

Integration of Natural Language Processing in a Web-Based Translanguaging System for Arabic-Indonesian Language Learning

Danang Bagus Reknadi¹, Mohammad Mansyur Abidin², Achmad Firman Choiri³

Department of Informatics Engineering, Faculty of Science and Technology, Universitas Islam Lamongan, Indonesia¹

Department of Informatics Engineering, Faculty of Engineering, Universitas Islam Darul Ulum, Indonesia²

Department of Information Technology, Institut Teknologi dan Bisnis Widya Gama Lumajang, Indonesia³

Corresponding Author: Danang Bagus Reknadi (danz.0907@unisla.ac.id)

ARTICLE INFO

Date of entry:
10 September 2025
Revision Date:
25 October 2025
Date Received:
31 October 2025

ABSTRACT

Arabic–Indonesian language learners often face challenges in understanding the contextual meaning of texts due to differences in morphological and syntactic structures between the two languages. To address this, this study proposes the development of a web-based translanguaging system integrated with Natural Language Processing (NLP) to help users understand and translate texts more meaningfully. This system was developed using the Waterfall model with stages of requirements analysis, design, implementation, testing, and maintenance. The implemented NLP module includes tokenization, part-of-speech tagging, and sentence structure analysis to produce translations that consider context, not just literal word equivalents. The implementation results show that the system is able to improve user comprehension of Arabic–Indonesian texts with a simple and accessible interface. Furthermore, the translation history feature supports continuous self-learning. Although the system still has limitations in handling idiomatic text and complex sentence structures, the NLP integration has proven effective in improving the quality of translanguaging. This research contributes to the development of bilingual learning technology and can be further developed using deep learning models such as BERT or mBERT to improve semantic and contextual accuracy.

Keywords: Translanguaging, Natural Language Processing (NLP), Web-Based Learning, Arabic–Indonesian, Bilingual Education.



Cite this as : Reknadi, D. B., Abidin, M. M., & Choiri, A. F. (2025). Integration of Natural Language Processing in a Web-Based Translanguaging System for Arabic-Indonesian Language Learning. *Journal of Informatics Development*, 4(1), 9–16. <https://doi.org/10.30741/jid.v4i1.1753>

INTRODUCTION

The development of information technology has brought significant changes to the world of education, particularly in language learning. The translanguaging approach has begun to be widely implemented as a learning strategy that leverages students' bilingual abilities to gain a deeper understanding of the material (Wulandari et al., 2025). In the context of Arabic-Indonesian language learning, translanguaging plays a crucial role in helping students bridge the structural and semantic differences between languages (Muchlis, 2024; Reknadi et al., 2025). However, digital applications of translanguaging are still limited, particularly in the form of web-based systems capable of automatic language processing.

Natural Language Processing (NLP) technology holds great potential in supporting translanguaging because it can computationally analyze, understand, and generate natural language. By integrating NLP into a translanguaging system, the learning process can become more interactive, adaptive, and contextual. Therefore, research focused on developing a web-based translanguaging system with NLP technology for Arabic–Indonesian is needed to improve the effectiveness and efficiency of language learning.

Although NLP technology has been widely used for global languages such as English, its application in the context of Arabic and Indonesian still faces several challenges.

1. The complex linguistic structure of Arabic, such as root-based morphology, makes language processing challenging.
2. The limited availability of parallel Arabic–Indonesian corpus datasets makes it difficult for automatic translation systems to achieve accurate results.
3. There is limited research integrating translanguaging and NLP in a single web-based learning system.
4. There is a lack of interactive platforms that allow students to practice translanguaging automatically and contextually.

Previous research has largely focused on the development of automatic translation systems (machine translation) or web-based language learning applications. However, most have not combined the concept of translanguaging—which emphasizes the dynamic use of two languages—with the analytical capabilities of NLP.

Furthermore, NLP research for Arabic–Indonesian remains scarce, especially in the context of technology-based language learning. This creates a research gap in the development of systems that not only translate but also facilitate a two-way learning process between Arabic and Indonesian.

