Implementation of Optical Character Recognition and Voice Recognition of House of Words (How) Dictionary Application on Android Platform

Hutami Septiana Raswaty[1], Nuryuliani[2*]

[1,2] Information System Department, Information System,
Gunadarma University,
Jakarta, Indonesia 11480
hutami.sraswaty@student.gunadarma.ac.id;
nryulia@staff.gunadarma.ac.id

*Correspondence: nryulia@staff.gunadarma.ac.id

Abstract *- People with different languages need to be assisted by translator to establish the communication between them. The technology development which exists to fulfill communication needs is digital dictionary as translator tool. The capability of digital dictionary to translate the languages yet has a weakness in putting the input. Through this research, Optical Character Recognition using Tesseract library and Voice Recognition technologies using Google Speech-To-Text are used to replace the previous input system. Based on the implementation and testing, the OCR and Voice Recognition have been successfully recognizing the text and voice input with the amount of similarity of 92,72% for OCR and 95,46% for Voice Recognition. The result of the implementation is expected to help a group of people with different language to communicate easily.*

Keywords: *Android; Dictionary; Optical Character Recognition; Speech-To-Text; Text-To-Speech; Voice Recognition*

I. INTRODUCTION

The development of technology in early decade has helped fulfilling the communication needs especially for different group of people with different languages. The example of communication technology development is digital dictionary as a translator tool. The existence of digital dictionary has replaced the existence of the conventional dictionary with its physical form. The digital dictionary easily support the communication needs between groups of people with different languages to translate not only a single word but also sentences without searching for the words manually through the alphabet.

The capability of current digital dictionary to translate the languages needs its user to type the input manually through keyboard, while the nature of language is a voice system and a tangible symbol [1]. Therefore, synchronizing the natural language concept with the current digital dictionary input is needed.

The current development of technology not only has helped fulfilling the communication needs but also has helped stimulating the growth of smartphone user. GSMA Intelligence, Worldometers, gather the data that 7.529 billion of people around the world are the active users of cellphone, while Research Institute of Digital Marketing, Emarketer, has estimated that there are more than 100 million of smartphone active users in Indonesia by 2018. The existence of smartphone as a mobile device opened up the opportunities for solving various problem in various fields through mobile application. Therefore, as the way to synchronize the current digital dictionary with the natural language concept and the growth of smartphone user, the mobile application of digital dictionary using Optical Characteristic Recognition and Voice Recognition for Android platform is made for supporting the current needs of communication between people by replacing the previous input process in the current digital dictionary.

The Optical Character Recognition (OCR) and Voice Recognition is used in the House of Words (HOW) Dictionary Application. The OCR is made by using Tesseract library and Leptonica Image while The Voice Recognition is made by using the Google Speech API as speech-to-text and text-to-speech tools, integrated with Google Translator API. Supporting the natural of language concept, OCR and Voice Recognition used in HOW Dictionary is intended to enable people to scan directly from text image with many special character or symbol and to be directly speech their words and/or sentences that will be translated.

Copyright © 2021

Literature Review

1. Optical Character Recognition (OCR)

Optical Character Recognition (OCR) is reading activity of the typed character, computer printouts, or handwriting text from any regular documents then translate those image document to be a form which can be processed by computer. The simple definition of OCR is recognition process of printed text, handwriting text, typewriting text, etc., from hard copy document to digital document [2]. OCR is used for any activities related to image documentation such as book pages image, legitimate document, medical record, and so on.

Morphological difference in character forms can be difficult or impossible to detect, particularly in the presence of printing and scanning artifact. OCR result may contain segmentation as well as classification error due to low image quality [3]. The better quality of the input data, then the OCR performance will approach the human ability for recognizing the character, on the other side, OCR will be very difficult and even fail on doing any detection [4].

The OCR systems have some common steps that can be seen on Figure 1 [5].

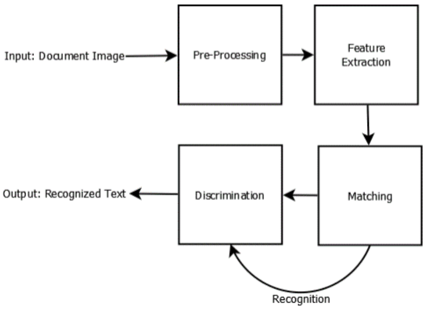


Figure 1. Flowchart of General Stages of OCR System [5]

2. Speech Recognition

Speech Recognition is defined as conversion process from sound signals into machine linguistic in the form of digital data (usually in the form of simple text). In the other words, speech recognition states the ability of pattern matching given from the vocabulary of sound signals into the right form. Another definition of speech recognition is a process where the computer (other types of machine) able to recognize the words spoken by human. This process is also called as interpretation of human speech in computer. The voice recognition programs are made for recognizing the vocabulary of words from pre-programmed words [6].

3. Tesseract

Google Tesseract is Optical Character Recognition (OCR) engine for any operating system which was originally developed by Hewlet Packard since 1985 to 1995 [7]. After 2006 Tesseract is released to be used freely, since the Tesseract is being developed widely by Google and released under Apache 2.0 license [7]. Smith in ICDAR Conference said that Tesseract is combining both matrix matching algorithm and feature extraction algorithm. Tesseract requires less training data and using both statistic classifier and adaptive classifier to enable the machine to train itself on the analyzed document [7]. Flow process of Tesseract engine shown in Figure 2.

