

# Pengembangan Aplikasi Sistem Peringkas Otomatis Jurnal Berbahasa Indonesia Berbasis Deep Learning

<sup>1</sup>AAIN Eka Karyawati

<sup>2</sup>Program Studi Informatika, FMIPA, Universitas Udayana  
Denpasar, Indonesia  
eka.karyawati@unud.ac.id

<sup>2</sup>Ngurah Agus Sanjaya ER, <sup>3</sup>LAA Rahning Putri,

<sup>2,3</sup>Program Studi Informatika, FMIPA, Universitas Udayana  
Denpasar, Indonesia

<sup>2</sup>agus\_sanjaya@unud.ac.id, <sup>3</sup>rahningputri@unud.ac.id

**Abstract**—Dewasa ini, ketersediaan referensi digital sangat berlimpah dan mudah diakses dari laman Internet. Ketersediaan referensi khususnya artikel jurnal yang sangat banyak tersebut sering membuat pembaca kebingungan untuk memilih referensi yang tepat atau sesuai dengan topik yang dialami. Pembaca sering menghabiskan banyak waktu untuk membaca artikel yang tidak tepat. Untuk mengatasi masalah ini, diperlukan sebuah aplikasi peringkasan otomatis yang mampu meringkas artikel secara otomatis sehingga mempersingkat waktu membaca artikel tersebut. Peringkasan teks otomatis (Automatic Text Summarization, ATS) adalah peringkasan teks yang dilakukan secara otomatis oleh komputer. Peringkasan dilakukan sesuai dengan fokus peringkasan. Ada dua pendekatan atau algoritma peringkasan teks otomatis, yaitu peringkasan ekstraksi dan peringkasan abstraksi. Metode peringkasan ekstraksi dilakukan dengan memilih unit teks (kalimat, segmen kalimat, paragraf, atau bagian), yang berisi informasi penting dari dokumen dan mengatur unit tersebut dengan benar. Sedangkan metode peringkasan abstraksi melibatkan proses parafrase (menguraikan dengan kata sendiri) dari dokumen asli. Metode peringkasan abstraksi lebih mendekati metode peringkasan yang dilakukan secara alami oleh manusia. Pada penelitian ini dikembangkan algoritma peringkasan abstraksi untuk peringkasan jurnal Ilmu Komputer/Informatika berbahasa Indonesia menggunakan metode Deep Learning Transformer. Dataset jurnal yang digunakan adalah data jurnal bidang Informatika berbahasa Indonesia yang diunduh dari laman Internet. Data ringkasan jurnal disusun secara manual mengikuti metodologi peringkasan jurnal ilmiah. Evaluasi dilakukan dengan mengukur tingkat akurasi dari model dan juga mengukur fungsionalitas dari prototipe aplikasi peringkas yang dikembangkan.

**Kata Kunci**— Automatic Text Summarization, Abstractive Summarization, Deep Learning, Transformer.

## I. PENDAHULUAN

Peringkasan teks otomatis (Automatic Text Summarization, ATS) adalah peringkasan teks yang dilakukan secara otomatis oleh komputer. Terdapat dua pendekatan atau algoritma peringkasan teks otomatis, yaitu peringkasan ekstraksi dan peringkasan abstraksi. Metode peringkasan ekstraksi dilakukan dengan memilih unit teks (kalimat, segmen kalimat, paragraf, atau bagian), yang berisi informasi penting dari dokumen dan mengatur unit tersebut dengan benar. Sedangkan metode peringkasan abstraksi melibatkan proses parafrase (menguraikan dengan kata sendiri) dari dokumen asli. Metode peringkasan abstraksi lebih mendekati metode peringkasan yang dilakukan secara alami oleh manusia. Pada penelitian ini dikembangkan algoritma peringkasan abstraksi untuk peringkasan jurnal Ilmu Komputer/Informatika berbahasa Indonesia menggunakan metode Deep Learning Transformer.

