

知识图谱课程项目

说明：围绕“知识如何表示和提取”这一问题，从论文、教材、标准协议文档（如 HTTP 协议的文档）、编程代码库和依赖、（或者其他任何感兴趣的领域，如梗、up 主等）中任选一个领域，定义本体、规则、schema 等并给出实际的例子。构建相应的知识图谱，并探索其在下游任务中的应用。完成一份技术报告，具体包括以下内容：

1. 组队安排：组长、组员、每位同学的研究方向、所在实验室等。最多 5 人一组。
2. 选择什么领域？
3. **该领域中有哪些知识？**
4. **如何表示这些知识？**
5. **如何利用大模型来帮助构建和抽取知识？**
6. 具体方案设计：
 - a) 如何获取或构造数据集，得到知识图谱？
 - b) 以下任务都需要考虑：
 - i. **知识抽取：**考虑到有些隐性、复杂知识很难抽取，那么如何有效地抽取这些知识？采用什么方法进行抽取？（可探索通过人为设计或由大模型自主规划的 Chain-of-Thought，引导大模型推理以提升抽取效果。）
 - ii. **下游应用：**选择什么下游应用？注意需要使用构造的知识图谱。（可探索多种形式的知识图谱与大模型结合方式，例如设计 workflow 流程或由大模型自主调用知识图谱。）

说明：技术报告需完整介绍整个工作。按论文格式，包含摘要，引言，数据集与知识图谱的构造，任务定义，方法，实验（针对知识抽取任务），原型系统（demo 的使用展示，针对下游任务），结论，参考文献，人员与工作安排，组内互评等部分，**除参考文献外字数不少于 5000 字。**

提交内容：汇报 PPT（需要保证内容完整，包含必要的文字说明、举例等）、完整的技术报告（Word/PDF 格式）、数据集、模型和/或 demo 的相关代码。

数据集规模：实体/属性数量不少于 2000，关系数量不少于 1000。需要确保所选用的算法在该数据集上能够产生有意义的输出。

材料提交截止时间：12 月 31 日 23:59

汇报时间：12 月 25 日