Implementasi Metode *Maximum Marginal Relevance (MMR)* dan Algoritma *Steiner Tree* Untuk Menentukan *Storyline* Dokumen Berita

Aji Teja Hartanto[1]

[1] Teknik Informatika – Universitas Komputer Indonesia
Jl. Dipatiukur 112-114 Bandung
Email : itera194@gmail.com[1]

ABSTRAK

Dalam proses pencarian kronologis suatu berita peristiwa secara online dengan menggunakan mesin pencari, pengguna harus mengakses berbagai situs yang memiliki relevansi peristiwa tersebut. Hal ini dikarenakan mesin pencari tidak memberikan hasil pencarian secara terstruktur. Proses pencarian membutuhkan kesesuaian hasil dengan *query* yang dimasukan sebagai kriteria pencarian. Relevansi berita tekstual dapat diperoleh dengan menggunakan cosine similiarity dari ringkasan-ringaksan berita dengan melakukan implementasi terhadap metode Maximum Marginal Relevance (MMR) yang ditentukan berdasarkan kemiripan terhadap query. Dalam konteks yang sama, Penelusuran lebih dari 90 sampel dokumen berita diterapkan algoritma *Steiner Tree* dalam menentukan jalur terbaik (lintasan) dalam kumpulan dokumen berita (verteks) yang terhubung sebagai sebuah bobot kemiripan (sisi) sekaligus sisi berarah terhadap dokumen tertentu (sisi berarah).

Berdasarkan metode pengujian *usability testing* yang ditujukan terhadap 6 responden sebanyak 66.7% responden menilai bahwa aplikasi ini mengerjakan tugas-tugas dengan lebih cepat dan 75% pengguna mampu mengatasi problem yang ada dengan proses pencarian berita yang menghemat waktu. Dari pengujian performansi terhadap algoritma yang diterapkan, diperoleh kompleksitas dari penerapan algortima *MMR* adalah $O(n^2)$ dan algoritma *Steiner Tree* adalah O(n). Dengan demikian, implementasi MMR dan Steiner Tree dalam aplikasi pencarian alur kronologis peristiwa ini dapat mengurangi tingkat keabsurdan berita yang diperoleh secara tidak terstruktur sehingga dapat mempermudah para pembaca dalam menentukan pemahaman tentang peristiwa yang tengah terjadi dengan cepat.

Kata kunci : peringkasan-teks, mmr, *steiner tree*, alur kronologis, kecerdasan buatan.

1. PENDAHULUAN

Berdasarkan data survei *Markplus Insight* dan Majalah *Online Marketeers 2-13,* terdapat 98% dari 2015 responden menggunakan internet sebagai sumber utama informasi. Sedangkan 58,2% dari persentasi responden tersebut, konten yang dijadikan rujukan informasi adalah berita [1]. Namun, survei yang dilakukan terhadap 41 responden menunjukan 58,84% pencarian berita membutuhkan waktu yang lama. Hal ini tidak terlepas dengan hasil pencarian yang tidak mengikut pola semantik dari sejumlah data yang diperoleh. Sehingga, hasil pencarian tidak terstruktur.

Penggunaan *MMR* umum digunakan untuk mendapatkan teks representatif (kesimpulan). Metode ini digunakan untuk mengurangi redudansi dalam perangkingan kalimat pada multi dokumen. Selain itu, pengurutan berita secara tersusun mengikuti alur peristiwa sejalan dengan penentuan konstruksi suatu jaringan dengan menentukan jaringan pipa gas dari suatu graf. Maka, penelitian ini bertujuan untuk mempermudah dan mempercepat proses pencarian yang dilakukan oleh pembaca media berita online dalam mencari alur kronologis sebuah peristiwa dnegan penerapan metode *maximum marginal relevance* dan algoritma *steiner tree.*

1.1 Metode MMR

Algoritma MMR merupakan metode ekstraksi ringkasan yang digunakan untuk meringkas dokumen tunggal maupun multi dokumen [2]. MMR meringkas dokumen dengan melakukan perhitungan kesamaan antara bagian teks. Pada peringkasan dokumen dilakukan proses segmentasi dokumen menjadi kalimat dan dilakukan pengelompokan sesuai dengan gendeer kalimat tersebut *MMR* digunakan dengan mengkomibasikan *query* yang diberikan *user.*

1.2 Text Processing

Dalam melakukan proses *text processing* terdapat beberapa tahap, diantaranya :

1. *Case Folding*, yaitu proses penyamaan *case* dalam sebuah dokumen teks.
2. *Tokenizing*, yakni proses pemotongan *string* masukan berdasarkan tiap kata yang menyusunnya [2].
3. *Filtering* dan Elimiminasi *Stopwords*, dilakukan dengan mengambil kata-kata penting dari hasil token. Dilakukan

pembuangan kata yang dianggap kurang penting *(stop list)*.

4. *Stemming*, adalah gunsi untuk mentransformasikan kata-kata dalam sebuah dokumen teks menjadi kata dasrnya. Pada umumnya kata dasar pada Bahasa Indonesia terdiri dari kombinasi [3]:

   Prefiks 1 + Prefiks 2 + Kata dasar + Sufiks 3 + Sufiks 2 + Sufiks 1

5. *Cosine Similiarity*, yakni berfungsi untuk menghitung pendekatan relevansi *query* terhadap dokumen. Penggunaannya merupakan bentuk kelanjutan dari proses pembobotan *TF-IDF*. Persamaannya dapat digambarkan sebagai berikut [4]:

$$COS(\Theta_{ij}) = \frac{\sum_k d_{ik} d_{jk}}{\sqrt{\sum_k d_{ik}^2} \sqrt{\sum_k d_{jk}^2}} \quad (1\text{-}1)$$

6. Pembobotan TF-IDF, *Term Frequency Factor* adalah faktor yang menentukan bobot term pada suatu dokumen berdasarkan jumlah kemunculannya dalam dokumen tersebut. Sedangkan, *IDF (Invers Document Frequency)*, yakni pengurangan dominasi term yang sering muncul diberbagai dokumen.

