REPOSITOR, Vol. 2, No. 1, Januari 2020, Pp. 107-112
ISSN : 2714-7975
E-ISSN : 2716-1382

Peringkasan Multi Dokumen Berita Dengan Pemilihan Kalimat Utama Berbasis Algoritma Cluster Importance Dengan Mempertimbangkan Posisi Kalimat

Syadza Anggraini*[1], Nur Hayatin[2], Gita Indah Marthasari[3]
[1,2,3]Teknik Informatika/Universitas Muhammadiyah Malang
sasaanggraini.sa@gmail.com*[1], noorhayatin@umm.co.id[2] , gita.voyager@gmail.com[3]

Abstrak

Peringkasan teks merupakan salah satu cara untuk mengurangi suatu dimensi dokumen yang besar untuk mendapatkan informasi penting dari dokumen tersebut. Berita adalah salah satu informasi yang biasanya dalam satu topik memiliki beberapa sub topik. Untuk dapat mengambil informasi penting dari satu topik secara cepat, peringkasan multi dokumen berita dapat menjadi solusi. Namun, peringkasan multi dokumen dapat menimbulkan redundansi. Oleh sebab itu, penelitian ini menerapkan algoritma cluster importance dengan mempertimbangkan posisi kalimat untuk mengatasi redundansi tersebut. Penelitian ini menggunakan 30 topik berita berbahasa Indonesia, dimana tiap topiknya terdiri dari 5 sub topik berita. Dari 30 topik berita yang diuji menggunakan Rouge-1, dimana terdapat 2 topik berita yang memiliki nilai Rouge-1 berbeda antara yang menggunakan algoritma cluster importance ditambah posisi kalimat dengan yang hanya menggunakan algoritma cluster. Namun dari 2 topik berita tersebut, nilai Rouge-1 yang menggunakan cluster importance ditambah posisi kalimat memiliki nilai yang lebih besar daripada yang hanya menggunakan cluster importance. Penggunaan posisi kalimat memiliki pengaruh terhadap urutan bobot kalimat pada setiap topiknya, namun hanya 2 topik berita yang berpengaruh terhadap hasil ringkasan.

Kata kunci: *Peringkasan Teks, Berita, Redundansi, Cluster importance, Posisi Kalimat*

Abstract

Text summarization is one of way to reduce large document dimension to get an important point of information. News is one of information which usually has some sub topics from one topic. In order to get the main information from one topic as fast as possible, multi document summarization is the solution. But sometimes it can create redundancy. So in this study, we applied cluster importance algorithm by considering sentence position to overcome the redundancy. This study used 30 topics of Indonesian news, where each topic consists 5 news sub topics. From 30 news topics where it has tested using Rouge-1, there are 2 news topics that have a Rouge-1 score differ between which used cluster importance algorithm by considering sentence position and which only used cluster importance. But, those 2 news topics which used cluster importance by considering sentence position have a greater score of Rouge-1 than which only used cluster importance. The use of sentence position had an effect on the order of sentence weights on each topic, but there was only 2 news topics that affect the outcome of the summary.

Keywords*: Text Summarization, News, Redundancy, Cluster Importance, Sentence Position*

1. Pendahuluan

Informasi adalah pemberitahuan mengenai berita ataupun kabar yang biasanya termuat dalam bentuk artikel, berita, makalah ilmiah dan buku. Namun,informasi yang disajikan biasanya bagi sebagian orang cukup sulit untuk dimengerti dikarenakan banyaknya informasi yang termuat atau biasa disebut *information overload* [1]. Dengan demikian, dibutuhkan peringkasan dokumen sebagai solusi untuk hal tersebut.

Peringkasan merupakan suatu proses dalam mereduksi ukuran dokumen yang asli menjadi ukuran tidak lebih dari setengah dari ukuran dokumen aslinya [2]. Dalam peringkasan dokumen ada dua jenis yaitu ekstraksi dan abstraksi. Dimana peringkasan secara ekstraksi merupakan ringkasan yang dihasilkan dengan mengambil kalimat asli dari suatu dokumen. Sedangkan peringkasan secara abstraksi merupakan peringkasan yang dihasilkan dari informasi suatu dokumen yang diubah kalimatnya namun masih tetap memiliki makna yang sama [3].

Penelitian dalam hal peringkasan dokumen kebanyakan dengan cara ekstraksi. Seperti halnya penelitian [4] yang melakukan penelitian mengenai pembangunan perangkat lunak terhadap multi dokumen berita menggunakan metode TF-IDF sebagai sentence scoring. Berdasarkan penelitian tersebut bahwa agar bisa mengoptimalkan pemilihan kalimat sebagai penyusun ringkasan, dibutuhkan suatu metode atau algoritma dalam mencari kemiripan antar kalimat sehingga tidak terjadi redundansi.

