



## Digital Receipt

This receipt acknowledges that Turnitin received your paper. Below you will find the receipt information regarding your submission.

The first page of your submissions is displayed below.

Submission author: Assignment title: Paper\_1  
Submission title: berita\_Covid19\_Lexical\_Chain\_Word\_Sense\_Disambiguation\_...  
File name: berita\_Covid19\_Lexical\_Chain\_Word\_Sense\_Disambiguation\_...  
File size: 780.13K  
Page count: 7  
Word count: 3,297  
Character count: 20,287  
Submission date: 24-Jul-2023 01:01PM (UTC+0700)  
Submission ID: 2135923201

Komputika: Jurnal Sistem Komputer  
Volume 10, Nomor 2, Oktober 2021, hlm. 145 - 151  
Terakreditasi Peringkat 3, SK No. 28/E/KPT/2019  
DOI: 10.34010/komputika.v10i2.4499

ISSN: 2252-9039 (print)  
ISSN: 2655-3198 (online)

### Peringkasan Berita Online Corona Virus dengan Metode Lexical Chain dan Word Sense Disambiguation

Nisrina Arintia Maghfiroh<sup>1</sup>, Galih Wasis Wickaksono<sup>2</sup>, Christian Sri Kusuma Aditya<sup>3</sup>

<sup>1</sup>Program Studi Informatika, Fakultas Teknik, Universitas Muhammadiyah Malang  
Jl. Raya Tlogomas No.246 Malang, Jawa Timur, Indonesia 65144

<sup>2</sup>\*email: galih.w@ummm.ac.id

(Naskah masuk: 29 Januari 2021; diterima untuk diterbitkan: 19 April 2021)

**ABSTRAK** – Peringkasan berita otomatis merupakan aktivitas mengekstraksi inti dari berita tanpa mengurangi makna penting yang terdapat dalam berita tersebut. Dalam peringkasan berita otomatis terdapat beberapa metode yang dapat digunakan salah satunya yaitu metode Lexical Chain. Metode ini memiliki kinerja yang baik dalam peringkasan teks dengan cara menentukan chain tertinggi. Namun, metode ini memiliki kelemahan yaitu tidak bisa mengidentifikasi kata ambigu yang terdapat pada kalimat berita. Oleh karena itu, untuk memperbaiki kekurangan dari kelemahan metode Lexical Chain maka pada penelitian ini diteliti dengan Word Sense Disambiguation untuk mengidentifikasi kata ambigu. Penelitian ini menggunakan 100 berita tentang Covid-19 yang bersumber dari portal berita online terpopuler. Peringkat akurasi peringkasan berita otomatis yang digunakan dalam penelitian ini menggunakan Recall-Oriented Understudy for Gisting Evaluation (ROUGE). Adapun evaluasi yang digunakan pada penelitian ini ada tiga macam yaitu precision, recall, dan f-measure. Hasil evaluasi diperoleh nilai rata-rata precision sebesar 0,62, recall sebesar 0,20, dan f-measure sebesar 0,30.

**Kata Kunci** – Peringkasan otomatis, berita, Covid-19, Lexical Chain, Word Sense Disambiguation, ROUGE

### Summary of Online News about Corona Virus with Lexical Chain Method and Word Sense Disambiguation

**ABSTRACT** – Automatic news summaries are the activity of extracting the core from the news without compromising the significance contained in the news. In the automatic news release there are several methods that can be used, one of which is the Lexical Chain method. This method performs well in summarizing text by determining the highest chain. However, this method has the disadvantage of not being able to identify ambiguous words contained in news sentences. Therefore, to correct the shortcomings of the Lexical Chain method weaknesses, the study was equipped with Word Sense Disambiguation to identify ambiguous words. This study used 100 news about Covid-19 sourced from the most popular online news portals. Accuracy testing of automatic news releases used in this study using Recall-Oriented Understudy for Gisting Evaluation (ROUGE). While the evaluation used in this study there are three kinds of precision, recall, and f-measure. The evaluation results obtained an average precision value of 0.62, recall of 0.20, and f-measure of 0.30.

