

P.O.Box 8718, Beijing 100080, China	Journal of Software, September 2007,18(9):2216-2225
E-mail: jos@iscas.ac.cn	ISSN 1000-9825, CODEN RUXUEW, CN 11-2560/TP
http://www.jos.org.cn	Copyright © 2007 by <i>Journal of Software</i>

一种P2P环境下基于用户行为的语义检索方案

邱志欢, 肖明忠, 代亚非

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

邱志欢, 肖明忠, 代亚非

(北京大学 计算机科学技术系, 北京 100871)

作者简介: 邱志欢(1982—),男,广东汕尾人,硕士生,主要研究领域为对等网络理论、技术.肖明忠(1970—),男,博士,主要研究领域为对等网络,信息检索技术.代亚非(1958—),女,教授,博士生导师,CCF高级会员,主要研究领域为分布式系统,网络存储,P2P计算.

联系人: 邱志欢 Phn: +86-10-62751799 ext 8013, E-mail: qzh@net.pku.edu.cn, http://net.pku.edu.cn

Received 2006-07-03; Accepted 2006-09-30

Abstract

Restricted by the diversity of resources and the complexity of search algorithms, current search mechanisms in peer-to-peer file sharing systems are based on file names and simple keyword matching. These mechanisms cannot recognize deeper relationships between keywords and resources; hence it cannot provide high search quality. This paper proposes a new search scheme, which is built on top of the current peer-to-peer network. It harnesses users' search behaviors and download behaviors to automatically discover the deeper relationships between keywords and resources, which is then used to improve the search quality. It has the advantages of low implementation cost, low complexity, self-evolving, and supports for semantic search. Simulations based on the Maze system show that this approach has high search hit rate and accuracy.

Qiu ZH, Xiao MZ, Dai YF. A user behavior based semantic search approach under P2P environment. *Journal of Software*, 2007,18(9):2216-2225.

DOI: 10.1360/jos182216

<http://www.jos.org.cn/1000-9825/18/2216.htm>

摘要

受资源类型多样化、搜索复杂度的制约,现有的P2P文件共享系统中的搜索机制是基于文件名的关键字匹配,这种方法不能发现关键字与资源内容之间的深层关系,因此不能实现语义检索.针对这个问题,提出一种新的搜索方案,该方案建立在已有的搜索机制之上,利用用户的搜索行为和下载行为的规律自动发现关键字和资源间的深层关系,在底层的P2P网络上构建一个元数据空间以辅助搜索.该方案具有实现代价小、时间复杂度低、可进化和支持语义搜索的优点.在Maze系统上的实验表明,该方案具有较高的查询命中率和查询准确率.

基金项目: Supported by the National Natural Science Foundation of China under Grant No.90412008 (国家自然科学基金); the National Basic Research Program of China under Grant No.2004CB318204 (国家重点基础研究发展计划(973))

References:

[1] Sripanidkulchai K, Maggs B, Zhang H. Efficient content location using interest-based locality in peer-to-peer systems. In: Proc. of the IEEE INFOCOM 2003. IEEE Press, 2003. 2166-2176.

[2] Asvanund A, Krishnan R, Smith M, Telang R. Interest-Based self-organizing peer-to-peer networks: A club economics approach. In: Proc. of the 13th Workshop on Information Technology and Systems. 2003. <http://www.business.uconn.edu/users/atung/seminar/fall2004/smith-paper.pdf>

[3] Crespo A, Garcia-Molina H. Semantic overlay networks for P2P systems. In: Moro G, Bergamaschi S, Aberer K, eds. Proc. of the 3rd Int'l Workshop on Agents and Peer-to-Peer Computing. Berlin: Springer-Verlag, 2004. 1-13.

- [4] Singh S, Ramabhadran S, Baboescu F, Snoeren AC. The case for service provider deployment of super-peers in P2P networks. In: Proc. of the Workshop on Economics of P2P Systems. Berkeley, 2003. <http://www.cs.ucsd.edu/~susingh/papers/tbs-p2pecon03.pdf>
- [5] Nejdil W, Wolf B, Ou C, Decker B, Sintek M, Naeve A, Nilsson M, Palmer M, Risch T. EDUTELLA: A P2P networking infrastructure based on RDF. In: Proc. of the 11th Int'l World Wide Web Conf. IEEE Press, 2002. 604-615.
- [6] Stoica I, Morris R, Karger D, Kaashoek MF, Balakrishnan H. Chord: A scalable peer-to-peer lookup service for Internet applications. In: Proc. of the ACM SIGCOMM 2001. San Diego: ACM Press, 2001. 149-160.
- [7] Zhao B, Kubiawicz J, Joseph A. Tapestry: An infrastructure for fault-tolerant wide-area location and routing. Technical Report, UCB/CSD-01-1141, Computer Science Division, U. C. Berkeley, 2001.
- [8] Ratnasamy S, Francis P, Handley M, Karp R, Shenker S. A scalable content-addressable network. In: Proc. of the ACM SIGCOMM 2001. San Diego: ACM Press, 2001. 168-175.
- [9] Gnutella. <http://www.gnutella.com>
- [10] Lü Q, Cao P, Cohen E, Li K, Shenker S. Search and replication in unstructured peer-to-peer networks. In: Proc. of the ACM SIGMETRICS 2002. ACM Press, 2002. 258-259.
- [11] eMule. <http://www.emule.net/>
- [12] BitTorrent. <http://www.bittorrent.com/>
- [13] Navigli R, Velardi P, Gangemi A. Ontology learning and its application to automated terminology translation. IEEE Intelligent Systems, 2003,18(1):22-31.
- [14] Liu HY. Analysis of resource characteristics and user behavior in P2P file sharing system maze [MS. Thesis]. Beijing: Peking University, 2005 (in Chinese with English abstract).
- [15] Has J, Kamber M. Data Mining: Concepts and Techniques. Morgan Kaufmann Publishers, 2000. 225-243.

附中文参考文献:

- [14] 刘翰宇.P2P文件共享系统Maze中资源及用户行为特征分析[硕士学位论文].北京:北京大学,2005.