IIM> Vol.3 No.6, November 2011

OPEN ⊙ ACCESS

A Rule Management System for Knowledge Based Data Cleaning

Author(s)
Louardi BRADJI, Mahmoud BOUFAIDA

ABSTRACT
In this paper, we propose a rule management system for data cleaning that is based on knowledge. This system combines features of both rule based systems and rule based data cleaning frameworks. The important advantages of our system are threefold. First, it aims at proposing a strong and unified rule form based on first order structure that permits the representation and management of all the types of rules and their quality via some characteristics. Second, it leads to increase the quality of rules which conditions the quality of data cleaning. Third, it uses an appropriate knowledge acquisition process, which is the weakest task in the current rule and knowledge based systems. As several research works have shown that data cleaning is rather driven by domain knowledge than by data, we have identified and analyzed the properties that distinguish knowledge and rules from data for better determining the most components of the proposed system. In order to illustrate our system, we also present a first experiment with a case study at health sector where we demonstrate how the system is useful for the improvement of data quality. The autonomy, extensibility and platform-independency of the proposed rule management system facilitate its incorporation in any system that is interested in data quality management.

KEYWORDS
Rule, Data Quality, Data Cleaning, Knowledge, Rule Management System, Rule Based System, Structure

Cite this paper
L. BRADJI and M. BOUFAIDA, "A Rule Management System for Knowledge Based Data Cleaning," *Intelligent Information Management*, Vol. 3 No. 6, 2011, pp. 230-239. doi: 10.4236/iim.2011.36028.

References
[1]   C. White, " Developing a Universal Approach to Cleansing Customer and Product Data," BI Research, 2008.

[2]   E. Friedman-Hill, " Jess in Action, Rule Based Systems in Java," Meanning Publications CO., 2003.

[3]   R. Whiting, " Ham-strung by Defective Data," InformationWeek, 2006.

[4]   A. Vavouras, " A Metadata-Driven Approach for Data Warehouse Refreshment," Ph.D. Thesis, Der Universit?t Zürich, 2002.

[5]   K. Duncan and D. Wells, " Rule Based Data Cleansing for Data Warehousing," San Diego, 2000.

[6]   J. L. Y. Koh, " Correlation-Based Methods for Biological Data Cleaning," Ph.D. Thesis, Singapore University, 2007.

[7]   T. T. P. Thi and M. Helfert, " Discovering Dynamic Integrity Rules with a Rules-Based Tool for Data Quality Analyzing," Proceedings of the 11th International Conference on Computer Systems and Technologies, Sofia, 2010, pp. 89-94.

[8]   L. Bradji and M. Boufaida, " Knowledge Based Data Cleaning for Data Ware-house Quality," Proceeding of ICDIPC 2011, Part 2, CCIS, Springer-Verlag Berlin Heidelberg, Vol. 189, 2011, pp. 373-384.

[9]   D. Batra and N. A. Wishart, " Comparing a Rule-Based Approach with a Pattern-Based Approach at

Dif-ferent Levels of Complexity of Conceptual Data Modelling Tasks," International Journal of Human-Computer Studies, Vol. 61, No. 4, 2004, pp. 397-419. doi:10.1016/j.ijhcs.2003.12.019

[10] S. Antony and R. San-thanam, " Could the Use of a Knowledge-Based System Lead to Implicit Learning?" Journal of Decision Support System, Vol. 43, 2007, pp. 141-151. doi:10.1016/j.dss.2006.08.004

[11] M. Hentea, " Intelligent System for Information Security Manage-ment: Architecture and Design Issues," Journal of Issues in Informing Science and Information Technology, Vol. 4, 2007, pp. 29-43.

[12] M. C. Lai, H. C. Huang and W. K. Wang, " Designing a Knowledge-Based System for Benchmarking: A DEA Approach," Journal of Knowledge-Based Systems, Vol. 24, 2011, pp. 662-671. doi:10.1016/j.knosys.2011.02.006

[13] A. Abraham, " Rule-Based Expert Systems," Handbook of Measuring System Design, John Wiley & Sons Ltd, 2005, pp. 909-919.

