

Basic Word Extraction Algorithm Based on Morphological Rules for Balinese Texts

I Made Wahyu Guna Negara^{a1}, Ngurah Agus Sanjaya ER^{a2}

^{a1}Informatics Engineering, Faculty of Math and Science, University of Udayana
South Kuta, Badung, Bali, Indonesia
¹wahyugunanegara@gmail.com
²agus_sanjaya@unud.ac.id

Abstract

Stemming is the process of extracting the root word of an affixed word. The process is intended to reduce the variations in the word. In this research, we are interested in applying stemming on Balinese language. Previous works on stemming of the Balinese language applied rule-based method but only prefix and suffix were considered. Moreover, the rules were constructed without providing much attention to the morphology of the Balinese language. Rule-based method can be verified and validated with ease on simple problem but fail to do so on problems with high complexity such as Balinese language. To overcome the weaknesses of rule-based stemming on Balinese language, we propose a method that reduce all variations of affix on Balinese language by combining the rule-based approach and the Balinese language morphology. Based on experiments carried out, our proposed method obtained an average stemming accuracy of 99% which is better than 96.67% achieved by the previous method.

Keywords: *Stemming, Balinese language, Rule-based*

1. Introduction

Indonesia is an archipelago with a variety of cultures, ethnicities and religions. Indonesia's diverse population consists of various ethnic groups with various regional languages and a variety of different cultural backgrounds. One of the riches of Indonesian culture is the regional language.

Balinese is one of the regional languages in Indonesia with a large number of users. Based on the number of users, the Balinese language can be classified as a large regional language because it is supported by a large community of users, which is used by approximately three million users [1]. Balinese language is still sustainable until now because it is still maintained, fostered and used by users in various aspects of life. Balinese as one of the regional languages is still used as an oral and written communication tool. As an oral language, Balinese is used in the communication process both on official and unofficial topics [2].

In language activities, communication can be divided into two types, namely communication with intermediaries of spoken language and communication with intermediaries of written language. Communication of spoken language is the process of delivering and receiving information from information givers to recipients of information without using any intermediaries, while written language communication is the process of delivering and receiving information from information givers to recipients of information using intermediaries [3]. One applies of language as a communication tool is the use of written language in print media, in this case specifically in the form of Balinese-language documents.

In relation to culture, Balinese language is the most appropriate tool to learn and explore Balinese culture. This is useful for fostering, maintaining and developing regional and national culture. Maintenance of the Balinese language can be carried out by utilizing it in everyday life. In Bali, the Balinese language is not only used as a medium of oral communication, but Balinese is also used in written forms, namely Balinese-language literary works. This literary work includes traditional literature and modern literature.

With the rapid information and communication technology, it is expected that it will provide convenience to obtain fast and accurate information, especially information in Balinese language.

Information Retrieval (IR), or information retrieval system, is a study area that studies the search method and the separation of material in the form of text documents from unstructured data that meets relevant information needs. As the available Balinese language documents are available, the resources needed by the IR algorithm are getting bigger, and the efficiency decreases. Therefore, an optimization process is needed to maintain the high level of effectiveness while maintaining efficiency. One way that can be done is by *stemming*. *Stemming* aims to reduce variations in words in the form of basic words [4], [3]. In Balinese the language morphology is known. Morphology is part of linguistics, especially grammar. The object of the analysis includes grammatical units at the morpheme level and words [1] that study the form, structure, and classification of words. The form of words and class of words in the Balinese language can change because they have an affix to the basic word. Affix words in the Balinese language can be distinguished according to their place attached to the basic form or origin, namely prefix, suffix, insertion, confix or simulfiks, and combinations of affix [5].

Nata and Yudiastra [6] previously stemming in Balinese using rule-based methods [5]. In this research, affix that are carried out are only the prefix and suffix, while the affix that cause disambiguation such as inserts, cnfixes or simulfiks and combinations of affix are not passed. Other research to stemming the language of Bali is done by Subali and friends [7]. The method used in the study is a combination of rule-based methods and *n-gram* to get the basic word in the Balinese language. In this study obtained an accuracy of 96.67% of the 10 queries tested.

