

Nazief & Adriani Stemming Algorithm With Cosine Similarity Method For Integrated Telegram Chatbots With Service

by Dedy R

Submission date: 27-Oct-2021 12:43AM (UTC+0700)

Submission ID: 1684816382

File name: Nazief_Adriani_Stemming_Algorithm_With_Cosine.pdf (828.57K)

Word count: 3345

Character count: 18229

PAPER • OPEN ACCESS

Nazief & Adriani Stemming Algorithm With Cosine Similarity Method For Integrated Telegram Chatbots With Service

4
To cite this article: IGLPE Prisma *et al* 2021 *IOP Conf. Ser.: Mater. Sci. Eng.* **1125** 012039

10
View the [article online](#) for updates and enhancements.

You may also like

- 13
- [Acacia's Revisi™ 9697 Multimedia CD: BOM for GCSE Science \(New Version 1.5\)](#)
Roger Hackett
- 12
- [Teacher's questions on mathematics learning based on ELPISA framework](#)
Rozalia, R Johar, M Duskri *et al*
- [Quality analysis of final test in the faculty engineering of Yogyakarta State University](#)
E Budiasuti, Sugiyono and MA Jerusalem

Nazief & Adriani Stemming Algorithm With Cosine Similarity Method For Integrated Telegram Chatbots With Service

IGLPE Prisma¹, DR Prehanto¹, DA Dermawan¹, AC Herlingga¹, SC Wibawa²

¹Informatics Department, Universitas Negeri Surabaya, Indonesia 60231

lanangprisma@unesa.ac.id

Abstract. Information systems have developed a number of questions that are often asked to customers service with a high level of question similarity. For help it create a FAQ related to information systems. The FAQ has a lot of information so that the user is confused and it takes time to find information. More users to ask questions to customer service. Chatbot is one way to help users and customers service in this matter. Customer service can answer questions automatically and users can ask questions as if to ask customer service live. In this study, researchers will apply Nazief & Adriani algorithm used to do stemming because this algorithm is an algorithm which is effective in stemming Indonesian, and method cosine similarity to find the similarity level of questions with FAQs that are in the FAQ database. With implement algorithms and methods in chatbot with the help of the Telegram Messenger service produce a relative answer according to that the user expected. This method is an effective way to answer questions automatically. And if the question asked did not find an answer or the answer is not suitable then the question will be conveyed in e-layanan Information technology development center of Universitas Negeri Surabaya.

Keywords: cosine similarity, chatbots, telegram.

1. Introduction

The information system has many features and many menus that information system users can use. This can lead to many questions about information systems. Questions asked by users often have similarities between users so customer service answer questions repeatedly. Documentation and videos can be used to provide guidance use of information systems, but users prefer to ask questions directly customer service rather than reading or viewing the documentation provided. Providing an FAQ is one way to help users when they encounter difficulties. The FAQ has a lot of information because of the many questions. This can confuse users and take time to find information. to overcome this, the media is needed to help customer service in answer questions automatically using Chatbot. Chatbots have been widely used in everyday life for example polls, games, etc. Taking advantage of a chatbot that can respond to dynamic queries requires natural language processing. Therefore we need an algorithm that is able to process words and match sentences in the database.

According to Wahyudi's research (2017) with the title "implementation and analysis of the Nazief & Adriani stemming algorithm and porters in Indonesian language documents." stemming Nazief & Adriani has better results than the algorithm stemming porter. However, processing time, Porter's

algorithm is better than Nazief & Adriani's algorithm. According to Suhada (2017) in a research entitled "Implementation of the Rabin Karp Algorithm and Stemming Najief Andriani for Document Plagiarism Detection" Algorithm stemming Nazief & Adriani is very suitable to be applied in the process of searching for basic words in Indonesian based on a basic word dictionary.

Relevant research on matching methods, according to research by Gunawan (2018) with the title " The Implementation of Cosine Similarity to Calculate Text Relevance between Two Documents " method cosine similarity has great results for matching 2 documents. According to Lahitani (2016) with the title " Cosine Similarity to determine similarity measure: Study Case in Online Essay Assessment " method cosine similarity can provide a similarity score in matching essay assessments online With the above relevant problems and research, the researcher wants to apply the algorithm stemming nazief & Adriani and methods cosine similarity to be used on the Telegram chatbot which is integrated with the service. The objectives of this study are: (1) Creating a chatbot that can answer questions automatically, (2) Calculating the accuracy of the suitability of the chatbot response to incoming questions with the Nazief & Adriani steaming algorithm and the similarity method using cosine similarity.

