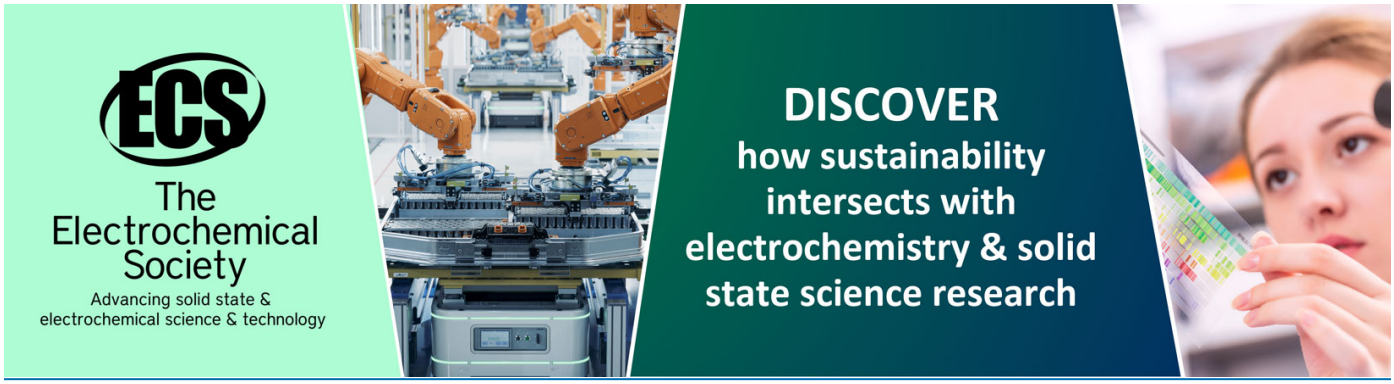
Preliminary Research Design on Sensor Data Gathering for Air Quality Text Generation



Preliminary Research Design on Sensor Data Gathering for Air Quality Text Generation

Indra Aulia [1,2], Ainul Hizriadi [1,3], Seniman [1,3], Muhibuddin [1,3]

[1]Department of Information Technology, Faculty of Computer Science and Information Technology, Universitas Sumatera Utara, Medan 20155, Indonesia
[2]Data Science and Intelligent System Research Group, Faculty of Computer Science and Information Technology, Universitas Sumatera Utara, Medan 20155, Indonesia
[3]Computer System Research Group, Faculty of Computer Science and Information Technology, Universitas Sumatera Utara, Medan 20155, Indonesia

E-mail: indraaulia@usu.ac.id , ainul.hizriadi@usu.ac.id, seniman@usu.ac.id, muhibuddinb@gmail.com

Abstract. Air pollution is one of the environmental problems faced by many countries (i.n. the developing countries). It often harms the health of people, plants, and animals as well as our environment. To decrease the risks, e ective and e cient technology for air quality monitoring is indispensable. In Indonesia, it is still limited in the case of the number of devices and the dissemination of air quality information. This paper describes the preliminary research design of air quality monitoring technology based on the Internet of Things (IoT) and Natural Language Generation (NLG) for solving these limitations. It de nes air quality knowledge in the form of a set of rules and presents the architecture design for data gathering and text generation. Data gathering uses the related sensors which are able to collect the values of parameters as air quality indicators. Then, the collected values are identi ed to be used for generating the textual information in Indonesian language to social media, such as twitter. So, in this research, the proposed solutions are able to solve the problems properly.

1. Introduction

Recently, air pollution is one of the serious problems faced by many countries in the world, especially the developing countries (e.g., Indonesia) [1]. It is often caused by harmful particles and gases, which are suspended in the Earth's atmosphere. These particles and gases may originate from industries, vehicle exhaust, dust, or wild res [1][2][3][4] in which vehicle exhaust and wild res often become the main source of urban air pollution in Indonesia [5][6][7][8][9][10]. Till now, several studies have reported that they have an adverse impact on the health of living organisms, such as allergies, harmful diseases (e.g., infections of the respiratory tract, cancers, lung diseases, and cardiovascular) and death [11][12][13]. Due to many adverse impacts, a system which is able to monitor air quality is needed.

