# **2023级本周 周报**

|  |  |
| --- | --- |
| **年级** | 2023级 |
| **时间** | 2024.1.20-2024.1.31 |
| **本周组长** | 郑健敏 |

**成员信息汇总**

|  |  |
| --- | --- |
| **姓名** | **吴中阳** |
| **本周工作** | 确定了基于知识图谱的旅游问答系统设计与实现这个方向，目前看论文看别人的实现步骤，但数据集和具体的旅游问答范围并没有确定。  之前是看了几篇藏文命名实体相关的论文以及知识，这周看了几篇知识图谱的问答系统的相关论文  主要步骤大概是  1.设计旅游知识查询模板并构建用户问句训练集  2.爬取相应数据集，进行标注，构建数据集（或者找是否有对应的数据集）  3.数据清洗，词性标注，提取文本特征  4.采用基于 BiLSTM-CRF /Bert\_bigru\_crf/Bert\_textcnn的命名实体识别模型，理解用户问句语义信息  5.依存关系解析，信息整合与查询构建  6.知识图谱查询，回答生成 |
| **下周工作计划** | 1.学习相关模型算法  2.了解提取文本特征的相关方法  3.继续看相关论文 |
| **姓名** | **顾佳凯** |
| **本周工作** | 广西民族大学聊天机器人项目： 完成了项目的PPT制作。  该项目旨在构建一个高效且互动性强的聊天机器人，用于提供自动化的客户服务和查询回答。  a.基础技术： 决定采用Mistral AI作为聊天机器人的基础大语言模型（LLM），并加之特定数据集进行微调，以期获得较好的聊天效果。  b.技术框架： 选用了LangChain进行Mistral AI的调用和集成，以便更好地管理和优化机器人的对话能力。  c.前端界面： 计划使用Gradio来构建用户友好的前端界面，以提高用户体验。 |
| **下周工作计划** | 争取在寒假期间学完《动手学深度学习》。 |
| **姓名** | **郑健敏** |
| **本周工作** | 1. 延续看深度学习，关于Tramsformer深度学习模型框架，重点理解了自注意力机制、编码器和解码器架构以及实现的全过程。 2. 看了一篇关于依存句法分析的文章，是一种研究词汇之间的依存关系的方法，如主谓、动宾等等。然后构建一个依存树，是一种有向无环图。文章还对四种句法分析模型进行了对比分析。 3. 目前研究方向是模型对中文理解的准确度，利用语法、语义等方面知识构建模型来提高依存的正确性。 |
| **下周工作计划** | 1. 继续看相关论文 2. 研究语句的切分，并用代码实现，为后续研究机器翻译做准备。 |
| **姓名** | **陈卓** |
| **本周工作** | 在确定智能问答作为我的研究方向后，这两周我巩固了这学期所学的nlp相关的知识，查阅了与大语言模型有关的资料，比较了目前比较火热的几个大语言模型。随后我综合考虑了自己目前的实践能力，定下了自己之后的项目方向及大致计划。  项目方向：  做智能问答系统，期望这个系统可以简单的回答日常生活中的各种问题，以及回答一些简单的学术问题。  开发此系统的大致计划：  一，大语言模型的选择：  选用Llama2作为自己智能问答系统的大语言模型。  二，应用开发框架的选择：  我选择使用LangChain作为自己的应用开发框架。  三，环境部署：  通过Anaconda部署Llama2，通过LangChain集成  Llama2，尝试使用steamlit编写友好交互界面，若实  在不行，就用黑框执行  四，数据集和知识库：  还在学习中 |
| **下周工作计划** | 1. 学习Llama2，LangChain等相关环境的部署 2. 学习LangChain的各种编程接口并尝试实战 3. 寻找训练数据集，学习知识库的构建 |
| **姓名** | **吕佳航** |
| **本周工作** | 想法：  基于大语言模型实现AI阅卷。  背景：   1. 文科学科的大题一般都是简答或者论述题，只要学生的答案中有涉及到得分点就给分，即答案围绕某些关键词或主题。 2. 大语言模型能够理解人类对相同问题的不同表述。   应用场景：   1. 现实中每个人在答题时的论述不尽相同，使用大语言模型对答案进行初步评判，可以节省教师阅卷时间。 |
| **下周工作计划** | 1. 研究方案可行性和实用性 2. 寻找方案实现方法 |
| **姓名** | **王纯熙** |
| **本周工作** | 确定了基于知识图谱的智能问答系统的研究方向，目前应用场景初步定为电影行业，并暂定了实现步骤（会结合后续看的论文更新）。这周大致看了BILSTM算法及几篇与本主题相关的论文。  1.设计电影知识查询模板并构建用户问句训练集  2.提取文本特征  3.爬取电影类新闻数据，标注电影名称命名实体语料数据集  4.采用基于 BiLSTM-CRF 的命名实体识别模型  5.采用Flask轻量级框架搭建web  （数据库没想好用图的还是用表的） |
| **下周工作计划** | 1.学习CRF算法  2.了解提取文本特征的相关算法  3.继续看相关论文 |