

PERINGKAS DOKUMEN OTOMATIS MENGGUNAKAN METODE FUZZY MODEL SISTEM INFERENSI MAMDANI

Achmad Ridok¹⁾, Tri Cahyo Romadhona²⁾

¹⁾ PTIIK Universitas Brawijaya Malang

²⁾ Ilmu Komputer FMIPA UB

email : acridokb@ub.ac.id¹⁾

Abstrak

Peringkasan dokumen otomatis adalah pembuatan versi ringkas dari sebuah teks secara otomatis tanpa mengurangi makna asli yang terkandung didalamnya. Secara umum terdapat 2 metode yaitu ekstraksi dan abstraksi. Pada penelitian ini diterapkan metode ekstraksi dengan cara melakukan pembobotan kalimat yang diperoleh 6 fitur pada setiap kalimat, hasil nilai tiap fitur tersebut digunakan sebagai input nilai fuzzy menggunakan inferensi model mamdani. Hasil akhir yang didapat berupa nilai akhir bobot setiap kalimat. Ringkasan dibuat berdasarkan kalimat-kalimat dengan nilai tertinggi hingga terendah sesuai dengan besar ringkasan yang diinginkan. Hasil uji coba sistem dengan besar ringkasan 25% dan 50% dibandingkan dengan AutoSummarize (Microsoft Word) terhadap 15 dokumen berita yang diambil secara acak, didapatkan hasil masing-masing untuk rata-rata precision dan rata recall untuk sistem adalah 0.453001, 0.468134 dan 0.589149, 0.601721. Sedangkan AutoSummarizer masing-masing untuk rata-rata precision dan rata-rata recall sebesar 0.349274, 0.354928 dan 0.512515, 0.521096. Dengan demikian hasil ringkasan yang diperoleh sistem lebih baik dibandingkan dengan AutoSummarize.

Kata kunci :

Ringkasan teks, ekstraksi kalimat, fuzzy logic, sistem inferensi mamdani

1. Pendahuluan

Seiring dengan berkembangnya teknologi, perkembangan informasi juga berkembang dari hari ke hari. Kebutuhan akan informasi yang tepat dan cepat menuntut adanya alat untuk membuat ringkasan secara otomatis, sehingga informasi yang didapat lebih cepat dan tepat tanpa kehilangan makna asli dari dokumen asal[2].

Peringkasan dokumen otomatis yaitu meringkas sebuah dokumen teks yang dilakukan oleh mesin untuk menampilkan versi yang lebih ringkas dari versi aslinya. [15].

Secara umum terdapat metode yaitu ekstraksi dan abstraksi. Metode ekstraksi meringkas dokumen teks dengan cara merangkai kata yang diambil secara utuh dari dokumen asal. Sedangkan metode abstraksi membuat ringkasan dengan merangkai kata baru yang tidak terdapat pada dokumen asal. Metode yang terakhir

ini menggunakan konsep NLP (*Natural Language Processing*) [10].

2. Tinjauan Pustaka

2.1 Konsep Peringkasan Teks

Peringkasan teks otomatis merupakan metode untuk mengubah sebuah dokumen teks yang panjang dan besar menjadi dokumen yang lebih ringkas dengan mengambil poin-poin penting yang terkandung didalamnya[1]. Tujuan dari peringkasan teks yaitu mengambil intisari informasi dari dokumen sumber dan menjadikannya lebih singkat tanpa menghilangkan topik utama sehingga dapat membantu pengguna lebih cepat memahami makna yang terkandung didalamnya[10].

2.2 Text Mining

Data mining yaitu suatu cara untuk memperoleh sedikit data yang berharga dari sejumlah banyak data yang tak berharga. Konsep data mining berasal dari filosofi menambang batuan permata (*mining*) yaitu mencari secuil batuan mulia dari sekumpulan ribuan batu yang tak berharga. Dalam implementasinya, *text mining* secara garis besar dipisahkan menjadi dua tahap yaitu *preprocessing* dan *processing*. Pada tahap *preprocessing* dilakukan beberapa proses yaitu: *segmentation*, *tokenization* *stopword* dan *stemming*. Tahap *processing* merupakan tahap ini untuk menggali informasi yang dibutuhkan dari korpus teks.

