Web-based Implementation of Finite State Automata Method on Lyrics Recognition System of Balinese Song "*Pupuh*"

A. A. K Oka Sudana
Department of Information Technology, Udayana University, Bali, Indonesia

Putu Wira Buana
Department of Information Technology, Udayana University, Bali, Indonesia

Titah Wulandari
Department of Information Technology, Udayana University, Bali, Indonesia

ABSTRACT
Bali has a various kinds of arts, one of them is Balinese Song (Indonesian: Tembang). Tembang is literary work presented in the form of vocal and instrumental sound. Implementation of technology information related to the local arts are less developed. Societal perspectives consider the local arts are not give any influence to the modern life causes decrease in history preservation. Based this case, Lyrics Recognition System of Balinese Song *Pupuh* is made using Finite State Automata Method to separate syllables. The system is able to recognize lyrics of *Pupuh* based on the rule of Padalingsa *Pupuh* such as the number of lines in one stanza, the number of syllables in each line and the last vowel in each line. The system is expected to be an e-learning for elementary, middle-school students and general.

General Terms
Implementation of Natural Languange Processing, Finite State Automata, Web-based Application, Parsing Syllables, Balinese Traditional Topic

Keywords
Pupuh, Finite State Automata, syllables, e-learning, web application

1. INTRODUCTION
Bali has a various kinds of arts and one of them is Tembang. Tembang is a literary work that expressed through the sound and contains the values of religion. There are several types of Balinese Tembang, one of them is *Pupuh*. *Pupuh* is a literary work that presented through a song and contains values of religious teachings, which is tied by the rules of number of lines, number of syllables and the last vowel. That rule is called "Padalingsa *Pupuh*" [1]. *Pupuh* is an artistic cultural heritage that should be preserved, however in this era, *Pupuh* is known by among the elders only. Young people especially also need to know more about *Pupuh*. Solutions that can be taken is to introduce *Pupuh* through technology (web). Recognition System of *Pupuh* produce some points that must be prepared, such as literature review, *Pupuh*'s types and examples, and rules of *Pupuh*. This system intended to compose *Pupuh*'s lyrics.

There are many studies of Finite State Automata, one of them is David L. Waltz and Jordan B. Pollack's journal in 1985 with the title Massively Parallel Parsing: A Strongly Interactive Model of Natural Language Interpretation regarding parsing words parallel with multiple meanings depend on the sentence used [2]. Finite Automata is implemented in the various fields of linguistics, one of them is George Anton journal entitled "Nonlinear Morphology multitiered Multitape Using Finite Automata: A Case Study on the Syriac and Arabic" that discussed about form of Arabic word and Syriac. Kessikbayeva, Gulshat & Cicekli, Ilyas, (2014) also conduct research related to FSA entitled Rule Based morphological analyzer of Kazakh Language. The study presented the details of computational analysis of the Kazakh language that attached to the rule [3]. Method of parsing words in Indonesia was made by Anny Yuniarti 2004 entitled Dictionary Arabic-Indonesian Online with Syllabics Solving Method Used Parsing "which discusses how separating the syllables in Arabic based on web [4]. Syllable parsing system has also been made by Sigit Wasista and Novita Astin with output sound (text to speech) entitled "Algorithms system of Indonesian Text Reader uses FSA method (Finite State Automata)" in 2010. The journal explains how they classify the syllable of Indonesian word using Finite State Automata method [5]. Related to the Balinese languange, research on Balinese dictionary ever made by A.A.K Oka Sudana et.al (2014) entitled Balinese Translator Android Based Indonesian into using Binary Search Method. This study discusses the language translator application Balinese to Indonesian using the binary search method [6].

