

KEYPHRASE EXTRACTION BERBAHASA INDONESIA MENGGUNAKAN
METODE *POSITIONRANK*

Diajukan Sebagai Syarat Untuk Menyelesaikan
Pendidikan Program Strata-1 Pada
Jurusan Teknik Informatika



Oleh :

Bintang Dwitama
NIM : 09021381924143

Jurusan Teknik Informatika
FAKULTAS ILMU KOMPUTER UNIVERSITAS SRIWIJAYA
2022

LEMBAR PENGESAHAN SKRIPSI

KEYPHRASE EXTRACTION BERBAHASA INDONESIA MENGGUNAKAN
METODE *POSITIONRANK*

Oleh :

Bintang Dwitama
NIM : 09021381924143

Palembang, 13 Desember 2022

Pembimbing I,



Novi Yusliani, M.T
NIP. 198211082012122001

Pembimbing II,



Muhammad Naufal Rachmatullah, M. T.
NIP. 199212012022031008



TANDA LULUS UJIAN KOMPREHENSIF

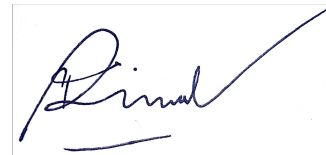
Pada hari Jumat tanggal 30 Desember 2022 Telah dilaksanakan ujian komprehensif skripsi oleh Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya

Nama : Bintang Dwitama
NIM : 09021381924143
Judul : *Keyphrase Extraction* Berbahasa Indonesia Menggunakan Metode *PositionRank*

Dan dinyatakan **LULUS**.

1. Ketua

Mastura Diana Marieska, M.T
198603212018032001



2. Penguji 1

Alvi Syahrini Utami, M.Kom
197812222006042003



3. Pembimbing 1

Novi Yusliani, M.T
198211082012122001



4. Pembimbing 2

Muhammad Naufal Rachmatullah, M.T
199212012022031008



Mengetahui,
Ketua Jurusan Teknik Informatika

Alvi Syahrini Utami, M.Kom
197812222006042003



HALAMAN PERNYATAAN

Yang bertanda tangan di bawah ini:

Nama : Bintang Dwitama
NIM : 09021381924143
Program Studi : Teknik Informatika Bilingual
Judul Skripsi : *Keyphrase Extraction* Berbahasa Indonesia Menggunakan Metode *PositionRank*

Hasil pengecekan *Software iThenticate/Turnitin*: *17%*

Menyatakan bahwa laporan skripsi saya merupakan hasil karya sendiri dan bukan hasil penjiplakan/plagiat. Apabila ditemukan unsur penjiplakan/plagiat dalam laporan skripsi ini, maka saya bersedia menerima sanksi akademik dari Universitas Sriwijaya sesuai dengan ketentuan yang berlaku.

Demikian pernyataan ini saya buat dengan sebenarnya dan tidak ada paksaan dari pihak manapun.



Palembang, 22 Desember 2022



Bintang Dwitama

NIM. 09021381924143

MOTTO DAN PERSEMBAHAN

Motto:

“Yesterday is history, tomorrow is a mystery, and today is a gift, that's why they call it present”

Kupersembahkan Karya Tulis ini kepada:

- Allah SWT
- Kedua orang tua dan keluarga besar
- Fakultas Ilmu Komputer
- Universitas Sriwijaya

ABSTRACT

Keyphrase determination is not a difficult task for information analysts. However, specify keyphrases for hundreds or even thousands of articles will take time and effort of the analysts. To ease the work, it needs to be made a system of automatic keyphrase extraction. The construction of this system passes the stages of preprocessing, scoring, and ranking the keyphrase candidates. This research was carried out by using 100 scientific papers taken from jtiik.ub.ac.id, jurnal.mdp.ac.id, and jurnal.untan.ac.id collections. This research uses *PositionRank* methodology. The result achieved suggested that the *PositionRank*-based keyphrase extraction can be used to extract keyphrase from Indonesian text, however the result are still unreliable and further improvements are necessarily suggested.