Another study on *stemming* using rule-based methods, but implemented in Indonesian is the algorithm Nazief and Andriani [8] the Porter algorithm [9]. The application of Porter's algorithm for *stemming* Indonesian language has advantages in terms of the time needed to complete the entire process. However, this algorithm produces better accuracy compared to the algorithm built by Nazief and Andriani. The Nazief and Adriani algorithms refer to Indonesia's morphological rules so that they get an accuracy that reaches 92.8 %

Rule-based methods have the advantage of being applied to a simple domain, so rule-based is easy to verify and validate. On the other hand, this method has weaknesses when applied to domains with a high level of complexity. If a rule-based system cannot recognize the rules given, there will be no results obtained by Grosan and Abraham [10]. To overcome the weaknesses of rule-based *stemming*, it is necessary to store data and create rules based on Balinese language morphology rules.

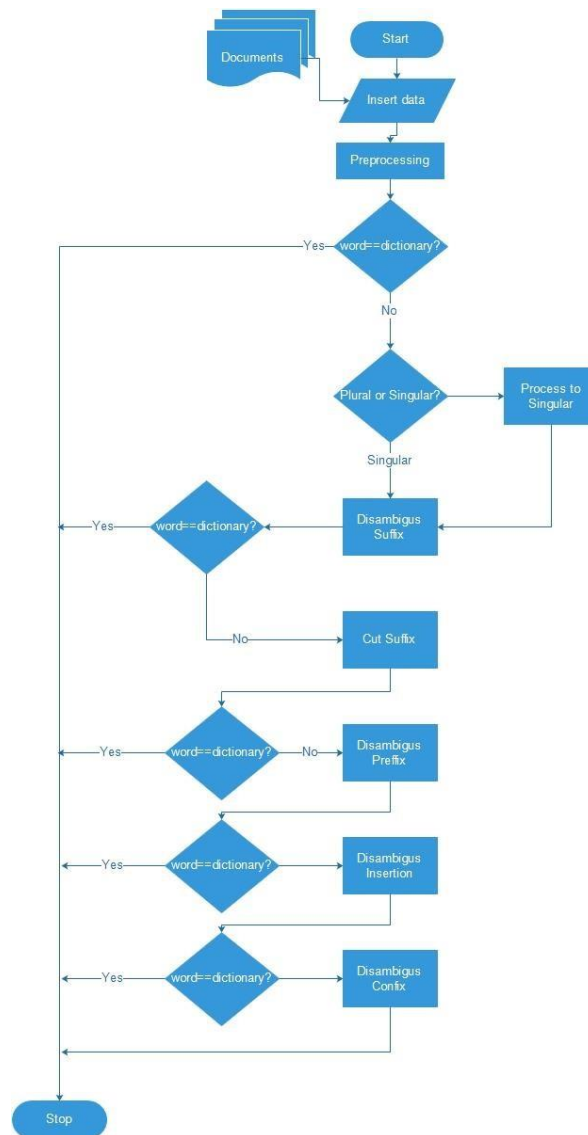
In this research, a stemming method was developed which overtook all variations of the affix on the Balinese language by combining a rule-based approach and paying attention to the Balinese language morphology rules. To prove the proposed method can provide optimal stemming accuracy, a series of tests were conducted. The test was conducted against the results of the proposed stemming method with the Balinese stemming method that had been done. Besides that, testing is done on the diversity of the number of test data which aims to see the stability of the accuracy of the method proposed.

2. Research Methods

The process stage in the development of morphological rule-based *stemming* extraction algorithms for Balinese language texts includes the initial processing (*preprocessing*) of data, *stemming* based on the morphological rules of the Balinese language, then producing outputs of *stemming* and outputs of data record if the data is not listed in the vocabulary. In the first stage data collection of basic Balinese words is carried out which is then stored in the Balinese vocabulary dictionary. The initial processing phase of the text includes *tokenization*, *data cleaning*, and *case folding*. The results of the initial processing phase of the text will be used as input for the *stemming* process based on the morphology of the Balinese language. The process flow from the development of this algorithm can be seen in Figure 1.

2.1. Dictionary of Balinese basic word

The collection of basic words of the Balinese language is obtained from the book entitled "Kumpulan Satua Bali" as well as the website www.kamuslengkap.com. On the website www.kamuslengkap.com the basic words of Balinese are obtained by scraping so that they get the basic word of Balinese in the amount of 1806 basic words of Balinese.