2.3 Pembobotan Kalimat

2.3.1 Kemiripan Judul

Secara umum judul dari sebuah bacaan menggambarkan isi dari bacaan tersebut, dengan demikian sebuah kalimat yang mempunyai kata mirip dengan judulnya maka mempunyai peluang besar sebagai kalimat ringkasan[15].

$$F1(K) = \frac{\text{jumlah kata yang sama dengan judul}}{\text{jumlah kata dalam judul}} \quad (1)$$

2.3.2 Panjang Kalimat

Kalimat yang terlalu singkat tidak diharapkan masuk dalam ringkasan, kalimat seperti ini biasanya berisi cuma *dateline* dan nama pengarang dalam artikel berita[15].

$$F2(K) = \frac{\text{jumlah kata dalam kalimat}}{\text{jumlah kata dalam kalimat terpanjang}} \quad (2)$$

2. 3. 3
Posisi

Kalimat

Dalam paragraf berbahasa Indonesia kalimat topik terletak pada awal kalimat, akhir kalimat maupun di tengah paragraf. Karena sulitnya mengidentifikasi secara pasti dan jaranganya penggunaan kalimat topik yang terletak ditengah, maka pada penelitian ini hanya digunakan sistem rangking berdasarkan pada posisi kalimat dimana nilai tertinggi terletak pada kalimat pertama, nilai tertinggi kedua terletak di akhir kalimat[4].

$$\begin{aligned}
 F3(K) = & 5/5 \text{ untuk posisi awal,} \\
 & 4/5 \text{ untuk posisi akhir,} \\
 & 3/5 \text{ untuk posisi awal + 1,} \\
 & 2/5 \text{ untuk posisi akhir - 1,} \\
 & 1/5 \text{ untuk posisi awal + 2,} \\
 & 0/5 \text{ untuk posisi yang lain.} \quad (3)
 \end{aligned}$$

2. 3. 4 Bobot Kata

Bobot kalimat dapat dihitung dari jumlah tiap kata penyusunnya, dengan menerapkan metode *tf.isf* (*Term Frequency Inverse Sentence Frequency*) dapat dihitung nilai tiap kata[15].

$$W_i = tf_i.isf_i = tf_i \cdot \log \frac{N}{n_i} \quad (4)$$

$$F_4(K) = \frac{\sum_{i=1}^j W_i(S_i)}{\text{Max} \left(\sum_{i=1}^j W_i(S_i^N) \right)} \quad (5)$$

Dimana :

tf_i = frekuensi kata pada dokumen

N = jumlah kalimat

n_i = jumlah kalimat dimana kata ke i muncul

j = jumlah kata dalam kalimat

2. 3. 5 Huruf Besar

Kata yang mengandung huruf besar mempunyai peluang termasuk dalam kata yang dipentingkan[5].

$$F5(K) = \frac{\text{banyaknya kata yang terdapat huruf besar}}{\text{banyak kata dalam kalimat}} \quad (6)$$

2. 3. 6 Data Numerik

Kalimat yang yang mengandung data berupa bilangan mempunyai kemungkinan termasuk dalam kalimat ringkasan[4].

$$F6(K) = \frac{\text{banyaknya kata yang terdapat angka}}{\text{banyak kata dalam kalimat}}$$

2. 4 Logika Fuzzy

Dalam bahasa Inggris logika *fuzzy* diartikan sebagai logika kabur atau tidak jelas, yaitu logika yang mengandung unsur ketidakjelasan. Logika fuzzy merupakan suatu cara untuk memetakan suatu ruang input ke dalam ruang output. Logika fuzzy digunakan untuk menterjemahkan suatu besaran yang dapat diekspresikan dengan bahasa (*linguistic*)[12], misalkan besaran data tentang suhu yang dapat diekspresikan dengan dingin, sedang, panas maupun sangat panas.

