Balinese Script Transliteration Prediction Result of Mining Machine Using Decision Tree Method

P.A.Darwata[1], G.Indrawan[2], N.Sukajaya[3]
{pandeagusd@gmail.com[1], gindrawan@undiksha.ac.id[2]}

Department of Computer Science, Graduate Program, Ganesha University of Education, Bali, Indonesia[123]

Abstract. Mining machine is algorithm that is created to be able to automatically mining the transliteration of Balinese script, especially for words that use the character "e". Until now the existing Balinese script transliteration algorithm still distinguishes the character input of e, ie for *pepet* and for *taleng*. Based on that, the aim of this research is to develop algorithm of transliteration into mining machine by applying Decision Tree method. This is a research development that builds transliteration algorithm of Latin text into Balinese script called LBtrans. The Decision Tree method is used to determine the parameter weight of each word in the sentence entered into the system. This research begins by making the design of each case Tree and make the test data in the form of sentences that one syllable contains the character "e". The Tree and tested sentences are validated by Balinese script experts. The result of testing machine mining algorithm has 98% accuracy value.

Keywords: mining machine, transliteration, balinese script, decision tree.

1   Introduction

Many previous studies have developed transliteration applications from Latin to Balinese script in desktop, web and mobile platforms in order to preserve Balinese script to be more interesting, easy to understand and one of the most recently developed form is transliteration by implementing unicode Noto. This application is called LBTrans (Latin to Balinese Transliteration script). This application has been able to translate aksara swalalita by utilizing database to process special words of Balinese Scripts.

Balinese script has syllabic characteristics or syllables and no spaces [1]. Silabic nature in question is one letter has one sound, while the Indonesian writing uses the phonetic system, which is the combination of 2 or more letters into a sound. This becomes a problem with pre-existing transliterations. An example is the word "read" in Latin text consists of 4 letters and in Balinese script consists of 2 letters. This is the obstacle to direct transliteration of fonts to produce less perfect output. This is in previous research was overcome by applying the database, but this still has problems when there is a word with a Latin form that is similar to the meaning and form of different Balinese script. For example "mekeber" has two meanings, namely flying and lined up. If inputted in the database, still have to have a different Latin "e".

Research conducted by Ayumitha, Fevi Henda entitled "Transliteration of Latin Letters to Javanese script using Decision Tree". The algorithm by applying Decision Tree is able to overcome the syllabic character of Javanese script [2]. In the research the database has not

been applied to address the special word. Based on the results of this study, then in this research it will be implemented the Decision Tree method.

In this research there will be done some things, they are designing algorithm regulation of written Balinese script by applying artificial intelligence, which is the Decision Tree method. Second is to implement algorithm in mobile platform, which is Android. Third is to perform accuracy testing based on test documents already provided.

2   Related Work

Various applications of Latin text transliteration to the Balinese script have been developed to attract the public interest in learning Balinese script. In 2012, Balinese script has been developed by Darthi programming language by using Sartini image to represent Balinese script [3]. Still in the same year, it was also developed the implementation of Microsoft Visual C # 2010 robot (Withdrawal Bot) control software and hardware that is used Lego Mindstrom NXT and giving results of robot that can write Balinese script [4]. As the first step of this research, we have tested the application of Balinese Converter (B.Delphi) and Bot V2 (C #) Removal. Test results show that Balinese script Converter (B.Delphi) still has limitations, such as the maximum words that can be input with the correct transliteration is 100 words and the process of displaying transliteration results of more than 50 words is more than 1 minute, which is 1.46 s . while the Bot V2 Withdrawal (C #) application is only capable of processing a maximum of 25 words. Then in 2017 the Balinese script converter application was developed by Sartini was integrated with robot by I Gusti Ngurah Putra Arimbawa [5]. In the year of 2013 it is developed a transliteration of Latin text into Balinese script with the title "Windows Phone 7-based Application Design of Latin Conversion Letters into Balinese Script " by Putu Suryadana. The results of this study only covers the basic sound script and basic wianjana script [6]. In the same year, a study entitled "Java-based Algorithm Design of Latin Text Conversion to Balinese Script" was made by I Made Dwi Asmara Putra, with the scope of voice script and wianjana script [7]. In 2016 I Kadek Aditya Kurniawan conducted a research entitled "Desktop-based Application Design of Latin Texts Conversion into Balinese Scripts". This study uses Balinese Simbar unicode and there is no transliteration that includes writing for swalalita scripts.

