An Interactive Platform to Track Global COVID-19 Epidemic

Zhenhe Pan*, Dhruv Mehta*, Anubhav Tiwari*, Siddhartha Ireddy*, Zhou Yang† and Fang Jin†
*Department of Computer Science, Texas Tech University
Email:{zhenpan, Dhruv.C.Mehta, Anubhav.Tiwari,siireddy}@ttu.edu
† Department of Statistics, George Washington University
Email: zhou_yang@gwmail.gwu.edu, fangjin@email.gwu.edu

***Abstract*—This project built a world-wide database of coronavirus cases, which helps to model the spread of the coronavirus disease (COVID-19), and to identify policy and social factors that impact the spread of COVID-19. Four essential tasks are implemented: 1) build a comprehensive database of coronavirus cases world-wide; 2) visualize the heatmap of confirmed cases for each country, provide detailed spreading trends for each countries and comparison among countries; 3) collect tweets about COVID-19 in real-time and extract people's daily concern flow; 4) integrate breaking news such as first confirmed/death case in each country. This demo will provide decision-makers with accurate data-driven representations in an easy to understand format that enables them to make more timely and cost-effective preparation and response plans.**

I. Introduction

COVID-19 is spreading rapidly and has already affected more than 169 countries, infecting near a million people and causing more tens of thousand deaths around the world (as of April 1, 2020). The collection of coronavirus cases information such as positive diagnoses, recovery, and death, is vital as it provides a foundation for modeling the disease evolution dynamics, allowing us to infer future trends, and analyze the factors that influence the speed at which the virus spreads. This information will be crucially important if we are to develop appropriate policies to manage this pandemic. A number of platforms [1]–[7] developed to predict and provide people an updated information about coronavirus as this outbreak unfolds. These work increased people's awareness of this disease and advanced the state of the art by developing accurate methods and effective tools to collect coronavirus cases from mining online platforms. However, most of the currently available coronavirus databases have some limitations: they either contain only the current numbers of cases and their locations, completely ignoring the historical information, or only cover certain countries/regions, or lack of timely update. Moreover, the pandemic-related data is stored in multiple unconnected resources (websites, news reports, social media posts, and ad-hoc databases) and recorded in many different languages. This lack of inter-database connectivity, along with the associated language barriers, significantly hamper our efforts to deal with this dangerous pandemic. There is thus an urgent need to develop a comprehensive database to identify and track coronavirus cases world-wide, with the ultimate goal being able to halt the spread of this and future unknown infectious disease outbreak more effectively and efficiently.

In this paper, we developed an interactive visualization platform [1] to closely monitor the global COVID-19 situation and the social media & news reports. Besides showing the latest COVID-19 cases numbers, the platform preserves the historical records and enable users to see the COVID-19 spreading trends. In addition, the platform collects and digests the tweets streams and grab people's top concern, providing users a One-stop experience of the overview COVID-19 situation. To summarize, our platform has the following contributions:

- Developed a comprehensive database of coronavirus cases worldwide, which would tremendously helps researchers and policy makers to model the spread of this infectious disease timely.
- Provided a Bird's-eye view of the global COVID-19 dynamics, not only showing the latest confirmed cases numbers, but also allow users to view the COVID-19 trends for each country and state/province.
- Developed practical tools to collect information on coronavirus cases world-wide by mining multiple online platforms.
- Integrated breaking events and people's concern flows into the map, providing a full picture of the COVID-19 dynamics.

II. System framework

Figure 1 illustrates the system framework of our platform. The framework is divided into three main components: (a) A front-end component which collects real-time data from multiple websites and social media platforms; (b) A database storage engine that enables real-time update and historical data retrieve; (c) A back-end component that includes multiple algorithms for tweets filtering, concern flow extraction, and breaking news identification.

A. Data Flow

As shown in Figure 2, our input data has three types: real-time COVID-19 cases for each country/state, coronavirus related tweets discussions, breaking news from multiple websites. The real-time COVID-19 case numbers were collected

[1] http://worldcovid-19stats.000webhostapp.com/covid.html#trends

IEEE/ACM ASONAM 2020, December 7-10, 2020
978-1-7281-1056-1/20/$31.00 © 2020 IEEE

