PERINGKASAN TEXT BAHASA INDONESIA MENGGUNAKAN AHP

Ramdan Chandra M[1], Nelly Indriani W., S.Si.,M.T.[2]

[1,2]Teknik Informatika – Universitas Komputer Indonesia
Jalan Dipatiukur 112-116 Bandung
E-mail : [1]ramdanandchandra@gmail.com, [2]nelly.indriani@email.unikom.ac.id

ABSTRAK

Peringkasan teks merupakan suatu proses penyaringan sebagian besar informasi penting untuk menghasilkan ringkasan bagi pemakainya. Pada peringkasan teks bahasa Indonesia menggunakan 6 fitur diantaranya frekuensi kalimat, kata kunci, kemiripan kata terhadap judul, kemiripan antara kalimat, posisi kalimat dan panjang kalimat.
Metode yang digunakan pada peringkasan teks bahasa Indonesia adalah metode AHP. Metode AHP dinilai cocok untuk peringkasan teks, karena metode ini akan membandingkan setiap fitur dan kriteria, dimana kriteria didalam peringkasan teks ini adalah kalimat yang dihitung berdasarkan ektraksi fitur, sehingga didapatkan bobot akhir. Pada metode AHP nilai fitur didapatkan dari skala perbandingan yang didapatkan dari pakar.
Pengujian sistem pada peringkasan teks bahasa indonesia menggunakan AHP dengan metode Rouge-1, metode pengujian fitur statistik dan pengujian konsistensi pakar dalam memberikan nilai bobot. Pada metode *rouge*-1 didapatkan akurasi *recall* 75,83, *precision* 63,61 dan F-*measure* 68,59. Untuk pengujian fitur statistik didapatkan bahwa fitur f1 dan fitur f3 memiliki akurasi yang paling kecil dan pengujian konsistensi pakar didapatkan nilai CR < 0,1, itu menandakan skala perbandingan yang didapatkan dari pakar sudah konsisten.

Kata kunci**:** Peringkasan Teks, Ringkasan, Metode AHP, Kalimat

1. Pendahuluan

Peringkasan teks merupakan proses penyaringan sebagian besar informasi penting untuk menghasilkan ringkasan bagi pemakainya. Peringkasan teks otomatis merupakan penyajian suatu dokumen menjadi lebih pendek, namun tetap memiliki informasi yang penting didalamnya[1]. Peringkasan teks umumnya terbagi atas 2 macam yaitu peringkasan ekstraksi dan abstraksi. Peringkasan ekstraksi pada dasarnya hanya mengambil kalimat – kalimat yang memiliki informasi penting saja. Sedangkan peringkasan abstraksi merupakan peringkasan yang dapat menghasilkan kalimat – kalimat baru yang menggambarkan inti dari dokumen yang diringkas [2].

Peringkasan teks bahasa Indonesia, sebenarnya sudah banyak diteliti dengan berbagai macam metode diantaranya metode *Support Vector Machine* (SVM) dengan metode *Maximal Marginal Importance* (MMI) [3], metode *Relevance Vector Machine* (RVM) [4]. Metode – metode tersebut merupakan metode yang digunakan dalam mesin *learning*, dimana metode tersebut sangat dipengaruhi oleh data *training* dan *testing*. Data *training* digunakan untuk melatih algoritma menemukan model yang cocok dan data *testing* digunakan untuk mengetahui performa dari model yang dipakai. Berdasarkan hal itu, penelitian ini akan mencoba menggunakan metode pengambilan keputusan *Analytical Hierarchy Process* (AHP) untuk digunakan sebagai metode peringkasan teks.

Metode *Analytical Hierarchy Process* (AHP) merupakan metode pengambil keputusan yang dapat membantu kerangka berfikir manusia. Menurut Bourgeois, metode AHP biasanya digunakan untuk menyusun prioritas dari berbagai alternatif / pilihan , dimana pilihan tersebut bersifat kompleks atau muti kriteria [5].

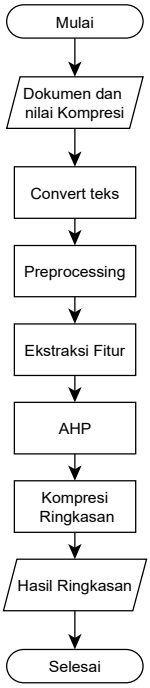
Penelitian peringkasan menggunakan AHP ini pernah dilakukan sebelumnya menggunakan untuk meringkas bahasa persian [6]. Pada bahasa persian AHP memiliki tingkat akurasi yang baik. Dengan akurasi yang baik tersebut AHP akan digunakan untuk peringkasan teks Bahasa Indonesia yang memiliki struktur tata bahasa yang berbeda dengan persian.