1.1 The Nazief & Adriani algorithm

The Nazief & Adriani algorithm was first developed by Bobby Nazief & Mima Adriani. This algorithm is based on the broad morphological rules of Indonesian, which are collected into one group and encapsulated into permissible additives and non-permitted additives. There is a basic word dictionary used to support word recording and word matching after word stemming. The stemming process in the Nazief & Adriani algorithm is as follows:

1. Before the next process, the words will be matched in the base word database. If there is a match then the algorithm stops.
2. Remove the suffix ("-lah", "-kah", "-tah" or "-pun") if it is a particle ("-lah", "-kah", "-tah" or "pun") this step is repeated again to get rid of the inflectional possessive pronoun suffixes ("-ku", "mu" or "-nya"). Check the word in the root word dictionary, if found, the algorithm stops.
3. Remove the Derivational Suffix ("-i" or "-an"). Check the words in the basic word dictionary. If the word is found in the base word dictionary, the algorithm stops. If not found then proceed to step 3a:
 - a. If there is an "-k" suffix and the last letter is "-k" then "-k" will be deleted. Check the words in the basic word dictionary. If the word is found in the dictionary, the process stops. If not, then go to step 3b.
 - b. Suffixes that are deleted ("-i", "-an", or "-kan") do not remove the affixes and go to step 4.
4. Remove the Derivational Prefix ("be-", "di-", "to-", "men", "pe-"; "se-" and "te-"). Check the words in the basic word dictionary. If the word is found in the root word database, the process stops. At this stage it stops if it meets the following conditions:
 - Unallowed combination of prefix and suffix was found
 - a. The prefixes found are the same as previously deleted prefixes.
 - b. Omitting the three prefixes
5. If the process has been carried out but the root word is not found in the base word dictionary database, the original word that has not been stemmed will be returned.

In their research, Asian et al. (Asian et al., 2005) developed several Nazief & Adriani algorithms as follows:

1. using a more complete word dictionary Add
2. the rules for words loop compound.
3. Adding the prefix and ending rules, as well as other rules, namely:
 - a. Adds a pericle (inflection suffix) "-pun".
 - b. Added prefix chopping rules.

- c. Changes to the dispatch rules for the prefix type "me".
4. Changes in the sequence of the stemming process, namely:
 - a. Words with the prefix "be-" and the suffix "-lah", remove the beginning first and then the ending. Words with prefix "be-"
 - b. and suffix "-an", remove the prefix first then the suffix.
 - c. words with the prefix "men" and the suffix "-i", remove the prefix first then the suffix.
 - d. words with prefix "di-" and suffix "-i", remove the prefix first then the suffix.
 - e. Words with prefix "pe-" and suffix "-i", remove the prefix first then the suffix
 - f. Words with the prefix "ter-" and the suffix "-i", remove the prefix first then the suffix.

1.2 Cosine Similarity

Ariantini (2016: 1) cosine similarity is used in positive space. So that the cosine similarity compares the document similarity level using the concept of cosine degree where the value has a limit between the values 0 and 1. If the result of cosine similarity is 0, it is said to be not similar. If the result of cosine similarity is 1, the document is considered similar

$$\text{Similarity} = \cos(\theta) = \frac{A \cdot B}{\|A\| \cdot \|B\|}$$

$$\cos(\theta) = \frac{\sum_{i=1}^n A_i \cdot B_i}{\sqrt{\sum_{i=1}^n (A_i)^2} \cdot \sqrt{\sum_{i=1}^n (B_i)^2}}$$

Information :

A = User message

B = FAQ in the database

A_i = The weight of the word i in blocks A_i

B_i = The weight of the word i in blocks B_i

i = The number of words in the sentence

n = Number of vectors

1.3 Chatbot

Chatbot is a service where a person can interact through a chat interface (chat). A person can interact using voice or text and can also get the same response using voice or text. In general, chatbots will respond in a conversational style, and chatbots can take actions in response to conversations being asked or ordered. Messaging applications that are often used to create chatbots include Facebook Messenger, Telegram, Whatsapp, Slack or SMS. There are 2 types of chatbots:

1. Chatbots that can operate based on a set of pre-made rules. The chatbot can respond to specific commands. If someone carries out an order that is not in the rules provided, the chatbot does not know what that person's command means.
2. Chatbots that use machine learning and artificial intelligence to provide a good response. (Cheryleaf 2017)

