

Stemmer Bahasa Melayu Riau Berdasarkan Aturan Morfologi

Yusra¹, Muhammad Fikry², Hendri³

^{1,2,3}Teknik Informatika, UIN Sultan Syarif Kasim Riau

Jl. H.R. Soebrantas no. 155 KM. 18 Simpang Baru, Pekanbaru, 28293

e-mail: yusra@uin-suska.ac.id¹, muhammad.fikry@uin-suska.ac.id², hendri6@students.uin-suska.ac.id³

Abstrak

Penggunaan Bahasa Melayu Riau semakin terpinggirkan. Saat ini, belum tersedia sumber daya bahasa Melayu Riau dalam bentuk digital yang dapat digunakan oleh peneliti dan pengembang perangkat lunak dalam tahapan analisis pemrosesan bahasa alami. Penelitian ini berfokus pada tahapan analisis pertama dalam pemrosesan teks, yaitu analisis leksikal. Salah satu tugas pemrosesan bahasa alami pada tahapan tersebut adalah stemming yang mereduksi suatu kata berimbuhan menjadi kata dasarnya. Hasil stemming dapat dimanfaatkan oleh tugas-tugas lain dalam pemrosesan teks. Dalam penelitian ini, telah dilakukan analisa dan perancangan algoritma stemming Melayu Riau berdasarkan aturan morfologi. Hasil implementasi sumber daya bahasa berupa stemmer bahasa Melayu Riau dalam bentuk PHP Extension. Berdasarkan hasil pengujian terhadap 1000 kata, diperoleh akurasi 100% yang memperlihatkan bahwa stemmer ini dapat digunakan untuk pemrosesan teks berbahasa Melayu Riau.

Kata kunci: bahasa melayu riau, kata dasar, morfologi, pemrosesan bahasa alami, stemmer

Abstract

Riau Malay language is increasingly marginalized. Currently, there are no Riau Malay language resources available in digital form that can be used by researchers and software developers in the analysis stage of natural language processing. This study focuses on the first stage of analysis in text processing, namely lexical analysis. One of the tasks at this stage is stemming which reduces an affixed word to its base word. The stemming results can be utilized by other tasks in text processing. In this study, an analysis and design of a stemming algorithm for Riau Malay has been carried out based on morphological rules. The implementation result is a Riau Malay language stemmer in the form of a PHP Extension.. Based on test results of testing on 1000 words, 100% accuracy is obtained which shows that this stemmer can be used for processing Riau Malay text.

Keywords: basic word, morphology, natural language processing, riau malay language, stemmer

1. Pendahuluan

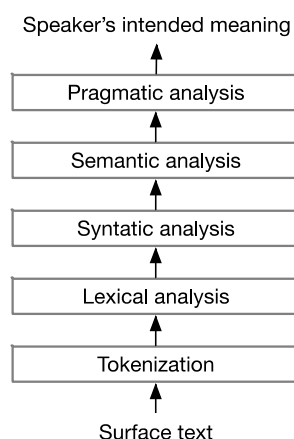
Bahasa Melayu memiliki banyak dialek (variasi bahasa) dan tersebar di berbagai daerah di Indonesia. Bahasa ini dituturkan mulai sepanjang pantai timur Sumatera, Kepulauan Riau, Kepulauan Bangka Belitung, Jambi, Sumatera Selatan, Bengkulu, hingga pesisir Pulau Kalimantan dan Jembrana, Bali. Berdasarkan hasil sensus penduduk tahun 2010 tentang bahasa sehari-hari penduduk Indonesia, bahasa Melayu berada pada peringkat ke empat setelah bahasa Jawa, Indonesia, dan Sunda dalam hal jumlah penuturnya [1]. Bahasa Indonesia umumnya digunakan sebagai alat komunikasi di kota-kota, sementara bahasa daerah semakin terpinggirkan.

Bahasa Melayu Riau merupakan bahasa daerah yang terbanyak digunakan di Provinsi Riau, baik sebelum pemekaran Provinsi Kepulauan Riau [2] maupun setelahnya [1]. Namun demikian, hanya 65% penduduk Riau bersuku bangsa Melayu yang menggunakannya sebagai bahasa sehari-hari [1]. Selain itu, bahasa Melayu Riau dapat ditemukan dalam bentuk sastra berbahasa Melayu, baik tertulis maupun lisan [3]. Karya sastra Melayu diantaranya pantun, syair, gurindam, mantra, peribahasa, seloka, hikayat, dongeng, legenda, cerita jenaka, dan puisi. Tulisan berbahasa Melayu Riau juga dapat ditemukan di media cetak dan daring seperti RiauPos (riaupos.jawapos.com) dan Tribun Pekanbaru (pekanbaru.tribunnews.com) berbentuk kutipan perkataan narasumber berita, opini, dan cerita fiksi meskipun jumlahnya tidak banyak.