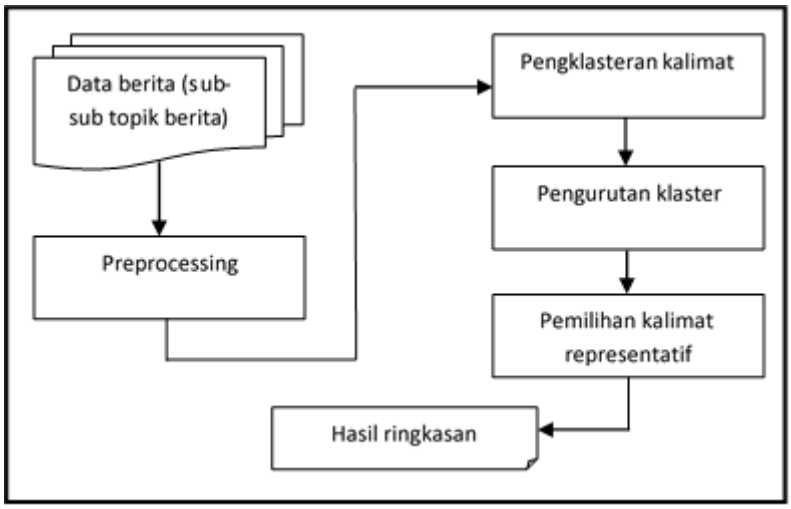
Peringkasan terhadap banyak dokumen atau multi dokumen dimungkinkan terjadinya redundansi yaitu munculnya kalimat yang mirip dalam suatu hasil ringkasan. Jika terjadi redundansi pada suatu hasil ringkasan maka menyebabkan banyaknya informasi penting yang terkandung didalamnya padahal memiliki makna yang sama. Oleh sebab itu, penelitian ini mengusulkan solusi untuk mengatasi redundansi yaitu dengan melakukan peringkasan multi dokumen berita menggunakan algoritma cluster importance. Berdasarkan penelitian [5] dimana peneltian tersebut menjelaskan bahwa dalam melakukan peringkasan tahap pertama yang dilakukan adalah pengklasteran kalimat terhadap sub-sub topik berita. Selanjutnya, tahap kedua adalah mengurutkan hasil klaster yang sudah terbentuk dengan membobot tiap klasternya. Kemudian, tahap terakhir yaitu memilih kalimat representatif dari tiap klaster dengan membobot setiap kalimat dan memilih satu kalimat dengan bobot tertinggi dari tiap klasternya sebagai bahan penyusun ringkasan.

Seperti yang sudah dijelaskan sebelumnya yaitu dalam mengatasi redundansi menggunakan algoritma cluster importance, penelitian ini juga menjadikan posisi kalimat sebagai bahan pertimbangan didalamnya. Berdasarkan penelitian [6] menjelaskan bahwa dalam suatu dokumen khususnya berita, posisi kalimat menjadi fitur penting, dimana dalam suatu berita kalimat yang terletak diawal mempunyai skor yang lebih besar daripada kalimat yang terletak diposisi akhir.

Dengan demikian penelitian ini menggunakan algoritma cluster importance sebagai solusi mengatasi redundansi ditambah dengan mempertimbangkan posisi kalimat dari berita itu sendiri.

2. Metode Penelitian

Peringkasan multi dokumen berita dalam penelitian digambarkan kedalam bagan sistem meliputi tahapan atau proses dalam membentuk hasil ringkasan. Gambar 1 berikut merupakan bagan sistem tersebut.



Gambar 1. Bagan Sistem Peringkasan Multi Dokumen Berita

Gambar 1 diatas merupakan bagan dari sistem peringkasan multi dokumen pada penelitian ini. Dari gambar tersebut terdapat beberapa tahapan utama dalam meringkas multi dokumen berita.

2.1 Data Berita

Peringkasan multi dokumen pada penelitian ini menggunakan data berita yang diambil secara *online* sebanyak 30 topik berita. 11 topik diantaranya dari penelitian [6] dan ditambah 19 topik berita lainnya. Dalam satu topik berita terdapat 5 berita yang berbeda-beda (5 sub topik

berita). Data berita yang digunakan dalam penelitian ini hanya megambil konten atau isinya saja tanpa melibatkan judul beritanya.

