

ISSN 2089-1083



EC-Council



Co-host:



PROSIDING Volume 04

SNATIKA 2017

Seminar Nasional Teknologi Informasi, Komunikasi dan Aplikasinya

Malang, 23 November 2017

diorganisasi oleh:

Lembaga Penelitian dan Pengabdian pada Masyarakat

Sekolah Tinggi Informatika dan Komputer Indonesia

SNATIKA 2017

**Seminar Nasional Teknologi Informasi, Komunikasi dan Aplikasinya
Volume 04, Tahun 2017**

PROGRAM COMMITTEE

Prof. Dr. R. Eko Indrajit, MSc, MBA (Perbanas Jakarta)
Tin Tin Hadijanto (Country Manager of EC-Council)
Dr. Eva Handriyantini, S.Kom, M.MT (STIKI Malang)

STEERING COMMITTEE

Laila Isyriyah, S.Kom, M.Kom
Sugeng Widodo, S.Kom, M.Kom
Daniel Rudiaman S., S.T, M.Kom
Subari, S.Kom, M.Kom
Jozua F. Palandi, S.Kom, M.Kom
Koko Wahyu Prasetyo, S.Kom, M.T.I
Nira Radita, S.Pd., M.Pd.

ORGANIZING COMMITTEE

Diah Arifah P., S.Kom, M.T
Meivi Kartikasari, S.Kom, M.T
Chaulina Alfianti O., S.Kom, M.T.
Eko Aprianto, S.Pd., M.Pd.
Saiful Yahya, S.Sn, M.T.
Mahendra Wibawa, S.Sn, M.Pd
Fariza Wahyu A., S.Sn, M.Sn.
Isa Suarti, S.Kom
Elly Sulistyorini, SE.
Roosye Tri H., A.Md.
Endah Wulandari, SE.
Ahmad Rianto, S.Kom
M. Syafiudin Sistiyanto, S.Kom
Muhammad Bima Indra Kusuma

SEKRETARIAT

Lembaga Penelitian dan Pengabdian Kepada Masyarakat
Sekolah Tinggi Informatika & Komputer Indonesia (STIKI) – Malang
SNATIKA 2017
Jl. Raya Tidar 100 Malang 65146, Tel. +62-341 560823, Fax. +62-341 562525
Website: snatika.stiki.ac.id
Email: snatika2017@stiki.ac.id

KATA PENGANTAR

Bapak/Ibu/Sdr. Peserta dan Pemakalah SNATIKA 2017 yang saya hormati, pertama-tama saya ucapkan selamat datang atas kehadiran Bapak/Ibu/Sdr, dan tak lupa kami mengucapkan terimakasih atas partisipasi dan peran serta Bapak/Ibu/Sdr dalam kegiatan ini.

SNATIKA 2017 adalah Seminar Nasional Teknologi Informasi, Komunikasi dan Aplikasinya yang diselenggarakan oleh STIKI Malang bekerjasama dengan EC-COUNCIL, APTIKOM Wilayah 7 dan Forum Dosen Kota Malang serta Perguruan Tinggi selaku Co-host: Universitas Nusantara PGRI Kediri dan STMIK Primakara Denpasar-Bali. Sesuai tujuannya SNATIKA 2017 merupakan sarana bagi peneliti, akademisi dan praktisi untuk mempublikasikan hasil-hasil penelitian, ide-ide terbaru mengenai Teknologi Informasi, Komunikasi dan Aplikasinya. Selain itu sesuai dengan tema yaitu "*Keamanan Informasi untuk Ketahanan Informasi Kota Cerdas*", topik-topik yang diambil disesuaikan dengan kompetensi dasar dari APTIKOM Wilayah 7 yang diharapkan dapat mensinergikan penelitian yang dilakukan oleh para peneliti di bidang Informatika dan Komputer. Semoga acara ini bermanfaat bagi kita semua terutama bagi perkembangan ilmu dan teknologi di bidang teknologi informasi, komunikasi dan aplikasinya.

Akhir kata, kami ucapkan selamat mengikuti seminar, dan semoga kita bisa bertemu kembali pada SNATIKA yang akan datang.

Malang, 20 November 2017
Panitia SNATIKA 2017

Daniel Rudiaman S., S.T, M.Kom

**SAMBUTAN KETUA
SEKOLAH TINGGI INFORMATIKA DAN KOMPUTER INDONESIA (STIKI) MALANG**

Yang saya hormati peserta Seminar Nasional SNATIKA 2017,

Puji & Syukur kita panjatkan kepada Tuhan Yang Maha Esa, atas terselenggarakannya Seminar Nasional ini sebagai rangkaian kerjasama dengan EC-COUNCIL, APTIKOM Wilayah 7 dan Forum Dosen Kota Malang serta Perguruan Tinggi selaku Co-host: Universitas Nusantara PGRI Kediri dan STMIK Primakara Denpasar-Bali. Kami ucapkan selamat datang kepada peserta Seminar Nasional serta rekan-rekan perguruan tinggi maupun mahasiswa yang telah berpartisipasi aktif sebagai pemakalah maupun peserta dalam kegiatan seminar nasional ini. Konferensi ini merupakan bagian dari 10 Flag APTIKOM untuk meningkatkan kualitas SDM ICT di Indonesia, dimana anggota APTIKOM khususnya harus haus akan ilmu untuk mampu memajukan ICT di Indonesia.

