

EKSTRAKSI KATA KUNCI ARTIKEL ILMIAH DI BIDANG
INFORMATIKA MENGGUNAKAN METODE *YAKE* DAN
KPMINER

*Diajukan Sebagai Syarat Untuk Menyelesaikan
Pendidikan Program Strata-1 Pada
Jurusan Teknik Informatika Fakultas Ilmu Komputer UNSRI*



Oleh :

Bayu Daru Pangestu

NIM : 09021282025085

Jurusan Teknik Informatika
FAKULTAS ILMU KOMPUTER UNIVERSITAS SRIWIJAYA
2024

LEMBAR PENGESAHAN SKRIPSI

EKSTRAKSI KATA KUNCI ARTIKEL ILMIAH DI BIDANG
INFORMATIKA MENGGUNAKAN METODE *YAKE* DAN
KPMINER

Oleh :

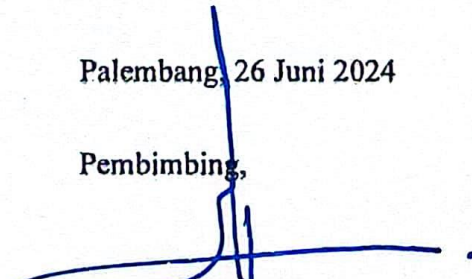
Bayu Daru Pangestu

NIM : 09021282025085



Palembang, 26 Juni 2024

Pembimbing,



Dr. Abdiansah, S.Kom., M.Cs.
NIP. 198410012009121005

TANDA LULUS UJIAN KOMPREHENSIF

Pada hari Rabu tanggal 5 Juni 2024 telah dilaksanakan ujian komprehensif skripsi oleh Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya

Nama : Bayu Daru Pangestu

NIM : 09021282025085

Judul : Ekstraksi Kata Kunci Artikel Ilmiah Di Bidang Informatika

Menggunakan Metode *Yake* Dan *KP Miner*

dan dinyatakan **LULUS**.

1. Ketua

Desty Rodiah, S.Kom., M.T.

NIP. 198912212020122011



2. Penguji

Hadipurnawan Satria., Ph.D

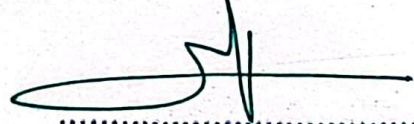
NIP. 198004182020121001



3. Pembimbing

Dr. Abdiansah, S.Kom., M.CS.

NIP. 198410012009121005



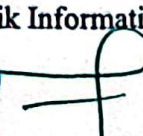
Mengetahui,

Ketua Jurusan Teknik Informatika



Dr. M. Fachrurrozi, S.Si., M.T.

NIP. 198005222008121002



HALAMAN PERNYATAAN

Yang bertanda tangan di bawah ini:

Nama : Bayu Daru Pangestu

NIM : 09021282025085

Program Studi : Teknik Informatika

Judul Skripsi : Ekstraksi Kata Kunci Artikel Ilmiah Di Bidang Informatika
Menggunakan Metode *Yake* Dan *KP Miner*

Hasil Pengecekan *Software iThenticate/Turnitin* : 17%

Menyatakan bahwa laporan proyek saya merupakan hasil karya sendiri dan bukan hasil penjiplakan/plagiat/ Apabila ditemukan unsur penjiplakan/plagiat dalam laporan proyek 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 mana pun.



Palembang, 12 Mei 2024



Bayu Daru Pangestu
NIM. 09021282025085

MOTTO DAN PERSEMBAHAN

Motto :

“Allah tidak membebani seseorang melainkan sesuai dengan kesanggupannya.”

