

PROSIDING

Konferensi Nasional Sistem Informasi



26 - 28 Februari 2015



Universitas Klabat
Pathway to Excellence

Dipublikasikan Tahun 2015 Oleh:
Fakultas Ilmu Komputer – Universitas Klabat
Airmadidi, Minahasa Utara, Sulawesi Utara

ISSN : 1906-9613

Panitia Tidak Bertanggung Jawab Terhadap Isi Paper dari Peserta



PROSIDING
KONFERENSI NASIONAL SISTEM INFORMASI 2015

Ketua Editor

Debby E. Sondakh, S.Kom, MT

Sekretaris Editor

Stenly R. Pungus, S.Kom, MT

Anggota Editor

Green F. Mandias, M.Cs

Oktoverano H. Lengkong, S.Kom, M.Ds

Jennifer Tambanua, S.Kom

KOMITE KNSI 2015

Steering Committee	:	Ir. Kridanto Surendro, M.Sc, Ph.D Dr. Ir. Rila Mandala, M.Eng Dr. Ir. Husni Sastramihardja, M.T Prof. Dr. Ir. Iping Supriana
Technical Committee	:	Ir. Kridanto Surendro, M.Sc, Ph.D (ITB) Dr. Ir. Rila Mandala, M.Eng (ITB) Dr. Ir. Husni Sastramihardja, M.T (ITB) Prof. Dr. Ir. Iping Supriana (ITB) Dr. Masayu Leyla Khodra (ITB) Dr. Djoko Soetarno (Univ. BINUS) Prof. Dr. A. Benny Mutiara (Univ. Gunadarma) Dr. Andrew Tanny Liem (Univ. Klabat) Stanley Karouw, ST, MTI (Univ. Sam Ratulangi)
Organizing Committee	:	
Penasihat	:	Amelius T. Mambu, MA, Ph.D Ronny H. Walean, MBA, Ph.D Marthen Sengkey, MBA, Ph.D Joppi Rondonuwu, MA, Ph.D Ir. Edson Yahuda Putra, M.Kom
Ketua Pelaksana	:	Debby E. Sondakh, S.Kom, MT
Sekretaris	:	Oktoverano Lengkong, S.Kom, M.Ds
Bendahara	:	Green Mandias S.Kom, MCs
PIC Acara	:	Stenly R. Pungus S.Kom, MT Andrew T. Liem, Ph.D Jacqueline M. Waworundeng, MT Reymond Rotikan, S.Kom, MS Jennifer Tambanua, S.Kom
Humas	:	Reynoldus Sahulata, MM
Publikasi	:	Steven Lolong, S.Kom, MT Stenly Adam, S.Kom
Multimedia	:	Andria Wahyudi, S.Kom, M.Eng Ryan Sael, S.Kom
Transportasi	:	Phaneendra Puppala, M.Sc.
Perlengkapan	:	Jimmy Moedjahedy, S.Kom, MM
Konsumsi	:	Meity Montolalu Jein M. Rewah, S.Kom, MBA

JADWAL PRESENTASI SESI II

Kamis, 26 Februari 2015

Waktu : 13.30 - 15.00 WITA

RUANG I			
NO	KODE	PENULIS	JUDUL
1	KNSI-009	Muhammad Fachrurrozi, Novi Yusliani	Analisis Sentimen Pengguna Jejaring Sosial Menggunakan Metode Support Vector Machine
2	KNSI-182	Neni Suryani, Ayi Purbasari and Agus Hexagraha	ANALISIS DAN PERANCANGAN MODEL DATA OPINION MINING MENGGUNAKAN ALGORITMA SUPPORT VECTOR MACHINE
3	KNSI-305	Jasman Pardede and Jordy Sinatria	Implementasi Maximum Marginal Relevance dan Matriks Cosine Similarity Pada Aplikasi Peringkasan Dokumen
4	KNSI-324	Achmad Ridok and Yusi Tyroni Mursityo	Optimasi Pemilihan Seeds Algoritma K-Means Pada Pengelompokan Dokumen Berbahasa Inggris
5	KNSI-165	Masayu L. Khodra, Yudi Wibisono, Ahmad Fauzan and Bagus Rahman	PEMBANGUNAN SISTEM CERDAS AGREGATOR BERITA BAHASA INDONESIA
6	KNSI-268	Dian Eka Ratnawati	Klasifikasi Buku Komputer Dengan K-Means LVQ

