Indonesian Journal of Electrical Engineering and Computer Science
Vol. 30, No. 2, May 2023, pp. 1214~1223
ISSN: 2502-4752, DOI: 10.11591/ijeecs.v30.i2.pp1214-1223

Jahai language repository: a mobile application

Nurazzah Abd Rahman[1], Masurah Mohamad[2], Itaza Afiani Mohtar[2], Saidi Adnan Md Nor[3]
[1]School of Computing Sciences, College of Computing, Informatics and Media, Universiti Teknologi MARA, Shah Alam, Malaysia
[2]Computing Sciences Studies, College of Computing, Informatics and Media, Universiti Teknologi MARA Perak Branch, Tapah Road, Malaysia
[3]Academy of Contemporary Islamic Studies, Universiti Teknologi MARA Perak Branch, Tapah Road, Malaysia

Article Info

Article history:

Received Sep 29, 2022
Revised Dec 21, 2022
Accepted Dec 26, 2022

Keywords:

Jahai
Language
Malay
Mobile application
Translation

ABSTRACT

The Jahai language is facing extinction primarily because it is among the least spoken language among minority groups in Malaysia. This is due to lesser speakers and the shift to using a more dominant language. The Jahai tribe is one of the community groups living in the Royal Belum Perak State Park. People who need assistance from Jahai people often face difficulties in communicating with them due to the language barrier. Therefore, a mobile translation system was developed to preserve the language. The system translates the Jahai terms into Malay language. This way, by using the system, other ethnics in Malaysia can understand the language especially when communicating with Jahai people. Three main steps are required in the translation process; first, key in the text input via special character keypad. Then, the system will search the matching word in the database. Finally, the meaning of the word will be displayed. The testing results have indicated this system is functional and accepted with the SUS score of 94/100. Several future recommendations could be made such as including voice search function and adding more Jahai terms in other categories so as to improve the functionality and usability of this proposed system.

Corresponding Author:

Masurah Mohamad
Computing Sciences Studies, College of Computing, Informatics and Media, Universiti Teknologi MARA Perak Branch, Tapah Campus, 35400 Tapah Road, Perak, Malaysia
Email: masur480@uitm.edu.my

1. INTRODUCTION

Jahai is a sub-ethnic group of the major indigenous peoples of the Negrito tribes [1]. They can be found in inland northeast Perak, either near river estuaries or lake coasts (https://www.perakgis.my/jakoa/index.php/kenali-orangasli/suku-kaum-orang-asli/kaum-jahai). They are natives of the Royal Belum Perak State Park. They live a nomad life, but with government funding, they now live in settlements with basic living necessities such as water and electricity. Their settlements are in Banun, Sungai Tiang and the coast of Temenggor Dam. The Jahai language, pronounced Jehai, is a Mon-Khemer (Austroasiatic) language spoken on the Mainland of Southeast Asia as well as various parts of India, Bangladesh, southern China, and Nepal [2]. The language is used among the Jahai people only. Living in Royal Belum Perak State Park, Jahai is one of the ethnic groups who are very well versed in the surrounding areas of the Royal Belum State Park. They often receive tourists and researchers eager to learn about Malaysian flora and fauna. However, language becomes a barrier [3] between the non speakers and Jahai speakers. A translator is needed to assist them in communicating which each other. Moreover, Jahai people also have difficulties in communicating with others, especially when they need basic services from hospitals, schools, and markets [4].

Recently, instead of direct communication, people connect with each other using internet technology. Communication is required not only for social activities, but also for business purposes. To ensure a smooth

communication process without having language barriers, both parties must speak the same language; otherwise, a translator must be hired to complete the mission. It can be difficult to find a perfect translator, especially when dealing with people who speak a relatively uncommon language [5]. Thus, a language translation system is required to overcome this issue. By implementing the language translation, a number of limitations can be reduced. For instance, there is no need to pay expensive translator charge [6], trustworthy issue can also be solved, and extra cost on transportation and accommodation for the translator could be waived [7].

Various applications, either web-based or mobile, have been developed to assist the people in need especially when travelling or on a business trip. One of example and the easiest to use is the application that has been offered by Google which is Google translate (https://www.g2.com/products/google-translate/reviews). This application covers almost all main languages around the world including *Bahasa* Malaysia. The user not only can input the text, but it can be extended into voice searching to ease the translation process. Other applications offer advanced technology such as artificial intelligence model (M2M-100 https://about.fb.com/news/2020/10/first-multilingual-machine-translation-model/), voice recognition (Translate Me https://www.g2.com/products/translate-me/reviews) and neural network to translate the languages [8], [9]. However, these applications do not include Jahai language as one of the options. Thus, it could not solve the communication problem between Jahai and non-Jahai speakers. Many research works have been conducted in this area so that more sophisticated applications could be developed to improve the existing applications. Besides, some researchers are more likely to preserve the language instead of to commercialise the application. Preserving the language by storing the text and pronunciation voice in a database either via web-based system or mobile application [10] could help our next generation to learn and understand the language.

The following are some of the existing recent mobile applications available in the market. Each of the applications has its own purpose and functions to fulfill the needs of specific users or requirements.

