



## 面向专利分析的法律状态分布式采集系统的设计与实现

吴红<sup>1</sup>, 王凤英<sup>2</sup>, 付秀颖<sup>1</sup>

1. 山东理工大学科技信息研究所 淄博 255049;
2. 山东理工大学计算机科学与技术学院 淄博 255049

Wu Hong<sup>1</sup>, Wang Fengying<sup>2</sup>, Fu Xiuying<sup>1</sup>

1. Institute of Scientific & Technical Information, Shandong University of Technology, Zibo 255049, China;
2. College of Computer Science and Technology, Shandong University of Technology, Zibo 255049, China

- 摘要
- 参考文献
- 相关文章

Download: PDF (727KB) [HTML \(KB\)](#) Export: BibTeX or EndNote (RIS) [Supporting Info](#)

**摘要** 针对专利分析对法律状态需求量大而法律状态需单项获取的矛盾,设计并实现基于网络的法律状态信息采集系统。系统通过分布式架构启动多台机器进行搜索,通过一致性Hash算法并引入虚拟客户节点,克服进程分配中的不平衡问题,配合以网页识别,实现包括法律状态在内的专利信息批量、经济、快速获取,有效解决专利现状分析中的数据获取瓶颈问题。实例表明,融入法律状态的专利分析,能使分析结果更为全面、客观和丰富,并更具针对性。

**关键词:** [分布式架构](#) [一致性哈希](#) [专利法律状态](#) [存活寿命](#) [信息采集](#)

**Abstract:** Aiming at the contradictions that patent analysis has enormous demands for legal status, while legal status is acquired singly, the paper designs and realizes a legal status information collection system. The system adopts the distributed framework, which conducts searches by multiple machines, and overcomes the unbalanced process distribution by consistent hashing algorithms as well as virtual client nodes. Besides, using identification of Web pages technology, it is able to acquire patent information that includes legal status in a batch, economic, quickly process. It also presents a solution to the information acquiring bottleneck problem in today's patent status analysis. The application example shows that combining patent analysis with legal status makes the analysis results more comprehensive, objective, multiple and specific.

**Keywords:** [Distributed architecture](#), [Consistent hashing](#), [Patent legal status](#), [Survival life span](#), [Information collection](#)

收稿日期: 2012-08-28;

基金资助: 本文系山东省社会科学规划研究项目“高校专利质量综合评价及对策研究”(项目编号: 12CTQJ03)和山东理工大学人文社会科学基金项目“Web信息挖掘与智能检索”(项目编号: 2010GGTD05)的研究成果之一。

通讯作者 吴红 Email: wuhong0256@163.com

引用本文:

吴红, 王凤英, 付秀颖. 面向专利分析的法律状态分布式采集系统的设计与实现[J]. 现代图书情报技术, 2012, V(12): 66-71

Wu Hong, Wang Fengying, Fu Xiuying. Design and Establishment of Legal Status Distributed Collection System Based on Patent Analysis[J]. *Modern Library and Information Service*, 2012, V(12): 66-71

链接本文:

<http://www.infotech.ac.cn/> 或 <http://www.infotech.ac.cn/Y2012/V/112/66>

[1] 中国知识产权网. 专利分析之专利法律状态分析[EB/OL]. [2011-10-05].

[http://www.cnipr.com/services/zlxyysw/wz/201105/t20110505\\_133565.html](http://www.cnipr.com/services/zlxyysw/wz/201105/t20110505_133565.html). (China Intellectual Property Net. Patent Legal State Analysis [EB/OL]. [2011-10-05]. [http://www.cnipr.com/services/zlxyysw/wz/201105/t20110505\\_133565.html](http://www.cnipr.com/services/zlxyysw/wz/201105/t20110505_133565.html).)

[2] Schankerman M. How Valuable is Patent Protection? Estimates by Technology Field Using Patent Renewal Data [J]. *Rand Journal of Economics*, 1998, 29(1): 77-107.

[3] Raghavan S, Garcia-Molina H. Crawling the Hidden Web[C]. In: *Proceedings of the 27th International Conference on Very Large Data Bases (VLDB' 01)*. San Francisco: Morgan Kaufmann Publishers Inc., 2001: 129-138.

