`

KEYPHRASE EXTRACTION MENGGUNAKAN *PRE-TRAINED LANGUAGE MODEL* ROBERTA DAN *TOPIC GUIDED GRAPH ATTENTION NETWORK*

Diajukan Untuk Memperoleh Gelar Strata-1

di Jurusan Teknik Informatika Fakultas Ilmu Komputer UNSRI

Oleh:

Muhammad Raihan Habibullah

NIM: 09021282025087

Jurusan Teknik Informatika

FAKULTAS ILMU KOMPUTER UNIVERSITAS SRIWIJAYA

2024

LEMBAR PENGESAHAN SKRIPSI

KEYPHRASE EXTRACTION MENGGUNAKAN *PRE-TRAINED LANGUAGE MODEL* ROBERTA DAN *TOPIC GUIDED GRAPH ATTENTION NETWORK*

Oleh:

Muhammad Raihan Habibullah
NIM: 09021282025087

Palembang, 2 Januari 2024

Pembimbing I,

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

Pembimbing II,

Junia Kurniati, M.Kom.
NIK. 1671046606890018

Mengetahui,
Ketua Jurusan Teknik Informatika

Alvi Syahrini Utami, M.Kom.
NIP. 197812222006042003

TANDA LULUS UJIAN KOMPREHENSIF SKRIPSI

Pada hari Jumat tanggal 29 Desember 2023 telah dilaksanakan ujian komprehensif skripsi oleh Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya

Nama : Muhammad Raihan Habibullah
NIM : 09021282025087
Judul : *Keyphrase Extraction* Menggunakan *Pre-Trained Language Model* RoBERTa dan *Topic Guided Graph Attention Network*

dan dinyatakan **LULUS**.

1. Ketua Penguji

   Desty Rodiah, M.T.
   NIP. 198912212020122011 ..... 

2. Penguji I

   Alvi Syahrini Utami, M.Kom.
   NIP. 197812222006042003 .....

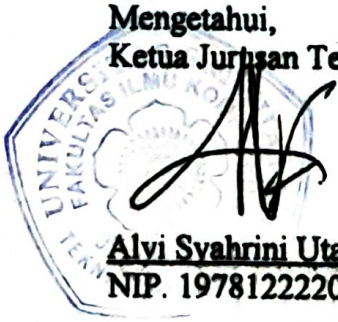
3. Pembimbing I

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

4. Pembimbing II

   Junia Kurniati, M.Kom.
   NIK. 1671046606890018 .....

Mengetahui,
Ketua Jurusan Teknik Informatika,



Alvi Syahrini Utami, M.Kom.
NIP. 197812222006042003

HALAMAN PERNYATAAN

Saya yang bertanda tangan di bawah ini:

Nama	: Muhammad Raihan Habibullah
NIM	: 09021282025087
Program Studi	: Teknik Informatika
Judul Skripsi	: *Keyphrase Extraction* Menggunakan *Pre-Trained Language Model RoBERTa* dan *Topic Guided Graph Attention Network*

Hasil pengecekan *Software iThenticate/Turnitin*: 12%

Menyatakan bahwa laporan proyek saya merupakan hasil kerja sendiri dan bukan hasil penjiplakan/plagiat. Apabila ditemukan unsur penjiplakan/plagiat dalam laporan proyek ini, 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, 5 Januari 2024

Muhammad Raihan Habibullah
NIM. 09021282025087

`

MOTTO DAN PERSEMBAHAN

Motto:

"Learn from yesterday, live for today, hope for tomorrow. The important thing is not to stop learning."

- Albert Einstein.

Kupersembahkan Karya Tulis ini kepada:

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

`

ABSTRACT

In the current digital information era, the vast amount of text from various sources such as websites and academic documents poses a challenge for humans to comprehend the information contained in readings. Keyphrase extraction can be one solution to determine the most relevant words from a scholarly publication. One of the keyphrase extraction methods is the Pre-trained Language Model RoBERTa (Robustly Optimized BERT Pretraining Approach) and TgGAT (Topic Guided Graph Attention Networks). This research aims to perform keyphrase extraction in the Indonesian language using RoBERTa and TgGAT. The dataset utilized for this study consists of 100 scholarly publications from previous research by Plakasa (2022), specifically in the field of Computer Science. Based on the research findings, the configuration of the parameter for the number of keywords or top keyphrases has an impact on the generated keyphrase f-score and accuracy. The evaluation results of this study obtained an f-score of 4.65% and an accuracy of 59.3% with a parameter configuration of 15 top keyphrases.

