**电工导实验报告4**

F1403023 5140309534 韩坤言

**一、实验目的**

1. **了解**web.py 利用web框架进行简单的web开发
2. 建立一个简单的搜索引擎

**二、实验内容**

1. 利用web框架web.py实现web开发的基本功能
2. 运用模版简化工作量，运用表单达到页面跳转的功能
3. 建立一个基于之前的搜索引擎，实现网页搜索和图片搜索

**三、实验环境**

1. **Firefox + Firebug插件或Chrome**
2. **Python 2.7 + easy\_install + BeautifulSoup**
3. **JCC + PyLucene**
4. **web.py**

**四、实验步骤**

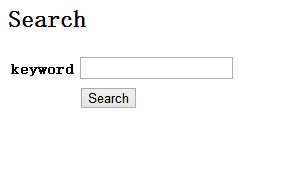
所谓web框架，就是某种应用的半成品，好处是减少重复开发的工作量，缩短开发时间，降低开发的成本。我们使用的是web.py,基于python进行web开发。通过一些例子来了解各个部分是如何运作的，了解了模版，以及如何处理表单。

中期整合则是在之前几次lab的基础上加加以整合，为了界面的美观与整洁，运用div+css来规范网页的样式，以及布局，调整各个模块的属性以达到满意的效果。

**五、问题及其解决**

**1. 使用web.py，结合前面学习的HTML, Lucene, 中文分词等知识点，根据上次实验爬取的网页，建立一个简单的搜索引擎。**

在了解了web.py的工作原理之后，其实这个问题也就引刃而解了。我们在做的无非就是把用户在input框中的keyword传入Search.py进行搜索，最后把结果通过模版的方式呈现给用户。

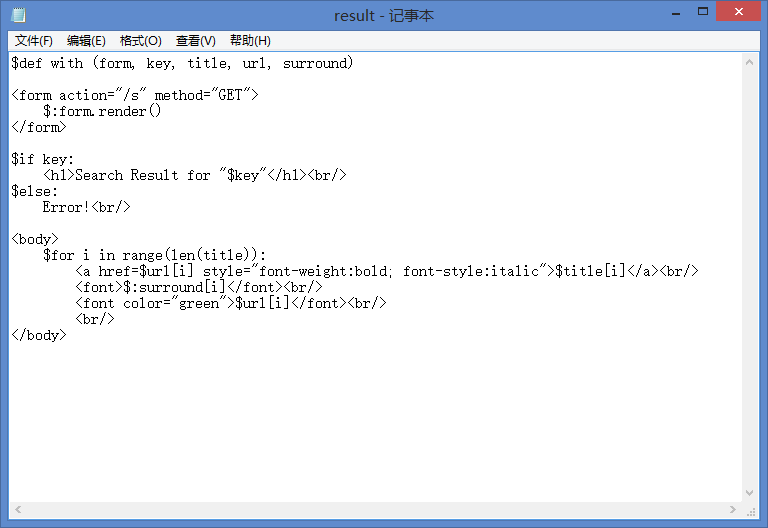


初始界面



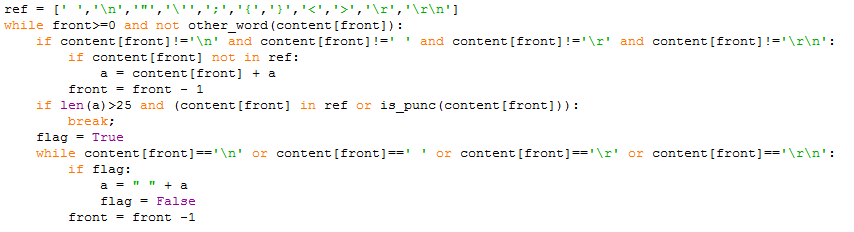
结果界面

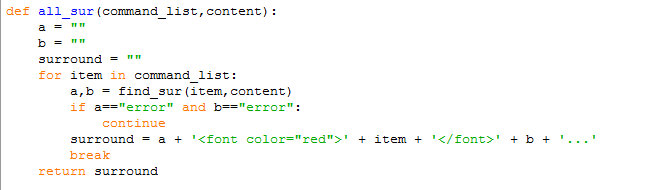
实现起来比较容易，展示搜索结果通过模版呈现，通过for循环，把结果依次现实。<a href=...>来创建超链接，通过改变参数来修改字体，颜色等，使其看起来像一个搜索引擎的结果页面。



