TEXT SUMMARIZATION UNTUK ARTIKEL BAHASA INDONESIA PADA KORPUS OBJEK BUDAYA DENGAN METODE LEXRANK

Skripsi

oleh:

Bernadus Maria Rosario Hardus Tukan
71180280

PROGRAM STUDI INFORMATIKA FAKULTAS TEKNOLOGI INFORMASI
UNIVERSITAS KRISTEN DUTA WACANA
2022

TEXT SUMMARIZATION UNTUK ARTIKEL BAHASA INDONESIA PADA KORPUS OBJEK BUDAYA DENGAN METODE LEXRANK

Skripsi

Diajukan kepada Program Studi Informatika Fakultas Teknologi Informasi
Universitas Kristen Duta Wacana
Sebagai Salah Satu Syarat dalam Memperoleh Gelar
Sarjana Komputer

Disusun oleh

Bernadus Maria Rosario Hardus Tukan
71180280

PROGRAM STUDI INFORMATIKA FAKULTAS TEKNOLOGI INFORMASI
UNIVERSITAS KRISTEN DUTA WACANA
2022

HALAMAN PENGESAHAN

TEXT SUMMARIZATION UNTUK ARTIKEL BAHASA INDONESIA PADA KORPUS OBJEK BUDAYA DENGAN METODE LEXRANK

Oleh: BERNADUS MARIA ROSARIO HARDUS TUKAN / 71180280

Dipertahankan di depan Dewan Penguji Skripsi
Program Studi Informatika Fakultas Teknologi Informasi
Universitas Kristen Duta Wacana - Yogyakarta
Dan dinyatakan diterima untuk memenuhi salah satu syarat memperoleh gelar
Sarjana Komputer
pada tanggal 25 Oktober 2022

Yogyakarta, 3 November 2022
Mengesahkan,

Dewan Penguji:

1. Gloria Virginia, S.Kom., MAI, Ph.D.
2. Budi Susanto, SKom.,M.T.
3. R. Gunawan Santosa, Drs. M.Si.
4. Yuan Lukito, S.Kom., M.Cs.

Dekan

(Restyandito,S.Kom.,MSIS.,Ph.D.)

Ketua Program Studi

(Gloria Virginia,Ph.D.)

HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI SKRIPSI/TESIS/DISERTASI UNTUK KEPENTINGAN AKADEMIS

Sebagai sivitas akademika Universitas Kristen Duta Wacana, saya yang bertanda tangan di bawah ini:

Nama	: Bernadus Maria Rosario Hardus Tukan
NIM	: 71180280
Program studi	: Informatika
Fakultas	: Teknologi Informasi
Jenis Karya	: Skripsi

demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Kristen Duta Wacana **Hak Bebas Royalti Noneksklusif** (*None-exclusive Royalty Free Right*) atas karya ilmiah saya yang berjudul:

"TEXT SUMMARIZATION UNTUK ARTIKEL BAHASA INDONESIA PADA KORPUS OBJEK BUDAYA DENGAN METODE LEXRANK"

beserta perangkat yang ada (jika diperlukan). Dengan Hak Bebas Royalti/Noneksklusif ini Universitas Kristen Duta Wacana berhak menyimpan, mengalih media/formatkan, mengelola dalam bentuk pangkalan data (*database*), merawat dan mempublikasikan tugas akhir saya selama tetap mencantumkan nama kami sebagai penulis/pencipta dan sebagai pemilik Hak Cipta.

Demikian pernyataan ini saya buat dengan sebenarnya.

Dibuat di : Yogyakarta
Pada Tanggal : 20 Desember 2022

Yang menyatakan

(Bernadus Maria Rosario Hardus Tukan)
NIM.71180280

PERNYATAAN KEASLIAN SKRIPSI

Saya menyatakan dengan sesungguhnya bahwa skripsi dengan judul:

TEXT SUMMARIZATION UNTUK ARTIKEL BAHASA INDONESIA PADA KORPUS OBJEK BUDAYA DENGAN METODE LEXRANK

yang saya kerjakan untuk melengkapi sebagian persyaratan menjadi Sarjana Komputer pada pendidikan Sarjana Program Studi Informatika Fakultas Teknologi Informasi Universitas Kristen Duta Wacana, bukan merupakan tiruan atau duplikasi dari skripsi kesarjanaan di lingkungan Universitas Kristen Duta Wacana maupun di Perguruan Tinggi atau instansi manapun, kecuali bagian yang sumber informasinya dicantumkan sebagaimana mestinya.