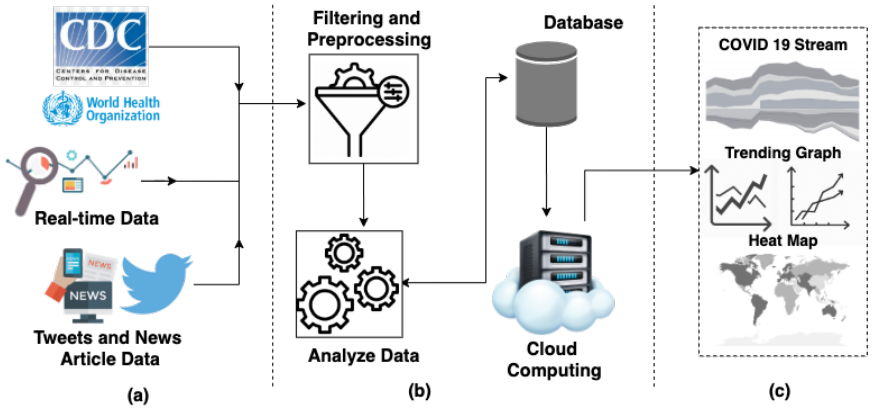

Fig. 1. Framework of the platform.

from John Hopkins Coronavirus Research Center [2] and the World Health Organization (WHO). We developed a web collector to fetch the dataset in a live stream. Specifically, it will browse the two websites (John Hopkins and WHO) to collect the most recent COVID-19 time series, e.g., new confirmed cases, new death case, and etc. WHO covers more countries' cases information, while John Hopkins provides the latest updated COVID-19 dynamics. The tweets were crawled from public Twitter API in real-time. We developed a twitter crawler which collects the tweets whose hashtag or content related to coronavirus, such as "#Coronavirus", "#COVID19", "#CoronavirusPandemic". These collected tweets are then fed into a python-based analyser, which segregates the tweets based on the location and the language, and also sorts them based on the retweets number, such as likes and user followers. After that the tweet content is compared against the news articles collected from trusted sources, such as the Centers for Disease Control and Prevention (CDC), WHO and etc. to test their credibility. The announcements or news articles are collected using the respective APIs from (CDC) and WHO and using a web scraper for the media outlets like CNN.

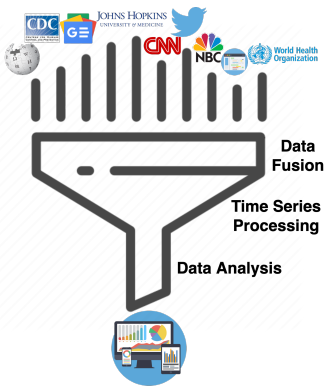

Fig. 2. Demonstration of data flow.

B. Database Design

We choose Mysql to store the confirmed case time series, and mongodb to store and search tweets and various news. Three tables are designed respectively. Confirmed case time series are stored in table "case_time_series_table", tweets are stored in table "tweet_table", and news information are stored in table "news_table".

- *Case_time_series_table*. The attributes in this table include "country", "state", "city", "zipcode", "date", "new infected number", "new death number", "total infected number", "total death number", "cured number", and etc. We store each country's time series into one single table, to reduce the data redundancy and increase database's robustness. Tables for countries with large number of infected patients, such as China, the U.S., Italy, Germany, and Spain, are normalized to third normal form/ For the rest countries, we decide to keep all the confirmed case information into one table, because it does not contain huge volume amount of data. However, because the frequent query and comparison, we create many views to connect these tables such as compare today's confirmed case numbers from each country.
- *tweets_table*. The attributes in this table include "country", "post_date", "state", "country", "city", "user", "contents", "hashtag", "url", and so on. Every one hour, the tweets crawler will capture the most recent coronavirus related tweets. These first hand raw twitters will be stored locally.
- *news_table*. The attributes in this table include "date", "title", "content", "sources", "url", "author", and "others". Similarly, very one hour, web crawler will capture the most recent news using some key words, such as "COVID-19", "coronavirus" and etc.

III. User Interface

This platform offers several main functions to demonstrate the covronavirus dynamics in every aspect. As shown in Figure 3, the geographical heat map shows the confirmed COVID-19 cases distribution. Trends of the confirmed cases