**Keywords** – Automatic Summarization, News, Lexical Chain, Word Sense Disambiguation, ROUGE, Covid-19

**1. PENDAHULUAN**  
Covid-19 diumumkan memasuki Indonesia pada bulan Maret 2020, dengan jumlah ± 900 ribu kasus sampai bulan Januari 2021. Seiring dengan penambuhan kasus harian, berdampak pada banyaknya pekarjan berita tentang Covid-19 banyak di internet. Hal ini berbanding lurus dengan

jumlah pengguna internet di Indonesia. Pada tahun 2019, pengguna aktif internet di Indonesia mencapai 150 juta[1]. Pengguna meningkat pada awal tahun 2020 yaitu mencapai 175,4 juta pengguna[2]. Di sisi lain peningkatan jumlah pengguna internet aktif juga secara tidak langsung memacu jumlah konten berita di internet.  
Teks pada berita online umumnya panjang dan

Tersedia di <https://ojs.ummm.ac.id/index.php/komputika>

# berita\_Covid19\_Lexical\_Chain\_ Word\_Sense\_Disambiguation\_R OUGE.pdf

*by*

---

**Submission date:** 24-Jul-2023 01:01PM (UTC+0700)

**Submission ID:** 2135923201

**File name:** berita\_Covid19\_Lexical\_Chain\_Word\_Sense\_Disambiguation\_ROUGE.pdf (780.13K)

**Word count:** 3297

**Character count:** 20287

## Peringkasan Berita Online Corona Virus dengan Metode *Lexical Chain* dan *Word Sense Disambiguation*

Nisrina Arintia Maghfiroh<sup>1</sup>, Galih Wasis Wicaksono<sup>2\*</sup>, Christian Sri Kusuma Aditya<sup>3</sup>

<sup>1</sup>Program Studi Informatika, Fakultas Teknik, Universitas Muhammadiyah Malang  
Jl. Raya Tlogomas No.246 Malang, Jawa Timur, Indonesia 65144

\*email: galih.w.w@umm.ac.id

(Naskah masuk: 29 Januari 2021; diterima untuk diterbitkan: 19 April 2021)

**ABSTRAK** – Peringkasan berita otomatis merupakan aktivitas mengekstraksi inti dari berita tanpa mengurangi makna penting yang terdapat dalam berita tersebut. Dalam peringkasan berita otomatis terdapat beberapa metode yang dapat digunakan salah satunya yaitu metode *Lexical Chain*. Metode ini memiliki kinerja yang baik dalam peringkasan teks dengan cara menentukan chain tertinggi. Namun, metode ini memiliki kelemahan yaitu tidak bisa mengidentifikasi kata ambigu yang terdapat pada kalimat berita. Oleh karena itu, untuk memperbaiki kekurangan dari kelemahan metode *Lexical Chain* maka pada penelitian ini dilengkapi dengan *Word Sense Disambiguation* untuk mengidentifikasi kata ambigu. Penelitian ini menggunakan 100 berita tentang Covid-19 yang bersumber dari portal berita online terpopuler. Pengujian akurasi peringkasan berita otomatis yang digunakan dalam penelitian ini menggunakan *Recall-Oriented Understudy for Gisting Evaluation (ROUGE)*. Adapun evaluasi yang digunakan pada penelitian ini ada tiga macam yaitu *precision*, *recall*, dan *f-measure*. Hasil evaluasi diperoleh nilai rata-rata *precision* sebesar 0,62, *recall* sebesar 0,20, dan *f-measure* sebesar 0,30.

**Kata Kunci** – Peringkasan otomatis, berita, Covid-19, *Lexical Chain*, *Word Sense Disambiguation*, *ROUGE*

## Summary of Online News about Corona Virus with *Lexical Chain* Method and *Word Sense Disambiguation*

**ABSTRACT** – Automatic news summarizes are the activity of extracting the core from the news without compromising the significance contained in the news. In the automatic news release there are several methods that can be used, one of which is the *Lexical Chain* method. This method performs well in summarizes text by determining the highest chain. However, this method has the disadvantage of not being able to identify ambiguous words contained in news sentences. Therefore, to correct the shortcomings of the *Lexical Chain* method weaknesses, the study was equipped with *Word Sense Disambiguation* to identify ambiguous words. This study used 100 news about Covid-19 sourced from the most popular online news portals. Accuracy testing of automatic news releases used in this study using *Recall-Oriented Understudy for Gisting Evaluation (ROUGE)*, while the evaluation used in this study there are three kinds of *precision*, *recall*, and *f-measure*. The evaluation results obtained an average *precision* value of 0.62, *recall* of 0.20, and *f-measure* of 0.30.

