

**EKSTRAKSI KATA KUNCI PADA ARTIKEL BERBAHASA  
INDONESIA MENGGUNAKAN METODE *RAKE***

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



Oleh :

Adys Syakhira Maharani

NIM : 09021382126153

**Jurusan Teknik Informatika  
FAKULTAS ILMU KOMPUTER UNIVERSITAS SRIWIJAYA  
2024**

# LEMBAR PENGESAHAN SKRIPSI

## EKSTRAKSI KATA KUNCI PADA ARTIKEL BERBAHASA INDONESIA MENGGUNAKAN METODE *RAKE*

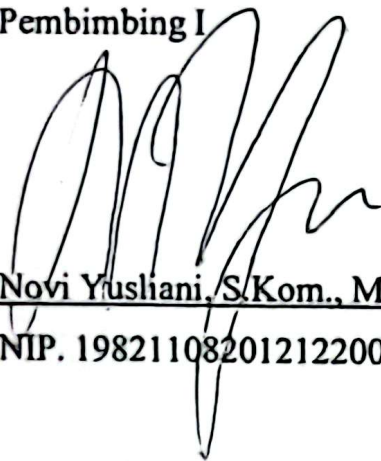
Oleh:

Adys Syakhira Maharani

NIM: 09021382126153

Palembang, 27 Desember 2024

Pembimbing I



Novi Yusliani, S.Kom., M.T.

NIP. 198211082012122001

Pembimbing II



Muhammad Naufal Rachmatullah, M.T

NIP. 199212012022031008

Mengetahui,

Ketua Jurusan Teknik Informatika



Hadipurnawan Satria, S.Kom., M.c., Ph.D.

NIP. 198004182020121001

## TANDA LULUS UJIAN KOMPREHENSIF

Pada hari Senin tanggal 23 Desember 2024 telah dilaksanakan ujian komprehensif skripsi oleh Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya

Nama : Adys Syakhira Maharani

NIM : 09021382126153

Judul : Ekstraksi Kata Kunci pada Artikel Berbahasa Indonesia Menggunakan Metode RAKE

Dan dinyatakan LULUS.

1. Ketua Penguji

Hadipurnawan Satria, S.Kom., M.Sc., Ph.D  
NIP. 198004182020121001



2. Penguji

Dian Palupi Rini, M.Kom., Ph.D  
NIP. 197802232006042002



3. Pembimbing I

Novi Yusliani, S.Kom., M.T.  
NIP. 198211082012122001



4. Pembimbing II

Muhammad Naufal Rachmatullah, M.T.  
NIP. 199212012022031008



Mengetahui,  
Ketua Jurusan Teknik Informatika



Hadipurnawan Satria, S.Kom., M.Sc., Ph.D.  
NIP. 198004182020121001

## HALAMAN PERNYATAAN

Yang bertanda tangan di bawah ini :

Nama : Adys Syakhira Maharani

NIM : 09021382126153

Program Studi : Teknik Informatika

Judul Skripsi : Ekstraksi Kata Kunci pada Artikel Berbahasa Indonesia

Menggunakan Metode RAKE

Hasil Pengecekan Software Turnitin : 15%

Menyatakan bahwa laporan tugas akhir 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, 25 November 2024



*Adys*  
Adys Syakhira Maharani

NIM. 09021382126153

## MOTTO DAN PERSEMBAHAN

“Allah tidak mengajarkan hidup ini mudah. Tetapi Allah berjanji, bahwa  
sesungguhnya bersama kesulitan ada kemudahan”

(Q.S. Al-Insyirah 5-6)

*“It will pass, everything you’ve gone through it will pass”*

(Rachel Venny)

Skripsi ini kupersembahkan kepada:

- Allah SWT
- Orang Tua dan Keluarga
- Dosen Pembimbing
- Universitas Sriwijaya
- Teman-teman

