

LAPORAN AKHIR  
PENELITIAN DOSEN PEMULA



INTEGRASI PERINGKAS DOKUMEN OTOMATIS DENGAN  
ALGORITMA LATENT SEMANTIC ANALYSIS (LSA) UNTUK  
PROSES CLUSTERING DOKUMEN

Tahun ke 1 dari rencana 1 tahun

**TIM PENGUSUL :**

<b>Ardytha Luthfiarta, M.Kom, M.CS</b>	<b>NIDN: 0625078504</b>
<b>Junta Zeniarja, M.Kom, M.CS</b>	<b>NIDN: 0606068802</b>
<b>Abu Salam, M.Kom</b>	<b>NIDN: 0629038402</b>

**UNIVERSITAS DIAN NUSWANTORO**

**OKTOBER 2013**

### HALAMAN PENGESAHAN

**Judul Kegiatan** : INTEGRASI PERINGKAS DOKUMEN OTOMATIS DENGAN ALGORITMA LATENT SEMANTIC ANALYSIS (LSA) UNTUK PROSES CLUSTERING DOKUMEN

**Peneliti / Pelaksana**  
**Nama Lengkap** : ARDYTHA LUTHFIARTA  
**NIDN** : 0629078504  
**Jabatan Fungsional** :  
**Program Studi** : Teknik Informatika  
**Nomor HP** :  
**Surel (e-mail)** : ardyt\_moslem@yahoo.com

**Anggota Peneliti (1)**  
**Nama Lengkap** : JUNTA ZENIARIA  
**NIDN** : 0606068802  
**Perguruan Tinggi** : UNIVERSITAS DIAN NUSWANTORO

**Anggota Peneliti (2)**  
**Nama Lengkap** : ABU SALAM  
**NIDN** : 0629038402  
**Perguruan Tinggi** : UNIVERSITAS DIAN NUSWANTORO

**Institusi Mitra (jika ada)**  
**Nama Institusi Mitra** :  
**Alamat** :  
**Penanggung Jawab** :  
**Tahun Pelaksanaan** : Tahun ke 1 dari rencana 1 tahun  
**Biaya Tahun Berjalan** : Rp. 11.500.000,00  
**Biaya Keseluruhan** : Rp. 14.990.000,00

Mengetahui  
Dekan Fakultas Ilmu Komputer  
  
(Dr. Abdul Syukur, MM)  
NIP/NIK 0686.11.1992.017



Semarang, 11-11-2012,  
Ketua Peneliti,  
  
(ARDYTHA LUTHFIARTA)  
NIP/NIK0686.11.2012.460

Menyetujui,  
Ketua Lembaga LPPM  
  
(S.Si. M. Komil)  
NIP/NIK 0686.11.1994.046



## **KRINGKASAN**

Clustering dokumen merupakan proses pengelompokan dokumen yang memiliki kesamaan topik, clustering dokumen memudahkan pengguna menemukan dokumen yang diinginkan. Dalam proses clustering dokumen, dokumen direpresentasikan menggunakan Vector Space Model (VSM). Masalah dalam VSM adalah matrik term-dokumen biasanya sangat jarang (banyak mengandung angka 0 dalam term- dokumen matrik) dan juga mempunyai dimensi tinggi, sehingga masalah-masalah ini dapat mengurangi kinerja clustering dokumen, dalam hal ini untuk mengatasi dimensi tinggi tersebut digunakan proses peringkasan dokumen otomatis sebagai proses reduksi kalimat sebelum dilakukan proses Clustering. Dalam penelitian ini diusulkan model integrasi peringkasan dokumen otomatis menggunakan algoritma LSA (Latent Semantic Analysis) untuk proses clustering dokumen. Algoritma ini akan dibandingkan dengan algoritma Feature Based yang sudah ada dalam penelitian sebelumnya untuk meningkatkan akurasi hasil clustering dokumen. Tujuan dari penelitian ini adalah untuk meningkatkan akurasi dari clustering dokumen dengan mengintegrasikan peringkasan dokumen otomatis. Proses peringkasan dokumen otomatis ditujukan untuk penyeleksian kalimat agar didapatkan ringkasan teks yang diperoleh dengan menyajikan kembali bagian tulisan yang dianggap topik utama tulisan dengan bentuk yang lebih disederhanakan, selanjutnya dilakukan proses clustering dokumen.

Keyword: Clustering dokumen, Latent Semantic Analysis (LSA), Peringkasan Dokumen Otomatis

## **PRAKATA**

Alhamdulillah, segala puji bagi Allah Yang Maha Kuasa karena dengan rahmat dan karunia-Nya dapat diselesaikan laporan kemajuan penelitian tahap pertama 70%, dengan judul Integrasi Peringkat Dokumen Otomatis Dengan Algoritma Latent Semantic Analysis (LSA) Untuk Proses Clustering Dokumen.

Ucapan terimakasih yang sebesar-besarnya kepada Direktorat Penelitian dan Pengabdian Kepada Masyarakat, Direktorat Jenderal Pendidikan Tinggi, Kementerian Pendidikan dan Kebudayaan yang telah memberikan kepercayaan dengan membiayai penelitian ini melalui program Penelitian Dosen Pemula tahun pembiayaan 2013. Kami ucapkan terima kasih juga kepada Kepala Lembaga Penelitian dan Pengabdian Masyarakat (LP2M) Universitas Dian Nuswantoro beserta staff yang telah memberikan dukungan dan pelatihan workshop pelatihan penelitian. Ucapan terima kasih kepada seluruh anggota tim peneliti yang telah bersemangat dan bekerja keras untuk menyelesaikan tahap pertama penelitian ini. Serta terimakasih kami ucapkan kepada semua pihak atas dukungannya pada penelitian ini.

Besar harapan kami bahwa hasil penelitian ini dapat berguna bagi akademik dan bagi ilmu pengetahuan khususnya dalam bidang Clustering Dokumen dan Peringkasan Dokumen. Tentunya tidak ada di dunia ini yang sempurna kecuali Allah Yang Maha Kuasa, demikian pula dengan penelitian ini. Peneliti senantiasa terbuka untuk kritik dan saran yang membangun.

Semarang, Oktober 2013

Ketua Tim Peneliti

Ardytha Luthfiarta, M.Kom, M.CS

# DAFTAR ISI

<b>RINGKASAN .....</b>	<b>III</b>
<b>PRAKATA.....</b>	<b>IV</b>
<b>DAFTAR ISI.....</b>	<b>V</b>
<b>DAFTAR TABEL .....</b>	<b>VII</b>
<b>DAFTAR GAMBAR.....</b>	<b>VIII</b>
<b>DAFTAR LAMPIRAN.....</b>	<b>IX</b>
<b>BAB 1. PENDAHULUAN .....</b>	<b>1</b>
1.1 LATAR BELAKANG MASALAH.....	1
1.2 RUMUSAN MASALAH .....	2
1.3 TUJUAN PENELITIAN.....	2
1.4 BATASAN MASALAH.....	2
1.5 LUARAN YANG DIHARAPKAN.....	2
<b>BAB 2. TINJAUAN PUSTAKA .....</b>	<b>3</b>
2.1 PENELITIAN TERKAIT.....	3
2.2 TEXT MINING .....	5
2.3 CLUSTERING DOKUMEN .....	6
2.4 PREPROCESSING .....	7
2.5 FEATURE-BASED AUTOMATIC SUMMARIZATION .....	7
2.6 LSA (LATENT SEMANTIC ANALYSIS).....	8
2.7 EVALUATION MEASURE .....	9
<b>BAB 3. TUJUAN DAN MANFAAT PENELITIAN .....</b>	<b>10</b>
3.1 TUJUAN PENELITIAN.....	10
3.2 MANFAAT PENELITIAN.....	10
<b>BAB 4. METODE PENELITIAN.....</b>	<b>11</b>
4.1 METODE PENELITIAN.....	11
4.2 MODEL YANG DIUSULKAN.....	11
4.3 DATASET.....	12
4.4 PREPROCESSING .....	12
4.5 EVALUASI.....	13
<b>BAB 5. HASIL YANG DICAPAI.....</b>	<b>14</b>
5.1 EKSPERIMEN.....	14
5.2 HASIL PENELITIAN.....	17

5.3	WAKTU EKSPERIMEN.....	18
5.4	TARGET DAN CAPAIAN PENELITIAN.....	19
5.5	HAMBATAN DAN PENANGANAN .....	21
<b>BAB 6. KESIMPULAN DAN SARAN.....</b>		<b>22</b>
<b>DAFTAR PUSTAKA.....</b>		<b>23</b>
<b>LAMPIRAN 1: TAMPILAN SAMPLE DATASET.....</b>		<b>25</b>
<b>LAMPIRAN 2 : SCRIPT CODING DI JAVA .....</b>		<b>26</b>
<b>LAMPIRAN 3 : PAPER SEMANTIK 2013 .....</b>		<b>27</b>
<b>LAMPIRAN 4 : LAPORAN PENGGUNAAN DANA .....</b>		<b>28</b>
<b>LAMPIRAN 5 : TANDA TERIMA HONOR PENELITI .....</b>		<b>32</b>
<b>LAMPIRAN 6 : KUITANSI DAN NOTA BAHAN HABIS PAKAI .....</b>		<b>33</b>
<b>LAMPIRAN 7 : KUITANSI DAN NOTA PERALATAN .....</b>		<b>33</b>
<b>LAMPIRAN 8 : KUITANSI DAN NOTA BIAYA KOMUNIKASI.....</b>		<b>35</b>
<b>LAMPIRAN 9: KUITANSI DAN NOTA KEBUTUHAN RAPAT KOORDINASI .....</b>		<b>36</b>
<b>LAMPIRAN 10: PELAPORAN DAN PUBLIKASI.....</b>		<b>37</b>

## DAFTAR TABEL

<b>TABEL 1. PENELITIAN TERKAIT .....</b>	<b>3</b>
<b>TABEL 2. TAHAPAN PERCOBAAN PENELITIAN.....</b>	<b>16</b>
<b>TABEL 3. TABEL HASIL PENELITIAN.....</b>	<b>18</b>
<b>TABEL 4. TARGET DAN CAPAIAN PENELITIAN .....</b>	<b>20</b>

## DAFTAR GAMBAR

<b>GAMBAR 1. MODEL YANG DIUSULKAN .....</b>	<b>11</b>
<b>GAMBAR 2. DOKUMEN ASLI SEBELUM PREPROCESSING .....</b>	<b>14</b>
<b>GAMBAR 3. HASIL DOKUMEN SETELAH PROSES PREPROCESSING .....</b>	<b>15</b>
<b>GAMBAR 4. PROSES PERINGKAS DOKUMEN OTOMATIS .....</b>	<b>16</b>
<b>GAMBAR 5. SAMPEL HASIL PROSES CLUSTERING DOKUMEN K-MEANS .....</b>	<b>17</b>
<b>GAMBAR 6. RATA-RATA HASIL KINERJA CLUSTERING DOKUMEN .....</b>	<b>18</b>
<b>GAMBAR 7. WAKTU PROSES CLUSTERING DOKUMEN DALAM DETIK .....</b>	<b>19</b>



## DAFTAR LAMPIRAN

<b>LAMPIRAN 1: TAMPILAN SAMPLE DATASET .....</b>	<b>25</b>
<b>LAMPIRAN 2 : SCRIPT CODING DI JAVA .....</b>	<b>26</b>
<b>LAMPIRAN 3 : PAPER SEMANTIK 2013 .....</b>	<b>27</b>
<b>LAMPIRAN 4 : LAPORAN PENGGUNAAN DANA .....</b>	<b>28</b>
<b>LAMPIRAN 5 : TANDA TERIMA HONOR PENELITI.....</b>	<b>32</b>
<b>LAMPIRAN 6 : KUITANSI DAN NOTA BAHAN HABIS PAKAI.....</b>	<b>33</b>
<b>LAMPIRAN 7 : KUITANSI DAN NOTA PERALATAN .....</b>	<b>33</b>
<b>LAMPIRAN 8 : KUITANSI DAN NOTA BIAYA KOMUNIKASI.....</b>	<b>35</b>
<b>LAMPIRAN 9: KUITANSI DAN NOTA KEBUTUHAN RAPAT KOORDINASI.....</b>	<b>36</b>
<b>LAMPIRAN 10: PELAPORAN DAN PUBLIKASI.....</b>	<b>37</b>

# **BAB 1. PENDAHULUAN**

## **1.1 Latar Belakang Masalah**

Teknologi pengklasteran dokumen memiliki peran yang signifikan dalam kemajuan teknologi informasi, diantaranya mempunyai peranan penting dalam pengembangan web di bidang akurasi kategorisasi keyword otomatis pada search engine, kategorisasi berita untuk surat kabar elektronik, peningkatan rating situs dengan teknologi Search Engine Optimization (SEO) dan sangat memungkinkan untuk diimplementasikan dalam berbagai teknologi informasi lainnya.

Oleh karena itu diperlukan penelitian untuk meningkatkan ketepatan akurasi dalam pengklasteran dokumen. Clustering dokumen adalah proses pengelompokan dokumen yang memiliki kesamaan topik, clustering dokumen memudahkan pengguna menemukan dokumen yang diinginkan [1].

