**软件工程实训题目申报**

|  |  |  |  |
| --- | --- | --- | --- |
| **课题名称** | 面向互联网文本的语义知识智能分析实验系统构建 | | |
| **指导老师** | 刘小明 | **联系方式** | 电话:15617875707;QQ:1028435020; Email: ming616@zut.edu.cn |
| **要求** | 1. 5-6人小组，设 项目负责人、需求分析、技术架构、质量保证、配置管理，(兼职) 2. 掌握软件开发的基础知识与实践技能，具有良好开发习惯与编码风格； 3. 自觉性强，能够自我激励与自我管理；遇到困难百折不回的精神； | | |
| **课题来源** | **纵向科研项目 ☑** | | |
| 项目的名称：基于关联数据的信息检索关键技术研究  项目编号：61672361  项目类型：自然科学基金面上项目  项目来源：国家自然科学基金委员会 | | |
| **横向合作项目 □** | | |
| 项目名称：  合作机构：  附证明材料(如项目合作协议或合同)； | | |
| **自拟项目 □** | | |
| 项目名称：  项目类型：竞赛 ☑； 先进技术 □； 社会实际应用 □； 创新创业 □；  其他 □ | | |
| **背景意义** | 互联网文本中包含有大量的语义知识信息，对这些大数据量的互联网文本进行智能化地分析，从中提取出关键词、命名实体、实体关系、语义角色等关键性信息，是对其进行理解的必要技术手段；  当前，自然语言处理相关理论技术，对规范性较好的书面性文本文的处理性能较好，但对大数量的不规则、杂乱的互联网文本的处理难以满足对这些文本所蕴含语义知识内容深入理解的需求。虽然已有学者对各个特定任务进行研究，取得了一定的效果，但还没有能够以统一的方式一体化的解决该问题的理论技术。  因此，众多领域的研究和应用（如互联舆情分析、网络）都迫切需要一种统一的针对不规范大数据互联网文本的语义知识智能分析框架，此课题将这些科研问题引入到本科高年级的实训课程中，以期本科生不仅能完成基本的科研实验验证程序的开发，并从中初步体会和感受科学研究的基本流程和方法，为今后的学习和深造打下良好基础。 | | |
| **主要目标** | 构建一个面向多种互联网文本（如 新闻、论坛、帖子、微博、微信）格式文本所蕴含的语义知识分析系统。前期以基于现有理论技术，完成初步实验结果（Baseline）；后期针对理论模型的改进方案，实现完成改进方案的实验验证； | | |
| **难点与挑战** | 1. 互联网文本不规范的特点分析； 2. 面向不规范文本的分析处理方案； 3. 统一的语义知识分析提取框架； | | |
| **预期成果** | 一期：  利用现有的规范化文本处理的理论技术，实现对互联网文本的初步处理，并分析存在的问题及原因，为后续进一步改进奠定基础；  二期：  基于一期发现的问题及其原因分析，提出一体化的解决方案，并开发实现试验验证，进一步分析所提解决方案的效果； | | |
| **验收标准** | 一期：基于现有理论技术的实验验证程序，并集成到统一开发平台，实验效果与分析；  二期：改进方案的实验验证程序，效果及分析； | | |