



Tersedia online di www.journal.unipdu.ac.id
Unipdu

Halaman jurnal di www.journal.unipdu.ac.id/index.php/register



Peringkasan multi-dokumen berita berdasarkan fitur berita dan *part of speech tagging*

Moch. Zawaruddin Abdullah^a, Chastine Fatichah^b

^{a,b} Teknik Informatika, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia

email: ^a zawaruddin017@gmail.com, ^b chastine.fatichah@gmail.com

INFO ARTIKEL

Sejarah artikel:

Menerima 7 Agustus 2018

Revisi 9 Agustus 2018

Diterima 9 Agustus 2018

Online 10 Agustus 2018

Kata kunci:

dokumen berita

fitur berita

informasi gramatikal

peringkasan multi-dokumen

part of speech tagging

Keywords:

grammatical information

multi-document summarization

news document

news feature

part of speech tagging

Style APA dalam mensitasi artikel ini:

Abdullah, M. Z., & Fatichah, C. (2018). Peringkasan multi-dokumen berita berdasarkan fitur berita dan *part of speech tagging*. *Register: Jurnal Ilmiah Teknologi Sistem Informasi*, 4(2), 56-66.

ABSTRAK

News Feature Scoring (NeFS) merupakan metode pembobotan kalimat yang sering digunakan untuk melakukan pembobotan kalimat pada peringkasan dokumen berdasarkan fitur berita. Beberapa fitur berita diantaranya seperti *word frequency*, *sentence position*, *Term Frequency-Inverse Document Frequency* (TF-IDF), dan kemiripan kalimat terhadap judul. Metode NeFS mampu memilih kalimat penting dengan menghitung frekuensi kata dan mengukur similaritas kata antara kalimat dengan judul. Akan tetapi pembobotan dengan metode NeFS tidak cukup, karena metode tersebut mengabaikan kata informatif yang terkandung dalam kalimat. Kata-kata informatif yang terkandung pada kalimat dapat mengindikasikan bahwa kalimat tersebut penting. Penelitian ini bertujuan untuk melakukan pembobotan kalimat pada peringkasan multi-dokumen berita dengan pendekatan fitur berita dan informasi gramatikal (NeFGIS). Informasi gramatikal yang dibawa oleh *part of speech tagging* (POS Tagging) dapat menunjukkan adanya konten informatif. Pembobotan kalimat dengan pendekatan fitur berita dan informasi gramatikal diharapkan mampu memilih kalimat representatif secara lebih baik dan mampu meningkatkan kualitas hasil ringkasan. Pada penelitian ini terdapat 4 tahapan yang dilakukan antara lain seleksi berita, *text preprocessing*, *sentence scoring*, dan penyusunan ringkasan. Untuk mengukur hasil ringkasan menggunakan metode evaluasi *Recall-Oriented Understudy for Gisting Evaluation* (ROUGE) dengan empat varian fungsi yaitu ROUGE-1, ROUGE-2, ROUGE-L, dan ROUGE-SU4. Hasil ringkasan menggunakan metode yang diusulkan (NeFGIS) dibandingkan dengan hasil ringkasan menggunakan metode pembobotan dengan pendekatan fitur berita dan *trending issue* (NeFTIS). Metode NeFGIS memberikan hasil yang lebih baik dengan peningkatan nilai untuk fungsi *recall* pada ROUGE-1, ROUGE-2, ROUGE-L, dan ROUGE-SU4 secara berturut-turut adalah 20,37%, 33,33%, 1,85%, 23,14%.

ABSTRACT

News Feature Scoring (NeFS) is a sentence weighting method that used to weight the sentences in document summarization based on news features. There are several news features including *word frequency*, *sentence position*, *Term Frequency-Inverse Document Frequency* (TF-IDF), and *sentences resemblance to the title*. The NeFS method is able to select important sentences by calculating the frequency of words and measuring the similarity of words between sentences and titles. However, NeFS weighting method is not enough, because the method ignores the informative word in the sentence. The informative words contained in the sentence can indicate that the sentence is important. This study aims to weight the sentence in news multi-document summarization with news feature and grammatical information approach (NeFGIS). Grammatical information carried by *part of speech tagging* (POS Tagging) can indicate the presence of informative content. Sentence weighting with news features and grammatical information approach is expected to be able to determine sentence

representatives better and be able to improve the quality of the summary results. In this study, there are 4 stages that are carried out including news selection, text preprocessing, sentence scoring, and compilation of summaries. Recall-Oriented Understanding for Gisting Evaluation (ROUGE) is used to measure the summary results with four variants of function; ROUGE-1, ROUGE-2, ROUGE-L, and ROUGE-SU4. Summary results using the proposed method (NeFGIS) are compared with summary results using sentence weighting methods with news feature and trending issue approach (NeFTIS). The NeFGIS method provides better results with increased value for recall functions in ROUGE-1, ROUGE-2, ROUGE-L, and ROUGE-SU4 respectively 20.37%, 33.33%, 1.85%, 23.14%.

© 2018 Register: Jurnal Ilmiah Teknologi Sistem Informasi. Semua hak cipta dilindungi undang-undang.

1. Pendahuluan

Ekstraksi kalimat merupakan proses pengambilan kalimat penting yang sering dilakukan dalam peringkasan dokumen. Salah satu metode dalam melakukan ekstraksi kalimat adalah dengan cara pembobotan kalimat. Pembobotan kalimat ditujukan untuk menyeleksi kalimat-kalimat penting sebagai dasar penyusunan ringkasan. Kalimat penting penyusun suatu ringkasan harus memiliki banyak informasi sebagai representasi dari dokumen sumber (He, Li, Shao, Chen, & Ma, 2008) dan juga memiliki kata-kata penting yang menjadi dasar penyusunan dari kalimat tersebut (Wan, Yang, & Xiao, 2007). Metode pembobotan kalimat untuk peringkasan dokumen berita sebagian besar memanfaatkan fitur dari berita itu sendiri atau disebut sebagai *News Feature Scoring* (NeFS) diantaranya *word frequency*, *Term Frequency-Inverse Document Frequency* (TF-IDF), posisi kalimat dan kemiripan kalimat terhadap judul (Hayatin, Faticah, & Purwitasari, 2014).

