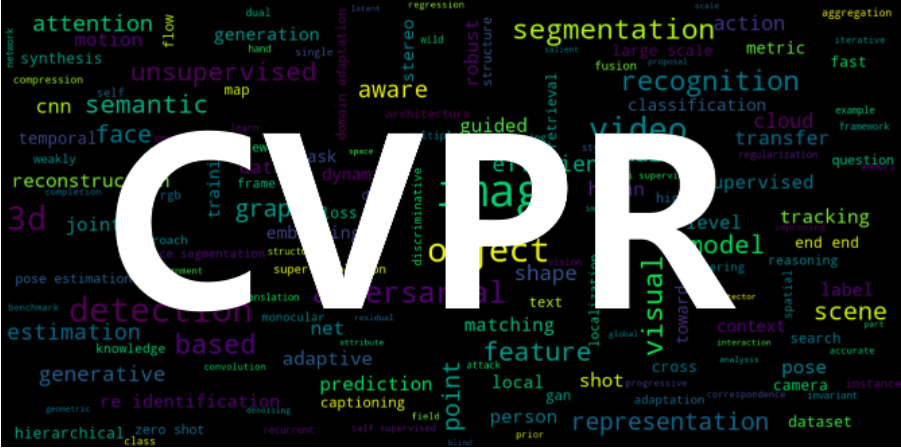
开题报告

顶会论文搜索和热词统计工具

东風何祺code组（黄宣祺 朱小东 廖一颿 何子昂）



1. 选题动机

计算机视觉技术的发展十分迅速，对于该领域的学者和研究人员而言，及时把握领域内的热点非常重要。本工具帮助学者了解近期ICCV, CVPR, ECCV三大顶级会议的热门话题，快速查找论文。

1. 功能
2. 爬取ICCV, CVPR, ECCV三大顶级会议近十年的论文的题目、摘要、关键词、原文链接。
3. 对爬取的信息进行结构化处理，分析top10热词，形成热词云图等直观查看方式
4. 可进行论文检索，当用户输入题目、关键词等关键信息，分析返回相关的论文、源代码、原文链接等信息
5. 对多年间、不同顶会的热词呈热度走势对比(这里将范畴限定在计算机视觉的三大顶会CVPR、ICCV、ECCV)
6. 可进行数据统计，例如每个国家录用文章的分析、每个学校录用文章的分析、哪个学校哪方面的研究方向比较强等。
7. 推荐：推荐热度高的论文
8. 技术路线&难点
9. **数据库**
10. 使用elasticsearch，便于搜索；

* Elasticsearch 是一个分布式搜索引擎, 底层基于 Lucene 实现。Elasticsearch 屏蔽了 Lucene 的底层细节，提供了分布式特性，同时对外提供了 Restful API。Elasticsearch 以其易用性迅速赢得了许多用户，被用在网站搜索、日志分析等诸多方面。
* 全文检索是指计算机索引程序通过扫描文章中的每一个词，对每一个词建立一个索引，指明该词在文章中出现的次数和位置，当用户查询时，检索程序就根据事先建立的索引进行查找，并将查找的结果反馈给用户的检索方式。维基百科的就使用了Elasticsearch的全文检索，可见其功能的强大。在本项目中，使用Elasticsearch引擎进行全文检索，可以方便的实现论文检索功能。

1. 使用redis缓存热词

* Redis是一个非常强大的分布式缓存引擎，是使用ANSI C编写的开源、支持网络、基于内存、分布式、可选持久性的键值对存储数据库。由于Redis基于内存，与在执行一个事务时将每个变化都写入硬盘的数据库系统相比就显得执行效率非常高，在本项目中，适合缓存热词。
* 在ASP.NET Core中使用Redis, 可以用到Microsoft.Extensions.Caching.Redis，是微软自己封装的Redis组件。

