

EKSTRAKSI KATA KUNCI PADA BAHASA INDONESIA  
MENGUNAKAN *JOINTLY MODELING LOCAL AND GLOBAL*  
*CONTEXT* DAN BERT

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



Oleh:

Muhammad Alfaris Oktavian

NIM: 09021282025109

**Jurusan Teknik Informatika**

**FAKULTAS ILMU KOMPUTER UNIVERSITAS SRIWIJAYA**

**2024**

LEMBAR PENGESAHAN PROPOSAL SKRIPSI

EKSTRAKSI KATA KUNCI PADA BAHASA INDONESIA  
MENGUNAKAN *JOINTLY MODELING LOCAL AND GLOBAL*  
*CONTEXT* DAN BERT

Oleh :

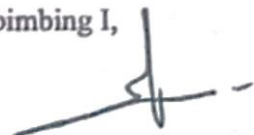
Muhammad Alfaris Oktavian

NIM : 09021282025109

Palembang, 28 Maret 2024



Pembimbing I,

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

## TANDA LULUS UJIAN KOMPREHENSIF

Pada hari Jum'at Tanggal 5 April 2024 Telah dilaksanakan ujian komprehensif skripsi oleh Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya.

Nama : Muhammad Alfaris Oktavian

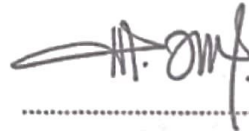
NIM : 09021282025109

JUDUL : Ekstraksi Kata Kunci Pada Bahasa Indonesia Menggunakan *Jointly Modeling Local And Global Context* dan BERT

Dan dinyatakan LULUS.

1. Ketua Penguji

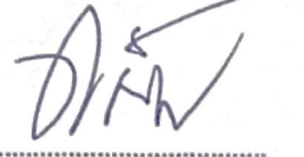
Annisa Darmawahyuni, M.Kom.  
NIP. 199006302023212044



---

2. Penguji

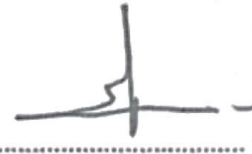
Samsuryadi, M.Kom., Ph.D.  
NIP. 197102041997021003



---

3. Pembimbing I

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 : Muhammad Alfaris Oktavian

NIM : 09021282025109

Program Studi : Teknik Informatika

Judul Skripsi : Ekstraksi Kata Kunci Pada Bahasa Indonesia Menggunakan  
*Jointly Modeling Local And Global Context* dan BERT

Hasil pengecekan Software Turnitin : 17%

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, 19 April 2024

Peny



Muhammad Alfaris Oktavian  
NIM. 09021282025109

## **MOTTO DAN PERSEMBAHAN**

Motto:

“Dan orang-orang yang bersungguh-sungguh untuk (mencari keridhaan) di jalan Kami, pasti akan Kami tunjukkan kepada mereka jalan-jalan Kami. Dan sungguh, Allah beserta orang-orang yang berbuat baik.”

- Al Ankabut : 69

Kupersembahkan Karya Tulis ini Kepada:

- Allah SWT
- Orang Tua
- Keluarga Besar
- Fakultas Ilmu Komputer
- Universitas Sriwijaya

## **ABSTRACT**

*Finding the right scientific journal from the vast array of journals available today is a significant challenge for researchers, academics and even students as it takes a considerable amount of time. Therefore, the use of keyphrases or keywords that reflect the content of the journal can be an effective solution to overcome this problem. The keyphrase extraction process is carried out to identify relevant keyphrases. One of the approaches used is Jointly Modeling Local and Global Context combined with a Pre-trained Language Model, namely Indobert, to obtain a language vector representation. This research aims to develop a keyphrase extraction system using Indobert's Pre-trained Language Model and Jointly Modeling Local and Global Context. In the test, 100 data sets consisting of titles and abstracts of scientific journals as well as experiments of 8 Pre-trained Language Model Indobert. The test results show that the best configuration is keyphrase extraction on the Indobert-Base-P1 model with the selection of the top 15 keyphrases, with an average f1-score of 0.226.*

*Keywords: Indobert, JMLGC, Jointly Modeling Local and Global Context, Keyphrase Extraction, Pre-trained Language Model, Scientific Journal*

## ABSTRAK

Menemukan jurnal ilmiah yang tepat dari beragam jurnal yang tersedia saat ini merupakan tantangan yang signifikan bagi para peneliti, akademisi bahkan mahasiswa karena membutuhkan waktu yang cukup lama. Oleh karena itu, penggunaan *keyphrase* atau kata kunci yang mencerminkan isi jurnal dapat menjadi solusi yang efektif untuk mengatasi masalah ini. Proses ekstraksi *keyphrase* dilakukan untuk mengidentifikasi *keyphrase* yang relevan. Salah satu pendekatan yang digunakan adalah *Jointly Modeling Local and Global Context* yang dipadukan dengan *Pre-trained Language Model*, yaitu Indobert, untuk memperoleh representasi vektor bahasa. Penelitian ini bertujuan untuk mengembangkan perangkat lunak ekstraksi *keyphrase* menggunakan *Pre-trained Language Model* Indobert dan *Jointly Modeling Local and Global Context*. Dalam pengujian, 100 set data yang terdiri dari judul dan abstrak jurnal ilmiah serta percobaan dari 8 *Pre-trained Language Model* Indobert. Hasil pengujian menunjukkan bahwa konfigurasi terbaik adalah ekstraksi *keyphrase* pada model Indobert-Base-P1 dengan pemilihan 15 *keyphrase* teratas, dengan nilai rata-rata *f1-score* sebesar 0,226.