3   Balinese Scripts

Balinese script has 47 characters, 14 characters of which are vowels (voice alphabet) and 33 characters are consonant letters (wianjana script). The voice or vowel may be subdivided into Hreswa (short voice) Dirgha characters (long voice). In writing wianjana script there are so-called *gantungan* and *gempelan*. Gantungan are a form of Balinese script when following consonants (Wianjana script) on closed syllables. These gantungan are written under the syllables or letters that precede them but the gantungan does not apply to consonants "sa" and "pa". The voice script consists of a, i, u, e, o, é. The character 'e' in the Latin text has two forms of Balinese script according to the phoneme, ie e in pepet shape and è for taleng. The existing transliteration application to date still differentiates the input form for pepet and

taleng so that it still utilizes the symbol character, the difference of its shape can be seen in Table 1.

Table 1. Table of pepet and taleng forms.

Latin Text	Pepet	taleng
e	ꦥꦶ	ꦼꦴꦩ
eling	ꦥꦶꦤꦶ (tangisan)	ꦼꦴꦩꦤꦶ (ingat)

4 Method

Mining machine that is built by using the decision tree method. The decission tree method is used to form an approximate tree by assigning weights based on the number of possible forms available from transliteration results. The Latin text which is the input of the system will be checked automatically by the system based on the flow of the tree. The result of decision tree as one of the example cases used can be seen in Figure 1.
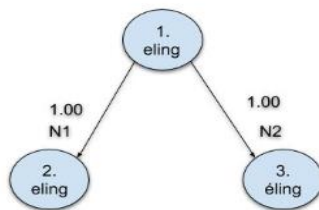

Fig. 1. Tree for eling word case.

Figure 1 shows the eling word tree (meaning: weeping) which has two possible outputs, they are eling (meaning weeping) and éling (meaning: remember). The default of that word is eling (meaning: weeping) with the pepet shape. The process begins with checking the word in the sentence entered, if in the application has one group of words in N1, then the output is eling. Conversely, if followed by N2 then the conclusion is éling (meaning: remember). The role flow of the tree can be displayed in table form like Table 2.

Table 2. Tree explaining table.

No	origin	purpose	weight	Filter	Result
1	0	1	1.00	eling	eling
2	1	2	0.50	Jerit, sigsigan, ulung, mata, medingehan, peningalan	eling
3	1	3	1.00	Leluhur, swadarma, tedun, padewekan	éling

5   Result

This research was conducted in several stages, namely making decision trees from each case used. The second form a weighting table by looking at the number of prediction results of existing case. The application of mining machines is distinguished based on the shape of the characters in vowels, characters in consonant letters and in the form of words that have the same Latin shape but the form of Balinese script is different. Examples of the application of mining machines to vowel characters in Balinese script, namely a u i e o ai, which can be seen in Figure 4. 1.
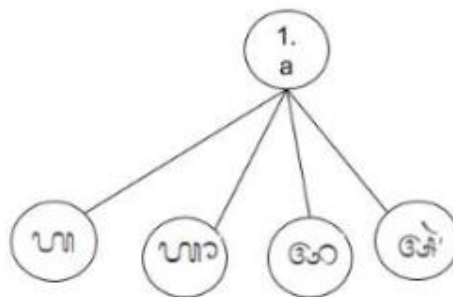

Fig. 2. Tree in a vowel character.

Figure 2 shows the Latin letters for characters that have four forms of Balinese script. The difference in shape is based on the type of word, such as the sacred keyword. Figure 2 using table form to be able to see directly the transliteration results that are most appropriate, can be seen in Table 3 Example of a decision tree can be seen in Figure 3

Table 3. Bobot table for the word mekeber case.

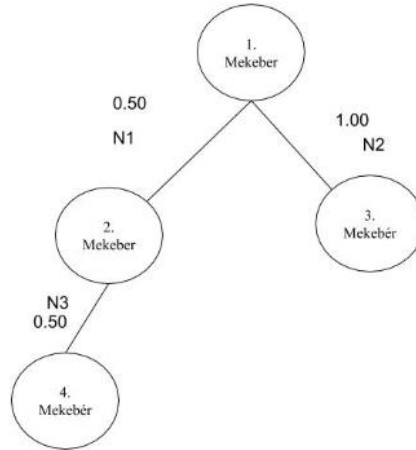| No	Source	Destiny	Bobot	Filter	Result
1	0	1	1.00	Ajeng, api, anggo, arit, dll	ᬆ
2	0	2	1.00	kadep, dll	ᬆᬑ
3	0	3	1.00	Akara tedong	ᬒᬵ
4	0	4	1.00	Arjuna, aksara dll	ᬓᬵ

Figure 3. Decission trees for the case of the word.

