

搜集的资料

4/16-4/17

数据方面：

- 腾讯疫情平台：<https://news.qq.com/zt2020/page/feiyang.htm#/>
- <https://cov19.cc/>
- 丁香园：<https://ncov.dxy.cn/ncovh5/view/pneumonia>
- 约翰霍普金斯大学疫情dashboard：<https://coronavirus.jhu.edu/vaccines/international>
- 约翰霍普金斯大学dashboard的数据仓库：<https://github.com/CSSEGISandData/COVID-19#covidd-19-data-repository-by-the-center-for-systems-science-and-engineering-csse-at-johns-hopkins-university>
- 参考：<https://article.itxueyuan.com/666glq>
- WindQuant 新冠肺炎数据Web API：<https://www.windquant.com/qntcloud/help/id-fdc2e335-7f50-4b63-b79c-07a2582cf15c#%E5%9B%BD%E9%99%85>
- 其他数据来源：WHO、卫健委、ourworldindata（疫苗）、其他新闻媒体发布平台等
- 新冠肺炎API：<https://github.com/ExpDev07/coronavirus-tracker-api>
- nCoV2019：<https://ncov2019.live/map>

问题：

数据来源很多，有官方网站、相似类型平台、Web API，

同类型的数据可能有多多个数据源，

我们是选一个来源？还是多个来源？

我们是爬取官网数据，还是直接调用他人的API？还是自行抽象建立API？

对于不同来源的数据我们该如何综合处理？出现同类型的多份数据源该怎么处理？是选择一份？还是多份综合分析？如果是只选一份的话，如何选择（如何确保数据的准确性）？如果是综合分析的话，如何综合分析（建模？设计一套算法？）？

技术方面：

- 爬虫爬取疫情大数据案例：<https://blog.csdn.net/Eastmount/article/details/104298388>
- 一个完整的疫情可视化平台（不涉及疫苗）的项目：<https://github.com/zhanqingheng/COVID-19>
介绍了项目开发的全过程：<https://zhanqingheng.com/2020/06/08/7covid19/#more>
可以借鉴技术路线与实现过程

其他同类作业（用于启发思路，参考价值不大，不必过于纠结）：

- 浙大某数据库作业，项目有所不同，有情感分析：<https://github.com/yunwei37/COVID-19-NLP-vis?spm=a2c4e.10696291.0.0.12e419a4LTGmM2>

可以参考爬虫与技术栈

作业部分相关技术栈

- 数据库的构建与 sql 语句查询
- pyEchart 可视化

- postgres sql 数据分析
- snowNLP 情感分析
- jieba
- flask
- 新冠疫情数据展示网站: https://github.com/DiangD/epidemic_COVID-19

技术栈

- SpringBoot + Thymeleaf + MyBatis
- ECharts 基于js的图表绘制
- Quartz 后台定时任务
- Jsoup爬虫
- 前端: AmazeUI (妹子UI)

功能介绍:

- 国内疫情的数据展示
 - 国外疫情的数据展示
 - 实时快讯的展示
 - 一点点的疫情数据可视化 (基于ECharts的图表)
- ncov-globe: 技术路线:
Back-end :
 - Flask
 - Python Scraper works every 2 hours. World data is from bing and Chinese data is from Dingxiangyuan.Front-end :
 - webgl globe
 - javascript

编程语言: python

pyecharts

爬虫爬取数据/直接调用API

web框架: flask

前端: vue, echarts

其他有用的资料

原型设计参考: <https://edu.cnblogs.com/campus/fzu/2020SpringW/homework/10400>