|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **项目名称** | **爬虫爬取信息** | | | | | | | |
| **小组名称** | **第六组** | **人数** | **7** | **组长** | **吴坤** | **电话** | **15287707645** |
| **成员姓名** | **学号** | **专业** | **年级** | **主要承担任务** | | | |
| **吴坤** | **20161120179** | **软件工程** | **2016** | **编写代码** | | | |
| **李金阳** | **20161060125** | **计算机科学与技术** | **2016** | **编写代码** | | | |
| **赵张毅** | **20161900036** | **材料物理** | **2016** | **测试填表** | | | |
| **董江枫** | **20161350114** | **网络与新媒体** | **2016** | **做讲解** | | | |
| **杨一龙** | 20161910136 | **统计学** | **2016** | **做PPT** | | | |
| **杨登富** | **20141060127** | **计算机科学与技术** | **2014** | **实验报告** | | | |
| **张基伟** | **20161120242** | **软件工程** | **2016** | **代码调试** | | | |

* 选题理由：

项目选题理由及意义：网络爬虫是一个自动提取网页的程序，它为搜索引擎从万维网上下载网页，是搜索引擎的重要组成部分，传统网络爬虫是一个从一个或很多初始网页上的URL，在抓取网页的过程中，不断从当前页面上抽取新的URL放入队列，直到满足系统的一定条件。

我们可以利用网络爬虫爬取网页中的信息，并且保存下来，以便查看或使用。使用python编写网络爬虫更加有效率并且简便。

* 作品简介:

python爬虫的用途非常广泛，它可以在互联网上爬去大量我们想要的数据，方便我们整理和分析他们，从而得到我们想要的结果，在这里我将编写一个小小的天气爬虫来学习python爬虫这个神奇的程序，学会爬虫后可以极大的方便了我们以后的学习生活中，面对极丰富的，繁杂的网络资源我们不在望洋兴叹。

应用场景：

在快节奏的生活中，迅速的获取自己需要的信息。从获得的天气中我们得知今天及未来几天的天气状况，可以提醒我们出门是否需要带雨伞，或对衣服的增减等。

* 总体设计：

1,首先我们需要找到我们需要爬取的网页和具体的对象，由于我们做的是爬取关于天气信息的爬虫，所以我们选择了中国天气网（<http://www.weather.com.cn/weather/101190401.shtml>）

找到我们所在城市的天气信息进行实时爬取当天天气信息。

2,工欲善其事必先利其器，在这里我们可以充分的利用python第三方库给我们提供的便利

。

Requests: requests是python的一个HTTP客户端库，跟urllib，urllib2类似,用来访问我们所需访问的网页，如: rep = requests .get ( url ,headers = header, timeout = timeout)

Beautiful Soup提供一些简单的、python式的[函数](http://cpro.baidu.com/cpro/ui/uijs.php?c=news&cf=1001&ch=0&di=8&fv=17&jk=d174a177de878ec7&k=%BA%AF%CA%FD&k0=%BA%AF%CA%FD&kdi0=0&luki=4&n=10&p=baidu&q=35091180_cpr&rb=0&rs=1&seller_id=1&sid=c78e87de77a174d1&ssp2=1&stid=0&t=tpclicked3_hc&tu=u1959022&u=http%3A%2F%2Fcuiqingcai%2Ecom%2F1319%2Ehtml&urlid=0)用来处理导航、搜索、修改分析树等功能。它是一个工具箱，通过解析文档为用户提供需要抓取的数据，因为简单，所以不需要多少代码就可以写出一个完整的应用程序。简单来说，Beautiful Soup是[python](http://cpro.baidu.com/cpro/ui/uijs.php?c=news&cf=1001&ch=0&di=8&fv=17&jk=d174a177de878ec7&k=python&k0=python&kdi0=0&luki=6&n=10&p=baidu&q=35091180_cpr&rb=0&rs=1&seller_id=1&sid=c78e87de77a174d1&ssp2=1&stid=0&t=tpclicked3_hc&tu=u1959022&u=http%3A%2F%2Fcuiqingcai%2Ecom%2F1319%2Ehtml&urlid=0)的一个库，最主要的功能是从网页抓取数据。

