2018.7.5 13652887467

1.一个是看计算机网络的书籍 一个是找python web框架的后台教程来做些项目（框架可以选flask 或者是Django）（2018.8达成）

2. git教程(后面在学，2018.11达成)

2018.9

学习django会用到前端知识，所以要学习html，css，js（2018.9.23已经学完，三个星期）

2018.9.18.

分布式爬虫（这个用到多台机器或者一台机器搭建多个服务，有多个端口，并且要搭建服务器，所以暂时不学习），异步（协程），数据库操作，框架爬虫（2018.10.14达成）

2018.9.24

继续学习django框架，包括了html，css，js等前端的内容（2018.10.13达成）

2018.10.1

学习html5的内容（目前不是很需要，达成）

2018.10.14  
爬取ajas加载的网页，学会代理ip反爬（2018.10.19达成）

2018.10.19

加深学习scrapy框架爬虫，结合各种提高爬虫效率的方法去爬虫（2018.10.30达成，scrapy本身就是异步爬虫）

做第二个django项目以及完善第一个django项目的html部分

2018.10.30

通过耗子叔的练级找到接下来要学习的方向（2018.10.31达成--学习数据库和分布式，**数据结构和算法**，web服务器端技术 （LAMP））

粗略学习mysql数据库，不知道这个是不是很重要的数据库（2018.11.03达成）

学习git分布式（2018.11.08达成）

2018.10.31

弄清调试的作用和如何运用，pycharm如何运用（2018.11.05达成）

2018.11.03

学习mysql的导出导入功能（2018.11.10达成）

2018.11.07

学习数据结构和算法，平时看公众号，到一定时候开始刷题

2018.11.10

爬取霞山租房信息（2018.11.12达成）

了解web服务器端技术 （LAMP）（2018.11.11达成，linux + apache + mysql /mongodb+php/python）

2018.11.11

学习自强学堂的基本的web知识（2018.11.14达成）

全面学习崔庆才的爬虫专题，内容有点多（2018.11.15达成）

2018.11.14

利用爬虫下载音乐视频等（爬虫视频或者音乐的url很难构造，需要解密，以后再学）

2018.11.15

学习linux基本操作（2018.11.21达成）

2018.11.18

学习破解爬虫验证码（2018.11.28达成，输入验证码破解，滑动验证码并未破解）

2018.11.21

学习linux shell的命令和在鸟哥的私房菜中加深学习linux（2018.11.27达成，鸟哥私房菜的东西太细了，先不看）

2018.11.24

寻找到日常可以学习计算机基础知识的网站（2018.11.24达成，鸟哥的私房菜和那个讲故事的公众号，补：鸟哥的私房菜先不看了，字体和内容有点看不下去）

2018.11.24

日常学习计算机基础知识

2018.11.27

学习Apache（2018.11.27，apache是一个应用程序，不需要像其他框架学习，懂得配置就行）

了解搭建网站（linux+Apache+mysql+python）需要学习linux的程度和其他知识的掌握程度（2018.11.28达成）

2018.11.28

学习pillow第三方库（2018.12.02达成）

学习爬虫下载网易云歌曲和下载视频（2018.12.03未达成，学习抓包软件和加密算法再来解决）

试一下之前搭过的django网站，确定别人是否可以访问（2018.12.04达成）

加深对lamp的理解（2018.12.06达成）

把文件上传到github（2018.12.05达成）

2018.12.02

了解学习抓包软件如Fiddler、Charles（2018.12.03达成）

学习崔庆才的爬虫可视化监控（2018.12.10达成，完成软件安装和基本操作，教程看不懂达不了监控爬虫）

把谷歌的收藏的链接转到搜狗收藏（2018.12.06达成）

2018.12.06

学习抓包工具charles（2018.12.24达成，需要多多接触）

爬取淘宝的网站（208.12.16达成，异步+mysql爬取成功，可是马上被封）

找到一本爬虫学习书--《用Python写网络爬虫》《 Python 网络数据采集》《Python3网络爬虫开发实战》（2018.12.09达成，先尽快看一下数据采集，再决定需不需要看开发实战）

