

io-port 05841630

Romero, Oscar; Abelló, Alberto

A framework for multidimensional design of data warehouses from ontologies.

Data Knowl. Eng. 69, No. 11, 1138-1157 (2010).

Summary: The data warehouse design task needs to consider both the end-user requirements and the organization data sources. For this reason, the data warehouse design has been traditionally considered a reengineering process, guided by requirements, from the data sources. Most current design methods available demand highly-expressive end-user requirements as input, in order to carry out the exploration and analysis of the data sources. However, the task to elicit the end-user information requirements might result in a thorough task. Importantly, in the data warehousing context, the analysis capabilities of the target data warehouse depend on what kind of data is available in the data sources. Thus, in those scenarios where the analysis capabilities of the data sources are not (fully) known, it is possible to help the data warehouse designer to identify and elicit unknown analysis capabilities. In this paper we introduce a user-centered approach to support the end-user requirements elicitation and the data warehouse multidimensional design tasks. Our proposal is based on a reengineering process that derives the multidimensional schema from a conceptual formalization of the domain. It starts by fully analyzing the data sources to identify, without considering requirements yet, the multidimensional knowledge they capture (i.e., data likely to be analyzed from a multidimensional point of view). Next, we propose to exploit this knowledge in order to support the requirements elicitation task. In this way, we are already conciliating requirements with the data sources, and we are able to fully exploit the analysis capabilities of the sources. Once requirements are clear, we automatically create the data warehouse conceptual schema according to the multidimensional knowledge extracted from the sources.

Keywords: OLAP; multidimensional design; ontologies

doi:10.1016/j.datak.2010.07.007