I.J. Intelligent Systems and Applications, **2014, 01, 1-3**
Published Online January 2014 in MECS (http://www.mecs-press.org/)
DOI: 10.5815/ijisa.2014.01.01

Weather Monitoring Information System With Data Acquisition Based On Web Scrapper Technique

Robi. Kurniawan
STMIK Handayani, Makassar, 90231, Indonesia
Email: robikurniawan@student.handayani.ac.id

Zahir. Zainuddin and Andani. Achmad
Department of Technical Information, Departement of Electrical Engineering
Hasanuddin University
Makassar, 90245, Indonesia
Email: {second.author, third.author}@hostname2.org

Abstract— This study aimed to monitor weather and display weather effective and accurate data information by creating a weather monitoring information system. This research applied a scrapper web-based data acquisition technique that is the process of taking data from the web pages target www.wunderground.com and www.weather.com through the internet and storing the data into a database to be used as a web service. Data acquired from the two websites are weather data; temperature, dew point, humidity, wind, wind speed, wind gust, pressure, precip, precip accum, and conditions which were then displayed in readable tabular and graphical form. The results of data acquisition on the website www.wunderground.com produced 3868 records starting on April 1 - October 5, 2019 while on the website www.weather.com as many as 40 records on October 4 - October 5, 2019. Based on the test results in the form of a questionnairetested on several respondents showing that web services provided accurate and real-time weather data information with a percentage value of 84%, 92% percentage values for interface design and informative weather presentations.

Index Terms— Weather Monitoring, Data Acquisition, Web Scrapper, Web Service.

I. INTRODUCTION

Weather & climate are condition or situation formed from interactionsaf several components or elements called weather & climate elements which interacts each other. The elements are radiation, temperature, humidity, wind pressure wind, clood, precipitation and evaporation [1]. Monitoring weather is able to show description about recent condition of climate.

It played a vital role in humans life, is wellaas date collection & information about temporal weather changing

Weather data has been spread in several websites & presents visual information about weather condition, temperature, humidity, wind velocity, rainfall & its prediction. The data is so vital for certain people : famers, fishermans, pilots & stakeholders.

Nowadays, weather data is hard to gain for stakeholders who needs it. In case that they require the data as reference in taking decision or conducting fusther analysis, they can easily find it in certain websites. On the other hand, if people need the data, they have to visit the instantion & they have to pay for it. Also the problem for stakeholders, most websites are not available to download & if it is available, the format is not suitable.

The problem above is surely vital and it needs a new website presenting weather data information that be easily used and cultivated by skateholders with appropiate format without visiting the agency to obtain the data and easily to download.

Based on explanation above, it can developed into a weather monitoring information system with data acquisition based on web scraper technique. Data acquisition is data used to collect, gain, and prepare on progress data then it will be cultivated in computer for certain purpose [2]. It is a process to change data from sensor into electricity signal and then be converted into digital form and analyzed by the computer. Data acquisition system consists of sensor, signal processing unit, data acquisition hardware and computer unit [3]. It needs sensor devices to convert phisical variable into electricity voltage variable [4].

Web Scrapping or web extraction is a technique of collecting information from internet which mostly gain from web pages and then the document analyzed and used fo certain pirposes [5]. In its practice, the use of Web Scrapping consists of four steps. There are: Creating Scrapping Tempaltes, This process needs observation about HTML website documents whose information will be gained or scrapped by conducting HTML tag to collect information; Exploring Navigation Sites, In creating Web Scrapper application, a programmer needs to understand navigation technique

on the website whose data or information will be copied; Automate Navigator and Extraction, In this section, application is build to optimalize collecting data/information from two previous sections; Data Extraction & History Saving, Information gained from step 3 then the data is stored in database table [6].

Web scraper operates by driving into website page, then conducting data extraction from the website and saves it into one file database. The using of web scraper technique is expected to produce a website that is able to present weather data information efficiently and accurately. It has goal that the data can be used well for certain people.

II. Related Work

Considering how difficult to gain the weather data. Several reserachers try to solve this problem. Some works have been done and these have inspired us to our research. A summary of the works highlighted is shown in the table below

