Peer-Reviewed Article

IOTA (Internet of Things and Artificial Intelegence) Journal
Volume 1, Issue 1 2021

Implementation of The Speech Recognition System Using a real time web Server Based

Moch. Viky Aditya [1*], Aries Boedi Setiawan[1], Ir. Abd.Rabi[1]
[1] *Department of Electrical Engineering, University of Merdeka Malang, Indonesia*

*Corresponding Author: viky.aditya@stu.unmer.ac.id

Abstract

a Speech Recognition is a system used to recognize word commands from human voices and then translated them into data that can be understood by a computer. The advantages of Speech Recognition include speed and ease of use. Merdeka Malang is a private university founded on January 29, 1964. The University of Merdeka Malang still doesn't have a strategy in recommending new students based on historical data calculations in choosing faculties. and majors. In calculating the history, if you use a manual system, they will be constrained by the time and accuracy of the data, causing the data to be less valid. Therefore, there is a need for a system that can exchange data in real-time so that the data displayed is always up to date without manual data updating. This process is useful for monitoring and accelerating the process of calculating student interest data, which can later be recommended to future new students. Moreover, The speech technique used is using speech recognition which is used for processing speech data on smartphones while sending data using the XAMPP web server which is used to exchange data from smartphones to servers.

Keywords: speech, recognition, realtime, web, server

1. Introduction

The University of Merdeka Malang is a private university established on January 29, 1964, which is organized by the Malang Central Merdeka Higher Education Foundation (YPTMPM) in Malang (Deed No. 5.a July 5, 1964). This educational institution was founded by R. Edwin Soedardji, Soekiman Dahlan, SH., Frasnsiscus Soetrisno, Soegondo, Soetikno, SH., Dharma. In 1972 the name YPTMPM was changed to the Malang Merdeka Higher Education Foundation, which was abbreviated as YPTM. Every year, prospective new students register through the test and non-test routes provided by the University of Merdeka Malang. However, these new students are often confused about choosing the faculties and majors to take. Meanwhile, the University of Merdeka Malang campus still does not have a strategy in recommending new students based on the calculation of the history of previous student enthusiasts in choosing faculties and departments. In calculating history, if you use the manual system, the University of Merdeka Malang will be constrained by the time and accuracy of the data, causing the data to be less valid.

Therefore, there is a need for a system that can exchange data in real-time so that the data displayed is always up-to-date without doing manual data update processes. This

process is useful for monitoring and accelerating the process of calculating student interest data, which can later be recommended to the new student. The use of smartphones at this time by various people has become an important necessity in life. This has prompted any university that manages the system to start focusing on developing systems on smartphones. Among the several mobile systems that exist today, Android is getting a lot of attention from many circles. The prospect of developing the android system in the future is felt to be even better. An example of a method for sending data is the Short Message Service (SMS). However, the Short Message Service (SMS) is an old system, so it needs further development, namely using Speech Recognition.

This speech recognition will later be used as a substitute for Short Message Service (SMS), where the University of Merdeka Malang enters data on the number of faculty and department students in the form of speech (voice) and then reads it by an android smartphone. After that, it is sent to the server using the internet network on each Android smartphone. Furthermore, the data is received by the server in realtime. Speech recognition is a system used to recognize word commands from a human voice and then translated them into data that can be understood by a computer. The advantages of this speech recognition include the speed and ease of use. The words that are captured and recognized can be the result, for an application such as command & control, data entry, and document preparation. The parameter being compared is the level of sound emphasis which will then be matched with the available database templates.

While the voice recognition system based on the person speaking is called speaker recognition [1]. The web server is a system that provides data-based services and functions to receive requests from HTTP or HTTPS to clients. This system is widely used in information systems, telecommunications, commerce, banking, and others. The main function of the Web server is to transfer user request files through predefined communication protocols, including text files, videos, images, and files. This webserver was developed by apache so that it has advanced features such as configurable error messages, database-based authentication, and more. Apache is also supported by a number of graphical user interfaces (GUIs) that allow easy server handling. Apache is open-source software (open source) developed by an open community consisting of developers under the Apache Software Foundation. Therefore, in this final project, the writer will do "Implementing Speech Recognition System Using Android-Based Realtime Web Server".

