KEYPHRASE EXTRACTION DENGAN MENGGUNAKAN *PRE-TRAINED LANGUAGE MODELS* BERT DAN *TOPIC-GUIDED GRAPH ATTENTION NETWORKS*

Diajukan Sebagai Syarat untuk Menyelesaikan
Pendidikan Program Strata-1 pada
Jurusan Teknik Informatika

Oleh :

AINI NABILAH
0902128025067

Jurusan Teknik Informatika
FAKULTAS ILMU KOMPUTER UNIVERSITAS SRIWIJAYA
2024

LEMBAR PENGESAHAN SKRIPSI

KEYPHRASE EXTRACTION* DENGAN MENGGUNAKAN *PRE-TRAINED LANGUAGE MODELS* BERT DAN *TOPIC-GUIDED GRAPH ATTENTION NETWORKS

Oleh :

AINI NABILAH
0902128025067

Palembang, 3 Januari 2024

Pembimbing 1

Novi Yusliani, M.T.
NIP. 198211082012122001

Pembimbing 2

Annisa Darmawahyuni, M.Kom.
NIP. 199006302023212044

Mengetahui,
Ketua Jurusan Teknik Informatika

Alvi Syahrini Utami, M.Kom.
NIP. 197812222006042003

TANDA LULUS UJIAN KOMPREHENSIF

Pada hari Kamis tanggal 21 Desember 2023 Telah dilaksanakan ujian komprehensif skripsi oleh Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya

Nama : Aini Nabilah

NIM : 09021282025067

Judul : *Keyphrase Extraction* dengan *Pre-Trained Language Models* BERT dan *Topic-Guided Graph Attention Networks*

Dan dinyatakan **LULUS**.

1. Ketua Penguji

   Kanda Januar Miraswan, M.T
   NIP. 199001092019031012 .....

2. Penguji

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

3. Pembimbing I

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

4. Pembimbing II

   Annisa Darmawahyuni, M.Kom.
   NIP. 199006302023212044 .....

Mengetahui,

Ketua Jurusan Teknik Informatika

Alvi Syahrini Utami, M.Kom.

NIP. 197812222006042003

HALAMAN PERNYATAAN

Yang bertanda tangan di bawah ini :

Nama : Aini Nabilah

NIM : 09021282025067

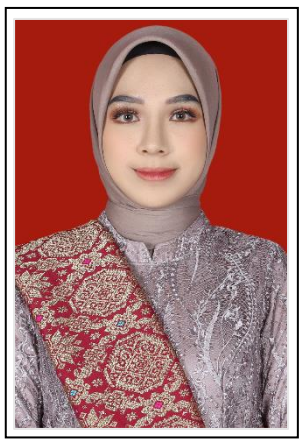
Program Studi : Teknik Informatika

Judul Skripsi : *Keyphrase Extraction* dengan *Pre-Trained Language Models* BERT dan *Topic-Guided Graph Attention Networks*

Hasil pengecekan Software Turnitin : 16%

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, 3 Januari 2024

Aini Nabilah

NIM. 09021282025067

MOTTO DAN PERSEMBAHAN

Motto:

"Waktu adalah Uang"

(Benjamin Franklin)

Tugas Akhir ini Kupersembahkan Kepada:

- Allah SWT.
- Kedua orang tua dan Keluarga saya
- Orang-Orang terdekat saya
- Fakultas Ilmu Komputer
- Universitas Srwijaya

ABSTRACT

The increase in the amount of information and documents has made it difficult for individuals to search for information that matches relevant keywords. This poses a challenge in managing and accessing the required information. One solution to address this challenge is the use of Keyphrase Extraction Systems. These systems aim to generate keyphrase that represent the content of documents, enabling users to find documents based on relevant keywords related to the topic they are searching for. This research use an approach to keyphrase extraction that combines the use of pre-trained language models such as BERT and Topic-guided Graph Attention Networks. This method allows for the capture of semantic relationships between words in documents based on their topics. Empirical studies were conducted on a dataset containing 100 accredited scientific journal publications from Sinta 2 and Sinta 3 to evaluate the system's performance. The experimental results show a precision value of 0.058, a recall value of 0.070, and an F1-score of 0.062 for the five generated keywords across the entire tested dataset.

