

## The Development of Indonesian POS Tagging System for Computer-aided Independent Language Learning

<https://doi.org/10.3991/ijet.v12.i11.7383>

Muljono<sup>(✉)</sup>, Umriya Afini, Catur Supriyanto, Raden Arief Nugroho  
Dian Nuswanto University, Semarang, Indonesia  
muljono@dsn.dinus.ac.id

**Abstract**—Word processing tool is a basic need in learning a language. One of the word processors needed by a language learner is part of speech (POS) tagging. While many POS Tagging tools for Indonesian language have been developed, no systems have been addressed specifically for language learners. This paper presents a study on an Indonesian part of speech (POS) tagging system developed as one of word processing tools for language learners. We use resources from previous Indonesian POS tagging research, such as MorphInd for the morphological analysis and IPOSTagger for part of speech tagging. Objective and subjective tests are employed to evaluate this system. In the objective test the part of speech tagging results use a system model developed from IPOSTagger in combination with MorphInd as the morphological analyzer, and compared with the results of part of speech tagging produced from the original IPOSTagger system model. The results show that the part of speech tagging accuracy using this system model is higher than other models. For its subjective evaluation, Mean Opinion Score (MOS) is used to the 24 participating respondents. The MOS results obtained reach 3,61 for test-1, 3,87 for test-2, and 3,72 for test-3. From the results, we expect that this POS tagging system could be used to help language learners in their Indonesian language self-learning process.

**Keywords**—Part of speech tagging, language learning, self-learning, morphology analysis

### 1 Introduction

The rapid development of Natural Language Processing (NLP) device has led to an increasingly higher need to develop a word processing application to help language learning processes. Previous studies suggest that learning a language with a computer aid namely word processor could significantly improve learner's proficiency than when using conventional method[1]. A word processing tool could provide them with important information such as language morphology, syntax, phonology, semantic and etymology which could be helpful for them in understanding and using language optimally[2].

The importance of the word processing tool will be even higher for learners desiring to learn foreign language who will likely find it hard thanks to the interference from their mother tongue. Word processing tools have been a basic need in learning a language. This is because these tools could enrich language learners with important aspects, namely writing, translation, and spelling.

One of word processors which could be helpful for language learners is part of speech (POS) tagging. Part of speech tagging is necessarily needed since manual tagging will take too much time and require language experts. The large number of ambiguous words, i.e. those words which have different POS tag when they are in different contexts and complicated language morphology have also been some obstacles for learners in POS tagging.

In previous studies, several POS tagger models, particularly for Indonesian language, have been developed using various approaches. Some of these approaches are statistic-based [3][4][5], rule-based [6] and transformation-based learning approaches[7]. From these numerous POS tagger types, we learn that while many POS tagging tools for Indonesian language have been studied and developed, no POS tagging tools for general audience have been widely available. Therefore, the access to these POS Tagging tools becomes the main issue for its language learners.

This study focuses on developing an Indonesian part of speech (POS) tagging computer-aided self-learning system addressed for language learners. However, it is not impossible that this tool could be used as one of word processing tools to support language learning in classroom. To make its use easier, the system is made in the form of a desktop application which could be installed and operated in the language learners' personal computer.

This POS tagging system for self-learning should have a high accuracy level in part of speech tagging. Hence, the best method from previous research on POS tagging for Indonesian language is selected. A POS Tagging with statistical approach using Hidden Markov Model (HMM) is selected to deal with ambiguity since it gives higher accuracy and fast processing time[8]. In addition, an Indonesian morphological analyzing system is applied to help reduce tagging error in unknown words (Out Of Vocabulary/OOV) during the tagging process [9].

The existing system resources to be used to build this system are MorphInd and IPOSTagger[4][10]. MorphInd system is an Indonesian morphological analyzing system. And IPOSTagger is an Indonesian POS tagger system which applies statistical approach using HMM. While IPOSTagger have applied such methods as Affix tree, Succeeding POS tag and Lexicon from KBBI-Kateglo, we use MorphInd system as a morphological analyzer in OOV words since it could deal with irregular characters such as affixation, reduplication and cliticization (proclitic and enclitic).

