

ANALISIS KUALITAS RINGKASAN TEKS BERBAHASA INDONESIA ANTARA PENGABUNGAN LINIER GRAPH BASED SUMMARIZATION ALGORITHM DIBANDINGKAN DENGAN PENGABUNGAN GRAPH BASED SUMMARIZATION ALGORITHM DENGAN FUZZY SYSTEM

Eny Widaryanti¹, Warih Maharani², Shaufiah³

¹Teknik Informatika, Fakultas Teknik Informatika, Universitas Telkom

Abstrak

Kebutuhan user akan informasi yang berupa artikel berita menyebabkan user menghabiskan waktu yang cukup lama untuk membaca keseluruhan isi artikel berita tersebut. Oleh karena itu, diperlukan adanya solusi agar user tidak perlu membaca keseluruhan isi artikel berita akan tetapi user tetap mendapatkan informasi yang singkat dan padat yang merepresentasikan isi artikel berita tersebut. Solusinya adalah sebuah sistem yang mampu menyaring informasi paling penting dari sebuah sumber atau beberapa sumber untuk membuat sebuah versi ringkas dari teks dengan memanfaatkan aplikasi yang dijalankan pada computer yaitu, sistem peringkasan teks otomatis (automatic text summarization).

Dalam Tugas Akhir ini diimplementasikan graph-based summarization algorithm dan similarity yaitu Textrank dan Similarity with Title yang mengambil konsep perankingan graf untuk melakukan perankingan kalimat dalam pembangunan aplikasi peringkasan teks otomatis. Aplikasi peringkasan teks otomatis menghasilkan dua jenis ringkasan ekstraksi, pertama yaitu hasil ringkasan dengan penggabungan linier Textrank dan Similarity with Title dan yang kedua adalah hasil ringkasan penggabungan Textrank dan Similarity with Title dengan fuzzy system. Peringkasan teks otomatis melakukan proses penghitungan skor setiap kalimat dalam sebuah teks dan memberikan rangking terhadap setiap kalimat berdasarkan skornya. Hasil keluaran aplikasi berupa ringkasan ekstraktif yang terdiri dari kalimat-kalimat beranking tinggi.

Pengujian dilakukan menggunakan ROUGE evaluation toolkit. Hasil pengujian menunjukkan bahwa hasil ringkasan penggabungan Textrank dan Similarity with Title dengan fuzzy system memiliki akurasi lebih baik daripada penggabungan linier TextRank dan similarity with title. Parameter-parameter fuzzy juga memberikan pengaruh terhadap hasil ringkasan teks. Selain itu, hasil ringkasan penggabungan Textrank dan Similarity with Title dengan fuzzy system memiliki akurasi lebih baik dengan kompresi 10% dibandingkan dengan kompresi 25%.

Kata Kunci : peringkasan teks, TextRank, similarity with title, fuzzy system

Telkom
University

Abstract

The users need of information such as news article make them spend a long time to read those whole news article. For that reason, a solution is needed so users do not need to read the whole news article but still the users get the brief information that represent the news article. The solution is a system which can filter the most important information from a source or some source to create a shortened version of a text by a computer program, that is automatic text summarization system.

This final assignment implements graph-based summarization algorithm and similarity they are textrank and similarity with title applying graph-based ranking concept for sentences to build an automatic text summarization application. Automatic text summarization application produce two kinds of extraction summary, the first is a summary with linear combination of TextRank and similarity with title and the second one is a summary with fuzzy system combined of TextRank and similarity with title. Automatic text summarization scoring each sentences of a text and rank them according to their score. The output of the application is an extractive summary which consist of sentences with high scores.

The evaluation of the summaries using ROUGE evaluation toolkit. The result of the experiment shows that the accuracy of a summary with fuzzy system combined of TextRank and similarity with title is better than a summary with linear combination of TextRank and similarity with title. The fuzzy system parameters also give influence on the summary result. Besides, a summary with fuzzy system combined of TextRank and similarity with title has a better accuration within 10% of compression rate than within 25% of compression rate.