Semakin banyaknya volume dokumen yang ada, dapat menyebabkan suatu permasalahan pada clustering dokumen yaitu besarnya matrik term-dokumen yang bisa menyebabkan proses kerja clustering dokumen tidak optimal. Hal ini bisa terjadi karena adanya data yang tidak relevan dan redundan. Oleh karena itu diperlukan suatu metode untuk bisa mengurangi dimensi dokumen tersebut sehingga bisa meningkatkan kinerja proses clustering dokumen tanpa mengurangi tingkat akurasi hasil clustering [3] [11].

Ringkasan dokumen dapat diartikan sebagai proses dari pembuatan intisari informasi terpenting dari sumber untuk menghasilkan versi yang lebih ringkas, terdapat dua tipe pembuatan suatu ringkasan yang mengambil bagian terpenting dari teks aslinya yaitu abstrak dan ekstrak. Abstrak menghasilkan sebuah interpretasi terhadap teks aslinya, dimana sebuah kalimat akan ditransformasikan menjadi kalimat yang lebih singkat, sedangkan ekstraksi merupakan ringkasan teks yang diperoleh dengan menyajikan kembali bagian tulisan yang dianggap topik utama tulisan dengan bentuk yang lebih disederhanakan [5].

Dalam penelitian ini akan digunakan fitur ringkasan ekstrak sebagai model peringkasan dokumen otomatis. Di dalam model peringkasan dokumen otomatis dapat digunakan algoritma Feature Based dan LSA (Latent Semantic Analysis) untuk proses reduksi kalimat. Penelitian yang sudah pernah dilakukan dengan menggunakan algoritma Feature Based dalam proses peringkasan dokumen otomatis sebagai feature reduction untuk proses clustering dokumen dihasilkan tingkat akurasi yang lebih baik dibandingkan dengan proses clustering menggunakan teknik feature reduction standar [18]. Peringkasan Dokumen menggunakan Algoritma LSA (Latent Semantic Analysis) diharapkan dapat melakukan proses reduksi kalimat dengan baik dibandingkan algoritma Feature Based sehingga dapat lebih meningkatkan akurasi proses clustering dokumen.

## 1.2 Rumusan Masalah

Dari latar belakang di atas, dapat disimpulkan bahwa dengan adanya pengaruh besarnya matrik term-dokumen terhadap kinerja dan akurasi proses clustering dokumen, diusulkan metode integrasi peringkasan dokumen otomatis dengan LSA (Latent Semantic Analysis) untuk proses clustering dokumen.

Sedangkan pertanyaan penelitian (research question) yang diangkat dalam penelitian ini adalah “Bagaimana pengaruh integrasi peringkasan dokumen otomatis dengan algoritma LSA dibandingkan dengan algoritma Feature based dalam meningkatkan kinerja tingkat akurasi clustering dokumen?”

## 1.3 Tujuan Penelitian

Berdasarkan latar belakang dan rumusan masalah di atas, maka dapat dirumuskan tujuan penelitian sebagai berikut :

Meningkatkan akurasi dari clustering dokumen dengan mengintegrasikan peringkasan dokumen otomatis menggunakan algoritma LSA (Latent Semantic Analysis).

## 1.4 Batasan Masalah

Agar fokus terhadap tujuan penelitian dan dapat diselesaikan sesuai dengan jadwal penelitian, maka perlu ada batasan masalah diantaranya adalah sebagai berikut:

- a. Area penelitian ini adalah pada pengembangan teknik clustering dokumen. Lebih khususnya pada proses peringkasan dokumen otomatis untuk reduksi kalimat.
- b. Hasil akhir yang dikembangkan dalam penelitian ini berupa model dan bukan merupakan *application*, sehingga penelitian ini hanya berfokus pada model yang menghasilkan akurasi lebih baik untuk proses clustering dokumen.

## 1.5 Luaran Yang Diharapkan

Penelitian ini memiliki target capaian luaran kegiatan berupa publikasi ilmiah dalam seminar atau conference nasional.

## BAB 2. TINJAUAN PUSTAKA

### 2.1 Penelitian Terkait

**Tabel 1. Penelitian terkait**

1.	<p>Judul : Improving the Performance of Text Categorization using Automatic Summarization [8].</p> <ul style="list-style-type: none"> <li>JIANG Xiao-Yu, FAN Xiao-Zhong, Wang Zhi-Fei and Jia Ke-Liang, <i>International Conference on Computer Modeling and Simulation IEEE 2009</i>.</li> </ul>		
	<b>DESCRIPTION</b>	<b>MAIN FEATURE</b>	<b>OUTCOME &amp; CONCLUSIONS</b>
	<p>Penelitian metode untuk meningkatkan kinerja <b>klasifikasi</b> dokumen.</p>	<p>Mengajukan 2 metode, yang pertama adalah peringkasan otomatis sebagai feature selection untuk pengklasifikasian dokumen, dan metode kedua adalah metode seleksi dan pembobotan untuk masing-masing hasil ringkasan dokumen, dan kemudian mengklasifikasi teks bebas menggunakan algoritma KNN.</p>	<p>Hasil penelitian membuktikan metode pertama memiliki kinerja yang lebih unggul, peringkasan dokumen masih tetap mempertahankan kata dan kalimat inti dari isi dokumen, dan dapat mengurangi volume dokumen untuk tahap proses klasifikasi, sehingga mempercepat kinerja, dan akurasi hasil klasifikasi masih tetap terjaga.</p>
2.	<p>Judul : A Feature Selection Algorithm for Document Clustering based on Word co-occurrence Frequency [9].</p> <ul style="list-style-type: none"> <li>Yuan-Chao Liu, Xiao-Long Wang &amp; Bing-Quan Liu , <i>Proceedings of the Third International Conference on Machine Learning and Cybernetics, Shanghai, 26-29 August 2004</i>.</li> </ul>		
	<b>DESCRIPTION</b>	<b>MAIN FEATURE</b>	<b>OUTCOME &amp; CONCLUSIONS</b>

	Dalam penelitian ini membahas model feature selection pada proses clustering dokumen.	Mengevaluasi dampak dari feature selection berdasarkan frekuensi kemunculan kata yang sama untuk proses clustering dokumen.	Dari hasil analisa pada hipotesis clustering bahwa dokumen dari kelas yang sama akan mirip satu sama lain ketika direpresentasikan dalam model Vector Space Model (VSM), dari hasil feature selection pada prinsipnya adalah untuk mengurangi dimensi pada saat proses clustering sehingga mempercepat kinerja dan akurasi hasil clustering tetap terjamin.
3.	Judul : An Evaluation on Feature Selection for Text Clustering [4]. <ul style="list-style-type: none"> <li>Tao Liu, Shengping Liu, Zheng Chen &amp; Wei-Ying Ma, <i>Proceedings of the Twentieth International Conference on Machine Learning (ICML-2003)</i>, Washington DC, 2003.</li> </ul>		
	<b>DESCRIPTION</b>	<b>MAIN FEATURE</b>	<b>OUTCOME &amp; CONCLUSIONS</b>
	Evaluasi metode unsupervised feature selection dalam proses dokumen clustering.	Dalam penelitian ini membandingkan beberapa metode unsupervised feature selection diantaranya DF, TS, En dan metode baru yang diusulkan yaitu TC.	Hasil yang diperoleh membuktikan bahwa TC dan TS lebih unggul daripada DF dan En, kemudian TC lebih disukai sebagai model unsupervised feature selection pada dokumen clustering.
4.	Judul : Automatic Text Summarization Using Feature Based Fuzzy Extraction [5]. <ul style="list-style-type: none"> <li>Ladda Suanmali, Naomie Salim &amp; M Salem Binwahlan, <i>Jurnal teknologi Maklumat jilid 20. Bil 2, 2008.</i></li> </ul>		
	<b>DESCRIPTION</b>	<b>MAIN FEATURE</b>	<b>OUTCOME &amp; CONCLUSIONS</b>
	Peringkasan dokumen otomatis menggunakan metode fuzzy logic.	Metode fuzzy logic dalam peringkasan dokumen menggunakan proses ekstraksi kalimat-kalimat yang dianggap mempunyai bobot dan penting, terdapat 8 fitur untuk penilaian yaitu : title feature, sentence length, term weight, sentence position, sentence to sentence similarity, proper noun, thematic word dan numerical data.	Hasil pengujian menunjukkan rata-rata ketelitian, recall dan f-mean yang lebih baik pada suatu ringkasan dokumen single

5.	Judul : A Comparative Study on Unsupervised Feature Selection Methods for Text Clustering [2]. <ul style="list-style-type: none"> <li>Luying LIU, Jianchu KANG, Jing YU &amp; Zhongliang WANG, <i>Natural Language Processing and Knowledge Engineering, 2005. IEEE NLP-KE '05. Proceedings of 2005 IEEE International Conference on..</i></li> </ul>		
	<b>DESCRIPTION</b>	<b>MAIN FEATURE</b>	<b>OUTCOME &amp; CONCLUSIONS</b>
	Studi perbandingan unsupervised feature selection.	Membandingkan beberapa metode unsupervised feature selection diantaranya metode DF, TC, TVQ dan TV.	Hasil ujicoba menyimpulkan bahwa pada kasus unsupervised feature selection untuk clustering dokumen metode TC dan TV lebih baik daripada

			menggunakan metode DF dan TVQ dalam hal kinerja dan akurasi.
6.	Judul : Integrating Feature-Based Document Summarization as Feature Reduction in Document Clustering[18]. <ul style="list-style-type: none"> <li>Catur Supriyanto, Abu Salam &amp; Abdul Syukur, <i>July 12,2012, Yogyakarta,Indonesia, CITEE 2012.</i></li> </ul>		
	<b>DESCRIPTION</b>	<b>MAIN FEATURE</b>	<b>OUTCOME &amp; CONCLUSIONS</b>
	Teknik peringkas dokumen otomatis sebagai fitur reduction pada clustering dokumen.	Mengusulkan teknik peringkas dokumen otomatis sebagai feature reduction dalam proses clustering dokumen.	Hasil ujicoba mendapatkan kenaikan tingkat akurasi yang signifikan dibandingkan tingkat clustering standard.

## 2.2 Text Mining

Text mining merupakan salah satu bidang khusus dari ilmu data mining. Perbedaan yang mendasar terdapat dalam sumber data, dalam data mining menggunakan data yang terstruktur, sedangkan dalam text mining memproses data yang tidak terstruktur. Tujuan dari text mining adalah untuk menemukan pola yang berguna dalam sebuah koleksi dokumen text. [12]. Dengan text mining informasi pada sebuah koleksi dokumen dapat diekstraksi, diklusterisasi, diklasifikasi serta dianalisis untuk tujuan tertentu.

## 2.3 Clustering Dokumen

Clustering dokumen adalah proses pengelompokan dokumen yang memiliki kesamaan topik [1]. Tujuan clustering dokumen adalah mempartisi dokumen menjadi beberapa kelompok dimana dokumen dalam kelompok yang sama adalah mempunyai kemiripan satu sama lain berdasarkan frekuensi kemunculan term [13][16]. Teknik clustering dokumen ini merupakan teknik yang lebih spesifik dari pengorganisasian unsupervised dokumen, ekstraksi topik otomatis serta pengambilan dan penyaringan informasi secara cepat, clustering dokumen juga telah digunakan secara otomatis untuk menghasilkan cluster hirarki dokumen. Sebagai contoh, metode clustering dapat digunakan oleh web search engine untuk mengelompokkan secara otomatis dokumen pada sebuah daftar kategori yang memudahkan untuk memperoleh keterkaitan informasi yang relevan, sebagaimana yang telah diterapkan oleh Enterprise search engine seperti Northem Light dan Vivisimo atau perangkat lunak opensource seperti Carrot2.

Beberapa tahapan dasar sebelum clustering dokumen dilakukan, yaitu preprocessing, term weighting baru kemudian penerapan algoritma clustering. Tahapan preprocessing akan menghasilkan kumpulan term atau kata yang nantinya akan diberikan bobot atau nilai dimana bobot tersebut mengindikasikan pentingnya sebuah term terhadap dokumen. Semakin banyak term tersebut muncul pada koleksi dokumen, semakin tinggi nilai atau bobot term tersebut. Pemberian bobot tersebut dinamakan term weighting. Setelah tahapan pemberian bobot selesai, maka akan dihasilkan sebuah matrik term-document dengan dimensi  $m \times n$ , dimana  $m$  adalah jumlah term dan  $n$  adalah jumlah dokumen. Model sistem temu kembali informasi seperti ini disebut model ruang vektor atau vector space model (VSM).

Algoritma yang umum digunakan pada proses clustering dokumen dan juga yang akan digunakan dalam penelitian adalah algoritma K-means, Prinsip utama dari teknik ini adalah menyusun  $k$  buah prototype / pusat massa (centroid) / rata-rata (mean) dari sekumpulan data berdimensi  $n$ . Teknik ini mensyaratkan nilai  $k$  sudah diketahui sebelumnya (a priori). Algoritma k-means dimulai dengan pembentukan prototipe cluster di awal kemudian secara iteratif prototipe cluster ini diperbaiki hingga konvergen (tidak terjadi perubahan yang signifikan pada prototipe cluster). Dasar algoritma K-means dapat disusun menjadi 4 tahapan sebagai berikut :

1. Inisialisasi titik pusat Cluster
2. Masukkan setiap dokumen ke cluster yang paling cocok berdasarkan ukuran kedekatan dengan centroid, dalam penelitian ini menggunakan cosines similiarity. Centroid adalah vektor term yang dianggap sebagai titik tengah cluster.
3. Setelah semua dokumen masuk ke cluster. Hitung ulang centroid cluster berdasarkan dokumen yang berada di dalam cluster tersebut.
4. Jika centroid tidak berubah (dengan treshold tertentu) maka stop. Jika tidak, kembali ke langkah 2.

$$\text{Sim}(d_x, d_y) = \frac{\sum_{k=1}^n x_k \times y_k}{\sqrt{\sum_{k=1}^n x_k^2} \times \sqrt{\sum_{k=1}^n y_k^2}}$$

Dimana k adalah nilai bobot setiap term dalam dokumen. Proses dimulai dengan mula-mula memilih secara random k buah dokumen sebagai pusat cluster awal.

