浙江大学计算机科学与技术学院

Java 程序设计课程报告

2020—2021 学年秋冬学期

题目	图书搜索引擎
学号	
学生姓名	
所在专业	
所在班级	

目　录

1 引言

本次开发的是一个图书搜索引擎，这是一个综合性的题目，可以对 Java 语言中的各项功能有更好的理解和使用，通过具体的程序来加深对 Java 语言的掌握，提高自己的编程水平，为以后的工作打下一定的基础。

1．1 设计目的

图书搜索引擎是一个 web 爬虫的典型案例。本文使用 Java 语言编写一个与其类似的扫雷游戏。具体功能如下：

(1) 程序可以在当当网进行图书信息爬取，包括书名，价格，作者，封面，出版社，内容简介，目录等信息。

(2) 程序会对所爬取的图书信息建立索引，包括书名，作者，价格，出版社，内容简介，目录。

(3) 用户可以通过选择索引项并输入相关的关键词进行查找。

1．2 设计说明

本程序采用 Java 程序设计语言，在 IntelliJ IDEA 平台下编辑、编译与调试。具体程序由 1 人组成的小组开发而成。小组成员的具体分工如表 1 所示：

表 1 各成员分工表

成员名称	完成的主要工作	
	程序设计	课程报告
康锦辉	负责整个程序前期的需求分析和整体功能的架构 程序中爬虫部分的实现 程序中索引部分的实现 程序后期的测试与运行	报告的全部内容

