

JAVA课程大作业 elective 报告

引入

心仪课程不好找？

选课计划好几页？

输入验证码费事？

课程表无设计感？

不用怕，elective来帮忙啦！

整体框架



如何运行

代码地址 <https://github.com/wongsingfo/course-ele>
需要Java SE 11

界面交互

下载系统对应的JavaFX11 SDK包，复制到项目根目录下，可能会覆盖当前已有的文件。
下载地址 <https://gluonhq.com/products/javafx/>。
编译和运行都要加参数 --module-path ./lib --add-modules=javafx.controls,javafx.fxml

爬虫功能

需要安装一个Selenium支持的浏览器，以及对应的jar包。

具体实现

爬虫功能

由于北大选课网有反爬虫机制，为了达到隐蔽性的要求，我们采用了selenium模拟浏览器登陆的方式爬取数据，爬取过程中使用xpath解析路径；在登陆伊始即爬取下该同学可以选择的培养计划、专业课程、通选课、英语课等一系列课程，分类保存在本地的csv文件中，此外ele_result类用于保存爬取到的课程表信息，包括行、列（用于表明是周几的第几节课）、课程信息等。

爬虫的主要函数都在Scrapy_elective.java中，各函数说明如下：

List<WebElement> get_menu(WebDriver driver): 用于得到选课网上的右上角的那一行按钮选项，以便进入各个分类爬取数据；

get_elective_plan(WebDriver driver,List<WebElement> choice): 用于获取选课计划的相关信息；

add_course_* (各科分类) (WebDriver driver,List<WebElement> choice)**: 用于获取各类课程的相关信息，并保存在本地csv文件中；

WebDriver log_in(String user_name,String password): 新建浏览器对象，参数正确时即可成功登陆上选课网主页，以便进行接下来的爬取工作；

List<elec_result> get_elec_result(WebDriver driver,List<WebElement> choice): 将目前的选课结果以elec_result对象的形式保存在列表中，数据用于实现课程表功能；

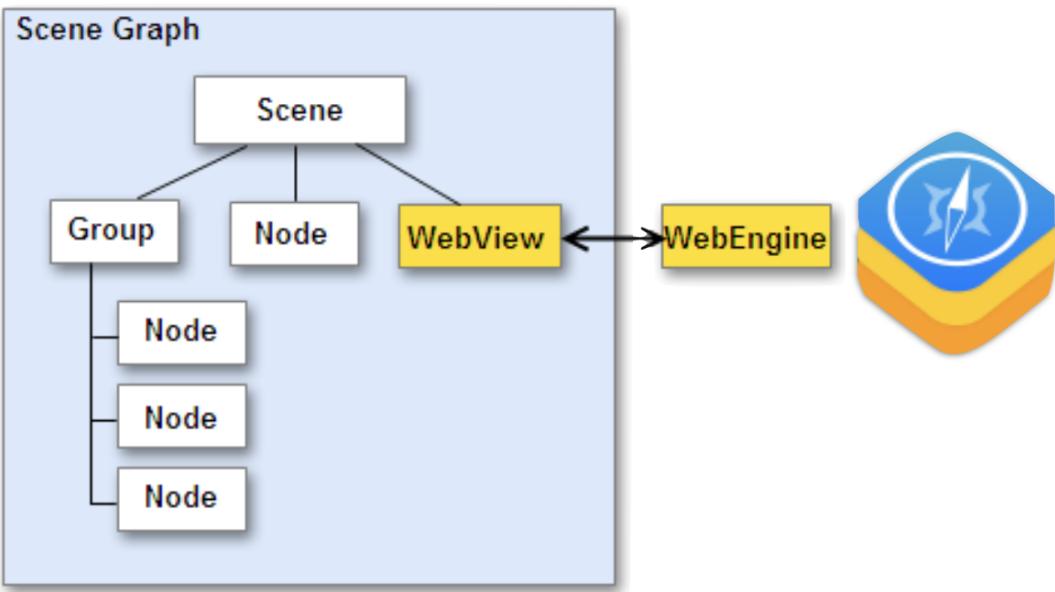
getData(String id,String password): 将以上功能打包成一个函数，用于在登陆时直接将所有信息爬取下来