Jika dikemudian hari didapati bahwa hasil skripsi ini adalah hasil plagiasi atau tiruan dari skripsi lain, saya bersedia dikenai sanksi yakni pencabutan gelar kesarjanaan saya.

Yogyakarta, 21 November 2022

Bernadus Maria Rosario Hardus Tukan
71180280

HALAMAN PERSETUJUAN

Judul Skripsi	:	TEXT SUMMARIZATION UNTUK ARTIKEL BAHASA INDONESIA PADA KORPUS OBJEK BUDAYA DENGAN METODE LEXRANK
Nama Mahasiswa	:	BERNADUS MARIA ROSARIO HARDUS TUKAN
NIM	:	71180280
Mata Kuliah	:	Skripsi (Tugas Akhir)
Kode	:	TI0366
Semester	:	Genap
Tahun Akademik	:	2022/2023

Telah diperiksa dan disetujui di
Yogyakarta,
Pada tanggal 5 Oktober 2022

Dosen Pembimbing I	Dosen Pembimbing II
Gloria Virginia, S.Kom., MAI., Ph.D	Budi Susanto, S.Kom., M.T.

HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR UNTUK KEPENTINGAN AKADEMIS \ SECARA ONLINE UNIVERSITAS KRISTEN DUTA \ WACANA YOGYAKARTA

Saya yang bertanda tangan di bawah ini:

NIM	:	71180280
Nama	:	Bernadus Maria Rosario Hardus Tukan
Prodi / Fakultas	:	Teknologi Informasi / Informatika
Judul Tugas Akhir	:	TEXT SUMMARIZATION UNTUK ARTIKEL BAHASA INDONESIA PADA KORPUS OBJEK BUDAYA DENGAN METODE LEXRANK

bersedia menyerahkan Tugas Akhir kepada Universitas melalui Perpustakaan untuk keperluan akademis dan memberikan **Hak Bebas Royalti Non Ekslusif** (*Non-exclusive Royalty-free Right*) serta bersedia Tugas Akhirnya dipublikasikan secara online dan dapat diakses secara lengkap (*full access*).

Dengan Hak Bebas Royalti Nonekslusif ini Perpustakaan Universitas Kristen Duta Wacana berhak menyimpan, mengalihmedia/formatkan, mengelola dalam bentuk *database*, merawat, dan mempublikasikan Tugas Akhir saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta.

Demikian pernyataan ini saya buat dengan sebenar-benarnya.

Yogyakarta, 24 November 2022

Yang menyatakan,

(…71180280 – BERNADUS MARIA ROSARIO HARDUS TUKAN…)

KATA PENGANTAR

Segala puji dan syukur kepada Tuhan yang maha kasih, karena atas segala rahmat, bimbingan, dan bantuan-Nya maka akhirnya Skripsi dengan judul TEXT SUMMARIZATION UNTUK ARTIKEL BAHASA INDONESIA PADA KORPUS OBJEK BUDAYA DENGAN METODE LEXRANK ini telah selesai disusun.

Penulis memperoleh banyak bantuan dari kerja sama baik secara moral maupun spiritual dalam penulisan Skripsi ini, untuk itu tak lupa penulis ucapkan terima kasih yang sebesar-besarnya kepada:

1. Tuhan yang maha kasih,
2. Orang tua yang selama ini telah sabar membimbing dan mendoakan penulis tanpa kenal untuk selama-lamanya.
3. Bapak Restyandito,S.Kom.,MSIS.,Ph.D. selaku Dekan Fakultas Teknologi Informasi Universitas Kristen Duta Wacana.
4. Ibu Gloria Virginia, S.Kom., MAI, Ph.D selaku Kepala Program Studi Informatika Fakultas Teknologi Informasi Universitas Kristen Duta Wacana.
5. Ibu Gloria Virginia, S.Kom., MAI, Ph.D selaku Dosen Pembimbing 1, yang telah memberikan ilmunya dan dengan penuh kesabaran membimbing penulis.
6. Bapak Budi Susanto, S.Kom., M.T. selaku Dosen Pembimbing 2 yang telah memberikan ilmu dan kesabaran dalam membimbing penulis.
7. Keluarga tercinta yang selalu mendukung penulis menyelesaikan skripsi ini.
8. Kristianto Pratama, Gratsia Theodorin, Elza Miyori, Stevani Dwi dan Maria Amanda yang menemani dan menyemangati penulis selama mengerjakan skripsi.
9. Dan juga berbagai pihak lain yang telah mendukung moral, spiritual, dan dana untuk belajar selama ini.