2 总体设计

2．1 功能模块设计

本程序需实现的主要功能有：

(1) 用户可以定义想要爬取的图书信息数量

(2) 用户可以基于索引进行图书搜索，并在支持的范围内选择关键词

程序的总体功能如图 1 所示：

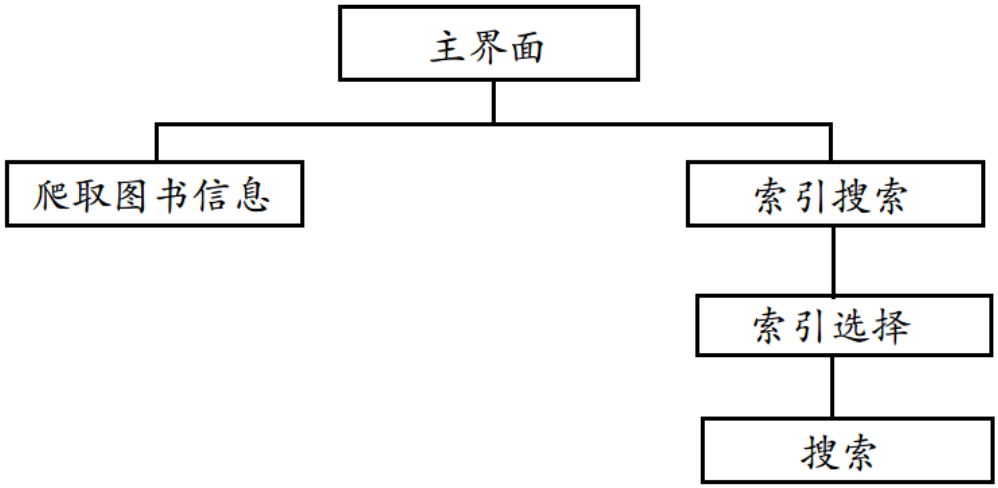

图 1 总体功能图

2. 2　流程图设计

程序总体流程如图 2 所示：

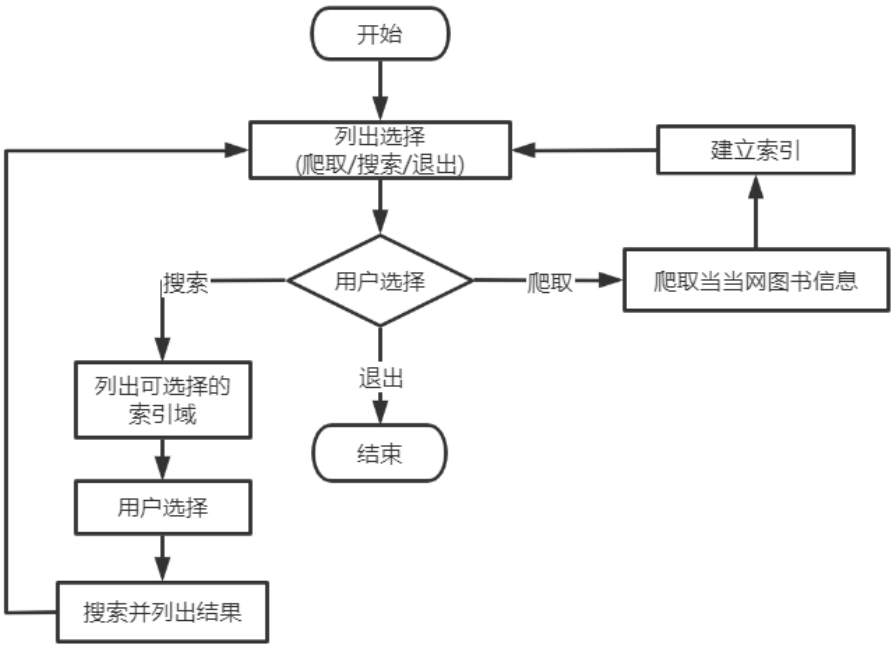

图 2 总体流程图

3　详细设计

3．1　爬虫部分设计

爬虫部分主要用 Crawler4j 来爬取当当网图书信息的页面，再通过 jsoup 进行 html 页面的解析，提取出需要的图书信息，最后保存为文件。

3．2　搜索部分设计

首先遍历爬取到的图书信息的文件，提取相关信息，然后通过 Lucene 建立索引，具体操作为，对每一本书都建立一个 Document 对象，然后将 Document 对象添加进索引。

搜索时，先选择索引域，然后输入关键词，以及期望的结果数量，Lucene 就会通过索引域和关键词进行搜索，并将结果输出。

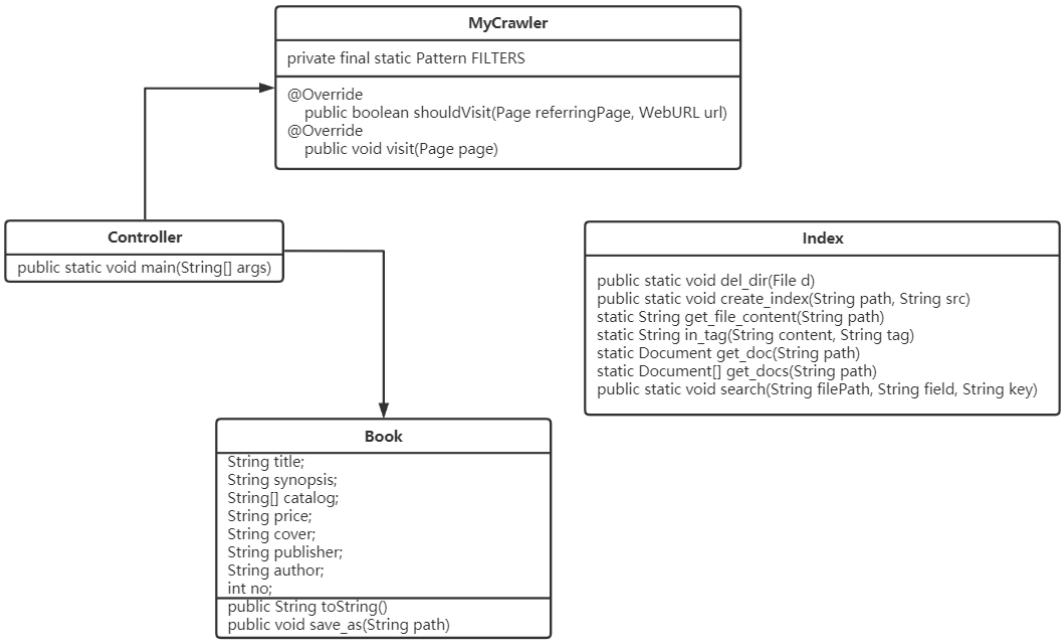

图 3 类图

3．3　类设计

以下是类图中有关数据和方法的详细说明：

Book 类的成员变量即为图书的基本信息，toString 方法可以把 Book 的所有数据输出为字符串，便于储存，sava_as 方法可以把图书信息按编号 no 保存为一个文件，便于建立索引。