ABSTRAK

Penentuan kata kunci merupakan pekerjaan yang tidak sulit bagi para analis informasi. Namun, menentukan kata kunci untuk ratusan bahkan ribuan artikel, akan memakan waktu dan tenaga yang tidak sedikit bagi para analis informasi. Untuk meringankan pekerjaan tersebut, dibangun sebuah sistem ekstraksi kata kunci otomatis. Pembangunan sistem melewati tahapan praproses, *scoring*, dan pemeringkatan kata kunci. Eksperimen dilakukan dengan menggunakan 100 makalah ilmiah yang diambil dari koleksi jtiik.ub.ac.id, jurnal.mdp.ac.id, dan jurnal.untan.ac.id. Eksperimen ini menggunakan metode *PositionRank*, Hasil yang didapatkan menunjukkan bahwasanya ekstraksi kata kunci berbasis *PositionRank* dapat dilakukan tetapi menghasilkan perolehan yang belum dapat diandalkan dan pengembangan lebih lanjut sangat di anjurkan.

KATA PENGANTAR

Puji syukur kepada Allah SWT atas rahmat dan nikmat Nya yang lebih diberikan kepada penulis sehingga dapat menyelesaikan skripsi ini dengan baik. Skripsi ini disusun sebagai salah satu syarat menyelesaikan Pendidikan program Strata-1 di Fakultas Ilmu Komputer Universitas Sriwijaya. Dalam menyelesaikan skripsi ini, penulis menerima bantuan, bimbingan dan dukungan dari banyak pihak, baik secara langsung maupun tidak langsung. Oleh karena itu, penulis ingin menyampaikan terima kasih kepada:

1. Allah SWT atas rahmat dan nikmat-Nya penulis dapat menyelesaikan skripsi ini dengan baik.
2. Kedua orang tua, serta keluarga besar yang telah mendoakan, memberi semangat, memotivasi, dan nasihat untuk menyelesaikan skripsi ini.
3. Bapak Abdiansah, S.Kom., M.Cs. selaku Dosen dan sekaligus pembimbing akademik
4. Ibu Alvi Syahrini Utami, M. Kom. selaku Ketua Jurusan Teknik Informatika Universitas Sriwijaya
5. Ibu Novi Yusliani, M.T. selaku Dosen Pembimbing I dan Bapak Muhammad Naufal Rachmatullah, M. T. selaku dosen Pembimbing II yang telah membimbing, memberikan motivasi serta arahan kepada penulis dalam proses pengerjaan skripsi.
6. Bapak Alm. Drs. Megah Mulya, M.T. selaku Dosen yang sangat menginspirasi saya.
7. Seluruh dosen program studi serta admin Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya.

8. Adisti, Alvin, Asyraf, Fadel, Raihan, teman-teman FOVV club, dan teman-teman penulis lainnya yang telah memberikan saran, motivasi, dan semangat selama mengerjakan skripsi ini.
9. Pihak – pihak lain yang tidak dapat penulis sebutkan satu-persatu.

Penulis menyadari bahwa dalam penyusunan skripsi ini masih banyak sekali kekurangan dikarenakan kurangnya pengalaman dan pengetahuan penulis. Oleh karena itu penulis mengharapkan saran dan kritik yang membangun guna kemajuan penelitian selanjutnya. Semoga tugas akhir ini dapat bermanfaat. Terima kasih.