Terdapat cukup banyak penelitian tentang peringkasan teks otomatis menggunakan metode abstraksi (Chopra et al., 2016; Duan et al., 2019; Adelia et al., 2019; Shi et al., 2020; Zhang et al., 2020; Zaki et al., 2020; Subramanian et al., 2020; Beltagy et al., 2020; Dou et al., 2021; Huang et al., 2021; Jiang et al., 2022). Penelitian ATS dengan metode abstraksi tersebut menggunakan metode Deep Learning Language Generation yaitu model LSTM, RNN dan Transformer. Tetapi, masih sedikit penelitian tentang peringkasan teks otomatis berbahasa Indonesia berbasis metode abstraksi yang dikembangkan. Penelitian peringkasan teks otomatis untuk artikel ilmiah sudah dilakukan oleh beberapa peneliti (Cohan & Goharian, 2017; Lauscher et al., 2017; Abura'ed et al., 2017; Slamet et al., 2018; Zhang, Li, & Yao, 2018; Marques, Jiang et al., 2019; Cozman, & Santos, 2019; Agrawal et al., 2019; Adelia et al., 2019). Sebagian besar peneliti (Lauscher et al., 2017; Abura'ed et al., 2017; Slamet et al., 2018; Agrawal et al., 2019) menggunakan metode peringkasan ekstraksi untuk meringkas artikel ilmiah. Penelitian peringkasan artikel ilmiah yang menggunakan metode abstraksi masih sangat sedikit (Jiang et al., 2019; Adelia et al., 2019). Disamping itu, masih sangat sedikit penelitian ATS artikel ilmiah yang berbahasa Indonesia (Slamet et al., 2018; Adelia et al., 2019). Pada penelitian ini dikembangkan Sistem Peringkasan Teks Otomatis berbasis Deep Learning Transformer untuk peringkasan Jurnal Informatika berbahasa Indonesia.

## II. METODE DAN PROSEDUR

### A. Data dan Metode Pengumpulan Data

Pada penelitian ini, data yang digunakan adalah pasangan data jurnal dan ringkasannya. Target jumlah data penelitian adalah 2000 pasangan data jurnal-ringkasan. Jurnal yang digunakan dalam penelitian ini adalah jurnal berbahasa bidang Ilmu Komputer/Informatika yang diunduh dari laman internet <https://scholar.google.co.id/>.

### B. Dataset Jurnal Ilmu Komputer/Informatika Berbahasa Indonesia

Dataset jurnal berbahasa Indonesia bidang Ilmu Komputer/Informatika yang sudah terkumpul adalah 1000 jurnal. Target jumlah jurnal yang digunakan sebagai data training untuk model Deep Learning Transformer peringkasan otomatis adalah 2000 jurnal. Data yang digunakan hanya data teks dari jurnal tersebut. Semua data selain teks seperti gambar dan tabel dieliminasi dari jurnal. Data jurnal Informatika berbahasa Indonesia yang, bisa dilihat pada link:

[https://drive.google.com/drive/folders/1s\\_3MgnhV6vvBLkJEOUTWDFFIYea0DPpT?usp=sharing](https://drive.google.com/drive/folders/1s_3MgnhV6vvBLkJEOUTWDFFIYea0DPpT?usp=sharing)

### C. Penyusunan Ringkasan Secara Manual

Data ringkasan dari setiap jurnal disusun secara manual berdasarkan metodologi peringkasan artikel jurnal ilmiah dengan fokus untuk studi literature. Tahapan dari peringkasan artikel jurnal ilmiah didasarkan pada format APA untuk menemukan poin utama. Ringkasan disusun dengan mengidentifikasi setiap bagian dari artikel dan menyusun ulang bagian tersebut secara singkat dengan kata-kata sendiri (paraphrasing).

Terdapat lima bagian utama dari sebuah artikel jurnal ilmiah sesuai format APA yaitu:

- 1) Pertanyaan penelitian dan dasar pemikiran penelitian (dinyatakan dalam bab Pendahuluan);
- 2) Hipotesis atau hipotesis yang diuji (pada bab Pendahuluan);
- 3) Bagaimana hipotesis diuji (pada bab Metode);
- 4) Temuan (pada bab Hasil, termasuk tabel dan gambar);
- 5) Bagaimana temuan diinterpretasikan (pada bab Diskusi);

Data ringkasan bisa dilihat pada link:

[https://drive.google.com/drive/folders/13BmXvljNvjymhIE65S\\_ZLI0huJl\\_WEs?usp=sharing](https://drive.google.com/drive/folders/13BmXvljNvjymhIE65S_ZLI0huJl_WEs?usp=sharing)