## 2.4 Preprocessing

Preprocessing merupakan tahapan untuk mengubah struktur isi dari suatu dokumen kedalam format yang sesuai berupa kumpulan term atau kata untuk diproses kedalam algoritma clustering [5]. Preprocessing biasanya dilakukan dengan menghilangkan term yang tidak signifikan (stopword) dan proses stemming [14], dalam penelitian ini ada 4 tahapan preprocessing yang akan di gunakan yaitu: Tokenization, Stopword, dan Stemming.

## 2.5 Feature-Based Automatic Summarization

Ringkasan dokumen merupakan proses dari pembuatan intisari informasi terpenting dari sumber untuk menghasilkan versi yang lebih ringkas [5][17], terdapat dua tipe pembuatan suatu ringkasan yang mengambil bagian terpenting dari teks aslinya yaitu abstrak dan ekstrak. Abstrak menghasilkan sebuah interpretasi terhadap teks aslinya, dimana sebuah kalimat akan ditransformasikan menjadi kalimat yang lebih singkat, sedangkan ekstrak merupakan ringkasan teks yang diperoleh dengan menyajikan kembali bagian tulisan yang dianggap topik utama tulisan dengan bentuk yang lebih disederhanakan. Dalam penelitian ini akan digunakan fitur ringkasan ekstrak.

6 fitur tahapan yang digunakan dalam penelitian ini adalah sebagai berikut:

### 1. Sentence length

Fitur ini berguna untuk menyaring kalimat pendek seperti datelines dan nama penulis yang biasa ditemukan pada artikel-artikel berita, dimana kalimat pendek tersebut tidak diharapkan muncul pada ringkasan dokumen. Adalah rasio dari jumlah kata dalam kalimat dengan jumlah kata yang terdapat pada kalimat terpanjang pada suatu dokumen.

$$\text{score}(S) = \frac{\text{No. Word occuring in } S}{\text{No. Word occuring in longest sentence}}$$

### 2. Term weight

Menghitung frekuensi munculnya sebuah term pada dokumen yang biasa digunakan untuk menentukan penting tidaknya posisi kalimat pada sebuah dokumen. Perhitungan rata-rata TF-ISF (Term Frequency, Inverse Sentence Frequency) adalah sebagai berikut:



$$score(S) = \frac{Sum\ of\ TF-IFS\ in\ S}{Max(Sum\ of\ TF-IFS)}$$

### 3. Sentence position

Apakah letak suatu kalimat ada pada akhir atau awal suatu paragraf dalam dokumen, karena hal ini mempengaruhi skor dari suatu kalimat dalam dokumen, skor 1 untuk kalimat pertama atau akhir dan 0 untuk kalimat yang lainnya.

$$score(s) = 1\ for\ first\ and\ last\ sentence.  
0\ for\ other\ sentences.$$

### 4. Sentence to sentence similiarity

Kesamaan antar kalimat, dimisalkan kalimat s, pengukuran kesamaan antara kalimat s dengan kalimat lainnya dengan menghitung rasio dari ringkasan kesamaan kalimat pada kalimat s tersebut dengan maksimum ringkasan jumlah dari keseluruhan kesamaan kalimat pada dokumen.

$$score(S) = \frac{Sum\ of\ Sentence\ Similiarity\ in\ S}{Max(Sum\ of\ sentence\ Similiarity)}$$

### 5. Thematic word

Thematic Word merupakan jumlah kata tematik yang ada dalam kalimat. Fitur ini penting karena term yang ada dalam dokumen sering terkait dengan topic suatu dokumen. Jumlah kata tematik menunjukkan kata-kata dengan relativitas maksimal. Skor dalam fitur ini dihitung dari rasio jumlah kata tematik dalam kalimat dengan panjang kalimat dalam dokumen.

$$score(s) = \frac{No.Thematic\ word\ in\ S}{Length(S)}$$

### 6. Numerical data

Adalah jumlah data numerik yang ada dalam kalimat, kedudukan kalimat yang mengandung data numeric adalah penting karena dimungkinkan akan masuk kedalam isi ringkasan dokumen [5]. Skor untuk fitur ini merupakan rasio dari jumlah kata numerik pada kalimat dengan panjang kalimat dalam dokumen.

$$score(s) = \frac{No.Numerical\ data\ in\ S}{Length(S)}$$

## 2.6 LSA (Latent Semantic Analysis)

Latent Semantic Analysis adalah salah satu teknik Information Retrieval yang terdiri dari algoritma matematika yang diterapkan untuk koleksi teks. Inisialisasi koleksi teks itu sebelum diproses dan direpresentasikan sebagai matrik term-by-file yang berisi term dan jumlah frekuensinya di dalam file. Transformasi matriks yang diterapkan sedemikian rupa sehingga nilai-nilai dari istilah dalam file disesuaikan tergantung pada seberapa sering mereka muncul di dalam dan diseluruh file didalam koleksi data.

Algoritma matematika yang disebut Singular Value Decomposition (SVD), terurai di dalam matriks term-by-file menjadi matriks terpisah yang menangkap kesamaan antara term dan diantara file di berbagai dimensi. Tujuannya adalah untuk mewakili hubungan asli antara term dalam mengurangi dimensi ruang kebisingan tersebut yang dihapus dari data dan oleh karena itu mengungkap hubungan penting antara term dan dokumen. LSA bertujuan untuk

menemukan pokok (laten) hubungan antara istilah yang berbeda yang memiliki arti yang sama tetapi tidak pernah terjadi pada file yang sama. Dalam konteks tekstual Information Retrieval akun noise untuk vaiabilitas penggunaan term.

## 2.7 Evaluation Measure

Dalam penelitian ini, digunakan F-measure untuk mengukur kinerja clustering. F-measure diperoleh dari pengukuran recall dan precision. Recall adalah rasio dokumen yang relevan yang terambil dengan jumlah seluruh dokumen dalam koleksi dokumen, sedangkan precision adalah rasio jumlah dokumen relevan terambil dengan seluruh jumlah dokumen terambil. Nilai recall dan precision yang tinggi menunjukkan tingkat keakuratan dari sebuah clustering [6][15]. Recall dan precision kategori  $i$  dalam cluster  $j$  diperoleh dari persamaan berikut:

$$Recall(i,j) = \frac{n_{ij}}{n_i}$$

$$Precision(i,j) = \frac{n_{ij}}{n_j}$$

Dinama  $n_{ij}$  merupakan jumlah dokumen kategori  $i$  dalam cluster  $j$ ,  $n_i$  adalah jumlah dokumen dalam kategori  $i$  dan  $n_j$  merupakan jumlah dokumen dalam cluster  $j$ . kemudian untuk menghitung F-measure yang digunakan adalah persamaan berikut:

$$F(i,j) = \frac{2 * (Precision * Recall)}{(Precision + Recall)}$$

Secara keseluruhan, rata-rata dari F-measure dapat dihitung dengan persamaan berikut:

$$F = \sum_i \frac{n_i}{n} \max_{j=1, \dots, k} F(i,j)$$

Dimana  $\max\{f(i,j)\}$  adalah nilai maksimum F-measure dari kategori  $i$  dalam cluster  $j$ . Semakin tinggi nilai F-measure semakin tinggi tingkat akurasi dari cluster [7].

## **BAB 3. TUJUAN DAN MANFAAT PENELITIAN**

### **3.1 Tujuan Penelitian**

Penelitian ini bertujuan untuk meningkatkan akurasi hasil clustering dokumen dengan mengintegrasikan teknik peringkasan dokumen otomatis menggunakan algoritma *Latent Semantic Analysis (LSA)* dibandingkan dengan teknik peringkasan dokumen dalam penelitian-penelitian sebelumnya, diantaranya adalah dengan algoritma *Feature Based*.

### **3.2 Manfaat Penelitian**

Penelitian ini diharapkan akan memberikan beberapa manfaat sebagai berikut :

#### **a. Manfaat Praktis**

Hasil penelitian ini diharapkan dapat digunakan sebagai teknik pengintegrasian peringkasan dokumen otomatis untuk proses clustering dokumen yang nantinya dapat digunakan dalam proses kategorisasi keyword otomatis pada search engine, kategorisasi berita untuk surat kabar elektronik, peningkatan rating situs dengan teknologi Search Engine Optimization (SEO) dan dapat diimplementasikan dalam berbagai teknologi informasi lainnya.

#### **b. Manfaat Akademis**

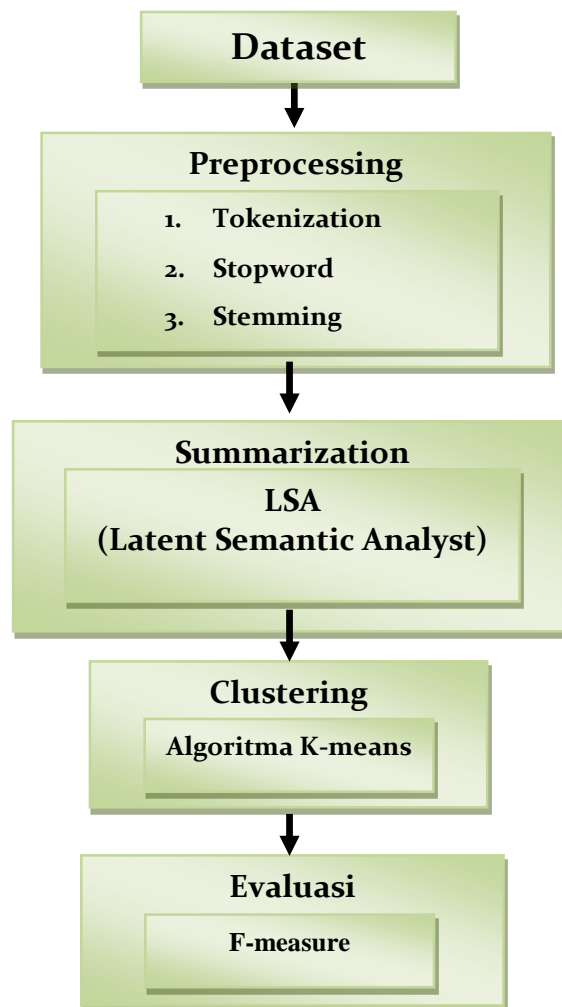
Penelitian ini memiliki luaran penelitian berupa makalah yang akan dipublikasikan melalui Seminar Nasional. Hasil publikasi dalam penelitian ini dapat memberikan kontribusi yaitu tambahan kahzanah ilmu pengetahuan dalam bidang teknologi informasi, khususnya dalam bidang clustering dokumen.

## BAB 4. METODE PENELITIAN

### 4.1 Metode Penelitian

Metode penelitian merupakan pendekatan secara konsisten untuk memecahkan suatu masalah di dalam penelitian. Ini dapat diartikan sebagai ilmu yang mempelajari bagaimana penelitian dilakukan secara teknis. Di dalam penelitian ini menggunakan metode penelitian eksperimen. Pada bagian ini akan membahas tentang rincian metode penelitian yang dilakukan meliputi dataset, preprocessing, summarization, clustering dan evaluasi.

### 4.2 Model yang diusulkan



Gambar 1. Model yang diusulkan

Algoritma yang diusulkan akan diimplementasikan secara umum dengan menggunakan pemrograman JAVA. Sistem akan menggunakan *Lucene3* sebagai *java library*. *Lucene* menyediakan fungsi untuk *stopword removal* dan *stemming* untuk tahapan *preprocessing*. *Lucene* juga menyediakan perhitungan pembobotan dengan metode *Term Frequency Invers Document Frequency* (TFIDF) dan perhitungan *cosines similarity* untuk menghitung kemiripan antar dokumen, selain itu *lucene* secara luas sudah diakui dalam penggunaannya untuk mesin pencari dan situs pencarian. Keunggulan lainnya adalah *lucene* merupakan software library yang open source.

### 4.3 Dataset

Penelitian ini memakai data yang berasal dari situs portal berita yahoo news Indonesia, jumlah dataset test sebanyak 150 dokumen berita berbahasa indonesia dari 5 kategori berita yaitu: Sport, Ekonomi, Hukum, Kriminal, dan Politik. Dataset tersebut di-transformasi untuk mendapatkan atribut yang relevan dan sesuai dengan format input algoritma clustering dokumen.

### 4.4 Preprocessing

Di dalam penelitian ini menggunakan 3 tahapan preprocessing yang akan di gunakan yaitu: Tokenization, Stopword, dan Stemming.

