

FUZZY LOGIC DAN LEXICAL CHAINS UNTUK PERINGKASAN TEKS OTOMATIS

Nelly Indriani Widiastuti¹⁾, Wulan Kamilia Afnan²⁾

Abstract— *Automatic text summarization (ATS) is a process that was required to produce a summary of a text that contains information with help from a computer. This system is needed to determine the subject matter of a document, so that readers can quickly understand. ATS systems need to process the documents resulting in important sentences from the document. In this study lexical chains is used to generate optimal value for the candidate most powerful word in each sentence. The document is extracted to produce features such as sentence length, the weight of the sentence, the position of the sentence, and the similarity between sentences. The value of the strongest chain will be combined into fuzzy parameters. These features Fuzzy logic predicted the results of a summary of the values of the parameters to be grouped based on the value of linguistic important and unimportant. Furthermore, the final value of the fuzzy will determine the final outcome text summary of the document that are input by the user lexical chains. Testing conducted by manual sentence summary results sourced from respondents and a summary of the results of the ATS system recall, precision and F-measure. Based on the results of research that includes the step of determining the problem, the analysis to implementation and testing that has been done before, it can be concluded that the results of the implementation of the method of lexical chains with fuzzy logic for automatic text summarization achieve fairly good.*

Index Terms— *Automatic text summarization, fuzzy logic, lexical chains, summarization.*

I. INTRODUCTION

Peringkasan teks otomatis atau Automated Text Summarization disingkat ATS adalah sebuah proses yang diperlukan untuk menghasilkan ringkasan dari teks yang berisi informasi dengan menggunakan bantuan komputer. Dengan munculnya Internet, banyaknya portal berita menyebabkan jumlah dokumen elektronik dan data tekstual menjadi besar. Manusia, tidak dapat secara manual menangani volume teks yang besar. Hal ini menyebabkan penelitian dalam peringkasan teks otomatis banyak dikembangkan. ATS Akan memungkinkan pengguna untuk memutuskan apakah suatu dokumen menarik atau tidak, tanpa membaca

seluruh dokumen, yaitu dengan mengekstraksi informasi singkat tanpa kehilangan makna dan informasi penting dalam teks asli. [1]

Peringkasan teks otomatis yang menggunakan rantai leksikal diperkenalkan pada Barzilay dan Elhadad pada tahun 1997. Metode ini menggunakan database pengetahuan WordNet untuk menentukan hubungan kohesi antara istilah kemudian menyusun rantai berdasarkan istilah-istilah tersebut. Nilai atau skor diberikan berdasarkan jumlah dan jenis hubungan dalam rantai leksikal. Hasil ringkasan adalah konsentrasi rantai kalimat yang terkuat. Sebuah metode yang sama dengan menggunakan basis pengetahuan WordNet dan Wikipedia disajikan oleh Pourvali dan Abadeh Mohammad [2]. Metode ini menemukan makna yang tepat dari setiap kata dalam teks menggunakan WordNet, kemudian membangun rantai leksikal dan menghilangkan kalimat yang memiliki skor yang lemah dibandingkan dengan yang lain. Struktur kohesi leksikal teks dapat dimanfaatkan untuk menentukan pentingnya kalimat [1]. Penelitian lain yang menggunakan lexical chain dilakukan penggabungan dengan genetic algorithm [3]. Skor kalimat yang tinggi dari fitur algoritma langsung terpilih sebagai ringkasan. Hasilnya kurang optimal dibandingkan dengan penelitian lain yang akurasi lebih baik. Penelitian ini menggunakan rantai leksikal sebagai fitur baru. Ini mempengaruhi kinerja sistem ATS meningkat sebesar 46%. Hasil penelitian menunjukkan dengan menggunakan kombinasi fitur yang berbeda yang meningkatkan kinerja meringkas.

Penelitian ATS juga dilakukan dengan membandingkan penerapan logika fuzzy dan fuzzy c-means (FCM) untuk mengoptimalkan ringkasan [4]. Hasil ringkasan menunjukkan logika fuzzy lebih baik dibandingkan dengan mengoptimalkan FCM. Dalam studi logika fuzzy lain [5] dan atau dikombinasikan dengan lexical chain [6]. Riset yang dilakukan oleh F. Kyoomarsi, H. Khosravi, E. dan M. Eslami Davoudi, peringkasan teks otomatis dilakukan dengan menggunakan rantai leksikal dan logika fuzzy [6]. Mereka menggunakan parameter kata kunci, lokasi, jenis, dan WordNet. Dalam penelitian lain yang dilakukan Farshad dan rekan [7], menggunakan pendekatan lain yaitu logika fuzzy. Metode ini memperhitungkan setiap fitur teks seperti frekuensi kata, kesamaan dengan kata kunci, kesamaan dengan judul, posisi kalimat, statistik dari co-occurrence rantai leksikal, ekspresi indikatif. Setelah mengekstrak fitur, nilai antara 0 sampai 1 digunakan pada output untuk menentukan tingkat kepentingan dari kalimat dalam

