







# A Balinese folklore digital portal with natural language processing framework

Ngurah Agus Sanjaya ER<sup>1</sup>, Ni Luh Nyoman Seri Malini<sup>2</sup>, I Made Widiartha<sup>3</sup>,  
Pijar Candra Mahatagandha<sup>4</sup>

<sup>1,3,4</sup> Informatics Department, Udayana University, Bali, Indonesia

<sup>2</sup> English Department, Udayana University, Bali, Indonesia

\*agus\_sanjaya@unud.ac.id

**Abstract.** Balinese story is one of the cultural heritages that are filled with local wisdom. These stories were often passed on orally from one generation to the next. The advancement of information technology and more accessible internet connection have opened opportunities for preserving these Balinese stories in a digital portal. In this research, we developed a digital portal which is equipped with advanced features built with a natural language processing framework. These features include grouping of stories with similar topics, summarizing the story by extracting meaningful sentences and applying a Genetic algorithm to maximize the features' weight, and automatically identifying the characters in each story. These features are integrated into a server-side Balinese natural language processing framework. The framework is applied to the collection of Balinese stories on the portal and any new story added to the collection. Experiment results on the portal show that as many as 77.83% of respondents gave a reasonable opinion of the portal.

**Keywords:** framework, digital portal, natural language processing, Balinese.

## 1 Introduction

Language is an essential element of culture that needs to be preserved [1]. The Balinese language itself is a regional language in the province of Bali which characterize the culture of the Balinese people. The Balinese language is essential for Balinese people to communicate daily and pass on Balinese stories orally from one generation to the next. Moreover, preserving Balinese stories is no less critical due to the local wisdom of Balinese people contained in the stories. The characterization of the characters in the stories implicitly portrays local wisdom. For example, the story “I Lutung teken I Kekua” communicates a hidden message to the readers, i.e., to always be respectful towards others and not over pride themselves. In this story, “I Kekua” is the moral of the story, i.e., the good character.

This research aims to digitally preserve the Balinese culture, specifically the local wisdom in Balinese stories. We develop a server-side Balinese natural language processing framework. In addition to implementing the framework for our digital portal of Balinese stories, we also provide advanced features, including clustering stories with

similar topics, summarizing, and automatically identifying characters in the story. We will first briefly explain our work in developing the Balinese NPL library for Python, then revisit our work which implements the library and provide the top features on our portal. We conclude the section with a summary of our contributions.

To assist in our goal of preserving the Balinese language, we begin with developing a natural language processing library for the Balinese language in Python. In the library, we provide stemming for the Balinese language, named entity recognition (NER) for person, location, and time expression, along with a part-of-speech (POS) tagger.

Previous research on Balinese stemmer implemented rule-based [2] and n-gram [3] methods. In [2], rules are developed to remove prefixes and suffixes from words to generate the “stem” words. However, the authors did not consider words generated by adding an infix or affix (combining prefix and suffix) into the stem words. We overcame the problem in [4] by developing rules for words generated by adding infix, affix, and simulfix (a type of affix that changes one or more existing phonemes).

Named entity recognition holds a vital role in character identification in stories. This subtask of information extraction locates, and highlights named entities (people, organizations, places, etc.) in unstructured data. NER has been applied to various languages, including Arabic [5], Greek [6], and Punjabi [7]. For the Indonesian language, NER research has been extensively developed, starting by using convolutional neural networks [8], co-training [9], and ensemble methods [10]. In addition, NER has also been applied to specific domains such as biology [11], medicinal plants [12], finance [6], accounting [13], and short text messages [14]. However, research on named entity recognition for the Balinese language has yet to receive similar attention. Due to this reason, we have implemented a rule-based name entity recognition for the Balinese language to identify time [15], location [16], and person [17].

The task of a POS tagger is to classify words in a corpus to their appropriate part of the speech tag. POS taggers for the Indonesian language have been built using rule-based methods and Hidden Markov Models (HMM) [18], conditional random fields [19], and Brill Tagger [20]. In [21], the authors have conducted a complete and detailed comparison of the previously mentioned POS tagger methods. In [22], we developed a POS tagger for the Balinese language, inspired by [18].

