Balinese Latin Text Becomes Aksara Bali Using Rule Base Method

I Putu Agus Eka Darma Udayana[1], Made Sudarma[2], and I Nyoman Satya Kumara[3]

[1](Magister Program of Electrical and Computer Engineering, Udayana University Graduate Program, Sudirman Denpasar, Bali-Indonesia)
[2,3](Department of Electrical and Computer Engineering, Faculty of Engineering, Udayana University Jimbaran Campus, Bali-Indonesia)

Abstract: *Lontar is one of the cultural heritage which has information about history of Balinese civilization in the past away. Problems encountered today are that the lontar are not well maintained. While, the lontar that used as letter of Balinese alphabet will be worn out soon, because it doesn't have good endurance for long time. The exploiting of technology is one of media that can be used as a solution to solve the problems. The digitalization process of letter Balinese alphabet can be done by rewrite the script of that lontar in Balinese alphabet by using translation script with Rule Base and Levenshtein Distance Approach. The exploiting of technology will make the lontar becoming digital form and it won't be worn out when the lontar is kept safe for long time and information that consisted in lontar can be protected for long time.*

Keywords: *Translitation, rule base, levenshtein distance, aksara bali.*

I. INTRODUCTION

Aksara Bali (Balinese Alphabet) is a symbol visual system that showed in a media, it has function to uncover the elements that expressing a language(Sudarma et al., 2016). Based on the pronouncing, *Aksara* Bali (Balinese Alphabet) devided itno some groups called *warga aksara*. This division is based by *Sansekerta Panini's principle* such as; *warga kanthya, talawya, murdhanya, dantya dan osthya* (Surada, 2007).

That's not easy for *Aksara* Bali (Balinese Alphabet) to hold in every development of new era. *Lontar* is one of traditional documentation media that has been endowed by the Balinese ancestors which has information about all culture in Bali by using *Aksara* Bali (Balinese alphabet) as the mediator (Ginarsa, 1976). So that's why, the *Aksara* Bali (Balinese alphabet) that written in *lontar* must be protected. But, there will be a lot of challenge that should be faced in trying this preservation. One of them is limited financing by the government to take care all these documents. While, that *lontar* which is used as media of *aksara* Bali (Balinese alphabet) text doesn't have good endurance to keep for long time.

The exploiting of this technology is one of great media to be used in preserve the Balinese culture which has in *lontar* written in *aksara* Bali (Balinese alphabet), this is appropriate with S.K. nomor 179 tahun 1995 and circulation letter no 01/1995 that explained to the society in using *aksara* Bali (Balinese Alphabet) (Mahendra et al., 1996). Other than that, one of the efforts from the government in preserve the Balinese culture in form of *lontar* is by doing the translation the *lontar* from *aksara* Bali (Balinese alphabet) become text in Balinese language.

For preserving the Balinese culture in form of *lontar* is not enough if you translate all the *lontar* only. The real *lontar* which is written in *lontar*s leaf in Balinese alphabet should be become modern. The step of Balinese alphabet being modern in that *lontar* is done by rewriting the result of *lontar* translation in Balinese language Latin text become Balinese alphabet in digital form or computerization. The result of rewriting process from Latin text in Balinese language become Balinese alphabet called translation

The same papers explain about translation had been done by using *Syllabification Approach* method (Joshi et al., 2013). In this paper, the method is used for solving the problems which is faced the similarity of two different languages. There are some other method that can be used for solving the translation problem, those methods are *Rule Base, Decision Tree*dan *Hybrid* (Bhalla et al., 2013)(Kaur et al., 2014)(Kaur et al., 2012). There is a paper which do reviewing about the solving of translation process, in the paper is explained the great method for doing translation with structure case in different language called *Rule Base*(Kaur and Kaur, 2014). Based on the support of paper and problems in doing translation Latin text become Balinese alphabet, so this research uses *Rule-Base* method which combined with the application of spell character uses Levenshtein Distance method on preprocessing step. This combination method is used to decrease the level of mistakes, because there are some words in Latin of Balinese language which has homonym meaning, so that will often do mistakes in translation progress when using Rule Base method.

Need of the Study

Lontar is a Balinese cultural heritage which has information about the history of Balinese civilization in the past away. Till today, that cultural heritage still tries to be kept its preservation by the society although it start to lose the application and comprehension about letter of Balinese alphabet in Balinese's society from generation to the next generation. As time passes the lontar will be worn out soon, because it doesn't have good endurance for long time. This technology is needed to rewrite existing manuscripts lontar in digital form, so that the manuscript lontar can hold for a long time.