This research is important because it addresses the need for innovative digital learning media, especially for educational institutions teaching Arabic. With the increasing use of web-based technology, NLP-powered translanguaging systems can provide a more interactive, flexible, and accessible learning experience. Furthermore, this research contributes to the development of NLP resources for Arabic–Indonesian, which are currently very limited compared to other languages such as English or Mandarin.

The novelty of this research lies in the integration of NLP technology with the concept of translanguaging into a single web-based system for Arabic-Indonesian language learning. This system not only translates text but also understands the linguistic context of both languages, provides interactive explanations, and facilitates users in the natural bilingual learning process. Thus, this research combines computational linguistics, language education, and web-based intelligent systems into one applicable innovation.

This research aims to design and develop a web-based translanguaging system that utilizes Natural Language Processing (NLP) technology to support Arabic-Indonesian language learning. Through this system, users are expected to learn and understand both languages interactively with the help of natural language processing that can recognize the structure, context, and meaning of the given text. Furthermore, this research also aims to implement an NLP model that matches the linguistic characteristics of Arabic and Indonesian, thus producing relevant and accurate output in the learning context. The final results of this research are expected to make a significant contribution to the development of AI-based learning technology and strengthen the NLP research ecosystem in the local language domain, particularly for the Arabic-Indonesian language pair.

Academically, this research contributes to the development of science in the fields of Natural Language Processing, translanguaging, and web-based learning systems, especially for rarely explored languages such as Arabic and Indonesian. From a practical perspective, the developed system can be utilized by teachers and students as an innovative learning medium that enables the

bilingual learning process to be more effective and adaptive to the user's context. Furthermore, socially, this research is expected to expand public access to language learning through interactive digital platforms, thereby supporting the transformation of technology-based education in the digital era. Thus, this research not only focuses on technological development, but also has a positive impact on improving the quality of language learning in Indonesia.

Translanguaging is a pedagogical approach that utilizes all of a student's language abilities to understand and construct meaning in the learning process (Afrita, 2023; Reknadi et al., 2022). In a bilingual context, translanguaging enables students to use two languages dynamically and complementary to each other, rather than simply translating them word for word (Rakhmat & Qohar, 2024). This approach, as explained by (Orús et al., 2014) can strengthen conceptual understanding and improve cross-language communication skills. In Arabic-Indonesian language learning, translanguaging is highly relevant because the two languages have fundamentally different linguistic systems, both in terms of morphology and syntax. To support its digital implementation, technology is needed that can automatically understand language structure and meaning, one of which is through the application of Natural Language Processing (NLP). NLP is a branch of Artificial Intelligence (AI) that enables computers to process and understand human language in a meaningful way (Fanni et al., 2023). Some of the main components of NLP include tokenization, stemming, part-of-speech tagging, named entity recognition, and machine translation (Chiche & Yitagesu, 2022).

This technology has been widely used in various applications such as chatbots, automatic translation systems, and language learning platforms (Adamopoulou & Moussiades, 2020). However, applying NLP to Arabic and Indonesian presents its own challenges. Arabic is known for its complex root-based morphology (Abu-Rabia, 2002), while Indonesian is agglutinative with a simpler structure but a high degree of ambiguity (Trismanto, 2018). This challenge necessitates integrating NLP into both languages, requiring specific and context-sensitive linguistic models. Previous research has largely focused on a single language, such as NLP systems for Arabic or Indonesian separately, without integrating both in a bilingual context. Therefore, developing a translanguaging system that integrates NLP for Arabic and Indonesian is a potential new area of research. Through a web-based approach, translanguaging systems can be designed to be interactive, accessible, and support more adaptive learning. Web-based technology allows direct integration between NLP models and user interfaces, allowing students to practice, translate, and understand the linguistic context of both languages directly.

