

# Komparasi Transformer T5 dan BART+BERT Untuk Pembangkitan Kata Kunci Otomatis Dari Abstrak Ilmiah

Johan\*, Erwin Panjaitan

Informatika, Magister Teknologi Informasi, Universitas Mikroskil, Medan, Indonesia

Email: <sup>1</sup>johantento12@gmail.com, <sup>2</sup>erwin@mikroskil.ac.id

Email Penulis Korespondensi: johantento12@gmail.com\*

Submitted: 06/11/2025; Accepted: 16/12/2025; Published: 31/12/2025

**Abstrak**—Peningkatan volume publikasi ilmiah menuntut sistem pengindeksan yang efisien, di mana kata kunci memegang peran krusial. Penelitian ini membandingkan efektivitas dua pendekatan berbasis model transformer—T5 dan kombinasi BART+BERT—dalam membangkitkan kata kunci secara otomatis dari abstrak artikel ilmiah. Eksperimen dilakukan menggunakan 100 artikel jurnal multidisipliner dengan evaluasi berbasis ROUGE (kesamaan leksikal) dan Semantic Similarity (kesamaan makna). Hasil menunjukkan bahwa T5 unggul dalam menghasilkan kata kunci eksplisit yang presisi dan cocok untuk pengindeksan teknis, dengan skor Semantic F1 0,5233 terhadap ground truth. Sebaliknya, BART lebih efektif menangkap tema keseluruhan abstrak, menghasilkan kata kunci yang representatif secara semantik dengan Semantic F1 0,7517, sehingga ideal untuk sistem rekomendasi atau klasifikasi topik. Filter BERT terbukti meningkatkan kualitas semantik, sementara Jaccard lebih baik dalam pencocokan leksikal. Temuan ini menekankan pentingnya pemilihan model dan strategi filtering yang disesuaikan dengan tujuan aplikasi sistem pembangkitan kata kunci otomatis.

**Kata Kunci:** Pembangkitan Kata Kunci Otomatis; Model Transformer; T5; BART; BERT; Abstrak Ilmiah

**Abstract**—The increasing volume of scientific publications demands efficient indexing systems, where keywords play a crucial role. This study compares the effectiveness of two transformer-based approaches—T5 and the BART+BERT combination—for automatic keyword generation from scientific article abstracts. Experiments were conducted using 100 multidisciplinary journal articles, evaluated based on ROUGE (lexical similarity) and Semantic Similarity (meaning similarity). Results show that T5 excels in generating precise, explicit keywords suitable for technical indexing, with a Semantic F1-score of 0.5233 against the ground truth. In contrast, BART is more effective at capturing the overall theme of the abstract, producing semantically representative keywords with a Semantic F1-score of 0.7517, making it ideal for recommendation systems or topic classification. The BERT filter proved to enhance semantic quality, while Jaccard was better for lexical matching. These findings emphasize the importance of selecting models and filtering strategies tailored to the intended application of automatic keyword generation systems.

**Keywords:** Automatic Keyword Generation; Transformer Models; T5; BART; BERT; Scientific Abstracts

## 1. PENDAHULUAN

Lonjakan volume publikasi ilmiah dalam beberapa tahun terakhir telah menciptakan tantangan signifikan dalam proses temu balik informasi dan pengelolaan metadata di repositori digital [1]. Setiap harinya, ribuan artikel dari berbagai disiplin ilmu diterbitkan, membanjiri sistem pencarian akademik dengan informasi yang semakin kompleks dan beragam. Dalam ekosistem akademik yang padat ini, kata kunci (*keywords*) memainkan peran yang tidak dapat dianggap remeh—mereka berfungsi sebagai pintu gerbang utama untuk memahami inti sebuah artikel secara cepat dan efisien, sekaligus menjadi komponen krusial untuk pengindeksan yang akurat di mesin pencari akademik seperti *Scopus* atau *Google Scholar* [2], [3]. Kata kunci yang tepat dan representatif tidak hanya membantu peneliti menemukan literatur yang relevan dengan kebutuhan mereka, tetapi juga sangat menentukan visibilitas dan dampak dari sebuah publikasi ilmiah.

Pendekatan otomatis tradisional seperti TF-IDF (*Term Frequency-Inverse Document Frequency*) dan RAKE (*Rapid Automatic Keyword Extraction*) telah lama digunakan untuk mengatasi keterbatasan metode manual [4], [5]. Namun, metode-metode ini memiliki keterbatasan mendasar karena hanya berfokus pada frekuensi kemunculan kata atau pola statistik lainnya tanpa mampu memahami konteks atau makna semantik yang lebih dalam dari sebuah teks. Akibatnya, metode statistik tradisional seringkali gagal menangkap istilah-istilah penting yang bersifat implisit atau tidak muncul secara eksplisit dalam abstrak [6]. Padahal, dalam banyak kasus, terutama dalam publikasi ilmiah yang kompleks, kata kunci yang paling penting justru tidak selalu dinyatakan secara langsung dalam teks, melainkan tersirat dalam pembahasan dan kesimpulan penelitian.

Kemajuan revolusioner dalam bidang Pemrosesan Bahasa Alami (*Natural Language Processing/NLP*), khususnya dengan kemunculan model berbasis arsitektur transformer, telah membuka peluang baru yang menjanjikan untuk mengatasi keterbatasan metode tradisional [7]. Model-model transformer seperti T5 (*Text-to-Text Transfer Transformer*) [8], BART (*Bidirectional and Auto-Regressive Transformers*) [9], dan BERT (*Bidirectional Encoder Representations from Transformers*) [10] menawarkan kemampuan yang jauh lebih unggul dalam menangkap konteks dan hubungan semantik antar kata dalam sebuah dokumen. T5 dikenal dengan pendekatan text-to-text-nya yang fleksibel, yang menyatukan berbagai tugas NLP dalam satu kerangka kerja terpadu. Sementara itu, BART unggul dalam tugas generatif seperti peringkasan teks berkat arsitektur *encoder-decoder*-nya yang menggabungkan pemahaman *bidirectional* dan generasi *autoregressive*. Di sisi lain, BERT,

dengan pemahaman *bidirectional*-nya yang mendalam, telah terbukti sangat efektif sebagai penyaring (*filter*) semantik untuk mengevaluasi relevansi kata kunci terhadap konten dokumen.