Dikutip dari republika.co.id, Lembaga Adat Melayu Riau menyatakan bahwa penggunaan bahasa Melayu terpinggirkan akibat pengaruh berbagai faktor, salah satunya adalah tuntutan zaman [4]. Apabila dibiarkan, kepunahan bahasa daerah dapat terjadi akibat kurangnya penghargaan terhadap bahasa daerah, kurangnya intensitas pemakaian bahasa daerah, pengaruh faktor ekonomi, dan pengaruh pemakaian bahasa Indonesia [5]. Oleh karena itu, Pemerintah Provinsi Riau mendorong upaya pelestarian kebudayaan Melayu di Riau. Sesuai Perda Provinsi Riau no. 9 tahun 2015 tentang Pelestarian Kebudayaan Melayu Riau, salah satunya diwujudkan melalui penggunaan bahasa dan huruf Melayu di lingkungan pendidikan, media massa, pasar, perkantoran, serta nama jalan dan nama tempat.

Sayangnya, belum tersedia sumber daya bahasa (*language resources*) Melayu Riau dalam bentuk digital yang dapat digunakan oleh peneliti dan pengembang perangkat lunak di bidang komputasi linguistik, pembelajaran mesin, dan pemrosesan bahasa alami. Menurut [6], sumber daya bahasa adalah kumpulan data dan deskripsi bahasa dalam bentuk yang dapat dibaca mesin, digunakan untuk membangun, meningkatkan, atau mengevaluasi bahasa alami. Sumber daya bahasa meliputi kumpulan data, berbagai alat, dan *best practice* untuk mengelola dan memproses data [7]. Data merupakan informasi apa pun yang mendokumentasikan atau menjelaskan bahasa, seperti korpus tulis, korpus lisan, dan kamus bahasa. Alat merupakan sumber daya komputasi yang memfasilitasi pembuatan, menampilkan, meminta, atau menggunakan data bahasa. Alat tidak hanya mencakup program perangkat lunak, tetapi juga sumber daya digital yang digunakan oleh program tersebut. Upaya penyediaan sumber daya bahasa, baik data maupun alat, akan memakan waktu dan biaya tinggi [8].

Pada bahasa daerah lain di Indonesia, ada banyak penelitian mengenai kamus bahasa digital. Beberapa di antaranya berbentuk aplikasi Android (Minang [9], Jawa [10], dan Sunda [11]), aplikasi *web* (Bugis [12] dan Massenrempulu [13]), dan aplikasi *desktop* (Bima [14] dan Tountemboan [15]). Namun demikian, kamus-kamus tersebut diperuntukkan bagi manusia, tidak dalam bentuk yang dapat digunakan oleh peneliti dan pengembang perangkat lunak. Selain kamus bahasa, penelitian mengenai bahasa daerah lain juga menghasilkan algoritma *stemming* atau *stemmer*. *Stemmer* yang dimaksud meliputi *stemmer* bahasa Madura dengan memodifikasi algoritma Enhanced Confix Stripping [16], *stemmer* bahasa Jawa [17] [18], *stemmer* Basa Sunda [19], *stemmer* teks sor-singgih bahasa Bali berdasarkan algoritma Porter Stemmer for Bahasa Indonesia [20], serta *stemmer* bahasa Bali berdasarkan algoritma Nazief dan Adriani [21].