## **ABSTRACT**

*The increasing number of scientific articles presents a challenge for researchers in quickly accessing relevant information. One of the main challenges is determining efficient keywords. To address this, an automatic keyword extraction system becomes an essential solution, aimed at developing and evaluating methods for keyword extraction to accelerate the search and management of information from scientific articles. This study will perform keyword extraction using a statistical approach through the Rapid Automatic Keyword Extraction (RAKE) method. The results show that the best configuration without cosine similarity in post-processing produced an f-score of 0.0527 (title and abstract) and 0.0066 (entire article), while with cosine similarity, it resulted in an f-score of 0.1312 (title and abstract) and 0.223 (entire article). Although the RAKE method was successfully applied, the results were not optimal.*

**Keywords:** *Keyword Extraction, Scientific Articles, Rapid Automatic Keyword Extraction.*

## ABSTRAK

Peningkatan jumlah artikel ilmiah menimbulkan tantangan bagi peneliti dalam mengakses informasi relevan secara cepat. Salah satu tantangan utama adalah penentuan kata kunci yang efisien. Untuk mengatasinya, sistem ekstraksi kata kunci otomatis menjadi solusi penting, yang bertujuan mengembangkan dan mengevaluasi metode ekstraksi kata kunci untuk mempercepat pencarian dan pengelolaan informasi dari artikel ilmiah.. Penelitian ini akan melakukan ekstraksi kata kunci menggunakan pendekatan statistik melalui metode Rapid Automatic Keyword Extraction (RAKE). Hasil penelitian menunjukkan bahwa konfigurasi terbaik tanpa cosine similarity pada post-processing menghasilkan f-score 0.0527 (judul dan abstrak) dan 0.0066 (seluruh artikel), sementara dengan cosine similarity menghasilkan f-score 0.1312 (judul dan abstrak) dan 0.223 (seluruh artikel). Meskipun metode RAKE berhasil diterapkan, hasilnya belum optimal.

**Kata Kunci :** Ekstraksi Kata Kunci, Artikel Ilmiah, *Rapid Automatic Keyword Extraction*.

## **KATA PENGANTAR**

Puji syukur kehadiran Allah SWT atas berkat dan rahmat-Nya yang diberikan kepada penulis sehingga dapat menyelesaikan tugas akhir ini dengan baik. Tugas akhir ini disusun untuk memenuhi salah satu syarat guna menyelesaikan pendidikan program Strata-1 pada Fakultas Ilmu Komputer Program Studi Teknik Informatika di Universitas Sriwijaya. Dalam menyelesaikan Tugas Akhir ini banyak pihak yang telah memberikan bantuan, bimbingan, dan dukungan baik secara langsung maupun secara tidak langsung. Oleh karena itu, penulis juga menyampaikan banyak rasa terima kasih kepada:

1. Allah SWT atas berkat, rahmat, dan karunia-Nya, sehingga penulis dapat menyelesaikan tugas akhir.
2. Papa Nasrun Efendi, SE. Ak, (Almh) Mama dr. Nur Akbari, Mama Debby Yulianti, Kakak Karina Auliya Maharani, S.T, Adek Dean, dan Rayyanza selaku keluarga penulis yang senantiasa memberikan motivasi, dukungan, dan doanya.
3. Bapak Prof. Dr. Erwin, S.Si., M.Si. selaku Dekan Fakultas Ilmu Komputer.
4. Bapak Hadipurnawan Satria, S.Kom., M.Sc., Ph.D. selaku Ketua Jurusan Teknik Informatika.
5. Ibu Novi Yusliani, S.Kom., M.T. dan Bapak Muhammad Naufal Rachmatullah, M.T selaku Dosen Pembimbing Skripsi di Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya.
6. Bapak Danny Matthew Saputra, M.Sc selaku Dosen Pembimbing Akademik di Jurusan Teknik Informatika Fakultas Ilmu Komputer



Universitas Sriwijaya.

