

Komparasi Hasil Transliterasi Huruf Latin ke Aksara Bali Menggunakan Konsep Pemadanan Kata Khusus Bahasa Bali

¹Made Sudarma

¹*Teknik Elektro, Universitas Udayana*
Bali, Indonesia
msudarma@unud.ac.id

²Ni Wayan Sri Ariyani, ³I Putu Agus Eka Darma Udayana, ⁴Lie Jasa,

⁵I Gede Tresna Agustina Putra
²*Teknik Elektro, Universitas Udayana*
Bali, Indonesia
sriariyani@unud.ac.id

³*Ilmu Teknik, Universitas Udayana*
Bali, Indonesia
agus.ekadarma@gmail.com

⁴*Ilmu Teknik, Universitas Udayana*
Bali, Indonesia
liejasa@unud.ac.id

⁵*Informatika, Institut Bisnis dan Teknologi Indonesia*
Bali, Indonesia
gdtresna.agustina@gmail.com

Abstract—Lontar merupakan salah satu warisan budaya Bali yang harus dijaga keberadaannya. Beberapa tahun terakhir, tim konservasi lontar dari Kabupaten Gianyar menemukan banyak koleksi lontar milik warga yang mengalami kerusakan. Kerusakan ini disebabkan oleh berbagai faktor, salah satunya adalah kurangnya pengetahuan dalam cara merawat lontar. Tidak hanya koleksi milik warga, beberapa koleksi rusak juga ditemukan di Dinas Kebudayaan Provinsi Bali. Sebagai solusinya, pada penelitian ini dikembangkan sistem digitalisasi lontar yang dapat mengubah teks latin kedalam bentuk lontar aksara Bali menggunakan metode *Jaro Winkler Distance*. Hasil dari pengembangan sistem digitalisasi lontar ini adalah sistem mampu menghasilkan output yang mendekati konten dari lontar asli. Dari 10 lembar yang diuji, di dapatkan tingkat akurasi sebesar 92.7%. Selain uji akurasi, pada penelitian ini juga dilakukan pengujian ahli media dengan melakukan penilaian struktur penulisan yang sesuai dengan aturan aksara Bali, yang mendapatkan tingkat akurasi sebesar 93.6%.

Kata Kunci—Aksara Bali, Digitalisasi Lontar, *Jaro Winkler Distance*, Transliterasi

I. PENDAHULUAN

Lontar merupakan salah satu warisan budaya Bali dan merupakan sumber literatur yang berisi ilmu pengetahuan [1] [2] yang masih digunakan oleh masyarakat Bali dalam kehidupan sehari-hari. Selain sebagai sumber literatur, lontar juga digunakan sebagai media penulisan dari masa Hindu-Budha dengan bahasa Bali kawi sebagai bahasa dasar dalam penulisan lontar [3]. Seiring berjalannya waktu, penggunaan lontar di Bali mulai mengalami penurunan. Penurunan ini disebabkan oleh keterbatasan linguistik masyarakat terhadap penulisan aksara Bali [4]. Hal ini menjadi masalah yang serius karena masalah ini membuat lontar hanya menjadi bukti sejarah dari sejarah peradaban yang terjadi di Bali.

