Subject Information Mediation for Integrated Access to Heterogeneous Collections

(Abstract)

Leonid A. Kalinichenko

Institute of Informatics Problems, Russian Academy of Science,

e-mail: leonidk@synth.ipi.ac.ru

A middleware layer formed by mediators providing a uniform query interface to multiple data sources is considered. The layer is intended to free users from having to locate the relevant collections, query each one in isolation, and combine manually the information extracted from them. Mediators provide for the integration of heterogeneous information sources. Well-known information mediation frameworks are briefly introduced as belonging to one of two main architectural classes of mediators known as Global as Views (GAV) and Local as Views (LAV). Basic mediators' features are surveyed.

Subject mediators are emphasized that support representation and access to specific subject domains. Infrastructure of the subject mediating environment aiming at semantic interoperability of heterogeneous collections of information is introduced. Subject mediator is consolidated by a respective community in a specific domain. Its definition includes ontological, structural and behavioral specifications. During operational phase mediator supports the process of systematic registration and classification of collections providing the uniform metainformation for integrated querying and access of heterogeneous sources. Advantages and disadvantages of subject mediation are briefly discussed.

Basic problems of mediators development are discussed mostly for the LAV mediator framework. Specifically, basic issues of query rewriting using views are presented. Fundamental problems of database querying, such as query containment, need to be solved. The state of the art is briefly surveyed including basic techniques for conjunctive queries containment test as well as their extensions for more comprehensive queries with negation, union, complex objects, semistructured data. Similarly the state of query rewriting algorithms for query languages with various capabilities are briefly surveyed.

The final part of the talk is given in context of a research project being developed at the Institute of Informatics Problems of the Russian Academy of Science. A canonical mediator's model and metainformation repository supporting the model are considered. Consolidation and operational phases of the subject mediator life cycle are discussed. Mediators' scalabiity measures are presented. Heterogeneous collection registration problems and respective issues of information contextualization are discussed. Registration of a collection at the mediator is considered as a process of compositional development in which specification of a collection is treated as a specification of requirements and the mediator's definition is treated as a specification of components (e.g., classes). Ontological integration leading to contextualization of the collection vocabularies and ontologies at the mediator is part of this process. Structure of the collection registration tool intended for the information collection providers support is presented.