We combine both systems into a sub-system of a self-learning system for Indonesian POS Tagger. Finally, a graphical user interface (GUI) is added for interface in order to facilitate the use of this system.

## 2 Indonesian Part of Speech Tagging

POS *tagging* is a task of tagging *part of speech* (POS) in each word in any sentence being inputted. POS, which is also called as part of speech, or syntax category, could provide information on the word and those around it. This part of speech information could be used to determine a word's structure syntax, making this POS tagging an important part of syntactic parsing. POS tagging is also used as one of features in named entity recognition and information extraction[11]. In addition to being an important part in Natural Language Processing (NLP) devices, POS tagging can also be used as an aid to help language learners.

POS tagging system could be used by language learners to discover the part of speech tag of a word in a sentence. Learners only need to input the sentence they want its POS is tagged on and the system will give the result immediately. Furthermore, the part of speech information they obtain from the system could be directly used in their language learning.

In developing the Indonesian POS tagging self-learning system, several resources such as Indonesian POS corpus, tagset and POS tagging sub-system are required.

### 2.1 Indonesian Data Preparation

Some supporting data are used for the Indonesian POS Tagging system such as tagset and POS corpus for Indonesian language.

**Tagset.** Tagset is a sequence of POS tag which may be assigned to a word. Some versions of tagset for Indonesian POS corpus have previously been developed. For the system we are working on, the tagset is taken from the tagset modification used in previous studies [3][4]. The tagsets used here are 31 POS tag types. Table 1 shows the tagset used by this Indonesian POS tagging system.

**Indonesian POS Corpus.** For the Indonesian POS corpus, we use Dinakaramani corpus, which manually labels 250,000 tokens using 23 tag types [12]. We take around 10,000 tokens from this corpus and re-tag with the tagsets in Table 1. The resulting corpus is then used for the data training process prior to POS tagging process.

**Table 1.** Tagset used by the system

NO	POS	POS Name	Example
1	OP	Open Parenthesis	{
2	CP	Close Parenthesis	}
3	GM	Slash	/
4	;	Semicolon	;
5	:	Colon	:
6	"	Quotation	"
7	.	Sentence Terminator	!?
8	,	Comma	,
9	-	Dash	-
10	CC	Coordinating Conjunction	dan, atau, tetapi, etc.
11	CD	Cardinal number	dua, juta, 7916 etc.
12	OD	Ordinal number	pertama, ketiga, ke-4 etc.
13	DT	Determiner	para, sang, si etc.
14	FW	Foreign Words	foreign, word, origin etc.
15	IN	Preposition	di, ke, dari etc.
16	JJ	Adjective	kaya, senang, bersih etc.
17	MD	Modal	bisa, boleh, harus etc.
18	NEG	Negation	bukan, tidak etc
19	NN	Common Noun	mobil, kertas, rupiah etc.
20	NNP	Proper Noun	Semarang, Indonesia, Jawa etc.
21	NND	Classifier, partitive, and measurement noun	orang, helai, lembar etc.
22	PR	Demonstrative pronoun	ini, itu, sini etc.
23	PRP	Personal Pronouns	saya, kamu, dia etc.
24	RB	Adverb	sementara, nanti, sangat etc.
25	RP	Particles	pun, kah, lah
26	SC	Subordinating conjunction	jika, ketika, supaya, etc.
27	SYM	Symbols	@#\$%^&
28	UH	Interjection	wah, aduh, Oi, oh, hai etc.
29	VB	Verb	membeli, memakan, tidur etc.
30	WH	WH-Pronouns	apa, siapa, kapan etc.
31	X	Unknown	bangedd, G0k!L, jus+ice etc.

## 2.2 Building Indonesian POS Tagging Sub-system

After the supporting data are completed, we make an Indonesian POS tagging sub-system. This sub-system is used for backend process. The main process in this sub-system is morphological analysis and tagging. We add preprocessing in the form of sentence cutting and tokenization before the sub-system main processes are executed. Figure 1 explains the processes in this sub-system.

