

ABSTRACT

Doctoral dissertation

Eliminating redundancy and duplicates in XML data

This thesis is dedicated to the analysis of occurrence and solving the issues that result from the presence of redundancy and duplicates in databases, in particular XML databases. The paper presents general methods useful for detecting redundancy and duplicates.

The main thesis of the paper is to show that improving the quality of both schema and status of the XML database can be obtained by detecting and eliminating redundancy and duplicates occurring in it. Methods of detecting redundancy and duplicates developed for the needs of relational databases can be effectively generalized to the case of XML databases. The proposed method of normalization to the XML normal form allows transformations, as a result of which redundancy is eliminated, while both data and functional dependencies are retained. The proposed method of detecting and eliminating duplicates in XML data does not cause the loss of the normal form if it was in the form of normal XML before normalization.

Important theoretical results presented in the paper relate to the methods of elimination of redundancy in the XML data through their standardization. The most important practical results concern the elimination of duplicates in XML data, where the methods of elimination of duplicates in XML databases are discussed, based on a review of selected solutions and own propositions.

A handwritten signature in blue ink, appearing to read 'TOMRIZ' followed by a stylized flourish.