# Integration of Transformer Model Text Summarization and Text-to-Speech in Helping Document Understanding in the Bukudio Application

**[1*]Ivana Lucia Kharisma, [2]Kamdan, [3]Anggun Fergina, [4]Tofik Hidayat, [5]Moh. Abd. Aziz Hidayat,**

**[6]Muhamad Muslih** iD **, [7]Adhitia Erfina** iD
[1,2,3,4,5]Informatics Engineering, Nusa Putra University, Indonesia
[6,7]Information Systems, Nusa Putra University, Indonesia
E-mail: [1*]ivana.lucia@nusaputra.ac.id, [2]kamdan@nusaputra.ac.id, [3]anggun.fergina@nusaputra.ac.id,
[4]tofik.hidayat_ti20@nusaputra.ac.id, [5]aziz.hidayat_ti22@nusaputra.ac.id,
[6]muhamad.muslih@nusaputra.ac.id, [7]adhitia.erfina@nusaputra.ac.id
**\*Corresponding author**

***Abstract***

*The need for effective, accurate and precise understanding of information will provide optimization of the decision-making process, increase knowledge and quality of life. Understanding information in relation to the document summarization process, if done manually, sometimes takes quite a long time. Text summarization techniques which are useful as document summarizers have been developed and applied to various things such as summarizing important documents, news texts or customer feedback. In this article, text summarization using the text rank method and transformer modeling integrated with text to speech techniques is developed in the Bukudio application, which is an application that provides audio versions of book documents in the application database. Based on the test results, the evaluation process was carried out using the Rouge method and gave the best results in calculating the Rouge 1 overlap monogram resulting in 0.523 for the F1 Score value, 0.434 for the precision value and 0.659 for the recall value. This research will be developed using other methods so that not only files in PDF document format can be processed, but other EPUB (Electronic Publication) files.*

*Keywords: Text Summarization, Text To Speech, Bukudio, Document*

## 1. Introduction

The need for information motivates individuals and organizations in the process of searching, obtaining and understanding [1]. The importance of information lies in the ability of information to be used in various things such as decision making, increasing knowledge and improving the quality of life in various aspects of life. Information is available in massive quantities and can be accessed in various ways thanks to advances in information technology in the dissemination process [2]. Books and documents are available not only in hard copy form, but also in soft copy versions. Efforts to gain understanding of the information in documents sometimes become obstacles for information users [3]. The process of understanding and filtering the most important information becomes a challenge in itself, especially if a large amount of information or documents are available. With this background, research was conducted with the title "Integration of Text Summarization and Text-to-Speech in Helping Document

Understanding in the Bukudio Application" [4]. Effective understanding of information needs is the key to optimizing individual and organizational potential. The integration between text summarization and text to speech functions as an instrument, can help users obtain information quickly [5]. An automatic text summary provides a summary that combines all the important information from the source material and includes. Due to this, information is transferred quickly while maintaining the original purpose of the document [6].

Several previous studies on the use of text summarization and also text to speech became the basis for writing this research, as a source of data needed at the research method stage. Making automatic text summarization using extractive methods, abstraction and also a combination of the two is able to produce good document summarization [7]. The use of the NLTK python library to summarize documents in text that is input directly in the form of speech by utilizing the Google API facility, was proposed in research by Vinnarasu [8] . Research that aims to find out how abstract texts can be summarized that can be used in German uses multi-language models such as BERT for comparison of making summaries using other models such as BART [9].

In this research, the use of text summarization and text to speech was integrated and added to the Bukudio application, an application that provides audio versions of book files and provides a summary of the book contents contained in the application database. The Bukudio application was designed as an effort to provide inclusive application access and prioritize users who need assistance and have hearing problems [10].

## 2. Research Method

This research is the development process of the Bukudio application. The Bukudio application was designed with the aim of providing convenience for users, in this case focused on blind people [11]. The appearance of the Bukudio application is shown in Figure 1 below:
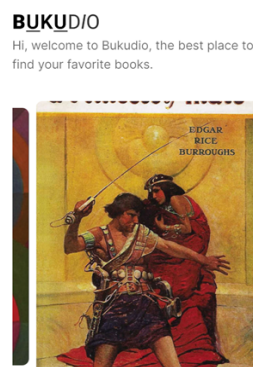


Figure 1. Bukudio Application Appearance

In Figure 1, we can see the front page display of the Bukudio application. The Bukudio application is equipped with a description or about feature, containing a description of the book, a summary feature, this feature provides a summary of the book, as well as audio. The audio menu is useful for changing text on book pages into audio files [12].