Konferensi ICT bertujuan untuk menjadi forum komunikasi antara peneliti, penggiat, birokrat pemerintah, pengembang sistem, kalangan industri dan seluruh komunitas ICT Indonesia yang ada didalam APTIKOM maupun diluar APTIKOM. Kegiatan ini diharapkan memberikan masukan kepada *stakeholder* ICT di Indonesia, yang meliputi masyarakat, pemerintah, industri dan lainnya, sehingga mampu sebagai penggerak dalam memajukan ICT Internasional.

Akhir kata, semoga forum seperti ini dapat terus dilaksanakan secara periodik sesuai dengan kegiatan tahunan APTIKOM. Dengan demikian kualitas makalah, maupun hasil penelitian dapat semakin meningkat sehingga mampu bersinergi dengan ilmuwan dan praktisi ICT internasional.

Sebagai Ketua STIKI Malang, kami mengucapkan terimakasih kepada semua pihak atas segala bantuan demi suksesnya acara ini.

“Mari Bersama Memajukan ICT Indonesia”

Malang, 20 November 2017
Ketua STIKI,

Dr. Eva Handriyantini, S.Kom, M.MT.

DAFTAR ISI

		Halaman	
	Halaman Judul	ii	
	Kata Pengantar	iii	
	Sambutan Ketua STIKI	iv	
	Daftar Isi	v	
1	<i>Erri Wahyu Puspitarini</i>	Analisa <i>Technological Content Knowledge</i> dengan menggunakan <i>Structural Equation Modeling</i>	1 - 5
2	<i>Ina Agustina, Andrianingsih, Ambi Muhammad Dzuhri</i>	Sistem Pendukung Keputusan Analisa Kinerja Tenaga <i>Marketing</i> Berbasis WEB Dengan Menggunakan Metode TOPSIS	6 - 14
3	<i>Ahmad Bagus Setiawan, Juli Sulaksono</i>	Sistem Pendataan Santri Berdasarkan Tingkat Pendidikan di Pondok Pesantren Al-Ishlah Bandar Kidul Kota Kediri	15 – 18
4	<i>Risa Helilintar, Siti Rochana, Risky Aswi Ramadhani</i>	Sistem Pakar Diagnosis Hepatitis Menggunakan Metode K-NN untuk Pelayanan Kesehatan Primer	19 - 23
5	<i>Mety Liesdiani, Enny Listiawati</i>	Sistem Kriptografi pada Citra Digital Menggunakan Metode Substitusi dan Permutasi	24 - 31
6	<i>Devie Rosa Anamisa, Faikul Umam, Aeri Rachmad</i>	Sistem Informasi Pencarian Lokasi Wisata di Kabupaten Jember Berbasis Multimedia	32 – 36
7	<i>Ardi Sanjaya, Danar Putra Pamungkas, Faris Ashofi Sholih</i>	Sistem Informasi Laboratorium Komputer di Universitas Nusantara PGRI Kediri	37 – 42
8	<i>I Wayan Rustana Putra Yasa, I Gusti Lanang Agung Raditya Putra, I Putu Agus Swastika</i>	Sistem Informasi Geografis Pemetaan Penyakit Kronis dan Demam Berdarah di Puskesmas 1 Baturiti Berbasis Website	43 - 49

9	<i>Ratih Kumalasari Niswatin, Ardi Sanjaya</i>	Sistem Informasi Berbasis Web untuk Klasifikasi Kategori Judul Skripsi	50 - 55
10	<i>Rina Firliana, Ervin Kusuma Dewi</i>	Sistem Informasi Administrasi dan Peramalan Stok Barang	56 - 61
11	<i>Patmi Kasih, Intan Nur Farida</i>	Sistem Bantu Pemilihan Dosen Pembimbing Tugas Akhir Berdasarkan Kategori Pilihan dan Keahlian Dosen menggunakan Naïve Bayes	62 – 68
12	<i>Teguh Andriyanto, Rini Indriati</i>	Rancang Bangun Sistem Informasi Sidang Proposal Skripsi di Universitas Nusantara PGRI Kediri	69 – 73
13	<i>Luh Elda Evaryanti, I Gusti Lanang Agung Raditya Putra, I Gede Putu Krisna Juliharta</i>	Rancang Bangun Sistem Informasi Perpustakaan Berbasis Website pada SMK N 1 Gianyar	74 – 80
14	<i>I Kadek Evayanto, I Gusti Lanang Agung Raditya Putra, I Putu Agus Swastika</i>	Rancang Bangun Sistem Informasi Geografis untuk <i>Monitoring</i> Kependudukan di Desa Ubung Kaja Denpasar	81 - 87
15	<i>I Gusti Ayu Made Widyari, I Gusti Lanang Agung Raditya Putra, I Gede Putu Krisna Juliharta</i>	Rancang Bangun Sistem Informasi Data Siswa Praktik Kerja Lapangan (PKL) Berbasis Web Responsive pada SMK TI Udayana	88 – 94
16	<i>Ni Putu Risna Diana Ananda Surya, I Gede Juliana Eka Putra, I Gede Putu Krisna Juliharta</i>	Rancang Bangun Sistem Informasi Akademik Berbasis Website pada Yayasan Perguruan Raj Yamuna	95 – 102
17	<i>Resty Wulanningrum, Ratih Kumalasari Niswatin</i>	Rancang Bangun Aplikasi Identifikasi Tanda Tangan Menggunakan Ekstraksi Ciri PCA	103 – 107