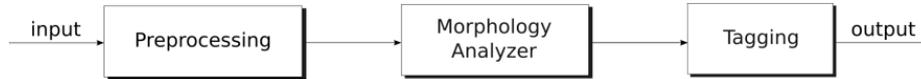


Fig. 1. Indonesian POS tagging sub-system

There are two types of input in this sub-system, i.e. training corpus and the sentence to be POS tagged. In the Tagging process, the training corpus input will be used for modeling the word with its POS tag using HMM algorithm. Meanwhile, the sentence to be tagged will be processed in the next processes.

**Preprocessing.** In this preprocessing, the system will perform tokenization and sentence detection. Tokenization is performed for word cutting with punctuation. It is done since POS tags are also used by punctuation such as comma, quotation, and also punctuation for sentence terminators: period; question mark.

Sentence detection is performed since the system will do the tagging on per sentence basis. Thus, if the inputs are more than one sentence, the system will detect them and tag every one of these sentence terminators.

**Morphological Analysis.** The Indonesian morphological analyzer system MorphInd is applied for morphological analysis process. MorphInd system is a finite state-based Indonesian morphological analyzer system [10]. MorphInd uses Foma toolkit for its compiling process. This system will analyze every token in a sentence on unigram basis, thus it does not depend on its surrounding token. Its output takes the form of morphemic segmentation, lemma morpheme position, lexical category, and morphological feature. Figure 2 shows an example of input and output of MorphInd system.

Input	Output
apa	^apa<w>_W--\$
yang	^yang<s>_S--\$
tidak	^tidak<g>_G--\$
membunuhmu	^meN+bunuh<v>_VSA+kamu<p>_PS2\$
,	^,<z>_Z--\$
membuatmu	^meN+buat<v>_VSA+kamu<p>_PS2\$
lebih	^lebih<d>_D--\$
kuat	^kuat<a>_ASP\$

Fig. 2. MorphInd input and output example

We will only take the results of lexical category of each token. However, since MorphInd has different lexical category or tagset from the tagset being applied, this output will be changed into the tagset used as indicated in table 1.

**Tagging.** The system applied for the tagging process is IPOStagger system. We apply 2 types of tagging model configuration, i.e. HMM bigram and HMM trigram methods as its basic model. IPOStagger system has also applied *smoothing* method to deal with sparse data issue, i.e. Jelinec-Mercer *smoothing* and Linier Interpolation *smoothing* methods. Figure 3 indicates the tagging process workflow in this sub-system.

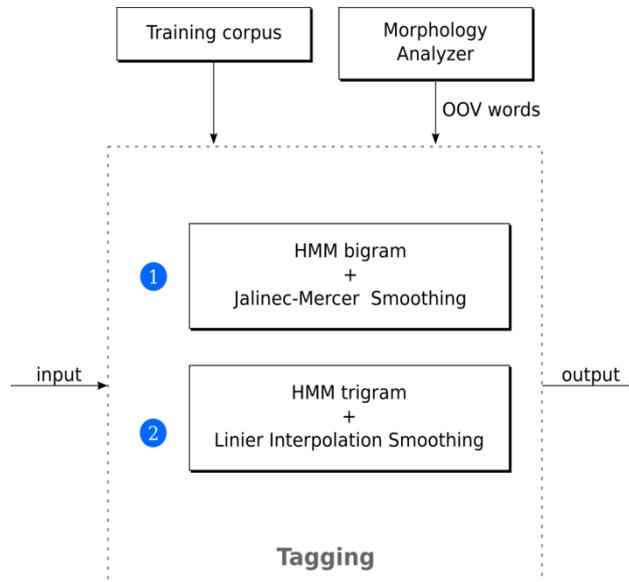


Fig. 3. Sub-system tagging process

The input in Figure 3 is the sequence of preprocessing results. Every word in this sequence of words will be figured out from the training corpus whether or not it is an OOV word. If it is not an OOV word, the tagging will be immediately performed, and if it is an OOV word, the POS tag assigned to it is obtained from the previous morphological analysis process.