**Keywords** - Automatic Summarization, News, *Lexical Chain*, *Word Sense Disambiguation*, *ROUGE*, Covid-19

### 1. PENDAHULUAN

Covid-19 diumumkan memasuki Indonesia pada bulan Maret 2020, dengan jumlah ± 900 ribu kasus sampai bulan Januari 2021. Seiring dengan penambahan kasus harian, berdampak pada banyaknya pencarian berita tentang Covid-19 banyak di internet. Hal ini berbanding lurus dengan

jumlah pengguna internet di Indonesia. Pada tahun 2019, pengguna aktif internet di Indonesia mencapai 150 juta[1]. Pengguna meningkat pada awal tahun 2020 yaitu mencapai 175,4 juta pengguna[2]. Di sisi lain peningkatan jumlah pengguna internet aktif juga secara tidak langsung memicu jumlah konten berita di internet.

Teks pada berita *online* umumnya panjang dan

tidak langsung tertuju pada inti berita, sedangkan pembaca lebih cepat memahami isi berita dengan membaca ringkasan daripada membaca keseluruhan berita tersebut. Akan tetapi, tidak semua berita memiliki ringkasan teks sehingga pembaca harus membaca keseluruhan dari berita[3]. Ringkasan dalam konteks dokumen berbasis teks adalah informasi-informasi penting yang dihasilkan dari dokumen. Meringkas berita secara manual oleh manusia memiliki keterbatasan, oleh karena itu solusi yang dibutuhkan yaitu peringkasan teks secara otomatis (*automatic text summarization*)[4].

Peringkasan teks otomatis adalah mengambil inti dari berita tanpa mengurangi konten dan makna penting dari berita[5]. Metode ekstraksi dan metode abstraksi merupakan dua metode yang digunakan untuk meringkas teks [6]. Metode ekstraksi (*shallower approaches*) merupakan pendekatan yang digunakan untuk menyalin informasi penting dan mengubahnya menjadi ringkasan. Ringkasan dibuat dengan mengekstraksi kalimat dari dokumen asli tanpa mengubah makna dari kalimat asli, langkah penting yang dilakukan adalah menilai pentingnya berbagai kalimat. Selanjutnya, subset dari kalimat-kalimat dengan skor teratas dipertahankan untuk meminimalkan redundansi. Metode abstraksi (*deeper approaches*) menyajikan informasi penting dari teks asli dengan bahasa yang berbeda tetapi tidak mengubah makna dari teks asli[7]. Peringkasan abstraksi membuat ringkasan yang berisi kalimat-kalimat baru yang tidak tersedia dalam dokumen asli, metode ini menggunakan frasa dan klausa dari representasi asli meskipun keseluruhan teks masih dianggap baru. Meski begitu, ringkasan yang dihasilkan belum tentu berisi kalimat yang mempunyai makna karena secara umum peringkasan abstraksi jauh lebih sulit [8].

Berdasarkan dua metode yang telah dijabarkan. teknik ekstraksi dapat dengan mudah diskalakan untuk memberikan ringkasan pada kasus berita

*online* terkait Covid-19 dibandingkan dengan metode abstraksi yang memiliki keterbatasan pada kamus input dalam hal ini berita [9]. Penelitian ini menggunakan pendekatan ekstraksi dengan memanfaatkan metode *Lexical Chain*. Selain itu untuk meningkatkan keakuratan metode, khususnya terhadap kata ambigu maka penelitian ini menambahkan *Word Sense Disambiguation* pada alur peringkasan berita.

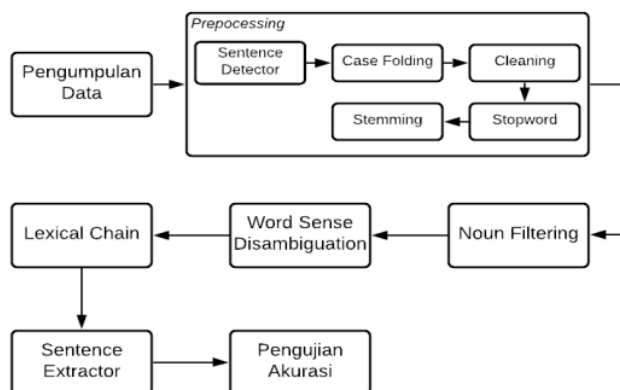
## 2. METODE DAN BAHAN

Implementasi metode *Lexical Chain* dengan *Word Sense Disambiguation* pada kasus peringkasan teks otomatis diperlukan sistem yang dapat meringkas teks dokumen secara otomatis. Metode *multi-document* akan digunakan akan digunakan dalam penelitian ini. Tahapan yang pertama yaitu pengumpulan data, *text preprocessing*, *noun filtering*, *word sense disambiguation*, *lexical chain*, *sentence extractor* dan tahap terakhir pengujian akurasi. Alur peringkasan teks pada penelitian ini dapat dilihat pada Gambar 1.