Berdasarkan uraian diatas, pada penelitian ini akan menggunakan metode pada sistem pengambilan keputusan yaitu AHP, dimana keputusan pemilihan prioritas fitur akan menggunakan salah satu pakar ahli untuk menjadi acuan dalam menentukan fitur mana yang lebih penting.

2. METODE

Pada penelitian ini akan menggunakan metode *Analytical Hierarchy Process* (AHP) untuk peringkasan teks bahasa Indonesia, metode ini merupakan metode sistem pengambil keputusan. Pada sistem peringkasan teks Bahasa Indonesia, masukan awalnya berupa file berbentuk *Portable Document Format* (PDF). Saat file dokumen dimasukan kedalam sistem, file akan diconvert dahulu dari binary pada pdf menjadi teks, untuk nantinya bisa diolah pada tahap *preprocessing*, setelah masuk ke tahap *preprocessing* dilanjut dengan proses ekstraksi fitur. Setelah fitur diekstraksi, nilai dari setiap fitur akan diolah dengan metode AHP untuk mendapatkan hasil ringkasan.

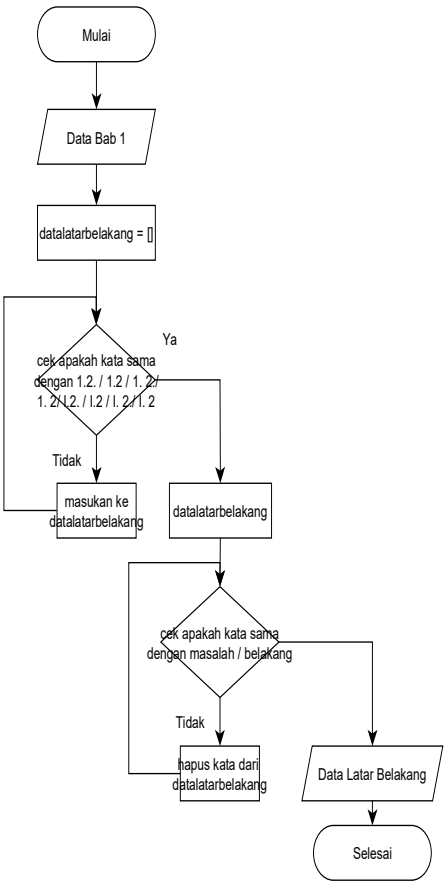
Pada penelitian ini, peringkasan teks memimiliki 5 tahapan utama pada arsitektur sistem diantaranya *convert* teks, *preprocessing*, Ekstraksi Fitur, AHP, Kompresi Ringkasan. Pada gambar 1 merupakan gambaran secara lengkap arsitektur yang akan dibangun.



Gambar 1 Gambaran Sistem

2.1. *Convert* Teks

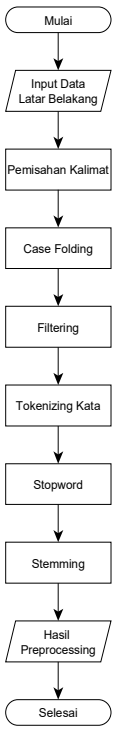
Pada tahap ini, akan dilakukannya proses *Convert* teks berupa *Encoding* pada PDF yang memiliki nilai binary menjadi teks. Proses *Convert* tersebut menggunakan *Library* Pdfparser. Setelah data menjadi teks tahapan selanjutnya adalah pengambilan data yang akan diringkas. Pada gambar 2 merupakan proses pengambilan data latar belakang sebagai data yang akan diringkas.



Gambar 2 Proses Pengambilan data Latar Belakang

2.2. *Preprocessing*

Pada tahap ini, data masukan setelah di *convert* akan diolah pada tahap *preprocessing*. *Preprocessing* sendiri memiliki fungsi untuk membuang kata – kata atau simbol yang tidak memiliki kepentingan dalam sistem peringkasan teks . proses ini terdiri atas pemisahan kalimat, *case folding, filtering, tokenizing* kata, *removal stopword*, dan *stemming* [7]. pada gambar 3 merupakan tahapan dari proses *preprocessing*.



Gambar 3 Proses *Preprocessing*

2.2.1 Pemisahan Kalimat

Langkah awal dalam *preprocessing* adalah pemisahan kalimat. Proses ini digunakan untuk memisahkan kalimat didalam dokumen yang akan diringkas, sehingga didapatkan sekumpulan kalimat. Pada tahap ini pemisahan kalimat dilakukan jika pada dokumen ditemukan tanda titik (.), tanda tanya (?), dan tanda seru (!) [7]. Untuk kalimat yang mengandung kata seperti (PT.), (CV.), dan lain – lain akan diubah dulu menjadi (PT,), (CV,) untuk terjadinya pemisahan kalimat.