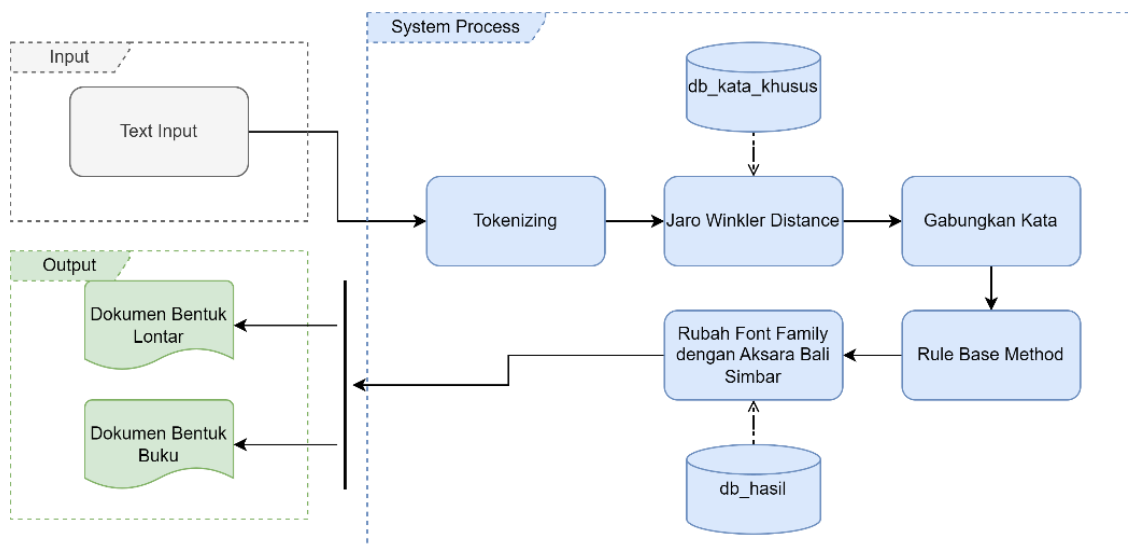
Pada bulan Februari 2023, menurut Bapak I Wayan Suparsa selaku koordinator Divisi Konservasi Lontar Penyuluh Bahasa Bali Kabupaten Gianyar, ditemukan 30 lontar milik masyarakat di Desa Pupuan, Kabupaten Gianyar. 10 diantaranya dalam kondisi rusak akibat pemilik lontar yang jarang membaca lontar miliknya, dilansir dari balinesia.id [5]. Tidak hanya di kabupaten/kota di Bali, koleksi lontar yang rusak juga ditemukan di Dinas Kebudayaan Provinsi Bali, dari hasil observasi penulis dengan Ibu Ni Putu Seni selaku Pustakawan Ahli Madya. Penulis menemukan beberapa koleksi lontar yang isi atau tulisannya sudah tidak dapat dibaca dan ada juga yang rusak secara fisik. Sebagai solusi dari permasalahan yang telah diuraikan, maka perlu dilakukan proses digitalisasi sebagai upaya menjaga dan melestarikan warisan budaya dengan memanfaatkan perkembangan teknologi. Upaya digitalisasi ini telah dicoba oleh banyak pihak dengan *output* sistem informasi transliterasi aksara Bali, namun sistem yang dihasilkan belum mampu memberikan *output* yang sesuai. Permasalahan ini muncul karena cara sistem memproses *input* yang tidak menggunakan kaidah penulisan aksara Bali. Aksara Bali memiliki banyak aturan dalam penulisannya sehingga perlu dibuatkan aturan yang sesuai dalam memproses teks masukan. Masalah lainnya adalah sistem yang ada tidak dapat mendeteksi homonim.

Homonim adalah dua kata yang terlihat dan/atau terdengar sama [6] namun memiliki makna yang berbeda. dan kata ini sering ditemukan dalam bahasa kawi Bali yang merupakan bahasa dasar penulisan lontar. Untuk itu, penelitian ini menerapkan sebuah metode pengolahan kata yaitu *Jaro Winkler Distance*. Metode ini merupakan metode yang membandingkan kemiripan karakter penyusun dan mengukur kemiripan dua buah *string* [7]. Metode ini dapat digunakan dalam pendeteksian kata ganda. Kemiripan antara dua buah *string* dinilai berdasarkan *Jaro Winkler Distance*, semakin besar nilainya maka kedua *string* tersebut dianggap sama [8]. Penelitian yang dilakukan oleh Syahputra [9] membandingkan metode *Jaro Winkler Distance* dengan metode lain yang sejenis dengan hasil metode *Jaro Winkler Distance* dapat membandingkan kata dengan cepat dibandingkan dengan metode lain yang sejenis.