Method of parsing of syllables uses Balinese language has not been made before, this is seen from some of the research that has been done, and then the study take the title Web-Based Implementation of Finite State Automata Method on Lyrics Recognition System of Balinese Song "*Pupuh*". The system is designed for detecting the text of *Pupuh* lyrics. Particular types of *Pupuh* can be identified by counting total of lines, syllables and the last vowel in each line. There are features of the application such as wrong input warning of lyrics according to the type of *Pupuh* and the Padalingsa rules. The system is expected to preserve the arts of Bali and an e-learning for young generation and students to get to know *Pupuh*.

2. LITERATURE REVIEW
2.1 Pupuh
Dharma Gita is the Hinduism song, which is part of a Sad dharma as a liability for preservation of Hindu culture. Dharma Gita in its development in Bali is a song literature. I G.B. Sugriva (alm.) grouping Balinese songs into four parts: (1) Gegendingan (Javanese: dolanan), (2) *Pupuh* (Sekar Alit) (3) Sekar Madia, (4) Kakawin [1].

Sekar Alit is Balinese song that can be called "Macapat" which contains religious teachings. *Pupuh* is bound to the rules of the lines in one stanza (guru gatra), the number of syllables in each line (guru wilang), anchoring the sound (lingsa) last vowel on each line (guru ding-dong). *Pupuh* in

Bali, known as the original macapat are *Pupuh* Sinom, Semarandhana, Pucung, Pangkur, Ginada, Ginanti, Durma, Dangdang Gula, Maskumambang, Mijil [1].

2.2 Finite State Automata

Automata is derived from the Greek, Automatos, which means something that works automatically like a machine. Automata is part of Natural Languange Processing. Finite Automata is an abstract machine in the form of mathematical model system that processes the inputs and outputs that can recognize regular language. Finite state machine has been used in many fields of computational linguistics. In Linguistic, Automata is easy to use because it can provide solutions to problems that are often encountered in the study of language and recognize string sequences. Each Finite State Automata has a finite set of (finite state) consisting of one state that serves as the initial state that can be represented as q0 and some state as the final state. FSA also has a set of input symbols, and a transition function that determines the next status of each pair status and an input symbol [7].

3. METHODOLOGY

This study uses The Finite State Automata (FSA). Finite state has been used in many fields of computational linguistics. In language, Automata are easy to use because it can provide solutions for any problems in linguistic [8]. There are three stages to parsing the syllable as the Finite State diagram as follows:

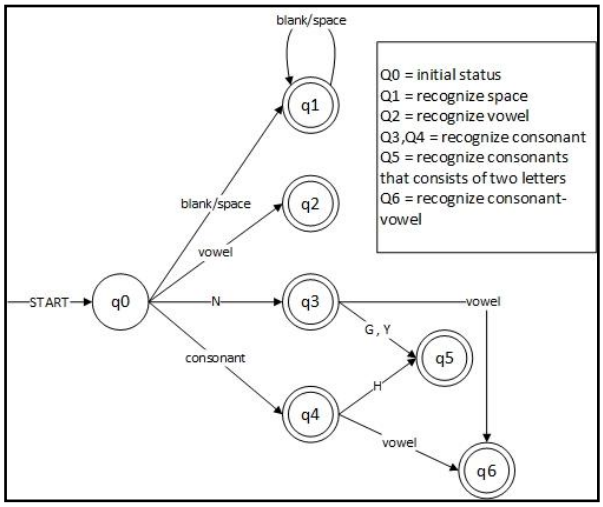

Fig. 1: FSA Stage 1

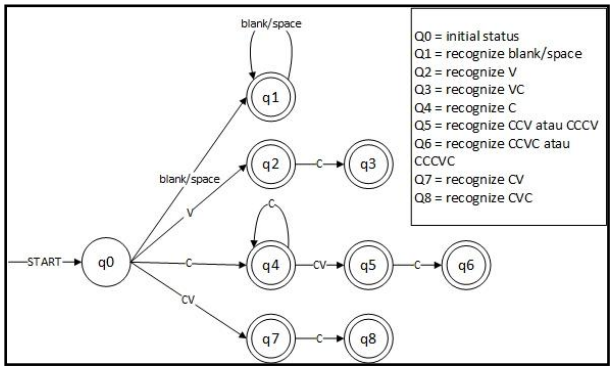

Fig. 2: FSA Stage 2

Each Finite State Automata has a finite set consists of one status that serves as a final state and the initial state that can be represented as q0 [9]. FSA also has a set of input symbols, and a transition function that determines the next status of each pair status [7]. Stage 1 describes the displacement state with q0 as an initial status and identify patterns of consonants (C), vocal (V), consonants and consonant-vowel through two states. Results of stage 1 will be input at the next stage. Implementation of the FSA in coding uses RegEx (Regular Expression) is an effective solution for the implementation of Automata which contains many iterations series [10].

FSA stage 2 recognize patterns VC, CCV, CCVC / CCCVC, CCV / CCCV and CVC which state with double lines of the circle is the state results. FSA stage 3 will recognize more complex patterns such as: VCC, CVCC, CCVCC, and CCV.

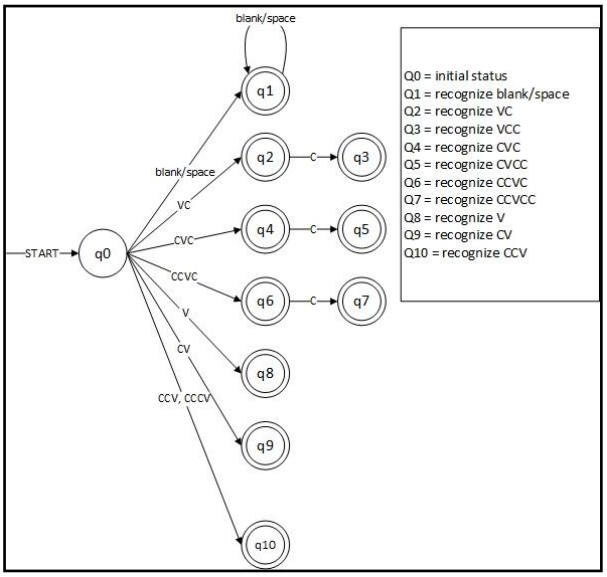

Fig. 3: FSA Stage 3

Stage 3 is the final stage in which the sum of the syllables will be determined based on the pattern of syllables using the FSA. Examples of the implementation of the FSA in the word recognition is described in the form of state diagram. Explanation displacement state is illustrated in Figure 4 below.

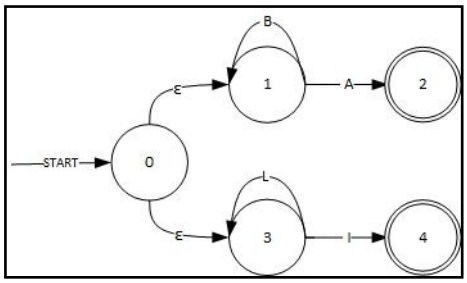

Fig. 4: Example of FSA Transition

Figure 4 illustrates the displacement state with the input of the word "Bali". The letters have been classified into consonants and vowels so that groups of consonants consists of several kinds, as well as vowels. State diagram designed in the system is a type of Non-deterministic Finite Automata, where ε symbol on the diagram indicates input that starts from the initial state is not determined only one type of characters, in the sense that the input is a set of consonants or vowels. Input can be the letter B | L depending on the input. Consonants are stored in state 1 and 3. The system checks the next character, if the character is a vowel A and I then combined the character of a state previously stored at the end of the state is

state 2 and 4. The system has been able to recognize the pattern of consonant-vowel is "ba-li ". Another example is the word "Tembang". The first phase state diagram described in Figure 5 below.