### D. Tahap Penelitian

Tahapan utama dari Sistem Peringkasan Otomatis Jurnal Berbahasa Indonesia yang diusulkan adalah: (1) Pengumpulan data dan pembangunan korpus; (2) Training model Deep Learning Transformer; (3) Validasi model; (4) Testing model terbaik; dan (5) Pengembangan prototipe. Tahap pertama, dilakukan pengumpulan dataset pasangan jurnal-ringkasan. Kemudian mengeliminasi gambar dan tabel yang ada dalam jurnal untuk membangun korpus data teks jurnal dan ringkasannya. Korpus ini dibagi menjadi 3 dataset: 80% menjadi data training; 10% menjadi data development; dan 10% menjadi data testing. Kemudian dataset tersebut di-preprocessing agar menjadi bentuk standar, yaitu dalam bentuk vektor angka numerik, yang siap diproses di dalam machine learning (deep learning).

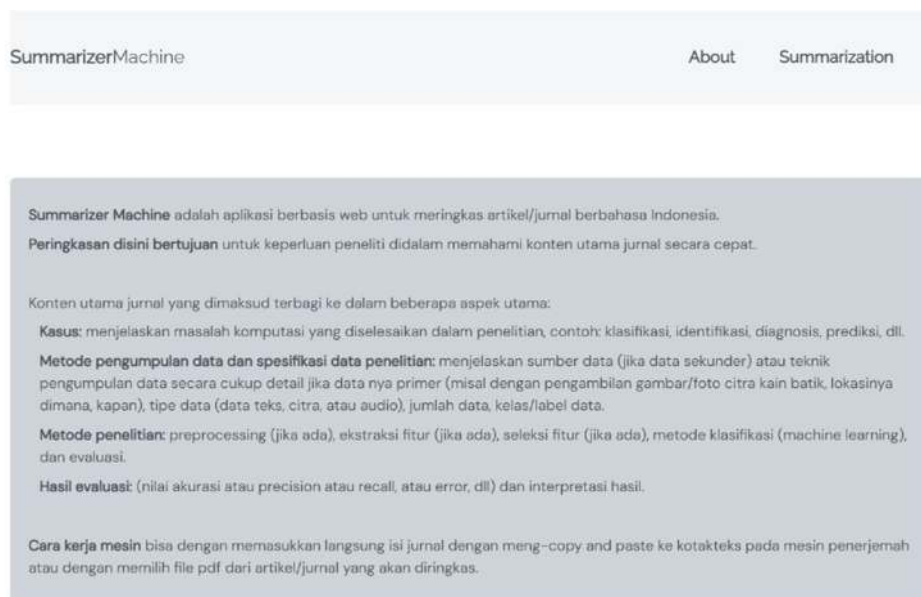
Tahap yang kedua adalah training data untuk menghasilkan model Deep Learning Transformer peringkasan otomatis jurnal berbahasa Indonesia. Pada tahap ini dataset training dan hyper-parameter dari model diinputkan ke mesin untuk mendapatkan model Deep Learning Transformer peringkasan otomatis jurnal berbahasa Indonesia. Model Deep Learning Transformer ini nantinya digunakan membangkitkan ringkasan secara otomatis dari jurnal yang diinputkan.

Pada tahap ketiga yaitu validasi model, model Deep Learning Transformer yang dihasilkan dari proses training divalidasi untuk memilih model Deep Learning Transformer terbaik. Metode validasi yang digunakan pada penelitian yang diusulkan adalah 10-fold cross validation. Validasi dilakukan dengan mengatur kombinasi nilai hyper-parameter model Deep Learning Transformer. Setiap satu kombinasi hyper-parameter dilakukan 10 kali percobaan. Model dengan kombinasi nilai hyper-parameter yang menghasilkan nilai rata-rata akurasi tertinggi yang dipilih sebagai model terbaik.

Selanjutnya pada tahap keempat, model terbaik hasil validasi dievaluasi untuk mengetahui tingkat akurasi menggunakan dataset yang baru (dataset testing). Nilai akurasi ini yang dijadikan dasar apakah model sudah cukup layak dikembangkan menjadi prototipe aplikasi Sistem Peringkasan Otomatis Jurnal Berbahasa Indonesia atau masih diperlukan pengembangan model lebih lanjut.

