

## APLIKASI PERINGKAS BERITA ONLINE OTOMATIS MENGUNAKAN METODE ORDINARY WEIGHTING PADA SITUS PENGUMPUL BERITA

**Wijakso Budiyo<sup>[1]</sup>, Firdaus Solihin<sup>[2]</sup>**

Program Studi Teknik Informatika, Universitas Trunojoyo  
Jl. Raya Telang, PO BOX 2, Kamal, Bangkalan  
E-mail: iniwijkak@yahoo.com<sup>[1]</sup>, fsolihin@if.trunojoyo.ac.id<sup>[2]</sup>

### ABSTRAK

Berita memiliki peranan penting dalam kehidupan manusia. Saat ini tren memperoleh informasi melalui situs berita *online* semakin populer. Kebanyakan pembaca ingin memperoleh informasi dari sebuah berita dengan cepat, ringkas dan tidak banyak membuang waktu. Tujuan penelitian ini untuk memberikan kemudahan bagi pembaca untuk mendapatkan informasi terpenting dari suatu berita dengan cepat. Pada penelitian ini, dibangun sebuah aplikasi peringkasan berita *online* otomatis yang dikembangkan menggunakan metode perhitungan *Ordinary Weighting*. Metode *Ordinary Weighting* merupakan perhitungan bobot term berdasarkan prinsip  $TF * IDF$  dengan menggunakan  $TF$  tanpa normalisasi. Tujuan aplikasi ini untuk menghasilkan sebuah ringkasan yang terdiri dari tiga kalimat dengan nilai bobot tertinggi sebagai ringkasan dari suatu artikel berita dengan tetap mempertahankan ide utama dari berita tersebut. Proses evaluasi dilakukan dengan cara membandingkan hasil ringkasan yang dihasilkan oleh sistem dengan ringkasan yang dibuat secara manual oleh koresponden melalui metode kuesioner, dengan prinsip *Majority Opinion*. Dari hasil penelitian diketahui bahwa kesesuaian hasil ringkasan sistem dengan pengguna yang diperoleh dari kuesioner menunjukkan hasil nilai rata-rata *recall* sebesar 0.47, *precision* 0.40 dan rata-rata *f-measures* sebesar 0.53.

Kata Kunci: *Peringkasan Berita Online Otomatis, Ordinary Weighting, Pengumpul Berita.*

### ABSTRACT

Today, news has an important role in human life. The current trend of obtaining information through the online news sites is become popular now. Most readers would like to get information on a news fast, quick and not wasted much time. The purpose of this study to provide convenience for the reader to get the most important information on a news fast. In this study, constructed an automatic online news summarizer applications developed using *Ordinary weighting* calculation method. *Ordinary* method of calculating the weight of term weighting is based on the principle of using the  $TF * IDF$   $TF$  without normalization. Purpose of this application to produce a summary that consists of three sentences with the highest weight value of sentence as a summary of a news article while maintaining the main idea of the story. The evaluation process is done by comparing the results of summaries produced by the system with summary generated manually by correspondence through the questionnaire method, the principle of *Majority Opinion*. The research revealed results that the suitability of the system to the user summary results obtained from the questionnaire showed the results of average value of recall is 0.47, precision 0.40 and average *f-measures* 0.53.

Keywords: *Automatic Online News Summarizer, Ordinary Weighting, News Crawler.*

## 1. PENDAHULUAN

Seiring perkembangan teknologi informasi mengakibatkan teknologi *internet* semakin pesat, sehingga memicu munculnya banyak situs berita *online*. Untuk membaca seluruh isi dari suatu berita memerlukan waktu yang lebih lama dibandingkan dengan memahami isi berita tersebut dengan hanya melalui ringkasan teksnya saja, sehingga ringkasan teks menjadi sangat penting dalam memahami suatu isi berita untuk memperoleh suatu informasi yang dibutuhkan secara cepat.

Dengan adanya ringkasan, pembaca dapat dengan cepat dan mudah memahami isi sebuah berita tanpa harus membaca keseluruhan isi berita. Hal ini dapat menghemat waktu pembaca karena dapat menghindari pembacaan pada suatu berita yang tidak relevan dengan informasi yang diharapkan oleh pembaca, apalagi saat ini informasi berita yang tersedia di *internet* cukup banyak.