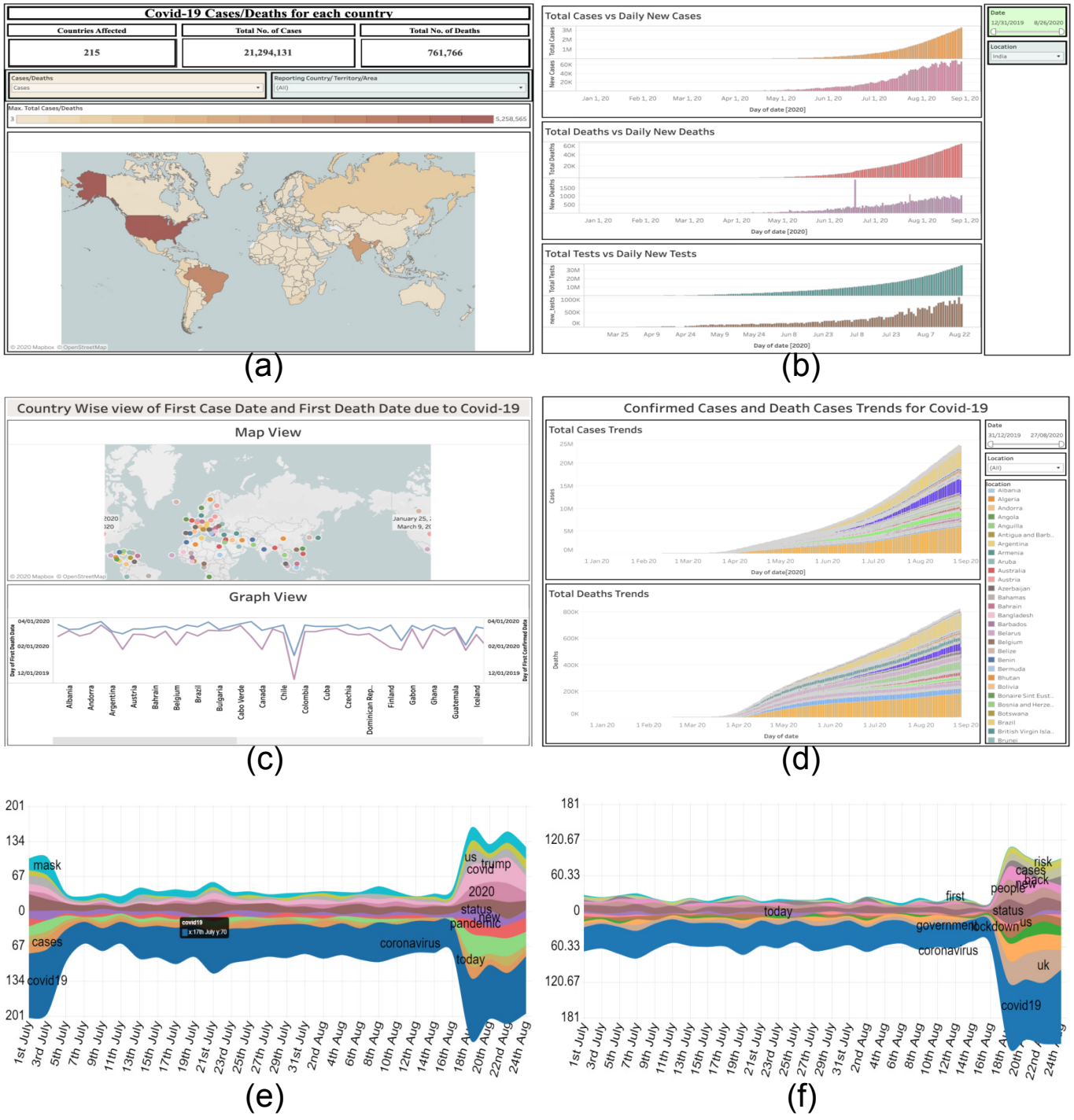

Fig. 3. Interface of our platform. (a).Heat Map of the confirmed COVID-19 cases; (b).Zoom into certain country/state of COVID-19 cases and Deaths; (c).Breaking news of first COVID-19 cases and Deaths in each country; (d).Trends of the confirmed COVID-19 cases distribution among countries; (e) and (f) are Tweets concern flow from USA and UK respectively.

provides users the perspective to look back, and users can choose a specific country to zoom in. The breaking news of first confirmed case and death for each country are visualized on map. Each country's contribution to the global COVID-19 are ranked in bar chat. Besides, people's daily concerns with the propagation of the virus are extracted from tweets.

A. Heatmap of the confirmed cases

As shown in Figure 3(a), the landing page gives an overview of the Covid-19 heatmap for each country, showing number of countries affected, total number of cases and deaths. With the color range from light yellow to dark green, it represents the accumulated case number levels, from light to serve. With the mouse hang over a county/region, it will highlight the contour of this country/region and a rectangle box will give the country/region's name and its confirmed cases/deaths.