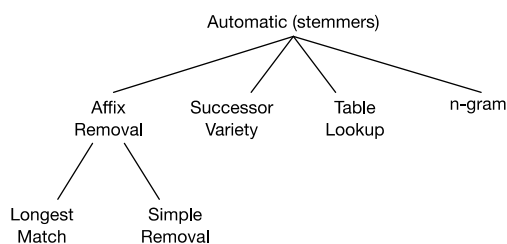


Gambar 1. Tahapan Analisis dalam Pemrosesan Teks [22]

Proses *stemming* mereduksi suatu kata berimbuhan menjadi kata dasarnya. Proses ini merupakan salah satu tugas pemrosesan bahasa alami (NLP *task*) pada tahapan analisis leksikal. Gambar 1 memperlihatkan tahapan-tahapan pemrosesan teks yang dilakukan untuk memahami maksud penulisnya, meliputi tokenisasi, analisis leksikal, sintaksis, semantik, dan pragmatis [22]. Hasil *stemming* diperlukan oleh tugas-tugas pemrosesan bahasa alami lainnya, baik di tahapan analisis leksikal maupun di tahapan-tahapan analisis selanjutnya. Proses

stemming juga digunakan di dalam sistem Information Retrieval untuk meningkatkan kualitas informasi yang didapatkan [23].

Proses *stemming* untuk setiap bahasa berbeda dengan bahasa yang lain karena perbedaan morfologi [24]. Gambar 2 memperlihatkan empat pendekatan yang dapat digunakan oleh *stemmer*. Pendekatan yang umum digunakan adalah penghapusan imbuhan (*affix removal*) berdasarkan aturan morfologi bahasa, dengan atau tanpa penggunaan kamus bahasa (*table lookup*) sebagai acuan. Sering kali, modifikasi dilakukan terhadap algoritma *stemming* bahasa lain agar sesuai dengan aturan morfologi bahasa yang diinginkan [16] [20] [21].



Gambar 2. Pendekatan *Stemmer* [25]

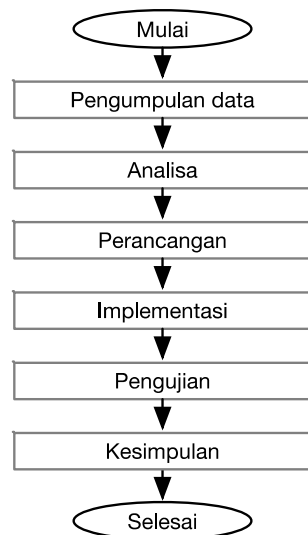
Penelitian mengenai leksikon, morfologi, dan sintaksis bahasa Melayu Riau sudah dilakukan oleh Pusat Pembinaan dan Pengembangan Bahasa (sekarang Badan Bahasa) sejak tahun 1983, jauh sebelum pemekaran Provinsi Kepulauan Riau dari Provinsi Riau pada tahun 2002. Secara garis besar berdasarkan leksikon dan morfologinya, bahasa Melayu Riau terdiri atas dua dialek yaitu dialek daratan dan dialek kepulauan [26]. Dialek daratan dipakai oleh penduduk di daerah Riau daratan, sedangkan dialek kepulauan dipakai oleh penduduk di daerah kepulauan dan pantai atau pesisir [27]. Setiap dialek pun mempunyai subdialek [28]. Sementara secara geografis, bahasa Melayu Riau terbagi menjadi tiga dialek yaitu dialek daratan, pesisir, dan kepulauan. Dialek pesisir berkembang sangat mirip dengan dialek kepulauan. Dialek pesisir digunakan oleh suku Melayu yang tinggal di pesisir, meliputi Bagan Siapi-api, Dumai, Sungai Pakning, Bengkalis, Tembilahan, Selat Panjang, Tangjungbatu, Tanjungbalai Karimun, dan Tanjungpinang [26]. Dengan demikian, bahasa Melayu Riau di Provinsi Riau dapat dibedakan menjadi dialek daratan dan dialek pesisir.

Penelitian ini bertujuan menyediakan sumber daya bahasa (*language resources*) Melayu Riau dalam bentuk kamus dan *stemmer* bahasa Melayu Riau. Bahasa Melayu Riau yang dibahas adalah leksikon dan morfologi dialek kepulauan, yang secara geografis meliputi dialek kepulauan dan dialek pesisir. Pendekatan yang digunakan yaitu penghapusan imbuhan berdasarkan aturan morfologi dengan melibatkan kamus bahasa. Adanya *stemmer* ini akan memungkinkan dilakukannya berbagai tugas pemrosesan bahasa alami terhadap teks berbahasa Melayu Riau yang selama ini belum ada.