MyCrawler 类继承了 WebCrwaler 类并重写了 shouldVisit 和 visit 两个方法，用于自定义要爬取的页面以及爬取规则，新定义了 FILTERS 成员变量用于过滤不重要的页面。

Controller 只有一个 main 方法，用于调用 WebCrawler 类进行爬取，其中定义了要爬取的初始链接，线程数，要爬取的页面的数量

Index 类主要的方法为 CreateIndex 和 search，分别用于建立索引和基于索引搜索。建立索引时会先检查清空索引文件夹，然后读取图书信息保存的目录，遍历所有图书并为每本图书构造 Document，然后添加进索引。搜索时会在索引文件夹下根据关键词和索引域进行查找。

4 测试与运行

4．1 程序测试

在程序代码基本完成后，经过不断的调试与修改，最后测试本次设计的图书搜索引擎能够正常运行，没有出现明显的错误和漏洞，但是在一些细节方面仍然需要完善，比如增加爬取数据方面的自定义。总的来说本次设计在功能上已经基本达到要求，其他细节方面有待以后完善。

4．2 程序运行

程序运行主界面：

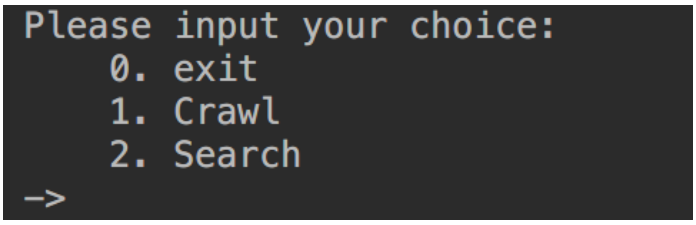

图 4 程序运行初始界面

直接退出：

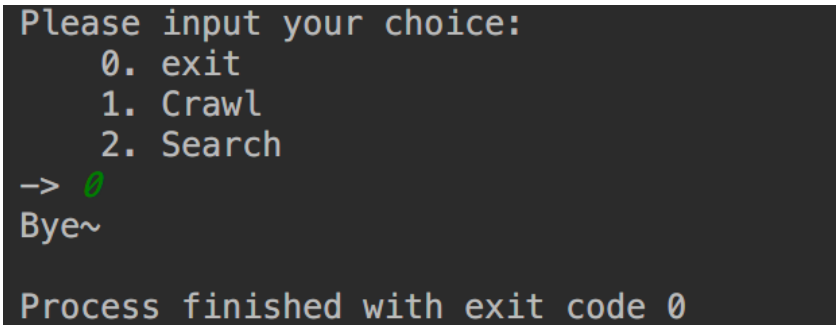

图 5 选择直接退出程序

选择爬取：

图 6 选择爬取