7. Seluruh dosen program studi serta staf Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya.
8. Kepada Muhammad Rusnan Arkan yang telah kebersamai penulis pada hari-hari yang tidak mudah selama proses penyusunan skripsi. Terima kasih atas dukungan dan semangat serta tempat untuk berkeluh kesah.
9. Kepada Andini dan Nabilla selaku teman seangkatan yang telah menemani, membantu, dan memberi dukungan penulis selama mengerjakan skripsi.
10. Pihak-pihak lain yang telah memotivasi dan memberikan dukungan namun tidak dapat disebutkan satu-persatu.

Penulis menyadari dalam penyusunan Tugas Akhir ini masih terdapat banyak kekurangan, Oleh karena itu, kritik dan saran yang membangun sangat diharapkan untuk kemajuan penelitian selanjutnya. Akhir kata semoga Tugas Akhir ini dapat berguna dan bermanfaat bagi kita semua.

Palembang, 22 November 2024

Adys Syakhira Maharani

## DAFTAR ISI

LEMBAR PENGESAHAN SKRIPSI .....	i
ABSTRACT.....	ii
ABSTRAK.....	iii
KATA PENGANTAR.....	iv
DAFTAR ISI.....	v
DAFTAR GAMBAR .....	vi
DAFTAR TABEL.....	vii
DAFTAR LAMPIRAN.....	viii
BAB I PENDAHULUAN .....	I-1
I.1    Pendahuluan.....	I-1
1.2    Latar Belakang .....	I-1
1.3    Rumusan Masalah .....	I-4
1.4    Tujuan Penelitian .....	I-4
1.5    Manfaat Penelitian.....	I-5
1.6    Batasan Penelitian .....	I-5
1.7    Sistematika Penulisan.....	I-5
1.8    Kesimpulan .....	I-7
BAB II KAJIAN LITERATUR .....	II-1
2.1    Pendahuluan .....	II-1
2.2    Landasan Teori .....	II-1
2.2.1    Ekstraksi Kata Kunci .....	II-1
2.2.2    RAKE .....	II-2
2.2.3    Pra-Pengolahan Teks.....	II-3
2.2.4 <i>Noun phrase chunking</i> .....	II-6
2.2.5 <i>Co-occurrence Matrix</i> .....	II-6

2.2.6	<i>Scoring Keyword</i> .....	II-6
2.2.7	<i>Confusion Matrix</i> .....	II-8
2.2.8	<i>Rational Unified Process</i> .....	II-10
2.3	Penelitian Lain yang Relevan.....	II-12
2.4	Kesimpulan .....	II-13
<b>BAB III METODOLOGI PENELITIAN</b> .....		<b>III-1</b>
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	Menentukan Kerangka Kerja Penelitian .....	III-3
3.3.2	Menentukan Kriteria Pengujian .....	III-21
3.3.3	Menentukan Format Data Pengujian .....	III-21
3.3.4	Menetapkan Alat Bantu Penelitian .....	III-22
3.3.5	Melakukan Pengujian Penelitian .....	III-23
3.3.6	Melakukan Analisis dan Menarik Kesimpulan Penelitian .....	III-24
3.4	Metode Pengembangan Perangkat Lunak.....	III-24
3.4.1.	Fase Insepsi.....	III-24
3.4.2.	Fase Elaborasi.....	III-25
3.4.3.	Fase Konstruksi .....	III-25
3.4.4.	Fase Transisi .....	III-26
3.5	Kesimpulan .....	III-26
<b>BAB IV PENGEMBANGAN PERANGKAT LUNAK</b> .....		<b>IV-1</b>
4.1	Pendahuluan .....	IV-1
4.2	Fase Insepsi .....	IV-1
4.2.1	Pemodelan Bisnis .....	IV-1
4.2.2	Kebutuhan Sistem .....	IV-2
4.3	Fase Elaborasi .....	IV-24
4.3.1	Pemodelan Bisnis .....	IV-24
4.3.2	Kebutuhan Sistem .....	IV-29
4.3.3	Analisis dan Perancangan .....	IV-xxix
4.4	Fase Konstruksi.....	IV-32
4.4.1	Kebutuhan Sistem .....	IV-33