Laporan proposal/skripsi ini tentunya tidak lepas dari segala kekurangan dan kelemahan, untuk itu segala kritikan dan saran yang bersifat membangun guna kesempurnaan skripsi ini sangat diharapkan. Semoga proposal/skripsi ini dapat bermanfaat bagi pembaca semua dan lebih khusus lagi bagi pengembagan ilmu komputer dan teknologi informasi.

Yogyakarta, 24 November 2022

Bernadus Maria Rosario Hardus Tukan

DAFTAR ISI

DAFTAR GAMBAR

DAFTAR TABEL

DAFTAR LAMPIRAN

INTISARI

TEXT SUMMARIZATION UNTUK ARTIKEL BAHASA INDONESIA PADA KORPUS OBJEK BUDAYA DENGAN METODE LEXRANK

Pemahaman akan pentingnya informasi dari sebuah laporan atau artikel semakin dibutuhkan di era teknologi saat ini dan orang cenderung tidak menghabiskan waktu untuk mencari substansi dari artikel yang mereka baca. Memberikan sistem peringkasan otomatis menjadi solusi yang bisa ditawarkan ke pembaca.

Pada penelitian ini menggunakan metode *LexRank* (*Lexical PageRank*) untuk melakukan peringkasan teks otomatis dengan teknik *extractive summarization*. Metode *LexRank* menerapakan konsep *Connectivity Matrix* yang memungkinkan suatu kalimat menjadi sangat kuat jika berkoneksi dengan beberapa kalimat lain. *LexRank* juga menggunakan pendekatan berbasis *graph* yaitu memodelkan teks kedalam bentuk *graph* dengan menjadikan teks sebagai *vertex* dan menambahkan *edges* pada *graph* berdasarkan koneksi antar unit teks untuk menentukan tingkat pentingnya kalimat berdasarkan struktur *graph* keseluruhan. Adapun metode-metode terkait untuk memaksimalkan peringkasan *LexRank* yaitu: *TF-IDF*, *Cosine Similarity* dan *Power Iterations* (*Eigen Vector Centrality*).

Kalimat yang diekstrak sebesar 25%, dengan ambang batas 0.1, 0.2, 0.3, 0.4, 0.5 dan proses evaluasi menggunakan metode *ROUGE* dengan menghitung *precision*, *recall* dan *f-score*. Selain itu penelitian ini juga melakukan evaluasi untuk menilai performa metode *LexRank* dan metode peringkasan teks lain yaitu *TextRank*.

Kata-kata kunci : Teknologi informasi, peringkasan teks *extractive*, *LexRank* , *ROUGE* , *multi documents, Connectivity Matrix,* Pendekatan *Graph*

ABSTRACT

TEXT SUMMARIZATION FOR INDONESIAN ARTICLES ON THE CORPUS OF CULTURAL OBJECTS USING THE LEXRANK METHOD

Understanding the importance of information from a report or article is increasingly needed in today's technology era and people tend not to spend time looking for the substance of the articles they read. Providing an automatic summary system is a solution that can be offered to readers.

In this study, the LexRank (Lexical PageRank) method is used to perform automatic text summarization with extractive summarization techniques. The LexRank method applies the Connectivity Matrix concept which allows a sentence to be very strong if it is connected to several other sentences. LexRank also uses a graph-based approach, namely modeling text into graph form by making text as a vertex and adding edges to the graph based on the connection between text units to determine the level of importance of sentences based on the overall graph structure. While related methods to maximize LexRank summary are: TF- IDF, Cosine Similarity and Power Iterations (Eigen Vector Centrality).

Sentences extracted by 25%, with a threshold of 0.1, 0.2, 0.3, 0.4, 0.5 and the evaluation process using the ROUGE method by calculating precision, recall and f-score. In addition, this study also evaluates the performance of the LexRank method and another text summary method, namely TextRank.