选择搜索：

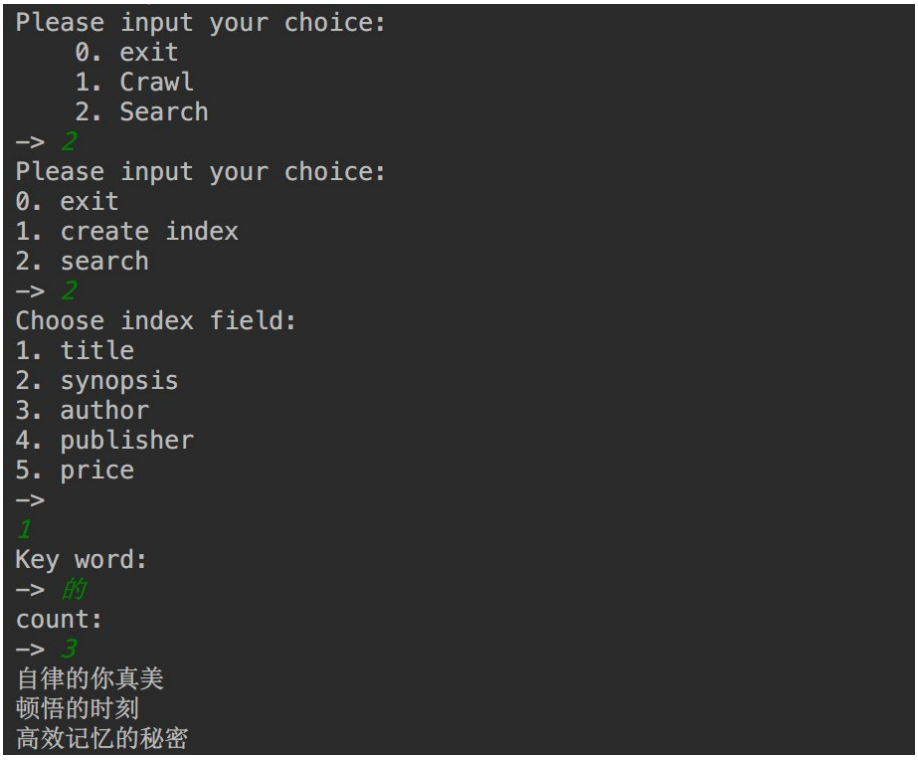

图 7 选择搜索

其他输入：

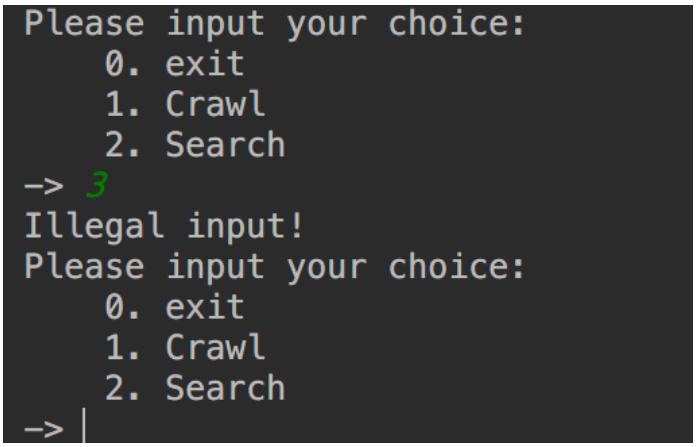

图 8 不合法输入

5．总结

这个图书搜索引擎其实并不难，主要在于使用第三方工具包，涉及到包导入还有包依赖管理等问题，实际上需要做的设计和需要写的代码并不多，但是依然涵盖了许多知识，想要做好并不容易。

经过编写这个图书搜索引擎，我认识到应该注意细节问题，虽然是很小的问题，但可以提高自己编程的能力，而且还可以培养自己编程的严谨性，同时还可以为以后的编程积累经验。

通过这次，可以全面系统的理解程序构造的一般原理和基本实现方法。把死板的课本知识变得生动有趣，激发了学习的积极性。把学过的知识强化，能够把课堂上学的知识通过自己设计的程序表示出来，加深了对理论知识的理解。

参考文献

[1] 耿祥义．Java 大学实用教程[M]．北京：清华大学出版社，2009．

[2] 耿祥义．Java 课程设计[M]．北京：清华大学出版社，2008．

[3] 王鹏．Java Swing 图形界面开发与案例详解[M]．北京：清华大学出版社，2008．

[4] 丁振凡．Java 语言实验教程[M]．北京：北京邮电大学出版社，2005．

[5] 郑莉．Java 语言程序设计[M]．北京：清华大学出版社，2006．