4.4.2	Implementasi .....	IV-34
4.4.3	Fase Transisi.....	IV-39
4.4.4	Pemodelan Bisnis .....	IV-39
4.4.5	Kebutuhan Sistem .....	IV-39
4.4.6	Rencana Pengujian .....	IV-39
4.4.7	Implementasi .....	IV-40
4.5	Kesimpulan .....	IV-41
BAB V HASIL DAN ANALISIS PENELITIAN.....		V-1
5.1	Pendahuluan .....	V-1
5.2	Hasil Penelitian .....	V-1
5.2.1	Skenario Pengujian.....	V-1
5.2.2	Data Hasil Pengujian Tanpa Menggunakan <i>Cosine Similarity</i> .....	V-2
5.2.3	Hasil Pengujian Data Uji Menggunakan <i>Cosine Similarity</i> .....	V-5
5.3	Analisis Hasil Penelitian .....	V-8
5.4	Kesimpulan.....	V-11
BAB VI KESIMPULAN DAN SARAN.....		VI-1
6.1	Pendahuluan .....	VI-1
6.2	Kesimpulan .....	VI-1
6.3	Saran.....	VI-1
DAFTAR PUSTAKA.....		x

## DAFTAR GAMBAR

<b>Gambar II- 1.</b> Arsitektur Sistem RAKE .....	II-2
<b>Gambar II- 2.</b> Tahapan Pra-Pengolahan Teks.....	II-4
<b>Gambar II- 3.</b> Tabel Confusion Matrix.....	II-8
<b>Gambar II- 4.</b> Arsitektur RUP .....	II-11
<b>Gambar III- 1.</b> Tahapan Penelitian.....	III-3
<b>Gambar III- 2.</b> Kerangka Kerja Penelitian .....	III-4
<b>Gambar IV- 1.</b> Diagram Use Case.....	IV-21
<b>Gambar IV- 2.</b> Rancangan Interface Halaman Ekstraksi Kata Kunci .....	IV-26
<b>Gambar IV- 3.</b> Rancangan Interface Halaman Hasil Ekstraksi Kata Kunci .....	IV-27
<b>Gambar IV- 4.</b> Rancangan Interface Halaman Hasil Evaluasi Kata Kunci .....	IV-28
<b>Gambar IV- 5.</b> Diagram Aktivitas Ekstraksi Kata Kunci.....	IV-30
<b>Gambar IV- 6.</b> Diagram Aktivitas Evaluasi Kata Kunci.....	IV-31
<b>Gambar IV- 7.</b> Sequence Diagram Ekstraksi Kata Kunci .....	IV-32
<b>Gambar IV- 8.</b> Sequence Diagram Evaluasi Kata Kunci .....	IV-32
<b>Gambar IV- 9.</b> Class Diagram .....	IV-34
<b>Gambar IV- 10.</b> Implementasi Antarmuka Halaman Ekstraksi.....	IV-37
<b>Gambar IV- 11.</b> Implementasi Antarmuka Halaman Setelah Ekstraksi .....	IV-38
<b>Gambar IV- 12.</b> Implementasi Antarmuka Halaman Evaluasi.....	IV-38
<b>Gambar V- 1.</b> Grafik Rata Rata Evaluasi Tanpa Menggunakan Cosine Similarity .....	V-4
<b>Gambar V- 2.</b> Grafik Rata Rata Evaluasi Menggunakan Cosine Similarity .....	V-8
<b>Gambar V- 3.</b> Grafik Perbandingan Kedua Konfigurasi Pengujian .....	V-9