The end results of this Indonesian POS Tagger sub-system is a sentence which has been assigned with POS tags. Figure 4 shows the system input and output.

**Input:** *Apa yang tidak membunuhmu, membuatmu lebih kuat.*

**Output:** *Apa (WH-Pronouns)  
yang (Subordinating conjunction)  
tidak (Negation)  
membunuh (Verb)  
mu (Common noun)  
. (Comma)  
membuat (Verb)  
mu (Common noun)  
lebih (Adverb)  
kuat (Adjective)  
. (Sentence terminator)*

Fig. 4. Indonesian POS tagging sub-system example

### **3 Building a System of Indonesian POS tagging for Self-Learning**

After preparing the Indonesian POS Tagging sub-system, we develop an Indonesian POS Tagging system for language self-learning.

In order for the system development to go effectively, we use Instructional Systems Design (ISD) with ADDIE model. ADDIE itself is a design framework containing a sequence of general processes using instructional design and training development [13][14]. This model has 5 (five) phases, namely analysis, design, development, implementation and evaluation phases. These phases constitute the steps towards building Indonesian POS Tagging system for language self-learning.

#### **3.1 Analysis Phase**

During this phase, we analyze the importance of developing POS tagging capable of producing information on part of speech for language learners. The fairly complicated Indonesian morphology has been quite a challenge for learners wanting to figure out the correct language structure[15]. The morphological information will be even more helpful for those learners speaking a foreign language in their effort of understanding the language structure since it has significant influence in the translation.

This POS tagging system is addressed towards language learners who learn the language independently. Learners do not necessarily possess specific skills; they are expected only to get accustomed to operating computer. We have also prepared a clear guideline for learners to install it themselves to make it more user friendly. They are expected only to follow the guidelines appropriately and they could immediately embark on the learning by themselves.

With the aid of this POS tagging system for Indonesia language, it is expected that learners who previously find it difficult to determine the POS tag of a word in a sentence could directly figure out the POS tags in that word. A system taking the form of desktop application compatible with many operating systems would eventually enable learners to learn at any time and any place as long as they have their personal computer devices.

#### **3.2 Design Phase**

The targeted users of this POS Tagging system are language learners who wish to use this system for self-learning. We consider many backgrounds and ranges of age these learners may come from and attempt to make the design of this POS tagging system easily operable to anyone.

In terms of its display, the system is designed as simple as possible. The window will directly show the main interface which contains text input field, and a main control button to run the system process. The final output of the system is shown in a field right below it.

On its right-hand side, a button to help user delete the input in the input field is available, returning the display to default. The system uses English as its default for its dialog. Figure 5 explains the environment of this Indonesian POS Tagging system to learn the language independently.