II. FUNDAMENTAL THEORY

Aksara Bali

Balinese alphabet is one of Balinese legacy by Balinese people which is valuable because it has so many culture values of society in Bali (Darma et al., 2015). The history of Balinese alphabet has close relation with the alphabet in India. Balinese alphabet is from India when Hindu and Buddha came to Indonesia (Sudarma and Surya, 2014). *Aksara* (Alphabet) can be meant as a visual symbol system which written on a media, it has function to shows the elements which express the language (Trieha, 2014). Based on its pronunciation *aksara Bali* (Balinese alphabet) is divided become some groups which is called *warga akasara* (Surada, 2007). Based on the *Pesamuhan Agung* which is done in 1963 that have decided kinds of letters which is made to write or presenting Balinese alphabet by Latin text (Sudarma and Sutramiani, 2014). *Aksara suara* is basically as same as vocal letter in Latin. Table 1 is *aksara* group which include in *Aksara suara* (Sutramiani et al., 2015). In Balinese alphabet (*aksara* Bali), *AksaraWianjana* is also called consonant. Although, its written without vocal letter in the way of writing. Table 2 is the classification of *AksaraWianjana* based on *Warga Aksara*.

Table-1: Aksara Suara

Number	Balinese script	Latin script
ᬳ		a
ᬇ		ê
ᬈ		i
ᬉ		u
ᬏ		e
ᬑ		o

Table-2: Aksara Wianjana

Number	Balinese script	Latin script	Number	Balinese script	Latin script
		h / a			l
		n			m
		c			g
		r			b
		k			ng
		d			p
		t			j
		s			y
		w			ny

Lontar

One of the heritages from the ancestors in the Indonesia Archipelago is a text which is called *lontar*. *Lontar* (Java: *ron tal, "daun tal"(*tal leaf*))* is siwalan leaf or tal (*Borassus flabellifer* atau *palmyra*) that had been dried. *Lontar* tree (*Borassus flabellifer*) is a kind of *palma*(palm) which grow in South East and South of Asia. In Indonesia Archipelago, the relic of *lontar* had been found in some area such as in Bali, Java, Lombok and Sulawesi. That *Lontar* is used as place for writing before paper (Sudarma, 2015).

Figure1: Capture Lontar Babad Brahmana
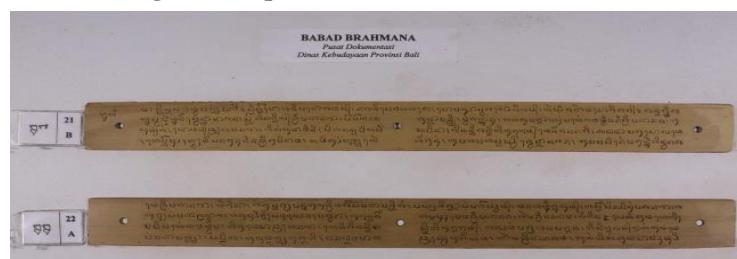
Other than *Lontar*, there was also another media which was used by our ancestors as writing media before paper like nipa's leaf, wood's skin and goat's skin. The old script is one of heritage by our ancestor which has important value. For now, The dominant old script which are in Indonesia are kept by a person and the other are kept in government instance.

Transliteration

Transliteration can be called as a progress of changing words in one language into the other language and still have the same structure (Kaur and Kaur, 2014). Transliteration can be as a words conversion progress which is written in one language into another language too by keeping the pronunciation. In this case, you will

hear about transliteration and translation. Transliteration can be used to translate the meaning and the great technique from language, but translation will interpret from the text or communicate the same message from original language.

Tokenizing

Tokenizing is a segmentation progress or divide a text documents become token, it means the structure of characters which represent the words by tokenizer (Indranandita et al., 2011). Tokenizer divides a string into some tokens based on the certain characters. In tokenizing progress, the characters usually divided as guidance for doing the segmentation sentence become token such as spasi, tab, new row (white space character). This is the example of tokenizing in the sentence "*tiang sampun ngajeng*".