II. METODE DAN PROSEDUR

A. Alur Kerja Sistem

Alur proses sistem dibagi menjadi tiga sub-proses, pertama proses *input user* menginputkan teks latin, Proses Sistem: dimulai dengan tokenisasi teks input, melakukan spelling checker dengan metode *Jaro Winkler Distance*, dan terakhir merubah urutan teks latin agar sesuai dengan kaidah penulisan aksara bali, proses terakhir adalah *ouput* file kata dengan tata letak lontar bali.



Gambar 1. Alur Kerja Sistem

Gambar 1 di atas menjelaskan setiap bagian dari proses yang terjadi pada sistem.

B. Tokenizing

Tokenisasi adalah proses pertama yang harus diselesaikan sebelum memasuki pemeriksaan ejaan dan proses berbasis aturan. Selama proses ini, kalimat disegmentasi menjadi kata-kata individual melalui berbagai langkah seperti pelipatan huruf besar-kecil (konversi ke huruf kecil), penghilangan spasi dan karakter khusus, dan token yang dihasilkan disimpan dalam sebuah larik [10][11].

Contoh teks : “Rahina Galungn”
Case folding : “rahina galungn”
Space remove : “rahina%galungn”

C. Jaro Winkler Distance

Setelah proses tokenizing selesai, selanjutnya masuk ke proses spell checker menggunakan *Jaro Winkler Distance*. Proses *spell checker* dimulai dari proses mencari *Jaro Winkler* dari kedua teks dan kemudian menghitung jarak *Jaro Winkler* dari kedua teks tersebut [12].

Hasil dari tokenizing (a): [rahina, galungn]
Teks pada database (b) : [rahina,galungan]

Untuk mencari nilai *Jaro Winkler* dapat menggunakan rumus:

$$dj = \frac{1}{3} \left(\frac{m}{|s1|} + \frac{m}{|s2|} + \frac{m-t}{m} \right) \quad (1)$$

Dan untuk mencari nilai *Jaro Winkler Distance* menggunakan rumus:

$$d_w = d_j + (l * p(1 - d_j)) \quad (2)$$

Untuk kata pertama ‘rahina’ memiliki nilai *Jaro Winkler Distance* = 1 dengan implementasi rumus sebagai berikut:

$$dj = \frac{1}{3} \left(\frac{6}{|6|} + \frac{6}{|6|} + \frac{4-0}{4} \right) = 1 \quad (1)$$

$$d_w = 1 + (0 * 0.1(1 - 1)) = 1 \quad (2)$$

Sehingga, dapat disimpulkan bahwa kata masukan ‘rahina’ dengan kata ‘rahina’ yang ada pada *database* bernilai sama. Untuk kata kedua ‘galungn’ memiliki nilai *Jaro Winkler Distance* = 0.916 dengan implementasi rumus sebagai berikut:

$$dj = \frac{1}{3} \left(\frac{7}{|7|} + \frac{7}{|8|} + \frac{7-2}{7} \right) = 0.86 \quad (1)$$

$$d_w = 0.86 + (4 * 0.1(1 - 0.86)) = 0.916 \quad (2)$$

Dari hasil perhitungan untuk kata kedua ‘galungn’ dengan kata ‘galungan’ yang ada pada *database*, mendapatkan nilai *Jaro Winkler Distance* = 0.916 yang berarti kedua kata tersebut memiliki sedikit perbedaan sehingga disarankan untuk melakukan perubahan dari ‘galungn’ yang merupakan teks masukan menjadi “galungan” yang merupakan kata target. Langkah terakhir adalah menggabungkan teks yang telah diproses sehingga hasilnya adalah “rahina galungan”.

D. Rule Based Method

Pada proses ini, bentuk teks yang telah diproses pada tahap *spell checker* diubah ke dalam format yang siap untuk diubah dengan menggunakan *font* Bali Simbar.

Teks masukan : “rahina galungan”

- Proses 1 : “rhin galungan”
- Proses 2 : “rhin glu\n/.”
- Proses 3 : “rhin glu\n/.”