Keyword : Keyphrase Extraction, RoBERTa, Topic Guided Graph Attention Networks, f-score, accuracy

`

ABSTRAK

Pada era informasi digital saat ini, jumlah teks yang besar dari berbagai sumber seperti situs web dan dokumen akademis menjadi tantangan bagi manusia dalam memahami informasi yang terkandung dalam bacaan. *Keyphrase extraction* dapat menjadi salah satu solusi untuk menentukan kata-kata yang paling relevan dari suatu publikasi ilmiah. Salah satu metode *keyphrase extraction* yaitu *Pre-trained Language Model* RoBERTa (*Robustly Optimized* BERT *Pretraining Approach*) dan TgGAT (*Topic Guided Graph Attention Networks*). Penelitian ini akan melakukan *Keyphrase extraction* pada Bahasa Indonesia menggunakan RoBERTa dan TgGAT. *Dataset* yang digunakan untuk melakukan penelitian ini adalah 100 publikasi ilmiah dari penelitian terdahulu oleh Plakasa (2022), terkhusus pada topik Ilmu Komputer. Berdasarkan hasil penelitian yang dilakukan, konfigurasi parameter jumlah kata kunci atau *top keyphrase* memiliki pengaruh terhadap kata kunci *f-score*, dan *accuracy* yang dihasilkan. Hasil evaluasi dari penelitian ini mendapat nilai *f-score* 4,65% dan *accuracy* 59,3% dengan konfigurasi parameter 15 *top keyphrase*.

Kata Kunci: *Keyphrase Extraction*, RoBERTa, *Topic Guided Graph Attention Network, f-score, accuracy*

`

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 saudara kandung yang telah mendoakan, memberi semangat, memotivasi, dan nasihat untuk menyelesaikan skripsi ini.
3. Prof. DR. Erwin, S.Si., M.Si. selaku Dekan Fakultas Ilmu Komputer Universitas Sriwijaya
4. Ibu Alvi Syahrini Utami, M.Kom. selaku Ketua Jurusan Teknik Informatika Universitas Sriwijaya
5. Ibu Novi Yusliani, S.Kom., M.T. selaku Dosen Pembimbing I dan Ibu Junia Kurniati, M.Kom. selaku dosen Pembimbing II yang telah membimbing, memberikan motivasi serta arahan kepada penulis dalam proses pengerjaan skripsi.

`

6. Ibu Dian Palupi Rini, M.Kom., Ph.D. selaku Pembimbing akademik dari selama masa perkuliahan di Fakultas Ilmu Komputer Universitas Sriwijaya.
7. Seluruh dosen program studi serta admin Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya.
8. Teman – teman penulis terutama 09031282025072 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, 15 Januari 2024