### A. Pengumpulan Data

Data yang digunakan pada penelitian ini diambil dari lima website berita *online* Indonesia yang terpopuler menurut [www.alexacom](http://www.alexacom), yaitu [www.detik.com](http://www.detik.com), [www.kompas.com](http://www.kompas.com), [www.tribunnews.com](http://www.tribunnews.com), [www.liputan6.com](http://www.liputan6.com), dan [www.okezone.com](http://www.okezone.com). Data yang digunakan merupakan berita tentang Covid-19 sejumlah 100 berita yang diambil menggunakan *web scraping*.

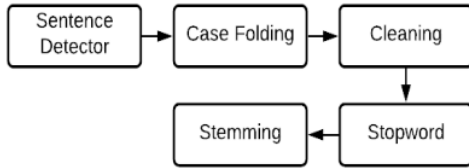
*Scraping* dilakukan dengan mencari link berita yang mengandung kata "Covid" yang ada pada lima website berita menggunakan *library request* dan *selector* dengan total link tiap website 20 link. Selanjutnya, pengambilan data tiap link menggunakan *library BeautifulSoup*. Data-data yang dikumpulkan kemudian dikelola dengan format url berita, judul berita dan isi berita.



Gambar 1. Alur Peringkasan Teks Otomatis

## B. Preprocessing

*Preprocessing* dilakukan dengan mengolah teks berita menjadi *term* yang akan dibobot[10]. Tahapan *preprocessing* diilustrasikan pada Gambar 2.



Gambar 2. Tahapan Preprocessing

1. *Sentence Detector*: tahap pemisah kalimat berita menjadi kumpulan kalimat-kalimat dengan tanda titik (.) sebagai delimeter untuk memotong *string* dokumen[7].
2. *Case Folding*: proses yang digunakan untuk mengubah semua huruf kapital menjadi huruf kecil (*lower case*)[11].
3. *Cleaning*: proses untuk menghilangkan karakter-karakter tidak penting pada berita seperti tanda baca, dan sebagainya[7].
4. *Stopword*: proses yang dilakukan untuk menghilangkan kata yang tidak penting. Pada proses ini, setiap hasil dari tahap *cleaning* dicek berdasarkan *stoplist* kata, dan jika ada kata yang memiliki kesamaan kata maka kata tersebut akan hilang. Kata tidak penting merupakan kata yang tidak memiliki makna seperti “yang”, “di”, “untuk”, “dari”, dan sebagainya[6].
5. *Stemming*: *Stemming* adalah proses yang dilakukan untuk menghilangkan kata yang memiliki awalan, sisipan dan akhiran menjadi kata dasar[11]. Pada tahap ini, menerapkan algoritma Nazief dan Andriani yang merupakan algoritma untuk mengubah kata ber-sufiks, prefiks dan konfiks menjadi kata dasar.

## C. Noun Filtering

*Noun filtering* merupakan proses yang dilakukan untuk memilih atau menyaring kata-kata yang memiliki kata bersifat benda (*noun*) dari kumpulan kata yang dihasilkan pada tahap *preprocessing*. Penyaringan kata benda dilakukan dengan cara pengecekan kata dengan database kata benda menggunakan Kateglo ([www.kateglo.com](http://www.kateglo.com)). Kateglo merupakan layanan web kamus, tesaurus dan glosarium bahasa Indonesia.

*Noun filtering* yang dihasilkan akan menjadi input untuk membentuk *lexical chain*. *Noun filtering* dapat meningkatkan kinerja dari peringkasan teks karena suatu kalimat memiliki komponen subjek yang bertipe *noun phrase* atau *gerund*, dan kata kerja (*verb*) digunakan sebagai predikatnya sehingga *noun* dianggap sebagai *head* dari sebuah kalimat. Selain itu,

relasi semantik yang dimiliki *noun* lebih banyak dibandingkan *verb* atau *adjective*[7].

## D. Word Sense Disambiguation

Metode *Lexical Chain* memiliki kelemahan dalam mengidentifikasi kata ambigu untuk pembentukan rantai leksikal. *Word sense disambiguation* dapat dimanfaatkan untuk membantu proses identifikasi makna kata pada suatu kalimat tertentu yang memiliki beberapa makna yang berbeda (ambigu) [7].