Picture 1. System Flow Chart

2.2. Balinese Morphology

Etymologically the word morphology comes from the word (*morph*) which means form and word (*logy*) which means science. So, the word morphology literally means the science of form. In linguistic studies, morphology means the linguistic branch in the ins and outs of words and their changes and the impact of these changes on the meaning and class of words [2].

Classification of words as can be done by looking at the behavior of grammatical words in a more complex level, namely at the level of phrases and sentences.

2.3. Ambiguous Suffix Validation

The Balinese language has eight suffixes which are located at the end of the Balinese word or in Balinese language is called (*pengiring*). The suffix in Balinese consists of -ang, -in, -an, -a, -n, -ing, -e, -ne. But eight suffixes there are have a derivative when meeting with vocal or consonant in accordance with the morphology book of the Balinese language [2]. The rules are divided into 3 ambiguous suffix rules that pay attention to sequences to improve accuracy based on observations made by researchers. The suffix rule is applied to eliminate every suffix in Balinese then is validated using a basic word dictionary. The suffix rule table can be seen in Table 1 -3.

Table 1. Rules for Ambiguity Suffix 1

Suffix	Rule	Examples of Words	Result
-e	Vocal and Consonant	dokare	dokar
-ne	Vocal and Consonant	lumurne	lumur
-nne	Vocal and Consonant	bajunne	baju

Table 2. Rules for Ambiguity Suffix 2

Suffix	Rule	Examples of Words	Result
-a	Vocal and Consonant	Jemaka	Jemak
-na	Vocal	Raina	Rai
-ina	Vocal and Consonant	Jemakina	Jemak

Table 3. Rules for Ambiguity Suffix 3

Suffix	Rule	Examples of Words	Result
-n	Vocal	Bukun	Buku
-in	Vocal and Consonant	Miluin	Milu
-nin	Vocal	Belinin	Beli
-nan	Vocal and Consonant	Gedenan	Gede

2.4. Remove suffix

After the ambiguous suffix validation is done and if no matching words are found in the Balinese word dictionary the processed word is returned to the initial word (recording). Then the suffix removal process is carried out before proceeding with the ambiguous prefix validation process. The rules for deleting suffix are divided into 3 rules that pay attention to sequences in order to improve accuracy based on observations made by researchers. The suffix rule is applied to delete each suffix in Balinese, the table of rules for suffix delete can be seen in Table 4 - 6.

Table 4. Rules for Removing Suffix 1

Suffix	Rule	Examples of Words	Result
-ing	Vocal and Consonant	Jeroing	Jero
-ning	Vocal and Consonant	Purnamaning	Purnama
-n	Vocal and Consonant	Bajun	Baju
-in	Vocal and Consonant	Miluin	Milu
-nin	Vocal and Consonant	Belinin	Beli
-ina	Vocal and Consonant	Jemakina	Jemak

Table 5. Rules for Removing Suffix 2

Suffix	Rule	Examples of Words	Result
-e	Vocal and Consonant	Dokare	dokar
-ne	Vocal and Consonant	Lumurne	lumur
-nne	Vocal and Consonant	Bajunne	baju

Table 6. Rules for Removing Suffix 3

Suffix	Rule	Examples of Words	Result
-ang	Vocal and Consonant	Jemakang	Jemak
-nang	Vocal and Consonant	Gedenang	Gede
-yang	Vocal and Consonant	Satuayang	Satua
-na	Vocal and Consonant	Abana	Aba
-nan	Vocal and Consonant	Gedenan	Gede
-a	Vocal and Consonant	Jemaka	Jemak

2.5. Ambiguous Prefix Validation

The Balinese language has thirteen prefix forms. In Balinese the prefix is called (*pengater*). The prefix in Balinese consists of N-(anasuara), ma-, ka-, sa-, pa-, pi-, a-, pra-, pari-, pati-, maka-, saka-, kuma-. However, of the thirteen prefixes, there are prefixes that have a derivative when they meet with vocal or consonant according to the Balinese morphology book [2].