2018.12.11

学习《Python 网络数据采集》（2018.12.16达成）

学习数据库redis（2018.12.25达成）

2018.12.13

每周六更新github

日常看计算机基础包括算法结构和码农翻身等的基础

每周至少写一个爬虫，档次要越来越高

2018.12.16

用另外一种方法爬取淘宝的评论（2018.12.22达成，多进程爬虫加代理还是会被封）

按照网上教程用lamp搭建服务器（2019.01.05达成）

2018.12.24

爬虫获取几篇小说的内容（2019.01.04达成）

结合算法公众号学习《数据结构 C语言版第2版 人民邮电出版社》这本书

2018.12.30

学习正则表达式，解决爬取小说遇到的问题（2018.12.31达成）

2018.12.31

6月份前要开始学数据分析（2019.02.12达成）

6月份前要开始做减法，学精（2019.02.12）

6月份前要开始刷算法题（2019.02.12达成）

6月份前要开始看源代码

2019.01.01

弄明白ping的原理，做笔记（2019.01.11达成）

加深学习python中的类，类中的init函数部分（2019.01.06达成）

2019.01.05

弄明白LAMP的基本原理（2019.01.08达成）

运用云服务器挂微信自动发送信息的程序（2019.01.10达成）

爬取斗鱼的弹幕（未达成，关系到socket等知识，以后接触再学）

2019.01.06

整理最近的知识，做好笔记（2019.01.08达成）

2019.01.08

学习TCP/IP协议，写出基本过程（2019.01.08达成）

运用云服务器搭建vpn（2019.01.12达成，需要买其他地区的服务器才能翻墙如vultr一个国外的服务器）

学习了解socket（2019.01.11达成）

学会终端连接数据库来管理数据库（2019.01.12达成，平时要多练习）

2019.01.11

学习了解MAC等基本协议（2019.01.21达成）

2019.01.12

做出requests爬虫模板，带有headers，cookies，代理，数据库保存，txt保存，excel保存（2019.01.16达成）

2019.01.13

使用云服务器找出一个有意义的项目（2019.01.13，爬取某个网站每天更新的东西）

过年前完成看完前面那本数据结构的书、所有的爬虫模板，然后准备开始崔大的书本学习或者鸟哥的私房菜的学习（2019.02.03达成）

使用云服务器爬取每天新增加的妹子图的图片（2019.01.20达成）

2019.01.14

找到几本好书或者专栏（2019.01.14达成，高性能mysql、极客时间的mysql的专栏、linux的鸟哥私房菜、unix操作系统设计）

了解前端的一些框架（2019.01.15达成，先jQuery、再Node.js）

2019.01.14

解决FZ工具下载上传文件失败的问题（2019.01.16达成，主要是文件权限问题，改成root用户或者chmod获取权限，有时候也是不知道什么原因一天行一天不行）

2019.01.16

设置Linux系统root用户的密码（2019.01.21达成）

学会设置pycharm的默认python运行版本（2019.01.21达成）

2019.01.20

补充最近一个星期的笔记（2019.01.21达成）

做出正则表达式的爬虫模板，带有headers，cookies，代理，数据库保存，txt保存，excel保存（2019.01.22达成）

2019.01.21

申请免费的腾讯云服务器（2019.02.02未达成，只有15天所以考虑亚马孙等）

弄明白网络传输的五个层（2019.01.23）

2019.01.22

爬取天眼查或者企业猫或者京东的数据（2019.01.29达成，爬取天眼查的企业信息）

了解是否可以利用终端工具连接本地电脑（2019.01.25达成，powershell）

2019.01.23

学会通过ip远程连接别人的电脑（2019.02.02未达成，找不到原因）

写出selenium+firefox爬虫的模板，带有headers，cookies，代理，数据库保存，txt保存，excel保存（2019.01.26达成，不过cookies还不太会也不太需要）

