

# Unlocking Insights: A Literature Review on Enhanced Confix Stripping and Nazief & Adriani Algorithm Modifications for Makassar Language Text Stemming

Muhammad Wahyu Ade Saputra<sup>1</sup>; Ema Utami<sup>2</sup>; Ainul Yaqin<sup>3</sup>

<sup>1,2</sup>Magister of Informatics, Universitas AMIKOM Yogyakarta

<sup>3</sup>Faculty of Computer Science, Universitas AMIKOM Yogyakarta

<sup>123</sup>Sleman, Daerah Istimewa Yogyakarta, Indonesia

**Abstract:-** This study investigates the Nazief and Adriani Algorithm and the Enhanced Confix Stripping Stemmer (ECS) in the context of Makassar language. Following a comprehensive investigation, the Nazief & Adriani Algorithm demonstrates proficiency in capturing the complexities of Makassar language by applying numerous morphological criteria. Meanwhile, the Enhanced Confix Stripping Stemmer (ECS) exhibits versatility in dealing with language obstacles, identifying opportunities for further improvement. Using Sastrawi, Confix Stripping, Enhanced Confix Stripping, and Nazief-Adriani, the study emphasizes the need of using linguistically suitable techniques for exact analysis. This work sheds light on improving text processing technology in Makassar language, opening the path for algorithms customized to the language's unique qualities.

**Keywords:-** Stemming; Makassar; Algorithm; Language; Linguistics.

## I. INTRODUCTION

Makassar language holds a significant position in Indonesia, particularly in South Sulawesi, with a rich historical background and widespread usage among the local populace. Despite its prevalence as a daily communication tool, text processing and information management in Makassar often lag the standards observed in Indonesia language, the widely adopted national language. This disparity poses multifaceted challenges, particularly in the realm of information technology, where text processing efficiency directly impacts communication quality in Makassar. Consequently, there arises a critical need for the development of resources and technologies tailored to support text and information processing in Makassar language.

Linguistic structure of Makassar language is characterized by 23 phonemes, encompassing 18 consonant phonemes and seven vowel phonemes, with five native vowel segments. Notably, consonant phonemes are distributed across various positions, while affixation plays a prominent role in word formation. Affixes in Makassar language include verbal prefixes, compound prefixes, suffixes, and various types of

infixes, each contributing to the morphological complexity of the language.

In addressing the linguistic intricacies of Makassar language, researchers have explored various stemming algorithms to facilitate text processing tasks. Stemming is the process of reducing inflections or derivations to their fundamental forms, similar to reducing the derivation "comfortable" to its base, "comfort"[1]. Stemming is commonly used for pre-processing in text-based applications[2]. Stemming algorithms were classified into two categories. There were two types of stemmers: statistic-based and rule-based. Statistic-based stemmers were unsupervised algorithms that used training data to construct models for stemming, whereas rule-based stemmers used a set of predefined rules to execute stemming[3]. The advantages of this stemming may be used to develop search engines[4]. However over stemming and under stemming become common difficulties during the stemming process[5]. Every language has its own unique characteristics and structure, particularly the affix structure, thus the stemming method will be altered in line with the language's characteristics[6].

Each language has a unique stemming algorithm that differs from those used in other languages[7]. There have only been two existing stemming algorithms for Indonesia language. Nazief and Adriani developed these algorithms, as well as Tala's algorithm[8]. Nazief Adriani's method is a stemming algorithm using a dictionary as its working principle, whereas the Tala algorithm is based on Porter's algorithm and operates on a rule basis[9]. Among these, the Nazief & Adriani Algorithm, based on extensive morphological rules of Indonesia language, and the Enhanced Confix Stripping Stemmer (ECS), designed to rectify errors in the Rule-Based Approach method, have garnered attention. These algorithms offer potential solutions to enhance the accuracy and efficiency of text processing in Makassar language, albeit with distinct methodologies and outcomes.

Nazief Adriani Algorithm developed for the first time by Bobby Nazief and Mirna Adriani. This takes a rudimentary word dictionary and executes the recording, writing back the words that underwent repeated stemming[10]. Stands as a

prominent stemming method grounded in extensive morphological rules derived from Indonesia language. This algorithm consolidates diverse rules into a comprehensive framework, encapsulating permitted and prohibited affixes. Following the stemming process, a foundational word dictionary facilitates the matching and recording of words, enhancing the accuracy and reliability of the algorithm.