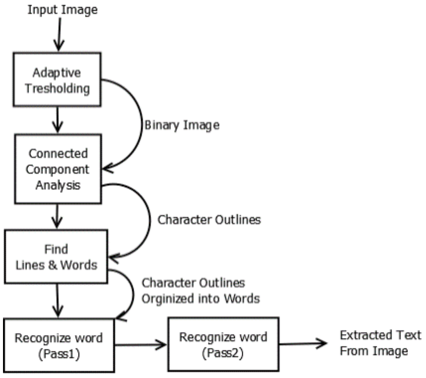


Figure 2. Flow Process of Tesseract Engine [8]

4. Google Speech API

Google speech API or Google Voice search was released on 2008 in USA for some type of smartphone. Google Speech API is a framework developed by Google to recognize the voice, convert it into string (text) and insert it to Google search page until the result appears based on the voice input. Voice recognition is done in Google server, which means the input that is accepted by the Android device (smartphone) will be sent to Google server, then Google server will recognize and convert it into the text using Hidden Markov Model Algorithm. After that, the result of voice conversion will be inserted into Google search page, then Google server will send the search result to the Android device [9].

The search page interface provides all types of information and service in internet, but user must search through the result list first to get the desire result. Development of voice-based search has some advantages. The advantages will be described as follows [10]:

1. Allows to call certain application on terminal or web service.
2. Using the language processing technology and classification formula to automatically determine which categories belong to user speech and suggest the right application.
3. Combine layers that provide easy access to other applications in other categories related to speech.

5. Google Translator API

Google Translate is an online multi language translation service. Google uses Googlebot to support this service. Compared with Google Dictionary, Google Translate able to translate a whole page of a book at a time. Google gives the discretion to the user to edit the translation result according to the standards language rules [11].

6. Java Programming Language

Java is a programming language which can be run on any kind of computers and operating systems including mobile phones. Java is developed by Sun Microsystems and released in 1995. Java is a software technology which is categorized as multi-platform. In addition, Java is a platform which has virtual machine and library needed for writing and running a program [12].

The advantages of Java programming language will be described as follows:

1. The main advantage of Java programming language is it can be run on multiple platform or computer operating system, along with the principle ‘writing once, run anywhere’.
2. Object Oriented Programming (OOP): Java is one of the natural object-oriented programming languages. All data types are being inherited from the base class called Object.
3. The completeness of the library and the existence of a great Java community which is constantly making new libraries for covering the needs of application developer.
4. Automatic Garbage Collection: Java has memory usage control so that the programmers do not need to directly set the memory [12].

II. METHODS

2.1 Requirement Definition

House of Words (HOW) dictionary is a digital dictionary application using Optical Character Recognition and Voice Recognition technology which allows the user to give an input in uncommon ways. User can easily give an input by scanning the text contained in an image or document, or they can simply speech the words or sentence to the application. Those input will be processed by the application to give the user a translated words or sentence from what they have given before. The output is divided into two types, the first one is a translated text which can be saved as .txt file, and the second one is performed as sound which allows the application to speech the translated word or sentence to the user.

To use the application, user must give an image or voice as an input. While user give an input, the application will catch the input, if it is voice, the input will be filtrated first through microphone, and if it is an image the application will get it from camera or gallery. Next, the image input will be continued to cropping process. After finishing those steps, the user must choose the origin language of a text in OCR feature or both origin language and translation language of a voice in Voice Recognition feature, this step require Google Translator API integration to get library language. In OCR feature, before the text can immediately be translated, the user must choose the segmentation mode of OCR they want to use, then the image will be continued to image processing stage using tesseract engine. After that, both OCR feature and Voice Recognition feature can be continued to translation process. In translation process, both features are integrated with Google Translator API. The result of both processes can be spoken out as the application integrated with Google Speech-To-Text API and Text-To-Speech.

2.2 Application Design

Activity diagram is used for describing the workflow and activity running between user and system while user use House of Words (HOW) Dictionary application. It is used to describe about how the system starts, the decision which happen during the process, and how the system stop.

The activity diagram of OCR feature can be seen on Figure 3.

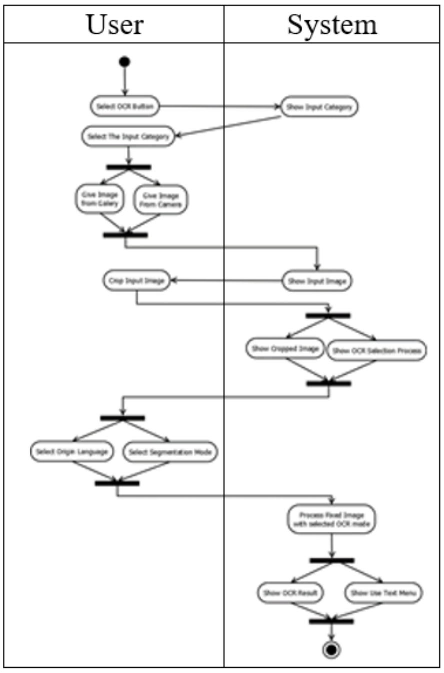

User
System
Select OCR Button
Show Input Category
Select The Input Category
Give Image from Galery
Give Image From Camera
Show Input Image
Crop Input Image
Show Cropped Image
Show OCR Selection Process
Select Origin Language
Select Segmentation Mode
Process Fixed Image with selected OCR mode
Show OCR Result
Show Use Text Menu

Figure 3. Activity Diagram of OCR Features

User must choose the way they want to give an input image. It can be from device camera or device gallery. After giving an input, the input will be continued to cropping process, the user can adjust the text lied in the image. Next, the user will be brought to segmentation-mode page. At this stage, the user can see the cropped image and select the segmentation mode they want to use in the OCR process along with the origin language of the text lied in image. Once the user chooses to process the image with its selected mode and origin language, the image will be processed using Tesseract engine. Last, the result will come up along with the other features in use text page.

The other features lied in OCR function is translation feature. Activity diagram of translation in OCR feature can be seen on Figure 4.