Pada tahap ini, makna kata diidentifikasi dengan menggunakan algoritma Lesk yang bekerja dengan cara membandingkan definisi kata ambigu dengan definisi kata tetangganya berdasarkan definisi dari kamus. Setelah pengambilan makna kata ambigu dari web Kateglo selanjutnya dilakukan proses menghitung bobot kata. Nilai dari pembobotan kata diperoleh dari perbandingan banyaknya kata yang sama pada makna kata ambigu dengan makna kata pembanding dan dihitung bobotnya[6].

## E. Lexical Chain

*Lexical Chain* adalah metode yang dilakukan untuk pembentukan rantai leksikal. Kata benda, kata sifat dan kata kerja pada kalimat yang berelasi harus berkaitan dengan *Lexical Chain*. Hubungan leksikal yang dibentuk harus kalimat yang berkaitan satu sama lain sehingga relasi antar leksikal juga berkaitan[7]. Untuk menghitung rantai leksikal, digunakan rumus sebagai berikut:

$$\text{Score (chain)} = \text{length} \times \text{homogeneity index} \quad (1)$$

$$\text{Homogeneity index} = 1 - \left( \frac{\text{distinct member}}{\text{length}} \right) \quad (2)$$

Keterangan:

- a. *Length*: total keseluruhan dari kemunculan seluruh anggota penyusun rantai.
- b. *Distinct member*: banyaknya anggota penyusun rantai yang berbeda
- c. *Homogeneity index*: nilai dari anggota-anggota penyusun rantai yang memiliki kemiripan.

Proses menunjukkan tingkat kemiripan anggota pembentuk rantai dilakukan melalui perhitungan *homogeneity index*. Agar dapat menentukan tingkat perbedaan anggota pembentuk rantai, *distinct member* dibagi dengan *length*. Setelah melewati tahap *noun filtering* dan *word sense disambiguation*, maka kandidat kata yang dipilih digunakan untuk pembentukan *lexical chain* [6].

## F. Sentence Extractor

Proses *sentence extractor* bertujuan untuk menemukan kalimat yang akan diekstrak sebagai penyusun hasil ringkasan. *Sentence extractor*

diperoleh dari nilai *score* pada *lexical chain* dibagi dengan total kalimat.

Sebagai upaya meningkatkan akurasi, maka pada proses *sentence extractor* ditambahkan juga metode Laplacian *Smoothing* yang bertujuan menghindari angka 0 akibat dari tidak ditemukannya data *test* pada data *training* [12]. Hasil akhir akan di rangking berdasarkan nilai tertinggi.

### G. Pengujian Akurasi

Perhitungan akurasi pada penelitian ini menggunakan metode *Recall-Oriented Understudy for Gisting Evaluation* (ROUGE). ROUGE adalah teknik yang digunakan untuk mengevaluasi ringkasan dengan membandingkan hasil ringkasan secara otomatis dengan hasil ringkasan oleh pakar [13].

Pada penelitian ini jenis evaluasi ROUGE yang digunakan ialah ROUGE-L (*Longest Common Subsequence*) ada tiga macam yaitu *precision*, *recall*, dan *F-measure* [14]. Rumus menghitung nilai *precision*, *recall* dan *F-measure* dijelaskan sebagai berikut:

$$Precision = \frac{hits}{noise} \quad (3)$$

$$Recall = \frac{hits}{misses} \quad (4)$$

$$F - measure = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (5)$$

Keterangan :

- hits*: jumlah kata penting pada ringkasan
- misses*: jumlah kata pada ringkasan manusia (pakar)
- noise*: jumlah kata pada ringkasan otomatis

## 3. HASIL DAN PEMBAHASAN

Prosedur *preprocessing* dilakukan dengan tahapan melakukan ekstraksi kalimat dari paragraf menggunakan *delimiter* pemisah titik (.). Kalimat-kalimat yang memiliki huruf kapital (*upper case*) diubah menjadi huruf kecil (*lower case*). Selanjutnya menghilangkan tanda baca serta kata yang tidak penting. Tahapan selanjutnya adalah mengubah kata yang berhasil diseleksi menjadi kata dasar. Berdasarkan contoh satu berita *online* yang telah di akuisisi dan dilakukan *preprocessing* diperoleh sepuluh kalimat dengan jumlah kata dasar 104 kata.

Setelah melakukan *preprocessing*, maka proses selanjutnya ialah melakukan seleksi kata benda (*noun filtering*). *Noun* dianggap sebagai *head* kalimat karena memiliki relasi semantik yang lebih banyak dibandingkan *verb* atau *adjective*. Berdasarkan hasil *noun filtering*, diperoleh 41 kata benda dari 104 kandidat kata.