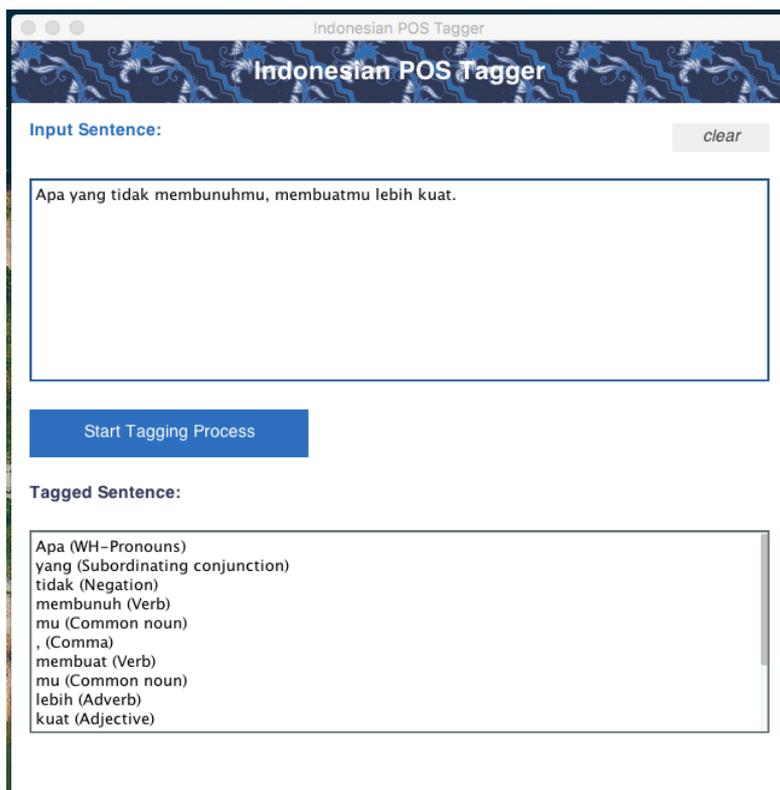


Fig. 5. Environment of Indonesian POS Tagging system for language self-learning

The procedure to use the system is as follows: Users will be asked to enter the sentence whose POS will be tagged into the main input field, and then they are asked to press the button to commence the tagging process. After the system begins the tagging process, it will show the results in the form of a sentence whose POS has been tagged in the existing output field. If users want to make another tagging, they merely need to press the 'clear' button at the upper right corner, and they could use the system for the tagging process.

From the way the system works, it has two parts, namely frontend and backend. In the frontend part, we apply a system interface as a bridge between users and the system. In the backend part, we apply Indonesia POS Tagging sub-system.

### 3.3 Development Phase

In this development phase, we apply both system design (backend) and interface design (frontend) which have been prepared during the previous phases. To facilitate the application of these two parts, we apply a modular model, i.e. making each system process into modules. Figure 6 shows the modules used in the self-learning of Indonesian POS Tagging system.

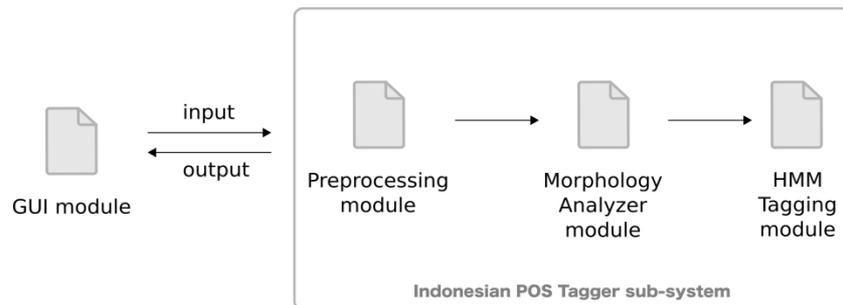


Fig. 6. Modules from Indonesian POS Tagging system for language self-learning

In the frontend part, the module made is GUI module. This module serves as the system interface. This interface is made using Java Swing, a toolkit in Java programming language for making GUI. Once the interface is made, we test it to ensure that each part functions well.

In the backend part, a module is made for each component of the Indonesian POS tagging sub-system processes. The Preprocessing module runs the input preprocessing task from GUI, the Morphology analyzer module contains input processing from the previous process using MorphInd system, and IPOSTagger module performs the part of speech tagging task using HMM algorithm.

We combine the modules from the frontend and backend parts into a unit. Afterwards, we test it to find out whether or not the system could work as expected. Once the system successfully meets the test feasibility, an executable file is then made for the system so that it can be easily installed in various systems.

### 3.4 Implementation Phase

During this implementation phase, we install the system we have developed into a device which will be used for self-learning. Several system software and module dependences should be fulfilled first prior to the installation of the system.

The software or modules required for the system to work appropriately are Java Runtime Environment (JRE) and Foma. For JRE, every operating system usually has provided its default version. Thus, all that is required is to update it to its latest version. As for Foma, one needs to download it first before installing.

Upon the installation of module dependence, the executable file of Indonesian POS Tagging system for language self-learning could be installed in the device.