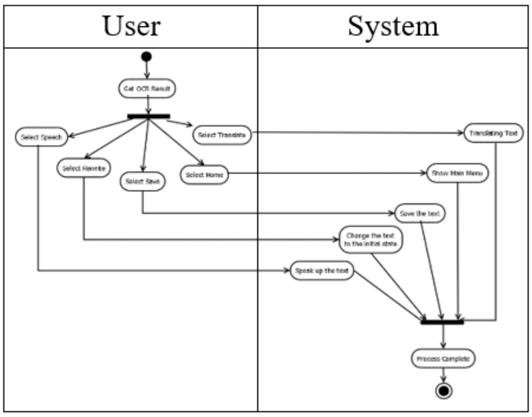

User
System
Get OCR Result
Select Speech
Select Translate
Select Favorite
Select Save
Select Home
Translating Text
Show Main Menu
Save the text
Change the text to the initial state
Speak up the text
Process Complete

Figure 4. Activity Diagram of Translation in OCR Feature

The activity diagram of voice recognition feature can be seen on Figure 5.

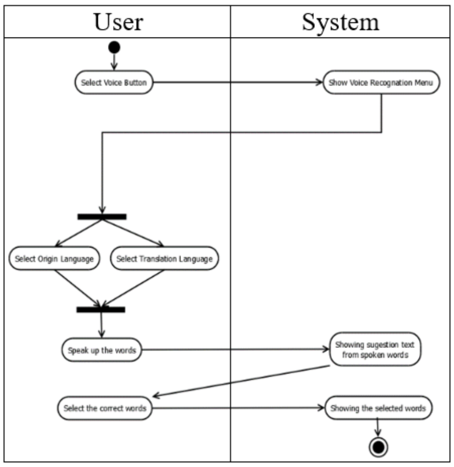

User
System
Select Voice Button
Show Voice Recognation Menu
Select Origin Language
Select Translation Language
Speak up the words
Showing sugestion text from spoken words
Select the correct words
Showing the selected words

Figure 5. Activity Diagram of Voice Recognition Feature

The other feature of the application is Voice Recognition. The user can access this feature through main menu page. First, user must choose the origin and translation language before speaking up the words/ sentences. Once the words is spoken up to the application the application will catch the sound and shows the list of suggestion words to the user. The most accurate word is put at the top of the list. Next, the user selects the most correct words/ sentences with the words/ sentences they have spoken before.

The activity diagram of translation in voice recognition feature can be seen on Figure 6.

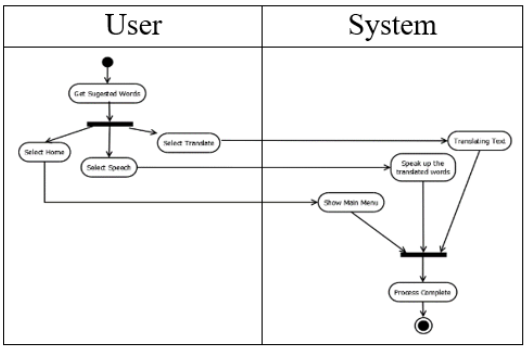

User
System
Get Sugested Words
Select Translate
Translating Text
Select Home
Select Speech
Speak up the translated words
Show Main Menu
Process Complete

Figure 6. Activity Diagram of Voice Recognition Feature

After the most accurate words/ sentences prediction is selected, the user can choose other feature lied in voice recognition page such as translate, and speech. Both features have the same function as they are in OCR page, which are used for translating and speak out the words/ sentences.

The navigation structure is made for describing the navigation process in HOW (House of Words) Dictionary application can be seen on Figure 7.

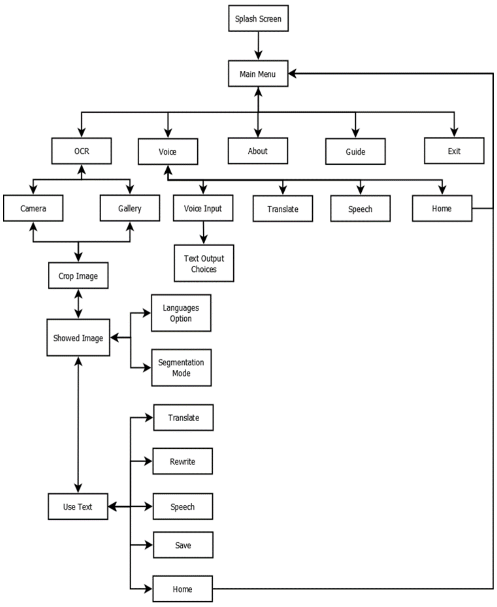

Splash Screen
Main Menu
OCR
Voice
About
Guide
Exit
Camera
Gallery
Voice Input
Translate
Speech
Home
Text Output Choices
Crop Image
Showed Image
Languages Option
Segmentation Mode
Translate
Rewrite
Use Text
Speech
Save
Home

Figure 7. Navigation Structure of HOW Dictionary Application

III. RESULT AND DISCUSSION

3.1 Implementation

The main menu page contains all the main feature of the application, which are OCR feature and Voice Recognition feature. Through this page, the user can start to translate using one of the main features. The main menu page itself contains 4 menus, which are OCR menu, Voice menu, Guide Menu, and About Menu. Main Menu Page can be seen on Figure 8.

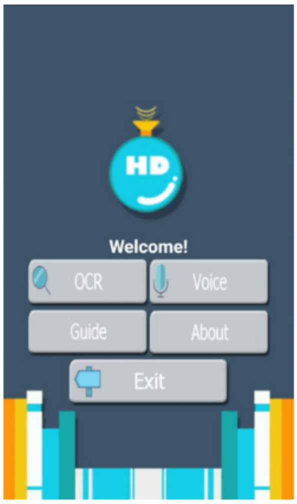

HD
Welcome!
OCR
Voice
Guide
About
Exit

Figure 8. Main Menu Page

One of main features in HOW Dictionary Application is lied on OCR page. The OCR page is the main feature page which can be accessed once the user arrived on main menu page. This page leads the user to give an input through the gallery or taking the picture directly from the camera button. The OCR Menu Page can be seen on Figure 9.