Table 1. A Summary of Existing Works

Work	Target	Metdhodology
Wireless sensor network (WSN) application as a web-based weather monitoring system [7]	To formed an url address combining weather data based on comparison from microcontroller and weather website	The research compared data from microcontroller monitor serial with data from website because they think that the existing website is only showed forecast of weather condition in general.
Design and build a weather monitoring information system [8]	To present the weather information system on the website with real time data matching with sending result of data from the sensor	The research created weather monitoring system by applying several weather sensor wtih wireless transmission system on 2,4 GHz frequence.
Web Syndication Using Web Scraping Techniques for UKM Handicraft Marketing Collaboration [9]	To create a program that can be used for collaboration between the UMKM web, so that UMKM can make the right product marketing decisions.	The Method is syndicate the web with web scraping techniques that taking the information needed from one web to be used on another web. The technique used is web scraping using simple HTML DOM Parser.
Data Analysis and Visualization of Continental Cancer Situation by Twitter Scraping [10]	To link patients with a particular illness (cancer) together and to provide researchers with enriched patient data that might be very useful for future analysis of this disease.	The research scraped tweets from over the last two years from all around the world. They used sentiment analysis and natural language processing to classify them into positive, negative and neutral tweets to determine which of the tweet means to have cancer and which don't. Then they analyzed the prepared dataset and visualized and compared them with veritable cancer-related information to ascertain if people's tweets are allied with actual cancer situation.
An Intelligent Survey of Personalized Information Retrieval using Web Scraper [11]	To do an intelligent background survey of Personalized Information Retrieval, a specialized and crucial subsection of Information Retrieval or IR	The method of IR as Web Scraping, a technique that is extremely popular and is proven to have multi-domain usage.

From Table 1, we can see that a lot of work has been done using web scrapper techniques. But to the best of our knowledge, there are very few works using data acquisition based on web scrapper that can present weather information systems efficiently and accurately so that it can be reused by stakeholders. While there is a lot of existing work using web scrapper technique to prediction pandemic diseases, product marketing, etc.

III. System Overview

A. Source of Data

In collecting valid data, source of data used in this research is from Makassar weather data, www.wunderground.com and weather.com collected periodically every day from web scrapping result. The website is from page showing weather parameter data that has been verified and valid. Page of website provides data history and real-time data. Parameter data which is collecter are: Temperature, Rainfall, Humidity, Wind speed, and Extreme weather

The whole data mentioned is presented in chart below:

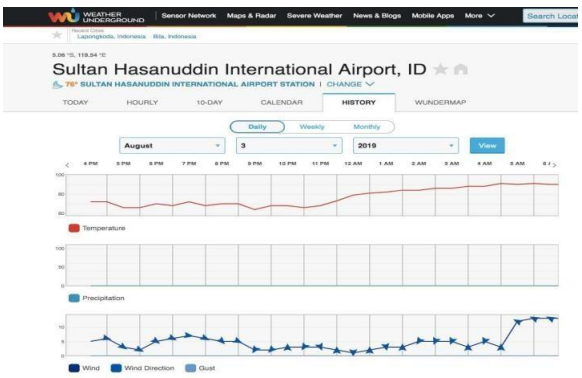
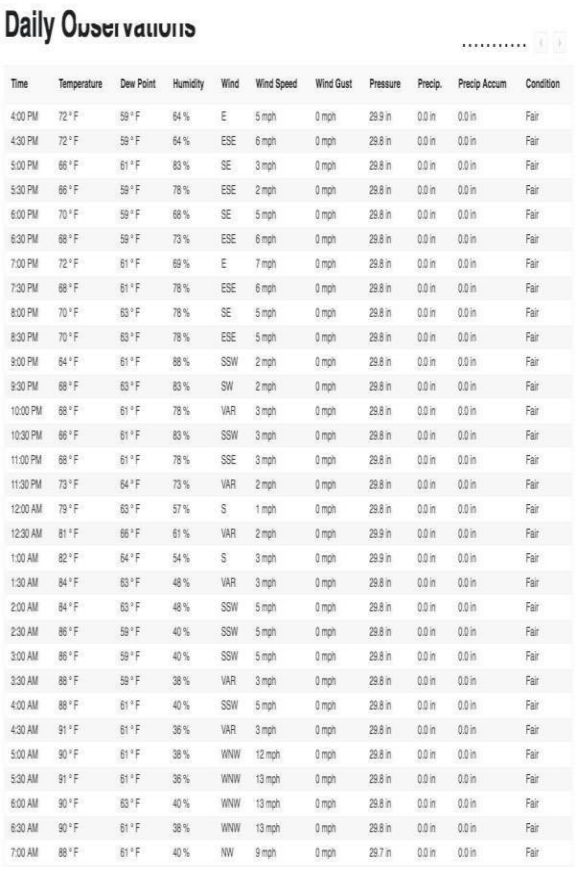