2. Metode Penelitian

Tahapan-tahapan yang dilakukan diperlihatkan pada Gambar 3. Pada tahapan pertama, dilakukan pengumpulan kata dasar, aturan morfologi, dan kata uji. Kumpulan kata dasar diambil dari kata-kata berbahasa Melayu Riau yang ditemukan dalam Kamus Bahasa Indonesia – Melayu Riau [29]. Kamus dwibahasa ini memuat entri kata berbahasa Indonesia dengan penjelasan makna dalam bahasa Melayu Riau beserta contoh pemakaiannya dalam kalimat. Oleh karena tidak adanya kamus lain dalam bentuk digital yang berisikan entri kata berbahasa Melayu Riau, maka kosakata diambil dari kata-kata berbahasa Melayu Riau yang ada di penjelasan makna dan contoh kalimatnya. Kamus diunduh dalam format PDF dari repositori Kementerian Pendidikan dan Kebudayaan. Pemeriksaan dan koreksi dilakukan secara manual terhadap kata-kata yang berhasil diperoleh untuk mendapatkan kata dasarnya. Jumlah kata yang berhasil dikumpulkan sebanyak 3587 kata. Hasil pengumpulannya didiskusikan dengan validator (budayawan LAMR) untuk memperoleh masukan.

Pada tahapan pertama, juga dilakukan pengumpulan aturan pembentukan kata (morfologi) yang diambil dari buku Morfologi dan Sintaksis Bahasa Melayu Riau [27]. Untuk keperluan pengujian, dipersiapkan sebanyak 1000 kata berimbuhan. Kumpulan kata yang digunakan dalam pengujian diperoleh dari buku cerita rakyat yang berjudul Wak si Dolan (1993) dan Putra Lokan (1995).



Gambar 3. Tahapan Penelitian

Pada tahapan analisa, dilakukan penyusunan langkah-langkah penghapusan imbuhan untuk mendapatkan kata dasarnya berdasarkan aturan pembentukan kata (morfologi) yang ada. Aturan morfologi yang didapatkan meliputi:

- Penggunaan prefiks (awalan) me- (termasuk mem-, men-, meng-, meny-), pe- (termasuk pem-, pen-, peng-, peny-), be- (termasuk bel-, ber-), di-, se-, dan ter-.
- Penggunaan sufiks (akhiran) -an, -kan, i, -nye, -ku, -mu, dan -man.
- Penggunaan infiks (sisipan) -er-, -el-, -em-, dan -in-.
- Penggunaan konfiks ke-an.

Langkah-langkah penghapusan imbuhan dibuat berdasarkan aturan morfologi tersebut di atas. Setiap imbuhan merupakan kumpulan dua, tiga, atau empat huruf. Contohnya awalan ter- yang terdiri atas huruf t, e, dan r secara berurutan. Setiap langkah akan mendeteksi ada atau tidaknya kumpulan huruf yang mewakili suatu imbuhan, disebut kandidat imbuhan. Apabila terdeteksi, penghapusan kandidat imbuhan dilakukan untuk mengetahui apakah hasilnya ditemukan di kamus kata.

Pendeteksian juga dilakukan dengan menyertakan sebuah huruf yang diletakkan setelah kandidat imbuhan. Hal ini bermanfaat untuk mengetahui terjadinya peluluhan (disebut juga peleburan) fonem di awal kata dasar. Contohnya kandidat imbuhan pen- yang diikuti oleh sebuah huruf vokal mengindikasikan adanya peleburan huruf t. Apabila terdeteksi, kandidat imbuhan beserta sebuah huruf vokal yang mengikutinya akan digantikan dengan sebuah huruf baru, kemudian diperiksa apakah hasilnya ditemukan di kamus kata. Jika suatu langkah tidak menemukan kata dasar, maka dilanjutkan ke langkah lainnya untuk mendeteksi huruf-huruf yang mewakili imbuhan yang lain.

Pada tahapan perancangan, digunakan *pseudocode* (kode semu) dan *flowchart* (diagram alir). Pada tahapan implementasi, dibangun *stemmer* dalam bentuk PHP Extension dengan menggunakan bahasa pemrograman Zephir. PHP terdiri dari aplikasi inti (*core*) dan fungsi tambahan (*extension*). PHP Extension merupakan *library* atau kumpulan fungsi tambahan yang memungkinkan PHP dapat mendukung berbagai fitur baru, salah satunya pemrosesan bahasa alami. Selain itu, dibangun suatu aplikasi berbasis *web* sederhana yang menggunakan PHP Extension tersebut untuk menemukan kata dasar.