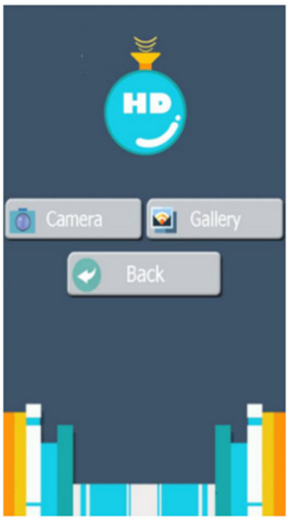


Figure 9. Main Menu Page

The users have to give an input by clicking the camera button or gallery button on OCR Page. Those buttons will bring user to two different ways for giving image input. If the camera is chosen, then the application will activate the camera on user device so that make the users able to take the picture directly. Camera feature can be seen on Figure 10.

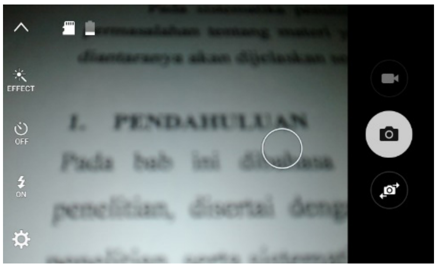
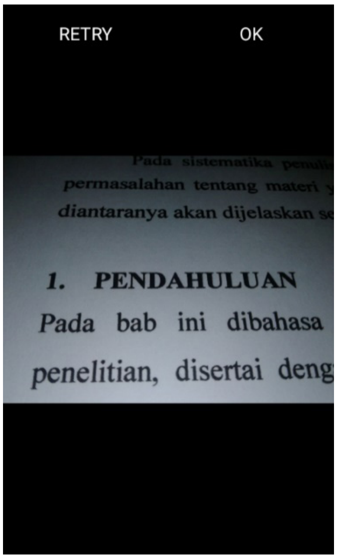


Figure 10. Input Image from Camera

The picture is taken with the flash mode on to make sure that the brightness is enough for making the text readable by the tesseract engine. After the picture is taken, then the application will ask you whether you want to retry taking the picture or not. It can be seen on Figure 11.

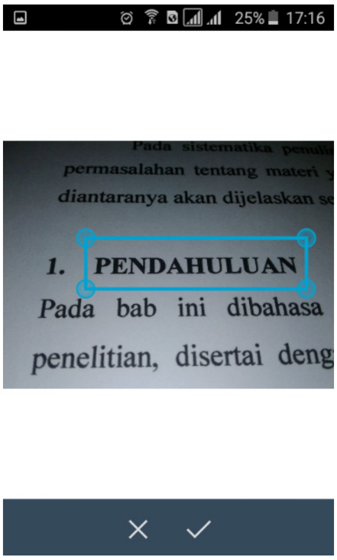


Pada sistematika penuli
permasalahan tentang materi y
diantaranya akan dijelaskan se

1. PENDAHULUAN
Pada bab ini dibahasa
penelitian, disertai deng

Figure 11. Fixing Input Image

If the user chooses RETRY then the application will open up the device camera to enable the user takes the picture again, but if the user chooses OK then the Crop Image page will appear, it can be seen on Figure 12.

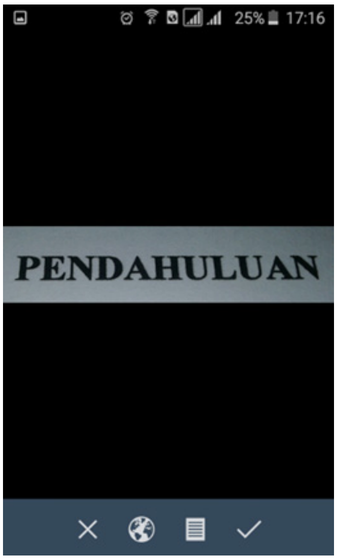


Pada sistematika penuli
permasalahan tentang materi y
diantaranya akan dijelaskan se

1. PENDAHULUAN
Pada bab ini dibahasa
penelitian, disertai deng

Figure 12. Crop Image Page

Crop Image page enables user to adjust the text in the picture that has been taken before. At this implementation, PENDAHULUAN is the text that will be taken to be translated to English language. If the adjustment is finish, then the user can click on cross button or check button. Cross button is intended to cancel the OCR process while in crop image page. Check button is intended to bring the user to the next page in OCR feature, which is The Segmentation Mode page that can be seen on Figure 13.



PENDAHULUAN

Figure 13. SegMod Page

There are two symbols to select the mode of OCR process, the globe symbol, and list symbol. Globe symbol contained language list and used for selecting the origin language of the cropped image, and list symbol contains the segmentation mode which is used for processing the image. The list language can be seen on Figure 14.

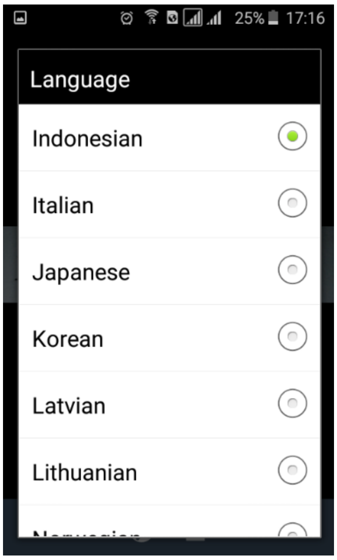

25% 17:16
Language
Indonesian
Italian
Japanese
Korean
Latvian
Lithuanian

Figure 14. Language List

The segmentation mode page shows the cropped picture taken from the previous page. The Segmentation Mode can be seen on Figure 15.