***Keywords**: Keyphrase Extraction, Topic-Guided Graph Attention Networks, Pre-Trained Language Models* BERT

ABSTRAK

Peningkatan jumlah informasi dan dokumen telah menyebabkan individu kesulitan untuk mencari informasi yang sesuai dengan kata kunci yang relevan. Hal ini menjadi tantangan dalam mengelola dan mengakses informasi yang diperlukan. Salah satu solusi untuk mengatasi tantangan ini adalah dengan menggunakan Sistem Ekstraksi Kata Kunci. Sistem ini bertujuan untuk menghasilkan kata kunci yang merepresentasikan isi dokumen, sehingga memungkinkan pengguna untuk menemukan dokumen berdasarkan kata kunci yang relevan dengan topik yang sedang dicari. Penelitian ini menggunakan pendekatan ekstraksi kata kunci yang menggabungkan penggunaan *pre-trained language models* BERT dan *Topic-guided Graph Attention Networks*. Metode ini memungkinkan penangkapan hubungan semantik antara kata-kata dalam dokumen berdasarkan topik. Studi empiris dilakukan pada dataset berisi 100 jurnal publikasi ilmiah terakreditasi sinta 2 dan sinta 3 untuk mengevaluasi kinerja sistem. Hasil penelitian menunjukkan nilai *precision sebesar* 0.058, *recall* sebesar 0.070, dan f1-*score* sebesar 0.062 untuk lima kata kunci yang dihasilkan terhadap seluruh dataset yang diuji.

Kata Kunci : Ekstraksi Kata Kunci, *Topic-Guided Graph Attention Networks*, *Pre-Trained Language Models* BERT

KATA PENGANTAR

Puji syukur kepada Allah atas berkat dan rahmat-Nya yang telah 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 ingin menyampaikan rasa terima kasih kepada:

1. Allah SWT atas rahmat, ridho, dan karunia-Nya sehingga penulis dapat menyelesaikan tugas ini dengan baik
2. Ayahku Alfatah dan ibuku Hermawansiah yang selalu memberi motivasi, semangat, perhatian, dan tak lupa selalu mendoakan sehingga penulis selalu kuat dan selalu termotivasi untuk segera menyelesaikan tugas akhir ini.
3. Kedua saudari-ku Desi Oktariana dan Fidia Lestari, serta seluruh keluarga besarku yang selalu memberikan dukungan dan sudut pandang baru.
4. Bapak Prof. Dr. Erwin, S.Si., M.Si. selaku Dekan Fakultas Ilmu Komputer Universitas Sriwijaya dan Ibu Alvi Syahrini, M.Kom selaku Ketua Jurusan Teknik Informatika.
5. Ibu Novi Yusliani, M.T selaku dosen pembimbing I sekaligus dosen Pembimbing Akademik dan Ibu Annisa Darmawahyuni, M.Kom. selaku pembimbing II, yang telah membimbing, mengarahkan dan memberikan motivasi pada penulis dalam proses perkuliahan dan pengerjaan tugas akhir.

6. Seluruh dosen program studi serta admin Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya.
7. Muhammad Fauzan yang selalu menemani, membantu, dan memberi dorongan kepada penulis dari awal hingga selesai proses pengerjaan tugas akhir.
8. Teman – teman seperjuangan, Raihan, Kurnia, Zhafira, dan Fiolinora yang telah menemani penulis selama mengerjakan skripsi serta Kak Rifqi yang memberikan saran dan masukan mengenai perkuliahan dari semester 3 hingga penyusunan tugas akhir.
9. Pihak-pihak lain yang telah memotivasi dan memberi dukungan namun tidak dapat disebutkan satu-persatu.

