

社交网络技术与应用--大作业要求

1、项目安排情况

总分占期末成绩的 40%，小组合作

项目演示日期：2025-06-04，第五、六节课

项目演示时间：每小组报告需在 5 分钟内完成（请严格控制时间，时间到了直接停止）

演示之前需要先提交小组的报告，演示的顺序按照 QQ 群里的在线文档“分组情况”顺序

（如果要调整顺序，请和其他组协商调换之后发信给任课老师或者助教进行调整）

报告模板见课程网络平台上的“社交网络技术与应用—项目报告模板”

2、项目开题（5 分）

学生先进行 15 分钟的开题研讨，具体研讨内容如下：大作业主题选择讨论（需要与教师核实）；分工讨论（根据大作业要求完成每个成员具体任务分工）；预计完成进度计划讨论（根据大作业时间节点完成阶段性计划表；完成开题汇报 PPT）。

PPT 内容要求如下：

第一页内容：小组名称

第二页内容：项目的选题背景描述以及任务分工

第三~六页内容：每个成员展示自己负责子任务的技术框架图

第七页内容：项目预计完成时间计划表

注意：项目总体展示时间每组不能超过 5 分钟

3、项目具体要求

1) 收集主题社交媒体问答数据（共 2 分）

可自行选定社交媒体网站，例如豆瓣，微博，知乎等，爬取大作业主题相关的问答数据（2 分）

要求至少爬取 10 万条数据

2) 数据预处理与存储（共 3 分）

a) 问答数据处理

数据预处理（中文分词、去停用词）与关键词提取（1 分）

数据索引（1 分）

b) 数据存储

将处理后的数据存储进数据库设计好的表里（1 分）

3) 实现社交媒体跨模态问答系统（共 13 分）

a) 在网站前端实现聊天窗口（1 分）

在页面上实现一个聊天窗口，允许用户进行问答（1 分）

b) 实现文本问答功能（5 分）

问题理解（1 分）

答案检索或推理（2 分）

基于提供的测试集分析问答算法的优劣（2 分）

c) 实现视觉问答功能（7 分）

视觉问题理解（1 分）

基于图像生成对应描述（2 分）

答案检索或推理（2 分）

基于提供的测试集分析问答算法的优劣（2 分）

4) 实现文档处理与分析系统（共 12 分）

a) 在网站前端聊天窗口实现提示功能（1 分）

- 实现提示语句，允许用户进行文档处理与分析（1分）
- b) 实现文档处理与分析功能（6分）
 - 允许上传文档（1分）
 - 对文档内容进行词性标注并显示部分结果（1分）
 - 对文档内容进行实体识别并显示部分结果（1分）
 - 实现文档摘要功能并显示摘要（2分）
 - 允许用户下载结果文件（1分）
- c) 实现文本挖掘功能（5分）
 - 允许上传数据集压缩包（1分）
 - 对数据集进行文本聚类并显示 t-SNE 的结果，分类显示词云图结果（4分）
 - 允许用户下载结果文件
- 5) 实现复杂文档处理系统（共 5 分）
 - a) 在网站前端聊天窗口实现提示功能
 - 实现提示语句，允许用户进行复杂文档处理
 - b) 实现文档处理与分析功能（5分）
 - 允许上传复杂类型文档（doc、pdf、png 等）
 - 对复杂文档内容进行布局分析（图像、表格、文本、公式的布局）（1分）
 - 对文档内容进行至少 2 种类型提取并显示部分结果（2分）
 - 根据布局和提取结果转换成用户想要的文档（2分）
- 6) 最后的报告和演示（5分）
 - 第一页内容：项目名称、小组名称
 - 第二页内容：项目组成员具体分工
 - 第三~六页内容：每个成员展示自己负责子任务的具体实现步骤、运行结果（可以添加必要的截图来辅助说明）
 - 第七页内容：项目演示与项目总结（可简单演示一下整体系统，简述项目实施过程中遇到的难点以及克服的过程）
 - 项目总体展示时间不能超过 5 分钟
- 4、项目可选主题
 - 1) 社交网络舆情分析问答
 - 2) 国内外公开武器装备知识问答
 - 3) 国内外领导人相关事件问答
 - 4) 金融知识问答
 - 5) 医疗知识问答
 - 6) 国家关于某个领域政策知识问答，例如，科技创新、绿色低碳发展、产业升级与现代化、区域协调发展、社会民生与公共服务、国家安全与国防现代化、改革开放与制度创新、文化繁荣与软实力提升、基础设施建设以及乡村振兴与农业现代化等
 - 7) 其他和任课老师商定的主题

注意：

- 1. 开题需要小组每个人的参与和努力，没有参与开题的同学将以 0 分计开题分数；
- 2. 最后的报告和演示 5 分是以个人进行打分，每个同学都需要展示和汇报自己的工作，未展示和汇报的同学将以 0 分计；
- 3. 所有源代码、数据和结果文件夹均放到命名为“附件材料”的文件夹里。最后将开题报告 PPT、项目报告、项目演示 PPT 和附件材料文件夹一起放到一个命名为

“project_小组名”的文件夹并打包成命名为 project_小组名.zip 的压缩包提交到课程中心平台上（每个小组提交一份即可）。项目最终材料提交期限：2023-06-23，23:59pm；

4. 如果项目文件过大，课程中心平台无法上传，请提交到 ftp 上的“大作业”文件夹里。