- Application 1: Morisia: A neural machine translation system to translate between Kreol Morisien and English. This system is proposed by Pudaruth *et al*. [11] which has a similar objective with the proposed work where it wanted to preserve the Kreol Morisien language and let other people learn Morisien language. It was developed using a deep learning approach based on the transformer model for the translation process. It is a user-friendly application that can be downloaded from translatekreol.mu and from the Google PlayStore.
- Application 2: Multilingual machine translation system based on decoder recurrent neural network. This system is proposed in order to overcome the traditional language translation system in terms of word searching speed. A recursive neural network has been used to develop this system which can translate English to Chinese words. The researchers have claimed that this work has improved the BLUE score 1.51-1.86 compared with the baseline system even though there are several limitations [12].
- Application 3: Hybrid Filipino-English machine translation system. This system implemented a hybrid approach that combined transfer rules and translation templates from the dataset collection. This work also preserved the subject-verb agreement and other linguistic issue during the development process [13].
- Application 4: Other mobile language translation applications. These applications can be downloaded from the provider or even from the application store. https://www.g2.com/categories/machine-translation has listed a number of applications that are popular among the users in language translation. Some of the applications are Google translate that obtained high review scores from the users in terms of ease of use, direct text translation, real-time translation and auto-detect language. Unbabel, which is focused on business users, offers artificial intelligence and human approaches in translating the language. It works with Facebook and Microsoft and many leading brands to help them to communicate effortlessly around the world. Translate Me, is the third popular application that helps users to translate using photos and voice recognition technology. These applications are really helpful in increasing the productivity of the users especially the translators [14].

This study proposed a mobile language translation application that could preserve the Jahai language for their next generation and communities. In addition, it is aimed to assist other people to learn and understand the Jahai language by communicating with the Jahai people without using a translator. Moreover, this proposed application is hoped to overcome the communication problems between Jahai people and other people who are unable to speak Jahai language but have direct contact with them like doctors, teachers, tourist and other tribes.

This paper is organized as follows in order to present the whole structure of the proposed work. Section 1 introduces the background study of the proposed work. Section 2 explains some literature works that might increase the readers understanding such as the Jahai language and several language translation methods. Section 3 presents the method used to develop the proposed system and section 4 presents the results and discussion of the proposed system. Finally, section 5 concludes the overall works.

2. RELATED LITERATURE

This section presents several terms and works that are related to this proposed work. It will be narrowed down to the definition of Jahai language, natural language processing, machine translation, information retrieval and a number of existing mobile language translation applications. It will increase the understanding of the focus issue and the problem that need to be solved.

2.1. Jahai language

Due to the elaborate processes of derivational affixation, reduplication, and rich pronominal and demonstrative distinctions, the characteristics of the Jahai language are extremely complex [15]. Between 2004 and 2008, Burenhalt N. published a series of studies on Jahai language, including descriptions of bodily parts [16], hydrological lexicon [17], landscape words and typonyms [18], and placement and removal events [19]. Later, other researchers expanded on Burenhalt's work, such as Wierzbicka [20] on semantic topology and Wegener [21] on Savosavo ethnic body parts. Table 1 presents some of the Jahai terms focusing on odours and their meaning in English. Jahai has a rich inventory of sounds, which some of them are not present within the English language. For example, most Jahai terms depend on how nasal a specific word is. For example, the term 'kis'. Translated directly without the nasal sound, 'kis' means 'to dig'. But *kĩs* which sounds like 'kiss' is actually spoken with air sent through the nose, carries the meaning 'ghost' [22].

Table 1. Odours terms in Jahai

Verb	Approximate translation	Prototypical sources
cŋəs	'to smell edible, tasty'	cooked food, sweets
crŋir	'to smell roasted'	roasted food
harɨm	'to be fragrant'	flowers, perfumes, soap
ltpɨt	'to be fragrant'	flowers, perfumes, soap
pʔus	'to be musty'	old dwellings, mushrooms, stale food
cŋɛs	'to have a stinging smell'	petrol, smoke, various plants and insects
sʔɪ̃ ŋ	'to have a smell of human urine'	human urine, village ground

2.2. Natural language processing

Natural language processing (NLP) is an area of artificial intelligence (AI) in which computers analyze, understand and extract meaning from human language in a useful and smart manner. By using NLP, AI developers can organize and structure information to perform tasks such as automated summarization, translation, entity recognition, relationship extraction, emotion analysis, speech recognition, and subject segmentation. NLP is used to analyze text, which allows machines to understand how humans talk and react. This human-computer interaction facilitates real-world applications such as automated text summarization, emotion analysis, subject extraction, named object recognition, speech tags, relationship extraction, stemming, and more. NLP is commonly used for text mining, machine translation, and automated question-response. With the exponential growth of AI and computational technology, current NLP research and implementation also includes AI-base machine learning, data mining, deep learning and agent ontology [23]. NLP can be divided into two broad areas: core or fundamental, and applications. The core areas address language modeling, morphological processing, syntactic processing, and semantic processing [24]. Meanwhile, the application areas involve topics, such as chatbot, speech recognition, language translation, recommender systems, spam detection, and sentiment analysis [25].