#### a. Tokenization

Tahap tokenizing adalah tahap pemotongan string input berdasarkan tiap kata yang menyusunnya, contoh dari tahapan ini adalah sebagai berikut :

**Tekts Input** : “Belajar membaca buku”.

**Hasil Token** :

Belajar
membaca
buku

## b. Stopword

Dalam tahap stopwords, kata-kata yang tidak relevan dalam penentuan topic sebuah dokumen akan dihilangkan, misal kata “adalah”, “dari”, “sebuah”, “atau” dan lain-lain dalam dokumen bahasa Indonesia.

## c. Stemming

Stemming merupakan tahap mencari root kata / kata dasar dari tiap kata hasil filtering, contoh dari tahap ini adalah sebagai berikut :

**Hasil Filter** :

Belajar  
membaca  
buku

**Hasil Stemming** :

ajar  
baca  
buku

## 4.5 Evaluasi

Evaluasi dilakukan dengan mengamati hasil clustering dari pengujian metode yang diusulkan dengan algoritma LSA (*Latent Semantic Analysis*). Dalam penelitian ini, digunakan F-measure untuk mengukur kinerja *clustering*. F-measure diperoleh dari pengukuran *recall* dan *precision*. *Recall* adalah rasio dokumen yang relevan yang terambil dengan jumlah seluruh dokumen dalam koleksi dokumen, sedangkan *precision* adalah rasio jumlah dokumen relevan terambil dengan seluruh jumlah dokumen terambil. Validasi hasil dengan membandingkan hasil evaluasi metode yang diusulkan.

## BAB 5. HASIL YANG DICAPAI

### 5.1 Eksperimen

Dalam penelitian ini dilakukan proses dari dataset yang sudah dikumpulkan untuk menguji kinerja dan akurasi dari desain yang diusulkan, pengukuran kinerja dari proses clustering dengan menggunakan metode peringkas dokumen otomatis sebagai algoritma pengurangan matrik term dokumen, kemudian hasil yang diperoleh dievaluasi untuk dibandingkan dengan rangkain proses clustering dokumen tanpa peringkas dokumen otomatis.

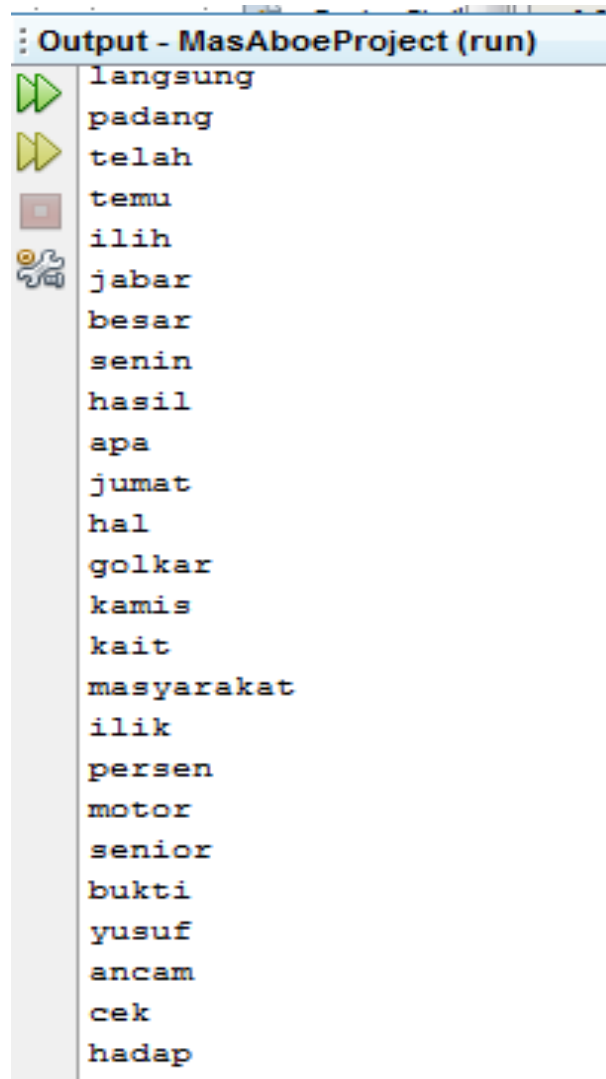
Ada 3 kategori model clustering pada percobaan yang dilakukan, yaitu:

- Model clustering standar
- Model clustering dengan menggunakan peringkas dokumen otomatis dengan metode LSA.
- Model clustering menggunakan peringkas dokumen otomatis dengan metode Fitur

Urutan langkah pada clustering standar pada percobaan yang dilakukan adalah: Tokenization, penghapusan Stopword, Stemming, dan proses k-means untuk tahap clustering dokumen. Berikutnya urutan langkah pada clustering menggunakan peringkas dokumen otomatis adalah: Tokenization, penghapusan Stopword, Stemming, proses peringkas dokumen otomatis dan proses k-means untuk tahap clustering dokumen. Gambar 2. menunjukkan dokumen asli sebelum preprocessing dan gambar 3 menunjukkan dokumen setelah tahap preprocessing.

```
<datasets>
<docID>SP002</docID>
<category>sport</category>
<title>Pelita Jaya Laporkan Masalah Diego ke FIFA dan AFC</title>
<content>Keputusan PSSI yang mengambil Diego Michiels secara sepihak membuat kubu Pelita Jaya FC berang dan akan melaporkan hal ini kepada FIFA dan AFC agar segera terselesaikan. Hal ini disampaikan oleh Manager Pelita Jaya, Lalu Mara kepada Sekjen Komite Penyelamat Sepakbola Indonesia (KPSI) di Kantor KPSI, Senayan, Jakarta, Senin (02/01) sore. Lalu mengungkapkan jika perbuatan PSSI yang mengancam Diego serta empat pemain naturalisasi dinilai tidak etis. "Saya mendapatkan laporan mengenai pengunduran Diego melalui Pak Iman Arif. Kemudian saya menghubungi Ketum Pelita Jaya Gunawan Tamzil untuk bertanya soal kebenaran surat tersebut dan ia mengatakan benar. Pak Gunawan bilang jika surat tersebut disiapkan oleh Penanggung Jawab timnas Bernard Limbong," ujarnya. Pelita Jaya mengontrak Diego Michiels selama tiga tahun mulai Februari 2011 hingga 2014. Lalu juga mengatakan dalam surat kontrak tersebut jika pemain ingin mengakhiri kontrak harus ditujukan kepada PT Liga Indonesia sebagai administrator Liga Super Indonesia. "Saya akan membawa masalah ini kepada FIFA dan AFC karena PSSI sudah melakukan tidak lazim karena telah mencampuri urusan kontrak yang sangat sakral antara dua pihak (Diego dan Pelita Jaya)," lanjutnya. Selain Diego, empat pemain naturalisasi lainnya yaitu Ruben Wuarbanaran, Greg Nwokolo, Victor Igbonefo dihubungi oleh PSSI dengan diikuti dengan ancaman. PSSI juga menjanjikan mereka (pemain naturalisasi) akan bermain di timnas Indonesia jika keluar dari Pelita Jaya. "Semua pemain dihubungi oleh PSSI dan diancam, namun hanya Diego yang memenuhi panggilan tersebut. PSSI juga menjanjikan kepada pemain naturalisasi tersebut bisa bermain di timnas Indonesia. Saya pikir pengurus PSSI sekarang sudah tidak lagi kredibel karena telah melakukan tindakan yang tidak etis," tutupnya. Sementara itu, Sekjen KPSI Hince Panjaitan akan segera membuat laporan mengenai empat orang yang dianggap sebagai biang keladi dalam masalah ini kepada FIFA dan AFC agar langsung adanya investigasi soal hal ini. Dalam surat pengaduan Pelita Jaya itu ada empat orang yang diadakan kepada FIFA dan AFC melalui KPSI. Mereka ialah Ketum PSSI Djoehar Arifin Husin, Ketua Kompetisi Sihar Sitorus, Penanggung Jawab timnas Bernard Limbong, dan Koordiantor timnas Bob Hippy.
</content>
</datasets>
```

Gambar 2. Dokumen asli sebelum preprocessing

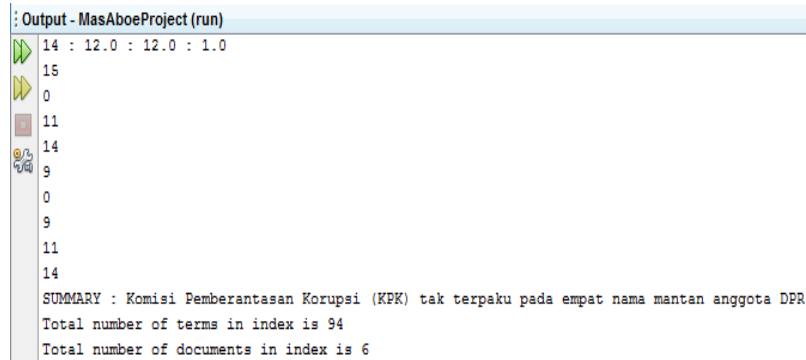


**Gambar 3. Hasil dokumen setelah proses preprocessing**

Setelah proses tokenization, stopwords dan stemming selesai, selanjutnya dilakukan proses pemenggalan kalimat baru kemudian tahap berikutnya adalah proses peringkasan dokumen otomatis. Dalam penelitian peringkasan dokumen otomatis menggunakan metode peringkasan Fitur dan LSA, metode peringkasan fitur merupakan ringkasan teks yang diperoleh dengan menyajikan kembali bagian tulisan yang dianggap topik utama tulisan dengan bentuk yang lebih disederhanakan. Terdapat 7 fitur tahapan yang digunakan dalam penelitian ini, yaitu: Fitur Judul, Panjang Kalimat, Bobot kata, Posisi kalimat, Kesamaan antar kalimat, Kata tematik, Data numerik. Sedangkan metode peringkasan LSA menggunakan SVD, dimana SVD digunakan untuk mengurangi matrik term dokumen ke dalam matrik yang mempunyai dimensi lebih rendah. Truncated SVD menggunakan pendekatan peringkat- $k$  untuk mengurangi SVD.  $Peringkat-k \leq \min(m, n)$ , dimana  $m$  adalah jumlah dokumen dan  $n$  adalah jumlah term.



Gambar 4. menunjukkan sampel hasil peringkas dokumen otomatis pada salah satu dokumen.



**Gambar 4. Proses peringkas dokumen otomatis**

Setelah dihasilkan dokumen dalam bentuk summary dokumen direpresentasikan kedalam bentuk matrik term-dokumen  $m \times n$  menggunakan Vector Space Model. Proses selanjutnya adalah proses algoritma clustering k-means, untuk titik pusat cluster dilakukan acak, sehingga dalam proses penelitian dilakukan 5 kali tahapan clustering untuk 1 langkah rangkaian algoritma.

**Tabel 2. tahapan percobaan penelitian**

Metode	Hasil 1	Hasil 2	Hasil 3	Hasil 4	Hasil 5
Original K-Means	Clustering	Clustering	Clustering	Clustering	Clustering
Feature Summary 20%	Clustering	Clustering	Clustering	Clustering	Clustering
Feature Summary 40%	Clustering	Clustering	Clustering	Clustering	Clustering
Feature Summary 60%	Clustering	Clustering	Clustering	Clustering	Clustering
Feature Summary 80%	Clustering	Clustering	Clustering	Clustering	Clustering
LSA Summary 20%	Clustering	Clustering	Clustering	Clustering	Clustering
LSA Summary 40%	Clustering	Clustering	Clustering	Clustering	Clustering
LSA Summary 60%	Clustering	Clustering	Clustering	Clustering	Clustering
LSA Summary 80%	Clustering	Clustering	Clustering	Clustering	Clustering

Dari gambar 4 diatas proses penelitian menjalankan 9 langkah percobaan yang dilakukan untuk mengukur tingkat akurasi kinerja clustering, masing-masing tahapan dilakukan uji coba sebanyak 5 kali eksekusi clustering, yang masing-masing mengindikasikan titik pusat clustering / centroid yang berbeda-beda.

Gambar 5 menunjukkan sampel hasil cluster dari proses clustering dokumen yang dilakukan.

