

**SISTEM DETEKSI KEMIRIPAN IDENTITAS UNTUK REKOMENDASI PERHITUNGAN
PAJAK PROGRESIF PADA KENDARAAN BERMOTOR MENGGUNAKAN
LATENT SEMANTIC ANALYSIS (LSA)**

Noor Syam Azwar^{*}, Wina Witanti, Faiza Renaldi

Jurusan Informatika, Fakultas MIPA, Universitas Jenderal Achmad Yani
Jl. Terusan Jendral Sudirman, Cimahi, Jawa Barat.

^{*}Email: aby.ga.punya.email@gmail.com

Abstrak

Ketersediaan data dan informasi yang lengkap, benar dan tepat sudah menjadi kebutuhan pokok bagi SAMSAT dalam menentukan nilai pajak bagi setiap wajib pajak, Perbandingan kemiripan untuk menguji beberapa dokumen yang diambil untuk mengetahui seberapa persis dari dokumen tersebut mirip ataukah tidak dengan algoritma yang efektif dan efisiensi. Sistem yang dibuat dan dirancang untuk melakukan pengujian kedua dokumen baik dokumen satu (latih) dengan dokumen dua (target) dalam bentuk penyimpanan databases yang telah terindeks, dengan prediksi apabila ada kemiripan dari teks satu dan teks dua maka mendekati sama atau tidak sama, secara garis besar sistem terdiri dari tiga proses yaitu proses awal (praprocessing with index), proses pencarian (searching) dan proses perhitungan presisi dengan penandaan dan exact match. Fungsi similarity merupakan proses penandaan yang dilakukan untuk mengecek kata apabila ketemu sama persis maka bernilai 1 (true) dan tidak bernilai 0 (false). Analisis yang dilakukan semakin mendekati angka 1 maka dokumen target dinyatakan semakin mirip tetapi jika mendekati angka 0 maka semakin tidak mirip. Jika di tengah-tengah 0,5 setengah mirip setengah tidak..

Kata kunci: dokumen, indeks, similarity, sumber, teks

1. LATAR BELAKANG MASALAH

Di Indonesia terdapat dua jenis pajak yang menggunakan tarif pajak progresif, yaitu Pajak Penghasilan (PPh) dan Pajak Kendaraan Bermotor (PKB), Provinsi Xyz diperkirakan mendapat pemasukan dari sektor pajak progresif yang dimasukkan menjadi Pendapatan Asli Daerah (PAD) hingga Rp1,6 triliun hanya dalam tiga bulan (Oktober-Desember 2015). Tahun lalu, Pajak kendaraan bermotor (PKB) menyumbang sekitar Rp 4,6 triliun bagi pemasukan kas APBN Provinsi Xyz, dengan asumsi jumlah kendaraan bermotor sebesar 4.780.893 unit. Salah seorang Ketua Komisi DPRD Provinsi Xyz menyebutkan bahwa tujuan kenaikan tarif pajak progresif PKB tersebut, selain untuk menambah potensi penerimaan PAD, juga ditujukan untuk mengurangi tingkat kemacetan yang makin meningkat. (Indonesia, 2016)

Sementara dokumen yang ada pun menimbulkan masalah baru, yaitu dokumen SAMSAT yang menjadi rujukan sekaligus menjadi data latih yang nanti harus dibandingkan dengan dokumen KTP dan KK mengenai kemiripan dokumen itu sendiri khususnya dengan kemiripan identitas diri dan alamat keluarga yang tertera didalamnya baik berdasarkan kesamaan kata atau frasa, kasus yang sudah diangkat sebelumnya menggunakan Metode Latent Semantic Analysis (LSA) untuk kesamaan atau kemiripan kata atau frasa. Kasus yang sudah diangkat sebelumnya yaitu sistem penilaian esai otomatis yang akan mencocokkan jawaban siswa dengan kunci jawaban (Aji, Baiza, & Firdaus, 17-18 JUNI 2011).