   Nilai *idf* sebuah sistem kata dapat dihitung dengan menggunakan persamaan:

$$IDF = log\left(\frac{D}{dfi}\right) \quad (1\text{-}2)$$

   D adalah jumlah dokumen yang berisi *term* dan *dfi* adalah jumlah kemunculan *term* terhadap D. adapun algoritma untuk *W* (bobot) masing-masing dokumen kata kunci yakni:

$$W_{d,t} = tf_{d,t} * IDF_t \quad (1\text{-}3)$$

Pengguna yang menginginkan ruang sampel informasi disekitar query, maka harus menetapkan λ pada nilai yang lebih rendah. Sedangkan bagi pengguna yang ingin fokus untuk memperkuat dokumen-dokumen lebih relevan, maka harus menetapkan pada nilai yang lebih dekat dengan λ. Kalimat dengan nilai MMR tertinggi dari setiap perhitungan iterasi akan diambil, kemudian dipilih sebagai ringkasan. Iterasi berhenti pada saat hasil MMR maksimum sama dengan 0 [2].

1.3 Teori Graf

Graf didefinisikan sebagai sistem yang terdiri dari 2 komponen, yaitu himpunan tak kosong V(G) yang anggotanya disebut titik dan himpunan sisi E(G) yang berupa himpunan pasangan tak terurut dari buah titik berbeda di V(G) [5]. Terdapat 2 jenis graf secara umum:

1. Graf tidak berarah, yakni sisinya tidak memiliki orientasi arah
2. Graf berarah, yakni sisinya memiliki orientasi arah.

Adapun graf bipartit merupakan graf yang simpulnya dapat dipisah menjadi dua himpunan, misalnya bagian U dan V, sedemikian sehingga setiap sisi pada graf tersebut menghubungkan sebuah simpul di U dengan sebuah simpul di V, graf tersebut dapat dinyatakan sebagai G(U,V). begitu pun dengan simpul di V [6].

1.4 Multiview Graph

Multi-view graph merupakan G = (V,E,A) yakni graf yang terdiri dari *verteks*, *edge* tidak berarah dan seperangkat *edge* terarah *(arcs).*

1.5 Himpunan Dominasi

Dominating Set (Himpunan Dominasi) adalah suatu konsep penentuan suatu titik pada graf dengan ketentuan titik sebagai dominating set meng-cover atau menutupi titik yang ada disekitarnya dan seminimal mungkin dengan ketentuan graf sederhana yang tidak memiliki loop dan sisi ganda.

1.6 Algoritma Steiner Tree

Algoritma *Steiner tree* adalah sekumpulan subset atau bagian tertentu dari *vertices* pohon rentang *(spanning tree).*

1.7 Algoritma Steiner Tree

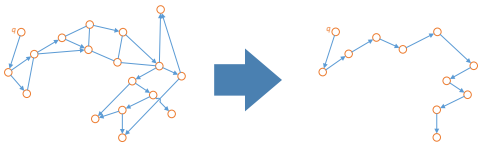
Alur kronologis adalah nama lain dari alur maju, alur lurus atau alur progresif. Peristiwa-peristiwa ditampilkan secara kronologis, maju, secara runtut dari awal tahap tengah hingga akhir. Dalam alur ini terdapat hitungan jam, menit, detik, hari dan sebagainya.

2. ISI PENELITIAN

2.1 Analisis Sistem

Berikut ini adalah uraian dari deskripsi sistem yang akan di bangun:

1. Pada tahap pertama terdapat beberapa dokumen berita dalam database.
2. User akan memberikan masukan atau query untuk memulai proses.
3. Kemudian dilakukan summarization terhadap masing-masing dokumen berita untuk mengambil kalimat inti yang merepresentasikan dokumen tersebut. Pada tahap ini digunakan metode maximum marginal relevance. Proses summarization dihitung dengan query yang diberikan oleh user.
4. Setelah mendapatkan kalimat yang merepresentasikan dokumen diantara dokumen-dokumen berita yang relevan maka dilakukan pengaitan sesuai dengan topik tersebut secara struktural. Setiap dokumen yang diwakili kalimat representatif merupakan node atau vertices (titik/vektor) dengan mengandung weight. Bila weight tinggi maka kemiripan dengan query pun tinggi.
5. Tahap selanjutnya adalah pembentukan multiview graph yakni graf yang dibentuk baik dengan sisi (edge) berarah maupun tidak berarah.
6. Selanjutnya ditentukan himpunan dominasi yang akan menutupi (covering) node tetangga dari titik yang menjadi bagian dari himpunan dominasi minimum dengan menggunakan bipartite graph
7. Maka tahap selanjutnya adalah menghubungkan titik-titik yang berada pada himpunan dominasi tersebut sehingga menjadi sebuah steiner tree.



Gambar 2.1 Proses pembentukan pohon steiner

8. Setelah terbentuk Steiner Tree, data ditampilkan dalam bentuk tabel secara berurut sesuai timeline waktu kejadian berita tersebut dikabarkan.
9.

