



Web Scraping- Covid-19 Pandemic Data

*Siva Ranjini C^{*1}, Siva Roopini C^{*2}*

^{*1}Student, Department of Data Science, Sri Krishna Arts And Science College, Coimbatore, Tamil Nadu, India

^{*2}Student, Department of Data Science, Sri Krishna Arts And Science College, Coimbatore, Tamil Nadu, India

ABSTRACT

The overview of this article comparatively of the net scraping at the covid-19 virus with aid of excel. Web Scraping (additionally termed as Screen Scraping, Web Data Extraction, Web Harvesting, etc.) is an idea applied to collect records from websites wherein the records are extricated and documented which is spared on your PC or to a database withinside the plot of a table (the spreadsheet). As human beings commenced exchange, share, and garage alternatives for the records at the net, difficulties emerged that the way to address the records overload and the way the purchaser gets or gets to the pleasant statistics in the least endeavors. The expert spots out a brand-new process known as Web Scraping to remedy these issues. Web scraping, or net scratching, is a process that's applied to create prepared records primarily based totally on reachable unstructured records at the net. The crawler leads the scrapper and extracts the asked statistics. Let us recognize approximately those additives of net scraping. The statistics on the websites are unstructured. Here, a Microsoft Excel spreadsheet is exercised for web scrapping on the covid-19 virus unfolding in India.

Keywords: web scraping, data extraction, web, analyze.

1. Introduction

Visualize that you must pull a massive volume of data from a website, and you want to make it as hastily as possible. How would you make it without manually going to each website and grasping the data? all right," Web Scraping is the contemporary kick for this stem ". Getting data from a particular website easier and more hastily is through web scraping. Gathering and parsing data from any Website of user choice is termed web scraping. The topmost source of information and misinformation on the earth is the Internet moderator. Multitudinous professions, analogous as data wisdom, business intelligence, and investigative reporting, can benefit from collecting and assaying data from websites.

- Usage of string styles and regular expressions on the website.
- Parse website data using an HTML parser.
- Interact with forms and another website element.

For illustration, suppose you are eager beaver with a design called "headphones comparing website," where you want to know model names, every headphone conditions, and the price of headphones to make comparisons between the different headphones those are belonging to different companies. Web scraping consists of two corridors a web crawler and a crawler and extracts the requested data. Let's now understand these two factors of web scraping. The data on the websites are formless. Web scraping helps collect these unshaped data and store it in a shaped form. The different ways to scrape websites through online Services, APIs, or writing your law. It makes the website business shaft and may beget the breakdown of the website garçon. Therefore, all people aren't allowed to scrape data from the web. How do you perceive which websites are allowed or not? You can look at the

Siva roopini C. Tel.:8838199058

E-mail address: sivaroopini2003@gmail.com

machine.txt' train of that website. The website URL that you want to scrape is followed by clearly "putmachine.txt". It is great to use the matching model in it. For the utmost part, occasionally it's easier to use an HTML parser that is explicitly designed for parsing out HTML runners. The effortless tool that is used for scraping data from any website is Microsoft Excel, it is smooth as well as helpful. It's a good one to start with web scraping. With ways like this, we can scrape data from any website that cyclically modernizes its data. Still, we should be apprehensive that demanding multitudinous a runner multiple times in a rapid-fire race can be seen as indeed Resentful or suspicious, website usage.

2. Methodology

The methodology used for the article is to gather all the data extracted from various sources by using Microsoft excel and further analyze it as per one requirement. The web crawler which is excel based also may help us extract the desired result as we analyze the process by specific steps and give the desired URL for performing iteration for scrapping the data from the source URL. There are two ways to extract data from any website, the Manual extraction technique, and the automated extraction technique. To extract data from the website the following steps should be followed,

Step 1:Open Microsoft Excel and select the Data menu in the top left corner of the spreadsheet.

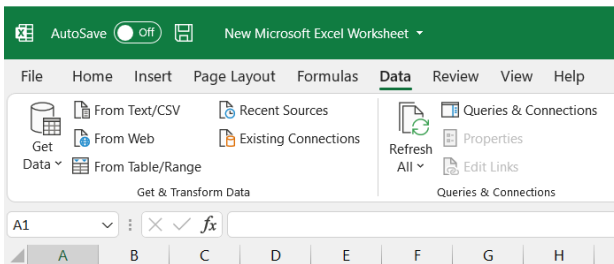


Fig 1:Data tab

Step 2:In the ribbon under the Get & Transform Data section, select from the website.