2.2.2 *Case Folding*

Setelah melalui tahap pemisahan kalimat dilanjutkan dengan proses *Case Folding*. Pada tahap ini setiap huruf kapital dalam kalimat akan diubah menjadi huruf kecil [7]. Contohnya kata "Kegiatan" diubah menjadi "kegiatan".

2.2.3 *Filtering*

Proses *Filtering* berguna untuk membuang tanda baca pada teks sehingga teks hanya terdapat huruf antara 'a-z', angka (0-9) dan spasi (" ").

2.2.4 *Tokenizing* kata

Proses ini digunakan untuk memisahkan kata pada teks. Proses ini akan memisahkan kata pada kalimat berdasarkan spasi (" ") sebagai pemisah [7].

2.2.5 *Stopword removal*

Stopword merupakan proses untuk menghilangkan kata yang tidak memiliki arti. Untuk mengetahui kata apa saja yang masuk ke dalam *stopword* digunakan kamus *stopword* [8]. Contoh kata yang termasuk kedalam stopword adalah ke, itu, ini dan lain – lain.

2.2.6 *Stemming*

Stemming merupakan proses untuk membuang imbuhan pada kata agar menjadi kata dasar [9]. Metode yang digunakan pada proses ini adalah metode *Porter Stemmer* Bahasa Indonesia [9].

2.3. Ekstraksi Fitur

Setelah tahap *preprocessing* masuk kedalam tahap Ekstraksi Fitur. Pada tahap ini, kalimat pada dokumen akan diekstrak berdasarkan fitur – fitur. Fitur yang dipakai diantaranya Frekuensi Kalimat, Kata Kunci, Kemiripan kata terhadap judul, kemiripan antar kalimat, lokasi kalimat dan panjang kalimat.

2.3.1 Frekuensi Kalimat

Fitur Frekuensi kalimat bisa dilakukan setelah proses *preprocessing*, dimana secara umum metode ini menggunakan metode tf-idf dalam menghitung frekuensi kata [9]. Menggunakan rumus (1).

$$wi = tfi * log \frac{\mathrm{N}}{\mathrm{n}} \qquad \textbf{(1)}$$

Dimana :

tfi : Banyaknya kata yang dicari dalam sebuah dokumen
N : Total Dokumen / Kalimat
n : Banyak dokumen yang mengandung kata yang dicari

Setelah diketahui nilai (wi) selanjutnya masuk ke proses perhitungan fitur 1 menggunakan rumus (2) [6].

$$F1 = \frac{\sum_{i=1}^{k} Wi}{Max\ (Wi)} \qquad \textbf{(2)}$$

Dimana :

k : jumlah kata dalam kalimat

2.3.2 Kata Kunci

Pada tahap ini kata kunci didapatkan dari kata yang muncul berulang kali, dimana kata yang masuk kedalam kata kunci akan diberi

nilai 0,1 dan kata yang tidak ditemukan pada kata kunci akan diberi nilai 0 [7]. Pada fitur kata kunci terdapat kondisi, kalimat dalam kata kunci tidak ditemukan dan menghasilkan nilai 0. Jika terdapat kalimat yang memiliki nilai 0 maka proses tidak bisa dilanjutkan pada proses perhitungan pada metode AHP. Pada fitur kata kunci, setiap kalimat akan dijumlahkan dengan nilai 0,1.

2.3.3 Kemiripan Kata terhadap Judul

Fitur ini akan menghitung kata mana saja pada kalimat yang mirip dengan kata yang ada pada judul. Untuk menghitung kemiripan kata tersebut digunakan rumus (3).

$$\text{F3} = \frac{Kata\ dalam(i) \cap Kata\ dalam\ judul}{Kata\ dalam(i) \cup Kata\ dalam\ judul} + 0{,}1 \quad \textbf{(3)}$$

Pada fitur ketiga ini terdapat nilai 0,1. Nilai 0,1 pada fitur ini berguna untuk mencegah terjadinya nilai fitur 0 pada kalimat.