Keywords : text summarization, Textrank, Similarity with Title, Fuzzy system

1. Pendahuluan

1.1 Latar belakang

Perkembangan teknologi dan ilmu pengetahuan mengakibatkan ketersediaan informasi meningkat, terutama informasi yang bersifat dinamis yang dapat diakses secara *online*. Informasi *online* dapat berupa artikel berita, dokumen-dokumen, video klip, film, musik, dan lain sebagainya. Kebutuhan user akan informasi yang berupa artikel berita menyebabkan user menghabiskan waktu yang cukup lama untuk membaca keseluruhan isi artikel berita tersebut. Oleh karena itu, diperlukan adanya solusi agar user tidak perlu membaca keseluruhan isi artikel berita akan tetapi user tetap mendapatkan informasi yang singkat dan padat yang merepresentasikan isi artikel berita tersebut. Solusinya adalah sebuah sistem yang mampu meringkas artikel berita yaitu, sistem peringkasan teks otomatis (*automatic text summarization*). Peringkasan untuk artikel berita sesuai diterapkan pada dokumen tunggal (*single document*) karena prosesnya yang cepat.

Summary atau ringkasan didefinisikan sebagai sebuah teks yang dihasilkan dari satu atau lebih teks sumber yang mengandung informasi signifikan/penting dari teks sumber dan panjangnya tidak lebih dari setengah teks sumber[5]. Ringkasan dapat diklasifikasikan dalam dua kategori, yaitu ekstraksi dan abstraksi. Ringkasan ekstraksi merupakan sejumlah kalimat atau frase yang terpilih dari teks sumber dengan nilai tertinggi dan diletakkan bersama pada sebuah teks baru yang lebih pendek tanpa mengubah isi teks. Sedangkan ringkasan abstraksi menggunakan metode linguistik untuk memeriksa dan menginterpretasikan teks. Sebagian besar sistem peringkasan teks otomatis menggunakan metode ekstraksi untuk menghasilkan ringkasan. Peringkasan teks otomatis bekerja optimal pada dokumen-dokumen yang terstruktur dengan baik seperti pada sejumlah artikel, berita, laporan, dan karya ilmiah.[17]

Graph-based summarization algorithm atau peringkasan teks berbasis graf merupakan suatu metode peringkasan teks yang *language independent* yang dapat menghasilkan ringkasan ekstraktif. Graf tekstual merupakan teks sumber yang direpresentasikan menjadi sebuah graf. Pembangunan graf tekstual menggunakan konsep *similarity* antar unit teks. *Vertex* pada graf tekstual dapat berupa unit-unit teks seperti kata-kata, kalimat-kalimat, atau paragraf-paragraf dalam teks. *Edge* dalam graf tekstual menunjukkan keterhubungan antar *vertex*. Keterhubungan antar *vertex* dapat berupa *similarity* antar kalimat ataupun hubungan leksikal atau gramatikal antar kata/frasa.[13]

Dewasa ini sudah banyak *tools* untuk peringkasan teks, tetapi belum banyak yang membahas kualitas ringkasan teks. Kualitas ringkasan teks sangat penting karena melihat tingginya tingkat kepentingan suatu dokumen. Dimana ketika menghadapi banyak dokumen yang harus dibaca dalam waktu singkat, ringkasan teks yang relevan dengan dokumen sangat diperlukan.[17] Teori *fuzzy set* dapat merepresentasikan dan menangani masalah ketidakpastian yang dalam hal ini dapat berarti keraguan, ketidaktepatan, kekuranglengkapan informasi, dan kebenaran yang bersifat sebagian[18]. Dalam tugas akhir ini, penulis menggunakan *fuzzy system* untuk melakukan ekstraksi kalimat-kalimat penting dalam peringkasan teks.

Dalam tugas akhir ini, penulis akan menganalisis kualitas ringkasan antara penggabungan linier *Graph-Based Summarization Algorithm* dibandingkan dengan *Graph-Based Summarization Algorithm* dengan *fuzzy system* dilihat dari *precision*, *recall*, dan *f-measure* menggunakan *ROUGE Evaluation toolkit*. Sehingga dari analisis perbandingan kedua metode ini akan diketahui kualitas ringkasan yang lebih baik.