Kata Kunci: Indobert, Jurnal Ilmiah, JMLGC, *Jointly Modeling Local and Global Context*, *Keyphrase Extraction*, *Pre-trained Language Model*.

## KATA PENGANTAR

Puji syukur kepada Allah SWT atas rahmat dan nikmat Nya yang telah 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 dan keluarga besar yang telah mendoakan, memberi semangat, memotivasi, dan nasihat untuk menyelesaikan skripsi ini.
3. Bapak Dr. M. Fachrurrozi, S.Si., M.T. selaku Ketua Jurusan Teknik Informatika Universitas Sriwijaya.
4. Ibu Rizki Kurniati, M.T. selaku Dosen Pembimbing Akademik yang telah memberikan banyak sekali bantuan dan arahan kepada penulis selama perkuliahan.
5. Dr. Abdiansah, S.Kom., M.CS. selaku Dosen Pembimbing I yang telah membimbing serta memberikan arahan kepada penulis selama proses pengerjaan skripsi.
6. Seluruh dosen program studi serta admin Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya.
7. Saudara M. Farhan Ghifari, Sheva Satrian, Alif Toriq Alkausar, Fadhil Zahran Muwafa, M. Bintang Khadafi, Anwaripasha Akbar, Bayu Daru Pangestu dan Fauzan Aziman Zohro sebagai teman yang selalu memotivasi penulis untuk semangat mengerjakan skripsi.
8. Saudara/i Fasya Pavita Rayhan, M. Ghaly Alghifari, Akbar Abdul Karim, Salman Ariiq, Tazkia Defitri Hasficinsia, Hanin Hanifah Nurzam, Kania Nurhafni, Syakina Zahra Sahary Siregar, Widyadari Dhiya Parahita, dan Fauziyyah Khoirunnisa Rusyidah sebagai teman yang selalu memotivasi dan menginspirasi penulis dalam mengerjakan skripsi.



9. Seluruh Staf Administrasi dan Pegawai Fakultas Ilmu Komputer yang telah membantu dalam urusan administrasi tugas akhir penulis.
10. Seluruh teman-teman yang telah memberikan saran, motivasi, dan semangat kepada penulis.
11. Pihak-pihak lain yang tidak dapat penulis sebutkan satu per satu.

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, 11 Maret 2024

Penulis



Muhammad Alfaris Oktavian

## DAFTAR ISI

|  |       |
|--|-------|
| LEMBAR PENGESAHAN PROPOSAL SKRIPSI .....                     | ii    |
| MOTTO DAN PERSEMBAHAN .....                                  | iii   |
| ABSTRACT.....  | vi    |
| ABSTRAK.....   | vii   |
| KATA PENGANTAR .....   | viii  |
| DAFTAR ISI.....  | x     |
| DAFTAR TABEL.....  | xiii  |
| DAFTAR GAMBAR .....  | xvi   |
| BAB I PENDAHULUAN .....                                      | III-1 |
| 1.1 Pendahuluan .....  | III-1 |
| 1.2 Latar Belakang Masalah .....                             | III-1 |
| 1.3 Rumusan Masalah .....                                    | III-3 |
| 1.4 Tujuan Penulisan.....                                    | III-3 |
| 1.5 Manfaat Penelitian .....                                 | III-3 |
| 1.6 Batasan Penelitian .....                                 | III-3 |
| 1.7 Sistematika Penulisan .....                              | III-4 |
| 1.8 Kesimpulan .....   | III-5 |
| BAB II KAJIAN LITERATUR .....                                | IV-5  |
| 2.1 Pendahuluan.....   | IV-5  |
| 2.2 Landasan Teori.....                                      | IV-5  |
| 2.2.1 <i>Keyphrase Extraction</i> .....                      | IV-5  |
| 2.2.2 <i>Pre-Train Language Model BERT</i> .....             | IV-5  |
| 2.2.3 <i>Jointly Modeling Local and Global Context</i> ..... | IV-8  |
| 2.2.4 <i>POS Tagging dan Penjelasan Tagger</i> .....         | IV-13 |
| 2.2.5 <i>Confusion Matrix</i> .....                          | IV-14 |
| 2.3 Penelitian Lain yang Relevan .....                       | IV-15 |
| BAB III METODOLOGI PENELITIAN .....                          | V-16  |
| 3.1 Pendahuluan.....   | V-16  |
| 3.2 Pengumpulan Data.....                                    | V-16  |
| 3.2.1 Jenis dan Sumber Data.....                             | V-16  |
| 3.2.2 Metode Pengumpulan Data.....                           | V-16  |
| 3.3 Tahapan Penelitan.....                                   | V-16  |