Terdapat beberapa penelitian yang menggunakan fitur berita untuk melakukan pembobotan kalimat. Penelitian Radev, Hovy, dan McKeown (2002) menggunakan fitur berita untuk melakukan pembobotan kalimat, yaitu posisi kalimat, *centroid*, dan kemiripan kalimat terhadap kalimat pertama dari berita. Metode yang diusulkan mampu memberikan hasil yang cukup relevan dalam peringkasan dokumen. Akan tetapi metode tersebut tidak memperhatikan relasi antara judul dengan isi dokumen sehingga pemilihan kalimat untuk peringkasan dokumen kurang koheren (keterpaduan informasi). Penelitian Ferreira, dkk (2013) menggunakan berbagai fitur berita untuk pembobotan kalimat, akan tetapi tidak semua kombinasi fitur menghasilkan nilai bobot kalimat yang efektif, hanya terdapat beberapa kombinasi metode yang paling efektif dalam melakukan pembobotan untuk peringkasan dokumen berita, yaitu *word frequency*, TF-IDF, posisi kalimat, dan kemiripan kalimat terhadap judul (*resemblance to the title*). Penelitian Hayatin, Faticah, dan Purwitasari (2014) melakukan pembobotan kalimat menggunakan pendekatan fitur berita (*word frequency*, TF-IDF, posisi kalimat, dan kemiripan kalimat terhadap judul) dan *trending issue* yang disebut NeFTIS. Metode NeFTIS mampu menyeleksi kalimat penting berdasarkan kata-kata penting yang sesuai dengan *trending issue* dari data Twitter. Akan tetapi metode tersebut tidak memperhatikan kata-kata informatif yang ada pada dokumen berita itu sendiri, dan melakukan seleksi *term* penting hanya berdasarkan frekuensi sehingga *term* yang memiliki frekuensi kemunculan tinggi akan dianggap sebagai *term* penting. Padahal tidak hanya sebatas frekuensi kemunculan suatu *term* dianggap penting, melainkan juga konten informatif yang dikandung dalam suatu kalimat. Kata-kata informatif yang terkandung pada kalimat dapat mengindikasikan bahwa kalimat tersebut penting.

Selain penggunaan fitur berita (NeFS) dalam pembobotan kalimat, informasi gramatikal (*grammatical information*) yang terdapat pada suatu kalimat juga dapat berpengaruh dalam pembobotan kalimat seperti contoh kelas kata. Kelas kata yang dimaksud adalah penggolongan kata dalam satuan bahasa berdasarkan kategori bentuk, fungsi, dan makna dalam sistem gramatikal. Dalam Bahasa Indonesia, kelas kata dibagi menjadi 5 kelompok besar yaitu kata kerja (*verb*), kata sifat (*adjective*), kata keterangan (*adverb*), kata benda (*noun*), dan kata tugas (*function word*) (Pisceldo, Adriani, & Manurung, 2009). Informasi gramatikal yang dibawa oleh *part of speech tagging* (POS Tagging) dapat menunjukkan adanya konten informatif. Terdapat beberapa penelitian yang menggunakan pendekatan *POS Tagging* dalam pembobotan kalimat, seperti penelitian Lioma dan Blanco (2017), Fachrurrozi, Yusliani, dan Yoanita (2013), dan Arifin, dkk (2018). Kelas kata yang terkandung dalam informasi gramatikal mampu memberikan peningkatan bobot kalimat, karena penggunaan *POS Tagging* mampu mendefinisikan

kelas kata pada suatu kalimat dan menentukan kata mana yang merupakan kata informatif. Kata-kata informatif yang terkandung pada kalimat dapat mengindikasikan bahwa kalimat tersebut penting. Kata informatif memberikan nilai lebih dalam pembobotan kalimat, sehingga proses pemilihan kalimat representatif untuk peringkasan dokumen lebih efektif.

Berdasarkan informasi dari beberapa penelitian yang telah disebutkan, maka penelitian ini bertujuan untuk melakukan peringkasan multi-dokumen berita dengan mengusulkan metode pembobotan dengan pendekatan fitur berita (NeFS) dan informasi gramatikal. Di mana informasi gramatikal ditujukan untuk mengetahui adanya kata informatif pada suatu kalimat yang dapat mengindikasikan bahwa kalimat tersebut penting. Pemilihan kalimat penting dengan skema tersebut diharapkan mampu memilih kalimat representatif sebagai dasar penyusunan ringkasan dan mampu meningkatkan kualitas hasil ringkasan.

2. State of the Art

Automatic text summarization adalah proses peringkasan dokumen teks yang dilakukan secara otomatis melalui mesin komputer. Otomatisasi ringkasan dapat dilakukan terhadap satu dokumen atau banyak dokumen (Aditya, Faticah, & Purwitasari, 2016). Peringkasan dokumen dapat dibedakan menjadi dua yaitu *abstractive* dan *extractive*. Peringkasan *abstractive* dilakukan dengan cara menulis ulang kalimat berdasarkan pemahaman terhadap dokumen sumber dan disampaikan dengan menghasilkan kalimat baru dari dokumen yang diringkas. Peringkasan *extractive* dilakukan dengan cara ekstraksi kalimat penting yang ada di dalam dokumen yang akan dijadikan sebagai kandidat ringkasan berdasarkan fitur statistik dan linguistik, kemudian menggabungkannya untuk membentuk sebuah ringkasan (Mandar & Gunawan, 2017). Berikut penjelasan tentang landasan teori yang diambil dari pustaka penelitian sebelumnya sebagai bahan acuan dalam melakukan penelitian ini.

2.1. Grammatical information by part of speech tagging

Part of Speech Tagging (POS Tagging) merupakan suatu proses yang mengidentifikasi informasi gramatikal yaitu kelas kata pada suatu kata dalam kalimat. *Part of Speech Tagging* merupakan bagian dari *Natural Language Processing* (NLP) dalam menentukan kelas kata (Rashel, Luthfi, Dinakaramani, & Manurung, 2014). Bentuk sederhana dari proses ini adalah identifikasi kata pada sebuah kalimat dan memberikan label kelas kata, baik sebagai kata benda, kata kerja, kata sifat, dll. Dalam penelitian Pisceldo, Adriani, dan Manurung (2009) terdapat 25 tag label kelas kata untuk Bahasa Indonesia dengan rincian 9 label untuk simbol dan tanda baca, dan 16 label untuk kata seperti pada Tabel 1.

