

**PERBANDINGAN PEMBANGKITAN KATA KUNCI PADA ABSTRAK ARTIKEL  
PENELITIAN MENGGUNAKAN T5 DAN KOMBINASI BART + BERT**

**TESIS**

**Oleh:**

**JOHAN**  
NIM. 231232157



**PROGRAM STUDI S-2 TEKNOLOGI INFORMASI  
FAKULTAS INFORMATIKA  
UNIVERSITAS MIKROSKIL  
MEDAN  
2025**

**COMPARISON OF KEYWORD GENERATION IN RESEARCH ARTICLE  
ABSTRACTS USING T5 AND THE BART + BERT COMBINATION**

**THESIS**

**By:**

**JOHAN  
ID NUMBER. 231232157**



**MAJOR OF S-2 INFORMATION TECHNOLOGY  
FACULTY OF INFORMATICS  
UNIVERSITAS MIKROSKIL  
MEDAN  
2025**

## LEMBARAN PENGESAHAN

### PERBANDINGAN PEMBANGKITAN KATA KUNCI PADA ABSTRAK ARTIKEL PENELITIAN MENGGUNAKAN T5 DAN KOMBINASI BART + BERT

#### TESIS

Diajukan untuk Melengkapi Persyaratan Guna  
Mendapatkan Gelar Magister  
Program Studi S-2 Teknologi Informasi

Oleh:

**JOHAN**  
**NIM. 231232157**

Disetujui Oleh:

Dosen Pembimbing,

  
Ir. Erwin Setiawan Panjaitan, M.M.S.I., Ph.D.

Medan, 23 Agustus 2025  
Diketahui dan Disahkan Oleh:

Ketua Program Studi  
S-2 Teknologi Informasi

  
Ir. Erwin Setiawan Panjaitan, M.M.S.I., Ph.D.

## KATA PENGANTAR

Penulis mengucapkan puji syukur kepada Tuhan Yang Maha Esa, atas karunia dan bimbingan-Nya sehingga penulis dapat menyelesaikan skripsi ini dengan baik.

Tesis ini berjudul **“Perbandingan Pembangkitan Kata Kunci pada Abstrak Artikel Penelitian Menggunakan T5 dan Kombinasi BART + BERT”** dan disusun sebagai syarat kelulusan untuk memperoleh Gelar Magister Komputer (M.Kom) tahap pendidikan Strata II di Program Studi Teknologi Informasi Universitas Mikroskil Medan.

Pada kesempatan ini, penulis menyampaikan rasa terima kasih yang sebesar-besarnya kepada:

1. Bapak Ir. Erwin Setiawan Panjaitan, M.M.S.I., Ph.D, selaku Dosen Pembimbing I.
2. Ibu Dr. Sofiana Nurjanah, S.Kom., M.T.I, selaku Dosen Pembimbing II.
3. Bapak Hardy, S.Kom., M.SC., PH.D, selaku Rektor Universitas Mikroskil Medan.
4. Bapak Sunaryo Winardi, S.Kom., M.T.I, selaku Dekan Fakultas Informatika Universitas Mikroskil Medan.
5. Bapak Ir. Erwin Setiawan Panjaitan, M.M.S.I., Ph.D, selaku Ketua Program Studi S-2 Teknologi Informasi Fakultas Informatika Universitas Mikroskil Medan.
6. Para Dosen di Universitas Mikroskil, baik S-1 maupun S-2
7. Secara Khusus, penulis sampaikan terimakasih kepada Orang tua, Keluarga, famili dan sanak saudara penulis yang telah memberikan dukungan moril dan materil.
8. Seluruh Teman-teman sekelas dan rekan-rekan mahasiswa/i Program Studi S-2 Teknologi Informasi yang selalu memberikan semangat dan bantuan dalam penulisan Tesis ini.