- Al Baqarah : 286

“Even in the darkest times, a single ray of light can illuminate the path forward”

- Anonymous

Kupersembahkan Karya Tulis ini kepada :

- Allah SWT
- Kedua orang tua dan saudara saya
- Fakultas Ilmu Komputer
- Universitas Sriwijaya

ABSTRACT

Currently, the abundance of scientific journals is continually growing. Keyword identification can represent the general topics discussed in a journal, but this process requires a deep understanding of the journal's content, which is time-consuming. To address the challenge of formulating keywords, the solution lies in developing an automatic keyword extraction system. One of the approaches used is statistical-based methods, such as Yake and KP Miner. This research aims to perform keyword extraction using the Yake (Yet Another Keyword Extraction) and KP Miner methods, along with an experimental method combining both approaches. The dataset consists of 100 scientific articles collected from various sources. Based on the research findings, the best parameter configuration involves selecting all top candidates using KP Miner, resulting in an F-Score of 0.556.

Keywords : Keyphrase Extraction, Yake, KP Miner, Scientific Articles.

ABSTRAK

Saat ini, jumlah jurnal ilmiah yang beredar sangat melimpah dan terus berkembang. Identifikasi kata kunci dapat mewakili topik pembahasan secara umum dalam sebuah jurnal, tetapi proses ini membutuhkan pemahaman mendalam terhadap isi jurnal yang memakan waktu. Untuk mengatasi tantangan dalam merumuskan kata kunci, solusinya adalah dengan mengembangkan sistem ekstraksi kata kunci otomatis. Salah satu pendekatan yang digunakan adalah metode berbasis statistik, seperti *Yake* dan *KP Miner*. Penelitian ini bertujuan untuk melakukan ekstraksi kata kunci menggunakan metode *Yake (Yet Another Keyword Extraction)* dan *KP Miner*, dan ditambah metode percobaan dengan Gabungan antara kedua Metode Tersebut. *Dataset* yang digunakan terdiri dari 100 artikel ilmiah yang diambil dari beberapa sumber. Berdasarkan hasil penelitian, konfigurasi parameter terbaik adalah dengan mengambil semua kandidat teratas menggunakan *KP Miner*, menghasilkan *F-Score* sebesar 0,556.

Kata Kunci : Ekstraksi Kata Kunci, *Yake*, *KP Miner*, Artikel Ilmiah.

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 proses penyelesaian tesis ini, penulis mendapat dukungan, bimbingan, dan bantuan dari berbagai pihak, baik secara langsung maupun tidak langsung. Dengan demikian, penulis ingin mengungkapkan rasa terima kasih kepada:

1. Allah SWT atas rahmat dan nikmat-Nya penulis dapat menyelesaikan skripsi ini dengan baik.
2. Kedua orang tua, saudara dan keluarga besar yang telah mendoakan, memberi semangat, memotivasi, dan nasihat untuk menyelesaikan skripsi ini.
3. Ibu Desty Rodiah, S.Kom., M.T. selaku Dosen dan sekaligus pembimbing akademik.
4. Bapak Dr. M. Fachrurrozi, S.Si., M.T. selaku Ketua Jurusan Teknik Informatika Universitas Sriwijaya.
5. Bapak Dr. Abdiansah, S.Kom, M.Cs. selaku Dosen Pembimbing I yang telah membimbing, memberikan motivasi serta arahan kepada penulis dalam proses pengerjaan skripsi.
6. Teman-teman L Gasos yang telah menemani penulis, memberikan saran, motivasi, dan semangat dalam mengerjakan skripsi ini

7. Teman-teman MHTSBA yang beranggotakan Miftah, Hani, Tharisa, Syafiq dan Aulia yang telah membantu memberikan dukungan, saran dan semangat selama mengerjakan skripsi ini.
8. Seluruh Staf Administrasi dan Pegawai yang telah membantu dalam urusan administrasi
9. Teman – Teman penulis yang telah memberikan saran, motivasi, dan semangat selama mengerjakan skripsi ini
10. 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, 5 Juni 2024