18	<i>Bimo Hario Andityo, Sasongko Pramono Hadi, Lukito Edi Nugroho</i>	Perancangan SOP Pemilihan Pengadaan Proyek TI Menggunakan Metode <i>E-purchasing</i> di Biro TI BPK	108 - 114
19	<i>Kadek Partha Wijaya, I Gede Juliana Eka Putra, I Gede Putu Krisna Juliharta</i>	Perancangan Sistem Informasi Media Pembelajaran Pramuka Berbasis Mobile Apps di Kwarcab Klungkung	115 – 120
20	<i>Ira Diana Sholihati, Irmawati, Dearisa Glory</i>	Aplikasi Data Mining Berbasis Web Menggunakan Algoritma Apriori untuk Data Penjualan di Apotek	121 – 126
21	<i>Sigit Riyadi, Abdul Rokhim</i>	Perancangan Aplikasi Tanggap Bencana Banjir Berbasis SMS Gateway di Desa Kedawung Wetan Pasuruan	127 – 132
22	<i>Fahrudin Salim</i>	Pengaruh <i>Information Technology Service Management (ITSM)</i> terhadap Kinerja Industri Perbankan	133 - 137
23	<i>Fajar Rohman Hariri, Risky Aswi Ramadhani</i>	Penerapan Data Mining menggunakan <i>Association Rules</i> untuk Mendukung Strategi Promosi Universitas Nusantara PGRI Kediri	138 - 142
24	<i>Johan Ericka W.P.</i>	Penentuan Lokasi <i>Road Side Unit</i> untuk Peningkatan Rasio Pengiriman Paket Data	143 – 147
25	<i>Irmawati, Sari Ningsih</i>	Pendeteksi Redundansi Frase pada Pasangan Kalimat	148 – 153
26	<i>Lilis Widayanti, Puji Subekti</i>	Pendekatan <i>Problem Based Learning</i> untuk Meningkatkan Pemahaman Konsep Mahasiswa Prodi Teknik Informatika	154 – 160
27	<i>Sufi Oktifiani, Adhistya Erna Permanasari, Eko Nugroho</i>	Model Konseptual Faktor-Faktor yang Mempengaruhi Literasi Komputer Pegawai Pemerintah	161 – 166
28	<i>Ervin Kusuma Dewi, Patmi Kasih</i>	Meningkatkan Keamanan Jaringan dengan Menggunakan Model Proses Forensik	167 - 172

29	<i>Aminul Wahib, Witarto Adi Winoto</i>	Menghitung Bobot Sebaran Kalimat Berdasarkan Sebaran Kata	173 – 179
30	<i>Evi Triandini, M Rusli, IB Suradarma</i>	Implementasi Model B2C Berdasarkan ISO 9241-151 Studi Kasus Tenun Endek, Klungkung, Bali	180 – 183
31	<i>Ina Agustina, Andrianingsih, Taufik Muhammad</i>	Implementasi Metode SAW (<i>Simple Additive Weighting</i>) pada Perancangan Sistem Pendukung Keputusan Penerimaan Beasiswa Berbasis Web	184 – 189
32	<i>Danar Putra Pamungkas, Fajar Rohman Hariri</i>	Implementasi Metode PCA dan <i>City Block Distance</i> untuk Presensi Mahasiswa Berbasis Wajah	190 – 194
33	<i>Lukman Hakim, Muhammad Imron Rosadi, Resdi Hadi Prayoga</i>	Deteksi Lokasi Citra Iris Menggunakan Threshold Linear dan Garis Horisontal Imajiner	195 – 199
34	<i>Hendry Setiawan, Windra Swastika, Ossie Leona</i>	Desain Aransemen Suara pada Algoritma Genetika	200 – 203
35	<i>Kartika Rahayu Tri Prasetyo Sari, Hisbuloh Ahlis Munawi, Yosep Satrio Wicaksono</i>	Aplikasi <i>Principal Component Analysis</i> (PCA) untuk Mengetahui Faktor yang Mempengaruhi Stres Kerja Perawat	204 – 208
36	<i>Dwi Harini, Patmi Kasih</i>	Aplikasi Bantu Sistem Informasi dan Rute Rumah Sakit di Kota Kediri dengan <i>Local Based Service</i> (LBS)	209 – 213
37	<i>Diah Arifah P., Daniel Rudiaman S.</i>	Analisa Identifikasi <i>Core Point</i> Sidik Jari	214 – 219
38	<i>Mochamad Subianto, Windra Swastika</i>	Sistem Kontrol Kolaborasi Java Programming dan MySQL pada Raspberry Pi	220 - 225
39	<i>Meme Susilowati, Hendro Poerbo Prasetya</i>	Hasil Analisis Proses Bisnis Sistem Informasi Pembiayaan Akademik sesuai Borang Akreditasi	226 – 230

40	<i>Mochamad Bilal, Teguh Andrianto</i>	Uji Kinerja Tunneling 6to4, IPv6IP Manual dan Auto	231 – 235
----	--	---	-----------

Menghitung Bobot Sebaran Kalimat Berdasarkan Sebaran Kata

Aminul Wahib¹, Witarto Adi Winoto²

Teknik Informatika
Politeknik Kota Malang
(POLTEKOM Malang)