2. Literature Review

Speech recognition is a system used to recognize word commands from a human voice and then translated them into data that can be understood by a computer. The advantages of this speech recognition include the speed and ease of use. The words that are captured and recognized can be the end result, for an application such as command & control, data entry, and document preparation. The parameter being compared is the level of sound emphasis which will then be matched with the available database templates. Meanwhile, the voice recognition system based on the person speaking is called speaker recognition. However, the complexity of the algorithm that is implemented by speech recognition is simpler than speaker recognition. Speech recognition is also a technique and system development that allows computers to receive input in the form of spoken words. This technology allows a device to recognize and understand spoken words by

digitizing words and matching the digital signal with a certain pattern stored in a device. The words spoken are transformed into digital signals by changing sound waves into a set of numbers which are then adjusted to certain codes to identify the words. The result of identifying the spoken word can be displayed in written form or can be read by a technology device as a command to do a job, for example pressing a button on a cell phone which is done automatically with a voice command. Currently, speech recognition is used to replace the role of input from the keyboard and mouse, in this case, it can replace the role of the touch screen on Android-based smartphones (Iizuka, S., 2012).

Sound is a type of longitudinal mechanical wave that is generated by the vibration of an object and can be detected by the ear. The sound occurs due to vibrations from an object and can be detected by the ear. The sound occurs because of the vibration of every substance. The substance can be air, water, wood, or other material. The only place where sound cannot propagate is the vacuum. When substances vibrate or move forward and backward rapidly, they produce sound. Our ears collect these vibrations and allow us to interpret them. A sound wave can be described by its frequency and amplitude. The frequency of 1 Hz means 1 complete wave cycle every one second. The unit of frequency is Hertz (Hz). The frequency that can be picked up by the human ear is between 20 Hz to 20 000 Hz. The amplitude of a wave refers to the amount of change in pressure and the loudness of the sound wave (Bhiksha, Raj dan *Rita Singh*. 2013).

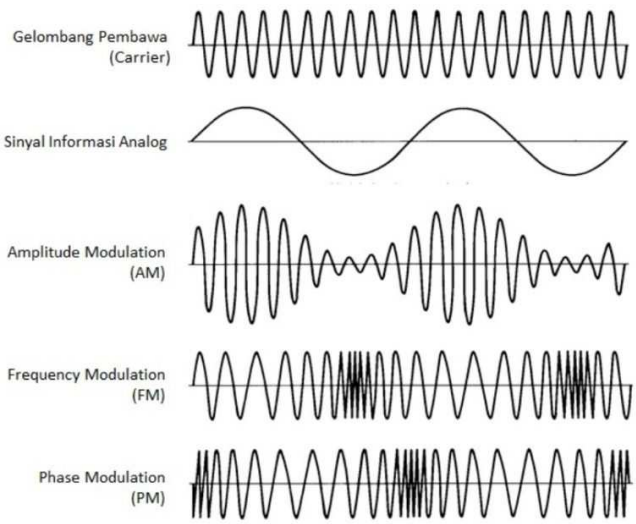

Figure 1. Signal Variation

Sound waves vary as well as variations in the pressure of an intermediate medium such as air. Sound is created by the vibrations of an object, which causes the surrounding air to vibrate. These air vibrations then cause the human eardrum to vibrate, which the brain then interprets the sound. Sound waves travel through the air in much the same way as water waves travel through water. However, because water waves are easy to see and understand, they are often used as an

analogy to illustrate how sound waves propagate. Figure 1 Sound Wave Propagation Process Each air molecule moves at a small distance as a vibration but causes adjacent molecules to vibrate all affected by traveling to the ear. All waves have three important properties for audio work including wavelength, amplitude, and frequency. Sound waves can also be shown in a graph as shown in Figure 2.