There are five general steps in natural language processing: i) morphological analysis; ii) syntactic analysis; iii) semantic analysis; iv) pragmatic analysis; and v) discourse analysis. Morphological analysis is analyzing words into their component parts, free morphs and bound morphs. It involves identifying and analyzing the structure or words where the entire chapter of text is divided into paragraphs, sentences and words. Syntactic analysis or known as Parsing is analysis of words in the sentence for grammar and arranging the words in the manner that shows the relationship among the words. For example, sentence such as "The shop goes to man" is rejected by English syntactic analyzer. Next is semantic analysis. It will draw the exact meaning from the text by mapping the syntactic structures with objects in the task domain. For example, sentences such as "cold fire" will be disregarded. Discourse integration is where the meaning of the sentence just before it is taken into consideration when deriving meaning of a text. This consequently can bring about the meaning of immediately succeeding sentence. Finally, yet importantly, the Pragmatic analysis re-interpretes the text on what it actually meant. It also involves deriving the aspects of the language, which require real world knowledge [26].

2.3. Machine translation

The automatic translation of one language into another using a software programme is known as machine translation. Large amounts of data can be translated in a short amount of time without the need for human interaction. The publication of Warren Weaver's Memorandum of Translation in 1949 marked the beginning of modern machine translation. Rule-based, statistical machine translation (SMT), corpus-based, and neural machine translation (NMT) are some of the machine translation methodologies employed [27]. In the 1970s, the rule-based method to machine translation was introduced. For each language pair, a vast number of linguistic rules and bilingual dictionaries are required to parse text. As a result, a sentence from a source language to a target language with complicated rule sets and huge lexicons would be produced. They succeeded in translating text into glossed sign language corpora, which is an example of research for this approach [28].

The statistical machine translation outcomes are more fluent and understood by the user than the rule-based technique. This is due to the fact that this method makes use of statistical models derived from the examination of monolingual and bilingual corpora. Unfortunately, this method requires a lot of CPU and hardware resources. The outcomes are similarly unpredictable and inconsistent [29]. However, the corpus-based technique compares the translated text to its original text in a parallel corpus or to another corpus built with similar design requirements in the same or another language. The drawback for this technique is that the translation is heavily dependent on the parallel corpus to achieve high quality results. Generalizing the translation would result in low quality translation.

The neural machine approach extends and simplifies the statistical machine translation approach since the result satisfies user expectations better than other approaches. This is accomplished by employing the developed statistical model to train neural networks to achieve translation without having to go through the statistical machine learning pipeline of specialist systems. Google Translate is the most well-known example. However, there are certain drawbacks to using these systems, such as domain mismatch, a large amount of training data, unusual words, extensive sentences, word alignment, and beam search [28].

2.4. Neural machine translation

2.4.1. Artificial neural network

Artificial neural network (ANN) is a structure made up of machine learning algorithm where the artificial neurons make the central computational unit. The neurons focus on uncovering the underlying patterns or associations within the data collection. This process is somewhat similar to how the human brain does during making a decision. An ANN have thousands or more of artificial neurons called processing units interconnected by nodes. These processing units consist of input and output units. The input units are provided with different types of knowledge structures based on an internal weighting system. The neural network will then find information to prepare an output report. Just as people need rules and instructions for a result, ANNs often use learning rules set which is backpropagation. ANN is known as feed-forward neural network because inputs are processed only in the forward direction. ANN can be used to solve problem related to tabular data, image data and text data [30]-[32].

2.4.2. Deep neural networks

Deep neural networks are multiple layers of ANN that carry out the machine learning process. The first layer will receive and process the raw input and passes it to the next layer. The processing will continue throughout all the layers until the network achieves the desired result. There are many types of deep neural network (deep learning) tools for natural language processing (NLP) such as recurrent neural network (RNN) and long short-term memory (LSTM) [11]. Recurrent neural network (RNN) is an ANN that processes temporal data. Standard neural machine translation system is an end-to-end neural network, in which the source sentence is encoded by an RNN called encoder, and the targeted sentence is decoded using another RNN which is decoder. The encoder will read a source sentence one word at a time and will summarize the entire source sentence in its last hidden state. Backpropagation is used by the RNN decoder to learn this summary and returns the translated version. RNNs is bound with the problem of vanishing or exploding gradient problem, where depending on the activation functions used, information rapidly will get lost over time. Consequently, the RNNs will experience difficulty memorizing previous words which are far from the sequences and could only make predictions based on the recent words only [33]. Long short-term memory (LSTM) is a network that solves the vanishing or exploding gradient problem. This network uses gates and an explicitly defined memory cell. The gates function to safeguard the information by stopping or allowing the flow of it. The input gate will determine how much is the information from the previous layers get stored in the cell. The output layer takes the job on the other end and determines how much of the next layer gets to know about the state of a cell. The forget gate will stop certain information from flowing into the cell so that the network can forget some characters from previous sentences [32].

2.5. Information retrieval

Information retrieval is a process of searching for knowledge or information that involves data acquisition, retrieval and organization. The information not only focuses on the text but has been extended into image, video, medical data and chemical structures. Due to the various type and characteristics of data, the new digital repositories have tremendously evolved [34]. An information retrieval system must enable three essential processes: the representation of document content, the representation of the user's information demand, and the comparison of the two representations. There are two main processes in information retrieval; indexing and query formulation. Indexing involves processing documents, while query formulation is a process of representation of user's need. Both processes will go through the matching phase or also known as comparing process between the query and documents available in the database. The matching query or document is then returned to the user as a result. There are many models and techniques introduced by recent researchers such as Boolean model and vector space model (VSM) for document searching and Signature file and inversion indices for indexing process [35]. Some of the applications that implement information retrieval are digital library, media search and search engines.