Palembang, 10 Desember 2022



Penulis, Bintang Dwitama

CS Scanned with CamScanner

DAFTAR ISI

	Halaman
TANDA LULUS UJIAN KOMPREHENSIF	ii
HALAMAN PERNYATAAN	iii
MOTTO DAN PERSEMBAHAN	iv
ABSTRACT	v
ABSTRAK	vi
KATA PENGANTAR	vii
BAB I PENDAHULUAN	I-1
1.1. Pendahuluan	I-1
1.2. Latar Belakang Masalah	I-1
1.3. Rumusan Masalah	I-2
1.4. Tujuan Penelitian	I-3
1.5. Manfaat Penelitian	I-3
1.6. Batasan Masalah	I-3
1.7. Sistematika Penulisan	I-4
1.8 Kesimpulan	I-5
BAB II KAJIAN LITERATUR	II-1
2. 1. Pendahuluan	II-1
2. 2. Landasan Teori	II-1
2. 2. 1. Keyphrase Extraction	II-1
2. 2. 2. Parts-of-Speech Tagging	II-4
2. 2. 3. PositionRank	II-4
2. 2. 4. <i>Confusion Matrix</i>	II-7
2. 3. Penelitian Relevan	II-8
2. 4. Kesimpulan	II-12
BAB III METODOLOGI PENELITIAN	III-1
3. 1. Pendahuluan	III-1
3. 2. Pengumpulan Data	III-1
3. 2. 1. Jenis dan Sumber Data	III-1
3. 2. 2. Metode Pengumpulan Data	III-2
3. 3. Tahap Penelitian	III-2
3. 3. 1. Mengumpulkan Data	III-2
3. 3. 2. Menetapkan Kerangka Kerja / Framework	III-3
3. 3. 3. Menetapkan Kriteria Pengujian	III-5

3. 3. 4. Format Data Pengujian.....	III-6
3. 3. 5. Menentukan Alat yang Digunakan dalam Pelaksanaan Penelitian	III-7
3. 3. 6. Melakukan Pengujian Penelitian.....	III-7
3.3.7. Melakukan Analisis Hasil Pengujian dan Membuat Kesimpulan Penelitian	III-7
3.4. Metode Pengembangan Perangkat Lunak.....	III-8
3. 4. 1. Fase Insepsi	III-8
3. 4. 2. Fase Elaborasi	III-8
3. 4. 3. Fase Konstruksi.....	III-9
3. 4. 4. Fase Transisi.....	III-9
3. 5. Manajemen Proyek Perangkat Lunak	III-10
3. 6. Kesimpulan	III-13
BAB IV PENGEMBANGAN PERANGKAT LUNAK.....	IV-1
4. 1. Pendahuluan	IV-1
4. 2. <i>Rational Unified Process</i>	IV-1
4. 2. 1. Fase <i>Inception</i>	IV-1
4. 2. 2. Elaborasi.....	IV-17
4. 2. 3. Fase Konstruksi.....	IV-21
4. 2. 4. Fase Transisi.....	IV-23
4. 3. Kesimpulan	IV-24
BAB V HASIL DAN ANALISIS PENELITIAN.....	V-1
5. 1. Pendahuluan	V-1
5. 2. Data Hasil Penelitian.....	V-1
5. 2. 1. Konfigurasi Percobaan	V-1
5. 2. 2. Data Hasil Pengujian Menggunakan <i>Cosine Similarity</i>	V-27
5. 2. 3. Data Hasil Pengujian Tanpa <i>Cosine Similarity</i>	V-28
5. 3. Analisis Hasil Penelitian	V-29
5. 4. Kesimpulan	V-31
BAB VI KESIMPULAN DAN SARAN	VI-1
6.1. Pendahuluan	VI-1
6.2. Kesimpulan	VI-1
6.3. Saran.....	VI-2
DAFTAR PUSTAKA	i
LAMPIRAN.....	vi

DAFTAR GAMBAR

Gambar II-1. <i>Tabel Confusion Matrix</i>	II-8
Gambar III-1. Rincian Kegiatan Penelitian	III-2
Gambar III-2 Kerangka kerja penelitian	III-3
Gambar III-3 Diagram Tahapan Pengujian.....	III-7
Gambar IV-1. Data yang dijadikan input	IV-6
Gambar IV-2 Diagram Use Case.....	IV-15
Gambar IV-3 Rancangan <i>Interface</i>	IV-18
Gambar IV-4 Activity Diagram	IV-20
Gambar IV-5 Diagram <i>Sequence</i>	IV-20
Gambar IV-6 Diagram <i>Class</i>	IV-21
Gambar IV-7 Implementasi Tampilan	IV-23
Gambar V-1 Visualisasi Hasil Evaluasi Metode Menggunakan <i>Cosine Similarity</i>	V-27
Gambar V-2 Visualisasi Hasil Evaluasi Metode Tanpa <i>Cosine Similarity</i>	V-29
Gambar V-3 Perbandingan Kedua Konfigurasi Pengujian.....	V-30