2. Method

The following is the flow of the research to conduct research "Nazief & Adriani stemming algorithm with methods cosine similarity for integrated telegram chatbots with services "in general depicted in Figure 1, as follows:

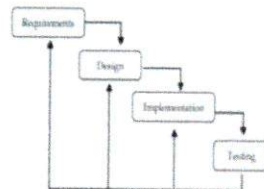


Figure 1 Flow Method [7]

Figure 1 discusses the research flow of the Nazief & Adriani stemming algorithm with the cosine similarity method for integrated telegram chatbots with services. The stages carried out in this study are needs analysis by conducting literature studies to find related literature to be used as a supporting reference in this study. The literature used is related to the Nazief & Adriani algorithm and the cosine similarity method. Then make a chatbot application design using the help of the telegram service. At the implementation stage, this research was written in a chatbot using the PHP language. In the testing phase the system is carried out to determine the suitability of the response to the answers to user questions by the bot.

2.1 Data collection

At this stage the authors conducted a literature study related to stemming theory using the Nazief & Adriani stemming algorithm and similarity methods Cosine Similarity. This research is focused on implementing and analyzing the Nazief & Adriani algorithm stemming theory and similarity methods Cosine Similarity on a chatbot. In this study, the authors used a dictionary of Indonesian basic words downloaded from the internet with a total of 28,530 words. These root words are stored in a database to be used in the process stemming. The author also uses the Tala stopword list published in the 2007 Asian research. The list of questions or FAQs that will be a response from the bot is obtained from the results of grouping questions that have been asked at the Surabaya State University services related to questions related to SIAKADU (Integrated Academic Information System) on January 1, 2020 to April 4, 2020. There are 15 frequently asked questions on services related to the Integrated Academic Information System of Universitas Negeri Surabaya.

2.2 System Design

In this system use the Telegram application service assistance as user interface for displays the response of the Nazief & Adriani algorithm process & adriani with metode cosine similarity. The flow of the system is described in Figure 2.

Figure 2 shows the flow of the Nazief & Adriani stemming algorithm system design using the method cosine similarity for chatbot telegram integrated with services in general. To start the process the user sends messages via chatbot using the Telegram application service. Then the bot will process the incoming sentences and respond or provide replies to questions submitted by users.

2.3 Design the Application

This study implements the Nazief & Adriani algorithm with the method cosine similarity using the PHP programming language and PostgreSQL database to create a website and bot. Website is used to enter FAQ whereas bot will provide responses or answers to incoming questions with the help of application services Telegram.

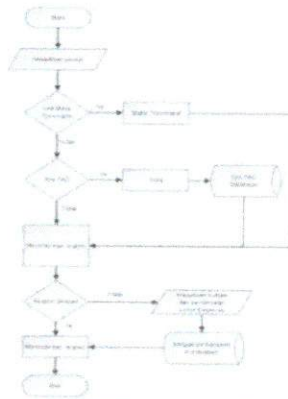


Figure 2 System design flow

2.4 Testing the Application

The trials in this study were conducted by 75 users service application telegram college student Universitas Negeri Surabaya. These 75 users ask questions to bots related to the Integrated Academic Information System (SIKADU)

3. Results and Discussion

Here is a master FAQ view:

(1) Display FAQ list page

To display a list of FAQs that have been entered by the admin.

Category	Title	Tag	Action
SIKADU	Perwakilan dari mata kuliah: Jawaban : SIKADU an membuat surat yang ditujukan ke Wakil Rektor 1	Deskripsi, mata, kuliah	[Edit] [Hapus]
SIKADU	Penghapusan dari mata kuliah: Jawaban : SIKADU an membuat surat yang ditujukan ke Wakil Rektor 1	Tagasi, mata, kuliah	[Edit] [Hapus]
SIKADU	Garis daftar beretiket SIKADU: Jawaban : SIKADU an membuat surat yang ditujukan ke Wakil Rektor 1	gato, jawab, beretiket, SIKADU	[Edit] [Hapus]
SIKADU	Lupa Password SIKADU: Jawaban : SIKADU an menggunakan akun: SIKADU	lupa, password, SIKADU	[Edit] [Hapus]

Figure 3 List FAQ

(2) Added FAQ page view

The form is used to add FAQ scenarios by entering question, question and answer categories.