In this research, which aims to integrate text summarization and text to speech features to help understand documents designed by applying qualitative research methods [13]. The qualitative research method is shown in Figure 2 below:
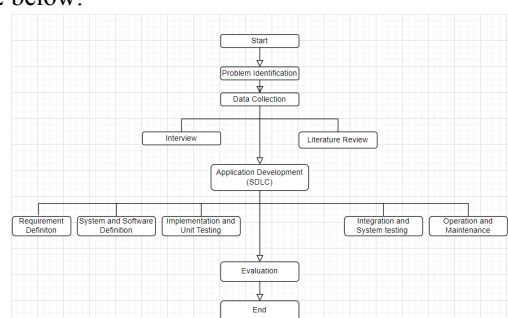


Figure 2. Research Methodology

In Figure 2, the research stages carried out in this writing consist of problem identification, where the process of formulating the problem is based on the data collection process by conducting interviews with Bukudio application users, followed by conducting literature studies related to text summarization and also text to speech [14]. The results of the interviews conducted showed that there is a need to understand not only the contents of books that have been uploaded to the Bukudio application, but there is also a need to understand other documents, especially documents that consist of quite a lot of words and sentences [15].

The software application development stage applies the software development life cycle waterfall method [16]. This method is a systematic and sequential information system development model [17], with the following stages: The PDF file is divided into several individual pages and the text of each page is extracted and examined, shown in Figure 3 below:
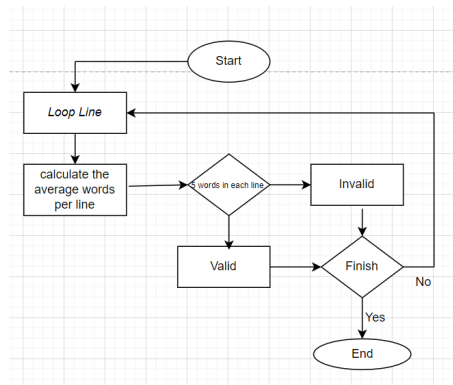


Figure 3. PDF to text stage

In Figure 3 above, each line is read and confirmed with the rule that if there are more than 5 words in one line excluding conjunctions, then the line is a valid word. The goal is to determine whether a line is a sentence or junk text such as headers and footers. If the statement is invalid then it is skipped and if it is valid then the next process continues [18].

The text summarization stage process is to create a summary of a document with a certain number of pages, in this case the first 10 pages. This process is carried out so that users can get a brief overview of the document without having to read all the pages simultaneously, shown in Figure 4 below:
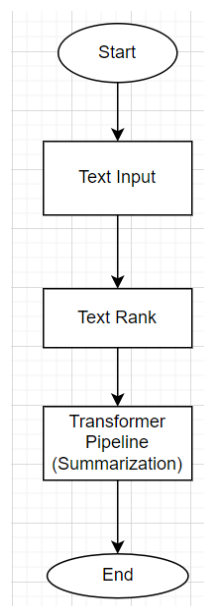


Figure 4. Text Summarization Stage

In Figure 4, we can see that the text summarization stage starts from the text input process, then the text will be processed using the text rank algorithm which is a graph-based compression algorithm according to Mihalcea, R.: Graph-based ranking algorithms for sentence extraction, applied to text summarization. Then the text will be processed using transformers. Transformer is designed to meet natural language processing machine learning model pipeline standards. This process applies the model, and makes predictions. Complete information about library functions can be seen in the documentation available at https://huggingface.co/transformers/ .

In text to speech, the extracted text and text summary are converted into mp3 audio files. The main process uses a Google text to speech converter and backup process. Text to Speech (TTS) is a technology that allows computers to convert written text into speech. There are two principles behind it, namely changing text into phonemes and changing phonemes into speech.

At the software design stage, the PHP, Node.JS and Python programming languages were used. PHP is used to perform simple operations and with databases such as lists of books and documents, PHP was chosen because it has a framework called Laravel which makes the development process more efficient, according to K. Wibowo [19]. Node.JS was selected and used for pipeline queue management and real-time processes such as summarization and text validation. Python was chosen to perform the summarization and text-to-speech processes. Python was chosen because the Transformer framework recommends using Python, and Python itself is suitable for machine learning and artificial intelligence. Based on the Python computer programming language, the NLTK library is a set of tools that can be used for natural language processing. With this library, researchers can carry out the entire process from corpus creation to research in one environment, so that there are no difficulties between different software and data conversion, and research work is further expanded [19].