Keywords*:* Information technology, extractive text summarization, LexRank, ROUGE, multi documents, Connectivity Matrix, Graph Approach

BAB I
PENDAHULUAN

1.1. Latar Belakang Masalah

Kebutuhan mengetahui inti informasi dari sebuah berita atau artikel semakin diperlukan pada era teknologi, masyarakat cenderung kurang suka menghabiskan waktu mereka hanya untuk mencari inti dari artikel yang dibaca. Penggunaan *Text Summarization* merupakan solusi untuk memberikan informasi ringkas ke pembaca tanpa harus menghabiskan banyak waktu. Informasi ringkas yang diberikan tidak semata-mata berupa teks yang dipotong agar terlihat lebih sedikit melainkan informasi ringkasan tersebut sudah diproses sehingga intisari dari artikel dapat tersampaikan ke pembaca. Pada *Text Summarization* terdapat beberapa teknik untuk meringkas informasi salah satunya adalah *extractive summary*. *Extractive summary* berupa penyalinan bagian-bagian teks yang dianggap penting oleh sistem lalu digabungkan menjadi suatu ringkasan. Teks yang disalin berupa kata, kalimat atau paragraf tanpa menambahkan kalimat baru kedalam teks aslinya (Uçkan & Karcı, 2020).

Khusus pada Website Korpus Artikel Objek Budaya UKDW "https://alunalun.info/korpus/" untuk beberapa waktu terakhir telah mengumpulkan berbagai macam artikel yang berkaitan dengan objek budaya yang ada di Indonesia, Artikel ini disimpan berdasarkan kategori-kategori untuk memudahkan mencari data artikel tertentu. Karena kategori yang telah dikumpulkan belum memiliki ringkasan atau belum menjelaskan intisari dari kategori tersebut maka perlu dilakukan *Text Summarization* sebagai solusi mendapatkan ringkasannya. Untuk mendapatkan intisari bisa diselesaikan dengan cara memilah kata atau kalimat pada setiap artikel lalu secara manual dikumpulkan menjadi satu versi yang lebih singkat untuk mendapatkan ringkasan atau intisari dari kategori artikel yang diinginkan. Umumnya langkah seperti ini bisa dilakukan namun sangat tidak efektif jika kategori artikel memiliki 10 macam artikel dan masing-masing artikelnya mempunyai 5 sampai 8 halaman atau bahkan lebih yang memakan banyak waktu

dan tidak efisien. Memberi *Automatic Text Summarization* dengan menekankan pemilihan kata kunci sehingga membentuk *gold summary* (Klymenko et al., 2020) pada kategori artikel korpus objek budaya UKDW dibutuhkan untuk memberi ringkasan berupa intisari kategori artikel sebelum ditambahkan di katalog repo pada portal web budaya "https://portal.alunalun.info/", tujuannya agar pembaca memahami apa saja yang terdapat pada setiap ketegori artikel yang dibaca. Dikarenakan data artikel yang digunakan berbahasa Indonesia, salah satu metode pendukung untuk ringkasan teks adalah metode *LexRank* dengan pendekatan berbasis *graph*. Metode *LexRank* membutuhkan algoritma tertentu untuk peringkasan teks otomatis yaitu algoritma *LexRank*. Peringkasan dengan algoritma *LexRank* bertujuan untuk meringkas dokumen yang banyak sekaligus (*multi documents*).

Pada penulisan ini solusi yang ingin dicapai adalah membuat sistem peringkasan teks otomatis yang diterapkan pada "Portal Web Budaya Indonesia". Tujuan sistem adalah memberi ringkasan pada setiap kategori artikel yang ada di Portal Web Budaya, sehingga memudahkan pengguna dalam memahami intisari dari kategori yang dikunjungi. Penelitian ini merupakan bagian dari Penelitian Terapan Unggulan Perguruan Tinggi (PTUPT) dengan dana hibah RISTEK-BRIIN berjudul "Portal Objek Budaya Berbasis Semantic Web".

Berdasarkan latar belakang yang sudah dijelaskan diatas, penelitian ini akan mengembangkan sistem peringkasan otomatis berbasis *multi documents* dengan teknik *extractive* terhadap kategori artikel Objek Budaya .