¹ Universitas Komputer Indonesia
email: indiwidi@gmail.com

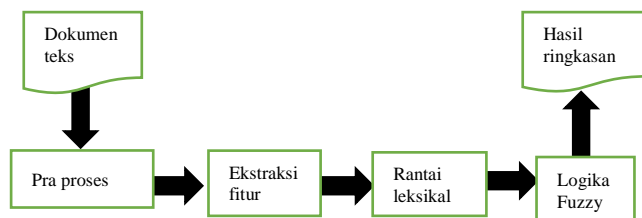
² Universitas Komputer Indonesia

ringkasan akhir. Hasil eksperimen menunjukkan penurunan precision yang tidak signifikan [1].

Berdasarkan penelitian sebelumnya, dalam penelitian ini akan dilakukan dengan menerapkan metode fuzzy logic dengan rantai leksikal. Penelitian ini akan menggunakan parameter yang berbeda dari penelitian sebelumnya. Parameter tersebut akan digunakan sebagai parameter fuzzy yang diharapkan akan memberikan nilai yang lebih optimal kepada calon kata terkuat.

I. METODE

Pada umumnya, sistem peringkasan mengalami beberapa proses sebelum menghasilkan ringkasan. Setiap prosesnya memiliki peran masing-masing dalam menjalankan proses. Proses-proses yang terlibat dalam peringkasan teks meliputi input dokumen teks, *preprocessing*, *extraction features*, implementasi *lexical chains*, dan implementasi metode logika fuzzy sehingga menghasilkan ringkasan teks.. Proses-proses tersebut selengkapnya dapat dilihat pada gambar 1.



Gambar 1. Diagram alir proses peringkasan teks otomatis

A. Data Masukan

Data masukan yang diperlukan untuk menjalankan sistem ATS berupa dokumen dengan format teks file (.txt). Sumber data diperoleh dari hasil unggah dokumen teks pada sistem.

B. Pra proses

Pra proses merupakan proses yang dilakukan untuk mengolah data mentah menjadi data yang berkualitas. Tahapan ini terdiri dari segmentasi kalimat, tokenization, removing stop word, dan stemming.

C. Ekstraksi fitur

Ekstraksi fitur merupakan proses pengolahan kalimat menggunakan beberapa fitur sehingga menghasilkan nilai-nilai yang nantinya akan digunakan untuk meringkas dokumen. Ekstraksi fitur akan dilakukan dari data hasil stemming. . Ekstraksi fitur bertujuan untuk memberikan nilai pada setiap kalimat untuk menentukan ringkasan teks. Ekstraksi fitur untuk peringkasan teks otomatis ini meliputi panjang kalimat, bobot kalimat, posisi kalimat, dan kemiripan antar kalimat.

D. Rantai Leksikal

Proses rantai leksikal merupakan tahapan dimana mencari rantai terkuat pada kata yang terdapat di dalam kalimat. Setiap kalimat memiliki kata yang mungkin saja memiliki kesamaan dengan kalimat lain. Hal ini dapat diketahui dengan mengelompokkan kata-kata

pada kalimat. Selain itu, setiap kalimat memiliki panjang rantai dan jumlah pasangan kata yang berbeda, sehingga mempengaruhi nilai dari masing-masing kalimat. Kata-kata yang memiliki kesamaan pada setiap kalimat, panjang rantai, dan jumlah pasangan.

E. Logika Fuzzy

Setelah rantai terkuat diperoleh dari proses rantai leksikal, proses selanjutnya dilakukan proses logika fuzzy. Proses ini merupakan tahapan akhir setelah seluruh proses yang terlibat dalam peringkasan sebelumnya telah dilakukan dan kalimat telah siap untuk memberikan hasil ringkasan melalui proses ini. Proses ini mengelompokkan nilai dari hasil ekstraksi yang telah dilakukan sebelumnya ke dalam beberapa variabel linguistik untuk diketahui nilai linguistiknya. Proses yang terjadi pada logika fuzzy terdiri dari *fuzzification*, *inference system*, *defuzzification*. Hasil dari ekstraksi fitur dikelompokkan dalam dua kategori yaitu kalimat penting dan kalimat tidak penting.

F. Hasil Ringkasan

Hasil yang diperoleh dari proses yang telah dilalui adalah dengan mengeluarkan ringkasan teks. Berdasarkan hasil dari *defuzzification*, kalimat kedua dan ketujuh merupakan ringkasan. Kemunculan ringkasan dihitung sebanyak 30% dari banyaknya kalimat dengan nilai *defuzzifikasi* terbesar.

