



ISSN 1000-1239

CODEN JYYFEY

Journal of Computer Research and Development

计算机研究与发展

第50卷增刊I 2013年8月

Vol.50 Suppl. I Aug. 2013

主办 中国科学院计算技术研究所 中国计算机学会 出版 科学出版社

中国计算机学会会刊

万方数据

计算机研究与发展

(Jisuanji Yanjiu Yu Fazhan)

第 50 卷 增刊 2013 年 8 月

目次

Web 与半结构化数据管理

XML 查询中具有位置谓词的树模式匹配方法	李发金 廖湖声 高红雨 (1)
面向 XQuery 语言的 XML 树模式识别方法	陈君鹏 廖湖声 苏航 高红雨 (12)
一种基于 HBase 的 RDF 数据存储模型	朱敏 程佳 柏文阳 (23)
面向 Twig 查询的 XML 类型推导	罗卿 廖湖声 杨红丽 (32)
一种基于 BPMN 的业务流程图到 BPEL 的映射方法	冯晓宁 李麒麟 王卓 (44)
含有丰富结构化数据的 Web 页面分类技术的研究	赵志滨 贾岩峰 姚兰 鲍玉斌 (53)
递归树模式查询	官肇宁 廖湖声 高红雨 (61)

查询处理与优化

查询性能预测方法的性能评测研究	徐赢 刘屹 阴红志 崔斌 (70)
一个高效的连续 k 近邻查询改进算法	孙圣力 林硕 (80)
基于图的音乐数据查询处理及优化方法	欧晓平 王朝坤 汪浩 陈俊 白易元 (90)
一种基于压缩策略的高维空间子空间 skyline 查询算法	孟熠 刘玉葆 李启睿 (101)
含排他性选择规则的主动规则集的汇流性判定方法	熊中敏 赵梦露 周雪楠 黄冬梅 (109)

近似和非确定性数据库

Attribute-or 模型下不确定关系的无损分解算法	叶杰敏 刘国华 袁慧 石丹妮 吴云龙 费凡 (117)
面向属性级不确定数据的 U-Top k 查询优化算法的研究	于洋 赵志滨 鲍玉斌 于戈 (125)
科学工作流中面向不确定数据源图的受限可达查询	胡海洋 刘占晨 胡华 (133)

空间和时态数据库系统

无线数据广播环境下的空间关键字查询	陈翀 陈楚南 孙未未 (145)
基于结构-属性的时空对象图聚类算法的研究	韩启龙 赵洪斌 潘海为 印桂生 常吉羽 (154)
DKR-Tree: 一种支持动态关键字的空间对象索引树	戴健 许佳捷 刘奎恩 武斌 丁治明 (163)
面向实际道路网络模型的空间资源 Top- k 查询算法	马武彬 刘明星 黄宏斌 邓苏 (171)

数据挖掘和知识发现

基于主题模型的微博重要话题发现与排序方法	姜晓伟 王建民 丁贵广 (179)
满足均匀分布的不确定数据关联规则挖掘算法	陈爱东 刘国华 费凡 周宇 万小妹 袁慧 (186)
基于 ELM 的室性早搏检测算法	王之琮 刘红艳 肖静 于戈 康雁 (196)
一种基于 RDB 中自身连接的 Web 服务聚类方法	刘建晓 王健 张秀伟 刘峰 李小霞 (205)
PEIF: 基于并行机群的大数据实体识别算法	李明达 王宏志 张佳程 李建中 高宏 (211)

数据质量管理与隐私安全

基于数据质量规则的缺失结果解释约减	张奥千 宋韶旭 王建民 (221)
-------------------------	-------------------

数据完整性的评估方法	刘永楠	邹兆年	李建中	王海洁	(230)
高效的实体匹配结果消解算法	刘显敏	李建中			(239)
基于事务型 k -anonymity 的动态集值属性数据重发布隐私保护方法	武毅	王丹	蒋宗礼		(248)

信息检索与数据库

利用术语本体关系扩展 SBN 检索模型	陈振亚	徐建民	吴树芳		(257)
一种基于曝光量和点击率的用户组优化策略	郭心语	何晓丰	官学庆	张蓉	周傲英 (264)
基于相关查询的关键词搜索优化技术	冯丽敏	杨艳	钟颖莉		(272)
基于属性权重的实体解析技术	甄灵敏	杨晓春	王斌	Ahmed A Hussein	(281)
基于 M-tree 的不等半径覆盖相异多样集求解方法	刘明星	马武彬	邓苏	黄宏斌	(290)

云计算与物联网数据管理

面向物联网传感器信息的数据分配策略	李昕龙	张玲	李坤	王永炎	(297)
基于 MapReduce 的连接聚集查询算法研究	陈勇旭	陈梦杰	刘雪冰	宋杰	(306)
基于 Map-Reduce 的大数据缺失值填充算法	金连	王宏志	黄沈滨	高宏	(312)
基于块级的分级存储数据特征模型及其应用研究	施光源	王恩东	张宇		(322)
共享的 MapReduce 环境下批量作业的调度算法研究	王习特	申德荣	聂铁铮	寇月	于戈 (332)
云计算系统中能量有效的数据摆放算法和节点调度策略	肖艳文	王金宝	李亚平	高宏	(342)
ASIC:一种适用于云数据管理的自适应辅助索引机制	牟雁超	苏汉宸	程序	李红燕	王腾蛟 (352)
一种 Hadoop 平台下的调度算法及混合调度策略	李千目	张晟骁	陆路	戚湧	张宏 (361)
一种基于动态划分的 MapReduce 负载均衡方法	周家帅	王琦	高军		(369)