Tabel 1. Label POS Tagging untuk Bahasa Indonesia

No	Tag	Deskripsi	No	Tag	Deskripsi
1	(Kurung buka	14	WH	5W1H / Kata tanya
2)	Kurung tutup	15	VB	Kata kerja
3	,	Koma	16	MD	Modal/pendandaian
4	.	Titik	17	JJ	Kata sifat
5	:	Titik dua	18	CD	Bilangan pokok
6	--	Tanda garis	19	NEG	Negasi
7	"	Tanda kutip pembuka	20	IN	Kata depan
8	"	Tanda kutip penutup	21	CC	Konjungsi koordinatif
9	SYM	Simbol	22	SC	Konjungsi subordinatif
10	NN	Kata benda umum	23	RB	Kata keterangan
11	NNP	Kata benda nama diri	24	WDT	WH-determiners
12	PRP	Kata ganti orang	25	FW	Kata asing
13	PR	Kata ganti umum			

2.2. Sentence length threshold

Konsep dari *Sentence Length Threshold* (W_{st}) adalah jumlah kata penyusun suatu kalimat harus sekurang-kurangnya sesuai dengan *threshold* (batas ambang) untuk diproses ke suatu sistem. Dalam Bahasa Indonesia, suatu kalimat harus memiliki minimal satu subjek dan satu predikat. Kalimat yang terlalu panjang ataupun terlalu pendek harus dihindari dalam sebuah ringkasan sehingga dibutuhkan batasan (*threshold*) panjang kalimat yang diproses dalam pembobotan kalimat (Meena & Gopalani, 2014) (Meena & Gopalani, 2015).

Berdasarkan penelitian Ferreira, dkk (2013), kalimat representatif yang terlalu panjang atau terlalu pendek tidak dianggap sebagai pemilihan optimal. Untuk itu dapat digunakan perhitungan dengan menentukan batas ambang (*threshold*) dari jumlah minimal dan jumlah maksimal kata yang dapat dilakukan proses pembobotan kalimat. Di mana kalimat dengan nilai 1 akan diproses pembobotan kalimat, sedangkan kalimat dengan nilai 0 akan diabaikan. Penentuan nilai dapat dihitung dengan menggunakan kondisi seperti Persamaan 1,

$$W_{slt} = \begin{cases} 1, & \text{jika } (L_{min} < \text{panjang}(s) < L_{maks}) \\ 0, & \text{lainnya} \end{cases} \quad (1)$$

di mana L_{min} adalah jumlah minimal kata pada kalimat, didefinisikan oleh *user*, sedangkan L_{maks} adalah jumlah maksimal kata pada kalimat, didefinisikan oleh *user*.

2.3. Term Frequency – Inverse Sentence Frequency (TF-ISF)

Skema pembobotan TF-ISF atau *Term Frequency – Inverse Sentence Frequency* adalah variasi dari pembobotan TF-IDF. TF-ISF merupakan fitur yang digunakan pada penelitian (Fachrurrozi, Yusliani, & Yoanita, 2013) (Meena & Gopalani, 2015). Di mana *Term Frequency* (TF) merupakan frekuensi kemunculan dari sebuah *term t* pada kalimat *s* pada satu dokumen. Sedangkan *Inverse Sentence Frequency* (ISF) merupakan sebuah perhitungan bagaimana *term t* terdistribusi secara menyeluruh pada koleksi kalimat dalam satu dokumen. Pembobotan berdasarkan TF-ISF dapat mengacu pada Persamaan 2,

$$\begin{aligned} TF.ISF(s, t) &= TF(s, t) \times ISF(t) \\ &= f(s, t) \times \log\left(1 + \frac{N}{sf(t)}\right) \end{aligned} \quad (2)$$

Di mana N adalah jumlah total kalimat dalam satu dokumen, sedangkan $sf(t)$ merupakan jumlah kalimat yang terdapat *term t*.

2.4. Term Frequency – Inverse Document Frequency (TF-IDF)

Skema pembobotan suatu *term* pada *text mining* yang secara luas digunakan adalah skema *Term Frequency – Inverse Document Frequency* (TF-IDF). *Term frequency* (TF) merupakan frekuensi kemunculan dari sebuah *term* pada suatu koleksi dokumen. Sedangkan *Inverse Document Frequency* (IDF) merupakan sebuah penghitungan bagaimana *term* terdistribusi secara menyeluruh pada koleksi dokumen yang diproses. IDF menunjukkan hubungan ketersediaan suatu *term* terhadap koleksi dokumen. Semakin sedikit jumlah dokumen yang mengandung *term*, maka nilai IDF semakin besar. Pembobotan berdasarkan TF-IDF dapat mengacu pada Persamaan 3,

$$\begin{aligned} TF.IDF(s, t) &= TF(s, t) \times IDF(t) \\ &= f(s, t) \times \log\left(1 + \frac{N}{df(t)}\right) \end{aligned} \quad (3)$$

Di mana N adalah jumlah total koleksi dokumen yang diproses, sedangkan $df(t)$ merupakan jumlah dokumen yang terdapat *term t*.

2.5. Sentence position

Sentence position atau posisi kalimat merupakan salah satu fitur berita yang digunakan pada pembobotan kalimat. Di mana penilaiannya berdasarkan letak kalimat dalam sebuah dokumen. Pembobotan ini mengadopsi penelitian Mei dan Chen (2012) yang menyatakan bahwa kalimat yang berada di awal dokumen memiliki bobot yang lebih besar dibanding dengan kalimat yang berada di akhir dokumen. Hal ini dipertegas oleh penelitian Verdianto, Arifin, dan Purwitasari (2016) yang melakukan evaluasi terhadap hasil peringkasan, di mana kalimat yang berada pada awal dokumen berita cenderung memiliki bobot yang tinggi daripada kalimat yang berada di akhir dokumen. Hal ini didasarkan atas pernyataan bahwa sebagian besar berita cenderung menyampaikan ide pokoknya pada awal-awal kalimat, sedangkan kalimat-kalimat selanjutnya merupakan penjelas atau bahkan informasi-informasi lain di luar pokok bahasan. $W_{pos}(S_i)$ adalah pembobotan kalimat berdasarkan posisi kalimat dirumuskan seperti pada Persamaan 4,

$$W_{pos}(S_i) = \frac{1}{\sqrt{Pos(S_i)}} \quad (4)$$

Di mana $Pos(S_i)$ adalah posisi kalimat pada sebuah dokumen.