2.2 Preprocessing

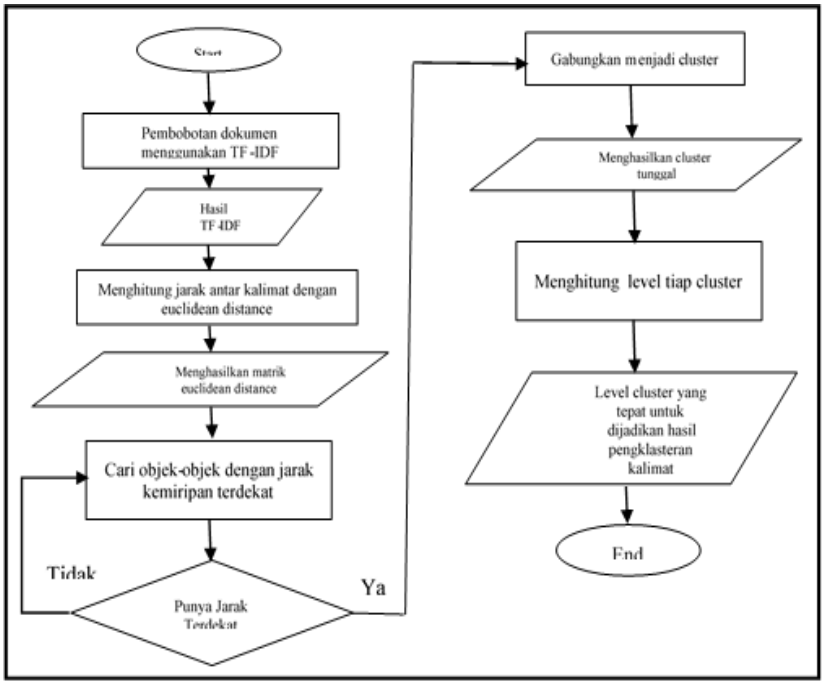
Pada tahap *preprocessing* data berita akan dilakukan beberapa proses seperti: pemecahan berita menjadi kalimat indicidu, *case folding*, *tokenizing*, dan *stopwords removal*. Beberapa proses tersebut dilakukan sebelum memasuki tahap utama yaitu algoritma *cluster importance*. Pemecahan berita menjadi kalimat individu merupakan proses pemecahan berita yang biasanya tertulis dalam bentuk paragraf dipecah menjadi kalimat individu pada tiap sub topiknya. Setelah dipecah menjadi kalimat individu, dilakukan *case folding* guna menyeragamkan semua huruf menjadi *lowercase* (huruf kecil) serta penghapusan *delimiter* atau tanda baca seperti: (.), (,), (!), (:), (;), dan sebagainya. Kemudian dilanjutkan dengan melakukan proses *tokenizing* yaitu memecah tiap kalimat tersebut menjadi kata, dimana proses tersebut dilakukan sebelum proses *stopwords removal*, agar mempermudah dalam menghapus kata-kata yang tidak penting. Daftar *stopwords* yang digunakan berdasarkan penelitian [7].

2.3 Algoritma Cluster Importance

Algoritma *cluster importance* meliputi pengklasteran kalimat, pengurutan klaster, dan pemilihan kalimat representatif. Secara lebih lanjut akan dijelaskan pada sub-sub bab dibawah ini.

2.3.1 Pengklasteran Kalimat

Pengklasteran kalimat merupakan tahapan pertama dari algoritma *cluster importance*. Secara lebih jelas mengenai pengklasteran kalimat tersebut akan digambarkan dalam *flowchart* dibawah ini.



Gambar 2 Flowchart Pengklasteran Kalimat Menggunakan Single Linkage

Berdasarkan Gambar 2 diatas, pengklasteran kalimat menggunakan metode *single linkage*. Namun sebelum masuk kedalam metode *single linkage*, terlebih dahulu dilakukan pembobotan menggunakan *TF-IDF*. *Term Frequency Inverse Document Frequency* (*TF-IDF*) merupakan pembobotan dengan menghitung *Term Frequency* (*TF*) yaitu frekuensi atau banyaknya kata yang muncul dalam suatu dokumen, serta dengan menghitung *Inverse Document Frequency* (*IDF*) yaitu untuk menghitung penting atau tidaknya suatu kata dalam dalam suatu dokumen yang dilihat dari sejumlah dokumen secara keseluruhan [8]. Berikut Persamaan 1 dan Persamaan 2 yang menunjukkan pembobotan *TF-IDF*.

$$IDF(t) = \log \frac{N}{df(t)} \quad (1)$$

$$TF.IDF = TF(dt) * IDF\ (t) \quad (2)$$

Setelah dilakukan pembobotan menggunakan *TF-IDF*, selanjutnya mengukur kedekatan antar kalimat menggunakan *euclidean distance*. Jarak antar kalimat diukur guna mengetahui seberapa dekat jarak antara kalimat yang satu dengan yang lainnya. Berikut Persamaan 3 yang menunjukkan *euclidean distance*.