Fig.1. Wunderground.com Web Displayed



Daily Observations

Time	Temperature	Dew Point	Humidity	Wind	Wind Speed	Wind Gust	Pressure	Precip.	Precip Accum	Condition
4:00 PM	72 °F	59 °F	64 %	E	5 mph	0 mph	29.9 in	0.0 in	0.0 in	Fair
4:30 PM	72 °F	59 °F	64 %	ESE	6 mph	0 mph	29.8 in	0.0 in	0.0 in	Fair
5:00 PM	66 °F	61 °F	83 %	SE	3 mph	0 mph	29.8 in	0.0 in	0.0 in	Fair
5:30 PM	66 °F	59 °F	78 %	ESE	2 mph	0 mph	29.8 in	0.0 in	0.0 in	Fair
6:00 PM	70 °F	59 °F	68 %	SE	5 mph	0 mph	29.8 in	0.0 in	0.0 in	Fair
6:30 PM	68 °F	59 °F	73 %	ESE	6 mph	0 mph	29.8 in	0.0 in	0.0 in	Fair
7:00 PM	72 °F	61 °F	69 %	E	7 mph	0 mph	29.8 in	0.0 in	0.0 in	Fair
7:30 PM	68 °F	61 °F	78 %	ESE	6 mph	0 mph	29.8 in	0.0 in	0.0 in	Fair
8:00 PM	70 °F	63 °F	78 %	SE	5 mph	0 mph	29.8 in	0.0 in	0.0 in	Fair
8:30 PM	70 °F	63 °F	78 %	ESE	5 mph	0 mph	29.8 in	0.0 in	0.0 in	Fair
9:00 PM	64 °F	61 °F	88 %	SSW	2 mph	0 mph	29.8 in	0.0 in	0.0 in	Fair
9:30 PM	68 °F	63 °F	83 %	SW	2 mph	0 mph	29.8 in	0.0 in	0.0 in	Fair
10:00 PM	68 °F	61 °F	78 %	VAR	3 mph	0 mph	29.8 in	0.0 in	0.0 in	Fair
10:30 PM	66 °F	61 °F	83 %	SSW	3 mph	0 mph	29.8 in	0.0 in	0.0 in	Fair
11:00 PM	68 °F	61 °F	78 %	SSE	3 mph	0 mph	29.8 in	0.0 in	0.0 in	Fair
11:30 PM	73 °F	64 °F	73 %	VAR	2 mph	0 mph	29.8 in	0.0 in	0.0 in	Fair
12:00 AM	79 °F	63 °F	57 %	S	1 mph	0 mph	29.8 in	0.0 in	0.0 in	Fair
12:30 AM	81 °F	66 °F	61 %	VAR	2 mph	0 mph	29.9 in	0.0 in	0.0 in	Fair
1:00 AM	82 °F	64 °F	54 %	S	3 mph	0 mph	29.9 in	0.0 in	0.0 in	Fair
1:30 AM	84 °F	63 °F	48 %	VAR	3 mph	0 mph	29.8 in	0.0 in	0.0 in	Fair
2:00 AM	84 °F	63 °F	48 %	SSW	5 mph	0 mph	29.8 in	0.0 in	0.0 in	Fair
2:30 AM	86 °F	59 °F	40 %	SSW	5 mph	0 mph	29.8 in	0.0 in	0.0 in	Fair
3:00 AM	86 °F	59 °F	40 %	SSW	5 mph	0 mph	29.8 in	0.0 in	0.0 in	Fair
3:30 AM	88 °F	59 °F	38 %	VAR	3 mph	0 mph	29.8 in	0.0 in	0.0 in	Fair
4:00 AM	88 °F	61 °F	40 %	SSW	5 mph	0 mph	29.8 in	0.0 in	0.0 in	Fair
4:30 AM	91 °F	61 °F	36 %	VAR	3 mph	0 mph	29.8 in	0.0 in	0.0 in	Fair
5:00 AM	90 °F	61 °F	38 %	WNW	12 mph	0 mph	29.8 in	0.0 in	0.0 in	Fair
5:30 AM	91 °F	61 °F	36 %	WNW	13 mph	0 mph	29.8 in	0.0 in	0.0 in	Fair
6:00 AM	90 °F	63 °F	40 %	WNW	13 mph	0 mph	29.8 in	0.0 in	0.0 in	Fair
6:30 AM	90 °F	61 °F	38 %	WNW	13 mph	0 mph	29.8 in	0.0 in	0.0 in	Fair
7:00 AM	88 °F	61 °F	40 %	NW	9 mph	0 mph	29.7 in	0.0 in	0.0 in	Fair

Fig.2. Daily Observation Wunderground.com

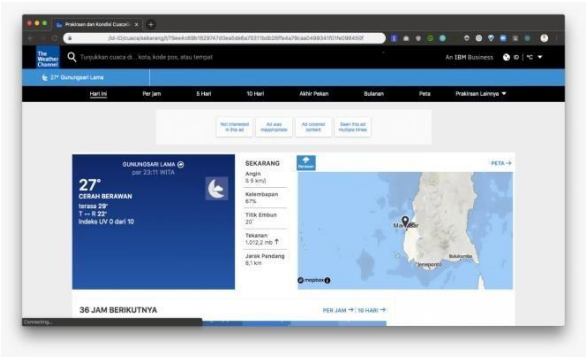


Fig.3. Weather.com Web Displayed

B. *Stages of Research*

This research consists of several stages, they are:

Literature Review

The aim of literature review is to assist in identifying the problems. Literature review is process to find out problem in this research.