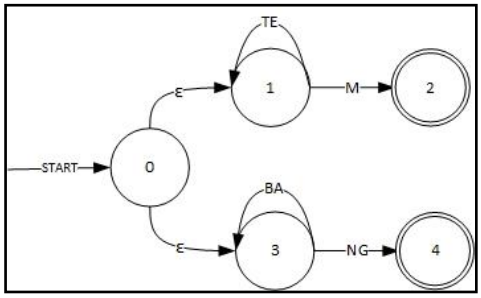

Fig. 5. Example of Word Transition "Tembang"

Figure 5 is an example of the displacement state recognition word "Tembang". Similar to the previous diagram, where the symbol ε not determine absolute input value and can be consonant or vowel in the initial state. State 2 is a state final consonant-vowel pattern recognition is "te-". State 3 is the introduction of the character "m" as a single consonant. State 5 is the introduction of the character "ba" and recognize the combined state 7 consonants of "ng". The resulting output is parsing "te-m-ba-ng". That separation is not according to the rules of language. FSA transition phase two needs to be done to enhance the syllables. Output at stage one becomes a new input to the second stage. Figure 6 illustrates the state diagram for second phase.

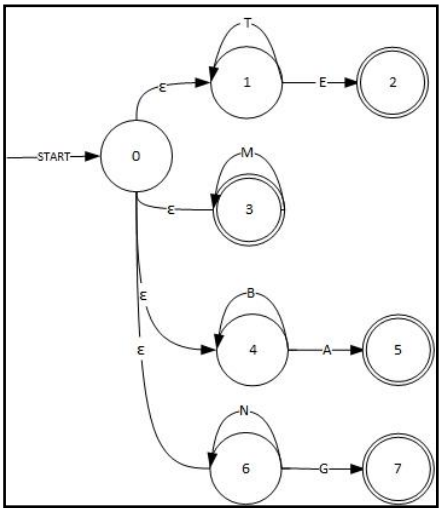

Fig. 6. Example of Word Transition "Tembang" Stage 2

The second stage is the stage of processing the results of phase 1. The word "te" and "ba" has been classified as a pattern of C-V system. When the system checks the next character is a consonant, the results of new word patterns stored in two states, as well as the syllable "ba" which met with a combination of consonants ("ng") which is considered as a group consonants. The output that resulted from the separation of the syllables of "tem-bang" and is the final stage character recognition with the FSA.

4. RESULTS

Lyrics of *Pupuh* recognition system are divided into two as follow the implementation of parsing syllable for validation and determination of *Pupuh*. The system is based on web, so that can be accessed via computer (PC). There are two features in this system, such as: the first feature serves as a validation if the user input is in conformity with the rules of *Pupuh* . Users are required to select the type of *Pupuh* that will be checked first. The second menu functions to determine the type of *Pupuh* which approach with input and give feedback correction. The Features in the search menu of types of *Pupuh* that closest have processing such as parsing syllable, then counting the lines, counting the syllables and the last vowel.

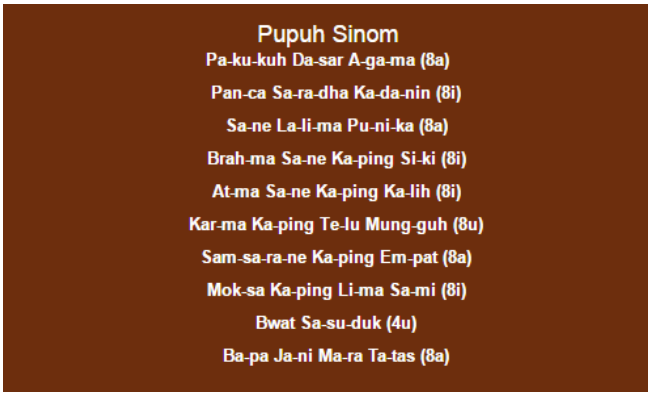

Fig. 7: Result Execution

The system compares rules of *padalingsa Pupuh* with input from users. If there is wrong input, it will display a warning and displays the best advice what the user types in accordance with the rules of *Pupuh* (see Fig. 8). Results of execution for both features in the lyrics of *Pupuh* recognition system display the types of *Pupuh*, parsing syllables, total of syllables and the last vowel input by the user.