$$Dis(x, y) = \sqrt{\sum(x_i - y_i)^2} \quad (3)$$

Proses selanjutnya setelah pembobotan *TF-IDF* dan menghitung *euclidean distance* adalah mengklaster kalimat dengan *single linkage*. *Single linkage* merupakan teknik *clustering* yang dilakukan dengan cara penggabungan secara berurutan yang diawali dengan mencari dua buah objek dengan jarak kedekatan minimum. Jika memiliki jarak kedekatan minimum maka akan bergabung menjadi satu *cluster*. Selanjutnya, mencari objek lain yang memiliki kedekatan minimum dengan *cluster* yang sudah terbentuk. Jika memiliki jarak kedekatan minimum maka akan bergabung dengan *cluster* yang sudah terbentuk tersebut, atau membentuk *cluster* baru dengan objek lainnya. Hal tersebut dilakukan hingga membentuk *cluster* tunggal. Setelah *cluster* tunggal terbentuk, selanjutnya memilih level *cluster* yang tepat untuk menentukan *cluster* sebagai hasil dari pengklasteran kalimat. Dalam menentukan level *cluster* tersebut dengan mengukur *dissimilarity* antar *cluster* [9]. Berikut Persamaan 4 untuk menghitung *dissimilarity*.

$$dissimilarity(cluster1, cluster2) = \frac{\sum Euclidean\ (d1,d2)}{size\ cluster1 \times size\ cluster2} \quad (4)$$

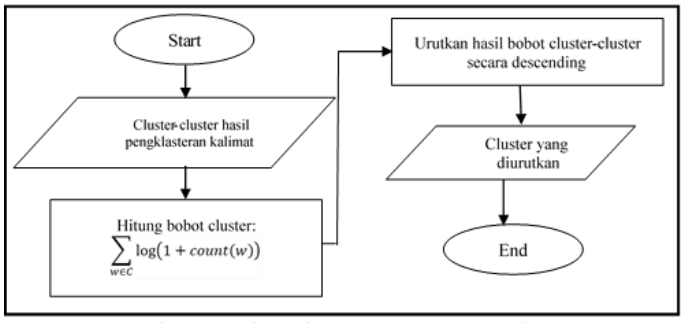
Sedangkan untuk nilai tengah dari *cluster* [9] ditunjukkan oleh Persamaan 5 berikut ini.

$$Sim(X) = \sum_{d \epsilon X} Euclidean(d, c) \quad (5)$$

Jadi, untuk menentukan level *cluster* atau penggabungan *cluster* sebagai hasil dari pengklasteran kalimat adalah dengan menggunakan Persamaan 4, dan level cluster yang dipilih adalah dengan nilai *dissimilarity* paling besar.

2.3.2 Pengurutan Cluster

Pengurutan *cluster* merupakan tahapan kedua dalam algoritma *cluster importance*. Setelah selesai dilakukan pengklasteran kalimat, dimana sejumlah *cluster* telah terbentuk maka selanjutnya dilakukan pengurutan terhadap *cluster-cluster* tersebut. Tujuan dari pengurutan *cluster* adalah untuk menunjukkan *information richness* dari tiap-tiap *cluster* yang terbentuk. Terkadang, tidak semua cluster dengan kalimat terbanyak menunjukkan *information richness*. Bisa saja cluster tersebut terdiri dari kalimat-kalimat yang kurang penting. Oleh sebab itu dibutuhkan pengurutan *cluster* [5]. Secara lebih jelas akan digambarkan melalui *flowchart* dibawah ini.