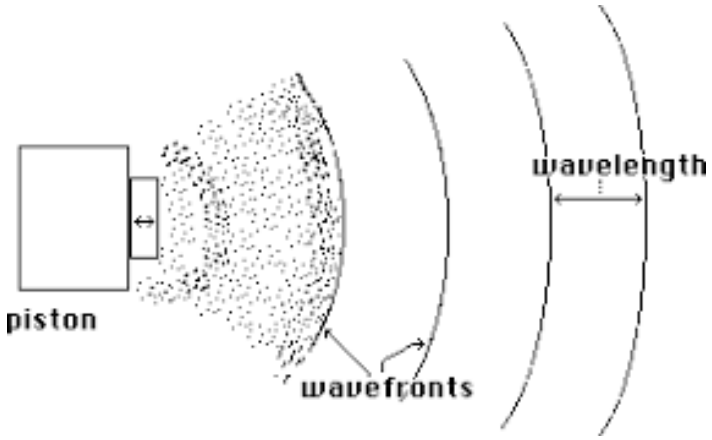

Figure 2. Sound Wave Propagation Process

Data processing used for this research is speech (voice). Speech data (voice) is used for input (input) before sending to the server. Speech (sound) detection is done using an android smartphone. Speech (voice) that is detected is in the form of characters or numbers. This data is needed in sending data to the server for real-time processing. Besides, when the data is sent, the process speed test will be carried out for the length of time to find out the traffic in data traffic from the user's smartphone to the server data. Furthermore, Real-Time Processing is a control mechanism, recording data, processing very fast so that the resulting output can be received at relatively the same time. The difference with the online system is that the time unit used in real-time is usually one hundredth or one-thousandth of a second, while on-line is still on a scale of seconds or sometimes even a few minutes. Another difference, online usually only interacts with users, while real-time interact directly with users and the mapped environment. Real-time processing will greatly simplify the company's cash cycle. Real-time systems with computer terminals connected to a central computer will reduce or even eliminate obstacles e.g., delays in data transmission.

3. Method, Data, and Analysis

From the block diagram in Figure 3, it can be seen that the user enters data in the form of data on the number of faculty and department students. Then it is processed using an Android smartphone to receive voice input from the user. After that, the data is sent to the server and processed in a database and then displayed on a PC to see the results of the comparison of the number of data from faculty and department students at the University of Merdeka Malang which are used for recommendations for future new student candidates.

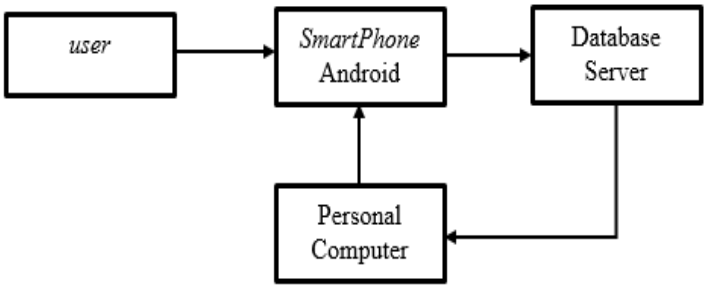

Figure 3. Block Diagram

This basic concept is illustrated through a flowchart as in the Figure 4, which explains the basic flow of research from the first step.

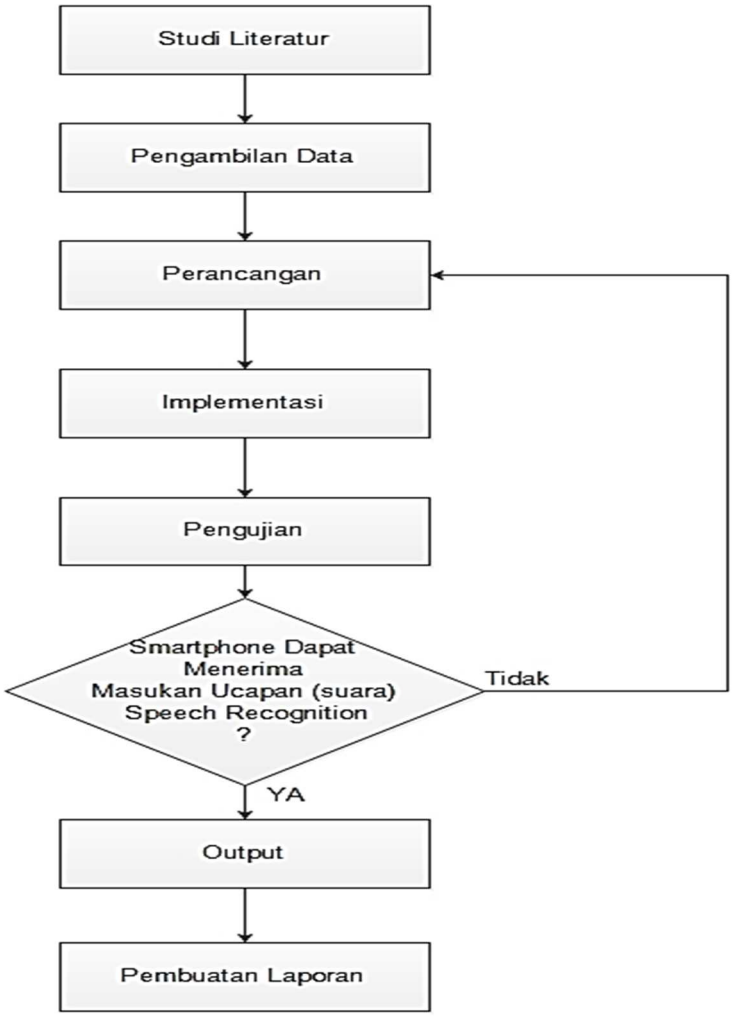

Figure 4. Flowchart system on this research

4. Results

This implementation and testing aim to determine whether the tools that have been designed and made run as planned. The tests carried out include hardware and software testing. (Figure 5).