### III. HASIL DAN PEMBAHASAN

Penelitian ini sudah melakukan pembangunan korpus dataset artikel dengan pasangan ringkasan untuk digunakan sebagai pengembangan model ringkasan yang lebih baik. Selain dataset, prototipe aplikasi juga sudah selesai diimplementasikan. Aplikasi dimulai dengan menampilkan penjelasan dari Aplikasi peringkasan otomatis, bisa dilihat pada Gambar 1.



Gambar 1. Tampilan awal aplikasi penjelasan tentang aplikasi peringkasan otomatis

Terdapat 3 fitur utama aplikasi yaitu pertama, meringkas dengan menginputkan teks pada *text box*, kemudian menekan tombol *Ringkas Teks* untuk melakukan peringkasan, seperti terlihat pada Gambar 2.

## Peringkasan Otomatis Jurnal Berbahasa Indonesia

Tujuan Peringkasan untuk Telaah Pustaka (Review Paper)



The screenshot shows the SummarizerMachine interface. At the top, there are three tabs: "Input Teks", "Pilih Dokumen", and "Pencarian Dokumen". The "Input Teks" tab is active. Below the tabs is a large text input area with the placeholder text "Masukkan Teks (Maksimum 2000 Karakter)". At the bottom left of the input area, it says "0/2000 Karakter". At the bottom right, there is a button labeled "Ringkas Teks".

Gambar 2. Tampilan memasukkan teks

Setelah tombol *Ringkas Teks* ditekan maka akan muncul hasil ringkasan seperti terlihat pada Gambar 3. Hasil ringkasan bisa disimpan dengan menekan tombol *Simpan ke .txt*.



The screenshot shows the SummarizerMachine interface displaying the summarized text. At the top, there are three tabs: "Input Teks", "Pilih Dokumen", and "Pencarian Dokumen". The "Input Teks" tab is active. Below the tabs is a large text area containing the summarized text. At the bottom right, there is a button labeled "Simpan ke .txt".

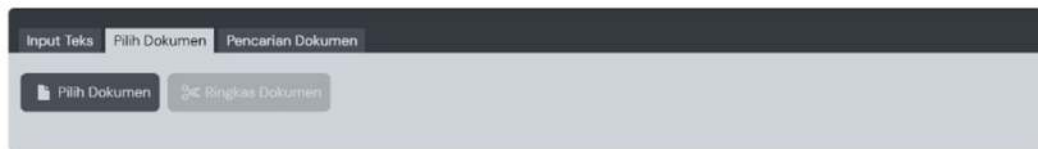
**Hasil Ringkasan**

Penelitian ini bertujuan untuk menganalisis nilai performa (akurasi, presisi, recall, dan f-measure) pada proses klasifikasi dataset citra penyakit pneumonia dan tidak pneumonia. Dataset citra penyakit pneumonia dan tidak pneumonia diperoleh dari Kaggle dataset library. Dataset citra Chest X-ray penyakit pneumonia di publikasikan oleh Paul Mooney pada tanggal 22 Maret 2018 dan terakhir diperbaharui di kaggle pada tanggal 25 Maret 2018, data tersebut dikumpulkan dari pasien anak berusia 1-5 tahun di Guangzhou Women and Children's Medical Center, Guangzhou. Tahapan yang dilakukan yaitu membagi dataset dengan berbagai simulasi rasio, deteksi tepi sobel, ekstraksi fitur moment invariant, klasifikasi metode KNN, nilai K=2 sampai K=900. Penelitian ini mencoba untuk mencari nilai performa terbaik dari dataset citra penyakit pneumonia dengan berbagai simulasi rasio yang terdiri dari rasio perbandingan 20:80 (200 citra penyakit pneumonia, 800 citra tidak pneumonia), 50:50 (500 citra penyakit pneumonia, 500 citra tidak pneumonia), 80:20 (800 citra pneumonia, 200 citra tidak pneumonia) serta beragam nilai K. Analisis dilakukan dengan menghitung performa (akurasi, presisi, recall, dan f-measure) dengan metode K-Nearest Neighbor (KNN) pada objek tersebut. Hasil Penelitian menunjukkan performa terbaik terdapat pada simulasi rasio 20:80 dengan memperoleh nilai akurasi 96%, presisi 97%, recall 97%, f-measure 97% dengan menggunakan nilai K=3.

