

復旦大學

本科毕业论文



论文题目： 基于网络爬虫和数据挖掘技术的
新冠疫情可视化分析

姓 名： 曲俊杰 学 号： 16307110125

院 系： 软件学院

专 业： 软件工程

指导教师： 徐迎晓 职 称： 高级工程师

单 位： 复旦大学

完成日期： 2020 年 月 日

目录

目录	1
摘要	2
ABSTRACT	3
第一章 引言	4

摘要

2020 开年，新冠疫情肆虐全球，给人们带来了空前的恐慌，短短三个月内，数以百万计的居民生命安全遭受威胁，亿万人的生活受到不同程度的影响，中国作为 14 亿人口的泱泱大国，充分使用了应对 2003 年非典疫情时的经验，采用地区强制隔离等政策抗击疫情，全国上下齐心协力，共同战疫。中国人民抗击疫情的坚定决心和显著成效给世界各国极大的鼓舞和信心。

项目将从当前世界关于新冠疫情的研究入手，简述该项目的知识背景和实现意义，通过网络爬虫的技术手段收集整理疫情数据，在保证数据准确性的基础上，使用可视化技术对数据进行多维度的直观呈现，并使用 echart 制图技术和 javascript 技术结合搭建每日疫情上报页面，使用机器学习相关技术做疫情发展趋势的简单预测。

此外，项目也对疫情中媒体的作用予以关注，使用相关度判别方法筛选整理出近千篇微信公众号文章，使用中文分词程序结合疫情中出现的具体问题作词频分析，利用 wordcloud 技术生成新冠疫情词图，揭示新冠疫情中媒体视角下的关注热点。

论文将对项目开发具体过程中遇到的问题和解决方案做一个总结，并对项目可供延伸改进的地方做一些猜想和尝试。

关键词 新冠疫情，网络爬虫，数据可视化，词频分析，机器学习

ABSTRACT

第一章 引言