penghitungan nilai pajak progresif masih memiliki celah untuk ditembus oleh para wajib pajak, oleh karena diperlukan sebuah sistem yang mana sistem tersebut dapat merelasikan beberapa sumber database diantaranya KTP, Kartu keluarga (KK) dan STNK dimana data tersebut dihitung kemiripannya (similarity). Penelitian ini berfokus pada pencocokan kesesuaian dokumen kartu keluarga (KK), kartu tanda penduduk (KTP) dan data dari SAMSAT, berdasarkan dari kata kunci yang didapat setelah dilakukan praproses, proses, dan perhitungan bobot kemudian akan dicek kesesuaian kemiripan identitas dari dokumen tersebut.

1.2 Metode Penelitian

Metode penelitian merupakan pendekatan secara konsisten untuk memecahkan masalah didalam penelitian. Ini dapat diartikan sebagai ilmu yang mempelajari bagaimana penelitian dilakukan secara teknis, Latent Semantic Analysis (LSA) merupakan teknik matematika/statistika untuk mengekstraksi dan menyimpulkan hubungan kontekstual arti kata yang diaplikasikan pada bagian teks yang dibutuhkan (sari & puspanigrum, 19 januari 2013). Pada LSA dilakukan *preprocessing* yang salah satunya berfungsi sebagai penentu kumpulan *term* untuk direpresentasikan dalam sebuah matriks semantik dan kemudian diolah secara matematis menggunakan teknik aljabar linier Singular Value Decomposition (SVD), sehingga dalam hal ini, query dapat dibandingkan dengan hasil SVD untuk menghitung similaritas antara *query-dokumen* (purwitasiri, sahputra, yuniar, yuhana, & siahaan, 2011), kemudian diproses dengan beberapa tahap menggunakan metode LSA pada *Text Mining* sehingga mengeluarkan keluaran berupa nilai persentase kesesuaian. Masukan berupa dokumen KK dan KTP, data *stopword* untuk tahap *pre-processing*, serta control terhadap *wordnet* yang berguna pada proses pembobotan *term* atau *concept* Gambar 1.1 merupakan alur proses menggunakan LSA untuk kasus dalam penelitian ini.

2. TINJAUAN PUSTAKA

2.2 Dokumen KK, SAMSAT DAN KTP

Dokumen KK, SAMSAT dan KTP merupakan dokumen yang digunakan dalam penelitian ini yang bertujuan untuk mendeteksi kemiripan identitas untuk rekomendasi perhitungan pajak progresif menggunakan Latent Semantic Analisis.

2.2.1 Kartu Keluarga

Kartu keluarga adalah Kartu Identitas Keluarga yang memuat tentang nama, susunan dan hubungan dalam keluarga serta identitas anggota keluarga.

2.2.2 SAMSAT (Kepolisian, DIPENDA Dan Jasa Raharja)

SAMSAT merupakan singkatan dari Sistem Administrasi Manunggal Satu Atap. Terdiri dari 3 Instansi (Data kendaraan – untuk Polda, Pajak Daerah-untuk Pemerintah daerah dan Asuransi – untuk pemilik).

2.2.3 Kartu Tanda Penduduk (KTP)

Kartu tanda penduduk adalah identitas resmi penduduk sebagai bukti diri yang diterbitkan oleh instansi pelaksana yang berlaku di seluruh wilayah Negara Kesatuan Republik Indonesia. (Kebumen, 2016)

2.3 Data Penelitian

Latent Semantic Analysis(LSA) adalah sebuah teknik informasi retrieval yang terdiri dari algoritma matematika yang diterapkan untuk koleksi teks, Secara sederhana proses dari LSA adalah sebagai berikut :