Pada tahapan pengujian, validator menentukan kata dasar dari sejumlah kata uji yang ada. Akurasi *stemmer* diketahui dengan cara membandingkan kata dasar yang diperoleh validator terhadap kata dasar yang diperoleh *stemmer*.

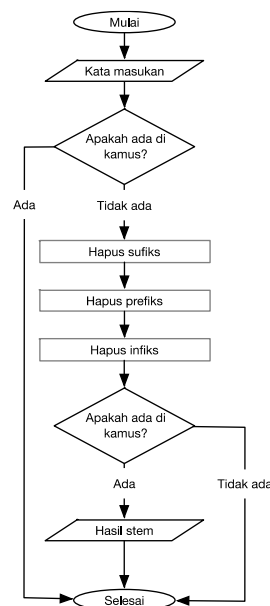
3. Hasil dan Pembahasan

Stemmer bahasa Melayu Riau dirancang bangun menggunakan pendekatan penghapusan imbuhan berdasarkan aturan morfologi dengan melibatkan kamus bahasa. Setelah dilakukan analisa berdasarkan data uji yang ada, maka urutan penghapusan yang digunakan yaitu penghapusan sufiks (termasuk bagian konfiks -an), diikuti penghapusan prefiks (termasuk bagian konfiks ke-), dan diakhiri penghapusan infiks.

Pertama kali, dilakukan pengecekan apakah masukan merupakan kata dasar. Apabila tidak ditemukan pada kamus bahasa, maka dilanjutkan ke penghapusan imbuhan-imbuhan dengan memeriksa apakah kata tersebut mengandung imbuhan. Sebelum melakukan penghapusan suatu imbuhan, diperiksa apakah hasil penghapusannya ditemukan pada kamus bahasa. Apabila tidak ditemukan, maka dilakukan penghapusan imbuhan lainnya. Akhirnya, apabila semua imbuhan berhasil dihapus, maka algoritma mengembalikan hasil stem berupa kata dasar. Sebaliknya, algoritma mengembalikan kata masukan. Sebagian *pseudocode* algoritma diperlihatkan sebagai berikut.

```
STEM(kata_masukan)
kamus_bahasa = load_dictionary()
word = kata_masukan
if word in kamus_bahasa then
    return word
word = remove_suffix(word)
word = remove_prefix(word)
word = remove_infix(word)
if word in kamus_bahasa then
    return word
else
    return kata_masukan
```

Flowchart algoritma diperlihatkan pada Gambar 4. Pada simbol proses hapus imbuhan (sufiks, prefiks, dan infiks), selalu dilakukan pengecekan apakah hasil penghapusannya ditemukan pada kamus bahasa, yang mana tidak diperlihatkan pada *flowchart* ini.



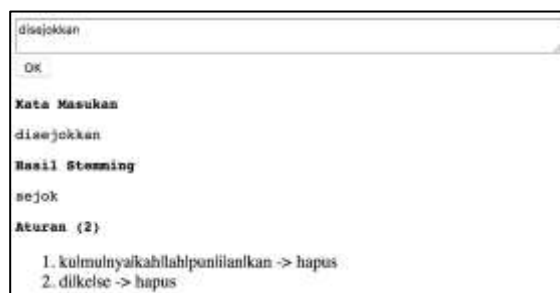
Gambar 4. *Flowchart* Algoritma

Implementasi *stemmer* menggunakan bahasa pemrograman Zephir, berupa PHP Extension. Oleh karena pengembangan dilakukan pada sistem operasi macOS, diperoleh dua file .so yaitu `php_zephir_parser.so` dan `niphy_melayu_riau.so`. *File-file* PHP Extension tersebut dimuatkan ke *web server* Apache. Untuk menggunakan PHP Extension ini, disediakan dokumentasi API baik untuk kamus maupun *stemmer*-nya sebagaimana diperlihatkan pada Gambar 5.



Gambar 5. Dokumentasi API untuk kamus dan *stemmer*

Untuk menggunakan PHP Extension tersebut, dibangun suatu aplikasi berbasis *web* sederhana yang menerima masukan berupa kata, kemudian mengembalikan kata dasarnya sebagaimana diperlihatkan pada Gambar 6.



Gambar 6. Aplikasi berbasis *web*

Selanjutnya, pengujian dilakukan dengan cara membandingkan kata dasar yang diperoleh validator terhadap kata dasar yang diperoleh *stemmer*. Berdasarkan hasil pengujian terhadap 1000 kata, diperoleh hasil akurasi *stemmer* sebesar 100%.