Form FAQ

Category:

Title:

Tag:

Answer:

Figure 4 Add FAQ

(3) Change FAQ page view

Form to change the FAQ scenario.

Ubah FAQ	
Caption	SIKADU
Title	Perbaikan nilai mata kuliah
Answer	Silahkan membuat surat yang ditujukan ke Wakil Rektor 1
<input type="button" value="Simpan"/> <input type="button" value="Batal"/>	

Figure 5 Change FAQ

The following is a page display for simulated questions and the results of the answers or responses from bot. This menu is used to make it easier for admins to try out scenarios FAQ by entering a sentence. In this menu, you will see the results of the step-by-step sentences being processed to get a response or answer.

3.1. System Testing

System testing was carried out by 15 users of the Surabaya State University student telegram application. These 15 users ask questions to bots related to the Integrated Academic Information System (SIKADU) which are based on questions that often appear on the Surabaya State University services. Each user is free to use their respective vocabulary, not fixated on the sentences that are in FAQ. Below Figure 6 is an example of a question asked by a user on a bot via the telegram application.

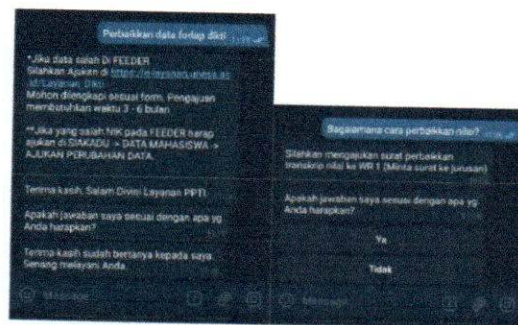


Figure 6 Trials 1

If the answer does not match, there will be a question to raise the problem to the E-service PPTI UNESA as figure 7.

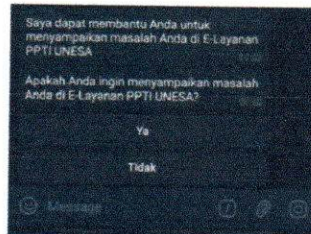


Figure 7. services

4. Conclusion

Testing is done by asking questions at random. After the user asks a question and answered by the bot, the bot will again ask about the response done by bots. From the trial there were 236 questions who entered. There were 217 responses which stated that they were with hope and there were 19 responses that stated no corresponding. From these results, the percentage can be more 91.95% answered accordingly and 8.05 who answered not accordingly. From these results it can be concluded that the algorithm Nazief & Adriani with the Cosine Similarity method can be used to help answer questions online automatically with answers that are relative to that expected by the user. Thus this way is effective way to assist customer service in answer questions automatically and the user is in asking questions via chatbot.

5. References

- [1] [1] Elcholiqi, Abidah dan Musdholifah, Aina. 2020. "Chatbot in Bahasa Indonesia Using NLP to Provide Banking Information". *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*. Vol. 14 (1): hal 81-90.
- [2] Wahyudi, Dwi DKK. 2018. "Implementasi Dan Analisis Algoritma Stemming Nazief & Adriani Dan Porter Pada Dokumen Berbahasa Indonesia". *Jurnal Ilmiah SINUS*. hal 49-56.
- [3] Suhada, Satia dan Bahri Saeful. 2017. " Implementasi Algoritma Rabin Karp Dan Stemming Najief Andriani Untuk Deteksi Plagiarisme Dokumen". *SWABUMI* Vol.5: hal. 84-89.
- [4] D Gunawan, dkk. 2017. "The Implementation of Cosine Similarity to Calculate Text Relevance between Two Documents", 2nd International Conference on Computing and Applied Informatics. hal. 1-6.
- [5] Salim, Mohammad A dan Anistiyasari, Yeni. 2017. " Pengembangan Aplikasi Penilaian Ujian Essay Berbasis Online Menggunakan Algoritma Nazief Dan Adriani Dengan Metode Cosine Similarity ". *Jurnal IT-EDU*. Vol. 2 (1): hal. 126-135.
- [6] Lahitani, Alfirna R, dkk. 2016. " Cosine similarity to determine similarity measure: Study case in online essay assessment". *International Conference on Cyber and IT Service Management*.
- [7] Benedictus, Ruspandi R, dkk. 2017. " Rancang Bangun Chatbot Helpdesk untuk Sistem Informasi Terpadu Universitas Sam Ratulangi". *E-Journal Teknik Informatika*. Vol. 11 (1)
- [8] Handoyo, Eko, dkk. 2018. "Proc. of 2018 5th Int. Conf. on Information Tech., Computer, and Electrical Engineering (ICITACEE)". hal. 325-330.
- [9] Mardiana, Tari, dkk. 2016. "Stemming Influence on Similarity Detection of Abstract Written in Indonesia ". *Jurnal TELKOMNIKA*. Vol. 14 (1): hal. 219-227.
- [10] Prihatini PM, dkk. 2017. "Stemming Algorithm for Indonesian Digital News Text Processing", *International Journal of Engineering and Emerging Technology*. Vol. 2 (2): hal. 1-7.