These rules are divided into 10 ambiguous prefix rules that pay attention to sequences to improve accuracy based on observations made by researchers. The prefix rule is applied to eliminate each prefix in Balinese which is then continued with validation based on a basic word dictionary. The table prefix rules can be seen in Table 7 - 16.

Table 7. Rule N-(anasuara) = ng

Prefix	Rule
ng-	Vocal and Consonant
ng-	(vocal) = kV
ng-	(vocal) = gV
-nga	Vocal and Consonant

Table 9. Rule N-(anasuara) = n

Prefix	Rule
n-	(vocal) = tV
n-	(vocal) = dV

Table 8. Rule N-(anasuara) = ny

Prefix	Rule
ny-	(vocal + a,y,r,l) = cV
ny-	(vocal) = jV
ny-	(voca) = gV

Table 10. Rule Ma-

Prefix	Rule
m-	(vocal) = bV
mam-	vocal and consonant
mam-	(vocal and consonant) = pV
mam-	(vocal and consonant) = bV

m-	vocal
m-	(vocal) = pV
ma-	consonant

Table 11. Rule Ka-

Prefix	Rule
ka-	Vocal and Consonant
k-	vocal
ko-	(vocal and consonant) = uV

Table 12. Rule Sa-

Prefix	Rule
sa-	Vocal and Consonant

Table 13. Rule Pi-, Pa-, Pra-, Pari-, Pati-

Prefix	Rule
pat-	Vocal and Consonant
pak-	Vocal and Consonant
pik-	Vocal and Consonant
pi-	(vocal) = bV
pati-	Vocal and Consonant
pari-	Vocal and Consonant
pa-	Vocal and Consonant

Table 14. Rule a-

Prefix	Rule
ng-	Vocal and Consonant

Table 16. Rule kuma-

Prefix	Rule
kuma-	Vocal and Consonant

Table 15. Rule pra-

Prefix	Rule
pra-	Vocal and Consonant

2.6. Infixes Validation

Balinese has four forms of infix. In Balinese the infixes is called (*sese/lan*). These rules are divided into 4 ambiguous rules that pay attention to sequences to improve accuracy based on observations made by researchers. The infix rule is applied to eliminate every infix in Balinese then validation is done in the basic word dictionary. Table of infix rules can be seen in Table 17.

Table 17. Rule Infix

Infix	Rule	Examples of Words	Result
-in-	Vocal and Consonant	Tinulung	Tulung
-um-	Vocal and Consonant	Rumaksa	Raksa
-el-	Vocal and Consonant	Telusuk	Tusuk
-er-	Vocal and Consonant	Gerigi	Gigi

2.7. Confix Validation

Balinese has four forms of confix or affix in the form of a combination of prefix and suffix. The following are the rules that are applied to eliminate every confix in Balinese which can be seen in Table 18.

Table 18. Rule Confix

Confix	Rule	Examples of Words	Result
pa-an	Vocal and Consonant	Pasirepan	Sirep
ma-an	Vocal and Consonant	Majemakan	Jemak
ka-an	Vocal and Consonant	Kasugihan	Sugih
bra-an	Vocal and Consonant	Bramahna	mah

2.8. System Evaluation

In this research, testing of the classification results from test data will be tested. The calculation is done by calculating the amount of data that gets the correct classification with the total test data using equation (1) Test data is classified as correct if the results of stemming are the same as those contained in the word dictionary.

$$\frac{\text{Correct Word Classification}}{\text{All Test Word Data}} \quad (1)$$

3. Result and Discussion

To prove the proposed method can work optimally, a series of tests are carried out, such as comparing the results of the (*stemming*) method proposed with the previous method, comparing the accuracy of each document of test data, as well as testing the stability of accuracy based on the large amount of data.

Implementation and testing of the system is done in the software development environment as follows: Windows 8 64-bit OS, Intel (R) Core (TM) i5-3210M processor, 4.00GB RAM, IDE pycharm Community, Language PYTHON 3.6 with Packages NLTK 3.4 and Regex.