Penulis

Muhammad Raihan Habibullah

`

DAFTAR ISI

Halaman

`

`

`

`

DAFTAR TABEL

`

`

`

DAFTAR GAMBAR

`

`

DAFTAR LAMPIRAN

`

DAFTAR ISTILAH, SINGKATAN DAN LAMBANG

Anchor	:	*Node* penting di dalam *Anchor Aware Graph*
Anchor Aware Graph	:	Representasi graf yang menggunakan konsep "*anchor*" sebagai *node* penting
BERT	:	Sebuah model bahasa yang menggunakan arsitektur *transformer* untuk memahami konteks kata dalam suatu kalimat
Case Folding	:	Proses merubah seluruh huruf dalam teks menjadi huruf kecil atau huruf besar
Noun Phrase Chunking	:	Proses mengelompokkan kata-kata untuk membentuk frasa benda (*noun phrase*)
Golden Keyphrase	:	Kata kunci yang dihasilkan oleh penulis dokumen
Keyphrase Extraction	:	Proses meng*keyphrase extraction* atau frasa yang mewakili inti dari suatu dokumen
Masked Language Model	:	Model bahasa di mana beberapa kata dalam suatu kalimat di-"*mask*" dan model diminta untuk memprediksi kata-kata yang di-*mask*
Neural Topic Module	:	Merujuk pada modul dalam suatu sistem yang bertanggung jawab untuk mengekstrak topik dari teks menggunakan pendekatan berbasis jaringan saraf
Next Sentence Prediction	:	Salah satu tugas dalam pelatihan model bahasa di mana model diminta untuk memprediksi apakah satu kalimat akan mengikuti kalimat lainnya
POS *Tagging*	:	Proses menandai setiap kata dalam suatu teks dengan kategori tata bahasa seperti nomina, verba, atau adjektiva
Pra-Pemrosesan Teks	:	Tahap pra-pemrosesan dalam analisis teks

`

		yang melibatkan langkah-langkah seperti *tokenisasi*, penghilangan *stopwords*, dan *stemming*
Pre-Trained Language Model	:	Model bahasa yang telah dilatih sebelumnya pada tugas-tugas tertentu dan dapat digunakan untuk tugas-tugas baru
RoBERTa	:	Sebuah model bahasa yang merupakan pengembangan dari arsitektur BERT dengan modifikasi tertentu untuk meningkatkan kinerja
Topic Guided Graph Attention Network	:	Merujuk pada jaringan perhatian graf yang dipandu oleh topik untuk analisis teks.
Tokenizing	:	Proses memecah suatu teks menjadi unit-unit kecil yang disebut *token*.
Unsupervised Keyphrase Extraction	:	Metode untuk mengekstraksi frasa kunci dari teks tanpa menggunakan anotasi atau pengawasan tambahan.

BAB I
PENDAHULUAN

1.1 Pendahuluan

Pada bab ini akan dijelaskan mengenai latar belakang, rumusan masalah, tujuan penulisan, manfaat penelitian, batasan penelitian, dan sistematika penulisan. Secara keseluruhan, skripsi ini menjelaskan mengenai bagaimana membangun sebuah sistem *keyphrase extraction* dengan menggunakan pra-pelatihan model bahasa (*Pre-Trained Language Model*), yakni RoBERTa dan menggunakan *Topic Guided Graph Attention Network*. Sistem ini bertujuan untuk menghasilkan informasi penting yang dibutuhkan.

1.2 Latar Belakang Masalah

Pada era informasi digital saat ini, volume teks yang besar dan konten yang dihasilkan oleh berbagai sumber seperti situs web, media sosial, artikel berita, dan dokumen akademis, menyulitkan manusia untuk memahami informasi ataupun isi konten yang disampaikan dalam sebuah teks. Di samping itu, hal ini juga menciptakan tantangan baru dalam mengelola dan mengakses informasi yang relevan. Untuk menghadapi tantangan tersebut, diperlukan suatu sistem yang dapat melakukan pencarian kata kunci secara otomatis. Dalam konteks ini, kata kunci menjadi sangat penting untuk mencari sebuah informasi yang dibutuhkan.

Menurut Li & Wang (2014), kata kunci (*keyword*) berfungsi untuk menemukan informasi yang dibutuhkan dari banyaknya informasi yang ada secara lebih cepat. Kata kunci ataupun *keywords* merupakan frasa penting atau kata tunggal yang menghubungkan fitur-fitur dokumen. Misalnya, sebuah dokumen tentang "membangun rumah" harus berisi kata kunci seperti "pondasi" atau "atap" (Abimbola, et al. 2022). Menurut Zhang, An dan Liu (2017), kata kunci memiliki peran penting untuk mempermudah pengguna dalam menemukan konten yang relevan dengan tujuan membantu pengguna untuk merangkum dan memahami topik yang dibahas oleh keseluruhan teks tersebut. Dalam menentukan kata kunci, diperlukan suatu sistem yang dapat melakukan *keyphrase extraction*.