Penulis

Bayu Daru Pangestu

DAFTAR ISI

LEMBAR PENGESAHAN SKRIPSI	ii
TANDA LULUS UJIAN KOMPREHENSIF	iii
HALAMAN PERNYATAAN	iv
MOTTO DAN PERSEMBAHAN	v
ABSTRACT	vi
ABSTRAK	vii
KATA PENGANTAR.....	viii
DAFTAR ISI	x
DAFTAR TABEL.....	xiii
DAFTAR GAMBAR	xvi
BAB I PENDAHULUAN	I-1
1.1. Pendahuluan	I-1
1.2. Latar Belakang	I-1
1.3. Rumusan Masalah	I-3
1.4. Tujuan Penulisan	I-3
1.5. Manfaat Penelitian.....	I-4
1.6. Batasan Masalah.....	I-4
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. <i>Keyword Extraction</i>	II-1
2.2.2. <i>KP Miner</i>	II-3
2.2.2.1. Candidate Keyphrase Selection.....	II-4
2.2.2.2. Candidate Keyphrases Weight Calculation	II-4
2.2.2.3. Final Candidate Phrase List Refinement	II-6
2.2.3. <i>Yake</i>	II-6
2.2.3.1. Text Pre-Processing	II-7

2.2.3.2.	Feature Extraction	II-7
2.2.3.3.	Individual Term Weighting.....	II-11
2.2.3.4.	Candidate Keyword List Generation	II-11
2.2.4.	<i>Levenshtein Distance</i>	II-12
2.2.5.	<i>Evaluation Matrix</i>	II-13
2.2.6.	<i>Rational Unified Process</i>	II-14
2.3.	Penelitian Lain yang Relevan.....	II-16
2.4.	Kesimpulan.....	II-17
BAB III METODE 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.	Tahapan Penelitian.....	III-2
3.3.1.	Mengumpulkan Data.....	III-3
3.3.2.	Menentukan Kerangka Kerja Penelitian	III-3
3.3.3.	Menentukan Kriteria Pengujian	III-4
3.3.4.	Menentukan Format Data Pengujian.....	III-4
3.3.5.	Menentukan Alat Bantu Penelitian.....	III-5
3.3.6.	Melakukan Pengujian Penelitian.....	III-6
3.3.7.	Melakukan Analisis dan Menarik Kesimpulan Penelitian	III-6
3.4.	Kesimpulan.....	III-7
BAB IV PENGEMBANGAN PERANGKAT LUNAK		IV-8
4.1	Pendahuluan	IV-8
4.2	Fase Insepsi	IV-8
4.2.1	Pemodelan Bisnis	IV-8
4.2.2	Kebutuhan Sistem	IV-9
4.2.3	Analisis dan Desain.....	IV-10
4.3	Fase Elaborasi.....	IV-32
4.3.1	Pemodelan Bisnis	IV-33
4.3.2	Kebutuhan Sistem	IV-34
4.3.3	Analisis dan Perancangan	IV-34

4.4	Fase Konstruksi	IV-36
4.4.1	Kebutuhan Sistem	IV-36
4.4.2	Implementasi	IV-37
4.5	Fase Transisi	IV-38
4.5.1	Pemodelan Bisnis	IV-38
4.5.2	Rencana Pengujian	IV-38
4.5.3	Implementasi	IV-38
4.6	Kesimpulan.....	IV-39
BAB V HASIL DAN ANALISIS PENELITIAN		V-1
5.1	Pendahuluan	V-1
5.2	Hasil Penelitian.....	V-1
5.3	Analisis Hasil Penelitian	V-32
5.4	Kesimpulan.....	V-35
BAB VI KESIMPULAN DAN SARAN		VI-1
6.1	Pendahuluan	VI-1
6.2	Kesimpulan.....	VI-1
6.3	Saran	VI-2
DAFTAR PUSTAKA		iii
LAMPIRAN.....		vi

