**项目计划问答记录**

# 基本信息

|  |  |
| --- | --- |
| **项目名称** | Scray |
| **项目组成员** | 郭炜锋，王益飞，胡勇，武丁泽宇 |
| **项目简介** | Scrapy是一个为了爬取网站数据，提取结构性数据而编写的应用框架。 可以应用在包括数据挖掘，信息处理或存储历史数据等一系列的程序中。Scrapy吸引人的地方在于它是一个框架，任何人都可以根据需求方便的修改。它也提供了多种类型爬虫的基类，如BaseSpider、sitemap爬虫等，最新版本又提供了web2.0爬虫的支持。 |
| **评审时间** | 2015.3.17 晚上18：00-21：00 |
| **评审地点** | A209 |
| **评审对象** | 《Scrapy项目计划》 |
| **评审方式** | 会议审查 |
| **评审员** | 刘超 教授  任健 教师  全体选课同学 |

# 评审意见

|  |  |  |
| --- | --- | --- |
| 序号 | 问题 | 回答 |
| 1 | 可以抓取动态网站么？ | 这是个web抓取框架，根据需求需要自己进行代码编写，直接使用抓取动态网站比较困难 |
| 2 | 代码改进？ | 需求研究不深，可以根据具体需求进行改进 |
| 3 | 会出现哪些高并发问题？ | Spider的数量过大，会出现高并发性能问题 |
| 4 | 高并发对性能有影响，根据什么评估？ | 根据抓取跟其他因素进行评估 |
| 5 | 怎么选取的，为什么选择此项目？ | 在自己写爬虫的时候会出现很多问题，根据开源爬虫项目使用，对工作很有帮助，并且根据小组成员讨论进行确定 |
| 6 | 四个组员的差异性如何处理？ | 进行讨论，根据成员的不同能力进行分析 |
| 7 | 对于后期实验中的改进与扩展工作如何进行？ | 再进行小组讨论，综合考虑项目的连续性 |