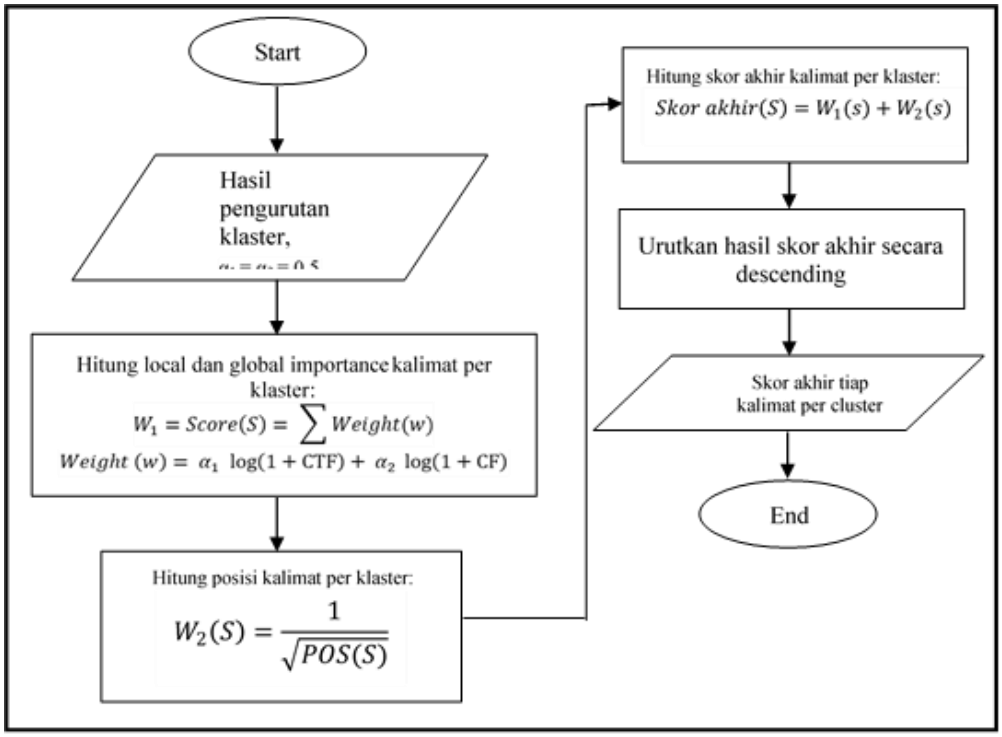


Gambar 3. Flowchart Pengurutan Cluster

Berdasarkan Gambar 3 diatas, setelah sejumlah *cluster* terbentuk dari hasil pengklasteran kalimat, tiap *cluster* dihitung bobotnya dengan menghitung bobot kata-kata didalam tiap *cluster* tersebut. Setelah didapatkan bobot dari setiap *cluster*, selanjutnya dilakukan pengurutan yaitu dari yang bobot terbesar hingga terkecil. Pengurutan *cluster* dilakukan untuk mengetahui *cluster* mana yang terlebih dahulu dijadikan sebagai bahan peringkasan.

2.3.3 Pemilihan Kalimat Representatif

Tahapan ketiga setelah pengurutan *cluster* yaitu pemilihan kalimat representatif. Pemilihan kalimat representatif merupakan tahap dalam memilih satu kalimat dari tiap *cluster* yang terbentuk dengan cara membobot tiap kalimatnya. Secara lebih lanjut akan dijelaskan melalui gambar *flowchart* dibawah ini.



Gambar 4. Flowchart Pemilihan Kalimat Representatif

Berdasarkan Gambar 4 diatas, pada tahap pemilihan kalimat representatif terdapat dua perhitungan pembobotan yaitu *W1* dan *W2*. *W1* merupakan perhitungan bobot pertama kalimat yaitu dengan menghitung *local* dan *global importance*. *Local importance* merupakan kata yang menunjukkan seberapa banyak kata tersebut dalam pembentukan *cluster* atau jumlah suatu kata tersebut dalam suatu *cluster*, perhitungannya menggunakan log(1+CTF), CTF (*Cluster Term Frequency*). Sedangkan *Global importance* merupakan jumlah *cluster* yang terdapat suatu kata, perhitungannya menggunakan log(1+CF), CF (*Cluster Frequency*) [5].

Selain menghitung bobot pertama yaitu *W1*, selanjutnya menghitung bobot kedua yaitu *W2*. *W2* merupakan perhitungan posisi kalimat. Dari gambar 4 diatas POS(S) menunjukkan *sentence index* dari kalimat yang muncul dalam suatu dokumen. Pertimbangan posisi kalimat dalam peringkasan multi dokumen berita karena berdasarkan penelitian [10] dimana kalimat yang letaknya diawal dokumen memiliki skor lebih besar daripada diakhir dokumen. Serta berdasarkan ilmu jurnalistik, teknik penulisan dalam suatu berita *online* yang digunakan adalah “piramida terbalik”, kalimat penting terletak diawal berita dan kalimat kurang penting berada diakhir berita [11].

Untuk mendapatkan bobot sebuah kalimat, yang dilakukan selanjutnya adalah dengan menjumlahkan bobot pertama dan kedua, dengan kata lain menghitung skor akhir kalimat. Setelah skor akhir kalimat didapatkan, skor tersebut diurutkan dari yang terbesar ke yang terkecil. Skor akhir kalimat paling besar dari tiap *cluster* dipilih sebagai perwakilan *cluster* tersebut sebagai penyusun ringkasan.

2.4 Perancangan Pengujian

Pada penelitian ini pengujian dilakukan terhadap hasil ringkasan yang dihasilkan. Berikut gambar dibawah ini menunjukkan rancangan pengujian.