DAFTAR TABEL

Tabel II-1 Perbedaan <i>Keyword</i> dan <i>Keyphrase</i> Extraction	II-2
Tabel III-1. Hasil kinerja seluruh <i>dataset</i>	III-5
Tabel III-2. Hasil Analisis Ekstraksi Kata Kunci.....	III-6
Tabel IV-1. Kebutuhan Fungsional	IV-9
Tabel IV-2. Kebutuhan Non-Fungsional	IV-10
Tabel IV-3. Contoh Data Judul.....	IV-11
Tabel IV-4. Contoh Data Abstrak	IV-11
Tabel IV-5. Hasil Penggabungan Teks Judul dan Abstrak.....	IV-12
Tabel IV-6. Hasil Tokenisasi	IV-13
Tabel IV-7. Hasil Penghapusan Karakter Spesial.....	IV-14
Tabel IV-8. Hasil <i>Term Frequency</i>	IV-15
Tabel IV-9. Hasil <i>Word Casing</i>	IV-16
Tabel IV-10. Hasil <i>Word Position</i>	IV-18
Tabel IV-11. Hasil <i>Word Frequency</i>	IV-19
Tabel IV-12. Hasil <i>Word Relatedness to Context</i>	IV-20
Tabel IV-13. Hasil <i>Word DifSentence</i>	IV-21
Tabel IV-14. Hasil <i>Individual Term Weighting</i>	IV-22
Tabel IV-15. Hasil <i>Candidate Keyword List Generation</i>	IV-24
Tabel IV-16. Hasil <i>Levensthein Distance</i>	IV-25
Tabel IV-17. Hasil <i>Candidate Keyphrase Selection</i>	IV-26
Tabel IV-18. Hasil <i>Candidate Keyphrases Weight Calculation</i>	IV-27
Tabel IV-19. Hasil <i>Final Candidate Phrase List Refinement</i>	IV-28

Tabel IV-21. Hasil <i>Cosine Similarity</i> Sampel Jurnal Menggunakan <i>KP Miner</i> ..	IV-
Error! Bookmark not defined.	
Tabel IV-22. Hasil <i>Evaluation Matrix Yake</i>	IV-30
Tabel IV-23. Hasil <i>Evaluation Matrix KP Miner</i>	IV-30
Tabel IV-24. Definisi <i>Actor</i>	IV-31
Tabel IV-25. Definisi <i>Use Case</i>	IV-31
Tabel IV-26. <i>Scenario Use Case</i>	IV-32
Tabel IV-27. Rencana Pengujian <i>Use Case</i> Ekstraksi Kata Kunci.....	IV-38
Tabel IV-28. Pengujian <i>Use Case</i> Ekstraksi Kata Kunci	IV-39
Tabel V-1. Sampel Data yang Digunakan pada Proses Pengujian.....	V-2
Tabel V-2. Candidate Keyphrase yang telah dihasilkan oleh Yake	V-6
Tabel V-3. Tabel <i>Selected Keyphrase</i> Yake untuk Top 5	V-6
Tabel V-4. Tabel <i>Selected Keyphrase</i> Yake untuk Top 10	V-7
Tabel V-5. Tabel <i>Selected Keyphrase</i> Yake untuk Top 20	V-8
Tabel V-6. Tabel <i>Selected Keyphrase</i> Yake untuk Top 50.....	V-9
Tabel V-7. Tabel <i>Evaluation Matrix</i> Yake untuk Top 5	V-12
Tabel V-8. Tabel <i>Evaluation Matrix</i> Yake untuk Top 10	V-12
Tabel V-9. Tabel <i>Evaluation Matrix</i> Yake untuk Top 20	V-13
Tabel V-10. Tabel <i>Evaluation Matrix</i> Yake untuk Top 50	V-13
Tabel V-11. Rata-rata pengukuran hasil kata kunci Yake	V-14
Tabel V-12. Candidate Keyphrase yang telah dihasilkan oleh <i>KP Miner</i>	V-15
Tabel V-13. Tabel <i>Selected Keyphrase</i> <i>KP Miner</i> untuk Top 5	V-15
Tabel V-14. Tabel <i>Selected Keyphrase</i> <i>KP Miner</i> untuk Top 10	V-16