The decision tree in Figure 3 is described in the table form to explain the sequentially flow words to get the most appropriate transliteration results, can be seen in Table 3

Table 4. Bobot table of case words

No	Resouce	Destiny	Bobot	Filter	Result
1	0	1	1.00	*mekeber*	mekeber
2	1	2	0.50	*Langit, kedis, layangan, dedalu, nyerit, mesliweran, kupu*	mekeber
3	1	3	1.00	*Rab, baju, mobil, banten, dagang, motor, sepeda, toko, murid*	mekebér
4	2	4	0.50	*Rab,bintang*	mekebér

Tests were carried out on 750 cases which included 400 word forms and 350 sentences with Latin words that have different forms of Balinese script. Testing according to the case shows 100% accuracy value. Details of the test results and the validated test documents are in the attachment. The most influential types of Balinese scripts in Balinese language are nominated by the letter "e", while the differences in Balinese characters are for the use of the vocals the name of a holy place. The numbers of vocal character tests, consonant characters and used word forms are described in Table 5

Table 5. Number of test data.

No	Type	input	Probabilitas output	case	unit	result (%)
1	Vowel characters	1	2 s.d 4	350	word	100

2	Consosnant character	1	2	50	word	100
4	word	1	2 s.d 3	350	sentence	100

The decision tree in the vowel character of the letter a has 4 forms of Balinese script, namely a according to hanacaraka, a tedong, akara tedong. Vocal e has two forms in general, namely taleng and pepet. Taleng has the form of hanacaraka and aikara. U vocals have a form of Balinese script in the form of suku, ilut and ukara, ukara tedong. The forms of Balinese vowel are in four forms, namely ulu, ulu sari, ikara, ikara tedong and vocal o have two forms, namely o taleng tedong and okara.

6   Conclusion

Based on the research that has been done, namely "Result of Prediction Balinese Script Transliteration with Mining Machine Using Decision Tree Method" as for the conclusions obtained, among others.
1. The characters that influence the Balinese script from the same Latin text are all vowels and 38.46% consonant letters, namely b, d, g, k, l, n, p, r, s, t.
2. The mining machine with decision trees can provide transliteration results for Balinese characters with the same Latin text input forms and different forms of Balinese script.

References

[1] I. M. Suatjana and I. B. A. Sudewa, "Babad Bali - Aksara Bali," 2000. [Online]. Available: https://www.babadbali.com/aksarabali/aksara_bali.htm.

[2] F. H. Ayumitha, "Transliterasi huruf latin ke dalam aksara Jawa dengan menggunakan decision tree," Universitas Islam Negeri (UIN) Maulana Malik Ibrahim Malang, 2014.

[3] I. D. A. M. Sartini, M. W. A. Kesiman, and I. G. M. Darmawiguna, "Pengembangan Text to Digital Image Converter Untuk Dokumen Aksara Bali," Pros. Semin. Nas. Pendidik. Tek. Inform. (SENAPATI 2012), pp. 280–296, 2012.

[4] I. G. N. P. Arimbawa, M. W. A. Kesiman, and I. G. M. Darmawiguna, "Pengembangan Robot Penulis Karakter Aksara Bali Berbasis NXT-G dengan Lego Mindstorm NXT," Pros. Semin. Nas. Pendidik. Tek. Informartika, pp. 160 – 174, 2012.

[5] I. G. N. P. Arimbawa, G. Indrawan, and I. N. Sukajaya, "PENGEMBANGAN ROBOT PENULIS AKSARA BALI BERBASIS LEGO MINDSTROM NXT," Proceeding SENARI; Vol 5 Senari 2017, vol. 5, pp. 8–15, 2017.

[6] P. SURYADANA, "Perancangan Aplikasi Konversi Huruf Latin Menjadi Aksara Bali Berbasis Windows Phone 7," STIKOM Bali, 2013.

[7] I. M. D. A. PUTRA, "Rancang Bangun Algoritma Konversi Huruf Latin Ke Aksara Bali Berbasis Java," STIKOM Bali, 2013.