Several studies, such as those by (Baazeem et al., 2021) and (Mustaqim et al., 2023), have demonstrated the effectiveness of NLP in language learning, but none have explicitly incorporated translanguaging aspects. Thus, there is a research gap in the development of learning systems that integrate translanguaging approaches with NLP technology to support the Arabic-Indonesian language learning process automatically and contextually. This research aims to address this gap by developing a web-based translanguaging system capable of intelligently understanding, translating, and displaying cross-language meaning through the use of NLP technology.

METHOD

This study uses the Waterfall system development model as the primary approach in designing and building a web-based translanguaging application (Petersen et al., 2009). As in Figure 1, the Waterfall Model was chosen because it has systematic and structured stages, making it easier for researchers to manage each development process sequentially and in a well-documented manner.

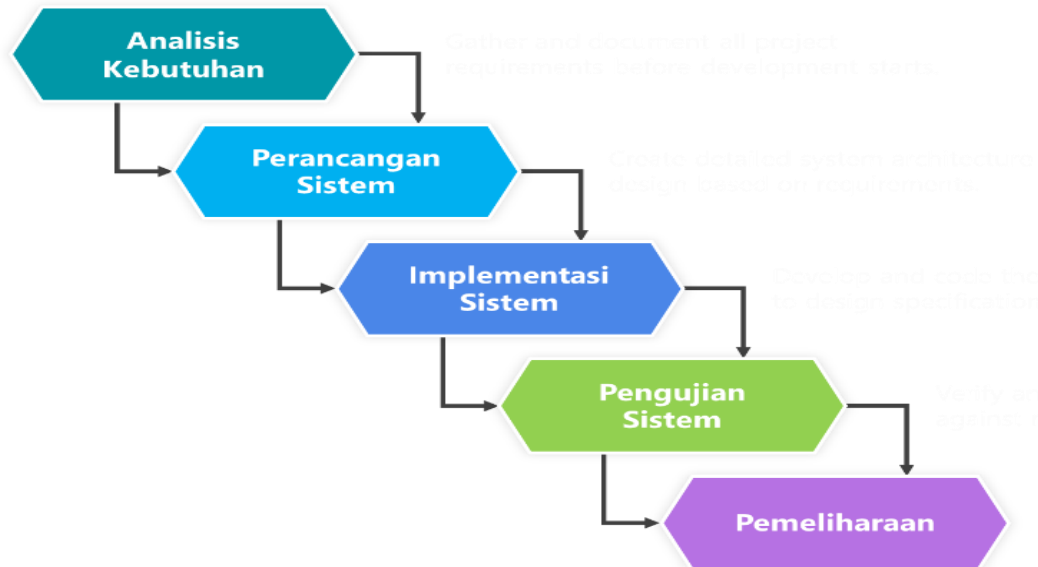


Figure 1. Model Waterfall

The following explains each stage of the waterfall model in developing this web-based translanguaging application:

1. Requirements Analysis, which is the process of identifying user and system needs, including linguistic analysis to determine relevant translanguaging features and the natural language processing requirements to be used.
2. System Design, which includes the creation of a user interface design, system architecture, database design, and an integration workflow between the NLP module and the web system.
3. Implementation (Coding), where the system design is translated into program code using a web programming language and NLP libraries to handle the translation and text analysis processes.
4. Testing to ensure each component functions as required, both in terms of functionality and the accuracy of the translanguaging results.
5. Maintenance, which aims to fix bugs, improve system performance, and adapt features to future user needs. With this approach, the development process is expected to produce a web-based translanguaging system that is effective, tested, and capable of optimally supporting the improvement of Arabic-Indonesian literacy.