DAFTAR TABEL

Tabel III-1 Tabel <i>confusion matrix</i>	III-6
Tabel III-2 Tabel hasil pengujian.....	III-6
Tabel III-3 <i>Grantt Chart</i>	III-10
Tabel IV-1 Kebutuhan Fungsional	IV-3
Tabel IV-2 Kebutuhan Non-Fungsional.....	IV-3
Tabel IV-3 Hasil proses <i>Tokenizing</i>	IV-7
Tabel IV-4 Hasil <i>POS Tagging</i>	IV-7
Tabel IV-5 Hasil <i>Filtering</i>	IV-8
Tabel IV-6 Kandidat <i>keyphrase</i>	IV-8
Tabel IV-7 <i>Adjacency Matrix</i>	IV-10
Tabel IV-8 Hasil Perhitungan <i>PositionRank</i>	IV-10
Tabel IV-9 <i>Score PositionRank</i> Kandidat <i>Keyphrase</i>	IV-12
Tabel IV-10 Hasil Pemeringkatan Kandidat <i>Keyphrase</i>	IV-14
Tabel IV-11 Hasil Evaluasi	IV-15
Tabel IV-12 Definisi <i>Actor</i>	IV-16
Tabel IV-13 Definisi <i>Use Case</i>	IV-16
Tabel IV-14 Skenario <i>Use Case</i> Demonstrasi.....	IV-16
Tabel IV-15 Implementasi <i>Class</i>	IV-22
Tabel IV-16 Rencana Pengujian <i>Use Case</i> Demonstrasi	IV-24
Tabel IV-17 Pengujian <i>Use Case</i> Demonstrasi.....	IV-24
Tabel V-1 Data Untuk Evaluasi Dengan <i>Similarity Threshold</i> 40%.....	V-4
Tabel V-2 Data Untuk Evaluasi Dengan <i>Similarity Threshold</i> 60%.....	V-6

Tabel V-3 Data Untuk Evaluasi Dengan <i>Similarity Threshold</i> 80%.....	V-9
Tabel V-4 Data Untuk Evaluasi Dengan Top 5 Kandidat.....	V-11
Tabel V-5 Data Untuk Evaluasi Dengan Top 10 Kandidat.....	V-14
Tabel V-6 Data Untuk Evaluasi Dengan Top 50 Kandidat.....	V-18
Tabel V-7 Data Untuk Evaluasi Dengan Top 100 Kandidat.....	V-22
Tabel V-8 Hasil Evaluasi Dengan Konfigurasi Cosine Similarity	V-27
Tabel V-9 Hasil Evaluasi Dengan Konfigurasi Tanpa Cosine Similarity	V-28

DAFTAR LAMPIRAN

Lampiran 1. Source Code Main.py	vi
Lampiran 2 Source Code KeyphraseExtraction.py	vi
Lampiran 3 Source Code Tokenizer.py	ix
Lampiran 4 Source Code Tagger.py.....	ix
Lampiran 5 Source Code Stemmer.py	x
Lampiran 6. Hasil Evaluasi Dengan <i>Similarity Threshold</i> 40%	xi
Lampiran 7. Hasil Evaluasi Dengan <i>Similarity Threshold</i> 60%	xv
Lampiran 8. Hasil Evaluasi Dengan <i>Similarity Threshold</i> 80%	xviii
Lampiran 9 Hasil Evaluasi Dengan 5 Kandidat Teratas	xxii
Lampiran 10 Hasil Evaluasi Dengan 10 Kandidat Teratas	xxvi
Lampiran 11 Hasil Evaluasi Dengan 50 Kandidat Teratas	xxx
Lampiran 12 Hasil Evaluasi Dengan 100 Kandidat Teratas	xxxiv

BAB I

PENDAHULUAN

1.1. Pendahuluan

Pada bab ini akan diuraikan tentang latar belakang penelitian *keyphrase extraction* berbahasa Indonesia dengan metode *PositionRank*, rumusan masalah, tujuan dan manfaat penelitian, batasan atau ruang lingkup masalah, sistematika penulisan, dan kesimpulan.

