PENERAPAN ALGORITMA SCORE-BASED PADA PERINGKASAN TEKS CERPEN OTOMATIS

Malak Diana D[1], Bagas Satya D. N[2], Faisal Rahutomo[3]

[1,2,3] Teknik Informatika, Teknologi Informasi, Politeknik Negeri Malang
[1,2,3] Program Studi Teknik Informatika, Jurusan Teknologi Informasi, Politeknik Negeri Malang
[1] malakdiana00@gmail.com, [2]bagasnugraha@polinema.ac.id, [3]faisal@polinema.ac.id

***Abstrak*—.Cerita pendek banyak dijadikan sebagai bacaan yang menghibur bagi seluruh kalangan usia. Dalam penyajian cerpen, seringkali penamaan judul dan pemberian kategori kurang mewakili apa isi dari cerita yang ditulis. Salah satu alternatif untuk memahami isi cerita tanpa harus membaca keseluruhan teks adalah dengan membaca ringkasan. Membaca ringkasan juga tidak membutuhkan waktu yang lama, sehingga mempermudah pembaca menentukan jenis cerita yang mereka cari. Untuk meringkas sebuah cerpen akan lebih efisien jika menggunakan suatu sistem peringkas teks otomatis. Dalam hal meringkas teks metode yang dapat digunakan yaitu algoritma *score-based*. Algoritma *score-based* tersebut digunakan untuk mencari inti dari teks dengan memberikan nilai untuk setiap kalimat dari perhitungan fitur yang telah ditentukan. Pengujian akurasi peringkasan teks cerpen menggunakan algoritma *score-based* dilakukan dengan membandingkan ringkasan sistem dengan ringkasan manual dari pakar dengan hasil rata-rata nilai *precision* sebesar 52, *recall sebesar* 44,88889, dan *f-measure sebesar* 46,65009. Pengujian kepuasan pembaca dilakukan menggunakan kuisioner dan diperoleh hasil sebesar 71,3%.**

***Kata kunci*—** ringkasan otomatis, cerpen, algoritma score-based

I. PENDAHULUAN

Dalam penyajian cerpen, seringkali penamaan judul dan pemberian kategori kurang mewakili apa isi dari cerita yang ditulis. Sehingga akan menimbulkan masalah ketika dalam mencari sebuah cerpen pembaca tidak memiliki dasar untuk menemukan cerita yang sedang dicari. Sehingga untuk mengetahui isi dari cerpen pembaca perlu memahami keseluruhan atau minimal sebagian dari cerpen tersebut yang tentunya membutuhkan waktu yang tidak singkat. Alternatif lainnya yaitu dengan membaca ringkasan atau sinopsis dari sebuah cerpen.

Ringkasan adalah sebuah penyajian peristiwa atau kejadian yang panjang disajikan secara singkat[1]. Membaca ringkasan juga tidak membutuhkan waktu yang lama, sehingga mempermudah pembaca menentukan jenis cerita yang mereka cari. Untuk meringkas sebuah cerpen akan lebih efisien jika menggunakan suatu sistem peringkas teks otomatis, yang perlu diperhatikan yaitu bahwa cerpen memiliki unsur-unsur instrinsik cerpen antara lain adalah tema, plot atau alur, penokohan, latar, sudut pandang, gaya bahasa dan pesan. Sehingga perlu memilih metode yang tepat dalam membangun sistem ini. Dalam hal meringkas teks metode yang dapat digunakan yaitu algoritma score-based.

Algoritma tersebut digunakan untuk mencari inti dari teks dengan mencari kedekatan-kedekatan antar kalimat dengan pemberian skor[2]. Berdasarkan uraian di atas, penelitian yang akan diajukan adalah suatu sistem yang memungkinkan untuk dapat meringkas teks cerpen secara otomatis menggunakan algoritma score-based yang dapat mewakili isi dari cerpen dan sekaligus untuk mengetahui tingkat keberhasilan peringkasan teks cerpen menggunakan metode tersebut.