Menurut Hasan dan Ng dalam Zhu et al. (2023), *Keyphrase Extraction* (KPE) adalah tugas *Natural Language Processing* (NLP) yang menyangkut pengekstraksian suatu kata kunci terkait topik utama yang dibahas dari sebuah dokumen. *Keyphrase Extraction* adalah langkah untuk menyelesaikan masalah seperti *Text Mining*, *Headline Generation*, dan *Automatic Essay Grading*. Proses *Keyphrase Extraction* memungkinkan pengguna untuk menentukan relevansi dokumen tanpa harus membaca seluruh isinya (Abimbola, et al. 2022).

Salah satu metode *Keyphrase Extraction* yaitu *Topic Guided Graph Attention Network*. Berdasarkan penelitian yang dilakukan oleh Zhu et al. (2023), *Topic Guided Graph Attention Network* adalah konsep yang menggabungkan dari dua area utama, yaitu GAT (*Graph Attention Networks*) berdasarkan penelitian (Velickovic, et al. 2018) dan pemahaman topik teks. *Topic Guided Graph Attention Network* dapat

mengintegrasikan informasi topik dan struktur graf dengan lebih baik, sehingga mampu meningkatkan kemampuan model untuk mengekstrak kata kunci dengan lebih kontekstual. Menurut Zhu et al. (2023), dari hasil penelitian yang didapatkan bahwa dengan mengabaikan desain pemodelan topik akan mengakibatkan penurunan kinerja yang sangat signifikan, dengan rata-rata penurunan F1@15 sebesar 5,46%. Hal ini secara langsung membuktikan pentingnya pemanfaatan informasi topik untuk tugas *Keyphrase Extraction*.

Dalam *Keyphrase Extraction* penggunaan *Pre-trained Language Model* dapat memberikan beberapa keuntungan utama seperti, pemahaman bahasa yang lebih baik, pemahaman kontekstual, dan dapat menghasilkan representasi vektor yang kaya informasi untuk setiap kata di dalam teks. Salah satu metode *Pre-trained Language Model* adalah RoBERTa (*Robustly Optimized BERT Pretraining Approach*). Menurut Azizah et al. (2023), dari hasil penelitiannya didapatkan bahwa selama proses pelatihan, RoBERTa berhasil mencapai tingkat akurasi 10% lebih tinggi daripada BERT sehingga memungkinkan sistem memperoleh hasil kinerja yang lebih baik karena menggunakan *Pre-trained Language Model* yang memiliki tingkat akurasi lebih tinggi.

Berdasarkan penjelasan di atas, maka akan dikembangkan sistem *Keyphrase Extraction* menggunakan *Pre-trained Language Model* RoBERTa dan *Topic Guided Graph Attention Networks*.

1.3 Rumusan Masalah

Berdasarkan latar belakang yang telah diuraikan, maka didapat perumusan masalah pada penelitian ini adalah

1. Bagaimana mengembangkan sistem *Keyphrase Extraction* menggunakan *Pre-Trained Language Model* RoBERTa dan *Topic Guided Graph Attention Networks*?
2. Bagaimana kinerja sistem *Keyphrase Extraction* menggunakan *Pre-Trained Language Model* RoBERTa dan *Topic Guided Graph Attention Networks* berdasarkan nilai *F-Score* dan *Accuracy*?

1.4 Tujuan Penulisan

Tujuan dari penelitian ini adalah

1. Menghasilkan sebuah sistem yang dapat melakukan *Keyphrase Extraction* dengan menggunakan metode *Pre-Trained Language Model* RoBERTa dan *Topic Guided Graph Attention Networks.*
2. Mengetahui kinerja sistem *Keyphrase Extraction* menggunakan *Pre-Trained Language Model* RoBERTa dan *Topic Guided Graph Attention Networks* berdasarkan *F-Score* dan *Accuracy*.

1.5 Manfaat Penelitian

Manfaat yang didapat dari penelitian ini adalah

1. Sistem dapat mengekstrak kata kunci pada teks berbahasa Indonesia.
2. Hasil penelitian dapat digunakan sebagai rujukan pada penelitian terkait.