B. Trends of the confirmed cases

With the COVID-19 outbreak unfold, people are not only interested in the latest case number, but also want to trace back the trends of the infections and work on possible ways to bend the curve. So we offer two scenarios to look back trends. This fist function is shown in Figure 3(b), which is a detailed trends description for each country/state about four categories time series: total confirmed cases/daily confrimed cases, total deaths/daily deaths, and total tests/daily tests. In the drop-down menu, users can select a country they are interested in, it will the corresponding visualization. The date range menu allows users to specify the trends start time and end time. The second is shown in Figure 3(d), we pack each country's case number into one bar chat, and visualize the overall time series of confirmed cases, but distinguish each country's contribution by different colors. Users can choose include or exclude mainland of China as its fight for this disease is getting close to end.

C. Breaking news for each country

To ingest breaking news articles, we have developed a web scraper which can crawl through some popular news portals and collect required data like first confirmed and death dates in different countries. Figure 3(c) shows the map integrating first case&death events for top 80 countries (ranked by case numbers). The colorful circle represents one country and detailed information will show up when the mouse hang over that country. The bottom diagram provides timeline of the first confirmed case and death for each country.

D. Concern flow from Tweets

The objective of concern flow [8] is to monitor people's concern and illustrate the most frequent words during the pandemic period. The tweets contained spatial information from city or state or country level, and also people's discussion related to COVID-19. Following the method in [8], we plot people's concern flow in USA and UK as shown in Figure 3(e, f). The graph represents repeated concerns over time. The more discussed concerns are presented relatively proportional with the thickness of represented country in the graph.

E. Accessing Interface for shared Database

To make the database accessible to users (government officer, disease expert, software developer, etc.), we designed an application programming interface (API). The API will provide interface for application developer with more advanced level of access to build applications related to coronavirus cases. It makes our database more utilized addressing other demand to support the community. The API will return output as JSON format for humans to read and write, and machines to parse and generate easily. This shared database will promote efforts by research communities, governmental bodies, schools, industries, and private sector organizations to work together to fight the spread of the virus.

IV. Conclusions

We developed an interactive platform to track global COVID-19 in real-time, at the same time, we monitor multiple social media and news report platform to integrate the latest events about coronavirus, providing users a bird's eye view of the global pandemic dynamics. This platform enables a variable ranking mechanism to look into the global COVID-19, also allow users to go into details of state/province level to check the latest COVID-19 cases, deaths, recoveries. What is more, it preserves historical data and show the trends. Besides, this platform is also one of the first data driven approaches in location based event identification for COVID-19 situation awareness from social media.

Given the popularity and impact of the dashboard to date, we plan to continue hosting and managing the tool throughout the entirety of the COVID-19 outbreak and to build out its capabilities to establish a standing tool to monitor and report on future outbreaks. Besides this, we plan to continue extend this platform by ingesting more data-driven analysis such as statistic models of infection types, infection paths, with the goal to enrich people's knowledge and understanding about the threaten disease. We believe our platform will provide more insight about people's concern about COVID-19, which helps decision-makers, such as government, community and organizations, to evaluate the urgency degree in the near future and make cost-effective policies. Beyond that, the data-driven analysis can help academic researchers to understand the disease spreading spatial dynamics and temporal evolution.

References

[1] "2019-ncov," https://yiqing.ahusmart.com/, 2019.

[2] E. Dong, H. Du, and L. Gardner, "An interactive web-based dashboard to track covid-19 in real time," *The Lancet Infectious Diseases*, 2020.

[3] Mistletoer, "China-covid-19," https://github.com/Mistletoer/COVID-19-historical-data-visualization-2019-nCoV-.

[4] tk, "covid19," https://covid19.tk/, 2019.

[5] Q. C. R. I. Tsinghua Database Research Group, "Covid-19 dashboard," https://ncov.deepeye.tech/, 2019.

[6] Z. Yang, J. Xu, Z. Pan, and F. Jin, "Covid19 tracking: An interactive tracking, visualizing and analyzing platform," *arXiv preprint arXiv:2008.04285*, 2020.

[7] Z. Pan, H. L. Nguyen, H. Abu-gellban, F. Jin, and Y. Zhang, "Google trends analysis of covid-19," *IEEE Transactions on Big Data*, 2020.

[8] L. Nguyen, Z. Yang, J. Li, Z. Pan, G. Cao, and F. Jin, "Forecasting people's needs in hurricane events from social network," *IEEE Transactions on Big Data*, 2019.