Covid-19 Datafetcher Using ParseHub Tool with Voice Assistance

Mariam Fatima[1], M Ganeshan[2]

[1]Master of Computer Application, [2]Associate Professor,
[1,2]Jain Deemed-to-be University, Bengaluru, Karnataka, India

ABSTRACT

Main purpose of Web Scraping is to abstract information from single or N websites and then process it into plain structures such as spreadsheets, CSV file or database. Although, in adding to be a very convoluted task, Web Scraping is source and tiresome, primarily when it is carried out manually. Now in this pandemic situation it is very important for all the people have a look at the live cases which are spreading rapidly. Parsehub will help us to fetch the data from the websites more accurately and easily. By using parsehub tool researchers can collect real-time data of covid-19 cases which can be easily extracted from anywhere. Parsehub uses machine learning for its relationship engine which is state-of-the-art.

KEYWORDS: Parsehub, web scraping, Covid-19, real-time data.

1. INTRODUCTION

What is going to be the worldwide effect of the coronavirus? Answering to this question needs exact prognostication the blowout of confirmed cases and as well as survey of the number of demises and convalescence. In paper [1] they have described the chronology of a live foretelling exercise with enormous probable inference for drafting and decision making and offers objective forecasts for the confirmed cases of coronavirus. Using this we can get the cumulative daily figures collected globally of the three main variables of interest: deaths, confirmed cases and recoveries. In paper [2] they have used social media as a means to collect the data of covid-19 cases. But it also says that social media framework is a platform for misinformation to spread rumors.

In paper [3] here they have used data driven prediction method to find the cases in China, Italy, South Korea and Iran. However, the data driven model might not work or perform reasonably if it is put in to a new epidemic which has a notably non-identical set of spreading features, in a populace or the country having a notably different topology of contact, community behavior, travelling forms, efficacy of control also as climate.

From Covid-19 DataFetcher Using ParseHub Tool With Voice Assistance we get the knowledge about novel disease coronavirus with the help of latest tool best known for its data gathering (Scrapping), with ParseHub tool we will be able to collect the data about the cases and situations going on in several countries through which public will be aware of increasing cases, deaths, active cases, predictions and so on.

Using ParseHub we can combine the data from different sites and get the result in the form of csv files, excel and Json format. This data after getting processed will be fetched as results to the system so when the user asks about any related information about the coronavirus cases the data will be fetched and the results will be in audio format. We will be utilizing voice recognition to tell the results from the user query.

With this we will be able to contribute the society with the accurate data through which we can stop the virus from spreading and reduce the over-all number of cases which are increasing day-by-day.

2. OVERVIEW OF PARSEHUB

Web scraping refers to the mining of data from a website. Web scraping can be done using many software tools. After the data is pulled and then it is exported in a more suitable format such as a spreadsheet, Excel or JSON.

In most cases, you would download a pre-built web scraping tool that addresses the requirements of web scraping. Some of the web scrapers generally come as either browser extensions or more compelling applications that are figuratively installed on your pc. Web scrapper can also run the scrapes right on your device or run within the cloud, freeing up your device resources.

ParseHub is a graphic data extraction tool that can be used by anyone to fetch data from the online. You will not have to

write an internet scraper again and the user can effortlessly create APIs from websites that do not have them. ParseHub offers users with both a free plan and custom enterprise plans for enormous data extraction.

Some companies like Endource, Mobile- movielala trailer and so on uses parsehub. Features of parsehub include:

i) It works with multi-page and single-page apps
ii) It uses machine learning for its relationship engine.
iii) It instantly shows sample data as you are working.

Fig 1: Simple working of Parsehub [A]

We just have to download parsehub tool and give the URL of the website we want to scrape the data from and select the fields it within few minutes your result gets ready to use[A] you can then use the data which will be in JSON, CSV or excel[B] format and use that in your project or research purpose.
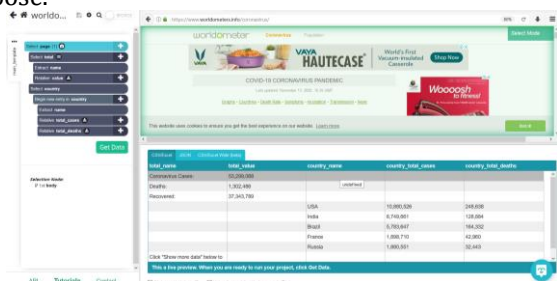

Fig 2: Parsehub Interface [B]

Pros of parsehub:

i] Easy to set-up
ii] Great support
iii] Complex websites
iv] Native Desktop App

3. PROBLEM STATEMENT
coronavirus has potential of high contagious, substantial fatal deaths in some high-risk groups, and potential to cause huge societal and economic disruption in the nation. As more and more cases are increasing, we need a solution to get the accurate data of total cases, active cases and so on from across the world.
We need a system that tells us what are the preventive controls, and give us accurate data of the cases.