## DAFTAR TABEL

<b>Tabel III- 1.</b> Contoh Dataset yang digunakan .....	III-2
<b>Tabel III- 2.</b> Case Folding.....	III-5
<b>Tabel III- 3.</b> Pembersihan Teks.....	III-6
<b>Tabel III- 4.</b> Tokenisasi.....	III-8
<b>Tabel III- 5.</b> Stopword Removal .....	III-9
<b>Tabel III- 6.</b> POS Tagging .....	III-11
<b>Tabel III- 7.</b> Noun Phrase Chunking.....	III-13
<b>Tabel III- 8.</b> Index Kata .....	III-15
<b>Tabel III- 9.</b> Co-occurrence Matrix .....	III-17
<b>Tabel III- 10.</b> Scoring Keyword.....	III-18
<b>Tabel III- 11.</b> Scoring Keyphrase.....	III-19
<b>Tabel III- 12.</b> Pemeringkatan kandidat keyphrase.....	III-20
<b>Tabel III- 13.</b> Hasil kinerja seluruh dataset .....	III-22
<b>Tabel IV- 1.</b> Kebutuhan Fungsional.....	IV-3
<b>Tabel IV- 2.</b> Kebutuhan Non-Fungsional .....	IV-3
<b>Tabel IV- 3.</b> Contoh Data Judul .....	IV-5
<b>Tabel IV- 4.</b> Contoh Data Abstrak .....	IV-5
<b>Tabel IV- 5.</b> Hasil Penggabungan Teks Judul dan Teks Abstrak.....	IV-6
<b>Tabel IV- 6.</b> Hasil Case Folding.....	IV-7
<b>Tabel IV- 7.</b> Pembersihan Teks.....	IV-8
<b>Tabel IV- 8.</b> Tokenisasi.....	IV-9
<b>Tabel IV- 9.</b> Stopword Removal .....	IV-10
<b>Tabel IV- 10.</b> POS Tagging.....	IV-11
<b>Tabel IV- 11.</b> Noun Phrase Chunking .....	IV-12
<b>Tabel IV- 12.</b> Kandidat Keyphrase.....	IV-xiii
<b>Tabel IV- 13.</b> Index Kata.....	IV-14
<b>Tabel IV- 14.</b> Co-occurrence Matrix .....	IV-14
<b>Tabel IV- 15.</b> Scoring Keyword .....	IV-17
<b>Tabel IV- 16.</b> Scoring Keyphrase.....	IV-18
<b>Tabel IV- 17.</b> Pemeringkatan kandidat keyphrase .....	IV-19
<b>Tabel IV- 18.</b> Tabel Golden dan Selected Keyphrase.....	IV-20
<b>Tabel IV- 19.</b> Evaluasi Hasil .....	IV-20

<b>Tabel IV- 20.</b> Tabel Definisi Aktor .....	IV-21
<b>Tabel IV- 21.</b> Tabel Definisi Use Case.....	IV-xxii
<b>Tabel IV- 22.</b> Skenario Use Case Mengekstrak Kata Kunci.....	IV-xxii
<b>Tabel IV- 23.</b> Skenario Use Case Mengevaluasi Kata Kunci .....	IV-23
<b>Tabel IV- 24.</b> Rancangan Data .....	IV-xxv
<b>Tabel IV- 25.</b> Implementasi Kelas.....	IV-35
<b>Tabel IV- 26.</b> Rencana Pengujian Use Case Ekstraksi .....	IV-39
<b>Tabel IV- 27.</b> Rencana Pengujian Use Case Evaluasi .....	IV-40
<b>Tabel IV- 28.</b> Implementasi Pengujian Use Case Mengekstrak Kata Kunci.....	IV-40
<b>Tabel IV- 29.</b> Implementasi Pengujian Use Case Mengevaluasi Kata Kunci .....	IV-41
<b>Tabel V- 1.</b> Sampel Hasil Evaluasi RAKE dengan Judul dan Abstrak.....	V-2
<b>Tabel V- 2.</b> Sampel Hasil Evaluasi RAKE dengan Keseluruhan Isi Artikel.....	V-3
<b>Tabel V- 3.</b> Sampel Hasil Evaluasi Cosine Similarity dengan Judul dan Abstrak .....	V-5
<b>Tabel V- 4.</b> Sampel Hasil Evaluasi Cosine Similarity dengan Keseluruhan Isi Artikel .	V-6

