INTERNATIONAL JOURNAL OF ADVANCE RESEARCH, IDEAS AND INNOVATIONS IN TECHNOLOGY

ISSN: 2454-132X
Impact factor: 4.295
(Volume 5, Issue 3)
Available online at: www.ijariit.com

Weather report on metropolitan cities in India using web scraping technique

Shubham Singh
shubhamsingh4874@outlook.com
Shri Ramdeobaba College of Engineering and Management, Nagpur, Maharashtra

Rainy Jain
jainrainy7044@gmail.com
Shri Ramdeobaba College of Engineering and Management, Nagpur, Maharashtra

ABSTRACT

In metropolitan cities of India, there are several climate measurement points that are owned by several different institutions such as Indian Institute of Tropical Meteorology, NCMRWF, India Meteorological Department (Pune) and many other institutes. However, obtaining the latest meteorological data in detail within a specific period has restricted the bureaucratic process to each institution. While the availability of meteorological data sets is necessary for conducting research in the field of data analysis to predict climate and analyzes for DSS that require weather patterns. On the other hand, some sites provide real-time weather data for some cities. For this, in this research, we use web scraping technology to collect meteorological data in metropolitan cities in India. Web scraping technology is a technique for retrieving the content of a web page individually. The data collected by this web scraping technique will form a database or data warehouse that can be used for further research on data mining for weather forecasting in metropolitan cities, which in the future can be developed a support application to based decision.

Keywords— *Metropolitan cities, Data analysis, Web scraping, Weather forecasting*

1. INTRODUCTION

Weather forecasts have a significant impact on people, especially in the economic and social sphere. By collecting weather data, we can analyse the data patterns of annual meteorological data that affect temperature patterns and precipitation. Some studies using climate patterns are used to study weather patterns in agriculture 1), in health 2), transport 3), urban planning. Climate assessment alone is not easy, because the climate is constantly changing and it is also influenced by the atmosphere or atmosphere dynamics. Approaches to meteorological forecasting depend to a large extent on the observed data and the methods and methods of meteorological forecasting used. Data is necessary to measure data not only at one point to enhance the weather forecast analysis. We need to measure the data at some points to see the atmosphere movement, the flow of clouds, wind, etc., to make the weather forecast important.

India itself has several different agencies, such as Indian Institute of Tropical Meteorology, NCMRWF, India Meteorological Department (Pune) and others. However, getting the latest weather in a bureaucratic process to each agency. On the other hand, some sites, such as www.WorldAtlas.com and www.timeanddate.com provides updated weather data online.

As a result, this study uses Web Scraping technology to collect data from sites that provide information on the web. The web-based method is a mechanism for allowing data to be extracted from the various websites that facilitate the compilation of data collected and viewed. The information gathered through this array of breakout methods will be based on a database or meta-logical dataset. This research is a preliminary study of the use of meteorological data to be used for future investigations and of survey data for India weather forecasts and for the study of meteorological shapes for decision-making.

The paper is divided into five sections. Section 2 describes several of the related studies in this research. Section 3 describes the methodology, Section 4 presents result, discussion, Section 5 is about Legal Issue Related to Web Scraping and Section 6 draws conclusions of this study.

2. RELATED WORK

Web queries are data collection techniques associated with APIs (application programming interfaces). Web-based technologies typically overwrite data generation agents (usual pages in HTML and other web pages) by creating a proxy for the web server

request and then extracting the data. It uses many programs and technical tools that can be used in web applications to handle data analysis, natural prediction, and information security.

Some research that explains the techniques of web scraping techniques such as Bo Zhao [3] discusses the web scraping techniques briefly. Research [2] mentions the Web Scrapping and Semantic Annotation briefly.

3. METHODOLOGY

In this case, there are a few steps to implement an ontology-based semantic website, web browsing for weather forecasting report. First, users collect data sets from a different institution in India. Figure 1 illustrates the step-by-step using the architectural diagram.

