

E-Translator Kawi to Bahasa

Oka Sudana^{1,*}, *Darma Putra*², *M.Sudarma*³, *Rukmi S.Hartati*³, *Ryan Pradnya Prastika*² and *Ayu Wirdiani*²

¹Study Program of Doctoral Engineering Science, Faculty of Engineering, Udayana University

²Information Technology Departement, Faculty of Engineering, Udayana University

³Electrical and Computer EngineeringDepartement, Faculty of Engineering, Udayana University
Jalan Raya Bukit Jimbaran, Kabupaten Badung - Bali, 80361

Abstract. Kawi Language is a kind of language which has developed in Java during the Hindu-Buddhist era in the ancient Indonesian archipelago. As the time passed, Kawi began to come to extinction. To overcome this, preservation of Kawi language can be done by enhancing the understanding of culture through the use of technology. Technology that can be used for cultural preservation is one of NLP based applications such as Kawi Language to Bahasa Translator Application which is developed in this research. The translator application is an application that translates from one language to another. The result of translation of the translator application is a language that is easy to understand by the people or commonly used language. This translator application is developed to processes Kawi Language text into bahasa based on search method which each word inserted. The stemming method used is Bobby Nazief & Mima Adriani Algorithm using Kawi Language rule. The translated Kawi Language text can be a word, sentence or paragraph with the result of the Text in Bahasa as well as words, sentences or paragraphs.

1 Introduction

Kawi language is a kind of language that once has developed on the island of Java in the era of the Hindu-Buddhist Kingdom in the ancient Indonesian Archipelago and used in the writing of literary works. Kawi Language is now approaching extinction because of its users are getting less.

Preservation of Kawi Language can be done by making translator which is the application of Natural Language Processing. Natural Language Processing (NLP) is a branch of Artificial Intelligence (AI) that focuses on natural language processing. Translator application is widely available, but there is no application that can translate Kawi Language into Bahasa (Indonesian). The Translator Application of Kawi Language to Bahasa is considered necessary due to the absence of an electronic-based dictionary or translation that still uses a conventional dictionary so that translation becomes longer and more difficult to learn by people who do not really understand Kawi Language.

* Corresponding author: agungokas@unud.ac.ic

Translator apps are widely available and can be easily found with the help of the Internet. Translator application can be built on desktop, web, and mobile. There are several translators that have been created using various methods such as the Translator Application of Balinese Language into Bahasa, based on Android with searching process using binary search [1], Electronic dictionary application of Bahasa and Javanese sentences using Markov Model [2], English to Bangla Language using rule-based methods [3], an English to Balinese text translator using a rule-based approach [4], an Android-based Bahasa to Javanese language translator [5], an Android-based Translator Application of Javanese Language to Bahasa [6], An Indian to English Translator using a transfer-based method [7], an English to Marathi translator using the rule-based method [8].

The translator application is an application that can be combined with the cultural aspects, especially language. Technology which is combined with the culture will have a very positive impact especially in terms of learning, especially in language learning. Language is an important part of the organization of human life in the aspects of manners and the value of humanity itself [9]. Humans can grow, learn, and can develop their culture, preserve their culture or even develop their culture through language. These reasons indicate the necessity of Translator Application the Kawi Language to Bahasa. This application can process input in the form of Kawi Language word to produce output in the form of translation in Bahasa. Translator apps created on a web basis. Translator application will help interpret and understand Kawi Language easier.

2 Methods

2.1 Natural Language Processing

The principle of natural language is a form of representation of a message which wishes to be communicated between humans. The main form of representation is in the form of sound or speech but often expressed in writing also. The essence of natural language processing is the sentence deciphering or often called parsers. This parser works to read sentences, word by word and determine what types of words may follow the respective word. There are several applications that can be created using NLP such as, machine translation, speech recognition, speech synthesis, natural language interfaces to databases, information retrieval, information extraction, question answering, text summarization, dialogue systems [10]. Machine translation refers to the translation of the text from one language to another. The Translator of Kawi to Bahasa is a type of machine translation in the NLP. The Translator Application of Kawi to Bahasa can be categorized as machine translation because the software serves to replace humans in translating the source language text into target. Human intervention is not necessary when the translation process is being done because all the processes have been pre-programmed with Kawi Language based rules.

2.2 Tokenization

Tokenization is the process of dividing text in the form of a sentence, paragraph or document, into a token or a specific section. This process is done by searching spaces after each word. All generated tokens are stored in a separate file for further processing. Tokenization is the task of cutting the order of characters and units of a document which is defined into sections and discarding certain characters such as punctuation. The tokenization process places the word limit. The endpoint of a word and the beginning of the next word are called boundary words. This process is also referred to as word segmentation [11].

2.3 Stemming

Stemming is a process or way of finding the base word of a word. Stemming itself serves to eliminate morphological variations attached to a word by removing affixes to the word so that later obtained a correct word according to the morphological structure of language [12]. Stemming itself is part of the Information Retrieval process. Information Retrieval has two processes, indexing and searching. The Indexing process consists of four sub processes: Word Token (converts document into a set of terms by deleting all characters in punctuation found on the document and converting the term set into a lowercase), Stop Word Removal (The process of deleting words that often appear in documents Such as: and, or, not etc.), Stemming is the process of converting a full word into a base word and Word Weighting is the process of weighting each term in a document [13].

