EKSTRAKSI KATA KUNCI MENGGUNAKAN *PRE-TRAINED LANGUAGE MODEL* BERT DAN *POSITION AWARE GRAPH*

Diajukan Untuk Menyusun Skripsi

Di Jurusan Teknik Informatika Fakultas Ilmu Komputer UNSRI

Oleh:

M. Farhan Ghifari

NIM: 09021282025064

JURUSAN TEKNIK INFORMATIKA

FAKULTAS ILMU KOMPUTER UNIVERSITAS SRIWIJAYA

2024

LEMBAR PENGESAHAN SKRIPSI

EKSTRAKSI KATA KUNCI MENGGUNAKAN *PRE-TRAINED LANGUAGE MODEL* BERT DAN *POSITION AWARE GRAPH*

Oleh:

M. Farhan Ghifari

NIM: 09021282025064

Inderalaya, 3 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

Pada hari selasa tanggal 19 Desember 2023 telah dilaksanakan ujian komprehensif skripsi oleh Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya

Nama : M. Farhan Ghifari

NIM : 09021282025064

Judul : Ekstraksi Kata Kunci Menggunakan *Pre-Trained Language Model* Bert Dan *Position Aware Graph*

dan dinyatakan **LULUS.**

1. Ketua

   Mastura Diana Marieska, M.T .....

   198603212018032001

2. Penguji 1

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

   198410012009121005

3. Pembimbing 1

   Novi Yusliani, M.T .....

   198211082012122001

4. Pembimbing 2

   Junia Kurniati, M.Kom .....

   1671046606890018

Mengetahui,

Ketua Jurusan Teknik Informatika

Alyi Syahrian Utami, M.Kom

197812222006042003

HALAMAN PERNYATAAN

Yang bertanda tangan di bawah ini:

Nama : M. Farhan Ghifari

NIM : 09021282025064

Program Studi : Teknik Informatika

Judul Skripsi : Ekstraksi Kata Kunci Menggunakan *Pre-Trained Language Model* Bert Dan *Position Aware Graph*

Hasil Pengecekan Software iThenticate/Turnitin : 14%

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, 13 Desember 2023

M. Farhan Ghifari

NIM.09021282025064

MOTTO DAN PERSEMBAHAN

Motto:

"Allah tidak membebani seseorang melainkan sesuai dengan kesanggupannya."

- Al Baqarah 286

Kupersembahkan Karya Tulis ini kepada:

- Allah SWT
- Kedua Orang Tua
- Fakultas Ilmu Komputer
- Universitas Sriwijaya

ABSTRACT

The number of scientific journals circulating today is very large and will continue to grow. If you want to see the words that are the subject of the journal in general, it can be represented by keywords. Determining the keywords of a scientific journal requires the ability to understand the contents of the journal in general and requires a lot of time. The solution to the problem of keyword formulation is to create an automatic keyword extraction system. One of the graph-based methods for keyword extraction is Position Aware Graph (PAG). This research will perform keyword extraction using the Pre-trained Language Model BERT and Position Aware Graph (PAG) with the help of the Pre-trained Language Model BERT. The dataset used is 100 scientific journals taken from the *jtiik, jatisi*, and *jepin* websites. Based on the research conducted, the best parameter configuration is in taking the top 15 candidates using PAG with an f1-score of 0.0089. The results obtained show that keyword extraction using BERT and PAG was successfully carried out but did not get maximum results.

Keywords: Keyphrase Extraction, BERT, PAG, Scientific Journal, *Pre-trained Language Model*

ABSTRAK

Jurnal ilmiah yang beredar saat ini jumlah sangat banyak dan akan terus berkembang. Jika ingin melihat kata-kata yang menjadi bahasan dari jurnal secara umum maka dapat di wakili oleh kata kunci. Dalam menentukan kata kunci dari sebuah jurnal ilmiah diperlukan kemampuan untuk memahami isi jurnal secara umum dan membutuhkan waktu yang tidak sedikit. Solusi dari permasalahan perumusan kata kunci adalah membuat sistem ekstraksi kata kunci secara otomatis. Salah satu metode *graph based* untuk ekstraksi kata kunci adalah *Position Aware Graph* (PAG). Penelitian ini akan melakukan ekstraksi kata kunci menggunakan *Pre-trained Language Model* BERT dan *Position Aware Graph* (PAG) dengan menggunakan bantuan dari *Pre-trained Language Model* BERT. Dataset yang digunakan sejumlah 100 jurnal ilmiah yang diambil dari website *jtiik, jatisi*, dan *jepin*. Berdasarkan penelitian yang dilakukan, konfigurasi parameter terbaik ada pada pengambilan 15 kandidat teratas menggunakan PAG dengan f1-score 0,121. Hasil yang didapatkan menunjukkan bahwa ekstraksi kata kunci menggunakan BERT dan PAG berhasil dilakukan tetapi belum mendapat hasil yang maksimal.