¹wahib@poltekom.ac.id, ²witarto@poltekom.ac.id

ABSTRAK

Algoritma pembobotan kalimat memiliki peranan penting dalam natural language processing (NLP) khususnya untuk menentukan topik dan ringkasan dokumen. Bobot kalimat dapat dihitung berdasarkan bobot unsur-unsur pembentuk kalimat yaitu bobot kata. Metode sebaran kata telah membuktikan bahwa kata tersebar merupakan factor utama yang dapat dijadikan ukuran untuk mengungkap kata-kata penting dalam sebuah dokumen. Jika kata penting merupakan kata tersebar maka kalimat penting dapat dihitung berdasarkan unsur-unsur sebarannya. Penelitian ini mengenalkan metode pembobotan kalimat menggunakan sebaran kalimat. Sebaran kalimat dihitung berdasarkan unsur-unsur pembentuknya yaitu sebaran kata. Metode ini telah di ujicobakan dalam peringkasan multi-dokumen. Data yang digunakan dalam penelitian adalah Document Understanding Conferences tahun 2004 dan evaluasi menggunakan metode ROUGE-1. Hasil Ujicoba pembobotan kalimat untuk peringkasan dokumen menunjukkan metode yang diusulkan memiliki performa lebih baik jika dibandingkan dengan metode LIGI. Hasil evaluasi menggunakan ROUGE-1 terbukti metode sebaran kalimat meningkat 5,1% dibanding dengan metode LIGI.

Kata Kunci: Pembobotan kalimat, Sebaran kalimat, Natural Language Processing

1. Pendahuluan

Pembobotan kalimat adalah suatu proses atau cara yang dilakukan untuk memberikan nilai terhadap kalimat sehingga dapat diukur kedudukan kalimat tersebut terhadap kalimat-kalimat yang lain baik dalam sebuah dokumen maupun dalam banyak dokumen. Metode pembobotan kalimat banyak dikembangkan oleh para peneliti terutama dalam bidang peringkasan dokumen, pencarian dokumen, analisis dokumen berita, ekonomi, pendidikan, bisnis maupun dalam bidang-bidang lainnya. Pembobotan kalimat bertujuan untuk memperoleh kalimat penting yang berada dalam dokumen.

Kalimat penting dapat diperoleh dengan melakukan ekstraksi dokumen dengan menghitung bobot kalimat berdasarkan frekuensi kemunculan kata (Sarkar 2009). Kalimat penting dari sebuah dokumen harus kaya dan memiliki kepadatan informasi sehingga mampu mewakili dokumen sumber (He 2008). Kalimat penting dari kumpulan banyak kalimat harus mampu merepresentasikan kelompok kalimat tersebut (Sarkar 2009). Menurut (Boros 2001) kedudukan kalimat menjadi penting jika mampu mencakup topik dan terhindar

dari redundansi. Selain itu kalimat dapat menjadi penting jika kata-kata yang menyusun kalimat tersebut juga penting (Wahib 2016). Kata penting adalah kata-kata yang tersebar pada setiap bagian dalam sebuah dokumen (Tian 2011).

Penelitian (Tian 2011) menghasilkan dua teori. Teori pertama disebut dengan sebaran kata lokal, teori ini menyatakan bahwa kata dalam sebuah dokumen yang memiliki sebaran pada beberapa bagian (tersebar luas dalam beberapa paragraf) seharusnya memiliki nilai yang lebih tinggi jika dibanding dengan kata yang memiliki frekuensi yang sama namun hanya tersebar dalam bagian (paragraf) tertentu, karena sebuah kata yang tersebar dalam sebuah dokumen lebih mempresentasikan topik dari dokumen tersebut. Teori kedua (Tian 2011) disebut dengan sebaran kata global, teori ini menyatakan bahwa, hubungan sebaran kata dalam kumpulan dokumen dengan bobot kata adalah negatif nonlinier, sehingga semakin jarang kata tersebut tersebar dalam kumpulan dokumen, maka bobot kata tersebut akan semakin tinggi.

Berdasarkan hipotesa bahwa kata penting adalah kata tersebar maka dapat disimpulkan kalimat penting adalah kalimat

yang tersebar. Penelitian ini mengenalkan metode baru pembobotan kalimat menggunakan sebaran kalimat. Sebaran kalimat dihitung berdasarkan pola sebaran kalimat lokal dan pola sebaran kalimat global.

Hasil eksperimen menunjukkan pembobotan kalimat menggunakan metode sebaran kalimat dalam peringkasan multi-dokumen memiliki hasil yang lebih baik jika dibandingkan dengan metode yang diusulkan oleh (Sarkar 2009).

2. Metode Penelitian

Metode penelitian ini meliputi pembentukan metode sebaran kalimat lokal dan sebaran kalimat global. Sebaran kalimat lokal menghitung bobot kalimat dalam sebuah *cluster* kalimat sedangkan Sebaran kalimat global menghitung bobot kalimat dalam keseluruhan *cluster* kalimat yang terlibat.

2.1 Metode Sebaran Kalimat Lokal

Metode sebaran kalimat lokal dikembangkan dari metode sebaran kata lokal. Metode sebaran kata lokal diperoleh dengan menghitung frekuensi kata yang sering muncul dengan memperhitungkan lokasi sebaran dalam sebuah dokumen. Lokasi sebaran yang dimaksud adalah paragraf (Tian 2011).