Berbagai metode telah diterapkan dan masih terus dikembangkan oleh para peneliti tentang peringkasan teks. Salah satunya adalah Peringkasan Teks Otomatis (*Automated Text Summarization*) yaitu sebuah proses untuk menghasilkan ringkasan (*summary*) dari suatu artikel dengan menggunakan komputer namun tetap memiliki gambaran yang akurat dari isi suatu artikel. Tujuannya adalah mengambil sumber informasi dengan mengutip sebagian besar isi yang penting dan menampilkan kepada pembaca dalam bentuk yang ringkas sesuai dengan kebutuhan pembaca. Dengan demikian teknologi ini dapat membantu pembaca untuk menyerap informasi yang ada dalam artikel melalui ringkasan tanpa harus membaca seluruh isi dokumen.

Penelitian mengenai *Automatic Text Summarization* telah dilakukan dengan memanfaatkan metode faktorisasi matriks nonnegatif, yang diperoleh hasil evaluasi bahwa ringkasan menggunakan metode NMF cenderung menghasilkan nilai Kappa yang tinggi dan terbukti lebih bagus dibandingkan dengan ringkasan menggunakan metode LSA. Berdasarkan hasil uji coba yang dilakukan dalam menentukan rentang bilangan acak terbaik untuk implementasi peringkasan menggunakan metode NMF dapat digunakan rentang bilangan acak antara 0.1 hingga 0.25. [1]

Pada penelitian lain yang mengangkat tema mengenai *Automatic Text Summarization* telah dilakukan sebuah penelitian dengan menggunakan metode *Term Frequency* dengan *Inverse Document Frequency* dapat menghasilkan suatu ringkasan secara otomatis meskipun tanpa menggunakan proses teks *preprocessing filtering* dan *stemming* pada penelitiannya. Pada penelitian yang dikembangkan tersebut dapat memiliki tingkat akurasi hingga 61%. [2]

Tujuan dari penelitian ini adalah untuk menghasilkan suatu perangkat lunak yang dapat memberikan kemudahan bagi para pengguna berita untuk mendapatkan informasi terpenting dari suatu berita dengan cepat tanpa harus untuk membaca keseluruhan isi teks berita.

Untuk lebih memfokuskan ruang lingkup penelitian dan menghindari meluasnya permasalahan serta untuk menanggulangi keterbatasan yang dihadapi oleh penulis maka disusun beberapa batasan masalah, antara lain :

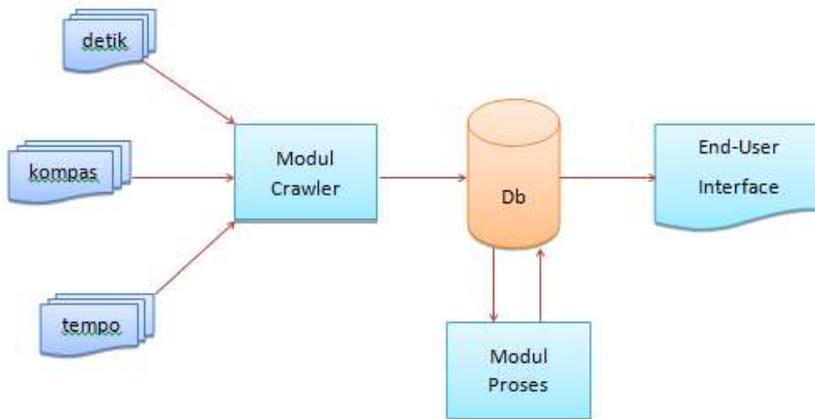
- a. Daftar situs berita yang akan dilakukan proses *crawling* hanya tiga situs berita *online* yang cukup populer saja yaitu Detik.com, Kompas.com dan Tempo.co.
- b. Jenis artikel berita yang akan digunakan pada penelitian ini hanya artikel berita yang berbahasa Indonesia saja.

## 2. METODE

Peringkasan teks otomatis (*Automatic Text Summarization*) adalah pembuatan versi yang lebih singkat dari sebuah teks dengan memanfaatkan aplikasi yang dijalankan pada komputer. Hasil peringkasan ini mengandung poin-poin penting dari teks asli. *Summary* atau ringkasan didefinisikan sebagai sebuah teks yang dihasilkan dari satu atau lebih teks, mengandung informasi dari teks asli dan panjangnya tidak lebih dari setengah teks asli. [3]