Gambar 3. Tampilan hasil ringkasan

Fitur kedua adalah meringkas dengan memilih file dari koleksi dokumen yang ada. Bisa dilihat pada Gambar 4.

**Peringkasan Otomatis Jurnal Berbahasa Indonesia**  
Tujuan Peringkasan untuk Telaah Pustaka (Review Paper)



Gambar 4. Tampilan awal pilih dokumen

Kemudian file dipilih dengan mengklik tombol *Pilih Dokumen*. Dokumen yang terpilih akan ditampilkan terlebih dahulu, seperti terlihat pada Gambar 5. Jika memang ingin meringkas tinggal menekan tombol *Ringkas Dokumen*.

**Peringkasan Otomatis Jurnal Berbahasa Indonesia**  
Tujuan Peringkasan untuk Telaah Pustaka (Review Paper)

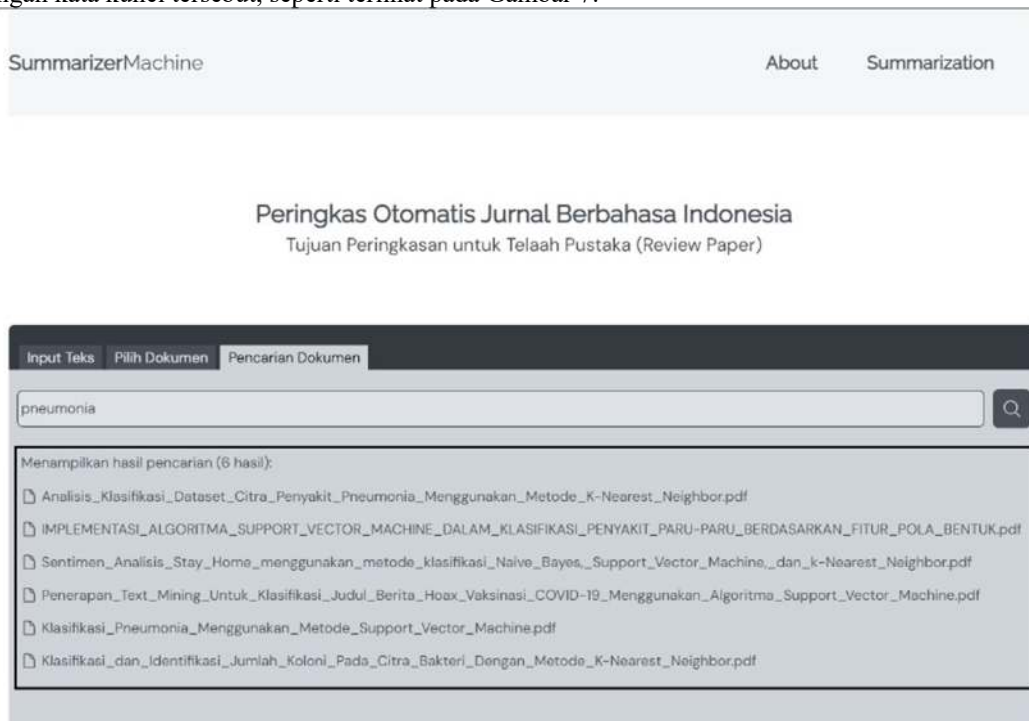


Gambar 5. Tampilan dokumen yang dipilih

Jika tombol *Ringkas Dokumen* ditekan maka akan muncul hasil ringkasan. Hasil ringkasan bisa disimpan dengan menekan tombol *Simpan ke .txt*, seperti terlihat pada Gambar 3. Fitur ketiga adalah dengan melakukan pencarian menggunakan kata kunci tertentu, jika belum yakin ada dokumen tertentu yang akan diringkaskan, seperti bisa dilihat pada Gambar 6.