1. **爬虫**
2. Scrapy 是用 Python 实现的一个为了爬取网站数据、提取结构性数据而编写的应用框架。包括Scrapy Engine(引擎)，Scheduler(调度器)，Downloader（下载器），Spider（爬虫），Item Pipeline(管道)，Downloader Middlewares（下载中间件），Spider Middlewares（Spider中间件）。
3. **服务器**
4. 在本项目中，需要定时爬取三大顶会的论文，需要用到定时任务框架。**使用Quartz做定时任务，用于定时爬虫和更新数据库**。Quartz.NET是一个Github开源项目，用于创建一个免费商业用途的作业调度框架。支持 .NET Core。
5. 使用HTTP服务器**Nginx**部署前端项目**。**Nginx是一款轻量级的HTTP服务器，采用事件驱动的异步非阻塞处理方式框架，这让其具有极好的IO性能，时常用于服务端的反向代理和负载均衡。
6. **数据分析**
7. 通过统计论文的关键词频率来获得热词
8. **数据可视化**
9. 采用apache echarts。一个使用 JavaScript 实现的开源可视化库，可以流畅的运行在 PC 和移动设备上，兼容当前绝大部分浏览器（IE9/10/11，Chrome，Firefox，Safari等），底层依赖矢量图形库 [ZRender](https://github.com/ecomfe/zrender" \t "_blank)，提供直观，交互丰富，可高度个性化定制的数据可视化图表。
10. **前端**
11. Vue.js : 国内最流行的构建用户界面的渐进式JavaScript框架
12. Pinia：[Pinia](https://link.juejin.cn/?target=https%3A%2F%2Fpinia.vuejs.org" \t "_blank" \o "https://pinia.vuejs.org)是一款全新的状态管理库（下一个版本的Vuex）
13. Element Plus：基于 Vue 3，面向设计师和开发者的组件库
14. Vue Router：是[Vue.js](http://v3.vuejs.org/" \t "_blank)的官方路由。它与 Vue.js 核心深度集成，让用 Vue.js 构建单页应用变得轻而易举。
15. Vite：是一种新型前端构建工具，能够显著提升前端开发体验
16. Axios：是一个基于Promise 用于浏览器和 nodejs 的 HTTP 客户端
17. Less：是一门 CSS 预处理语言，它扩展了 CSS 语言，增加了变量、Mixin、函数等特性，使 CSS 更易维护和扩展。
18. TypeScript：是 Microsoft 开发和维护的一种开源的面向对象的编程语言，通过在JavaScript的基础上添加静态类型定义构建而成。它是 JavaScript 的超集，可以被编译成 Javascript；它包含了 JavaScript 的所有元素，并扩展了 JavaScript 的语法；它还可以载入 JavaScript 代码运行。
19. Jest：Jest是一个Javascript测试框架，由Facebook开源，致力于简化测试，降低前端测试成本，已被create-react-app、@vue/cli等脚手架工具默认集成。Jest主打开箱即用、快照功能、独立并行测试以及良好的文档和Api.
20. **接口管理与测试**
21. Swagger：Swagger 是一个规范和完整的框架，用于生成、描述、调用和可视化 RESTful 风格的 Web 服务。
22. Eolink APIKit：结合 API 设计、文档管理、自动化测试、监控、研发管理和团队协作的一站式 API 生产平台，从个人开发者到跨国企业用户，Eolink 帮助全球超过30万开发者和数万家企业更快、更好且更安全地开发和使用 API。
23. 成员分工

朱小东：

* 使用**Nginx**搭建和维护前端项目;
* 负责部分前端代码编写
* 使用**Jest**测试框架进行前端自动化测试

何子昂：负责前端部分代码编写

* Webpack打包

编写Axios http客户端

黄宣祺：负责后端部分代码编写

* 实现爬虫
* 服务器和数据库交互的部分（包括elasticsearch, redis, 定时任务)

廖一颿：负责后端部分代码编写

* 使用asp.net向前端提供RESTful接口
* 通过数据分析得到热词