3. METHOD

The development of the system follows the mobile application life-cycle. The cycle consists of four phases which are requirements specification, development modeling, design and development, and testing and deployment. The specification of the system is obtained from the literature review of existing systems and also from the existing web-based system. It is an enhanced version of the web-based system that requires the development of a mobile application to improve user experience and functionality. After obtaining the specific requirements, the application starts to be modelled, designed and developed according to user requirements. Finally, the final version of the proposed application was tested and deployed.

3.1. The workflow of the system

Figure 1 presents the architecture of the proposed application. The dataset consists of Jahai and Malay languages and being included into a cloud database for text searching purposes. Currently, Jahai words are taken from the previous version [36] where the words have been originally collected from the research works done by Burenhalt [16]–[18], [22] and from the interview session with the Jahai ethnic itself. There are two types of users, end user and admin. The end user needs to key in the text either in Jahai or Malay for the translation process. The output will be displayed after the search engine finds the matching word. Meanwhile, the admin is responsible for managing the application such as inserting, deleting and updating the words. The application is developed for mobile users and can be implemented in either Android or IOS platform.

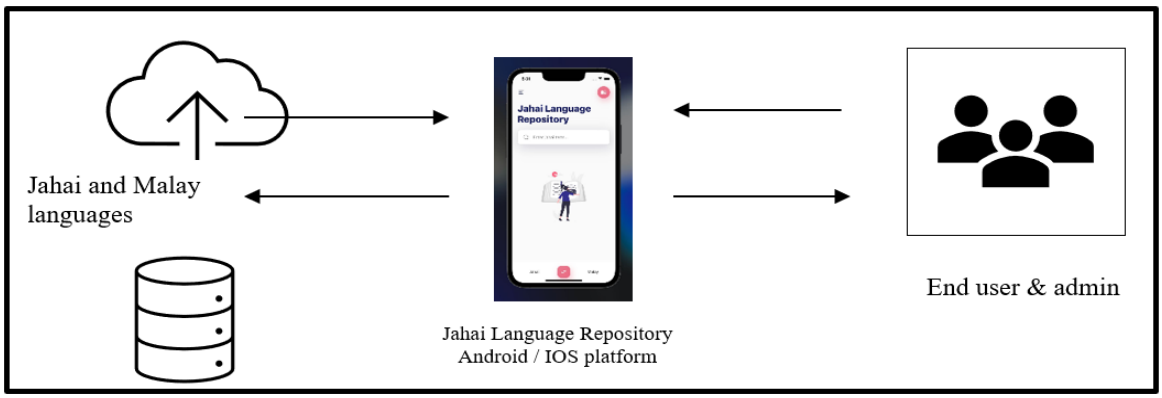


Figure 1. System architecture of the proposed application

The flow of the translation process is as follows. The user displays the main page of the Jahai language repository application and chooses the mode of language that he or she wants to translate. Then, the user needs to key in the terms to start the translation process. If the user is an administrator, he or she needs to login to the system to manage the terms that have been stored in the repository. The management tasks involve view word, update word, delete word and insert new word into the system. Both users can view the output of the translation process at the display page. The details of the pages are shown in the following sections. Figure 2 presents the basic flow of the translation process for both users.

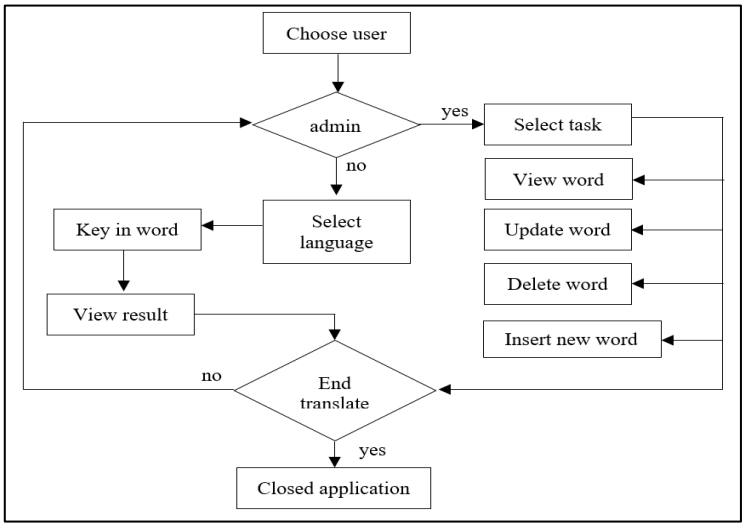
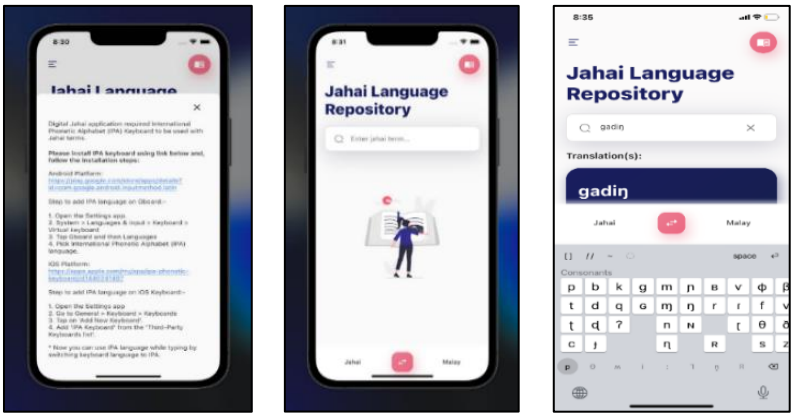