Numerous automatic text summarization systems have been developed for English [23], [24], and Indonesian [25], [26]. These automated systems have been applied to various fields, including scientific articles [25] and news [26], [27]. The techniques used to perform the summarization are widely varied, including artificial neural networks [24], deep learning [23], Latent Semantic Analysis (LSA) [28], Latent Dirichlet Allocation (LDA) [29] or genetic algorithm [30], [31]. In [32], we implemented the genetic algorithm to optimize the features’ weight of sentences in a text. The weight combination of the features is directed towards the ROUGE-N [33] objective function. The genetic algorithm applies selection, mutation, and cross-over of the chromosomes iteratively to find the optimal weight combination.

Documents often have similar underlying topics that are hidden from the readers. These latent topics can be exploited to group the collection of documents. For the Indonesian language, grouping documents based on their topics is achieved by applying

classification [34], [35] or clustering [36], [37] method. The main difference between the two methods is that the former assumes that a document exclusively belongs to a group, while the latter relaxes this constraint. Previous research on grouping documents in the Indonesian language implemented LDA [38], PLSA [39], and K-Means [40]. In [41], we first identified the topic’s distribution of each document in our corpus and then grouped them by applying an agglomerative hierarchical clustering method.

To conclude the section, we list our contributions as follows. We first develop a natural language processing framework for Balinese documents. The framework is the culmination of our previous research, which can process a Balinese document by utilizing our Balinese library. The documents can then be automatically summarized and grouped based on their topic’s distribution. We then implement the framework on the server and develop a portal for Balinese stories. The portal acts as a crowd-sourced platform in which anyone who has signed up can contribute new stories.

## 2 Research Methodology

This research applies a naive sequential method, as shown in Figure 1. We begin with a literature study to identify previous research on methods necessary to provide advanced features for our Balinese stories’ portal. Most of the research we investigate is related to natural language processing for the Balinese language and has been covered in the previous section. We also summarized our works aimed at improving the result of the related research and providing tools that are not available but necessary to the portal.



**Fig. 1.** Research Method

There are two sub-stages in the development stage, i.e., framework development and portal development. The former sub-stage is focused on developing a workflow for stories consisting of preprocessing them by employing the Balinese natural language library we have developed in our previous research, creating a summary of each story, identifying the story's main characters, and grouping stories based on their latent topics. After programming the framework, we launched it on the server as the basis for the Balinese stories’ portal. The portal acts as a gateway for readers interested in reading Balinese stories and contributing new stories to the portal.

The subsequent stage is the trial and testing stage. The testing of the entire system is carried out by adding a new story directly from the portal and investigating each process in the workflow. Tests on each process are conducted to identify issues, bugs, or errors. They are immediately addressed upon finding issues, bugs, or errors until no more problems are identified. It is critical to remove any issues before we enter the evaluation stage. In the evaluation stage, we involve the general public, who are asked to assess the developed system. The assessment is carried out online, and respondents must respond to several questions related to system features. The evaluation

stage is necessary to obtain an unbiased opinion and receive feedback to improve the system. The conclusions of the research and suggestions for further research development are based on the result of the evaluation stage.

### 3 Results and Discussions

The objective of this research is two-fold. The first objective is to develop a framework for processing documents in Balinese, particularly Balinese stories. The framework then applies the necessary preprocessing before producing a summary of the story and identifying the story's characters. Grouping the Balinese stories is another responsibility of the framework, carried out when specific criteria are satisfied, i.e., the number of stories added to the portal and the number of days after the last grouping occurred. The second objective of the research is to develop a portal for Balinese stories with the developed framework as the backend. Within the portal, visitors can read the collection of Balinese stories or register to the portal and become registered users who can contribute new stories to the collection. The portal is evaluated by conducting an online survey of the users. The following sections will highlight details of the developed framework and portal.

#### 3.1 Framework

We develop a framework for processing any documents written in Balinese. In this research, the Balinese document is in the form of a story that contains local wisdom. Three steps are carried out sequentially for each Balinese document in the collection, namely preprocessing, summarization, and character identification. Previously unseen documents added to the collection also experience the same process. The papers are then grouped based on their latent topics in a bottom-up manner [41] since the last grouping occurred.

We begin the preprocessing step by conducting tokenization of the documents. Texts on the documents are scanned and split into sentences. Each sentence is then broken up into tokens, i.e., words in the Balinese language, for further processing. Tokens on each sentence are then compared to a stop words list. This predefined list contains ubiquitous words in the Balinese language that are deemed “not important” and can thus be ignored. Any token which appears in the list is removed and will not be processed further. The remaining tokens of each sentence are then fed into our Balinese stemmer to identify the stem words of each token. Tokens within a sentence originating from the exact stem words will only be represented by a single word. The result of the preprocessing step is sentences with distinct tokens.