result

过程中比较麻烦的就是得到关键词周围的文字信息。这势必需要把content保存下来，放到index里，考虑到这样查询的效率会很低下，索性要找周围信息的时候重新打开一次文件速度会快得多。得到前后的信息涉及到的是字符串的处理，由于用过bs的get\_text()后任然有许多html的tag残留，直接向前向后找是不可取的。我过滤掉了许多明显的具有tag标志的字符，然后先前寻找标点符号或者空格，当字数达到一定的数量以后就停止，把这一段存下来，向后同理，再把中间的关键词前后加上html的语言使其显示出来的时候呈红色。



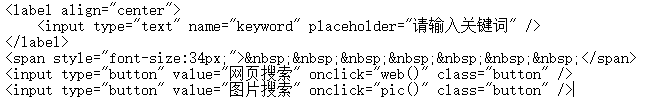


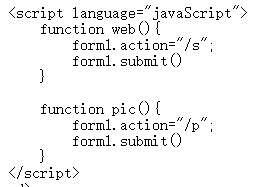
部分代码

虽然界面很朴素，但是基本的功能都已经实现。

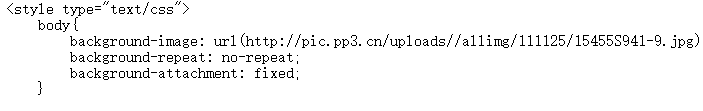
**2.** **制作一个图片加文字的搜索引擎，作为中期整合。在上次的基础上，加入图片搜索，使用css制定样式**

首先是把图片的搜索功能加入。我的想法是一个搜索框对应两个button,一个搜索website,一个搜索picture。经过大量资料的查询，我实现了我的想法。我舍弃了之前的form模版，自己创建了一个type是text的输入框和两个type是button的按钮，不直接选择按钮是应为这样能添加更多的属性。通过添加鼠标的click事件来进行跳转。很不错的达到了预期的效果。