RUANG II			
NO	KODE	PENULIS	JUDUL
1	KNSI-306	Meri Azmi, Hidra Amnur and Adil Arrasyid	Perancangan Sistem Pendukung Keputusan Pemilihan Pegawai Teladan Seksi Berita dan Bagian Umum Pada LPP TVRI Sumatera Barat
2	KNSI-66	Jane Sumilat, Meydi Wuwung	Sistem Pendukung Keputusan Menentukan Jenis Usaha di Bidang Food and Beverage
3	KNSI-197	Rofiqoh Dewi - and Wiwi Verina -	Sistem Pendukung Keputusan Penilaian Kinerja Dosen Menggunakan Metode Fuzzy SAW (Studi Kasus Universitas Potensi Utama)
4	KNSI-196	Dahriani Hakim Tanjung	Pemilihan Sepeda Motor Second dengan Metode Analytical Hierarchy Process (AHP) (Studi Kasus: PT. XYZ)
5	KNSI-167	Dinnardian Winandya and Sali Alas M	MODEL MEDIA AJAR PENDIDIKAN USIA DINI
6			

Implementasi SAAS Pada Aplikasi Travelling Recomender Berbasis Android ----- Jasman Pardede and Raden Arie Erlangga	1280
Perancangan Aplikasi Dokter untuk Si Kecil ----- Puji Sularsih and Deasy Indayanti	1286
Pembangunan Aplikasi Generate User Interface (UI) Menggunakan Teknologi Freemarker ----- Ichsan Budiman, Ayi Purbasari and Hendra Komara	1292
Implementasi Algoritma Minmax Dengan Alpha Beta Pruning Pada Permainan Gomoku10 ----- Margaretha Desiany, Samuel Lukas and Pujianto Yugopuspito	1296
Perancangan Sistem Informasi E-Recruitment Karyawan Perusahaan ----- Susanti Margaretha Kuway and Sandy Kosasi	1300
Sistem Informasi eRecruitment Dosen Pada Perguruan Tinggi Swasta ----- Tertiaavini - and Suzan Agusti	1308
Analisis Sistem Informasi Pemantauan Status Gizi Balita di Lombok Barat ----- Helna Wardhana	1314
Aplikasi Penulisan Huruf Bali Dengan Input Teks dan Suara Berbasis Android ----- I Ketut Dedy Suryawan and Yudi Agusta	1318
Taksonomi XBRL Untuk Solusi Keragaman Informasi Pada Perbankan di Indonesia ----- Romdhoni Susiloatmadja, I Wayan Simri Wicaksana and Lintang Yu- niar Banowosari	1323
Analisa dan Rancang Bangun Sistem Absensi Mengajar di STIKOM Indonesia ----- Aniek Suryanti Kusuma and I Made Gede Sri Artha	1327
Perancangan Sistem Informasi Kamus Digital Bahasa Isyarat Tangan Secara Visual di SLB YKS Mangunreja Kabupaten Tasikmalaya ----- Egi Badar Sambani, Dani Rohpandi and Atep Kurniawan	1332
Sistem Informasi Geografis Pemantau Kualitas Udara Di Kota Cilegon ----- Anggoro Suryo Pramudyo	1342
Implementasi Maximum Marginal Relevance dan Matriks Cosine Similarity Pada Aplikasi Peringkasan Dokumen ----- Jasman Pardede and Jordy Sinatra	1348
Perancangan Sistem Pendukung Keputusan Pemilihan Pegawai Teladan Seksi Berita dan Bagian Umum Pada LPP TVRI Sumatera Barat ----- Meri Azmi, Hidra Amnur and Adil Arrasyid	1353

IMPLEMENTASI MAXIMUM MARGINAL RELEVANCE DAN MATRIKS COSINE SIMILARITY PADA APLIKASI PERINGKASAN DOKUMEN

Jasman Pardede¹, Jordy Sinatra²

^{1,2} Jurusan Teknik Informatika, Fakultas Teknik Industri, Itenas Bandung

Jln. PKH. Hasan Mustapa No.23 Bandung 40124

¹ jasman@itenas.ac.id, ²sinarta25@gmail.com

Abstract

Automatic Text Summarization are sorting paragraphs into shorter forms using a computer operated applications. This automatic summarization technique works in a computer to summarize the text inputted by user. A document entered into the computer application, then processed and summarized to produce a summary from original text. In research document extraction method used Maximum Marginal Relevance algorithm. MMR is a summary document extraction method used to summarize a single document or multiple documents. MMR summarizing document by calculating the similarity between the sentences text in a paragraph. In this summarization contents for document segmentation process is carried out using a combination of gender-based matrix cosine similarity. Cosine similarity is used to calculate the relevance of the query approach on a document. The determination of relevance in a query against document is considered as measurement of similarity between queries with vectors of documents. The results from the automatic summary application is a sequential list words according to the results obtained algorithms. The closeness results with query assessed from the lambda limit values applied in the cosine similarity is 0 to 1.

