# 大规模知识图谱增强语言模型系统

指导老师：张文

## 系统背景:

知识图谱通过结构化的方式表示实体及其之间的关系，能够为语言模型提供丰富的、上下文相关的背景信息，从而增强模型的理解与推理能力。近年来，语言模型如GPT-3、BERT等在生成、分类和问答等任务上表现出色，但它们通常依赖大规模数据训练，缺乏结构化知识的引导，容易产生事实错误或缺乏对实体之间关系的深度理解。

为了解决这些问题，研究人员尝试通过将知识图谱嵌入语言模型，使得系统不仅能够生成高质量的自然语言文本，还可以在回答问题时准确地利用知识图谱中的信息。这种结合能够显著提升语言模型的知识性、准确性和可解释性，特别是在知识密集型任务中，如医学、法律、科研等领域。通过知识图谱，系统可以获得更强的推理能力，实现从简单的文本生成向复杂问题解决的转变。

本实训拟开发大规模知识图谱增强的语言模型系统，借用大规模知识图谱中包含的丰富知识数据，提升语言模型的生成效果。

## 主要功能

大规模知识图谱增强的语言模型系统的主要功能有：

1. 给定一个自然语言输入，根据当前数据从大规模知识图谱中检索相关知识；
2. 将检索得到的相关知识以合适的方式输入语言模型中

大规模知识图谱增强的语言模型系统的核心模块包括：

1. 大规模知识图谱存储模块
2. 大规模知识图谱检索模块
3. 检索信息序列化模块
4. 生成模块
5. 前端界面

## 开发环境

编程语言：后端Python等，前端javascript等