

Finding, Managing and Enforcing CFDs and ARs via a Semi-Automatic Learning Strategy

Katalin Tünde Jánosi-Rancz

Sapientia Hungarian University of Transylvania, Tirgu Mures, Romania

tsuto@ms.sapientia.ro

This paper describes our strategy, which finds Conditional Functional Dependencies (CFDs) and Association Rules (ARs), and instead of using them to clean dirty data we use them to prevent their appearance in the database. We achieve this by differentiating permanent CFDs/ARs from temporary CFDs/ARs. If we know about a CFD/AR that it will be valid in the future, we can rely on them by creating constraints which guarantee that the CFD-rule will not be breached by insertions or modifications. Along with complete management of CFDs/ARs our implemented application called DependencyManager also uses Formal Concept Analysis (FCA) methods to analyze the permanent CFDs/ARs and draw useful conclusions, helping the users of the application to prevent inconsistencies, fix bugs and optimize their queries and applications by providing a lattice of CFDs/ARs, using usefulness as the relation. We consider a CFD/AR to be more useful than the other if it needs less information to determine more information.

References

- [1] G. Cong, W. Fan, F. Geerts, X. Jia, and S. Ma. Improving data quality: Consistency and accuracy. In C. Koch, J. Gehrke, M. N. Garofalakis, D. Srivastava, K. Aberer, A. Deshpande, D. Florescu, C. Y. Chan, V. Ganti, C.-C. Kanne, W. Klas, and E. J. Neuhold, editors, *VLDB*, pages 315–326. ACM, 2007. ISBN 978-1-59593-649-3.
- [2] G. Cormode, L. Golab, F. Korn, A. McGregor, D. Srivastava, and X. Zhang. Estimating the confidence of conditional functional dependencies. In U. Çetintemel, S. B. Zdonik, D. Kossmann, and N. Tatbul, editors, *SIGMOD Conference*, pages 469–482. ACM, 2009. ISBN 978-1-60558-551-2.
- [3] W. Fan, F. Geerts, X. Jia, and A. Kementsietsidis. Conditional functional dependencies for capturing data inconsistencies. *ACM Trans. Database Syst.*, 33 (2), 2008.