1.6 Batasan Penelitian

Agar permasalahan tidak menyimpang dari batasan yang telah ditetapkan, adapun batasan dari penelitian ini adalah

1. Dokumen yang digunakan sebagai data uji adalah dokumen artikel ilmiah berbahasa Indonesia.
2. Data yang digunakan merupakan Publikasi Ilmiah dari *website* jtiik[1], jatisi[2], jepin[3] .
3. Metode RoBERTa yang digunakan adalah Indonesia RoBERTa *Base*.

1.7 Sistematika Penulisan

Sistematika penulisan yang digunakan pada penelitian ini mengikuti standar operasional penulisan tugas akhir Fakultas Ilmu Komputer Universitas Sriwijaya yakni:

BAB I. PENDAHULUAN

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

[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. TINJAUAN PUSTAKA

Bab ini menjelaskan mengenai landasan teori yang digunakan dalam menunjang penelitian. Pada bab ini dimuat mengenai literature dan penelitian terkait sebelumnya yang berkaitan dengan penelitian ini, seperti penjelasan mengenai *Pre-Trained Language Model* RoBERTa, *Topic Guided Graph Attention Network*, *Neural Topic Module* (NTM), *Anchor Aware Graph* serta penjelasan lain terkait.

BAB III. METODOLOGI PENELITIAN

Bab ini akan menjelaskan mengenai tahapan-tahapan atau proses yang dilakukan selama penelitian seperti metode pengumpulan data hingga metode dalam perancangan perangkat lunak. Setiap tahapan penelitian akan dijelaskan secara rinci sesuai dengan kerangka kerja yang telah ditetapkan.

BAB IV. PENGEMBANGAN PERANGKAT LUNAK

Bab ini akan membahas mengenai perancangan perangkat lunak mulai dari analisis kebutuhan perangkat lunak hingga pengujian pada perangkat lunak guna mengevaluasi pengembangan perangkat lunak.

BAB V. HASIL DAN ANALISIS PENELITIAN

Bab ini memaparkan hasil penelitian berdasarkan langkah dan metode yang telah direncanakan sebelumnya. Analisis tersebut diberikan sebagai dasar kesimpulan yang akan diambil dari penelitian ini.

BAB VI. KESIMPULAN DAN SARAN

Bab ini memaparkan kesimpulan dari penelitian yang dilakukan berdasarkan uraian pada bab-bab sebelumnya dan memuat saran yang diharapkan dapat membuat sistem lebih baik lagi kedepannya.

1.8 Kesimpulan

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

DAFTAR PUSTAKA

Abimbola, Rilwan O., Iyabo O. Awoyelu, Folasade O. Hunsu, Bodunde O. Akinyemi, and Ganiyu A. Aderounmu. "A Noun-Centric Keyphrase Extraction Model." *Journal of Advances in Information Technology* Vol. 13, No 6 (Desember 2022): 578-589.

Azizah, Shafna Fitria Nur, Hasan Dwi Cahyono, Sari Widya Sihwi, and Wisnu Widiarto. "Performance Analysis of Transformer Based Models (BERT, ALBERT and RoBERTa) in Fake News Detection." Agustus 2023.

Cai, Fengze, Qiang Hu, Renjie Zhou, and Neal Xiong. "REEGAT: RoBERTa Entity Embedding and Graph Attention Networks Enhanced Sentence Representation for Relation Extraction." (electronics) Mei 2023.

Chowdhury, Jishnu Ray, Cornelia Caragea, and Doina Caragea. "Keyphrase Extraction from Disaster-related Tweets." 2019: 1555–1566.

Devlin, Jacob, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding." Mei 2019.

Dhruva Sahrawat, Debanjan Mahata, Haimin Zhang, Mayank Kulkarni, Agniv Sharma, Rakesh Gosangi, Amanda Stent, Yaman Kumar, Rajiv Ratn Shah, and Roger Zimmermann. "Keyphrase Extraction as Sequence Labeling Using Contextualized Embeddings." (Springer) April 2020: 328-335.

Gulla, Jon Atle, Hans Olaf Borch, and Jon Espen Ingvaldsen. "Unsupervised Keyphrase Extraction for Search." *Natural Language Processing and IInformation Systems, 11th International Conference on Applications of Natural Language to Information Systems.* Berlin Heidelberg: Springer-Verlag , 2006. 25 – 36.