Figure 2. Flow chart of the proposed work

4. RESULTS AND DISCUSSION

The Jahai-Malay translation application can be used by anyone interested in learning the Jahai language. The first version of the proposed application [36] only offers translation from Jahai to Malay or from Malay to Jahai Languages. However, this application also includes the English term when it displays the translation result. Besides, the result page also provides a description of the search word to increase the understanding of the user. To use this application, users are required to install a special character keypad known as the international phonetic alphabet (IPA) keypad. This is because, Jahai terms consist of several alphabets that are not commonly used in Malay or English words. Figure 3 shows the main interfaces of the application which (a) indicates the instruction setting for IPA keypad, (b) page to enter the Jahai or Malay term, and (c) the of IPA keypad. Figure 3(a) shows the page that instructs the user to install the IPA keypad for two different users either using Andriod platform or iOS platform. Meanwhile, Figure 3(b) will receive the input from the user. The user can enter the Jahai term and also select the translation mode either from Jahai to Malay or from the Malay to Jahai at the bottom of the page. Then, Figure 3(c) displays the page with the IPA keypad to ease the user in using special character to write the Jahai term for the translation process.

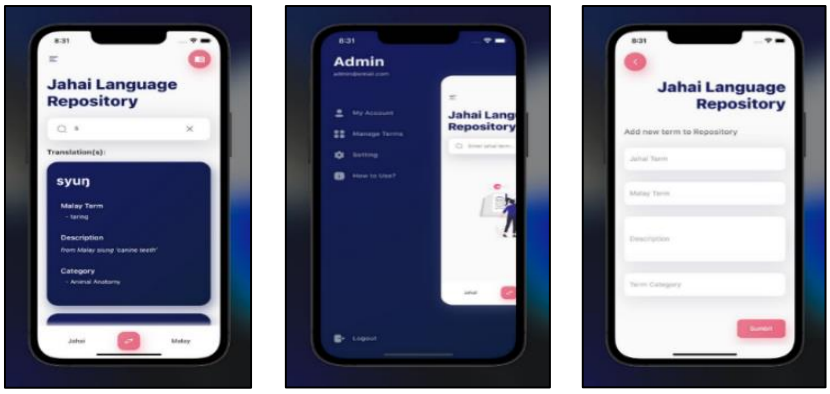


(a) (b) (c)

Figure 3. Jahai language repository main page s where (a) is an IPA keypad setting page, (b) the keyword searching page, and (c) the IPA keypad

Figure 4 demonstrates other pages in the proposed application where (a) shows the example of the translation output, (b) the main page for admin, and (c) setting to add a new term in the repository. Figure 4(a) demonstrates the example output of the translation process that provides the Malay term, description and the category of the entered Jahai word. For example, when user keys in character “s” in the search box, the page

will display the matching result under Translation section. The results will display the selected term with other descriptions such as the term in Malay definition, description of the term and also the category of the term. Figure 4(b) is created for the admin user to manage the words stored in the repository such as inserting and deleting words in the repository. For example, if the admin wants to add a new term, he or she needs to choose "Manage Terms" function and add the required term. Figure 4(c) demonstrates how the function of adding new term to the repository is done when the admin chooses "Manage Terms" function. Four input details are needed to add a new term to the repository which are i) Jahai term, ii) Malay term, iii) Description, and iv) Term category. Other additional functions available for admin are, to check account in "My Account", to alter settings in "Setting" and to learn more on how to use the application in "How to use" function.



(a) (b) (c)

Figure 4. Other available pages where (a) the translation output page, (b) the admin page, and (c) adding new term page

4.1. Results discussion

Two evaluation measures namely black box functionality test and system usability scale (SUS) test have been conducted to evaluate the performance of the proposed application. The black box functionality test is specifically used to evaluate the functionality of the application [37]. Table 2 indicates the results of the functionality test on the proposed application's main functions. It showed that all provided functions are well-functioning and can be accepted for this version. Meanwhile, the SUS test was conducted to evaluate the feedback received from the end-users regarding the usability of the specified requirements, especially on the interface. These tests are conducted by providing several questions that are constructed based on the standard specified questions.

Table 2. The functionality test results of the proposed application

No	Features	Expected results	Observation	Results
1.	Jahai-Malay language translation menu	User can select Jahai language option	User can select the Jahai language option	Accepted
2.	Malay-Jahai language translation menu	User can select Malay language option	User can select the Malay language option	Accepted
3.	International phonetic alphabet (IPA) keypad	User can install the IPA keypad User can use the IPA keypad	User can install the IPA keypad User can use the IPA keypad	Accepted
4.	Administrator login page	Admin can login to admin page	Admin can login to admin page	Accepted
5.	View word	Admin can view existing words stored in the repository	Admin can view existing words stored in the repository	Accepted
6.	Update word	Admin can update existing words in the repository	Admin can update existing words in the repository	Accepted
7.	Delete word	Admin can delete existing words in the repository	Admin can delete existing words in the repository	Accepted
8.	Insert word	Admin can insert words in the repository	Admin can insert new words in the repository	Accepted
9.	Translation page	User can view the result of the search word	The description of the search word is displayed	Accepted