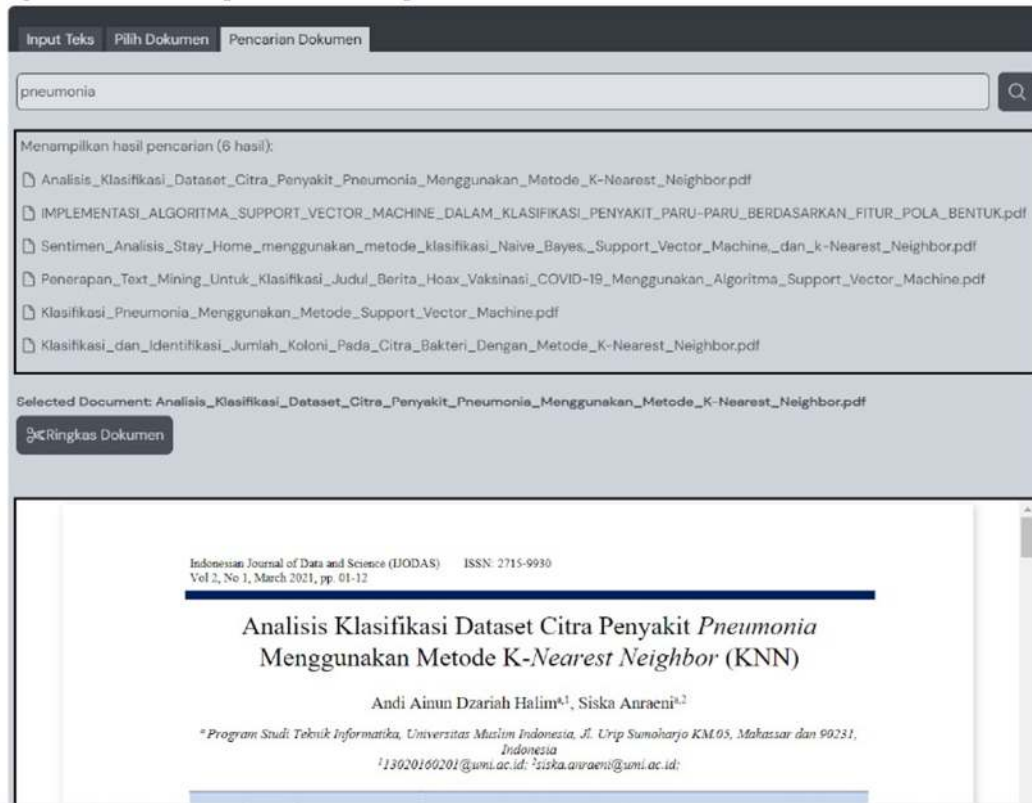


**Gambar 6.** Tampilan memasukkan kata kunci untuk pencarian dokumen  
Setelah memasukkan kata kunci dan mengklik tombol pencarian, maka akan muncul beberapa dokumen yang sesuai dengan kata kunci tersebut, seperti terlihat pada Gambar 7.



**Gambar 7.** Tampilan hasil pencarian dokumen

Jika satu dokumen diklik maka dokumen tersebut akan ditampilkan. Jika ingin meringkas bisa menekan tombol *Ringkas Dokumen*, seperti bisa dilihat pada Gambar 8.



**Gambar 8.** Tampilan meringkas hasil pencarian dokumen

Jika tombol *Ringkas Dokumen* ditekan maka akan muncul hasil ringkasan. Hasil ringkasan bisa disimpan dengan menekan tombol *Simpan ke .txt*, seperti terlihat pada Gambar 3.

#### IV. KESIMPULAN

Penelitian ini sudah berhasil membangun dataset untuk keperluan pengembangan model peringkas otomatis. Dataset yang berupa 1000 pasangan Jurnal berbahasa Indonesia dengan ringkasannya sudah berhasil dibangun. Model peringkas otomatis masih perlu dikembangkan untuk mendapatkan akurasi yang lebih tinggi. Prototipe dari aplikasi peringkas otomatis sudah berhasil dikembangkan, tetapi masih diperlukan pengujian untuk meningkatkan tingkat kepuasan user dalam menggunakan aplikasi peringkas otomatis tersebut.