Data Collecting Procedure

This step is conducted to gain as much as possible information needed in order to achieve the goal of this research.

C. *Drafts of System*

The drafts of system are presented based web scrapper using data acquisition below:

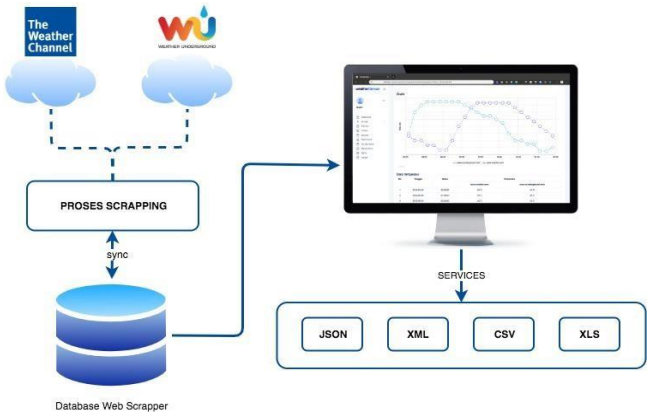


Fig. 4. Steps in Drafting System

Table 2. Description of Web Scrapping Process

Process	Explanation
Website Data Source	Source of website which will be processed by scrapper
Scrapping	Data Scrapped Processing from website
Sync	Storage for scrapper data syncronisation
Monitoring System	Web presenting weather data in cultivable data
Users	Stakeholders who need data for need analysis, developing system, or public need in JSON, XML, CSU, and XLS format.

IV. IMPLEMENTATION

A. *Scrapping Process*

Web Scraapper Technique used in collecting data from wonderfround com and weather, com due to this method can be applied to conduct data acquisition contained in the website or document can is conduoted in order to conduct acquisition value. Historc data them will be saved in Json file in order to make it easier gaining information. These are steps :

Validating the URL

URL from https://www.wunderground.com/hourly/id/makassar dan https://weather.com/idID/cuaca/sekarang/l/d9a5585a9439eb18d5467df8a8a6f38dbde21d0d5ed03ecbe65261d2dbb80957 is initialized in the beginning of programing to validate the website.

Conducting Semantical Structure Render

In this stage, website page that has been decided then respresented in semantical structure then it is changed into object array series. The next step is deciding & verifying tag, selector & it becomes document objects (DOM)

Conducting Website Page Parsing

In this stage, result of DOM gained from previous step is choosen which data will be taken. Data Paesing result in this step then will be extracted. Field whch will be extracted can be viewed in picture below :

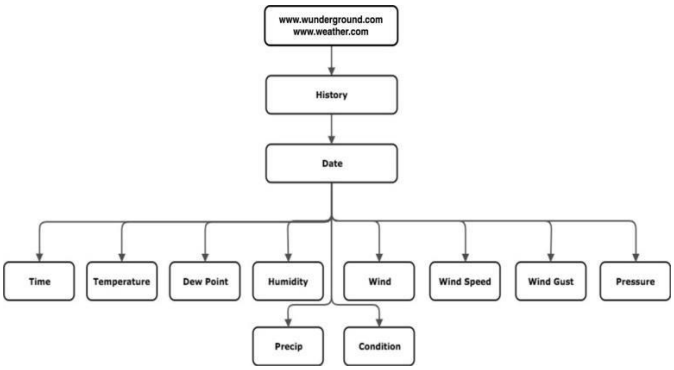


Fig.5. Taken Semantical Data Structure

Field Description

History	:Weather data history information
Date	: Weather data on date
Time	: Weather data on time
Temperature	: Information about temperature in celcius
Dew Point	: Information about place where temperature starting to condence in Celcius
Humidity	: Information about humidity in percentage
Wind	: Information about wind direction
Wind Speed	: Information about wind speed in KM/h
Wind Gust	: Information about wind gust or increasing of wind speed
Pressure	: Information about air pressure in hPa (hetopascal)
Precip	: Information about rainfall in mm
Precip Accum	:Information about rainfall predicted accumulation
Condition	:Information about weather condition

Extracting Value From Class Element HTML

Part of website whose data will be taken is marked with tag class wu-value. In this class, there are several fields: time, remperature, dew point, humidity, wind, wind speed, wind gust, pressure, precip, precip accum, and condition.

Data Checking

In the system, error always exists in the process of data automatitation saving, so checking needs to conduct in order to avoid duplicated data

Data Acquisition

Data in array then will be arranged in JSON formin order to make processing data easier.