|   |               |
|---|---------------|
| 3.3.1 Mengumpulkan Data.....  | V-17          |
| 3.3.2 Menentukan Kerangka Kerja Penelitian .....                                    | V-18          |
| 3.3.3 Menentukan Kriteria Pengujian .....   | V-19          |
| 3.3.4 Menentukan Format Data Pengujian.....   | V-19          |
| 3.3.5 Menentukan Alat Bantu Penelitian .....  | V-20          |
| 3.3.6 Melakukan Pengujian Penelitian.....   | V-21          |
| 3.3.7 Melakukan Analisis dan Menarik Kesimpulan Penelitian .....                    | V-21          |
| 3.4 Kesimpulan .....  | V-22          |
| <b>BAB IV PENGEMBANGAN PERANGKAT LUNAK.....</b>                                     | <b>VI-23</b>  |
| 4.1 Pendahuluan .....   | VI-23         |
| 4.2 Fase Insepsi.....   | VI-23         |
| 4.2.1 Pemodelan Bisnis.....   | VI-23         |
| 4.2.2 Kebutuhan Sistem .....  | VI-23         |
| 4.2.3 Analisis dan Desain.....  | VI-24         |
| 4.3 Fase Elaborasi .....  | VI-39         |
| 4.3.1 Pemodelan Bisnis.....   | VI-39         |
| 4.3.2 Kebutuhan Sistem .....  | VI-41         |
| 4.3.3 Analisis dan Perancangan .....  | VI-42         |
| 4.4 Fase Konstruksi.....  | VI-43         |
| 4.4.1 Kebutuhan Sistem .....  | VI-43         |
| 4.4.2 Implementasi.....   | VI-44         |
| 4.5 Fase Transisi .....   | VI-47         |
| 4.5.1 Pemodelan Bisnis.....   | VI-47         |
| 4.5.2 Rencana Pengujian.....  | VI-47         |
| 4.5.3 Implementasi.....   | VI-47         |
| 4.6 Kesimpulan .....  | VI-48         |
| <b>BAB V HASIL DAN ANALISIS.....</b>  | <b>VII-47</b> |
| 5.1 Pendahuluan.....  | VII-47        |
| 5.2 Hasil Penelitian .....  | VII-47        |
| 5.2.1 Konfigurasi Percobaan.....  | VII-47        |
| 5.2.2 Data Hasil Pengujian Menggunakan <i>Jointly Local and Global Context</i> .... | VII-51        |
| 5.2.3 Data Hasil Pengujian Menggunakan <i>Traditional Centrality</i> .....          | VII-70        |
| 5.3 Analisis Hasil Penelitian .....   | VII-88        |

|  |                |
|--|----------------|
| 5.4 Kesimpulan .....                     | VII-91         |
| <b>BAB VI KESIMPULAN DAN SARAN .....</b> | <b>VIII-47</b> |
| 6.1 Pendahuluan .....                    | VIII-47        |
| 6.2 Kesimpulan .....                     | VIII-47        |
| 6.3 Saran .....                          | VIII-48        |
| <b>DAFTAR PUSTAKA .....</b>              | <b>xv</b>      |