Cities in Indonesia have been implementing air monitoring system, called as Air Pollutant Standard Index Measuring Instrument. The instrument often placed in urban or industry area, such as the side of crossroads. It determines air quality index by calculating the average of air pollutant values for 24 hours. After that, it presents the air quality index using a color and simple text through its screen. Unfortunately, it does not work properly in several cities (e.g., Medan). If it works properly, the information about the air pollutant index is only conveyed on

the instrument screen so that the information is only known by people who look at the screen directly. In addition to that, the number of instruments is still limited due to high cost of procurement and maintenance.

Based on the above limitations, this study developed air quality monitoring technology called URAQUM (Urban Air Quality Monitoring). URAQUM is designed to be a low-cost, mobile, fast, and real-time technology for gathering the air pollutant data using sensors installed to URAQUM. The data is useful for AQu-TweetBot System in generating and disseminating the textual information about the air quality index. AQu-TweetBot system will be designed by adopting Data-to-Text Architecture based on Natural Language Generation Approaches. One of the approaches used in this study is a template-based method for Indonesian language. By using this method, the system will selects one of templates by considering the air pollutant data. The selected template becomes the sentence pattern of the air quality index interpretation, where it has a pattern similar to the sentence pattern written by experts in conveying the related information.

This paper is structured as follows: Section 2 presents the previous research related to this study. In Section 3, this paper describes how to develop URAQUM and to design AQu-TweetBot System. Finally, this paper is closed by the nal remarks of this study.

2. Related Work

In recent years, there have been several studies aimed at developing air monitoring technology based on Internet of Things using several sensors.

Amli et al [1] have developed an air quality measurement system using several sensors based on micro controller. These sensors are used to acquire 5 (ve) parameters related to air quality, such as dust particles, sulfur dioxide, carbon monoxide, ozone, and nitrogen dioxide. The acquired values will be sent to server using GSM/GPRS module, and then they are analyzed for getting Air Pollution Standard Index (ISPU). In this study, the presentation of ISPU is still a graphical representation, which is informed via a web application. In other words, only readers who know the address of web application can read the related information.

Kumar and Jasuja [12] presented a system for monitoring air quality based on various parameters, such as PM 2.5, carbon monoxide, carbon dioxide, temperature, humidity, and air pressure. In order to collect the parameters' value, they use the sensor DSM501A, DHT22, BMP180, MQ9 and MQ135, which are connected to Arduino Board and Raspberry Pi. Communication between sensor and client is based on Message Queuing Telemetry Transport (MQTT), where the client can only access the data that is being displayed on the dashboard using the device id without getting modi cation access.

Enigella and Shahnasser [14] performed a research on the creation of a system that could real-time monitor the air pollutant level. The system was designed by utilizing some microcontroller and sensors. The use of these devices and sensors can be alternative low-cost and low-power technology, when the sensor data can be analyzed and processed in other side (e.g. cloud server). After data analysis and processing, Air Quality Index (AQI) is presented by using geospatial representation. Even though this representation is powerful in providing valuable real-time AQI for a certain region, it is still limited in the case of information dissemination, such as via social media. Rahmat also did similar research using Arduino Pro Mini, with DHT11, MQ-135, MQ-7, IMU10DOF, and GPS Dongle [15]

The di erence of this study with the previous research works is that in this study, we implement Natural Language Generation (NLG) approaches to interpret the Air Quality Index (AQI) in the form of a textual representation that will be disseminated to social media. For achieving the main goal of our research, we design the preliminary study for generating AQI interpretation using AQu-TweetBot system based on the air pollutant data collected by URAQUM.

3. Proposed Methodology

This study proposed URAQUM as the technology for measuring and collecting air pollutant data and AQu-TweetBot as the system for generating the textual interpretation of Air Quality Index based on air pollutant data. It illustrates the basic concept of the proposed research, as shown in Figure 1.