Table 3 indicates the results of the system usability scale (SUS) test that has been conducted to 10 randomly selected respondents aged 30 to 50 years old Malaysian citizens who volunteered to test the proposed

application. The respondents were selected without considering educational qualification, personal background, or expertise in any specific field. The aim is to obtain their response towards the design of the interface, which will help to improve the usability and functionality of the proposed application. A Lickert scale from 1 to 5 which represents strongly disagree until strongly agree is used to obtain the response given by the respondents. Table 3 lists the 10 questions presented to the respondents. As can be seen, the questions consist of positive and negative statements. The response to these statements, once calculated will produce the SUS score and the overall result of the user acceptance towards this application. The SUS score obtained is divided into three categories; not acceptable if the score is below 50, marginal if the score is above 50 and below 68 and acceptable if the score is above 68.

Table 3. SUS questions

Number	Questions
1	I think that I would like to use this project frequently.
2	I found the project unnecessarily complex.
3.	I thought the project was easy to use.
4.	I think that I would need the support of a technical person to be able to use this project.
5.	I found the various functions in this project were well integrated.
6.	I thought there was too much inconsistency in this project.
7.	I imagine that most people would learn to use this project very quickly.
8.	I found the project very cumbersome to use.
9.	I felt very confident using the project.
10.	I needed to learn a lot of things before I could get going with this project.

According to the SUS score obtained from the responses, out of 100, the proposed application obtained 94 points which is above the average value. This score indicates 'acceptable to be applied'. Figure 5 presents the score for even and odd questions for each respondent and Figure 6 shows the SUS score of each respondent. The overall result indicates all respondents accept the design and are satisfied with the interface of the proposed application. However, only one respondent, which is respondent 6 gave a low score of 87.5 compared to other respondents. This is maybe because the respondent did not really want to use this application, or the available features did not fulfill their needs.

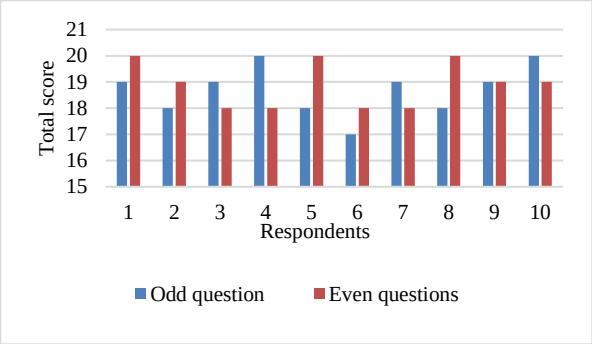


Figure 5. Score on odd and even questions for each respondent

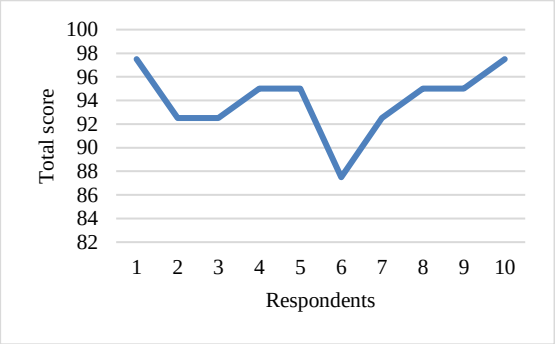


Figure 6. SUS score for each respondent

5. CONCLUSION

As a conclusion, the Jahai language repository application does give a positive impact to the Malaysian community, especially to the people who wants to learn and preserve Jahai language for their next generation. Moreover, this application could become a mobile translator to anyone who needs to communicate with Jahai speakers efficiently. Thus, the number of human translators could be reduced and no additional cost is required. In addition, according to the evaluation results, a lot of improvement can be made in order to increase the interest of people in using this application. One of the major improvements that can be made is to apply voice recognition technology, so that people with disabilities in entering the text or having other difficulties using the IPA keypad can make use of this feature. Furthermore, this application can include more foreign languages such as English, Chinese and Tamil and other international languages to get higher volume of users.

ACKNOWLEDGEMENTS

The authors would like to acknowledge UiTM for the funding of this project under the LESTARI Sustainable Goal Development Triangle @ UiTM Resesarch Grant 600-RMC/LESTARI SDG-T 5/3 (144/2019).

REFERENCES

[1] O. Zainon, "Management of information and infrastructure of indigenous community at Royal Belum State Park using geographical information system: a conceptual design," *The Asian Journal of Technology Management (AJTM)*, vol. 8, no. 1, pp. 47–55, 2015, doi: 10.12695/ajtm.2015.8.1.5.

[2] S. Likin, N. Zainun, and F. B. Jaafar, "Traditional knowledge of Jahai Tribe at Kampung Semelor, Gerik, Perak," in *In Proceedings of the 7th International Conference on Multidisciplinary Research (ICMR 2018)*, 2020, pp. 676–680, doi: 10.5220/0008892306760680.

[3] J. Espinoza and S. Derrington, "How should clinicians respond to language barriers that exacerbate health inequity?," *AMA Journal of Ethics*, vol. 23, no. 2, pp. E109–E116, 2021, doi: 10.1001/AMAJETHICS.2021.109.