Sebagai contoh, sebuah token judul adalaj "giat ajar" dengan kalimat pertama adalah "giat, ajar, giat, individu". Untuk menghitung nilai fitur pada kalimat pertama dapat dilihat dibawah ini :

$$F3 = \frac{(giat, ajar, giat, individu \cap giat, ajar)}{(giat, ajar, giat, individu \cup giat, ajar)}$$

$$F3 = \frac{(giat, ajar)}{(giat, ajar, giat, individu)}$$

$$F3 = \frac{2}{3} = 0{,}76667$$

pada tabel 1 merupakan perhitungan fitur kemiripan judul pada kalimat yang lainnya.

Tabel 1 Perhitungan Kemiripan kata terhadap judul

Kalimat	F3
1	$F3 = \frac{2}{3} + 0{,}1 = 0{,}7667$
2	$F3 = \frac{2}{4} + 0{,}1 = 0{,}6$
3	$F3 = \frac{2}{7} = 0{,}385714$
4	$F3 = \frac{2}{7} + 0{,}1 = 0{,}385714$
5	$F3 = \frac{2}{5} + 0{,}1 = 0{,}5$
6	$F3 = \frac{2}{9} + 0{,}1 = 0{,}32222$

2.3.4 Kemiripan Antar Kalimat

Fitur ini akan menghitung kemiripan kata pada kalimat dengan kata pada kalimat lainnya. Fitur ini dapat dihitung dengan rumus (4).

$$F4 = \frac{Kata\ dalam(i) \cap Kata\ antar\ kalimat}{Kata\ dalam(i) \cup Kata\ antar\ kalimat} (4)$$

Sebagai contoh, diketahui kata pada kalimat pertama adalah "giat, ajar, individu" dan kata dikalimat yang lainnya adalah "arah, ajar, bicara didik, efektif, individu, jalan, giat, kembang, komponen, langsung, lihat, lintas, media, peserta, rencana, sekolah, strategi, tenaga". Untuk mendapatkan nilai fitur kemiripan antar kalimat pada kalimat pertama bisa dilihat dibawah ini :

$$F4 = \frac{(giat, ajar, individu \cap \begin{matrix} arah, ajar, bicara, didik, \\ efektif, individu, jalan, giat, \\ kembang, komponen, langsung, \\ lihat, lintas, media \\ peserta, rencana, sekolah, \\ strategi, tenaga \end{matrix})}{giat, ajar, individu \cup \begin{matrix} arah, ajar, bicara, didik, \\ efektif, individu, jalan, giat, \\ kembang, komponen, langsung, \\ lihat, lintas, media \\ peserta, rencana, sekolah, \\ strategi, tenaga \end{matrix})}$$

$$F4 = \frac{(giat, ajar, individu)}{\begin{pmatrix} arah, ajar, bicara, didik, \\ efektif, individu, jalan, giat, \\ kembang, komponen, langsung, \\ lihat, lintas, media \\ peserta, rencana, sekolah, \\ strategi, tenaga \end{pmatrix}}$$

$$F4 = \frac{3}{21} = 0{,}142857$$

Pada tabel 2 merupakan perhitungan fitur kemiripan antar kalimat pada kalimat lainnya.

Tabel 2 Perhitungan Kemiripan antar kalimat

Kalimat	F4
S1	$F4 = \frac{3}{21} = 0{,}142857$
S2	$F4 = \frac{4}{21} = 0{,}190476$
S3	$F4 = \frac{7}{21} = 0{,}333333$
S4	$F4 = \frac{7}{21} = 0{,}333333$
S5	$F4 = \frac{5}{21} = 0{,}2380952$
S6	$F4 = \frac{9}{21} = 0{,}4285714$

2.3.5 Posisi Kalimat

Pada fitur posisi kalimat didasarkan asumsi bahwa kalimat pertama dan kalimat terakhir pada merupakan bagian paling penting pada sebuah dokumen. Nilai fitur pada posisi kalimat bisa dihitung menggunakan rumus (5).

$$F5 = Max\left(\frac{1}{i}, \frac{1}{n - i + 1}\right) \quad \textbf{(5)}$$

Dimana i : posisi kalimat

n : jumlah kalimat

Proses perhitungan pada fitur posisi kalimat dapat dilihat pada tabel 3.