Input   : *tiang sampun ngajeng*
Output   : *[tiang] [sampun] [ngajeng]*

Rule Base

Rule Base method is a technique that used basic language rules in doing transliteration progress (Kaur and Kaur, 2014). Other than rules, dictionary data is needed too for every word in two languages. So, every word is translated one by one, and then arranged again based on the basic language rule. The system depends on the linguistic knowledge. Their benefits are able to analyze till syntaxes level and semantic in deeply. The weakness needs good language knowledge and impossible to write the rule in all languages.

III. RESEARCH METHOD

Design and Implementation

Transliteration scheme system Latin text become *Aksara Bali* is arranged based on some steps, they are Latin text input progress, saving the result of tokenization data, spell checker, transliteration and show the result of translation in *Aksara* Bali. Picture 2 is visualization progresses which happen in transliteration Latin text become *Aksara* Bali in block diagram.

Figure2: Block Diagram of System
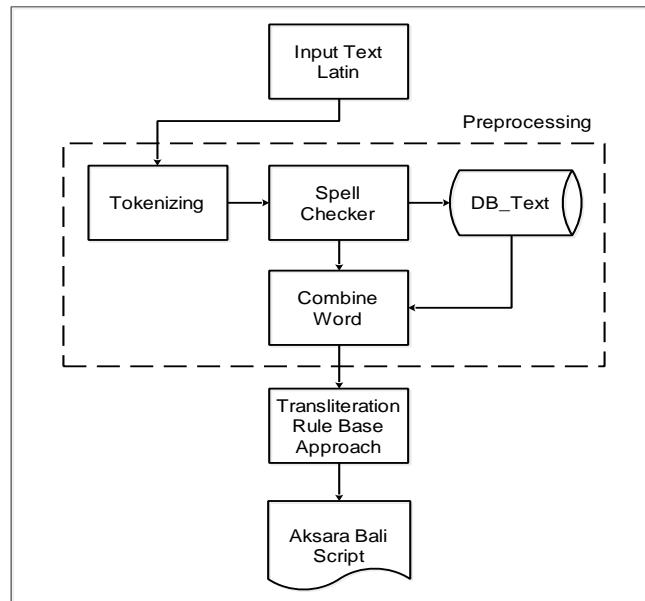
Tokenizing Process

Tokenizing progress which is done on preprocessing transliteration Latin text becomes Aksara Bali. In this progress, Latin text which is inputted will be read by system. When all texts have read by system then that texts will be divided into some groups and the space will be deleted. On the next progress, every words which have past the deleting the space progress will be given unique sign. That sign will be used to arrange word when there will be done the fixing text progress. When all the progresses have done well, all those words will be saved to go to the next progress.

Spell Checker

Spell Checker progress is done to do the checking homonym words which are in Balinese language. Spell checker progress will be done by matching every word in data base which is saved by the system. On the first progress will be done theinitializationthe total of words in the text which is inputted by the system.Those words will be inputted into Levenshtein Distance algorithm to do counting the space between source words and destination words in data base. The word which doesn't have space will be showed as suggestion word which become the purpose of the user and will be included with the meaning of that suggestion word by using direct

mapping. The suggestion word can be meant as wordthat has more than one meaning or homonym, so, when it comes in to transliteration progress, there will not happen mistake in the *Aksara* Bali output. This is because of the writing of *Aksara* Bali, the word which notabene has same writing and may different result in Balinese language. Talking about spell checker, this progress will continue well from the checking word or checking the space of word by Levenshtein Distance method till all words which has inputted complete.

Levenshtein Distance

In this progress uses *Levenshtein Distance*in approach, which is done the equivalent in every word that input with word that is in transliteration database system. Every word that come into this progress will be faced with some treatments, those treatments which in *Levenshtein Distance* function are changing the character, increasing character and deleting character. In changing character operation, a character will be changed into other character. In increasing character operation will be increased character into string. In deleting character operation is done by deleting character in the string. Based on that condition, if the last word from the first text more or less from the second text, so it will be done the deleting and increasing string, if the string on the first text and second text has difference, so there will be changing progress. Those progresses will always recur till all string in second text which is equivalent come into Levenshtein Distance progress.