## DAFTAR LAMPIRAN

Lampiran 1.....	xiv
Lampiran 2.....	xvii
Lampiran 3.....	xx
Lampiran 4.....	xxiv
Lampiran 5.....	xxviii



# **BAB I**

## **PENDAHULUAN**

### **I.1 Pendahuluan**

Pada bab ini akan membahas latar belakang masalah, rumusan masalah, tujuan penelitian, manfaat penelitian, batasan penelitian, dan sistematika penulisan. Secara keseluruhan, skripsi ini menjelaskan mengenai bagaimana mengembangkan sebuah sistem ekstraksi kata kunci dengan menggunakan metode *Rapid Automatic Keyword Extraction* (RAKE).

### **1.2 Latar Belakang**

Dalam beberapa tahun terakhir, peningkatan volume artikel ilmiah yang signifikan menimbulkan tantangan besar bagi para peneliti dalam mengakses dan mengelola informasi yang relevan dengan cepat. Salah satu tantangan besarnya adalah penentuan kata kunci. Penentuan kata kunci untuk ratusan bahkan ribuan artikel secara manual akan memakan waktu yang cukup lama dan seringkali tidak efisien, mengingat volume data dalam dokumen yang terbilang besar. Untuk mengatasi permasalahan tersebut, diperlukan sistem ekstraksi kata kunci otomatis. *Keyphrase Extraction* dapat membantu peneliti dalam proses pengelolaan informasi dan navigasi literatur yang kompleks menjadi lebih efektif dan konsisten.

Kata kunci atau *keyword* merupakan satu kata atau frasa yang menonjol (*significant*) pada judul, tajuk subjek, isi, abstrak atau teks entri dalam katalog online dan basis data bibliografi, yang dapat digunakan sebagai istilah pencarian dalam pencarian bebas untuk menemukan seluruh cantuman yang memuat kata kunci tersebut (Reitz, 2012). Dalam pengimplementasiannya, kata kunci seringkali digunakan untuk merepresentasikan informasi pada sistem *search engine*, sedangkan pada artikel publikasi ilmiah, kata kunci seringkali digunakan untuk merepresentasikan isi dari dokumen yang bertujuan untuk mempermudah pembaca dalam mendapatkan bayangan dari dokumen tersebut.

Ekstraksi kata kunci atau *keyword extraction* biasa juga orang menyebutnya *keyphrase extraction* diartikan sebagai sebuah tahapan untuk dapat mengidentifikasi berbagai kumpulan teks pada suatu dokumen dan menemukan kata kunci yang tepat sesuai dengan topik pembahasan dari dokumen yang telah diproses (Li & Wang, 2014). Secara umum, ekstraksi kata kunci merupakan sebuah sistem yang mengacu pada proses otomatis mengambil kata atau frasa yang paling penting dan representatif dari sekumpulan dokumen teks.

Terdapat dua pendekatan yang digunakan dalam ekstraksi kata kunci antara lain *supervised learning* dan *un-supervised learning*. Penelitian ini menggunakan pendekatan *un-supervised learning*, dikarenakan pendekatan secara *supervised* memerlukan data yang berlabel, sedangkan pendekatan *un-*

*supervised learning* lebih fleksibel hanya dengan mengambil informasi dari dokumen itu sendiri (Zhang et al., 2023). Salah satu pendekatan yang dapat diterapkan dalam *un-supervised learning* adalah pendekatan statistik. Penelitian ini memanfaatkan model *statistical* karena dilakukan langsung berdasarkan fitur frekuensi kemunculan kata. Pendekatan ini memungkinkan pengukuran tingkat pentingnya setiap frasa kandidat, sehingga menjadi landasan bagi berbagai model lanjutan (Xie et al., 2023).