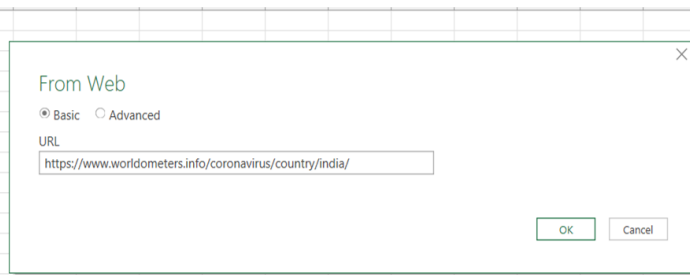


Fig 2:From the web tab

Step 3:This will open the Web Query window. Enter or paste the URL to the page with the data into the URL field. Select OK.

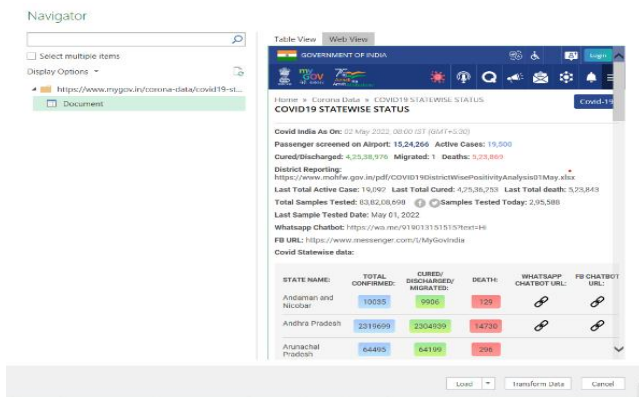


Fig 3: Website

Step 4:Microsoft Excel will reach out and pull the webpage content. Select each table result on the left of the Navigator window to view a preview of the table in the right panel.

Step 5: When you view the table preview display the table from the website with the data you want to scrape, select the Load button which is situated at the bottom of the window.

Step 6: You'll see the data from the website table appear in the first shell of the active Excel spreadsheet. Web scraper also scrapes data using python language.

2.1 Tools and platforms used to scrape data from web

- OS
- Windows
- Linux
- Python IDE: python 2.7.x
- Beautiful soup
- Browser
- Django
- MySQL

2.2 Hardware used

- RAM:4GB
- Processor: Intel i3
- Hard Disk:500GB

3. Result and Discussion

The overall result turned out to be helpful to understand the use and benefits of scraping data from the web. The web scrap extracted the data from a webpage and save it in an excel sheet for further operations such as creating pie charts, graphs, etc. Moreover, the analysis was done to show infected cases and deaths related to covid -19. The below table was the result of web scraping.

STATE	INFECTED	INFECTED TODAY	DEATHS	DEATHS TODAY	RECOVERED TODAY	ACTIVE	TESTS
Maharashtra	7876841	144	147834	2	0	7729007	0
Kerala	6539432	281	68843	24	0	6470589	0
Karnataka	3946934	60	40057	0	0	3906877	0
Tamil Nadu	3453552	52	38025	0	0	3415527	0
Andhra Pradesh	2319662	3	14730	0	0	2304932	0
Uttar Pradesh	2072892	212	23505	2	0	2049387	0
West Bengal	2018012	41	21201	0	0	1996811	0
Delhi	1874876	1083	26167	1	0	1848709	0
Odisha	1287989	9	9124	0	0	1278865	0
Rajasthan	1283332	15	9552	0	0	1273780	0
Gujarat	1224245	12	10943	0	0	1213302	0
Chhattisgarh	1152235	2	14034	0	0	1138201	0
Madhya Pradesh	1041319	6	10735	0	0	1030584	0
Haryana	989292	417	10618	0	0	978674	0
Bihar	830524	2	12256	0	0	818268	0
Telangana	791793	21	4111	0	0	787682	0
Punjab	759422	21	17744	0	0	741678	0

Assam	724207	2	6639	0	0	717568	0
Jammu and Kashmir	453990	12	4751	0	0	449239	0
Uttarakhand	437388	12	7693	0	0	429695	0