II. HASIL DAN PEMBAHASAN

Sistem peringkasan otomatis teks terus dikembangkan untuk memenuhi kebutuhan pembaca dalam mendapatkan esensi dari membaca dokumen teks dengan cepat dan akurat. Berbagai metode yang diterapkan untuk mendapatkan ringkasan untuk akurasi yang lebih baik daripada penerapan metode yang ada.

E-learning has created new markets for teaching and learning material and equipment, attracting the attention of academic institutions as well as companies supplying them in different sectors – computer manufacturers, software producers, publishing houses and special training providers. It has also led to the reorientation of government policy in particular, towards encouraging the spread of e-learning techniques and developing the skills and know-how required for their use. Therefore, e-learning was assigned a key role in the pursuit of the policy objective, announced at the Lisbon Summit in March 2000, of making the 'the most competitive and dynamic knowledge-driven economy in the world'. The e-learning initiative was launched by the European Commission to encourage its spread. This was followed by its incorporation into the European employment strategy in the form of a specific objective of 'developing e-learning for all citizens'. To achieve this, States committed themselves to ensuring that 'all education and training institutions have access to the Internet and multimedia resources by the end of 2001 and that all the teachers and trainers concerned are skilled in the use of these technologies by the end of 2002 in order to provide all pupils with a broad digital literacy'. Despite its central importance in government policy and significant interest in the scale of the actual and potential market, there is an acute shortage of quantitative information on the extent of e learning in providing initial and continuing vocational education and training and on the rate at which it is growing

Gambar 2. Dokumen sampel

Dalam penelitian ini data yang digunakan adalah dokumen teks berformat .txt, Sebagai gambaran proses yang dilakukan dapat dilihat data sampel pada gambar

2. Data sampel ini akan diproses sesuai dengan tahap-tahap yang dilakukan dalam penelitian ini.

A. Pra Proses

Dalam tahap pertama pra proses dilakukan segmentasi kalimat, tokenisasi, *removing stop word* dan *stemming*.

1) Segmentasi kalimat

Proses ini memisahkan kalimat-kalimat dalam dokumen berdasarkan kemunculan tanda titik (.). Hasil segmentasi kalimat untuk dokumen sampel dapat dilihat pada tabel 1. Tahap segmentasi untuk dokumen pada gambar 1 menghasilkan 7 kalimat.

kalimat	Isi kalimat
S ₁	E-learning has created new markets for teaching and learning material and equipment, attracting the attention of academic institutions as well as companies supplying them in different sectors – computer manufacturers, software producers, publishing houses and special training providers.
S ₂	It has also led to the reorientation of government policy, in particular, towards encouraging the spread of e-learning techniques and developing the skills and know-how required for their use.
S ₃	Therefore, e-learning was assigned a key role in the pursuit of the policy objective, announced at the Lisbon Summit in March 2000, of making the 'the most competitive and dynamic knowledge-driven economy in the world'.
S ₄	The elearning initiative was launched by the European Commission to encourage its spread.
S ₅	This was followed by its incorporation into the European employment strategy in the form of a specific objective of 'developing e-learning for all citizens'.
S ₆	To achieve this, States committed themselves to ensuring that 'all education and training institutions have access to the Internet and multimedia resources by the end of 2001 and that all the teachers and trainers concerned are skilled in the use of these technologies by the end of 2002 in order to provide all pupils with a broad digital literacy'.
S ₇	Despite its central importance in government policy and significant interest in the scale of the actual and potential market, there is an acute shortage of quantitative information on the extent of e learning in providing initial and continuing vocational education and training and on the rate at which it is growing.

Tabel 1. Hasil segmentasi kalimat

2) Tokenisasi

Setelah proses segmentation kalimat, selanjutnya dilakukan proses tokenisasi. Proses ini merupakan proses memisahkan deretan kata dari suatu kalimat atau paragraf menjadi potongan kata tunggal berdasarkan karakter spasi yang ditemukan. Proses tokenisasi juga berperan untuk menghilangkan beberapa karakter seperti tanda baca dan angka.

Gambaran proses yang terjadi dalam tahap tokenisasi. Sistem akan mencari karakter spasi kemudian memisahkan kata dengan kata yang lain ke dalam tabel token. Dalam proses token karakter numerik dan tanda baca akan diabaikan.

3) Proses *Removing Stop Word*

Setelah proses tokenisasi, selanjutnya dilakukan proses *removing stop word*. Proses ini merupakan

proses menghilangkan kata-kata yang sering muncul, kata-kata yang memiliki makna tidak berkaitan langsung dengan pokok pikiran dalam dokumen atau kata-kata yang tidak memiliki makna jika dilihat sebagai kata tunggal termasuk kata ganti orang. Dalam penelitian ini daftar stop word berjumlah 677 kata.

4) Proses Stemming

Pada proses stemming. Proses ini merupakan proses menghilangkan awalan, sisipan dan akhiran dari setiap kata dan mengubahnya menjadi kumpulan kata dasar.

Proses stemming dalam penelitian ini menggunakan algoritma yang sudah ada yaitu Porter Stemmer.