II. RELAVANSI

Penelitian terkait dengan peringkasan teks otomatis dilakukan oleh I Wayan Adi Setyadi pada tahun 2018 yang membahas mengenai implementasi *automatic text summarization* menggunakan metode graph dan *ant colony optimization*. Pada penelitian tersebut pengujian hasil ringkasan dengan mencari kesamaan hasil ringkasan sistem dengan hasil ringkasan secara manual menggunakan cosine similarity memperoleh persentase kesamaan 76.3% [3].

Penelitian lain yang terkait peringkasan teks otomatis dilakukan oleh Ni Komang Widyasanti mengenai seleksi fitur bobot kata dengan metode TFIDF untuk ringkasan bahasa Indonesia. Penelitian tersebut dilakukan untuk meringkas 8 dokumen dengan kategori teks diantaranya fiksi cerita rakyat, fiksi dongeng, non fiksi berita, dan non fiksi artikel. Hasil dari penelitian tersebut adalah dimana secara keseluruhan uji coba mengandung rata-rata 31,875 kalimat dan 387,375 kata. Compression rate rata-rata untuk kalimat adalah 37,645% dan 41,2525% untuk kata. Berdasarkan uji coba, ringkasan cenderung menghasilkan hasil akhir lebih banyak untuk dokumen non-fiksi. Durasi rata-rata untuk seluruh dokumen uji yakni 68,25 detik. Rata-rata kuesioner memiliki nilai 61,94%, dengan nilai dokumen non-fiksi kembali lebih tinggi dibanding dengan dokumen fiksi [4].

Penelitian lainnya yaitu dilakukan oleh oleh Aziz Qaroush tahun 2018, mengenai perbandingan peringkasan teks Bahasa Arab menggunakan metode scored-based dan

Machine Learning. Metode Machine Learning menghasilkan hasil recall dan F-score dengan rata-rata 33% dan 14%. Sedangkan, metode scored-based menghasilkan hasil recall, presisi, dan F-Score dengan peningkatan rata-rata masing-masing 23%, 23%, dan 24% [2].

III. LANDASAN TEORI

A. Peringkasan Teks

Peringkasan teks (text summarization) merupakan suatu proses penyulingan sebagian besar informasi penting dari sumber (beberapa sumber) untuk 4 menghasilkan suatu ringkasan bagi pemakai atau pekerja tertentu (Indriani, 2014). Ketika proses peringkasan teks dilakukan oleh computer secara otomatis , maka disebut sebagai Automatic Text Summarization (peringkas teks otomatis – PTO). Peringkasan teks otomatis menghasilkan 2 (dua) keluaran yaitu ekstraksi dan abstraksi. Adapun pengertian peringkasan Ekstraksi adalah suatu proses untuk memilih beberapa kalimat tertentu dari sekumpulan data teks dokumen yang dimana nantinya kalimat yang terpilih ini akan digunakan sebagai ringkasan dari keseluruhan dokumen [5]

B. Cerpen

Cerpen (cerita pendek) adalah jenis karya sastra berbentuk prosa dan bersifat fiktif yang menceritakan/menggambarkan suatu kisah yang dialami oleh suatu tokoh secara ringkas disertai dengan berbagai konflik dan terdapat penyelesaian atau solusi dari masalah yang dihadapi [6].

C. Algoritma Score-based

Dalam Algoritma score-based, kalimat-kalimat penting diekstraksi berdasarkan total skor yang diberikan pada kalimat tersebut [2]. Kalimat yang memiliki skor tertinggi akan mewakili konten terpenting dari dokumen tersebut.