4. Kesimpulan

Dalam penelitian ini, dihasilkan sumber daya bahasa berupa *stemmer* bahasa Melayu Riau berdasarkan aturan morfologi. Hasil implementasinya berbentuk PHP Extension dapat digunakan oleh peneliti dan pengembang perangkat lunak di bidang pemrosesan bahasa alami. Berdasarkan pengujian, diperoleh tingkat akurasi 100% yang memperlihatkan bahwa *stemmer* ini dapat digunakan untuk pemrosesan teks berbahasa Melayu Riau.

Kedepannya, penelitian ini akan dilanjutkan terhadap dialek lainnya di Provinsi Riau, yaitu dialek daratan. Selain itu, penelitian ini juga dapat dilanjutkan ke NLP *task* lainnya seperti di tahapan analisis leksikal, seperti *spelling correction* dan *part-of-speech (POS) tagging*, atau di tahapan-tahapan analisis selanjutnya, seperti *phrase chunking*, *machine translation*, dan *transliteration*.

Daftar Pustaka

- [1] Naim, A. & Syaputra, H., (2011). Kewarganegaraan, Suku Bangsa, Agama, dan Bahasa Sehari-hari Penduduk Indonesia - Hasil Sensus Penduduk 2010. Jakarta: Badan Pusat Statistik.
- [2] Dahlan, S., Syair, A., Manan, A., Jakub, I., Sabrin, A., Razak, A., Mujiono, & Yulisman, (1985). Pemetaan Bahasa Daerah Riau dan Jambi. Jakarta: Pusat Pembinaan dan Pengembangan Bahasa, Departemen Pendidikan dan Kebudayaan.
- [3] Jamil, T. I., (2018). Pendidikan Budaya Melayu Riau. Pekanbaru, Riau, Indonesia: Lembaga Adat Melayu Riau (LAMR).
- [4] Maharani, E., (2015). Bahasa Melayu Mulai Terpinggirkan. Accessed: Dec. 23, 2020. [Online].