Medan, 06 Agustus 2025

Penulis,



Johan

## HALAMAN PERNYATAAN

Saya yang membuat pernyataan ini adalah mahasiswa Program Studi S-2 Teknologi Informasi Universitas Mikroskil Medan dengan identitas mahasiswa sebagai berikut:

Nama : Johan

NIM : 231232157

Saya telah melaksanakan penelitian dan penulisan Tesis dengan judul dan tempat penelitian sebagai berikut:

Judul Tesis \*) : Perbandingan Pembangkitan Kata Kunci pada Abstrak Artikel Penelitian Menggunakan T5 dan Kombinasi BART + BERT

Sehubungan dengan Tesis tersebut, dengan ini saya menyatakan dengan sebenar-benarnya bahwa penelitian dan penulisan Tesis tersebut merupakan hasil karya saya sendiri (tidak menyuruh orang lain yang mengerjakannya) dan semua sumber, baik yang dikutip maupun dirujuk, telah saya nyatakan dengan benar. Bila di kemudian hari ternyata terbukti bahwa bukan saya yang mengerjakannya (membuatnya), maka saya bersedia dikenakan sanksi yang telah ditetapkan oleh Universitas Mikroskil Medan, yakni pencabutan ijazah yang telah saya terima dan ijazah tersebut dinyatakan tidak sah.

Selain itu, demi pengembangan ilmu pengetahuan, saya menyetujui untuk memberikan kepada Universitas Mikroskil Medan Hak Bebas Royalti Non-eksklusif (Non-exclusive Royalty Free Right) atas Tesis saya beserta perangkat yang ada (jika diperlukan). Dengan hak ini, Universitas Mikroskil Medan berhak menyimpan, mengalihmedia/formatkan, mengelola dalam bentuk pangkalan data (database), merawat, dan mempublikasikan Tesis saya, secara keseluruhan atau hanya sebagian atau hanya ringkasannya saja dalam bentuk format tercetak dan/atau elektronik, selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik hak cipta. Menyatakan juga bahwa saya akan mempertahankan hak eksklusif saya untuk menggunakan seluruh atau sebagian isi Tesis saya guna pengembangan karya di masa depan, misalnya dalam bentuk artikel, buku, ataupun perangkat lunak/sistem informasi.

Demikian pernyataan ini saya perbuat dengan sungguh-sungguh, dalam keadaan sadar dan tanpa ada tekanan dari pihak manapun.

Medan, 06 Agustus 2025

Saya yang membuat pernyataan,



Johan

## ABSTRACT

The increasing volume of scientific publications presents challenges in information retrieval and metadata management. Keywords are essential for indexing, yet manual keyword assignment is inconsistent and labor-intensive. This study investigates the effectiveness of transformer-based models—T5 and a BART-BERT combination—for automatic keyword generation from scientific abstracts. Using a dataset of 100 journal articles, the models were fine-tuned and evaluated on accuracy, precision, recall, F1-score, and runtime. T5, particularly when paired with BERT filtering, produced the most precise keywords, closely matching author-assigned terms with a ROUGE F1-score of 0.2533 and a Semantic F1-score of 0.5233. In contrast, BART combined with either Jaccard or BERT filtering achieved higher semantic relevance to the full abstract content, with a Semantic F1-score of 0.7517. The findings suggest T5 is better suited for generating explicit, technical keywords for indexing, while BART is more effective for capturing thematic content, making it ideal for applications like content recommendation. Filtering methods also impacted results: BERT improved semantic quality, while Jaccard enhanced lexical alignment. These insights highlight the need to tailor model selection and filtering strategies based on the intended use of keyword generation in academic systems.