Kata kunci :summarization, MMR, queries, similarity, cosine similarity.

1. Pendahuluan

Buku adalah sumber informasi konvensional yang sering kita temukan di perpustakaan. Dalam perkembangannya buku lebih banyak dibuat secara digital karena semakin banyak sumber informasi yang dapat diakses sehingga jauh lebih efektif. Perubahan inilah yang memicu banyaknya pembuatan perpustakaan digital yang dianggap lebih baik karena dapat menghemat tempat dan lebih efisien. Perpustakaan digital kini juga dapat menampung banyak referensi yang dapat diakses oleh pembaca. Namun semakin banyaknya sumber referensi mengakibatkan pembaca kesulitan untuk mencari sumber mana yang sesuai pokok bahasan yang dipelajari. Untuk dapat mengetahui isi pokok bahasan membutuhkan waktu yang lama, karena setidaknya pembaca harus membacanya halaman demi halaman dari sumber referensi. Permasalahan ini dapat diringankan jika sebelum membaca sebuah buku, pembaca dapat langsung mengetahui isi pokok melalui *resume* yang telah disediakan.

Resume dapat diartikan sebagai hasil dari kegiatan merangkum suatu tulisan atau pembicaraan menjadi sebuah rangkaian inti kalimat yang lebih singkat dengan memperlihatkan beberapa gagasan pokok yang dibahas. Dalam membuat ringkasan pada umumnya melalui beberapa langkah yang

harus dilakukan oleh pembaca yaitu, membaca naskah asli sekitar satu atau dua kali, mencatat gagasan utama, dan menceritakan gagasan utama itu dengan menggunakan kata-kata sendiri.

Peringkasan teks atau *Automatic Text Summarization* adalah penyortiran beberapa paragraf menjadi bentuk yang lebih singkat menggunakan aplikasi yang dioperasikan dalam komputer[1,7,8]. Sebuah teks yang dimasukkan ke dalam komputer diolah dan diringkas untuk menghasilkan ringkasan dari teks asli. Teknik ini dipelopori oleh Luhn sejak tahun 1958 dan telah dikembangkan hingga kini menggunakan berbagai metode[6,8]. Beberapa metode yang sering digunakan diantaranya adalah metode genetika, Lexrank, Maximum Marginal Relevance, graph based. Metode ekstraksi dokumen yang digunakan dalam penelitian ini adalah metode Maximum Marginal Relevance.

MMR merangkas dokumen dengan menghitung *similarity* antara bagian kalimat dalam teks[2]. Perbedaan dalam penggunaan metode ini adalah *cosine similarity* yang digunakan untuk menghitung pendekatan relevansi query terhadap sebuah dokumen. Penggunaan metode ini dipilih karena dapat mengurangi redundansi dalam perangkaian kalimat pada dokumen dengan tahapan teks *preprocessing* dan penggunaan matriks *cosine*

similarity. Pada peringkasan ini dilakukan proses segmentasi isi dokumen berdasarkan gender atau koefisien λ yang digunakan untuk mengatur kombinasi bahwa kalimat tersebut relevan dan dapat mengurangi redundansi nilai menggunakan kombinasi dari matrik *cosine similarity*.

Penentuan relevansi query terhadap suatu dokumen dianggap sebagai pengukuran kesamaan antara vektor query dengan vektor dokumen. Semakin besar sebuah nilai kesamaan vektor query dengan vektor dokumen maka *query* tersebut dipandang relevan dengan dokumen yang diekstraksi. Penggunaan pendekatan relevansi menggunakan *matrik cosine similarity* diharapkan menghasilkan hasil ringkasan dari sebuah dokumen yang lebih baik.

Berdasarkan uraian latar belakang diatas, maka peneliti menemukan rumusan masalah sebagai berikut :

1. Bagaimana membuat aplikasi peringkasan dokumen dengan hasil ringkasan yang berisi gagasan pokok dari dokumen sesuai dengan *query* yang dimasukan oleh user.
2. Bagaimana menerapkan algoritma Maximum Marginal Relevance dalam ekstraksi sebuah dokumen.
3. Bagaimana melakukan pembobotan untuk mencari *cosine similarity* terhadap sebuah dokumen.
4. Bagaimana melakukan pengelompokan *similarity* sesuai dengan kalimat yang ada.