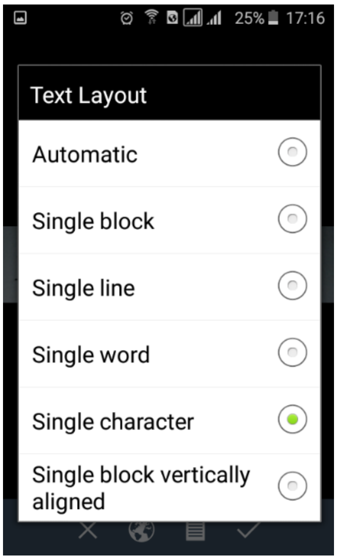

25% 17:16
Text Layout
Automatic
Single block
Single line
Single word
Single character
Single block vertically aligned

Figure 15. Segmentation Mode

After selecting the origin language and segmentation mode. If user click on the check symbol, then the picture will be processed by the application and the use text page will appears along with the OCR result in Figure 16.

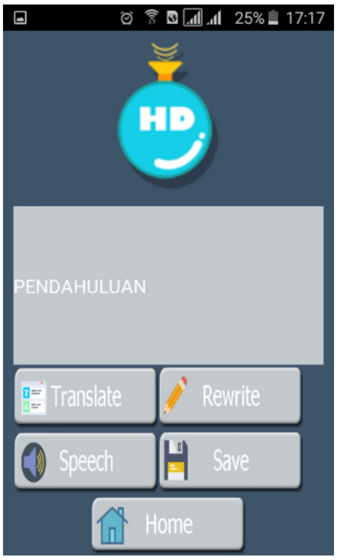
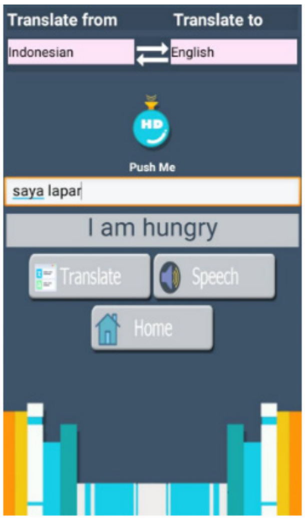

PENDAHULUAN
25% 17:17
HD
Translate
Rewrite
Speech
Save
Home

Figure 16. OCR Translation Page

Voice Recognition Page appears when user click on Voice Button at Main Menu Page. Voice Recognition Page contains some features inside such as spinner, button, image view, and text view. Spinner is divided into two types, the Spinner 1 is intended to give a list of languages which can be used as the original or source language used for translation, while the Spinner 2 is intended to give a list of languages which can be used as the destination language for translation. Reverse button is intended to swap the source language in Spinner 1 with destination language in Spinner 2. Speech button is intended to give a voice as an input which will be converted into text, then the converted text will be translated, after that the translated text will be spoken out by the application. Voice Recognition Page can be seen on Figure 17.


Translate from: Indonesian
Translate to: English

saya lapar

I am hungry
HD
Push Me
Translate
Speech
Home

Figure 17. Voice Recognition Page

Guide Page appears when user click on Guide Button at Main Menu Page. Guide page is intended to show the instruction about how to use HOW (House of Words) Dictionary application that can be seen on Figure 18.

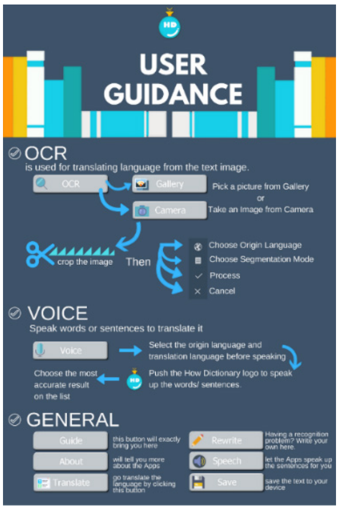

USER
GUIDANCE
OCR
is used for translating language from the text image.
OCR
Gallery
Pick a picture from Gallery
or
Take an Image from Camera
Camera
crop the image
Then
Choose Origin Language
Choose Segmentation Mode
Process
Cancel
VOICE
Speak words or sentences to translate it
Voice
Select the origin language and translation language before speaking
Push the How Dictionary logo to speak up the words/ sentences.
Choose the most accurate result on the list
GENERAL
Guide
this button will exactly bring you here
About
will tell you more about the Apps
Translate
go translate the language by clicking this button
Rewrite
Having a recognition problem? Write your own here.
Speech
let the Apps speak up the sentences for you
Save
save the text to your device

Figure 18. Guide Page

About Page appears when user click on About Button at Main Menu Page. About page is intended to give such a short description of the application. It contains the name of the application, the logo, the version, the year of release, and the author. The About Page can be seen on Figure 19.

HOW Dictionary
House of Words
DICTIONARY
HD
Version 1.1.01
2017
Hutami Septiana Raswaty

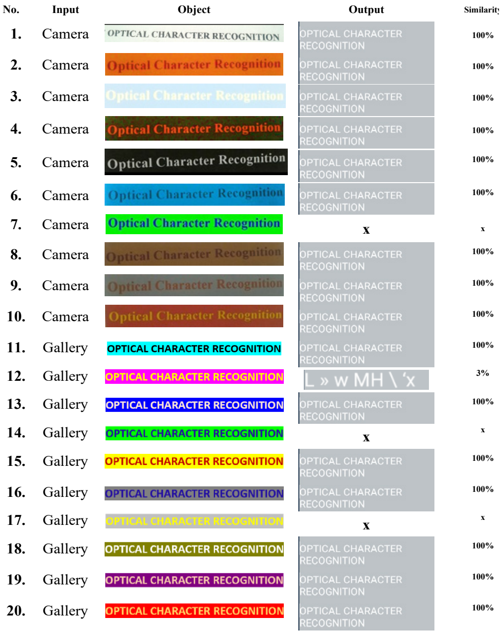
Figure 19. About Page

3.2 TESTING

Testing is divided into 2 types, which are OCR Testing and Voice Recognition Testing. The OCR testing is done by using 40 pictures and Voice Recognition testing is done by using 15 voices. The picture category itself is divided to 20 pictures with different variety of background and text colors, 10 pictures with different letter of alphabet (script), 10 pictures with different number of characters, and 15 voices with different sp oken languages and different number of spoken words.