1. Merepresentasikan teks dalam matriks, dimana baris menunjukkan kata yang un (Indonesia, 2016)ik dan kolom adalah dokumen yang bersangkutan. Setiap *cell* akan menunjukkan jumlah/frekuensikata pada setiap dokumen. Matriks akan didapat dari tahap ini Selanjutnya LSA melakukan *singular value decomposition* (SVD) terhadap matriks di atas. matriks akan didekomposisike dalam produk dari tiga matriks. Prosesini melakukan dekomposisi matriks $\{X\} = \{W\} \{S\} \{P\}$.
2. Perhitungan *similarity* akan mengikuti proses metode *cosine similarity* digunakan untuk menghitung nilai cosinus dari kedua sudut koordinat i dan j, rumus sebagai berikut:

$$\sin(i, j) = \cos(\vec{i}, \vec{j}) = \frac{\vec{i} \cdot \vec{j}}{\|\vec{i}\|_2 * \|\vec{j}\|_2} \dots\dots\dots (3)$$

2.4 Singular Value Decopotision (SVD)

SVD adalah sebuah metode untuk mengidentifikasi dan mengurutkan dimensi yang menunjukan data mana yang menunjukan variasi yang paling banyak, didasarkan teorema dari aljabarlinear yang mengatakan bahwa persegi panjang matrik A dapat dipecah menjadi produk dari tiga matrik-matrik *orthogonal* U, *diagonal* matrik S, dan *transpose* dari matrik V *orthogonal*, dengan rumus:

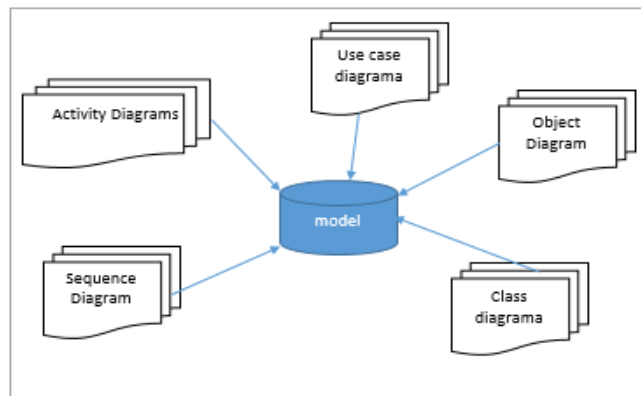
$$A_{m \times n} = U_{m \times n} S_{m \times n} V^T_{m \times n} \dots\dots\dots (4)$$

2.5 Perhitungan Nilai Kemiripan

Perhitungan nilai kemiripan merupakan proses terakhir untuk menghasilkan dokumen yang relevan dengan *query* yang diberikan pengguna. Pada proses ini dokumen akan diberikan peringkat sesuai tingkat relevansinya dengan *query*, kemudian disajikan secara terurut mulai dari peringkat teratas. Nilai kemiripan dihitung dengan rumusan sebagai berikut:

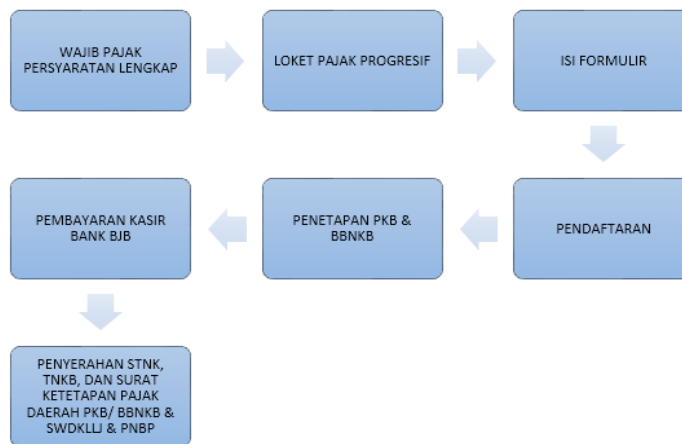
$$sim(Q, D) = sin_{onto}(Q, D) \times sim_{Lucene}(Q, D) \dots\dots\dots(11)$$