The vocabulary data used is 10,279 lists of basic words collected from the book with the title "Kumpulan Satua Bali" as well as the websites www.kamuslengkap.com and www.dictionary.basabali.org. At the testing stage we used 60 queries. The 50 queries used are pieces of Balinese language stories "I Belog" and "Pan Belog", and the remaining 10 queries obtained from Subali et al. Research, to compare accuracy with the proposed method. The query can be seen in the table 19, where in the 51st to 60th rows is a query used to compare the accuracy of the proposed method.

Table 19. Query

No	Query	No	Query
1	Ada tutur satua anak belog.	33	Nangingke ia tusing pesan bani tulak teken pamunyin kurenanne Kenken ja panguduh kurenanne setata ia takut dogen.
2	Baan belogne ia adan I Belog.	34	Ditu lant as ia ka peken.
3	Sedek dina ia tonden meli bebek ka peken teken meme.	35	Kacrita sanapake ia di peken, nglaut ia ngojog dagang bebek.
4	Ditu lant as ia jemak meme pis.		
5	Jero niki jinah, tiang meli bebek dua.		
6	Bebek ukud aji Rp. 4000.		

<p>7 Lantas meme buin ngomong, kema jani cai enggal ka peken, terus meli be dua di tongos dagang bebek.</p> <p>8 suba I Belog neked di peken, kema-mai ia ningal dagang bebek sakewala ia enjuh pipis dasa tali rupiah.</p> <p>9 Lantas dagang bebek maang I Belog susuk bui Rp. 2000.</p> <p>10 suba maan meli bebek lantas I Belog mulih.</p> <p>11 crita ia mulih, tur liwat tukad linggah</p> <p>12 Ditu lantas bebek leb. Maka dua bebek ngelangi di tukad.</p> <p>13 I Belog bengong ningal bebek kambang tur ia ngrengkeng kene.</p> <p>14 Beh, bebek puyung bakat beli.</p> <p>15 Awak nagih bebek mokoh tur baat, sakewala bebek puyung baang.</p> <p>16 I Dewek belog.</p> <p>17 Lantas bebek tusing ejuk tur kalah mulih.</p> <p>18 suba I Belog neked jumah, ajin baan meme tuara ngaba bebek.</p> <p>19 Meme ngomong, ih belog cen bebek? saut I Belog, "maan ja icang meli bebek, nanging puyung icang adepin teken dagang bebek.</p> <p>20 Lantas bebek leb di tukad, tur ngelangi.</p> <p>21 Buin laut ulah icang sawireh meli bebek puyung tuara ada guna.</p> <p>22 Ditu lantas I Belog welang baan meme.</p> <p>23 Keto upah anake belog, tuara ngresep teken munyi.</p> <p>24 Bebeke mula kambang yan ia lebang di tukake dalem</p> <p>25 Sedek dina anu Pan Belog tundena ka peken teken kurenanne meli bebek dadua, lakar tampaha anggona banten, krana matuane buin maninne tutug abulan pitung dina.</p> <p>26 Kene munyinne teken Pan Belog, "Ih, Bapanne, kene cai suba nawang, buin mani I Bapa tutug abulan pitung dina, buina icang repot pesan magarapan, tusing icang maan magedi kija-kija.</p> <p>27 Kema jani cai ka peken meli bebek dadua, pilihan men meli bebék ane mokoh-mokoh, tur baat-baat beli.</p> <p>28 Ne pipis aba. Nah kema suba cai majalan ka peken!", kéto abet kurenanne.