B. Ekstraksi Fitur

1) Panjang Kalimat

Panjang kalimat merupakan hasil pembagian dari jumlah kata di dalam sebuah kalimat dengan jumlah kata terpanjang di dalam satu kalimat.

Untuk dokumen sampel yang digunakan hasil perhitungan panjang kalimat dapat dilihat pada tabel 2.

S_F1(S _i)	Hasil
S_F1(S ₁)	0.913
S_F1(S ₂)	0.4782
S_F1(S ₃)	0.7391
S_F1(S ₄)	0.3043
S_F1(S ₅)	0.4348
S_F1(S ₆)	0.913
S_F1(S ₇)	1

Tabel 2. Hasil perhitungan panjang kalimat

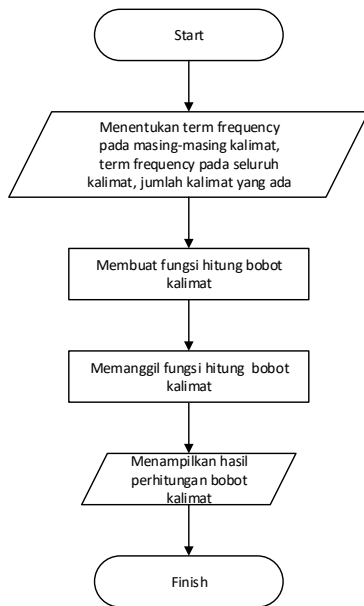
2) Bobot Kalimat

Bobot kalimat digunakan dengan menggunakan perhitungan *tf-isf* (*term frequency – inverse sentence frequency*). Alur proses bobot kalimat dapat dilihat pada gambar 3.

Ts-isf didapat dari frekuensi kejadian kata yang sering muncul pada kalimat beserta dengan jumlah dari frekuensi banyak kata yang sering muncul dalam satu dokumen teks. Frekuensi banyaknya kata dalam dokumen dinyatakan dengan *tf*, sedangkan jumlah kalimat yang terdapat di dalam sebuah dokumen dinyatakan dengan *N*, dan jumlah kata yang memiliki kesamaan dalam kalimat dinyatakan dengan *n*. Adapun perhitungan bobot (*weight*) diperoleh dari persamaan (2).

$$S_F2(S) = w_i = tf_i \times isf_i = tf_i \times \left(\log \frac{N}{n_i} \right) \quad (2)$$

dimana, *tf_i* = frekuensi banyak kata di dalam dokumen
N = jumlah kalimat
n_i = jumlah kalimat yang mengandung kata



Gambar 3. Proses hitung bobot kalimat

Perhitungan bobot (*weight*) ini merupakan jumlah keseluruhan dari *tf-isf* yang telah dihitung. Hasil perhitungan bobot (*weight*) dapat dilihat pada Tabel 3.

Bobot Kalimat	
S_F2(S _i)	Hasil
S_F2(S ₁)	0.9010
S_F2(S ₂)	0.3699
S_F2(S ₃)	0.7689
S_F2(S ₄)	0.2419
S_F2(S ₅)	0.3965
S_F2(S ₆)	0.9816
S_F2(S ₇)	1

Tabel 3. Hasil perhitungan bobot kalimat

3) Posisi Kalimat

Posisi kalimat merupakan letak kalimat dalam paragraf. Pada penelitian ini diasumsikan bahwa kalimat pertama tiap paragraf adalah kalimat penting. Perhitungan posisi kalimat diperoleh dari persamaan (3).

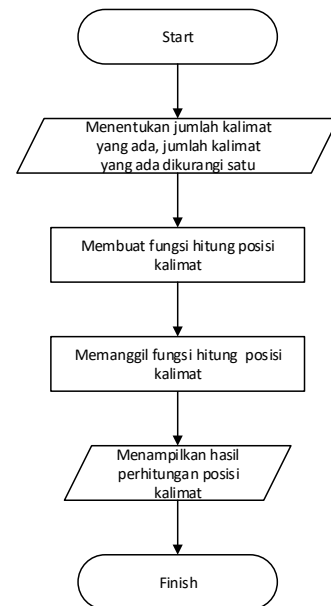
$$S_F3(S) = \frac{\text{Kalimat}}{\text{Paragraf}} \quad (3)$$

Dimana,

Kalimat = urutan kalimat yang menunjukkan kepentingan suatu kalimat

Paragraf = jumlah kalimat dalam dokumen

Alur proses perhitungan posisi kalimat dapat dilihat pada **Error! Reference source not found.4.**



Gambar 4. Proses Hitung Posisi Kalimat

Hasil perhitungan posisi kalimat untuk dokumen sampel dapat dilihat pada tabel 3.