2.2 Analisis MMR

Diberikan sebuah dokumen berita untuk melakukan analisa terhadap algoritma MMR dalam menentukan ringkasan dokumen sebagai berikut:

Washington (ANTARA News) - Presiden AS George W. Bush akan mengirim hingga 40.000 tentara lagi ke Irak saat ia mengumumkan revisi kebijakan soal Irak, tulis media massa AS, Kamis. Laporan-laporan media memberikan perkiraan antara 9.000 dan 40.000 tambahan tentara akan dikirim ke Irak. Saat ini menurut sumber militer ada 130.000 tentara AS di Irak. Rencana tersebut mungkin suatu kontroversial karena kebijakan perang Irak makin tidak populer di mata publik AS. Jaringan televisi CNN melaporkan, Bush bakal mengirim 20.000 hingga 40.000 tentara lagi. Pengumumannya kemungkinan pada awal pekan depan. Suatu peningkatan kekuatan pasukan merupakan hal yang secara aktif dibicarakan, kata seorang pejabat senior di pemerintahan kepada CNN. Sementara itu CBS News yang mengutip sumber militer AS, mengatakan bahwa Bush tengah mempersiapkan pengiriman sekitar 9.000 tentara dan marinir ke Irak, sementara 11.000 lainnya bersiaga di Kuwati dan AS. Dua brigade tentara dengan jumlah sekitar 7.500 personil akan berangkat ke Baghdad, sedangkan sekitar 1.500 marinir ke provinsi Al-Anba, lapor CBS. Pasukan lainnya bersiaga di Kuwait, dan dua lagi di AS. Surat kabar The McClatchy melaporkan bahwa Bush sedang mempertimbangkan pengiriman tiga hingga empat brigade tempur AS, atau antara 15.000 dan 20.000

Dengan asumsi *query* yang dimasukan oleh *user* yaitu *"Bagaimana terjadinya peperangan di suriah dan irak?"*. Dari dokumen tersebut, diperoleh segmentasi kalimat sebagai berikut:

Tabel 2.1 Segmentasi Kalimat

No	Kalimat
1	Washington (ANTARA News) - Presiden AS George W
2	Bush akan mengirim hingga 40.000 tentara lagi ke Irak saat ia mengumumkan revisi kebijakan soal Irak, tulis media massa AS, Kamis
3	Laporan-laporan media memberikan perkiraan antara 9.000 dan 40.000 tambahan tentara akan dikirim ke Irak
4	Saat ini menurut sumber militer ada 130.000 tentara AS di Irak
5	Rencana tersebut mungkin suatu kontroversial karena kebijakan perang Irak makin tidak populer di mata publik AS.
6	Jaringan televisi CNN melaporkan, Bush bakal mengirim 20.000 hingga 40.000 tentara lagi
7	Pengumumannya kemungkinan pada awal pekan depan
8	Suatu peningkatan kekuatan pasukan merupakan hal yang secara aktif

	dibicarakan, kata seorang pejabat senior di pemerintahan kepada CNN
9	Sementara itu CBS News yang mengutip sumber militer AS, mengatakan bahwa Bush tengah mempersiapkan pengiriman sekitar 9.000 tentara dan marinir ke Irak, sementara 11.000 lainnya bersiaga di Kuwati dan AS
10	Dua brigade tentara dengan jumlah sekitar 7.500 personil akan berangkat ke Baghdad, sedangkan sekitar 1.500 marinir ke provinsi Al-Anba, lapor CBS
11	Pasukan lainnya bersiaga di Kuwait, dan dua lagi di AS
12	Surat kabar The McClatchy melaporkan bahwa Bush sedang mempertimbangkan pengiriman tiga hingga empat brigade tempur AS, atau antara 15.000 dan 20.000 tentara, demikian AFP

Tahap textprocessing dilakukan untuk mendapatkan kata dasar dari susunan kata yang terdapat pada artikel tersebut. Berikut adalah tahapannya:

1. Case Folding, pada tahap ini, seluruh kalimat ditransformasikan kedalam bentuk huruf kecil tanpa adanya huruf kapital.
2. *Tokenizing*, diambil segmen 8 sebagai contoh hasil dari *tokenizing*.

Tabel 2.2 *Tokenizing*

No	Segmen 8
1	suatu
2	peningkatan
3	kekuatan
4	pasukan
5	merupakan
6	hal
7	yang
8	secara
9	aktif
10	dibicarakan
11	kata
12	seorang
13	pejabat
14	senior
15	di
16	pemerintahan
17	kepada
18	cnn

3. *Filtering* dan eliminasi *stopwords*, kata-kata yang tidak terlalu penting untuk dihilangkan atau dengan istilah lain eliminasi stopwords
4. *Stemming*, pada tahap ini digunakan *ECS Stemmer* dalam melakukan tahap *stemmin* yakni merubah setiap string kedalam bentuk kata dasar. Maka diperleh sebagai berikut:

Tabel 2.3 *Stemming*

No	Contoh Kalimat Segmen 8
1	Tingkat
2	Kuat
3	Pasukan
4	aktif
5	bicara
6	kata
7	jabat
8	senior
9	pemerintah
10	Cnn

Maka, dari tahap pemrosesan teks, diperoleh hasil dari *text processing*:

Tabel 2.4 Hasi dari *Text Processing*

No	Segmen 1	Segmen 2	Segmen 3	Segmen 4	Segmen 5	Segmen 6
1	*Washington*	*Bush*	*lapor*	*sumber*	*rencana*	*Jaring*
2	*antaranews*	*Kirim*	*media*	*militer*	*kontroversial*	*televisi*
3	*Presiden*	*40000*	*kira*	*130000*	*bijak*	*Cnn*
4	*As*	*tentara*	*9000*	*tentara*	*perang*	*Lapor*
5	*George*	*Irak*	*40000*	*as*	*irak*	*Bush*
6	*W*	*Umum*	*tambah*	*irak*	*populer*	*Kirim*
7		*Revisi*	*tentara*		*mata*	*20000*
8		*Bijak*	*kirim*		*publik*	*40000*
9		*Irak*	*irak*		*as*	*tentara*
10		*Tulis*				
11		*Media*				
12		*Massa*				
13		*As*				
14		*Kamis*				

No	Segmen 7	Segmen 8	Segmen 9	Segmen 10	Segmen 11	Segmen 12
1	*Umum*	*Tingkat*	*cbs*	*Brigade*	*pasukan*	*Surat*
2	*Pekan*	*Kuat*	*news*	*Tentara*	*siaga*	*Kabar*
3		*Pasukan*	*kutip*	*7500*	*kuwait*	*The*
4		*Aktif*	*sumber*	*personil*	*as*	*mcclatchy*
5		*Bicara*	*militer*	*berangkat*		*Lapor*
6		*Kata*	*As*	*baghdad*		*Bush*
7		*Jabat*	*kata*	*1500*		*timbang*
8		*Senior*	*bush*	*Marinir*		*Kirim*
9		*Pemerintah*	*siap*	*provinsi*		*brigade*
10		*Cnn*	*kirim*	*al-anba*		*tempur*
11			*9000*	*Lapor*		*As*
12			*tentara*	*Cbs*		*15000*
13			*marinir*			*20000*
14			*irak*			
15			*11000*			
16			*siaga*			
17			*Kuwait*			
18			*As*			

Perhitungan TF dan IDF

Term	Q	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	SF	LOG(n/sf)
As		1	1		1	1				2		1	1	$\sum TF = 8$	Log(13/8)=0.211

Setiap jumlah kemunculan term pada masing-masing segmen dikalikan dengan nilai *IDF*. Maka diperoleh bobot term sebagai berikut:

Tabel 2.5 Perhitungan Bobot Term “AS”

	Q	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12
TF	0	1	1	0	1	1	0	0	0	2	0	1	1
IDF	0.211	0.211	0.211	0.211	0.211	0.211	0.211	0.211	0.211	0.211	0.211	0.211	0.211
W(TF) = TF*IDF	0	0.211	0.211	0	0.211	0.211				0.422		0.211	0.211

Setelah itu, lakukan penjumlahan keseluruhan bobot term pada segmen tertentu, pada S1 diperoleh sebagai berikut:

$\sum$ W(TF)	5.781

Sehingga, bobot ssegmen secara keseluruhan, dapat dilihat sebagai berikut;

Tabel 2.6 Hasil keseluruhan bobot segmen.

Q	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12
3.377	5.781	9.502	6.09	3.924	7.743	6.266	1.751	10.237	12.981	11.085	2.65	11.074

Selanjutnya perhitungan perkalian skalar antara *query* dengan setiap dokumen (segmentasi kalimat) pada segmen S2:

$$\sum(W(Q) * W(S2)) = 0.226$$

Tahap kedua menghitung panjang setiap dokumen termasuk *query* dengan melakukan kuadrat terhadap bobot term setiap dokumen, lakukan penjumlahan nilai kuadrat tersebut dan kemudian diakarkan sehingga menghasilkan penyebut dalam persamaan 2.1. Sebagai contoh diambil nilai segmen kalimat S1:

$\sqrt{\sum W(TF)^2}$ pada segmen 2 adalah $\sqrt{6.25}$ = 2.5

Maka, diperoleh hasil secara keseluruhan, sebagai berikut:

Tabel 2.7 perhitungan Panjang Vektor

Q	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12
1.804	2.5	2.853	2.207	1.768	2.772	2.24	1.283	3.266	3.33	3.323	1.424	3.279

Setelah itu, terapkan rumus *Cosine Similiarity* untuk menghitung kemiripan *query* dengan segmentasi kalimat lainnya. Pada S2, sebagai berikut:

$$\text{COS (Q, S2)} = \frac{0.672}{1.804 * 2.853} = 0.044$$

Maka, Diperoleh:

Tabel 2.8 Nilai Kemiripan Antar Segmen

	Q	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12
Q	**1**	**0**	**0.044**	**0.028**	**0**	**0.155**	**0**	**0**	**0**	**0.019**	**0**	**0**	**0**
S1	**0**	1	0.006	0	0.01	0.006	0	0	0	0.011	0	0.013	0.005
S2	**0.044**	0.006	1	0.251	0.112	0.118	0.149	0.111	0	0.091	0.012	0.011	0.051
S3	**0.028**	0	0.251	1	0.029	0.018	0.193	0	0	0.144	0.051	0	0.06
S4	**0**	0.01	0.112	0.029	1	0.009	0.029	0.179	0	0.259	0.019	0.018	0.008
S5	**0.155**	0.006	0.118	0.018	0.009	1	0	0	0	0.022	0	0.011	0.005
S6	**0**	0	0.149	0.193	0.029	0	1	0	0.09	0.073	0.05	0	0.185
S7	**0**	0	0.111	0	0.179	0	0	1	0	0	0	0	0
S8	**0**	0	0	0	0	0	0.09	0	1	0.061	0	0.142	0
S9	**0.019**	0.011	0.091	0.144	0.259	0.022	0.073	0	0.061	1	0.13	0.298	0.048
S10	**0**	0	0.012	0.051	0.019	0	0.05	0	0	0.13	1	0	0.085
S11	**0**	0.013	0.011	0	0.018	0.011	0	0	0.142	0.298	0	1	0.01
S12	**0**	0.005	0.051	0.06	0.008	0.005	0.185	0	0	0.048	0.085	0.01	1