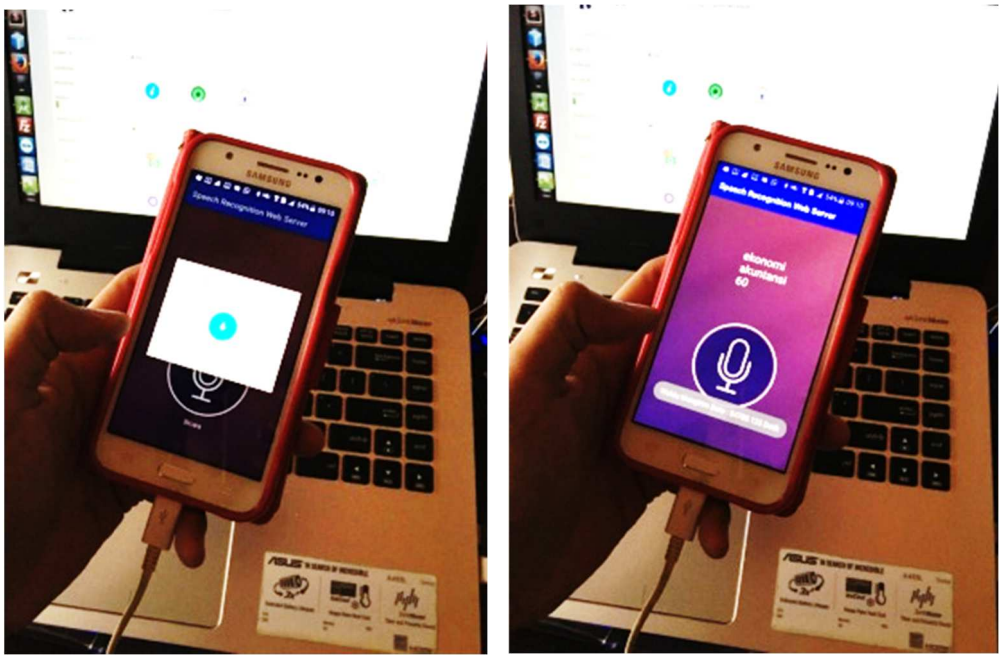

Figure 5. Testing Speech To Text Android, Speech Receiving Results (Voice) on a Smartphone Device

Speech to Text Distance Testing

At this stage, the test is carried out on an Android smartphone to make the process of changing voice to text based on the distance (cm) between user input and the Android smartphone. This is done, to know the performance of the device when receiving voice input. Can be seen the test results in Table 1.

Table 1. Results of Testing Distance (cm) Between User and Smartphone

No	The spoken word	Distance (cm)	Display
1.	Teknik Elektro 120	20	legible
		40	legible
		60	legible
		80	legible

		100	Can't be read

Based on Table 1, it can be concluded that the effective distance in this test is at least 20 cm and a maximum of 80 cm. The system success rate of 80%. Voice input must be clear so that the system can convert voice to text based on user input.

Testing Changes from Voice to Text

At this stage, the test is carried out on an android smartphone to make the process of changing the voice to text. This test is carried out by entering sound parameters in the form of Electrical Engineering 100 speech, you can see the test results in Figure 6 and Figure 7.