Posisi Kalimat	
S_F3(S _i)	Hasil
S_F3(S ₁)	1
S_F3(S ₂)	0.8571
S_F3(S ₃)	0.7143
S_F3(S ₄)	0.5714
S_F3(S ₅)	0.4286
S_F3(S ₆)	0.2857
S_F3(S ₇)	0.1429

Tabel 3. Hasil Perhitungan Posisi kalimat

4) Kemiripan Antar Kalimat

Kemiripan antar kalimat adalah kata yang muncul dalam kalimat sama dengan kata yang muncul dalam kalimat lain. Alur proses perhitungan kemiripan kalimat dapat dilihat pada Gambar 5.

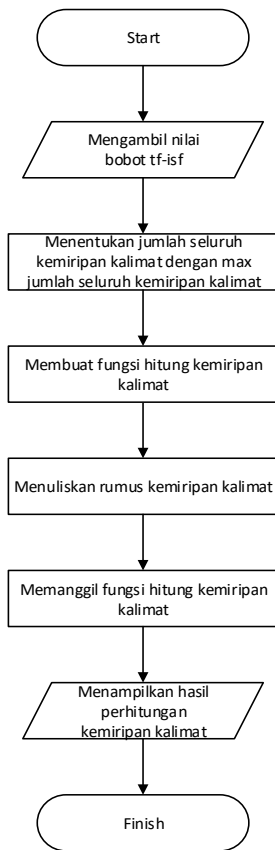
Kemiripan antar kalimat diperoleh dari jumlah jarak dari kedua kalimat dibagi dengan jumlah maksimum jarak dari kedua kalimat. Perhitungan jarak kalimat diperoleh dari persamaan (4).

$$S_F4(S) = \frac{\sum Sim(S_i, S_j)}{\text{Max}(\sum Sim(S_i, S_j))} \quad (4)$$

dimana,

$Sim(S_i, S_j)$ = kemiripan antara kalimat ke-i dan kalimat ke-j

$\sum Sim(S_i, S_j)$ = jumlah kemiripan antara kalimat ke-i dan kalimat ke-j



Gambar 5. Proses Perhitungan Kemiripan Antar Kalimat

Dengan S_i, S_j diperoleh dari rumus (5) jarak antar kalimat yaitu:

$$S_i, S_j = \frac{\sum_{t=1}^n W_{it} \times W_{jt}}{\sqrt{\sum_{t=1}^n W_i^2} \times \sqrt{\sum_{t=1}^n W_j^2}} \quad (5)$$

dimana,

$\sum_{t=1}^n W_{it} \times W_{jt}$ = jumlah perhitungan skalar dari pembobotan kalimat

$\sqrt{\sum_{t=1}^n W_i^2} \times \sqrt{\sum_{t=1}^n W_j^2}$ = jumlah perhitungan vektor dari pembobotan kalimat

Perhitungan jarak kalimat dapat dilihat pada Tabel 4.

S_F4(S_i)	Hasil
S_F4(S_1)	0.8016
S_F4(S_2)	1
S_F4(S_3)	0.8038
S_F4(S_4)	0.9242
S_F4(S_5)	0.8805
S_F4(S_6)	0.8305
S_F4(S_7)	0.8305

Tabel 4. Hasil Perhitungan Bobot Kalimat

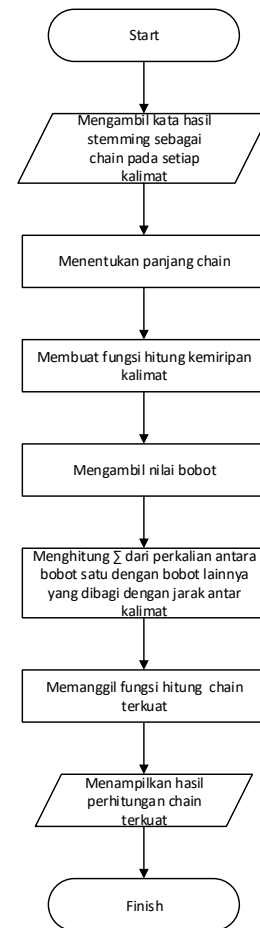
Kesimpulan dari keseluruhan pra proses adalah nilai-nilai fitur yang akan diproses lebih lanjut sebagai masukan dalam parameter logika fuzzy. Hasil kesimpulan tersebut dapat dilihat pada tabel 5.

Kalimat (S_i)	Nilai			
	Panjang Kalimat	Bobot Kalimat	Posisi Kalimat	Kemiripan Antar Kalimat
S_1	0.9130	0.9010	1	0.8016
S_2	0.4782	0.3699	0.8571	1
S_3	0.7391	0.7689	0.7143	0.8038
S_4	0.3043	0.2419	0.5714	0.9242
S_5	0.4348	0.3965	0.4286	0.8805
S_6	0.9130	0.9816	0.2857	0.8305
S_7	1	1	0.1429	0.8305

Tabel 5. Hasil Ekstraksi Fitur

C. Rantai Leksikal

Proses lexical chains merupakan tahapan dimana mencari chain terkuat pada kata yang terdapat di dalam kalimat. Alur proses perhitungan chain terkuat dapat dilihat pada Gambar 6.