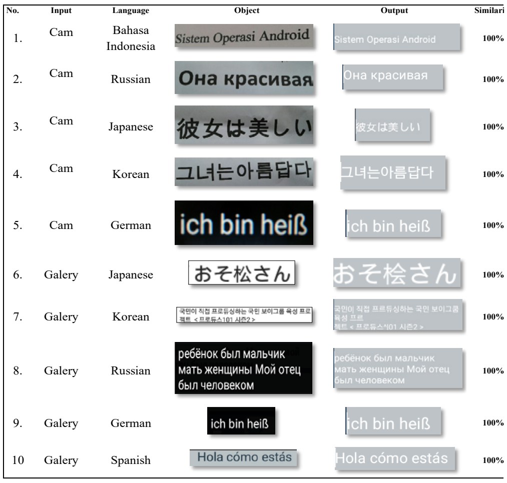
3.2.1 OCR Testing

OCR testing is divided into 3 types of image testing. Those are, different background and font colors which can be seen on Figure 20, different letter of alphabet (script) which can be seen on Figure 21, and different number of characters which can be seen on Figure 22.



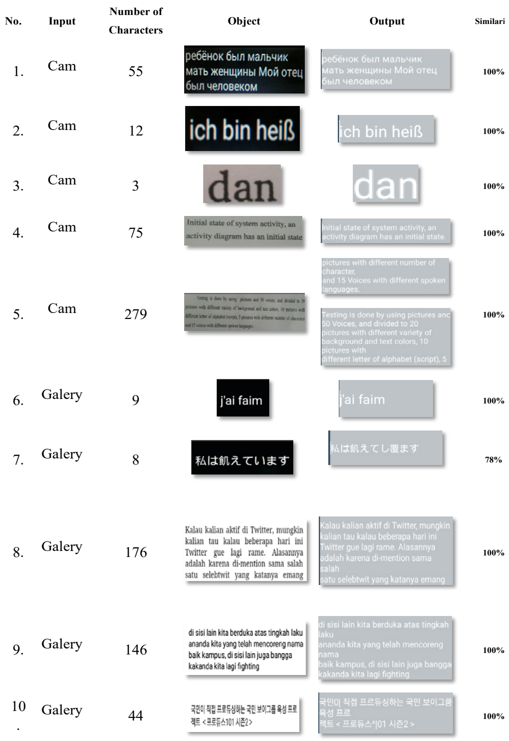
No.	Input	Object	Output	Similarity
1.	Camera	OPTICAL CHARACTER RECOGNITION	OPTICAL CHARACTER RECOGNITION	100%
2.	Camera	Optical Character Recognition	OPTICAL CHARACTER RECOGNITION	100%
3.	Camera	Optical Character Recognition	OPTICAL CHARACTER RECOGNITION	100%
4.	Camera	Optical Character Recognition	OPTICAL CHARACTER RECOGNITION	100%
5.	Camera	Optical Character Recognition	OPTICAL CHARACTER RECOGNITION	100%
6.	Camera	Optical Character Recognition	OPTICAL CHARACTER RECOGNITION	100%
7.	Camera	Optical Character Recognition	x	x
8.	Camera	Optical Character Recognition	OPTICAL CHARACTER RECOGNITION	100%
9.	Camera	Optical Character Recognition	OPTICAL CHARACTER RECOGNITION	100%
10.	Camera	Optical Character Recognition	OPTICAL CHARACTER RECOGNITION	100%
11.	Gallery	OPTICAL CHARACTER RECOGNITION	OPTICAL CHARACTER RECOGNITION	100%
12.	Gallery	OPTICAL CHARACTER RECOGNITION	L » w MH \ 'x	3%
13.	Gallery	OPTICAL CHARACTER RECOGNITION	OPTICAL CHARACTER RECOGNITION	100%
14.	Gallery	OPTICAL CHARACTER RECOGNITION	x	x
15.	Gallery	OPTICAL CHARACTER RECOGNITION	OPTICAL CHARACTER RECOGNITION	100%
16.	Gallery	OPTICAL CHARACTER RECOGNITION	OPTICAL CHARACTER RECOGNITION	100%
17.	Gallery	OPTICAL CHARACTER RECOGNITION	x	x
18.	Gallery	OPTICAL CHARACTER RECOGNITION	OPTICAL CHARACTER RECOGNITION	100%
19.	Gallery	OPTICAL CHARACTER RECOGNITION	OPTICAL CHARACTER RECOGNITION	100%
20.	Gallery	OPTICAL CHARACTER RECOGNITION	OPTICAL CHARACTER RECOGNITION	100%