The stemming process in text of Bahasa is different from stemming in English text. The process required in English text stemming is only the process of removing suffixes, whereas in text of Bahasa it eliminates suffixes, prefixes, and confixes. Stemming algorithms for several languages have been developed, such as Porter Algorithm for English text, Algorithm Porter for text in Bahasa, Bobby Nazief and Mirna Adriani Algorithms for text in Bahasa [14]. The Bobby Nazief and Mirna Adriani algorithm refers to the morphological rules in Bahasa, which classify the affixes, ie allowable affixes or unauthorized affixes [15]. These groupings include prefixes, back affixes (suffixes), middle affix (infix) and the combination of prefix and suffix (confix) [16].

The stemming process of Kawi language is a process or way of finding the basic word of a Kawi word. Stemming Kawi word serves to eliminate morphological variations inherent in a Kawi word by removing affixes to the Kawi words according to the rules of Kawi language so that later on can be a correct word according to the structure of morphology of the true Kawi language. The stemming of the Kawi word is based on the stemming algorithm of Bobby Nazief and Mirna Adriani using the prefix, infix, and Kawi language suffixes. The method of Bobby Nazief and Mirna Adriani stemming algorithm is used in this application because this algorithm always checks to the database every word which is succeed omitted prefix, infix or suffix. Checking into the database of each cutting process makes the results of Bobby Nazief and Mirna Adriani algorithms close to perfect. The stemming process of the Kawi word will take the word which has been tokenized before which then eliminates the prefix, infix, and suffix contained in Kawi language rules to produce the base word. The process of stemming the word Kawi which takes the word that has been tokenized initially.

2.4 The Word Searching Process

The word searching process is a process to search the word in Bahasa, at the database based on the base Kawi word that has been obtained by the results of the stemming process. Word search process using looping to find the until the following word is found out. The process of searching a word which originally took the base Kawi language that has been stemmed for matching the result in Bahasa.

2.5 The Word Searching Process

The word combining process is a process to combine words that have been translated into sentences. The process of combining words begins with taking the Bahasa word that has been searched in the database which then combined using the rules of Bahasa.

3 Experiment and Result

The testing phase will be conducted by taking samples of Kakawin Ramayana *Sargah* 1 consisting of 64 stanzas. The Kakawin will be inserted into the text area "Kawi Language Text" in the application as in Fig. 1.



Fig. 1 Kawi Text

The text which has already inputted will be executed for tokenizing and stemming process. The stemming process applies Bobby Nazief and Mirna Adriani Algorithms that are enhanced by Kawi language rules. Bobby These algorithms combined with Kawi language rules produce stemming products with a level of 82.27% accuracy.



Fig. 2 Tokenizing and Stemming Result

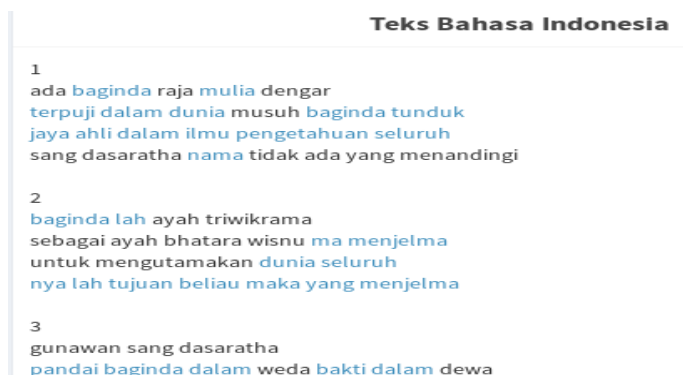


Fig. 3 Text Result

Fig. 2 is the result of the tokenizing and stemming process. The first sentence will be divided to produce a word or phrase. The result of a word that has been altered will be stemmed to find the base word of a word. Stemming eliminates morphological variations embedded in a word by removing affixes such as prefixes, infixes, and suffixes to the word, to obtain a correct word according to the correct Indonesian morphology structure.

After going through some process then The text result in Bahasa will appear as shown in Fig. 3. The translation in Bahasa results have more than one meaning such as the blue word. The blue colour is the link used to select the word. The selection feature is used to select another word which is the meaning of the source text. The selection of the other words can be used if there are some meanings of the word Kawi which are more than one. Selection of other words also serves to adjust the word on the sentence so as to produce a good sentence in Bahasa.