Adapun tujuan dari penelitian ini adalah mengembangkan aplikasi peringkasan dokumen yang dapat menghasilkan ringkasan yang mewakili gagasan pokok dari dokumen menggunakan algoritma MMR dengan pembobotan *cosine similarity*.

2. Tinjauan Pustaka

Pada subbab ini akan membahas tentang pengertian ringkasan, MMR, cosine similarity, term frequency dan inverse term frequency.

2.1 Ringkasan [2]

Ringkasan adalah sari karangan tanpa hiasan. Fungsi sebuah ringkasan adalah memahami atau mengetahui pokok pikiran sebuah buku atau karangan. Dengan membuat ringkasan, kita mempelajari cara seseorang menyusun pikirannya dalam gagasan-gagasan yang diatur dari gagasan yang besar menuju gagasan penunjang, melalui ringkasan kita dapat menangkap pokok pikiran dan tujuan penulis. Langkah membuat ringkasan bacaan adalah sebagai berikut :

1. Membaca bacaan atau karangan asli untuk mengetahui kesan umum dan maksud pengarang.
2. Mencatat gagasan utama atau pokok pikiran dalam tiap paragraf.

3. Menyusun pokok pikiran atau gagasan pokok bacaan menjadi satu paragraf atau lebih.

2.2 Maximum Marginal Relevance [2]

Maximum Marginal Relevance (MMR) adalah sebuah algoritma yang digunakan untuk metode ringkasan ekstrasi yang dapat meringkas satu dokumen atau lebih. Metode ini pertama kali dikemukakan oleh Carbonell dan Goldstein pada tahun 1998. Metode MMR digunakan untuk memilih kalimat dengan mempertimbangkan aspek kerelevan kalimat dengan *query* dan keterbaruan informasi, seperti yang dinyatakan pada persamaan (1).

$$MMR = \arg \max \left[\lambda * \text{Sim}_1^{(D_i, Q)} - (1-\lambda) * \text{Sim}_2^{(D_i, D)} \right] \dots (1)$$

Dimana:

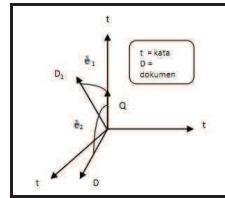
λ : Koefisiensi nilai penekan kalimat relevan
 $\text{Sim}_1, \text{Sim}_2$: Matriks dari kesamaan kalimat
 D_i : Kalimat dalam dokumen
 D : Kalimat yang telah diekstrak
 Q : *Query*

Parameter λ memiliki nilai mulai dari 0 sampai dengan 1 artinya ketika $\lambda = 1$ maka nilai MMR yang diperoleh cenderung relevan terhadap dokumen asal. Sedangkan jika $\lambda = 0$ maka nilai MMR cenderung relevan terhadap dokumen yang sudah diekstrak sebelumnya. Nilai λ yang digunakan untuk peringkasan dokumen yang paling efektif $\lambda=0.3$ untuk memberi penekanan lebih, dan kemudian meningkatkannya $\lambda=0.7$ untuk fokus pada dokumen yang paling relevan [2].

2.3 Cosine Similarity [5]

Cosine Similarity digunakan untuk menghitung pendekatan relevansi *query* terhadap dokumen. Semakin besar nilai kesamaan vektor *query* dengan vektor dokumen maka *query* tersebut dipandang semakin relevan dengan dokumen. Cosinus dari 0° adalah 1, dan kurang dari 1 untuk setiap sudut lainnya. Dengan demikian dua vektor dengan orientasi yang sama memiliki kesamaan cosinus dari 1, dua vektor pada 90° memiliki kesamaan 0. *Cosine similarity* terutama digunakan dalam ruang positif, dimana hasilnya dibatasi dengan (0,1). *Cosine similarity* kemudian memberikan tolok ukur seberapa mirip dua dokumen [8].

Cosinus dari dua vektor dapat diturunkan dengan menggunakan *dot product* Euclidean yaitu $a \cdot b = \|a\| \|b\| \cos \theta$ dimana vector atribut, A dan B kesamaan cosinus $\cos \theta$, diwakili menggunakan dot produk dan besarnya ditunjukkan dengan persamaan (2).