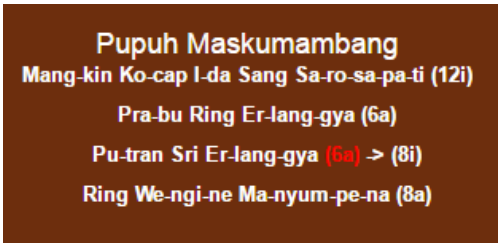

Fig. 8: Wrong Input Warning

Table 1 is example results of parsing syllables test. Parsing syllables testing is done by entering one hundred sentences in Balinese.

Table 1. Table Parsing Syllables

No	Input	Output	Consonant-Vowel Parsing
1	Mangkin kocap Ida Sang Sarosapati	Mang-kin ko-cap I-da Sang Sa-ro-sa-pa-ti	CVCC-CVC CV-CVC V-CV CVCC CV-CV-CV-CV-CV
2	Prabu ring Erlanggya	Pra-bu ring Er-lang-gya	CCV-CV CVCC VC-CVCC-CCV
3	Wengine manyumpena	We-ngi-ne ma-nyum-pe-na	CV-CCV-CV CV-CCVC-CV-CV
4	Ne dadi prabotang sai	Ne da-di pra-bo-tang sa-i	CV CV-CV CCV-CV-CVCC CV-V
5	Panca sradha	Pan-ca sra-dha ka-da-	CVC-CV

	kadanin	nin	CCV-CCV CV-CV-CVC
6	Om swastyastu	Om swas-tyas-tu	VC CCVC-CCVC-CV
7	Swadharmaning Wesya	Swa-dhar-ma-ning We-sya	CCV-CCVC-CV-CVCC CV-CCV
8	Om Hayu Wredhiyasa Wredhi	Om Ha-yu Wre-dhi-ya-sa Wre-dhi	CV CV-CV CCV-CCV-CV-CV CCV-CCV
9	Sang sisyawan matur nimbal	Sang si-sya-wan ma-tur nim-bal	CVCC CV-CCV-CVC CV-CVC CVC-CVC
10	Jani melahang mangrungu indik Tri Parartha Tatwa	Ja-ni me-la-hang mang-ru-ngu in-dik Tri Pa-rar-tha Tat-wa	CV-CV CV-CV-CVCC CVCC-CV-CCV VC-CVC CCV CV-CVC-CCV CVC-CV

The second testing is identification Pupuh. Testing is done by entering 100 Pupuh lyrics into the system. Table 2 is example results for Pupuh identification with wrong input. System can identify the types of Pupuh correctly and calculate the nearest value corresponding user input.

