

可选题目

项目一：三国演义知识图谱

简介：可包含人物、国家、地点、战役等实体

项目二：哈利波特知识图谱

简介：可包含人物、学院、魔法、组织机构、物品等实体

项目三：漫威宇宙知识图谱

简介：可包含人物、组织机构、地点、装备、重大事件等实体

项目四：DC 宇宙知识图谱

简介：同漫威宇宙

项目五：自行确定一个领域构建知识图谱

简介：军事、医疗、金融和事理知识图谱除外

项目需求

- (1) **知识抽取**：从网页或者其他来源获取文本，实现从文本中抽取实体，实现关系分类；选做：爬虫。
 - a) 可利用的文本来源如：对应内容的 wiki 网站（如 https://harrypotter.fandom.com/wiki/Main_Page）、小说、介绍性文章、通过搜索引擎搜索得到的内容等。
- (2) **知识存储**：任选关系数据库、内存数据库或者图数据库。
- (3) **向量表示与推理**：实现 TransX 系列算法或其它你觉得更好的算法来学习节点和边的语义向量表示。并基于向量表示实现知识推理：给定一种关系和一个头（尾）实体，预测满足关系的尾（头）实体。

项目要求

- (1) 自由组队，每队 2~4 人，任选一项目
- (2) 每个项目最多允许 8 个队伍
- (3) 提交内容：项目报告、讲解视频、代码和数据
 - a) 报告内容需包括：数据来源；各部分的关键技术、关键代码说明、结果与分析；项目分工
 - b) 讲解视频：讲解和演示代码运行，说明如何运行、输入输出
- (4) 提交日期：2024 年 6 月 20 日
- (5) 提交方式：网盘提交，后面会公布提交链接

其他说明

使用中文数据还是英文数据：
都可

可否使用大模型：

可以，但项目需求中的向量表示部分必须实现，且实体抽取和关系分类两项任务不能都只用基于大模型的方法。

可否使用基于规则的方法：

可以，但得分很可能比基于神经网络的方法低

训练数据怎么获取：

可考虑的获取方式：利用公开数据集、自行标注、使用基于规则或远程监督的方法标注

效果不好怎么办：

最终效果不是评分重点。如果效果不好，可在报告中分析原因。

是否需要实现实体消歧或实体链接：

可简单处理，无需实现相关技术。例如，直接手动将“哈利”改为“哈利波特”。

可否参考网上的代码：

可以，需要在报告中对关键代码进行说明