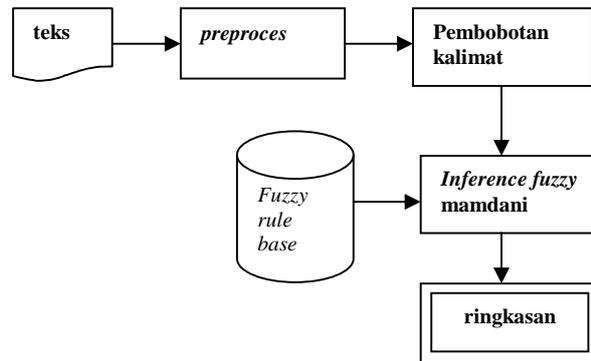
2. 5 Sistem Inferensi Fuzzy Metode Mamdani

Metode *mamdani* diperkenalkan oleh Ibrahim Mamdani pada tahun 1975. Metode *Mamdani* sering juga disebut metode Max-Min dengan aturan sebagai berikut :

1. Pembentukan Himpunan Fuzzy, pada Metode *Mamdani*, baik variable input maupun output dibagi satu atau lebih himpunan fuzzy.
2. Aplikasi fungsi Implikasi (Aturan), pada metode *Mamdani*, fungsi implikasi yang digunakan adalah minimum.
3. Komposisi Aturan, inferensi diperoleh dari kumpulan dari beberapa korelasi antar aturan. Aturan yang digunakan dalam melakukan inferensi sistem fuzzy yaitu menggunakan metode *MAX* (*Maximum*).
4. Penegasan (*Defuzzy*), suatu proses pengubahan *crisp input* menjadi *fuzzy input*.

3. Metode Penelitian

Dalam pelaksanaanya, sistem dibagi dalam tiga tahap. Tahap pertama dan kedua preproses dan proses. Tahap yang ketiga tahap perhitungan logika *fuzzy* dimana keluaran dari tahap proses ini digunakan sebagai inputan bagi *fuzzy* yang hasilnya akan diolah menggunakan mode *mamdani* dan hasil akhirnya akan didapat nilai bobot akhir pada tiap kalimat.



Gambar 1. Diagram Proses Umum Sistem

3. 1. PreProses

Proses awal dilakukan untuk mempersiapkan sebuah dokumen teks sebelum proses inti dijalankan. Dalam preproses ini terdapat beberapa tahap :

- a. Segmentasi, yaitu pemacahan dokumen teks menjadi paragraph, kalimat maupun kata.
- b. Hapus Tanda Baca, karakter-karakter selain huruf A-Z maupun 0-9 dihilangkan dari teks.
- c. Hapus kata-kata yang tidak penting seperti kata penghubung dan kata tanya yang dikenal dengan stopword. Daftar stopword yang dihilangkan pada penelitian ini menggunakan sumber dari Tala[17].
- d. Stemming, yaitu proses untuk menentukann kata dasar dari suatu kata-kata di dalam dokumen. Pada Penelitian ini digunakan algoritma Porter yang telah diadokpsi untuk stemming bahasa Indonesia oleh Tala [17].

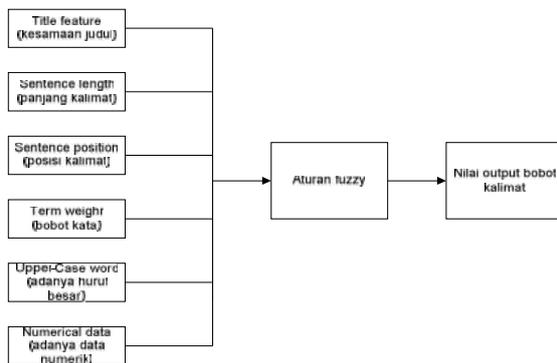
3. 1. 2 Proses

Pada bagian ini dilakukan perhitungan nilai bobot setiap kalimat dengan kriteria fitur berikut: fitur judul (*title feature*), panjang kalimat (*sentence length*), bobot kata (*term weight*), posisi kalimat (*sentence position*), kesamaan antar kalimat (*sentence to sentence similarity*) dan data numerik (*numerical data*).

- Hitung kemiripan masing-masing kalimat dengan judul menggunakan persamaan 1.
- Hitung perbandingan jumlah kata dalam kalimat menggunakan persamaan 2.
- Hitung posisi kalimat dalam suatu paragraph menggunakan persamaan 3.
- hitung bobot kata pada masing-masing kalimat menggunakan persamaan 5.
- Hitung peluang kata yang mempunyai huruf besar dengan jumlah kata dalam kalimat menggunakan persamaan 6.
- Hitung peluang data numeric menggunakan persamaan 7.