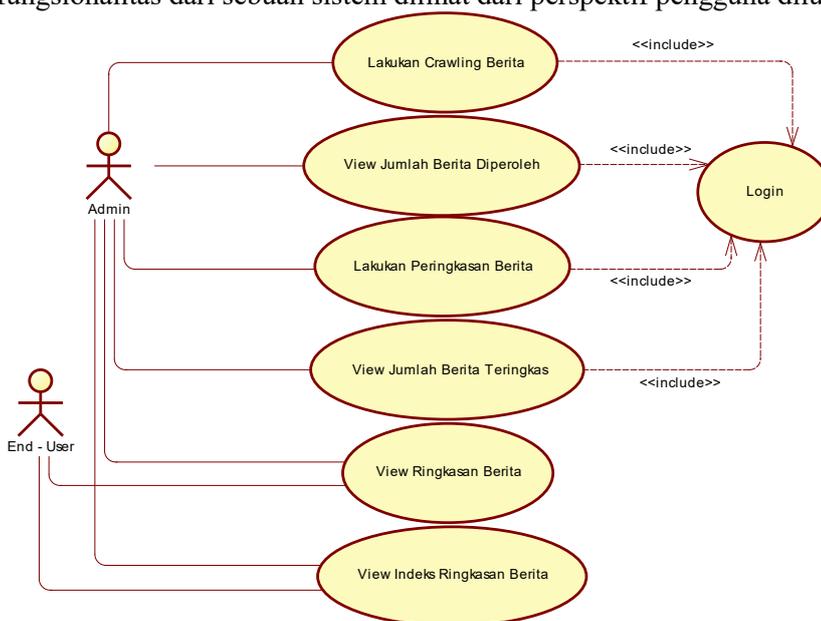
Secara garis besar kebutuhan sistem terdiri atas tiga modul utama seperti yang ditampilkan pada gambar 1, yaitu modul pengumpul artikel berita (*Web Crawler*), modul proses yang membuat peringkasan teks secara otomatis dan modul *user interface* yang menampilkan hasil ringkasan berita. Berikut penjabaran dari ketiga modul :

1. Modul *Input*, merupakan modul masukan yang berfungsi untuk mengirim sebagian atau keseluruhan teks yang akan diproses. Proses pemasukan data akan berjalan otomatis sesuai dengan jadwal yang ditentukan dengan memanfaatkan kinerja dari sebuah *web crawler*. Inputan akan berasal dari tiga situs berita online yaitu Detik.com, Kompas.com dan Tempo.co dengan lima kategori berita yang berbeda antara lain berita nasional, berita internasional, berita sepakbola, berita olahraga dan berita teknologi.
2. Modul *Proses*, merupakan kumpulan dari beberapa modul yang berfungsi untuk mengolah teks masukan menjadi suatu ringkasan. Pada modul ini terdapat beberapa proses yang akan dilakukan, yakni proses pemecahan teks menjadi kalimat, proses pemisahan kata dari kalimat, proses *stemming*, proses penghilangan *stop word*, proses penghitungan *Terms Frequency*, proses penghitungan *Document Frequency*, proses penghitungan *Inverse Document Frequency*, proses perhitungan nilai bobot untuk setiap kalimat menggunakan metode perhitungan *Ordinary Weighting* dan proses perangkaian kalimat berdasarkan nilai bobot kumulatif pada setiap kalimat.
3. Modul *Output*, merupakan bagian yang berfungsi untuk merangkai dan menampilkan hasil dari proses dalam bentuk ringkasan untuk setiap artikel berita.



Gambar 1. Modul Sistem

Pada Gambar 2 merupakan suatu bentuk *usecase* diagram yang dapat menggambarkan fungsionalitas dari sebuah sistem dilihat dari perspektif pengguna diluar sistem.



Gambar 2. Use Case Diagram

Terdapat 2 aktor utama dalam *use case diagram*, yaitu *admin* dan *end-user* (pengguna/pembaca berita). Berikut ini adalah penjelasan dari masing-masing *use case*, yaitu:

- a. Lakukan *Crawling* Berita, *admin* dapat melakukan proses *crawling* berita dengan cara memilih menu *WebCrawler* setelah melakukan *login* ke halaman *admin*.
- b. *View* Jumlah Berita Diperoleh, setelah melakukan proses *crawling* berita, *admin* dapat melihat hasil dari proses *crawling* yang dilakukan sebelumnya berupa jumlah *link* berita yang diperoleh pada masing-masing situs target beserta total *link* yang diperoleh secara keseluruhan dalam satu kali proses *crawling* dan total waktu yang sudah dilalui dalam proses *crawling* berita tersebut.
- c. Lakukan Peringkasan Berita, yaitu proses yang dapat dilakukan oleh *admin* untuk mendapatkan ringkasan dari masing-masing artikel teks berita yang diperoleh dari proses *crawling* sebelumnya.
- d. *View* Jumlah Berita Teringkas, *admin* dapat melihat hasil dari proses peringkasan berita *online* otomatis berupa jumlah berita yang teringkas beserta total waktu yang sudah dilalui selama proses peringkasan tersebut.
- e. *View* Ringkasan Berita, pengguna akhir maupun *admin* dapat melihat/ membaca hasil dari keseluruhan proses yaitu dalam bentuk ringkasan berita melalui halaman utama/ *home* yang terdiri dari 5 kategori berita dari 3 situs berita *online* yang berbeda.
- f. *View* Indeks Ringkasan Berita, yaitu pengguna akhir maupun *admin* juga dapat membaca ringkasan berita yang pernah dimuat sebelumnya melalui tautan halaman indeks berita yang terdapat di masing-masing kategori berita pada halaman utama/ *home*.

Pada penelitian ini, peringkasan teks otomatis yang di kembangkan merupakan sistem peringkasan dengan inputan berupa single dokumen dan secara otomatis menghasilkan ringkasan (*summary*). Proses teks *preprocessing* yang dilakukan pada peringkasan teks otomatis ini hanya proses *tokenizing* yaitu proses pemotongan *stringinput* berdasarkan tiap kata yang menyusunnya. Pemecahan kalimat menjadi kata -kata tunggal dilakukan dengan *me-scan* kalimat dengan pemisah (*delimiter*) *whitespace* (spasi, *tab* dan *newline*).[4]

Setelah proses pemisahan kata, maka proses selanjutnya adalah proses penghilangan *stopword*. Proses ini berfungsi untuk melakukan proses penghilangan kata tidak penting dari daftar kata yang telah dihasilkan pada proses pemisahan kata. Proses ini akan menghasilkan daftar kata kunci dari artikel berita tersebut. Kata kunci didapat dengan membandingkan kata pada daftar kata dengan daftar *stop list* yang telah diinput sebelumnya. Jika kata tersebut tidak ditemukan pada daftar *stop list* maka kata tersebut diasumsikan sebagai kata kunci. Kata tersebut kemudian akan dihitung jumlahnya dalam daftar kata. Hasil daftar kata kunci akan digunakan pada saat perhitungan nilai dari setiap kalimat.

Proses selanjutnya adalah proses *stemming*, yaitu proses untuk mengubah kata berimbuhan menjadi bentuk kata dasarnya. Pencarian bentuk kata dasar bertujuan untuk meningkatkan ketelitian pada saat proses pencarian daftar kata kunci dari keseluruhan artikel.

Pada penelitian ini algoritma *stemming* yang digunakan adalah algoritma *stemming* Nazief & Adriani yang memiliki nilai keakuratan lebih baik dari pada algoritma Porter. Algoritma ini juga memiliki akurasi nilai presisi yang bergantung pada kelengkapan kamus yang digunakan.[5]

Metode *Ordinary Weighting* merupakan perhitungan bobot *term* berdasar prinsip  $TF * IDF$  dengan menggunakan  $TF$  tanpa normalisasi[6]. Untuk dokumen tunggal tiap kalimat dianggap sebagai dokumen. Metode ini menggabungkan dua konsep untuk perhitungan bobot, yaitu *Term Frequency* (TF) merupakan frekuensi kemunculan kata (t) pada kalimat (d).

*Sentence frequency* adalah banyaknya kalimat dimana suatu kata (t) muncul. Frekuensi kemunculan kata di dalam dokumen yang diberikan menunjukkan seberapa penting kata itu di dalam dokumen tersebut. Frekuensi dokumen yang mengandung kata tersebut menunjukkan seberapa umum kata tersebut. Bobot kata semakin besar jika sering muncul dalam suatu dokumen dan semakin kecil jika muncul dalam banyak dokumen [7]. Pada Metode ini pembobotan kata dalam sebuah dokumen dilakukan dengan mengalikan nilai  $TF$  dan  $IDF$ .

