

**信 息 工 程 系**

**2018年暑期实训报告**

|  |  |
| --- | --- |
| **班 级：** | 计算机科学与技术2班 |
| **学 号：** | 20170180 |
| **姓 名：** | 文岚清 |
| **指导老师：** | 何青 |

二零一八年七月

**目 录**

[一、实训任务书 1](#_Toc524516833)

[二、实训日志 2](#_Toc524516834)

[三、项目报告 3](#_Toc524516835)

[四、实训心得 4](#_Toc524516836)

[五、实训评价 5](#_Toc524516837)

整个报告写完后，记得更新页码

# 

# 一、实训任务书

**1、实训目的**

本学年我们学习了Python语言程序开发课程，学习、掌握了Python语言的基础语法。通过学习我们了解到Python语言常用的场景包括：科学计算、web项目开发、人工智能、网络爬虫。为了巩固Python基础知识，扩展Python编程技能，提高学生发现问题、解决问题、独立思考的能力，以Python网络爬虫的设计与实现为题开展为期10天的暑期实训，介绍网络爬虫的原理与实现方法、技术，最终由学生独立设计实现自己的爬虫软件。

目标：

1. 培养学生运用所学课程Python语言程序设计的理论知识和技能，分析解决计算机实际应用中的问题的能力；
2. 培养学生在Python语言程序设计的基础上，开发和实现网络爬虫项目的能力；
3. 培养学生调查研究、查阅技术文献、资料手册以及编写技术文献的能力。

**2、实训时间**

2018年7月5日-2018年7月14日

上午：8:00 - 11:30

下午：14:00 - 17:30

**3、实训地点**

计算机专业实训室1,2,3

**4、实训主要内容**

|  |  |
| --- | --- |
| 7月5日-6日 | Python面向对象编程 |
| 7月7日 | Requests模拟网络请求 |
| 7月8日 | Beautifulsoup解析 |
| 7月9日 | Python操作数据库 |
| 7月10日 | 爬虫项目实现 |
| 7月11日-7月13日 | 小组项目实施 |
| 7月14日 | 小组项目展示 |

# 二、实训日志

# 三、项目报告

|  |  |
| --- | --- |
| **项目名称** |  |
| **项目组成员** | 学号，姓名 |
| **项目背景** | |
| 描述为什么做这个题目，有什么意义，比起别人做的有什么创新 | |
| **作品描述** | |
| 描述作品具体实现了什么样的功能，开发的流程，用了哪些库，哪些方法来实现等等 | |
| **作品截图** | |
| 截图要求清晰，能清楚反应项目的成果 | |
| **项目总结** | |
| 你在项目中的分工是什么，整个项目完成的如何，有哪些需要改进的 | |

# 四、实训心得

|  |
| --- |
| 通过这次的python爬虫实训的学习，让我有增长了一门学问，并且意识到python爬虫有好处也有局限性。先从局限性来说吧，1.访问局限性。比如我访问一个网站，要获取网站的连接。如果我们抓取的数据无法访问其类容，或者无法转到下一个页面，爬虫无法解决这类问题。 2.针对性太专一。爬虫去爬取下载网页可以的到一个统一的模板，但是关于网页的解析是不行的，网站的格式及模板却不是统一的。解析操作是在我们下载的网页中提取我们需要的内容，网站的设计不同，提取的内容不同，相关的操作也存在着很大的差异。3. .时效性：作为一个网络爬虫开发者，我们应该随时关注我们正在爬取的网站的更新动态，一旦网页的结构发生了变化，我们的解析规则可能也需要变化，及时更新爬虫的解析规则，以确保爬虫的正常运行。  随着不断的学习使用，渐渐的接受了python的编码方式，python简单的语法的确带来的极大的便利。我们们不需要去纠结变量的类型问题，了解了python的基本4大组件之后就可以开始我们自己的编程之路。特殊的语法规则大大的减少了代码的数量，严格的缩进要求使得代码的格式更加统一，交接更加便捷。针对于轻量级的任务而言，代码透明影响不大，网络上提供了大量的开源包，使得编程者的工作量骤减。我认为好的爬虫的应该具备以下的特点：1.友好性。爬虫开发者与网站开发者有永无止境的“战争”，爬虫开发者应当充分考虑网站服务器的压力，有好的去反问网站。2. .高性能：每秒钟能够抓取的网页数量是衡量爬虫性能的重要指标。这也是爬虫开发者需要好好考虑的一个问题。如和在保证友好性的同时尽可能的提高自身的效率，还需集体问题具体分析。3.拓展性。.扩展性：爬虫需要抓取的网页数最巨大，即使单个爬虫的性能很高，要将所有网页都下载到本地，仍然需要相当长的时间周期，为了能够尽可能缩短抓取周期，爬虫系统应该有很好的可扩展性，即很容易通过增加抓取服务器和爬虫数量来达到此目的。  学生签字：  日期： |

# 五、实训评价

|  |
| --- |
| **实训成绩：** |
| **教师评语：**  教师签字：  日期： |