4. PROPOSED SYSTEM
Using covid-19 datafetcher using parsehub tool with voice assistance we are collecting data from different websites and giving the users the accurate data of the total cases, active cases which is getting extracted from the tool called parsehub and then it is being imported in JSON format from which the program written is taking the data.

As we are using parsehub tool which is Cloud based the extracted data will be saved on site's servers so each run will be saved and speed of this tool is far faster than traditional web scraping methods, where inspecting the page for each command can grow tedious and difficult. Below is the Architecture which shows the working of proposed system[C].
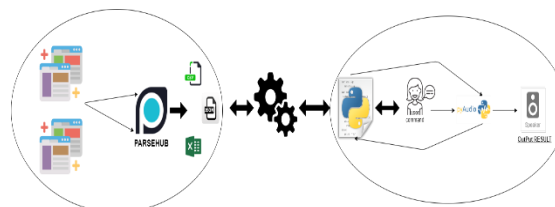

Fig 3: Working of proposed system[C]

From Covid-19 DataFetcher Using ParseHub Tool With Voice Assistance we get the information about novel disease coronavirus with the help of new tool best known for its data gathering (Scrapping), with ParseHub tool we will be able to collect the data about the cases and situations going in several countries through which public will be aware of increasing cases, deaths, active cases, predictions and so on. Using ParseHub we can combine the data from different sites and get the result in the form of csv files, excel and Json format. This data after getting processed will be fetched as results to the system so when the user asks about any related information about the coronavirus cases the data will be fetched and the results will be in audio format [4.b]. Here we will be using voice recognition to tell the results from the user query.
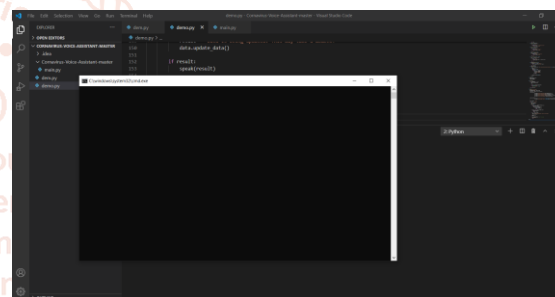
5. RESULTS

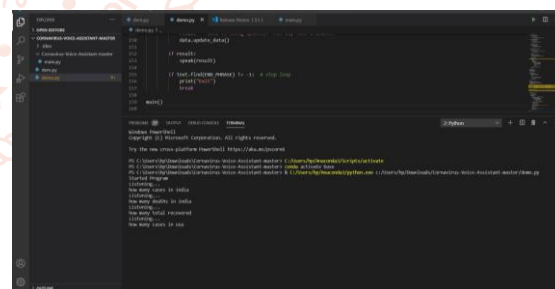**Fig 4[a]: Running the application**

Fig 4[b]: Application listening to User commands.

5. CONCLUSION AND FUTURE ENHANCEMENT
By this application it can be helpful in identifying the number of cases, precautionary measures and spread knowledge among the people about the day-to-day increasing cases around the world.
As we can get live counts of the cases from the website user can see the data and make the decisions whether they want to go to that particular place or not. In future we can enhance this application by adding more number of websites and scrape all the data and store it in a database which will be helpful in identifying the number of cases, precautionary measures and spread knowledge among the people about the day-to-day increasing cases around the world.

REFERENCE

[1] F. Petropoulos and S. Makridakis, "Forecasting the novel coronavirus COVID-19," PLOS ONE, 2020.

[2] R. Murri, F. V. Segala, P. D. Vecchio, A. Cingolani, E. Taddei, G. Micheli and M. Fantoni, "Social media as a tool for scientific updating at the time of COVID pandemic," PLOS ONE, 2020.

[3] C. Zhan, C. . K. Tse, Z. Lai, T. Hao and J. Su, "Prediction of COVID-19 spreading profiles in South Korea, Italy and Iran by data-driven coding," PLOS ONE, 2020.

[4] "Stackshare," [Online]. Available: https://stackshare.io/parsehub.

[5] "Parsehub," [Online]. Available: https://www.parsehub.com/.

[6] M. S. Parvez, S. S. Rajendra, A. S. K. Tasneem and K. R. Bodke, "Analysis Of Different Web Data Extraction Techniques," International Conference on Smart City and Emerging Technology (ICSCET), 2018.

[7] V. Singrodia, A. Mitra and S. Paul, "A Review on Web Scrapping and its Applications," International Conference on Computer Communication and Informatics (ICCCI -2019), 2019.

[8] R. DIOUF, E. N. SARR, O. SALL, B. BIRREGAH, M. BOUSSO and S. N. MBAYE, "Web Scraping: State-of-the-Art and Areas of Application," IEEE International Conference on Big Data (Big Data), pp. 6040-6042, 2019.