Pembobotan diperoleh berdasarkan jumlah kemunculan *term* dalam kalimat (TF) dan jumlah kemunculan *term* pada seluruh kalimat dalam dokumen (IDF). Bobot suatu istilah semakin besar jika istilah tersebut sering muncul dalam suatu dokumen dan semakin kecil jika istilah tersebut muncul dalam banyak dokumen.[8]

Nilai IDF sebuah term dihitung menggunakan Persamaan 1.

$$IDF = \log \left( \frac{N}{n_{f_i}} \right) (1)$$

dengan:

N = jumlah seluruh kalimat

$n_{f_i}$  = jumlah kemunculan kata (*term*) terhadap dokumen kalimat

Pada Metode ini pembobotan kata dalam sebuah dokumen dilakukan dengan mengalikan nilai TF dan IDF. Penghitungan bobot (W) masing-masing dokumen dilakukan dengan menggunakan Persamaan 2.[9]

$$W_{d,t} = TF_{d,t} * IDF_t (2)$$

dengan :

d = kalimat ke-d

t = kata(*term*) ke -t

TF = *term frequency*

W = bobot kalimat ke-d terhadap kata(*term*)ke- t

IDF = *inverse document frequency*

Setelah dilakukan penghitungan bobot maka akan dilakukan proses pengurutan (*sorting*) nilai kumulatif dari W untuk setiap kalimat. Dari hasil *sorting* maka dihasilkan kalimat-kalimat utama yang akan dijadikan sebagai hasil dari ringkasan atau sebagai *output* dari peringkasan teks otomatis.

Pada Gambar 3 ditampilkan *pseudocode* untuk proses peringkasan berita *online* otomatis menggunakan metode perhitungan *ordinary weighting*, dimana algoritma ini merupakan perhitungan bobot *term* berdasar prinsip  $tf * idf$  dengan menggunakan *tf* tanpa normalisasi.

```

Start

Read data artikel_berita dari tb_berita.
While not empty ( isi_berita ){
    array kalimat = tokenisasi_kalimat (isi_berita)

    Foreach( kalimat ){
        array kata = tokenisasi_kata (kalimat)
        array kata = stopwords (kata)
        array kata = stemming (kata)

        Foreach( kata ){
            /* Hitung Nilai TF */
            array tf_kata = hitung_tf (kata)

            /* Hitung Nilai DF */
            array df_kata = hitung_df (kata)
        }
        End Foreach
    }
    End Foreach
}
End While

/* Hitung Nilai IDF */
N = sizeof ( kalimat )
Foreach (df_kata ) {
    array nilai_idf = hitung_idf ( log10(N/df_kata) )
}
End Foreach

/* Hitung Nilai W */
Foreach( kalimat ){
    array bobot_kal= hitung_bobot ( tf_kata * nilai_df )
}
End Foreach

Sort by value ( bobot_kal )
kalimat_tertinggi = Slice ( 3, bobot_kal )
kalimat_tertinggi_urut = Sort by index ( bobot_kal )

/* Gabungkan kalimat */
Foreach( kalimat_tertinggi_urut ){
    ringkasan =
    kalimat_tertinggi_urut[0].kalimat_tertinggi_urut[1].kalimat_
    tertinggi_urut[2]
}
End Foreach

End

```

**Gambar 3.** Pseudocode proses peringkasan berita online otomatis

Pada penelitian ini proses evaluasi hasil ringkasan dilakukan dengan metode kuesioner. Sampel data yang digunakan sebanyak 100 artikel berita yang diperoleh dari tiga situs target yaitu Detik.com, Kompas.com dan Tempo.co. Data sampel yang digunakan adalah artikel berita yang dipublikasikan pada tanggal 18 Juni 2013 dalam 5 kategori berita yang berbeda.

Proses evaluasi dilakukan dengan membandingkan kesesuaian antara kalimat ringkasan yang dihasilkan oleh sistem dengan kalimat ringkasan manual yang diperoleh dari metode kuesioner. Proses pengambilan data kuesioner dilakukan dengan bantuan lima belas orang mahasiswa yang dianggap mampu memahami isi suatu berita dengan baik.

Proses pengambilan data kuesioner dilakukan dengan memberikan satu orang koresponden sebanyak 20 artikel berita dengan petunjuk untuk memilih tiga kalimat pada setiap artikel berita yang dianggap dapat mewakili keseluruhan isi berita. Untuk setiap satu artikel berita dilakukan data kuesioner oleh tiga orang koresponden, dengan kata lain jumlah total data sampel kuesioner yang diberikan sebanyak 300 data artikel berita dengan rincian 100 berita yang berbeda diambil dari tiga situs berbeda dengan lima kategori yang sama.