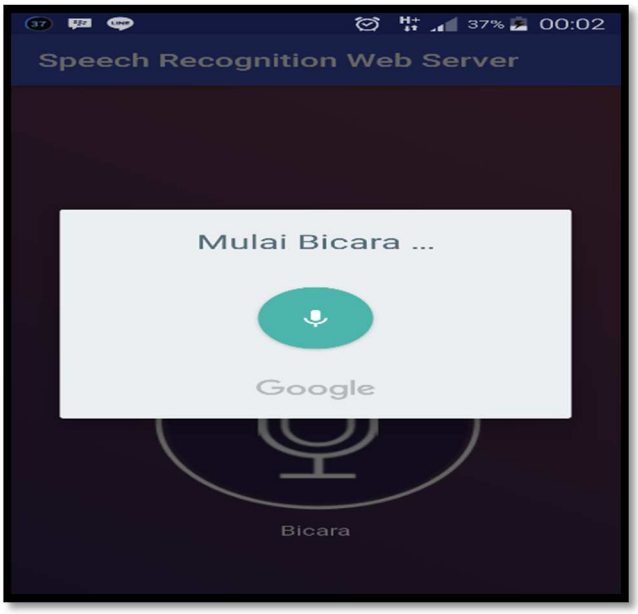

Figure 6. Voice Input Form button

In Figure 7, the initial step for the user to press the button provided by the system then the system displays the voice input form. After that, the user enters the sound parameters which include faculties, majors, and numbers. The test this time was the words uttered by Electrical Engineering 100, Faculty of Psychology 127, and Information Technology 183. Furthermore, the system carried out the process of reading the voice and changing the voice into text.

Figure 7. Voice Reading and Voice Change to Text

Table 2. Test Results of Voice Changes to Text

The spoken word	Result Display
Teknik Elektro 100	

Fakultas Psikologi 127	
Teknologi Informasi 183	

Based on Table 2. it can be concluded that the results of testing the input process, reading voices, and changing the voice to the text were successfully carried out by the system. It can be concluded that the tools made work optimally when carrying out the processes that have been tested. The first step is to create a Speech Input prompt function than to declare a variable using the intent command. After that call the speech library using the ACTION_RECOGNIZE_SPEECH command. The next process is to extract the language used on the Android smartphone using the EXTRA_LANGUAGE_MODEL and EXTRA_LANGUAGE commands. The locale. getdefault () command is to set the language by default used on the smartphone. The EXTRA_PROMPT command accepts input and voice readings from the user.

Testing the Web Server to Receive Data from an Android Smartphone

Figure 8. Testing with a Web Server

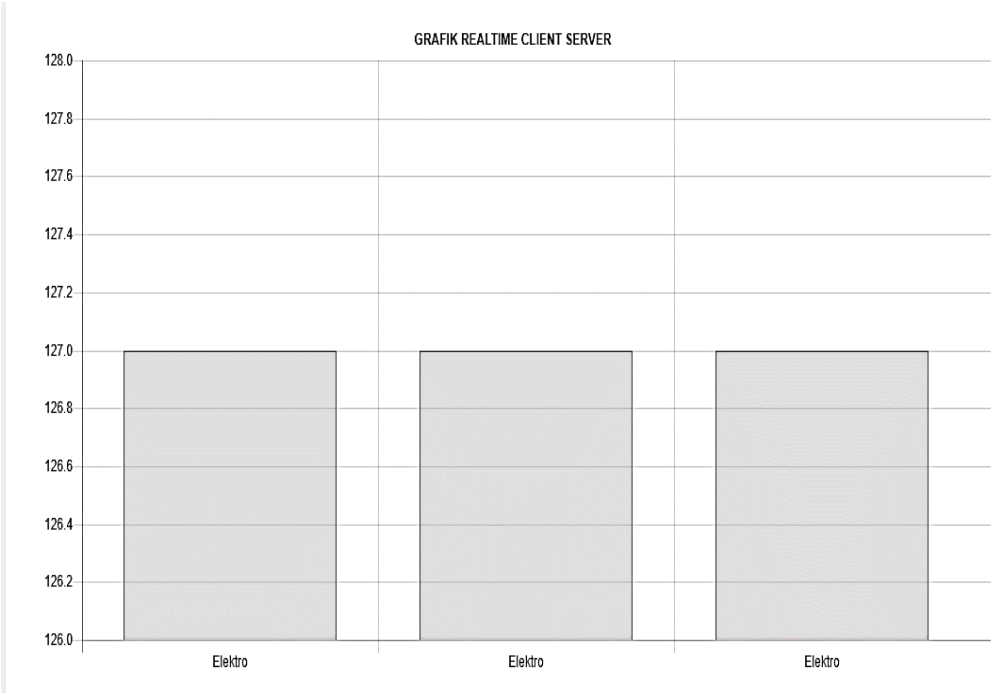

Figure 9. Receive Data from Android Smartphone with Speech, Same Amount and Sent Simultaneously