</p> <p>29 Pan Belog anak mula ia jlema kaliwat belog pesan, turin mawuwuh-wuwuh kabeloganne, krana ia tusing pesan taen bareng-bareng ngajak anak ririh mapaomongan.</p> <p>30 Kalingan ke ia maan mapaomongan ngajak anak lenan, kadirasa ia ngenot dogenan ia suba takut.</p> <p>31 Teked ditu, tusing ja ia makeneh nakonang ajin bebek, wiadin nawah, sakewala kene kone munyinne teken i dagang bebek, "Jero dagang bebek, niki jinah, icen tiang bebek kakalih!" Ditu Pan Belog ngenjuhang ringgit aketeng, nangingke Pan Belog tusing nawang ento madan ringgit.</p> <p>32 I dagang bebek ngon ia teken tingkah anake mablanja buka keto, tuara nakonang aji malu, jag maang pipis, tur nagih bebek. Nanging mara tawanga teken i dagang bebek Pan Belog jlema deeng, ditu lantas makenyir tur encol lantas ia maang Pan Belog bebek, ane mokoh-mokoh tur baat-baat dadua.</p>	<p>36 Pan Belog nyemakin bebeke ento tur lantas ia mlipetan mulih, tusing ia buin nagihin i dagang panyusuk.</p> <p>37 Kacrita pajalan Pan Beloge ngamulihang, ngentasin tukad linggah.</p> <p>38 sawetara mara ia neked di tengah tukade, laut ngeleb bebeke maka dadua, tur lantas nglangi.</p> <p>39 Pan Belog bengong ia ngenot tingkah bebeke buka keto, laut ngrengkeng padidiana.</p> <p>40 Béh, aeng ja jailné dagang bebeke ento teken deweke.</p> <p>41 awake nagih bebek maisi, nget bulu dogen awake adepina.</p> <p>42 Aéng ja dueg dagange ento melog-melog deweke.</p> <p>43 Suud keto, lantas bebeke ento ulaha.</p> <p>44 Suba ia neked jumahne, ajinanga lantas ia teken kurenanne tuara ngaba bebek, laut ia matakon, "Ih Bapanne, ento dadi cai matalang mulih dija bebeke, sing maan cai meli bebek keto?" Mesaut Pan Belog, "Maan ja icang meli bebek, nanging bebek puyung adepina teken dagange.</p> <p>45 Jani ia suba kakutang bebeke ento di tukade.</p> <p>46 Buin matakon kurenanne nyesedang krana ia tusing ngerti baana kurenanne ngorahang bebek puyung. Kene munyinne, "Puyung-puyung kenken ja bebeke Bapanne?" Ditu Pan Belog nuturang saunduk-undukne di tengah jalan.</p> <p>47 Baane bebeke tuara nyilem, ento krananne dadi bebeke orahanga puyung.</p> <p>48 Bengong turing sebet kurenanne ningehang tutur Pan Beloge buka keto.</p> <p>49 Ditu lantas ia ngeling, manganang dewekne ngelah somah kaliwat belog ludin lacur.</p> <p>50 Pipis ilang bebek tuara bakat.</p> <p>51 i meme ngajak i bape negakin sepeda semengan kuluke ngongkong.</p> <p>52 palajahin made nganti mamuduh teken i luh.</p> <p>53</p> <p>54 telapak liman made beseh ulian dibi majaguran.</p> <p>55 made lan i luh makurenan duang dasa tiban.</p> <p>56 nyen ento menyuling di jabe tengah?</p> <p>57 sire sane maborbor lulune?</p> <p>58 mangorahang isin hati beline</p> <p>59 dadong dauh ngelah siap putih lan sampi aukud.</p> <p>60 ngiring lestariang basa baline.</p>
---	--

