ABSTRACT

VIEW EVOLUTION SUPPORT FOR INFORMATION INTEGRATION SYSTEMS OVER DYNAMIC DISTRIBUTED INFORMATION SPACES

by

Anisoara Nica

Chair: Elke A. Rundensteiner

Challenging issues for creating and maintaining tailored information gathering systems over large-scale information spaces (e.g., Digital Libraries, the World Wide Web) include the diversity of the information sources (ISs) in terms of their structures, query interfaces and search engines, as well as the dynamics of sources continuously being added, removed or upgraded. Current information integration systems are often based on static apriori defined views that gather information from heterogeneous information sources and provide the user with a uniform view of the information space for browsing and querying. This dissertation addresses one of the largely unexplored issues that such information integration systems raise, namely, the evolution and maintenance of data warehouses when the underlying information sources change their capabilities, i.e., schema level changes.

The overall solution approach that this dissertation puts forth consists of defining the problem of view evolution triggered by capability changes of ISs and designing evolution algorithms that achieve synchronization of the affected views in the presence of these types of changes. The contributions made by this dissertation include (1) an extension of the SQL view definition language that allows the user to apriori specify evolution preferences, e.g., whether dropping or changing a view component is acceptable; (2) a formal definition of what is a legal view rewriting under capability changes, i.e., semantics of view evolution; (3) algorithms for view synchronization that find a modified view definition in response to a capability change of an IS; (4) algorithms for the maintenance of materialized views after the view synchronization process; (5) experimental evaluations comparing the maintenance strategies after view synchronization with alternative maintenance techniques; and (6) the development of an working system incorporating some of the proposed evolution algorithms.