### 3.1. Word Sense Disambiguation

Tahapan ini dilakukan untuk menemukan kata ambigu dengan cara memetakan makna dari kata ambigu dari website Kateglo.

Tabel 1 Makna Kata Ambigu

Kata Ambigu	Makna Kata
Negara	1. Organisasi dalam suatu wilayah yang mempunyai kekuasaan tertinggi yang sah dan ditaati oleh rakyat. 2. Kelompok sosial yang menduduki wilayah atau daerah tertentu yang diorganisasi di bawah lembaga politik dan pemerintah yang efektif, mempunyai kesatuan politik, berdaulat sehingga berhak menentukan tujuan nasionalnya.

Aktivitas selanjutnya adalah pembobotan kata. Nilai dari pembobotan kata diperoleh dari perbandingan kesamaan kata antara makna kata ambigu dengan makna kata pembanding, proses pembobotan contoh berita ditampilkan pada Tabel 2.

Tabel 2 Proses Pembobotan Kalimat

Kata Ambigu: Negara		
Makna Kata	Kata Pembanding	Makna Pembanding
Organisasi dalam suatu wilayah yang mempunyai kekuasaan tertinggi yang sah dan ditaati oleh rakyat.	Karantina	Tempat penampungan yang lokasinya terpencil guna mencegah terjadinya penularan suatu penyakit.
Kelompok sosial yang menduduki wilayah atau daerah tertentu yang diorganisasi di bawah lembaga politik dan pemerintah yang efektif, mempunyai kesatuan politik, berdaulat	Virus	Mikroorganisme yang tidak dapat dilihat dengan menggunakan mikroskop biasa, hanya dapat dilihat dengan menggunakan mikroskop elektron, penyebab dan penular penyakit. Tempat penampungan yang lokasinya terpencil guna mencegah terjadinya penularan suatu penyakit. Mikroorganisme yang tidak dapat dilihat dengan menggunakan mikroskop biasa, hanya dapat

Kata Ambigu: Negara		
Makna Kata	Kata Pembeding	Makna Pembeding
sehingga berhak menentukan tujuan nasionalnya.		dilihat dengan menggunakan mikroskop elektron, penyebab dan penular penyakit.
<b>Bobot Kalimat</b>	4	3

Dari perhitungan di atas dengan hasil yang tertera, maka makna kata ambigu yang tertinggi terdapat pada makna 1 dengan nilai bobot 4.

### 3.2. Lexical Chain

Pada tahap ini diperlukan *score* untuk menyusun rantai leksikal, untuk mencari *score* diperlukan nilai *homogeneity index* yang digunakan untuk menunjukkan tingkat kemiripan anggota pembentuk rantai. Sebelumnya perlu mengukur nilai *distinct member* dan *length*. Perhitungan *distinct member* pada contoh berita dapat dilihat pada Tabel 3.

Tabel 3 Proses Perhitungan *distinct member*

Kata Ambigu: Negara				
S-ke	Lexical Chain	Makna	Nilai	Length
1	Spanyol dan Iran telah melonggarkan aturan karantinanya, meski wabah virus masih berlanjut di kedua negara tersebut.	Organisasi dalam suatu wilayah yang mempunyai kekuasaan tertinggi yang sah dan ditaati oleh rakyat.	15	16

Setelah menghitung *distinct member*, akan dilakukan perhitungan menggunakan metode *lexical chain*. Contoh perhitungannya bisa dilihat pada Tabel 4.

Tabel 4 Perhitungan Lexical Chain

Lexical Chain	Proses
Spanyol dan Iran telah melonggarkan aturan karantinanya, meski wabah virus masih berlanjut di kedua negara tersebut.	$Homo = 1 - \left(\frac{15}{16}\right) = 0,0625$

### 3.3. Sentence Extractor

Pada *sentence extractor* terdapat tahapan-tahapan yakni, *sentence scoring* dan *sentence extractor*. Perhitungan *sentence extractor* membutuhkan nilai *score* yang dibagi dengan total kalimat. Tabel 5 menunjukkan proses *sentence scoring*.

Tabel 5 Perhitungan <i>Sentence Scoring</i>		
Indeks Kalimat	Proses	$\sum$ Kalimat Dokumen
Spanyol dan Iran telah melonggarkan aturan karantinanya, meski wabah virus masih berlanjut di kedua negara tersebut.	$Score = 16 \times 0,0625 = 1$	10

Proses perhitungan *sentence extractor* kemudian dioptimalkan dengan metode *Laplacian Smoothing*. Optimalisasi dimaksudkan untuk menghilangkan angka 0 pada data *training* [15]. Perhitungan *sentence extractor* pada contoh berita dengan metode *Laplacian Smoothing* diuraikan pada Tabel 6.