Perhitungan dengan metode MMR ini dilakukan dengan menghitung iterasi antara kombinasi dua matrik cosine similiarity yakni query relevance dan similiarity kalimat. Bila akan diambil ruang sampel informasi disekitar query, maka harus menetapkan λ pada nilai yang lebih rendah. Namun, jika ingin fokus untuk memperkuat dokumen yang relevan maka menetapkan λ pada nilai yang lebih dekat dengan λ. Kalimat dengan nilai MMR tertinggi itulah yang akan menjadi kalimat representatif dari artikel berita tekstual.

$$\boldsymbol{MMR} = \boldsymbol{arg} \max_{\boldsymbol{D_1 \epsilon R \backslash S}} \left[\boldsymbol{\lambda\, Sim_1(D_i, Q)} - \boldsymbol{(1-\lambda)} \max_{\boldsymbol{D_j \epsilon S}} \boldsymbol{Sim_2\,(D_i, D_j)} \right] \quad (2\text{-}1)$$

Dari analisa yang dilakukan, diperoleh hasil iterasi sebagai berikut:

Tabel 1.1 iterasi

	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12
Q1	0	0.044	0.028	0	0.155	0	0	0	0.019	0	0	0
S1	0		0	-0.002	-0.0012							
S2	0.0352		-0.015		0.0116							
S3	0.0224				0.0188							
S4	0		-0.0058		-0.0018							
S5	0.124											
S6	0		-0.0386		0							
S7	0		0		0							
S8	0		0		0							
S9	0.0152		-0.0136		0.0108							
S10	0		-0.0102		0							
S11	0		0		-0.0022							
S12	0		-0.012		-0.001							

Tabel 2.9 Hasil iterasi *MMR*

Iterasi ke	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12
1	0	0.035	0.022	0	**0.124**	0	0	0	0.015	0	0	0
2	-	0.012	**0.0188**	-		0	0	0	0.0108	0	-	-
3		-		-		-	0	0	-	-	0	-

Tabel 2.10 Bobot *MMR* maksimum iterasi *MMR*

Iterasi ke	**Kalimat**	**Bobot ArgMax *MMR***
***MMR*MAX(1)**	S5	0.124
***MMR*MAX(2)**	S3	0.0188

Dari hasil perhitungan *MMR*, Maka diperoleh hasil ekstraksi kalimat sebagai berikut:

Tabel 2.11 Hasil Ekstraksi *MMR*

Kalimat ke	Kalimat
S5	rencana tersebut mungkin suatu kontroversial karena kebijakan perang irak makin tidak populer di mata publik as.
S3	laporan-laporan media memberikan perkiraan antara 9.000 dan 40.000 tambahan tentara akan dikirim ke irak

2.3 Analisis Algoritma Steiner Tree

Sebagai kelanjutan dari perhitungan MMR, diberikan kasus bahwa terdapat beberapa dokumen berita yang merupakan dokumen berita dalam topik yang sama yakni peristiwa internasional. Diasumsikan, bahwa hasil yang diperoleh dari peringkasan adalah dengan query yang sama dengan kasus sebelumnya. Berita-berita ini diperoleh dari beberapa sumber portal berita sebagaimana telah disebutkan sebelumyan. Kumpulan berita dapat disajikan sebagai berikut:

Tabel 2.12 Dokumen *Sumamry*

No.	*Timestamp*	Konten Berita Tekstual
1	4 Januari 2007 22:43 WIB	*laporan-laporan media memberikan perkiraan antara 9.000 dan 40.000 tambahan tentara akan dikirim ke irak*
2	14 Des 2003; 23:14 WIB	*Presiden Irak Saddam Hussein ditangkap dalam sebuah operasi bersandi "Red Dawn" atau Fajar Merah*
3	13 Nov 2002 02:18 WIB	*Presiden Amerika Serikat George Walker Bush mengancam akan mengerahkan seluruh kekuatan militer AS untuk menggempur Irak.*
4	20 Maret 2003; 19:35 WIB	*Pasukan Amerika Serikat dan sekutunya telah memulai perang dengan menggempur daerah selatan Kota Baghdad, Irak, Kamis (20/3) pagi waktu setempat*
5	18 Juli 2003; 19:57 WIB	*Peringatan 35 tahun kekuasaan Partai Baath di Kota Fallujah, Irak, berlanjut.*
6	on 23 Nov 2004 06:59 WIB	*Komisi Pemilihan Umum Ukraina menyatakan Perdana Menteri Viktor Yanukovych sebagai pemenang pemilihan umum. Yanukovych yang didukung pemerintah Rusia meraup suara 49,42 persen.*
7	on 09 Des 2000 08:07 WIB	*Dua ledakan bom yang terjadi hampir bersamaan mengguncang sebuah kota di perbatasan Rusia-Chechnya, Jumat (8/12)*
8	on 05 Okt 2001 at 19:00 WIB	*Sebuah pesawat jenis Tupolev 154 milik Maskapai Penerbangan Rusia Sibir Airlines meledak dan jatuh di atas Laut Hitam, Kamis (4/10).*
9	14-4-2003 / 20:52 WIB	Tentara Amerika Serikat tiba di Mosul sejak Jumat dengan 11 mobil, diikuti oleh sekitar 300 tentara Kurdi.
10	17 Oktober 2007 14:20	*Pemerintah wilayah otonom Kurdi di Irak Utara memperingatkan para anggota parlemen Turki bahwa campur tangan apapun adalah ilegal.*
11	5 April 2006 09:46	*Pemerintah pusat Baghdad, yang menyebut operasi militer itu sebagai upaya meredam pemberontakan gerilyawan Kurdi, menyatakan wilayah itu sebagai kawasan terlarang.*
12	22 Oktober 2015 20:09	*Assad bertemu dengan Vladimir Putin di Moscow Selasa waktu setempat dan berterimakasih kepada Putin karena telah melancarkan serangan udara kepada lawan-lawannya di Suriah.*
13	25 Oktober 2015, 13:34 WIB	*AS diketahui memang telah menolak proposal yang diajukan Rusia untuk menyatukan kekuatan melawan ISIS.*
14	15-4-2003 / 17:11 WIB	*Saat berpidato di Gedung Putih, Presiden AS George W. Bush mengatakan, Suriah memiliki senjata pemusnah massal.*
15	30/09/2015 18:31 WIB	*Parlemen Rusia akhirnya membulatkan suara, memberikan kewenangan bagi pemerintah Presiden Vladimir Putin untuk mengerahkan kekuatan militer negara ke Suriah.*
16	Sabtu, 04 Peb 2012 08:02:16 WIB	*Sekitar 130.000 prajurit NATO bekerja sama dengan lebih dari 300.000 anggota pasukan keamanan Afghanistan.*