2019.01.26

写出多进程、异步requests+lxml爬虫（2019.01.29达成）

2019.01.27

解决time函数的问题（2019.01.30达成）

2019.01.28

做好上周的笔记、总结（2019.01.30达成）

2019.01.30

写出scrapy爬虫模板，带有headers，cookies，代理，数据库保存，txt保存，excel保存（2019.01.31达成）

整理笔记，做减法（2019.01.31达成）

把谷歌的收藏转移到搜狗（2019.01.31达成）

关注极客时间的mysql的专栏（2019.02.02未达成，需要钱应考虑）

2019.02.11

学完崔庆才的网络爬虫书籍

开始在LeetCode刷题和看数据结构和算法公众号（2019.02.11达成）

把煎蛋妹子图的项目挂在另一个服务器运行（2019.02.14未达成，没有服务器，之前也已经达成了）

申请亚马逊或者其他的免费服务器（2019.02.13未达成，需要信用卡）

2019.02.14

爬取去哪儿的信息，运用urllib库来爬取（2019.03.03未达成，selenium爬取加载不了）

开始详细学习鸟哥的私房菜的linux

复习一遍笔记（2019.02.20）

2019.02.15

完善代理ip的py文件（2019.02.21达成）

2019.02.16

完成网络爬虫开发实战中猫眼电影的爬虫（2019.02.26达成，熟悉正则表达式的目的达成）

完成网络爬虫开发实战中第三章的笔记补充（2019.02.18达成）

2019.02.17

下载安装崔大推荐的几个工具，包括关于xpath的工具（2019.02.25达成）

2019.02.18

写出利用爬虫协议robots.txt判断哪些部分是否可以爬取（2019.02.23达成）

安装插件将鸟哥的私房菜转化为简体字（2019.02.20达成）

2019.02.19

完成网络爬虫开发实战第四章笔记补充（2019.02.22达成）

2019.02.20

完成网络爬虫开发实战第五第六章的笔记补充（2019.02.23达成）

2019.02.21

完善代理ip的笔记以及爬虫模板（2019.02.22达成）

2019.02.22

整理C盘的py文件（2019.02.25达成）

2019.02.26

安装可以打开json的软件（2019.02.27达成，nope、pycharm都可以）

结合网络爬虫开发实战，运用selenium爬取淘宝的相关物品信息（2019.03.02达成）

2019.03.02

补充上个星期的笔记（2019.03.03达成）

2019.03.03

总结爬取过的网址的特点（2019.03.19）

继续完善爬虫模板特别是保存部分（2019.03.26达成）

2019.03.04

复习一下数据结构这本书

2019.03.05

完成tesserocr的安装（2019.03.10未达成，用了pytesseract来代替，效果差不多）

爬取一亩三分地这个论坛（2019.03.10达成）

学习完4种破解验证码的方法（2019.03.14达成，最后一种微博9宫格没有了，不需要再学习）

2019.03.06

复习一遍写过的爬虫（2019.03.19达成）

2019.03.09

写出正则表达式匹配测试的模板（2019.03.10达成）

整理最近的笔记（2019.03.10达成）

2019.03.10

学习try…except中的报出具体错误内容（2019.03.15）

2019.03.13

爬取珍爱网的数据（219.03.23达成，开始尝试爬取尽可能多的数据）

2019.03.15

完成验证码的笔记补充（2019.03.25达成）

2019.03.17

写出生成词云的模板（2019.03.19）

2019.03.18

爬取搜狗微信站点的微信文章（2019.03.30没有达成，ip池不高效）

2019.03.26

爬取咸鱼和转转的在售产品类别，价格，过往价格（部分有），地理位置，词云

完成网络爬虫开发实战的ip池维护章节的学习（2019.03.30达成）

2019.03.30

补充最近的笔记（2019.03.31达成）

2019.03.31

学习flask（2019.04.14达成）

完成网络爬虫开发实战的cookies模拟登陆章节的学习（2019.04.02达成）

2019.04.07

学习fiddler和charles（2019.04.10达成）

完成pyspider的学习（2019.04.09达成）

2019.04.08

学习pyquery，了解基本的功能（2019.04.09达成）

2019.04.09

补充最近的笔记（2019.04.10达成）

2019.04.10

学会获取路径创建文件夹的函数（2019.04.11达成）

Pyspider爬取淘宝妹子图（2019.04.14达成）

2019.04.14

再次系统学习scrapy这个框架（2019.04.28达成，已经读懂部分源码）

补充最近的笔记（2019.04.15达成）

2019.04.16

学一种画思维导图的软件（2019.04.26达成，xmind）

2019.04.27

补充scrapy对接selenium的模板（2019.04.28达成）

复习redis（2019.05.05达成）

学习分布式爬虫（2019.05.05达成）

完成scrapy的笔记（2019.04.28未达成，直接看scrapy模板比较好，容易理解）

做好字典的笔记（2019.04.27达成）

2019.04.28

了解js逆向反爬、破解和app脱壳（2019.05.06达成）

了解notion软件（2019.04.28达成）

2019.05.06

开始学习数据分析（2019.05.06达成）

学习哈希base64加密解密（2019.05.11达成）

学习js（2019.05.14达成）

复习一遍爬虫（2019.05.07达成）

加深mysql的学习

开始对爬虫的各个重要模块的加深学习（2019.05.24达成）

找出很棘手的几个网站，以后尝试一个一个攻破爬取（2019.05.09达成，有**淘宝，拉勾，咸鱼，转转，12306，京东，拼多多，新浪，微信站点，煎蛋妹子图，协程，后面陆续增加，大众点评，微博，私募排排，66ip**）

2019.05.09

弄清楚常用的编码方式（2019.05.15达成）

复习一下C语言（2019.05.12达成）

继续爬取咸鱼网，数据要尽可能大（2019.05.19达成，已基本爬取全部数据）

补充最近爬取过的网站到笔记（2019.05.15达成）

掌握破解煎蛋妹子图的图片链接的方法（2019.05.25达成）

2019.05.11

补充哈希加密和base64编码解码的笔记（2019.05.15达成）

加深学习爬虫的ip代理、ip池模块（2019.05.31达成）

2019.05.13

学习pandas和numpy这两个模块（2019.05.22达成）

了解jupyter（2019.05.14达成）

学会创建不同环境（2019.05.22达成，Anaconda实现）

2019.05.19

在多进程或者异步中实现计数功能（2019.05.22未达成，因为多进程或者异步各自都是独立的进程相互数据不连通，可以从主进程设置一个判断函数实现但是不现实）

2019.05.20

爬取转转网能够爬取的数据

2019.05.22

完成最近的笔记（2019.05.25达成）

2019.05.25

开始使用notion做笔记（2019.05.31达成）

2019.05.28

找出最常用的免费代理网站（2019.05.30未达成，常用的免费代理其实都用过了，可用率最多10%）

不使用selenum的情况下爬取66ip的ip和端口（2019.05.31达成，此前是因为多次爬取被封了ip）

2019.05.29

完成Anaconda的笔记（2019.05.31达成）

试一下熊猫代理

2019.05.31

完成ip代理部分的笔记与总结以及爬虫模板的补充

开始加深学习爬虫的储存部分

2019.06.03

学会浏览器的调试功能