Tabel 5 Perhitungan *Sentence Extractor*

Rank	Kalimat	Nilai <i>Sentence Extractor</i>
1	S10	$\left(\frac{1,998}{10}\right) + 1 = 0,1998 + 1 = 1,1998$
2	S9	$\left(\frac{1,001}{10}\right) + 1 = 0,1001 + 1 = 1,1001$
3	S1	$\left(\frac{1}{10}\right) + 1 = 0,1 + 1 = 1,1$
4	S2	$\left(\frac{0}{10}\right) + 1 = 0 + 1 = 1$
5	S5	$\left(\frac{1}{10}\right) + 1 = 0,1 + 1 = 1,1$
6	S6	$\left(\frac{1}{10}\right) + 1 = 0,1 + 1 = 1,1$
7	S7	$\left(\frac{1}{10}\right) + 1 = 0,1 + 1 = 1,1$
8	S8	$\left(\frac{1}{10}\right) + 1 = 0,1 + 1 = 1,1$
9	S3	$\left(\frac{0,999}{10}\right) + 1 = 0,0999 + 1 = 1,0999$
10	S4	$\left(\frac{0,999}{10}\right) + 1 = 0,0999 + 1 = 1,0999$

Setelah mendapatkan hasil dari perhitungan *sentence extractor*, selanjutnya dilakukan perankingan, yakni dengan mengurutkan bobot kalimat dari yang terbesar ke yang terkecil bobot kalimatnya. Proses terakhir dari tahap-tahap di atas yaitu mengetahui hasil ringkasan yang dihitung 30% dari banyaknya kalimat dokumen berita [4].

Pada tahap menentukan ringkasan berita, kalimat dipulihkan dan siap ditampilkan menjadi sebuah ringkasan [6]. Hasil ringkasan pada contoh berita ditunjukkan pada Tabel 7.

Tabel 6 Hasil ringkasan berita

Hasil Ringkasan
Spanyol menerapkan kebijakan tersebut usai mencatatkan peningkatan harian terendah dalam jumlah kasus dan korban meninggal beberapa hari terakhir. Pantauan dari jurnalis Reuters memperlihatkan, orang-orang di pusat transportasi utama diberikan masker oleh polisi saat mereka berangkat kerja pada Senin pagi. Spanyol dan Iran telah melonggarkan aturan karantananya, meski wabah virus masih berlanjut di kedua negara tersebut.

### 3.4. Pengujian Akurasi

Pengujian akurasi dengan ROUGE-L akan menghitung akurasi dengan persentase data peringkasan yang digunakan sebanyak 20%, 25% dan 30% tertinggi dari total dataset berita *online*. Hasil perhitungan akurasi pada 100 berita ditunjukkan pada Tabel 8.

Tabel 7 Perhitungan Akurasi

	20%	25%	30%
<i>Precision</i>	0,702148	0,662789	0,627974
<i>Recall</i>	0,159374	0,180446	0,207058
<i>F-measure</i>	0,257079	0,281278	0,308948

Merujuk pada Tabel 8, ditemukan bahwa, hasil akurasi tertinggi berada pada persentase 30% dari keseluruhan data yaitu nilai rata-rata *precision* sebesar 0,62, nilai rata-rata *recall* sebesar 0,20 dan nilai rata-rata *f-measure* sebesar 0,30. Maka dari hasil tersebut, persentase 30% lebih optimal.

## 4. KESIMPULAN

Penelitian ini telah berhasil menemukan bahwa dengan metode *lexical chain* dan *word sense disambiguation* nilai terbaik peringkasan dapat diperoleh dengan menggunakan 30% rangking tertinggi dari *dataset* yang digunakan. Hal tersebut menunjukkan bahwa semakin banyak data yang digunakan akan meningkatkan persentase akurasi dari hasil peringkasan. Penelitian lebih lanjut diharapkan dapat menyelesaikan ambiguitas kata tanpa bergantung pada kamus kata. Selain itu, perlu dilakukan pengujian terhadap kombinasi metode lain yang dapat meningkatkan akurasi peringkasan multi-dokumen, serta mengembangkan perhitungan *homogeneity index* pada *lexical chain*.