This research was conducted through several stages referring to the Waterfall system development model. The first stage is Needs Analysis, which is the process of collecting data and identifying user needs, both in terms of application functionality and linguistic needs in the context of Arabic-Indonesian translanguaging. At this stage, a study of relevant natural language processing (NLP) technologies was also conducted to understand the structure and context of the two languages. The second stage is System Design, which includes the creation of a system architecture design, a web-based user interface design, and an integration design between the NLP module and the database and application server. In addition, a translanguaging algorithm was also designed to dynamically translate, adapt, and combine contexts between languages. The third stage is Implementation, which is the process of translating the system design into program code using web programming languages such as Python and the Flask framework, as well as the application of NLP libraries to support syntactic and semantic analysis of text. The fourth stage is Testing, where the system is tested to ensure functionality runs according to predetermined requirements, including testing translanguaging accuracy, system performance, and user interface appearance. The final stage is Maintenance, which includes evaluation activities, bug fixes, and advanced feature development to improve the system's performance and accuracy in processing bilingual text. By following these

stages, the resulting web-based translanguaging system is expected to be highly reliable and contribute to the development of Arabic-Indonesian language learning technology.

The data design in this study focuses on supporting a web-based translanguaging process with the integration of Natural Language Processing (NLP). The data used consists of a corpus of bilingual Arabic-Indonesian texts obtained from various sources, such as educational articles, religious texts, and Arabic language materials translated into Indonesian. This data is used to train and test the NLP model, ensuring its ability to recognize grammatical structures and meanings across languages contextually. Furthermore, user data is designed to store translation results, learning activities, and interactions between users and the system. The system's database structure is designed with several main tables as follows:

Table 1. database design

No	Tabel	Description	Main Attributes
1	user	Storing the identity of the translanguaging system user	id_user, nama, email, password
2	text_input	Stores input text submitted by the user for processing.	id_input, id_user, teks_asal, bahasa_asal, waktu_input
3	translation_result	Saving the results of the translanguaging process from the NLP module	id_hasil, id_input, teks_hasil, bahasa_tujuan, tingkat_akurasi
4	log_activity	Record user and system activity for evaluation purposes	id_log, id_user, aktivitas, waktu_aktivitas

This study used several supporting materials and tools to build and test a web-based translanguaging system. The primary material used was a bilingual Arabic–Indonesian text dataset, which served as the data source for training the NLP model and testing the translanguaging system. The explanation of the hardware and software required is as in the BLA table below:

Table 2. Supporting Materials and Devices

No	Category	Description	Function
1	Hardware	PC/Laptop (Intel Core i5, RAM 8 GB, SSD 512 GB)	Used for development, model training, and system testing.
2	Software	Python 3.x	The primary programming language for building translanguaging systems and integrating NLP modules.
3		Flask Framework	A Python-based framework for web application development.
4		MySQL Database	Stores user data, input text, and translanguaging results.
5		Library NLP (NLTK, spaCy, Transformers)	For text processing, syntactic analysis, and NLP model deployment.
6		Visual Studio Code	The primary code editor for system development.
7		Postman	A tool for testing APIs and endpoints in web systems.
8	Supporting Devices	Browser Web (Firefox)	For running and testing application interfaces.

With this data design, the system is expected to be able to manage the translanguaging process efficiently, maintain data integrity, and support the development of adaptive learning features for users.

RESULTS AND DISCUSSION

Based on implementation and testing results, this web-based language translation system has proven effective in helping users understand the meaning of Arabic texts from Indonesian to Arabic and vice versa, contextually. The Natural Language Processing (NLP) integration module enables the system to analyze language patterns more deeply than conventional dictionary-based translation. Using the NLP approach, the system not only translates word-by-word but also considers the relationships between phrases, sentence context, and Arabic morphological structure.

Figure 2 below shows the main interface of the developed web-based translanguaging application. This application is designed with a simple, responsive interface that is easy to use for both Arabic language students and teachers. On the main page, users can enter text in Arabic or Indonesian, and the system will automatically detect the language input and generate translanguaging results with the appropriate contextual meaning. This process is carried out with the help of an NLP module that recognizes words, phrases, and the semantic meaning of the input text.