Tabel V-15. Tabel <i>Selected Keyphrase</i> KP Miner untuk Top 20	V-17
Tabel V-16. Tabel <i>Selected Keyphrase</i> KP Miner untuk Seluruh Kata Kunci...	V-18
Tabel V-17. Tabel <i>Evaluation Matrix</i> KP Miner untuk Top 5	V-20
Tabel V-18. Tabel <i>Evaluation Matrix</i> KP Miner untuk Top 10	V-20
Tabel V-19. Tabel <i>Evaluation Matrix</i> KP Miner untuk Top 20	V-21
Tabel V-20. Tabel <i>Evaluation Matrix</i> KP Miner untuk Seluruh Kata Kunci	V-21
Tabel V-21. Rata-rata pengukuran hasil kata kunci KP Miner	V-22
Tabel V-22. Candidate Keyphrase yang telah dihasilkan oleh Yake-KP Miner	V-23
Tabel V-23. Tabel <i>Selected Keyphrase</i> Yake-KP Miner untuk Top 5	V-23
Tabel V-24. Tabel <i>Selected Keyphrase</i> Yake-KP Miner untuk Top 10	V-24
Tabel V-25. Tabel <i>Selected Keyphrase</i> Yake-KP Miner untuk Top 20	V-25
Tabel V-26. Tabel <i>Selected Keyphrase</i> Yake-KP Miner untuk Top 50	V-26
Tabel V-27. Tabel <i>Evaluation Matrix</i> Yake-KP Miner untuk Top 5	V-29
Tabel V-28. Tabel <i>Evaluation Matrix</i> Yake-KP Miner untuk Top 10	V-29
Tabel V-29. Tabel <i>Evaluation Matrix</i> Yake-KP Miner untuk Top 20	V-30
Tabel V-30. Tabel <i>Evaluation Matrix</i> Yake-KP Miner untuk Top 50	V-30
Tabel V-31. Rata-rata pengukuran hasil kata kunci Yake KP Miner	V-31

DAFTAR GAMBAR

Gambar II-1. Arsitektur Sistem <i>KP Miner</i>	II-3
Gambar II-2. Arsitektur Sistem <i>Yake</i>	II-6
Gambar II-3. Perbandingan <i>String</i> menggunakan GLD	II-12
Gambar II-4. Arsitektur RUP (IBM, 1998)	II-15
Gambar III-1. Rencana Kegiatan Penelitian	III-15
Gambar III-2. Kerangka Kerja Penelitian	III-3
Gambar IV-1. Diagram <i>Use Case</i>	IV-30
Gambar IV-2. Desain Antarmuka Aplikasi.....	IV-33
Gambar IV-3. Diagram <i>Activity</i>	IV-35
Gambar IV-4. <i>Sequence</i> Diagram Ekstraksi Kata Kunci.....	IV-36
Gambar IV-5. Implementasi Tampilan	IV-37
Gambar V-1. Diagram evaluasi <i>Yake</i>	V-14
Gambar V-2. Diagram Evaluasi <i>KP Miner</i>	V-22
Gambar V-3 Diagram Evaluasi <i>Yake-KPMiner</i>	V-31
Gambar V-4. Matrix Pengukuran Top 5	V-32
Gambar V-5. Matrix Pengukuran Top 10	V-33
Gambar V-6. Matrix Pengukuran Top 20	V-34

BAB I

PENDAHULUAN

1.1. Pendahuluan

Pada bab ini akan membahas garis besar pokok-pokok pikiran dalam penelitian. Pokok-pokok pikiran yang akan dibahas antara lain latar belakang masalah, rumusan masalah, tujuan penelitian, manfaat penelitian, dan batasan masalah. Pokok-pokok pikiran yang dijelaskan ditujukan sebagai acuan dalam penelitian ini. Skripsi ini menguraikan cara membangun sistem ekstraksi kata kunci dengan mengadopsi metode *Yake* dan *KP Miner*. Dalam sistem ini, metode *Yake* dan *KP Miner* digunakan untuk menghasilkan kata kunci yang relevan, sehingga informasi yang diperoleh sesuai dengan konten keseluruhan yang dibahas, dan dapat memenuhi kebutuhan informasi yang penting.