## DAFTAR TABEL

|   |        |
|---|--------|
| <b>Tabel III-1.</b> Hasil Evaluasi.....   | V-20   |
| <b>Tabel III-2.</b> Alat Bantu yang Digunakan Dalam Penelitian.....                                     | V-20   |
| <b>Tabel III-3.</b> Hasil Analisis Candidate Keyphrase dengan Model Lokal dan Global ...                | V-22   |
| <b>Tabel III-4.</b> Tabel Hasil Analisis Candidate Keyphrase dengan Traditional Graph....               | V-22   |
| <b>Tabel IV-1.</b> Kebutuhan Fungsional.....  | VI-24  |
| <b>Tabel IV-2.</b> Kebutuhan Non-Fungsional.....  | VI-24  |
| <b>Tabel IV-3.</b> Contoh Data Judul .....  | VI-25  |
| <b>Tabel IV-4.</b> Contoh Data Abstrak .....  | VI-25  |
| <b>Tabel IV-5.</b> Hasil Penggabungan Judul dan Abstrak .....   | VI-26  |
| <b>Tabel IV-6.</b> Hasil Penghapusan Karakter Spesial dan Tokenisasi .....                              | VI-27  |
| <b>Tabel IV-7.</b> Hasil Stopword Removal.....  | VI-28  |
| <b>Tabel IV-8.</b> Hasil Post Tagging .....   | VI-28  |
| <b>Tabel IV-9.</b> Hasil Noun Phrase Chunking .....   | VI-29  |
| <b>Tabel IV-10.</b> Hasil Embedding Menggunakan Indobert-Large-P1 .....                                 | VI-30  |
| <b>Tabel IV-11.</b> Hasil Embedding Menggunakan Indobert-Large-P2.....                                  | VI-31  |
| <b>Tabel IV-12.</b> Hasil Skoring Menggunakan JMLGC Pada Indobert-Large-P1 .....                        | VI-32  |
| <b>Tabel IV-13.</b> Hasil Skoring Menggunakan JMLGC Pada Indobert-Large-P2.....                         | VI-32  |
| <b>Tabel IV-14.</b> 5 Kandidat dengan Skor Terbesar Menggunakan Indobert-Large-P1 ...                   | VI-33  |
| <b>Tabel IV-15.</b> 5 Kandidat dengan Skor Terbesar Menggunakan Indobert-Large-P2 ...                   | VI-33  |
| <b>Tabel IV-16.</b> Hasil Evaluasi Confusion Matrix .....   | VI-34  |
| <b>Tabel IV-17.</b> Hasil Confusion Matrix dengan sampel jurnal .....                                   | VI-37  |
| <b>Tabel IV-18.</b> Definisi User.....  | VI-38  |
| <b>Tabel IV-19.</b> Definisi Use Case .....   | VI-38  |
| <b>Tabel IV-20.</b> Skenario Use Case Ekstraksi Kata Kunci.....   | VI-38  |
| <b>Tabel IV-21.</b> Implementasi Kelas .....  | VI-45  |
| <b>Tabel IV-22.</b> Rencana Pengujian Use Case Ekstraksi Kata Kunci .....                               | VI-47  |
| <b>Tabel IV-23.</b> Pengujian Use Case Ekstraksi Kata Kunci .....                                       | VI-48  |
| <b>Tabel V-1.</b> Sampel Data Uji.....  | VII-48 |
| <b>Tabel V-2.</b> Data Pengujian dan Kandidat Kata Kunci dengan Indobert-Large-P1 ...                   | VII-48 |
| <b>Tabel V-3.</b> Data Pengujian dan Kandidat Kata Kunci dengan Indobert-Large-P2 ...                   | VII-49 |
| <b>Tabel V-4.</b> Top 5 Selected Keyphrase Indobert-Large-P1 dan Indobert-Large-P2...                   | VII-51 |
| <b>Tabel V-5.</b> Top 5 Selected Keyphrase Indobert-Base-P1 dan Indobert-Base-P2.....                   | VII-52 |
| <b>Tabel V-6.</b> Top 5 Selected Keyphrase Indobert-Lite-Base-P1<br>dan Indobert-Lite-Base-P2.....      | VII-53 |
| <b>Tabel V-7.</b> Top 5 Selected Keyphrase Indobert-Lite-Large-P1<br>dan Indobert-Lite-Large-P2 .....   | VII-53 |
| <b>Tabel V-8.</b> Top 10 Selected Keyphrase Indobert-Large-P1 dan Indobert-Large-P2                     | VII-54 |
| <b>Tabel V-9.</b> Top 10 Selected Keyphrase Indobert-Base-P1 dan Indobert-Base-P2...                    | VII-55 |
| <b>Tabel V-10.</b> Top 10 Selected Keyphrase Indobert-Lite-Base-P1<br>dan Indobert-Lite-Base-P2 .....   | VII-56 |
| <b>Tabel V-11.</b> Top 10 Selected Keyphrase Indobert-Lite-Large-P1<br>dan Indobert-Lite-Large-P2 ..... | VII-57 |
| <b>Tabel V-12.</b> Top 15 Selected Keyphrase Indobert-Large-P1 dan Indobert-Large-P2                    | VII-58 |
| <b>Tabel V-13.</b> Top 15 Selected Keyphrase Indobert-Base-P1 dan Indobert-Base-P2..                    | VII-59 |

|   |        |
|---|--------|
| <b>Tabel V-14.</b> Top 15 Selected Keyphrase Indobert-Lite-Base-P1<br>dan Indobert-Lite-Base-P2 .....                 | VII-60 |
| <b>Tabel V-15.</b> Top 15 Selected Keyphrase Indobert-Lite-Large-P1<br>dan Indobert-Lite-Large-P2.....                | VII-62 |
| <b>Tabel V-16.</b> Hasil Evaluasi Top 5 Selected Keyphrase Indobert-Large-P1<br>dan Indobert-Large-P2.....            | VII-63 |
| <b>Tabel V-17.</b> Hasil Evaluasi Top 5 Selected Keyphrase Indobert-Base-P1<br>dan Indobert-Base-P2 .....             | VII-63 |
| <b>Tabel V-18.</b> Hasil Evaluasi Top 5 Selected Keyphrase Indobert-Lite-Base-P1<br>dan Indobert-Lite-Base-P2 .....   | VII-64 |
| <b>Tabel V-19.</b> Hasil Evaluasi Top 5 Selected Keyphrase Indobert-Lite-Large-P1<br>dan Indobert-Lite-Large-P2.....  | VII-65 |
| <b>Tabel V-20.</b> Hasil Evaluasi Top 10 Selected Keyphrase Indobert-Large-P1<br>dan Indobert-Large-P2.....           | VII-65 |
| <b>Tabel V-21.</b> Hasil Evaluasi Top 10 Selected Keyphrase Indobert-Base-P1<br>dan Indobert-Base-P2 .....            | VII-66 |
| <b>Tabel V-22.</b> Hasil Evaluasi Top 10 Selected Keyphrase Indobert-Lite-Base-P1<br>dan Indobert-Lite-Base-P2 .....  | VII-66 |
| <b>Tabel V-23.</b> Hasil Evaluasi Top 10 Selected Keyphrase Indobert-Lite-Large-P1<br>dan Indobert-Lite-Large-P2..... | VII-67 |
| <b>Tabel V-24.</b> Hasil Evaluasi Top 15 Selected Keyphrase Indobert-Large-P1<br>dan Indobert-Large-P2.....           | VII-67 |
| <b>Tabel V-25.</b> Hasil Evaluasi Top 15 Selected Keyphrase Indobert-Base-P1<br>dan Indobert-Base-P2 .....            | VII-68 |
| <b>Tabel V-26.</b> Hasil Evaluasi Top 15 Selected Keyphrase Indobert-Lite-Base-P1<br>dan Indobert-Lite-Base-P2 .....  | VII-68 |
| <b>Tabel V-27.</b> Hasil Evaluasi Top 15 Selected Keyphrase Indobert-Lite-Large-P1<br>dan Indobert-Lite-Large-P2..... | VII-69 |
| <b>Tabel V-28.</b> Rata-rata pengukuran tiap ketentuan pengambilan .....  | VII-69 |
| <b>Tabel V-29.</b> Top 5 Selected Traditional Keyphrase Indobert-Large-P1<br>dan Indobert-Large-P2.....               | VII-71 |
| <b>Tabel V-30.</b> Top 5 Selected Traditional Keyphrase Indobert-Base-P1<br>dan Indobert-Base-P2 .....                | VII-71 |
| <b>Tabel V-31.</b> Top 5 Selected Traditional Keyphrase Indobert- Lite-Base-P1<br>dan Indobert-Lite-Base-P2 .....     | VII-72 |
| <b>Tabel V-32.</b> Top 5 Selected Traditional Keyphrase Indobert- Lite-Large-P1<br>dan Indobert-Lite-Large-P2.....    | VII-72 |
| <b>Tabel V-33.</b> Top 10 Selected Traditional Keyphrase Indobert-Large-P1<br>dan Indobert-Large-P2.....              | VII-73 |
| <b>Tabel V-34.</b> Top 10 Selected Traditional Keyphrase Indobert-Base-P1<br>dan Indobert-Base-P2 .....               | VII-74 |
| <b>Tabel V-35.</b> Top 10 Selected Traditional Keyphrase Indobert- Lite-Base-P1<br>dan Indobert-Lite-Base-P2 .....    | VII-75 |
| <b>Tabel V-36.</b> Top 10 Selected Traditional Keyphrase Indobert- Lite-Large-P1<br>dan Indobert-Lite-Large-P2.....   | VII-76 |
| <b>Tabel V-37.</b> Top 15 Selected Traditional Keyphrase Indobert-Large-P1<br>dan Indobert-Large-P2.....              | VII-77 |

