

**网络爬虫工具的设计与实现**

**阶段报告（一）**

|  |  |  |
| --- | --- | --- |
| 小组成员 | 姓名 | 学号 |
| 组长 | 王奕超 | 1501211000 |
| 组员 | 张晗 | 1501211029 |
| 组员 | 胡玉 | 1501210913 |

‌

2015年12月9日 星期六

1. **项目题目**

技术类博客汇总分类

1. **项目简介**

技术类博客对于技术人员的学习是非常有帮助的，可以快速的帮助程序员发现能够解决所遇到问题的方案或途径。但是网络上的技术类博客数量庞大而杂乱，在查阅的过程中会严重的影响问题的解决速度以及准确度，进而对项目任务会造成一定程度的不利影响。

本项目针对这一现实问题进行设计与实现，通过编写网络爬虫，抓取网络上的技术类博客数据，并进行较为细致的分类，方便技术人员的查阅，提高问题的解决速度与质量。

1. **涉及技术**

Java IO，Java 文件操作，HTML的解析，Web开发相关技术等