Research findings on Javanese and Madurese languages underscore the applicability and limitations of these algorithms in specific linguistic contexts, shedding light on their efficacy and areas for improvement. Despite promising results, challenges persist in adapting these algorithms to Makassar language, necessitating further investigation and modification to optimize their performance in this linguistic domain.

In light of the foregoing, this literature review aims to provide a comprehensive survey and comparative analysis of modified ECS and Nazief & Adriani algorithms for text stemming in Makassar language. By synthesizing existing research findings and identifying gaps in knowledge, this review seeks to contribute to the advancement of text processing technologies tailored to the unique linguistic characteristics of Makassar language. Through critical

evaluation and synthesis of relevant literature, this review aims to elucidate the strengths, weaknesses, and potential applications of these algorithms in the context of Makassar language text processing, ultimately informing future research directions in this field.

## II. RESEARCH METHODS

A Systematic Literature Review (SLR) study attempts to identify essential relevant studies, obtain the necessary data, then evaluate and synthesize the results to acquire greater insight into the research topic[1].

Regardless of the unique topic matter, disciplinary concentration, or philosophical position, Systematic Literature Review (SLR) is an organized procedure that consists of six different and crucial components, which are described below.

### A. Research Questions

When using the systematic literature review (SLR) approach, it is necessary to develop a set of research questions (RQs). The questions offered in Table 1 are critical in providing a more clear, goal-oriented, and efficient framework for the research project. This thorough method helps to improve and concentrate the research process.

Table 1 Research Question

ID	Research Question
RQ1	What techniques are employed by researchers for data collection in studies related to stemming?
RQ2	What methodologies are applied in the field of stemming research?
RQ3	What findings emerge from the investigation into stemming within the research context?

### B. Research Strategy

The investigator performs a thorough search for scientific papers in major databases such as ScienceDirect, IEEE, Springer, Semantic Scholar, Google Scholar, and Elsevier. This investigation is driven using two keywords, which include terminology in both Indonesian and English, to guarantee a complete and inclusive retrieval of relevant material:

- “Stemming” and “stemmer algorithm.”
- “Nazief & Adriani algorithm” and “enhanced confix stripping algorithm.”

### C. Study Selection

Establishing criteria is essential when assessing manuscripts. The researcher employs two distinct types of criteria applicable to paper composition: inclusion criteria and exclusion criteria. Presented below are specific inclusion criteria employed in the context of this study:

- Research paper is research conducted from 2019 to 2024.
- Research paper selected are written either in English or Indonesian.
- Main topic of the research study must be Indonesian stemming.

➤ And below are Exclusion Criteria for this Study:

- Research that is not included in the inclusion criteria.
- The research does not clearly describe its flow or methodology.
- Research that fails to meet research objectives.

➤ Criteria used in this Research is Shown Like Diagram in Figure 1 below

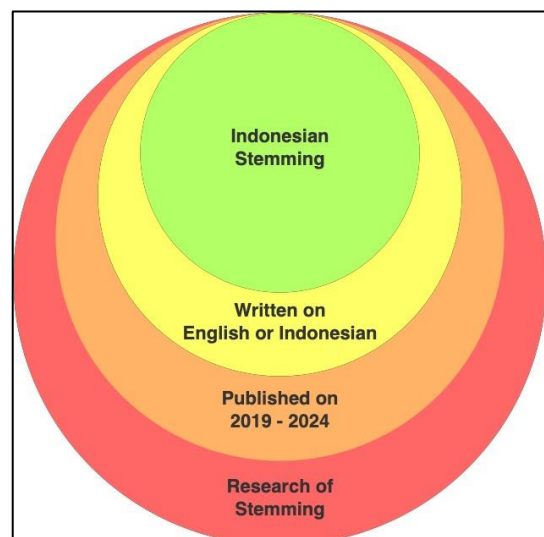


Fig 1 Criteria in this Research

**D. Quality Assessment**

To gain a comprehensive understanding of the study's overall quality, a careful quality evaluation must be conducted. This evaluation step is critical to determining if the identified data is relevant and appropriate for inclusion in the research. Within the context of this study, the collected data will be subjected to a thorough evaluation, led by a set of specified criteria meant to assess its quality. The use of these quality assessment standards guarantees a systematic and objective review, which contributes to the study's strength and dependability.

