

Early Integration Testing for Entity Reconciliation in the Context of Heterogeneous Data Sources

Raquel Blanco¹, José G. Enríquez², Francisco J. Domínguez-Mayo², M. J. Escalona²
and Javier Tuya¹

¹Departamento de Informática, Universidad de Oviedo

²Departamento de Lenguajes y Sistemas Informáticos, Universidad de Sevilla
{rblanco, tuya}@uniovi.es, jose.gonzalez@iwt2.org, {fjdominguez, mjescalona}@us.es

Palabras Clave: Early testing, entity reconciliation, heterogeneous data sources, model-driven engineering, software testing, specification-based testing

Lugar de publicación: IEEE Transactions on Reliability, 67(2), pp. 538-556

Impact factor: 2,729

DOI: <https://doi.org/10.1109/TR.2018.2809866>

Resumen(Abstract). Entity reconciliation (ER) aims to combine data from different sources for a unified vision. The management of large volumes of data has given rise to significant challenges to the ER problem due to facts such as data becoming more unstructured, unclean, and incomplete or the existence of many datasets that store information about the same topic. Testing the applications that implement the ER problem is crucial to ensure both the correctness of the reconciliation process and the quality of the reconciled data. This paper presents an approach based on model-driven engineering that allows the creation of test models for the early integration testing of ER applications, contributing in three main aspects: the description of the elements of the proposed framework, the definition of the testing model, and the validation of the proposal through two real-world case studies. This validation verifies that the early integration testing of the ER application is capable of detecting a series of deficiencies, which a priori are not known and that will help to improve the final result that the ER application offers.