GUI设计

在GUI设计中，我们主要选取了JavaFX和JavaFX Scene Builder作为开发工具，来搭建elective的前端界面。

JavaFX是一个强大的图形和多媒体处理工具包集合，它允许开发者来设计、创建、测试、调试和部署富客户端程序。由于JavaFX库被写成了Java API，因此JavaFX应用程序代码可以调用各种Java库中的API，并且已经完整地集成到了JRE和JDK之中，所以JavaFX应用程序也是跨平台的，使用JDK7及以后版本编译的JavaFX应用程序也可以在所有的主要桌面平台上运行。JavaFX还可以自定义程序外观，使用层级样式表（CSS）可以方便地定制程序的外观和样式。

JavaFX Scene Builder 是一个可视化布局工具，可快速设计JavaFX应用程序用户界面，无需编写代码。用户可以拖放UI组件到工作区，修改组件的属性，应用样式表，而且在后台自动生成所创建布局的FXML代码。FXML是一种基于XML的声明式标记语言，用于描述JavaFX应用程序的用户界面。最后得到的是一个可以与Java项目整合到一起的FXML文件，从而将UI与应用程序逻辑绑定起来。从而我们可以通过JavaFX Scene Builder 来交互式地设计GUI。



我们的主要设计理念是：

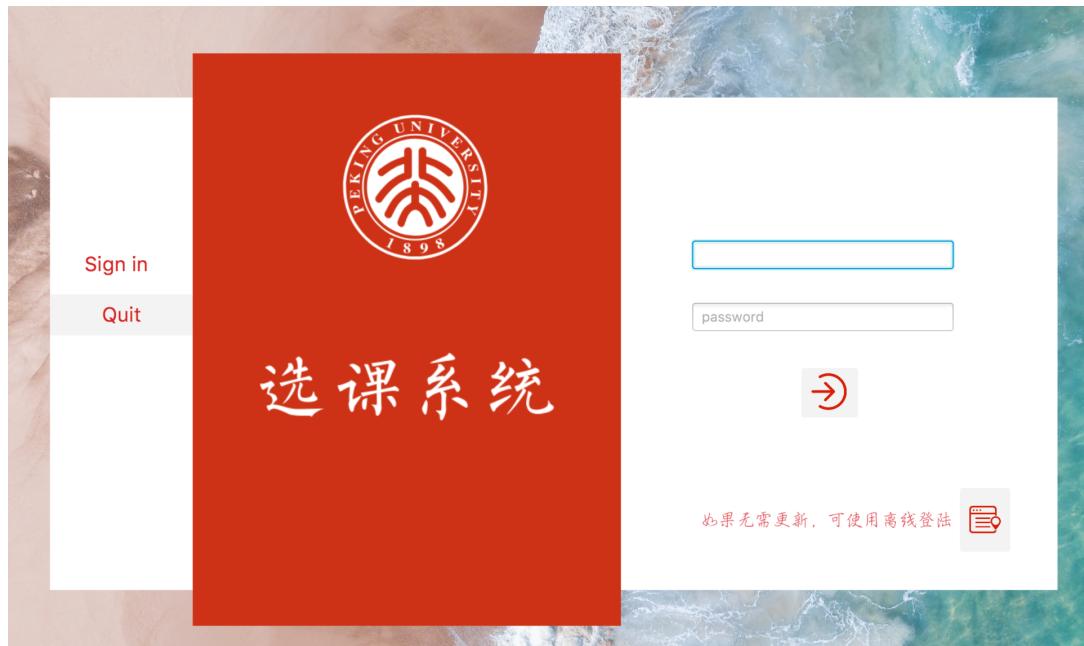
扁平化设计

北大红加上专业配色！！！

下面开始介绍我们设计的各个界面。

登录界面在initStage时设置Stage为透明，因此登录界面可以更加个性化；

登录界面设置了几个按钮以进行页面交互：



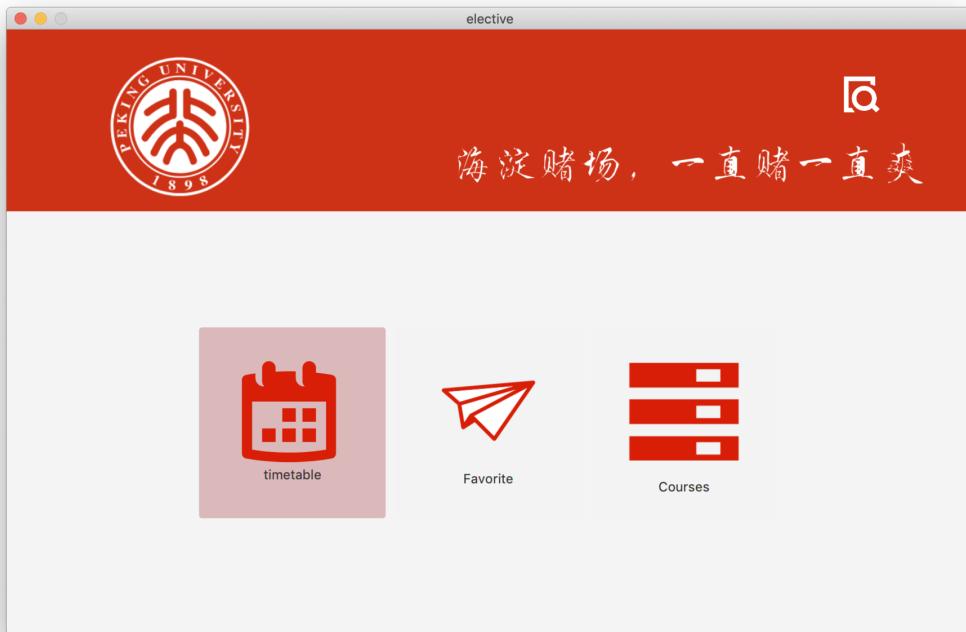
Quit：退出按钮，用于终止整个程序；

登录文本框和密码框：用于输入学号和密码，从选课网爬取数据；

箭头按钮：用于登陆模式下进入该程序；

离线按钮：当没有数据更新时（例如不是选课阶段），可以点击此按钮用保存在本地的数据登陆；

当登陆以后，进入以下主界面：



此界面作为导航界面有四个可选择的项目：查询、课程表、收藏和所有课程查看，进入每一个项目都不会打开新的窗口，而是在此窗口上重新布置Scene；