Tabel 3 Perhitungan pada fitur posisi kalimat

Kalimat	F5
1	$F5 = Max\left(\frac{1}{1}, \frac{1}{6-1+1}\right) = 1$
2	$F5 = Max\left(\frac{1}{2}, \frac{1}{6-2+1}\right) = 0{,}5$
3	$F5 = Max\left(\frac{1}{3}, \frac{1}{6-3+1}\right) = 0{,}33$
4	$F5 = Max\left(\frac{1}{4}, \frac{1}{6-4+1}\right) = 0{,}33$
5	$F5 = Max\left(\frac{1}{5}, \frac{1}{6-5+1}\right) = 0{,}5$
6	$F5 = Max\left(\frac{1}{6}, \frac{1}{6-6+1}\right) = 1$

2.3.6 Panjang Kalimat

Fitur ini didasari bahwa semakin panjang kalimat berarti semakin banyak kata penting pada kalimat tersebut. Fitur ini dapat dihitung dengan menggunakan rumus (6).

$$F6 = \frac{panjang\ kalimat\ (i)}{kalimat\ terpanjang\ dalam\ dokumen} \quad (6)$$

Dimana i : posisi kalimat

proses perhitungan panjang kalimat dapat dilihat pada tabel 4.

Tabel 4 perhitungan pada fitur panjang kalimat

Kalimat	F6
1	$F6 = \frac{4}{13} = 0{,}3077$
2	$F6 = \frac{4}{13} = 0{,}3077$
3	$F6 = \frac{9}{13} = 0{,}69231$
4	$F6 = \frac{7}{13} = 0{,}5385$
5	$F6 = \frac{5}{13} = 0{,}48462$
6	$F6 = \frac{13}{13} = 1$

2.4. Metode AHP

Langkah awal pada metode ini adalah melakukan pembobotan skala fitur, dimana hasil dari skala fitur didapatkan dari pakar ahli. Setelah didapatkan nilai perbandingan fitur, nilai fitur dimasukan kedalam matriks. Pada tabel 5 merupakan matriks perbandingan fitur berdasarkan skala perbandingan yang diberikan oleh pakar.

Tabel 5 Matriks skala perbandingan fitur

Fitur	F1	F2	F3	F4	F5	F6
F1	1	0,619	0,236	0,260	0,382	0,25
F2	1,613	1	0,381	0,420	0,617	0,403
F3	4,226	2,619	1	1,100	1,617	1,056
F4	3,839	2,380	0,908	1	1,469	0,959
F5	2,613	1,6199	0,6183	0,680	1	0,653
F6	4	2,479	0,946	1,041	1,530721	1

Setelah nilai skala perbandingan yang dimasukan kedalam matriks, langkah selanjutnya adalah melakukan normalisasi. Fungsi dari normaslisasi ini bertujuan agar nilai berada pada range antara 0 – 1.

Pada tabel 6 merupakan hasil dari normalisasi berdasarkan matriks pada tabel 5.

Tabel 6 Hasil Bobot Fitur

	F1	**F2**	**F3**	**F4**	**F5**	**F6**	**W**
F1	0,052	0,085	0,041	0,038	0,038	0,038	0,049
F2	0,263	0,426	0,291	0,465	0,465	0,465	0,396
F3	0,052	0,0609	0,0416	0,031	0,031	0,031	0,041
F4	0,21	0,142	0,208	0,155	0,155	0,155	0,171
F5	0,21	0,142	0,208	0,155	0,155	0,155	0,171
F6	0,21	0,142	0,208	0,155	0,155	0,155	0,171

Setelah nilai bobot fitur didapatkan, dilanjutkan dengan melakukan pencarian bobot kalimat pada setiap fitur. pada tabel 7 merupakan cara perhitungan untuk mencari bobot kalimat dari setiap fitur.

Tabel 7 Perhitungan Bobot Kalimat setiapo fitur

Fke(1..6)	S1	S2	...	Sn	W
S1	1	α			W1
S2	1/ α	1			W2
...			1		...
Sn				1	Wn

Pada tabel 8 merupakan perhitungan matriks pada bobot fitur pertama.

Tabel 8 Matriks Bobot Kalimat pada Fitur 1

F1	S1	S2	S3	S4	S5	S6
S1	1	0,619	0,236	0,260	0,382	0,25
S2	1,6131	1	0,3816	0,420	0,617	0,403
S3	4,2262	2,6199	1	1,100	1,617	1,056
S4	3,8394	2,3800	0,9084	1	1,469	0,959
S5	2,6131	1,6199	0,6183	0,680	1	0,653
S6	4	2,4796	0,9464	1,041	1,530	1
$\sum sj$	17,29	10,719	4,091	4,503	6,61	4,328

Nilai 1 pada kolom pertama dan baris pertama didapat dari hasil bagi antara nilai kalimat pertama dengan nilai kalimat kedua. Nilai 0,619 didapatkan dari nilai kalimat pertama dibagi dengan nilai kalimat kedua dan seterusnya sampai kolom ke enam dan baris ke enam. Setelah hasil matriks didapatkan, dilakukan normalisasi matriks agar nilai bobot ada antara range 0 – 1.

