|  |  |
| --- | --- |
| **2020年短学期项目实训小组会议记录** | |
| 时间 | 2020.7.3 09:00 |
| 地点 | 钉钉视频会议 |
| 与会人员 | 应承峻、张佳瑶、贺婷婷、杨建伟、戴陈威 |
| 会议内容 | 1.讨论了《需求规格说明书》中数据字典和ER图等数据部分的内容，明确这些内容需要等待数据爬取的结构化信息以及接口确定后统一编写。  2.讨论了系统的数据流图的设计，应分为客户端、服务器、搜索引擎服务器等，明确绘制的时候需要注意正确表达。  3.讨论了数据爬取所需的结构化信息以及一些非结构化信息，爬取这些信息对应的网站或者数据库，以及如何爬取这些信息。针对数据爬取的数据量，结合时间的限制，进行了折中考虑。  4.讨论了结果聚类多维排序在搜索结果页面的表现形式，明确了对于搜索结果的列表页面，左侧设置针对电影、影评、资讯、解说、资源等栏目以实现对于搜索结果的筛选；  5.明确了爬取数据的ddl，在7.6之前基本实现完整的爬取代码，先将爬取的部分数据分享给前后端进行功能设计，后续可以持续爬取。对于数据分析等其他深度搜索内容，明确了在7.10之前完成实现。  6.讨论了极限挑战中深度搜索技术的可行性，明确了对用户查询的意图进行分析，实现查询分发与结果整合以及结果聚类、关联分析的多维排序以及热度等信息的可视化。我们也讨论了对非结构化文档进行结构化信息抽取，例如影评的关键字提取生成词云。  7.讨论了图片搜索与识别的可能性以及实现方法，目前暂时列为待定的需求，考虑从人的维度进行内容抽取。  8.明确了今天前端开发的任务是基本开发出界面原型图，先由爬取数据的同学绘制参考图，前端初步设计完成后，后续讨论进行模块的增删改。 |
| 备注 | 第四次组会 |
| 记录人 | 戴陈威 |