In this research the authors conducted several tests by providing input in the form of a document with a variety of the number of sentences so that the basic word results obtained from the words in the sentence entered. The results of the sentence document with the basic word are then tested for similarity with the document word sentence that is built manually.

The first test, the authors conducted a test of the stability of accuracy with a variety of the number of queries each document. In this test the author uses 5 documents with a number of different queries. The first document contains 10 queries, 1 - 10 queries in the table 19, the second document contains

20 queries, 1 - 20 queries in the table 19, the third document contains 30 queries, 1 - 30 queries in the table 19, the fourth document contains 40 queries, 1 - 40 queries in the table 19, and the fifth document contains 50 queries, 1 - 50 queries in the table 19. From the test, the accuracy of the proposed method shown in the table 20.

Table 20. Testing Accuracy Results 1

Documents	Number of Sentences	Result
Doc1.txt	10 Sentences	99.51%
Doc2.txt	20 Sentences	99.47%
Doc3.txt	30 Sentences	99.29%
Doc4.txt	40 Sentences	99.23%
Doc5.txt	50 Sentences	99.22%

The second test compares the method developed by Subali et al. With the proposed method of the author. In this test the author uses 10 queries that are used in research Subali et al. From these tests the results obtained are shown in table 21.

Table 21. Testing Accuracy Results 2

Query	Subali et al	Proposed Method
Q51	100%	100%
Q52	100%	100%
Q53	66,67%	100%
Q54	100%	100%
Q55	100%	100%
Q56	100%	87.49%
Q57	100%	83.33%
Q58	100%	100%
Q59	100%	100%
Q60	100%	100%
Average	96,67%	97,4%

From the table above it can be seen the range of accuracy results by testing the similarity of the results of documents built by the system and documents that are built by the system have an average accuracy of 99%. Accuracy results in increasing number of sentences result in smaller accuracy, but stable at 99% percentage. When compared with the research *stemming* conducted by Subali et al [7] with the same testing data, then the method developed in this research gets higher accuracy. However, in Subali et al research there is a difference in the stages of calculating accuracy, where in Subali et al research only calculates accuracy based on words that have affixes only, but in this study calculates accuracy based on all words in the query. This aims to find out whether the system developed can convert affixed words into Balinese root words, and not change non-affixed words.

4. Conclusion

In this research, researchers combined the rule-based method and the morphology of the Balinese language. The rule-based method is used to form rules that encompass all variations of affix. Based on a list of basic and fiftieth words *query* given between the proposed method and the previous method, the proposed method obtains accuracy *stemming* is better, which is 99% compared to the Nata and Yudiastra methods that obtain 75%. This is because the rules in the Nata and Yudiastra methods only cause two variations of the affix, namely the prefix and the suffix [11]. And this research has a better accuracy compared to the research of Subali et al who obtained an accuracy of 96.67%, this is because in the research of Subali et al only used 1000 word dictionaries.

References

- [1] W. S. Gitananda, "Serba nasalisasi n-(atau ng-?) dalam bahasa bali," *DHARMASMRTI: Jurnal Ilmu Agama dan Kebudayaan*, vol. 16, no. 01, pp. 1-7, 2017.
- [2] I. W. Bawa and I. W. Jendra, *Struktur Bahasa Bali*, vol. 70, Pusat Pembinaan dan Pengembangan Bahasa, Departemen Pendidikan dan Kebudayaan, 1981.
- [3] Y. D. Pramudita, S. S. Putro and N. Makhmud, "Klasifikasi Berita Olahraga Menggunakan Metode Naive Bayes dengan Enhanced Confix Stripping Stemmer," *Jurnal Teknologi Informasi dan Ilmu Komputer*, vol. 5, no. 3, pp. 269--276, 2018.
- [4] H. B. Patil and A. S. Patil, "MarS: A rule-based stemmer for morphologically rich language Marathi," in *2017 International Conference on Computer, Communications and Electronics (Comptelix)*, IEEE, 2017, pp. 580-584.
- [5] F. Z. Tala, "A study of stemming effects on information retrieval in Bahasa Indonesia," *Institute for Logic, Language and Computation, Universiteit van Amsterdam, The Netherlands*, 2003.
- [6] G. N. M. Nata and P. P. Yudiastra, "Stemming teks sor-singgih Bahasa Bali," *E-Proceedings KNSI&I STIKOM Bali*, pp. 608--612, 2017.
- [7] M. A. P. Subali and C. Fatchah, "Kombinasi Metode Rule-Based dan N-Gram Stemming untuk Mengenali Stemmer Bahasa Bali," *Jurnal Teknologi Informasi dan Ilmu Komputer*, vol. 6, no. 2, pp. 219--228, 2019.
- [8] M. Adriani, J. Asian, B. Nazief, S. M. Tahaghoghi and H. E. Williams, "Stemming Indonesian: A confix-stripping approach," *ACM Transactions on Asian Language Information Processing (TALIP)*, vol. 6, no. 4, pp. 1--33, 2007.
- [9] M. F. Porter, "An algorithm for suffix stripping," *Program*, vol. 14, no. 3, pp. 130--137, 1980.
- [10] A. Abraham, "Rule-Based expert systems," *Handbook of measuring system design*, 2005.
- [11] V. Levenshtein, "Binary Codes Capable of Correcting Deletions, Insertions and Reversals," *Soviet Physics Doklady*, vol. 10, p. 707, 1966.
- [12] I. W. O. Granoka, *Tata Bahasa Baku Bahasa Bali*, Pemerintah Propinsi Tingkat I Bali, 1996.