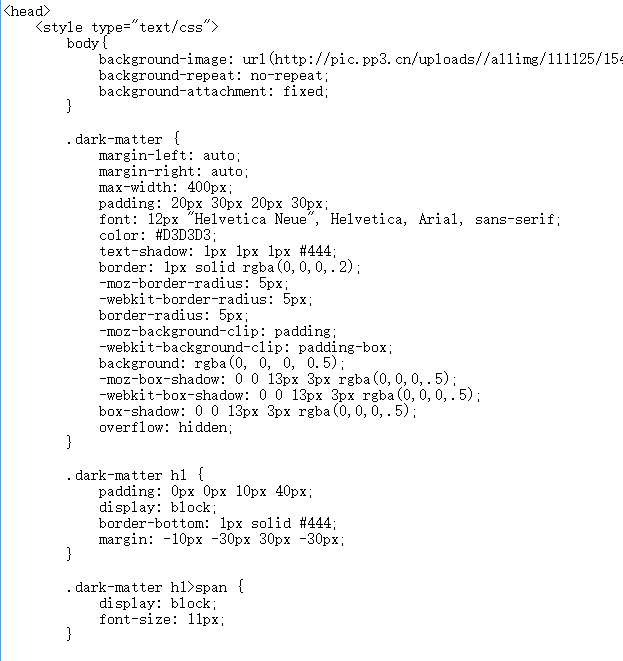




为了美观，通过div+css的方式来使网页中的每个元素更加漂亮。

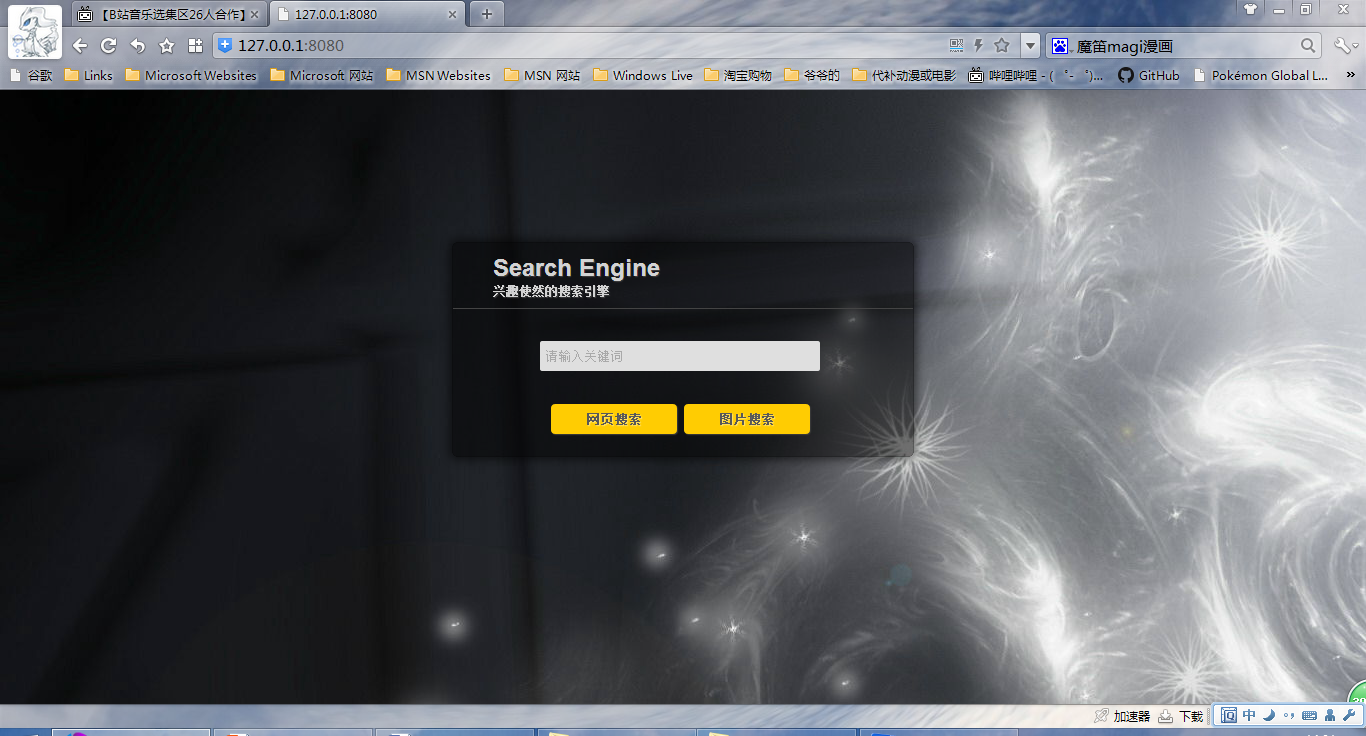
我定下的基调是黑色和金色的高雅风格，我选择了一张合适的背景，为了防止背景重复以及变动，我进行了如下设置