IV. METODE PENELITIAN

A. Teks Prepocessing

Tahap ini adalah tahap awal dalam hampir semua metode ringkasan. Tujuan utamanya adalah untuk menyiapkan dokumen input teks untuk diproses di tahap lain. Proses *preprocessing* ini meliputi *tokenizing, filtering, stopword removal dan stemming*

B. Ekstraksi Fitur

Untuk memaksimalkan hasil ringkasan dilakukan serangkaian fitur yaitu

- Fitur Frase-Kunci, fitur ini digunakan untuk menghitung daftar singkat istilah-istilah penting yang mewakili topik-topik utama dokumen. Persamaan yang digunakan dalam perhitungan ini adalah

  Fitur 1 $= \sum_{KPi \in Si}(KPFi * \sqrt{KPL} * PNV)$ (1)

  Dimana KPF adalah frekuensi frase kunci, KPL adalah panjang dari frase kunci dan PNV adalah nilai dari frase kunci. Setiap frase kunci yang terdapat dalam KBBI bernilai 2 dan apabila tidak bernilai 1.

- Lokasi kalimat, Dalam penulisan suatu teks, pentingnya suatu kalimat tergantung pada lokasinya dalam paragraf tersebut. Sehingga terbentuk suatu rumus yang diberikan sesuai dengan letak kalimat yaitu seperti berikut.

$$\text{Fitur 2}=\begin{cases} 3, & kata\ pertama\ di\ awal\ paragraf \\ 2, & kata\ pertama\ di\ akhir\ paragraf \\ 1, & kata\ pertama\ ditengah\ paragraf \\ \frac{1}{\sqrt{S_{in}}} & awal\ atau\ akhir\ pargraf \\ \frac{1}{\sqrt{S_{in}+P_{in}^2}} & kecuali\ awal\ dan\ akhir \end{cases} \quad (2)$$

- Kemiripan dengan judul, pentingnya fitur ini berasal dari gagasan bahwa jika sebuah kalimat terdiri dari kata-kata yang muncul dalam judul, maka kalimat tersebut memilik kemungkinan menjadi kalimat yang penting. Untuk menghitung skor kemiripan dengan judul menggunakan persamaan berikut

  $$fitur\ 3 = Kedekatan(Judul, S_i) * \sqrt{1 + (KPT\ \cap KPS_i)} \quad (3)$$

  Sedangkan untuk menghitung kesaaman dengan judul menggunakan persamaan cosinus seperti rumus 4.

  $$Kedekatan = \frac{\sum_{w \in S,T} tf_{w,s} * tf_{w,T}(isf_w)^2}{\sqrt{\sum_{S_i \in S}(tf_{S_i s} * isf_{S_i})^2}\sqrt{\sum_{T_i \in T}(tf_{T_i T} * isf_{T_i})^2}} \quad (4)$$

- Sentralitas Kalimat, Fitur ini didefinisikan untuk menghitung nilai kesamaan antara satu kalimat dengan kalimat lain dalam dokumen. Dengan demikian, dapat diketahui jika suatu kalimat mungkin menjadi pusat dari dokumen, dan ada kalimat mungkin menjadi kalimat penjelas. Persamaan yang digunakan yaitu

  $$fitur\ 4 = \frac{Kesamaan\ S_i}{kesamaan\ maksimal\ dalam\ dokumen} \quad (5)$$

- Panjang kalimat, Panjang kalimat mungkin memengaruhi arti pentingnya, semakin panjang kalimat tersebut memungkinkan kalimat untuk mengandung informasi lebih banyak. Skor kalimat dihitung dan dinormalisasi menggunakan persamaan berikut ini

  $$fitur\ 5 = \frac{\#\ kata\ dalam\ kalimat}{\#\ kata\ dalam\ kalimat\ terpanjang} \quad (6)$$

- Frasa isyarat, yang dimaksud adalah seperti pada kata “kesimpulannya”, “hasilnya”, “tepat”. Kalimat-kalimat yang mengandung kata isyarat akan diberikan skor yang lebih tinggi. Skor dari fitur ini dihitung dan dinormalisasi menggunakan persamaan

  $$fitur\ 6 = \frac{\#\ kata\ isyarat\ pada\ kalimat}{\#\ kata\ isyarat\ pada\ dokumen} \quad (7)$$