Gambar 1. Skalar vektor [6]

$$CS(b_1, b_2) = \frac{\sum_{t=1} W_{t,b1} W_{t,b2}}{\sqrt{\sum_{t=1}^n W_{t,b1}^2} \sqrt{\sum_{t=1}^n W_{t,b2}^2}} \dots \dots \dots (2)$$

Dimana:

t : term dalam Kalimat

w_{t, b_1} : bobot term t dalam blok b_1

2.4 Term Frequency dan Inverse Term

Frequency [4]

Term Frequency dan *Inverse Term Frequency* adalah statistik numerik yang dimaksudkan untuk mencerminkan betapa pentingnya sebuah kata dalam dokumen [6].

tf-idf banyak digunakan sebagai faktor bobot (*W*) dalam pencarian informasi dan text mining. Pembobotan diperoleh dari jumlah kemunculan term dalam sebuah dokumen *term frequency* (*tf*) dan jumlah kemunculan term dalam koleksi dokumen *inverse document frequency* (*idf*). Variasi dari skema pembobotan *tf-idf* sering digunakan oleh mesin pencari sebagai alat eksekusi. *tf-idf* dapat berhasil digunakan stop-words dalam penyarangan di berbagai bidang, termasuk *test summarization* dan klasifikasi. Stop-words gunanya untuk menghentikan kata-kata atau kata-kata yang disaring sebelum atau setelah pengolahan data bahasa alami [6].

Bot suatu istilah semakin besar jika istilah tersebut sering muncul dalam suatu dokumen dan semakin kecil jika istilah tersebut muncul dalam banyak dokumen [8].

Dalam tf frekuensi term pilihan paling sederhana adalah dengan menggunakan frekuensi baku dalam dokumen, yaitu berapa kali term (t) terjadi dalam dokumen (d). Nilai idf sebuah term (kata) dapat dihitung menggunakan persamaan (3).

D adalah jumlah dokumen yang berisi *term* (t) dan *dfi* adalah jumlah kemunculan (frekuensi) *term* terhadap D . Adapun algoritma yang digunakan untuk menghitung bobot (W) masing-masing dokumen terhadap kata kunci (*query*) menggunakan persamaan (4).

Dimana:

D : dokumen ke-*d*

t : term ke-t dari kata kunci

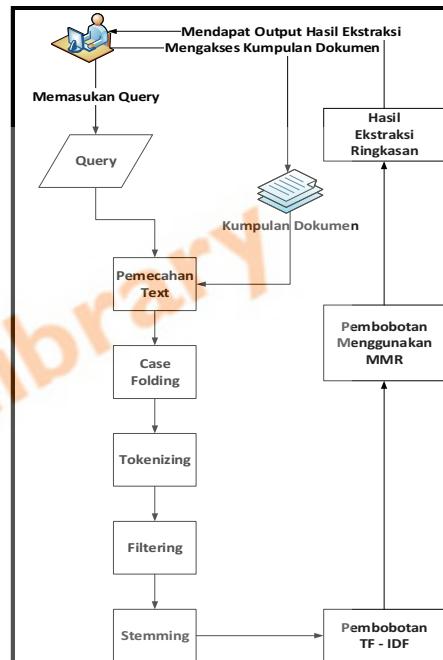
tf : term frekuensi/frekuensi kata

W : bobot dokumen ke- d terhadap term ke- t

3. Hasil Penelitian

Pada subbab berikut ini akan membahas tentang analisis sistem, use case diagram, implementasi dan teknik pengujian.

3.1 Analisis Sistem



Gambar 2. Alur kerja sistem peringkas dokumen

Peringkasan dokumen otomatis yang dibuat merupakan sistem yang dapat membaca satu teks dokumen dan secara otomatis menghasilkan sebuah ringkasan. Proses yang dilakukan adalah proses text preprocessing, pembobotan $tf-idf$, pembobotan *similarity* kalimat dan ekstraksi MMR. Dokumen yang telah dipilih oleh *user* kemudian dibaca secara keseluruhan oleh sistem untuk mempersiapkan dokumen kepada proses pertama yang dijalankan.

Tahapan analisa terhadap kebutuhan perancangan aplikasi peringkas dokumen dengan tahapan proses yang dilakukan adalah :

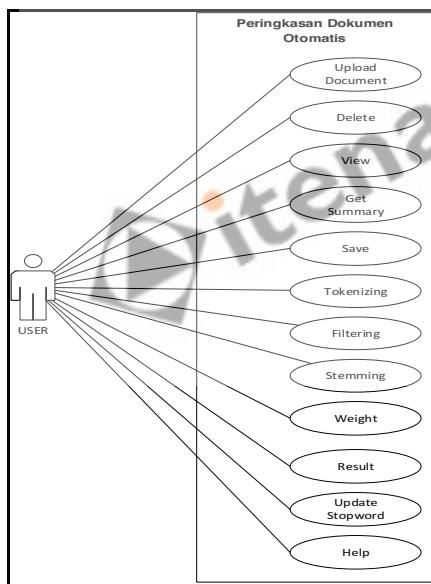
1. Pilih dokumen yang diringkas dan masukkan kalimat *query*.
2. Sistem melakukan text preprocessing yaitu pemecahan kalimat, case folding, filtering, tokenizing dan stemming[3].
3. Hitung bobot tf-idf, guna mencari bobot relevansi antar *query* terhadap seluruh kalimat yang ada dalam dokumen.