Gambar 6. Proses Perhitungan Chain Terkuat

Setiap kalimat memiliki kata yang mungkin saja memiliki kesamaan dengan kalimat lain. Hal ini dapat diketahui dengan mengelompokkan kata-kata pada kalimat. Selain itu, setiap kalimat memiliki panjang chain dan jumlah pasangan kata yang berbeda, sehingga mempengaruhi nilai dari masing-masing kalimat. Kata-

kata yang memiliki kesamaan pada setiap kalimat, panjang chain, dan jumlah pasangan dapat dilihat pada Tabel 6.

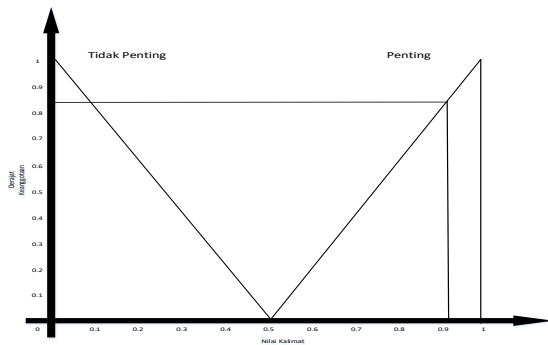
Kalimat (S_i)	Hasil
S_1	10.7663
S_2	15.8559
S_3	10.0992
S_4	10.3518
S_5	10.5711
S_6	7.7251
S_7	16.0597

Tabel 6. Hasil hitung rantai terkuat

D. Logika Fuzzy

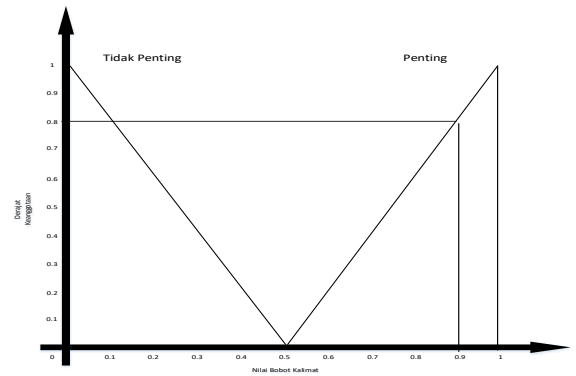
Setelah rantai terkuat diperoleh dari proses lexical chains, proses selanjutnya dilakukan proses logika fuzzy. Proses ini merupakan tahapan akhir setelah seluruh proses yang terlibat dalam peringkasan sebelumnya telah dilakukan dan kalimat telah siap untuk memberikan hasil ringkasan melalui proses ini. Proses ini mengelompokkan nilai dari hasil ekstraksi yang telah dilakukan sebelumnya ke dalam beberapa variabel linguistik untuk diketahui nilai linguistiknya. Proses yang terjadi pada logika fuzzy terdiri dari *fuzzification*, *inference system*, *defuzzification*. Hasil dari fitur ekstraksi dikelompokkan dalam dua kategori yaitu kalimat penting dan kalimat tidak penting. parameter yang digunakan dalam sistem ini adalah sebagai berikut:

- a. Panjang kalimat (f_1) \rightarrow Tidak Penting(TP) dan Penting(P)
fungsi keanggotaan untuk fitur panjang kalimat dapat dilihat pada gambar 7.



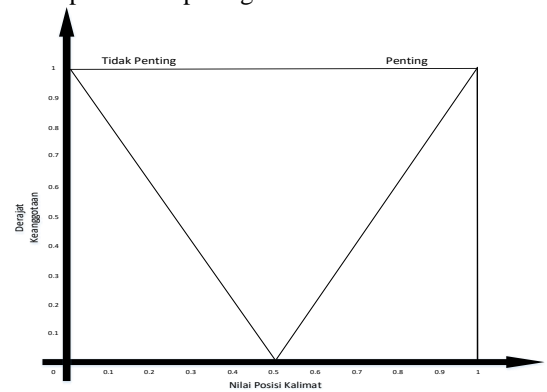
Gambar 7. Fungsi keanggotaan panjang kalimat

- b. Bobot kalimat (f_2) \rightarrow Tidak Penting(TP) dan Penting(P)
Fungsi keanggotaan untuk fitur bobot kalimat dapat dilihat pada gambar 8.



