# 会议记录（更新至4.28）

|  |  |
| --- | --- |
| 时间 | 记录 |
| 3/10 | 1. 开首次组讨论会，简单介绍了各自的情况。 2. 分配寻找自己感兴趣的开源项目。 |
| 3/12 | 决定开源项目为：Scrapy，要求组内查找关于项目的相关资料。 |
| 3/13 | 1. 查找关于项目的相关资料，进行进一步的分析， 2. 讨论本周工作要求，编写计划书和计划表， 3. 由郭炜锋进行编写第一版任务进度，郭炜锋、王益飞编写初步计划书，胡勇写ppt。 4. 由武丁泽宇进行ppt的演讲。 |
| 3/17 | 1. 小组决定每周五晚上进行一次线下讨论。 2. 对项目需求进行具体分析。 |
| 3/19 | 1. 郭炜锋进行本周会议记录。 2. 胡勇进行本周ppt及汇报。 3. 武丁泽宇进行本周需求规格说明书初稿编写。 4. 王益飞进行计划进度更新。 |
| 3/24 | 小组在课后开了简单会议，针对下周任务简单安排组员进行相应。的工作，在之后进行详细讨论 |
| 3/27 | 1. 小组四人分别画用例图。 2. 武丁泽宇进行需求说明书的更新。 3. 王益飞进行计划进度的更新。 4. 胡勇进行ppt编写与演讲。 5. 郭炜锋进行问题清单整理。 6. 评审在之后会进行讨论。 |
| 3/29 | 1. 讨论检查单编写和提交问题。 2. 每个人编写自己的检查单并且提交上github。 3. 将问题汇总至组长，进行统一处理。 |
| 3/31 | 1. 讨论下周工作的主要任务。 2. 分配学习使用RUCM，并且画出主要需求的RUCM用例图。 3. 对需求说明书进行修改。 4. 武丁泽宇进行需求说明书的整体完善 |
| 4/5 | 1. 讨论RUCM用例图的分配 2. 郭炜锋画爬虫引擎 和 数据存储 3. 胡勇画业务调度和数据下载 4. 王益飞画爬虫和运行设置 |
| 4/7 | 1. 开会讨论本组本周需要完成的任务，同时讨论之后的安排计划如下： 2. 每周五我们进行一次线下会议；目的：明确本周工作任务和大致分配； 3. 周五晚上到周六我进行本周任务分配的具体调整；（因为现在要做好以后的分配，所以只能是以后调整） 4. 下一周的周二我们微信碰个头。目的：协调一下进度，如果有特殊情况，可以重新分配。（没什么事一人说一句就完了） 5. 郭炜锋跟新微信讨论进度 6. 下一周周四到周五，胡勇负责汇报ppt编写 7. 周五晚上汇报ppt（暂时计划如此，可能会有改动） 8. 讨论周日之前的需求评审，由胡勇提供需求评审格式表 |
| 4/14 | 1. 由于缺少个人日志，要求从这周开始编写个人日志 2. 进度计划中可能缺少某些已经做过却未记录的工作，组员查看进度计划与之前工作进行修改和更新 3. 针对评审内容进行需求说明书的修改 4. 暂时考虑实验8的5个目标，周二进行深入讨论 |
| 4/16 | 分配实验八工作内容  1组队立项、2项目计划、3需求分析、4软件需求评审、5文档变更与追踪、6需求分析结果的质量评估、7确定实验目标、8实验要点说明、9实验计划、10试验统计分析的阶段报告、11实验总结报告、12分开写完以后的汇总  武：5,6,12  郭：1,3,4  胡：7,8,11  王：2,9,10 |
| 4/18 | 讨论的主要内容为：  1.第七个实验，git的事情（补一次每个人的git学习的任务，但是以后就不单独设置任务了，或者是否需要单独设置git管理的任务）  2.重点需求的改进（确定我们改进的点有哪些）  讨论结果：  补一次每个人的git学习的任务，但是以后就不单独设置任务了  关于scrapy的改进可以从这几方面考虑：  1.scrapy针对已经下载过的网页的过滤器用的是Python集合，可以做改进。  2.关于反爬，可以在下载中间件添加IP代理、动态修改UserAgent、设置延迟、禁用cookie等。  3.关于动态网页。selenium+一个浏览器内核。1.下载器要改成用浏览器内核去下载。2.提供模拟浏览器操作的接口给爬虫开发者等等，细节之后继续讨论 |
| 4/21 | 小组进行线下讨论，讨论下周的工作分配，以及周日晚上之前的需求修改中具体的细化内容和老师提出的修改要求，但是由于时间关系以及组员之间的讨论没有确定具体内容，需要在之后的微信上继续讨论。 |
| 4/28 | 小组线下讨论改进的详细内容以及计划：   1. 胡勇进行爬虫代码的编写，根据组内提出的需求进行编写 2. 郭炜锋进行爬虫代理中部分代码的改进 3. 王益飞进行过滤代码部分的改进 4. 武丁根据小组成员的工作进行需求报告的编写（注：武丁缺席） |