1.2 Perumusan masalah

Permasalahan yang menjadi objek penelitian pada tugas akhir ini adalah

1. Bagaimana menerapkan penggabungan *Graph-Based Summarization Algorithm* dengan *Fuzzy System* terhadap dokumen tunggal berbahasa Indonesia, khususnya artikel berita?
2. Bagaimana menganalisa pengaruh parameter-parameter *fuzzy* untuk ringkasan teks dengan penggabungan *Graph-Based Summarization Algorithm* dengan *Fuzzy System*?
3. Bagaimana kualitas ringkasan teks antara penggabungan linier *Graph-Based Summarization Algorithm* dibandingkan dengan penggabungan *Graph-Based Summarization Algorithm* dengan *Fuzzy System*?

Dalam tugas akhir ini, peringkasan teks dilakukan dengan batasan masalah sebagai berikut:

1. Teks sumber yang akan diringkas adalah artikel berita berbahasa Indonesia yang diperoleh dari *website* berita atau koran yang berupa *softcopy* dokumen teks.
2. Peringkasan teks dilakukan secara *offline*.
3. Peringkasan teks yang dilakukan adalah peringkasan dokumen tunggal.
4. Hasil peringkasan berupa ekstraksi dari teks sumber.
5. Tidak melakukan proses *stemming* terhadap teks masukan.
6. Metode yang digunakan adalah *Fuzzy Based Method*.
7. Evaluasi dilakukan dengan membandingkan *content overlap* antara hasil peringkasan otomatis dan ringkasan referensi dengan menggunakan *ROUGE (Recall-Oriented Understudy for Gisting Evaluation) evaluation toolkit* [3].

1.3 Tujuan

Tujuan tugas akhir ini adalah:

1. Menerapkan penggabungan *Graph-Based Summarization Algorithm* dengan *fuzzy system* terhadap dokumen tunggal berbahasa Indonesia, khususnya artikel berita.
2. Menganalisa pengaruh parameter-parameter *fuzzy* untuk ringkasan teks dengan penggabungan *Graph-Based Summarization Algorithm* dengan *fuzzy system*.
3. Membandingkan kualitas ringkasan teks antara penggabungan linier *Graph-Based Summarization Algorithm* dengan penggabungan *Graph-Based Summarization Algorithm* dengan *Fuzzy system*. Kualitas ringkasan teks dilihat dengan *ROUGE evaluation toolkit*.

1.4 Metodologi penyelesaian masalah

Metodologi yang dilakukan untuk menyelesaikan permasalahan dengan menggunakan langkah-langkah sebagai berikut:

1. Studi literatur dan pengumpulan data

Mengumpulkan bahan-bahan referensi yang menunjang proses penelitian, yaitu yang berhubungan dengan peringkasan teks dan *fuzzy system*. Pengumpulan data - data yang diperlukan untuk mendukung penyelesaian masalah, seperti teks berita untuk digunakan sebagai input.

2. Analisis dan perancangan sistem

Melakukan analisis dan perancangan terhadap sistem yang dibangun, menganalisis metode yang akan digunakan untuk menyelesaikan permasalahan, termasuk menentukan bahasa pemrograman yang digunakan, arsitektur, fungsionalitas, dan antarmuka sistem. Input sistem berupa dokumen teks artikel berita. Output dari sistem berupa teks hasil ringkasan artikel berita.

3. Implementasi dan pembangunan sistem

a. Membangun sistem peringkasan teks otomatis dengan mengimplementasikan *Graph-Based Summarization Algorithm* berdasarkan analisis dan perancangan yang dibuat, dengan tahapan:

- Membangun sistem untuk mencari skor ringkasan teks dengan *Textrank*.
- Membangun sistem untuk mencari skor ringkasan teks dengan *Similarity with title*.
- Membangun sistem untuk mencari skor ringkasan teks dengan penggabungan linier skor *TextRank* dan *similarity with title*.

b. Membangun sistem peringkasan teks otomatis dengan penggabungan *Graph-Based Summarization Algorithm* dengan *fuzzy system* berdasarkan analisis dan perancangan yang dibuat, dengan tahapan:

- Membangun sistem *fuzzy* untuk mencari skor gabungan antara *Textrank* dan *Similarity with title*.

