**南 京 大 学**

**本科生毕业论文（设计、作品）指导情况记录**

|  |  |
| --- | --- |
| **开**  **题**  **简**  **况** | **题目：分布式爬虫的设计与实现** |
| **1、选题质量（简述选题与专业培养目标、专业要求关系、题目难度、工作量、创新性、理论性、实用性）**  **在大数据分析中，网络爬虫具有重要的地位，只有搜集的数据够准确和全面，才能支持接下来数据分析的工作。然而在如今信息大爆炸的时代，互联网中的信息传播迅速，信息资源的增长速度前所未有。如此大规模的信息，使单机爬取数据并加以分析的方法成为过去。相比而言，爬虫的分布式架构是一种合理的实现方式。单机爬虫在规模不断扩大的互联网中无论效率还是工作质量都无法同分布式爬虫相媲美。**  **本人准备采用java的jms技术，构建一个爬虫云平台，让主服务器分配给各个从服务器爬取任务（一系列url），从服务器执行完后反馈信息，最后将结果汇总。题目由于涉及到网络通讯和多线程，因此有一定难度，工作量适中。利用jms及其他技术亲自实现一个轻便的分布式系统不仅让自己对于分布式系统有了更深刻的理解，也具有一定创新性。另外，从理论和实际两方面来看，该项目都是具有挑战性但是可行的。**  **2、开题意见：**  **指导教师签名：**  **年 月 日** |
| **中期检查** | **指导教师检查论文的进展情况：（指导和培养学生查阅文献资料、综合运用知识、研究方案设计、研究方法和手段运用和外文应用等能力简况）**  **指导教师签名：**  **年 月 日** |