## DAFTAR PUSTAKA

- [1] K. Simon, "Digital in 2016," 2016. <https://wearesocial.com/special-reports/digital-in-2016>.
- [2] K. Simon, "Digital in 2020," 2020. <https://wearesocial.com/digital-2020>.
- [3] B. S. Pratiwi, Shaufiah, and M. A. Bijaksana, "Implementasi Word Sense Disambiguation Dengan Metode Maximal Marginal Relevance Pada Peringkasan Teks," *eProceedings of Engineering*, vol. 4, no. 1. pp. 1152-1157, 2017.
- [4] N. I. Widiastuti and W. K. Afnan, "Fuzzy Logic Dan Lexical Chains Untuk Peringkasan Teks Otomatis," *J. Sist. Komput.*, vol. 7, no. 1, pp. 5-12, 2017, doi: 10.14710/jsk.v7i1.130.
- [5] M. Fahmi, M. Arif, A. Fatchul, and S. Gunung, "Implementation of Automatic Text Summarization with TextRank Method in the Development of Al-Qur'an Vocabulary Encyclopedia Method in the Development of Al-Qur'an Vocabulary Encyclopedia," *Procedia Comput. Sci.*, vol. 179, no. 2020, pp. 391-398, 2021, doi: 10.1016/j.procs.2021.01.021.
- [6] D. M. Fazar and N. I. Widiastuti, "Lexical Chain dan Word Sense Disambiguation Untuk Peringkasan Artikel Berbahasa Indonesia," *Techno.Com*, vol. 16, no. 2, pp. 195-207, 2017, doi: 10.33633/tc.v16i2.1413.
- [7] D. M. Fazar, "Implementasi Metode Lexical Chain dengan Word Sense Disambiguation untuk Peringkasan Artikel Kesehatan," *Ilm. Komput. dan Inform.*, vol. 1, no. 1, pp. 1-6, 2016.
- [8] C. C. Aggarwal, *Machine learning for text*, 1st ed. New York: Springer International Publishing, 2018.
- [9] A. Mahajani, V. Pandya, I. Maria, and D. Sharma, "A Comprehensive Survey on Extractive and Abstractive Techniques for Text Summarization," *Ambient Communications and Computer Systems*, pp. 339-351, 2019, doi: 10.1007/978-981-13-5934-7.
- [10] R. Widiastutik and J. Santoso, "Peringkasan Teks Ekstraktif pada Dokumen Tunggal Menggunakan Metode Restricted Boltzmann Machine," *J. Intell. Syst. Comput.*, vol. 1, no. 2, pp. 58-64, 2019, doi: 10.52985/insyst.v1i2.84.
- [11] H. Irmayanti, "Analisis Algoritma Fuzzy Logic dalam Pengklasifikasian Tugas Akhir," *Komputika: Jurnal Sistem Komputer*, vol. 7, no. 2, pp. 71-77, 2018, doi: 10.34010/komputika.v7i2.1401.
- [12] I. Listiowarni, "Implementasi Naïve Bayesian dengan Laplacian Smoothing untuk Peminatan dan Lintas Minat Siswa SMAN 5 Pamekasan," *J. Sisfokom (Sistem Inf. dan Komputer)*, vol. 8, no. 2, p. 124, 2019, doi: 10.32736/sisfokom.v8i2.652.



- [13] F. Anshari, "Peringkasan Multi Dokumen Berbahasa Indonesia Menggunakan Metode Vikor," *Skripsi Universitas Komputer Indonesia*, 2019.
- [14] A. N. Ammar and Suyanto, "Peringkasan Teks Ekstraktif Menggunakan Binary Firefly Algorithm," *Indones. J. Comput.*, vol. 5, no. September, pp. 31-42, 2020, doi: 10.21108/indojc.2020.5.2.440.
- [15] K. E. Dewi and E. Widiastuti, Nelly Indriani Rainarli, "Evaluasi Sentence Extraction pada Peringkasan Dokumen Otomatis," *SNIA (Seminar Nasional Informatika dan Aplikasinya)*, vol. 3, pp. 8-12, 2017.

ORIGINALITY REPORT

---

**20%**

SIMILARITY INDEX

**18%**

INTERNET SOURCES

**10%**

PUBLICATIONS

**10%**

STUDENT PAPERS

---

MATCH ALL SOURCES (ONLY SELECTED SOURCE PRINTED)

---

3%

★ Submitted to Universitas International Batam

Student Paper

---

Exclude quotes      On

Exclude matches      Off

Exclude bibliography      On