- [11] Sari, Bunga dan Sibaroni, Yuliant. 2019. "Deteksi Kemiripan Dokumen Bahasa Indonesia Menggunakan Algoritma SmithWaterman dan Algoritma Nazief & Andriani". *Ind. Journal on Computing*. Vol. 4 (3): hal. 88-98.
- [12] F. B. Djafar, A. Lahinta and H. Lillyan , "Penerapan Algoritma Smith-Waterman Dalam Sistem Pendeteksi Kesamaan Dokumen".
- [13] M. Y. Rimaldo and S."Sistem Pendeteksi Plagiat Dokumen Teks Menggunakan Algoritma SmithWaterman Serta Algoritma Arifin dan Setiono," 2012.
- [14] Ariantini, Dewa A, dkk.2016. "Pengukuran Kemiripan Dokumen Teks Bahasa Indonesia Menggunakan Metode Cosine Similarity". *Jurnal Inovtek Polbeng-Seri Informatika. E-Journal Teknik Informatika*. Vol. 9 (1): hal. 1-8.
- [15] Deolika, Agatha, dkk.2019. "Analisis Pembobotan Kata Pada Klasifikasi Text Mining". *Jurnal Teknologi Informasi*. Vol. 3 (2): hal. 179-184.
- [16] Diki Susandi, Pemanfaatan Vector Space Model pada Penerapan Algoritma Nazief Adria ni, KNN dan Fungsi Similarity Cosine untuk Pembobotan IDF dan WIDF pada Prototipe Sistem Klasifikasi Teks Bahasa Indonesia. *Jurnal ProTekInfo* Vol. 3 No. 1 ISSN: 2406-7741 , 2016.
- [17] Mandias, F. A., Penerapan Data Mining Untuk Evaluasi Kinerja Akademik Mahasiswa Di Universitas Klabat Dengan Metode Klasifikasi, *Konferensi Nasional Sistem & Informatika 2015, STIKOM Bali*, 9-10 Oktober 2015, 2015.
- [18] M. Syukri Mustafa, Implementasi Data Mining untuk Evaluasi Kinerja Akademik Mahasiswa Menggunakan Algoritma Naive Bayes Classifier, *Citec Journal*, Vol. 4, No. 2 2460- 4259, 2017.
- [19] R. Shah, S. Lahoti, and Lavanya, "An Intelligence Chat-bot using Natural Language Processing," *International Journal of Engineering Research*, vol no. 6, Issue no. 5, pp: 281-286, 2017.
- [20] Ariantini, Dewi Ayu Rai. 2016. Pengukuran Kemiripan Dokumen Teks Bahasa Indonesia Menggunakan Metode Cosine Similarity. Vol.9 ISBN:2301-8364: pp 1-8
- [21] Shutharland Smith. *Plagiarism, the internal and student learning: Improving Academic integrity*. New York: Routledge. 2014. [22] Manopo, R. I. 2016. "Perancangan Aplikasi Help Desk di UPTTIK Unsrat", *Tekno*, vol. 8, no.1, 2016.
- [22] Gibson, Darril. 2015. *Effective Help Desk Specialist Skills*. USA: Pearson Education, Inc.
- [23] Dewi, A. & Setiaji, B. 2014. "Pemanfaatan Sentence Similarity Measurement untuk Proses Pencarian Pola pada Chatbot Berbasis Pattern-Matching," *Seminar Nasional Teknologi Informasi dan Multimedia 2014*, ISSN : 2302-3805, 8 Februari 2014.
- [24] Baiti, Zifora.,Nugroho, Fresy. 2013. "Aplikasi Chatbot IM3 untuk Informasi Jurusan Teknik Informatika Berbasis Sistem Pakar menggunakan Forward Chaining", *MATICS* vol.5 no.3, September 2013.
- [25] Britz, Denny. 2016. "Deep Learning for Chatbots, Part 1 – Introduction" [Online]. Available: <http://www.wildml.com/2016/04/deep-learning-forchatbots-part1-introduction> Diakses tanggal 12 Juni 2017