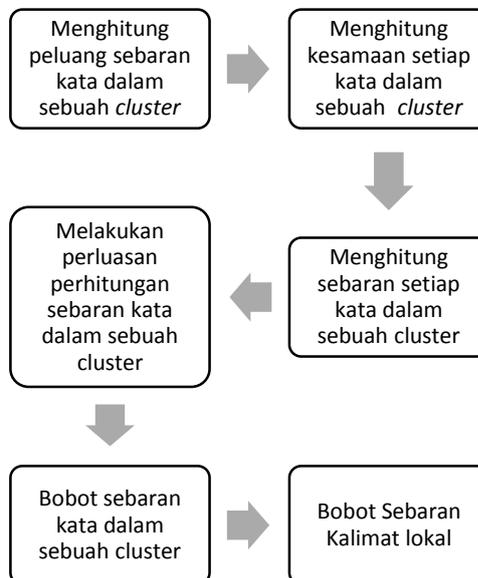
Tabel 1 menunjukkan perbedaan definisi metode sebaran kata lokal dengan metode sebaran kalimat lokal dalam menggunakan objek data perhitungan bobot.

Tabel 1
Objek yang digunakan Pembobotan Kata lokal dan Pembobotan kalimat lokal

Pembobotan Sebaran Kata	Pembobotan Sebaran Kalimat
$d_j = \{p_{1j}, p_{2j}, p_{3j}, \dots, p_{nj}\}$ adalah dokumen ke- j yang terdiri dari paragraf p sebanyak n buah.	$c_j = \{s_{1j}, s_{2j}, s_{3j}, \dots, s_{nj}\}$ adalah <i>cluster</i> ke- j yang terdiri dari kalimat s sebanyak n buah.
$W_{d-1} (jk)$ adalah bobot lokal kata ke- j pada dokumen ke- k	$W_{is} (s_{ik})$ adalah bobot lokal kalimat ke- i pada <i>cluster</i> ke- k

Metode Sebaran kalimat lokal digunakan untuk menentukan kedudukan setiap kalimat dalam sebuah *cluster*. Dengan asumsi kalimat yang memiliki unsur-unsur kata paling tersebar dalam sebuah *cluster* akan memiliki kedudukan paling tinggi

dalam *cluster* tersebut. Dengan metode ini diharapkan terpilih satu kalimat yang paling representatif yang mampu mewakili *cluster* tersebut. Detail metode sebaran kalimat lokal dapat dilihat pada Gambar 1.



Gambar 1 Metode menghitung bobot sebaran kalimat lokal

Proses awal dalam pembobotan sebaran kalimat lokal adalah menghitung peluang sebaran kata dalam sebuah *cluster* yang dapat dilakukan melalui persamaan 1.

$$r_{ij} = \frac{|s_{ik}|_{dt}}{|c_k|_{dt}} \quad \dots(1)$$

dimana r_{ij} adalah peluang sebaran seragam kata ke- j dalam kalimat ke- i , $|s_{ik}|_{dt}$ merupakan jumlah kata (berbeda) dari kalimat ke- i pada *cluster* ke- k dan $|c_k|$ adalah jumlah $|s_{ik}|_{dt}$ dalam *cluster* ke- k .

Dengan menghitung jumlah perbedaan antara *frekuensi* kata dengan *frekuensi* sebaran seragam kata ke- j dalam kalimat ke- i maka tingkat kesamaan antar kata dalam sebuah *cluster* dapat diperoleh menggunakan persamaan 2.

$$\chi_{jk}^2 = \sum_{j=1}^{|c_k|_{dt}} \frac{(v_{ij} - n_{jk} r_{ij})^2}{n_{jk} r_{ij}}, \quad \dots(2)$$

dimana χ_{jk}^2 adalah sebaran seragam kata ke- j dalam *cluster* ke- k , n_{jk} adalah *frekuensi* kata ke- j dalam *cluster* ke- k , v_{ij} merupakan *frekuensi* kata ke- j dalam kalimat ke- i , dan $|c_k|$ adalah jumlah kata (berbeda) dalam *cluster* ke- k .

Formula 2 menunjukkan semakin kecil nilai χ^2_{jk} maka kata ke- j semakin mendekati penyebaran seragam yang mana nilai tersebut bertentang dengan hubungan bobot dan penyebaran kata yang memiliki korelasi positif non linier (Tian 2011) sehingga formula ini dikembangkan seperti persamaan 3.

$$U_{jk} = \frac{1}{1 + \sum_{j=1}^{|c_k|_{dt}} \frac{(v_{ij} - n_{jk} r_{ij})^2}{n_{jk} r_{ij}}} \dots(3)$$

dimana, U_{jk} adalah nilai(tingkat) sebaran kata ke- j seragam dalam cluster ke- k .

Untuk mengukur sebaran kata ke- j dalam cluster secara optimal dilakukan perluasan perhitungan sehingga diperoleh persamaan 4.

$$St_{jk} = \log_2 \left(1 + \frac{P_{jk}}{P_k} \right) \dots(4)$$

dimana p_{jk} adalah jumlah kalimat yang mengandung kata ke- j pada cluster ke- k , P_k adalah jumlah total kalimat dalam cluster ke- k . Sehingga bobot kata ke- j dalam sebuah cluster dapat dihitung dengan persamaan 5 dan persamaan 6.

$$Wt_{l,jk} = \log_2 (1 + U_{jk} * St_{jk}) \dots(5)$$

$$Wt_{l,jk} = \log_2 \left(1 + \frac{\log_2 \left(1 + \frac{P_{jk}}{P_k} \right)}{1 + \sum_{j=1}^{|c_k|_{dt}} \frac{(v_{ij} - n_{jk} r_{ij})^2}{n_{jk} r_{ij}}} \right) \dots(6)$$

dimana, $Wt_{l,jk}$ adalah bobot sebaran kata lokal ke- j dalam cluster ke- k .