Data Cleaning

Process of updating data is process to gaming the best data in several ways. In obtaining proper data & prossessing few of noise, data cleaning applied to detect & erase errors & inconsistency of data chunk to increase quality of data

B. Implementation of Interface Design

In implementing interface design, this application consists of several homepages; dashboard, prediotion & history page

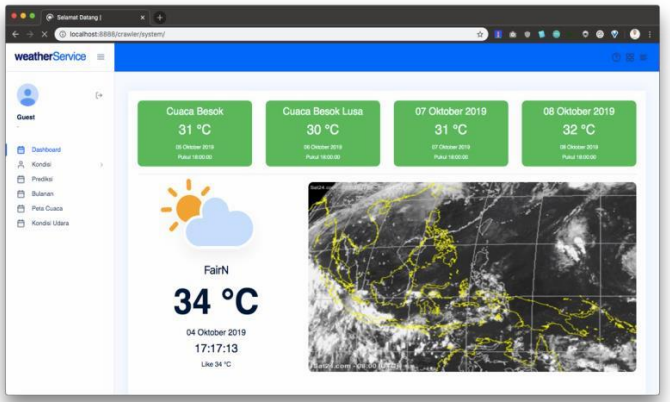


Fig.6. Dashboard Page

Dashboard page showing today weather from wunderground.com & weather.com on real time. This page also shows information about temperature, humidity, rainfall and wind.

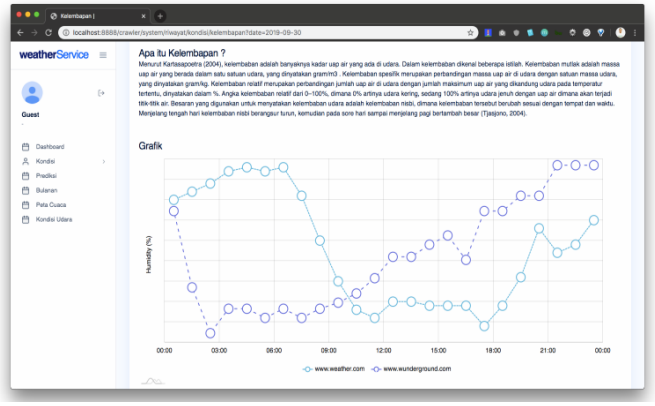


Fig.7. Humidity Graphic

Humidity Graphic showing information about humidity from wunderground.com & weather.com. Data about humidity is presented on humidity data per hours.

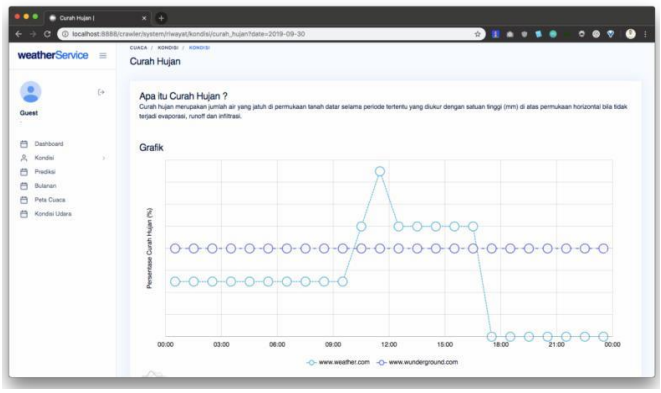


Fig.8. Rainfall Graphic

Rainfall Graphic whose data taken in a day from wunderground and weather.com

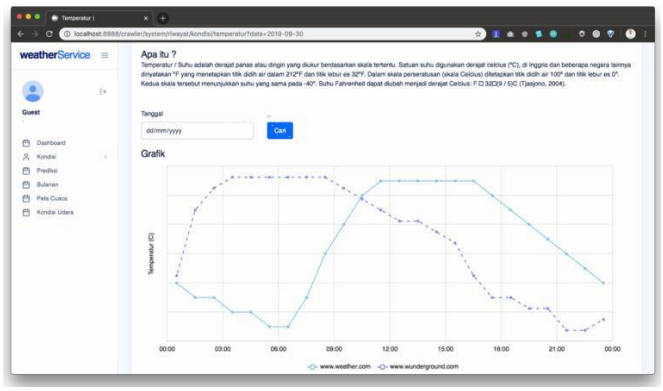


Fig.9. Temperature Graphic

Temperature Graphic showing increasing and decreasing of temperature from wunderground and wheather.com

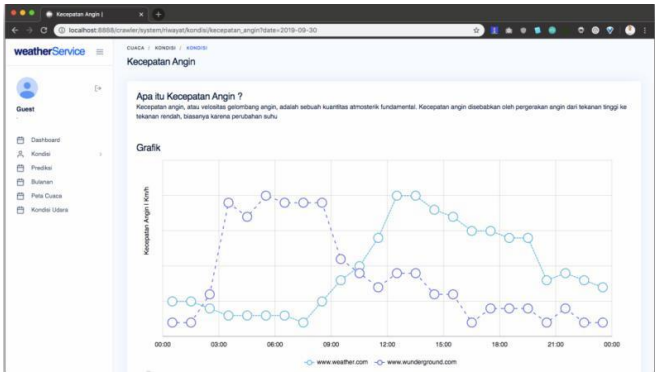


Fig. 10. Wind Speed Graphic

Wind speed graphic taken in a day from wunderground and weather.com

Fig.11. Daily Temperature Graphic

Daily Temperature Graphic in a month where it reahes its peak on Mei 20th 2019.