2.6. Kemiripan kalimat terhadap judul (*Resemblance to the title*)

Konsep dari pembobotan kalimat berdasarkan kemiripan kalimat terhadap judul (W_{sim}) adalah pembobotan yang dilakukan berdasarkan tingkat kemiripan kalimat terhadap judul. Metode pembobotan ini mengacu pada penelitian Ferreira, dkk (2013), di mana kata-kata dalam kalimat yang terdapat pada judul merupakan kalimat penting. Semakin tinggi kemiripan kalimat dengan judul, maka semakin tinggi nilai bobot kalimat. Pembobotan berdasarkan kemiripan kalimat terhadap judul dapat diformulasikan seperti Persamaan 5,

$$W_{sim} = \frac{NTW}{T} \quad (5)$$

di mana NTW adalah jumlah *term* pada judul yang terdapat pada kalimat, sedangkan T adalah jumlah total *term* yang ada pada judul.

2.7. Part of Speech Tagging Local Distribution

Part of Speech Tagging Local Distribution (W_{POS-l}) adalah pembobotan kalimat berdasarkan pendekatan *POS Tagging* pada satu dokumen. Pembobotan ini ditujukan untuk menghitung tingkat persebaran *term* yang berlabel *POS p* pada satu dokumen. Mengadopsi dari penelitian Arifin, dkk (2018) yang menyebutkan bahwa sebuah *term* yang tersebar luas pada suatu dokumen memiliki nilai yang tinggi dibandingkan *term* lain, karena sebuah *term* yang tersebar dalam dokumen lebih mempresentasikan topik dari dokumen tersebut. Dari pernyataan tersebut dapat diformulasikan mengenai persebaran *term* dengan pendekatan *term frequency* (TF) dan *POS Tagging* seperti pada Persamaan 6,

$$W_{POS-l}(t) = TF_{d,p,t} \times PW_p \quad (6)$$

di mana $TF_{d,p,t}$ adalah *Term frequency* untuk *term j* yang berlabel *POS p* pada dokumen d , sedangkan PW_p adalah nilai bobot untuk label *POS p*.

2.8. Part of Speech Tagging Global Distribution

Part of Speech Tagging Global Distribution (W_{POS-g}) adalah pembobotan kalimat berdasarkan pendekatan *POS Tagging* pada koleksi dokumen. Pembobotan ini ditujukan untuk menghitung tingkat persebaran *term* yang berlabel *POS p* pada koleksi dokumen dengan menggunakan pendekatan *term frequency* (TF) dan *POS Tagging* pada seluruh dokumen yang diproses. Pembobotan kalimat ini dapat dirumuskan seperti Persamaan 7,

$$W_{POS-g}(t) = TF_{p,t} * PW_p \quad (7)$$

di mana $TF_{p,t}$ adalah *Term frequency* untuk *term j* yang berlabel *POS p*, sedangkan PW_p adalah nilai bobot untuk label *POS p*.

3. Metode Penelitian

Secara garis besar ada empat fase yang dilakukan pada perancangan sistem. Empat fase tersebut adalah seleksi berita, *text preprocessing*, *sentence scoring*, dan penyusunan ringkasan.

3.1. Seleksi berita

Fase seleksi berita ini dilakukan untuk mengelompokkan berita berdasarkan kategori dan tiap kategori berita memiliki beberapa topik berita. Data berita yang akan diproses terlebih dahulu akan diseleksi berdasarkan topik berita dan tiap topik berita berisi beberapa dokumen berita.

3.2. Text preprocessing

Fase *Text preprocessing* adalah fase yang dilakukan untuk mengolah dokumen berita menjadi token/*term* untuk dijadikan *input* dalam proses pembobotan kalimat. Proses *text preprocessing* dilakukan terhadap isi dokumen berita maupun judul berita. Terdapat 7 proses dalam *text preprocessing* yaitu:

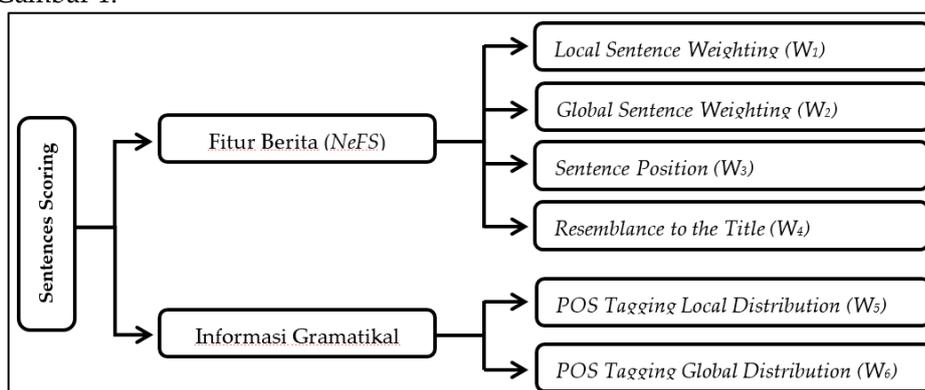
- 1) *Sentences segmentation* adalah proses pemenggalan teks dokumen menjadi kumpulan kalimat. Pemilihan kalimat representatif untuk peringkasan multi-dokumen berita tidak luput dari

segmentasi kalimat yang diambil dari dokumen. Proses segmentasi kalimat sangat penting dalam proses pembobotan kalimat. Kesalahan dalam segmentasi kalimat dapat mengakibatkan kesalahan dalam menentukan kalimat representatif sehingga hasil peringkasan tidak sesuai. Segmentasi kalimat yang baik mampu memilah kalimat dalam suatu teks secara benar dan terstruktur sehingga pemilihan kalimat menjadi utuh dan dapat meningkatkan proses pemilihan kalimat representatif untuk peringkasan dokumen. Dalam hal ini proses *sentence segmentation* menggunakan *library sastrawi sentence-detector for Bahasa Indonesia*. *Library* ini mampu melakukan segmentasi kalimat dengan benar dan utuh.

