

论文

Azszp: 一种基于规则的数据清洗方案

李俊奎, 王元珍, 李 专

华中科技大学计算机科学与技术学院, 湖北 武汉 430074

摘要:

数据清洗是提高数据集成质量的一个重要手段. 提出了一种基于动态规则的数据清洗方案AzszpClean, 这种方法对各种清洗规则进行动态编译, 将数据转换和数据清洗两者结合起来, 强化清洗的描述能力. 实际应用表明, AzszpClean方法可以完成硬编码的功能, 具有更高的实现效率.

关键词: 数据清洗 动态规则编译 描述能力

AzszpClean: a rule-based solution to data cleaning

LI Jun-kui,WANG Yuan-zhen,LI Zhuan

College of Computer Science & Technology, Huazhong University of Science & Technology, Wuhan 430074, Hubei, China

Abstract:

Data cleaning is one of the important methods to improve the data quality in data integration. An approach to data cleaning called AzszpClean was proposed, which combines data transformation and filtering, and dynamically compiles the cleaning rules. The experimental results show that AzszpClean outperforms hard coding with more efficiency of implementation.

Keywords: data cleaning dynamic rule compiling descriptive power

收稿日期 1900-01-01 修回日期 1900-01-01 网络版发布日期 2006-10-24

DOI:

基金项目:

通讯作者: 李俊奎

作者简介:

本刊中的类似文章

1. 李人和, 宫学庆, 常建龙, 周游弋, 周红福, 周傲英. RealMon: 处理低质量SNMP数据流的实时监测系统[J]. 山东大学学报(理学版), 2007,42(11): 32-36

扩展功能

本文信息

Supporting info

PDF(298KB)

[HTML全文](0KB)

参考文献[PDF]

参考文献

服务与反馈

把本文推荐给朋友

加入我的书架

加入引用管理器

引用本文

Email Alert

文章反馈

浏览反馈信息

本文关键词相关文章

▶ 数据清洗

▶ 动态规则编译

▶ 描述能力

本文作者相关文章

▶ 李俊奎

▶ 王元珍

▶ 李 专