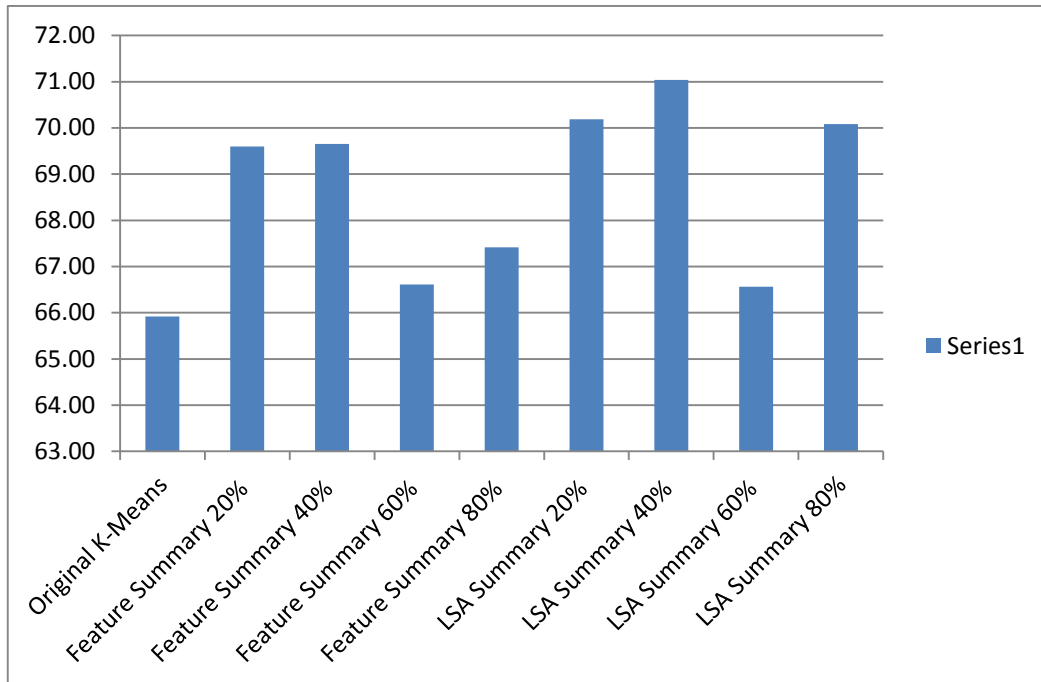
```
Output - MasAboeProject (run)
=== Clusters from K-Means Algorithm ===
sport:[SP001, SP002, SP004, SP006, SP007, SP008, SP009, SP010, SP011, SP012, SP
economy:[EC001, EC002, EC003, EC004, EC005, EC006, EC007, EC008, EC009, EC010,
hukum:[EC024, HK001, HK002, HK003, HK004, HK005, HK006, HK007, HK008, HK009, HK
kriminal:[SP003, HK025, HK029, KR001, KR003, KR004, KR005, KR006, KR007, KR008,
politik:[SP005, HK014, HK023, KR011, KR016, KR019, KR020, KR023, KR024, KR025,
```

Gambar 5. Sampel hasil proses clustering dokumen k-means

## 5.2 Hasil Penelitian

Dari hasil penelitian yang dilakukan dapat dibuktikan bahwa rata-rata hasil proses clustering dokumen menggunakan peringkasan dokumen otomatis dengan metode Latent Semantic Analysis (LSA) dapat meningkatkan akurasi hasil clustering paling baik pada dokumen teks berbahasa Indonesia.

Tingkat akurasi rata-rata tertinggi diperoleh pada saat menggunakan peringkasan dokumen otomatis dengan metode LSA mencapai 71,04 % yang diperoleh pada tingkat peringkasan dokumen otomatis LSA 40% dibandingkan dengan tanpa menggunakan peringkasan dokumen otomatis yang hanya mencapai rata-rata tingkat akurasi 65,92 %, dari gambar di atas juga dapat dilihat hasil rata-rata proses clustering dokumen dengan menggunakan teknik peringkasan dokumen otomatis secara keseluruhan mengalami peningkatan kinerja dari pada kinerja clustering dokumen tanpa menggunakan teknik peringkasan dokumen otomatis. Hasil penelitian lebih lengkap dapat dilihat pada gambar 6 rata-rata hasil kinerja clustering dokumen dan table dibawah ini:



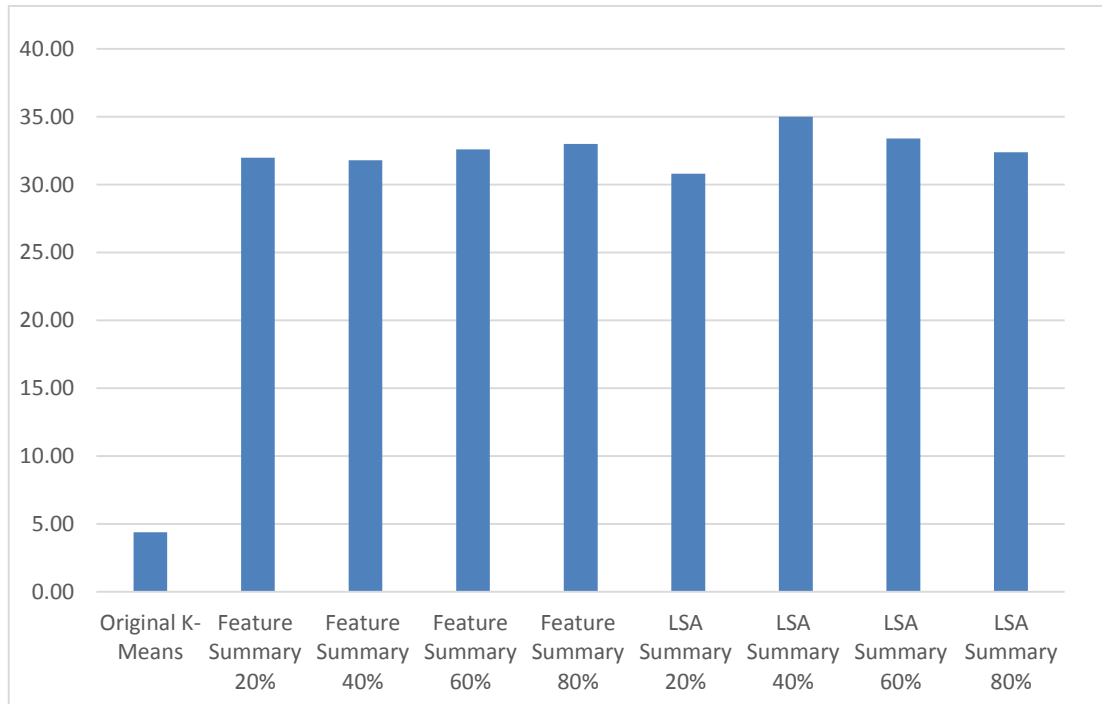
**Gambar 6. rata-rata hasil kinerja clustering dokumen**

**Tabel 3. tabel hasil penelitian**

Metode	F-Measure	F-Measure	F-Measure	F-Measure	F-Measure	Rata-rata	%
Original K-Means	0.62400	0.68800	0.65333	0.60533	0.72533	0.65920	65.92
Feature base 20%	0.67733	0.81867	0.67200	0.66400	0.64800	0.69600	69.60
Feature base 40%	0.71733	0.79200	0.67467	0.61333	0.68533	0.69653	69.65
Feature base 60%	0.65867	0.74400	0.60800	0.65067	0.66933	0.66613	66.61
Feature base 80%	0.71200	0.67467	0.63733	0.65600	0.69067	0.67413	67.41
LSA 20%	0.71200	0.68000	0.68800	0.74933	0.68000	0.70187	70.19
LSA 40%	0.70400	0.66933	0.76267	0.71200	0.70400	0.71040	71.04
LSA 60%	0.67733	0.65333	0.63200	0.72800	0.63733	0.66560	66.56
LSA 80%	0.82667	0.66667	0.70667	0.63200	0.67200	0.70080	70.08

### 5.3 Waktu Eksperimen

Waktu rata-rata yang diambil diukur mulai dari proses preprocessing sampai dengan hasil clustering diperoleh.



**Gambar 7. Waktu proses clustering dokumen dalam detik**

Gambar 7 menunjukkan bahwa adanya tambahan proses summary pada proses clustering dokumen mempengaruhi keseluruhan waktu proses dengan cukup signifikan jika dibandingkan dengan hanya menggunakan Original K-Means tanpa summary. Akan tetapi jika dibandingkan waktu proses antara Feature Summary dengan LSA Summary tidak terjadi perbedaan waktu yang signifikan.

Dengan demikian jika dibandingkan antara Feature Summary dengan LSA Summary dimana waktu proses tidak terlalu jauh berbeda, maka parameter waktu dapat diabaikan dan fokus pada permasalahan akurasi.

Kesimpulan dari hasil eksperimen penelitian ini menunjukkan bahwa teknik LSA Summary memiliki akurasi yang lebih baik jika dibandingkan dengan teknik Feature Summary pada keseluruhan proses clustering dokumen.

#### **5.4 Target dan Capaian Penelitian**

Merujuk pada metode penelitian, terdapat beberapa capaian yang ditargetkan pada penelitian ini. Target output tersebut kemudian di-jabarkan ke dalam beberapa tahap sesuai jadwal kegiatan yang tercantum dalam proposal. Adapun realisasi dari target output dan capaian tersebut, dapat diamati pada tabel di bawah ini:

**Tabel 4. Target dan Capaian Penelitian**

No	Jenis Kegiatan	Target Output	Realisasi	
			Capaian	Keterangan
<b>1</b>	<b>Identifikasi Masalah</b>			
	a. Analisis Masalah	List permasalahan penelitian	100%	List permasalahan sistem terlampir pada lampiran 1
	b. Studi Literatur		100%	
	c. Penyusunan Metode Penelitian	Metode Penelitian	100%	List Permasalahan
<b>2</b>	<b>Collecting Data</b>			
	a. Crawl Data dari yahoo.co.id	Dataset	100%	Dataset siap diolah
	b. Studi Literatur	State of The Art dan Daftar Pustaka	100%	
<b>3</b>	<b>Preprocessing</b>			
	a. Tokenization	Framework Java	100%	Terlaksana
	b. Stopword		100%	
	c. Stemming		100%	
<b>4</b>	<b>Clustering</b>			
	a. K-Means	Clustering Dokumen	100%	Clustering dokumen
<b>5</b>	<b>Eksperimen dan Evaluasi</b>			
	a. Eksperimen	Java	100%	Terlaksana
	b. Evaluasi	F-Measure	100%	Tterlaksana
<b>6</b>	<b>Pengambilan Kesimpulan dan Penulisan Laporan</b>			
	a. Pengambilan kesimpulan	Kesimpulan penelitian, beserta kendala yang dihadapi selama penelitian	100%	Teknik LSA Summary lebih baik dibandingkan Feature Summary
	b. Penulisan Laporan	Laporan dan usulan topik penelitian selanjutnya yang relevan	100%	Selesai

## 5.5 Hambatan dan Penanganan

Dalam menjalankan penelitian ini, terdapat beberapa hambatan baik internal maupun eksternal. Beberapa hambatan tersebut telah ditangani dan dapat diselesaikan.

No	Hambatan	Penanganan
1	Kurangnya jurnal, buku teks, literatur teknik-teknik clustering dokumen dan teknik peringkasan dokumen	Diskusi dan bergabung dengan forum atau grup research serta melengkapi buku-buku rujukan.
2	Kurangnya data, informasi dan pengetahuan tentang seminar-seminar publikasi nasional dan jurnal nasional terakreditasi	Penugasan dan pembagian job des yang baik.
3	Kurangnya data, informasi dan pengetahuan tentang teknik pelaporan kemajuan dan laporan akhir ke Simlitabmas.	Mempelajari dengan seksama format penulisan laporan di panduan penelitian dan diskusi ke LPPM Udinus.

## **BAB 6. KESIMPULAN DAN SARAN**

### **Kesimpulan**

Dari hasil penelitian yang dilakukan dapat dibuktikan bahwa rata-rata hasil proses clustering dokumen menggunakan peringkasan dokumen otomatis dengan metode Latent Semantic Analysis (LSA) dapat meningkatkan akurasi hasil clustering pada dokumen teks berbahasa Indonesia.

Tingkat akurasi rata-rata tertinggi diperoleh pada saat menggunakan peringkasan dokumen otomatis dengan metode LSA mencapai 71,04 % yang diperoleh pada tingkat peringkasan dokumen otomatis LSA 40% dibandingkan dengan tanpa menggunakan peringkasan dokumen otomatis yang hanya mencapai rata-rata tingkat akurasi 65,92 %, Hasil rata-rata proses clustering dokumen dengan menggunakan teknik peringkasan dokumen otomatis secara keseluruhan mengalami peningkatan kinerja dari pada kinerja clustering dokumen tanpa menggunakan teknik peringkasan dokumen otomatis.

Tambahan proses summary pada proses clustering dokumen mempengaruhi keseluruhan waktu proses dengan cukup signifikan jika dibandingkan dengan hanya menggunakan Original K-Means tanpa summary. Akan tetapi jika dibandingkan waktu proses antara Feature Summary dengan LSA Summary tidak terjadi perbedaan waktu yang signifikan.

Dengan demikian jika dibandingkan antara Feature Summary dengan LSA Summary dimana waktu proses tidak terlalu jauh berbeda, maka parameter waktu dapat diabaikan dan fokus pada permasalahan akurasi.

Kesimpulan dari hasil eksperimen penelitian ini menunjukkan bahwa teknik LSA Summary memiliki akurasi yang lebih baik jika dibandingkan dengan teknik Feature Summary pada keseluruhan proses clustering dokumen.

### **Saran**

Penggunaan Feature Reduction pada proses clustering dokumen, memiliki dampak yang signifikan dalam peningkatan akurasi model, dalam penelitian ini dibuktikan penggunaan teknik peringkasan dokumen otomatis juga dapat meningkatkan akurasi pada proses secara keseluruhan. Pada penelitian selanjutnya dapat dikembangkan dengan menggabungkan (hybrid) teknik feature reduction dengan teknik peringkasan otomatis.