Composing the summary of each story automatically is one of the several advanced features provided by our portal. We implement an extractive method to create the summary of the story by identifying the most critical sentence in each paragraph. To accomplish the objective, we first define features of a sentence being examined, i.e., the most frequent words (positive keywords), the least frequent words (negative keywords), words in the sentence which also appear in the title, words in the sentence that

commonly occur in other sentences within the same document, and cosine similarity between the sentence and the remaining sentences within the same document. Each sentence feature is basically the normalized count of words that satisfy the corresponding condition. The score for a sentence is calculated as the multiplication of the feature value with the weight of the feature. Since the importance of the features is different, we must then investigate the true significance of the features. We approach the issue as an optimization problem that must satisfy an objective function. In this research, we utilize the genetic algorithm to find the correct combination of the features' weight, where the solution is guided toward optimizing the average accuracy of the summaries. The baseline of the summaries is created manually beforehand with the help of an expert. The expert applies the same extractive method when composing the summaries. In this research, we define the accuracy of a summary as the unigram ROUGE score. The ROUGE score is calculated as the number of words that appear in the manual summary created by the expert and the synthetic one produced by our method. According to the genetic algorithm, we define the weight of a feature and the combination of five features' weight as gen and chromosome, respectively. The genetic algorithm iteratively seeks the best combination of features' weight through mutation and crossover of genes between pair of chromosomes. Finally, after the iterative process has been concluded, we utilize the best combination of features' weight for summarizing new stories added to the collection. We refer the readers to a more detailed explanation of the summarization process in [32].

Named entity recognition in Balinese documents has yet to receive considerable attention. Consequently, we have developed a rule-based Balinese name entity recognition to identify a person [17] in Balinese text. A naïve approach to generating the rules is directly observing the corpus and recognizing the similarities of persons' names in Balinese. Several morphological structures in the Balinese language can be exploited to create the rules [42]. "I" and "Ni" are widespread articles in Balinese names, for example, "I Gusti Putu Ardana" and "Ni Komang Ayu". The articles "I" and "Ni" are used to differentiate male and female gender, respectively. A different morphological structure that can be exploited is using articles to express birth order in a family. These articles are "Wayan", "Putu", and "Gede" for the first child, "Made", "Nengah", and "Kadek" for the second, "Nyoman" or "Komang" for the third, and "Ketut" for the fourth. In addition, Balinese names are also derived from the person's social class, i.e., "Brahmana", "Ksatria", "Waisya", and "Sudra". These social classes represent religious leaders, kings and nobles, traders, and peasants. Each social class possesses its distinctive naming, for example, "Ida Bagus" or "Ida Ayu" for "Brahmana", "Anak Agung" or "Gusti" for "Ksatria", and "Dewa" or "Pasek" for "Waisya". The "Sudra" class applies the birth order rule to name their children. Based on these morphological structures, we developed a ruled-based NER to identify persons in Balinese text [17] and embedded it into the preprocessing stage of our portal.

Balinese stories often possess similarities in the messages or topics which the author wish to convey to the readers. In an extensive collection, it would be convenient to the readers if stories with similar topics were grouped together, enabling them to read stories that meet their preferences. In [41] we achieved this goal by first applying Latent Dirichlet Allocation (LDA) to our collection of Balinese stories and then representing

each story by its top-T words with regards to the word and topic distribution found as the results of the previous step, and finally grouping the stories in a bottom-up process. Applying LDA to a corpus and supplying k number of topics to find would generate a topic distribution of each document and corpus-wide word-to-topic distribution. After this step, we represent a document with T number of words which are extracted from each topic based on the document's topic distribution. We choose the top-N words from each topic proportional to the topic's percentage in that document. We also calculate a distance matrix M which shows the cosine similarities between documents. Each document is considered an independent group at the beginning of the grouping process. We start to merge pair of documents with the highest similarities. We update M with refers to the proximity measure to reflect the latest distance between documents. We repeat the process of selecting a pair of documents with the highest score, grouping them, and continuing the process until the number of groups has been satisfied. At the end of the process, we managed to group the documents based on their topic. We present this group of documents as categories in our portal.