Salah satu metode yang bisa digunakan untuk melakukan ekstraksi kata kunci dengan pendekatan statistik adalah *Rapid Automatic Keyword Extraction* (RAKE). Metode ini memanfaatkan pola statistik alami dalam teks tanpa memerlukan data berlabel atau proses pelatihan, sehingga mampu diaplikasikan pada dokumen baru. Dengan pendekatan ini, RAKE menerapkan prinsip *statistical base models* dalam proses heuristiknya untuk menghasilkan kata kunci secara otomatis dan efisien.

Kelebihan dari metode RAKE yaitu tidak bergantung pada domain dan tidak mempertimbangkan bahasa untuk mengekstrak kata kunci dari tiap dokumen (Rinartha & Kartika, 2021). Berdasarkan hasil penelitian terbukti bahwa ekstraksi kata kunci menggunakan metode RAKE mencapai hasil skor kandidat frasa kata kunci yang lebih baik dibandingkan dengan metode TextRank (Baruni & Sathiaseelan, 2020). RAKE mencapai *precision* yang lebih tinggi dan *recall* yang serupa dengan TextRank karena kinerja RAKE

dalam menilai kata kunci dalam sekali jalan, sedangkan TextRank memerlukan iterasi berulang untuk mencapai konvergensi pada peringkat kata. Selain itu, RAKE secara efektif menggunakan serangkaian parameter masukan yang sederhana dan secara otomatis mengekstrak kata kunci dalam satu proses yang mengungguli dalam hal *precision*, *eficiency*, dan *simplicity*, sehingga cocok untuk berbagai jenis dokumen teks (Stuart et al., 2010).

Oleh karena itu, penelitian ini akan mengembangkan ekstraksi kata kunci pada artikel ilmiah berbahasa Indonesia menggunakan metode RAKE. Metode ini akan diterapkan untuk mengidentifikasi kata kunci yang paling relevan dan penting dalam teks akademik, guna mempermudah pencarian informasi dan analisis konten dalam konteks penelitian ilmiah di Indonesia.

### **1.3 Rumusan Masalah**

Rumusan masalah pada penelitian ini adalah

1. Bagaimana mengembangkan sistem Ekstraksi Kata Kunci menggunakan metode RAKE untuk artikel ilmiah berbahasa Indonesia?
2. Bagaimana kinerja metode RAKE dalam melakukan Ekstraksi Kata Kunci berdasarkan nilai *F-Score*?

### **1.4 Tujuan Penelitian**

Tujuan dari penelitian ini adalah

1. Menghasilkan sebuah sistem yang dapat melakukan Ekstraksi Kata Kunci pada Artikel berbahasa Indonesia dengan menggunakan metode RAKE.
2. Mengetahui kinerja metode RAKE dalam mengekstraksi kata kunci berdasarkan *F1-Score*.

### **1.5 Manfaat Penelitian**

Manfaat yang didapat dari penelitian ini adalah

1. Sistem dapat menghasilkan *keyword* secara otomatis pada artikel ilmiah berbahasa Indonesia.
2. Diharapkan dapat dijadikan referensi untuk penelitian terkait.

### **1.6 Batasan Penelitian**

1. Data Publikasi Ilmiah terbatas pada topik Ilmu Komputer.
2. Data yang digunakan adalah 100 data Artikel Ilmiah berbahasa Indonesia dari *website* *jtiik*<sup>1</sup>, *jatisi*<sup>2</sup>, *jepin*<sup>3</sup> berupa judul, abstrak, dan kata kunci.

### **1.7 Sistematika Penulisan**

Sistematika penulisan tugas akhir mengikuti standar penulisan tugas

---

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

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

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

akhir Fakultas Ilmu Komputer Universitas Sriwijaya yaitu sebagai berikut :

## **BAB I. PENDAHULUAN**

Pada bab ini membahas latar belakang, rumusan masalah, tujuan dan manfaat penelitian, batasan masalah serta sistematika penelitian yang akan dijadikan sebagai pokok pikiran penelitian ini.