Beberapa penelitian terdahulu telah mengeksplorasi model-model transformer secara terpisah dalam berbagai konteks ekstraksi informasi. Raffel et al. memperkenalkan T5 sebagai kerangka kerja serbaguna untuk berbagai tugas NLP [8], yang kemudian dievaluasi lebih lanjut oleh Mastropaolo et al. pada tugas berbasis kode perangkat lunak [11]. Di sisi lain, Lewis et al. mendemonstrasikan keefektifan BART dalam tugas generasi teks [9], yang didukung oleh studi Hartawan et al. yang membuktikan kemampuannya menghasilkan ringkasan koheren dalam bahasa Indonesia [12]. Penelitian yang lebih spesifik oleh Pezik et al. mengaplikasikan T5 untuk pembangkitan kata kunci dari abstrak ilmiah [13], sementara Babayigit & Sattuf memanfaatkan BERT untuk penyaringan semantik kata kunci [14]. Studi oleh Nugroho et al. mengkaji aspek efisiensi BERT [15], dan Glazkova & Morozov mengeksplorasi ketangguhan model generatif untuk kata kunci pada domain yang berbeda [16], [17]. Meskipun demikian, pendekatan statistik tradisional seperti TF-IDF dan RAKE yang dikaji oleh Gries masih menjadi pembanding penting, meski memiliki keterbatasan dalam memahami konteks semantik. Namun, celah penelitian yang masih mencolok adalah belum adanya perbandingan langsung yang komprehensif antara pendekatan generatif murni (T5) dengan pendekatan hybrid yang menggabungkan generator dan penyaring semantik (BART+BERT) dalam konteks pembangkitan kata kunci dari abstrak ilmiah multidisipliner.

Namun, celah penelitian (*research gap*) yang masih mencolok adalah belum adanya perbandingan langsung yang komprehensif antara pendekatan generatif murni (T5) dengan pendekatan *hybrid* yang menggabungkan *generator* dan penyaring semantik (BART+BERT) dalam konteks pembangkitan kata kunci dari abstrak ilmiah multidisipliner. Sebagian besar penelitian fokus pada evaluasi model tunggal atau kombinasi yang terbatas, tanpa mempertimbangkan perbandingan menyeluruh yang mencakup aspek leksikal dan semantik secara bersamaan. Selain itu, analisis mengenai efektivitas metode filtering yang berbeda (seperti *Jaccard Similarity* yang berbasis leksikal versus *BERT semantic similarity*) dalam menyaring hasil keluaran model generatif juga masih terbatas. Pemahaman yang mendalam tentang interaksi antara model generatif dan metode *filtering* ini sangat penting untuk mengoptimalkan kualitas kata kunci yang dihasilkan.

Berdasarkan identifikasi celah penelitian tersebut, penelitian ini bertujuan untuk secara empiris membandingkan kinerja model T5 dan kombinasi BART+BERT dalam tugas pembangkitan kata kunci otomatis dari abstrak artikel penelitian. Penelitian ini juga menganalisis pengaruh pemilihan metode *filtering* (Jaccard vs. BERT) terhadap kualitas hasil akhir, serta mengevaluasi konsistensi performa model pada berbagai topik ilmiah. Diharapkan, temuan dari penelitian ini dapat menjadi dasar pertimbangan yang kuat dalam pengembangan sistem pengindeksan otomatis yang lebih cerdas, akurat, dan adaptif untuk repositori ilmiah di tingkat nasional maupun internasional.

## 2. METODOLOGI PENELITIAN

### 2.1 Tahapan Penelitian

Penelitian ini dilakukan melalui empat tahap utama yang disusun secara sistematis sebagai berikut:

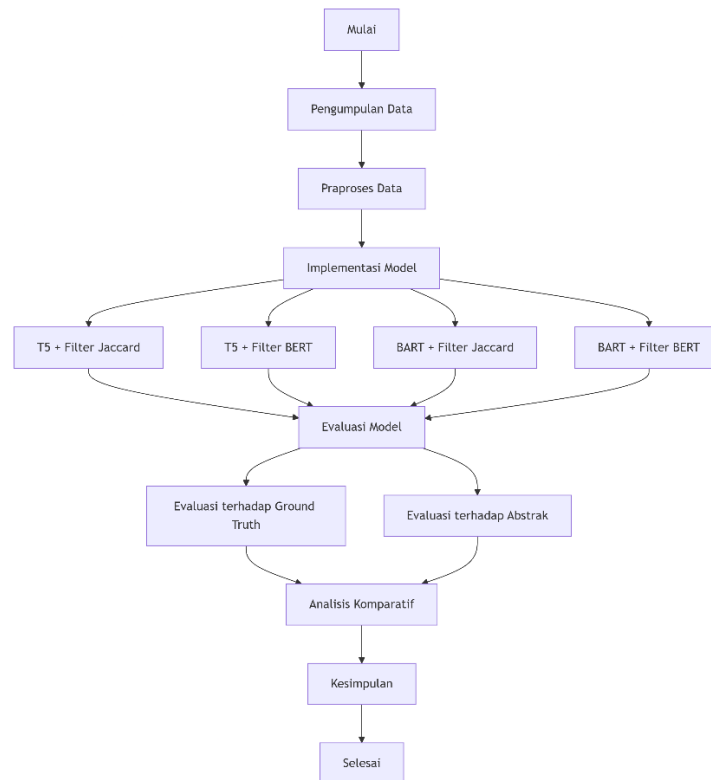
1. Pengumpulan dan Praproses Data.  
Tahap pertama berfokus pada persiapan dataset sebagai fondasi penelitian.
  - a. Mengumpulkan 100 artikel ilmiah dari jurnal bereputasi (*Springer, Elsevier, IEEE*).
  - b. Mengambil abstrak dan kata kunci asli sebagai ground truth.
  - c. Melakukan praproses data: pembersihan teks, normalisasi huruf, tokenisasi kata kunci, penyimpanan data terstruktur dalam format Excel.
 Tahapan ini memastikan bahwa data yang digunakan konsisten dan siap diproses oleh model.
2. Penerapan Metode (Implementasi Empat Pendekatan).  
Pada tahap ini, model dan teknik filtering diterapkan untuk menghasilkan kata kunci. Ada empat metode kombinasi yang diuji:
  - a. T5 + Jaccard
  - b. T5 + BERT
  - c. BART + Jaccard
  - d. BART + BERT
 Seluruh pipeline diimplementasikan menggunakan *HuggingFace Transformers* dan *Sentence-Transformers*, dengan teknik beam search untuk menghasilkan kandidat kata kunci.
3. Pengujian dan Evaluasi Model  
Tahap ini bertujuan menilai performa keempat pendekatan berdasarkan dua perspektif evaluasi:
  - a. Evaluasi terhadap *Ground Truth*  
Mengukur sejauh mana kata kunci prediksi mirip dengan kata kunci asli penulis.
  - b. Evaluasi terhadap Abstrak  
Mengukur apakah kata kunci prediksi mewakili topik dan isi abstrak secara keseluruhan.
 Kedua perspektif evaluasi ini memberikan gambaran menyeluruh: kesesuaian literal vs. kesesuaian makna terhadap dokumen.

#### 4. Analisis Komparatif

Tahap akhir membandingkan hasil evaluasi keempat metode:

- Menentukan model mana yang unggul secara leksikal, mana yang unggul secara semantik.
- Menilai peran *filter* Jaccard vs BERT.
- Menarik implikasi praktis mengenai pemilihan model sesuai tujuan aplikasi.