4. Hitung bobot MMR kalimat dengan kombinasi matrik *cosine similarity* dari bobot relevansi dan bobot similarity kalimat.
5. Ekstraksi menghitung bobot MMR kalimat dan kemudian menyimpulkan hasil ekstraksi dan menyajikannya terhadap user.

Output dari aplikasi peringkasan dokumen adalah sebuah paragraf yang mendekati isi pokok dari dokumen yang dimasukan oleh user yang memiliki bobot tertinggi yang sesuai dengan tingkat kemiripan *query*. Kalimat tersebut disusun sesuai dengan nilai bobot dari yang tertinggi hingga kalimat yang terendah. Kalimat yang memiliki bobot yang kurang dari parameter yaitu dari 0 hingga 1 akan dihapus. Adapun alur kerja sistem peringkas dokumen seperti yang dinyatakan pada Gambar 2.

3.2 Use case diagram

Berdasarkan analisis sistem yang dilakukan, fungsionalitas yang dibutuhkan di dalam peringkasan dokumen dengan MMR adalah fungsionalitas upload, delete, view, meringkas, save, tokenizing, filtration, stemming, weight, result, update stop word list, help, seperti yang dinyatakan pada Gambar 3.



Gambar 3. Use case diagram aplikasi

3.3 Implementasi Sistem

Untuk mengimplementasikan perancangan sistem yang dinyatakan pada subbab 3.2 membutuhkan bahasa pemrograman java JDK1.7 atau versi yang lebih tinggi. Selain itu juga menggunakan software pendukung lainnya, yaitu Play Framework 2.2.2 [10].

3.4 Teknik Pengujian

Pada aplikasi *automatic summary*, dilakukan pengujian black box testing yang berfungsi untuk

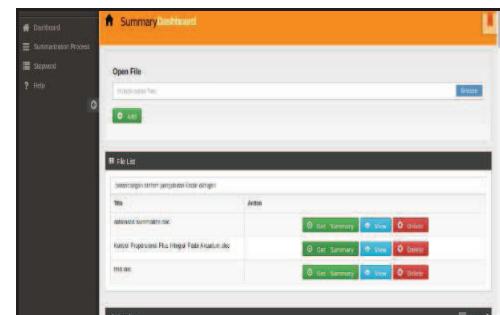
menguji setiap fungsionalitas aplikasi yang telah dibuat. Tujuan dari proses pengujian ini dilakukan untuk memeriksa apakah aplikasi ini telah melakukan fungsionalitas sesuai dengan awal perancangan pembuatan. Proses pengujian dilakukan berdasarkan Tabel 1, untuk fungsionalitas *Get Summary*.

Tabel 1. Pengujian Summary

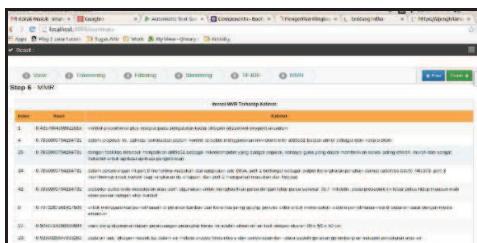
Identifikasi	TASA-04												
Nama Butir Uji	Summary												
Tujuan	Aktor dapat melihat hasil dari peringkasan dokumen yang telah dimasukan.												
Deskripsi	Pengguna memilih dokumen yang ada untuk diringkas atau meng-upload dokumen baru yang hendak diringkas, kemudian memasukkan query. Aplikasi akan melakukan proses peringkasan dokumen dengan metode MMR berdasarkan query yang diberikan oleh pengguna.												
Kondisi Awal	Aktor berada pada tab form Dashboard index.scala.html menu utama aplikasi peringkasan.												
Pengujian	<p>Skenario Uji</p> <p>1. Tekan tombol Get Summary</p> <p>Kriteria Evaluasi Hasil</p> <p>Sistem dapat meringkas dokumen serta menampilkannya kepada pengguna.</p> <table border="1"> <thead> <tr> <th colspan="4">Kasus dan Hasil Uji (Data Normal)</th> </tr> <tr> <th>Input</th> <th>Harapan</th> <th>Pengamatan</th> <th>Kesimpulan</th> </tr> </thead> <tbody> <tr> <td>Query dan dokumen yang akan diringkas</td> <td>Sistem menampilkan hasil peringkasan dokumen dari dalam tabel.</td> <td>Sistem menampilkan hasil peringkasan dokumen dari dalam tabel.</td> <td><input checked="" type="checkbox"/> Terima <input type="checkbox"/> Tolak</td> </tr> </tbody> </table>	Kasus dan Hasil Uji (Data Normal)				Input	Harapan	Pengamatan	Kesimpulan	Query dan dokumen yang akan diringkas	Sistem menampilkan hasil peringkasan dokumen dari dalam tabel.	Sistem menampilkan hasil peringkasan dokumen dari dalam tabel.	<input checked="" type="checkbox"/> Terima <input type="checkbox"/> Tolak
Kasus dan Hasil Uji (Data Normal)													
Input	Harapan	Pengamatan	Kesimpulan										
Query dan dokumen yang akan diringkas	Sistem menampilkan hasil peringkasan dokumen dari dalam tabel.	Sistem menampilkan hasil peringkasan dokumen dari dalam tabel.	<input checked="" type="checkbox"/> Terima <input type="checkbox"/> Tolak										