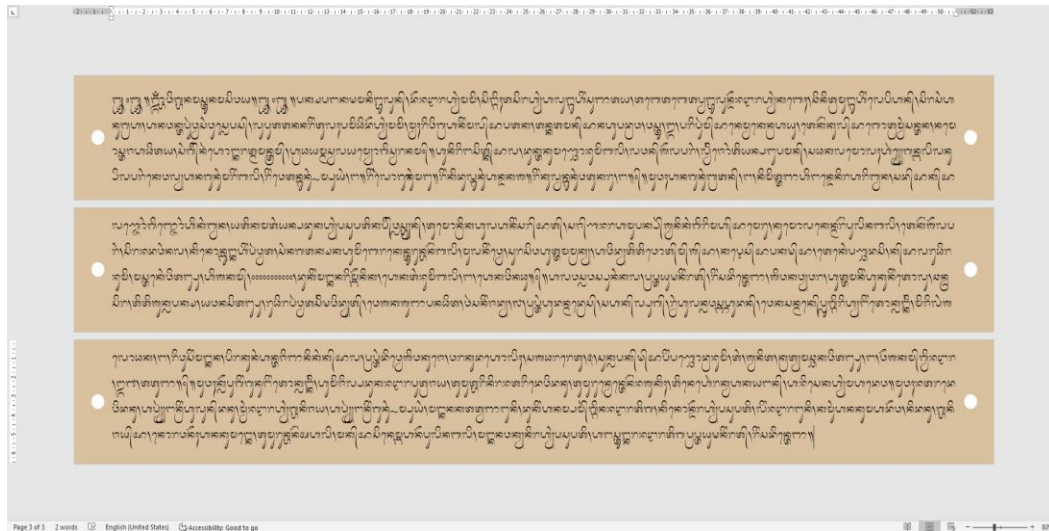
Pada proses 1 dan 2 yang melakukan perubahan dan penambahan titik di akhir kalimat untuk menghasilkan teks yang siap dikonversi ke dalam aksara Bali dengan menggunakan font simbar Bali dengan hasil seperti pada proses 3

III. HASIL DAN PEMBAHASAN

Berdasarkan hasil implementasi dan pengujian, sistem yang dikembangkan telah mampu menghasilkan *output* menyerupai lontar dan memiliki akurasi yang cukup baik. Berikut ini adalah penjelasan dari hasil yang telah dilakukan oleh peneliti.

A. File Output

Tata letak lontar yang digunakan dalam file *output* adalah tata letak lontar bali kuno.



Gambar 2. File Output

File *output* yang dihasilkan dalam sistem ini adalah file .docx yang dapat diakses menggunakan aplikasi *Microsoft Word*

B. Pengujian Akurasi

Pengujian akurasi pada penelitian ini dilakukan dengan membandingkan *output* yang dihasilkan dari sistem dengan citra asli Lontar Pasek Sapta Rsi. Perbandingan dilakukan dengan membandingkan setiap karakter dari keluaran sistem dengan setiap karakter pada gambar lontar asli. Hasil pengujian akurasi adalah sebagai berikut:

Kode Lontar	Total Karakter	Benar	Salah	Akurasi
1B	436	404	32	92.7%
2A	450	417	33	92.7%
2B	426	403	23	94.6%
3A	398	373	23	93.7%
3B	440	400	40	91.0%
4A	420	380	40	90.4%
4B	445	417	28	93.7%

5A	416	387	29	93.0%
5B	439	405	34	92.2%
6A	505	472	33	93.4%
Total				92.7%

Jumlah aksara yang salah paling banyak ditemukan pada lembar lontar 3B dan 4A, sedangkan jumlah kesalahan paling sedikit ditemukan pada lembar lontar 2B, 3A, dan 4B. Faktor-faktor yang menyebabkan kesalahan antara lain perbedaan penggunaan aksara khusus antara aksara Babad Pasek Sapta Rsi dan Aksara Bali Simbar, termasuk perbedaan dalam merepresentasikan bunyi 'pepet', yaitu dengan menggunakan aksara 'é' dan 'e'. Teks Latin Babad Pasek Sapta Rsi yang dipindai menghadirkan tantangan transliterasi, dengan ketidakkonsistenan seperti 't' yang disalahartikan sebagai 't'. Selain itu, kesalahan pengetikan dalam Babad Pasek Sapta Rsi, seperti salah eja dan kata yang hilang, menyebabkan ketidaksesuaian antara hasil keluaran sistem dengan teks lontar asli.

