# 复评审意见反馈

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **序号** | **问题位置** | **问题描述** | **评审人** | **处理人** | **处理意见**  **（除接受外，其他意见需具体描述）** | **评审人意见**  **（填写是否认可）** |
| 1 | 图3.1 | 解析页面使用了3个extend用例，即可以三个都不选，但是按照图3.6应当是三选一，这里是否有矛盾？（同样的，处理项目使用了两个extends，按照下文的意思，应该也是必须二选一） | F-麦梓健 | 赵正阳 | 部分接受：“解析页面”扩展的3个用例确实可以都不选，因为不是所有页面都有需要提取的数据，已修改RUCM图中的相关表述 | 认可 |
| 2 | 3.1 | RUCM图例过小，阅读非常吃力 | F-麦梓健  F-洪治凑 | 赵正阳 | 接受 | 认可 |
| 3 | 3.2 | “..编程者可以通过它提供的JSON API来部署…”  JSON API太笼统了，能否简单的解释下 | F-麦梓健 | 梁远志 | 接受，已增加描述和术语解释 | 认可 |
| 4 | 3.1、3.2 | 3.2一共出现的Actor指代名词包括：编程者、使用者、Scheduler（用例图中），再加上3.1出现的爬虫程序开发者，是否应当统一指称 | F-麦梓健 | 沈一聪 | 接受，在3.1.3中明确了各个外部框架或者子系统的Actor，现在各个部分的参与者已经被统一和明确 | 认可 |
| 5 | 3 | 后端与前端划分不清，首先没有区分是Web前后端还是服务器前后端  （1）在第二章中，图2.2 Web前端只包含了前端界面UI，将剩余的Model（引擎等）和Controller（JavaScript等）全部划进了服务器后端？或者说如果认为图片只说明了Web前后端的分离，数据库是否不应当作为该部分出现，它应当是服务器后端的部分  （2）按照3.1.2的说法，“…后端服务器…”，如果认为该节“后端”指的是服务器后端，但是处理的又是前端的一些JS交互和POST、GET交互这些Web后端的东西  （3）3.3.3 非功能需求-后端，几个特性从Web前端到服务器后端都有 | F-麦梓健 | 沈一聪 | 接受，现已将问题描述中的相关部分进行了合并 | 认可 |
| 6 | 2.4 | 位置2.4中，“…前端使用JavaScript  ES6、JQuery 4.3.1作为编程语言…”前端页面一般是HTML+CSS+JS，其次Jqery是JS库，不应当算作编程语言 | F-麦梓健 | 宋冰晨 | 接受 | 认可 |
| 7 | 2.4 | 2.4中的运行工具定义是什么？如果将运行工具定义为引擎，那么除去JS的模版渲染引擎外，Django作为重量级框架应当是也是有引擎的，更不用说数据库引擎 | F-麦梓健 | 宋冰晨 | 接受 | 认可 |
| 8 | 1.3 | 提供的维基百科Scrapy url资源404无法访问 | F-洪治凑 | 宋冰晨 | 部分接受，在参考资料页面添加脚注 | 认可 |
| 9 | 1.4 | 后文出现的一些英文简写在此处没有说明，如HTML、URL等 | F-洪治凑 | 赵正阳 | 接受 | 认可 |
| 10 | 4 | 希望运行环境中添加详细说明，如操作系统：windows 7、windows  10等 | F-洪治凑 | 赵正阳 | 接受 | 认可 |
| 11 | 全文 | RUCM图里文字过小难以看清 | F-洪治凑 | 赵正阳 | 接受 | 认可 |
| 12 | 3.2.3 | 主流新闻、电商类的网站定义模糊，建议举例 | F-洪治凑 | 宋冰晨 | 接受 | 认可 |
| 13 | 1.2 | “Scrapy还提供了一个网络爬虫外壳”一句中的“外壳”容易引起误解 | F-王伟民 | 宋冰晨 | 接受 | 认可 |