- 2) *Part of Speech Tagging (POS Tagging)* adalah proses untuk mengidentifikasi kelas kata yang ada dalam suatu kalimat. Dalam proses mengidentifikasi kelas kata, sistem menggunakan *Application Program Interface (API)* dari *Part of Speech* khusus Bahasa Indonesia yang beralamat di <http://bahasa.cs.ui.ac.id/postag/tagger>.
- 3) *Case Folding* adalah proses konversi seluruh karakter huruf pada suatu teks atau kalimat menjadi bentuk format yang sama (menjadi huruf capital atau huruf kecil). Dalam hal ini *case folding* yang diterapkan adalah dengan menjadikan karakter huruf berformat huruf kecil (*lowercase*).
- 4) *Tokenizing* adalah proses pemenggalan teks atau kalimat menjadi unit-unit kecil pembentuk kalimat yaitu *term/kata*.
- 5) *Stopword* adalah proses filtrasi kata-kata umum (*common words*) yang biasanya muncul dalam frekuensi besar dan dianggap tidak memiliki makna dalam suatu kalimat. Dalam penelitian ini tidak semua *list stopwords* dipakai untuk proses filtrasi, *list* yang dipakai adalah kata-kata umum yang menjadi kata penghubung.
- 6) *Stemming* adalah proses yang dilakukan untuk merubah kata bentukan menjadi kata dasar.
- 7) *Sentences length threshold* adalah proses filtrasi terhadap kalimat yang akan diproses apabila kalimat memiliki batas ambang (*threshold*), baik pada batas ambang jumlah kata minimal maupun batas ambang jumlah kata maksimal. Kalimat yang kurang dari ambang minimal atau kalimat yang lebih dari ambang maksimal akan diabaikan dari proses pembobotan kalimat. Hal ini ditujukan untuk optimasi pemilihan kalimat agar kalimat representatif yang menjadi kalimat ringkasan tidak terlalu pendek atau kalimat tidak terlalu panjang.

3.3. Sentence Scoring

Fase *sentences scoring* adalah tahapan proses pembobotan kalimat yang dilakukan untuk menentukan kalimat representatif sebagai kalimat penyusun ringkasan. Terdapat dua komponen utama dalam *sentences scoring* yaitu fitur berita (*News Feature Scoring / NeFS*) dan *grammatical information* seperti yang ditampilkan Gambar 1.



Gambar 1. Desain *Sentence Scoring*

Berdasarkan Gambar 1, pada komponen Fitur Berita terdapat empat subkomponen dalam melakukan pembobotan kalimat yaitu *Local Sentence Weighting (W₁)*, *Global Sentence Weighting (W₂)*, *Sentence Position (W₃)*, dan *Resemblance to the Title (W₄)*. Pada komponen Informasi Gramatikal terdapat dua subkomponen yaitu *POS Tagging Local Distribution (W₅)*, dan *POS Tagging Global Distribution (W₆)*. Keenam subkomponen inilah yang akan digunakan dalam melakukan pembobotan kalimat untuk peringkasan multi-dokumen berita.

1) *Local Sentence weighting* (W_1)

Local sentence weighting (W_1) merupakan pembobotan kalimat yang ditujukan untuk memberikan bobot kalimat terhadap dokumen tunggal, di mana kalimat yang memiliki bobot paling tinggi merupakan kalimat representatif yang mewakili konten dalam satu dokumen. Pembobotan kalimat ini menggunakan pendekatan *Term Frequency – Inverse Sentence Frequency* (TF-ISF) dengan menghitung jumlah bobot *term t* yang muncul pada kalimat *i* (S_i) seperti yang ditunjukkan pada Persamaan 8,

$$W_1(S_i) = \log \left(1 + \frac{\sum_{t=1}^k TF.ISF_{it}}{k} \right), \quad (8)$$

di mana k adalah jumlah total term pada kalimat i (S_i).

2) *Global Sentence Weighting* (W_2)

Global sentence weighting (W_2) merupakan pembobotan kalimat yang ditujukan untuk memberikan bobot kalimat terhadap kata yang muncul pada koleksi dokumen. Kata yang sama dan tersebar pada koleksi dokumen, mengindikasikan kata tersebut penting dan merepresentasikan tingkat similaritas dokumen. Kata penting pada kalimat tersebut dapat mewakili sebagian dokumen dalam menentukan kalimat representatif untuk peringkasan dokumen. Pembobotan ini menggunakan pendekatan *Term Frequency-Inverse Document Frequency* (TF-IDF) dengan menghitung jumlah bobot *term t* yang muncul pada kalimat i (S_i) seperti yang ditunjukkan pada Persamaan 9,

$$W_2(S_i) = \log \left(1 + \frac{\sum_{t=1}^k TF.IDF_{it}}{k} \right), \quad (9)$$

di mana k adalah jumlah total term pada kalimat i (S_i).

3) *Sentence Position* (W_3)

Sentence position (W_3) merupakan pembobotan kalimat berdasarkan posisi kalimat dalam suatu dokumen. Proses pembobotan ini mengadopsi penelitian Mei dan Chen (2012), di mana kalimat yang berada di awal dokumen memiliki skor yang tinggi dibanding kalimat dengan posisi terakhir. Pembobotan kalimat berdasarkan posisi kalimat dapat dirumuskan seperti pada Persamaan 10,

$$W_3(S_i) = \frac{1}{\sqrt{Pos(S_i)}}, \quad (10)$$

di mana $Pos(S_i)$ merupakan posisi kalimat pada dokumen.