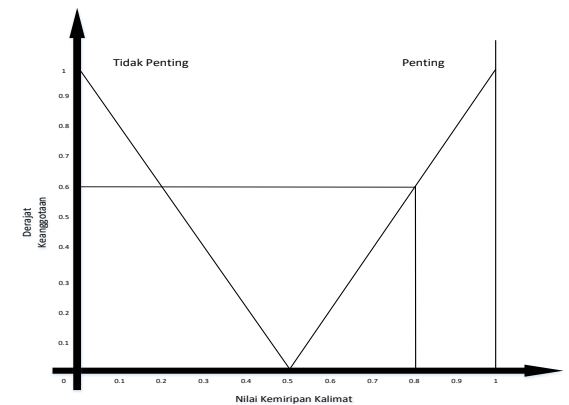
Gambar 8. Fungsi keanggotaan bobot kalimat

- c. Posisi kalimat (f_3) \rightarrow Tidak Penting(TP) dan Penting(P)
Fungsi keanggotaan untuk fitur posisi kalimat dapat dilihat pada gambar 9.



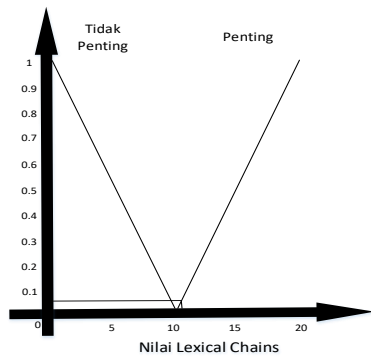
Gambar 10. Fungsi keanggotaan posisi kalimat

- d. Kemiripan antar kalimat (f_4) \rightarrow Tidak Penting(TP) dan Penting(P)
Fungsi keanggotaan untuk fitur bobot kalimat dapat dilihat pada gambar 10.



Gambar 10. Fungsi keanggotaan kemiripan antar kalimat

- e. Lexical chains (f_5) \rightarrow Tidak Penting(TP) dan Penting(P)
Fungsi keanggotaan untuk fitur bobot kalimat dapat dilihat pada gambar 11.



Gambar 11. Fungsi keanggotaan lexical chain

Proses terakhir merupakan kesimpulan dari hasil variabel linguistik yang diubah dengan menggunakan nilai crisp yang berfungsi untuk mewakili nilai akhir kalimat. Nilai-nilai diketahui dari hasil kemunculan nilai yang termasuk ke dalam kelompok ringkasan dan bukan ringkasan. Hasil perhitungan dari defuzzification dapat dilihat pada tabel 7.

Sentence (S_i)	Result
S_1	3.7851
S_2	11.0145
S_3	1.6601
S_4	1.4741
S_5	1.4737
S_6	4.1287
S_7	12.3816

Tabel 7. Hasil Defuzzification

Hasil yang diperoleh dari proses yang telah dilalui adalah dengan mengeluarkan ringkasan teks. Berdasarkan hasil dari defuzzification, kalimat kedua dan ketujuh merupakan ringkasan. Kemunculan ringkasan dihitung sebanyak 30% dari banyaknya kalimat dengan nilai defuzzifikasi terbesar. Ringkasan teks tersebut dapat dilihat pada gambar 8.

It has also led to the reorientation of government policy in particular, towards encouraging the spread of e-learning techniques and developing the skills and know-how required for their use. Despite its central importance in government policy and significant interest in the scale of the actual and potential market, there is an acute shortage of quantitative information on the extent of e learning in providing initial and continuing vocational education and training and on the rate at which it is growing.

Gambar 8. Hasil Peringkasan

III. PENGUJIAN DAN KESIMPULAN

Pengujian sistem merupakan proses untuk menampilkan sistem ATS dengan tujuan untuk mencari dan menemukan kesalahan pada sistem sebelum digunakan oleh pengguna. Keuntungan yang diperoleh dari hasil pengujian adalah mengetahui apakah sistem

ATS ini sudah baik atau belum. Pengujian yang akan dilakukan meliputi pengujian akurasi yang berarti menghitung keakuratan dari penerapan metode lexical chains dengan fuzzy.

A. Pengujian Akurasi

Pengujian akurasi merupakan ukuran yang menunjukkan kedekatan hasil analisis dengan hasil sebenarnya. Pada pengujian akurasi ini kualitas ringkasan ditentukan oleh *precision*, dan *f-measure* yang merupakan gabungan dari keduanya.

Skenario pengujian pada penelitian sistem peringkasan otomatis ini, dilakukan pengujian akurasi terhadap hasil ringkasan pada sistem dan hasil ringkasan manual. Hasil ringkasan manual diperoleh dari 3 orang ahli yang merupakan lulusan Sarjana Sastra Inggris, mengingat studi kasus yang diambil di dalam penelitian ini yaitu dokumen teks berbahasa Inggris.