| 14 | 1.2 | “它使您可以使用JSON API部署（上载）您的项目并控制爬虫。  ”中使用了第二人称 | F-王伟民 | 宋冰晨 | 接受 | 认可 |
| 15 | 1.3 | 反爬虫措施中“IP请求速度限制”表达有误 | F-王伟民 | 赵正阳 | 接受 | 认可 |
| 16 | 1.3 | 有关Scrapy中的特定术语解释部分内容不充分，没有说明该术语为Scrapy的组件，让人不容易理解。  比如术语“Scheduler”应该是Scrapy中的调度组件，未加说明的情况下容易让人以为“Scheduler”是和“RUCM”这样的通用术语 | F-王伟民 | 赵正阳 | 接受 | 认可 |
| 17 | 1.4 | 参考资料格式不统一：7、8、9参考资料缺少“[OL].”标注 | F-王伟民 | 赵正阳 | 接受 | 认可 |
| 18 | 2.3.1 | 爬虫工作过程的描述中术语中英文混用 | F-王伟民 | 赵正阳 | 接受 | 认可 |
| 19 | 2.4 | “编程语言”部分表达不准确：JQuery不是编程语言 | F-王伟民 | 宋冰晨 | 接受 | 认可 |
| 20 | 3.2.3 | 该小节一开始列举的内容缺少说明。让人无法确是该部分是描述功能需求还是其他。 | F-王伟民 | 宋冰晨 | 接受 | 认可 |
| 21 | 3.3.2 | “unobtrusive JavaScript”首字母没有大写 | F-王伟民 | 宋冰晨 | 接受 | 认可 |
| 1 | 封面 | 封面的北航logo没有居中，而是左对齐，而其他内容都是居中对齐，所以导致logo有点不好看 | G-王康明 | 赵正阳 | 接受 | 认可 |
| 2 | 2.1 目标 | 句子存在语病。  “在生成爬虫脚本时，加入应对反爬虫机制来增强爬虫程序的鲁棒性，通过增加分布式爬虫机制来提高爬取效率。”这句话连主语都没有，让读者感到非常困惑。 | G-王康明 | 赵正阳 | 接受 | 认可 |
| 3 | 2.2用户特点 | 3.1.3 参与者说明  里说明EasySpider有user、administrator和tasker这3类参与者，  但是 2.2 用户特点 里没有对这3类参与者的特点分别进行分析。 | G-王康明 | 沈一聪 | 部分接受，2.2章的作用主要是想专门用于介绍软件的服务对象。也就是3.1.3中若干参与者中的“用户”，其余的参与者在3.1.3节中有介绍。现已在2.2用特点中进行了说明。 | 认可 |
| 4 | 3.5  数据存储 | “以及4个关系（Relation）：”后面的文字，在中文文字中使用了英文的逗号、英文的括号和中文的冒号，标点符号不统一。 | G-王康明 | 梁远志 | 接受，已改用全角符号 | 认可 |
| 5 | 3.5  数据存储 | “（1）任务作业数据存储:用于存储用户信息、爬虫任务信息等结构化的数据；  （2）爬虫结果数据存储：用于存取非结构化的爬虫结果数据。”这段话里面，既有中文的冒号，又有英文的冒号，标点符号不统一。 | G-王康明 | 梁远志 | 接受，已改用全角冒号 | 认可 |
| 6 | 3.3.1  Scrapy框架 | “内置的中间件及扩展为下列功能提供支持:Cookies和session处理”这里面使用了一个英文的冒号，而全文除了网址之外的其他地方，都是使用中文的冒号，格式不统一。 | G-王康明 | 赵正阳 | 接受 | 认可 |
| 7 | 3.2.4.3 | “然后需要再数据库中获取该任务所在的服务器节点”，这句话里的“需要再”含有错别字，或是语义不明。 | G-王康明 | 沈一聪 | 接受 | 认可 |
| 8 | 3.3.3 后端  （7）应对反爬虫 | “场景：增加动态网站，数据通过JS动态加载，增加网络分析复杂度；”内容不完整，“；”之后就没有内容了。 | G-王康明 | 沈一聪 | 接受，标点符号已修正 | 认可 |