1.2. Latar Belakang

Banyaknya dokumen yang tersebar di internet menyebabkan pencarian, penyusunan, pemeliharaan dokumen menjadi sulit untuk dilakukan. Kata kunci merupakan ringkasan singkat dari sebuah dokumen yang memberikan solusi dalam mencari, Menyusun dan memelihara sebuah dokumen. Kata kunci digunakan secara luas di dalam perpustakaan digital ataupun dalam mencari informasi (Liu *et al.*, 2009).

Pemberian daftar kata kunci untuk setiap dokumen merupakan hal yang cukup sulit, terutama jika dilakukan oleh orang lain yang bukan pengarang dari

dokumen tersebut contohnya petugas perpustakaan *online* atau administrator web penyedia dokumen makalah (Satia *et al.*, 2012).

Saat ini, di mana banyak informasi tersedia dalam bentuk teks di internet. Menjadikan penggunaan ekstraksi kata kunci lebih penting untuk memangkas waktu dan tenaga (Baruni & Sathiascelan, 2020). Ekstraksi kata kunci diartikan sebagai kegiatan yang dilakukan secara otomatis untuk mengidentifikasi sekumpulan istilah yang menggambarkan isi dari dokumen (Arini *et al.*, 2019).

Ada berbagai algoritma untuk ekstraksi kata kunci, termasuk pendekatan berbasis statistik, graf, dan *embedding*. *PageRank* dan *TextRank* adalah contoh pendekatan graf, sementara *TF-IDF*, *Rake*, *Yake*, dan *KP-Miner* menggunakan pendekatan statistik. Menurut (Ying *et al.*, 2017) metode seperti *PageRank*, *TextRank*, *Rake*, dan *TF-IDF* memiliki kelemahan yaitu sering mengabaikan struktur dokumen, seperti pentingnya posisi kata dalam kalimat atau paragraf, di mana kata-kata di awal atau akhir dokumen bisa lebih signifikan dibandingkan yang di tengah.

Berdasarkan temuan penelitian, terdapat bukti yang menunjukkan bahwa metode gabungan dari *Yake* dan *KP-Miner* menghasilkan kinerja yang lebih unggul dibandingkan dengan keenam algoritma lainnya (Piskorski *et al.*, 2021). Oleh karena itu, penulis akan menjalankan studi tentang Proses Ekstraksi Kata Kunci pada bahasa Indonesia, di mana pendekatan yang akan digunakan adalah metode *Yake* dan *KP Miner* dengan menggunakan data dari publikasi ilmiah.

Sebagai perbandingan dari penelitian sebelumnya, penulis menggunakan *dataset* berbahasa Indonesia dan Levenshtein Distance untuk menghilangkan kata kunci yang serupa. Dimana pada penelitian sebelumnya menggunakan dataset dari 7 bahasa yang berbeda yaitu Inggris, Prancis, Jerman, Italia, Polish, Rumania dan Spanyol. Serta menggunakan *Weighted Longest Common Substrings* untuk menghilangkan kata kunci yang serupa (Piskorski et al., 2021).

1.3. Rumusan Masalah

Berdasarkan latar belakang yang telah dijelaskan, rumusan masalah pada penelitian ini adalah :

1. Bagaimana mengembangkan model Ekstraksi Kata Kunci menggunakan algoritma *Yake* dan *KP Miner*?
2. Bagaimana nilai *F-Score* dari model Ekstraksi Kata Kunci menggunakan algoritma *Yake* dan *KP Miner*?