Berdasarkan hasil pengujian yang dilakukan pengguna terhadap Fungsionalitas Summary dengan mengikuti skenario yang dinyatakan pada Tabel 1 diperoleh hasil pengujian seperti yang ditunjukkan pada Gambar 4 dan Gambar 5.

Pada pengujian ini, dokumen yang diringkas adalah dokumen dengan id ke-2 dengan **judul** “*Kontrol Proporsional Plus Integral Pada Akuarium*”. Dengan **query** “*perancangan sistem pengaturan kadar oksigen*” menghasilkan isi ringkasan seperti yang ditunjukkan pada Gambar 5.



Gambar 4. Pengujian Fungsionalitas Get Summary



Gambar 5. Hasil Pengujian Fungsionalitas Get Summary

Berdasarkan hasil pengujian fungsionalitas get summary didapatkan hasil dari pengolahan bobot kalimat diurutkan menurut hasil relevansi terhadap *query* seperti yang dinyatakan pada Tabel 2.

Tabel 2. Hasil Ekstraksi Dokumen

Index	Hasil	Kalimat
9	0.03588492535708987	pada pr oposisi i ni perancangan si stem pengaturan kadar oksi gen adal ah dengan mengat ur put ar an mot or aer at or / pompa oksi gen
10	0.47574949662024757	adapun t ujian peneliti an ini adal ah membuat suatu perancangan si stem pengaturan kadar oksi gen yang sesuai dengan habi t asli i kan kerapu
29	0.658993177423812	untuk proses pengaturan terjadi penambahan kadar mel al ui pompa aer at or / pompa gel p t yang mengatur angka kadar oksi gen dengan cara memi tkan pengetahuan yang sebel umnya t el ah di akun penberi gar am di dasar akurasi um karena keru kan kadar gar am dapat membantu per urusan kadar oksi gen
5	0.9839067609588149	pengaturan kadar oksi gen di gunakan untuk memenuhi sal ah sat u unsur bagi pengembangan akan i kan kerapu, di sampi ng tingkat sal i nitas, i level arus, kadar ph dan per ubahan temperatur atau

Dengan cara yang sama untuk pengujian 10 dokumen, hasil peringkasan sangat dipengaruhi oleh banyaknya kalimat pada dokumen yang diringkas. Semakin banyak jumlah kalimat maka waktu yang dibutuhkan untuk meringkas dokumen akan semakin lama, dengan kebutuhan waktu setiap kalimat hampir sama, seperti pada Tabel 3.

Tabel 3. Hasil Uji Kecepatan

No	Dokumen	Jumlah Kalimat	Waktu Meringkas (detik)	Rata-Rata Tiap Kalimat
1	SUMBER 3.pdf	17	3.58	0.21
2	Abstract.doc	38	5.43	0.14
3	Instalasi Listrik.doc	49	7.14	0.15
4	Mobile Application Development.docx	66	9.21	0.14
5	SUMBER 2.pdf	82	11.23	0.14
6	Mobil Pintar.docx	101	13.43	0.13
7	SNETE-2011.pdf	108	15.32	0.14
8	Peringkas Berita Otomatis.pdf	132	17.87	0.14
9	ipi 109847.pdf	153	18.12	0.12
10	Monitoring Jaringan Listrik.doc	168	19.53	0.12
Rata-Rata Kemampuan Peringkasan Kalimat			0.14	