## DAFTAR PUSTAKA

- [1] H. Al-mubaid and A.S. Umair, "A new text categorization technique using distributional clustering and learning logic," *IEEE Trans. Knowl. Data Eng.*, vol. 18, 2006, pp. 1156-1165.
- [2] Luying LIU, Jianchu KANG, Jing YU & Zhongliang WANG, "A Comparative Study on Unsupervised Feature Selection Methods for Text Clustering," *Natural Language Processing and Knowledge Engineering, 2005. IEEE NLP-KE '05. Proceedings of 2005 IEEE International Conference on.*
- [3] R. Peter, S. G, D. G, & S. Kp, "Evaluation of SVD and NMF Methods for Latent Semantic Analysis," *International Journal of Recent Trends in Engineering*, vol. 1, 2009, pp. 308-310.
- [4] Tao Liu, Shengping Liu, Zheng Chen & Wei-Ying Ma, "An Evaluation on Feature Selection for Text Clustering," *Proceedings of the Twentieth International Conference on Machine Learning (ICML-2003)*, Washington DC, 2003.
- [5] Ladda Suanmali, Naomie Salim & M Salem Binwahlan, "Automatic text summarization using feature based fuzzy extraction," *Jurnal teknologi Maklumat jilid 20. Bil 2, 2008.*
- [6] L. Muflikhah & B. Baharudin, "Document Clustering using Concept Space and Cosine Similarity Measurement," *International Conference on Computer Technology and Development*, Kota Kinabalu: 2009, pp. 58 - 62.
- [7] A. Güven, Ö.Ö. Bozkurt, & O. Kalıpsız, "Advanced Information Extraction with n-gram based LSI," *World Academy of Science, Engineering and Technology*, vol. 17, 2006, pp. 13-18.
- [8] JIANG Xiao-Yu, FAN Xiao-Zhong, Wang Zhi-Fei & Jia Ke-Liang, "Improving the Performance of Text Categorization using Automatic Summarization," *International Conference on Computer Modeling and Simulation IEEE 2009.*
- [9] Yuan-Chao Liu, Xiao-Long Wang & Bing-Quan Liu, "A Feature Selection Algorithm for Document Clustering based on Word co-occurrence Frequency," *Proceedings of the Third International Conference on Machine Learning and Cybernetics, Shanghai, 26-29 August 2004.*



- [10] Changqiu Sun, Xiaolong Wang & Jun Xu, "Study on Feature Selection in Finance Text Categorization," *International Conference on Systems, Man, and Cybernetics Proceedings of the 2009 IEEE*.
- [11] Rakesh Peter, Shivapratap G, Divya G & Soman KP, "Evaluation of SVD and NMF Methods for Latent Semantic Analysis," *International Journal of Recent Trends in Engineering*, Vol 1, No. 3, May 2009.
- [12] Zhang, Y., & Jiang, M., "Chinese Text Mining Based on Subspace Clustering," *International Conference on Fuzzy Systems and Knowledge Discovery*, 1617-1620., *IEEE 2010*.
- [13] Wu, R., "Improved K-Modes Clustering Method Based on Chi-square Statistics," *International Conference on Granular Computing*. doi: 10.1109/GrC.2010.66. *IEEE 2010*.
- [14] Mohammed Abdul Wajeed, & Adilakshmi, T., "Text Classification Using Machine Learning," *Journal of Theoretical and Applied Information Technology*, 119-123. 2009.
- [15] Yoo, I., & Hu, X., "A Comprehensive Comparison Study of Document Clustering for a Biomedical Digital Library MEDLINE," *In JCDL'06 (pp. 220-229)*. 2006.
- [16] Anna Hung, "Similarity Measures for Text Document Clustering," *NZCSRSC 2008, April 2008, Christchurch, New Zealand*.2008.
- [17] Krysta M. Svore, Lucy V., & Christopher J.C. Burges, "Enhancing Single-document Summarization by Combining RankNet and Third-party Sources," *Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, pp. 448–457, Prague, June 2007.
- [18] S. Catur, S. Abu, and S. Abdul, "Integrating Feature-Based Document Summarization as Feature Reduction in Document Clustering," *Proceedings of International Conference on Information Technology and Electrical Engineering*, July 2012, pp. 39-42.

## Lampiran 1: Tampilan Sample Dataset

```
C:\Users\ardytha\Desktop\myClusterDatasetAsliSVD.xml - Notepad++
File Edit Search View Encoding Language Settings Macro Run Plugins Window ?
[CW.java] [weather.numeric.arff] [weather.nominal.arff] [menghitung selisih dua lingkaran.c] [14_oper3.c] [coba.c] [LuasSegitiga.java] [AbstractSimilarity.java] [2.c] [myClusterDatasetAsliSVD.xml] [contoh.mod]
10 <datasets>
11   <docID>SP001</docID>
12   <category>sport</category>
13   <title>Alasan Diego Mundur karena Ingin Masuk Timnas Indonesia</title>
14   <content>Diego Michiels akhirnya buka suara mengenai keputusannya mengundurkan diri dari Klub Liga Super Indonesia (LSI) Pelita Jaya.
15   Pemain berusia 21 tahun ini mengaku keputusannya meninggalkan Pelita Jaya adalah demi mempertahankan kesempatan bermain untuk tim nasional Indonesia.
16   Adanya surat FIFA yang melarang pemain dari LSI merumpuk untuk tim nasional jadi dasar keputusan pria kelahiran Deventer, Belanda tersebut. "Alasan saya
17   Saya mau masuk ke timnas Indonesia. Kalau saya ke Pelita, saya tidak bisa bermain untuk timnas," ujar Diego dalam keterangan pers di Jakarta, Selasa (3/
18   Dia mengaku, ambisi mengenakan kostum Merah Putih merupakan satu satunya alasan yang membuatnya hijrah dari Belanda menuju Indonesia. Karena itu, wajar
19   "Karena bermain untuk tim nasional itu saya memutuskan ke Indonesia," tambah Diego. Seperti diberitakan sebelumnya, Diego memutuskan kontrak dengan Peli
20   </content>
21 </datasets>
22 <datasets>
23   <docID>SP002</docID>
24   <category>sport</category>
25   <title>Pelita Jaya Laporkan Masalah Diego ke FIFA dan AFC</title>
26   <content>Keputusan PSSI yang mengambil Diego Michiels secara sepihak membuat kubu Pelita Jaya FC berang dan akan melaporkan hal ini kepada FIFA dan AFC
27   Pelita Jaya mengontrak Diego Michiels selama tiga tahun mulai Februari 2011 hingga 2014. Lalu juga mengatakan dalam surat kontrak tersebut jika pemain i
28   "Semua pemain dihubungi oleh PSSI dan diancam, namun hanya Diego yang memenuhi panggilan tersebut. PSSI juga menjanjikan kepada pemain naturalisasi ters
29   </content>
30 </datasets>
31 <datasets>
32   <docID>SP003</docID>
33   <category>sport</category>
34   <title>Carlos Tevez Ingin Gabung AC Milan</title>
35   <content>Striker Manchester City Carlos Tevez dikabarkan lebih ingin bergabung dengan AC Milan ketimbang dengan Inter Milan. Seperti yang diketahui, Man
36 </datasets>
37
38 <datasets>
39   <docID>SP004</docID>
40   <category>sport</category>
41   <title>Exco PSSI Sepakat Tolak KLB</title>
42   <content>Komite Eksekutif (Exco) PSSI menyatakan sepakat menolak permintaan Kongres Luar Biasa (KLB) yang digulirkan Komite Penyelamat Sepak bola Indone
43   Verifikasi KLB, KPSI telah memberikan verifikasi berkas dukungan KLB kepada PSSI. Dari 460 berkas yang diterima, hanya 320 berkas yang lolos ve
44   </content>
eXtensible Markup Language file | length : 285004 | lines : 1183 | Ln : 1 | Col : 1 | Sel : 0 | 0 | UNIX | ANSI as UTF-8 | IN5
```

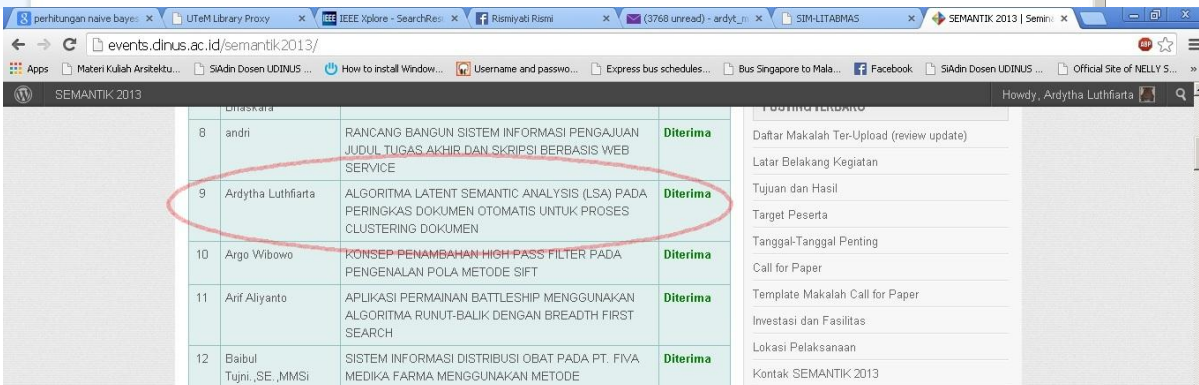
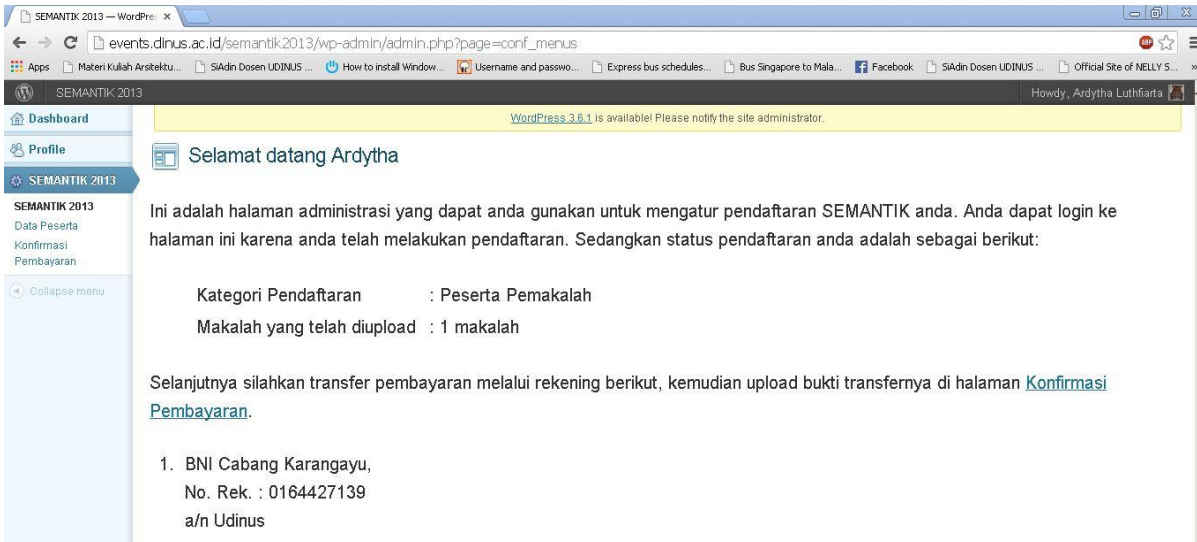
## Lampiran 2 : Script Coding di Java

The image shows a Windows file explorer window for 'MasAboeProject' and a Notepad++ editor window displaying Java code. The file explorer shows a directory structure with folders like 'build', 'lib\_', 'nbproject', 'src', and 'test'. The Notepad++ editor shows the following Java code:

```
124 // "Raw Term Frequencies 2", lsi.getReducedDocumentVector().transpose(),
125 // null, null, new PrintWriter(System.out, true));
126
127 KMeansClustererIna kmeansClusterer = new KMeansClustererIna();
128 //penentuan manual untuk pusat cluster pada dokumen untuk masing2 kategori...
129 String[] initial = {"SPO01", "ECO01", "HK001", "KRO01", "PLO01"};
130 kmeansClusterer.setInitialClusterAssignments(initial);
131 List<Cluster> clusters = kmeansClusterer.cluster(documentCollection);
132 System.out.println("=== Clusters from K-Means Algorithm ===");
133
134 HashMap<String, String> predictedMap = new HashMap<String, String>();
135 for (Cluster cluster : clusters) {
136 // System.out.println(cluster.toString() + cluster.getDocumentName(1));
137 for (int i = 0; i < cluster.size(); i++) {
138 predictedMap.put(cluster.getDocumentName(i), cluster.getId());
139 }
140 }
141
142 double tp = 0, tn = 0, fp = 0, fn = 0;
143 for (Cluster cluster : clusters) {
144
145 System.out.println(cluster.toString() + cluster.getDocumentName(1));
146 for (int i = 0; i < indexReader.numDocs(); i++) {
147 if (indexReader.document(i).getField("kcat").stringValue().equalsIgnoreCase(cluster.getId()) && predictedMap.get(indexReader.document(i).getF
148 tp++;
149 } else if (!indexReader.document(i).getField("kcat").stringValue().equalsIgnoreCase(cluster.getId()) && !predictedMap.get(indexReader.documen
150 tn++;
151 } else if (!indexReader.document(i).getField("kcat").stringValue().equalsIgnoreCase(cluster.getId()) && predictedMap.get(indexReader.document
152 fn++;
153 } else if (indexReader.document(i).getField("kcat").stringValue().equalsIgnoreCase(cluster.getId()) && !predictedMap.get(indexReader.document
154 fp++;
155 }
156 }
157 // System.out.println(tn);
158 }
```