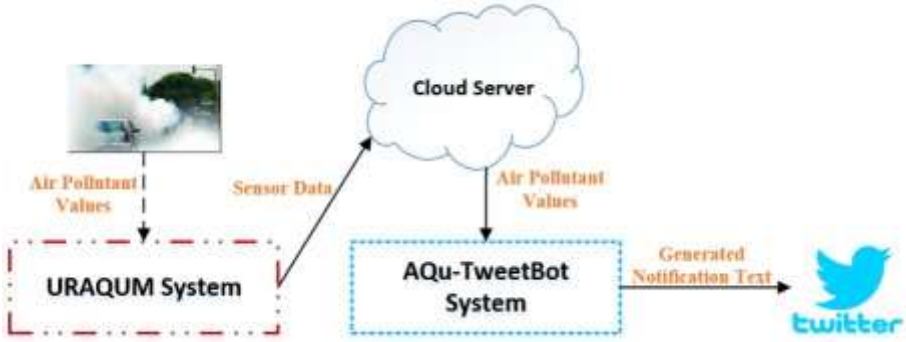


Figure 1: The Basic Concept of Proposed System

3.1. URAQUM System

URAQUM consists of some sensors which are used to measure and collect data of the particles and gases, such as Dust, Carbon Monoxide, Sulfur Dioxide, Smoke, Temperature, and Humidity. These sensors are MQ136, MQ2, MQ7, and DHT11. MQ136, MQ2, and MQ7 are installed to measure Sulfur Dioxide, smoke, and Carbon Monoxide, respectively. Meanwhile, DHT11 is a sensor used for measuring dust level. This study proposed ESP32 for connecting all sensors used in URAQUM. The reason why this study used ESP32 is that ESP32 is one of the microcontrollers, which is low-cost and low-power so that ESP32 is considered to be suitable for the case of this study.

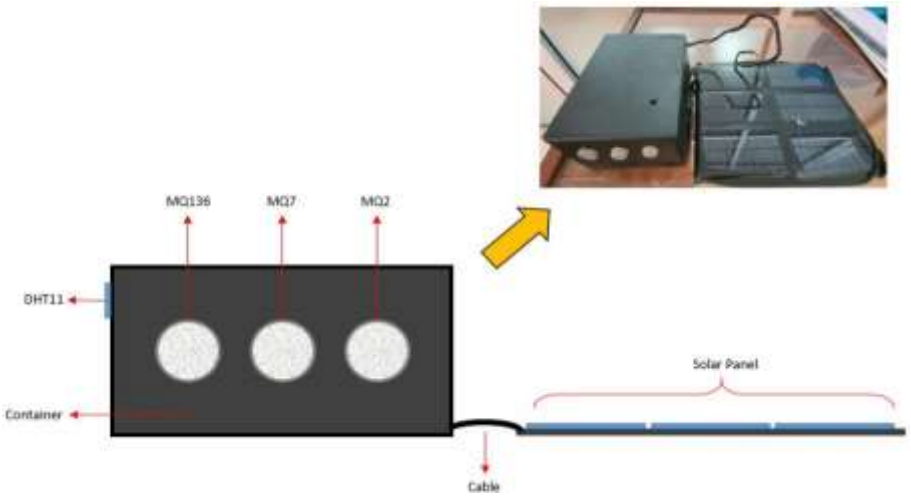


Figure 2: The Prototype of URAQUM

This study has designed the prototype of URAQUM as depicted in Figure 2. The related sensors are placed on the side of the container, where the power of URAQUM originates from Solar Panel which is separated from the container. The data collected by sensors will be continuously trasmitted to cloud using GSM module stored in the container.

3.2. AQu-TweetBot System

AQu-TweetBot system employs Natural Language Generation (NLG) approaches in generating the textual interpretation of air quality index into Indonesian Language. The textual interpretation is generated by selecting the sentence template and then replacing the template's slots with the related data. This method has a limitation in terms of diversity and exibility, so that we will be formulating some text generation rules. For example, if there are two or more air pollutant variables that have the same information content, AQu-TweetBot is able to convey their information in only one sentence. The detailed of the proposed system can be depicted as in Figure 3.