Gambar 5. Skenario Pengujian Hasil Ringkasan

Berdasarkan Gambar 5 diatas, gambar tersebut menunjukkan skenario pengujian terhadap hasil ringkasan multi dokumen berita pada penelitian ini, dimana pengujian dilakukan antara hasil ringkasan yang hanya menggunakan *cluster importance* saja dengan yang hasil ringkasan yang menggunakan *cluster importance* ditambah posisi kalimat. Pengujian dilakukan menggunakan *Recall Oriented Understudy for Gisting Evaluation* (ROUGE). ROUGE-N digunakan untuk menghitung *recall* N-gram antara ringkasan referensi dan ringkasan sistem [12]. Nilai N yang digunakan adalah 1. Berikut Persamaan 6 yang menunjukkan ROUGE-N.

$$ROUGE - N = \frac{\sum_{S \epsilon Summ_{ref}} \sum_{N-gram \epsilon S} Count_{match}\,(N-gram)}{\sum_{S \epsilon Summ_{ref}} \sum_{N-gram \epsilon S} Count(N-gram)} \quad (6)$$

Oleh karena penelitian menggunakan multi dokumen maka perhitungan akhir ROUGE-N yang digunakan sebagai berikut [12].

$$ROUGE - N_{multi} = argmax_i ROUGE - N(r_i, S) \quad (7)$$

Persamaan 7 diatas untuk menentukan nilai akhir dari ROUGE-N, dimana nilai akhir diambil dari nilai ROUGE-N yang paling besar.

3. Hasil Penelitian dan Pembahasan

Pada hasil penelitian dan pembahasan ini menjelaskan hasil penelitian berupa hasil pengujian dari ringkasan berdasarkan rancangan pengujian yang telah dijelaskan sebelumnya. Berikut akan dijelaskan lebih lanjut mengenai hasil pengujian tersebut:

Pada pengujian ringkasan menjelaskan hasil pengujian dari rancangan pengujian yang telah dijelaskan sebelumnya. Tabel 1 dibawah ini menunjukkan hasil pengujian ringkasan multi dokumen berita pada penelitian ini.

Tabel 1. Nilai ROUGE-1 Pengujian Ringkasan

		Nilai Rata-Rata Max Rouge-1	
No	Topik	Cluster Importance + Posisi Kalimat	Cluster Importance
1	Air Asia	0,51705	0,51705
2	Angkot Tabrak Grab	0,24103	0,24103
3	Banjarnegara	0,56796	0,56796
4	BBM	0,37888	0,37888
5	BPJS	0,47429	0,42286
6	Countdown Asian Games	0,53498	0,53498
7	Demo Angkot Malang	0,46018	0,46018
8	Dokter Letty	0,43182	0,43182
9	Dolly	0,48831	0,48831
10	Ebola	0,76812	0,76812
11	Gempa Korea Selatan	0,66447	0,66447
12	Gunung Agung	0,49032	0,49032
13	Habib Rizieq	0,38462	0,38462
14	Hari Raya Nyepi	0,41341	0,41341
15	Konser boyband SHINee	0,44872	0,44872

16	Kunjungan Obama	0,50691	0,50691
17	Kurikulum 2013	0,47511	0,47511
18	Ledakan Gudang Mercon	0,54040	0,54040
19	Mahasiswi UI	0,34574	0,34574
20	Palestina	0,41905	0,41905
21	Penasehat KPK	0,35673	0,35673
22	Penutupan Hotel Alexis	0,40462	0,40462
23	Penyanderaan Angkot	0,24725	0,24725
24	Penyiraman Novel Baswedan	0,50459	0,50459
25	Pemilihan Presiden	0,36290	0,36290
26	Pria Pencuri Amplifier	0,64737	0,64737
27	Saksi Kunci E-KTP	0,61290	0,61290
28	Sinabung	0,24171	0,15640
29	Tora Sudiro	0,53333	0,53333
30	U19	0,54113	0,54113

Keterangan :
: topik dengan nilai ROUGE-1 berbeda

Pengujian ringkasan dilakukan terhadap 30 topik berita menggunakan dua sumber *groundtruth*. Berdasarkan Tabel 1 diatas dari 30 topik berita yang diuji terdapat 2 topik yang memiliki nilai *ROUGE-1* berbeda antara yang menggunakan *cluster importance* dengan yang menggunakan *cluster importance* ditambah posisi kalimat. Topik-topik tersebut adalah "BPJS" dan "Sinabung". Namun, nilai *ROUGE-1* dari *cluster importance*+posisi kalimat memiliki nilai yang lebih besar daripada yang menggunakan *cluster importance* saja. Sedangkan 28 topik berita lainnya memiliki nilai *ROUGE-1* yang sama.

Adanya perbedaan nilai *ROUGE-1* tersebut disebabkan karena pengaruh bobot kalimat pada tahap pemilihan kalimat representatif, yang secara otomatis mempengaruhi urutan kalimat dengan bobot tertinggi yang dijadikan sebagai penyusun ringkasan. Berikut contoh perubahan urutan kalimat dari topik "BPJS" karena adanya pengaruh bobot posisi kalimat.