4. Kesimpulan

Berdasarkan pengujian sistem yang telah dilakukan peneliti pada aplikasi peringkas dokumen diperoleh kesimpulan sebagai berikut :

1. Telah berhasil mengimplementasikan algoritma *Maximum Marginal Relevance* dan matriks *cosine similarity* untuk meringkas dokumen yang berformat*.doc, *.docx, *.pdf.
2. Aplikasi *Automatic Summary* mampu menemukan kata yang relevansi dengan *query* yang diinginkan oleh pengguna seperti hasil dari pengujian pada Tabel 2, dengan hasil kalimat yang dimunculkan memiliki tingkat relevansi yang sesuai dengan batas *lambda* yang ditetapkan.
3. Penggunaan similaritas yang diterapkan pada aplikasi ini sangat dipengaruhi oleh *query* yang dimasukan oleh *user*.
4. Kebutuhan waktu untuk meringkas dokumen sangat dipengaruhi oleh jumlah kalimat yang dimiliki oleh dokumen, seperti yang dinyatakan pada Tabel 3.

Daftar Pustaka:

- [1] Aristoteles. (2013) Penerapan Algoritma Genetika pada Peringkasan Teks Dokumen Bahasa Indonesia, Universitas Lampung, Lampung.
- [2] Carbonell dan Goldstein (1998), The use of MMR, diversity-based reranking for reordering documents and producing summaries.
- [3] Dyan, Rian, Agus (2012) Algoritma Nazief dan Adriani.
- [4] Mulyana,Iyan. Ramadona,Sena. Herfina. (2013) Penerapan Terms Frequency-Inverse Document Frequency Pada Sistem Peringkasan Teks Otomatis Dokumen Tunggal Berbahasa Indonesia, Universitas Pakuan, Bogor.
- [5] Poo, D., Kiong, D., Ashok, S., (2008), *Object-Oriented Programming and Java*, Second Edition, Springer-Verlag, London.
- [6] Rajaraman, A. Ullman, J. D. (2011),*Mining of Massive Datasets*.
- [7] Ridok,Achmad. Putra, Widhy. (2014) Peringkasan Otomatis Dokumen Teks Berbahasa Indonesia Dengan Metode LexRank
- [8] Singhal, Amit (2001), *Modern Information Retrieval*
- [9] _____, Alur Proses Metodologi Prototype, <http://rhs13.blog.ugm.ac.id/>, diunduh pada tanggal 11 Mei 2014,pukul 21:00
- [10] _____, Play Framework, <https://www.playframework.com/>, diunduh pada tanggal 21 Desember 2014,pukul 3.43 WIB

Konferensi Nasional Sistem Informasi

KNSI 2015

CERTIFICATE

given to

JASMAN PARDEDE

has participated as

PRESENTER

of the paper entitled

**Implementasi Maximum Marginal Relevance dan Matriks Cosine Similarity
Pada Aplikasi Peringkasan Dokumen**

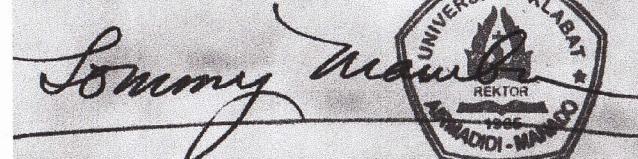
In the Konferensi Nasional Sistem Informasi 2015

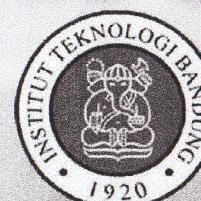
“Bridging the gap between Theories & Practices”

Conducted by Fakultas Ilmu Komputer Universitas Klabat.

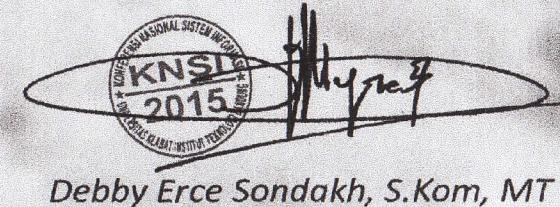
Airmadidi, Manado, on 26-28 Februari 2015

ector of Universitas Klabat,


Amelius Tommy Mambu, Ph.D.



Chairman of the Committee,


Debby Erce Sondakh, S.Kom, MT



the way to Excellence



Universitas Klabat

Jl. Arnold Mononutu Airmadidi, Minahasa Utara,
Sulawesi Utara, Indonesia, 95371
Telp.: +62 431 891035, 891041/ 42
Website: www.unklab.ac.id