|   |        |
|---|--------|
| <b>Tabel V-38.</b> Top 15 Selected Traditional Keyphrase Indobert-Base-P1<br>dan Indobert-Base-P2 .....                             | VII-78 |
| <b>Tabel V-39.</b> Top 15 Selected Traditional Keyphrase Indobert- Lite-Base-P1<br>dan Indobert-Lite-Base-P2 .....                  | VII-79 |
| <b>Tabel V-40.</b> Top 15 Selected Traditional Keyphrase Indobert- Lite-Large-P1<br>dan Indobert-Lite-Large-P2.....                 | VII-80 |
| <b>Tabel V-41.</b> Hasil Evaluasi Top 5 Selected Traditional Keyphrase<br>Indobert-Large-P1 dan Indobert-Large-P2.....              | VII-81 |
| <b>Tabel V-42.</b> Hasil Evaluasi Top 5 Selected Traditional Keyphrase<br>Indobert-Base-P1 dan Indobert-Base-P2.....                | VII-81 |
| <b>Tabel V-43.</b> Hasil Evaluasi Top 5 Selected Traditional Keyphrase<br>Indobert-Lite-Base-P1 dan Indobert-Lite-Base-P2 .....     | VII-82 |
| <b>Tabel V-44.</b> Hasil Evaluasi Top 5 Selected Traditional Keyphrase<br>Indobert-Lite-Large-P1 dan Indobert-Lite-Large-P2 .....   | VII-82 |
| <b>Tabel V-45.</b> Hasil Evaluasi Top 10 Selected Traditional Keyphrase<br>Indobert-Large-P1 dan Indobert-Large-P2.....             | VII-83 |
| <b>Tabel V-46.</b> Hasil Evaluasi Top 10 Selected Traditional Keyphrase<br>Indobert-Base-P1 dan Indobert-Base-P2.....               | VII-83 |
| <b>Tabel V-47.</b> Hasil Evaluasi Top 10 Selected Traditional Keyphrase<br>Indobert- Lite-Base-P1 dan Indobert-Lite-Base-P2 .....   | VII-84 |
| <b>Tabel V-48.</b> Hasil Evaluasi Top 10 Selected Traditional Keyphrase<br>Indobert- Lite-Large-P1 dan Indobert-Lite-Large-P2 ..... | VII-85 |
| <b>Tabel V-49.</b> Hasil Evaluasi Top 15 Selected Traditional Keyphrase<br>Indobert-Large-P1 dan Indobert-Large-P2.....             | VII-85 |
| <b>Tabel V-50.</b> Hasil Evaluasi Top 15 Selected Traditional Keyphrase<br>Indobert-Base-P1 dan Indobert-Base-P2.....               | VII-86 |
| <b>Tabel V-51.</b> Hasil Evaluasi Top 15 Selected Traditional Keyphrase<br>Indobert- Lite-Base-P1 dan Indobert-Lite-Base-P2 .....   | VII-86 |
| <b>Tabel V-52.</b> Hasil Evaluasi Top 15 Selected Traditional Keyphrase<br>Indobert- Lite-Large-P1 dan Indobert-Lite-Large-P2 ..... | VII-87 |
| <b>Tabel V-53.</b> Rata-rata Pengukuran Tiap Pengambilan Menggunakan Traditional<br>Centrality .....                                | VII-87 |