Dengan mengikuti pola pemrosesan *cosine similiarity*, maka diperoleh bobot kemiripan antar dokumen sebagai berikut:

Tabel 2.13. Bobot kemiripan antar dokumen *summary*

Q	R1	R2	R3	R4	R5	R6	R7	R8	R9	R10	R11	R12	R13	R14	R15	R16
1	0.026	0.024	0.028	0.163	0.026	0	0	0	0	0.022	0	0.057	0	0.061	0.069	0
0.026	1	0.012	0.014	0.012	0.013	0	0	0	0.096	0.011	0	0	0	0	0	0
0.024	0.012	1	0.047	0.011	0.012	0	0	0	0	0.01	0.058	0	0	0.028	0.032	0
0.028	0.014	0.047	1	0.188	0.014	0	0	0	0.103	0.012	0.045	0	0.115	0.225	0.228	0
0.163	0.012	0.011	0.188	1	0.012	0	0	0.054	0.089	0.01	0.059	0	0	0	0	0.085
0.026	0.013	0.012	0.014	0.012	1	0	0.081	0	0	0.078	0	0	0	0	0	0
0	0	0	0	0	0	1	0.017	0.013	0	0.02	0.02	0	0.018	0	0.092	0
0	0	0	0	0	0.081	0.017	1	0.08	0.076	0	0	0	0.028	0	0.026	0
0	0	0	0	0.054	0	0.013	0.08	1	0	0	0	0	0.02	0.051	0.019	0
0	0.096	0	0.103	0.089	0	0	0.076	0	1	0.041	0.04	0	0	0	0	0
0.022	0.011	0.01	0.012	0.01	0.078	0.02	0	0	0.041	1	0.116	0	0	0	0.097	0.079
0	0	0.058	0.045	0.059	0	0.02	0	0	0.04	0.116	1	0	0	0	0.071	0
0.057	0	0	0	0	0	0	0	0	0	0	0	1	0.068	0.025	0.176	0
0	0	0	0.115	0	0	0.018	0.028	0.02	0	0	0	0.068	1	0.047	0.081	0
0.061	0	0.028	0.225	0	0	0	0	0.051	0	0	0	0.025	0.047	1	0.062	0
0.069	0	0.032	0.228	0	0	0.092	0.026	0.019	0	0.097	0.071	0.176	0.081	0.062	1	0
0	0	0	0	0.085	0	0	0	0	0	0.079	0	0	0	0	0	1

Bobot dari edge tidak berarah adalah nilai similiarity dari dua edge yang saling berkaitan. Bobot masing-masing edge tersebut dapat dilihat dalam bentuk *adjcency matrix*.

Tabel 2.14 Adjacency Matrix dari *Undirected Graph*

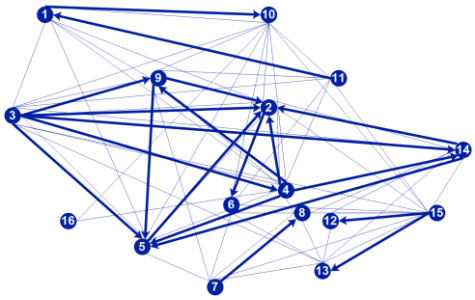
Q	R1	R2	R3	R4	R5	R6	R7	R8	R9	R1	R11	R12	R13	R14	R15	R16
1	1	1	1	1	1					1		1		1	1	
1	1	1	1	1	1				1	1						
1	1	1	1	1	1					1	1			1	1	
1	1	1	1	1	1				1	1	1		1	1	1	
1	1	1	1	1	1			1	1	1	1					1
1	1	1	1	1	1		1			1						
						1	1	1		1	1		1		1	
					1	1	1	1	1				1		1	
				1		1	1	1					1	1	1	
	1		1	1			1		1	1	1					
1	1	1	1	1	1	1			1	1	1				1	1
		1	1	1		1			1	1	1				1	
1												1	1	1	.1	
			1			1	1	1				1	1	1	1	
1		1	1					1				1	1	1	1	
1		1	1			1	1	1		1	1	1	1	1	1	
				1						1						1

Berdasarkan tenggang waktu yang ditentukan dan nilai *indeks komulatif kejadian*, maka diperoleh Graf berarah dari tabel kemiripan dalam bentuk *adjacency matrix*, sebagai berikut:

Tabel 2.15 Adjacency Matriks dari *Directed Graph*

	Q	R1	R2	R3	R4	R5	R6	R7	R8	R9	R1	R11	R12	R13	R14	R15	R16
Q																	
R1											1						
R2							1										
R3			1		1	1									1		
R4			1			1				1							
R5			1														
R6																	
R7								1									
R8																	
R9			1			1											
R10																	
R11		1															
R12																	
R13																	
R14			1			1											
R15													1	1			
R16																	

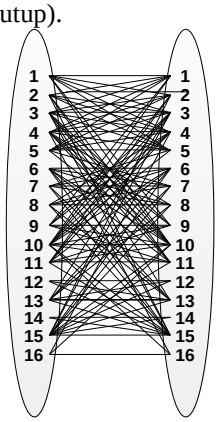
Dari tabel diatas dapat digambarkan *multiview graph* sebagai berikut:



Gambar 2.2. Gambar Representasi *Multiview Graph*

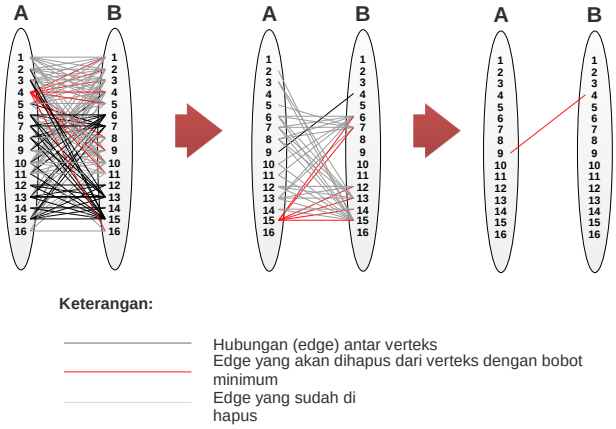
2.4. Menentukan Himpunan Dominasi Minimum

Dalam menentukan himpunan dominasi, terlebih dahulu ubah graf diatas kedalam bipartite graph diasumsikan sebagai graf R. Hal ini dilakukan untuk menghitung covering numbers (sejumlah titik/verteks penutup).



Gambar 2.3. Bentuk Bipartite Graf

tahap selanjutnya lakukan kembali sebagaimana tahap di awal yakni mencari verteks dengan bobot terbesar pada himpunan A kemudian menghapus *edge* yang berhubungan dengan verteks terpilih. Maka dapat dilihat hasilnya sebagai berikut pada graf R:

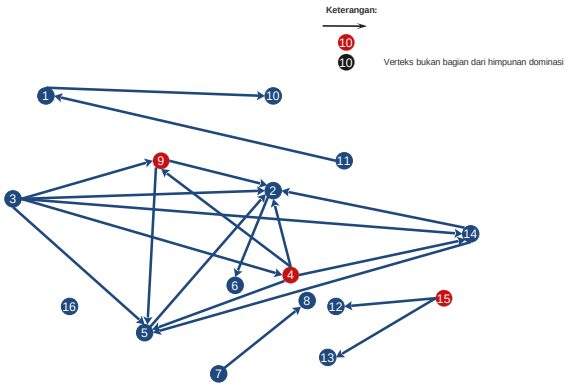


Gambar 2.4. Pencarian Himpunan Dominasi

DS (*Dominating Set*) = {R4, R15, R9}

2.5. Menentukan Steiner Tree

Setelah diperoleh titik-titik yang terdominasi, maka selanjutnya dilakukan aproksimasi steiner tree yakni untuk menghubungkan titik-titik (vertex) yang menjadi bagian dari himpunan dominasi. Berikut ini adalah gambar graf dengan edge yang ditentukan oleh timestamps beserta himpunan verteks dominasi berbobot.



Gambar 2.5. Graf Berarah

Untuk membentuk sebuah steiner tree, diperlukan matriks berbobot dan matriks lintasan terpendek.

Berikut ini adalah tabel kemiripan berdasarkan query

Tabel 2.16 Tabel Kemiripan Query

	R1	R2	R3	R4	R5	R6	R7	R8	R9	R10	R11	R12	R13	R14	R15	R16
Q	0.026	0.024	0.028	0.163	0.026	0	0	0	0	0.022	0	0.057	0	0.061	0.1	0
1-w(v)	0.974	0.976	0.972	0.837	0.974					0.978		0.943		0.939	0.931	

Dari perolehan bobot yang dihitung berdasarkan query. Maka, nilai 1-W(v) adalah bobot yang diberikan terhadap setiap vertex yang mewakili bobot dari sebuah directed edge (A). Sehingga, hasilnya dapat dilihat sebagai berikut:

Tabel 2.17 Tabel Lintasan (A) dengan Bobot Verteks

No	v	R7	R8	R3	R4	R9	R14	R5	R2	R6	R11	R1	R10	R16	R15	R12	R13
1	**R7**		**0.92**														
2	**R8**																
3	**R3**				**0.812**		**0.775**	**0.986**	**0.953**								
4	**R4**					**0.911**		**0.988**	**0.989**								
5	**R9**							1	1								
6	**R14**							1	**0.972**								
7	**R5**								**0.988**								
8	**R2**									1							
9	**R6**																
10	**R11**											1					
11	**R1**												**0.989**				
12	**R10**																
13	**R16**																
14	**R15**															**0.824**	**0.919**
15	**R12**																
16	**R13**																

Berikut ini adalah penerapan algoritma *Steiner Tree*:

Algoritma Steiner Tree

Input:

a. V adalah verteks, W adalah total Weight dan A adalah total Arc (directed edge).