➤ For Every Papers used in this Research is Selected by using this Criteria below:

- Was the research published between 2019 and 2024?
- Is the research written in Indonesian or English?
- Is the research main topic is corelated with Indonesian stemming?

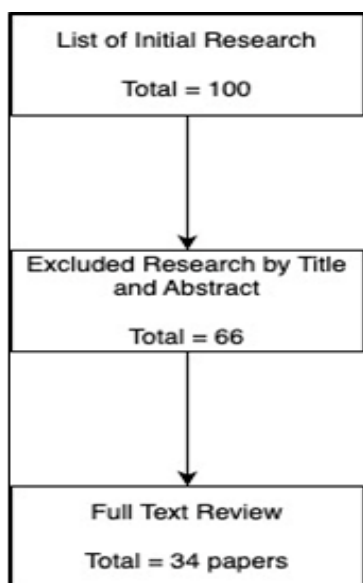


Fig 2 Data Selection

**E. Data Extraction**

At this point, the data extracted from the examined paper delves into key aspects such as the publication year of the research, the dataset used in the study under review, the methodologies used for data collection, the specific approach taken for stemming within the scrutinized research, and the resulting implications of stemming on the study. Following that, all relevant data was meticulously entered into a spreadsheet document. This diligent record-keeping serves as the foundation for a thorough analysis of the collected data. The plain structuring of this material provides an efficient and systematic study that adheres to known scholarly research practices.

**F. Data Synthesizing**

At this point in the research process, we've gathered a substantial pool of 100 studies, carefully sifting through titles and abstracts for relevance. This initial screening narrowed down our selection to 66 papers. However, our scrutiny didn't stop there. Using stringent inclusion and exclusion criteria, we meticulously handpicked 66 papers that aligned with our research goals. If a paper met our inclusion criteria, it earned a spot in our literature review; conversely, those meeting exclusion criteria were omitted. This discerning process resulted in a final set of 34 papers, which underwent a detailed review and analysis. The gleaned data and primary findings from these papers underwent a thorough examination, and their synthesis is methodically presented in Table 2. This approach adheres to the scholarly standards of research, ensuring a comprehensive and well-structured exploration of the literature.

Table 2 Reviewed Paper (A)

Research	Year	Language	Collection
[11]	2024	Indonesian	data from web
[12]	2024	English	data from web
[13]	2023	English	printed documents
[14]	2023	Indonesian	printed documents
[15]	2023	English	corpus or dictionary
[16]	2023	English	printed documents
[17]	2023	English	data from web
[18]	2023	English	data from web
[19]	2022	English	corpus or dictionary
[20]	2022	English	corpus or dictionary
[21]	2022	English	corpus or dictionary
[22]	2022	English	corpus or dictionary
[23]	2022	English	data from web
[24]	2022	English	corpus or dictionary
[25]	2022	English	data from web
[26]	2021	English	data from web
[27]	2021	English	printed documents

[28]	2021	English	data from web
[29]	2021	English	printed documents
[30]	2021	English	data from web
[31]	2021	English	data from web
[32]	2020	English	corpus or dictionary
[33]	2020	Indonesian	corpus or dictionary
[34]	2020	English	corpus or dictionary
[35]	2020	English	printed documents
[36]	2020	English	print document
[37]	2020	English	corpus or dictionary
[38]	2020	English	corpus or dictionary
[39]	2020	Indonesian	printed documents
[40]	2019	English	printed documents
[41]	2019	English	printed documents
[42]	2019	English	corpus or dictionary
[43]	2019	English	data from web
[44]	2019	English	printed documents

### III. RESULTS AND DISCUSSION

#### ➤ Data Collection Techniques

When conducting a study, researchers need data to focus their inquiry. Data collection is the systematic process of obtaining and reviewing exact data from many sources with the goal of finding answers to research questions, discovering patterns, exploring options, and evaluating prospective results. Throughout the data gathering process, researchers must define the nature of the data, identify its sources, and describe the methodology used. It is clear that data gathering methods vary, and this diversity is reflected in different strategies. Notably, the scientific, commercial, and governmental spheres rely heavily on effective data gathering procedures to inform their respective undertakings. This comprehensive approach is consistent with the rigorous standards observed in scientific research.

Previous studies employed a number of data collecting methods. Research [11], [12], [17], [18], [23], [25], [26], [28], [30], [31] and [43]. Get data used for research from various online site like twitter or any online news portal from Indonesia.