- Fitur kata kuat, beberapa kata digunakan untuk menekankan atau fokus pada ide inti dalam kalimat, sehingga kalimat yang mengandung kata inti perlu diberikan skor yang lebih tinggi. Skor dari fitur ini dihitung menggunakan persamaan berikut.

  $$fitur\ 7 = \frac{\#\ kata\ kuat\ pada\ kalimat}{\#\ kata\ kuat\ pada\ dokumen} \quad (8)$$

- Data numerik, keberadaan data numerik seperti angka, tanggal, dan waktu dapat mempengaruhi pentingnya kalimat. Skor dari fitur ini dihitung menggunakan persamaan berikut.

$$fitur\ 8 = \frac{\#\ kemunculan\ angka\ dalam\ kalimat}{\#\ kemunculan\ angka\ dalam\ dokumen} \quad (9)$$

- Kata lemah, dalam sebuah kalimat jika didalamnya memuat kata "misalnya","sebagai tambahan" adalah tanda bahwa kalimat tersebut berfungsi sebagai penjelas. Kata seperti itu melemahkan kalimat karena menyiratkan bahwa kalimat selanutnya adalah informasi tambahan, sehingga berpengaruh juga terhadap skor kalimat. Skor dari fitur ini dihitung menggunakan persamaan berikut

$$Skor\ kata\ lemah = \frac{\#\ kata\ lemah\ pada\ kalimat}{\#\ kata\ pada\ kalimat} \quad (10)$$

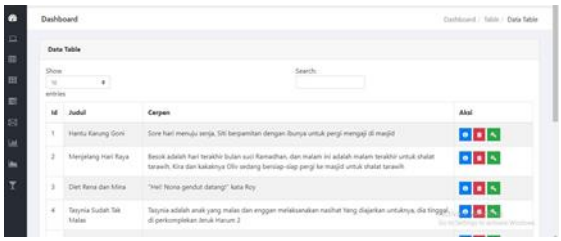
C. *Normalisai Kalimat.*

Setelah memperoleh nilai setiap fitur maka dilanjutkan menghitung nilai normalisasi yaitu dengan menjumlahkan seluruh nilai fitur yang telah dikalikan dengan bobot fitur. Pada penelitian ini, seluruh fitur dianggap memiliki bobot yang sama sehingga diberika nilai 1. Selanjutnya akan diperoleh skor perkalimat yang akan di rangking untuk mengetahui kalimat mana yang memiliki skor tinggi dan dianggap sebagai kalimat yang mewakili isi cerpen. Dari kalimat yang memiliki kor tinggi tersebut diurutkan sesuai letak kalimat dan terbentuklah sebuah ringkasan cerpen.

V. IMPLEMENTASI DAN PENGUJIAN

A. *Implementasi Interface Sistem*

Berikut ini adalah implementasi sistem ke dalam halaman web.

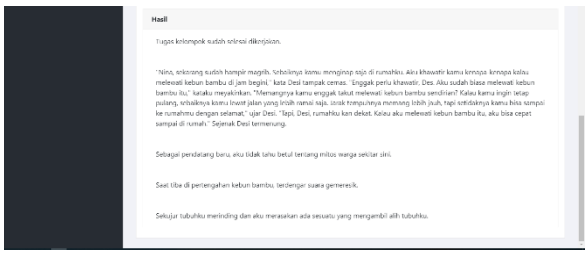


Gambar 1 Tampilan Daftar Cerpen

Cerpen yang telah didapatkan dari proses scraping akan ditampilkan di halaman cerpen seperti pada gambar 1. Admin dapat melihat detail cerpen, menghapus cerpen juga menambahkan kata kunci.

Gambar 2 Tampilan nilai fitur

Pada gambar 2 ditampilkan halaman ektraksi fitur yang menampilkan detail perhitungan dari masing-masing fitur ringkasan.



Gambar 3 Tampilan hasil ringkasan

Pada gambar 3 ditampilkan halaman hasil ringkasan. Pada halaman ini terdapat list cerpen yang sudah melalui proses ekstraksi fitur dan bisa dilihat hasil ringkasannya.