3. 1. 2 Perhitungan Logika Fuzzy

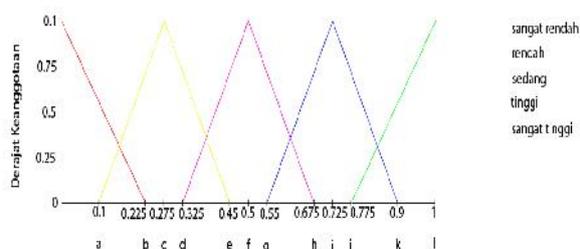
Setelah nilai fitur pada masing masing kalimat diperoleh, selanjutnya dilakukan perhitungan logika fuzzy untuk menentukan bobot akhir setiap kalimat sebagaimana digambarkan pada gambar 2.



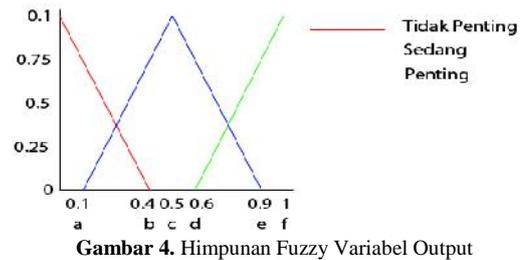
Gambar 2. Sistem Fuzzy Penentuan Nilai Bobot Kalimat

a. Proses Fuzzyfikasi

Pada proses ini nilai *crisp input* dari fitur kalimat diubah menjadi *fuzzy input* berdasarkan fungsi keanggotaan masing – masing fitur sebagaimana pada gambar 3.



Gambar 3 Grafik Keanggotaan Himpunan Fuzzy



Gambar 4. Himpunan Fuzzy Variabel Output

b. Rule Fuzzy

Aturan (*rule*) yang dipakai dalam *inferensi fuzzy*. contohnya:

IF Kesamaan Judul = Sangat Tinggi AND Panjang Kalimat = Sedang AND Posisi Kalimat = Sangat Tinggi AND Bobot Kata = Sedang AND Huruf Besar = Sangat Rendah AND Data Numerik = Sangat Rendah THEN Kalimat = Penting

c. Inferensi Mamdani

Pada proses inferensi mamdani, aturan tersebut digunakan dalam pengambilan keputusan menggunakan metode *Min-Max*. Kemudian fungsi komposisi yang digunakan adalah fungsi *Max*,

d. Defuzzifikasi

Defuzzifikasi adalah proses perubahan *crisp input* menjadi *fuzzy input* menggunakan metode *centroid* (*center of gravity*).

4. Hasil dan Pembahasan

4. 1. Skenario Uji Coba

Pengujian ini dilakukan terhadap 15 dokumen berita yang dipilih secara random masing – masing akan dibuat ringkasannya sebesar 50% dan sebesar 25%. Selanjutnya hasil sistem dibandingkan dengan hasil ringkasan *Autosummarizer*. Kedua hasil uji coba dibandingkan dengan hasil ringkasan manusia dalam hal ini seorang guru bahasa Indosia sebagai *benchmark*.

4. 2 Hasil Uji Coba

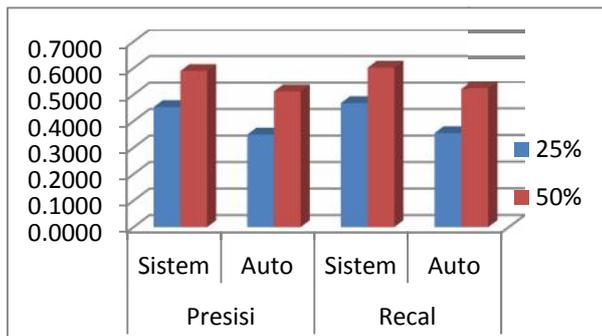
4. 2. 1 Hasil Pengujian

Hasil pengujian nilai precision dan recall pada masing – masing dokumen yang dengan besar ringkasan sebesar 25% dan 50% dari dokumen asal dengan nilai pembandingnya berasal dari hasil ringkasan buatan manusia. Maka didapat nilai rata-rata sebagaimana pada table 1.