Meanwhile research like [13], [14], [16], [27], [29], [35], [36], [39], [40], [41] and [44]. All of that research done using data directly taken from printed document source. Either a book or paper exam result.

As for research [15], [19], [20], [21], [22], [24], [32], [33], [34], [37], [38] and [42]. All this one research is bases on dictionary, either it is Indonesian language only or with another option.

Detail of this source material classification can be seen in detail in table 2 about reviewed paper[A].

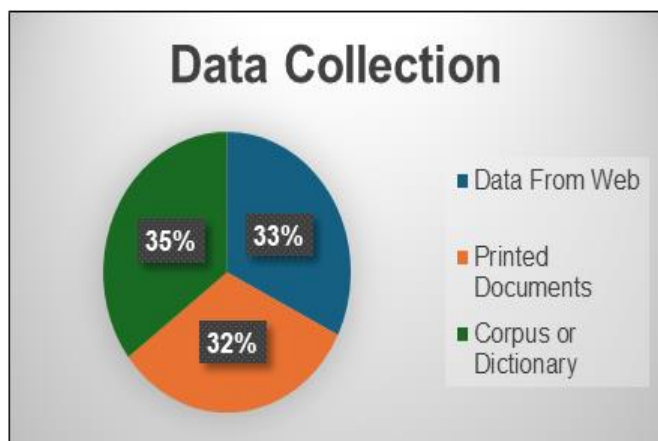


Fig 3 Data Collection Technique

#### ➤ Stemming Methods

Stemming, a key process in language analysis, is the reduction of inflections or derivations to their essential root forms, as demonstrated by reducing "comfortable" to its root form, "comfort." It is important to note that stemming does not simply reduce a word to its dictionary stem; rather, it uses algorithms to identify the proper truncation of words. This is in contrast to lemmatization, a more complex procedure that reduces words to their dictionary roots, requiring a high level of linguistic skill for proper implementation. This distinction emphasizes the need of selecting the right language analysis approach depending on the individual needs and peculiarities of a given research.

Stemming is a difficult operation in the context of the Indonesian language, given its huge lexicon, which includes around 127,000 fundamental terms described in the comprehensive Big Indonesian Dictionary. The complexity of stemming resides in the extraction of root words from attached words, which requires the removal of numerous affixes like as prefixes, infixes, suffixes, and combinations thereof. This method is important because it has a considerable impact on the quality of analytical results. To negotiate this linguistic complexity, a variety of stemming algorithms have been created. These include the Nazief-Andriani algorithms, Confix Stripping, Enhanced Confix

Stripping, and Sastrawi, which each offer unique techniques to dealing with the intricacies of Indonesian language stemming. Choosing an appropriate algorithm is critical to assuring the correctness and dependability of the stemming process inside the analysis framework.

List of regular distribution about stemming method that used in this research can be seen in table 3 reviewed paper (B).

Table 3 Reviewed Paper (B)

Research	Stemming Method
[11]	Sastrawi and Nazief-Adriani
[12]	Sastrawi, Nazief-Adriani, and Arifin Setiono
[13]	Sastrawi and Nazief-Adriani
[14]	Porter Stemmer
[15]	Confix Stripping (CS)
[16]	Nazief-Adriani
[17]	Nazief-Adriani
[18]	Sastrawi, Nazief-Adrian, Confix Stripping (CS) and Enhanced Confix Stripping (ECS)
[19]	Enhanced Confix Stripping (ECS) and New Enhances Confix Stripping (NECS)
[20]	Confix Stripping (CS)
[21]	Enhanced Confix Stripping (ECS)
[22]	Sastrawi
[23]	Nazief-Adriani
[24]	Confix Stripping (CS)
[25]	Nazief-Adriani
[26]	Nazief-Adriani
[27]	Nazief-Adriani
[28]	Confix Stripping (CS)
[29]	Nazief-Adriani
[30]	Nazief-Adriani
[31]	Nazief-Adriani
[32]	Nazief-Adriani
[33]	Enhanced Confix Stripping (ECS)
[34]	Nazief-Adriani
[35]	Nazief-Adriani
[36]	Sastrawi
[37]	Nazief-Adriani
[38]	Nazief-Adriani
[39]	Enhanced Confix Stripping (ECS)
[40]	Enhanced Confix Stripping (ECS)
[41]	Enhanced Confix Stripping (ECS)
[42]	Enhanced Confix Stripping (ECS)
[43]	Sastrawi
[44]	Nazief-Adriani

#### ➤ *Stemming Usage*

Examining the results reported in Table 3 from the evaluated studies indicates a variety of stemming algorithms that provide useful insights into understanding the complexities of the Makassar language. The Makassar language's distinguishing traits, particularly affixes and suffixes, lurk under the surface, offering obstacles to linguistic research.