Figure 20. OCR Testing Result Using Different Background and Font Color



No.	Input	Language	Object	Output	Similari
1.	Cam	Bahasa Indonesia	Sistem Operasi Android	Sistem Operasi Android	100%
2.	Cam	Russian	Она красивая	Она красивая	100%
3.	Cam	Japanese	彼女は美しい	彼女は美しい	100%
4.	Cam	Korean	그녀는아름답다	그녀는아름답다	100%
5.	Cam	German	ich bin heiß	ich bin heiß	100%
6.	Galery	Japanese	おそ松さん	おそ桧さん	100%
7.	Galery	Korean	국민이 직접 프로듀싱하는 국민 보이그룹 육성 프로젝트 <프로듀스101 시즌2>	국민이 직접 프르듀싱하는 국민 보이그룹 육성 프르 젝트 < 프로듀스시01 시즌2 >	100%
8.	Galery	Russian	ребёнок был мальчик мать женщины Мой отец был человеком	ребёнок был мальчик мать женщины Мой отец был человеком	100%
9.	Galery	German	ich bin heiß	ich bin heiß	100%
10	Galery	Spanish	Hola cómo estás	Hola cómo estás	100%

Figure 21. OCR Testing Result Using Different Letter of Alphabet (Script)

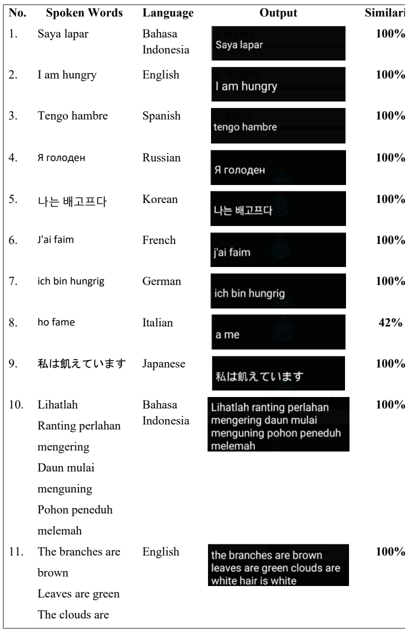


No.	Input	Number of Characters	Object	Output	Similari
1.	Cam	55	ребёнок был мальчик мать женщины Мой отец был человеком	ребёнок был мальчик мать женщины Мой отец был человеком	100%
2.	Cam	12	ich bin heiß	ich bin heiß	100%
3.	Cam	3	dan	dan	100%
4.	Cam	75	Initial state of system activity, an activity diagram has an initial state	Initial state of system activity, an activity diagram has an initial state.	100%
5.	Cam	279	Testing is done by using pictures and 50 voices, and divided to 20 pictures with different variety of background and text colors, 10 pictures with different letter of alphabet (script), 5 pictures with different number of characters and 15 voices with different spoken languages.	pictures with different number of character, and 15 Voices with different spoken languages. Testing is done by using pictures anc 50 Voices, and divided to 20 pictures with different variety of background and text colors, 10 pictures with different letter of alphabet (script), 5	100%
6.	Galery	9	j'ai faim	j'ai faim	100%
7.	Galery	8	私は飢えています	私は飢えてし覆ます	78%
8.	Galery	176	Kalau kalian aktif di Twitter, mungkin kalian tau kalau beberapa hari ini Twitter gue lagi rame. Alasannya adalah karena di-mention sama salah satu selebtwit yang katanya emang	Kalau kalian aktif di Twitter, mungkin kalian tau kalau beberapa hari ini Twitter gue lagi rame. Alasannya adalah karena di-mention sama salah satu selebtwit yang katanya emang	100%
9.	Galery	146	di sisi lain kita berduka atas tingkah laku ananda kita yang telah mencoreng nama baik kampus, di sisi lain juga bangga kakanda kita lagi fighting	di sisi lain kita berduka atas tingkah laku ananda kita yang telah mencoreng nama baik kampus, di sisi lain juga bangga kakanda kita lagi fighting	100%
10.	Galery	44	국민이 직접 프로듀싱하는 국민 보이그룹 육성 프로젝트 <프로듀스101 시즌2>	국민이 직접 프르듀싱하는 국민 보이그룹 육성 프르 젝트 < 프로듀스^|01 시즌2 >	100%

Figure 22. OCR Testing Result Using Different Number of Characters

3.2.2 Voice Recognition Testing

Google Text-To-Speech engine is used to speak up every tested word in this voice recognition testing. The result of this testing can be seen on Figure 23.



No.	Spoken Words	Language	Output	Similari
1.	Saya lapar	Bahasa Indonesia	Saya lapar	100%
2.	I am hungry	English	I am hungry	100%
3.	Tengo hambre	Spanish	tengo hambre	100%
4.	Я голоден	Russian	Я голоден	100%
5.	나는 배고프다	Korean	나는 배고프다	100%
6.	J'ai faim	French	j'ai faim	100%
7.	ich bin hungrig	German	ich bin hungrig	100%
8.	ho fame	Italian	a me	42%
9.	私は飢えています	Japanese	私は飢えています	100%
10.	Lihatlah Ranting perlahan mengering Daun mulai menguning Pohon peneduh melemah	Bahasa Indonesia	Lihatlah ranting perlahan mengering daun mulai menguning pohon peneduh melemah	100%
11.	The branches are brown Leaves are green The clouds are	English	the branches are brown leaves are green clouds are white hair is white	100%

Figure 23. Voice Recognition Test Result

3.3 TESTING RESULT

3.3.1 OCR Testing Result

The similarity percentage is obtained from Levenshtein Distance equation for calculating the similarity between actual text in the image and OCR testing result. All similarity percentage are averaged using Mean Arithmetic Formula. Mean Arithmetic, or simply called as Mean, of a set of N numbers $X_1, X_2, X_3, \dots\dots X_n$, can be declared by $\bar{X}$, as well as defined as [13]:

$$\bar{X} = \frac{X_1 + X_2 + X_3 \dots\dots + X_n}{N} = \frac{\sum_{j=1}^{N} X_j}{N} = \frac{\sum X}{N}$$

The result of OCR similarity test can be seen on Figure 24.