1.2. Perumusan Masalah

Katagori artikel pada objek budaya yang belum di-intisari menjadi penyebab utama belum bisa dimasukkan kedalam katalog repo pada *project* alun-alun " https://app.alunalun.info/" . Pemberian intisari bisa dilakukan dengan menerapkan *Automatic Text Summarization*. *Automatic Text Summarization* digunakan untuk mempercepat proses ringkasan pada setiap kategori artikel yang banyak dengan bantuan *library LexRank* yang tersedia pada bahasa pemrograman *python* untuk program komputasinya. Berdasarkan masalah yang sudah dijabarkan, penelitian ini akan menerapkan metode *LexRank* dari *library python* untuk melakukan peringkasan teks. Oleh karena dokumen yang diolah adalah berbahasa Indonesia, maka diperlukan juga beberapa *pra-processing* untuk program komputasinya. *Pre-processing* yang digunakan antara lain *case folding, tokenization, lemmatization, labelling* dan *phrase detection* dengan basis kata. Hasil dari penerapan algoritma *LexRank* ini akan dievaluasi menggunakan *ROUGE* untuk menguji kalimat buatan sistem dan kalimat buatan responden.

1.3. Batasan Masalah

Dalam penulisan ini terdapat beberapa batasan yang digunakan yaitu:

1) Teks yang digunakan adalah teks yang berbahasa Indonesia.
2) Dokumen ringkasan dikhusukan pada kategori artikel objek budaya.
3) Sistem hanya akan mengeluarkan output berupa ringkasan untuk setiap kategori artikel.
4) Dataset artikel yang diuji berdasarkan data yang didapat dari *Website* Korpus Objek Budaya.
5) Ketegori yang diringkas ada 3 yaitu Dalang, Tari Tradisional dan Wayang dengan masing-masing kategori sebanyak 20 artikel.
6) Hasil evaluasi dengan *ROUGE* menghitung *recall* dan *precision* dari responden tanpa memperhatikan indeks kalimat.
7) Pembuatan *Gold Standar* menggunakan *Kappa Score*
8) Pemberian Label Kata disesuaikan dengan sumber data

1.4. Tujuan Penulisan

Tujuan penulisan ini adalah membuat sistem peringkasan teks otomatis dengan menerapkan metode *LexRank* yang disediakan oleh *python* untuk memudahkan pengguna untuk mengetahui instisari dari keseluruhan dokumen. Sebagai tools peringkasan teks otomatis untuk portal website budaya Indonesia (https://portal.alunalun.info/) dan memberikan sinopsis untuk berbagai korpus budaya yang ada di portal website budaya Indonesia.

1.5. Manfaat Penulisan

Manfaat yang diperoleh dari penelitian ini adalah dengan *Text Summarization* pada korpus artikel objek budaya untuk bahasa Indonesia dapat menyelesaikan masalah berupa abstraksi pada kategori artikel dengan memberi kemudahakan bagi pembaca untuk melakukan peringkasan otomatis di portal *web* budaya Indonesia (https://portal.alunalun.info/). Sementara pada sisi penulis, mempertajam proses berpikir yang terstruktur dan menambah pengetahuan pada bidang NLP (*Natural Language Processing*).

1.6. Metodologi Penulisan

Beberapa metode yang digunakan untuk mendukung penelitian ini meliputi: Studi literatur, Pengumpulan data, *Pre-processing*, *Processing*, Evaluasi. Tahapan – tahapan penelitian dapat dijabarkan sebagai berikut :

a. Studi Literatur

   Studi literatur dilakukan dengan mencari sumber-sumber terkait judul penulisan "*Text Summarization Untuk Artikel Bahasa Indonesia Pada Korpus Objek Budaya Dengan Metode LexRank*". Studi literatur yang digunakan berupa artikel, jurnal dan skripsi-skripsi terdahulu sebagai referensi dan bahan belajar agar lebih memahami proses dan alur kerja dari Metode *LexRank*.

b. Pengumpulan data

Data yang digunakan berdasarkan kategori untuk setiap dokumen objek budaya yang di input oleh kontributor kedalam website Korpus Artikel Objek Budaya. Kategori yang digunakan yaitu : wayang, tarian tradisional dan dalang.

c. *Pre-processing*

Pre-processing dilakukan pembersihan data terlebih dahulu berupa *case folding*, *tokenization, lemmatization*, *labelling* kata dan *phrase detection* sebelum diringkas. *Pre-processing* menggunakan *library NLTK* (*Natural Language Tool Kit*) dan Sastrawi yang disediakan oleh *python*.

