



# 社交网络技术与应用

## 大作业要求

# 项目安排情况

---

- 总分占期末成绩的**40%**， 小组合作
- 项目开题研讨日期
  - ❖ **2025-04-16**, 第五、六节课
- 项目演示日期
  - ❖ **2025-06-04**, 第五、六节课（有时间冲突的同学请尽快反馈）
- 项目演示时间
  - ❖ 每组不能超过**5分钟**（时间到了直接停止）
  - ❖ 注意控制时间，没有演示完将会影响分数
- 项目最终材料提交期限
  - ❖ **2025-06-23, 23:59pm**
  - ❖ 请将开题报告PPT、项目报告、期末汇报ppt、源代码、数据文件和结果按照小组提交到课程平台/ftp

# 项目开题研讨-开题讨论

---

□开题讨论时间为：5分钟

- ❖ 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）

□完成开题报告PPT

□大作业开题（5分）

# 项目开题研讨-开题演示

---

- 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： 总体展示时间每组不能超过5分钟
- 大作业开题（5分）

# 项目具体要求一

---

## □ 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

## □ 数据预处理与存储（共3分）

### ❖ 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）

### ❖ 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）

# 项目具体要求二

---

## □ 实现社交媒体跨模态问答系统（共13分）

### ❖ 请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，允许用户进行多模态问答（  
1分）

### ❖ 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）

### ❖ 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）

# 项目具体要求三

---

## □ 实现文档处理与分析系统（共12分）

### ❖ 请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）

### ❖ 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）

### ❖ 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

# 项目具体要求四（选做模块）

---

## □ 实现复杂文档处理系统（共5分）

❖ 请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

❖ 实现复杂文档处理功能（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）

允许用户下载结果文件

# 项目报告与演示（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，简述项目实现过程中遇到的难点以及克服的过程）
- 注意：项目总体展示时间不能超过**5分钟**



谢谢！