Tabel 1. Perbandingan Hasil Sistem dengan *AutoSummarizer*

%	Presisi		Recal	
	Sistem	Auto	Sistem	Auto
25%	0.4530	0.3493	0.4681	0.3549
50%	0.5891	0.5125	0.6017	0.5211

Penyajian dengan representasi Grafik sebagaimana pada gambar 5.



Gambar 5. Perbandingan Hasil Sistem dengan AutoSummarizer

4. 2 Analisa hasil

Secara keseluruhan sistem peringkas dokumen yang menggunakan metode ekstraksi kalimat dan digabungkan dengan logika Fuzzy model mamdani menghasilkan nilai rata – rata dari precision sebesar 0.453001 dan recall sebesar 0.468134 untuk besar ringkasan sebesar 25% dari dokumen asal. Sedangkan untuk besar ringkasan sebesar 50% menghasilkan nilai rata - rata precision 0.589149 dan recall 0.601721. Nilai tersebut masih lebih baik jika dibandingkan dengan hasil ringkasan yang dibuat dari Autosummarizer yang menghasilkan nilai rata – rata dari precision sebesar 0.349274 dan recall 0.354928 untuk besar ringkasan sebesar 25% dan untuk hasil ringkasan sebesar 50% menghasilkan nilai rata – rata precision sebesar 0.512515 dan recall 0.521096.

Dari hasil uji yang dilakukan, dapat diketahui bahwa nilai rata-rata precision dan recall untuk hasil ringkasan sistem lebih besar dibandingkan nilai precision dan recall dari hasil ringkasan AutoSummarizer. Sedangkan untuk hasil besar ringkasan sebesar 50% nilai rata – rata precision dan recall lebih besar dibandingkan dengan hasil ringkasan sebesar 25%.

Ada beberapa hal yang mempengaruhi hasil uji coba tersebut, antara lain metode stemming yang digunakan, karena metode stemming yang digunakan hanya memotong awalan dan akhiran tanpa adanya pengecekan berupa kata dasar atau bukan. Misalkan kata “tabrakan” akan menghasilkan kata dasar “tabra” karena dipotong berupa akhiran “kan”, kata tersebut akan mempunyai arti yang berbeda dengan “menabrak” padahal bentuk kata dasar keduanya sama. Kekurangan pada metode ini akan berpengaruh pada fitur kesamaan judul dan fitur bobot kata, karena kedua fitur ini sama-sama menghitung nilai kesamaan antar kata.

Jumlah kalimat dalam satu paragraf juga berpengaruh pada hasil uji coba sistem, pada data uji dokumen berita banyak ditemui beberapa paragraf yang cuma terdiri dari 1 kalimat. Sehingga kalimat tersebut mempunyai nilai yang tinggi pada fitur posisi kalimat. Kerja sistem akan optimal pada fitur posisi kalimat apabila dalam satu paragraf terdiri lebih dari 5 kalimat. Hal lain yang juga berpengaruh pada hasil uji coba sistem yaitu pemilihan kandidat kalimat ringkasan. Sistem ini akan mengenali kalimat yang terdiri dari kata penting dari dokumen, sehingga kerja sistem akan

optimal jika kalimat tersebut mengandung deskripsi panjang mengenai sebuah dokumen yang terdiri dari banyak kata penting. Dalam hal ini sistem tidak bisa mengekstrak kalimat kandidat ringkasan yang berupa kalimat deskripsi singkat.

5. KESIMPULAN DAN SARAN

5. 1 Kesimpulan

Kesimpulan yang dihasilkan dalam skripsi ini adalah:

1. Peringkasan dokumen teks otomatis berita bahasa Indonesia dengan menggunakan gabungan pembobotan kalimat dan fuzzy model inferensi mamdani secara umum menghasilkan rata – rata precision dan recall untuk ringkasan sebesar 25% dan 50% masing-masing adalah 0.453001, 0.468134 dan 0.589149, 0.601721.
2. Setelah dilakukan perbandingan dengan sistem peringkas lain yaitu AutoSummarize (Microsoft Word 2007) dihasilkan kesimpulan bahwa sistem peringkas dokumen ini lebih baik.