## DAFTAR GAMBAR

|   |        |
|---|--------|
| <b>Gambar II-1.</b> BERT .....  | IV-6   |
| <b>Gambar III-1.</b> Rincian Kegiatan Penelitian .....  | V-17   |
| <b>Gambar III-2.</b> Kerangka Kerja Penelitian.....   | V-18   |
| <b>Gambar IV-1.</b> Diagram Use Case .....  | VI-37  |
| <b>Gambar IV-2.</b> Desain Antarmuka Aplikasi.....  | VI-40  |
| <b>Gambar IV-3.</b> Desain Antarmuka Pembahasan Aplikasi.....                                     | VI-41  |
| <b>Gambar IV-4.</b> Desain Antarmuka Evaluasi Aplikasi .....                                      | VI-41  |
| <b>Gambar IV-5.</b> Activity Diagram Ekstraksi Kata Kunci .....                                   | VI-42  |
| <b>Gambar IV-6.</b> Sequence Diagram Ekstraksi Kata Kunci .....                                   | VI-43  |
| <b>Gambar IV-7.</b> Class Diagram Ekstraksi Kata Kunci.....                                       | VI-44  |
| <b>Gambar IV-8.</b> Implementasi Antarmuka.....   | VI-46  |
| <b>Gambar IV-9.</b> Implementasi Antarmuka Pembahasan.....  | VI-46  |
| <b>Gambar IV-10.</b> Implementasi Antarmuka Hasil .....   | VI-47  |
| <b>Gambar V-1.</b> Diagram Perbandingan JMLGC Menggunakan Model Indobert .....                    | VII-70 |
| <b>Gambar V-2.</b> Diagram Evaluasi Traditional Centrality<br>Menggunakan Indobert-Large-P1 ..... | VII-88 |
| <b>Gambar V-3.</b> Perbandingan Konfigurasi JMLGC dan Traditional .....                           | VII-90 |



# **BAB I**

## **PENDAHULUAN**

### **1.1 Pendahuluan**

Bab awal ini akan menguraikan konteks di balik topik, perumusan isu, tujuan riset, keunggulan riset, lingkup masalah, dan struktur penulisan. Bagian ini bertujuan untuk memberikan gambaran keseluruhan tentang apa yang akan dibahas di bab-bab berikutnya. Pendahuluan akan dimulai dengan mengidentifikasi masalah yang ada serta mencari solusi yang tepat untuk mengatasinya.

### **1.2 Latar Belakang Masalah**

Dalam bidang penelitian, banyak sekali publikasi ilmiah yang tersedia melalui jurnal-jurnal yang sering diakses oleh berbagai kalangan, termasuk pelajar, mahasiswa, dosen, dan bahkan pegawai kantor, sebagai sumber informasi yang dapat diandalkan mengenai hasil penelitian dan data yang sah. Di Indonesia, upaya untuk meningkatkan indeksasi publikasi hasil penelitian sedang giat dilakukan, tidak hanya di tingkat internasional tetapi juga di tingkat nasional. Tujuan dari publikasi hasil penelitian ini adalah untuk menyebarkan informasi tersebut ke berbagai kalangan, baik masyarakat umum maupun kalangan peneliti dan akademisi (Arini et al., 2019). Dengan banyaknya publikasi ilmiah berbahasa Indonesia yang ada, dibutuhkan waktu bagi para *reviewer* untuk memilih publikasi yang sesuai dengan bidang keilmuan yang mereka kuasai.

Pentingnya kata kunci dalam menentukan topik penelitian tidak dapat diabaikan. Kata kunci sering digunakan dalam konteks jurnal ilmiah untuk mencerminkan substansi dari isi jurnal tersebut (Mothe et al., 2018). Proses ekstraksi kata kunci melibatkan identifikasi kata atau frasa penting yang mewakili inti dari dokumen atau teks (Zhang et al., 2023). Dalam ekstraksi otomatis kata kunci, tujuannya adalah untuk mengenali frasa atau kata-kata kunci yang paling relevan untuk mencerminkan inti dari dokumen tersebut (Saxena et al., 2020). Secara keseluruhan, ekstraksi kata kunci otomatis merupakan bagian dari bidang

Pengolahan Bahasa Alami yang fokus pada pengembangan metode yang efektif dan efisien dengan menggunakan frase kunci dalam dokumen.

Dalam ekstraksi kata kunci, terdapat dua pendekatan umum yang digunakan *supervised learning* dan *unsupervised learning*. Pendekatan *supervised learning* membutuhkan dataset yang sudah diberi label dengan jumlah yang cukup besar, sering kali terbatas pada domain atau dataset tertentu (Zhang et al., 2023). Sebaliknya, pendekatan *unsupervised learning* lebih fleksibel karena mampu mengekstrak informasi intrinsik dari dokumen tanpa memerlukan label yang sudah ditentukan sebelumnya (Zhang et al., 2023).

Dalam ekstraksi otomatis kata kunci, penggunaan *pre-trained language model* sangat penting. Dalam penelitian ini, penulis memanfaatkan salah satu jenis *pre-trained language model*, yaitu BERT, untuk menghasilkan representasi kontekstual dinamis dari kalimat dan kandidat frasa kunci (Zhang et al., 2023). BERT dikenal dapat memberikan representasi teks yang berkualitas tinggi dengan melakukan *encoding* kata dan kalimat secara dinamis. Secara spesifik, penelitian ini menggunakan model IndoBert (Wilie et al., 2020) karena model tersebut telah dilatih dengan dataset bahasa Indonesia yang melibatkan miliaran kata, sehingga siap digunakan dalam konteks penelitian.