系统演示

HaoLap:基于 Hadoop 的海量数据 OLAP 系统	郭朝鹏	王智	韩峰	张一川	宋杰 (378)
基于 D-Ocean 的金融风险监控系统	冯普超	邵春翡	黄忠东	孙建伶	(384)
SmartMS:Android 平台下灵活的海监执法取证及分析系统	黄冬梅	张喆昱	赵丹枫	王建	(389)
TPCHSuite:一个 TPC-H 自动化测试工具的设计与实现	康强强	江舟	金澈清	周傲英	(394)
MMDVis:一个基于微博用户的多博文传播分析及可视化系统	郝竞超	王朝坤	司德潭	汪浩	王巍 王建民 (399)
基于众包的电子商务数据实体分类系统	叶晨	王宏志	周小田	李建中	高宏 (405)
生物实验信息管理系统	王黎维	付祖发	刘东明	王梁	李佳瑾 彭智勇 (410)
一种分布式系统上的元数据管理系统	魏光辉	李杰斌	王程玉	何震瀛	汪卫 (416)
XML 数据上支持查询扩展的关键词检索系统	胡昊	王君伟	常橙	何震瀛	汪卫 (421)
基于 R k NN 的村镇基础设施选址分析系统	许景科	于戈	孙焕良	刘小溪	(426)
CloudIndexEval:面向云平台上多维索引的评测系统	周新	王延昊	刘辰	宋少华	张孝 王珊 (431)
TPCDroid:面向移动智能终端的数据库查询能耗测试	杨濮源	金培权	岳丽华		(435)

JOURNAL OF COMPUTER RESEARCH AND DEVELOPMENT

Vol. 50 Supplement August 2013

CONTENTS

Web and Semi-Structured Data Management

- Twig Pattern Matching with Positional Predicates in XML Queries *Li Fajin et al.* (1)
- An Identification Algorithm of XML Tree Pattern for XQuery Language *Chen Junpeng et al.* (12)
- A Storage Model for RDF Data Based on HBase *Zhu Min et al.* (23)
- Type Inference on XML Twig Query *Luo Qing et al.* (32)
- A Mapping Method from BPMN Models to BPEL *Feng Xiaoning et al.* (44)
- Research on Classification of Web Pages with Rich Structured Data *Zhao Zhibin et al.* (53)
- Recursive Twig Pattern Query *Guan Zhaoning et al.* (61)

Query Processing and Optimization

- An Empirical Study of the Performance Evaluation of Query Performance Predictors ... *Xu Ying et al.* (70)
- A Improved Algorithm For Efficient Continuous KNN Queries *Sun Shengli et al.* (80)
- An Approach to Music Data Query Processing and Optimization Based on Graph ... *Ou Xiaoping et al.* (90)
- A Compression Method Based Algorithm for Subspace Skyline Computation in High Dimensional Space
..... *Meng Yi et al.* (101)
- A Confluence Decison Method for an Active Rule Set with Exclusive Rules ... *Xiong Zhongmin et al.* (109)

Approximate and Uncertain Database

- Lossless Decomposition of Uncertain Relation in Attribute-or Model *Ye Jiemin et al.* (117)
- Optimizing U-Top k Queries on Attribute-Level Uncertain Data *Yu Yang et al.* (125)
- Answering Label-Constraint Reachability Queries over Uncertain Data Provenance in Scientific
Workflows *Hu Haiyang et al.* (133)

Spatial and Temporal Database

- Spatial Keyword Queries in Wireless Broadcast Environment *Chen Chong et al.* (145)
- Research on Spatio-Temporal Object Graph Clustering Algorithm Based on Structure and Attribute
..... *Han Qilong et al.* (154)
- DKR-Tree: A Dynamic-Keyword-R Tree *Dai Jian et al.* (163)
- Spatial Resource Top- k Query Algorithm Facing Real Road Network Model *Ma Wubin et al.* (171)

Data Mining and Knowledge Discovery

- Detection and Ranking of Significant Topics on Sina Weibo Based on Topic Model
..... *Jiang Xiaowei et al.* (179)
- Association Rules Mining Algorithm for Uncertain Data with the Uniform Distribution
..... *Chen Aidong et al.* (186)
- The ELM-based Algorithm on the Detection of Premature Ventricular Contraction
..... *Wang Zhiqiong et al.* (196)
- A Web Service Clustering Method Based on Self-Join in RDB *Liu Jianxiao et al.* (205)
- PEIF: Parallel Entity Resolution on Big Data *Li Mingda et al.* (211)