5. 2 Saran

Beberapa saran mengenai pengembangan lebih lanjut yang dapat diberikan penulis antara lain:

1. Penanganan metode stemming yang digunakan mampu mengenali sisipan dan dapat membadakan apakah kalimat tersebut merupakan kata dasar atau bukan.
2. Dokumen yang akan diringkas tidak sebatas hanya dokumen yang memiliki judul dan isi saja, namun dapat diperluas pada dokumen yang mempunyai bab maupun subbab.

Daftar Pustaka

- [1] Al-Hashemi, Rafeq. 2010. Text Summarization Extraction System (TSES) using Extracted Keyword. International Arab Journl of e-Technology, Vol. 1, No. 4, June 2010.
- [2] Azrifah, Masrah dan Martin, Trevor. 2009. Similarity-Based Estimation for Document Summarization using Fuzzy Sets. International Journal of Computer Science and Security, Volume (1) : Issue (4)
- [4] Fatah, Mohamed Abdel. Dan Ren, Fuji. 2008. Automatic Text Summarization. World Academy of Science, Engineering and Technology 37 2008.
- [5] Gupta, Vishal dan Singh Lehal, Gurpreet. 2010. A Survey of Text Summarization Extractive Techniques Journal of Emerging Technologies in Web Intelegence, Vol. 2, No. 3, August 2010
- [7] Hovy, E. 2003. Text Summarization. Dalam R. Mitkov, The Oxford Handbook of Computational Linguistic. Oxford: Oxford University Press
- [8] Isfahani, Fariba Rahimi. Kyoomarsi, Farshad. Khosravi, Hamid. Eslami, Esfandiar. Tajodin, Asgar dan Dehkordy, Pooya Khosravayan. 2008. Application of Fuzzy logic in the Improvement of Text Summarization. IADIS International Conference Informatics.

- [9] Jagadeesh, j. Pingali, Prasad. Varma, Vasudeva. 2005. Sentence Extraction Based Single Document Summarization. Workshop on Document Summarization, 19th and 20th March, 2005, IIT Allahabad.
- [10]Jazek, Karel dan Stainberger, Josef. 2008. Automatic Text Summarization (The state of the art 2007 and new challenges). Václav Sná'el (Ed.): Znalosti 2008, pp. 1{12, ISBN 978-80-227-2827-0.
- [11]Kusumadewi, Sri. 2003. Artificial Intelegence (Teknik dan Aplikasinya). Graha Ilmu. Yogyakarta
- [12]Kusumadewi, Sri dan Purnomo, Hari . 2004. Aplikasi Logika Fuzzy Untuk Pendukung Keputusan. Graha Ilmu. Yogyakarta
- [13]Kyoomarsi, Farshad. Khosravi, Hamid. Eslami, Esfandiar dan Dehkordy, Pooya Khosravyan. 2009. Optimizing Machine Learning Approach Based on Fuzzy Logic in Text Summarization International Journal of Hybrid Information Technology Vol.2, No.2, April, 2009
- [15]Suanmali, Ladda. Salim, Naomie dan Binwahlan, Mohammed Salem Fuzzy Logic Based Method for Improving Text Summarization. 2009. (IJCSIS) International Journal of Computer Science and Information Security, Vol. 2, No. 1, 2009
- [16] Saelan, Athia. 2009. Logika Fuzzy. Program Studi Teknik Informatika, Sekolah Teknik Elektro dan Informatika. Institut teknologi Bandung.
- [17]Tala, Fadillah Z. 2003. A Study of Stemming Effects on Information Retrieval in Bahasa Indonesia. Master of Logic Project. Institute for Logic, Language and Computation. Universiteit van Amsterdam. The Netherlands.

Biodata Penulis

Achmad Ridok, memperoleh gelar Sarjana Matematika (Drs) dari program studi matematika FMIPA Unibraw tahun 1992, Tahun 1999 mendapatkan gelar MKom dari program studi Ilmu Komputer Fasilkom UI. Saat sebagai staf Pengajar program Ilmu Komputer dan Sistem Informasi pada PTIIK Universitas Brawija.

Tri Cahyo Romadhona, memperoleh gelar Sarjana Komputer (SKom) pada program studi Ilmu Komputer Jurusan Matematika F.MIPA UB pada tahun 2011.