Kata Kunci: Ekstraksi Kata Kunci, BERT, PAG, Jurnal Ilmiah, *Pre-trained Language Model*

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, yang telah mendoakan, memberi semangat, memotivasi, dan nasihat untuk menyelesaikan skripsi ini.
3. Bapak Muhammad Qurhanul Rizqie, S.Kom,.M.T., PH.D. selaku Dosen dan sekaligus pembimbing akademik
4. Ibu Alvi Syahrini Utami, M. Kom. selaku Ketua Jurusan Teknik Informatika Universitas Sriwijaya
5. Ibu Novi Yusliani, M.T. selaku Dosen Pembimbing I dan Ibu Junia Kurniati, M.Kom selaku dosen Pembimbing II yang telah membimbing, memberikan motivasi serta arahan kepada penulis dalam proses pengerjaan skripsi.
6. Fadhil Zahran Muwafa sebagai teman yang bersedia memberikan tempat tinggal sementara pada masa skripsi
7. Seluruh dosen program studi serta admin Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya.

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 12 Desember 2023

Penulis

M. Farhan Ghifari

DAFTAR ISI

Halaman

DAFTAR TABEL

DAFTAR GAMBAR

DAFTAR ISTILAH

Embedding	:	Dalam konteks NLP, embedding kata (word embedding) adalah representasi vektor dari kata-kata dalam ruang vektor berdimensi tinggi.
Encoding	:	Proses dalam penciptaan pesan melalui kode-kode tertentu agar dapat dibaca oleh sistem
Fine Tuning	:	Proses dalam pembelajaran mesin (machine learning) di mana model yang telah dilatih sebelumnya (pretrained model) disesuaikan atau disempurnakan pada tugas tertentu.
F1-Score	:	F1-Score adalah matriks evaluasi yang menyatukan presisi (precision) dan recall (sensitivitas) dalam model klasifikasi.
Graph	:	Representasi visual dari objek dan hubungan antara objek tersebut. Dalam matematika dan ilmu komputer, graf digunakan untuk menggambarkan dan menganalisis struktur data yang terdiri dari simpul (node) yang terhubung oleh sisi (edge) yang menghubungkan simpul-simpul tersebut.
Keyphrase	:	Keyphrase, atau sering disebut juga sebagai kata kunci atau frasa kunci, adalah kata, frasa, atau rangkaian kata tertentu yang diidentifikasi sebagai representasi penting atau signifikan dalam sebuah teks atau dokumen.
Keyphrase Extraction	:	Proses dalam pemrosesan bahasa alami (Natural Language Processing, NLP) yang bertujuan untuk mengidentifikasi kata-kata atau frasa-frasa tertentu yang dianggap penting atau signifikan dalam sebuah teks atau dokumen
Np Chunking	:	Proses dalam pemrosesan bahasa alami (Natural Language Processing, NLP) yang bertujuan untuk mengidentifikasi dan mengelompokkan frasa benda (noun phrase) dalam sebuah teks atau kalimat
Pos-tagging	:	Proses dalam pemrosesan bahasa alami (Natural Language Processing, NLP) yang melibatkan penentuan jenis kata (part-of-speech) dari setiap kata dalam sebuah teks atau kalimat
Pre-train language model	:	Pretrain language model adalah jenis model pemrosesan bahasa alami (Natural Language Processing, NLP) yang telah dilatih pada volume besar data teks sebelum digunakan untuk tugas tertentu
Stopword removal	:	Salah satu langkah dalam pemrosesan teks yang bertujuan untuk menghilangkan kata-kata stop (stopwords) dari sebuah teks atau dokumen
Tokenisasi	:	Proses mengubah teks atau dokumen menjadi potongan-potongan kecil yang disebut "token.

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 ekstraksi kata kunci dengan menggunakan pra-pelatihan model bahasa (*Pre-Trained Language Models*) yakni BERT dan menggunakan *Position Aware Graph*. Sistem ini dapat digunakan dalam pengekstrakan kata kunci guna mendapatkan informasi penting yang dibutuhkan serta sesuai dengan kata kunci keseluruhan yang dibahas.

1.2 Latar Belakang Masalah

Arus informasi yang mengalir dengan sangat cepat dengan didukung oleh perkembangan teknologi informasi serta komunikasi membuat jumlah informasi menjadi semakin banyak. Pada bidang penelitian misalnya, informasi dalam bentuk jurnal-jurnal ilmiah menjadi semakin banyak. Salah satu hal yang dapat mewakili jurnal secara umum dapat terlihat dari kata kunci yang diberikan dari jurnal tersebut oleh penulis. Pencarian kata kunci secara manual dengan cara membaca jurnal dan memahaminya secara keseluruhan membutuhkan waktu yang lama dan terkadang kurang akurat.