Hermawan, Latiatus, and M. B. Ismiati. "Pembelajaran Text Preprocessing berbasis." *Transformatika* 17 (2020): 177-199.

Khairunnisa, Syifa, Adiwijaya, and Said Al Faraby. "Pengaruh Text Preprocessing terhadap Analisis Sentimen Komentar Masyarakat pada Media Sosial Twitter (Studi Kasus Pandemi COVID-19)." *JURNAL MEDIA INFORMATIKA BUDIDARMA* 5 (April 2021): 406-414.

Kruchten, Philippe B. *The Rational Unified Process: An Introduction THIRD EDITION*. 3nd. Addison-Wesley Professional, 2003.

Li, Guangyi, and Houfeng Wang. "Improved Automatic Keyword Extraction Based." *Natural Language Processing and Chinese Computing*, 2014: 403-414.

Liu, Chao, et al. "A-GNN: Anchors-Aware Graph Neural Networks for Node Embedding." *Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering* (Springer) 300 (1 2020).

Liu, Yinhan, et al. "RoBERTa: A Robustly Optimized BERT Pretraining Approach." Juli 2019.

Mu, Funan, et al. "Keyphrase Extraction with Span-based Feature Representations." Februari 2020.

Nasar, Zara, Syed Wagar Jaffry, and Muhammad Kamran Malik. "Textual Keyword Esxtraction and Summarization: State-of-the-art." *Information Processing & Management*, 2019.

Nata, G. N. M., and P. P. Yudiastra. "Preprocessing Text Mining Pada Email Box Berbahasa Indonesia." *Konferensi Nasional Sistem & Informatika 2017* 1 (2017): 479-483.

Nurlayli, Akhsin, and Moch. Ari Nasichuddin. "Topic Modeling Penelitian Dosen JPTEI UNY pada Google Scholar." *ELINVO (Electronics, Informatics, and Vocational Education)*, November 2019: 154-161.

Plakasa, Gerald. "*KEYPHRASE EXTRACTION* PADA BAHASA INDONESIA MENGGUNAKAN METODE YAKE." *Skripsi* (Universitas Sriwijaya. Palembang), 2022.

Siddiqi, Sifatullah, and Aditi Sharan. "Keyword and Keyphrase Extraction Techniques: A Literature Review." *International Journal of Computer Applications*, Januari 2015.

Sivakumar, A, and R Gunasundari. "A Survey on Data Preprocessing." *International Journal of Pure and Applied Mathematics* 117 (2017): 785-794.

Togatorop, Parmonangan R, Rezky Prayitno Simanjuntak, Siti Berliana Manurung, and Mega Christy Silalahi. "PEMBANGKIT ENTITY RELATIONSHIP DIAGRAM DARI SPESIFIKASI KEBUTUHAN MENGGUNAKAN

NATURAL LANGUAGE PROCESSING UNTUK BAHASA INDONESIA." *Jurnal Komputer & Informatika*, Oktober 2021: 196-206.

Velickovic, Petar, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Lio, and Yoshua Bengio. "GRAPH ATTENTION NETWORKS." *ICLR*, Februari 2018.

Violos, J., K. Tserpes, I. Varlamis, and T. Varvarigou. "Text Classification Using the N-Gram Graph Representation Model Over High Frequency Data Streams." *Frontiers in Applied Mathematics and Statistics* 4 (2018).

Xu, Bing, Naiyan Wang, Tianqi Chen, and Mu Li. "Empirical Evaluation of Rectified Activations Convoluiton Network." November 2015.

Zhang, Xuekun, Jing An, and Wen Liu. "Research and implementation of keyword extraction algorithm based on professional background knowledge." *10th International Congress on Image and Signal Processing.* BioMedical Engineering and Informatics (CISP-BMEI), 2017.

Zhu, Xun, Yinxia Lou, Jing Zhao, Wang Gao, and Hongtao Deng. "Generative non-autoregressive unsupervised keyphrase extraction with." *Engineering Applications of Artificial Intelligence* (ELSEVIER), January 2023.