1.4. Tujuan Penulisan

Tujuan utama dari pelaksanaan penelitian ini adalah:

1. Mengembangkan model yang dapat mengekstraksi kata kunci dari Jurnal Informatika menggunakan metode *Yake* dan *KP Miner*.
2. Mengetahui nilai *F-Score* dari model Ekstraksi Kata Kunci pada Jurnal Informatika menggunakan metode *Yake* dan *KP Miner*.

1.5. Manfaat Penelitian

Manfaat yang diperoleh dari penelitian ini adalah:

1. Model mampu melaksanakan Ekstraksi Kata Kunci pada Jurnal Informatika dengan menggunakan metode *Yake* dan *KP Miner*.
2. Hasil penelitian ini dapat dijadikan referensi untuk penelitian-penelitian terkait kedepannya.

1.6. Batasan Masalah

Batasan masalah dari penelitian ini adalah :

1. Objek penelitian berfokus pada publikasi ilmiah berbahasa Indonesia.
2. Data yang digunakan diperoleh dari 100 publikasi ilmiah yang diunduh dari situs web *jtiik*¹, *jatisi*², *jepin*³.

1.7. Sistematika Penulisan

Dalam penelitian ini, penulis menyusun sistematika penulisan sebagai berikut :

BAB I. PENDAHULUAN

Bab ini menjelaskan latar belakang, perumusan masalah, tujuan penelitian, manfaat penelitian, batasan masalah, serta sistematika penulisan.

¹ <https://jtiik.ub.ac.id/index.php/jtiik/index>

² <https://jurnal.mdp.ac.id/index.php/jatisi/issue/archive>

³ <https://jurnal.untan.ac.id/index.php/jepin/index>

BAB II. KAJIAN LITERATUR

Pada bab ini akan dibahas mengenai landasan teori yang digunakan dalam penelitian, seperti definisi dari *Keyphrase Extraction*, metode *Yake* dan Metode *KP Miner* serta penelitian lain yang relevan dengan penelitian ini.

BAB III. METODOLOGI PENELITIAN

Bab ini akan membahas tahapan dan data yang akan digunakan dalam penelitian ini. Selain itu, akan dijelaskan secara rinci rencana tahapan penelitian, metode pengembangan perangkat lunak, serta manajemen proyek penelitian.

1.8. Kesimpulan

Pada Bab ini telah dibahas mengenai dasar-dasar penelitian serta acuan penting yang meliputi latar belakang, rumusan masalah, tujuan penelitian, manfaat penelitian, batasan masalah, dan sistematika penulisan.

DAFTAR PUSTAKA

- Arini, N. W. S., Widja, I. B. P., & Negara, I. K. R. Y. (2019). Analisis Frekuensi Kata untuk Mengekstrak Kata Kunci dari Artikel Ilmiah Berbahasa Indonesia. *Eksplora Informatika*, 8(2), 80–84. <https://doi.org/10.30864/eksplora.v8i2.162>
- Baruni, J. S., & Sathiaseelan, Dr. J. G. R. (2020). *Keyphrase Extraction from Document Using RAKE and TextRank Algorithms*.
- Behara, K. N. S., Bhaskar, A., & Chung, E. (2020). A novel approach for the structural comparison of origin-destination matrices: Levenshtein distance. *Transportation Research Part C: Emerging Technologies*, 111, 513–530. <https://doi.org/10.1016/j.trc.2020.01.005>
- Campos, R., Mangaravite, V., Pasquali, A., Jorge, A., Nunes, C., & Jatowt, A. (2020). YAKE! Keyword extraction from single documents using multiple local features. *Information Sciences*, 509, 257–289. <https://doi.org/10.1016/j.ins.2019.09.013>
- Denny, M. J., & Spirling, A. (2018). Text Preprocessing For Unsupervised Learning: Why It Matters, When It Misleads, And What To Do About It. *Political Analysis*, 26(2), 168–189. <https://doi.org/10.1017/pan.2017.44>
- El-Beltagy, S. R., & Rafea, A. (2009). KP-Miner: A keyphrase extraction system for English and Arabic documents. *Information Systems*, 34(1), 132–144. <https://doi.org/10.1016/j.is.2008.05.002>
- El-Beltagy, S. R., & Rafea, A. (2010). *KP-Miner: Participation in SemEval-2*.