下面以课程界面为例简述子界面的设计：

公选课 文计	名称	学分	授课老师	院系	信息	收藏课程
乐基础	2.0	马清(教授)	艺术学院	1~4周 每周一~周五 二教107 1~4周 每周三~周五 二教107	<input type="button" value="点击收藏"/>	
中国文化史	2.0	王锦贵(教授)	信息管理系	1~2周 每周一~周五 二教407 1~2周 每周二~周五 二教407 1~2周 每周一~周五 二教407 1~2周 每周一~周五 二教407	<input type="button" value="点击收藏"/>	
跨文化交流学	2.0	许静(教授) 李臻怡	新闻与传播学院	1~2周 每周一~周五 二教211 1~2周 每周一~周五 二教211 1~2周 每周一~周五 二教211 1~2周 每周一~周五 二教211	<input type="button" value="点击收藏"/>	
公共物理学	2.0	王晓钢	物理学院	1~4周 每周一~周五 二教316 1~4周 每周一~周五 二教316 1~4周 每周一~周五 二教316 1~4周 每周一~周五 二教316	<input type="button" value="点击收藏"/>	
美国史通论	2.0	王希(教授)	历史学系	1~4周 每周一~周五 二教205 1~4周 每周一~周五 二教205 1~4周 每周一~周五 二教205 1~4周 每周一~周五 二教205	<input type="button" value="点击收藏"/>	

back按钮：用于回到主导航界面；

查询按钮：用于打开查询界面；

右上角为导航栏，可以快速导航到其他功能；

使用点击收藏按钮可以收藏此门课程；

点击下拉滚动条可以选择课程大类进行查看。

考虑到学校的选课网上只有简单的检索功能，每次需要手动填写关键词来完成搜索，感觉十分僵硬，因此我们想要实现输入一句包含了多个关键词的话就可以完成课程的检索。同时配合上收藏功能，可以方便的把自己想要筛选的课加入收藏表，避免多次重复的查找。

