or object-oriented wrappers for SQL databases, and generation of wrappers for web sites. Other classical meta data problems are information resource management, design tool support and integration, and schema evolution and data migration.

Despite its longevity and continued importance, there is no widely-accepted conceptual framework for the meta data field, as there is for many other database topics, such as access methods, query processing, and transaction management. In this seminar, we propose such a conceptual framework. It consists of three layers: applications, design patterns, and basic operators. Applications are the end-user problems to be solved, like those listed in the previous paragraph. Design patterns are generic problems that need to be solved in support of many different applications, such as meta modeling (for all meta data problems), answering queries using views (for data integration and the semantic web), and change propagation (for data translation, schema evolution, and round-trip engineering). Basic operators are procedures that are needed to support multiple design patterns and applications, such as matching schemas to produce a mapping, merging schemas based on a mapping, and composing mappings.

We will describe several meta data management problems, and for each, we will explain the design patterns and operators that are needed to solve it. We will summarize the main approaches to each design pattern and operator – the main choices of language, data structures, and algorithms – and will highlight the relevant papers that address it.

This seminar is targeted at both practicing engineers and researchers. The former will learn about the latest solutions to important meta data problems and the many dif-

ficult unsolved problems that are best to avoid. Database researchers, especially professors, will benefit from considering the conceptual framework that we propose, since no database textbooks treat meta data management as a separate topic as far as we know.

Speakers

Philip A. Bernstein is a researcher at Microsoft Corporation (<http://www.research.microsoft.com/~philbe>). Over the past 25 years, he has been a product architect and industrial researcher at Microsoft and at Digital Equipment Corp., a professor at Harvard University and Wang Institute of Graduate Studies, and a VP Software at Sequoia Systems. During that time, he has published over 100 articles on the theory and implementation of database systems, and coauthored three books. He was architect of Meta Data Services (formerly Microsoft Repository), which is the meta data manager in Microsoft SQL Server. Over the past three years, he has been developing a new approach to meta data management, called Model Management. He is an ACM Fellow, a recipient of the ACM SIGMOD Innovations Award, and a member of the National Academy of Engineering.

Sergey Melnik is a Ph.D. candidate in Computer Science at Leipzig University, Germany (<http://www-db.stanford.edu/~melnik/>). He serves as an invited expert in the RDF Core Working Group at the World-Wide Web Consortium. He spent three years as a visiting researcher in the Stanford Database Group where he worked on a variety of topics including meta data management, database optimization, information retrieval, and Semantic Web. He has published papers in EDBT, SIGMOD, TODS, TOIS, and WWW, and received the best student paper award at ICDE 2002 for his paper on schema matching. His Ph.D. thesis is on Generic Model Management, which includes an implementation of many of the scenarios and operators to be presented in this seminar. He will be joining Microsoft Research in late 2003.