### 3.2 Portal

Katurangsatawa.com is an online portal to provide Balinese stories to readers. We implement our Balinese natural language processing framework into the server to process unseen stories to achieve this goal. Readers who have registered to the portal can upload these unseen stories. In general, a reader on the portal can utilize four main features: searching/browsing for particular stories, reading a story, registering, and writing unseen stories at the portal. The last feature can only be accessed by registered readers.



Fig. 2. Landing Page

Figure 2 shows the landing page of the portal, where readers are presented with a text box (1), categories of stories (2), characters of the story (3), and a summarized version of the story (4). Categories of stories (“Kategori1”, “Kategori2”, ..., “Kategori14”), identification of the characters (“Ni Bawange”, “I

Kedis Crukcuk Kuning”, ..., “Ni Kesunane”), and summary of the story are all the result of our Balinese natural language processing framework. These processes are carried out automatically at the server that utilizes our Balinese language processing framework.



Fig. 3. Search Result

Searching for a particular story can be achieved effortlessly throughout the portal. The reader must only provide the keywords related to the story and type them in at the given text box (shown as number 1 in Figure 2). Stories with the search keywords in their title will be presented promptly to the reader upon clicking the search button. Figure 3 shows the search result of the keyword “Kekua”. The search feature is one of the improvements we will address in future work.

The reader can also browse the collection of stories on the portal by clicking on any of the labels (shown as number 2 in Figure 2). When a reader clicks on a category, all stories under that particular category will be presented to the reader, as shown in Figure 4. In each story, the characters and summary of the story will also be ready for viewing. Clicking on any story will display the full text (Figure 5).

The reader can also browse the collection of stories on the portal by clicking on any of the labels (shown as number 2 in Figure 2). When a reader clicks on a category, all stories under that particular category will be presented to the reader, as shown in Figure 4. In each story, the characters and summary of the story will also be ready for viewing. Clicking on any story will display the full text (Figure 5).



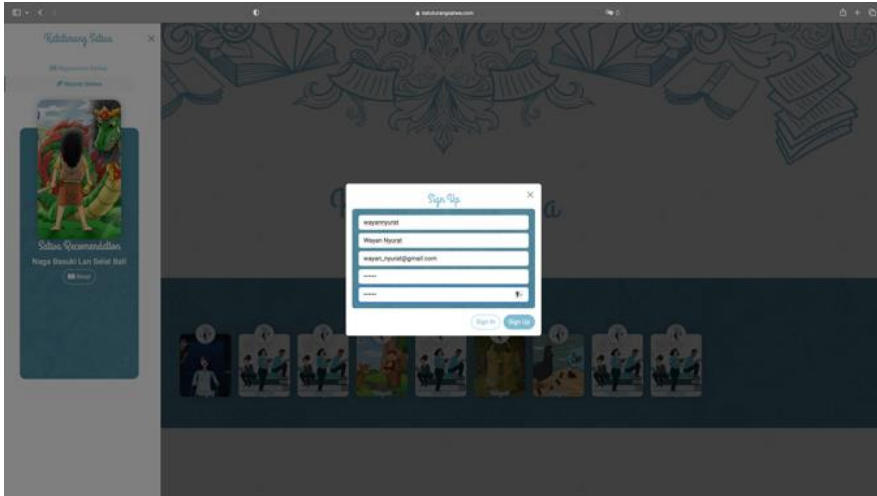
Fig. 4. Browsing the Collection

Upon browsing the collection of stories in the portal or searching for a particular story, readers may need help finding what they are looking for. We address this issue by allowing registered readers to write unseen stories to the portal. Readers can register to the portal by clicking the icon on the left side of the search text box (shown as number 5 in Figure 2), choosing the “Nyurat Satwa” menu, and finally clicking on the “Sign Up” button.



Fig. 5. Full Text of a Story





**Fig. 6.** Registration Window

Figure 6 shows the registration window where readers must provide personal details, including username, full name, email address, and their chosen password. The reader must click the “Sign Up” button to finalize the registration process. After completing the registration process, the reader can finally log in to the portal. When the readers have successfully logged into the portal, they may update their profile, modifying their full name and uploading their profile picture. Readers must click the “Update” button to finish editing their profile.