A significant discovery is the frequency of academics using stemming algorithms that are specifically designed to handle the intricacies of the Indonesian language or traditional Indonesian languages. Sastrawi, Confix Stripping, Enhanced Confix Stripping, and Nazief-Adriani are some of the popular methods used for Makassar language analysis. These algorithms were intentionally picked for their ability to

navigate the complexity of Makassar, notably its affixes and distinctive language traits.

The strategic use of these stemming techniques is critical in understanding the structure and semantics of the Makassar language. This advanced technique not only improves knowledge of Makassar, but also emphasizes the significance of using contextually relevant linguistic tools to conduct a nuanced and correct analysis. The use of these precise algorithms reflects a dedication to methodological rigor in linguistic research, ensuring that the approaches employed are appropriate for the complexities of the language being studied.

#### IV. CONCLUSION

Finally, this study investigates the complexities of stemming algorithms, namely the Nazief & Adriani Algorithm and the Enhanced Confix Stripping Stemmer (ECS), in the context of Makassar language. The study identifies each algorithm's strengths and weaknesses using a comprehensive technique that includes data collecting, algorithm installation, and rigorous assessment.

Nazief & Adriani Algorithm emerges as a strong contender, using comprehensive morphological criteria to correctly capture the intricacies of Makassar language. Meanwhile, the Enhanced Confix Stripping Stemmer (ECS) demonstrates flexibility and potential in resolving the difficulties of the Makassar language, but with room for improvement.

The comparative analysis and assessment metrics offered in this study provide useful insights for academics and practitioners working to improve text processing technology in Makassar language. This study provides the groundwork for future efforts to improve and build algorithms adapted to the particular linguistic features of the Makassar language, ultimately enhancing the area of information processing in this linguistic domain.