#### DAFTAR PUSTAKA

- [1] Chopra, S., Auli, M., and Rush, A.M. 2016. Abstractive Sentence Summarization with Attentive Recurrent Neural Networks. Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies.
- [2] Duan, X., et al. 2019. Contrastive Attention Mechanism for Abstractive Sentence Summarization. In Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing
- [3] Adelia, R., Suyanto, S., Wisesty, N. 2019. Indonesian Abstractive Text Summarization Using Bidirectional Gated Recurrent Unit. In Proceedings of the 4th International Conference on Computer Science and Computational Intelligence (ICCSICI) 2019.
- [4] Shi, T, Keneshloo, Y., Ramakrishnan, N., and Reddy, C. K. 2020. Neural Abstractive Text Summarization with Sequence-to-Sequence Models. ACM Trans. Data Sci., Vol. 1, No. 1.
- [5] Zhang, J., Zhao, Y., Saleh, M., and Liu, P. J. 2020. PEGASUS: Pre-training with Extracted Gap-sentences for Abstractive Summarization. In Proceedings of the 37th International Conference on Machine Learning, Online, 2020.
- [6] Subramanian, S., Li, R., Pilault, J., and Pal, C. 2020. On Extractive and Abstractive Neural Document Summarization with Transformer Language Models. arXiv:1909.03186 [cs.CL]

- [7] Beltagy, I., Peters, M. E., and Cohan, A. 2020. Longformer: The Long-Document Transformer. arXiv:2004.05150 [cs.CL]
- [8] Dou, Z.-Y., Liu, P., Hayashi, H., Jiang, Z., and Neubig, G. 2021. GSum: A General Framework for Guided Neural Abstractive Summarization. arXiv:2010.08014 [cs.CL]
- [9] Huang, L., Cao, S., Parulian, N., Ji, H., and Wang, L. 2021. Efficient Attentions for Long Document Summarization. arXiv:2104.02112 [cs.CL]
- [10] Jiang, Y., et al. 2022. Enriching Transformers with Structured Tensor-Product Representations for Abstractive Summarization. arXiv:2106.01317 [cs.CL]
- [11] Cohan, A. and Goharian, N. 2017. Contextualizing Citations for Scientific Summarization using Word Embeddings and Domain Knowledge. SIGIR '17, Shinjuku, Tokyo, Japan.
- [12] Lauscher, A., Glavač, G., and Eckert, K. 2017. University of Mannheim@ CLSciSumm-17: citation-based summarization of scientific articles using semantic textual similarity. In: CEUR Workshop Proceedings, pp. 33–42.
- [13] Abura'ed, A., Chiruzzo, L., Saggion, H., Accuosto, P and Bravo, A. 2017. LaSTUS/TALN @ CLSciSumm-17: Cross-document Sentence Matching and Scientific Text Summarization Systems. In: Proceedings of the Computational Linguistics Scientific Summarization Shared Task (CL-SciSumm 2017), pp. 55–66.
- [14] Slamet, C. Atmadja, A. R., Maylawati, D. S., Lestari, R. S., Darmalaksana, W., and Ramdhani, M. A. 2018. Automated Text Summarization for Indonesian Article Using Vector Space Model. Proceedings of the 2nd Annual Applied Science and Engineering Conference (AASEC 2017).
- [15] Zhang, J. Li, K and Yao, C., 2018. Event-based Event-based Summarization Summarization for for Scientific Scientific Literature Literature in in Chinese. Procedia Computer Science, Vol. 129, No. 2018, Pp. 88–92.
- [16] Jiang, X.-J., Mao, X.-L., Feng, B.-S., Wei, X., Bian, B.-B., & Huang, H. 2019. HSDS: An Abstractive Model for Automatic Survey Generation. In Proceedings of the Database Systems for Advanced Applications, Cham.
- [17] Marques, J. M. C., Cozman, F. G., & Santos, I. H. F. d. 2019. Automatic Summarization of Technical Documents in the Oil and Gas Industry. Proceedings of the 2019 8th Brazilian Conference on Intelligent Systems (BRACIS).
- [18] Agrawal, K., Mittal, A., and Pudi, V. 2019. Scalable, Semi-Supervised Extraction of Structured Information from Scientific Literature. In Proceedings of the Workshop on Extracting Structured Knowledge from Scientific Publications 2019, pp. 11–20