1.2. Latar Belakang Masalah

Jumlah dokumen, terkhususnya dokumen publikasi ilmiah di dunia maya sangat banyak. Karena banyaknya dokumen proses pencarian dokumen yang diinginkan memakan waktu yang tidak sedikit. Untuk mengatasi permasalahan tersebut, dimanfaatkanlah *keyphrase* yang merepresentasikan dokumen, dengan begitu proses pencarian hanya perlu dilakukan dengan mencocokkan *keyphrase* yang dimiliki dokumen dengan hal yang ingin dicari.

Keyphrase extraction adalah proses mengekstrak *keyphrase* dari suatu teks bacaan. *Keyphrase* itu sendiri adalah unit terkecil yang dapat merepresentasikan suatu dokumen atau teks (Kurniawan, 2021). Telah banyak pendekatan baik *supervised* maupun *unsupervised* yang di usulkan untuk *keyphrase extraction* (Hasan & Ng, 2014). Florescu & Caragea (2017) mengajukan metode *PositionRank* yang mana metode tersebut mengimplementasikan informasi lokasi pada metode *SingleRank* berdasarkan gagasan bahwa semakin awal sebuah kandidat kata muncul

dalam sebuah dokumen, maka semakin penting kata tersebut. *PositionRank* berfokus pada kata atau frasa yang sering muncul dan posisinya dalam sebuah dokumen, Zhang et al. (2007) menambahkan bahwa posisi relatif dalam sebuah dokumen merupakan fitur yang sangat efektif dalam ekstraksi kata kunci dengan pendekatan *supervised*. *PositionRank* merupakan sebuah algoritma pemeringkatan berbasis graph yang memanfaatkan posisi dari sebuah kata dan frekuensi kemunculan kata tersebut dalam sebuah dokumen untuk menghitung nilai *biased PageRank* dari setiap kandidat kata kunci.

Florescu & Caragea (2017) membuktikan bahwa metode *PositionRank* memiliki performa hingga 29,09% lebih baik dibandingkan metode *TF-IDF*, *TextRank*, *SingleRank*, *ExpandRank*, dan *TPR*. Karena itu, peneliti akan melakukan penelitian *Keyphrase extraction* berbahasa Indonesia menggunakan metode *PositionRank* dan data makalah ilmiah.

1.3. Rumusan Masalah

Berdasarkan latar belakang masalah yang telah dijelaskan, di bagian sebelumnya, penelitian ini akan berfokus pada ekstraksi frase kunci dari teks berbahasa Indonesia. Untuk mendapatkan kata kunci yang tepat cukup sulit dilakukan secara manual karena banyak faktor-faktor yang perlu diperhitungkan rumusan masalah dalam penelitian ini adalah bagaimana melakukan *keyphrase extraction* menggunakan metode *PositionRank*.

Untuk menyelesaikan masalah tersebut, maka disusun dua pertanyaan penelitian, yaitu:

1. Bagaimana mengembangkan sistem *keyphrase extraction* berbahasa Indonesia dengan metode *PositionRank* ?
2. Bagaimana kinerja metode *PositionRank* dalam mengekstrak *keyphrase* pada teks berbahasa Indonesia

1.4. Tujuan Penelitian

Tujuan dari penelitian ini adalah sebagai berikut.

1. Menghasilkan *keyphrase* yang diekstraksi menggunakan metode *PositionRank* untuk teks berbahasa Indonesia.
2. Mengetahui kinerja metode *PositionRank* dalam mengekstrak *keyphrase* pada teks berbahasa Indonesia.

1.5. Manfaat Penelitian

Manfaat dari penelitian ini adalah sebagai berikut.

1. Hasil penelitian dapat digunakan untuk mengekstrak kata kunci pada teks berbahasa Indonesia
2. Dapat menjadi rujukan untuk penelitian terkait

1.6. Batasan Masalah

Batasan masalah dari penelitian ini adalah sebagai berikut.

1. Jenis data yang digunakan dalam penelitian ini adalah artikel ilmiah berbahasa Indonesia

2. Keyphrase yang dihasilkan tersusun dari setidaknya satu kata yang memiliki kelas kata nomina dan diikuti oleh maksimal satu kata yang memiliki kelas kata adjektiva.