Data yang digunakan dalam penelitian ini berjumlah 20 dokumen yang diperoleh dari situs sekolah bahasa Inggris. Peringkasan dokumen dilakukan dengan mengambil 30% kalimat dalam dokumen, baik untuk hasil yang dilakukan secara manual maupun hasil dari sistem. Pengujian ringkasan dapat dilihat pada tabel 9 berikut:

Dok	Jumlah Kalimat	KR	KRS	$\frac{\sum KRS}{\sum KR}$	Akurasi
1	7	1, 7	1, 7	2	100%
2	19	4, 14, 16, 13, 19, 18	4, 8, 11, 13, 18, 19	6	50%
3	15	3, 4, 6, 7	3, 4, 7, 15	4	75%
4	27	5, 6, 21, 16, 17, 22, 23, 24	5, 6, 15, 16, 17, 21, 22, 24	8	75%
5	29	1, 6, 8, 13, 15, 16, 18, 22, 27	1, 6, 8, 13, 15, 16, 17, 18, 27	9	88%
6	20	4, 8, 10, 12, 13, 20, 22	4, 7, 8, 10, 11, 13, 20	7	71%
7	19	1, 4, 6, 10, 11, 14	1, 4, 6, 10, 11, 14	6	100%
8	25	4, 5, 8, 15, 10, 17, 22	3, 4, 5, 7, 10, 15, 22	7	71%
9	14	7, 8, 11, 22	8, 9, 11, 22	4	75%
10	26	1, 6, 7, 8, 12, 15, 17, 26	1, 6, 7, 10, 12, 14, 17, 26	8	85%
11	15	7, 9, 10, 13	3, 7, 9, 10	4	75%
12	35	1, 3, 6, 18, 21, 24, 25, 27, 28, 30	1, 7, 8, 16, 18, 21, 24, 25, 26, 28	10	60%
13	22	6, 8, 11, 12, 16, 17, 22	3, 6, 8, 13, 16, 17, 22	7	57%
14	33	4, 8, 9, 10, 15, 17, 20, 22, 24, 30	3, 4, 8, 9, 10, 11, 15, 17, 22, 25	10	60%
15	13	2, 6, 7, 13	2, 6, 8, 12	4	50%
16	22	3, 5, 10, 12, 16, 21	3, 5, 6, 10, 11, 12	6	66%
17	26	2, 4, 7, 18, 19, 24, 26	2, 4, 13, 14, 18, 24, 26	7	71%
18	31	5, 9, 13, 18, 21, 25, 26, 31	5, 6, 11, 13, 18, 22, 25, 26, 31	9	67%
19	28	7, 6, 12, 11, 17, 14, 15, 20	6, 7, 11, 12, 14, 15, 17, 20	8	100%
20	58	2, 3, 7, 8, 9, 18, 23, 34, 45, 56	3, 4, 7, 8, 9, 18, 23, 43, 45, 56	10	80%

Tabel 9. Hasil pengujian akurasi

Keterangan :

KR = Kalimat relevan
KRS = Kalimat relevan dengan sistem
 Σ KR = Jumlah kalimat relevan
 Σ KRS = Jumlah kalimat relevan dengan sistem

Kesimpulan yang diperoleh dari hasil uji menyatakan bahwa kesesuaian kalimat penting yang dipilih oleh responden dengan sistem peringkasan otomatis mencapai akurasi diantara 50% hingga 100%. Sedangkan hasil Precision 67% dan F-measure sebesar 78%.

Dalam penelitian ini pengujian dengan human judgment belum dapat mewakili hasil peringkasan sistem. Hal ini disebabkan oleh sumberdaya (knowledge representation dan algoritma) belum memadai.

REFERENCES

- [1] N. Alami, M. Mekkassi and N. Rais, "AUTOMATIC TEXTS SUMMARIZATION: CURRENT STATE OF THE ART," *Journal of Asian Scientific Research*, pp. 1-15, 2015.
- [2] M. Pourvali and S. A. Mohammad, "Automated text summarization base on lexicales chain and graph using of word net and wikipedia knowledge base," *IJCSI International Journal of Computer Science Issues*, Vols. 9, No. 3, 2012.
- [3] M. Berker and T. Gungor, "Using Genetic Algorithms With Lexical Chains For Automatic Text Summarization," *International Conference on Agents and Artificial Intelligence (ICAART)*, 2012.
- [4] D. A. Riandayani and I. K. G. D. Putra, "Comparing Fuzzy Logic And Fuzzy C-Means (FCM) On Summarizing Indonesian Language Document," *Journal of Theoretical and Applied Information Technology*, vol. 59, no. 3, 2014.
- [5] L. Suanmali, N. Salim and M. S. Binwahlan, "Fuzzy Logic Based Method for Improving Text Summarization".
- [6] F. Kyoomarsi, H. Khosravi, E. Eslami and M. Davoudi, "EXTRACTION-BASED TEXT SUMMARIZATION USING FUZZY," *Iranian Journal of Fuzzy System*, vol. 7 no 3, pp. 15 - 32, 2010.
- [7] K. Farshad, K. Hamid, E. Esfandiari and K. D. Pooya, "Optimizing text summarization based on fuzzy logic," in *IEEE/ACIS International Conference on Computer and Information Science*, Kerman, 2008.