### Lampiran 3 : Paper SEMANTIK 2013



## Lampiran 4 : Laporan Penggunaan Dana

### LAPORAN PENGGUNAAN DANA PELAKSANAAN PENELITIAN HIBAH KOMPETENSI DOSEN PEMULA TAHUN ANGGARAN 2013

NO	Komponen	Satuan				Pagu	Pajak	Realisasi 100%	Ket
<b>1. Gaji dan Upah</b>									
No	Pelaksana	Jumlah Pelaksana	Jumlah Jam/Minggu	Honor/Jam (Rp)	Jumlah (Rp)	Pajak	Diterima		
1	Peneliti Utama	1	5	32	7,187	1,150,000	69,000	1,081,000	PPH 21
2	Anggota Peneliti 1	1	5	32	3,593	575,000	28,750	546,250	PPH 21
3	Anggota Peneliti 2	1	5	32	3,593	575,000	28,750	546,250	PPH 21
						<b>2,300,000</b>	<b>126,500</b>	<b>2,173,500</b>	

#### 2.a Bahan Habis Pakai

No	Nama alat	Jumlah	Harga Satuan (Rp)	Jumlah (rupiah)	Pajak	Jumlah	Ket		
1	ATK1 (Kertas A4, Box File, Stapler dll)	1	192,700	192,700	20,146	192,700	PPN & PPh 22		
2	ATK2 (Paper Clips, Map, Buku Folio, dll)	1	51,700	51,700	5,405	51,700	PPN & PPh 22		
3	Materai	7	6,500	45,500	4,757	45,500	PPN & PPh 22		
4	Kertas A4 Gold	2	33,500	67,000	7,005	67,000	PPN & PPh 22		
5	ATK 3 (Materai dll)	1	41,300	41,300	4,130	41,300	PPN & PPh 22		
6	ATK 4 (Folder Map, Kertas)	1	84,550	84,500	8,455	84,500	PPN & PPh 22		
<b>Sub total</b>						<b>482.700</b>	<b>37,312</b>	<b>482,700</b>	

**2.b Peralatan**

No	Nama alat	Kegunaan	Jumlah	Harga Satuan (Rp)	Beli/sewa	Jumlah (rupiah)	Pajak	Jumlah	Ket
1	Flashdisk	Menyimpan file, source code program dan laporan.	3	280,000	Beli	840,000	87,818	840,000	PPN & PPh 22
2	Modem	Untuk sarana koneksi internet dan rapat koordinasi dimanapun.	3	285,000	Beli	855,000	89,386	855,000	PPN & PPh 22
3	Printer+Scanner	Untuk laporan dan dokumentasi.	1	480,000	Sewa	480,000	41,818	480,000	PPN & PPh 23
4	Buku Penunjang	Sebagai rujukan dan pustaka dalam penyusunan publikasi dan laporan.	1	84,375	Beli	84,375	8,821	84,375	PPN & PPh 22
5	Cetak Buku	Sebagai rujukan dan pustaka dalam penyusunan publikasi dan laporan.	2	145,000	Beli	290,000	30,318	290,000	PPN & PPh 22
<b>Sub total</b>						<b>2,549,375</b>	<b>258,162</b>	<b>2,549,375</b>	

**2.c Biaya Komunikasi**

No	Nama alat	Kegunaan	Jumlah	Harga Satuan (Rp)	Lama (bulan)	Jumlah (rupiah)	Pajak	Jumlah	Ket
1	Pulsa Telepon	Komunikasi anggota dalam berinteraksi di penelitian.	3	25,000	2	163,500	17,093	163,500	PPN & PPh 22
2	Pulsa Internet	Kebutuhan internet dalam mengakses jurnal dan materi.	3	100,000	2	611,000	63,877	611,000	PPN & PPh 22
3	Pulsa + Internet	Komunikasi dan Akses Internet	3	200.000	2	1.200.000	120,000	1.200.000	
<b>Sub total</b>						<b>1.974,500</b>	<b>80,970</b>	<b>1.974,500</b>	

### 2.d Kebutuhan Rapat Koordinasi

No	Jenis Pengeluaran	Frekuensi	Jumlah	Harga Satuan (Rp)	Jumlah	Pajak	Jumlah	Ket
					(Rp)			
1	Rapat Koordinasi di Waroeng Steak		2	78,000	156,000	16,309	156,000	PPh22
2	Rapat Koordinasi di Sate House		1	266,200	266,200	27,830	266,200	PPh22
3	Minuman Softdrink		1	18,900	18,900	1,976	18,900	PPh22
4	Minuman Aqua		1	23,500	23,500	2,457	23,500	PPh22
5	Rapat Koordinasi di Sate House		1	290,400	290,400	26,400	290,400	PPh22
<b>Sub total</b>					<b>755.000</b>	<b>48,572</b>	<b>755.000</b>	

### 3. Pelaporan dan Publikasi

No	Jenis Pengeluaran	Jumlah	Harga Satuan (Rp)	Jumlah	Pajak	Jumlah	Ket
				(Rp)			
1	Publikasi Ilmiah Periode 1	1	500,000	500,000	-	500,000	-
2	Biaya Proseding Publikasi Ilmiah Periode 1	1	200,000	200,000	-	200,000	-
3	Publikasi Ilmiah Periode 2	1	500,000	500,000	-	500,000	-
4	Biaya Proseding Publikasi Ilmiah Periode 2	1	200,000	200,000	-	200,000	-
5	Fotocopy dan Jilid	1	59,500	59,500	5.950	59,500	PPh 22
<b>Sub total</b>				<b>1,459,500</b>	<b>19,902</b>	<b>1,459,500</b>	

#### 4. Perjalanan dan Akomodasi

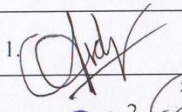
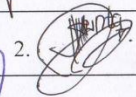
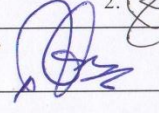
No	Jenis Pengeluaran	Jumlah	Harga Satuan (Rp)	Jumlah	Ket
				(Rp)	
1	Perjalanan dan Akomodasi Publikasi Seminar (2 Orang)	2	450,000	900,000	
<b>Sub total</b>				<b>900.000</b>	

REALISASI 100%	
Gaji dan Upah + Bahan Habis Pakai + Peralatan + Komunikasi + Rapat Koordinasi + Pelaporan	10.505.162
PPN 10%	1,045,454
PPH 2%	146,364
PPH 21	126,500
<b>Total</b>	<b>11.509.844</b>

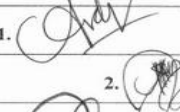

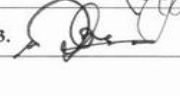


**Lampiran 5 : Tanda Terima Honor Peneliti**

**TANDA TERIMA HONOR PENELITI  
PENELITIAN DOSEN PEMULA 2013**

No.	Nama	Fungsi	Honor	Pajak	Diterima 70%	Tanda Tangan
1.	Ardytha Luthfiarta, M.Kom, M.Cs	Ketua	1.150.000	69.000	756.700	1. 
2.	Junta Zeniarja, M.Kom, M.Cs	Anggota 1	575.000	28.750	382.375	2. 
3.	Abu Salam, M.Kom	Anggota 2	575.000	28.750	382.375	3. 
<b>Total</b>			<b>2.300.000</b>	<b>126.500</b>	<b>1.521.450</b>	

**TANDA TERIMA HONOR PENELITI TAHAP 2 30%  
PENELITIAN DOSEN PEMULA 2013**

No.	Nama	Fungsi	Honor (Rp)	Diterima 30%	Tanda Tangan
1.	Ardytha Luthfiarta, M.Kom, M.Cs	Ketua	313.616	313.616	1. 
2.	Junta Zeniarja, M.Kom, M.Cs	Anggota 1	156.818	156.818	2. 
3.	Abu Salam, M.Kom	Anggota 2	156.818	156.818	3. 
<b>Total</b>			<b>627.273</b>	<b>627.273</b>	

Semarang, 8 Oktober 2013

Mengetahui,  
Ketua Peneliti

  
Ardytha Luthfiarta

Semarang, 11 Desember 2013

Mengetahui,  
Ketua Peneliti

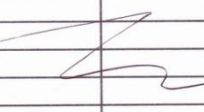
  
Ardytha Luthfiarta, M.Kom, M.Cs

**Lampiran 6 : Kuitansi dan Nota Bahan Habis Pakai**

**Ohnida** Semarang, 25-09-13

alat tulis alat market buku Islam aksesoris komputer aksesoris wanita busana muslim  
 JL. SIROJUDIN NO.14 TEMBALANG - SEMARANG (024-70799368)      NOTA : No.

Jumlah	Nama Barang	Harga Satuan	Harga Total
1	paper clips warna		2500
1	peng. Maped 40 cm		8800
10	Map B101a 5002	1300	13000
1	Buku folio BOKA 5K100		13200
1	Buku folio KAS 0-K		14200
	2 KOLAM		
TOTAL			Rp. 51.700

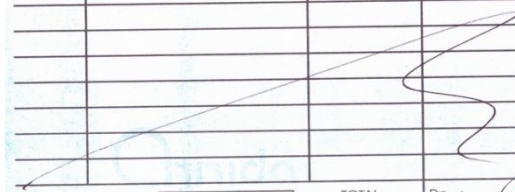
Hormat Kami, 

PERHATIAN  
Barang yang sudah dibeli tidak dapat dikembalikan kecuali ada pesanan.

**Ohnida** Semarang, 31/10/2013

alat tulis alat market buku Islam aksesoris komputer aksesoris wanita busana muslim      NOTA : No.  
 JL. SIROJUDIN NO.14 TEMBALANG - SEMARANG (024-70799368)

Jumlah	Nama Barang	Harga Satuan	Harga Total
2 Rim	HVS A4 80g Bold	33.500	67.000
TOTAL			Rp. 67.000

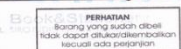
Hormat Kami, 

PERHATIAN  
Barang yang sudah dibeli tidak dapat dikembalikan kecuali ada pesanan.

**Ohnida** Semarang, 25-09-13

alat tulis alat market buku Islam aksesoris komputer aksesoris wanita busana muslim      NOTA : No.  
 JL. SIROJUDIN NO.14 TEMBALANG - SEMARANG (024-70799368)

Jumlah	Nama Barang	Harga Satuan	Harga Total
2 RIM	A4 80 BOLD	33.500	67.000
2	Box file Bantex	22.500	45.000
1	stapler kenko HD 50		15.000
1	stapler kenko HD 10		5.300
1	isi stapler 6-w 369		2.200
1	isi stapler 6-w 10-10		1.400
1 PK	Binder clips no.260		11.300
1 PK	Binder clips no.200		4.500
5	stabilo swing cool	8.200	41.000
TOTAL			Rp. 192.700

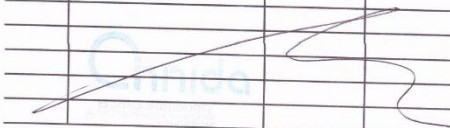
Hormat Kami, 

PERHATIAN  
Barang yang sudah dibeli tidak dapat dikembalikan kecuali ada pesanan.

**Ohnida** Semarang, 11-12-13

alat tulis alat market buku Islam aksesoris komputer aksesoris wanita busana muslim      NOTA : No.  
 JL. SIROJUDIN NO.14 TEMBALANG - SEMARANG (024-70799368)

Jumlah	Nama Barang	Harga Satuan	Harga Total
1 RIM	A4 80 Gold		33.500
2	Box file Bantex	22.500	45.000
1	cutter kenko glide		6.000
TOTAL			Rp. 84.500

Hormat Kami, 

PERHATIAN  
Barang yang sudah dibeli tidak dapat dikembalikan kecuali ada pesanan.