Table 2. Table *Pupuh* Identification

No	***Pupuh* Lyrics**	**Syllables Parsing**	**Types of *Pupuh***	**Wrong Input Warning**
1	Mangden adung Becik waspadayang malu Suba ke sanggurua Tekek ngamong sila yukti Ento tuhu Sikut becik ngawe sadya	Mang-den a-dung (4u) Be-cik was-pa-da-yang ma-lu (8u) Su-ba ke sang-gu-ru-a (7a) -> (6a) Te-kek nga-mong si-la yuk-ti (8i) En-to tu-hu (4u) Si-kut be-cik nga-we sa-dya (8a)	*Pupuh* Pucung	Wrong input for number of syllables
2	Buin pidan manyi padi kuning manyidayan kadi makunyit di alas katemu lamun idepe sarin tanah tiang ibuk blahan payuk bas bebeki beruk tanah sarat pisan dakin karna uling ilu daluang bisa ngumbara mangulayang kayun ira tumas manik jeron dewa ampurayang	Bu-in pi-dan ma-nyi pa-di ku-ning ma-nyi-da-yan (14a) -> (8a) ka-di ma-ku-nyit di a-las (8a) -> (8i) Mungkin maksud anda "kadi makunyit di alis"? ka-te-mu la-mun i-de-pe (8e) -> (8a) Mungkin maksud anda "katemu lamun idepa"? sa-rin ta-nah ti-ang i-buk (8u) -> (8i) Mungkin maksud anda "sarin tanah tiang ibik"? bla-han pa-yuk bas be-be-ki (8i) be-ruk ta-nah sa-rat pi-san (8a) -> (8u) Mungkin maksud anda "beruk tanah sarat pisun"? da-kin kar-na u-ling i-lu (8u) -> (8a) Mungkin maksud anda "dakin karna uling ila"? da-lu-ang bi-sa ngum-ba-ra (8a) -> (8i) Mungkin maksud anda "daluang bisa ngumbari"? ma-ngu-la-yang (4a) -> (4u) Mungkin maksud anda "mangulayung"? ka-yun i-ra tu-mas ma-nik (8i) -> (8a) Mungkin maksud anda "kayun ira tumas manak"? je-ron de-wa am-pu-ra-yang (8a)	*Pupuh* Sinom	Wrong input for number of lines, syllables, and the last vowel
3	Nanak Bagus Pyanak Bapa Mai Malu Nampekang Cening Malinggih Jani Melahang Mangrungu Indik Tri Parartha Tattwa Mangden Sinah Becik Artinnya Pang Weruh	Na-nak Ba-gus Pya-nak Ba-pa (8a) Ma-i Ma-lu (4u) Nam-pe-kang Ce-ning Ma-ling-gih (8i) Ja-ni Me-la-hang Mang-ru-ngu (8u) In-dik Tri Pa-rar-tha Tat-twa (8a) Mang-den Si-nah (4a) Be-cik Ar-tin-	*Pupuh* Pangkur	Wrong input for last vowel determination

	Tri Tatelu Keartiang Para Jagat Keartiang	nya Pang We-ruh (8u) Tri Ta-te-lu Ke-ar-ti-ang (8a) Pa-ra Ja-gat Ke-ar-ti-ang (8a) -> (8i) Mungkin maksud anda "Para Jagat Keartiing"?		
4	Mangkin kocap ida sang sarosapati Prabu ring erlanggya Putran sri erlanggya aji Ring wengine manyumpena	Mang-kin ko-cap i-da sang sa-ro-sa-pa-ti (12i) Pra-bu ring er-lang-gya (6a) Pu-tran sri er-lang-gya a-ji (8i) Ring we-ngi-ne ma-nyum-pe-na (8a)	*Pupuh* Masku-mam-bang	
5	Saking tuhu manah guru Mitutur cening jani Kawruhe luwir sanjata Ne dadi prabotang sai Kaanggen ngaruruh merta Saenun ceninge urip	Sa-king tu-hu ma-nah gu-ru (8u) Mi-tu-tu-rin ce-ning ja-ni (8i) Ka-wru-he lu-wir san-ja-ta (8a) Ne da-di pra-bo-tang sa-i (8i) Ka-ang-gen nga-ru-ruh mer-ta (8a) Sa-e-nun ce-ni-nge u-rip (8i)	*Pupuh* Ginanti	

Based on the results table testing of 100 Pupuh there are 23 wrong input, consisting of the number of lines, number of syllables, and the last vowel.