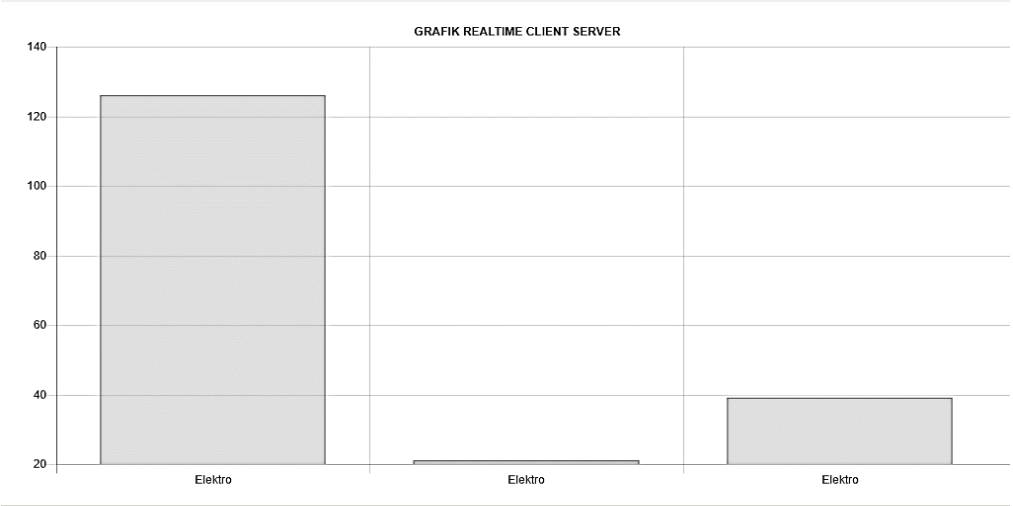

Figure 10. Receive Data from Android Smartphone with the same Speech, Different Amounts and Sent Simultaneously

At this stage, it is to test the web server that has been created. In figure 8. Users using 3 smartphones enter the Electrical Engineering speech command 127 simultaneously. Furthermore, the experiment step is a User using 3 smartphones to enter Electrical Engineering greeting commands with different numbers and are pronounced simultaneously. And the data sent can be received on the webserver. Analyze result on the Figure 9 and 10.

5. Conclusion

Android smartphone can be used as a tool to receive input sound parameters from the user based on the distance that has been tested and can send these parameters on a laptop using an online web server, therefore, that the system that has been created on a laptop can receive data transmission from an Android smartphone in real-time. Furthermore, from the test results, the speed in receiving voice parameter input data from the user on an Android smartphone is based on the distance between the user and the Android smartphone. If the distance between the user and the Android smartphone is getting closer, the faster the speech will be read by the Android smartphone. While sending data from an Android smartphone to a laptop is based on the influence of the internet connection used. This is because the longer the internet connection is used, the longer it will take to send data.

6. Suggestions

A good internet connection or high bandwidth is required in conducting voice recognition experiments using an Android smartphone.

References

Mikio Nakano, et.al., 2014, "Speech recognition apparatus, speech recognition method, and speech recognition robot", Patent :US8886534B2, https://www.uspto.gov/terms-use-uspto-websites

Dhairya Desai, et.al, 2020, "Visual Speech Recognition", International Journal of Engineering Research & Technology (IJERT),Vol. 9 Issue 04, April-2020, DOI : 10.17577/IJERTV9IS040509

J Gowri, N Lakshmipriya,2020,"A Voice Recognition Controlled Web Browser", The International Journal of Scientific Development and Research - IJSDR (IJSDR), March 2020, Volume 5, Issue 3, ISSN: 2455-2631

Iizuka, S., Tsujino, K., Oguri, S., & Furukawa, H. 2012. Speech Recognition Technology and Applications for Improving Terminal Functionality and Service Usability. NTT DOCOMO Technical Journal Vol. 13 No. 4 , 79-84.

Bhiksha, Raj dan Rita Singh. 2013. Design and Implementation of SpeechRecognition Systems. Carnegie Melon University. United States.

Nazaruddin Safaat H. 2012. Android Pemrograman Aplikasi Mobile Smartphone dan Tablet PC Berbasis Android. Informatika. Bandung.

Adi Nugroho, “Pemrograman Java Menggunakan IDE Eclipse”, Andi Publisher, Jakarta, 2008

Iskak, Ahmad dan Yustinah. 2008. Bahasa Indonesia. Erlangga.Kumar, C. S., & Rao, P. M. 2011

Martadipura,Bambang Avip Priatna, 2013. Populasi dan Sampel.

Kismtianti, 2010. Pengumpulan data dengan quick count dan exit poll