**Keywords:** Automatic Keyword Generation, Transformer Models, T5 and BART-BERT, Scientific Abstracts, Semantic Filtering

## ABSTRAK

Peningkatan jumlah publikasi ilmiah secara signifikan menimbulkan tantangan dalam pengambilan informasi dan manajemen metadata. Kata kunci berperan penting dalam proses pengindeksan, namun penetapannya secara manual masih bersifat inkonsisten dan memakan waktu. Penelitian ini mengkaji efektivitas model berbasis transformer—T5 dan kombinasi BART-BERT—untuk tugas pembangkitan kata kunci otomatis dari abstrak artikel ilmiah. Dengan menggunakan dataset berisi 100 artikel jurnal, kedua pendekatan dimodelkan dan dievaluasi berdasarkan akurasi, presisi, recall, skor F1, dan waktu proses. Model T5 yang dipadukan dengan filter BERT menghasilkan kata kunci paling presisi, mendekati kata kunci yang ditetapkan oleh penulis artikel, dengan skor ROUGE F1 sebesar 0,2533 dan Semantic F1 sebesar 0,5233. Sementara itu, BART yang dikombinasikan dengan filter Jaccard atau BERT menunjukkan relevansi semantik yang lebih tinggi terhadap isi abstrak secara keseluruhan, dengan Semantic F1 mencapai 0,7517. Temuan ini menunjukkan bahwa T5 lebih tepat digunakan untuk menghasilkan kata kunci eksplisit dan teknis, sedangkan BART lebih unggul dalam menangkap tema dan konteks secara menyeluruh. Pemilihan metode filtering juga memengaruhi hasil: BERT meningkatkan kualitas semantik, sementara Jaccard lebih baik dalam pencocokan leksikal. Hasil penelitian ini menekankan pentingnya pemilihan model dan strategi filtering berdasarkan tujuan penggunaan dalam sistem akademik.

**Kata Kunci:** Pembangkitan Kata Kunci Otomatis, Model Transformer, T5 dan BART-BERT, Abstrak Ilmiah, Penyaringan Semantik