Diagram alur penelitian disajikan pada Gambar 1 untuk memperjelas hubungan dan urutan setiap tahap.



**Gambar 1.** Diagram Alur Penelitian

## 2.2 Tahapan Pengumpulan Data

Tahap pengumpulan data merupakan fondasi utama dalam penelitian ini. Dataset yang digunakan terdiri dari 100 artikel ilmiah yang diambil secara manual dari jurnal-jurnal internasional bereputasi di penerbit ternama seperti *Springer*, *IEEE*, dan *Elsevier*. Pemilihan sumber ini dilakukan untuk memastikan kualitas dan keakuratan konten, termasuk abstrak dan kata kunci yang disediakan oleh penulis atau editor.

Setiap entri dalam dataset mewakili satu artikel yang lengkap dengan dua komponen utama:

- Abstrak (Abstract)  
Teks abstrak dengan panjang bervariasi antara 150 hingga 250 kata, yang merepresentasikan ringkasan inti dari artikel.
- Kata Kunci Referensi (*Ground Truth Keywords*)  
Daftar kata kunci asli yang terdiri dari 3 hingga 7 kata atau frasa, yang ditetapkan oleh penulis artikel aslinya. Kata kunci inilah yang menjadi acuan kebenaran (*ground truth*) dalam proses evaluasi model. Untuk memastikan data siap diproses oleh model, dilakukan serangkaian tahap praproses yang ketat. Tahapan ini meliputi:
  - Pembersihan Teks  
Menghilangkan karakter non-alfabet, simbol khusus, dan format yang tidak konsisten [18].
  - Normalisasi  
Melakukan *lowercasing* (penyeragaman huruf kecil) pada seluruh teks untuk menghindari duplikasi [18].
  - Tokenisasi Kata Kunci  
Memisahkan string kata kunci yang digabung dengan koma atau titik koma [18] menjadi daftar kata kunci individual.

Setelah melalui proses praproses, dataset disimpan dalam format terstruktur (file Excel) dengan kolom *processed\_abstract* dan *processed\_keywords*. Contoh sampel dataset yang telah dipraproses dapat dilihat pada Tabel 1 di bawah ini.

**Tabel 1.** Contoh Sampel Dataset yang Telah Dipraproses

Abstrak (Potongan)	Kata Kunci Referensi
"The synergy between deep learning and meta-heuristic algorithms presents a promising avenue for tackling the complexities of energy-related modeling and forecasting tasks..."	['meta-heuristics', 'deep learning', 'energy applications', 'renewable energy']
"Data science and material informatics are gaining traction in alloy design. This is due to increasing infrastructure, computational capabilities and established open-source databases..."	['alloy development', 'computational alloy design', 'machine learning', 'material informatics']

Dengan dataset yang telah dikurasi dan dipraproses ini, penelitian dapat dilanjutkan ke tahap penerapan metode untuk membangkitkan kata kunci secara otomatis.

### 2.3 Tahapan Penerapan Metode

Pada tahap ini, dilakukan implementasi dan perbandingan empat pendekatan berbeda yang menggabungkan model generatif (T5 dan BART) dengan metode penyaringan (*Jaccard Similarity* dan *BERT Semantic Similarity*). Keempat pendekatan tersebut adalah:

1. T5 + Jaccard

Model T5 yang telah dilatih sebelumnya (*pretrained*) khusus untuk tugas ekstraksi frasa kunci ("ml6team/keyphrase-generation-t5-small-openkp") digunakan sebagai *generator*. Kata kunci mentah yang dihasilkan kemudian disaring dengan menghitung *Jaccard Similarity* antara setiap kata kunci dan teks abstrak. Kata kunci dengan skor tertinggi (top-5) yang memiliki tumpang tindih leksikal terbesar dengan abstrak dipilih sebagai output akhir.

2. T5 + BERT

Model T5 yang sama digunakan sebagai *generator*. Namun, penyaringan dilakukan secara semantik menggunakan model *Sentence-BERT* ('all-MiniLM-L6-v2'). Kata kunci dan abstrak diubah menjadi vektor embedding, kemudian dihitung cosine similarity di antara keduanya. Kata kunci dengan skor kesamaan semantik tertinggi ( $\geq 0.7$ ) terhadap abstrak yang dipilih.

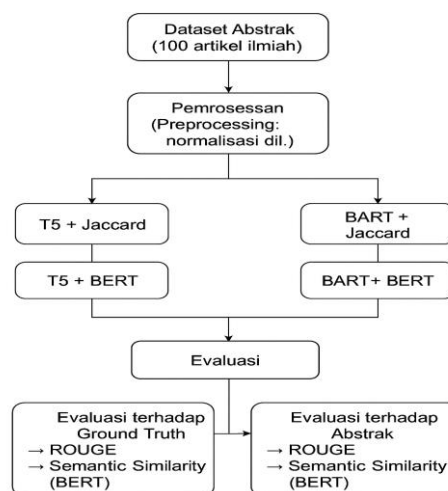
3. BART + Jaccard

Model BART dasar ("facebook/bart-base") berfungsi sebagai *generator* kata kunci. Output mentahnya kemudian disaring menggunakan metode *Jaccard Similarity* (mirip dengan pendekatan 1) untuk mempertahankan kata kunci yang secara leksikal paling mirip dengan isi abstrak.

4. BART + BERT

Model BART digunakan sebagai *generator*, yang kemudian dikombinasikan dengan penyaring semantik BERT. Ini merupakan pendekatan *hybrid* yang memadukan kemampuan generatif BART dengan pemahaman kontekstual BERT untuk memilih kata kunci yang paling relevan secara makna.

Pipeline dari keempat pendekatan ini diimplementasikan menggunakan pustaka *Transformers* dan *Sentence-Transformers* dari *Hugging Face*. Proses generasi menggunakan teknik *beam search* untuk mengeksplorasi berbagai kemungkinan urutan kata kunci yang optimal. Alur penerapan keempat metode ini dapat dilihat pada Gambar 2.

**Gambar 2.** Diagram Alur Penerapan Keempat Metode

Dengan menerapkan keempat variasi metode ini, penelitian ini bertujuan untuk mengisolasi dan menganalisis kontribusi dari masing-masing komponen (jenis model dan jenis *filter*) terhadap kualitas kata kunci akhir.

## 2.4 Tahapan Evaluasi Model

Tahap evaluasi merupakan proses kritis untuk menilai kualitas dan efektivitas kata kunci yang dihasilkan oleh keempat pendekatan. Evaluasi dilakukan dengan dua perspektif yang berbeda untuk mendapatkan analisis yang komprehensif:

### 1. Evaluasi terhadap *Ground Truth* (Kata Kunci Referensi)

Perspektif ini mengukur sejauh mana kata kunci prediksi cocok dengan kata kunci yang ditetapkan oleh penulis asli artikel. Evaluasi ini penting untuk menilai akurasi model dalam mereplikasi pilihan pakar. Metrik yang digunakan adalah:

#### a. ROUGE (*Recall-Oriented Understudy for Gisting Evaluation*)

Metrik ini mengukur kesamaan leksikal (tumpang tindih kata) [19] antara kata kunci prediksi dan ground truth.