#### REFERENCES

- [1]. Y. Karuniawati, E. Utami, and A. Yaqin, "A Systematic Literature Review of Stemming in Non-Formal Indonesian Language," vol. 8, no. 1, 2023.
- [2]. A. T. Ni'mah, D. A. Suryaningrum, and A. Z. Arifin, "Autonomy Stemmer Algorithm for Legal and Illegal Affix Detection use Finite-State Automata Method," *EPI Int. J. Eng.*, vol. 2, no. 1, pp. 46–55, Jun. 2019, doi: 10.25042/epi-ije.022019.09.
- [3]. A. S. Rizki, A. Tjahyanto, and R. Trialih, "Comparison of stemming algorithms on Indonesian text processing," *TELKOMNIKA Telecommun. Comput. Electron. Control*, vol. 17, no. 1, p. 95, Feb. 2019, doi: 10.12928/telkomnika.v17i1.10183.
- [4]. Enni Lindrawati, Ema Utami, and A. Yaqin, "ANoM STEMMER: Nazief & Andriani Modification for Madurese Stemming," *J. RESTI Rekayasa Sist. Dan Teknol. Inf.*, vol. 7, no. 6, pp. 1341–1347, Dec. 2023, doi: 10.29207/resti.v7i6.5086.
- [5]. E. Lindrawati, E. Utami, and A. Yaqin, "Comparison of Modified Nazief&Adriani and Modified Enhanced Confix Stripping algorithms for Madurese Language Stemming," *INTENSIF J. Ilm. Penelit. Dan Penerapan Teknol. Sist. Inf.*, vol. 7, no. 2, pp. 276–289, Aug. 2023, doi: 10.29407/intensif.v7i2.20103.
- [6]. J. Jumadi, D. S. Maylawati, L. D. Pratiwi, and M. A. Ramdhani, "Comparison of Nazief-Adriani and Paice-Husk algorithm for Indonesian text stemming process," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 1098, no. 3, p. 032044, Mar. 2021, doi: 10.1088/1757-899X/1098/3/032044.
- [7]. I. O. Suzanti and A. Jauhari, "COMPARISON OF STEMMING AND SIMILARITY ALGORITHMS IN INDONESIAN TRANSLATED AL-QUR'AN TEXT SEARCH," *J. Ilm. Kursor*, vol. 11, no. 2, p. 91, Jan. 2022, doi: 10.21107/kursor.v11i2.280.
- [8]. W. G. S. Parwita, "A document recommendation system of stemming and stopword removal impact: A web-based application," *J. Phys. Conf. Ser.*, vol. 1469, no. 1, p. 012050, Feb. 2020, doi: 10.1088/1742-6596/1469/1/012050.
- [9]. N. Pamungkas *et al.*, "Comparison of Stemming Test Results of Tala Algorithms with Nazief Adriani in Abstract Documents and National News," *Inf. J. Ilm. Bid. Teknol. Inf. Dan Komun.*, vol. 8, no. 1, pp. 33–41, Jan. 2023, doi: 10.25139/inform.v8i1.5569.
- [10]. S. Tuhpatussania, E. Utami, and A. D. Hartanto, "COMPARISON OF PORTERS STEMMING ALGORITHM AND NAZIEF & ADRIANI'S STEMMING ALGORITHM IN DETERMINING INDONESIAN LANGUAGE LEARNING MODULES," *J. Pilar Nusa Mandiri*, vol. 18, no. 2, pp. 203–210, Sep. 2022, doi: 10.33480/pilar.v18i2.3940.
- [11]. Muhammad Daffa Al Fahreza, Ardytha Luthfiarta, Muhammad Rafid, and Michael Indrawan, "Analisis Sentimen: Pengaruh Jam Kerja Terhadap Kesehatan Mental Generasi Z," *J. Appl. Comput. Sci. Technol.*, vol. 5, no. 1, pp. 16–25, Feb. 2024, doi: 10.52158/jacost.v5i1.715.
- [12]. L. Cahyaningrum, A. Luthfiarta, and M. Rahayu, "Sentiment Analysis on the Impact of MBKM on Student Organizations Using Supervised Learning with Smote to Handle Data Imbalance," 2024.
- [13]. Y. Purwati, F. S. Utomo, N. Trinarsih, and H. Hidayatulloh, "Feature Selection Technique to Improve the Instances Classification Framework Performance for Quran Ontology," *JOIV Int. J. Inform. Vis.*, vol. 7, no. 2, p. 615, Jul. 2023, doi: 10.30630/joiv.7.2.1195.
- [14]. G. N. M. Nata, "Pengembangan Algoritma Stemmer Bilingual Bali-Indonesia Dengan Rule-Base," 2023.
- [15]. S. I. Melia, J. Sholihah, D. Nisak, I. S. Juniaristha, and A. T. Ni'mah, "The Ngoko Javanese Stemmer uses the Enhanced Confix Stripping Stemmer Method," *Rekayasa*, vol. 16, no. 1, pp. 107–112, Apr. 2023, doi: 10.21107/rekayasa.v16i1.19308.
- [16]. D. S. Maylawati, Y. J. Kumar, and F. B. Kasmin, "Combination of Graph-based Approach and Sequential Pattern Mining for Extractive Text Summarization with Indonesian Language," vol. 9, no. 2, 2023.
- [17]. D. S. Maylawati, Y. J. Kumar, and F. Binti Kasmin, "Feature-based approach and sequential pattern mining to enhance quality of Indonesian automatic text summarization," *Indones. J. Electr. Eng. Comput. Sci.*, vol. 30, no. 3, p. 1795, Jun. 2023, doi: 10.11591/ijeecs.v30.i3.pp1795-1804.

