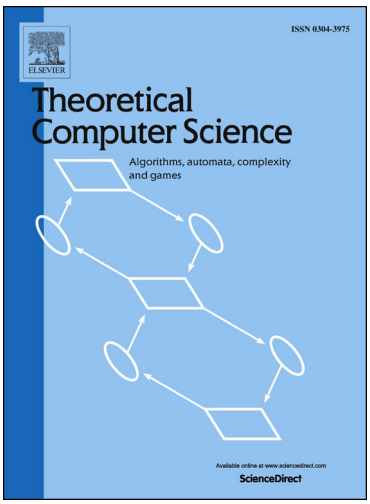
Accepted Manuscript



RQL: A query language for rule discovery in databases

Brice Chardin, Emmanuel Coquery, Marie Pailloux, Jean-Marc Petit

PII: S0304-3975(16)30594-1
DOI: http://dx.doi.org/10.1016/j.tcs.2016.11.004
Reference: TCS 10964

To appear in: *Theoretical Computer Science*

Received date: 18 March 2015
Revised date: 19 July 2016
Accepted date: 4 November 2016

Please cite this article in press as: B. Chardin et al., RQL: A query language for rule discovery in databases, *Theoret. Comput. Sci.* (2016), http://dx.doi.org/10.1016/j.tcs.2016.11.004

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

ACCEPTED MANUSCRIPT

RQL: A Query Language for Rule Discovery in Databases

Brice Chardin[1], Emmanuel Coquery[2], Marie Pailloux[3], Jean-Marc Petit[4]

Abstract

Promoting declarative approaches in data mining is a long standing theme, the main idea being to simplify as much as possible the way data analysts interact with their data. This paper goes into this direction by proposing a well-founded logical query language, $SafeRL$, allowing the expression of a wide variety of rules to be discovered against a database. By rules, we mean statements of the form "if ... then ...", as defined in logics for "implications" between boolean variables. As a consequence, $SafeRL$ extends and generalizes functional dependencies to new and unexpected rules. We provide a query rewriting technique and a constructive proof of the main query equivalence theorem, leading to an efficient query processing technique. From $SafeRL$, we have devised RQL, a user-friendly SQL-like query language. We have shown how a tight integration can be performed on top of any relational database management system. Every RQL query turns out to be seen as a query processing problem, instead of a particular rule mining problem. This approach has been implemented and experimented on sensor network data. A web prototype has been released and is freely available (`http://rql.insa-lyon.fr`). Data analysts can upload a sample of their data, write their own RQL queries and get answers to know whether or not a rule holds (if not, a counter example from the database is displayed) and much more.

keywords Query Languages, Formal Concept Analysis, Implications, Functional dependencies, Query Optimization, Relational Calculus

1. Introduction

The relational database management systems (DBMS) market is already huge and continues to grow since it is expected to nearly double by 2016 [41]. As a trivial consequence for the data mining community, it makes sense – more

[1]LIAS, ISAE-ENSMA, France
[2]Department of Computer Science, University Lyon 1, CNRS, France
[3]Department of Computer Science, University Clermont-Ferrand 2, CNRS, France
[4]Université de Lyon, CNRS, INSA-Lyon, LIRIS, France

ACCEPTED MANUSCRIPT

Download English Version:

https://daneshyari.com/en/article/4952380

Download Persian Version:

https://daneshyari.com/article/4952380

Daneshyari.com