#### b. *Semantic Similarity*

Untuk mengatasi keterbatasan ROUGE yang hanya melihat kesamaan literal, metrik ini mengukur kedekatan makna menggunakan *cosine similarity* [20] antara vektor embedding dari kata kunci prediksi dan ground truth. *Vector embedding* dihasilkan oleh model Sentence-BERT ('all-MiniLM-L6-v2').

### 2. Evaluasi terhadap Abstrak

Perspektif ini menilai sejauh mana kata kunci prediksi merepresentasikan keseluruhan isi dan tema dari abstrak. Ini berguna untuk menilai kegunaan kata kunci dalam sistem seperti pencarian semantik atau rekomendasi. Metrik yang digunakan serupa dengan perspektif pertama:

#### a. ROUGE

Menghitung kesamaan leksikal antara kata kunci prediksi dan teks abstrak lengkap.

#### b. *Semantic Similarity*

Menghitung kedekatan semantik antara kata kunci prediksi dan teks abstrak lengkap.

Dengan dua perspektif evaluasi ini, performa setiap model dapat dianalisis baik dari segi kesesuaian literal (ROUGE) maupun relevansi makna (*Semantic Similarity*). Hasil dari perhitungan metrik-metrik ini kemudian dirangkum dalam tabel perbandingan untuk dianalisis lebih lanjut.

## 3. HASIL DAN PEMBAHASAN

### 3.1 Hasil

Eksperimen komparatif terhadap empat pendekatan pembangkitan kata kunci otomatis menghasilkan data kinerja yang komprehensif. Analisis dilakukan berdasarkan dua perspektif evaluasi yang berbeda untuk memberikan gambaran menyeluruh tentang karakteristik masing-masing model.

#### 3.1.1 Evaluasi terhadap *Ground Truth* (Kata Kunci Referensi)

Evaluasi terhadap ground truth mengukur kemampuan model dalam mereplikasi kata kunci yang ditetapkan oleh penulis asli artikel. Hasil detailnya disajikan dalam Tabel 2.

**Tabel 2.** Hasil Evaluasi Berdasarkan *Ground Truth* (Kata Kunci Referensi)

Model + <i>Evaluator</i>	ROUGE <i>Precision</i>	ROUGE <i>Recall</i>	ROUGE <i>F1</i>	<i>Semantic</i> <i>Precision</i>	<i>Semantic</i> <i>Recall</i>	<i>Semantic</i> <i>F1</i>
T5 + Jaccard	0.2510	0.2658	0.2479	0.5558	0.4978	0.5202
T5 + BERT	0.2573	0.2714	0.2533	0.5597	0.5003	0.5233
BART + Jaccard	0.0492	0.4629	0.0879	0.5351	0.3271	0.3996
BART + BERT	0.0493	0.4637	0.0881	0.5351	0.3271	0.3996

Dari Tabel 2, dapat diidentifikasi beberapa pola penting:

#### 1. Dari Aspek Leksikal (ROUGE):

- T5 menunjukkan performa yang stabil dengan ROUGE F1 sekitar 0.25, dengan *precision* dan *recall* yang seimbang.
- BART memiliki ROUGE *Precision* yang sangat rendah (~0.049) namun ROUGE *Recall* yang tinggi (~0.46), mengindikasikan kecenderungan menghasilkan banyak kata kunci dengan akurasi leksikal rendah.
- Selisih ROUGE *F1* antara T5 dan BART mencapai 0.16, menunjukkan keunggulan signifikan T5 dalam kesesuaian leksikal.

#### 2. Dari Aspek Semantik:

- T5 + BERT mencapai *Semantic F1* tertinggi (0.5233), diikuti oleh T5 + Jaccard (0.5202).
- BART konsisten pada *Semantic F1* 0.3996 untuk kedua *filter*.



- c. Peningkatan *Semantic F1* dibandingkan ROUGE *F1* pada semua model (rata-rata +0.27) mengkonfirmasi bahwa model mampu menangkap makna meski bentuk katanya berbeda.

Berdasarkan Tabel 2, dapat disimpulkan bahwa model T5 secara konsisten unggul dalam menghasilkan kata kunci yang secara leksikal dan semantik lebih mendekati kata kunci pilihan penulis (*ground truth*) dibandingkan dengan BART.

Dari sisi kesamaan leksikal (ROUGE), model T5 mencapai ROUGE *F1* sekitar 0.25, yang secara signifikan lebih tinggi daripada BART yang hanya mencapai sekitar 0.088. Nilai ROUGE *Precision* dan *Recall* yang seimbang pada T5 (masing-masing ~0.25) menunjukkan kemampuannya yang baik dalam menghasilkan kata kunci yang tidak hanya tepat tetapi juga mencakup sebagian dari kata kunci referensi. Sebaliknya, BART menunjukkan ROUGE *Precision* yang sangat rendah (~0.049) meskipun *Recall*-nya tinggi (~0.46), yang mengindikasikan bahwa BART menghasilkan terlalu banyak kata kunci yang tidak sesuai secara leksikal dengan *ground truth*, meski mampu menangkap beberapa konsep yang relevan.

Dari sisi kesamaan semantik, model T5 kembali unggul dengan *Semantic F1* sekitar 0.52, dibandingkan BART yang mencapai 0.40. Yang menarik, kedua model menunjukkan peningkatan yang signifikan dalam *Semantic F1* dibandingkan ROUGE *F1*, mengkonfirmasi bahwa meskipun kata kunci yang dihasilkan tidak identik secara leksikal dengan referensi, maknanya tetap relevan. Penggunaan *filter* BERT memberikan peningkatan marginal (~0.003) pada *Semantic F1* model T5, sementara pada BART tidak memberikan perbedaan yang berarti.

### 3.1.2 Evaluasi terhadap Abstrak

Evaluasi terhadap abstrak mengukur kemampuan model dalam menghasilkan kata kunci yang merepresentasikan keseluruhan konten dokumen. Hasil detailnya disajikan dalam Tabel 3.