[14] G. Dondossola, " Formal Methods in the Development of Safety Critical Knowledge-Based Components," Proceedings of the KR' 98 European Workshop on Validation and Verification of KBS, 1998, pp. 232-237.

[15] A. Onisko, P. Lucas and M. J. Druzdzel, " Comparison of Rule-Based and Bayesian Network Approaches in Medical Diagnostic Systems," Proceeding of AIME 2001, LNAI, Springer-Verlag Berlin Heidelberg, 2001, pp. 283-292.

[16] M. S. Abdullah, C. Kimble, I. Benest and R. Paige, " Kn- owledge-Based Systems: A Re-Evaluation," Journal of K- nowledge Management, Vol. 10, No. 3, 2006, pp. 127-142. doi:10.1108/13673270610670902

[17] G. J. Nalepa and A. Ligeza, " Prolog-Based Analysis of Tabular Rule-Based Systems with the XTT Approach," Proceedings of the Nineteenth International Florida Artificial Intelligence Research Society Conference, 2006, pp. 426-431.

[18] E. Liorens, J. Comas, E. Marti, J. L. Riera, F. Sabater and M. Poch, " Integrating Empirical and Heuristic Knowle- dge in a KBS to Approach Stream Eutrophication," Journal of Ecological Modelling, Vol. 220, 2009, pp. 2162-2172. doi:10.1016/j.ecolmodel.2009.06.012

[19] S. M. M. Soe and M. P. Zaw, " Design and Implementation of Rule-Based Expert System for Fault Management," Journal of World Academy of Science, Engineering and Technology, Vol. 48, 2008, pp. 34-39.

[20] C. Angeli, " Diagnostic Expert Systems: From Expert' s Knowledge to Real-Time Systems," TMRF e-Book Advanced Knowledge Based Systems: Model, Applications & Research, Sajja & Akerkar, Eds., Vol. 1, 2010, pp. 50-73.

[21] G. J. Nalepa and A. Ligeza, " The HEKATE Methodology Hybrid Engineering of Intelligent Systems," International Journal of Appllied Mathematics and Computer Science, Vol. 20, No. 1, 2010, pp. 35-53. doi:10.2478/v10006-010-0003-9

[22] A. Paschke and A. Kozlenkov, " A Rule-Based Middleware for Business Process Execution," Proceedings of Multikonferenz Wirtschaftsinfor-matik, MKWI 2008, Mun- chen, GITO-Verlag, Berlin, 2008, pp. 1409-1420.

[23] M. Ribari?, D. Ga?evi? and M. Milanovi?, " A Rule-Based Approach to Modeling of Semantically-Enriched Web Services," Web4Web Workshop 2008, International work-shop on Semantic Web Technologies, Belgrade, 2008.

[24] F. Tip, " A Survey of Program Slicing Techniques," Jour- nal of Programming Languages, Vol. 3, 1995, pp. 121-189.

[25] M. L. Lee, T. W. Ling and W. L. Low, " IntelliClean: A Knowl-edge-Based Intelligent Data Cleaner," Proceedings of the sixth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2000, pp. 290-294. doi:10.1145/347090.347154

[26] J. Rodic and M. Baranovic, " Generating Data Quality Ru- les and Integration into ETL Proc-ess," Proceeding of DO- LAP' 09, Hong Kong, 2009, pp. 65-72.

[27] J. Liu, Z, Lu, L. M. López and Y. Xu, " Preference Crite- rion and Consistency in the Rule-Based System Based on a Lattice-Valued Logic," Proceeding of Eurofuse Work- shop New Trends in Preference Modelling, 2006, pp. 99-104.

[28] Y. Xu, J. Liu, D. Ruan and T. T. Lee, " On the Consis-tency of Rule Bases Based on Lattice-Valued First-Order Logic LF(X)," International Journal of Intelligent Sys- tems, Vol. 21, No. 4, 2006, pp. 399-424. doi:10.1002/int.20129

[29] S. Hedman," A First Course in Logic: An Introduction to Model Theory, Proof Theory, Computability and Complexity," Oxford University Press Inc., Edition, 2006.

[30] A. Ligeza and G. J. Nalepa, " A Study of Methodological Issues in Design and Development of Rule-Based Systems: Proposal of a New Approach," Wiley Interdisciplinary Reviews: Data Mining and