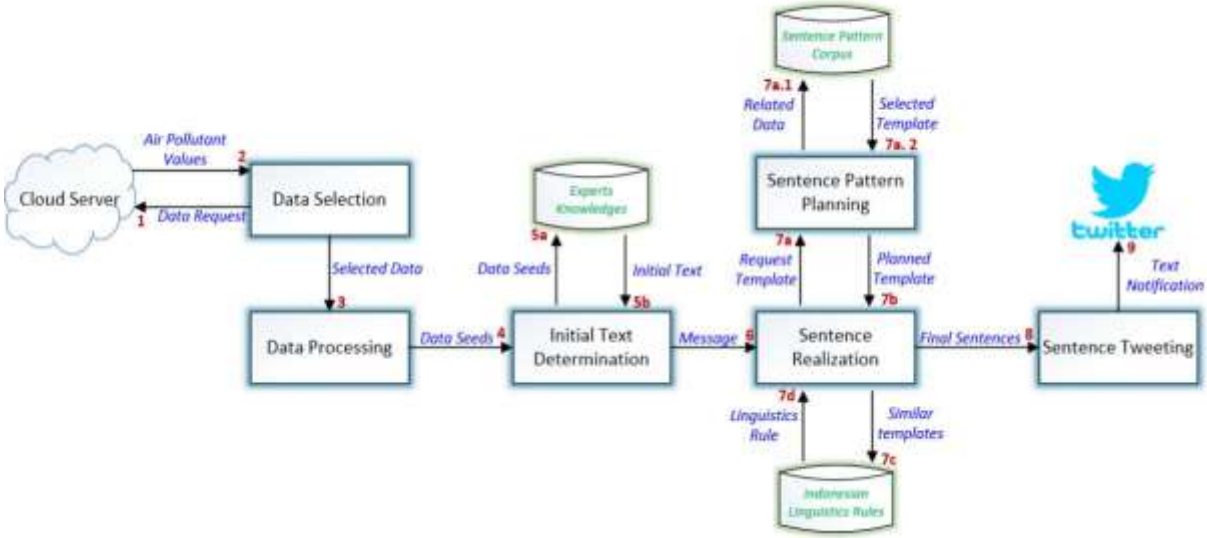


Figure 3: Architecture of AQu-TweetBot System

Based on Figure 3, AQu-TweetBot system consists of Data Selection, Data Processing, Initial Text Determination, Sentence Realization. Sentence Pattern Planning and Sentence Tweeting. Initial Text Determination, Sentence Realization, and Sentence Pattern Planning are the main modules to generate the textual interpretation of the air quality index. First, Data Selection requests the related data to Cloud Server so that Cloud Server provides data of air pollutant values, which has been measured and collected by URAQUM. These data values are called Selected Data, which are structured in the form of JSON format. Second, Data Processing calculates every value of air pollutants to acquire the air quality index. The output of this process is called as Data Seeds. Data Seeds still consist of the data value in numerical representation, so that they should be interpreted into an initial text interpretation by Initial Text Determination. In order to determine the initial text, this module involves the experts' knowledge, such as Good, Moderate, Unhealthy, Very Unhealthy, and dangerous (Air Quality Index can be seen in Table 1. Indeed, this knowledge is based on the Air Quality Index Standard made by credible institutions. After all values have been interpreted, the next process is generating the textual interpretation in the form of a sentence. This process involves two modules, where Sentence Realization will determine the appropriate template for every component of the message based on the list of templates that have been provided by the corpus. It combines the sentence templates which have similar content into one sentence template based on Linguistics Rules. The nal output of this process is nal sentences that will be posted to twitter using Sentence Tweeting.

Table 1: Air Quality Index (AQI)[16]

Range of AQI	Initial Representation (Category)	Description
0-50	Good	Air quality is not harmful to humans, animals, plants, and aesthetic value.
51 - 100	Moderate	Air quality is not harmful to humans and animals, but it is detrimental to sensitive plants and aesthetic value.
101 - 199	Unhealthy	Air quality is harmful to humans and animals, or it can be detrimental to sensitive plants and aesthetic value.
200 - 299	Very Unhealthy	Air quality is harmful to several population segments of living things.
300 - 500	Dangerous	Air Quality is very harmful to a population of living things so that it can provide a severe health impact.