C. Pengujian Ahli Media

Pengujian ahli media pada penelitian ini dibantu oleh Bapak I Wayan Suadnyana, S.Pd., M.Pd., penelitian ini dilakukan dengan melakukan penilaian terhadap kesesuaian kaidah penulisan aksara Bali. Pengujian ini menggunakan output yang melibatkan proses spelling checker dengan menggunakan metode *Jaro Winkler Distance*. Hasil dari pengujian dapat dilihat pada tabel di bawah ini.

Tabel 2. Hasil Pengujian Ahli Media

Kode Lontar	Total Karakter	Benar	Salah	Akurasi
1B	436	415	21	95.6%
2A	450	420	30	93.3%
2B	426	403	23	94.6%
3A	398	379	19	95.2%
3B	440	401	39	91.1%
4A	420	382	38	90.9%
4B	445	416	29	93.4%
5A	416	395	21	94.9%
5B	439	413	26	94.0%
6A	505	477	28	94.4%
Total				93.6%

Kesalahan paling banyak ditemukan pada lembar 3B, sedangkan lembar 3A memiliki kesalahan paling sedikit dalam uji ahli media. Faktor-faktor yang menyebabkan kesalahan adalah masalah yang muncul dalam penggunaan aksara Swalalita (*Akara, Ikara, dan Ukara*) yang menyimpang dari norma-norma yang telah ditetapkan, sehingga mengganggu kemiripan yang diinginkan dengan huruf kapital dalam bahasa Indonesia. Penempatan *adeg-adeg* yang tidak tepat, yang dirancang untuk meredam suara ocal, terjadi di luar parameter yang ditentukan, sehingga menimbulkan ketidakkonsistenan dalam aksara. Demikian pula, penanganan bisah yang tidak tepat, berdampak pada akurasi pada kata-kata dengan bunyi 'h' di akhir, menekankan perlunya kepatuhan terhadap pedoman yang tepat. Kesalahan pada output yang dihasilkan sistem disebabkan oleh beberapa faktor seperti banyaknya karakter yang diidentifikasi salah pada proses scan buku Babad Pasek Sapta Rsi, ini menyebabkan kesalahan dalam proses encode pada metode rule base, kemudian terdapat perbedaan dalam penggunaan karakter untuk penulisan pepet dan taleng antara buku Babad Pasek Sapta Rsi dengan lontar asli Babad Pasek Sapta Rsi sebagaimana dijelaskan pada pengujian akurasi pada poin pertama. Masalah berikutnya diakibatkan karena beberapa special character atau karakter khusus yang tidak dapat diidentifikasi oleh font Aksara Bali Simbar sehingga menyebabkan kesalahan pada output yang dihasilkan sistem.

IV. KESIMPULAN

Metode *Jaro Winkler Distance* secara efektif mendeteksi kata-kata masukan yang mirip dengan kata-kata khusus dalam *database* atau kata-kata homonim yang sudah ada. Meskipun ada beberapa masalah pengenalan dalam *dataset* "Babad Pasek Sapta Rsi" yang dipindai, memastikan pemeriksaan menyeluruh pada karakter khusus sebelum pengujian sangat penting. Penerapan metode *Jaro Winkler Distance* dalam mengembangkan sistem digitalisasi lontar mencapai akurasi 93,6%. Dampaknya terutama terlihat ketika mengidentifikasi kata-kata yang mirip dengan kata-kata yang ada dalam *database* sistem, menandai kemajuan yang signifikan dalam akurasi dan fungsionalitas sistem.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada DRPTM Kemdikbutristek yang telah memberikan pendanaan sehingga penelitian ini dapat terselesaikan, penulis juga mengucapkan terima kasih kepada LPPM Universitas Udayana yang telah memfasilitasi proses penelitian ini.