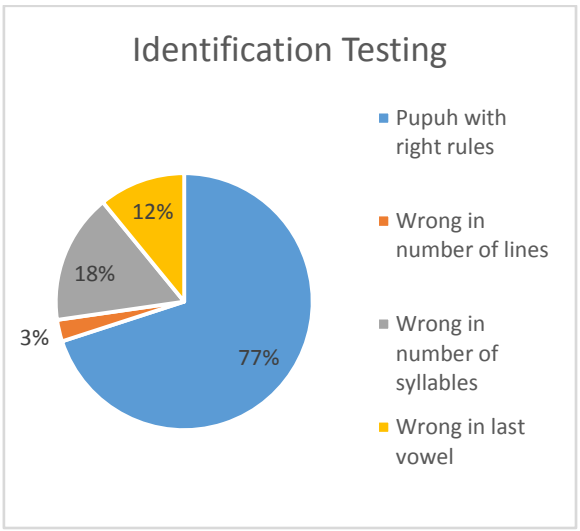

Fig. 9: Percentage Results of Identification Testing

Based on Pupuh testing table, the system can detect faults and keep identify the type of Pupuh. Figure 9 displays the percentage of which as many as 77% have been in accordance with Padalingsa rules, while 23% of them have wrong in inputting the number of rows, number of syllables, and the last vowel. The details of the 23% are wrong in inputting number of rows by 3%, wrong input for number of syllables as much as 18% and 12% wrong in inputting last vowel. The factors that cause wrong input are less conscientious author in making lyrics of Pupuh.

The results that have been achieved, these are: the system capable to performing parsing syllables, counting the syllables and determine final vowel correctly based on types of *Pupuh*. If there is an wrong input, then the system will give a warning and a suggestion for wrong input on last vowel. The system can identify lyrics by calculating the accumulated value of each of the factors, such as total of lines, syllables, and the last vowel.

5. CONCLUSION

Web-Based Implementation of Finite State Automata Method on Lyrics Recognition System of Balinese Song "*Pupuh*" obtained some conclusions based on system testing, as follow that system can be executed correctly and can separate syllables correctly. Types of *Pupuh* also can be identified by a corresponding input from the user. This system is very useful for students and general, due to increase knowledge about Tembang Sekar Alit and at the same time preserving the local culture of Indonesia. Further improvements of this system are develops additional features text to speech and speech to text for input and output of the system. Otherwise the system can be developed by creating a mobile version that can be used more flexibly. In the future, Finite State Automata method can be adapted for parsing syllables in other languages with any improvements.

6. ACKNOWLEDGMENTS

Thank you for Department of Information Technology Udayana University for supports and guidance on project preparation.

7. REFERENCES

[1] Tanjung Turaeni, Ni Nyoman. 2011. Dharmagita. Metasastra. Vol. 4 No. 2. h. 171-172.

[2] Waltz, David & Pollack, Jordan, 1985, Massively Parallel Parsing: A Strongly Interactive Model of Natural Language Interpretation, Cognitive Science.

[3] Kiraz, George Anton. 2000. Multitiered Nonlinear Morphology Using Multitape Finite Automata: A Case Study on Syriac and Arabic. Association for Computational Linguistics. Vol. 26 No. 1. h. 77-105.

[4] Yuniarti, Anny; Tjahyanto, Aris & Kuswardayan, Imam. 2004. Kamus Bahasa Arab-Indonesia Online dengan Pemecahan Suku Kata Menggunakan Metode Parsing. Jurnal Ilmiah Teknologi Informasi. Vol. 3 No. 1. h. 9-16.

[5] Wasista, Sigit & Astin, Novita. 2010. Algoritma Sistem Pembaca Teks Bahasa Indonesia Menggunakan Metode FSA (Finite State Automata). Politeknik Negeri Bandung. Vol. 15 No. 2. h. 1-8.

[6] Oka Sudana, AA. Kompiang; Adi Purnawan, I Ketut; Riana Mahlia Dewi, Ni Made. 2014. Android Based Translator of Balinese into Indonesian using Binary Search Method. International Journal of Software

Engineering and Its Applications. Vol. 8, No. 6. h. 165-182.

[7] Dar Aziz, Amal., Cackler, Joe & Yung, Raylene, *Basics of Automata Theory*, Stanford University.

[8] Mohri, Mehryar, 1997, Finite-State Transducers in Language and Speech Processing, Association for Computational Linguistics.

[9] References for Natural Languange Processing: http://galaxy.eti.pg.gda.pl/katedry/kiw/pracownicy/Jan.Daciuk/personal/thesis/node12.html

[10] Berry, Gerard; Sethi, Ravi, 1986, From Regular Expressions to Deterministic Automata, Theoretical Computer Science 48