Dari data yang diperoleh menggunakan metode kuesioner kemudian dilakukan proses perhitungan *Recall*, *Precision* dan *F-Measures* untuk mengetahui kualitas hasil ringkasan.

Perhitungan tersebut akan dilakukan berdasarkan nilai *correct*, *wrong* dan *missed* yang diperoleh setelah melakukan rekap data kuesioner. Dimana *correct* adalah jumlah kalimat yang diekstrak oleh sistem dan manusia, *wrong* merupakan jumlah kalimat yang diekstrak oleh sistem tetapi tidak diekstrak oleh manusia dan *missed* adalah jumlah kalimat yang diekstrak oleh manusia tetapi tidak diekstrak oleh sistem.

Dalam metode intrinsik, *Precision / Recall* dan *F-Measures* digunakan untuk mengukur kualitas suatu *Text Summarization*, dengan cara membandingkan ringkasan otomatis dan ringkasan manual (buatan manusia).[3]

$$Precision = \frac{correct}{(correct + wrong)} \quad (3)$$

$$Recall = \frac{correct}{(correct + missed)} \quad (4)$$

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

dengan :

*correct* = jumlah kalimat yang diekstrak oleh sistem dan manusia.

*wrong* = jumlah kalimat yang diekstrak oleh sistem tetapi tidak diekstrak oleh manusia.

*missed* = jumlah kalimat yang diekstrak oleh manusia tetapi tidak diekstrak oleh sistem.

Masalah yang dapat muncul dalam metode ini adalah dalam menentukan kalimat relevan karena pasti terdapat perbedaan pendapat antara subyek pembuat ringkasan dalam memilih kalimat yang dianggap penting. Untuk mengatasi hal ini digunakan beberapa metode seperti suara terbanyak (*majority opinion*), gabungan (*union*), atau irisan (*intersection*).[10]

### 3. HASIL DAN PEMBAHASAN

#### Pengujian Web Crawler

Pada bagian ini akan dilakukan pengujian dengan proses perhitungan data statistik rata-rata jumlah artikel berita yang dapat diperoleh dalam satu waktu. Pengujian ini bertujuan untuk mengetahui waktu yang paling tepat untuk melakukan proses *crawling* data artikel berita agar diperoleh jumlah artikel berita yang maksimal atau terukur dalam setiap proses *crawling*.

Data yang akan digunakan pada proses ini adalah jumlah data artikel berita *online* yang dapat dihasilkan melalui proses *web crawler*, yang berasal dari tiga situs berita *online* berbahasa Indonesia yaitu Detik.com, Kompas.com dan Tempo.co dalam lima kategori berita yang diambil artikel beritanya dalam kurun waktu 2 minggu terhitung selama 14 hari dimulai dari tanggal 10 Januari 2013 hingga tanggal 23 Januari 2013. Jumlah data artikel yang dapat dihitung akan dibatasi dengan hanya memperhitungkan artikel berita yang tersedia dalam rentang waktu 12 jam, yaitu sejak pukul 06.00 sampai dengan pukul 18.00.

Data statistik rata-rata jumlah artikel berita yang tersedia pada setiap jam dalam kurun waktu 12 jam dapat dilihat pada Tabel 1.

**Tabel 1. Tabel rata-rata jumlah artikel berita per jam**

Tanggal	Jam												Jumlah Per Hari	
	06.00	07.00	08.00	09.00	10.00	11.00	12.00	13.00	14.00	15.00	16.00	17.00		18.00
10-06-13	45	4	13	19	26	35	42	19	38	35	41	39	36	392
11-06-13	61	13	13	17	15	26	34	36	43	39	31	44	27	399
12-06-13	68	10	11	20	21	33	37	29	40	34	34	43	42	422
13-06-13	54	7	13	15	14	32	33	26	34	40	43	57	42	410
14-06-13	45	10	14	20	20	34	26	28	32	34	35	61	48	407
15-06-13	39	6	5	5	11	14	14	14	13	19	11	20	16	187
16-06-13	71	12	8	8	9	12	10	12	13	10	17	16	13	211
17-06-13	50	12	11	10	19	36	44	47	41	41	50	51	58	470
18-06-13	55	12	12	11	16	27	38	27	33	45	45	42	34	397
19-06-13	50	8	18	17	15	30	38	29	25	35	38	48	36	387
20-06-13	64	9	11	18	24	28	31	34	39	43	37	34	44	416
21-06-13	66	11	4	13	11	25	39	15	43	33	35	44	36	375
22-06-13	44	4	4	7	10	8	10	12	14	14	15	22	20	184
23-06-13	53	10	7	13	6	10	10	21	18	20	10	17	10	205
<b>Rata-rata</b>	<b>55</b>	<b>9</b>	<b>10</b>	<b>14</b>	<b>16</b>	<b>25</b>	<b>29</b>	<b>25</b>	<b>30</b>	<b>32</b>	<b>32</b>	<b>38</b>	<b>33</b>	<b>347</b>

