**湖南科技大学**

**毕业设计（论文）开题报告**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 题 目 | 基于ElasticSearch的多国科学基金数据  检索平台 | | | 课题类别 | 设计☑ 论文□ |
| 作者姓名 | 沈晨彬 | 学 院 | 计算机科学与工程学院 | 专 业 | 网络工程 |
| 学 号 | 1605020326 | 指导教师 | 石林 | 开题日期 | 2020年2月24日 |
| **选题的目的和意义**  科学基金制是由出资人设置基金，采取自主申请、专家评审、择优支持的机制，资助特定科学技术研究的制度。这项制度在发达国家取得了显著成效，我国自1981年设立自然科学基金以来，各类各级科学基金对推动我国基础研究发展起到了重大作用。但是不同国家的科学基金数据彼此隔离，形成信息孤岛，不利于全球化背景下的科研合作、科研选题。随着国际范围内科学基金数据的不断增多,信息化管理要求越来越高,客观上要求对不同国家之间公开的科学基金进行整合。  本项目利用网络爬虫技术，结合网站搭建和数据分析，实现一个基于ElasticSearch的多国科学基金数据检索平台，有助于帮助科研人员找到新的方向和切入点；有助于保证科研方向在领域的独创性和先进性；有助于避免课题重复和冗余研究、经费浪费；有助于新加入的科研从业人员了解科研现状。  基于ElasticSearch的多国科学基金数据综合平台，用户可通过主题词、申报机构、项目负责人等不同的关键词，在多个国家和组织的科研项目数组中查询是否有重复研究，了解行业最新进展，学习借鉴类似科研项目的研究思路。并且可以根据特定研究方向，获取相关领域的历年研究走势，直观的了解研究方向的热度变化，如:立项数量、资助金额等。 | | | | | |
| **主要内容和拟解决的关键问题（根据任务要求进一步具体化）**  **主要内容：**  本课题主要实现一个基于ElasticSearch的多国科学基金数据系统，采用B/S架构，后端使用Python为主要语言采用其Django框架搭建，前端使用Vue构建页面，mysql作为数据库，ElasticSearch作为检索引擎。   1. 数据采集模块 2. 数据清洗模块 3. 检索模块 4. 用户交互模块   **拟解决的关键问题：**   1. 数据的采集 2. 将数据转为ES可用的数据格式和类型 3. 分词字典选择 4. 可视化处理 | | | | | |

|  |
| --- |
| **重点与难点，拟采取的研究方法、步骤、技术路线（或主要措施）1**  **重点与难点：**   1. 多平台数据的爬取 2. 采集到的海量数据处理 3. 指标的可视化处理   **拟采取的研究方法**   1. Python的web框架Django是开放源代码的Web应用框架，采用了MTV的框架模式，即模型M，视图V和模板，核心组件有用于创建模型的对象关系映射、为最终用户设计的完美管理界面、一流的 URL 设计、设计者友好的模板语言和缓存系统。 2. Vue是一个构建数据驱动的web界面的渐进式框架，其目标是通过尽可能简单的 API 实现响应的数据绑定和组合的视图组件。采用自底向上增量开发的设计，只关心视图层，可以搭建美观的界面。 3. ElasticSearch是一个基于[Lucene](https://baike.baidu.com/item/Lucene/6753302" \t "/Users/shenchenbin/Documents\\x/_blank)的搜索服务器，它提供了一个分布式多用户能力的全文搜索引擎，基于RESTful web接口。它能很方便的使大量数据具有搜索、分析和探索的能力。   **步骤：**   1. 使用爬虫获取各个国家公开的科学基金数据的原始数据 2. 预处理数据，将数据转存到ElasticSearch中 3. 界面设计与后台数据库导入 4. 测试系统的各模块功能是否实现   **技术路线（或主要措施）：**   1. 通过爬虫技术获取2013-2018年的各国公开的项目数据 2. 将mysql中的数据选取有用的指标同步到ES中 3. 使用Django框架搭建系统后端，Vue编写界面 4. 进行功能测试、集成测试、性能测试 |
| **进度安排**  1. 2020年2月下旬-3月上旬：编写开题报告；  2. 2020年3月中旬-3月下旬：项目调研、查找资料(计算机技术资料和相关应用领域资料)  3. 2020年4月上旬-4月中旬：系统总体设计和详细设计；  4. 2020年4月中旬-5月中旬：实现系统，并不断完善和改进系统；  5. 2020年5月下旬-6月上旬：系统测试、论文整理和准备答辩。 |
| **参考文献（资料）**  [1] 徐伟杰, 王挺, 薛婉婷. 基于ElasticSearch的搜索引擎设计与实现[J]. 智库时代, 2019, 191(23):232+244.  [2] 刘晓强.[基于ElasticSearch的车型搜索引擎在保险系统中的设计和实现](https://kns.cnki.net/kcms/detail/detail.aspx?filename=GZDN201905014&dbcode=CJFQ&dbname=CJFD2019&v=" \t "/Users/shenchenbin/Documents\x/kcmstarget)[J].电脑与电信. 2019(05):51-55.  [3] 韦立梅,苏兵.Django框架下Python网站开发过程综述[J].电脑与电信,2019(10):54-56.  [4] 王伊,王韶红,刘晋泽,吕佳宪.Vue.js与Django组合框架的网络社交系统单页面架构方案设计[J].信息技术与信息化,2020(01):121-123.  [5]钟旭.中国与美英德法日五国科学基金项目资助强度研究[J].中国科技坛,2010(05):145-150. |
| 导师意见（在选题意义、技术指标或研究内容、是否同意开题等方面提出具体意见）  研究内容具体详实，研究方法成熟可行，选题有一定实用价值，同意开题。  导师签名： 系主任审核签名：  年 月 日 年 月 日 |

备注：

1.开题报告是本科生毕业设计（论文）的一个重要组成部分。学生应根据毕业设计（论文）任务书

的要求和文献调研结果，在开始撰写论文之前写出开题报告。

2.参考文献按下列格式（A为期刊，B为专著）

A：[序号]、作者（外文姓前名后，名缩写，不加缩写点，3人以上作者只写前3人，后用“等”

代替）、题名、期刊名（外文可缩写，不加缩写点）年份、卷号（期号）：起止页码；

B：[序号]、作者、书名、版次、（初版不写）、出版地、出版单位、出版时间、页码。

3.表中各项可加附页。