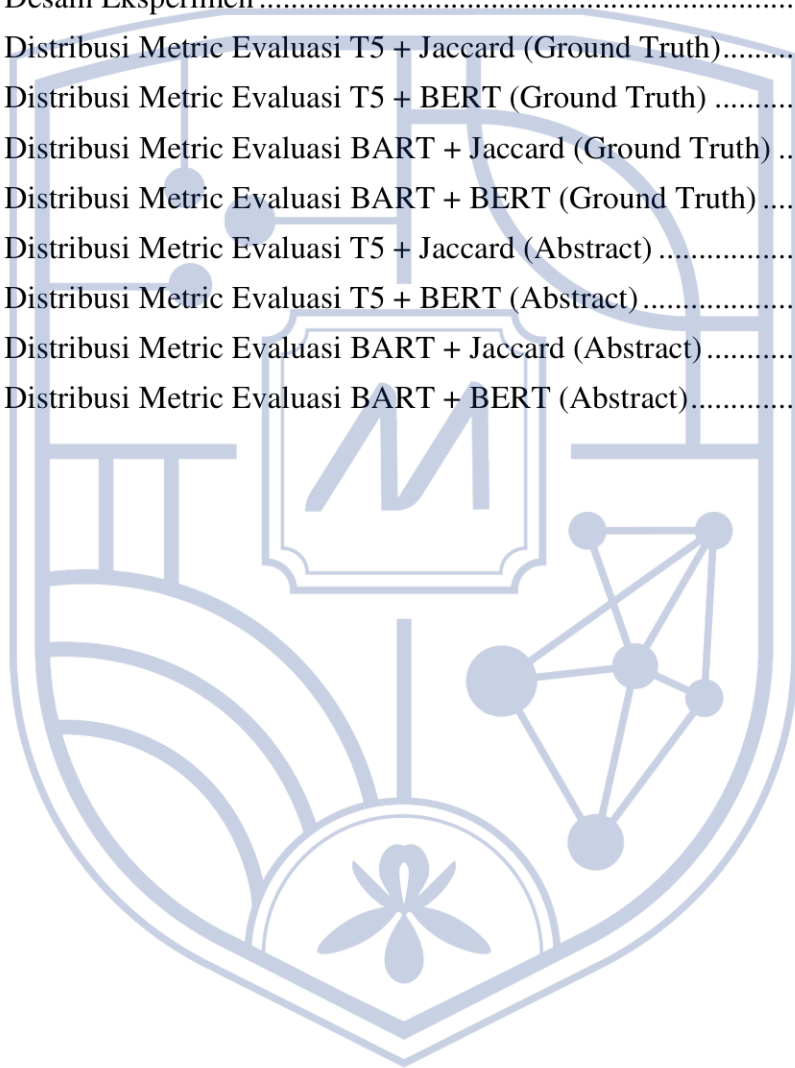
# DAFTAR ISI

|                                                                     |            |
|---------------------------------------------------------------------|------------|
| <b>ABSTRACT .....</b>                                               | <b>i</b>   |
| <b>ABSTRAK.....</b>                                                 | <b>ii</b>  |
| <b>KATA PENGANTAR .....</b>                                         | <b>iii</b> |
| <b>DAFTAR ISI .....</b>                                             | <b>iv</b>  |
| <b>DAFTAR GAMBAR .....</b>                                          | <b>vi</b>  |
| <b>DAFTAR TABEL .....</b>                                           | <b>vii</b> |
| <b>BAB I PENDAHULUAN .....</b>                                      | <b>1</b>   |
| <b>1.1 Latar Belakang.....</b>                                      | <b>1</b>   |
| <b>1.2 Rumusan Masalah .....</b>                                    | <b>3</b>   |
| <b>1.3 Tujuan.....</b>                                              | <b>3</b>   |
| <b>1.4 Manfaat.....</b>                                             | <b>3</b>   |
| <b>1.5 Ruang Lingkup .....</b>                                      | <b>4</b>   |
| <b>BAB II KAJIAN LITERATUR.....</b>                                 | <b>5</b>   |
| <b>2.1 Pembangkitan Kata Kunci Otomatis.....</b>                    | <b>5</b>   |
| 2.1.1 Pendekatan Pembangkitan Kata Kunci Otomatis.....              | 5          |
| 2.1.2 Kelemahan Metode Ekstraktif Tradisional .....                 | 6          |
| <b>2.2 Pemanfaatan Model Transformer dalam NLP.....</b>             | <b>7</b>   |
| 2.2.1 Arsitektur dasar T5 (Text-to-Text Transfer Transformer) ..... | 7          |
| 2.2.2 Arsitektur BART .....                                         | 9          |
| 2.2.3 Arsitektur BERT .....                                         | 11         |
| 2.2.4 Kombinasi Model dalam Pembangkitan Kata Kunci .....           | 14         |
| <b>2.3 Evaluasi Kinerja Model .....</b>                             | <b>15</b>  |
| <b>2.4 Gap Penelitian .....</b>                                     | <b>17</b>  |
| <b>2.5 Kerangka Konseptual.....</b>                                 | <b>18</b>  |
| <b>BAB III METODOLOGI PENELITIAN.....</b>                           | <b>22</b>  |
| <b>3.1 Desain Penelitian.....</b>                                   | <b>22</b>  |
| <b>3.2 Data Penelitian .....</b>                                    | <b>23</b>  |
| 3.2.1 Sumber dan Pengumpulan Data .....                             | 23         |
| 3.2.2 Praproses Data.....                                           | 24         |
| <b>3.3 Implementasi Model .....</b>                                 | <b>25</b>  |
| <b>3.4 Evaluasi dan Analisis.....</b>                               | <b>33</b>  |
| 3.4.1 Evaluasi Hasil .....                                          | 33         |
| 3.4.2 Perbandingan dan Analisis .....                               | 34         |