Kemudian, dalam proses perankingan *keyphrase extraction* menggunakan metode baru yang disebut *Jointly Modeling Local and Global Context*, pendekatan ini memperhitungkan konteks lokal dan global secara simultan untuk menentukan kata kunci (Liang et al., 2021). Pendekatan ini berusaha menggabungkan informasi dari sekitar kata atau frasa (konteks lokal) dan informasi yang lebih luas yang mencakup seluruh dokumen atau teks (konteks global) (Liang et al., 2021).

Penelitian ini akan membangun program yang melakukan ekstraksi kata kunci untuk teks berbahasa Indonesia menggunakan *pre-trained language model* BERT dan *Jointly Modeling Local and Global Context*.

### 1.3 Rumusan Masalah

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

1. Bagaimana mengembangkan Ekstraksi Kata Kunci pada Bahasa Indonesia menggunakan metode *Jointly Modeling Local and Global Context* dan BERT?
2. Bagaimana evaluasi kinerja dari Ekstraksi Kata Kunci Bahasa Indonesia menggunakan pendekatan *Jointly Modeling Local and Global Context* dan BERT?

### 1.4 Tujuan Penulisan

Tujuan dari penelitian ini adalah sebagai berikut:

1. Menghasilkan sebuah perangkat lunak yang dapat melakukan Ekstraksi Kata Kunci pada Bahasa Indonesia dengan menerapkan pendekatan *Jointly Modeling Local and Global Context* dan BERT.
2. Mengetahui kinerja dari Ekstraksi Kata Kunci Bahasa Indonesia dengan menerapkan pendekatan *Jointly Modeling Local and Global Context* dan BERT.

### 1.5 Manfaat Penelitian

Manfaat yang didapat dari penelitian ini adalah sebagai berikut:

1. Perangkat lunak dapat melakukan Ekstraksi Kata Kunci pada Bahasa Indonesia menggunakan metode *Jointly Modeling Local and Global Context* dan BERT.
2. Temuan dari penelitian ini dapat digunakan sebagai acuan untuk penelitian yang relevan.

### 1.6 Batasan Penelitian

Agar permasalahan tidak menyimpang dari batasan yang telah ditetapkan, maka adapun batasan dari penelitian ini adalah sebagai berikut:

1. Data yang digunakan adalah 100 data Publikasi Ilmiah dari website JTIK<sup>1</sup>, JATISI<sup>2</sup>, JEPIN<sup>3</sup>.
2. Data publikasi paper jurnal ilmiah yang digunakan adalah judul, abstrak dan kata kunci penulis.

### **1.7 Sistematika Penulisan**

Struktur penulisan yang diterapkan dalam penelitian ini mengacu pada pedoman penulisan tugas akhir yang ditetapkan oleh Fakultas Ilmu Komputer Universitas Sriwijaya, yang meliputi:

## **BAB I. PENDAHULUAN**

Bagian ini mengulas mengenai konteks, perumusan masalah, tujuan studi, manfaat penelitian, batasan penelitian, dan struktur penulisan yang diterapkan dalam penyusunan dokumen akhir ini.

## **BAB II. KAJIAN LITERATUR**

Bagian ini memberikan penjelasan tentang dasar teori yang mendukung penelitian ini. Bab ini mencakup tinjauan literatur dan studi terkait sebelumnya yang relevan dengan topik penelitian, termasuk pembahasan tentang *Pre-Trained Language Model BERT*, *Jointly Modeling Local and Global Context*, serta aspek lain yang terkait.

## **BAB III. METODOLOGI PENELITIAN**

Bagian ini akan menguraikan langkah-langkah atau proses yang dilakukan selama penelitian, mulai dari metode pengumpulan data hingga metode dalam perancangan perangkat lunak. Setiap tahapan penelitian akan dijelaskan secara terperinci sesuai dengan kerangka kerja yang telah ditetapkan.

---

<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>

## **1.8 Kesimpulan**

Dalam bagian ini, telah dipaparkan tentang konteks, perumusan masalah, tujuan dan manfaat penelitian, batasan penelitian, dan struktur penelitian yang akan menjadi fokus perhatian peneliti.