JL. MELAYUBAH BRBT I MEL. 20, MEL. AGANS - TEMBALANG

Bank NISAR 405-08100002 Kasif : HIFTRHU

MATERAI 6000	5	5.500	15.500
TOTAL	1	15.500	15.500
TUNJUK			15.500
Kebijakan PPN ( )			( )

Tgl: 09/10/2013 Jam : 13:26:30

NPWP : 03.290.526.4577.031  
 CV. LEBDA PUTRA JAYA  
 JL. MELAYUBAH BRBT I MEL. 20 RT 03  
 HARDA SGP SEDAHA TEMBALANG PPN  
 WAJAFANLINE.COM  
 024.76.08.207300

ALPABEST KULLAHORAH  
 PUTRA USMA -

FOTO COPY ANUGRAH  
 JL. IMAM BONJOL 191B  
 Semarang-  
 Tanggal: 15-9-13

**ANUGRAH**  
 JL. IMAM BONJOL No. 191B  
 SEMARANG

No.	Jenis	Jumlah	Harga
1	A3		
2	HVS		
3	A4		
4	Buram		
5	Jilid	✓	
6	Transparan		
7	Lain-lain		
Total			Rp 5000

FOTO COPY ANUGRAH  
 JL. IMAM BONJOL 191B  
 Semarang-  
 Tanggal: 2-9-13

**ANUGRAH**  
 JL. IMAM BONJOL No. 191B  
 SEMARANG

No.	Jenis	Jumlah	Harga
1	A3		
2	HVS		
3	A4		
4	Buram		
5	Jilid	✓	
6	Transparan		
7	Lain-lain		
Total			Rp 28.500





**Lampiran 8 : Kuitansi dan Nota Biaya Komunikasi**

ALFAMART MULARAYAN  
CV. LERY PUTRA JAYA

Jl. MELAKSAMA BARAT I NO. 20  
SIRAHAN, ---  
NPPN : 02.236.536.3-517.001  
No. 1632-001-0910002 Kasir : MIFTAKHU

E-VCH TELE TSEL 1	1	100.000	100.000
E-VCH ISM XL 100K	1	102.500	102.500
E-VCH ISM THREE 1	1	102.000	102.000
(+)			
Total Item	3	304.500	
Pajak		304.500	
(-)			
Keuntungan		0	
PN ( 27.481)			

Tgl 09/10/2013 Jan : 19:26:27  
rekening 03025013134 (Member A-Ku)

HARGA RP SUDAH TERMASUK PPN  
WAN ALFAMART LME. COM  
024.23.07.209300

Akti Transaksi Voucher  
No : 124001 E-VCH TELE TSEL 100K  
No. 1632-001-0910002  
kode Voucher : 8215040

Akti Transaksi Voucher  
No : 104030 E-VCH ISM XL 100K  
No. 1632-001-0910002  
kode Voucher : 8225252

Akti Transaksi Voucher  
No : 205045 E-VCH ISM THREE 100K  
No. 1632-001-0910002  
kode Voucher : 8225098

ALFAMART MULARAYAN  
CV. LERY PUTRA JAYA

Jl. MELAKSAMA BARAT I NO. 20  
SIRAHAN, ---  
NPPN : 02.236.536.3-517.001  
No. 1632-001-0910002 Kasir : MIFTAKHU

E-VCH ISM THREE 1	1	102.000	102.000
E-VCH ISM XL 100K	1	102.500	102.500
E-VCH ISM THREE 1	1	102.000	102.000
(+)			
Total Item	3	306.500	
Pajak		306.500	
(-)			
Keuntungan		0	
PN ( 27.481)			

Tgl 09/10/2013 Jan : 19:26:13  
03025013134 (JUNTA DEHARTIA)

ALFAMART MULARAYAN  
CV. LERY PUTRA JAYA

Jl. MELAKSAMA BARAT I NO. 20  
SIRAHAN, ---  
NPPN : 02.236.536.3-517.001  
No. 1632-001-0910002 Kasir : MIFTAKHU

E-VCH TELE TSEL 2	1	27.000	27.000
E-VCH ISM XL 100K	1	27.000	27.000
E-VCH ISM THREE 2	1	27.000	27.000
(+)			
Total Item	3	81.000	
Pajak		81.000	
(-)			
Keuntungan		0	
PN ( 8.043)			

Tgl 09/10/2013 Jan : 19:26:13  
03025013134 (Member A-Ku)

ALFAMART MULARAYAN  
CV. LERY PUTRA JAYA

Jl. MELAKSAMA BARAT I NO. 20  
SIRAHAN, ---  
NPPN : 02.236.536.3-517.001  
No. 1632-001-0910002 Kasir : MIFTAKHU

E-VCH TELE TSEL 2	1	100.000	100.000
E-VCH ISM XL 100K	1	100.000	100.000
E-VCH ISM THREE 2	1	100.000	100.000
(+)			
Total Item	3	300.000	
Pajak		300.000	
(-)			
Keuntungan		0	
PN ( 27.481)			

Tgl 09/10/2013 Jan : 19:26:13  
03025013134 (Member A-Ku)

Jl. MELAKSAMA BARAT I NO. 20  
SIRAHAN, ---  
NPPN : 02.236.536.3-517.001  
No. 1632-001-0910002 Kasir : MIFTAKHU

E-VCH TELE TSEL 2	1	27.000	27.000
E-VCH ISM THREE 2	1	21.000	21.000
(+)			
Total Item	2	48.000	
Pajak		48.000	
(-)			
Keuntungan		0	
PN ( 4.564)			

Tgl 09/10/2013 Jan : 19:26:13

Jl. MELAKSAMA BARAT I NO. 20  
SIRAHAN, ---  
NPPN : 02.236.536.3-517.001  
No. 1632-001-0910002 Kasir : MIFTAKHU

E-VCH ISM THREE 3	1	32.000	32.000
E-VCH ISM XL 30K	1	27.000	27.000
E-VCH TELE TSEL 2	1	21.000	21.000
E-VCH ISM THREE 3	-1	32.000	-32.000
E-VCH TELE TSEL 2	-1	21.000	-21.000
(+)			
Total Item	5	27.000	
Pajak		27.000	
(-)			
Keuntungan		0	
PN ( 1.255)			

Tgl 09/10/2013 Jan : 19:25:55  
03025013134 (Member A-Ku)

Jl. MELAKSAMA BARAT I NO. 20  
SIRAHAN, ---  
NPPN : 02.236.536.3-517.001  
No. 1632-001-0910002 Kasir : MIFTAKHU

E-VCH ISM THREE 1	1	12.500	12.500
GRAB SAND COM 20	1	1.000	1.000
Diskon		-1.000	
(+)			
Total Item	2	13.500	
Pajak		13.500	
(-)			
Keuntungan		0	
PN ( 1.255)			

Tgl 09/10/2013 Jan : 19:25:55  
03025013134 (Member A-Ku)

5-11-2013  
Tuan Toko *Junta*

NOTA NO. 137

BANYAKNYA	NAMA BARANG	HARGA	JUMLAH
1	Voucher Pulsa Smp	100.000	100.000
1	Voucher Pulsa M3	100.000	100.000
1	Voucher Pulsa XL	100.000	100.000
1	Voucher Internet (Paket Dab)	100.000	100.000
1	Voucher Internet 11 M3	100.000	100.000
1	Voucher Internet 11 XL	100.000	100.000

TANDA TERIMA PERHATIAN : Barang-barang yang sudah dibeli tidak dapat dikembalikan. Jumlah Rp. 600.000,-  
Normal Kami.  
**G "Ghiyatsi Cell"**  
Agen & Reseller Pulsa  
Tollmest XL, Indosat (sis. 3 Axis PLN)


2-12-2013  
Tuan Toko *Junta*

NOTA NO. 246

BANYAKNYA	NAMA BARANG	HARGA	JUMLAH
1	Voucher Pulsa M3	100.000	100.000
1	Voucher Pulsa XL	100.000	100.000
1	Voucher Pulsa Smp	100.000	100.000
1	Voucher Internet XL	100.000	100.000
1	Voucher Internet M3	100.000	100.000
1	Voucher Internet Smp	100.000	100.000

TANDA TERIMA PERHATIAN : Barang-barang yang sudah dibeli tidak dapat dikembalikan. Jumlah Rp. 600.000,-  
Normal Kami.  
**G "Ghiyatsi Cell"**  
Agen & Reseller Pulsa  
Tollmest XL, Indosat (sis. 3 Axis PLN)

**Lampiran 9: Kuitansi dan Nota Kebutuhan Rapat Koordinasi**


**RM. Sate House & Seafood**  
**"SRIWIJAYA"**  
 Jl. Imam Bonjol 184 Semarang  
 Telp. 024-3541675



---

No Bill : 02041013110005  
 No Meja : A4 / JMI Tamu : 1  
 Nama Tamu : None

---

# SATE BUNTEL KAMBING		
0006 2X 20,000	40,000	
# SATE BUNTEL SAPI		
0011 2X 20,000	40,000	
# IGA SALIS BBQ		
0071 1X 65,000	65,000	
# Bekmi Special		
0030040005 1X 30,000	30,000	
# NASI ROJO LEBE		
5000019 3X 5,000	15,000	
# ICE JUICE		
0130 1X 13,000	13,000	
# MIX JUICE (2 Item)		
0141 2X 17,500	35,000	
# MINERAL BOTOL		
0116 1X 4,000	4,000	
<b>QTY</b>	<b>13 Sub Total</b>	<b>242,000</b>
	Tax 10%	24,200
	<b>Total</b>	<b>266,200</b>

04-10-2013/001/020410131  
 Terima Kasih  
 Atas Kunjungan Anda

**WAROENG STEAKANDSHAKE**  
 www.waroengsteakandshake.com

NO	PESANAN	JUMLAH	HARGA
1			
2	Pendataan daging	2.	5150
3			
4	Alpukat	3.	2700
5			
6			
7			
8			
9			
10			
11			
12			
13			
14			
15			
16			
17			

No. Meja :  **TOTAL 7800**  
 No. Nota :


**RM. Sate House & Seafood**  
**"SRIWIJAYA"**  
 "SRIWIJAYA"  
 Jl. Imam Bonjol 184 Semarang  
 Telp. 024-3541675

---

No Bill : 02111213110004  
 No Meja : A5 / JMI Tamu : 2  
 Nama Tamu : None

---

# IGA SALIS BBQ		
0071 1X 65,000	65,000	
# SATE KAMBING MUDA		
0007 12X 5,000	60,000	
# BAKMI CORENG SEAFOOD		
0051 1X 30,000	30,000	
# KURUNGUNG CAPI SEAFOOD		
0042 1X 25,000	25,000	
# NASI ROJO LEBE		
5000019 3X 5,000	15,000	
# ICE JUICE		
0130 1X 13,000	13,000	
# MINERAL BOTOL		
0116 3X 4,000	12,000	
# MIX JUICE (3 Item)		
0141 1X 17,500	17,500	
# SOUP BUAH		
0147 1X 17,500	17,500	
# LEMON TEA ICE/ HOT		
0150 1X 3,000	3,000	
<b>QTY</b>	<b>25 Sub Total</b>	<b>354,000</b>
	Tax 10%	35,400
	<b>Total</b>	<b>389,400</b>

11-12-2013/001/021112131  
 Terima Kasih  
 Atas Kunjungan Anda

## Lampiran 10: Pelaporan dan Publikasi

semantik		UNIVERSITAS DIAN NUSWANTORO			
Telah diterima dari :	Ardytha Luthfiarta				
Banyaknya uang :	Tujuh Ratus Ribu Rupiah				
Guna membayar :	Membayar kepesertaan seminar dan call paper Membeli tambahan 1 proceeding Membayar kepesertaan seminar <small>*hari tanda (2) pada pilihan</small>	<table border="1"><tr><td>V</td></tr><tr><td>V</td></tr><tr><td></td></tr></table>	V	V	
V					
V					
Terbilang: Rp. 700.000		 Dra. Cicik Harini, MM			

semantik		UNIVERSITAS DIAN NUSWANTORO			
Telah diterima dari :	Junta Zenlarja				
Banyaknya uang :	Tujuh Ratus Ribu Rupiah				
Guna membayar :	Membayar kepesertaan seminar dan call paper Membeli tambahan 1 proceeding Membayar kepesertaan seminar <small>*hari tanda (2) pada pilihan</small>	<table border="1"><tr><td>V</td></tr><tr><td>V</td></tr><tr><td></td></tr></table>	V	V	
V					
V					
Terbilang: Rp. 700.000		 Dra. Cicik Harini, MM			

Semarang, 02 - Oct - 2013

No : 018/B.23/UDN-01/SEMANTIK/IX/2013  
Lampiran : 3 Lembar  
Hal : Hasil Review

Kepada Yth. Abu Salam  
di tempat

Berdasarkan hasil review oleh tim reviewer, paper Bp/lbu yang berjudul INTEGRASI PERINGKAS DOKUMEN OTOMATIS DENGAN PENGABUNGAN METODE FITUR DAN LATENT SEMANTIC ANALYSIS (LSA) SEBAGAI FEATURE REDUCTION PADA PROSES CLUSTERING DOKUMEN, dengan kode paper, dinyatakan DITERIMA. Hasil penilaian reviewer kami lampirkan pada lembar berikutnya. Selanjutnya kami mohon agar Bapak/lbu untuk segera memperbaiki paper sesuai dengan hasil review dan sesuai dengan format penulisan yang telah ditentukan. Selanjutnya, paper hasil revisi yang disertai dengan bukti pembayaran mohon dikirim kembali ke [semantik.dn@gmail.com](mailto:semantik.dn@gmail.com) dengan subject "SEMANTIK-41" dengan nama file "SEMANTIK-41\_Abu Salam\_[nama\_institusi]" paling lambat 15 Oktober 2013.

Pembayaran sebesar Rp 500.000,00 dapat dilakukan melalui Bank BNI Cabang Karangayu dengan nomor rekening 0164427139 a/n Universitas Dian Nuswantoro. Jika ada kelebihan halaman dalam paper Anda, maka akan kami kenakan biaya tambahan sebesar Rp 50.000,00 untuk setiap halamannya. Pembayaran kami terima paling lambat 15 Oktober 2013.

Demikian surat pemberitahuan dari kami, atas perhatian dan kerjasamanya kami sampaikan terima kasih.

Ketua Panitia


Juli Ratnawati S.E., M.Si





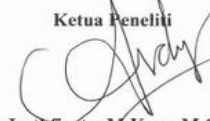
**Lampiran 11: Perjalanan dan Akomodasi**

**TANDA TERIMA UANG SAKU DAN AKOMODASI SEMINAR PUBLIKASI**

NO.	NAMA PENELITI	KEPERLUAN	NOMINAL (RP.)	TANDA TANGAN
1.	Ardytha Luthfiarta, M.Kom, M.Cs	Uang Saku	450.000	1. 
		Perjalanan		
		Akomodasi		
2	Junta Zeniarja, M.Kom, M.Cs	Uang Saku	450.000	2. 
		Perjalanan		
		Akomodasi		

Semarang, 16 November 2013

Mengetahui,  
Ketua Peneliti



Ardytha Luthfiarta, M.Kom, M.CS