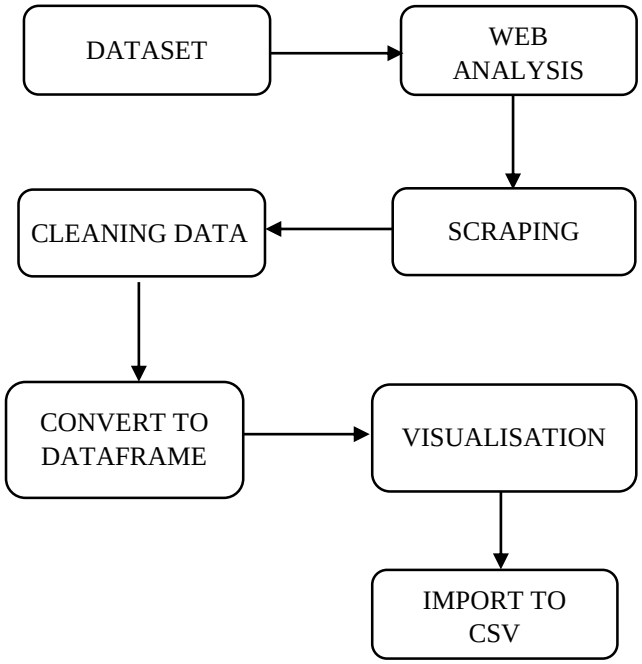

Fig. 1: Methodology

3.1 Datasets

This study uses data from the most reliable weather information from reputed websites such as timeandate.com and WorldAtlas.com. Where we pick some parameter upon such as Precipitation, Time, Feels Like, Wind, Humidity, Temperature from this we made further analysis and names of the metropolitan cities.

3.2 Scraping

Web scraping is the retrieval of a document from the internet, usually in a markup language such as an HTML or XHTML document, such as a web page and specific data taken from that page are analyzed for other interests. Web scraping is only aimed at retrieving data and retrieving data with different data sizes.

3.3 Cleaning

The data extraction contains HTML tags so that the data must be in the cleaning to remove the HTML tags and generate the underlying data if necessary where we use certain python libraries such as re, compile, sub for cleaning purpose.

3.4 Import to CSV

After clearing the data from HTML tags, the information is converted to data frame by using panda's library then we can export the data as a dataset in different forms such as CSV, Excel file and pdf file.

4. RESULT

Web Weather forecasting Scraping Application is developed using HTML Parsing created in Python programming language in Anaconda software platform running on Windows Operating System 10.1 Script is created using Beautiful Soup 4 library and the Requests library. The site in scraping is mentioned in the introduction.

- The first script will contain 10 Metropolitan cities in India, covering the central national capital region, Mumbai metropolitan region, Kolkata metropolitan region, Bangalore metropolitan region, Hyderabad metropolitan region, Chennai metropolitan region, Pune metropolitan region, Kanpur metropolitan region, Visakhapatnam metropolitan region and Nagpur metropolitan region in the web site of the WorldAtlas.com
- The Second script contains the parameter on the function of the first script parameters will be decided by these parameters Precipitation, Time, Feels Like, Wind, Humidity, and Temperature.
- To run Script data automatically and done periodically every five hours, so in this research, the researcher does hosting Python script in Python server anywhere. We also created task scheduler that will run the script continuously every five hours. This task scheduler will run both scripts automatically in every five hours.

5. LEGAL ISSUE RELATED TO WEB SCRAPING

The legitimacy and fair use of the use of web resources is often a problem. There are two aspects of debugging techniques, including copyright and unauthorized access.

In the first case, you will submit without permission. The web-crawling process to make the website accessible free of charge. Not for host data parameter. It is not the IP problem that is used, and all of the accomplished data is the limitless proxy server and is not the destination of the process. Therefore, the web page processing process does not require any unprofessional behaviour. Stored data is not being used in the collection of information, but its purpose. The research process does not repeat or miss, but to analyse...

First on copyright issues as in the case of Craigslist Inc. v. 3Taps Inc. 2013)), Federal Court's decision to disrupt the copyright infringement and general information. This web page provides clear disclosure information for web scraping techniques. It also uses this research on a regular basis, it does not download the site (only one hour per hour) and does not hurt the websites of the information you receive

6. CONCLUSION AND FUTURE WORKS

In this study, Web Scraping techniques were successfully used to collect weather data for many Metropolitan cities in India. Web Scraping is to collect data from many websites that provide weather data. The Web Fragmentation Model runs in the Python programming language and automatically collects data continuously every five hours. The data obtained is very detailed and can be used for further data analysis which was beyond the scope of this research paper.

The data collection process is continuing to generate weather data sets for the metropolitan cities in India. After collecting the datasets, a number of studies will be conducted to predict weather and weather analysis for agricultural, transport decisions and airways.

7. REFERENCES

[1] Amalia Amalia, Rizky Maulidya Afifa and Herriyance Herriyance “Resource Description Framework Generation for Tropical Disease Using Web Scraping” Available:- 2018 IEEE International Conference on Communication, Networks and Satellite
[2] Sanjay Kumar Malik, SAM Rizvi “Information Extraction using Web Usage Mining, Web Scrapping and Semantic Annotation” Available:- 2011 International Conference on Computational Intelligence and Communication Systems
[3] Bo Zhao “ Web Scraping” Available:- Springer International Publishing AG (outside the USA) 2017
[4] Fatmasari, Yesi Novaria Kunang and Susan Dian Purnamasari “Web Scraping Techniques to Collect Weather Data in South Sumatera” Available:- International conference on electrical engineering and computer science 2018, IEEE

