基于图数据库的文献检索方法优化与实现

林启胜 1,2,3，王 磊 1,2，周 喜 1，2，赵 凡 1,2，马 博 1,2
(1 中国科学院 新疆理化技术研究所，新疆 乌鲁木齐 830011；2 新疆民族语音语言信息处理实验室，新疆 乌鲁木齐 830011；3 中国科学院大学 计算机与控制学院，北京 100049)

摘 要：针对目前文献检索系统对于查询结果仅以文本形式呈现的问题，提出了一个基于图数据库的文献信息检索系统.该采用图数据库 Neo4j 进行存储，用户通过关键词查询，系统以图节点的可视化形式将信息呈现给用户，可以直观地了解文献的相关信息.实验结果表明，该系统比关系基于数据库的系统更快，执行的时间减少了 71％（对于 3-节点查询），88％（对于 4 节点的查询），以及 99％的（5-节点查询）.

关键词：Neo4j；数据库；数据存储；信息检索；Cypher

Design and Implementation of a Literature Retrieval System Based on Graph Database

LIN Qi-sheng 1,2,3,WANG Lei 1,2,ZHOU Xi 1,2,ZHAO Fan 1,2， MA Bo 1,2
(1 Xinjiang Technical Institute of Physics & Chemistry, Chinese Academy of Sciences, Urumqi 830011, China;2 Xinjiang Laboratory of Minority Speech & Language Information Processing, Urumqi 830011, China;3 College of Computer and Control Engineering, University of Chinese Academy of Sciences, Beijing 100049, China)

Abstract：For the current literature retrieval system for the results presented in the form of text only problem, a graph database based literature Information Retrieval System. The use graph database Neo4j storage, user query literature by keyword, the system visually in the form of graph nodes of information presented to the user can intuitively understand literature related information. Experimental results show that the system is faster than relational database system, the implementation time is reduced by 71% (for 3- node query) ,88% (for 4-node queries), and 99% (for 5-node query).

Key words：Neo4j; graph database; data store; information retrival; cipher

作者简介：
林启胜 男，（1992-），硕士研究生.研究方向为大数据存储应用、大数据分析.Email:a2950240@163.com.
王 磊 男，（1974-），博士，研究员.研究方向为计算机及应用、大数据分析.
周 喜 男，（1978-），博士，研究员.研究方向为物联网应用技术、大数据分析.
赵 凡 男，（1980-），博士研究生，副研究员.研究方向为信息安全、大数据分析.
马 博 男，（1984-），博士研究生,副研究员.研究方向为计算机应用技术.