- [18]. S. A. H. Bahtiar, C. K. Dewa, and A. Luthfi, "Comparison of Naïve Bayes and Logistic Regression in Sentiment Analysis on Marketplace Reviews Using Rating-Based Labeling," *J. Inf. Syst. Inform.*, vol. 5, no. 3, pp. 915–927, Aug. 2023, doi: 10.51519/journalisi.v5i3.539.
- [19]. S. H. Wibowo, R. Toyib, M. Muntahanah, and Y. Darnita, "Time complexity in rejang language stemming," *J. INFOTEL*, vol. 14, no. 3, pp. 174–179, Aug. 2022, doi: 10.20895/infotel.v14i3.764.
- [20]. S. Suyanto, A. Sunyoto, R. N. Ismail, E. Rachmawati, and W. Maharani, "Stemmer and phonotactic rules to improve n-gram tagger-based indonesian phonemicization," *J. King Saud Univ. - Comput. Inf. Sci.*, vol. 34, no. 6, pp. 3807–3814, Jun. 2022, doi: 10.1016/j.jksuci.2021.01.006.
- [21]. R. Sovia, S. Defit, and Yuhandri, "Development of the Minangkabau Local Language Translation Machine Based on Stemming," in *2022 International Symposium on Information Technology and Digital Innovation (ISITDI)*, Padang, Indonesia: IEEE, Jul. 2022, pp. 195–198. doi: 10.1109/ISITDI57734.2022.9944457.
- [22]. S. I. G. Situmeang, "Impact of Text Preprocessing on Named Entity Recognition Based on Conditional Random Field in Indonesian Text," vol. 6, no. 36, 2022.
- [23]. T. H. Jaya Hidayat, Y. Ruldeviyani, A. R. Aditama, G. R. Madya, A. W. Nugraha, and M. W. Adisaputra, "Sentiment analysis of twitter data related to Rinca Island development using Doc2Vec and SVM and logistic regression as classifier," *Procedia Comput. Sci.*, vol. 197, pp. 660–667, 2022, doi: 10.1016/j.procs.2021.12.187.
- [24]. H. Dwiharyono and S. Suyanto, "Stemming for Better Indonesian Text-to-Phoneme," *Ampersand*, vol. 9, p. 100083, 2022, doi: 10.1016/j.amper.2022.100083.
- [25]. A. Amalia, M. S. Lidya, A. Andrian, E. M. Zamzami, and S. M. Hardi, "OLCBot: Dissemination of Interactive Information Related To Indonesia's Omnibus Law With The Implementation of Fuzzy String Matching Algorithm and Sastrawi Stemmer," in *2022 6th International Conference on Electrical, Telecommunication and Computer Engineering (ELTICOM)*, Medan, Indonesia: IEEE, Nov. 2022, pp. 178–181. doi: 10.1109/ELTICOM57747.2022.10037966.
- [26]. R. Tjut Adek, R. Kesuma Dinata, and A. Ditha, "Online Newspaper Clustering in Aceh using the Agglomerative Hierarchical Clustering Method," *Int. J. Eng. Sci. Inf. Technol.*, vol. 2, no. 1, pp. 70–75, Nov. 2021, doi: 10.52088/ijesty.v2i1.206.
- [27]. I. Prisma, D. Prehanto, D. Dermawan, A. Herlingga, and S. Wibawa, "Nazief & Adriani Stemming Algorithm with Cosine Similarity Method for Integrated Telegram Chatbots With Service," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 1125, no. 1, p. 012039, May 2021, doi: 10.1088/1757-899X/1125/1/012039.
- [28]. F. Limansyah, Mokh. Suef, and V. Ratnasari, "Visitors Needs Analysis in Mall XYZ with Text Mining Analysis," *IPTEK J. Proc. Ser.*, vol. 0, no. 1, p. 152, Nov. 2021, doi: 10.12962/j23546026.y2020i1.11321.
- [29]. R. R. Et.al, "The Similarity of Essay Examination Results using Preprocessing Text Mining with Cosine Similarity and Nazief-Adriani Algorithms," *Turk. J. Comput. Math. Educ. TURCOMAT*, vol. 12, no. 3, pp. 1415–1422, Apr. 2021, doi: 10.17762/turcomat.v12i3.938.
- [30]. A. Amalia, D. Gunawan, and K. Nasution, "Sentiment analysis of GO-JEK services quality using Multi-Label Classification," *J. Phys. Conf. Ser.*, vol. 1830, no. 1, p. 012003, Apr. 2021, doi: 10.1088/1742-6596/1830/1/012003.
- [31]. M. Alfian, A. R. Barakbah, and I. Winarno, "Indonesian Online News Extraction and Clustering Using Evolving Clustering," *JOIV Int. J. Inform. Vis.*, vol. 5, no. 3, p. 280, Sep. 2021, doi: 10.30630/joiv.5.3.537.
- [32]. A. P. Wibawa, F. A. Dwiyanto, I. A. E. Zaeni, R. K. Nurrohman, and A. Afandi, "Stemming javanese affix words using nazief and adriani modifications," *J. Inform.*, vol. 14, no. 1, p. 36, Jan. 2020, doi: 10.26555/jifo.v14i1.a17106.
- [33]. N. W. Wardani and P. G. S. C. Nugraha, "Stemming Teks Bahasa Bali dengan Algoritma Enhanced Confix Stripping," *Int. J. Nat. Sci. Eng.*, vol. 4, no. 3, pp. 103–113, Dec. 2020, doi: 10.23887/ijnse.v4i3.30309.
- [34]. D. Soyusiawaty, A. H. S. Jones, and N. L. Lestariw, "The Stemming Application on Affixed Javanese Words by using Nazief and Adriani Algorithm," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 771, no. 1, p. 012026, Mar. 2020, doi: 10.1088/1757-899X/771/1/012026.
- [35]. M. S. Simanjuntak, J. Panjaitan, and S. A. Syahputra, "Using Preprocessing Text Mining With Nazief-Adriani Algorithms Similarity Of Essay Final Exam Semester," vol. 4, no. 36, 2020.
- [36]. M. A. Rosid, A. S. Fitriani, I. R. I. Astutik, N. I. Mulloh, and H. A. Gozali, "Improving Text Preprocessing For Student Complaint Document Classification Using Sastrawi," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 874, no. 1, p. 012017, Jun. 2020, doi: 10.1088/1757-899X/874/1/012017.
- [37]. R. A. Ramadhani, I. K. G. D. Putra, M. Sudarma, and I. A. D. Giriantari, "Stemming Algorithm for Indonesian Signaling Systems (SIBI)," *Int. J. Eng. Emerg. Technol.*, vol. 5, no. 1, p. 57, Jul. 2020, doi: 10.24843/IJEET.2020.v05.i01.p11.
- [38]. M. A. Nq, L. P. Manik, and D. Widiyatmoko, "Stemming Javanese: Another Adaptation of the Nazief-Adriani Algorithm," in *2020 3rd International Seminar on Research of Information Technology and Intelligent Systems (ISRITI)*, Yogyakarta, Indonesia: IEEE, Dec. 2020, pp. 627–631. doi: 10.1109/ISRITI51436.2020.9315420.

