**电 子 科 技 大 学**

**2018级本科毕业设计（顶岗实习）任务书**

**拟题单位**\_\_\_北京字节跳动科技有限公司\_\_**审题人（校内导师签名）**\_\_\_\_\_\_\_

**题目及副标题**\_基于BERT和BM25的智能问答系统设计与实现\_\_\_

**课题来源:**  1.科研 √2.生产 3.创新创业。

**主要任务及支撑毕业要求指标情况：**

问答系统(Question Answering System, QA)是信息检索系统的一种高级形式，它能用准确、简洁的自然语言回答用户用自然语言提出的问题。其研究兴起的主要原因是人们对快速、准确地获取信息的需求。问答系统是人工智能和自然语言处理领域中一个倍受关注并具有广泛发展前景的方向。现要求基于BERT和BM25算法，设计并实现一个智能问答系统，并要求该软件系统能够实现对实现文本的深度理解，以及能满足用户的问答需求。

|  |  |  |
| --- | --- | --- |
| **主要任务** | **支撑毕业**  **要求指标** | **考核点** |
| 对文本问答系统的前沿技术，包括Query分析，文本理解，搜索模型等进行分析。 | GR10.4 | 对毕设课题相关的国际前沿技术进行分析。 |
| 基于现有技术条件和开发周期分析实现需求的可行性。 | GR3.2 | 从技术、经济等角度进行可行性分析。 |
| 制定问答系统开发计划。 | GR11.3 | 制定项目计划。 |
| 对问答系统进行总体设计以及各个系统模块的详细设计。 | GR3.3 | 进行总体设计和详细设计。 |
| 针对Query分析，文本理解，搜索模型等具体问题自学相关算法知识；针对请求转发，实时部署等具体问题学习后端知识 | GR12.3 | 针对具体工程问题，能够通过自学掌握相关知识。 |
| 针对模型的实时性保障问题提出解决方案，并能通过分析文献提出多种解决方案，能对多种方案进行分析。 | GR2.3  GR2.4 | 针对具体工程问题，能通过文档描述解决方案；能分析文献寻求可替代的解决方案并证实方案合理。 |
| 基于模型的实时性保障问题的解决方案进行流程设计，提出不同设计方案并进行优选。 | GR3.4 | 针对具体工程问题进行流程设计，并能对不同设计方案进行优选。 |
| 选择Linux操作系统，Tensorflow、sk-learn深度学习框架、Flask后端框架，MySQL数据进行基于机器学习的智能问答系统的软件模型开发。 | GR5.2 | 选择合适的开发环境、工具与技术标准进行软件开发。 |
| 针对模型的质量评估问题，参考业界的统一评估指标，进行评估。 | GR4.2 | 为了解决具体工程问题，能够进行实验，并能对实验结果分析。 |
| 项目完成后提出改进方案，并分析该项目可能带来的社会、经济影响。 | GR7.2 | 评价项目的后续改进以及可能的对社会、经济的影响。 |
| 制作答辩PPT、现场答辩、回答答辩小组的问题。 | GR10.2 | 能够进行陈述发言，清楚表达对复杂软件工程问题的看法与见解。 |

**预期成果或目标：**

软件能正常运行，功能较完善，系统较为稳定可靠。

**成果形式：硬件、硬件＋软件、√软件（选择其中一种）**

**起止日期： 2021年10月20日至2022年5月31日**

**学生姓名\_\_ 廖梓尧 \_\_ 专业\_ \_\_软件工程\_ \_学号\_\_2018091620013\_\_\_\_**

**指导单位\_\_\_\_\_\_\_\_**北京字节跳动科技有限公司**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**校外导师姓名、职称\_\_\_\_\_\_于光华 高级工程师\_\_\_\_ \_\_\_\_\_**

**校内导师姓名、职称\_\_\_\_\_\_吴劲 副教授\_\_\_\_\_\_\_ \_\_**

**设计地点\_\_\_\_\_\_\_\_\_\_\_\_\_北京市海淀区知春路甲48号\_\_\_\_\_\_\_\_**

**2021年11月12日**

**备注：**1. 此任务书应由校外导师、校内导师协助学生填写。

2.此任务书应在学生毕业设计前下达给学生。

3.此任务书必须双面打印。