太　原　理　工　大　学

毕业设计（论文）任务书

第1页

|  |
| --- |
| 毕业设计（论文）题目：  网络爬虫的设计与实现  毕业设计（论文）要求及原始数据（资料）：  1．综述国内外网络爬虫发展历史及研究现状；  2．深入了解网络爬虫的相关技术；  3．训练编写基本爬虫的能力；  4．深入分析网络爬虫中的关键技术与编写语言；  5．设计并实现网络爬虫的简单应用；  6．训练检索文献资料和利用文献资料的能力；  7．训练撰写技术文档与学位论文的能力。 |

第2页

|  |
| --- |
| 毕业设计（论文）主要内容：  1．介绍网络爬虫的相关技术；  2．介绍网络爬虫的开发环境；  3．深入分析网络爬虫中的关键技术与相关编写语言的设计流程；  4. 分析一个基本的爬虫程序的运行原理与过程；  5. 设计与实现抓取某一信息并进行分析的爬虫程序；  6．综述爬虫技术在数据采集中的应用。  学生应交出的设计文件（论文）：  1. 内容完整、层次清晰、叙述流畅、排版规范的毕业设计论文；  2. 包括毕业设计论文、源程序等内容在内的毕业设计电子文档及其它相关材料。 |

第3页

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 主要参考文献（资料）：   1. 徐远超, 刘江华, 刘丽珍,等. 基于Web的网络爬虫的设计与实现[J]. 微计算机信息, 2007, 23(21) 2. 吴黎兵, 柯亚林, 何炎祥,等. 分布式网络爬虫的设计与实现[J]. 计算机应用与软件, 2011, 28(11). 3. 姜梦稚. 基于Java的多线程网络爬虫设计与实现[J]. 微型电脑应用, 2010, 26(7). 4. 周德懋, 李舟军. 高性能网络爬虫:研究综述[J]. 计算机科学, 2009, 36(8). 5. 刘金红, 陆余良. 主题网络爬虫研究综述[J]. 计算机应用研究, 2007, 24(10). 6. 张敏, 高剑峰, 马少平. 基于链接描述文本及其上下文的Web信息检索[J]. 计算机研究与发展, 2004, 41(1). 7. 周立柱, 林玲. 聚焦爬虫技术研究综述[J]. 计算机应用, 2005, 25(9). 8. Chau M, Xu J. Mining communities and their relationships in blogs: A study of online hate groups[J]. International Journal of Human-Computer Studies, 2007, 65(1). 9. Pant G, Srinivasan P. Learning to crawl: Comparing classification schemes[J]. Acm Transactions on Information Systems, 2005, 23(4). 10. Shkapenyuk V, Suel T. Design and implementation of a high-performance distributed Web crawler[J]. 2001. 11. Briangoetz. Java并发编程实战[M]. 机械工业出版社, 2012. 12. ElliotteRustyHarold, 哈诺德, 李帅,等. Java网络编程[M]. 中国电力出版社, 2014 13. 米切尔, 陶俊杰, 陈小莉. Python网络数据采集[M]. 人民邮电出版社, 2016..   14、麦金尼，唐学韬. 利用Python进行数据分析[M]. 机械工业出版社, 2014.   |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | --- | | 专业班级 | 软件1301班 | | | | | 学生 | 杨泽鑫 | | 要求设计（论文）工作起止日期 | | | | | 2017年3月13日~2017年6月25日 | | | | 指导教师签字 | |  | | | | 日期 | 2017年3月13日 | | 教研室主任审查签字 | | | |  | | 日期 |  | | 系主任批准签字 | | |  | | | 日期 |  | |