Dari data statistik yang dihasilkan dapat diperoleh beberapa informasi, diantaranya bahwa nilai rata-rata jumlah artikel berita terbanyak dapat diperoleh pada pagi hari tepatnya pada pukul 06.00 yaitu sebanyak 55 artikel berita, sedangkan nilai terendah terdapat pada pukul 07.00. Sedangkan nilai rata-rata jumlah artikel berita yang terbilang cukup tinggi dan stabil dapat diperoleh pada kisaran waktu pukul 14.00 hingga 18.00, dimana nilai rata-rata jumlah berita berada pada kisaran angka 30an artikel berita.

Nilai rata-rata total jumlah artikel berita yang dapat diperoleh dalam satu waktu (12 jam) adalah sebanyak 347 artikel berita. Sedangkan untuk memperoleh interval waktu yang tepat untuk siklus proses *crawling* dengan jumlah artikel berita ter-target, dapat dilakukan dengan membagi nilai rata-rata total jumlah artikel berita dalam satu waktu kedalam beberapa partisi sesuai dengan jumlah artikel berita yang diinginkan. Misalnya untuk memperoleh jumlah artikel berita sebanyak 50 artikel berita untuk setiap kali proses *crawling*, maka siklus waktu proses *crawling* yang paling tepat adalah dilakukan pada pukul 06.00, 10.00, 12.00, 14.00, 16.00 dan terakhir pada pukul 18.00.

### Evaluasi Hasil Ringkasan

Proses ini merupakan sebuah proses untuk mengukur kualitas suatu ringkasan yang dihasilkan oleh sistem peringkas berita *online* otomatis, dengan cara membandingkan hasil ringkasan yang dihasilkan oleh sistem dengan ringkasan yang dibuat secara manual oleh koresponden melalui metode kuesioner.

Pada proses ini akan dilakukan proses evaluasi dengan menggunakan 100 data artikel berita yang diambil pada tanggal 18 Juni 2013. Sedangkan data ringkasan manual yang digunakan sebagai parameter evaluasi, diperoleh dengan bantuan lima belas orang mahasiswa yang dianggap mampu memahami isi suatu berita dengan baik.

Berdasarkan data yang diperoleh dari tabel hasil rekap form kuesioner, telah disepakati bahwa suatu kalimat yang dianggap benar (*Correct*, *Missed* dan *Wrong*) bila terdapat 2 atau lebih koresponden yang menyatakan kalimat tersebut sebagai bagian dari ringkasan berita. Berdasarkan hasil rekapitulasi dari perhitungan pada data jumlah kalimat yang *correct*, *wrong* dan *missed*, diperoleh nilai rata-rata *Recall* sebesar 0.47, nilai rata-rata *Precision* sebesar 0.40 dan nilai *F-Measures* sebesar 0.53. Dari hasil perolehan nilai evaluasi tersebut, mengindikasikan bahwa metode *Ordinary Weighting* yang digunakan pada sistem peringkas berita *online* otomatis dapat menghasilkan ringkasan dengan kualitas cukup baik meski hanya terdiri dari tiga kalimat.

Menurut pengamatan penulis nilai hasil evaluasi yang dirasa kurang maksimal tersebut diperoleh karena penggunaan prinsip perhitungan *Majority Opinion* dalam mencari nilai (*correct*, *wrong* dan *missed*) dengan metode kuesioner yang kurang memperhatikan jumlah kalimat ringkasan, dimana pada form kuesioner penulis membatasi responden dengan opsi

hanya memilih tiga kalimat saja sesuai dengan jumlah kalimat pada ringkasan yang dihasilkan oleh sistem. Padahal variasi jumlah kalimat yang dapat dianggap sebagai bagian dari ringkasan pada suatu artikel berita dapat berbeda-beda jumlahnya, bahkan jumlahnya dapat lebih dari hanya tiga kalimat saja. Hal tersebut berdampak pada menurunnya jumlah kalimat yang bernilai *correct* pada setiap ringkasan artikel berita yang di-evaluasi.