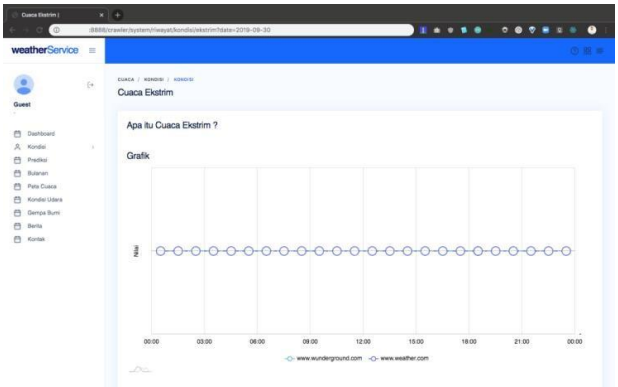


Fig.12. Extreme Weather Graphic

Extreme Weather Graphic showing straight line meaning that there has been no extreme weather in Makassar.

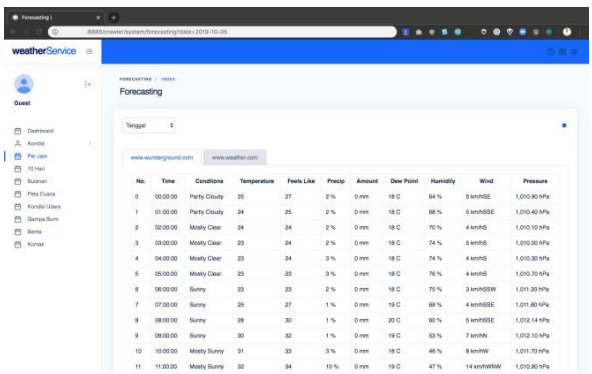


Fig.13. shows per-hours weather prediction (wunderground.com)

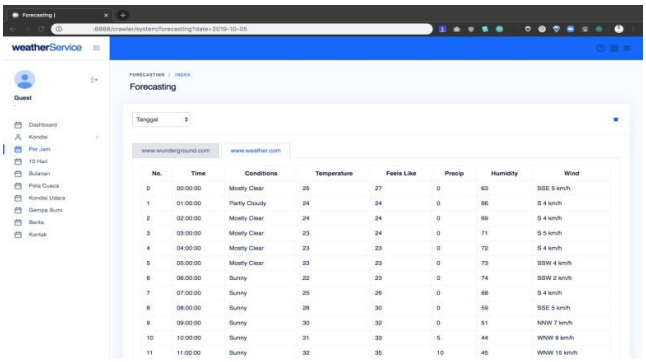


Fig.14. shows per-hours weather prediction (weather.com)