Bobot sebaran kalimat lokal diperoleh dengan menjumlahkan seluruh bobot lokal kata pembentuk kalimat ke- i dalam cluster ke- k dengan memperhatikan panjang kalimat atau jumlah kata pembentuk kalimat, hal ini dilakukan untuk menghindari hanya kalimat yang memiliki jumlah kata paling banyak yang akan muncul sebagai kalimat penting tanpa memperhatikan makna yang terkandung dalam kalimat tersebut, hal ini sesuai dengan persamaan 7.

$$W_{ls}(s_{ik}) = \frac{1}{|s_{ik}|} \sum_{Wt_{l,jk} \in s_{ik}} Wt_{l,jk} \dots(7)$$

dimana, $W_{ls}(s_{ik})$ adalah bobot sebaran kalimat lokal kalimat ke- i pada cluster ke- k

dan $|s_{ik}|$ merupakan jumlah kata penyusun kalimat ke- i pada cluster ke- k .

Untuk mengimplementasikan metode sebaran kalimat lokal kedalam program dapat menggunakan pseudocode seperti ditunjukkan pada Gambar 2.

```

ON ← (Ordered cluster list)
Wls ← Empty List (local sentences
score list)
LS ← Empty List (important local
sentences)
for each cluster c in ON do
    Calculate |ck|dt, njk, P
for each sentence s in cluster c do
    Calculate |sik|dt
    Calculate rij = |sik|dt / |ck|dt
for each term t in sentence s do
    Calculate pjk, vij
    Calculate xij = (vij - njk.rij)2 / njk.rij
end for
for each term t in cluster c do
    Calculate Xjk2 = sum(xij)
    Calculate Ujk = 1 / (1 + Xjk2)
    Calculate Stjk = log(1 + pjk / Pk)
    Calculate Wtjk = log(1 + Ujk * Stjk)
end for
Calculate W =  $\frac{1}{|s_{ik}|} \sum_{Wt_{jk} \in s} Wt_{jk}$ 
Add W to Wls
Add s to LS
end for

```

Gambar 2 Pseudocode Metode sebaran kalimat lokal

2.2 Metode Sebaran Kalimat Global

Metode sebaran kalimat global jika dibandingkan dengan metode sebaran kata global memiliki perbedaan dalam pemilihan objek data seperti ditunjukkan pada tabel 2.

Tabel 2
Objek Pembobotan Kata global dan kalimat global

Pembobotan Sebaran Kata	Pembobotan Sebaran Kalimat
$D = \{d_1, d_2, d_3, \dots, d_N\}$ adalah kumpulan dokumen sebanyak N buah.	$C = \{c_1, c_2, c_3, \dots, c_N\}$ adalah kumpulan cluster sebanyak N buah.
$W_{d-g}(j)$ adalah bobot global	$W_{gs}(s_{ik})$ adalah bobot global kalimat ke- i

kata ke- j pada koleksi dokumen	pada kumpulan $cluster$
-----------------------------------	-------------------------

Sebaran kalimat global adalah sebaran kata kata penting penyusun kalimat dalam kumpulan $cluster$. Posisi bobot sebaran kalimat global dapat diilustrasikan seperti pada Tabel 2.

Tabel 3
Analogi bobot sebaran kalimat global pada sekumpulan $cluster$

Cluster	Kalimat 1	Kalimat 2	Kalimat 3	Kalimat n
Cluster 1	$W_{\Xi}(1)$	$W_{\Xi}(2)$	$W_{\Xi}(3)$	$W_{\Xi}(n)$
Cluster 2				
Cluster 3				
Cluster 4				

Jika kita perhatikan pada Tabel 3 posisi bobot sebaran kalimat global adalah memberikan nilai bobot kalimat ke- i dimana ($i = 1,2,3 \dots n$) secara global pada sekumpulan $cluster$. Perbedaan dengan metode sebaran kalimat lokal adalah perhitungan bobot kalimat lokal memberikan bobot setiap kalimat pada sebuah $cluster$ ($cluster$ tertentu) sedangkan metode sebaran kalimat global memberikan bobot setiap kalimat pada keseluruhan $cluster$. Detail metode sebaran kalimat global ditunjukkan pada Gambar 3.



Gambar 3. Metode menghitung bobot sebaran kalimat global

Langkah awal untuk menghitung bobot kalimat global adalah mendapatkan nilai peluang sebaran kata dalam

keseluruhan $cluster$. Peluang sebaran dapat diperoleh menggunakan perhitungan yang ditunjukkan pada persamaan 8.

$$r_{jk} = \frac{|c_k|_{dt}}{|c|_{dt}}, \quad \dots(8)$$

dimana, r_{ij} adalah peluang sebaran seragam kata ke- j dalam $cluster$ ke- k , $|c_k|_{dt}$ adalah jumlah kata berbeda pada $cluster$ ke- k , dan $|c|_{dt}$ merupakan jumlah kata berbeda dalam kumpulan $cluster$.

Jumlah perbedaan antara frekuensi kata dengan frekuensi sebaran seragam kata ke- j dalam $cluster$ ke- k dapat digunakan untuk menentukan kata seragam ke- j dalam kumpulan $cluster$ seperti ditunjukkan pada persamaan 9.

$$\chi_j^2 = \sum_{j=1}^{|c|_{dt}} \frac{(v_{jk} - n_j r_{jk})^2}{n_j r_{jk}}, \quad \dots(9)$$

dimana, χ_j^2 adalah sebaran seragam kata ke- j dalam kumpulan $cluster$, n_j merupakan frekuensi kata ke- j dalam koleksi $cluster$ dan v_{jk} adalah frekuensi kata ke- j dalam $cluster$ ke- k .