[4] Ntoulas A, Cho J, Olston C. What's New on the Web? The Evolution of the Web from a Search Engine Perspective [C]. In: *Proceedings of the 13th International Conference on World Wide Web*. New York, NY, USA: ACM, 2004: 1-12.

## Service

- ▶ 把本文推荐给朋友
- ▶ 加入我的书架
- ▶ 加入引用管理器
- ▶ Email Alert
- ▶ RSS

## 作者相关文章

- ▶ 吴红
- ▶ 王凤英
- ▶ 付秀颖

- [5] 吴琳,魏星,霍翠婷.基于Web的专利双语语料自动获取研究及实现——以esp@cenet数据库为例[J]. 现代图书情报技术, 2009(9):57-63.(Wu Lin, Wei Xing Huo Cuiting. Research and Implement of Automatic Patent Bilingual Corpus Extraction from Web—Taking esp@cenet as an Example [J]. *Technology of Library and Information Service*, 2009(9):57-63.) 浏览
- [6] 梁莹,徐福缘. 基于多Agent的专利资源协同获取模型研究[J]. 情报理论与实践, 2009,32(8):118-120.(Liang Ying, Xu Fuyuan. On Collaborative Acquisition of Patent Resources Based on Multi-agent[J]. *Information Studies: Theory & Application*, 2009,32(8):118-120.)
- [7] 张红,杨祖国. 利用Internet免费检索并获取中国专利全文的方法[J]. 情报科学,2001(9):984-985.(Zhang Hong, Yang Zuguo. Free Retrieval Full Text Chinese Patent in Internet[J]. *Information Science*, 2001(9):984-985.)
- [8] 托姆舒,施穆策.下一代光网络:IP层的智能与光层技术的融合[M].北京:人民邮电出版社,2003.(Tomsu P, Schmutzer C. Next Generation Optical Network The Convergence of IP Intelligence and Optical Technologies[M].Beijing: Posts & Telecom Press,2003.)
- [9] 国家知识产权局规划发展司.2010年中国有效专利年度报告 [J]. 专利统计简报, 2011(6):1-27.(Planning and Development Division of SIPO. China Effective Patent Annual Report on 2010[J]. *Patent Statistical Briefing*, 2011(6):1-27.)
- [10] 《国家知识产权局2009年工作要点》解读[EB/OL].[2012-08-12].[http://www.gov.cn/zwhd/2009-04/02/content\\_1276132.htm](http://www.gov.cn/zwhd/2009-04/02/content_1276132.htm). (Explain on 2009 Work Emphasis of SIPO[EB/OL].[2012-08-12]. [http://www.gov.cn/zwhd/2009-04/02/content\\_1276132.htm](http://www.gov.cn/zwhd/2009-04/02/content_1276132.htm).)
- [1] 许鑫 黄仲清 邓三鸿.互联网侨情信息采集系统设计与实现\*[J]. 现代图书情报技术, 2010,26(7/8): 95-101
- [2] 陈诗琴 李文江.基于.Net的农产品市场行情信息采集 ——以重庆农产品市场行情查询网为例[J]. 现代图书情报技术, 2010,26(6): 88-92
- [3] 黄进.图书馆应用系统监控的设计与实现[J]. 现代图书情报技术, 2010,26(3): 90-94
- [4] 许鑫,黄仲清.垂直搜索引擎应用中的若干策略探讨\*——以12580餐饮垂直搜索为例[J]. 现代图书情报技术, 2009,3(2): 62-70
- [5] 钱爱兵.基于主题的网络舆情分析模型及其实现[J]. 现代图书情报技术, 2008,24(4): 49-55
- [6] 徐德智,王庆涛,王斌.基于本体的Web信息采集\*[J]. 现代图书情报技术, 2007,2(2): 53-55
- [7] 刘莉,肖诗斌,王涛,施水才.基于RSS的博客采集系统的设计与实现\*[J]. 现代图书情报技术, 2007,2(11): 45-48
- [8] . . . . . [J]. . . . . 2007.2(1): 44-48