Nazief & Adriani Stemming Algorithm With Cosine Similarity Method For Integrated Telegram Chatbots With Service

ORIGINALITY REPORT

21 %
SIMILARITY INDEX

17 %
INTERNET SOURCES

18 %
PUBLICATIONS

0 %
STUDENT PAPERS

PRIMARY SOURCES

- | | | |
|---|---|----|
| 1 | coa.ctu.edu.vn
Internet Source | 4% |
| 2 | etheses.whiterose.ac.uk
Internet Source | 2% |
| 3 | turcomat.org
Internet Source | 2% |
| 4 | Mohammad Idhom, Ronggo Alit, Akhmad Fauzi. "Implementation of Web Server Security Against Denial of Service (DoS) Attacks", IOP Conference Series: Materials Science and Engineering, 2021
Publication | 2% |
| 5 | journal.portalgaruda.org
Internet Source | 1% |
| 6 | Yan Watequlis Syaifudin, Pramana Yoga Saputra, Dwi Puspitasari. "The implementation of web service based text preprocessing to measure Indonesian student | 1% |

thesis similarity level", MATEC Web of Conferences, 2018

Publication

7	Mohammad Arifin Nq, Lindung Parningotan Manik, Dany Widiyatmoko. "Stemming Javanese: Another Adaptation of the Nazief-Adriani Algorithm", 2020 3rd International Seminar on Research of Information Technology and Intelligent Systems (ISRITI), 2020	1 %
Publication		
8	Proceeding of LPPM UPN "Veteran" Yogyakarta Conference Series 2020 – Engineering and Science Series, 2020	1 %
Publication		
9	journal.ugm.ac.id	1 %
Internet Source		
10	lppm.ub.ac.id	1 %
Internet Source		
11	Reza Gunawan, Ichan Taufik, Edi Mulyana, Opik T Kurahman, Muhammad Ali Ramdhani, Mahmud. "Chatbot Application on Internet Of Things (IoT) to Support Smart Urban Agriculture", 2019 IEEE 5th International Conference on Wireless and Telematics (ICWT), 2019	1 %
Publication		

12 Rozalia, R Johar, M Duskri, C Khairunnisak, T Zubaidah. "Teacher's questions on mathematics learning based on ELPSA framework", Journal of Physics: Conference Series, 2020
Publication 1 %

13 iopscience.iop.org
Internet Source 1 %

14 Dewi Soyusiawaty, Anna Hendri Soleliza Jones, Nora Lestari Lestariw. "The Stemming Application on Affixed Javanese Words by using Nazief and Adriani Algorithm", IOP Conference Series: Materials Science and Engineering, 2020
Publication 1 %

15 Alfirna Rizqi Lahitani, Adhistya Erna Permanasari, Noor Akhmad Setiawan. "Cosine similarity to determine similarity measure: Study case in online essay assessment", 2016 4th International Conference on Cyber and IT Service Management, 2016
Publication <1 %

16 Anggit Dwi Hartanto, Andy Syaputra, Yoga Pristyanto. "Best Parameter Selection Of Rabin-Karp Algorithm In Detecting Document Similarity", 2019 International Conference on Information and Communications Technology (ICOIACT), 2019 <1 %

17 journal.uad.ac.id <1 %
Internet Source

18 Julianto Lemantara, M.J. Dewiyani Sunarto, Bambang Hariadi, Tri Sagirani, Tan Amelia. "Prototype of Online Examination on MoLearn Applications Using Text Similarity to Detect Plagiarism", 2018 5th International Conference on Information Technology, Computer, and Electrical Engineering (ICITACEE), 2018 <1 %
Publication

19 Wahyu Hidayat, Ema Utami, Anggit Dwi Hartanto. "Effect of Stemming Nazief & Adriani on the Ratcliff/Obershelp algorithm in identifying level of similarity between slang and formal words", 2020 3rd International Conference on Information and Communications Technology (ICOIACT), 2020 <1 %
Publication

20 Seonghun Hong, Taeyun Kim, Jinwhan Kim. "Underwater visual SLAM with loop-closure using image-to-image link recovery", OCEANS 2015 - Genova, 2015 <1 %
Publication