B. *Pengujian*

1) *Pengujian akurasi*

Untuk menguji hasil penerapan algoritma digunakan pengujian *precision*, *recall*, *f-measure*. *Recall* merupakan evaluasi untuk mengetahui tingkat keberhasilan kinerja sistem dalam menghasilkan teks ringkasan otomatis, untuk menghitung recall dapat menggunakan persamaan 11

$$\text{Recall} = \frac{\sum S1\ \cap \sum S2}{\sum S1} \quad (11)$$

Precision (P) adalah tingkat ketepatan hasil ringkasan yang dihasilkan oleh sistem peringkas otomatis, untuk menghitung nilai precision menggunakan persamaan 12

$$\text{Precision} = \frac{\sum S1\ \cap \sum S2}{\sum S2} \quad (12)$$

Dimana :
S1 : kalimat ringkasan sistem
S2 : kalimat ringkasan manual

Sedangkan *f-measure* (F) adalah nilai yang digunakan untuk menentukan akurasi dari ringkasan, *f-measure* diperoleh dari menggabungkan nilai dari *precision* dan *recall* , untuk menghitung nilai *f-measure* dapat menggunakan persamaan 13

$$F - measure = 2\ \text{x}\ \frac{Precison\ x\ Recall}{(Precison + Recall)} \quad (13)$$

Untuk menghitung tingkat keakuratan hasil ringkasan dilakukan pengujian dengan membandingkan hasil ringkasan sistem dengan ringkasan manual yang dibuat oleh pakar. Data yang digunakan berupa 10 data cerpen yang diambil secara acak. Berikut adalah hasil perbandingan hasil ringkasan:

Tabel 1 Hasil akurasi

No	Judul	*Precision*	*Recall*	*F-measure*
1	Memang untukku	60,00	50,00	54,54545
2	Kukira Ucapanmu Telah Usai	80,00	44,44	57,14286

3	Senyum Kakak Di Langit Malam	80,00	80,00	80
4	Kembang Api Pembawa Petaka	40,00	11,11	17,3913
5	Maaf	40,00	50,00	44,44444
6	Stange Thing	20,00	8,33	11,76471
7	Kebun Bambu	20,00	20,00	20
8	Hantu Karung Goni	60,00	60,00	60
9	Dan Ketika Itu	60,00	50,00	54,54545
10	Lukisan Hantu	60,00	75,00	66,66667
Rata-rata		52	44,88889	46,65009

Berdasarkan perhitungan nilai akurasi diatas didapatkan hasil rata-rata nilai *precision* sebesar 52, *recall* sebesar 44,88 dan *f-measure* sebesar 46,65. Dapat disimpulkan bahwa peringkasan teks cerpen otomatis menggunakan algoritma *score-based* mempunyai tingkat akurasi cukup.

2) Pengujian Kepuasan Pembaca

Untuk menghitung tingkat kepuasan pembaca mengenai hasil ringkasan cerpen dilakukan pengambilan nilai melalui kuisioner. Terdapat 25 cerpen beserta masing-masing ringkasan yang telah dihasilkan oleh sistem, masing-masing cerpen akan dibaca oleh 10 responden terpilih dan akan memberikan nilai dari rentang angka 1-10 yang memiliki arti bahwa semakin besar nilai maka semakin cocok ringkasan tersebut. Berikut adalah hasil rata-rata kuisioner :