4) *Resemblance to the Title* (W_4)

Resemblance to the Title (W_4) merupakan pembobotan yang dilakukan berdasarkan tingkat kemiripan kalimat terhadap judul. Metode pembobotan ini mengacu pada penelitian Ferreira, dkk (2013), di mana kata-kata dalam kalimat yang terdapat pada judul merupakan kalimat penting. Pembobotan berdasarkan kemiripan kalimat terhadap judul dapat dirumuskan seperti Persamaan (11),

$$W_4(S_i) = \frac{NTW}{T}, \quad (11)$$

di mana NTW adalah jumlah *term* pada judul yang terdapat pada kalimat, sedangkan T adalah jumlah total *term* yang ada pada judul.

5) *POS Tagging Local Distribution* (W_5)

POS Tagging Local Distribution (W_5) ditujukan untuk menghitung tingkat persebaran *term* yang berlabel *POS p* pada satu dokumen. Pembobotan ini menggunakan Persamaan 6 dengan menghitung jumlah bobot *term t* yang muncul pada kalimat i (S_i) seperti yang ditunjukkan pada Persamaan 12,

$$W_5(S_{d,i}) = \log \left(1 + \frac{\sum_{t=1}^k (TF_{d,p,t} \times PW_p)}{k} \right), \quad (12)$$

di mana $TF_{d,p,t}$ adalah *Term frequency* untuk *term* t yang berlabel *POS* p pada dokumen d , PW_p adalah nilai bobot untuk label *POS* p , dan k adalah jumlah total *term* pada kalimat i .

6) POS Tagging Global Distribution (W_6)

POS Tagging Global Distribution (W_6) digunakan untuk mengetahui tingkat persebaran *term* yang belabel *POS* p pada koleksi dokumen. Pembobotan ini menggunakan Persamaan 7 dengan menghitung jumlah bobot *term* t yang muncul pada kalimat i (S_i) seperti yang ditunjukkan pada Persamaan 13,

$$W_6(S_i) = \log \left(1 + \frac{\sum_{t=1}^k (TF_{p,t} \times PW_p)}{k} \right), \quad (13)$$

di mana $TF_{p,t}$ adalah *Term frequency* untuk *term* t yang berlabel *POS* p , PW_p adalah nilai bobot untuk label *POS* p , dan k adalah jumlah total *term* pada kalimat i .

Setelah keenam pembobotan dilakukan, tahapan selanjutnya adalah akumulasi bobot pada tiap kalimat. Pembobotan W_1 dan W_5 sama-sama menghitung pada lingkup *local* (satu dokumen), maka dalam akumulasi ini dilakukan proses perkalian ($W_1 \times W_5$). Begitu pula dengan W_2 dan W_6 yang melakukan perhitungan pada lingkup *global* (koleksi dokumen), sehingga dalam perhitungan ini dilakukan proses perkalian ($W_2 \times W_6$). Proses ini terinspirasi atas penelitian (Wahib, Arifin, & Purwitasari, 2016), di mana proses perkalian antara W_1 dengan W_5 , dan W_2 dengan W_6 dilakukan agar bobot yang dimiliki pada masing-masing lingkup (*local* dan *global*) dapat saling menguatkan. Jika pada lingkup *local* W_1 atau *global* W_2 sebuah kalimat memiliki bobot yang besar, namun dalam lingkup *local* W_5 atau *global* W_6 ternyata kalimat tersebut memiliki bobot yang kecil, maka dengan perkalian ini akan menurunkan nilai kalimat tersebut, begitu juga sebaliknya. Skor atau total bobot akhir dalam pembobotan kalimat ($Score(S_i)$) didapat dengan cara menjumlahkan bobot yang telah dilakukan menggunakan Persamaan 14. Kombinasi pembobotan inilah yang selanjutnya disebut NeFGIS (*News Feature and Grammatical Information Score*),

$$Score(S_i) = (W_1 * W_5) + (W_2 * W_6) + W_3 + W_4 \quad (14)$$

3.4. Penyusunan Ringkasan

Proses penyusunan ringkasan dilakukan berdasarkan keenam pembobotan kalimat W_1 sampai dengan W_6 yang menghasilkan total bobot kalimat ($Score(S_i)$). Proses pengambilan kalimat representatif dilakukan dengan cara mengurutkan secara *descending* terhadap total bobot kalimat yang kemudian diambil sebanyak *top-n* kalimat. Hal ini diasumsikan bahwa semakin besar total bobot yang dimiliki oleh sebuah kalimat maka kalimat tersebut merupakan kalimat penting.

Tabel 2. Dataset uji coba

No	Topik	Jumlah Artikel	Total Kalimat
1	Air-Asia	7	88
2	Banjarnegara	15	195
3	BBM	12	204
4	BPJS	17	295
5	Dolly	9	180
6	Ebola	7	86
7	Kurikulum 2013	23	403
8	Palestina	17	186
9	Pilpres	18	231
10	Sinabung	6	87
11	U19	9	142
Total:		140	2.097

4. Hasil dan Pembahasan

4.1. Dataset uji coba

Dataset uji coba yang digunakan pada penelitian ini berasal dari *dataset* berita berbahasa Indonesia dari penelitian Hayatin, Faticah, dan Purwitasari (2014) yang diambil secara manual dari beberapa situs berita *online* berdasarkan rekomendasi *link* yang ada pada *tweets* dengan 11 topik berbeda. Topik-topik yang dipakai yaitu “*air asia*”, “*Banjarnegara*”, “*BBM*”, “*bpjs*”, “*dolly*”, “*pilpres*”, “*ebola*”, “*kurikulum 2013*”, “*sinabung*”, “*palestina*”, dan “*u19*”. Di mana total *dataset* sebanyak 140 berita. Untuk lebih detail dapat mengacu pada Tabel 2. Data *training* yang digunakan pada penelitian ini menggunakan 4 topik dari *dataset* yang telah ada yaitu topik “*BBM*”, “*BPJS*”, “*Kurikulum 2013*”, dan “*pilpres*”. Sedangkan untuk proses *testing* dilakukan terhadap ke-11 *dataset* yang ada.