Penulis menyadari dalam penyusunan Tugas Akhir ini masih terdapat banyak kekurangan disebabkan keterbatasan pengetahuan dan pengalaman, 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, 3 Januari 2023

Aini Nabilah

DAFTAR ISI

DAFTAR GAMBAR

DAFTAR TABEL

DAFTAR LAMPIRAN

BAB I
PENDAHULUAN

1.1 Pendahuluan

Bab ini akan menjelaskan 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 Model*) yakni BERT dan menggunakan *Topic-Guided Graph Attention Network*. 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

Era digital informasi berkembang dengan sangat pesat dan berdampak terhadap kehidupan, salah satunya adalah banyaknya *platform* yang menyediakan informasi dan dokumen yang dibutuhkan manusia. Hal ini memudahkan manusia untuk mencari informasi maupun dokumen yang dibutuhkan. Namun, semakin bertambahnya informasi dan dokumen tersebut membuat manusia menghabiskan banyak waktu untuk mencari informasi yang relevan berdasarkan kata kunci yang dicari sehingga menjadi tantangan dalam mengelola dan mengakses informasi. Selain itu, pemilihan kata kunci pada sebuah dokumen masih diproses secara manual sehingga bersifat kurang efektif dan tidak merepresentasikan isi dari dokumen.

Untuk menanggulangi permasalahan tersebut, dibutuhkan suatu sistem yang dapat melakukan ekstraksi kata kunci secara otomatis. Sistem *keyphrase extraction* menghasilkan suatu kata kunci yang merepresentasikan isi dari suatu dokumen dan mewakili poin-poin penting dari sebuah dokumen pencarian sehingga pengguna dapat menemukan dokumen berdasarkan kata kunci yang relevan dengan topik atau isu yang pengguna minati atau cari dan proses pencarian dokumen lebih efektif.

Kata kunci ataupun *keywords* merupakan kata-kata singkat yang dapat menggambarkan isi suatu artikel ataupun dokumen (Figueroa & Chen, 2014).Kata kunci dapat mempermudah para pembaca untuk mengetahui garis besar topik yang dibahas pada suatu dokumen. *Keyphrase Extraction (KPE)* atau ekstraksi kata kunci adalah tugas *Natural Language Processing (NLP)* yang menyangkut pengekstraksian suatu kata kunci terkait topik utama yang dibahas dari sebuah dokumen (Hasan & Ng, 2014).

Secara garis besar, *keyphrase extraction* dapat digolongkan dalam *supervised learning* dan *unsupervised learning*. *Supervised learning* membutuhkan data pelatihan berlabel yang cukup besar, sedangkan *unsupervised learning* lebih fleksibel dan mudah beradaptasi dengan mengekstraksi kata kunsi berdasarkan informasi dari dokumen itu sendiri (Zhang et al., 2023).

Penggunaan *pre-trained language model* memiliki memiliki beberapa keunggulan karena model *pre-trained language model* telah dilatih dengan kumpulan data yang besar sehingga dapat menciptakan pemahaman kontekstual data dan efisien dalam segi data (Wu et.al., 2022). *Bidirectional*

Encoder Representations from Transformers (BERT) adalah salah satu model *pre-trained language models* populer yang menggunakan *bidirectional self-attention* (Singla & Ramachandara, 2020). Berbeda dengan *pre-trained language models* lainnya yang memroses teks secara *unidirectional* (satu arah), BERT dirancang untuk memroses secara dua arah sehingga memungkinkan BERT memahami konteks kata berdasarkan kata disekitarnya (Devlin et al., 2018).

Zhu et al. (2023) dalam penelitiannya mengajukan penggunaan *anchor-aware graph* yang merupakan pemodelan graf berdasarkan *topic-guided graph attention network* untuk memperkuat cakupan frasa kunci guna menangkap konteks dari suatu dokumen. *Graph attention networks* sendiri merupakan salah satu jenis *graph neural networks* yang mempertimbangkan pentingnya setiap node tetangga dan menetapkan faktor pembobotan untuk setiap koneksi node (Veličković et al., 2017).

Dari penelitian yang dilakukan Zhu et al. pada tahun 2023 membuktikan bahwa dengan menggunakan BERT sebagai PLM memiliki hasil kinerja yang cukup baik dibandingkan tanpa menggunakan PLM karena dengan menggunakan PLM dapat memperkaya fitur semantik dari sebuah dokumen. Selain itu, penggunaan TGGAT dalam proses pengekstraksian kata kunci dapat meningkatkan kinerja sistem, hal ini dibuktikan dengan turunnya kinerja sistem sebesar 2.58% apabila TGGAT tidak digunakan. Oleh karena itu, peneliti melakukan sebuah studi mengenai keyphrase extraction dengan menggunakan *Pre-Trained Language Models* BERT 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* dengan menggunakan *Pre-Trained Language* BERT dan *Topic-Guided Graph Attention Networks* ?
2. Bagaimana kinerja sistem keyphrase extraction dengan menggunakan *Pre-Trained Language* BERT dan *Topic-Guided Graph Attention Networks* berdasarkan nilai F1-*Score?*