3.整理出我们想要的信息数据后，需要对这些信息数据做一些分析和整理。最后就是将得到的最终信息直观的展现在我们的眼前，由于天气的实时性和数据的特殊性我们选择了将这些天气数据按时间分类存储在.csv的电子表格文件中，实例代码如下：

def write\_ data(data, name):

file\_ name = name

with open (file\_ name, 'a', errors='ignore', newline='') as f:

f\_ csv = csv. writer(f)

f\_ csv. Writerows(data)

* 小组成员心得：

杨登富:

在本次小组活动的过程中，我们用了分工与合作的方式，每个人互责一定的部分，同时在一定的阶段共同讨论，以解决分工中个人不能解决的问题，在交流中大家积极发言，和提出意见，了解到了爬虫的相关信息和功能，学到了如何运用爬虫从网页抓取信息并进行编码，学习能力得到了提高。这次的项目让我认识到团队合作的重要性，每个人如果都能各司其职，互相帮助,发挥其最大效能，达到事半功倍的效果。

董江枫：

通过一学期Python的学习，对Python有了一个大概的了解。  
通过小组合作学习的方式，使不同专业的我们对彼此有了一些了解，也发挥了不同专业同学的各自优势，小组项目的共同完成，也提升了同学之间的合作能力，同时也增加了同学们的成就感。  
当然，由于投入的时间不够，自己学习本门课没有取得自己预期的效果。自己以后也会依靠网络对Python会进行持续不断的学习。也希望独自开发出几个小项目。当然，由于投入的时间不够，自己学习本门课没有取得自己预期的效果。

吴坤：

在此次实验中，了解了爬虫实现的方式，以及如何编写一个爬虫。通过查阅资料，翻阅书本进行了解爬虫的python写法，了解爬虫中各个算法，爬虫主要的三个内容:1、进行特定的网页的下载。2、抓取网页特定数据。3、保存想要的数据。在学习爬虫的过程中也更深刻的了解了python的强大之处，编写网络爬虫可以用scrap框架，这个提供更简便的网络爬虫，但是为了更深入的了解网络爬虫，在这次写网络爬虫选择了网络爬虫实现的基本写法。

赵张毅：

本次的小组活动中，每个人分工明确，负责认真的完成各自的任务。在小组合作中，大家一起思考，交流，营造出和谐的学习氛围，我从中也是了解到关于爬虫的一些知识。大家能通过合作，共同解决在设计时遇到的一些困难，取长补短，促进知识技能的掌握。而且，通过本次活动，我更是锻炼了自己的沟通交流能力，在合作完成项目的过程中，成员之间互相启迪，分享想法，小组的项目凝聚了所以成员的智慧。

李金阳：

团队合作和自我完成任务；我们有好几次线下任务，写代码，注释，写小程序，这些都有效的锻炼我们自我能力，让我们从实践中真正体会python的有趣之处;还有合作项目,线上我们有好几次要合作的任务，一方面让我们相互学习，另一反面也无形中锻炼了我们的团队意识。从中我感受到了python的有趣之处，也想好好把它学好。

最后真的好开心选了这门课，真的获益匪浅。

张基伟：

通过这个学期的学习，掌握了python的基础知识和一些简单的应用，阅读一个良好的python程序就感觉像是在读英语一样。python同时也极其容易上手，丰富的库文件也让开发变得十分简单。这次的期末作业主要用到urllib，request等几个主要模块来实现爬虫的，过查看文档和网上搜索前人的经验最后都还算成功完成了。开发一个程序很多时候不光是一个人的事情，是一个团队的工作，所以团队之间的交流沟通和合作也是软件开发的一个重要方面。

杨一龙：

在编程中，如果有好的编程习惯可以达到事半功倍的效果，例如在编写一个程序时可以首先搭建好一个基本框架，然后对框架进行调试，之后每写一个部分试着调试一次，这样可以大大减低后期调试程序的难度。在动手操作的过程中学到了很多新的方法、思路，让我对编程有了更深的理解。次小组作业让我学会了很多在过程中也收获了知识和快乐。