Table 1. Table data scraped from the website

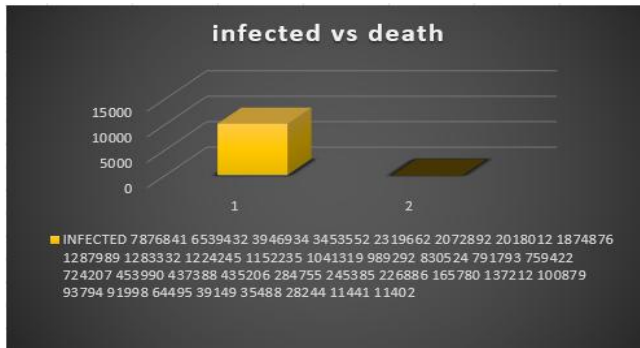


Figure 4: Bar Graph

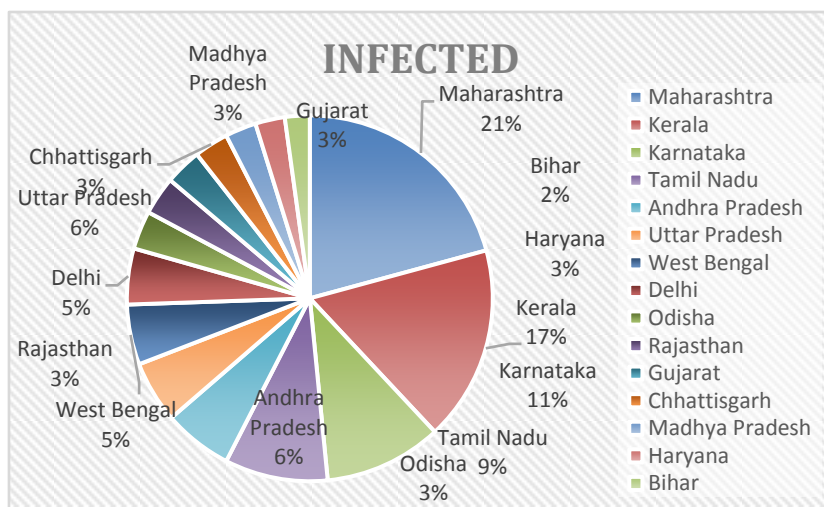


Figure 5: Pie Chart

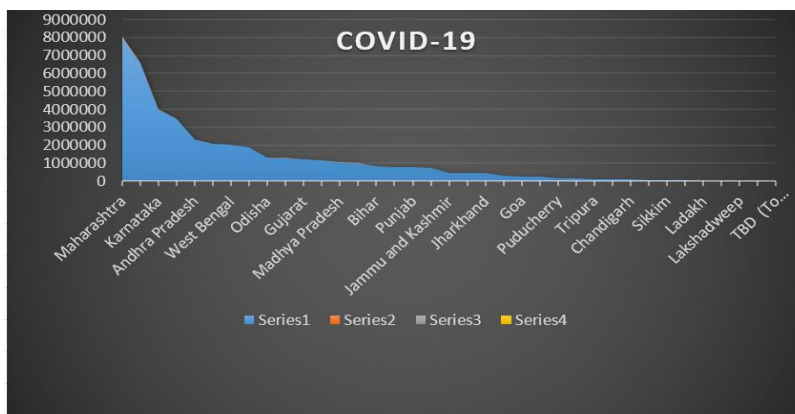


Figure 7: Fall in Covid-19 wave.

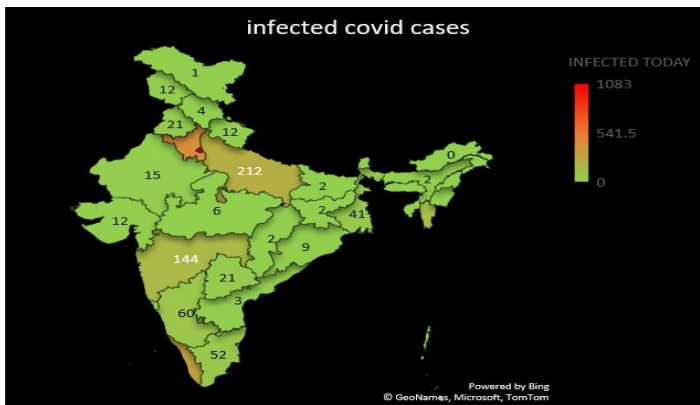


Figure 6: Data represented in the India map

4. CONCLUSION

The hidden web data are heterogeneous so, the extraction of data hidden in web data is a major challenge. It becomes an ineffective way to search for any kind of data nowadays. Web scraping is this user-friendly where users can efficiently deal with indexing, query processing, and effective data extraction technique based on web structure from the submission plan is one of the flop. The Web data which is hidden needs synthetic and semantic matching to fully achieve automatic integration in this automatic and domain-conditional prototype system is proffered that extracts and incorporated the data which is situated at the rear of the search form. Indeed, even with every one of the web information it still presents use openings where users realize how to put it at the proper usage.

REFERENCES

- [1] Fatma sari, Yesi Novaria Kunang, and Susan Dian Purnamasari," Web Scraping Techniques to Collect Weather Data in South Sumatera",pp1-4,2018
- [2] Vlad Krotov," Legality and Ethics of Web Scraping", pp.1-5,2018
- [3] Buchanan, E. 2017. "Internet Research Ethics: Twenty Years Later," in Internet Research Ethics for the Social Age: New Challenges, Cases, and Contexts, M. Zimmer and K. Kinder-Kurlanda (eds.), Bern, Switzerland: Peter Lang International Academic Publishers, pp. xxix-xxxiii.
- [4] Jaime Lopez," Web Scraping", pp.1-5, Jan 2018.
- [5] Tahseen Zaid," Web scraping from the web site", pp.1-4,2018
- [6] Salim Satriaajait, Satria Panuntum, Setia Pramana," Implementation of web scraping in criminal news collection during covid-19 pandemic", pp.300-308,2020.