**Fig. 7.** Writing Unseen Story

In the “Nyurat Satwa” window, as shown in Figure 7, the registered reader can begin adding unseen stories to the portal. The reader has to fill in the story's details,

including the title, the complete text, and artwork that reflects the story. The category part of this unseen story can be left out as it will be filled in automatically later. When the reader clicks on the “Submit” button, the unseen story will undergo the processes in the Balinese natural language processing framework, which we have described in section 3.1. After the processes have been completed, the unseen story will be presented in the portal under the category to which it has been assigned.

**Table 1.** Questionnaire Results

No	Features	Rating			
		Very Difficult	Difficult	Easy	Very Easy
1	Searching for a story	6.5%	23%	43%	27.5%
2	Registering to the portal	6%	15%	54.5%	24.5%
3	Submitting a new story	7%	14%	52.5%	26.5%
4	Summarization	3.5%	12.5%	49%	35%
5	Grouping of stories	5.5%	16%	40%	38.5%
6	Characters identification	6.5%	17.5%	44%	32%

Evaluation on the portal is carried out by validation method. The process involved 200 respondents with various educational backgrounds and interests. Since one of the goals of developing the Balinese natural language processing framework in addition to the portal [katuturangsawatwa.com](http://katuturangsawatwa.com) is to preserve the Balinese language for the younger generation, we then focus the evaluation process on students from junior high to tertiary education level. If the students acknowledge the portal, we are one step closer to achieving the goal. We ask a series of questions regarding the students’ experience browsing through the portal. Before answering the questions, we allow them to exercise the main features of the portal. For each question, the students were asked to provide a rating from 1 to 4, with one and four representing “Very Difficult” and “Very Easy” respectively.

Table 1 summarizes the questionnaire results. From Table 1, most students perceive that the main features provided at the portal are easy to use. However, some students experienced difficulties when searching for a particular story. The utilization of non-standard characters in the title and body of the story is likely one of the reasons for the issue. We will address this issue in our future work by improving the preprocessing step in the framework.

**Table 2.** Rating of System

Overall Rating			
1	2	3	4
		22.17%	77.83%

To improve the overall system, students were also asked to rate the portal from 1 to 4, with one being “Very Bad” and four “Very Good”. Table 2 shows the questionnaire results. From Table 2, we can conclude that most students agreed that the developed portal is already adequate. However, some respondents have also voiced constructive feedbacks. One of the respondents found difficulties in navigating through the

portal, specifically when submitting an unseen story. When submitting an unseen story (as seen in Figure 7), there is a category section but with no available options to choose. This section should be left empty as it will be filled in automatically when the grouping process is started. Nevertheless, the story can still be submitted regardless of the empty category. Some respondents have also asked about translation to other languages. We will address this issue in our future research by including a translator into the portal.

## 4 Conclusion

We present in this paper a portal for Balinese stories. The portal utilizes the Balinese natural language processing framework, which can automatically summarize, identify characters, and group the stories based on similarities of its topics. We evaluate the main features of the portal by sharing questionnaires with students. From the questionnaire results, most students agreed that the portal is easy to use. Students have also acknowledged the portal to be satisfactory. However, further improvements can be achieved by removing unnecessary category field in story submission as well as implementing translator into the portal.

## Acknowledgment

This research was funded by the Udayana University's Institute for Research and Community Service (LPPM Unud), under Udayana's Flagship Research (Penelitian Unggulan Udayana) scheme and grant number B/78.153/UN14.4.A/PT.01.03/2022, April 20<sup>th</sup>, 2022.

## References

1. Hudman, L. E. & Hawkins, D. E. *Tourism in Contemporary Society: An Introductory Text*. (Prentice-Hall, 1989).
2. Nata, G. N. M. & Yudiastra, P. P. Stemming teks sor-singgih Bahasa Bali. *E-Proceedings KNS&I STIKOM Bali* 608–612 (2017).
3. Subali, M. A. P. & Faticah, C. Kombinasi Metode Rule-Based dan N-Gram Stemming untuk Mengenali Stemmer Bahasa Bali. *Jurnal Teknologi Informasi dan Ilmu Komputer* **6**, 219 (2019).
4. Negara, I. M. W. & ER, N. A. S. Basic Word Extraction Algorithm Based on Morphological Rules for Balinese Texts. *JELIKU (Jurnal Elektronik Ilmu Komputer Udayana)* **8**, 401–410 (2020).
5. Shaalan, K., Khaled, Raza & Hafsa. Person name entity recognition for Arabic. in (2007). doi:10.3115/1654576.1654581.
6. Farnakiotou, D., Karkaletsis, V., Koutsias, J., Sigletos, G. & Spyropoulos, C. Rule-Based Named Entity Recognition For Greek Financial Texts. (1970).
7. Gupta, V. & Lehal, G. S. Article: Named Entity Recognition for Punjabi Language Text Summarization. *Int J Comput Appl* **33**, 28–32 (2011).
8. Gunawan, W., Suhartono, D., Purnomo, F. & Ongko, A. Named-Entity Recognition for Indonesian Language using Bidirectional LSTM-CNNs. *Procedia Comput Sci* **135**, 425–432 (2018).