4.2. Hasil uji coba

Uji coba dilakukan untuk mengukur performa dari hasil ringkasan antara metode pembobotan NeFGIS dengan metode pembobotan berdasarkan fitur berita/*news feature* dan *trending issue* (NeFTIS). Di mana untuk mengukur performa kedua metode pembobotan tersebut digunakan metode evaluasi ROUGE. ROUGE (*Recall-Oriented Understudy for Gisting Evaluation*) adalah metode yang digunakan untuk mengukur kualitas dari sebuah ringkasan berdasarkan penelitian (Lin, 2004). ROUGE akan membandingkan antara rangkuman yang dihasilkan oleh sistem terhadap rangkuman ideal (*Groundtruth*) yang dibuat oleh pakar. Berdasarkan penelitian penelitian Lin (2004) untuk evaluasi ringkasan multi-dokumen ROUGE-1, ROUGE-2, ROUGE-L dan ROUGE-SU4 memiliki performa yang baik. Oleh karena itu penelitian ini menggunakan keempat fungsi evaluasi tersebut dengan tiga pengukuran yaitu *recall*, *precision*, dan *f-score*.

Kelas kata (label *POS*) yang ada pada informasi gramatikal memiliki peran penting dalam menentukan adanya kata informatif dalam suatu kalimat. Nilai label *POS* diperoleh berdasarkan penelitian Lioma dan Blanco (2017) yang menjelaskan tentang Jespersen’s Rank Theory, di mana label *POS* dapat diberi peringkat ke dalam empat tingkat dari bobot terbesar hingga terkecil, yaitu: 1) Kata benda (*noun*) dengan bobot 1; 2) Kata sifat (*adjective*), kata kerja (*verb*), dan partisipan (*participles*) dengan bobot 0,75; 3) Kata keterangan (*adverb*) dengan bobot 0,5; dan 4) Semua label *POS* yang tersisa dengan bobot 0,25. Sehingga dalam penelitian ini, bobot label *POS* (*PW*) berdasarkan empat tingkat tersebut seperti Persamaan 15,

$$PW = \{1; 0,75; 0,5; 0,25\} \quad (15)$$

Uji coba dilakukan dengan menggunakan sampel topik “*pilpres*” untuk kedua metode pembobotan dengan tiga variasi *groundtruth*. Topik “*pilpres*” dipilih karena mengacu pada penelitian Hayatin, Faticah, dan Purwitasari (2014), bahwa topik “*pilpres*” adalah topik yang memiliki skenario dan ringkasan terbaik daripada topik lainnya. Hasil dari pembobotan akan dipilih 10 kalimat yang memiliki bobot tertinggi ($n = 10$) sebagai kalimat penyusun ringkasan.

Tabel 3. Hasil evaluasi NEFGIS dan NEFTIS terhadap *groundtruth* untuk $N = 10$

ROUGE	NeFGIS			NeFTIS		
	<i>Recall</i>	<i>Precision</i>	<i>F-score</i>	<i>Recall</i>	<i>Precision</i>	<i>F-score</i>
ROUGE-1	0,3279	0,4781	0,3890	0,2724	0,4061	0,3261
ROUGE-2	0,1770	0,2844	0,2182	0,1328	0,1841	0,1543
ROUGE-L	0,3418	0,4406	0,3849	0,3356	0,3128	0,3238
ROUGE-SU4	0,1959	0,3138	0,2413	0,1591	0,2250	0,1864

Tabel 3 menunjukkan hasil evaluasi NeFGIS unggul secara keseluruhan dibandingkan dengan metode NeFTIS terhadap *groundtruth*. Di mana NeFGIS memiliki nilai yang lebih tinggi daripada NeFTIS untuk perhitungan *recall*, *precision*, dan *f-score* pada evaluasi ROUGE-1, ROUGE-2, ROUGE-L, maupun ROUGE-SU4. Hal ini dapat dikatakan bahwa hasil ringkasan dari NeFGIS lebih relevan dibandingkan dengan hasil ringkasan NeFTIS. Lebih jelas tentang perbedaan nilai peningkatan metode NeFGIS terhadap metode NeFTIS dapat dilihat dengan mengacu pada Tabel 4.

Tabel 4. Peningkatan hasil evaluasi NeFGIS terhadap NeFTIS untuk topik “*pilpres*”

ROUGE	<i>Recall</i>	<i>Precision</i>	<i>F-score</i>
-------	---------------	------------------	----------------

ROUGE-1	0,05550 (↑20,37%)	0,07201 (↑17,73%)	0,06292 (↑19,30%)
ROUGE-2	0,04425 (↑33,33%)	0,10027 (↑54,46%)	0,06393 (↑41,43%)
ROUGE-L	0,00622 (↑ 1,85%)	0,12782 (↑40,87%)	0,06119 (↑18,90%)
ROUGE-SU4	0,03683 (↑23,14%)	0,08878 (↑39,46%)	0,05483 (↑29,41%)
Rata-rata	0,03570 (↑19,67%)	0,09722 (↑38,13%)	0,06072 (↑27,26%)

Tabel 5. Contoh *term* dengan multi-label POS.

<i>Term</i>	Label POS	Contoh Kata
	<i>Noun</i>	Pelaku
Laku	<i>Verb</i>	Melakukan, berlaku, memperlakukan
	<i>Adjective</i>	Laku

Berdasarkan Tabel 4 dapat disimpulkan performa dari metode yang diusulkan yaitu NeFGIS memiliki kualitas hasil ringkasan lebih baik dibandingkan dengan metode NeFTIS. Hal ini dimungkinkan karena seleksi *term* penting pada metode NeFTIS hanya berdasarkan frekuensi, sehingga *term* yang memiliki frekuensi kemunculan tinggi akan dianggap sebagai *term* penting. Padahal ada kemungkinan *term* yang frekuensi kemunculannya kecil juga dapat menjadi *term* penting. Di sisi lain NeFGIS menyeleksi *term* berdasarkan fitur berita dan informasi gramatikal, sehingga pemilihan *term* penting cakupannya luas dan tidak hanya sebatas frekuensi kemunculannya melainkan juga konten informatif yang dikandung dalam suatu kalimat. Metode NeFGIS yang diusulkan tidak memperlakukan setiap kata yang sama, karena kata yang sama tidak selalu memiliki fungsi yang sama, seperti yang ditunjukkan pada Tabel 5. Label POS akan mempengaruhi bobot setiap *term* dalam kalimat.