Pada tabel 9 merupakan hasil normalisasi dan bobot pada fitur pertama.

Tabel 9 Normalisasi matriks bobot fitur pertama

F1	S1	S2	S3	S4	S5	S6
S1	0,057	0,057	0,057	0,057	0,057	0,057
S2	0,09	0,09	0,09	0,09	0,09	0,09
S3	0,24	0,24	0,24	0,24	0,24	0,24
S4	0,22	0,22	0,22	0,22	0,22	0,22
S5	0,15	0,15	0,15	0,15	0,15	0,15
S6	0,23	0,23	0,23	0,23	0,23	0,23

Setelah didapatkan matriks bobot fitur selanjutnya dihitung nilai bobot. Pada tabel 10 merupakan hasil bobot pada fitur 1.

Tabel 10 Bobot pada Fitur 1

Kalimat	W(F1)
S1	0,057
S2	0,09
S3	0,24
S4	0,22
S5	0,15
S6	0,23

Untuk perhitungan pada fitur yang lainnya mirip seperti perhitungan yang dilakuka pada fitur 1. Lalu setelah didapatkan nilai bobot kalimat setiap fitur, proses selanjutnya adalah mencari bobot akhir, untuk mendapatkan nilai bobot akhir digunakan rumus (7).

$$W_{Si} = \sum_{j=1}^{6} W_{Fj} * W_{ij} \quad (7)$$

Pada tabel 11 merupakan hasil akhir bobot setiap kalimat pada dokumen yang dicontohkan.

Tabel 11 Hasil Bobot akhir

Kalimat	Wsi
S1	0,149693
S2	0,118922
S3	0,185164
S4	0,164302
S5	0,12924
S6	0,252679

2.5. Kompresi

pada tahap ini akan dilakukan pemilihan kalimat untuk diambil sebagai kalimat ringkasan. Nilai kompresi yang dipakai antar 30-40 % [6]. Kalimat yang dipilih berdasarkan nilai bobot paling tinggi. Pada dokumen yang dicontohkan akan menggunakan kompresi 30%. Sehingga, kalimat yang akan dipilih untuk ringkasan adalah 2 kalimat dan kalimat yang dipilih berdasarkan tabel 11 adalah kalimat 3 dan kalimat 6. Dan kalimat ringkasan yang terpilih dapat dilihat pada tabel 12.

Tabel 12 Hasil Ringkasan yang terpilih

Kalimat	Hasil Ringkasan
S3	Ketika kita membicarakan tentang kegiatan pembelajaran ini maka kita mungkin akan langsung mengarahkan pandangan kita pada kegiatan pembelajaran yang berlangsung di sekolah.
S6	Beberapa diantaranya adalah mengenai tenaga pendidik, para peserta didik, media pembelajaran, strategi pembelajaran, dan perencanaan kegiatan pembelajaran

3. PEMBAHASAN

Pengujian berguna untuk mengetahui hasil dari akurasi ketika menggunakan metode AHP pada peringkasan teks bahasa Indoensia. Pada pengujian ini menggunakan 50 data pengujian pada latar belakang tugas akhir. Dibawah ini ada 3 pengujian yang dipakai yaitu pengujian Akurasi menggunakan *Rouge* dengan n-*gram* = 1, pengujian fitur statistik dan pengujian konsistensi dari pakar.

3.1 Pengujian akurasi *Rouge*

Pada pengujian ini menggunakan n-*gram*. Pada metode ini akan menghitung nilai *recall*, *precision*, dan F-*measure*. Pada tabel 13 merupakan hasil pengujian menggunakan *Rouge*.

Tabel 13 Hasil Pengujian menggunakan *Rouge*

Dokumen	Recall	Precision	F-measure
1	55,435	59,302	57,303
2	73,529	54,945	62,893
3	68	62,044	64,885
4	89,61	76,67	82,635
5	76,74	66	70,968
6	96,67	50,87	66,67
7	84,615	49,749	62,658
8	97,826	72	82,949
9	71,579	81,928	76,404
10	72,864	66,514	69,544
11	78,378	67,442	72,5
12	86,301	92,647	89,362
13	86,184	66,497	75,072
14	74,576	57,895	65,185
15	73,563	83,117	78,049
16	100	100	100
17	94,167	84,962	89,328
18	53,425	34,513	41,935
19	74,684	63,441	68,605
20	84,524	66,981	74,737
21	56,835	33,193	41,91
22	73,016	63,014	67,647
23	71,765	52,586	60,697
24	75	50,336	60,0241
25	62,617	51,538	56,54
26	65	54,545	59,316
27	77,647	48,175	59,459
28	88,235	76,923	82,192
29	67,826	65	66,383
30	69,065	55,491	61,538
31	100	100	100
32	100	100	100
33	74,528	54,483	62,948
34	70,543	56,875	62,976
35	77,174	52,985	62,832
36	83,516	51,351	63,598
37	71,429	50	58,824
38	54,688	34,314	42,169
39	60,274	44,898	51,462
40	70,37	64,045	67,059
41	58,667	52,381	55,346
42	58,621	50	53,968
43	64,912	39,785	49,333
44	76,923	82,353	79,545
45	68	72,34	70,103
46	67,568	51,02	58,14
47	79,675	63,226	70,504
48	100	100	100
49	60,976	65,789	63,291
50	93,836	86,164	89,836
Rata - rata	75,82752	63,60654	68,58644