4. Conclusions

The objective of this study is the development of a text interpretation system for conveying the information about the air quality index based on the data of air pollutant to people. The system is implemented by Natural Language Generation (NLG) Approach which has been considered as the good approach for generating textual interpretation. For realizing this objective, this study develops two system, called URAQUM and AQu-TweetBot.

The rst step, URAQUM has been developed properly based on Internet of Things (IoT) technology. However, URAQUM technology is needed to consider the installation place so that it can resist weather conditions, such as rain. As future work, this study will improve the container resistance of URAQUM. The nal step is to create and implement all functions of AQu-TweetBot based on the data collected by URAQUM. Besides, the reliability of proposed method should be measured properly so that this study can discover the e ective and e cient value in this proposed study.

Acknowledgement

We would like to thank the Ministry of Research and Technology and Higher Education Republic of Indonesia who has supported and funded this research through the grant research of TALENTA Universitas Sumatera Utara (USU) 2019, Contract Number 4167/UN5.1.R/PPM/2019 on April 10, 2019.

References

[1] Amli M S, Yuliarto B and Nugraha N 2015 Jurnal Otomasi Kontrol dan Instrumentasi 7 1{8
[2] Air pollution accessed: 2019-07-25 URL https://www:who:int/health-topics/air-pollution
[3] What causes air pollution? accessed: 2019-07-25 URL https://climatekids:nasa:gov/air-pollution/
[4] Penyebab utama polusi udara di perkotaan accessed: 2019-11-25 URL https://www:liputan6:com/health/ read/4030320/penyebab-utama-polusi-udara-di-perkotaan
[5] Sabtu pagi, polusi udara jakarta terburuk di dunia accessed: 2019-11-25 URL https:// www:cnnindonesia:com/teknologi/20190810092804-199-420056/sabtu-pagi-polusi-udara-jakarta-terburuk-di-dunia

[6] Jakarta ada di peringkat 2 udara terkotor dunia pada selasa pagi accessed: 2019-11-25 URL https://www:liputan6:com/news/read/4053328/jakarta-ada-di-peringkat-2-udara-terkotor-dunia-pada-selasa-pagi
[7] Kusminingrum N and Gunawan G 2008 Jurnal, Jakarta, Puslitbang Jalan dan Jembatan
[8] Kebakaran hutan: Pakar kesehatan peringatkan kualitas udara 'sangat berbahaya' akibat asap accessed: 2019-11-12 URL https://www:bbc:com/indonesia/majalah-49738855
[9] Kebakaran hutan: Polusi udara palangkaraya tembus 20 kali lipat batas normal accessed: 2019-11-12 URL https://www:bbc:com/indonesia/indonesia-49711360
[10] So ati I 2010 Berita Dirgantara 11
[11] Marinov M B, Topalov I, Gieva E and Nikolov G 2016 2016 39th International Spring Seminar on Electronics Technology (ISSE) (IEEE) pp 443{448
[12] Kumar S and Jasuja A 2017 2017 International Conference on Computing, Communication and Automation (ICCCA) (IEEE) pp 1341{1346
[13] Kelly F J and Fussell J C 2015 Environmental geochemistry and health 37 631{649
[14] Enigella S R and Shahnasser H 2018 10th International Conference on Knowledge and Smart Technology (KST) (IEEE) pp 182{185
[15] Rahmat R F, Yusuf A, Lini, T.Z 2019 Real-Time Sensor Application for the Measurement of the Vertical Profiles of Atmosphere, 3rd International Conference on Electrical, Telecommunication and Computer Engineering, ELTICOM 2019 – Proceedings, 8943853, pp. 182-188
[16] Hidup K L 1997 Keputusan menteri negara lingkungan hidup nomor: Kep-45/menlh/10/1997 tentang indeks standar pencemar udara