b. S adalah kumpulan terminal (verteks terdominasi), q adalah verteks query dan k adalah nilai posisi dari v yang merupakan bagian dari S ataupun tidak.

c. UG adalah *Undirected Graph* atau Graf tak Berarah

d. DG adalah *Directed Graph* atau Graf Berarah

Output: Steiner tree T akar pada q menutupi paling sedikit verteks k dalam S

$A_i(k, v_0, UG, DG, S)$
T ← ϕ (kosong)
harga(T) ← 0
if k < |**S**| do
    Tbest ← ϕ (kosong)
    harga(Tbest) ← 0
    T' ← ϕ (kosong)
    harga(T') ← 0
    harga ← ∞
    for v ∈ UG ≠ 0, (v_0, v) ∈ DG, 0 ≤ k' < |**V**| do
        T' ← $A_i(k, v, UG, DG, S)$ ∪ {(v0, v)}
        if harga > cost(T') then
            Tbest ← T'
    T ← T ∪ Tbest
return T

Maka diperoleh hasil konstruksi *Steiner Tree*



Gambar 2.6. *Storyline*

Graf Steiner Tree yang diperoleh diintepretasikan sebagai sebuah alur kronologis dari kumpulan verteks dengan edge berarah berdasarkan waktu kejadian. Verteks yang terseleksi sebagai bagian dari graf steiner tree merupakan bobot akumulasi verteks dengan bobot kemiripan terhadap query paling tinggi. Sehingga, alur kronologis (Storyline) yang diperoleh adalah sebagai berikut:

Tabel 2.18 Alur Kronologis *(storyline)*

V	*Timestamp*	*Ringkasan Dokumen Query Focussed*	*Bobot W(v)*
R4	*20 Maret 2003; 19:35 WIB*	*Pasukan Amerika Serikat dan sekutunya telah memulai perang dengan menggempur daerah selatan Kota Baghdad, Irak, Kamis (20/3) pagi waktu setempat*	*0.163*
R9	*14-4-2003 / 20:52 WIB*	*Tentara Amerika Serikat tiba di Mosul sejak Jumat dengan 11 mobil, diikuti oleh sekitar 300 tentara Kurdi.*	

2.5. Pengujian

Dalam pengujian yang dilakukan dengan mengguakan metode *usability testing*, dilakukan pengukuran berdasarkan pada rekomendasi yang ditentukan terhadap responden dalam aspek *usability*. Dari tingkat *learnability*, diperoleh sebanyak 66.65% memberikan kemudahan dalam menyelesaikan tugas-tugasnya. Berdasarkan aspek *Efficiency*, 66.7% responden menilai bahwa aplikasi ini mengerjakan tugas-tugas dengan lebih cepat. Dari sisi *memorable*, sebanyak 100% responden merasakan kemudahan dalam mengakses informasi. Pada aspek *Error*, pengguna mampu mengatasi problem yang ada dengan proses pencarian berita yang menghemat waktu dengan akumulasi persentase sebanyak 75%. Sedangkan dari aspek *satisfication*, sebanyak 83.33% merasa puas dengan aplikasi ini karena sesuai dengan kebutuhan pengguna.

Dari perolehan data tersebut, maka akumulasi dari tingkat skala usability diperoleh persentasi total sebanyak 75%, yakni berdasarkan tabel kuantitatif masuk kedalam kategori kualifikasi yang berhasil, artinya pembangunan sistem ini telah **berhasil**.

Selain itu, Berdasrkan pengujian performansi terhadap algoritma yang diterapkan, diperoleh kompleksitas dari penerapan algortima *MMR* adalah $O(n^2)$ dan algoritma *Steiner Tree* adalah O(n).

5. KESIMPULAN

Berdasarkan hasil implementasi dan pengujian dari metode MMR dan algoritma Steiner Tree, maka diperoleh kesimbulan bahwa pencarian alur kronologis memberikan kemudahan secara efesien kepada pengguna dalam mencari berita-berita secara runut dalam bentuk alur kronologis.

6. DAFTAR PUSTAKA

[1] E. Lukman, "Laporan: Inilah Yang Dilakukan 74,6 Juta Pengguna Internet Indonesia Ketika Online," Techinasia, 31 Oktober 2013. [Online]. Available: Https://Id.Techinasia.Com/Tingkah-Laku-Pengguna-Internet-Indonesia/. [Accessed 7 10 2015].

[2] M. Mustaqhfiri, Z. Abidin And R. Kusumawati, "Peringkasan Teks Otomatis Berita Berbahasa Indonesia Menggunakan Metode Maximum Marginal Relevance," *Matics*, Vol. Iv, No. 4, Pp. 1-14, 2011.

[3] D. Nopiyanti And K. A. Sekarwati, "Aplikasi Pencarian Kata Dasar Dokumen Berbahasa Indonesia Dengan Metode Stemming Porter Menggunakan Php & Mysql," In Seminar Ilmiah Nasional Komputer Dan Sistem Intelijen (Kommit 2014) , Depok, 2014.

[4] R. V. Imbar, A. M. Ayub And A. Rehatta, "Implementasi Cosine Similarity Dan Algoritma Smith-Waterman Untuk Mendeteksi Kemiripan Teks," Informatika, Vol. 10, No. 1, Pp. 1-12, 2014.

[5] R. Nicholas, "Aplikasi Graf Berbobot Dalam Menentukan Jalur Angkot (Angkutan Kota) Tercepat," Itb, Bandung, 2010/2011.

[6] Z. Fathoni, "Algoritma Penentuan Graf Bipartit," Makalah If2091, P. 2, 2009.