3.2 Pengujian Fitur Statistik

Pada fitur ini berguna untuk mengetahui bagaimana fitur dapat mempengaruhi nilai akurasi yang dihasilkan oleh sistem. Pengujian fitur statistik akan mengambil 5 sample dari 50 dokumen dengan nilai F-*measure* terkecil, ini berfungsi untuk mengetahui fitur mana saja yang membuat dokumen memiliki akurasi kecil. Pada tabel 14 merupakan hasil pengujian fitur statistik.

Tabel 14 Hasil Pengujian Fitur Statistik

Fitur yang diuji	Recall	Precision	F-measure
1	52,803	32,5514	40,244
1,2	64,552	39,54	48,98234
1,2,3	62,052	38,2746	47,27714
1,2,3,4	67,1028	41,7874	51,46714
1,2,3,4,5	63,5298	38,589	47,9908
1,2,3,4,5,6	61,3048	38,8026	47,4224

Berdasarkan hasil pengujian pada tabel 4.8. Ketika menggunakan satu fitur yaitu fitur frekuensi kalimat (f1) nilai *recall* 52.803 %, *precision* 32.5114 % dan F-*measure* 40.244 %. Selanjutnya ditambah fitur kata kunci (f2) hasilnya meningkatkan nilai *recall*, *precision* dan F-*measure* masing – masing sebesar 64.552, 39.54, dan 48.98324. lalu ditambah dengan nilai fitur kemiripan kata terhadap judul (f3), menghasilkan penurunan nilai *recall, precision* dan F-*measure* yaitu 62.052, 38.2746, 47.27714. Setelah itu ditambah fitur kemiripan antar kalimat (f4), setelah ditambahkan terdapat peningkatan nilai yang cukup signifikan pada *recall*, *precision* dan F-*measure* masing – masing 67.1028, 41.7874, 51.46714. lalu ditambah dengan fitur ke lima yaitu posisi kalimat, pada pengujian fitur ini terdapat penurunan nilai yaitu *recall* 63.5298, *precision* 38.589, dan F-*measure* sebesar 47.9908. Pada tahap akhir ditambahkan fitur ke 6 yaitu panjang kalimat, pada fitur ini nilai *recall* dan F-*measure* mengalami penurunan dan nilai *precision* mengalami sedikit peningkatan dengan nilai *recall* 61.308, *precision* 38.8036 dan F-*measure* 47.4224. Pada pengujian ini

dapat disimpulkan bahwa fitur yang memiliki dampak nilai yang paling besar ada pada penambahan fitur kata kunci (f2) dan fitur kemiripan antar kalimat (f4) dan fitur yang kurang relevan adalah fitur frekuensi kalimat (f1) serta kemiripan kata terhadap judul (f3) karena memiliki nilai akurasi yang lebih kecil dibanding yang lainnya.

3.3 Pengujian Konsistensi Pakar

Pada tahap ini akan dilakukan pengujian konsistensi pakar dalam memberikan nilai fitur, apakah nilai yang diberikan oleh pakar bisa dipakai atau tidak. Konsistensi pakar dikatakan konsisten, jika nilai CR < 0,1. Untuk mendapatkan nilai CR berdasarkan penjelasan yang sudah dilakukan pada bab sebelumnya [5].

Tahap awal adalah menghitung nilai $\lambda_{maks.}$ Untuk mendapatkan nilai λ_{maks} diawali dengan kalikan setiap nilai pada kolom pertama dengan nilai bobot fitur pertama lalu jumlahkan .Nilai pada kolom kedua dengan nilai bobot fitur kedua lalu jumlahkan dan seterusnya. Sehingga didapatkan nilai pada tabel 15.