- [39]. E. Y. Hidayat and M. A. Rizqi, "Klasifikasi Dokumen Berita Menggunakan Algoritma Enhanced Confix Stripping Stemmer dan Naïve Bayes Classifier," *J. Nas. Teknol. Dan Sist. Inf.*, vol. 6, no. 2, pp. 90–99, Aug. 2020, doi: 10.25077/TEKNOSI.v6i2.2020.90-99.
- [40]. T. Yusnitasari, I. Humaini, L. Wulandari, and D. Ikasari, "Information Retrieval for Popular Words in Bahasa Translation of Al Quran and Hadith Bukhori Using Enhance Confix Stripping (ECS) Stemming," *Am. J. Softw. Eng. Appl.*, vol. 8, no. 1, p. 18, 2019, doi: 10.11648/j.ajsea.20190801.13.
- [41]. W. Rifai and E. Winarko, "Modification of Stemming Algorithm Using A Non Deterministic Approach To Indonesian Text," *IJCCS Indones. J. Comput. Cybern. Syst.*, vol. 13, no. 4, p. 379, Oct. 2019, doi: 10.22146/ijccs.49072.
- [42]. M. A. Muchtar *et al.*, "Separation of Basic Words in Angkola Batak Text Documents using Enhanced Confix Stripping Stemmer Case: Mandailing Ethnic," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 648, no. 1, p. 012024, Oct. 2019, doi: 10.1088/1757-899X/648/1/012024.
- [43]. I. G. M. Darmawiguna, G. A. Pradnyana, and G. S. Santyadiputra, "The Development of Integrated Bali Tourism Information Portal using Web Scrapping and Clustering Methods," *J. Phys. Conf. Ser.*, vol. 1165, p. 012010, Feb. 2019, doi: 10.1088/1742-6596/1165/1/012010.
- [44]. M. H. Ali and F. Rahutomo, "MANHATTAN DISTANCE AND DICE SIMILARITY EVALUATION ON INDONESIAN ESSAY EXAMINATION SYSTEM," *JUPI J. Ilm. Penelit. Dan Pembelajaran Inform.*, vol. 4, no. 2, p. 156, Dec. 2019, doi: 10.29100/jupi.v4i2.1398.