[4] M. Xiao, "Chinese international graduate students at Canadian universities: language barriers, cultural identities and perceived problems of engagement," *International Journal of Inclusive Education*, vol. 0, no. 0, pp. 1–18, 2021, doi: 10.1080/13603116.2021.1941318.

[5] R. N. Rosa, "Strategies in solving equivalent-related problems in translation: implications to cross-cultural understanding," *International Journal of Comparative Literature and Translation Studies*, vol. 9, no. 3, p. 8, 2021, doi: 10.7575/aiac.ijclts.v.9n.3p.8.

[6] R. Khelifa, T. Amano, and M. A. Nuñez, "A solution for breaking the language barrier," *Trends in Ecology and Evolution*, vol. 37, no. 2, pp. 109–112, 2022, doi: 10.1016/j.tree.2021.11.003.

[7] V. Kirov and B. Malamin, "Are translators afraid of artificial intelligence?," *Societies*, vol. 12, no. 2, 2022, doi: 10.3390/soc12020070.

[8] M. Singh, R. Kumar, and I. Chana, "Corpus based machine translation system with deep neural network for sanskrit to Hindi translation," *Procedia Computer Science*, vol. 167, no. 2019, pp. 2534–2544, 2020, doi: 10.1016/j.procs.2020.03.306.

[9] D. A. Hameed, T. A. Faisal, A. K. Abbas, H. A. Ali, and G. T. Hasan, "DIA-English-Arabic neural machine translation domain: sulfur industry," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 27, no. 3, pp. 1619–1624, 2022, doi: 10.11591/ijeecs.v27.i3.pp1619-1624.

[10] M. A. Dootio and A. I. Wagan, "Development of Sindhi text corpus," *Journal of King Saud University - Computer and Information Sciences*, vol. 33, no. 4, pp. 468–475, 2021, doi: 10.1016/j.jksuci.2019.02.002.

[11] S. Pudaruth *et al.*, "Morisia: A neural machine translation system to translate between Kreol Morisien and English," *inTRAlinea*, vol. 23, no. February, 2021, doi: 10.13140/RG.2.2.17456.69125.

[12] F. Deng, "Multilingual machine translation system based on decoder recurrent neural network," *Wireless Communications and Mobile Computing*, vol. 2022, 2022, doi: 10.1155/2022/9714800.

[13] G. K. A. Fontanilla and R. E. O. Roxas, "Hybrid Filipino-English machine translation system," Master's Thesis, Computer Science, College of Computer Studies, 2022.

[14] A. Baihaqi and A. Mulyana, "Reviewing the result of machine translation: a case for indonesian translation version By Google translate and imtranslator," *PROJECT (Professional Journal of English Education)*, vol. 4, no. 1, p. 1, 2021, doi: 10.22460/project.v4i1.p1-9.

[15] N. Burenhult, *A grammar of Jahai*, Pacific Linguistics, The Australian National University, Canberra, Australia, 2015.

[16] N. Burenhult, "Body part terms in Jahai," *Language Sciences*, vol. 28, no. 2–3, pp. 162–180, 2006, doi: 10.1016/j.langsci.2005.11.002.

[17] N. Burenhult, "Streams of words: Hydrological lexicon in Jahai," *Language Sciences*, vol. 30, no. 2–3, pp. 182–199, 2008, doi: 10.1016/j.langsci.2006.12.005.

[18] N. Burenhult, "Landscape terms and toponyms in Jahai: a field report," *Working Papers in Linguistics*, vol. 51, pp. 17–29, 2009.

[19] J. H. Petersen, "Events of putting and taking," *Typological Studies in Language*, pp. 349–366, 2012.

[20] A. Wierzbicka, "Bodies and their parts: An NSM approach to semantic typology," *Language Sciences*, vol. 29, no. 1, pp. 14–65, 2007, doi: 10.1016/j.langsci.2006.07.002.

[21] C. Wegener, "Savosavo body part terminology," *Language Sciences*, vol. 28, no. 2–3, pp. 344–359, 2006, doi: 10.1016/j.langsci.2005.11.005.

[22] A. Majid and N. Burenhult, "Odors are expressible in language, as long as you speak the right language," *Cognition*, vol. 130, no. 2, pp. 266–270, 2014, doi: 10.1016/j.cognition.2013.11.004.

[23] N. Bhadwal, P. Agrawal, and V. Madaan, "A machine translation system from hindi to sanskrit language using rule based approach," *Scalable Computing*, vol. 21, no. 3, pp. 543–553, 2020, doi: 10.12694:/scpe.v21i3.1783.

[24] D. W. Otter, J. R. Medina, and J. K. Kalita, "A survey of the usages of deep learning for natural language processing," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 32, no. 2, pp. 604–624, Feb. 2021, doi: 10.1109/TNNLS.2020.2979670.

[25] B. D. Shivahare, S. Ranjan, A. M. Rao, J. Balaji, D. Dattattrey, and M. Arham, "Survey paper: study of sentiment analysis and machine translation using natural language processing and its applications," *Proceedings of 3rd International Conference on Intelligent Engineering and Management, ICIEM 2022*, pp. 652–656, 2022, doi: 10.1109/ICIEM54221.2022.9853044.