**Tabel 3.** Hasil Evaluasi Berdasarkan Abstrak

Model + <i>Evaluator</i>	ROUGE <i>Precision</i>	ROUGE <i>Recall</i>	ROUGE <i>F1</i>	<i>Semantic</i> <i>Precision</i>	<i>Semantic</i> <i>Recall</i>	<i>Semantic</i> <i>F1</i>
T5 + Jaccard	0.2510	0.2658	0.2479	0.5558	0.4978	0.5202
T5 + BERT	0.2573	0.2714	0.2533	0.5597	0.5003	0.5233
BART + Jaccard	0.0492	0.4629	0.0879	0.5351	0.3271	0.3996
BART + BERT	0.0493	0.4637	0.0881	0.5351	0.3271	0.3996

Analisis Tabel 3 menangkap dinamika yang berbeda:

1. Dari Aspek Leksikal (ROUGE):
  - a. T5 memiliki ROUGE *Precision* sangat tinggi (>0.69) namun ROUGE *Recall* sangat rendah (<0.047), menghasilkan ROUGE *F1* yang minimal (~0.08).
  - b. BART menunjukkan keseimbangan yang lebih baik dengan ROUGE *F1* 0.44, didukung oleh *precision* (0.63) dan *recall* (0.35) yang moderat.
  - c. Selisih ROUGE *F1* antara BART dan T5 mencapai 0.36, menunjukkan keunggulan absolut BART dalam cakupan leksikal terhadap abstrak.
2. Dari Aspek Semantik:
  - a. BART mencapai *Semantic F1* 0.7517 yang sangat tinggi dan konsisten untuk kedua filter.
  - b. T5 hanya mencapai *Semantic F1* ~0.37, kurang dari separuh pencapaian BART.
  - c. Nilai *precision*, *recall*, dan *F1* yang identik pada evaluasi semantik mengindikasikan konsistensi model dalam menghasilkan kata kunci yang relevan.

Tabel 3 menunjukkan pola yang sangat berbeda. Ketika dievaluasi berdasarkan keseluruhan isi abstrak, model BART jauh lebih unggul dalam menghasilkan kata kunci yang representatif.

Dari sisi kesamaan leksikal dengan abstrak, BART mencapai ROUGE *F1* 0.44, sementara T5 hanya 0.08. Nilai ROUGE *Precision* T5 yang sangat tinggi (~0.73) menunjukkan bahwa kata kunci yang dihasilkannya memang banyak yang muncul secara literal dalam abstrak, namun ROUGE *Recall* yang sangat rendah (~0.04) mengungkapkan bahwa cakupannya sangat terbatas—T5 gagal menangkap sebagian besar konsep penting dalam abstrak.

Dari sisi kesamaan semantik dengan abstrak, keunggulan BART semakin tampak dengan *Semantic F1* yang mencapai 0.75, hampir dua kali lipat dari T5 yang hanya 0.37. Nilai *Semantic Precision*, *Recall*, dan *F1* yang identik dan tinggi pada BART mengindikasikan konsistensinya dalam menghasilkan kata kunci yang tidak hanya relevan secara semantik tetapi juga mencakup tema utama abstrak secara menyeluruh.

Perbedaan performa yang kontras ini mengindikasikan bahwa setiap model memiliki karakteristik dan kekuatan yang berbeda, yang sangat dipengaruhi oleh tujuan evaluasinya.

### 3.1.3 Analisis Komparatif Metode Filtering

Berdasarkan kedua tabel evaluasi, pengaruh metode *filtering* dapat dianalisis:

1. *Filter* BERT memberikan peningkatan +0.0031 pada *Semantic F1* model T5 terhadap *ground truth*.

2. *Filter Jaccard* cenderung lebih baik untuk ROUGE *Precision*, terlihat pada T5 + *Jaccard* yang mencapai 0.7351.
3. Pada model BART, pemilihan *filter* tidak memberikan dampak signifikan terhadap hasil akhir. Temuan ini menunjukkan pola yang konsisten: T5 unggul untuk replikasi *ground truth*, sementara BART unggul untuk representasi tema abstrak.

### 3.2 Pembahasan

Hasil eksperimen yang disajikan pada sub-bab sebelumnya mengungkap perbedaan kinerja yang signifikan dan konsisten antara model T5 dan BART, serta pengaruh metode penyaringan yang digunakan. Pembahasan ini akan menginterpretasikan temuan-temuan kunci tersebut dengan lebih mendalam, menghubungkannya dengan karakteristik mendasar dari setiap pendekatan, serta mengkontekstualisasikan dengan penelitian terdahulu.

#### 3.2.1 Interpretasi Kinerja Model: T5 vs. BART

Perbedaan mendasar dalam kinerja kedua model dapat diatribusikan kepada desain arsitektur, filosofi pre-training, dan kemampuan linguistik yang berbeda.

1. T5 sebagai "Task-Specialist" dengan Pendekatan Extractive-Implicit  
Keunggulan T5 ketika dievaluasi terhadap kata kunci referensi (Tabel 2) konsisten dengan karakteristiknya sebagai model text-to-text yang dilatih secara terawasi (supervised) untuk tugas transformasi teks-ke-teks yang spesifik [20]. Model ini cenderung bersifat lebih deterministik dan belajar untuk memetakan pola input (abstrak) ke output (kata kunci) yang sering kali bersifat eksplisit dan literal. Secara linguistik, T5 menunjukkan kecenderungan untuk mengambil noun phrases yang muncul secara eksplisit dalam teks, seperti "convolutional neural network," "support vector machine," atau "renewable energy." Pendekatan ini mirip dengan metode ekstraktif tradisional, namun dengan kemampuan generatif yang memungkinkannya menghasilkan variasi frasa yang tetap berakar pada leksikon input.  
Temuan ini sejalan dengan penelitian Pezik et al. [9] yang menunjukkan keefektifan T5 untuk ekstraksi frasa kunci, namun penelitian kami memperluasnya dengan menunjukkan bahwa keunggulan ini terutama terlihat dalam konteks replikasi kata kunci ahli (ground truth). Kelemahan T5 terletak pada sifatnya yang terlalu "patuh" pada teks input, sehingga kurang mampu melakukan abstraksi dan inferensi konsep implisit yang tidak secara langsung disebutkan dalam abstrak. Hal ini menjelaskan mengapa T5 memiliki ROUGE Precision yang sangat tinggi terhadap abstrak (Tabel 3) namun ROUGE Recall dan Semantic F1-nya rendah—model ini hanya mengambil apa yang secara harfiah ada, bukan apa yang dimaksudkan.
2. BART sebagai "Context-Understander" dengan Pendekatan Abstractive-Thematic  
Sebaliknya, keunggulan BART yang sangat mencolok dalam evaluasi berbasis abstrak (Tabel 3) berakar pada pre-training-nya sebagai denoising autoencoder [13]. Pre-training BART yang melibatkan rekonstruksi teks asli dari versi yang dirusak (corrupted) membekalinya dengan kemampuan pemahaman kontekstual dan kemampuan inferensi yang lebih kuat. Dalam tugas pembangkitan kata kunci, BART tidak hanya mencari kata-kata yang sering muncul, tetapi melakukan "pemahaman membaca" terhadap abstrak untuk menyimpulkan tema sentral, konsep kunci, dan hubungan antar-entitas.  
Kemampuan ini memungkinkan BART menghasilkan kata kunci yang lebih abstrak dan tematik, seperti "system performance," "data integrity," atau "computational efficiency," meskipun frasa-frasa tersebut tidak selalu muncul secara eksplisit dalam teks abstrak. Secara linguistik, BART menunjukkan kemampuan conceptual blending—menggabungkan konsep-konsep yang tersebar dalam teks menjadi frasa kunci yang representatif. Temuan ini mendukung penelitian Lewis et al. [13] yang menyoroti kekuatan BART dalam tugas generatif berbasis konteks, sekaligus memperluasnya ke domain pembangkitan kata kunci ilmiah.