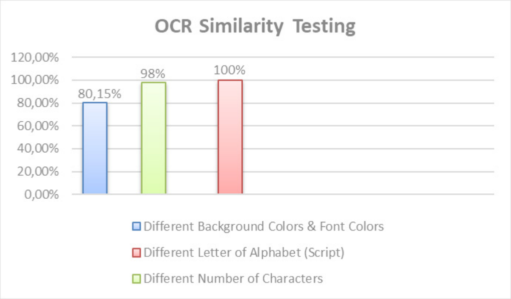


Figure 24. OCR Similarity Test

There are 3 of 20 background colors which cannot be read, two of them is because of the green background used in the image, and the other one is because of the yellow color of the font. On different background colors and font colors type of testing, it found out that the fuscia background used in one of the input images gave a minimum similarity, which is only 3%. The percentage of different number of character testing is on the second higher value on OCR similarity testing, it is 98%. According to the data test, there is one of the output images whose string character total is 8 and using Japanese language which has 78% similarity with its input. The last which hold the highest similarity percentage is similarity using nine different languages with 10 test images. All the percentage obtained in OCR similarity testing is calculated to find the grand percentage of the similarity using the same formula, Mean Arithmetic, and the result can be seen on Figure 25.

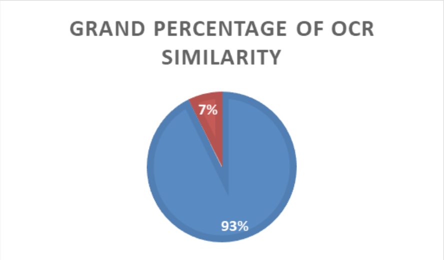


Figure 25. Grand Percentage of OCR Similarity Testing

3.3.1 Voice Recognition Test Result

According to the test which has been done, it is encountered that the percentage of voice recognition similarity is 95%. One of the tested voices using French is only has 42% out of 100%, and the other one using Russian has 90% out of 100%. Grand percentage of voice recognition test result can be seen on Figure 26.

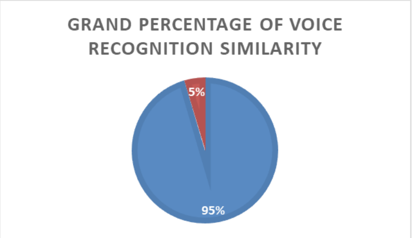


Figure 26. Grand Percentage of Voice Recognition Test

IV. CONCLUSION

4.1 CONLUSION

Based on the implementation and testing which have been done on this research, some conclusions can be drawn, include:

1. Implementation of optical character recognition and voice recognition of House of Words (HOW) Dictionary Application on Android platform have been successfuly done with average similarity between OCR input and output is 92,72% and average similarity of Voice Recognition is 95,46% which is encountered from Levenshtein Distance, and being averaged by Mean Arithmetic.
2. Tesseract can recognize the optical character with not dominant background colour, and Google Speech-To-Text and Google Text-To-Speech can recognize voice.

4.2 FUTURE WORKS

Suggestions for further development and refinement of Optical Character Recognition and Voice Recognition research are as follows:

1. Adding the latest and sophisticated image processing algorithm in Optical Character Recognition feature which can improve image input quality and reduce noises in order to read all colors and alphabet.
2. Adding voice filtering technology or algorithm to improve the voice control quality in voice recognition feature.
3. Add more languages option in voice recognition feature.
4. Create the offline one on other platform.

REFERENCES

[1] I. Oka and S. , Linguistik Umum, Jakarta: Proyek Pembinaan dan Peningkatan Mutu Tenaga Kependidikan: DIKTI, 1994.

[2] W.-s. W. Lian, Heuristic-Based OCR PostCorrection, Chapel Hill, 2009.

[3] H. Bunke and P. S. P. Wang, Handbook of Character Recognition and Document Image Analysis, USA: World Scientific Publishing Co. Pte. Ltd., 2000.

[4] L. Eikvil, Optical Character Recognition, Oslo, Norwegia: Norsk Regnesentral (Norwegian Computing Center) Blindern, 1993.

[5] A. Jain, A. Dubey, R. Gupta, N. Jain and P. Tripathi, “Fundamental Challanges to Mobile Based OCR,” Howarh: Tuhin Mallick, vol. 2, no. 8, 2013.

[6] A. Charisma, “Sistem Verifikasi Penutur Menggunakan Metoda Melfrequency Cepstral Coefficients-Vector Quantisation (Mfcc-Vq) Serta Sum Square Error (Sse) Dan Pengenalan Kata Menggunakan Metoda Logika Fuzzy” 2013.

[7] R. Smith, “An Overview of The Tesseract OCR Engine,” in Int. Conference on Document Analysis and Recognition (ICDAR), Washington DC, 2007.

[8] C. Patel, A. Patel and P. Dharmendra, “Optical Character Recognition by Open Source OCR Tool Tesseract: A Case Study,” International Journal of Computer Application, vol. 55, no. 10, 2012.

[9] R. B. Reddy and E. Mahender, “Speech To Text Conversion Using Android Platform.,” International Journal of Engineering Research and Application (IJERA), vol. III, no. 1, 2013.

[10] S. Iizuka, K. Tsujino, S. Oguri and H. Furukawa, “Speech Recognition Technology and Applications for Improving Terminal Functionality and Service Usability,” vol. 13, no. 4, p. NTT DOCOMO Technical Journal, 2012.

[11] A. Zenal, Buku Pintar Google, Jakarta Selatan: Media Kita, 2011.

[12] W. Komputer, The 40 Best Java Applications, Jakarta: PT. Elex Media Komputindo, 2010.

[13] M. R. Spiegel and L. J. Stephens, Statistics, Jakarta: Erlangga, 2004.