Semakin kecil nilai χ_j^2 menunjukkan bahwa kata ke- j semakin mendekati sebaran seragam yang mana nilai tersebut selaras dengan hubungan bobot dan penyebaran kata yang memiliki korelasi negatif non linier [5] sehingga tingkat sebaran kata seragam dalam keseluruhan $cluster$ dapat dihitung berdasarkan persamaan 10.

$$U_j = 1 + \sum_{j=1}^{|c|_{dt}} \frac{(v_{jk} - n_j r_{jk})^2}{n_j r_{jk}}, \quad \dots(10)$$

dimana, U_j adalah nilai (tingkat) penyebaran kata seragam dalam kumpulan $cluster$.

Untuk mengukur penyebaran kata ke- j dalam kumpulan $cluster$ secara optimal dilakukan perluasan perhitungan menggunakan persamaan 11.

$$St_j = \log_2 \left(1 + \frac{P}{p_j} \right), \quad \dots(11)$$

dimana p_j adalah jumlah $cluster$ yang mengandung kata ke- j dan P adalah jumlah total $cluster$. Jika pada penelitian [5] formula yang diusulkan dalam perluasan perhitungan sebaran kata seragam menggunakan

$$S' = \log_2(1 + \frac{P_j'}{P'})$$

namun formula tersebut tak sesuai dengan analisis manual yang dilakukan oleh peneliti [5] yang menyatakan hubungan sebaran kata dengan bobot kata adalah negatif non linier, sehingga semakin jarang kata tersebut tersebar dalam kumpulan dokumen semakin tinggi bobot kata tersebut, sehingga formula 11 pada penelitian ini diusulkan. Formula ini juga sesuai dengan konsep IDF (*inverse document frequency*) yang memiliki formula $\log_2(P/p)+1$ yang artinya semakin jarang kata ke- j dalam kumpulan dokumen maka bobot kata ke- j semakin besar.

Setelah diperoleh formula penyusunan bobot dalam kumpulan *cluster* maka bobot kata ke- j dalam kumpulan *cluster* dapat dihitung dengan persamaan 12 atau persamaan 13.

$$W_{t_{g,j}} = \log_2(1 + U_j' * St_j') \quad \dots(12)$$

$$W_{t_{g,j}} = \log_2(1 + (1 + \sum_{j=1}^{|c_k|} \frac{(v_{jk}' - n_j' r_{jk}')^2}{n_j' r_{jk}'})) \cdot \log_2(1 + \frac{P'}{p_j'}) \dots(13)$$

dimana $W_{g,j}$ adalah bobot global kata ke- j dalam kumpulan *cluster*.

Bobot kalimat global diperoleh dengan menjumlahkan seluruh bobot global kata pembentuk kalimat ke- i dalam *cluster* ke- k dengan memperhatikan panjang kalimat atau jumlah kata pembentuk kalimat, seperti ditunjukkan persamaan 14. Hal ini dilakukan untuk menghindari hanya kalimat yang memiliki jumlah kata paling banyak yang akan muncul sebagai kalimat penting tanpa memperhatikan makna yang terkandung dalam kalimat tersebut.

$$W_{gs}(s_{ik}) = \frac{1}{|s_{ik}|} \sum_{W_{t_{g,j}} \in s_{ik}} W_{t_{g,j}}, \quad \dots(14)$$

dimana $W_{gs}(s_{ik})$ adalah bobot global *sentence* kalimat ke- i pada kumpulan *cluster* dan $|s_{ik}|$ merupakan jumlah kata penyusun kalimat ke- i pada *cluster* ke- k .

Untuk mengimplementasikan metode sebaran kalimat global kedalam program dapat menggunakan pseudocode seperti ditunjukkan pada Gambar 4.

$ON \leftarrow$ (*Ordered cluster list*)

$W_{gs} \leftarrow$ *Empty List*(*local sentences score list*)

$GS \leftarrow$ *Empty List* (*important global sentences*)

for each cluster c in ON do

for each term t in cluster c do

Calculate $|c|_{dt}, n_j', P'$

Calculate p_j, v_{jk}

Calculate $|c_k|_{dt}$

Calculate $r_{jk}' = |c_k|_{dt} / |c|_{dt}$

Calculate $x_{jk}' = (v_{jk}' - n_j' r_{jk}')^2 / n_j' r_{jk}'$

Calculate $X_j^2' = \text{sum}(x_{jk}')$

Calculate $U_j' = 1 / 1 + X_j^2'$

Calculate $St_j' = \log(1 + p' / P_j')$

Calculate $Wt_j' = \log(1 + U_j' * St_j')$

end for

for each sentences s in cluster c do

$$\text{Calculate } W = \frac{1}{|s_{ik}|} \sum_{W_{t_j} \in s_{ik}} W_{t_j}'$$

Add W to W_{gs}

Add s to GS

end for

end for

Gambar 4. Pseudocode Metode sebaran kalimat global

Metode sebaran kalimat dirumuskan dari sebaran kalimat lokal dan sebaran kalimat global dapat dilihat pada persamaan 15.

$$Weight_{(s_{ik})} = W_{ts}(s_{ik}) \times W_{gs}(s_{ik}), \dots(15)$$

3. Hasil Penelitian dan Pembahasan

Untuk mengukur efektifitas metode yang diusulkan, maka dilakukan uji coba pembobotan kalimat pada peringkasan multi-dokumen. Inti dari peringkasan multi-dokumen adalah mencari kalimat-kalimat penting yang melibatkan banyak dokumen (Na 2016). Data analisis yang digunakan untuk *eksperimen* adalah data *Document Understanding Conferences (DUC) Tahun 2014*. Uji coba dilakukan terhadap dua puluh lima dokumen DUC. Hasil eksperimen dibandingkan dengan metode LIGI yang dikenalkan oleh Sarkar (Sarkar 2009).

