Techniques for the Analysis of Conceptual Schemas

Silvana Castano[1,*], Valeria De Antonellis[2], Maria Grazia Fugini[3] and
Barbara Pernici[3]

[1]Università degli Studi di Milano, Department of Computer Science, Via Celoria, 18 - 20133 Milano, Italy
[2]Università degli Studi di Brescia, Italy
[3]Politecnico di Milano, Department of Electronics, Information and Bioengineering, Via Ponzio, 34 - 20133 Milano, Italy

Abstract

The problem of analyzing and classifying conceptual schemas is becoming more and more important due to the availability of large sets of schemas from existing applications.The purpose of analysis and classification activities can be that of extracting information on schemas of legacy systems in order to migrate them to new architectures, to build libraries of reference conceptual components to be used in building new applications in a given domain, to analyze large sets of schemas in an organization to identify information flows and possible replication of data. The paper proposes a set of techniques to be adopted for schema classification and analysis: indexing techniques to associate a description with a schema, techniques for abstracting reference conceptual schemas based on schema clustering, and techniques for schema comparison. The application of these techniques in the context of reuse of conceptual components is briefly presented.

We started from the observation that a huge number of database conceptual schemas were accumulated over years of design activities and that schema analysis activities, usually performed manually by a schema analyst providing additional information about the contents of the schema at hand, demanded new techniques for automating as much as possible the analysis process to be performed on the large scale. The paper proposed a set of methodological instruments and associated techniques, to be used separately or in combination, to support a number of schema analysis activities, which are: schema indexing, to associate descriptors and keywords/features with schemas; schema matching, to evaluate schema similarity; schema abstraction, to derive abstract representation of schema contents. Moreover, we proposed a methodology for systematic schema analysis with the purpose of identifying and abstracting the similar of a set of schemas into reference components, to be (re)used in database and ontology design.

The techniques and methodology have been subsequently experimented to analyze a large inventory of ER schemas of Information Systems in the Italian Public Administration [1], and the final and consolidated work on schema analysis techniques has been published

on the ACM Transactions on Database Systems [2].

Since that time, the idea of having a set of techniques for schema and data analysis on large scale has dramatically gained importance and it has been the starting base for the development of a new generation of data analysis and data science techniques dealing with huge datasets of unstructured and textual data for effective data sharing and exploratory analysis on the global scale in the era of big data and cloud computing.

References

[1] C. Batini, S. Castano, V. D. Antonellis, M. G. Fugini, B. Pernici, Analysis of an inventory of information systems in the public administration, Requirements Engineering 1(1) (1996) 47–62.

[2] S. Castano, V. D. Antonellis, M. G. Fugini, B. Pernici, Conceptual schema analysis: Techniques and applications, ACM Transactions on Database Systems 23(3) (1998) 286–332.