d. *Processing*

Untuk penulisan ini tidak merancang ulang algoritma dari *LexRank*, melainkan menggunakan *library LexRank* yang disediakan *python* untuk proses komputasinya. Penulisan ini lebih menekankan pada proses *Pre-processing* untuk mendapatkan ringkasan yang berkualitas.

e. Evaluasi

Evaluasi yang dilakukan menggunakan *ROUGE* untuk menilai hasil ringkasan secara sistematis dengan menghitung *precision, recall* dan *f-score* dari hasil buatan sistem dan hasil buatan responden. Hasil evaluasi *ROUGE* yang telah didapat dihitung menggunakan *kappa score* untuk mendapatkan *gold standar* berupa nilai valid sebagai nilai akhir evaluasi peringkasan teks. Adapun evaluasi *review* berdasarkan pendapat pribadi dari responden, dimana responden akan melakukan evaluasi pada *google-form* terkait dengan dokumen yang sama, *review* tersebut akan dijadikan sebagai kesimpulan dan saran untuk mengetahui keefektifan sistem yang sudah dibuat.

1.7. Sistematika Penulisan

Sistematika penulisan ini dibagi menjadi 5 bagian utama, dimana setiap bab-nya meliputi :

Bab I Pendahuluan, pada bab ini memaparkan gambaran umum penulisan terkait *Text Summarization* dengan metode *LexRank* yang cakupan sub-bab terdiri dari latar belakang masalah, rumusan masalah, batasan masalah, tujuan penulisan, manfaat penulisan, metodologi penulisan dan sistematika penulisan.

Bab II Tinjauan Pustaka dan Landasan Teori, pada bab ini dijabarkan berbagai jurnal pendukung, definisi yang digunakan, serta formula yang dijadikan acuan untuk penulisan.

Bab III Metodologi Penulisan, pada bab ini berupa cakupan metode yang digunakan pada penelitian ini meliputi: deskripsi penulisan, analisis kebutuhan, diagram alur, tahapan penulisan dan tahapan evaluasi.

Bab IV Implementasi dan pembahasan, pada bab ini meliputi : Antarmuka aplikasi, *unique token processing*, hasil *pre-processing* seperti *original sentences*, *pre-processing sentences* dan *phrase sentences*, *processing Lexrank* : TF-IDF, *Cosine Similarity* (*idf-modified-cosine*), *power methods*, *evaluation* (*ROUGE*) dan *gold standar (kappa score)* .

Bab V Kesimpulan dan Saran, pada bab ini berupa pernyataan singkat dari penulis berdasarkan penjabaran dari hasil penelitian berupa implementasi program peringkasan yang sudah dilakukan dan saran-saran ke depan terkait penelitian serupa.

BAB V
KESIMPULAN DAN SARAN

5.1 Kesimpulan

Berdasarkan penelitian yang sudah dilakukan, ada beberapa hal yang bisa disimpulkan diantaranya sebagai berikut:

1. Dari 60 artikel yang diambil dari website korpus dokumen penulis mengambil tiap 5 artikel pada 3 kategori (dalang, tari tradisional dan wayang) untuk proses peringkasan teks. Artikel tersebut telah dievaluasi dan divalidasi sebelum di input ke system peringkasan.

2. Ketiga sentences yang dibuat yaitu *original sentence*, *pre-processing sentence* dan *phrase sentence*, dua diantaranya yaitu *pre-processing sentence* dan *phrase sentence* hanya sebagai fitur "**eksperiman**" dalam penulisan ini dan tidak direkomendasikan sebagai hasil output untuk dibaca user.

3. Evaluasi menggunakan *ROUGE* menunjukkan bahwa, original sentence memiliki hasil *f-score* yang mendekati ringkasan responden dengan nilai range 0.5-0.6 untuk setiap kategori (dalang, tari tradisional dan wayang).

4. Representasi berbasis grafik pada metode *LexRank* dan *TextRank* untuk peringkasan *single documents* dengan menghitung bobot per-kalimat serta pengurutan ringkasan berdasarkan indeks dinilai penulis sebagai faktor utama hasil evaluasi pembanding yang tidak terlalu jauh , meliputi *LexRank* dengan *f-score* yang rangenya 0.4 -0.6 untuk *ROUGE* – 1 dan *ROUGE* – 2

5. Keragaman hasil evaluasi disebabkan karena ringkasan ideal tidak dari satu responden yang sama namun dari beberapa responden sehingga hasil *ROUGE*-nya berbeda pada tiap artikelnya.