Analisis hasil diukur menggunakan metode *ROUGE-1*. *ROUGE-1* merupakan pengukuran dengan konsep *unigram matching*, dimana pengukuran ini dihitung berdasarkan jumlah setiap satu kata (*unigram*) yang sesuai antara ringkasan hasil sistem dengan ringkasan referensi yang dibuat manual oleh pakar (Lin 2014). Hasil Uji coba menggunakan metode *ROUGE-1*

terhadap dua puluh lima dokumen secara lengkap dapat dilihat pada Tabel 7.

Tabel 7
Hasil eksperimen

No	Data Uji (ID Dokumen)	Sebaran Kalimat	LIGI
1	d30042t	0.41048	0.36771
2	d30044t	0.43269	0.38679
3	d30045t	0.39462	0.29493
4	d30046t	0.4717	0.44976
5	d30047t	0.47489	0.41096
6	d30048t	0.32227	0.35514
7	d30049t	0.46226	0.42254
8	d30050t	0.35294	0.29146
9	d30051t	0.32836	0.42365
10	d30053t	0.36364	0.38144
11	d30055t	0.46364	0.47343
12	d30056t	0.35294	0.33175
13	d30059t	0.45192	0.43
14	d31001t	0.38182	0.34653
15	d31008t	0.43439	0.39423
16	d31009t	0.4455	0.4537
17	d31013t	0.34862	0.3871
18	d31022t	0.34978	0.35616
19	d31026t	0.4188	0.4569
20	d31031t	0.30986	0.32075
21	d31032t	0.36199	0.43318
22	d31033t	0.4433	0.34518
23	d31038t	0.40179	0.327277
24	d31043t	0.51741	0.43137
25	d31050t	0.40976	0.33981
Rata-Rata		0.4042	0.3845

Tabel 7 menunjukkan bahwa metode sebaran kalimat memiliki rata-rata nilai *ROUGE-1* lebih tinggi jika dibandingkan dengan metode *LIGI*. Nilai rata-rata yang diperoleh pada pengujian *ROUGE-1* dengan menggunakan metode sebaran kalimat adalah 0,404, dan pada metode *LIGI* didapatkan hasil rata-rata 0,384. Artinya metode sebaran kalimat lebih baik atau terjadi peningkatan 5,1% dibandingkan dengan metode *LIGI*.

4. Kesimpulan

Pembobotan kalimat menggunakan metode sebaran kalimat terbukti mampu

diimplementasikan dengan baik. Metode sebaran kalimat terbukti mampu mengekstraksi kalimat penting dalam kumpulan dokumen, sehingga dapat meningkatkan kualitas hasil ringkasan dengan capaian *ROUGE-1* 0,404. Hasil pemilihan kalimat dengan metode yang diusulkan memberikan peningkatan sebesar 5,1% jika dibandingkan dengan metode *LIGI*.

5. Referensi

- [1] Boros, E. Kantor, P. B. and Neu, D. J. 2011, "A Clustering Based Approach to Creating Multi-Docment Summaries". In *Proceedings of the 24th ACM SIGIR Conference*, Eds: Kraft, D. H. et al., ACM, New Orleans, Los Angeles, pp. 1-4.
- [2] He, T., Li F., Shao, W., Chen, J., and Ma, L. 2008, "A New Feature-Fusion Sentence Selecting Strategy for Query-Focused Multi-document Summarization", *Proceeding of International Conference Advance Language Processing and Web Information Technology*, Eds: Ock C. et al., University of Normal, Wuhan, China, pp. 81-86.
- [3] Lin, C. Y. 2014, "*ROUGE*: a Package for Automatic Evaluation of Summaries", In *Proceedings of Workshop on Text Summarization*.
- [4] Na Liu, Ying Lu, Xiao-jun Tang, Hai-wen Wang, Peng Xiao, and Ming-xia Li. 2016, "Multi-document Summarization Algorithm based on Significance Sentences", In *Proceedings IEEE of 28th Chinese Control and Decision Conference (CCDC)*, pp. 3847-3852.
- [5] Sarkar, K. 2009, "Sentence Clustering-based Summarization of Multiple Text Documents", *International Journal of Computing Science and Communication Technologies*, Vol. 2, No. 1, pp. 325-335.
- [6] Tian X. and Chai Y. 2011, "An Improvement to TF-IDF : Term Distribution based Term Weight Algorithm", *Journal of Software*, Vol. 6, No.3, pp. 413-420.

- [7] Wahib A., Arifin Z. A., and Purwitasari D. 2016, "Improving Multi-Document Summary Method Based on Sentence Distribution", *TELKOMNIKA (Telecommunication Comput. Electron. Control*, Vol.14, No.1, pp 286-293.