### 3.5 Evaluation Phase

Formative and summative evaluations are applied in the development of this Indonesian POS tagging system. The formative evaluation is held at each phase in this ADDIE model. This evaluation aims at revising system errors found and or improving the system quality. The summative one is performed after the latest version of the system is obtained. The main purpose of this summative evaluation is to figure out the quality of service (QoS) and quality of experience (QoE).

During the summative evaluation, we use objective and subjective measurements towards the system being made. In the subjective measurement, Mean Opinion Score (MOS) from respondents is used, where the value indicator is determined within the predetermined scale to measure the subject’s opinion about a system’s performance [16][17].

The objective measurement is performed to discover the POS tagging system accuracy level. We use calibration model, i.e. by comparing the result of the tagging system model accuracy measurement with that of another tagging model. The comparison is made with the original model of IPOStagger system, namely the basic models of HMM bigram, HMM trigram, Affix tree, and Lexicon. We use 3 corpora, the first one contains 10% OOV words, the second one contains 20% OOV words and the last one contains 30% OOV words. Table 2 shows that the results of HMM trigram and Morphology Analyzer (MA) models are higher than other models tested at an average reaching to 92.25%.

**Table 2.** Results of objective measurement of tagging system model accuracy with that of other tagging models

Model	Corpus 1	Corpus 2	Corpus 3
HMM Bigram (baseline)	90.16%	84.44%	76.33%
HMM Trigram (baseline)	90.65%	84.66%	76.53%
HMM Bigram + MA	95.32%	92.81%	88.33%
HMM Trigram + MA	95.68%	92.81%	88.26%
HMM Bigram + Affixtree	93.20%	88.90%	84.28%
HMM Trigram + Affixtree	93.79%	89.34%	84.44%
HMM Bigram + Affixtree + Lexicon	93.36%	89.72%	85.14%
HMM Trigram + Affixtree + Lexicon	94.15%	90.17%	85.24%

Meanwhile, the subjective measurement is held to discover the system quality standard based on the perception of those users directly interacting with the system. We prepare several criteria to test the subjective measurement from respondents’ opinions. We test these criteria to 24 respondents. They will be asked to assign a score to each criterion based on predetermined rules in each criterion. The final results of this evaluation will determine the system accuracy level for users.

The criteria tested in this evaluation include the design of the user interface menu, the user- friendliness of the system and properness of the system

**The design of the user interface menu (Test-1).** We use MOS scale to find out the level of system design user interface attractiveness. The score scale used is 1 - 5.

Scale 1: Bad (extremely uninteresting), scale 2: Poor (uninteresting), scale 3: Fair (fairly interesting), scale 4: Good (interesting), and scale 5: Excellent (absolutely interesting). Figure 7 shows the average results of participants' tests reached 3.61, which indicates that the user interface design used is fairly interesting.

**The user-friendliness of the system (Test-2).** In this user-friendliness test, the participants will be given MOS score scales 1 - 5, respectively 1: Bad (very hard to operate), 2: Poor (hard to operate), 3: Fair (easy enough to operate), 4: Good (easy to operate), and 5: Excellent (very easy to operate). In this evaluation, the final result of the system's user-friendliness test reaches 3.87, indicating that the system is easy enough to operate.

**Properness of the system (Test-3).** This test is held to discover whether or not the system is appropriate to be used for self-learning. MOS scale 1 - 5 is provided to participants to find out the subjective score of function properness of the system. 1: Bad (highly improper), 2: Poor (improper), 3: Fair (proper enough), 4: Good (proper), and 5: Excellent (highly proper). Figure 7 shows the participants' average final score, i.e. 3.72, showing that the system is proper enough for self-learning use.