1.4 Tujuan Penelitian

Tujuan dari penelitian ini adalah

1. Menghasilkan sebuah sistem yang dapat melakukan ekstraksi kata kunci dengan menggunakan *Pre-Trained Language* BERT dan *Topic-Guided Graph Attention Networks*
2. Mengetahui kinerja sistem melalui nilai F1-*Score* sebuah sistem ekstraksi kata kunci dengan menggunakan *Pre-Trained Language* BERT dan *Topic-Guided Graph Attention Networks*

1.5 Manfaat Penelitian

Manfaat yang didapat dari penelitian ini adalah

1. Diharapkan dapat menjadi referensi untuk penelitian atau pengembangan terkait selanjutnya.
2. Sistem dapat digunakan untuk mengekstrak kata kunci

1.6 Batasan Penelitian

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

1. Jenis data yang digunakan merupakan dokumen berbahasa Indonesia
2. Data yang digunakan merupakan 100 data publikasi ilmiah berupa abstrak, judul, dan kata kunci.
3. Model BERT yang digunakan merupakan IndoBERT.
4. Model Neural Topic Modeling yang digunakan adalah Latent Dirichlet Allocation (LDA)

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 menguraikan mengenai latar belakang, rumusan masalah, tujuan penelitian, manfaat penelitian, batasan masalah, dan sistematika penulisan yang diterapkan dalam penyusunan laporan akhir ini.

BAB II. TINJAUAN PUSTAKA

Bab ini membahas mengenai landasan teori yang 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, Topic-Guided Graoh Attention Network, serta penjelasan lain terkait.

BAB III. METODOLOGI PENELITIAN

Bab ini akan menjelaskan mengenai tahapan-tahapan atau proses yang dilakukan selama penelitian seperti metode pengumpulan da````ta 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 menyajikan hasil penelitian yang disusun sesuai dengan langkah dan metode yang telah ditetapkan sebelumnya. Analisa 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 *Topic-Guided Graph Attention Network.*

DAFTAR PUSTAKA

Cheng, H., Liu, S., Sun, W., & Sun, Q. (2023). A Neural Topic Modeling Study Integrating SBERT and Data Augmentation. *Applied Sciences*, *13*(7), 4595. https://doi.org/10.3390/app13074595

Devika, R., Vairavasundaram, S., Mahenthar, C. S. J., Varadarajan, V., & Kotecha, K. (2021). A Deep Learning Model Based on BERT and Sentence Transformer for Semantic Keyphrase Extraction on Big Social Data. *IEEE Access*, *9*, 165252–165261. https://doi.org/10.1109/ACCESS.2021.3133651

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2018). *BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding*.

Engineering Applications of Artificial Intelligence, *120*, 105934. https://doi.org/10.1016/j.engappai.2023.105934

Figueroa, G., & Chen, Y.-S. (2014). *Collaborative Ranking between Supervised and Unsupervised Approaches for Keyphrase Extraction*.

Fitria, A., & Widowati, H. (2017). Implementasi Metode Rational Unified Process Dalam Pengembangan Sistem Administrasi Kependudukan. Jurnal Ilmiah Teknologi Dan Rekayasa, *22*.

Haddi, E., Liu, X., & Shi, Y. (2013). *The Role of Text Pre-processing in Sentiment Analysis*. *Procedia Computer Science*, *17*, 26–32. https://doi.org/10.1016/j.procs.2013.05.005

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

Hidayat, I. R., & Maharani, W. (2022). *General Depression Detection Analysis Using IndoBERT Method*. *International Journal on Information and Communication Technology (IJoICT)*, *8*(1), 41–51. https://doi.org/10.21108/ijoict.v8i1.634

Kaur, K., & Gupta, V. (2011). *Keyword Extraction For Punjabi Laguage*. *Indian Journal of Computer Science and Engineering*, *2*.

Koto, F., Rahimi, A., Lau, J. H., & Baldwin, T. (2020). *IndoLEM and IndoBERT: A Benchmark Dataset and Pre-trained Language Model for Indonesian NLP*.