1.7. Sistematika Penulisan

BAB I. PENDAHULUAN

Bab ini membahas tentang latar belakang, rumusan masalah, tujuan penelitian, manfaat penelitian, batasan masalah, dan sistematika penulisan yang digunakan dalam menyusun laporan akhir ini.

BAB II. TINJAUAN PUSTAKA

Bab ini menjelaskan landasan teori yang digunakan dalam penelitian. Bab ini memuat penjelasan tentang penelitian terdahulu yang relevan dengan penelitian ini, penjelasan tentang PositionRank, serta penjelasan lain yang berkaitan dengan penelitian ini.

BAB III. METODOLOGI PENELITIAN

Bab ini membahas tahapan-tahapan yang akan dilakukan dalam penelitian ini. Setiap rencana tahapan penelitian dijelaskan secara rinci dengan mengacu pada kerangka kerja. Pada bagian akhir bab ini berisi tentang perancangan manajemen proyek dalam melakukan penelitian.

BAB IV. PENGEMBANGAN PERANGKAT LUNAK

Pada bab ini akan dibahas mengenai perancangan perangkat lunak yang akan dibangun pada penelitian ini.

BAB V. HASIL DAN ANALISIS PENELITIAN

Bab ini akan memaparkan hasil pengujian berdasarkan langkah-langkah yang telah direncanakan. Analisis tersebut diberikan sebagai dasar kesimpulan yang ditarik dalam penelitian ini.

BAB VI. KESIMPULAN DAN SARAN

Bab ini berisi kesimpulan dari semua uraian pada bab-bab sebelumnya dan juga berisi saran-saran yang diharapkan dapat berguna dalam pengembangan perangkat lunak ini selanjutnya.

1.8 Kesimpulan

Penelitian ini akan membahas tentang ekstraksi kata kunci dengan menggunakan *PositionRank*

DAFTAR PUSTAKA

- Hu, X., & Wu, B. (2006, December). Automatic keyword extraction using linguistic features. In *Sixth IEEE International Conference on Data Mining-Workshops (ICDMW'06)* (pp. 19-23). IEEE.
- Sari, D. P., & Purwarianti, A. (2014). Ekstraksi Kata Kunci Otomatis Untuk Dokumen Bahasa Indonesia Studi Kasus: Artikel Jurnal Ilmiah Koleksi Pdiilipi. *BACA: Jurnal Dokumentasi dan Informasi*, 35(2), 139-147.
- Hasan, K. S., & Ng, V. (2014, June). Automatic keyphrase extraction: A survey of the state of the art. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)* (pp. 1262-1273).
- Frank, E., Paynter, G. W., Witten, I. H., Gutwin, C., & Nevill-Manning, C. G. (1999). Domain-specific keyphrase extraction. In *16th International joint conference on artificial intelligence (IJCAI 99)* (Vol. 2, pp. 668-673). Morgan Kaufmann Publishers Inc., San Francisco, CA, USA.
- Turney, P. D. (2000). Learning algorithms for keyphrase extraction. *Information retrieval*, 2(4), 303-336.
- Nguyen, T. D., & Kan, M. Y. (2007, December). Keyphrase extraction in scientific publications. In *International conference on Asian digital libraries* (pp. 317-326). Springer, Berlin, Heidelberg.
- Romary, P. L. L. (2010). Automatic key term extraction from scientific articles in grobid. In *SemEval 2010 Workshop* (p. 4).

- Chuang, J., Manning, C. D., & Heer, J. (2012). “Without the clutter of unimportant words” Descriptive keyphrases for text visualization. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 19(3), 1-29.
- Caragea, C., Bulgarov, F., Godea, A., & Gollapalli, S. D. (2014, October). Citation-enhanced keyphrase extraction from research papers: A supervised approach. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)* (pp. 1435-1446).
- Liu, Z., Li, P., Zheng, Y., & Sun, M. (2009, August). Clustering to find exemplar terms for keyphrase extraction. In *Proceedings of the 2009 conference on empirical methods in natural language processing* (pp. 257-266).
- Mihalcea, R., & Tarau, P. (2004, July). TextRank: Bringing order into text. In *Proceedings of the 2004 conference on empirical methods in natural language processing* (pp. 404-411).
- Liu, Z., Huang, W., Zheng, Y., & Sun, M. (2010, October). Automatic keyphrase extraction via topic decomposition. In *Proceedings of the 2010 conference on empirical methods in natural language processing* (pp. 366-376).
- Lopez, P., & Romary, L. (2010, July). HUMB: Automatic key term extraction from scientific articles in GROBID. In *Proceedings of the 5th international workshop on semantic evaluation* (pp. 248-251).
- Wicaksono, A. F., & Purwarianti, A. (2010, August). HMM based part-of-speech tagger for Bahasa Indonesia. In *Fourth International MALINDO Workshop, Jakarta*.