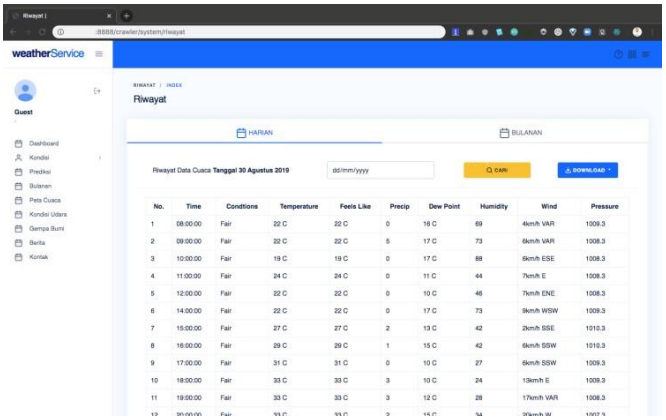


Fig.15. shows Daily History Page

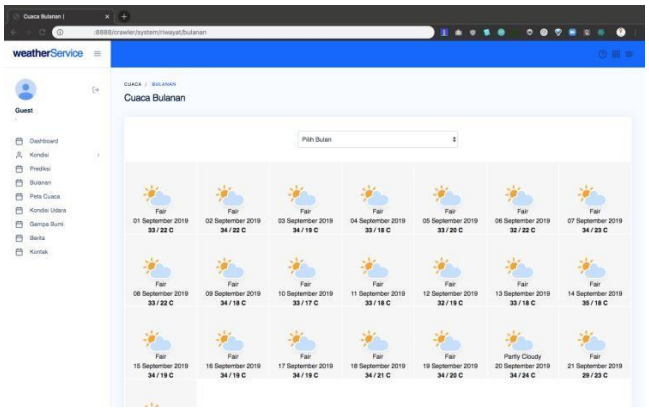


Fig.16. shows Monthly History Page

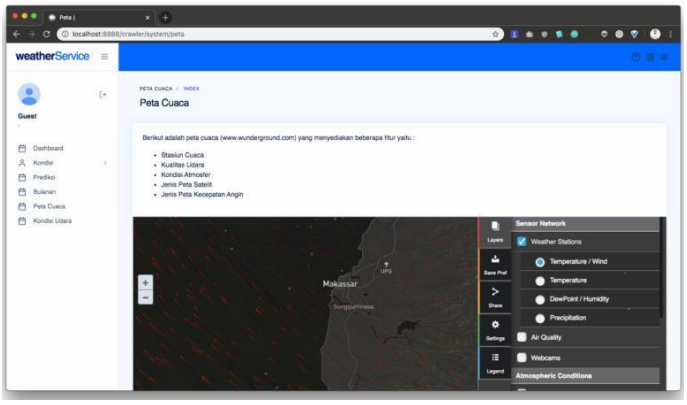


Fig.17. Map Page

Display weather condition in scrapping map from wunderground.com

C. Information Serving

In order to simplify the data, there is a description showing data in several format.

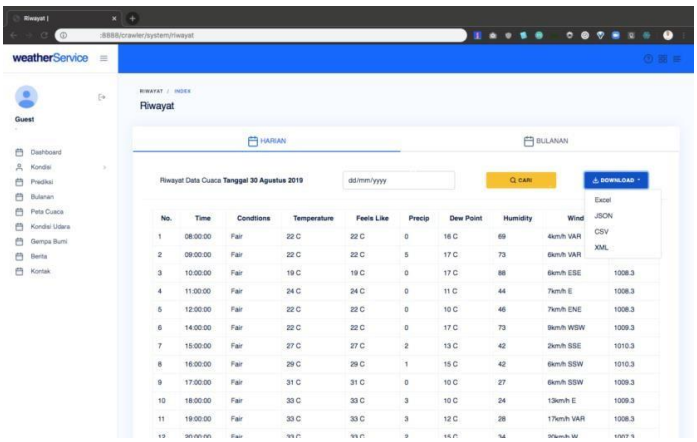


Fig. 18. Download by format

The picture above is a presentation of weather data information by selecting the available formats, namely Excel, JSON, CSV and XML. Following is an explanation of each data format.

JSON Format

It is format to share data its name, JSON is from Javascript programming terminology, it is available in other Termology, python, Ruby and Java. The data is web service to exchange data without considering history of data, created from an application which consuming data.

Excel Format

In Excel format, weather data will appear in Excel format based on the date of the search results.

XML Format

XML (eXtensible Markup Language) is markup language like HTML designed to store & share data. It is used to from structure on desktop programming like Delphi, lazaros, vb, gtk, etc. EML is used to create android application layout, it is to from project structure on Java using Maven, SV6, DOCSX, est .

CSV Format

After deciding CSV format, weather data will appear in CSV format based on searching date.

D. Quistionnare Checking

Questionnare Checking consists of 3 questions spread to 10 questionnares using likeast from 1-5. Based on data from questionnares, the researcher calculated using need for agree or disagree answer from respondents. To calculate the maximum score of each answer by timing score with total of respondents : skor x 10 respondents. Skor maximum can be seen on table 3.

Table 3. Quistionnare Checking

Answer	Score	Score Maximum (Skor * total Responden)
Strongly Agree	5	150
Agree	4	120
Quite Agree	3	90
Not agree	2	60
Disagree	1	30

After finishing previous step, the researcher figured out the percentage of each answer by using formula below:

Where:
Y : Percentage
TS : Total of Respondents = ∑ score of respondents
Ideal Score : Score x total of respondents = 5 x 10 = 50

Score Criteria can be seen on the table 4

Table 4. Score Criteria

Category	Information
0% - 20%	Disagree
21% – 40%	Not agree
41% - 60%	Quite Agree
61% - 80%	Agree
81% - 100%	Strongly Agree

The result of percentage of each answers for 10 respondents:

Table 5. First Question

Question	ANSWER	SCORe	RESPONDEN	TOTAL	PERCENTAGE (%)
Is the weather data displayed on realtime web services?	Strongly Agree	5	3	15	84
	Agree	4	6	24	
	Quite Agree	3	1	3	
	Not agree	2	0	0	
	Disagree	1	0	0	
		JUMLAH	10	42	

Table 6. Second Question

Question	ANSWER	SCORe	RESPONDEN	TOTAL	PERCENTAGE (%)
Is the weather web service display informative?	Strongly Agree	5	6	30	92
	Agree	4	4	16	
	Quite Agree	3	0	0	
	Not agree	2	0	0	
	Disagree	1	0	0	
		JUMLAH	10	46	