Tabel 15 Perhitungan nilai bobot

Fitur	Nilai bobot
F1	0,298222
F2	2,471845
F3	0,249927
F4	1,04939
F5	1,04939
F6	1,04939

Setelah nilai bobot ditemukan, langkah selanjutnya membagi nilai bobot dengan nilai bobot fitur pada tabel 19. sehingga didapatkan hasil pada tabel 16.

Tabel 16 Nilai

Fitur	Nilai
F1	6,046188
F2	6,239394
F3	6,039387
F4	6,135282
F5	6,135282
F6	6,135282

Lalu dari tabel 16 jumlahkan masing masing kolom dan dibagi dengan jumlah fitur yang ada sehingga didapatkan nilai λ_{maks} sebesar 6,121802. Untuk menghitung nilai Ci dan CR dapat dilihat dibawah ini :

Dik λ_{maks} = 6,121802
n = 6
Ri = 1,24

Dit Ci, CR ?

Jawab

$$Ci = \frac{\lambda_{\text{maks}} - \text{n}}{n - 1}$$

$$Ci = \frac{6{,}121802 - 6}{6 - 1}$$

$$Ci = 0{,}02436$$

$$CR = \frac{\text{Ci}}{Ri}$$

$$CR = \frac{0{,}02436}{1{,}24}$$

$$CR = 0{,}0196$$

Berdasarkan perhitungan diatas, diketahui nilai CR < 0,1, itu berarti nilai fitur yang diberikan pakar sudah konsisten.

4. Kesimpulan dan Saran

4.1 Kesimpulan

Berdasarkan hasil dari penelitian yang sudah dilakukan sebelumnya seperti penentuan masalah, analisis masalah dan implementasi serta pengujian, maka dapat disimpulkan bahwa hasil akurasi dari metode AHP untuk peringkasan teks Bahasa Indonesia, mempunyai hasil akurasi ,menggunakan Rouge1 menghasilkan rata – rata sebesar *recall* 75.83 %, *precision* 63.61 %, dan F-*measure* sebesar 68.59 %. .

4.2 Saran

Berdasarkan hasil analisis dan kesimpulan, terdapat saran untuk penelitian selanjutnya yaitu, Ekstraksi fitur yang kurang penting saat dibandingkan oleh pakar diganti dengan ekstraksi fitur yang lain. Salah satu Ekstraksi fitur yang kurang penting pada penelitian ini adalah kemiripan kata terhadap judul. Sehingga, ada baiknya diganti dengan ekstraksi fitur lainnya.

DAFTAR PUSTAKA

[1] Suyanto, Ririn Dwi Agustin Nengsih, "PERINGKASAN TEKS OTOMATIS UNTUK DOKUMEN TUNGGAL BERBAHASA INDONESIA MENGGUNAKAN ALGORITMA TERM FREQUENCY DAN LEAD," Sekolah Tinggi Teknologi Telkom, Bandung, 2006.

[2] B. Zamandan E. Winarko, "Analisis Fitur Kalimat untuk Peringkas Teks Otomatis pada," *Peringkasan Text*, vol. 5, Juli 2011.

[3] Novriana, "Implementasi Support Vector Machine (SVM) Dan Maximal Marginal Importance (MMI) Untuk Peringkasan Multi Dokumen Artikel Berbahasa Indonesia," Unikom, Bandung, 2017.

[4] Arief Budiman Eka Putra, "Implementasi Metode Relevance Vector Machine Dalam Peringkasan Teks Otomatis," Unikom, Bandung, 2017.

[5] R Bourgeois, "*Analytical Hierarchy Process*," an Overview UNCAPSA - UNESCAP., Bogor, 2005.

[6] Ram Gopal Raj, Hamid Haj Seyyed Javadi Seyyed Mohsen Tofighy, "AHP Techniques for Persian Text Summarizatio," *Malaysian Journal of Computer Science*, vol. 26, pp. 1-8, 2013.

[7] R. Feldman and J. Sanger, "*The Text Mining Handbook*," Cambridge, 2007.

[8] Pandu Adikara, Putra, *Kamus Kata Dasar dan Stoplist Bahasa Indonesia,* http://hikaruyuuki.lecture.ub.ac.id/kamus-kata-dasar-dan-stopword-list-bahasa-indonesia/, 21 Agustus 2017 21.00

[9] Fadillah Z Tala, "A Study of Stemming Efects on Information Retrieval in Bahasa Indonesia," Institute for Logic, Language and Computation Universite itvan Amsterdam The Netherlands, Amsterdam, 2003.

[10] Sari Irene Nanda, "Analisis dan Implementasi Peringkasan Teks Otomatis dengan Metode Maximal Marginal Importance (MMI)," Univeristas Telkom, Skripsi 2011.