# 武汉大学国际软件学院 实验报告

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 课程名称 | 面向对象程序设计(Java) | | | | | | | | | |
| 专业年级 | 软件工程2015级 | | | | | | | | | |
| 姓名 | 石宝荣 | | | | | | | | | |
| 学号 | 2015302580090 | | | | | | | | | |
| 协作者 | 无 | | | | | | | | | |
| 实验学期 | 2016-2017 | | | 学年 | 1 | | | | 学期 | |
| 课堂时数 | 3 | | | 课外时数 | | | 0 | | | |
| 填写时间 | 2017 | 年 | 1 | | | 月 | | 6 | | 日 |

|  |  |  |  |
| --- | --- | --- | --- |
| 实验概述 | | | |
| 【实验项目名称】 | | | |
| Implementing a Bilibili video webcrawler | | | |
| 【实验目的】 | | | |
| 1. Understand HTTP requests, responses, and redirections. 2. Use apache-httpclient to make up requests and decompose responses. 3. Use jsoup to traverse between and inside HTML tags. 4. Know SQL grouping and sorting. 5. Use Maven to manage project dependencies. | | | |
| 【实验环境（使用的软件）】 | | | |
| 硬件：华硕  操作系统：win10  （使用的软件）：Eclipse、MySQL Server 5.7 | | | |
| 【参考资料】 | | | |
| 1. 助教给的参考资料 2. 《Java语言程序设计》课本 3. 网络查询 | | | |
| 实验内容 | | | |
| 【实验方案设计】 | | | |
| 整个实验分为以下步骤，一、创建maven依赖的工程并添加需要的maven依赖；二、创建并连接数据库；三、在b站上爬取数据并将数据插入数据库；四、排序；五、下载每个类别收藏数top3的视频。下面我来分别说明。   1. 创建maven依赖的工程： 2. 下载maven并设置配置环境变量；   C:\Users\bero-\AppData\Roaming\Tencent\Users\1819246385\QQ\WinTemp\RichOle\($SJG%U6VN{P88N~U5~H2EG.png  C:\Users\bero-\AppData\Roaming\Tencent\Users\1819246385\QQ\WinTemp\RichOle\5T%L7W8)~]7LK[(07$SQ1%3.png   1. 安装maven for eclipse插件并配置；   3、构建maven项目；  C:\Users\bero-\AppData\Roaming\Tencent\Users\1819246385\QQ\WinTemp\RichOle\B~5FW2TC)$FCPG]{B91NPM6.png  4、 添加maven依赖；C:\Users\bero-\AppData\Roaming\Tencent\Users\1819246385\QQ\WinTemp\RichOle\7L}FQ0BVOO4@(I_@ELX{5SO.png   1. 创建并连接数据库 2. 在mysql中根据需要创建数据库和表； 3. 在maven中导入mysql依赖； 4. 使用java连接mysql数据库；   C:\Users\bero-\AppData\Roaming\Tencent\Users\1819246385\QQ\WinTemp\RichOle\OCVQC%CO~TFUA8147BTN%U9.png   1. 在b站上爬数据 2. 利用fiddler4软件在B站上抓包，获取不同分类页面的链接；   C:\Users\bero-\AppData\Roaming\Tencent\Users\1819246385\QQ\WinTemp\RichOle\VKS@S2IH5LSP8434]T5I[UB.png  C:\Users\bero-\AppData\Roaming\Tencent\Users\1819246385\QQ\WinTemp\RichOle\VQMF0{VE4V74ZB1IY}CL3C5.png   1. 采用HttpClient获取json格式的内容；   C:\Users\bero-\AppData\Roaming\Tencent\Users\1819246385\QQ\WinTemp\RichOle\935WBLP$(~G$B2Q5721092B.png   1. 利用Json对象提取内容，并将提取到的内容插入数据库；   C:\Users\bero-\AppData\Roaming\Tencent\Users\1819246385\QQ\WinTemp\RichOle\8~T2RU%YX8~4S@JJ2ZLDIH5.png   1. 一页页提取数据，将一个类别的数据全部提取出来插入数据库； 2. 排序，   利用sql语句对一个类中的数据进行排序；  C:\Users\bero-\AppData\Roaming\Tencent\Users\1819246385\QQ\WinTemp\RichOle\HZR7]_]})@ZEVDK_{3~~E8R.png   1. 下载视频   利用java IO 和ibilibili的下载地址，对指定视频进行下载。  C:\Users\bero-\AppData\Roaming\Tencent\Users\1819246385\QQ\WinTemp\RichOle\%%H`BS6~1KX@VIZ[SWTP4RH.png  C:\Users\bero-\AppData\Roaming\Tencent\Users\1819246385\QQ\WinTemp\RichOle\DGXW}IQH[T]H5GVF728YA4K.png   1. 清空数据库，重复进行三到五，下载每类的top3视频。   七、最后对代码进行补充，修整，优化，完成整个实验。 | | | |
| 【结果】  结果：在写实验的过程中遇到了很多问题，如：安装maven时，repository文件丢失，创建项目总是失败；用java连接数据库的时候总是报错，不是驱动程序找不到就是sql语法错误等等，但是最终都通过严密的逻辑推理，程序调试或者是在网上搜索解决方案解决了问题，最终成功的完成了实验，爬取了50万条数据。（因代码逻辑原因：采用的是爬一类，排序一类，下载一类top3，清空数据库再爬下一类的方法。所以没有办法获取Mysql中50万条数据的截图~但是代码逻辑中设置了50万break语句，所以爬够了50万条数据。）下载视频部分的截图如下：  C:\Users\bero-\homework_5\视频下载截图.png  每个都可以正常播放：  C:\Users\bero-\homework_5\视频播放截图.png  结论与收获：在这次实验中，我掌握了maven的配置与操作，理解并学会了response，request 以及Httpclient的相关知识，也学会了Json等技术的使用，并对SQL语句和mysql | | | |
| 数据库的使用有了进一步的掌握，收获很多。 | | | |
| 【小结】 | | | |
| 1、不要只局限于老师所讲的知识，这样是不够的，同时还要多在课下学知识，多往后看，多深度拓展，这样才能在编程上更加熟练更加轻松。  2、培养自己的灵活思维意识，要结合老师的思维进行思考创新，追求高效率，高品质的代码。不应死搬硬套，也不应闭门造车。  3、 牢记代码格式，这样才能更透彻理解其中逻辑关系，以便解题。  4、 不要忘记注释，一方面是为了美观，一方面也是为了人能看懂，这是一个好习惯。  5、 做完实验之后，感觉自己的代码水平明显提高很多，看来实践出真知，只有通过实践，才能掌握更多。  6、 需要课下继续多找题，多做题，提高自己的编程能力。  7、 必要的时候要问同学助教，这样不仅节省时间，还能学到很多。  8、需要多利用网络，这样可以学到更多。 | | | |
| 指导教师评语及成绩 | | | |
| 【评语】 | | | |
|  | | | |
| 成绩： |  | 指导教师签名： |  |
| 批阅日期： | | | |