The result of every string equivalent function from the text which compare d will has appropriate value with a lot of *Levenshtein Distance* function is happen in those second word, and if all string from equivalent word are same, so automatically there will be no Levenshtein Distance progress, so, the space between the words is zero(nol)

Rule Base Approach

The main method in transliteration progress on this system is the using of rule base approach. Every character which is in user input will be done the checking by rule which has decided to change the Latin text to *Aksara* Bali. When the first character come into rule base progress, that character will be done the checking for the first rule. When that character is not in the first rule, that character will be done the checking by the other rule in the system. When the Latin character has come into one of the rule which has prepared by the system, then that character will be saved in work memory, and it will be made become same with the next result of character.

IV. RESEARCH METHOD

White Box Testing

The White Box Testing in this research is done to know the result level of system to the function. This trial will be done by the application user by doing the transliteration progress Latin text to *Aksara* Bali. The system will be tested in every logic that done well like condition or repetition, whitebox trial will be done in spell checker process and transliteration in method rule base of condition or repetition and levenshtein distance method.

Table3: Flowgraph Rule Base Process

No	Name of Process/Function
1	Input Latin text in Balinese language
2	Latin alphabet is not in rule
3	Latin alphabet is in rule
4	System doesn't write the result of transliteration alphabet
5	System writes the result of transliteration alphabet.
6	System gives addition attribute on alphabet for next transliteration alphabet
7	Done

Table3 is flowgraph from transliteration process using Rule Base method, that the user will input Latin text in Balinese language, and then the system will do transliteration in every alphabet that consist in Balinese language Latin text to *Aksara* Bali form. These are the results of edge and node for this process; it can be seen in picture 3.
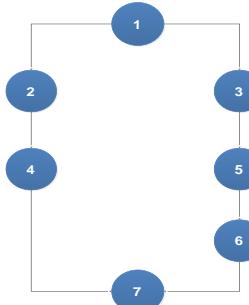

Figure3: Edge and Node in Transliteration Process

From edge and node in transliteration process that can be seen on figure 3, it can be formed into two ways or path which is in transliteration process such as 1 – 2 – 4 – 7 and 1 – 3 – 5 – 6 – 7 that will be tested on table4.

Table4: The Result Of Testing In Transliteration Process

Path	The Output That Is Wanted	Output System	Explantion
1	The application doesn't write the result of transliteration Latin alphabet in Balinese language.	The application doesn't write the result of transliteration Latin alphabet in Balinese language.	Valid
2	The application writes the result of transliteration in Latin alphabet	The application writes the result of transliteration in Latin alphabet	Valid

From table4 can be seen that all the transliteration process results valid hope to the output system which has been tested.

Testing of Transliteration

The next testing is about how the application which had been developed doing transliteration process by using Rule Base method. Figure 5 is figure from system which will be used for transliteration.
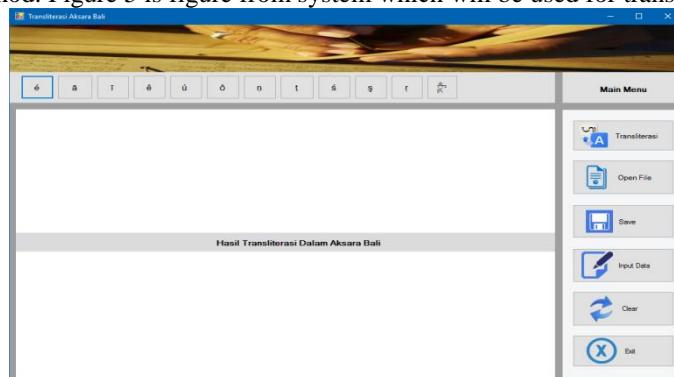
Figure5: Main Menu Interface System

In homepage of the application has some menu to support the transliteration process. For inputting the translation bundle of *lontar* uses open file menu.
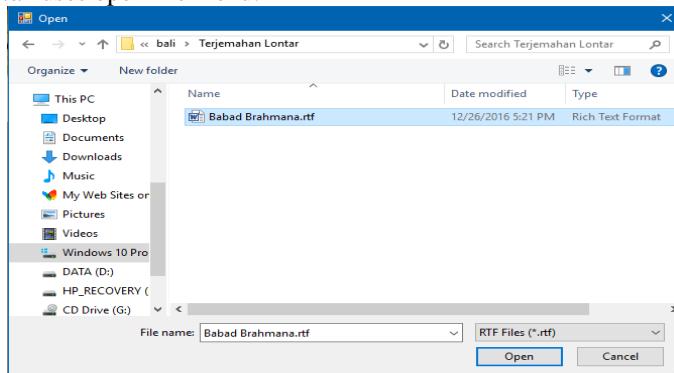
Figure6: Open Tranlation File of Lontar