Kata kunci sering dikaitkan dengan jurnal ilmiah dimana kata kunci dapat menggambarkan isi dari jurnal ilmiah (Mothe et al., 2018). Ekstraksi kata kunci merupakan proses mengekstrak kata-kata atau frasa penting yang dapat mewakili ide-ide utama dalam sebuah dokumen atau manuskrip (Zhang et al., 2023). Ekstraksi kata kunci otomatis merupakan proses pengidentifikasian frasa atau kata penting yang paling menggambarkan isi dari dokumen tertentu (Saxena et al., 2020). Secara umum Ekstraksi kata kunci otomatis adalah salah satu aplikasi dari bidang *Natural Language Processing* yang penerapannya berfokus pada proses pembangunan metode yang efektif dan efisien melalui pemanfaatan *keyphrase* pada dokumen.

Pendekatan yang biasa digunakan dalam *keyphrase extraction* adalah *supervised learning* dan *unsupervised learning*. *Supervised Learning* membutuhkan data terlabeli dengan jumlah banyak, dimana sering kali memiliki batasan pada domain atau dataset tertentu (Zhang et al., 2023). Berbeda dengan *supervised learning*, pendekatan *unsupervised learning* lebih fleksibel dan mudah beradaptasi dengan cara mengekstrak informasi intrinsik dari dokumen (Zhang et al., 2023).

Didalam proses ekstraksi kata kunci secara otomatis, penggunaan *pre-trained language model* menjadi peran yang penting. Dalam penelitian ini penulis menggunakan *pre-trained language model* (BERT) salah jenis dari *pre-trained language model* untuk menghasilkan representasi kontekstual yang dinamis dari kalimat dan kandidat frasa kunci (Zhang et al., 2023). BERT dapat memberikan kualitas *embedding* teks dengan kualitas tinggi dengan melakukan proses *encoding*

pada kata dan kalimat secara dinamis. Secara spesifik penulis menggunakan model Indo Bert (Wilie et al., 2020) untuk melakukan proses *embedding* karena model tersebut dengan dliatih dengan menggunakan miliaran data kosakata bahasa Indonesia dan juga sudah siap digunakan untuk penelitian.

Salah satu metode perankingan yang digunakan dalam proses *keyphrase extraction* adalah *Position Aware Graph*. *Position Aware Graph* merupakan metode berbasis *graph* yang sadar akan posisi dengan tingkat konsistensi menggunguli *Graph Neural Network* yang ada sebanyak 66% dalam skor ROC AUC (Zhang et al., 2023). Dalam penelitian yang dilakukan oleh (Zhang et al., 2023) berhasil mengungguli metode-metode ekstraksi kata kunci lain seperti UKERank (Liang et al., 2021), SIFRank (Sun et al., 2020), Embed Rank (Bennani-Smires et al., 2018) untuk dataset Inspec dengan *f1-score* pada 5 *keyphrase* 34,59, *f1-score* pada 10 *keyphrase* 40,90, dan *f1-score* 15 pada *keyphrase* 42,19.

Penelitian ini akan membangun sistem yang melakukan ekstraksi kata kunci untuk teks berbahasa Indonesia menggunakan *pre-trained language model* BERT dan *Position Aware Graph* (PAG).

1.3 Rumusan Masalah

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

1. Bagaimana mengembangkan sistem *keyphrase extraction* dengan menggunakan *Pre-Trained Language Model* BERT dan *Position Aware Graph*?

2. Bagaimana nilai F1-*score* sistem *keyphrase extraction* dengan menggunakan *Pre-Trained Language Model* BERT dan *Position Aware Graph*?

1.4 Tujuan Penulisan

Tujuan dari penelitian ini adalah sebagai berikut:

1. Menghasilkan sebuah sistem yang dapat melakukan ekstraksi kata kunci dengan menggunakan *Pre-Trained Language Model* BERT dan *Position Aware Graph*.
2. Mengetahui nilai F1-*score* sistem ekstraksi kata kunci dengan menggunakan *Pre-Trained Language Model* BERT dan *Position Aware Graph*.

1.5 Manfaat Penelitian

Manfaat yang didapat dari penelitian ini adalah sebagai berikut:

1. Perangkat lunak yang dihasilkan dapat digunakan untuk melakukan operasi *keyphrase extraction* menggunakan BERT dan *Position Aware Graph*.
2. Hasil penelitian dapat dijadikan sebagai rujukan untuk penelitian terkait.
3. Sistem dapat menghasilkan keyword secara otomatis pada teks berbahasa Indonesia.

1.6 Batasan Penelitian

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

1. Publikasi Ilmiah yang digunakan adalah publikasi berbahasa Indonesia dengan data spesifik berupa judul dan abstrak.
2. Data yang digunakan adalah 100 data Publikasi Ilmiah dari website jtiik[1], jatisi[2], jepin[3].

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.

BAB II. KAJIAN LITERATUR

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* BERT, *Position Aware Graph*, serta penjelasan lain terkait.

[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 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

Dengan uraian yang telah dijelaskan pada subbab sebelumnya, penelitian ini akan membahas mengenai ekstraksi kata kunci dengan *Pre-Trained Language Model* BERT dan *Position Aware Graph.*

DAFTAR PUSTAKA

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.

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.

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