NO.	CLUSTER	KALIMAT	W1	W2	BOBOT KALIMAT
		4.1 kepala bagian kepesertaan badan penyelenggara jaminan sosial bpjs sumatera bagian utara manna lubis mengatakan 2015 bpjs hadir gunungsitoli tapaktuan aceh	9.261	1	10.261
		1.5 melakukan aksi demo kantor bpjs pelayanan bpjs buruk diskriminasi kata ketua federasi serikat metal indonesia kota depok wido praktikno senin 1122014	9.28	0.447	9.727
		5.8 sesuai peraturan perundangan untuk pekerja penerima upah ppu bumn bumd badan usaha skala besar mapun wajib mendaftarkan pegawainya lambat 1 januari 2015	9.18	0.354	9.534
		2.1 ketua asosiasi pengusaha indonesia apindo kota medan rusmin lawin mengatakan program asuransi badan penyelenggara jaminan kesehatan bpjs bagus	8.437	1	9.437
		3.1 pelaksanaan program bpjs badan penyelenggara jaminan sosial kesehatan diajukan kamar dagang industri kadin tahun 2019 mendatang ditunda	7.373	1	8.373
		4.4 kata bpjs cabang tapaktuan kedepannya melayani masyarakat kawasan aceh bagian barat kawasan berdekatan unit usaha disana	7.201	0.5	7.701
		5.6 bpjs kesehatan menghimbau masyarakat untuk mendaftarkan menjadi peserta bpjs kesehatan selagi sehat	7.006	0.408	7.414
		2.6 rusmin menjelaskan hari sopirnya mengurus bpjs datang pukul 0600 wib antrean kantor bpjs mencapai 100 orang	6.904	0.408	7.312

Gambar 5. Pemilihan Kalimat Representatif Suatu Cluster + Posisi Kalimat

NO.	CLUSTER	KALIMAT	W1	BOBOT KALIMAT
		1.5 melakukan aksi demo kantor bpjs pelayanan bpjs buruk diskriminasi kata ketua federasi serikat metal indonesia kota depok wido praktikno senin 1122014	9.28	9.28
		4.1 kepala bagian kepesertaan badan penyelenggara jaminan sosial bpjs sumatera bagian utara manna lubis mengatakan 2015 bpjs hadir gunungsitoli tapaktuan aceh	9.261	9.261
		5.8 sesuai peraturan perundangan untuk pekerja penerima upah ppu bumn bumd badan usaha skala besar mapun wajib mendaftarkan pegawainya lambat 1 januari 2015	9.18	9.18
		2.1 ketua asosiasi pengusaha indonesia apindo kota medan rusmin lawin mengatakan program asuransi badan penyelenggara jaminan kesehatan bpjs bagus	8.437	8.437
		3.1 pelaksanaan program bpjs badan penyelenggara jaminan sosial kesehatan diajukan kamar dagang industri kadin tahun 2019 mendatang ditunda	7.373	7.373
		4.4 kata bpjs cabang tapaktuan kedepannya melayani masyarakat kawasan aceh bagian barat kawasan berdekatan unit usaha disana	7.201	7.201
		5.6 bpjs kesehatan menghimbau masyarakat untuk mendaftarkan menjadi peserta bpjs kesehatan selagi sehat	7.006	7.006
		2.6 rusmin menjelaskan hari sopirnya mengurus bpjs datang pukul 0600 wib antrean kantor bpjs mencapai 100 orang	6.904	6.904

Gambar 6. Pemilihan Kalimat Representatif Suatu Cluster Tanpa Posisi Kalimat

Dari Gambar 5 dan Gambar 6 diatas dikarenakan adanya perubahan urutan kalimat pada tahap pemilihan kalimat representatif maka secara otomatis mempengaruhi hasil ringkasan yang juga berdampak pada nilai rouge-1 yang dihasilkan.

4. Kesimpulan dan Saran

Dari penelitian yang telah dilakukan dapat diambil kesimpulan bahwa terdapat 2 topik yaitu topik "BPJS" dan "Sinabung" yang memiliki nilai *ROUGE-1* berbeda antara yang menggunakan *cluster importance* dan menggunakan *cluster importance*+posisi kalimat dari keseluruhan data uji yaitu 30 topik berita. Dari 2 topik tersebut nilai *ROUGE-1* yang menggunakan *cluster importance*+posisi kalimat memiliki nilai lebih tinggi daripada menggunakan *cluster importance*.