Data Quality Managment and Privacy Security

- Reducing Explanations of Non-Answers Using Data Quality Rules *Zhang Aoqian et al.* (221)

Evaluation of Data Completeness *Liu Yongnan et al.* (230)

Efficient Algorithms for Resolving Conflicts in Entity Matching Results *Liu Xianmin et al.* (239)

Privacy Preserving in Re-Publication of Dynamic Set-Valued Data Based on Transactional *k*-Anonymity
..... *Wu Yi et al.* (248)

Information Retrieval and Database

Using Ontology Relationships Among Terms to Extend SBN Retrieval Model *Chen Zhenya et al.* (257)

An Optimization Strategy Based on Impression and Click-Through Rate for User Segments
..... *Guo Xinyu et al.* (264)

Relevant-Query Based Keyword Search Optimization *Feng Limin et al.* (272)

An Entity Resolution Approach Based on Attributes Weights *Zhen Lingmin et al.* (281)

Dissimilarity and Coverage Diversification with Different Radius Based on M-tree
..... *Liu Mingxing et al.* (290)

Cloud and IoT Data Management

A Data Allocation Strategy for Sensor Information of Internet of Things *Li Xinlong et al.* (297)

MapReduce Based Aggregate-Join Query Algorithms *Chen Yongxu et al.* (306)

Missing Value Imputation in Big Data Based on Map-Reduce *Jin Lian et al.* (312)

Research on the Tier Model and Application of Extent-Based Data Feature *Shi Guangyuan et al.* (322)

Batch-Job Scheduling in Shared MapReduce Environment *Wang Xite et al.* (332)

Energy Efficient Data Placement Algorithm and Node-Scheduling Policies in Cloud Computing Systems
..... *Xiao Yanwen et al.* (342)

ASIC: An Adaptive Secondary Index for Data Management in Cloud Computing Environment
..... *Mou Yanchao et al.* (352)

A Job Scheduling Algorithm and Hybrid Scheduling Method on Hadoop Platform ... *Li Qianmu et al.* (361)

An Approach for Load Balancing in MapReduce via Dynamic Partitioning *Zhou Jiashuai et al.* (369)

Demo System

HaoLap: An Hadoop Based OLAP System for Massive Data *Guo Chaopeng et al.* (378)

Financial Risk Monitoring System Based on D-Ocean *Feng Puchao et al.* (384)

SmartMS: A Flexible Android-Based Marine Surveillance Evidence Collection and Analysis System
..... *Huang Dongmei et al.* (389)

TPCHSuite: The Design and Implementation of an Automatic Test Tool for TPC-H
..... *Kang Qiangqiang et al.* (394)

MMDVis: A Weibo User Based Multiple Microblogs Analysis and Visualization System
..... *Hao Jingchao et al.* (399)

Codesourcing-Based E-Commerce Entity Classification System *Ye Chen et al.* (405)

Biological Experiments Information Management System *Wang Liwei et al.* (410)

A Meta-data Management System on Distributed System *Wei Guanghui et al.* (416)

A Keyword Search System Supporting Query Expansion on XML Data *Hu Hao et al.* (421)

A Village Infrastructure Location Analysis System Based on R^kNN *Xu Jingke et al.* (426)

CloudIndexEval: A Prototype for Evaluating Multi-Dimensional Indexes on Cloud Platform
..... *Zhou Xin et al.* (431)

TPCDroid: Measuring Energy Consumption of Database Queries on Smart Devices
..... *Yang Puyuan et al.* (435)

编辑 《计算机研究与发展》编辑部
中国科学院计算技术研究所
地址: 北京中关村科学院南路6号
邮政编码: 100190
电话: +86 (10) 62620696 (兼传真)
+86 (10) 62600350
E-mail: crad@ict.ac.cn
http://crad.ict.ac.cn

主编 徐志伟
主管 中国科学院
主办 中国科学院计算技术研究所
中国计算机学会

出版 科学出版社
地址: 北京东黄城根北街16号
邮政编码: 100717

印刷装订 北京华耀印刷有限责任公司
订购处 《计算机研究与发展》编辑部
国外总发行 中国国际图书贸易总公司
北京399信箱
邮政编码: 100044

Journal of Computer
Research and Development
(Monthly, Started in 1958)
Vol.50 Suppl. I Aug. 2013

Edited by Editorial Office of *Journal of
Computer Research and Development*
Institute of Computing Technology,
Chinese Academy of Sciences
Add: 6 Kexueyuan South Road,
Zhongguancun, Beijing
100190, China
Tel: +86(10) 62620696 (also Fax)
+86(10) 62600350
E-mail: crad@ict.ac.cn
http://crad.ict.ac.cn

Editor-in-Chief Xu Zhiwei
Supervised by Chinese Academy of Sciences
Sponsored by Institute of Computing Technology,
Chinese Academy of Sciences
China Computer Federation

Published by Science Press
Add: 16 Donghuangchenggen North
Street, Beijing 100717, China

Printed by Beijing Huayao Printing Limited Company
Domestic Editorial Office of *Journal of
Computer Research and Development*
Foreign China International Book
Trading Corporation
P.O.Box 399, Beijing 100044, China