9. Aryoyudanta, B., Adji, T. & Hidayah, I. Semi-supervised learning approach for Indonesian Named Entity Recognition (NER) using co-training algorithm. in 7–12 (2016). doi:10.1109/ISITIA.2016.7828624.
10. Wibawa, A. & Purwarianti, A. Indonesian Named-entity Recognition for 15 Classes Using Ensemble Supervised Learning. *Procedia Comput Sci* **81**, 221–228 (2016).
11. Wulandari, D. W., Adikara, P. P. & Adinugroho, S. Named Entity Recognition (NER) Pada Dokumen Biologi Menggunakan Rule Based dan Naïve Bayes Classifier. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer* **2**, 4555–4563 (2018).
12. Setiyoadji, A., Muflikhah, L. & Fauzi, M. Named Entity Recognition Menggunakan Hidden Markov Model dan Algoritma Viterbi pada Teks Tanaman Obat. **1**, (2017).
13. Iswandi, I., Supriana, I. & Maulidevi, U. Perancangan Named Entity Recognition dalam Akuntansi untuk Identifikasi Transaksi berdasarkan Teks Indonesia. in (2015).
14. Ek, T., Kirkegaard, C., Jonsson, H. & Nugues, P. Named entity recognition for short text messages. *Procedia Soc Behav Sci* **27**, 178–187 (2011).
15. Wahyuni, N. M. S. & ER, N. A. S. Rule-based Named Entity Recognition (NER) to Determine Time Expression for Balinese Text Document. *JELIKU (Jurnal Elektronik Ilmu Komputer Udayana)* **9**, 555–562 (2021).
16. S, N. P. A. S. A. & ER, N. S. Location Named-Entity Recognition using Rule-Based Approach for Balinese Texts. *JELIKU (Jurnal Elektronik Ilmu Komputer Udayana)* **9**, 435–442 (2021).
17. Kurniadi, K. & Er, N. A. S. Person Named Entity Recognition in Balinese. *JELIKU (Jurnal Elektronik Ilmu Komputer Udayana)* (2021).
18. Widhiyanti, K. & Harjoko, A. POS Tagging Bahasa Indonesia Dengan HMM dan Rule Based. *Jurnal Informatika* **8**, (2013).
19. Pisceldo, F., Adriani, M. & Manurung, R. Probabilistic Part Of Speech Tagging for Bahasa Indonesia. (2009).
20. M, V. C., Pragantha, J. & Purnamasari, E. Implementasi Brill Tagger Untuk Memberikan. *Jakarta: Jurnal Teknik dan Ilmu Komputer* 301–315 (2012).
21. Yuliani, A. & Pardede, H. On part of speech tagger for Indonesian language. in 369–372 (2017). doi:10.1109/ICITISEE.2017.8285530.
22. Pradipta, I. G. M. H. & ER, N. A. S. Building Balinese Part-of-Speech Tagger Using Hidden Markov Model (HMM). *JELIKU (Jurnal Elektronik Ilmu Komputer Udayana)* **9**, 303–308 (2020).
23. Yousefi-Azar, M. & Hamey, L. Text summarization using unsupervised deep learning. *Expert Syst Appl* **68**, 93–105 (2017).
24. Nallapati, R. *et al.* Sequence-To-Sequence Rnns For Text Summarization. *Iclr* 4–7 (2016).
25. Anggraini, D. & Wulandari, L. Peringkasan Teks Artikel Ilmiah Berbahasa Indonesia Menggunakan Teknik Ekstraktif dan Fitur Kalimat Untuk Dokumen Tunggal. *Seminar Nasional Rekayasa Komputer dan Aplikasinya* 126–130 (2015).
26. Saputra, J. S. & Fachrurrozi, M. Peringkasan Teks Berita Berbahasa Indonesia Menggunakan Metode Latent Semantic Analysis (LSA) dan Teknik Steinberger & Jezek. *Prosiding Annual Research Seminar vol. 3* (2017).
27. Mustaqhfiri, M., Abidin, Z. & Kusumawati, R. Peringkasan Teks Otomatis Berita Berbahasa Indonesia Menggunakan Metode Maximum Marginal Relevance. *Matics* (2012) doi:10.18860/mat.v0i0.1578.
28. Saputra, Jerry. Fachrurrozi, Muhammad. Y. Peringkasan Teks Berita Berbahasa Indonesia Menggunakan Metode Latent Semantic Analysis (LSA) dan Teknik Steinberger & Jezek. *Computer Science and ICT* **3**, 215–219 (2017).