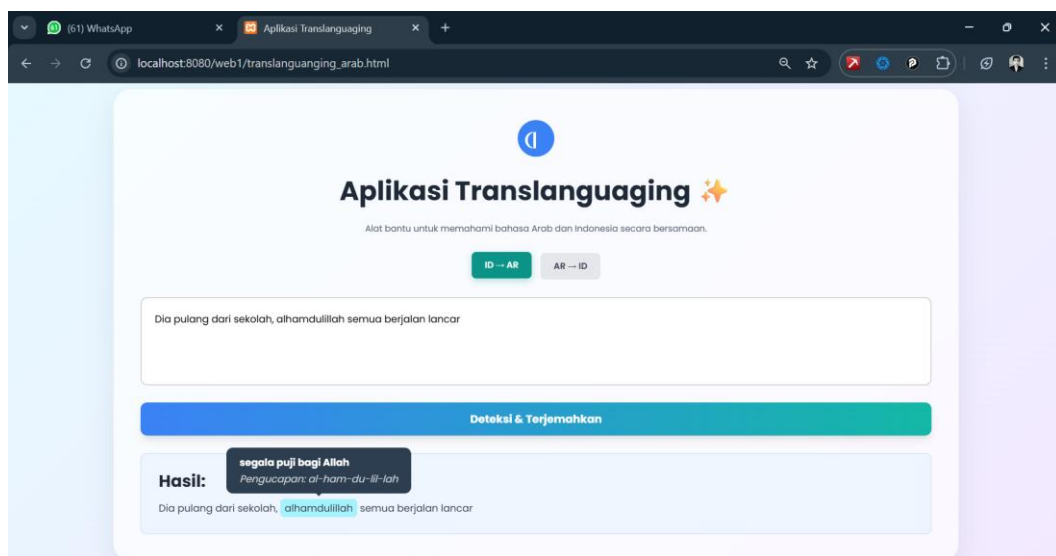


Figure 2. Translanguaging Application View

In this example, the user enters Indonesian text containing the word "alhamdulillah." The system automatically recognizes the Arabic context within the sentence and displays the translanguaging result, an explanation of the meaning of "all praise be to Allah," complete with the transliteration "al-ham-du li-l-lāh." This feature is one of the system's added values, as it not only translates literally but also provides contextual understanding between languages.

In terms of usability, the simple web interface makes the system easily accessible to a wide range of users, including teachers and students looking to improve their bilingual skills. Furthermore, the translation history storage feature allows users to engage in continuous, independent learning. Although test results show quite good performance, several challenges exist, such as the system's difficulty handling idiomatic text and complex sentence structures. This is due to the limitations of the bilingual corpus used and the lack of implementation of a deep learning model for advanced semantic understanding. Therefore, further development can be directed at integration transformer-based models such as BERT or mBERT to improve translation quality and expand the scope of text domains. Overall, this research demonstrates that the application of NLP technology in translation systems can be an innovative solution to support more interactive, contextual, and adaptive Arabic-Indonesian language learning in the digital age.

CONCLUSION

Based on the implementation results and discussions, it can be concluded that the developed web-based translanguaging system is capable of assisting users in understanding Arabic and Indonesian texts contextually. The integration of Natural Language Processing (NLP) technology makes the translation process more accurate because the system can recognize sentence structure and phrase meaning, rather than simply translating word-for-word.

This system also has a simple and interactive interface, making it easy for teachers and students to use in bilingual learning activities. The translation history storage feature provides added value in supporting continuous and independent learning processes.

However, the system still has limitations in translating idiomatic text and complex sentences due to limited corpus data. For further development, the implementation of transformer-based deep learning models such as BERT or mBERT is recommended to improve semantic capabilities and translation quality.