DAFTAR PUSTAKA

- [1] I. B. G. Sarasvananda, P. G. S. C. Nugraha, and I. B. A. I. Iswara, "The Balinese Lontar Manuscript Metadata Model: An Ontology-Based Approach," *Jurnal Multidisiplin Madani*, vol. 3, no. 9, pp. 1964–1971, 2023.
- [2] M. W. A. Kesiman and G. A. Pradnyana, "Image Patch Extraction in Text Area of Balinese Palm Leaf Manuscripts with Gabor Filters," in *3rd International Conference on Innovative Research Across Disciplines (ICIRAD 2019)*, Atlantis Press, 2020, pp. 19–23.
- [3] G. Indrawan, I. G. Aris Gunadi, M. Santo Gitakarma, and I. K. Paramarta, "Latin to Balinese script transliteration: lessons learned from the computer-based implementation," in *2021 The 4th International Conference on Software Engineering and Information Management*, 2021, pp. 171–175.
- [4] K. Teguh, Y. Aditya, M. Windu, A. Kesiman, and G. A. Pradnyana, "PENGEMBANGAN GAME EDUKASI TEMATIK AKSARA DAN BAHASA BALI PADA LONTAR BERBASIS MOBILE," *Kumpulan Artikel Mahasiswa Pendidikan Teknik Informatika (KARMAPATI)*, vol. 8, no. 3, 2019.
- [5] E. Ariana, "Hanya Diupacarai saat Saraswati, Sepertiga Lontar Koleksi Masyarakat di Pupuan Tak Bisa Dibaca," <https://balinesia.id/>, Feb. 23, 2023. Accessed: May 27, 2023. [Online]. Available: <https://bali.tribunnews.com/2022/11/18/seribuan-lontar-gianyar-telah-dikonservasi-sebagaimana-besar-mengulas-wariga>
- [6] Z. S. Mukhtoriyevna, "System for Teaching Homonyms Words in Primary School," *EUROPEAN JOURNAL OF INNOVATION IN NONFORMAL EDUCATION*, vol. 3, no. 3, pp. 1–4, 2023.
- [7] J. Frando, I. Ruslianto, and R. Hidayati, "Penerapan Jaro Winkler Distance dalam Aplikasi Pengoreksi Kesalahan Penulisan Bahasa Indonesia Berbasis Web," *Coding Jurnal Komputer dan Aplikasi*, vol. 7, no. 03, 2019.
- [8] B. V. Indriyono, "Kombinasi Damerau Levenshtein dan Jaro-Winkler Distance Untuk Koreksi Kata Bahasa Inggris," *Jurnal Teknik Informatika dan Sistem Informasi*, vol. 6, no. 2, 2020.
- [9] I. Syahputra and F. Syakti, "Perbandingan Algoritma Levenshtein dan Jaro Winkler Pada Sistem Informasi Pencarian Dokumen Perundang-Undangan (Studi Kasus: Diskominfo Lahat)," *SMATIKA JURNAL: STIKI Informatika Jurnal*, vol. 12, no. 02, pp. 176–186, 2022.
- [10] Srinivas Chakravarthy, "Tokenization for Natural Language Processing | by Srinivas Chakravarthy | Towards Data Science," *Towards Data Science*. Accessed: May 22, 2023. [Online]. Available: <https://towardsdatascience.com/tokenization-for-natural-language-processing-a179a891bad4>
- [11] A. Aziz, F. Fauziah, and I. Fitri, "Analisis Sentimen Terhadap Kebijakan Pemerintah Tentang Larangan Mudik Hari Raya Idulfitri di Indonesia Tahun 2021 Menggunakan Metode Naïve Bayes," *J-SAKTI (Jurnal Sains Komputer dan Informatika)*, vol. 5, no. 2, pp. 842–851, 2021.
- [12] M. A. Yulianto and N. Nurhasanah, "The hybrid of jaro-winkler and rabin-karp algorithm in detecting Indonesian text similarity," *Jurnal Online Informatika*, vol. 6, no. 1, pp. 88–95, 2021.