## DAFTAR PUSTAKA

- Arini, Ni Wayan Sri, Ida Bagus Putu Widja, and I. Komang Rinatha Yasa Negara. "Analisis Frekuensi Kata untuk Mengekstrak Kata Kunci dari Artikel Ilmiah Berbahasa Indonesia." *Jurnal Eksplora Informatika* 8.2 (2019): 80-84.
- Denny, Matthew J., and Arthur Spirling. "Text preprocessing for unsupervised learning: Why it matters, when it misleads, and what to do about it." *Political Analysis* 26.2 (2018): 168-189.
- Bennani-Smires, K., Musat, C., Hossmann, A., Baeriswyl, M., & Jaggi, M. (2018). Simple Unsupervised Keyphrase Extraction using Sentence Embeddings. *Proceedings of the 22nd Conference on Computational Natural Language Learning*, 221–229. <https://doi.org/10.18653/v1/K18-1022>
- Clark, K., Khandelwal, U., Levy, O., & Manning, C. D. (2019). *What Does BERT Look At? An Analysis of BERT's Attention* (arXiv:1906.04341). arXiv. <http://arxiv.org/abs/1906.04341>
- Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). *BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding*.
- 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).
- Koto, F., Rahimi, A., Lau, J. H., & Baldwin, T. (2020). IndoLEM and IndoBERT: *A Benchmark Dataset and Pre-trained Language Model for Indonesian NLP* (arXiv:2011.00677). arXiv. <http://arxiv.org/abs/2011.00677>
- Kurniawan, A. (2021). Aplikasi Sistem Ekstraksi Kata Kunci Berbahasa Indonesia Menggunakan Algoritma TextRank Studi Kasus Data Wikipedia Indonesia.
- Lamasigi, Zulfrianto Yusrin. (2021). "DCT Untuk Ekstraksi Fitur Berbasis GLCM Pada Identifikasi Batik Menggunakan K-NN." *Jambura Journal of Electrical and Electronics Engineering* 3.1 (1-6).

- Liang, X., Wu, S., Li, M., & Li, Z. (2021). Unsupervised Keyphrase Extraction by Jointly Modeling Local and Global Context. *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, 155–164. <https://doi.org/10.18653/v1/2021.emnlp-main.14>
- 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, Zara, Syed Waqar Jaffry, and Muhammad Kamran Malik. "Textual keyword extraction and summarization: State-of-the-art." *Information Processing & Management* 56.6 (2019): 102088.
- Peters, M., Neumann, M., Iyyer, M., Gardner, M., Clark, C., Lee, K., & Zettlemoyer, L. (2018). *Deep Contextualized Word Representations*. *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, 2227–2237. <https://doi.org/10.18653/v1/N18-1202>
- Plakasa, Gerald. (2022). Ekstraksi Kata Kunci Pada Bahasa Indonesia Menggunakan Metode Yake.
- Radford, A., Narasimhan, K., Salimans, T., & Sutskever, I. (2018). *Improving Language Understanding by Generative Pre-Training*.
- Santoso, J., Setiawan, E. I., Ferdinandus, F. X., Gunawan, G., & Hernandez, L. (2022). *Indonesian Language Term Extraction using Multi-Task Neural Network*. *Knowledge Engineering and Data Science*, 5(2), 160, <https://doi.org/10.17977/um018v5i22022p160-167>
- Saxena, A., Mangal, M., & Jain, G. (2020). KeyGames: A Game Theoretic Approach to Automatic Keyphrase Extraction. *Proceedings of the 28th International Conference on Computational Linguistics*, 2037–2048. <https://doi.org/10.18653/v1/2020.coling-main.184>

- Smith, J. D. (2020). *The Art of Writing Texts: A Comprehensive Guide*. Publisher.
- Sun, Y., Qiu, H., Zheng, Y., Wang, Z., & Zhang, C. (2020). SIFRank: A new baseline for unsupervised keyphrase extraction based on pre-trained language model. *IEEE Access*, 8, 10896–10906.
- de Vries, W., Wieling, M., & Nissim, M. (2022). Make the Best of Cross-lingual Transfer: Evidence from POS Tagging with over 100 Languages. In S. Muresan, P. Nakov, & A. Villavicencio (Eds.), *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)* (pp.7676-7685). Association for Computational Linguistics (ACL).
- Wilie, B., Vincentio, K., Winata, G. I., Cahyawijaya, S., Li, X., Lim, Z. Y., Soleman, S., Mahendra, R., Fung, P., Bahar, S., & Purwarianti, A. (2020). *IndoNLU: Benchmark and Resources for Evaluating Indonesian Natural Language Understanding*.
- Winkie Setyono, Donni Richasdy, & Mahendra D. Purbolaksono. 2022. POS Tagger Improvisation using HMM with the Addition of Foreign Word Labels on Telkom University News. 588-594. DOI 10.47065/bits.v4i2.1983.
- You, J., Ying, R., & Leskovec, J. (2019). *Position-aware Graph Neural Networks*.
- Yunefri, Y., Fadrial, Y. E., & Sutejo. 2021. Chatbot Pada Smart Cooperative Oriented Problem Menggunakan Natural Language Processing dan Naive Bayes Classifier. *Journal of Information Technology and Computer Science (INTECOMS)*, 4(2), 131–141.
- Yusuf, R., Saputri, T. A., & Wicaksono, A. A. (2022). Penerapan Natural Language Processing Berbasis Virtual Assistant Pada Bagian Administrasi Akademik Stmik Dharma Wacana. *International Research on Big-Data and Computer Technology: I-Robot*, 5(1), 33–47  
<https://doi.org/10,53514/ir.v5i1.228>



Zhang, Z., Liang, X., Zuo, Y., & Lin, C. (2023). *Improving unsupervised keyphrase extraction by modeling hierarchical multi-granularity features*. *Information Processing & Management*, 60(4), 103356. <https://doi.org/10.1016/j.ipm.2023.103356>.

Zheng, H., & Lapata, M. (2019). Sentence Centrality Revisited for Unsupervised Summarization. *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, 6236–6247. <https://doi.org/10.18653/v1/P19-1628>.