| 9 | 3.3.3 后端  （7）应对反爬虫 | 应对反爬虫的很多个场景，以及其解决方案更多像是功能性需求，而非非功能性需求。 | G-王康明 | 沈一聪 | 部分接受，这个问题在上次评审中已经将解决方案等容易引起误解的说法删去，但有一处疏漏，现已删去已避免引起误解。我们经过反复讨论还是觉反爬虫并不是一个用户可选的，必要的功能，而是保证软件功能完整性和可靠性的一个需要。 | 认可 |
| 10 | 3.3.3  后端 | 内容不一致，可能有些冲突。  3.6 故障处理  里提到“当Scrapyd节点的网络故障导致爬虫作业运行中断时，管理员应在1小时内检查并排除网络故障，并恢复Scrapyd节点的运行。当服务器硬件故障导致本平台无法访问时，管理员应在2小时内将系统重新部署到备用服务器上。”  而3.3.3 后端 “如果出现故障，要求在2小时内将系统重新部署到备用服务器。”  3.6的意思是对两类不同的故障，要求分别作出两类不同的应对措施；3.3.3的意思是对所有的故障，要求只做出同一个应对措施 | G-王康明 | 赵正阳 | 接受 | 认可 |
| 11 | 2.3.1 | 框架图中无步骤9图例，而步骤9是重复1操作 | G-刘佳恒 | 赵正阳 | 接受 | 认可 |
| 12 | 3.3.2 | “对于Javascript而言”，大小写应与前面统一 | G-牟秋宇 | 宋冰晨 | 接受 | 认可 |
| 13 | 3.1.2（2） | “开放了”不适合作为需求的措词 | G-张祥国 | 宋冰晨 | 接受 | 认可 |
| 14 | 3.2.1 | “爬虫程序开发者”与前面的参与者类别不对应 | G-张祥国 | 赵正阳 | 接受 | 认可 |
| 15 | 3.2.1 | 用例图中，“处理项目”与“存储项目”、“丢弃项目”间不知道extend关系是否合适。因为似乎“处理”和“丢弃”、“存储”间，更加类似于继承。  同样“系统设置”与“增加设置”、“读取设置”间也是如此。 | G-潘安佶 | 赵正阳 | 部分接受：“处理项目”是一个具体的功能（调用Pipeline.process\_item方法），但可能不会执行“存储项目”和“丢弃项目”中的任何一个（例如仅输出日志），符合“可选执行”的特点，因此将其作为extends关系；“系统设置”确实不是一个具体的用例，已将其删除 | 认可 |
| 16 | 3.2.1 | 对于用例图，可以考虑导出矢量图，或是visio直接复制，可以让图片更加清晰。 | G-潘安佶 | 赵正阳 | 部分接受：已尝试导出矢量图，但未注册版的StarUML会给图片加上"UNREGISTERED"背景 | 认可 |
| 17 | 3.2.2.1 | “以及节点名称作为节点的标识”与前文衔接不通顺 | G-张祥国 | 梁远志 | 接受，已删去该描述 | 认可 |
| 18 | 3.2.4.2 | “系统可以”和“MYSQL”之间好像缺少“从” | G-潘安佶 | 沈一聪 | 接受 | 认可 |
| 19 | 3.2.2.2 | “列出可用爬虫”在RUCM中是英文，文字中是中文 | G-张祥国 | 梁远志 | 接受，已将所有用例名称改为中文名称 | 认可 |
| 20 | 3.5 | 在UML数据库图中，“Task”、“Job”、“Node”都使用了“状态”对其内容进行了概括，不知道是否需要对其内容进行细化说明。 | G-潘安佶 | 梁远志 | 接受，已在图下方加上了状态的说明。此处的状态都是原子性的概念，已将可能的状态列出。 | 认可 |
| 21 | 表1.1 | 负责处理Spider中获取到的Item，并进行进行后期处理（详细分析、过滤、存储等）中两个”进行”重复 | G-马广洲 | 赵正阳 | 接受 | 认可 |
| 22 | 表1.1 | “处理发送到Spider的响应和从Spider生成的请求和项目过程中的钩子框架，用于在其中插入自定义功能”较为赘述 | G-马广洲 | 赵正阳 | 接受 | 认可 |