考虑到用户在使用搜索功能时输入的大多是课程号，课程名，教室，老师名，开课院系，学分，时间等内容，我们便把爬虫获得的所有课程信息读入到内存中，并将其中的每一节课封装到一个Class类中，在Class类使用正则表达式（主要用在课程时间和教室信息的提取上）完成了对上述易被搜索到的关键词的提取。

在匹配方面，有两种方案可以采用，一种是根据用户输入的数据进行中文分词，然后把分词后的结果和每一节课去匹配，得到每一节课与输入内容的匹配程度的评分，最后输出评分较高的课程。另一种方案是把每节课的关键词去与用户的输入内容进行匹配，给出匹配程度然后进行排序。考虑到中文分词需要下载比较大的中文词库，用户体验不佳，而且针对奇怪的课程缩写的输入，即使加入自定义的词典，中文分词效果也不好，我们采用了第二种匹配方案。

匹配方面我们采用了简单的正则表达式匹配。对于学分，教室等内容，我们采用的是完全匹配，只有用户输入的信息和课程中的信息完全相同时才会提高匹配相似度。对于课程名，开课院系等内容，为了能够对它们的缩写实现匹配，我们把它们的正则表达式写成非完全匹配，比如上面的“当代机器学习实践”，其对应的正则表达式是“当?代?机?器?学?习?实?践?”，这样可能会导致输入内容与很多课中的某些字相匹配，我们用两种方法来减少这种误匹配，首先我们去尝试匹配必须完整匹配的内容，比如“3学分”这样的输入，我们会将它先匹配掉，防止其匹配课程名等信息，其次我们给匹配长度更长的课赋更高的匹配分，也对第一个字能匹配上的赋更高的分。

为了处理否定输入的内容，我们做了这样简单的处理：某节课和输入内容匹配完后，如果输入内容中未被匹配的部分中出现了“不”字，我们将“不”字之后的匹配分全部改为负分。简单的实现效果如下：

The screenshot shows a list of elective courses. Each row contains the following information:

课程名称	学分	授课老师	院系	信息	收藏课程
信息科学中的数学	2.0	孔雨晴(助理教授)	信息科学技术学院	1~3周 每周一~4节 二教105 1~3周 每周三~4节 二教105 1~3周 每周五~4节 二教105	点击收藏
可编程逻辑电路设计(I)	2.0	蒋伟(副教授)	信息科学技术学院		点击收藏
3S野外综合实习	1.0	田原(副教授)	地球与空间科学学院		点击收藏
"一带一路"综合实习	2.0	柴彦威(教授)	城市与环境学院		点击收藏
阿尔卑斯野外综合地质国际...	2.0	韩宝福(教授) 吴朝东(...)	地球与空间科学学院		点击收藏
城乡地域空间认知实习	1.0	阴勤(讲师)	城市与环境学院		点击收藏
地貌实习	2.0	莫多闻(教授) 张家富(...)	城市与环境学院		点击收藏
航空航天工业实习	3.0	郑殿峰(高级工程师)	工学院		点击收藏
环境综合实习二	1.0	刘兆荣(副教授)	环境科学与工程学院		点击收藏
环境综合实习一	1.0	刘兆荣(副教授)	环境科学与工程学院		点击收藏

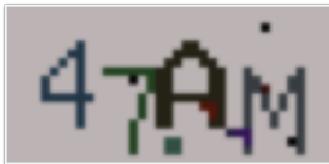
为了解决名字输入困难的问题，我们还添加了拼音首字母查找授课老师的功能。具体实现方法：注意到Java中String类的字符使用UTF-16编码，也就是说，每个汉字是一个字符，而且对应一个整数。我们只需要找到该整数到拼音首字母的映射就可以了。映射可以通过网上Unicode转拼音的数据生成。另外还有多音字的问题，因为名字一般只有三个字符，而且多音字也不会特别多，所以我们只需要将所有可能的组合都枚举一遍。

