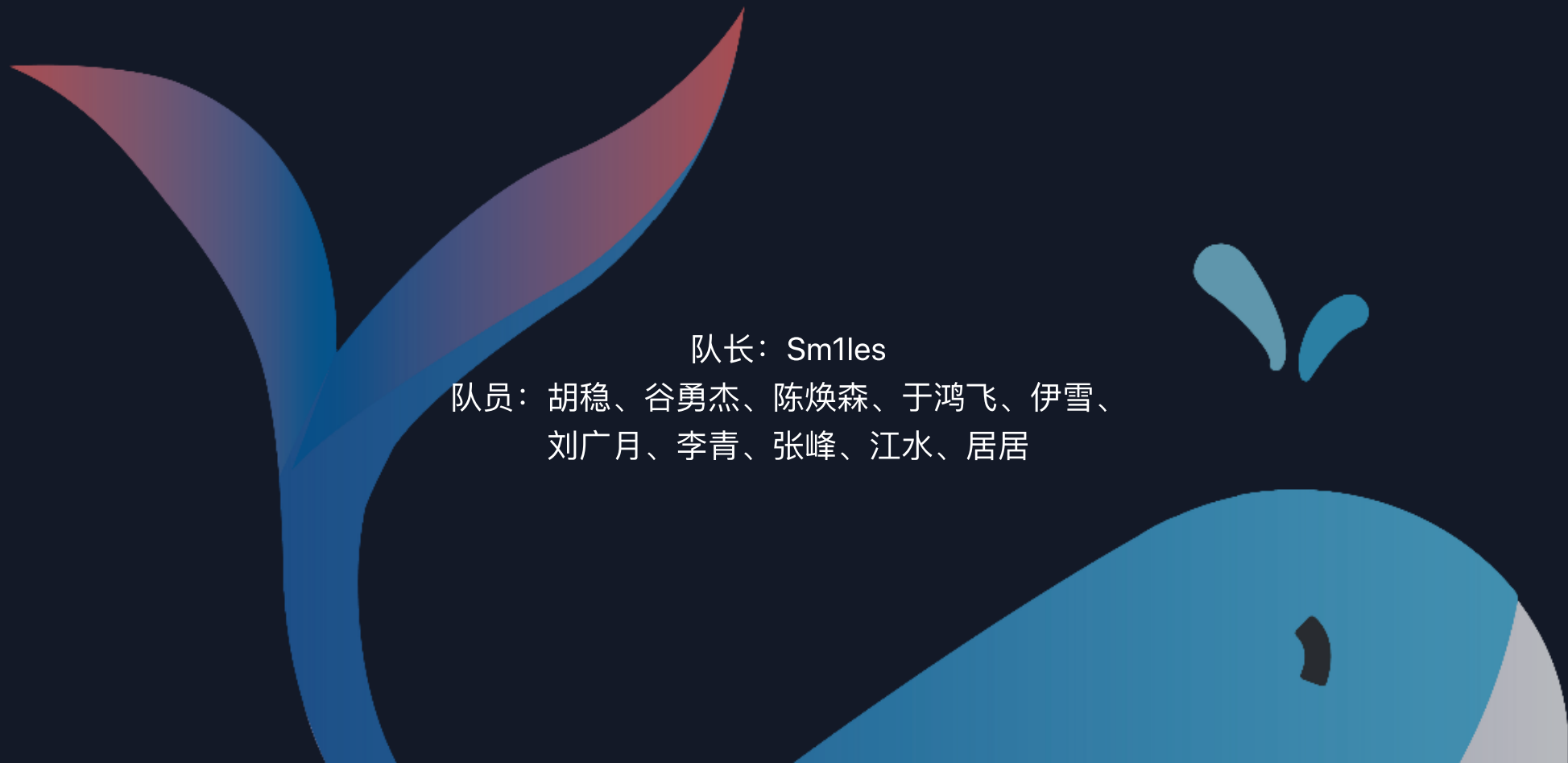


一站式疫情数据开放平台

队长: Sm1les

队员: 胡稳、谷勇杰、陈焕森、于鸿飞、伊雪、
刘广月、李青、张峰、江水、居居





目录

contents

项目背景

拟解决的问题

可行性分析

作品展示

关于我们



公开透明的疫情数据无论是用来警醒世人，还是用作科研分析，其本身都具有极高的实用价值，哪怕是仅仅当做一段历史，它也有记录的价值，毕竟你现在还能找到2003年的那段历史吗？



- 问题1：国家卫健委和WHO公布的结构化疫情数据粒度太粗，无法深入分析，细粒度的疫情数据由各地卫健委等官方渠道发布，但是各地发布的渠道和格式不一，无法用爬虫自动化爬取结构化的数据，甚至有些数据必须人工推理才能得到，因此收集数据成为阻碍众多开发者和科研工作者的第一个绊脚石
- 解决方案：我们采用的是开源协作的方式，将所有对数据感兴趣的人汇集到本平台，然后由本平台统一制定出数据结构化的方案，接下来大家按照统一的方案分工协作就能快速收集并将全国各地的数据结构化到一个统一的格式，最后使用数据的时候由本平台按照大家的需求统一分发



- 问题2：不同方向的研究人员对数据的格式要求不同
- 解决方案：尽管不同方向的研究人员对数据的格式要求不同，但是其原始数据都是一样的，都是来自官方通报的内容，所以在本平台在收集结构化数据的时候也要求将非结构化的原始数据也收集进来，并且本平台还提供了自定义标签的功能，这样的话，有特殊需求的研究人员可以用此功能结合原始数据进行二次标注



- 方案可行性：据调研了解，目前国内还没有开源免费的细粒度疫情数据集公开，其主要原因是人工标注成本太高，所以数据持有方不愿意免费放出，而很多研究人员无奈于个人能力有限，不得不放弃自行收集，我们在Datawhale的社群进行数据收集人员招募时也印证了我们的猜想。
- 技术可行性：本平台采用Web开发的方式即可满足所有需求，目前我们已经采用Django+Vue的技术架构完成了整个平台的开发工作；
- 商业可行性：本项目属于开源的公益项目，目前无商业目的；



网址: <http://47.93.61.57:8080>

测试账号: 15768658888

测试密码: 123456



一个专注于AI领域的开源组织