Perbedaan ini juga tercermin dalam analisis kualitatif output model. Misalnya, untuk abstrak tentang "deep learning applications in renewable energy forecasting," T5 menghasilkan: "deep learning, renewable energy, forecasting models" (eksplisit dan teknis), sedangkan BART menghasilkan: "AI-driven sustainability, predictive analytics, energy optimization" (lebih tematik dan konseptual).

#### 3.2.2 Analisis Mendalam Peran Strategi Filtering

Pemilihan metode penyaringan (filtering) berperan sebagai tahap post-processing kritis yang dapat memperkuat atau melemahkan kekuatan model generator.

1. Filter Jaccard: Penguat Kesetiaan Leksikal  
Filter Jaccard beroperasi dengan prinsip yang sederhana: semakin banyak tumpang tindih kata antara kandidat kata kunci dan dokumen, semakin tinggi skornya. Pendekatan ini sangat sinergis dengan T5, yang memang cenderung menghasilkan kandidat yang secara leksikal dekat dengan abstrak. Filter Jaccard kemudian memilih kandidat-kandidat yang paling "setia" kepada teks sumber, sehingga menghasilkan ROUGE Precision yang sangat tinggi (0.7351 pada Tabel 3).  
Namun, kekakuannya menjadi kelemahan utama. Filter ini akan secara tidak memihak menyingkirkan kata kunci yang sangat relevan secara semantik tetapi menggunakan sinonim atau parafrasa. Contohnya, untuk konsep "neural network," filter Jaccard mungkin mempertahankan frasa yang persis sama dalam abstrak, tetapi

menyaring keluar varian seperti "deep learning model" atau "artificial neural network" yang secara semantik setara tetapi leksikal berbeda. Inilah sebabnya kombinasi T5+Jaccard memiliki Semantic F1 yang relatif rendah dibandingkan dengan T5+BERT.

## 2. Filter BERT: Penyelamat Relevansi Semantik

Filter BERT mengatasi keterbatasan Jaccard dengan bekerja di ruang vektor semantik. Dengan memanfaatkan model Sentence-BERT, filter ini mengukur kedekatan makna, bukan kedekatan bentuk. Ini sangat bermanfaat untuk menyelamatkan kandidat kata kunci dari T5 yang secara semantik brilian namun kurang cocok secara leksikal.

Peningkatan marginal Semantic F1 pada T5+BERT (0.5233 vs 0.5202 pada Tabel 2) membuktikan efektivitasnya. Dalam analisis kualitatif, kami menemukan contoh di mana filter BERT berhasil mempertahankan frasa seperti "neural architecture search" meskipun frasa tersebut tidak muncul secara literal dalam abstrak, tetapi tersirat dalam pembahasan tentang "optimizing AI model structures." Namun, filter ini tampaknya kurang berdampak pada BART. Hal ini terjadi karena BART, berkat pemahaman kontekstualnya yang sudah bawaan, secara inherent telah menghasilkan kandidat kata kunci yang sudah sangat selaras secara semantik dengan abstrak. Proses penyaringan tambahan oleh BERT karenanya tidak banyak mengubah peringkat kandidat yang sudah baik tersebut.

### 3.2.3 Analisis Kesalahan Kualitatif dan Studi Kasus

Untuk memahami batasan masing-masing pendekatan, kami melakukan analisis kesalahan kualitatif terhadap beberapa kasus representatif.

#### 1. Kasus 1: Kegagalan T5 pada Abstrak dengan Konsep Implisit

Pada abstrak tentang "ethical considerations in algorithmic decision-making systems," kata kunci ground truth mencakup: "algorithmic bias, fairness, accountability, transparency." T5 + BERT menghasilkan: "algorithmic decision-making, ethics, machine learning" — secara semantik relevan tetapi gagal menangkap konsep spesifik seperti "bias" dan "fairness" yang tidak disebutkan eksplisit. BART + BERT menghasilkan: "algorithmic fairness, ethical AI, bias mitigation, transparent algorithms" — lebih komprehensif menangkap tema implisit.

#### 2. Kasus 2: Kegagalan BART pada Output Terlalu Umum (Overly Generic)

Pada abstrak teknis tentang "GPU-accelerated quantum circuit simulation," BART menghasilkan: "high-performance computing, simulation, quantum technology" — terlalu umum dan kurang informatif untuk pengindeksan spesifik. T5 justru menghasilkan: "GPU acceleration, quantum circuits, simulation algorithms" — lebih tepat secara teknis.

#### 3. Kasus 3: Tantangan pada Ground Truth yang Tidak Ideal

Analisis juga mengungkap bahwa kualitas ground truth sendiri bervariasi. Beberapa artikel memiliki kata kunci yang terlalu umum ("analysis, system, model") atau tidak konsisten dengan konten abstrak. Dalam kasus seperti ini, baik T5 maupun BART kesulitan menghasilkan kata kunci yang bermakna, menunjukkan pentingnya kurasi dataset yang ketat.

### 3.2.4 Kontekstualisasi dengan Penelitian Terdahulu

Temuan penelitian ini tidak hanya mengkonfirmasi tetapi juga memperluas wawasan dari penelitian sebelumnya:

#### 1. Dukungan dan Perluasan atas Pezik et al. [9]

Penelitian kami mendukung temuan Pezik et al. tentang keefektifan T5 untuk ekstraksi frasa kunci, namun memberikan nuansa penting: keunggulan T5 paling terlihat dalam konteks replikasi kata kunci ahli, bukan dalam menangkap tema keseluruhan dokumen. Ini menunjukkan bahwa efektivitas model sangat tergantung pada definisi "kualitas" kata kunci yang diadopsi.

#### 2. Konvergensi dengan Lewis et al. [13] tentang Kapabilitas Generatif BART

Temuan superioritas BART dalam representasi tematik sejalan dengan klaim Lewis et al. tentang kemampuan BART sebagai denoising autoencoder yang unggul dalam tugas generatif. Namun, penelitian kami mengidentifikasi batasan praktisnya: kecenderungan menghasilkan output yang terlalu umum pada teks yang sangat teknis.

#### 3. Dialog dengan Studi Filtering Semantik

Temuan tentang efektivitas filter BERT untuk T5 tetapi tidak untuk BART memperkaya diskusi tentang strategi post-processing. Studi oleh Babayigit & Sattuf [14] menunjukkan manfaat BERT untuk penyaringan semantik, namun penelitian kami menunjukkan bahwa manfaat ini tidak universal—tergantung pada karakteristik generator yang mendasarinya.