我修改了很多属性值，使得form的表单呈现半透明的效果，部分代码如下：



搜索网页的结果界面比较容易，多弄几个框，放进去就行了，设置沿用之前的即可，图片搜索的结果页面为了使一排现实多个图片，我设置了float，图片就会多张并排显示了。

最终的界面如下：



起始界面



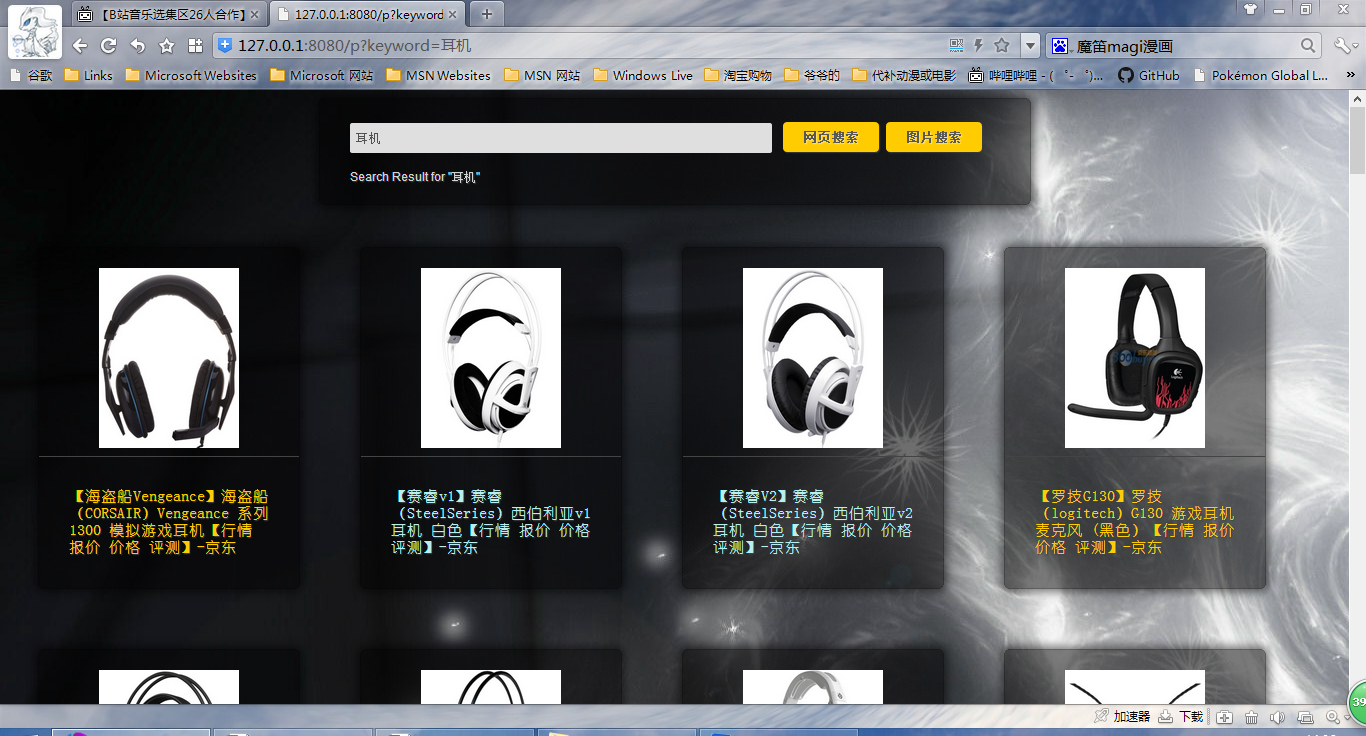
起始界面细节展示



网页搜索结果信息界面



网页搜索结果信息界面细节展示



图片搜索结果页面

网页我爬取了1000个，图片我爬取了500个来建立索引。网页从sjtu开始爬，图片从京东的某游戏展区开始爬的，如果需要测试可以测试一些相关关键词。

我发现用户点击网页搜索后到搜寻结果出来有一段等候时间，这时间随我爬取的网页数增多而增加，如果有方法能提高处理的速度就更加完美了。

**六、实验总结**

这次的作业把之前所有的成果整合在一起，虽说遇到历史遗留问题很麻烦，但是当自己的成果展现在网页上呈现在眼前的时候，能真切的体会到自己努力后的兴奋和喜悦。其中不乏被报错信息烦的半死，也有根本不知从何处下手的情况。可到最后，一切都引刃而解的时候，成就感满满。通过这半个学期的学习和练习，我对计算机网络，网页等方面都有了极大的兴趣，学习了很多很多。我很期待后半个学期的进一步的学习。