After choose the *lontar* translation document which will rewrite in *aksara* Bali. So, the application will show the home page likes in figure 7.
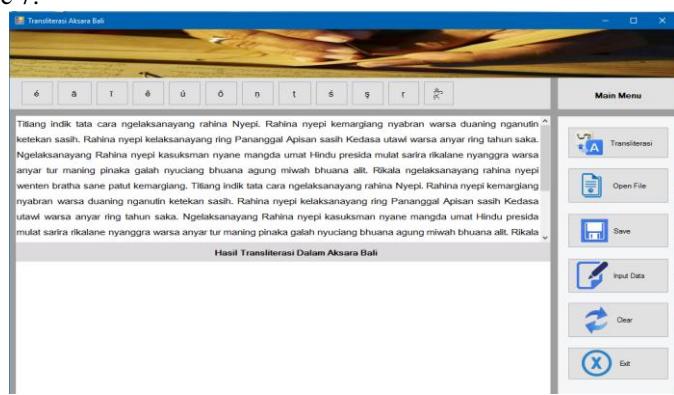

Figure7: Import The Translation Text Lontar

In figure 7, *Lontar* transalation text has been imported to transliteration application. The next step is doing the change process of *aksara* from translation *lontar* to *aksara* Bali by pushing transliteration menu.
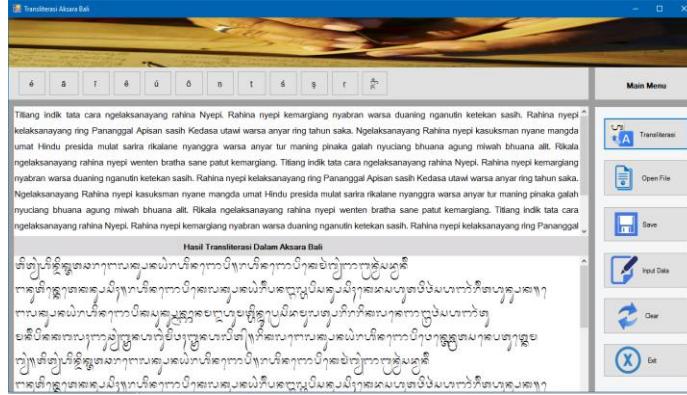

Figure8: The Result of Translation

In figure 8 is seen that the result of translation latin text in *lontar* has been completed to do transliteration or rewrite in *aksara* Bali form. To save the result of *aksara* Bali click save menu.
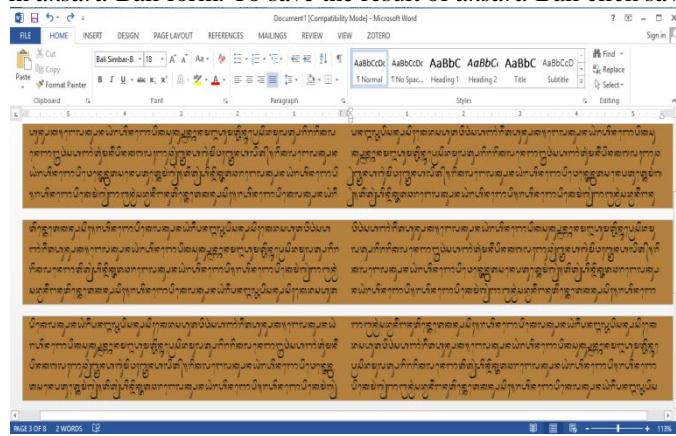

Figure9: The Result of Rewrite Lontar in Document Form

Figure 9 is the result of transliteration translation *lontar* document in *aksara* Bali. For the next testing, the result of this document will be made same to the original document of *lontar* to assess the accuracy of transliteration application in using rule base method. In this testing is used 10 pieces of *lontar* documents to match with the result of transliteration documents. The result of that testing is the accuracy value which is produced from the result of transliteration by using rule base method and supported with levenshtein distancemethod is 90.67%. From the result has 9.33% incorrect, where the mistakes is caused of the words in Balinese language which are translated with special writing in *aksara* Bali.