29. Hidayat, E. Y., Firdausillah, F., Hastuti, K., Dewi, I. N. & Azhari. Automatic text summarization using latent dirichlet allocation (LDA) for document clustering. *International Journal of Advances in Intelligent Informatics* **1**, 132–139 (2015).
30. Aristoteles. Penerapan Algoritma Genetika pada Peringkasan Teks Dokumen Bahasa Indonesia. *Semirata FMIPA Universitas Lampung* 29–33 (2013).
31. Jezek, K. & Steinberger, J. Automatic Text Summarization (The state of the art 2007 and new challenges). *Proceedings of Znalosti* 1–12 (2008).
32. Abimanyu, C. G., ER, N. & Karyawati, A. A. I. N. E. BALINESE AUTOMATIC TEXT SUMMARIZATION USING GENETIC ALGORITHM. *JITK (Jurnal Ilmu Pengetahuan dan Teknologi Komputer)* **6**, 13–20 (2020).
33. Suputra, I. P. G. H. Peringkasan teks otomatis untuk dokumen bahasa Bali berbasis metode ekstraktif. *Jurnal Ilmiah Komputer* **X**, 1–6 (2017).
34. Suharno, C., Fauzi, M. & Perdana, R. Klasifikasi Teks Bahasa Indonesia Pada Dokumen Pengaduan Sambat Online Menggunakan Metode K-Nearest Neighbors Dan Chi-square. *systemic information system and informatics journal* **3**, 25–32 (2017).
35. Pratiwi, N. & Widodo, W. Klasifikasi Dokumen Karya Akhir Mahasiswa Menggunakan Naive Bayes Classifier (NBC) Berdasarkan Abstrak Karya Akhir di Jurusan Teknik Elektro Universitas Negeri Jakarta. **1**, 31–38 (2017).
36. Sholehudin, M., Fauzi Ali, M. & Adinugroho, S. Implementasi Metode Text Mining dan K-Means Clustering untuk Pengelompokan Dokumen Skripsi ( Studi Kasus : Universitas Brawijaya ). *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer* **2**, 5518–5524 (2018).
37. Kurnia Bakti, V. & Indriyatno, J. Klasterisasi Dokumen Tugas Akhir Menggunakan K-Means Clustering, Sebagai Analisa Penerapan Sistem Temu Kembali. *KOPERTIP : Jurnal Ilmiah Manajemen Informatika dan Komputer* **1**, 31–34 (2017).
38. Prihatini, P. M., Suryawan, I. K. & Mandia, I. N. METODE LATENT DIRICHLET ALLOCATION UNTUK EKSTRAKSI TOPIK DOKUMEN. *Logic : Jurnal Rancang Bangun dan Teknologi* **17**, 153–157 (2017).
39. Suhartono, D. Probabilistic Latent Semantic Analysis (PLSA) untuk Klasifikasi Dokumen Teks Berbahasa Indonesia. (2015).
40. Guan, P., Wang, Y., Chen, B. & Fu, Z. K-means Document Clustering Based on Latent Dirichlet Allocation. in *Western Decision Sciences Institute (WDSI)* (2016).
41. ER, N. A. S. Implementasi Latent Dirichlet Allocation (LDA) untuk Klasterisasi Cerita Berbahasa Bali. *Jurnal Teknologi Informasi dan Ilmu Komputer* **8**, (2021).
42. Temaja, I. G. B. W. B. SISTEM PENAMAAN ORANG BALI. *HUMANIKA* **24**, 60–72 (2018).

**Open Access** This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