5.2 Saran

Adapun saran-saran yang penulis sampaikan untuk pengembangan penulisan serupa kedepennya yaitu :

1. Karena penulisan ini hanya menggunakan 3 kategori (dalang, tari tradisional dan wayang) dirasa perlu dilakukan penulisan untuk ketegori lain khususnya untuk budaya Indonesia.
2. Proses unique token pada penulisan ini dinilai penulis masih manual sehingga diharapkan untuk penulisan selanjutnya dapat menerapkan proses *mechine learning* untuk menentukan *unique token* secara otomatis.

DAFTAR PUSTAKA

Erkan, G., & Radev, D. R. (2004). LexRank: Graph-based lexical centrality as salience in Text Summarization. Journal of artificial intelligence research, 22, 457-479.

Maysyarah, N. (2019). Peringkasan Informasi Otomatis Berita Kriminal Online Bahasa Indonesia Menggunakan LexRank Algorithm

Vashisht, A. (n.d.). opengenus.org. Retrieved from LexRank method for Text Summarization: https://iq.opengenus.org/LexRank-text-summarization/

Kumar, Y. J., & Salim, N. (2012). Automatic multi document summarization approaches. In KS Gayathri, Received BE degree in CSE from Madras University in 2001 and ME degree from Anna University, Chennai. She is doing Ph. D. in the area of Reasoning in Smart.

Uçkan, T., & Karcı, A. (2020). Extractive multi-document Text Summarization based on graph independent sets. *Egyptian Informatics Journal*, *21*(3), 145-157.

Agastya, I. M. A. (2018). Pengaruh Stemmer Bahasa Indonesia Terhadap Peforma Analisis Sentimen Terjemahan Ulasan Film. *Jurnal Tekno Kompak*, *12*(1), 18-23.

Marsyah, Y. Perbandingan Kinerja Algoritma TextRank dengan Algoritma LexRank pada Peringkasan Dokumen Bahasa Indonesia.

Widman, M. (2020, September 21). *Cohen's Kappa: what it is, when to use it, how to avoid pitfalls*. Retrieved from Knime: https://www.knime.com/blog/cohens-kappa-an-overview

Oktriwina, A. S. (2021, Feb 02). glints.com. Retrieved from NLP: Kecerdasan Buatan yang Bantu Komputer Pahami Bahasa Manusia: https://glints.com/id/lowongan/natural-language-processing-adalah/

Putranto, H. A., Setyawati, O., & Wijono, W. (2016). Pengaruh Phrase Detection dengan POS-Tagger terhadap Akurasi Klasifikasi Sentimen menggunakan SVM. Jurnal Nasional Teknik Elektro dan Teknologi Informasi (JNTETI), 5(4), 252-259.

Setyaningsih, E. R. (2017). Part of Speech Tagger Untuk Bahasa Indonesia Dengan Menggunakan Modifika

Nugroho, K. S. (2019, Juni 18). *Medium.com*. Retrieved from Dasar Text *Pre-processing* dengan Python: https://ksnugroho.medium.com/dasar-text-*pre-processing*-dengan-python-a4fa52608ffe

Zalwert, M. (2021, May 5). Retrieved from LexRank algorithm explained: a step-by-step tutorial with examples: https://maciejzalwert.medium.com/lexrank-algorithm-explained-a-step-by-step-tutorial-with-examples-3d3aa0297c57

Kupiec J, Pedersen J, Chen F. 1995. *A trainable document summarize*r. Proceedings of the 18th annual international ACM SIGIR conference on Research and development in information retrieval; 1995 Jul 9-13; Seattle, Amerika Serikat. Seattle (US): ACM. hlm 68-73

Hermawan, L., Ismiati, M. B., Bangau, J., 60, N., & Charitas, M. (2020). Pembelajaran Text Preprocessing berbasis Simulator Untuk Mata Kuliah Information Retrieval. *TRANSFORMATIKA*, *17*(2), 188–199.

Klymenko, O., Braun, D., & Matthes, F. (2020). Automatic text summarization: A state-of-the-art review. *ICEIS 2020 - Proceedings of the 22nd International Conference on Enterprise Information Systems*, *1*, 648–655. https://doi.org/10.5220/0009723306480655

Lin, C.-Y., & Hovy, E. (n.d.). *From Single to Multi-document Summarization: A Prototype System and its Evaluation*.