# 新冠病毒项目 相关资料总结

## 1.爬取数据

1. Python 爬虫原理和数据抓取的介绍

<https://www.cnblogs.com/miqi1992/p/7828889.html>

1. 资料一，获取政府网站公示数据并且保存到MongoDB数据库。

<https://blog.csdn.net/xHibiki/article/details/84138968>

这里使用requests和lxml爬取, re清洗数据, multiprocessing实现多进程爬虫, 共爬取了99页的数据, 通过Pymongo存放到MongoDB中。

1. A screenshot of text

   Description automatically generated找到的几个表（可以看看梳理下）

A screenshot of a cell phone

Description automatically generated

A close up of a map

Description automatically generated

1. 正则表达式

<https://www.runoob.com/regexp/regexp-syntax.html#:~:text=%E6%AD%A3%E5%88%99%E8%A1%A8%E8%BE%BE%E5%BC%8F%E6%98%AF%E7%94%B1,%E7%9A%84%E5%AD%97%E7%AC%A6%E4%B8%B2%E8%BF%9B%E8%A1%8C%E5%8C%B9%E9%85%8D%E3%80%82>