SemreX:一种基于语义相似度的P2P覆盖网络

陈汉华, 金 海, 宁小敏, 袁平鹏, 武 浩, 郭志鑫

陈汉华, 金 海, 宁小敏, 袁平鹏, 武 浩, 郭志鑫
(集群与网格计算湖北省重点实验室(华中科技大学),湖北 武汉 430074)
作者简介: 陈汉华(1978－),男,湖北武汉人,博士生,主要研究领域为网格计算,对等计算,语义Web.金海(1966－),男,博士,教授,博士生导师,CCF高级会员,主要研究领域为计算机体系结构,集群计算与网格计算,并行与分布式计算,对等计算,高性能网络存储和并行I/O,Web和网络安全,多媒体技术,移动计算,普适计算.宁小敏(1978－),男,博士生,主要研究领域为网格计算,对等计算,语义Web.袁平鹏(1972－),男,博士,副教授,主要研究领域为CSCW,网格计算,语义Web.武浩(1979－),男,博士生,主要研究领域为网格计算,对等计算,语义Web.郭志鑫(1977－),男,博士生,讲师,主要研究领域为网格计算,语义Web.
联系人: 金 海 Phn: +86-27-87543529, E-mail: hjin@hust.edu.cn, http://grid.hust.edu.cn/hjin

Abstract
The decentralized structure together with the self-organization and fault-tolerant features makes P2P network an effective model for information sharing, however, the content location remains a serious challenge of large scale P2P networks. In this paper, the SemreX is introduced, which is a P2P system for literature retrieval. Semantic-similarity-based P2P overlay and routing algorithms are proposed for SemreX. Experimental results show that searching in semantic overlay greatly improves the efficiency of search.

摘要
对等(peer-to-peer)网络的非集中结构、良好的自治性及容错性等特征,使其可能成为Internet上有效的信息共享模型.然而,内容定位问题仍然是大规模P2P网络中信息共享所面临的挑战.SemreX系统是一种P2P网络环境下的文献检索系统.针对SemreX系统,提出一种基于语义相似度的P2P

拓扑管理和查询路由算法.仿真实验结果表明,语义拓扑能够有效地提高系统的搜索效率.

References:

[1] Shen HT, Shu Y, Yu B. Efficient semantic-based content search in P2P network. IEEE Trans. on Knowledge and Data Engineering, 2004,16(7):813-826.

[2] Stoica I, Morris R, Karger D, Kaashoek MF, Balakrishnan H. Chord: A scalable peer-to-peer lookup service for Internet applications. In: Govindan, ed. Proc. of the ACM SIGCOMM 2001. ACM Press, 2001. 149-160.

[3] Ratnasamy S, Francis P, Handley M, Karp R, Shenker S. A scalable content-addressable network. In: Govindan, ed. Proc. of the ACM SIGCOMM 2001. ACM Press, 2001. 162-172.

[4] Castro M, Costa M, Rowstron A. Debunking some myths about structured and unstructured overlays. In: Proc. of the 2nd USENIX Symp. on Networked Systems Design and Implementation (NSDI 2005). USENIX, 2005.

[5] Gkantsidis C, Mihail M, Saberi A. Random walks in peer-to-peer networks. In: Proc. of the IEEE INFOCOM 2004. New York: IEEE Press, 2004. 120-130.

[6] Broekstra J, Kampman A, Harmelen FV. Sesame: A generic architecture for storing and querying RDF and RDF schema. In: Horrocks I, Hendler JA, eds. Proc. of the ISWC 2002. Berlin: Springer-Verlag, 2002. 54?68.

[7] Milojicic DS, Kalogeraki V, Lukose R, Nagaraja K, Pruyne J, Richard B, Rollings S, Xu Z. Peer-to-Peer Computing. Palo Alto: HP Laboratories: Hewlett-Packard Company, 2002. 1-52.

[8] Budanitsky A, Hirst G. Semantic distance in WordNet: An experimental, application-oriented evaluation of five measures. In: Proc. of the Workshop on WordNet and other Lexical Resources. 2001.

[9] Rada R, Mili H, Bicknell E, Blettner M. Development and application of a metric on semantic nets. IEEE Trans. on System, Man, and Cybernetics, 1989,19(1):17-30.

[10] Yuhua L, Bandar ZA, McLean D. An approach for measuring semantic similarity between words using multiple information sources. IEEE Trans. on Knowledge and Data Engineering, 2003,15(4):871-882.

[11] Resnik P. Semantic similarity in a taxonomy: An information-based measure and its application to problems of ambiguity in natural language. Journal of Artificial Intelligence Research, 1999,11:95-130.

[12] Jiang JJ, Conrath DW. Semantic similarity based on corpus statistics and lexical taxonomy. In: Proc. of the Int'l Conf. Research on Computational Linguistics (ROCLING X). 1997.

[13] BatagelJ V, Mrvar A. Pajek—Analysis and visualization of large networks. In: Mutzel P, Jünger M, Leipert S, eds. Proc. of the 9th Int'l Symp. on Graph Drawing. Berlin: Springer-Verlag, 2001. 477-478.

[14] Matei R, Lamnitchi A, Foster I. Mapping the Gnutella network. IEEE Internet Computing, 2002,6(1):50-57.

[15] Haase P, Broekstra J, Ehrig M, Menken M, Mika P, Plechawski M, Pyszlak P, Schnizler B, Siebes R, Staab S, Tempich C. Bibster?A semantic-based bibliographic peer-to-peer system. In: McIlraith SA, Plexousakis D, Harmelen FV, eds. Proc. of the ISWC 2004. Berlin: Springer-Verlag, 2004. 122-136.~42.