Kondeti, B., Jyothirani,S.A.,& Haragopal (2022). *Keyword Extraction – Comparison of Latent Dirichlet Allocation and Latent Semantic Analysis*. EJ-*MATH, European Journal of Mathematics and Statistics*.

http://dx.doi.org/10.24018/ejmath.2022.3.3.119

Latief, M., Kandowangko, N., & Yusuf, R. (2017). Pengembangan Sistem Informasi Tanaman Obat Daerah Gorontalo Berbasis Web dan Mobile. *Jurnal Rekayasa Elektrika*, *13*(3), 152. https://doi.org/10.17529/jre.v13i3.8532

Liu, C., Li, X., Zhao, D., Guo, S., Xiaojun, K., Dong, L., & Yao, H. (2020). *A-GNN: Anchors-Aware Graph Neural Networks for Node Embedding* (pp. 141–153). https://doi.org/10.1007/978-3-030-38819-5_9

Liu, R., Lin, Z., & Wang, W. (2020). *Keyphrase Prediction With Pre-trained Language Model*.

Menaka S, & Radha N. (2013). An Overview of Techniques Used for Extracting Keywords from Documents. *International Journal of Computer Trends and Technology*, *4*(7). http://www.ijcttjournal.org

Muttaqin, Firdaus. A., dan Bachtiar, Adam. M., Implementasi Teks Mining pada Aplikasi Pengawasan Penggunaan Internet Anak "Dodo Kids Browser".

Normawati, Dwi & Prayogi, Surya ALLIT (2021). Implementasi Naïve Bayes Classifier Dan Confusion Matrix Pada Analisis Sentimen Berbasis Teks Pada Twitter. Jurnal Sains Komputer & Informatika (J-SAKTI). ISSN ISSN: 2548-9771/EISSN: 2549-7200

Rahayu, W. I., Prianto, C., & Novia, E. A. (2021). Perbandingan Algoritma K-Means Dan Naïve Bayes Untuk Memprediksi Prioritas Pembayaran Tagihan Rumah Sakit Berdasarkan Tingkat Kepentingan Pada Pt. Pertamina (Persero). *Jurnal Teknik Informatika*, *13*.

Sari, D. P., & Purwarianti, A. (2014). Ekstraksi Kata Kunci Otomatis Untuk Dokumen Bahasa Indonesia Studi Kasus: Artikel Jurnal Ilmiah Koleksi Pdii Lipi. BACA: Jurnal Dokumentasi dan Informasi, 35(2), 139-147.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2017). *Attention Is All You Need*.

Veličković, P., Cucurull, G., Casanova, A., Romero, A., Liò, P., & Bengio, Y. (2017). *Graph Attention Networks*.

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

Wang, H., Li, J., Wu, H., Hovy, E., & Sun, Y. (2022). Pre-Trained Language Models and Their Applications. *Engineering*. https://doi.org/ https://doi.org/10.1016/j.eng.2022.04.024

Wardani, P. K. (2017). Penerapan Metode Rational Unified Process pada Aplikasi Monitoring Periodic Service Alat Berat. *Indonesian Journal of Applied Informatics*, *1*(2), 1. https://doi.org/10.20961/ijai.v1i2.11002

Wu, D., Ahmad, W.U., & Chang, K,-W.(2022). Pre-Trained Language Models for Keyphrase Generation: A Thorough Empirical Study. University of California, Los Angeles. arXiv:2212.10233v1[cs.CL].

Wu, J., Yao, M., Wu, D., Chi, M., Wang, B., Wu, R., Fu, X., Meng, C., & Wang, W. (2023). *DEDGAT: Dual Embedding of Directed Graph Attention Networks for Detecting Financial Risk*.

Ying, Y., Qingping, T., Qinzheng, X., Ping, Z., & Panpan, L. (2017). A Graph-based Approach of Automatic Keyphrase Extraction. *Procedia Computer Science*, *107*, 248–255. https://doi.org/10.1016/j.procs.2017.03.087

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

Zhu, X., Lou, Y., Zhao, J., Gao, W., & Deng, H. (2023). Generative non-autoregressive unsupervised keyphrase extraction with neural topic modeling.