Based on the research results and implementation of this web-based translanguaging system, several recommendations can be used as a reference for future development and application, namely:

1. Language Corpus Enhancement: Expanding the Arabic–Indonesian bilingual dataset is necessary so that the system can handle a wider variety of contexts, including idioms, figurative expressions, and complex sentence structures.
2. Deep Learning Model Integration: Using transformer models such as BERT, mBERT, or MarBERT can improve semantic understanding and translation accuracy, especially for texts with deeper context.
3. Adaptive Learning Feature Development: The system can be developed into an interactive learning platform with practice features, automatic evaluation, and user feedback to make language learning more personalized and effective.
4. Interface Optimization and Responsiveness: Optimizing the display is recommended so that the system can run smoothly on various devices, including smartphones and tablets, to increase user accessibility.
5. Collaboration with Linguists and Educators: Further development should involve Arabic language teachers and linguists to ensure that translation results and translanguaging applications align with linguistic rules and learning objectives.

REFERENCES

- Abu–Rabia, S. (2002). Reading in a root–based–morphology language: the case of Arabic. *Journal of Research in Reading*, 25(3), 299–309.
- Adamopoulou, E., & Moussiades, L. (2020). An overview of chatbot technology. *IFIP International Conference on Artificial Intelligence Applications and Innovations*, 373–383.
- Afrita, J. (2023). Peran artificial intelligence dalam meningkatkan efisiensi dan efektivitas sistem pendidikan. *COMSERVA: Jurnal Penelitian Dan Pengabdian Masyarakat*, 2(12), 3181–3187.
- Baazeem, I., Al-Khalifa, H., & Al-Salman, A. (2021). Cognitively driven arabic text readability assessment using eye-tracking. *Applied Sciences*, 11(18), 8607.
- Chiche, A., & Yitagesu, B. (2022). Part of speech tagging: a systematic review of deep learning and machine learning approaches. *Journal of Big Data*, 9(1), 10.
- Fanni, S. C., Febi, M., Aghakhanyan, G., & Neri, E. (2023). Natural language processing. In *Introduction to artificial intelligence* (pp. 87–99). Springer.
- Muchlis, M. (2024). *Translanguaging as a strategy in fostering multilingual students' comprehension at Mambaus Sholihin 2 High School Blitar*. Universitas Islam Negeri Maulana Malik Ibrahim.
- Mustaqim, M., Gunawan, A., Pratama, Y. B., & Zaliman, I. (2023). Pengembangan Chatbot Layanan

- Publik Menggunakan Machine Learning Dan Natural Language Processing. *Journal of Information Technology and Society*, 1(1), 1–4.
- Orús, R., Wei, T.-C., Buerchaper, O., & García-Sacaz, A. (2014). Topological transitions from multipartite entanglement with tensor networks: a procedure for sharper and faster characterization. *Physical Review Letters*, 113(25), 257202.
- Petersen, K., Wohlin, C., & Baca, D. (2009). The waterfall model in large-scale development. *International Conference on Product-Focused Software Process Improvement*, 386–400.
- Rakhmat, M., & Qohar, H. A. (2024). Pengaruh Bilingualisme dalam Bahasa Indonesia. *Ranah Research: Journal of Multidisciplinary Research and Development*, 6(6), 3057–3072.
- Reknadi, D. B., Kristian, Y., & Harianto, R. A. (2022). Classification of Criticisms and Suggestions on Public Services at RSI Nashrul Ummah Lamongan Using K-Competitive Autoencoder. *Proceeding International Conference on Environment Health, Socioeconomic and Technology*, 1, 151–161.
- Reknadi, D. B., Rohman, M. G., & Utomo, A. F. L. (2025). Adaptation of Contrastive Learning and Augmentation for Indonesian Product Review Classification on Unbalanced Data Using Deep Learning and NLP. *Generation Journal*, 9(2), 105–116.
- Trismanto, T. (2018). Ambiguitas dalam bahasa Indonesia. *Bangun Rekaprima*, 4(1), 42–48.
- Wulandari, D. A., Widagdo, A., Shafira, H., Maulida, A., Wardani, R. W., Sarnita, S., & Rahmawati, A. (2025). Analisis Penerapan Strategi Pembelajaran Bilingual pada Video Praktik Pembelajaran di Sekolah Dasar. *Jurnal Inovasi Penelitian Ilmu Pendidikan Indonesia*, 177–183.