Schema Interpretation

Dr. Nayyar Masood

Department of Computer Sciences B.Z. University Multan

ABSTRACT

Interoperation between pre-existing, autonomous, heterogeneous, and possibly distributed databases has been an important research area since the beginning of last decade. The federated database systems (FDBS) [SL90] is one of the different architectures [LMR90, HM85, SM9I] proposed in this regard. The schema integration (SI) process by which a federated schema is derived is critical to the effectiveness of an FDBS. It requires an understanding of, and the ability to capture and reason with, the semantics of the CDBSs, but current data models do not represent schema semantics. In this paper I have discussed idea of schema interpretation, where by the semantics of component schemas is made explicit by developing a component concept model (CCM) for each of them. In a CCM, the schema elements of a component system are replaced by their corresponding concepts [YSDK9I], and the relationships between these schema elements are replaced by underlying semantic relationships [C86, WHC87, S93]. This helps in the identification of semantically similar schema elements, hence results in the greater automation of the SI process. The method has been developed in the framework of an FDBS architecture, in which the Object Model proposed by the Object Data Management Group (ODMG) [C96] is used as the canonical data model (CDM).

Semantic Aspects of Schema Integration

Schema integration is an essential and critical phase in the design of FDBSs [SL90]. The effectiveness of a system is dependent on synthesis of federated schemas which correctly represent and map into the relevant component schemas elements. The syntactic aspect to this problem concerns the design of federated schema elements, which simultaneously represent structures of corresponding component schema elements that they integrate. This form of reconciliation is straightforward once component schemas have been mapped into a CDM and correspondences between schema elements have been established. However, the semantic aspect, the identification of those correspondences, is problematic for following two reasons.