## **BAB II. KAJIAN LITERATUR**

Pada bab ini membahas landasan teori yang digunakan dalam penelitian, seperti definisi *Keyphrase Extraction* dan metode RAKE, serta beberapa literatur yang relevan dengan penelitian ini.

## **BAB III. METODOLOGI PENELITIAN**

Pada bab ini membahas proses yang akan dilaksanakan selama penelitian, Seperti pengumpulan data, analisis data dan perancangan pengembangan perangkat lunak. Setiap tahapan proses akan dijelaskan berdasarkan kerangka kerja yang sudah ditentukan.

## **BAB IV. PENGEMBANGAN PERANGKAT LUNAK**

Pada bab ini akan membahas mengenai pengembangan perangkat lunak dimulai dari analisis kebutuhan perangkat lunak hingga pengujian pada

perangkat lunak guna mengevaluasi hasil pengembangannya.

## **BAB V. HASIL DAN ANALISIS PENELITIAN**

Pada bab ini menyajikan hasil penelitian yang disusun berdasarkan langkah dan metode yang telah ditentukan sebelumnya, yang kemudian dianalisis untuk menjadi dasar dalam kesimpulan penelitian.

## **BAB VI. KESIMPULAN DAN SARAN**

Pada bab ini menyajikan hasil penelitian yang disusun sesuai dengan langkah dan metode yang telah ditetapkan sebelumnya, yang kemudian dianalisis sebagai dasar untuk kesimpulan penelitian.

### **1.8 Kesimpulan**

Pada bab ini telah dijelaskan latar belakang, rumusan masalah, tujuan dan manfaat penelitian, batasan penelitian serta sistematika penelitian yang akan dijadikan sebagai pokok pikiran peneliti.

## DAFTAR PUSTAKA

- Figueroa, G., & Chen, Y.-S. (2014). *Collaborative Ranking between Supervised and Unsupervised Approaches for Keyphrase Extraction*.
- Hasan, K. S., & Ng, V. (2014). *Automatic Keyphrase Extraction: A Survey of the State of the Art. Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 1262–1273.  
<https://doi.org/10.3115/v1/P14-1119>
- Sari, D. P., & Purwarianti, A. (2014). Ekstraksi Kata Kunci Otomatis Untuk Dokumen Bahasa Indonesia Studi Kasus: Artikel Jurnal Ilmiah Koleksi Pdi Lipi. *BACA: Jurnal Dokumentasi dan Informasi*, 35(2), 139-147.
- Verma, Y., Jangra, A., Saha, S., Jatowt, A., & Roy, D. (2022). MAKED: Multi-lingual Automatic Keyword Extraction Dataset. *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, 6170–6179.  
<https://aclanthology.org/2022.lrec-1.664>
- Kaur, K., & Gupta, V. (2011). *Keyword Extraction For Punjabi Language. Indian Journal of Computer Science and Engineering*, 2.
- 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.

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.

Wicaksono, A. F., & Purwarianti, A. (2010, August). *HMM based part-of-speech tagger for Bahasa Indonesia*. In *Fourth International MALINDO Workshop*, Jakarta

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)

Plakasa, Gerald. (2022). *Ekstraksi Kata Kunci Pada Bahasa Indonesia Menggunakan Metode Yake*

Nabilah, Aini. (2023). **KEYPHRASE EXTRACTION DENGAN MENGGUNAKAN PRE-TRAINED LANGUAGE MODELS BERT DAN TOPIC-GUIDED GRAPH ATTENTION NETWORKS**

Firdausillah, Fahri, and Erika Devi Udayanti. (2021). "*Keyphrase Extraction on Covid-19 Tweets Based on Doc2Vec and YAKE.*" *Journal of Applied Intelligent System* 6.1 (23-31).

Kurniawan, A. (2021). *Aplikasi Sistem Ekstraksi Kata Kunci Berbahasa Indonesia Menggunakan Algoritma TextRank Studi Kasus Data Wikipedia Indonesia.*

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).