接口是PinyinConverter类里的List<String> getAbbr(String name)方法。该方法的参数是中文名字，返回所有可能的拼音首字母缩写。

The screenshot shows a list of elective courses. Each row contains the following information:

课程名称	学分	授课老师	院系	信息	收藏课程
ACM/ICPC竞赛训练	2.0	郭炜(讲师)	教务部	3~4周 每周一~5~8节 二教105 3~4周 每周二~5~8节 二教105 3~4周 每周三~5~8节 二教105 3~4周 每周四~5~8节 二教105 3~4周 每周五~5~8节 二教105	点击收藏
中国传统表演艺术	3.0	张新亚	教务部	3~4周 每周一~2~4节 二教408 3~4周 每周二~2~4节 二教408 3~4周 每周三~2~4节 二教408 3~4周 每周四~2~4节 二教408 3~4周 每周五~2~4节 二教408	点击收藏
函数式语言介绍	2.0	郝丹(长聘副教授)	信息科学技术学院	1~2周 每周一~2~6节 二教319 1~2周 每周二~2~6节 二教319 1~2周 单周周三~2~6节 二教319 1~2周 单周周四~2~6节 二教319 1~2周 单周周五~2~6节 二教319	点击收藏
前沿科学研究实用方法	2.0	杨玉超(助理教授)	信息科学技术学院	4~5周 每周一~5~8节 三教104 4~5周 每周二~5~8节 三教104 4~5周 每周三~5~8节 三教104 4~5周 每周四~5~8节 三教104 4~5周 每周五~5~8节 三教104	点击收藏

验证码识别



选课网的验证码有以下几个特点：

- 噪声少
- 图片没有经过变形
- 仅有一种字体
- 每个字的颜色不一样

针对这些特点，我们事先对26（字母）+10（数字）个符号进行了人工标注，并设计了如下的识别流程：

第一步，通过颜色加上一些特殊判断，将四个符号分开

第二步，对于每个符号，与事先标注好的符号进行匹配，选择最像的那一个作为答案。

那么如何获取并填入验证码呢？只需要用JavaScript对验证码的加载事件进行监听，然后将验证码图片画在一个Canvas上就可以获得每个像素点的内容了。接着，通过JavaScript代码识别并填入验证码。

由于补退选开放到本课程作业提交这段时间比较短，提交的版本还没有加入该功能。

课程收藏

课程名称	学分	授课老师	院系	信息	已收藏课程
高级英语口语	2.0	TOM MANSON	英语语言文学系	1~4周 每周周一~周五 1~4周 每周一~周五 1~4周 每周一~周五 1~4周 每周一~周五	<input type="button" value="取消收藏"/>
棒、垒球	1.0	焦晨曦(教学助理)	体育教研部	1~2周 每周一~周五 1~2周 每周一~周五 1~2周 每周一~周五 1~2周 每周一~周五 1~2周 每周一~周五 1~2周 每周一~周五	<input type="button" value="取消收藏"/>
太极拳	1.0	杜军明(讲师)	体育教研部	1~2周 每周一~周五 1~2周 每周一~周五 1~2周 每周一~周五 1~2周 每周一~周五 1~2周 每周一~周五 1~2周 每周一~周五	<input type="button" value="取消收藏"/>

在课程收藏模块中，我们支持对课程的动态收藏和取消收藏，在每次登录时加载，并在退出elective时保存所做的收藏操作。

课程表



在课程表模块中，我们支持将实时更新的已选课程生成简洁的课程表，将鼠标放在某课程上就会弹出黑色方框显示备注信息。将收藏课程与已选课程同框且有区分地显示的功能有待未来进一步实现。

总结

本次课程作业，我们主要使用JavaFX和Selenium实现了用户与选课网之间的一个友好的交流通道。旨在通过这次试验，熟悉Java编程的方法和思想。