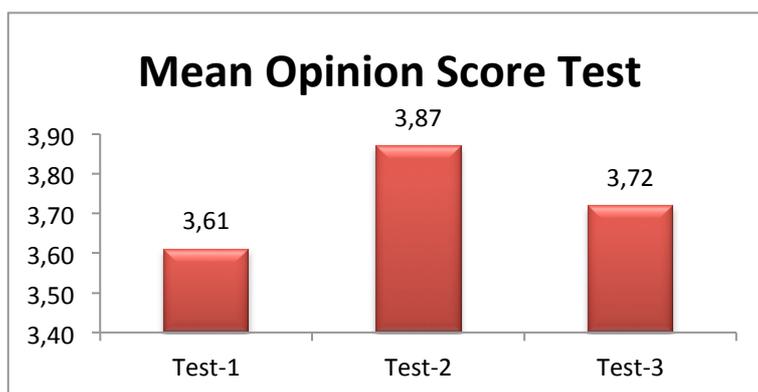


Fig. 7. Results of Mean Opinion Score Test (Test 1: The design of the user interface menu, Test 2: The user-friendliness of the system, Test 3: Properness of the system)

## 4 Conclusion

This research shows that the Indonesian POS Tagger system could be applied to assist learners of Indonesian language with independent learning. The system is designed to help assign tags to a sentence with 31 POS tags applied by the system.

During the evaluation phase, objective and subjective measurements are used. The objective measurement is used to assess the accuracy level of POS tagging in the method used by the system. The objective measurement results show that HMM trigram and Morphology Analyzer (MA) methods applied here show higher accuracy than other methods being tested. In the subjective measurement, 3 evaluation criteria are tested to participants, namely user interface design, user-friendliness of the sys-

tem, and properness of the system. The results of subjective evaluation from 24 participants indicate quite promising results.

We expect that this system could be applied for word processing devices for POS tagging which could be used to help learners to acquire the language skills through self-learning.

## 5 Acknowledgment

The authors would like to express sincere gratitude to the Directorate General of Higher Education of Indonesia for the financial support of this research under PENELITIAN PRODUK TERAPAN scheme 2017

## 6 References

- [1] B. Wang, "Empirical Study on the Computer-aided College English Translation Teaching," *Int. J. Emerg. Technol. Learn. IJET*, vol. 11, no. 12, pp. 68–71, Dec. 2016. <https://doi.org/10.3991/ijet.v11i12.6012>
- [2] A. Hamouda, "A study of dictionary use by Saudi EFL students at Qassim university," *Study Engl. Lang. Teach. Sch. CO LTD*, vol. 1, no. 1, Feb. 2013.
- [3] F. Pisceldo, M. Adriani, and R. Manurung, "Probabilistic Part Of Speech Tagging for Bahasa Indonesia," in *Proceeding In Third International MALINDO Workshop*, 2009.
- [4] A. F. Wicaksono and A. Purwarianti, "HMM based Part Of Speech Tagger for Bahasa Indonesia," in *Proceeding of 4th International Malindo (Malay and Indonesian Language) Workshop (2010)*, Jakarta, 2010.
- [5] A. Nurwidyanoro and E. Winarko, "Parallelization of Maximum Entropy POS Tagging for Bahasa Indonesia with MapReduce," *Int. J. Comput. Sci. Issues IJCSI*, vol. 9, no. 4 No. 2, Jul. 2012.
- [6] F. Rashel, A. Luthfi, A. Dinakaramani, and R. Manurung, "Building an Indonesian rule-based part-of-speech tagger," in *2014 International Conference on Asian Language Processing (IALP)*, 2014, pp. 70–73. <https://doi.org/10.1109/IALP.2014.6973521>
- [7] S. Sari, H. Hayurani, M. Adriani, and S. Bressan, "Developing Part of Speech Tagger for Bahasa Indonesia Using Brill Tagger.," in *Proceedings In The International Second MALINDO Workshop*, 2008.
- [8] T. Brants, "TnT: A Statistical Part Of Speech Tagger," in *ANLC '00 Proceedings of the sixth conference on Applied natural language processing*, Seattle, Washington, 2000, pp. 224–231. <https://doi.org/10.3115/974147.974178>
- [9] J. A. Bakar, K. Omar, M. F. Nasrudin, and M. Z. Murah, "Morphology Analysis In Malay POS Prediction," in *Proceeding of the International Conference on Artificial Intelligence in Computer Science and ICT (AICS 2013)*, 2013, vol. 25–26.
- [10] S. D. Larasati, V. Kuboň, and D. Zeman, "Indonesian Morphology Tool (MorphInd): Towards an Indonesian Corpus," in *Systems and Frameworks for Computational Morphology*, 2011, pp. 119–129.
- [11] D. Jurafsky and J. H. Martin, *Speech and Language Processing, 2nd Edition*, 2nd edition. Upper Saddle River, N.J: Prentice Hall, 2008.

