

P.O.Box 8718, Beijing 100080, China	Journal of Software, July 2005,16(7):1270-1281
E-mail: jos@iscas.ac.cn	ISSN 1000-9825, CODEN RUXUEW, CN 11-2560/TP
http://www.jos.org.cn	Copyright © 2005 by The Editorial Department of Journal of Software

## SEEKER:基于关键词的关系数据库信息检索

文继军, 王 珊

[Full-Text PDF](#) [Submission](#) [Back](#)

文继军, 王 珊

(中国人民大学 信息学院,北京 100872)

作者简介: 文继军(1971—),男,四川西充人,博士生,主要研究领域为数据库,信息检索;王珊(1944—),女,教授,博士生导师,主要研究领域为数据库,知识工程.

联系人: 文继军 Phn: +86-10-84909436, E-mail: wenjj@ruc.edu.cn, <http://www.ruc.edu.cn>

Received 2005-01-04; Accepted 2005-03-11

### Abstract

Traditionally, SQL is the main interface to access data from relational databases. However, it is difficult for inexperienced end users to learn the complicated syntax of SQL. Enabling keyword-based information retrieval over relational databases will allow users to acquire information from databases without any knowledge of SQL and the underlying database schema, just like the way of common search engines. This paper describes the design and implementation of SEEKER, a system supporting keyword-based information retrieval over relational databases. While there have been some existing systems that support searching text attributes in relational databases, SEEKER can also search database metadata and numeric attributes. Moreover, SEEKER employs an improved ranking function and supports Top-k queries. Experimental results show that SEEKER can achieve good retrieval performance.

Wen JJ, Wang S. SEEKER: Keyword-Based information retrieval over relational databases. *Journal of Software*, 2005,16(7):1270-1281.

DOI: 10.1360/jos161270

<http://www.jos.org.cn/1000-9825/16/1270.htm>

### 摘要

传统上,SQL是存取关系数据库中数据的主要界面.但是,对于没有经验的用户来说,学习复杂的SQL语法是一件困难的事情.实现基于关键词的关系数据库信息检索,将使用户不需要任何SQL语言和底层数据库模式的知识,用搜索引擎的方式来获取数据库中的相关数据.描述了一个基于关键词的关系数据库信息检索系统SEEKER的设计和实现.现有的关系数据库关键词查询系统只能检索关系数据库中的文本属性,而SEEKER还可以检索数据库元数据以及数字属性.并且,SEEKER采用了更合理的排序公式,支持Top-k查询.实验结果显示,SEEKER具有良好的查询性能.

基金项目: Supported by the National Natural Science Foundation of China under Grant Nos.60473069, 60496325 (国家自然科学基金); the National High-Tech Research and Development Plan of China under Grant No.2003AA4Z3030 (国家高技术研究发展计划(863)); the Science and Technology Plan of Beijing of China under Grant No.H030130060011 (北京市科技计划)

### References:

- [1] Oracle text. 2004. <http://otn.oracle.com/products/text/index.html>
- [2] DB2 text information extender. 2004. <http://www-3.ibm.com/software/data/db2/extenders/textinformation/index.html>
- [3] SQL server full-text query. 2004. [http://msdn.microsoft.com/library/default.asp?url=/library/en-us/architec/8\\_ar\\_sa2\\_0ehx.asp](http://msdn.microsoft.com/library/default.asp?url=/library/en-us/architec/8_ar_sa2_0ehx.asp)
- [4] Halotia G, Hulgeri A, Nakhey C, Chakrabarti S, Sudar-Shan S. Keyword searching and browsing in databases using BANKS. In: Agrawal R, et al., eds. Proc. of the 18th Int'l Conf. on Data Engineering. San Jose: IEEE Press, 2002. 431-440.

- [5] Agrawal S, Chaudhuri S, Das G. DBXplorer: A system for keyword-based search over relational databases. In: Agrawal R, et al., eds. Proc. of the 18th Int'l Conf. on Data Engineering. San Jose: IEEE Press, 2002. 5-16.
- [6] Hristidis V, Papakonstantinou Y. DISCOVER: Keyword search in relational databases. In: Bernstein PA, et al., eds. Proc. of the 28th Int'l Conf. on Very Large Data Bases. Hong Kong: Morgan Kaufmann Publishers, 2002. 670-681.
- [7] Hristidis V, Gravano L, Papakonstantinou Y. Efficient IR-style keyword search over relational databases. In: Freytag JC, et al., eds. Proc. of the 29th Int'l Conf. on Very Large Data Bases. Berlin: Morgan Kaufmann Publishers, 2003. 850-861.
- [8] Su Q, Widom J. Indexing relational database content offline for efficient keyword-based search. Technical Report, Stanford: Stanford University, 2003. <http://dbpubs.stanford.edu/pub/2003-13>
- [9] Balmin A, Hristidis V, Papakonstantinou Y. ObjectRank: Authority-Based keyword search in databases. In: Nascimento MA, et al., eds. Proc. of the 30th Int'l Conf. on Very Large Data Bases. Toronto: Morgan Kaufmann Publishers, 2004. 564-575.
- [10] Goldman R, Shivakumar N, Venkatasubramanian S, Garcia-Molina H. Proximity search in databases. In: Gupta A, et al., eds. Proc. of the 24th Int'l Conf. on Very Large Data Bases. New York: Morgan Kaufmann Publishers, 1998. 564-575.
- [11] Florescu D, Kossmann D, Manolescu I. Integrating keyword search into XML query processing. In: Albert V, et al., eds. Proc. of the 9th Int'l World Wide Web Conf. Amsterdam: ACM Press, 2000. 119-136.
- [12] Hristidis V, Papakonstantinou Y, Balmin A. Keyword proximity search on XML graphs. In: Dayal U, et al., eds. Proc. of the 19th Int'l Conf. on Data Engineering. Bangalore: IEEE Press, 2003. 367-378.
- [13] Guo L, Shao F, Botev C, Shanmugasundaram J. XRANK: Ranked keyword search over XML documents. In: Halevy AY, et al., eds. Proc. of the 2003 ACM SIGMOD Int'l Conf. on Management of Data. San Diego: ACM Press, 2003. 16-27.
- [14] Baeza-Yates R, Riberio-Neto B. Modern Information Retrieval. New York: Addison-Wesley, 1999. 29.
- [15] DBLP bibliography. 2004. <http://www.informatik.uni-trier.de/~ley/db/index.html>