### 3.2.5 Implikasi Teoretis dan Praktis

Temuan ini memiliki implikasi yang signifikan baik untuk penelitian teoretis maupun aplikasi praktis.

#### 1. Implikasi Teoretis



- a. Dikotomi Ekstraktif-Generatif dalam Pembangkitan Kata Kunci: Penelitian ini memperkuat bukti bahwa pembangkitan kata kunci berada pada spektrum antara ekstraksi dan generasi, dengan T5 mendekati ujung ekstraktif dan BART mendekati ujung generatif.
  - b. Importance of Task Definition: Kinerja model sangat bergantung pada bagaimana "kata kunci yang baik" didefinisikan—apakah kesetiaan pada istilah ahli atau kemampuan representasi tematik.
  - c. Sinergi Arsitektur dan Post-processing: Efektivitas strategi filtering bergantung pada kecocokan dengan karakteristik generator, menunjukkan perlunya pendekatan holistik dalam desain sistem.
2. Implikasi Praktis
- a. Sistem Hybrid untuk Aplikasi Komprehensif: Hasil ini menunjukkan potensi sistem hybrid yang menggabungkan T5 (untuk kata kunci eksplisit) dan BART (untuk kata kunci tematik) dalam satu pipeline. Misalnya, T5 dapat digunakan untuk menghasilkan kata kunci teknis untuk pengindeksan, sementara BART untuk kata kunci tematik untuk rekomendasi.
  - b. Rekomendasi Berbasis Tujuan Aplikasi:  
Untuk Sistem Pengindeksan Metadata Presisi Tinggi: T5 + Filter BERT adalah pilihan optimal, menawarkan keseimbangan antara kesesuaian leksikal dan kualitas semantik.  
Untuk Sistem Rekomendasi dan Pemetaan Topik: BART + Filter Jaccard cukup memadai, karena kemampuan tematik BART sudah dominan.  
Untuk Aplikasi dengan Sumber Daya Terbatas: T5 + Jaccard menawarkan efisiensi komputasi dengan kinerja yang masih baik untuk pengindeksan dasar.
  - c. Panduan Kurasi Dataset: Dataset kata kunci ilmiah sebaiknya tidak hanya berisi frasa teknis, tetapi juga konsep tematik. Anotasi ganda—oleh ahli domain untuk ketepatan teknis dan oleh ilmuwan informasi untuk representasi tematik—dapat meningkatkan kualitas dataset pelatihan.

### 3.2.6 Keterbatasan Penelitian

Meskipun memberikan wawasan yang berharga, penelitian ini memiliki beberapa keterbatasan yang membuka peluang untuk penelitian lanjutan.

1. Keterbatasan Penelitian Ini
  - a. Generalizability Domain: Dataset yang digunakan meskipun multidisiplin, masih didominasi oleh topik ilmu komputer dan teknik. Kinerja model perlu divalidasi pada domain seperti biomedis, ilmu sosial, atau humaniora yang memiliki karakteristik terminologi dan gaya penulisan yang berbeda.
  - b. Kedalaman Model Terbatas: Penelitian ini hanya menguji varian base dan small dari model. Eksperimen dengan model yang lebih besar (seperti T5-large atau BART-large) mungkin akan menghasilkan pola kinerja yang berbeda, terutama dalam hal kemampuan menangkap nuansa semantik.
  - c. Evaluasi Otomatis vs. Human-Centric: Metrik evaluasi otomatis (ROUGE, Semantic Similarity) meskipun objektif, tidak dapat sepenuhnya menangkap aspek kegunaan (utility) kata kunci bagi pengguna akhir seperti peneliti, pustakawan, atau sistem rekomendasi.
  - d. Aspek Temporal dan Dinamika Penelitian: Dataset tidak memperhitungkan evolusi terminologi ilmiah. Kata kunci yang relevan dapat berubah seiring waktu dengan munculnya sub-bidang baru atau paradigma penelitian.
2. Arah Penelitian Mendatang
  - a. Eksperimen Multi-Domain dengan Evaluasi Manusia: Validasi temuan pada domain yang lebih beragam dilengkapi dengan evaluasi manusia untuk menilai utilitas praktis kata kunci yang dihasilkan.
  - b. Eksplorasi Arsitektur Hybrid dan Ensemble: Penelitian tentang arsitektur yang secara organik menggabungkan kekuatan T5 dan BART, atau sistem ensemble yang memanfaatkan kedua model.
  - c. Integrasi Konteks Beyond Abstract: Mengeksplorasi penggunaan keseluruhan artikel (full-text) atau metadata tambahan (referensi, afiliasi) untuk pembangkitan kata kunci yang lebih kontekstual.
  - d. Studi Longitudinal tentang Evolusi Kata Kunci: Menganalisis bagaimana model menangkap tren penelitian yang muncul dan evolusi terminologi ilmiah.
  - e. Optimasi untuk Sumber Daya Terbatas: Mengembangkan varian model yang dioptimalkan untuk lingkungan dengan kendala komputasi atau memori, seperti repositori institusi kecil.
  - f. Interpretability dan Explainability: Mengembangkan metode untuk memahami mengapa model tertentu menghasilkan kata kunci tertentu, meningkatkan transparansi dan kepercayaan pada sistem otomatis.

## 4. KESIMPULAN

Berdasarkan hasil penelitian komparatif antara model T5 dan kombinasi BART+BERT dalam pembangkitan kata kunci otomatis dari abstrak artikel ilmiah, dapat disimpulkan bahwa pemilihan arsitektur model dan strategi filtering sangat berpengaruh terhadap kualitas hasil yang dihasilkan. Model T5 menunjukkan kinerja yang lebih unggul ketika dievaluasi terhadap kata kunci referensi (ground truth) yang ditetapkan oleh penulis, dengan skor Semantic F1 tertinggi sebesar 0.5233 pada konfigurasi T5+BERT. Hal ini menegaskan bahwa T5 efektif digunakan pada aplikasi yang menuntut presisi terminologi dan kesesuaian teknis, seperti sistem pengindeksan metadata

otomatis dan repositori ilmiah. Sebaliknya, model BART menunjukkan keunggulan signifikan dalam merepresentasikan konteks dan tema abstrak secara menyeluruh, dengan skor Semantic F1 mencapai 0.7517. Kemampuan ini menjadikan BART lebih sesuai untuk aplikasi yang menekankan pemahaman semantik global, seperti sistem rekomendasi artikel, klasifikasi topik otomatis, dan analisis tren penelitian. Temuan ini menunjukkan bahwa perbedaan arsitektur model berimplikasi langsung pada karakteristik kata kunci yang dihasilkan, baik dari sisi presisi terminologi maupun keluasan makna. Penelitian ini juga menegaskan pentingnya peran strategi filtering dalam meningkatkan kualitas keluaran. Filter berbasis Jaccard Similarity efektif untuk mempertahankan kesesuaian leksikal, sementara filter berbasis Semantic Similarity dengan BERT mampu meningkatkan relevansi makna, khususnya pada model T5, dengan mempertahankan kata kunci yang relevan secara semantik meskipun tidak identik secara tekstual. Meskipun memberikan hasil yang signifikan, penelitian ini masih memiliki keterbatasan, antara lain dominasi domain ilmu komputer dalam dataset, keterbatasan varian model yang diuji, serta evaluasi yang sepenuhnya bersifat otomatis. Oleh karena itu, penelitian selanjutnya disarankan untuk memperluas domain dataset, mengeksplorasi model berukuran lebih besar atau arsitektur hybrid, melibatkan evaluasi berbasis penilaian manusia, serta menganalisis aspek efisiensi komputasi. Secara keseluruhan, penelitian ini menegaskan bahwa pemilihan model dan strategi filtering harus disesuaikan dengan tujuan aplikasi, karena tidak terdapat satu pendekatan yang optimal untuk seluruh skenario pembangkitan kata kunci otomatis.

