2023年秋Java期末大作业

1、训练/微调一个机器学习（SVM）和一个深度学习（BERT）文本分类模型，然后将模型保存到本地，搭建B/S模式系统，用户通过前端输入一个新的文本（需要支持文本框输入和数据库导入两种模式），系统调用模型对文本进行分类，将分类结果在前端呈现给用户，注意用户可通过下拉框等方式自行选择调用哪个模型。

【数据可以但不限于用现成的电影评论文本集 [MOVIE REVIEWS IMDB (kaggle.com)](https://www.kaggle.com/datasets/shrutinikkigupta/movie-reviews-imdb)】⭐⭐⭐

2、搭建一个B/S模式系统，为系统配置一个大模型（ChatGLM-6B或ChatGPT等），允许用户通过前端输入一个新的文本（需要支持文本框输入和数据库导入两种模式），可以对文本实现两个操作，一是可以对文本进行翻译（中译英，英译中），二是可以对文本进行事件三元组抽取，操作结果通过前端呈现给用户。

【可参考：https://github.com/binary-husky/gpt\_academic】⭐⭐

3、搭建数据库系统，允许用户通过前端界面将TXT、CSV、XLSX等格式的数据上传到数据库，上传时允许用户选择特定字段；上传完毕后，允许用户通过前端界面进行检索，再将检索结果导出为TXT、CSV、XLSX等格式，导出时允许用户进行定制化导出（例如只导出特定几个字段）

【具体数据库表和字段后期由老师提供】⭐

4、搭建一个B/S模式爬虫系统，**动态追踪**给定的网站（暂定BBC、CNN、搜狐新闻、企查查，后期根据情况调整），将每天更新的与中国（美国）相关的新闻信息保存到数据库（包括新闻题目、正文、发布时间等），并通过邮件提醒用户今日更新了多少条，同时允许用户通过前端界面检索和浏览爬取的新闻和企业信息，用户对某新闻或企业信息感兴趣时，可以将浏览页面导出为PDF。

【在BBC、CNN追踪与中国相关的新闻，在搜狐追踪与美国相关的新闻，在企查查追踪与老师指定的主题、机构相关的企业信息（可尝试调用API）】⭐⭐⭐⭐

5、基于Lucene或ElasticSearch搭建一个B/S模式信息检索系统，**动态追踪**解析xml文档（联邦公报）并存储到数据库，调用百度或Google等翻译软件的API翻译文本，形成中英文两个版本文本后，进而基于前述框架实现中英文全文检索。

【[Bulk Data - Federal Register | GovInfo](https://www.govinfo.gov/bulkdata/FR/2023/11)】⭐⭐

6. 老师给定科技文档MySql数据表，基于该数据表建立术语之间的关系网络。在关系网络的基础上，基于Lucene或ElasticSearch搭建一个B/S模式信息检索系统，实现全文检索和术语关系展示。

【数据表由老师后期提供】⭐⭐

7. 老师给定一个科技术语词集，查询维基百科、百度百科等网络词典，获得该科技术语的解释文本，对解释文本进行术语抽取，找到与给定术语相关的术语，并构建术语之间的关系网络。在关系网络的基础上，基于Lucene或ElasticSearch搭建一个B/S模式信息检索系统，实现全文检索和术语关系展示。

【科技术语集由老师后期提供】⭐⭐⭐

模型的训练、微调等可利用Python等语言，系统的构建需要使用Java，具体来说即构建Java Web系统。