4. Pengujian dan analisis

Pengujian dan analisis dilakukan oleh penulis tanpa melibatkan pihak lain, meliputi:

- a. Pengujian untuk mengetahui kualitas ringkasan teks dengan penggabungan linier *Graph-Based Summarization Algorithm*.
- b. Pengujian untuk mengetahui kualitas ringkasan teks dengan penggabungan *Graph-Based Summarization Algorithm* dengan *fuzzy system*.
- c. Analisis pengaruh parameter *fuzzy* terhadap hasil ringkasan.
- d. Analisis perbandingan kualitas ringkasan teks antara teks dengan *Graph-Based Summarization Algorithm* dengan teks dengan penggabungan *Graph-Based Summarization Algorithm* dengan *fuzzy system* berdasarkan hasil pengujian *precision* dan *recall* dengan *ROUGE evaluation toolkit*.

5. Pengambilan kesimpulan dan saran, serta penyusunan laporan Tugas Akhir.

5. Penutup

5.1 Kesimpulan

Berdasarkan analisis terhadap hasil pengujian, diperoleh kesimpulan sebagai berikut:

1. Parameter *range* pada fungsi keanggotaan dalam *fuzzy system* mempengaruhi hasil ringkasan karena menentukan inputan untuk proses *fuzzy inference* dengan *fuzzy rules* untuk mendapatkan skor akhir kalimat.
2. *Fuzzy rules* pada *fuzzy system* mempengaruhi hasil ringkasan karena *fuzzy rules* digunakan untuk menentukan variabel output linguistik pada proses *defuzzification* yang menentukan skor akhir kalimat untuk peringkasan.
3. Penggabungan *Graph based summarization algorithm* dengan *fuzzy system* memberikan akurasi yang lebih baik dibandingkan dengan penggabungan linier karena berdasarkan hasil pengujian menunjukkan bahwa nilai rata-rata *recall (R)*, *precision (P)*, dan *f-measure (F)* pada penggabungan dengan *fuzzy system* lebih besar dibandingkan dengan pada penggabungan linier.
4. Hasil uji coba terhadap 26 data uji menunjukkan bahwa akurasi hasil peringkasan lebih baik pada kompresi 10% daripada kompresi 25%.

5.2 Saran

Berdasarkan hasil analisis dan kesimpulan, terdapat beberapa saran untuk perbaikan pada penelitian peringkasan teks sebagai berikut:

1. Mendefinisikan *fuzzy rules* dan *fuzzy set* menggunakan *Evolutionary Algorithm (EA)* atau *Artificial Neural Network (ANN)* untuk mendapatkan *fuzzy rules* dan *fuzzy set* yang optimum.