[26] S. S. Pandya and N. B. Kalani, "Preprocessing phase of text sequence generation for gujarati language," *Proceedings - 5th International Conference on Computing Methodologies and Communication, ICCMC 2021*, no. Iccmc, pp. 749–752, 2021, doi: 10.1109/ICCMC51019.2021.9418046.

[27] V. A. Gandhi, V. B. Gandhi, D. V. Gala, and P. Tawde, "A study of machine translation approaches for Gujarati to English translation," *Proceedings - 1st International Conference on Smart Technologies Communication and Robotics, STCR 2021*, no. October, 2021, doi: 10.1109/STCR51658.2021.9588859.

[28] N. K. Kahlon and W. Singh, *Machine translation from text to sign language: a systematic review*, no. July 2021. Springer Berlin Heidelberg, 2021.

[29] L. Benkova, D. Munkova, Ľ. Benko, and M. Munk, "Evaluation of English–Slovak neural and statistical machine translation," *Applied Sciences (Switzerland)*, vol. 11, no. 7, p. 2948, 2021, doi: 10.3390/app11072948.

[30] P. V. Terlapu, M. Yugandhar, B. Ramesh, B. V. Kumar, and R. Pemula, "Student cognitive learning capability (SCLC) prediction system using PCA-ANN based Model," *Proceedings - 2022 International Conference on Computing, Communication and Power Technology, IC3P 2022*, pp. 11–18, 2022, doi: 10.1109/IC3P52835.2022.00013.
[31] A. F. Hasan, H. U. Masud, T. Anzar, and P. Hasan, "Hand gesture recognition based on surface electromyogram signal (sEMG) with Muscular Contraction level and real time implementation on an artificial prosthetic wrist using artificial neural network (ANN)," *2020 IEEE Region 10 Symposium, TENSYMP 2020*, no. June, pp. 1648–1651, 2020, doi: 10.1109/TENSYMP50017.2020.9231038.
[32] M. G. M. Abdolrasol *et al.*, "Artificial neural networks based optimization techniques: A review," *Electronics (Switzerland)*, vol. 10, no. 21, 2021, doi: 10.3390/electronics10212689.
[33] J. A. Nasir, O. S. Khan, and I. Varlamis, "Fake news detection: A hybrid CNN-RNN based deep learning approach," *International Journal of Information Management Data Insights*, vol. 1, no. 1, 2021, doi: 10.1016/j.jjimei.2020.100007.
[34] W. R. Hersh, "Information retrieval and digital libraries," in *Biomedical Informatics: Computer Applications in Health Care and Biomedicine: Fourth Edition*, London: Springer London, 2014, pp. 613–641.
[35] A. El-Ansari, A. Beni-Hssane, M. Saadi, and M. El Fissaoui, "PAPIR: privacy-aware personalized information retrieval," *Journal of Ambient Intelligence and Humanized Computing*, vol. 12, no. 10, pp. 9891–9907, 2021, doi: 10.1007/s12652-020-02736-y.
[36] I. A. Mohtar, M. Mohamad, P. N. Azzuri, N. A. Rahman, and S. A. M. Nor, "Development of a web-based Jahai – Malay language repository," in *2021 Fifth International Conference on Information Retrieval and Knowledge Management (CAMP)*, Jun. 2021, pp. 14–18, doi: 10.1109/CAMP51653.2021.9497863.
[37] F. F. Laksana, Istianah, T. bin Sansuwito, and I. S. Kep, "Development of a survey instrument mobile app to determine the map location of a drugstore," *Malaysian Journal of Medicine and Health Sciences*, vol. 18, pp. 188–194, 2022.

BIOGRAPHIES OF AUTHORS

Nurazzah Abd Rahman received the B.Sc Computer Science & Mathematics from Heidelberg College, Tiffin, Ohio, U.S.A. 1983-1987, the M.Sc Computer Science from Ohio University, Athens, Ohio, U.S.A. 1987-1989 and Ph.D. in Information Technology and Quantitative Sciences (Information Retrieval) from Universiti Teknologi MARA, Shah Alam, in 2011. She can be contacted at email: nurazzah@uitm.edu.my.

Masurah Mohamad he is currently a Senior Lecturer at Universiti Teknologi MARA Perak Branch, Tapah Campus (UiTM Perak). Hehas received a B.Sc. (Hons.) in Computer Science majoring in Software Engineering in 2004, and M.Sc. in Computer Science from Universiti Teknologi Malaysia (UTM) in 2006. She received her Ph.D. in Computer Science from UTM in May 2021. She can be contacted at email: masur480@uitm.edu.my.

Itaza Afiani Mohtar received the B.Sc. (Hons.) in Intelligent Systems from Universiti Teknologi Mara (UiTM), Malaysia, the MIT (Computer Science) from the Universiti Kebangsaan Malaysia (UKM), Malaysia. She is currently a Senior Lecturer with the Department of Computer Science. Universiti Teknologi Mara Perak Branch. Her research interests include machine learning, and intelligent systems. She can be contacted at email: itaza328@uitm.edu.my.

Saidi Adnan Md Nor received the BA. in Syariah from Universiti Malaya (UM), Malaysia, the MA. degree in Syariah from Universiti Malaya (UM), Malaysia. He is currently a Lecturer with the Academy of Contemporary Islamic Studies, Universiti Teknologi Mara Perak Branch. His research interests include Islamic Law and Islamic Family Law. He can be contacted at email: saidi845@uitm.edu.my.