- Firdausillah, F., & Udayanti, E. D. (2021). Keyphrase Extraction on Covid-19 Tweets Based on Doc2Vec and YAKE. *Journal of Applied Intelligent System*, 6(1), 23–31. <https://doi.org/10.33633/jais.v6i1.4454>
- Hakim, Z., & Rizky, R. (2018). *Analisis Perancangan Sistem Informasi Pembuatan Paspur Di Kantor Imigrasi Bumi Serpong Damai Tangerang Banten Menggunakan Metode Rational Unified Process*. 6(2).
- Kurniawan, A. (2021). *Aplikasi Sistem Ekstraksi Kata Kunci Berbahasa Indonesia Menggunakan Algoritma Textrank Studi Kasus Data Wikipedia Indonesia*. <https://repository.uinjkt.ac.id/dspace/handle/123456789/56134>
- Liu, Z., Li, P., Zheng, Y., & Sun, M. (2009). Clustering to find exemplar terms for keyphrase extraction. *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing Volume 1 - EMNLP '09*, 1, 257. <https://doi.org/10.3115/1699510.1699544>
- Lubis, A. H., Putera Utama Siahaan, A., Aryza, S., Hariyanto, E., Rusiadi, R., Ikhwan, A., & Len Eh Kan, P. (2018). Combination of levenshtein distance and rabin-karp to improve the accuracy of document equivalence level. *International Journal of Engineering & Technology*, 7(2.27), 17. <https://doi.org/10.14419/ijet.v7i2.27.12084>
- Mothe, J., Ramiandrisoa, F., & Rasolomanana, M. (2018). Automatic keyphrase extraction using graph-based methods. *Proceedings of the 33rd Annual ACM Symposium on Applied Computing*, 728–730. <https://doi.org/10.1145/3167132.3167392>

- Nasar, Z., Jaffry, S. W., & Malik, M. K. (2019). Textual keyword extraction and summarization: State-of-the-art. *Information Processing & Management*, 56(6), 102088. <https://doi.org/10.1016/j.ipm.2019.102088>
- Paukkeri, M.-S., & Honkela, T. (2010). *Likey: Unsupervised Language-Independent Keyphrase Extraction*.
- Piskorski, J., Stefanovitch, N., Jacquet, G., & Podavini, A. (2021). *Exploring Linguistically-Lightweight Keyword Extraction Techniques for Indexing News Articles in a Multilingual Set-up*.
- Satia, B. G., Noertjahyana, A., & Susilo, R. Y. (2012). *Ekstraksi Kata Kunci Otomatis untuk Dokumen Berbahasa Indonesia menggunakan metode Genitor-plus-Extractor (GenEx)*.
- Suciadi, J. (2001). *Studi Analisis Metode-Metode Parsing dan Interpretasi Semantik Pada Natural Language Processing*. 2(1). <https://doi.org/10.9744/informatika.2.1.pp.%2013-22>
- Ying, Y., Qingping, T., Qinzheng, X., Ping, Z., & Panpan, L. (2017). A Graph-based Approach of Automatic Keyphrase Extraction. *Procedia Computer Science*, 107, 248–255. <https://doi.org/10.1016/j.procs.2017.03.087>
- Yunmar, R. A., Setiawan, A., & Tantriawan, H. (2020). The Combination of YAKE and Language Processing for Unsupervised Term Extraction Ontology Learning. *IOP Conference Series: Earth and Environmental Science*, 537(1), 012023. <https://doi.org/10.1088/1755-1315/537/1/012023>