**软件工程实训题目申报**

|  |  |  |  |
| --- | --- | --- | --- |
| **课题名称** | 面向互联网大数据的分布式知识库智能构建 | | |
| **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** | 众多行业领域的研究和应用（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信息的知识库。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** | 构建面向互联网文本数据(新闻、论坛、帖子、微博、微信……)的语义知识智能感知与提取系统。前期基于现有理论技术完成初步实验结果(Baseline)，作为初步的信息化构建方案；后期针对理论模型的改进方案，实现智能化的改进并实验验证。系统主要包括：   1. 互联网信息的智能采集与存储子系统； 2. 一体化的智能检索子系统； 3. 面向不同行业领域的主题发现与分析子系统； | | |
| **难点与挑战** | 1. 互联网信息的智能化采集方案与实施； 2. 互联网信息的分布式检索方案； 3. 互联网语义知识的快速分析框架； | | |
| **预期成果** | 一期：利用现有的规范化文本处理的理论技术，实现对互联网信息的采集、存储和检索，以及主题发现和分析，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。申请相关专利和学术论文。 | | |
| **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：对既有方案进行智能化改进，开发智能化原型系统并实验验证其效果分析； | | |