The Bukudio application will be implemented, especially the text summarization and text to speech menus in documents that have been added to the application. Implementation is carried out on the admin side, by adding new documents and carrying out document summarization processes and adding text to speech. Meanwhile, on the user side, testing will be carried out by opening the document as well as a summary and audio version of the document.
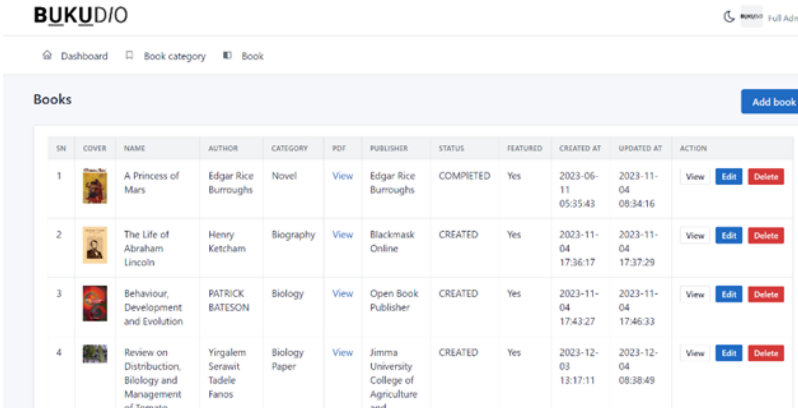
At the testing stage, the focus is on testing the results of document summarization. There are many methods that can be used to test document summarization results. In this research, we will test the Rouge method. A set of metrics called learning-oriented recall of Digest Evaluation (ROUGE) introduced by Lin (2004) provides a score based on the similarity of word sequences between human writing and model summaries and machine summaries [20].

## 3. Results and Analysis

Based on the results that have been developed in the Transformer Model Text Summarization and Text-to-Speech Integration process in Helping Document Understanding in the Bukudio Application, it consists of two pages for admins and application users. The implementation results that have been developed are as follows.

### 3.1. Admin Page

On the Admin menu, admins can add books or documents to add to the application. The admin menu display is shown in Figure 5 below:



Figure 5. Admin Page

In Figure 5, we can see the Dashboard, Book Category and Book menus, we can see that several books and documents have been added to the application. Admin can add books and documents in the available Add Book menu. For each book or document item, the admin can view the book description via the view menu, edit data and also delete data. The page display when adding a book is shown in Figure 6 below:



Figure 6. Add Book Page

In Figure 6 the admin can enter data on book title, category, book author, publisher, upload book files and cover images, manually enter summaries or summaries contained in documents and books and book explanations. When the Add menu is clicked, the audio and summary versions will be activated automatically with text summarization and text to speech. This aims to increase efficiency, avoiding additional configurations or operations.

The Book Category menu contains category data for books or documents added to the application. Shown in Figure 7 below:



Figure 7. Book Categories Page

Figure 7 shows that several types of books and documents have been classified into several categories such as novel, book, biography. Existing data can be edited and deleted by the admin. To add category types, you can use the Add new menu, as shown in Figure 8 below:

Figure 8. Add New Book Category Page

In Figure 8 it is shown that the admin can add book category types, category names, types, book descriptions and category thumbnails.

### 3.2. User Page

In this menu, the user will see a display of documents that have been added to the Bukudio application database, an example of which is a Biology research paper document with the title Review on Distribution, Biology and Management of Tomato Powdery Mildew (Oldium Nelycopersici). This menu is shown in Figure 9 below:
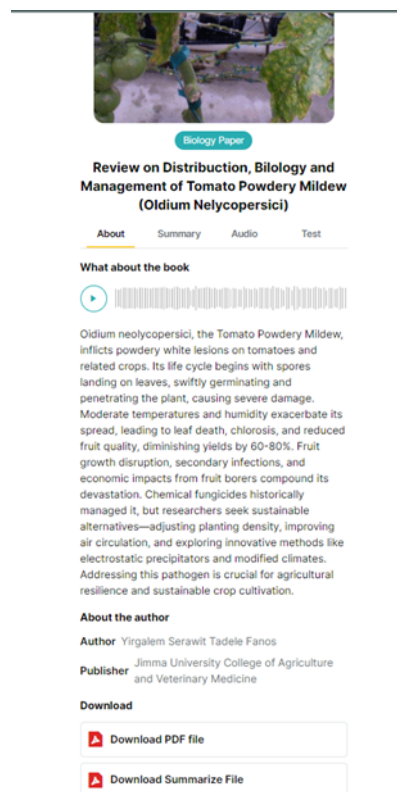


Figure 9. User Page

On the User page, we can see the About, Summary, Audio and Test menus. In the About menu, users can select the Audio version by clicking the play menu. The document summary file resulting from Transformer Text Summarization modeling can be downloaded on this page. Users can also download pdf files of this document.