Table 7. Thirth Question

Question	ANSWER	SCORe	RESPONDEN	TOTAL	PERCENTAGE (%)
Can weather data be reused by users in various formats?	Strongly Agree	5	8	40	96
	Agree	4	2	8	
	Quite Agree	3	0	0	
	Not agree	2	0	0	
	Disagree	1	0	0	
		JUMLAH	10	48	

IV. CONCLUSION

After conducting the research and creating the application, it could be concluded that implementation of Web scrapping using data acquisition from www.wunderground.com and www.weather.com. Also it is able to present weather data; temperature, dew point, humidity, wind, wind speed, wind gust, pressure, precip, precips accum, and condition in readable tables and graphics. The result of data acquisition on www.wunderground.com showed that 3868 records started from 1 April until 5 October 2019, on the other hand on www.weather.com showed 40 records started from 4 October – 5 October 2019.Based on result of checking questionnaires, it showed that Web Server presented accurate and real-time weather data with number of percentage was 84%, 92 % for Interface Design and 96% for Web Service owing to the capability in presenting information in several formats; JSON, excel, XML, and CSV.

REFERENCES

[1] L. Sabaruddin, *Agroklimatologi*, Cetakan 1. Bandung: Alfabeta, 2012.

[2] Husein, "Weather Monitoring Telemetry System Prototipe Based On XBEE Pro IEE.804.15.4," *J. Apl. Fis. Univ. Haluoleo*, vol. 6, no. 2, pp. 97–103, 2010.

[3] W. Bolton, *Sistem Instrumentasi Dan Sistem Kontrol*, Cetakan 1. Jakarta: Erlangga, 2009.

[4] E. Nasrullah and Y. Raharjo, "ELECTRICIAN Jurnal Rekayasa dan Teknologi Elektro," vol. 3, no. 3, pp. 1–9, 2009.

[5] M. Turland, *PHP|Architect's Guide to Web Scrapping with PHP*. Alexandria: Musketeers.me,LLC, 2010.

[6] A. Josi, L. A. Abdillah, and Suryayusra, "Penerapan teknik web scraping pada mesin pencari artikel ilmiah," 2014.

[7] S. K. Risandriya and Safrizal, "Aplikasi Wirelees Sensor Network (WSN) Sebagai Sistem Monitoring Cuaca Berbasis WEB," no. 1, pp. 1–6, 2014.

[8] A. Wijayanti, H. Mahmudah, N. Adi S, O. P, and H. Wijayanti, Ari Mahmudah, Hanuah Adi S, Nur P, Okkie Alfian, "Rancang Bangun Sistem Informasi Monitoring," *Jsik*, vol. 5, no. 1, pp. 1–7, 2016.

[9] S. Mulyani, F. Andreas, and H. Yulianton, "Sindikasi Web menggunakan Teknik Web Scrapping untuk Kolaborasi Pemasaran UMKM Handicraft," vol. 21, no. 1, pp. 56–62, 2016.

[10] M. H. Al Walid, D. . Anisuzzaman, and A. F. . Saifuddin Saif, "Data Analysis and Visualization of Continental Cancer Situation by Twitter Scraping," *Int. J. Mod. Educ. Comput. Sci.*, vol. 11, no. 7, pp. 23–31, 2019.

[11] B. Ghosh Dastidar, D. Banerjee, and S. Sengupta, "An Intelligent Survey of Personalized Information Retrieval using Web Scraper," *Int. J. Educ. Manag. Eng.*, vol. 6, no. 5, pp. 24–31, 2016.

Authors' Profiles

Firstname A. Lastname and the other authors may include biographies and photographs at the end of regular papers. Photographs, if provided, should be cropped into 26mm in width and 32mm in height. The first paragraph may contain a place and/or date of birth (list place, then date). Next, the author's educational background is listed. The degrees should be listed with type of degree in what field, which institution, city, state or country, and year degree was earned. The author's major field of study should be lower-cased.

The second paragraph uses the pronoun of the person (he or she) and not the author's last name. It lists military and work experience, including summer and fellowship jobs. Job titles are capitalized. The current job must have a location; previous positions may be listed without one. Information concerning previous publications may be included. Try not to list more than three books or published articles. The format for listing publishers of a book within the biography is: title of book (city, state: publisher name, year) similar to a reference. Current and previous research interests end the paragraph.

The third paragraph begins with the author's title and last name (e.g., Dr. Smith, Prof. Jones, Mr. Kajor, Ms. Hunter). List any memberships in professional societies like the IEEE. Finally, list any awards and work for professional committees

and publications. Personal hobbies should not be included in the biography.

Firstname B. Lastname includes the biography here.

Firstname C. Lastname includes the biography here.