The structure of sentences which have been resulted from the translation process is still not so perfect, for it needs some correction of sentences according to grammar or rules of Bahasa. However, this application does not use rule based methods such as English to Bangla Translator Application, English to Balinese text translator, English to Marathi Translator who keeps the basic rules that contain all knowledge of the problems occurred [3, 4, 8]. The Kawi Language Translator Application to Bahasa does not use rule based method because it only corrects spelling on the result of a bad sentence. Sentences generally form a series of words arranged in accordance with applicable rules. The word is the smallest element of the syntactic structure of a language, while the sentence is its greatest element [16]. The E-Translator Kawi to Bahasa can translate words, sentences and even paragraphs with a maximum word count of 400 words. It was used to translate the Kawi language, which is mostly used in Kakawin. The structure of the Kakawin's sentence is different from the Balinese, Bahasa or other language sentence structure. This is because Kakawin is a story which blended poetry that can be sung, for that the translator has an improvement process of sentence structure after successfully translated to Bahasa. Different from Android based translator application from Balinese into Bahasa using Binary Search Method, the Android based translator application from Balinese into Bahasa using Binary Search Method is a translator that translates Balinese into Bahasa. The translated source language is the Balinese which has the same sentence structure as Bahasa so it easily translates and produces a good sentence even though it translates for each word [1].

4 Conclusion

The E Translator Application of Kawi to Bahasa is made with web-based. The process begins with tokenization, stemming, word translation to database and merging the words translation into a sentence. The Stemming process uses an algorithm developed by Bobby Nazief and Mima Adriani which is enhanced by using the rules in Kawi. Bobby Nazief and Mirna Adriani's algorithm which is enhanced by Kawi language rules is resulting a level of 82.27% accuracy in stemming phase. The E Translator Application of Kawi to Bahasa which has been made is able to translate words, sentences and even paragraphs with a maximum word count of 400 words. Accuration of word translation (dictionary) is close to 100%. However, the result of the sentence translation, there are still much to be adjusted, because the actual meaning of the sentence or verse in Kawi Language often has the allegory or metaphor meaning.

References

1. Sudana, A.K.O., I.K.A. Purnawan, and N.M.R.M. Dewi, Android Based Translator of Balinese into Indonesian using Binary Search Method. *International Journal of Software Engineering and Its Applications*, **8(6): p. 165-182** (2014)
2. Afifah, N., T.B. Santoso, and M. Yuliana, Pembuatan Kamus Elektronik Kalimat Bahasa Indonesia dan Bahasa Jawa untuk Aplikasi Mobile Menggunakan Interpolation Search. *EEPIS Final Project*, (2010)
3. Francisca, J., M.M. Mia, and S.M. Rahman, Adapting rule based machine translation from english to bangla. *Indian Journal of Computer Science and Engineering (IJCSE)*, **2(3): p. 334-342** (2011)
4. Deni Pratama, I.P. and A. Muliantara, Perancangan dan Implementasi Sistem Penerjemah Teks Bahasa Inggris ke Bahasa Bali dengan Menggunakan Pendekatan Berbasis Aturan (Rule Based). *Jurnal Ilmu Komputer*, **5(1)** (2012)
5. Rohman, F. and P.W. Buana, Rancang Bangun Penerjemah Bahasa Indonesia ke Bahasa Jawa Berbasis Android. *Merpati*. **3(1)**
6. Budi Setiawan, A., P. Wira Buana, and I. Sukarsa, Aplikasi Translator Bahasa Jawa Ke Bahasa Indonesia Berbasis Android. *Merpati*, **2(3)** (2014)
7. Gehlot, A., et al., Hindi to English Transfer Based Machine Translation System. *arXiv preprint arXiv:1507.02012*, (2015)
8. Godase, A. and S. Govilkar, A Novel Approach for Rule Based Translation of English to Marathi. *Advanced Computational Intelligence: An International Journal*, **2(4)** (2015)
9. Chaer, A., *Morfologi bahasa Indonesia: pendekatan proses*, Rineka Cipta (2008)
10. Preeti and B.K. Sidhu, Natural Language Processing. *International Journal Computer Technology and Applications*, **4(5): p. 8** (2014)
11. Munot, N. and S.S. Govilkar, Rule Based Approach for Abstractive Text Summarization. *Journal of Research, Science and Technologies*, **1(6)** (2016)
12. Rajput, B.S. and A. NilayKhare, A survey of Stemming Algorithms for Information Retrieval, *IOSR Journal of Computer Engineering*, **1(17): p.76-80** (2015)
13. Jayanthi, R. and C. Jeevitha, An approach for effective text pre-processing using improved Porters stemming algorithm. *Int J Innov Sci Eng Technol*, **2(7): p.797-807** (2015)
14. Pramudita, H.R., Penerapan Algoritma Stemming Nazief & Adriani dan Similarity pada Penerimaan Judul Thesis. *Data Manajemen dan Teknologi Informasi (DASI)*, . **15(4): p. 15** (2014)
15. Asian, J., H.E. Williams, and S.M. Tahaghoghi. Stemming indonesian. in *Proceedings of the Twenty-eighth Australasian conference on Computer Science-Volume 38*. Australian Computer Society, Inc (2005)
16. Adriani, M., et al., Stemming Indonesian: A confix-stripping approach. *ACM Transactions on Asian Language Information Processing (TALIP)*, **6(4): p. 1-33** (2007)
17. Sulastra, J.S., Perancangan Penganalisis Struktur Kalimat Bahasa Indonesia dengan menggunakan Constraint-Based Formalism. *Lontar Komputer: Jurnal Ilmiah Teknologi Informasi*, **5(2)** (2014)