- Florescu, C., & Caragea, C. (2017, July). Positionrank: An unsupervised approach to keyphrase extraction from scholarly documents. In *Proceedings of the 55th annual meeting of the association for computational linguistics (volume 1: long papers)* (pp. 1105-1115).
- Zhang, Y., Milios, E., & Zincir-Heywood, N. (2007). A comparative study on key phrase extraction methods in automatic web site summarization. *J. Digit. Inf. Manag.*, 5(5), 323-332.
- Wan, X., & Xiao, J. (2008, July). Single document keyphrase extraction using neighborhood knowledge. In *AAAI* (Vol. 8, pp. 855-860).
- Siswadi, I. (2013). Mengenal konsep penetapan kata kunci. *Jurnal Pustakawan Indonesia*, 12(2).
- Hulth, A. (2003). Improved automatic keyword extraction given more linguistic knowledge. In *Proceedings of the 2003 conference on Empirical methods in natural language processing* (pp. 216-223).
- Lamasigi, Z. Y. (2021). DCT Untuk Ekstraksi Fitur Berbasis GLCM Pada Identifikasi Batik Menggunakan K-NN. *Jambura Journal of Electrical and Electronics Engineering*, 3(1), 1-6.
- Nasar, Z., Jaffry, S. W., & Malik, M. K. (2019). Textual keyword extraction and summarization: State-of-the-art. *Information Processing & Management*, 56(6), 102088.
- Gollapalli, S. D., & Caragea, C. (2014, June). Extracting keyphrases from research papers using citation networks. In *Proceedings of the AAAI conference on artificial intelligence* (Vol. 28, No. 1).

- Lahiri, S., Choudhury, S. R., & Caragea, C. (2014). Keyword and keyphrase extraction using centrality measures on collocation networks. *arXiv preprint arXiv:1401.6571*.
- Martinez-Romo, J., Araujo, L., & Duque Fernandez, A. (2016). Sem Graph: Extracting keyphrases following a novel semantic graph-based approach. *Journal of the Association for Information Science and Technology*, 67(1), 71-82.
- Bougouin, A., Boudin, F., & Daille, B. (2013, October). Topicrank: Graph-based topic ranking for keyphrase extraction. In *International joint conference on natural language processing (IJCNLP)* (pp. 543-551).
- T. H. Haveliwala. (2003). Topic-sensitive PageRank: a context-sensitive ranking algorithm for Web search, in *IEEE Transactions on Knowledge and Data Engineering*, vol. 15, no. 4, pp. 784-796.
- Danesh, S., Sumner, T., & Martin, J. H. (2015, June). Sgrank: Combining statistical and graphical methods to improve the state of the art in unsupervised keyphrase extraction. In *Proceedings of the fourth joint conference on lexical and computational semantics* (pp. 117-126).
- Le, T. T. N., Nguyen, M. L., & Shimazu, A. (2016, December). Unsupervised keyphrase extraction: Introducing new kinds of words to keyphrases. In *Australasian Joint Conference on Artificial Intelligence* (pp. 665-671). Springer, Cham.
- Adar, E., & Datta, S. (2015, July). Building a scientific concept hierarchy database (schbase). In *Proceedings of the 53rd Annual Meeting of the Association for*

Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers) (pp. 606-615).

Wang, R., Liu, W., & McDonald, C. (2014, July). Corpus-independent generic keyphrase extraction using word embedding vectors. In *Software engineering research conference* (Vol. 39, pp. 1-8).