#### 4. KESIMPULAN DAN SARAN

##### Kesimpulan

Berdasarkan hasil dari penelitian yang sudah dilakukan dapat diambil beberapa kesimpulan, antara lain:

1. Waktu yang paling tepat untuk memperoleh sebanyak 50 artikel berita pada setiap proses *web crawling*, adalah pukul 06.00, 10.00, 12.00, 14.00, 16.00 dan 18.00.
2. Metode *OrdinaryWeighting* dapat digunakan untuk menghasilkan ringkasan dari 100 artikel berita yang memiliki kualitas cukup baik dengan rata-rata nilai *recall* sebesar 0.47, *precision* sebesar 0.40 dan *f-measures* sebesar 0.53.

##### Saran

Berikut beberapa saran yang dapat diperoleh dari hasil penelitian ini, antara lain adalah sebagai berikut:

1. Pada penelitian ini jumlah kalimat ringkasan yang digunakan pada bagian proses evaluasi dinilai kurang tepat oleh penulis, sehingga untuk penelitian selanjutnya diharapkan sebelum menentukan jumlah kalimat ringkasan dapat dilakukan sebuah perhitungan statistik terlebih dahulu untuk menentukan jumlah kalimat ringkasan yang paling tepat agar dapat diperoleh nilai evaluasi yang paling maksimal dari penggunaan metode peringkasan teks yang dipakai.
2. Pada proses pemotongan kalimat masih terdapat beberapa kekurangan dimana masih terdapat pemotongan yang bukan merupakan suatu kalimat yang terpisah, misalnya pada tanda titik yang digunakan pada singkatan nama.

#### 5. DAFTAR PUSTAKA

- [1] Yuliawati, A., Purwitasari, D., dan Yuhana, U.L. *Implementasi Peringkasan Otomatis Pada Dokumen Terstruktur Dengan Metode Faktorisasi Matriks Nonnegatif*. Teknik Informatika, Fakultas Teknologi Informasi, ITS.Surabaya. 2011.
- [2] Mulyana, I., Ramadana, S., dan Herfina. *Penerapan Terms Frequency – Inverse Document Frequency Pada Sistem Peringkasan Teks Otomatis Dokumen Tunggal Berbahasa Indonesia*. Program Studi Ilmu Komputer, FMIPA, Universitas Pakuan. 2012.
- [3] Hovy, E.H. Automated Text Summarization. In R. Mitkov (Ed.), *Handbook of computation linguistics*. Oxford: Oxford University Press. 2001.
- [4] Tala dan Fadilah Z. *A Study of Stemming Effects on Information Retrieval in Bahasa Indonesia*. Institute for logic, Language and Computation University itvan Amsterdam the Netherlands. 2003.
- [5] Agusta, L. *Perbandingan Algoritma Stemming Porter Dengan Algoritma Nazief & Adriani untuk Stemming Dokumen Teks Bahasa Indonesia*. Konferensi Nasional Sistem dan Informatika 2009. Fakultas Teknologi Informasi. Universitas Kristen Satya Wacana. Bali. 2009.
- [6] Lee, J-H., Park, S., Ahn, C.-M., dan Kim, D. *Automatic Generic Document Summarization Based On Non-negative Matrix Factorization*. Information Processing and Management, 45, 20-34. 2009.
- [7] Robertson, S. "Understanding Inverse Document Frequency: On theoretical arguments for IDF". *Journal of Documentation*, Vol.60, no.5, pp. 503-520. 2004.
- [8] Grossman, D., dan Ophir, F. *Information Retrieval: Algorithm and Heuristics*. Kuwer Academic Publisher. 1998.

- [9] Mustaqhfiri, M., Abidin Z., dan Kusumawati, R. *Peringkasan Teks Otomatis Berbahasa Indonesia Menggunakan Metode Maximum Marginal Relevance*. Ejournal Matics, 4, 4, 135-147. Dimbali 5 Januari 2012 dari basis data saintek. 2011.
- [10] Kurniawan, A. *Perancangan dan Pembuatan Perangkat Lunak Peringkasan Dokumen Berita Berbahasa Indonesia Dengan Metode Ekstraksi*. Skripsi Jurusan Teknik Informatika, Fakultas Teknologi Informasi, Institut Teknologi Sepuluh Nopember. Surabaya. 2003.