The Summary menu display, which is a summary of the document, is shown in Figure 10 below:



Figure 10. Summary Page

On the Summary page, users can view a document summary as well as listen to an audio version of the summary.

To listen to all document files in audio version, users can select the Audio menu, shown in Figure 11 below:
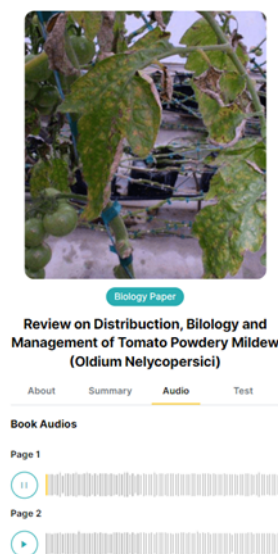


Figure 11. Audio Page

### 3.3. Testing Page

As explained in the previous chapter, testing will prioritize functional text summarization using the ROUGE calculation method. From existing document examples, users can see the text summarization evaluation in the Test menu, which is shown in Figure 12 below:
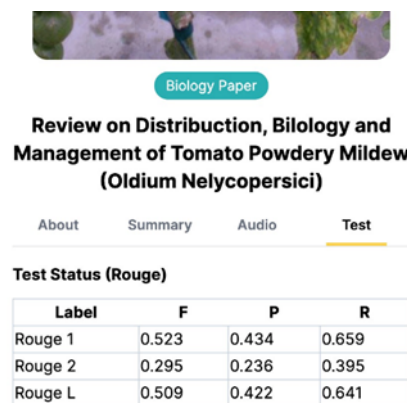
Figure 12 Test display

From the evaluation results using the Rouge method, it can be concluded that the results of text summarization in the Bukudio application are as follows:

The results of Label Rouge 1 (overlap monogram) compare the monogram (one word) in the resulting summary with the monogram in the reference summary, the Bukudio application gives a value of 0.523 for the F1 Score which means the mean value of precision and recall, 0.434 precision which means the proportion of n-grams which the model predicted correctly, and 0.659 recall value means the proportion of reference n-grams predicted by the model. The results of Label Rouge 2 (overlap bigram) compare the bigram (two consecutive words) in the resulting summary with the bigram in the reference summary. The Bukudio application gives a value of 0.295 for the F1 Score which means the mean value of precision and recall, 0.236 precision which means the proportion of n- grams that the model predicted correctly, and 0.395 recall value means the proportion of reference n-grams predicted by the model. The results of the Rouge L (longest common subsequence) label compare the longest series of common words in the summary produced in the reference summary. The Bukudio application gives a value of 0.509 for the F1 Score which means the mean value of precision and recall, 0.434 precision which means the proportion of predicted n-grams. correct model, and 0.641 recall value means the proportion of reference n-grams predicted by the model.

## 4. Conclusion

The ability to understand information effectively, quickly and accurately makes it easy for Bukudio Application users. The integration between Transformer modeling, especially in text summarization tasks, by combining the text rank method has been able to summarize documents that have quite a lot of pages so that it can reduce the time needed if you want to summarize documents manually. The added text to speech method can also make things easier for users, especially for those with visual impairments and for users who are more suited to learning methods via audio media. This application has been tested using the Rouge method and gives the best results in calculating Rouge 1 monogram overlap resulting in 0.523 for the F1 Score value, 0.434 for the precision value and 0.659 for the recall value. This research will be developed using other methods so that not only files in PDF document format can be processed, but other EPUB (Electronic Publication) files.

## References

[1] Dewi KE, Widiastuti NI. Automatic Summarization of Indonesian Texts Using a Hybrid Approach. Jurnal Teknologi Informasi Dan Pendidikan 2021;15. https://doi.org/10.24036/tip.v15i1.

[2] L. K. Choi, K. B. Rii, and H. W. Park, "K-Means and J48 Algorithms to Categorize Student Research Abstracts," *Int. J. Cyber IT Serv. Manag.*, vol. 3, no. 1, pp. 61–64, 2023.