- [12] A. Dinakaramani, F. Rashel, A. Luthfi, and R. Manurung, "Designing an Indonesian part of speech tagset and manually tagged Indonesian corpus," in *2014 International Conference on Asian Language Processing (IALP)*, 2014, pp. 66–69.
- [13] G. R. Morrison, S. M. Ross, J. E. Kemp, and H. Kalman, *Designing Effective Instruction*, 6 edition. Hoboken, NJ: Wiley, 2010.
- [14] \_ Muljono, S. Sumpeno, D. Arifianto, K. Aikawa, and M. H. Purnomo, "Developing an Online Self-learning System of Indonesian Pronunciation for Foreign Learners," *Int. J. Emerg. Technol. Learn. IJET*, vol. 11, no. 04, pp. 83–89, Apr. 2016. <https://doi.org/10.3991/ijet.v11i04.5440>
- [15] P. J. Hyun, "Potensi dan Tantangan Bahasa Indonesia Menuju Bahasa Internasional," *J. Sosioteknologi*, vol. 14, no. 1, pp. 12–21, Apr. 2015. <https://doi.org/10.5614/sostek.itbj.2015.14.1.2>
- [16] M. Muljono, S. Sumpeno, A. Arifin, D. Arifianto, and M. H. Purnomo, "Indonesian Text to Audio Visual Speech with Animated Talking Head," *Int. Rev. Comput. Softw. IRECOS*, vol. 11, no. 3, pp. 261–269, Mar. 2016. <https://doi.org/10.15866/irecos.v11i3.8706>
- [17] "P.10 : Vocabulary for performance and quality of service," 2006. [Online]. Available: <https://www.itu.int/rec/T-REC-P.10/en>. [Accessed: 24-Jun-2017].

## 7 Authors

**Muljono** is with Informatics Engineering Department, Dian Nuswantoro University, Semarang, Jawa Tengah, Indonesia (e-mail : [muljono@dsn.dinus.ac.id](mailto:muljono@dsn.dinus.ac.id)). He received his Doctorate Degree in Electrical Engineering Department, Institut Teknologi Sepuluh Nopember (ITS), Surabaya, Indonesia in 2016. He joined an Internship Program at School of Media Science, Tokyo University of Technology Japan in 2014. His research interests are speech processing, artificial intelligence and natural language processing.

**Umriya Afini** is with Informatics Engineering Department, Universitas Dian Nuswantoro, Semarang - Indonesia (e-mail : [afini.umriya@gmail.com](mailto:afini.umriya@gmail.com)). She received bachelor degree in Informatics Engineering Department at Universitas Dian Nuswantoro, Semarang, in 2016. Her research interest focused is natural language processing..

**Catur Supriyanto** is with Informatics Engineering Department, Universitas Dian Nuswantoro, Semarang - Indonesia (e-mail : [catur.supriyanto@dsn.dinus.ac.id](mailto:catur.supriyanto@dsn.dinus.ac.id)). He received Master Degree at Universiti Teknikal Malaysia Melaka in 2011. His research interests is artificial intelligence and information retrieval.

**Raden Arief Nugroho** obtained his Doctorate Degree at Universitas Sebelas Maret in 2017. He is a lecturer in the Departement of English Language Studies, Faculty of Humanities, Universitas Dian Nuswantoro, Semarang, Indonesia. His main research interests include translation process, computational linguistics, and disabilities studies.

Article submitted 04 July 2017. Published as resubmitted by the authors 23 August 2017.