- Available: <https://republika.co.id/berita/nasional/umum/15/03/19/nlgl6x-bahasa-melayu-mulai-terpinggirkan>
- [5] Tondo, F. H., (2009). Kepunahan Bahasa-Bahasa Daerah: Faktor Penyebab dan Implikasi Etnolinguistik, *J. Masy. dan Budaya*, vol. 11, no. 2, pp. 277–296.
 - [6] ELRA, (2015). What is a Language Resource? <http://www.elra.info/en/about/what-language-resource/> (accessed Jul. 07, 2021).
 - [7] Bird, S. & Simons, G. G., (2003). Extending Dublin Core Metadata to Support the Description and Discovery of Language Resources, *Comput. Hum.*, vol. 37, no. 4, pp. 375–388.
 - [8] Irmawati, B., Aranta, A., Wedhaswara, W., Putra, M. I. D., & Khairunnisa, S. O., (2019). Language Resources and Tools Development for Indonesian Languages, in *Language Technologies for All (LT4All)*, pp. 231–234.
 - [9] Sari, N., (2020). Perancangan Aplikasi Kamus Bahasa Minang Indonesia dan Indonesia Minang Menggunakan Algoritma Levenshtein, *J. Mhs. FTIK (Fakultas Tek. dan Ilmu Komputer)*, vol. 1, no. 1, pp. 1113–1124.
 - [10] Kuncoro, M. A., (2014). Pengembangan Aplikasi Kamus Bahasa Jawa Berbasis Android, Universitas Muhammadiyah Surakarta.
 - [11] Sidik & Suryoprayogo, A., (2016). Implementasi Algoritma Brute Force Pada Perancangan Aplikasi Kamus Bahasa Sunda Berbasis Android, *J. Inform. dan Bisnis*, vol. 5, no. 1, pp. 1–12.
 - [12] Anna, (2016). Aplikasi Kamus Bahasa Indonesia – Bugis Berbasis Web Dengan Metode Sequential Search, *Repos. Univ. Bina Sarana Inform.*, vol. 3, no. September, pp. 246–258, [Online]. Available: <https://repository.bsi.ac.id/index.php/repo/viewitem/1215>
 - [13] Agusriandi, (2018). Rancang Bangun Aplikasi Kamus Bahasa Massenrempulu Berbasis Web, *Edumaspul - J. Pendidik.*, vol. 2, no. 1, pp. 60–69, doi: 10.33487/edumaspul.v2i1.15.
 - [14] Ginting, S. B. P., Marbun, N., Zarlis, M., & Hartama, D., (2019). Penerapan Algoritma Horspool Perancangan Aplikasi Kamus Bahasa Bima – Indonesia Kamus, in *Seminar Nasional Matematika dan Terapan*, vol. 1, pp. 887–891.
 - [15] Mahastavira, B. D. S., Sengkey, R., & Sambeka, K., (2015). Aplikasi Kamus Bahasa Indonesia-Tountemboan Menggunakan Borland Delphi 7, *Bul. Sariputra*, vol. 151, no. 1, pp. 24–39.
 - [16] Maulidi, R., (2016). Stemmer untuk Bahasa Madura dengan Modifikasi Enhanced Confix Stripping Stemmer, in *Prosiding Seminar Nasional FDI 2016*, no. December, pp. 12–15.
 - [17] Madia, M., (2016). Stemming Bahasa Jawa untuk Mencari Akar Kata dalam Bahasa Jawa dengan Aturan Analisis Kontrasif Afiksasi Verba, Universitas Islam Negeri Maulana Malik Ibrahim.
 - [18] Fauziyah, M., (2019). Stemming Bahasa Daerah Menggunakan Algoritma Levenshtein dan Analisa Morfologi, Universitas Islam Negeri Maulana Malik Ibrahim.
 - [19] Junaedi, D., Herlistiono, O., & Akbar, D., (2010). Stemmer for Basa Sunda, in *Seminar Nasional Ilmu Komputer Universitas Diponegoro*, pp. 275–278.
 - [20] Nata, G. N. M. & Yudiasra, P. P., (2017). Stemming Teks Sor-singih Bahasa Bali, in *Konferensi Nasional Sistem & Informatika*, pp. 608–612. [Online]. Available: <http://knsi.stikom-bali.ac.id/index.php/e-proceedings/article/view/111>
 - [21] Wirayasa, I. P. M., Wirawan, I. M. A., & Pradnyana, I. M. A., (2019). Algoritma Bastal: Adaptasi Algoritma Nazief & Adriani Untuk Stemming Teks Bahasa Bali, *J. Nas. Pendidik. Tek. Inform.*, vol. 8, no. 1, p. 60, doi: 10.23887/janapati.v8i1.13500.
 - [22] Indurkha, N. & Damerau, F. J., (2010). Handbook of Natural Language Processing, 2nd ed. Chapman & Hall/CRC.
 - [23] Simarangkir, M. S. H., (2017). Studi Perbandingan Algoritma - Algoritma Stemming Untuk Dokumen Teks Berbahasa Indonesia, *J. Inkofar*, vol. 1, no. 1, pp. 41–47.
 - [24] Agusta, L., (2009). Perbandingan Algoritma Stemming Porter Dengan Algoritma Nazief & Adriani Untuk Stemming Dokumen Teks Bahasa Indonesia, *Konf. Nas. Sist. dan Inform. 2009*, no. KNS&I09-036, pp. 196–201.
 - [25] Frakes, W. B. & Baeza-Yates, R. A., Eds., (1992). Information Retrieval: Data Structures & Algorithms. Prentice-Hall.
 - [26] Dahlan, S., Ramli, B., Razak, A., Zainal, A., & Bahri, S., (1991). Struktur Bahasa Melayu Riau Dialek Pesisir. Jakarta: Departemen Pendidikan dan Kebudayaan.
 - [27] Hasan, K., Yunus, M., Sugio, & Nurbaiti, (1983). Morfologi dan Sintaksis Bahasa Melayu Riau. Pusat Pembinaan dan Pengembangan Bahasa, Departemen Pendidikan dan Kebudayaan.
 - [28] Rachmawati, R., Syahrir, E., Saleh, R., Agus, S., & Riswara, Y., (2010). Persebaran dan Kekerabatan Bahasa-Bahasa di Provinsi Riau dan Kepulauan Riau, 1st ed. Pekanbaru: Balai Bahasa Provinsi Riau.
 - [29] Lubis, I., Ramli, B., Dasri, Hakim, N., & Mustafa, M. N., (1997). Kamus Bahasa Indonesia - Melayu Riau. Jakarta: Pusat Pembinaan dan Pengembangan Bahasa, Departemen Pendidikan dan Kebudayaan.