Tabel 2 Hasil kuisioner kepuasan pembaca

No	Judul	Rata- rata Nilai
1	Hantu Karung Goni	7
2	Buah Dari Kesabaran	7,1
3	Dan Ketika itu	8,1
4	Cerita Mistis Hari ini	6,3
5	Kehidupan Diluar Alam Sadarku	7,3
6	Memang Untukku	8,3
7	Pilihan Hati	5,7
8	Lukisan Hantu	7
9	Kematian Di Waktu Senja	7
10	Kembang Api Pembawa Petaka	7,8
11	Kukira Ucapanmu Telah Usai	7,8
12	Kehampaan	6,4
13	Kebun Bambu	7,2
14	Maaf	7,8
15	Strange Thing	5,5
16	Mendekap Patahan Rindu	7,6
17	Anggraeni	6,8
18	Senyuman Kakak Di Langit Malam	8,3
19	Pelangi Hitam Putih	7,1
20	Janji Untuk Ibu	7,1
21	Dibalik Jendela	6,7
22	Fine	7,2
23	Kembali Bersama	7,3
24	Pertemuan Sesungguhnya	6,8
25	Melted	7,2
Rata-rata Akhir		7,136

Berdasarkan nilai perhitungan diatas didapatkan nilai persentase sebesar 71,36 % hal ini berarti hasil ringkasan yang dibuat oleh sistem ini dinilai dapat mewakili isi dari cerita pendek.

VI. KESIMPULAN DAN SARAN

A. Kesimpulan

Berdasarkan pembahasan dari penelitian ini dapat ditarik kesimpulan sebagai berikut :

1) Sistem yang dibuat dapat melakukan peringkasan teks cerpen secara otomatis
2) Pemilihan fitur berpengaruh pada hasil ringkasan.
3) Hasil pengujian untuk mengetahui akurasi hasil ringkasan menghasilkan nilai *precision* sebesar 52, *recall* sebesar 44,88 dan *f-measure* sebesar 46,65. Sedangkan pengujian dengan menggunakan kuisioner untuk mengetahui tingkat kepuasan pembaca adalah sebesar 71,3 % yang berarti ringkaan yang dihasilkan sistem cukup mewakili isi cerita pendek.

B. Saran

Berdasarkan penelitian yang dilakukan beserta masalah yang terjadi dapat diberikan saran sebagai berikut:

1) Menambahkan atau mengganti dengan fitur kalimat yang berbeda sesuai dengan kebutuhan teks yang diringkas.
2) Melakukan pengujian bobot untuk masing-masing fitur sehingga dapat diketahui fitur mana yang lebih penting.
3) Menggunakan metode lain sehingga dapat dibandingkan tingkat keakuratannya.
4) Mengembangkan proses peringkasan hingga menjadi ringkasan abstraktif.

DAFTAR PUSTAKA

[1] Pendidikan, S., & Lestari, E. P. (2016). *Pengaruh kemampuan membaca pemahaman terhadap hasil belajar menulis ringkasan siswa kelas iv sdn gugus robert wolter monginsidi kaliwungu kabupaten kendal*.

[2] Qaroush, A., Abu Farha, I., Ghanem, W., Washaha, M., & Maali, E. (2019). An efficient single document Arabic text summarization using a combination of statistical and semantic features. *Journal of King Saud University - Computer and Information Sciences*, (xxxx). https://doi.org/10.1016/j.jksuci.2019.03.010

[3] Setyadi, I. W. A., Khrisne, D. C., & Suyadnya, I. M. A. (2018). Automatic Text Summarization Menggunakan Metode Graph dan Metode Ant Colony Optimization, *17*(1), 124–130.

[4] Widyasanti, N. K., Gede, I. K., Putra, D., Kadek, N., & Rusjayanthi, D. (2018). Seleksi Fitur Bobot Kata dengan Metode TFIDF untuk Ringkasan Bahasa Indonesia. *Merpati*, *6*(2), 119–126.

[5] Prabowo, D. A., Fhadli, M., Najib, M. A., Fauzi, H. A., & Cholissodin, I. (2016). TF-IDF-Enhanced Genetic Algorithm Untuk Extractive Automatic Text Summarization. *Jurnal Teknologi Informasi Dan Ilmu Komputer*, *3*(3), 208. https://doi.org/10.25126/jtiik.201633217

[6] CERPEN. (2019, November 26). Diambil kembali dari yuksinau: https://www.yuksinau.id/cerpen-pengertian-ciri-unsur-struktur-fungsi/