Penggunaan posisi kalimat terhadap algoritma *cluster importance* sebagai pertimbangan dalam meringkas dokumen berita, tidak memberikan hasil signifikan berbeda dari yang menggunakan algoritma pada penelitian ini. Hal tersebut ditunjukkan dari 30 topik yang diuji, 28 topik diantaranya memiliki nilai *ROUGE*-1 yang sama antara yang menggunakan *cluster importance*+posisi kalimat dengan yang menggunakan *cluster importance*. Namun, penggunaan posisi kalimat tetap memperlihatkan perbedaan hasil urutan kalimat pada tahap pemilihan kalimat representatif, tetapi tidak berdampak pada hasil ringkasan akhir.

Berdasarkan penjelasan diatas penggunaan posisi kalimat untuk meringkas, dipengaruhi oleh data berita itu sendiri dimana tidak ada kalimat yang sama persis antara yang satu dan lainnya. Sehingga menyebabkan munculnya pengaruh terhadap bobot dari tiap kalimat.

Adapun saran yang dapat diberikan yaitu dengan melakukan penelitian selain menggunakan data berupa dokumen berita. Dengan demikian dapat diketahui penting atau tidaknya posisi kalimat sebagai bahan pertimbangan dalam meringkas suatu dokumen.

Daftar Notasi

N	: jumlah dokumen dalam koleksi
df(t)	: jumlah dokumen berisi term t
x	: dokumen x
y	: dokumen y
x_i	: dokumen x ke i
y_i	: dokumen y ke i
d	: dokumen pada cluster X
c	: nilai tengah dari cluster X
count(w)	: jumlah kata dalam suatu cluster
score(s)	: skor kalimat s
weight(w)	: bobot kata w
α_1 ,α_2	: nilai yang ditetapkan yaitu 0,5
POS(S)	: sentence index (posisi kalimat)
$Count_{match}$	: jumlah maksimum N-gram yang muncul pada ringkasan kandidat dan ringkasan referensi
r_i	: ringkasan kandidat (sistem)
S	: ringkasan referensi

Referensi

[1] W. E. Waliprana and M. L. Khodra, "Update Summarization Untuk Kumpulan Dokumen Berbahasa Indonesia," *J. Cybermatika*, pp. 6–10, 2009.

[2] N. Munot and S. S. Govilkar, "Comparative Study of Text Summarization Methods," *Int. J. Comput. Appl.*, vol. 102, no. 12, pp. 975–8887, 2014.

[3] A. Agrawal and U. Gupta, "Extraction Based Approach for Text Summarization Using K-means Clustering," *Int. J. Sci. Res. Publ.*, vol. 4, no. 11, pp. 9–12, 2014.

[4] F. H. Evan, Y. S. Purnomo, and Pranowo, "Pembangunan Perangkat Lunak Peringkas Dokumen Dari Banyak Sumber Menggunakan Sentence Scoring Dengan Metode Tf-Idf," *Semin. Nas. Apl. Teknol. Inf.*, pp. 17–22, 2014.

[5] K. Sarkar, "Sentence Clustering-Based Summarization of Multiple Text Documents," *Tech. – Int. J. Comput. Sci. Commun. Technol.*, vol. 2, no. 1, pp. 974–3375, 2009.

[6] N. Hayatin, C. Fatichah, and D. Purwitasari, "Pembobotan Kalimat Berdasarkan Fitur Berita dan Trending Issue Untuk Peringkasan Multi Dokumen Berita," *J. Ilm. Teknol. Inf.*, vol. 13, pp. 38–44, 2015.

[7] F. Z. Tala, “A Study of Stemming Effects on Information Retrieval in Bahasa Indonesia,” 2003.

[8] E. Purwanti, “Klasifikasi Dokumen Temu Kembali Informasi dengan K-Nearest Neghbour Information Retrieval Document Classified with K-Nearest Neighbor,” *Rec. Libr. J.*, vol. 1, pp. 187–196, 2015.

[9] Annisa, Y. Munarko, and Y. Azhar, “Peringkasan Tweet Berdasarkan Trending Topic Twitter Dengan Pembobotan TF-IDF dan,” *J. Kinet.*, vol. 1, no. 1, pp. 9–16, 2016.

[10] J. P. Mei and L. Chen, “SumCR: A New Subtopic-Based Extractive Approach for Text Summarization,” *Knowl. Inf. Syst.*, vol. 31, no. 3, pp. 527–545, 2012.

[11] S. Verdianto, A. Z. Arifin, and D. Purwitasari, “Strategi Pemilihan Kalimat Pada Peringkasan Multi Dokumen,” *J. Tek. ITS*, vol. 2, no. 7, pp. 1–5, 2016.

[12] C. Y. Lin, “Rouge: A Package for Automatic Evaluation of Summaries,” *Proc. Work. Text Summ. Brances Out Assoc. Comput. Linguist. Barcelona*, no. 1, pp. 25–26, 2004.