|                                                                            |           |
|----------------------------------------------------------------------------|-----------|
| <b>BAB IV HASIL DAN PEMBAHASAN .....</b>                                   | <b>35</b> |
| <b>4.1 Hasil .....</b>                                                     | <b>35</b> |
| 4.1.1 Deskripsi Umum Eksperimen .....                                      | 35        |
| 4.1.2 Evaluasi Model Berdasarkan Ground Truth.....                         | 43        |
| 4.1.3 Evaluasi Model Berdasarkan Abstrak .....                             | 54        |
| 4.1.4 Perbandingan Antar Metode .....                                      | 63        |
| <b>4.2 Pembahasan.....</b>                                                 | <b>67</b> |
| 4.2.1 Interpretasi Perbedaan Model (T5 vs BART) .....                      | 67        |
| 4.2.2 Interpretasi Perbedaan Filter (Jaccard vs BERT).....                 | 69        |
| 4.2.3 Interpretasi Perbedaan Acuan Evaluasi (Ground Truth vs Abstrak)..... | 70        |
| 4.2.4 Implikasi Temuan .....                                               | 72        |
| 4.2.5 Keterbatasan Penelitian.....                                         | 72        |
| <b>BAB V KESIMPULAN DAN SARAN .....</b>                                    | <b>74</b> |
| <b>5.1 Kesimpulan.....</b>                                                 | <b>74</b> |
| <b>5.2 Saran .....</b>                                                     | <b>74</b> |
| <b>DAFTAR PUSTAKA .....</b>                                                | <b>76</b> |
| <b>DAFTAR RIWAYAT HIDUP.....</b>                                           | <b>82</b> |

## DAFTAR GAMBAR

|                                                                            |    |
|----------------------------------------------------------------------------|----|
| Gambar 2. 1 Arsitektur T5.....                                             | 8  |
| Gambar 2. 2 Arsitektur BART .....                                          | 10 |
| Gambar 2. 3 Arsitektur BERT .....                                          | 12 |
| Gambar 2. 4 Kerangka Konseptual.....                                       | 21 |
| Gambar 3. 1 Diagram Alur Penelitian .....                                  | 22 |
| Gambar 3. 2 Hasil praproses dataset .....                                  | 25 |
| Gambar 4. 1 Desain Eksperimen .....                                        | 43 |
| Gambar 4. 2 Distribusi Metric Evaluasi T5 + Jaccard (Ground Truth).....    | 45 |
| Gambar 4. 3 Distribusi Metric Evaluasi T5 + BERT (Ground Truth) .....      | 48 |
| Gambar 4. 4 Distribusi Metric Evaluasi BART + Jaccard (Ground Truth) ..... | 50 |
| Gambar 4. 5 Distribusi Metric Evaluasi BART + BERT (Ground Truth) .....    | 53 |
| Gambar 4. 6 Distribusi Metric Evaluasi T5 + Jaccard (Abstract) .....       | 55 |
| Gambar 4. 7 Distribusi Metric Evaluasi T5 + BERT (Abstract).....           | 58 |
| Gambar 4. 8 Distribusi Metric Evaluasi BART + Jaccard (Abstract).....      | 60 |
| Gambar 4. 9 Distribusi Metric Evaluasi BART + BERT (Abstract).....         | 62 |



## DAFTAR TABEL

|             |                                                     |    |
|-------------|-----------------------------------------------------|----|
| Tabel 2. 1  | Ringkasan Penelitian Terdahulu .....                | 17 |
| Tabel 4. 1  | Sampel Dataset.....                                 | 36 |
| Tabel 4. 2  | Evaluasi T5 + Jaccard (Ground Truth) .....          | 44 |
| Tabel 4. 3  | Evaluasi T5 + BERT (Ground Truth) .....             | 46 |
| Tabel 4. 4  | Evaluasi BART + Jaccard (Ground Truth) .....        | 49 |
| Tabel 4. 5  | Evaluasi BART + BERT (Ground Truth) .....           | 52 |
| Tabel 4. 6  | Evaluasi T5 + Jaccard (Abstract) .....              | 54 |
| Tabel 4. 7  | Evaluasi T5 + BERT (Abstract).....                  | 57 |
| Tabel 4. 8  | Evaluasi BART + Jaccard (Abstract).....             | 59 |
| Tabel 4. 9  | Evaluasi BART + BERT (Abstract).....                | 61 |
| Tabel 4. 10 | Perbandingan Evaluasi Berdasarkan Ground Truth..... | 64 |
| Tabel 4. 11 | Perbandingan Evaluasi Berdasarkan Abstract .....    | 66 |