[3] Q. Aini, I. Sembiring, A. Setiawan, I. Setiawan, and U. Rahardja, "Perceived Accuracy and User Behavior: Exploring the Impact of AI-Based Air Quality Detection Application (AIKU)," *Indones. J. Appl. Res.*, vol. 4, no. 3, pp. 209–218, 2023.

[4] L. Philip and T. Pradiani, "Influence Brand Experience, Viral Marketing and Brand Image to Brand Loyalty to Service Users Streaming Spotify in Indonesia," *ADI J. Recent Innov.*, vol. 5, no. 2, pp. 127–135, 2024.

[5] Sasmito GW, Informatika JT, Bersama H, Mataram J, 09 N, Lor P. Penerapan Metode Waterfall Pada Desain Sistem Informasi Geografis Industri Kabupaten Tegal 2017;2.

[6] B. Bhima, A. R. A. Zahra, T. Nurtino, and M. Z. Firli, "Enhancing Organizational Efficiency through the Integration of Artificial Intelligence in Management Information Systems," *APTISI Trans. Manag.*, vol. 7, no. 3, pp. 282–289, 2023.

[7] Arrizqi N', Santoso I, Yosua D, Soetrisno AA. IMPLEMENTASI GOOGLE TEXT TO SPEECH PADA APLIKASI PENDETEKSI UANG BERBASIS ANDROID. vol. 10. 2021.

[8] Y. N. Qintharah and F. L. Utami, "Determinants on Environmental Disclosure Moderating by Integrated Corporate Governance," *Aptisi Trans. Technopreneursh.*, vol. 5, no. 1Sp, pp. 22–37, 2023.

[9] Wang M, Hu F. The application of nltk library for python natural language processing in corpus research. Theory and Practice in Language Studies 2021;11:1041–9. https://doi.org/10.17507/tpls.1109.09.

[10] Y. I. Maulana and I. Fajar, "Analysis of Cyber Diplomacy and its Challenges for the Digital Era Community," *IAIC Trans. Sustain. Digit. Innov.*, vol. 4, no. 2, pp. 169–177, 2023.

[11] Vinnarasu A, Jose D V. Speech to text conversion and summarization for effective understanding and documentation. International Journal of Electrical and Computer Engineering 2019;9:3642–8. https://doi.org/10.11591/ijece.v9i5.pp3642-3648.

[12] I. S. Fadli, N. P. Lestari, and A. A. Putri, "Implementation of White Hat SEO-Based Digital Incubator Platform," *Int. Trans. Educ. Technol.*, vol. 1, no. 2, pp. 122–130, 2023.

[13] Wolf T, Debut L, Sanh V, Chaumond J, Delangue C, Moi A, et al. HuggingFace's Transformers: State-of-the-art Natural Language Processing 2019.

[14] Endra RY, Aprilinda Y, Dharmawan YY, Ramadhan W. Analisis Perbandingan Bahasa Pemrograman PHP Laravel dengan PHP Native pada Pengembangan Website. EXPERT: Jurnal Manajemen Sistem Informasi Dan Teknologi 2021;11:48. https://doi.org/10.36448/expert.v11i1.2012.

[15] Saziyabegum S. REVIEW ON TEXT SUMMARIZATION EVALUATION METHODS. / Indian Journal of Computer Science and Engineering (IJCSE) n.d.;8.

[16] Nina Adlini M, Hanifa Dinda A, Yulinda S, Chotimah O, Julia Merliyana S. METODE PENELITIAN KUALITATIF STUDI PUSTAKA. vol. 6. 2022.

[17] Alcantara THM, Krütli D, Ravada R, Hanne T. Multilingual Text Summarization for German Texts Using Transformer Models. Information (Switzerland) 2023;14. https://doi.org/10.3390/info14060303.

[18] D. S. S. Wuisan, R. A. Sunardjo, Q. Aini, N. A. Yusuf, and U. Rahardja, "Integrating Artificial Intelligence in Human Resource Management: A SmartPLS Approach for Entrepreneurial Success," *Aptisi Trans. Technopreneursh.*, vol. 5, no. 3, pp. 334–345, 2023.

[19] O. Candra et al., "Energy simulation and parametric analysis of water cooled thermal photovoltaic systems: energy and exergy analysis of photovoltaic systems," *Sustainability*, vol. 14, no. 22, p. 15074, 2022.

[20] P. A. Sunarya, "The Impact of Gamification on IDU (ILearning Instruction) in Expanding Understudy Learning Inspiration," *Int. Trans. Educ. Technol.*, vol. 1, no. 1, pp. 59–67, 2022.