## REFERENCES

- [1] R. Dwi Handayani, Z. Zulkarnaini, W. Wasilah, and R. Herwanto, "Pendampingan Penerapan Intitutional Repository Digital Untuk Peningkatan Current Research Information System," *Dedik. J. Pengabd. Kpd. Masy.*, vol. 1, no. 2, pp. 74–84, Aug. 2022, doi: 10.53276/dedikasi.v1i2.31.
- [2] L. Corrin, K. Thompson, G. J. Hwang, and J. M. Lodge, "The importance of choosing the right keywords for educational technology publications," *Australas. J. Educ. Technol.*, vol. 38, no. 2, pp. 1–8, 2022, doi: 10.14742/ajet.8087.
- [3] P. Pottier *et al.*, "Title, abstract and keywords: a practical guide to maximize the visibility and impact of academic papers," *Proc. R. Soc. B Biol. Sci.*, vol. 291, no. 2027, Jul. 2024, doi: 10.1098/rspb.2024.1222.
- [4] R. N. Rath and A. Mustafi, "The importance of Term Weighting in semantic understanding of text: A review of techniques," *Multimed. Tools Appl.*, vol. 82, no. 7, pp. 9761–9783, 2023, doi: 10.1007/s11042-022-12538-3.
- [5] W. Lu, Z. Liu, Y. Huang, Y. Bu, X. Li, and Q. Cheng, "How do authors select keywords? A preliminary study of author keyword selection behavior," *J. Informetr.*, vol. 14, no. 4, pp. 1–17, 2020, doi: 10.1016/j.joi.2020.101066.
- [6] T. Bin Sarwar, N. M. Noor, and M. S. U. Miah, "Evaluating keyphrase extraction algorithms for finding similar news articles using lexical similarity calculation and semantic relatedness measurement by word embedding," *PeerJ Comput. Sci.*, vol. 8, 2022, doi: 10.7717/peerj-cs.1024.
- [7] D. B. Hier *et al.*, "Evaluation of standard and semantically-augmented distance metrics for neurology patients," *BMC Med. Inform. Decis. Mak.*, vol. 20, no. 1, pp. 1–15, 2020, doi: 10.1186/s12911-020-01217-8.
- [8] S. Wang, L. Thompson, and M. Iyyer, "Phrase-BERT: Improved Phrase Embeddings from BERT with an Application to Corpus Exploration," in *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, Stroudsburg, PA, USA: Association for Computational Linguistics, 2021, pp. 10837–10851. doi: 10.18653/v1/2021.emnlp-main.846.
- [9] P. Pęzik, A. Mikołajczyk, A. Wawrzyński, B. Nitoń, and M. Ogrodniczuk, "Keyword Extraction from Short Texts with a Text-to-Text Transfer Transformer," *Commun. Comput. Inf. Sci.*, vol. 1716 CCIS, pp. 530–542, 2022, doi: 10.1007/978-981-19-8234-7\_41.
- [10] M. Mars, "From Word Embeddings to Pre-Trained Language Models: A State-of-the-Art Walkthrough," *Appl. Sci.*, vol. 12, no. 17, 2022, doi: 10.3390/app12178805.
- [11] A. Mastropaolo *et al.*, "Studying the usage of text-to-text transfer transformer to support code-related tasks," *Proc. - Int. Conf. Softw. Eng.*, pp. 336–347, 2021, doi: 10.1109/ICSE43902.2021.00041.
- [12] A. Ninkov, J. R. Frank, and L. A. Maggio, "Bibliometrics: Methods for studying academic publishing," *Perspect. Med. Educ.*, vol. 11, no. 3, pp. 173–176, 2022, doi: 10.1007/s40037-021-00695-4.
- [13] M. Lewis *et al.*, "BART: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension," *Proc. Annu. Meet. Assoc. Comput. Linguist.*, pp. 7871–7880, 2020, doi: 10.18653/v1/2020.acl-main.703.
- [14] J. Devlin, M. W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," *NAACL HLT 2019 - 2019 Conf. North Am. Chapter Assoc. Comput. Linguist. Hum. Lang. Technol. - Proc. Conf.*, vol. 1, no. M1m, pp. 4171–4186, 2019.
- [15] K. S. Nugroho, A. Y. Sukmadewa, and N. Yudistira, "Large-Scale News Classification using BERT Language Model: Spark NLP Approach," *ACM Int. Conf. Proceeding Ser.*, pp. 240–246, 2021, doi: 10.1145/3479645.3479658.
- [16] A. Glazkova and D. Morozov, "Cross-Domain Robustness of Transformer-Based Keyphrase Generation," *Commun. Comput. Inf. Sci.*, vol. 2086 CCIS, pp. 249–265, 2024, doi: 10.1007/978-3-031-67826-4\_19.
- [17] A. Glazkova and D. Morozov, "Multi-task fine-tuning for generating keyphrases in a scientific domain," in *2023 IX International Conference on Information Technology and Nanotechnology (ITNT)*, IEEE, Apr. 2023, pp. 1–5. doi: 10.1109/ITNT57377.2023.10139061.
- [18] L. Gazit and M. Ghaffari, *Mastering NLP from Foundations to LLMs*. Packt Publishing, 2024.
- [19] N. F. Ali, J. Uddin Tanvin, M. R. Islam, J. Ahmed, and M. Akhtaruzzaman, "ROUGE Score Analysis and Performance Evaluation Between Google T5 and SpaCy for YouTube News Video Summarization," in *2023 26th International Conference on Computer and Information Technology (ICCIT)*, IEEE, Dec. 2023, pp. 1–6. doi: 10.1109/ICCIT60459.2023.10441408.



- [20] A. Bali, A. Bhagwat, A. Bhise, and S. Joshi, "Semantic Similarity Detection and Analysis For Text Documents," in *2024 Second International Conference on Emerging Trends in Information Technology and Engineering (ICETITE)*, IEEE, Feb. 2024, pp. 1–9. doi: 10.1109/ic-ETITE58242.2024.10493834.