Telkom
University

Referensi

- [1] Binwahlan, M.S. dan Naomi Salim dan L. Suanmali. *Intelligent Model for Automatic Text Summarization*. *Information Technology Journal* 2009. Didownload dari <http://scialert.net/pdfs/itj/2009/1249-1255.pdf> pada tanggal 23 Oktober 2009.
- [2] Chin-Yew Lin. 2004. *ROUGE: A Package for Automatic Evaluation of Summaries*. Didownload dari <http://www.aclweb.org/anthology-new/W/W04/W04-1013.pdf> pada tanggal 18 September 2008.
- [3] Chin-Yew Lin. 2004. *ROUGE Working Note*. Didownload dari research.microsoft.com/en-us/um/.../ROUGE-Working-Note-v1.3.1.pdf pada tanggal 4 Mei 2011.
- [4] Dragomir R Radev. 2004. *Text summarization: Tutorial ACM SIGIR*. CLAIR: Computational Linguistics And Information Retrieval group University of Michigan. Didownload dari www.summarization.com/sigirtutorial2004.ppt pada tanggal 16 Mei 2008.
- [5] E. H. Hovy. 2001. *Automated Text Summarization*. Handbook of computation linguistics, Oxford University Press. Didownload dari <http://www.isi.edu/natural-language/people/hovy/papers/05Handbook-Summ-hovy.pdf> pada tanggal 16 Oktober 2008.
- [6] Firdaus, Yanuar. 2008. *Introduction to Information Retrieval*. Institut Teknologi Telkom Bandung.
- [7] H. P. Edmondson. 1969. *New methods in automatic extracting*. JACM: Journal of the ACM. Didownload dari <http://courses.ischool.berkeley.edu/i256/f06/papers/edmonson69.pdf> pada tanggal 25 Januari 2009.
- [8] Hongyan Jing, et. Al. 1998. *Summarization Evaluation Methods: Experiments and Analysis*. Didownload dari <https://eprints.kfupm.edu.sa/67426/1/67426.pdf> pada tanggal 25 Januari 2009.
- [9] Inderjeet Mani, Mark T. Maybury. 2001. *Automatic Summarization: Tutorial Notes*. Didownload dari <http://mitre.org/resources/centers/it/maybury/summarization/summarization.htm> pada tanggal 6 Juli 2008.
- [10] Lin Ziheng. 2007. *Graph-Based Methods for Automatic Text Summarization*. Department of Computer Science School of Computing National University of Singapore. Didownload dari <http://wing.comp.nus.edu.sg/~linzihen/publications/HYP-thesis.pdf> pada tanggal 17 September 2008.
- [11] Marjan, Ali. 2009. *ROUGE-NW Sebagai Penggabungan ROUGE-N dan ROUGE-W untuk Mengevaluasi Automatic Text Summarization (ATS)*. Informatika Institut Teknologi Telkom Bandung.
- [12] Purwasih ,Nurzaitun. 2009. *Peringkasan Teks Otomatis Dokumen Tunggal Berbahasa Indonesia Menggunakan Graph-Based Summarization Algorithm dan Similarity(Study Kasus Artikel Berita)*. Fakultas Informatika Institut Teknologi Telkom Bandung.
- [13] Rada Mihalcea. 2004. *Graph-based Ranking Algorithms for Sentence Extraction, Applied to Text Summarization*. Didownload dari <http://www.cs.unt.edu/~rada/papers/mihalcea.acl2004.pdf> pada tanggal 23 Juni

- 2008.
- [14] Rada Mihalcea. 2004. *Language Independent Extractive Summarization*. Didownload dari <http://www.aclweb.org/anthology-new/P/P05/P05-3013.pdf> pada tanggal 16 Oktober 2008.
- [15] Rada Mihalcea, Paul Tarau. 2004. *TextRank: Bringing Order into Texts*. Didownload dari <http://www.cs.unt.edu/~rada/papers/mihalcea.emnlp04.pdf> pada tanggal 16 Oktober 2008.
- [16] Sergey Brin, Lawrence Page. 1998. *The Anatomy of a Large-Scale Hypertextual Web Search Engine*. *Computer Networks and ISDN Systems*, 30(1-7). Didownload dari <http://infolab.stanford.edu/pub/papers/google.pdf> pada tanggal 16 Oktober 2008.
- [17] Suanmali, Ladda dan Naomi Salim dan Mohammed Salem Binwahlan. *Fuzzy Logic Based Method for Improving Text Summarization*. *International Journal of Computer Science and Information Security* 2009. Didownload dari <http://arxiv.org/ftp/arxiv/papers/0906/0906.4690.pdf> pada tanggal 9 Oktober 2009.
- [18] Suyanto. 2007. *Artificial Intelligence : Searching, Reasoning, Planning, and Learning*. Bandung: Informatika.
- [19] Taher H. Haveliwala. 2003. *Topic-Sensitive PageRank: A Context-Sensitive Ranking Algorithm for Web Search*. Didownload dari <http://www-cs-students.stanford.edu/~taherh/papers/topic-sensitive-pagerank-tkde.pdf> pada tanggal 25 Januari 2009.
- [20] _____. Automatic summarization. Didownload dari http://en.wikipedia.org/wiki/Automatic_summarization.htm pada tanggal 31 Mei 2008.
- [21] _____. Graph (mathematics). Didownload dari [http://en.wikipedia.org/wiki/Graph_\(graph_theory\).htm](http://en.wikipedia.org/wiki/Graph_(graph_theory).htm) pada tanggal 10 Januari 2009.