Di sisi lain, metode yang diusulkan masih memiliki kelemahan pada bagian penyusunan ringkasan, di mana kalimat representatif yang diambil masih terdapat redundansi dengan kalimat representatif lain, dan ringkasan yang dihasilkan tidak mempertimbangkan urutan atau kesesuaian kalimat untuk kemudahan pembacaan.

5. Kesimpulan

Metode usulan NeFGIS menghasilkan ringkasan dengan evaluasi ROUGE yang lebih tinggi dibandingkan dengan metode NeFTIS. Berdasarkan evaluasi ROUGE-1, ROUGE-2, ROUGE-L, dan ROUGE-SU4 metode pembobotan NeFGIS mampu memberikan hasil yang lebih baik dibandingkan metode NeFTIS dengan rata-rata peningkatan nilai untuk *recall*, *precision*, dan *f-score* secara berturut-turut adalah 0,03570 (↑19,67%), 0,09722 (↑38,13%), 0,06072 (↑27,26%). Peningkatan nilai pada metode NeFGIS dikarenakan metode NeFGIS menggunakan pendekatan informasi gramatikal yang ditujukan untuk mengetahui adanya kata informatif pada suatu kalimat yang dapat mengindikasikan bahwa kalimat tersebut penting. Sehingga pemilihan kalimat penting dengan skema tersebut mampu memilih kalimat representatif sebagai dasar penyusunan ringkasan dan mampu meningkatkan kualitas hasil ringkasan.

Kedepannya kami mengusulkan untuk meningkatkan metode yang kami usulkan untuk menangani redundansi hasil ringkasan dan pertimbangan akan kesesuaian kalimat untuk kemudahan pembacaan ringkasan berita. Sehingga perlu dilakukan penelitian untuk melihat pengaruhnya terhadap metode NeFGIS.

6. Referensi

- Aditya, C. S., Fatichah, C., & Purwitasari, D. (2016). Ekstraksi trending issue dengan pendekatan distribusi kata pada pembobotan term untuk peringkasan multi-dokumen berita. *JUTI (Jurnal Ilmiah Teknologi Informasi)*, 14(2), 180-189.
- Arifin, A. Z., Abdullah, M. Z., Rosyadi, A. W., Ulumi, D. I., Wahib, A., & Sholikah, R. W. (2018). Sentence Extraction Based on Sentence Distribution and Part of Speech Tagging for Multi-document Summarization. *TELKOMNIKA*, 16(2), 843-851.

- Fachrurrozi, M., Yusliani, N., & Yoanita, R. U. (2013). Frequent Term based Text Summarization for Bahasa Indonesia. *International Conference on Innovations in Engineering and Technology (ICIET2013)*. Bangkok: IIENG.
- Ferreira, R., Cabral, L. d., Lins, R. D., Silva, G. P., Freitas, F., Cavalcanti, G. D., . . . Favaro, L. (2013). Assessing sentence scoring techniques for extractive text summarization. *Expert Systems with Applications*, 40(14), 5755-5764.
- Hayatin, N., Fatichah, C., & Purwitasari, D. (2014). *Penentuan trending issue data twitter menggunakan cluster importance untuk peringkasan multi dokumen berita*. Teknik Informatika. Surabaya: Institut Teknologi Sepuluh Nopember.
- He, T., Li, F., Shao, W., Chen, J., & Ma, L. (2008). A new feature-fusion sentence selecting strategy for query-focused multi-document summarization. *Proceeding of International Conference Advance Language Processing and Web Information Technology* (hal. 81-86). Dalian Liaoning: IEEE.
- Lin, C.-Y. (2004). Rouge: A package for automatic evaluation of summaries. In *Text Summarization Branches Out: Proceedings of the ACL-04 Workshop* (hal. 74-81). Barcelona: Association for Computational Linguistics.
- Lioma, C., & Blanco, R. (2017). Part of Speech Based Term Weighting for Information Retrieval.
- Mandar, G., & Gunawan, G. (2017). Peringkasan dokumen berita Bahasa Indonesia menggunakan metode Cross Latent Semantic Analysis. *Register: Jurnal Ilmiah Teknologi Sistem Informasi*, 3(2), 94-104.
- Meena, Y. K., & Gopalani, D. (2014). Analysis of Sentence Scoring Methods for Extractive Automatic Text Summarization. *Proceedings of the 2014 International Conference on Information and Communication Technology for Competitive Strategies*. Udaipur: ACM.
- Meena, Y. K., & Gopalani, D. (2015). Evolutionary algorithms for extractive automatic text summarization. *Procedia Computer Science*, 48, 244-249.
- Mei, J. -P., & Chen, L. (2012). SumCR: A new subtopic-based extractive approach for text summarization. *Knowledge and information systems*, 31(3), 527-545.
- Pisceldo, F., Adriani, M., & Manurung, R. (2009). Probabilistic Part Of Speech Tagging for Bahasa Indonesia. *Third International MALINDO Workshop*. Cyberjaya.
- Radev, D. R., Hovy, E., & McKeown, K. (2002). Introduction to the special issue on summarization. *Computational linguistics*, 28(4), 399-408.
- Rashel, F., Luthfi, A., Dinakaramani, A., & Manurung, R. (2014). Building an Indonesian rule-based part-of-speech tagger. *Asian Language Processing (IALP)* (hal. 70-73). Kuching: IEEE.
- Verdianto, S., Arifin, A. Z., & Purwitasari, D. (2016). Strategi pemilihan kalimat pada peringkasan multi dokumen. *NJCA: Nusantara Journal of Computer Applications*, 1(2).
- Wahib, A., Arifin, A. Z., & Purwitasari, D. (2016). Improving Multi-Document Summary Method Based on Sentence Distribution. *TELKOMNIKA*, 14(1), 286-293.
- Wan, X., Yang, J., & Xiao, J. (2007). Towards an Iterative Reinforcement Approach for Simultaneous Document Summarization and Keyword Extraction. *Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics* (hal. 552-559). Prague: Association for Computational Linguistics.