V. CONCLUCION

The conclusion that based form the result which is got that transliteration application Latin text in Balinese language into *aksara* Bali by using Rule Base method is able to rewrite the transaltation of latin text of *lontar* into *aksara* Bali as like in the original *lontar* document with 90,67% accuracy. Based on the result of white box testing, all the necessity which are in the system have been filled or fulfilled from the system. The next study can be done the addition word data bank in Balinese language which has special writing in *aksara* Bali or making the best method for higher result.

Limitations

One shortcoming in this research is when words in the manuscript *lontar*use special characters and these words are not listed on the system database then the system will be wrong in representing Balinese script transliteration.

Future Enhancement

The next study is requires more definitions of words with special Balinese characters for transliteration acceleration to have a higher percentage.

V. REFERENCES

[1] M. Sudarma, S. Ariyani, and M. Artana. Balinese Script's Character Reconstruction Using Linear Discriminant Analysis. Indones. J. Electr. Eng. Comput. Sci. 2016; 4(2); 479-485.
[2] I.M. Surada. Kamus Sansekerta-Indonesia. Surabaya: Paramitha. 2007.

[3]   K. Ginarsa. The Palm (Palmyra) Palm. Edisi Dua Bahasa. Denpasar. 1976.

[4]   N. Mahendra, M. B. Suasta, I. W. O. Granoka, and I. W. Japa. Pembinaan Bahasa Aksara Dan Sastra Bali Pedoman Penulisan Papan Nama Dengan Aksara Bali. Bali: Dinas Kebudayaan Propinsi Daerah Tingkat Bali. 1996.

[5]   H. Joshi, A. Bhatt, and H. Patel. Transliterated Search using Syllabification Approach. Forum for Information Retrieval Evaluation. 2013.

[6]   D. Bhalla, N. Joshi, and I. Mathur. Rule Based Transliteration Scheme for English to Punjabi. Int. J. Nat. Lang. Comput. 2013; 2(2); 67–73.

[7]   V. Kaur, A. kaur Sarao, and J. Singh. Hybrid Approach for Hindi to English Transliteration System for Proper Nouns. IJCSIT. 2014; 5(5); 6361-6366.

[8]   E. S. Kaur, M. R. Kaur, and E. N. Bhalla. Architecture of the Transliteration System from ENGLISH to PUNJABI using Hybrid approach. International Journal of Computer Science and Information Technology & Security (IJCSITS). 2012; 2(2); 482-485.

[9]   D. Kaur and E. R. Kaur. English To Punjabi Script Converter System For Proper Nouns Using Hybrid Approach. IJARCSSE. 2014; 4(7); 551–554.

[10] Darma, I. K. Putra, and M. Sudarma. Ekstraksi Fitur Aksara Bali Menggunakan Metode Zoning. Maj. Ilm. Teknol. Elektro. 2015; 14(2); 44-49.

[11] M. Sudarma and I. W. A. Surya. The Identification of Balinese Scripts' Characters based on Semantic Feature and K Nearest Neighbor. Int. J. Comput. Appl. 2014; 91(4); 14-18.

[12] http://ensiklo.com/2014/09/istilah-aksara-berasal-dari-bahasa-sanskerta-yang-berarti-tidak-musnah, cited: September 12, 2014.

[13] M. Sudarma and N. P. Sutramiani. The Thinning Zhang-Suen Application Method in the Image of Balinese Scripts on the Papyrus. Int. J. Comput. Appl. 2014; 91(1); 9-13.

[14] N. P. Sutramiani, Ik. G. Darmaputra, and M. Sudarma. Local Adaptive Thresholding Pada Preprocessing Citra Lontar Aksara Bali. Maj. Ilm. Teknol. Elektro J. Electr. Technol. 2015; 14(1); 27-30.

[15] M. Sudarma. Identifying of the Space Color CIELab for the Balinese Papyrus Characters. Int. J. Soft Comput. 2015; 11(2). 64-69.

[16] Indranandita, B. Susanto, and A. Rahmat. Sistem Klasifikasi dan Pencarian Jurnal Dengan Menggunakan Metode Naive Bayes dan Vector Space Model. J. Inform. 2011; 4(2).

[17] N. M. M. Adriyani. Implementasi Algoritma Levenshtein Distance Dan Metode Empiris Untuk Menampilkan Saran Perbaikan Kesalahan Pengetikan Dokumen Berbahasa Indonesia. JELIKU-J. Elektron. Ilmu Komput. Univ. Udayana. 2012; 1(1).