云顶书院Python方向一期培训内容

1月19日 简单认识爬虫，做技术铺垫，了解HTML基本结构

1月20日 学习爬虫基础，Urllib模块的使用

1月21日 爬虫第一次基础练习：百度贴吧内容爬取练习

1月22日 深入了解http请求，cookie，session等原理

1月23日 掌握re，Beautifulsoup等模块的使用

1月24日 爬虫提升练习：豆瓣书籍top100爬取

1月25日 mangoDB的基础安装与使用

1月26日 python对于mango的基本操作

1月27日 python与fiddler学习

1月28日 巩固练习：手机app内容爬取并设计数据表结构，把信息存入表中

1月29日 Selenium模块的学习

1月30日 python模拟浏览器行为

1月31日 为期两天的一期综合练习：

2月1日 对QQ空间进行爬取，并设计自己的数据库确定信息之间的属性，将信息存入数据库候整理输出

学习材料

爬虫：<https://www.cnblogs.com/xin-xin/p/4297852.html>

链接：<https://pan.baidu.com/s/1MrztIyGhQwPcVui4leIN1Q> 提取码：urxp

上面教程中应该学到的爬虫部分

<https://www.cnblogs.com/miqi1992/p/7828886.html>

mango：<https://blog.csdn.net/mmayanshuo/article/details/79051768>

<https://www.cnblogs.com/shhnwangjian/p/6247473.html>

fiddler：<https://blog.csdn.net/a877415861/article/details/79447440>

<https://blog.csdn.net/qq4165498/article/details/77802642>

Selenium：<https://blog.csdn.net/hgdzw/article/details/82789889>

<https://blog.csdn.net/hgdzw/article/details/82789889>

<https://www.cnblogs.com/hanxiaobei/p/6108677.html>

模拟浏览器：<https://www.cnblogs.com/sesshoumaru/p/python-selenium-webdriver.html>

<https://blog.csdn.net/weixin_38641983/article/details/80830129>