Pada dasarnya UML memuat diagram-diagram pemodelan sistem yang terdiri dari *Use casediagram* (diagram kasus), *Class diagram* (diagram kelas), *Object diagram* (diagram objek), *Statechart diagram* (diagram keadaan), *Activity diagram* (diagram aktivitas), *Sequence diagram* (diagram urutan), *Component diagram* (diagram komponen), *Deployment diagram* (diagram penyebaran), *Collaboration diagram* (diagram kolaborasi). Diagram pemodelan sistem pada UML dapat dilihat pada Gambar 2.1

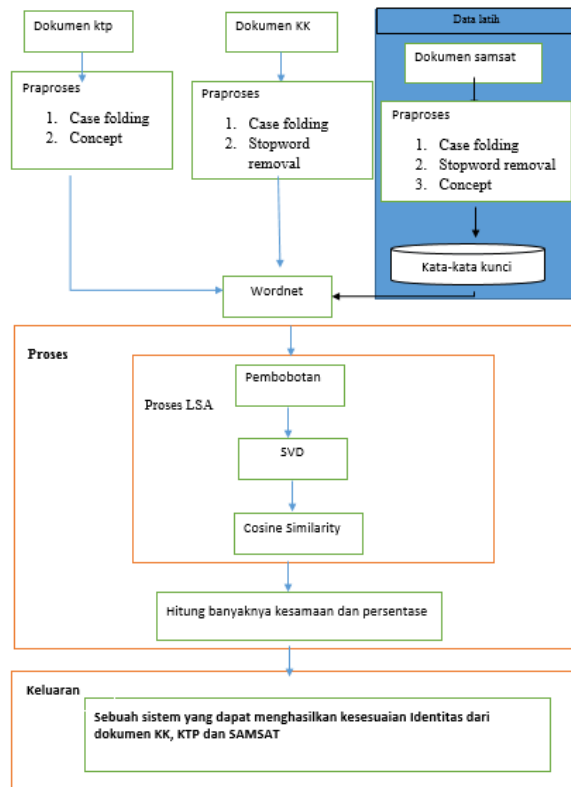


Gambar 2.1 Diagram pemodelan sistem UML

3. ANALISA DAN PERANCANGAN



Gambar 3.1 merupakan Alur pembayaran pajak kendaraan bermotor SWKDKLLJ & pengesahan STNK tahunan (Roda 2) dan (Roda 4).



Gambar 3.2 Metode Penelitian Kemiripan Dokumen KK, KTP, SAMSAT

Langkah utama yang perlu diselesaikan adalah mendekomposisikan matrik A menjadi tiga matrik lain menggunakan SVD. Untuk perhitungan pada tiap langkah, penggunaan *Blue Bit Matrix Calculator* dan mengikuti (Yustiana, 2015).

Langkah 1: Hitung $A^T A$.

3.000 1.000 2.000
 1.000 4.000 0.000
 2.000 0.000 3.000

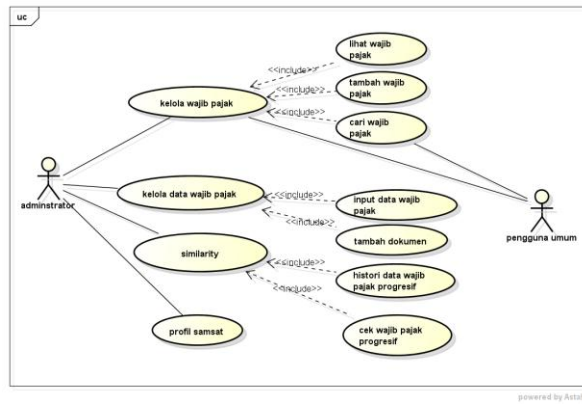
4. ANALISIS DAN PERANCANGAN

Tabel 3.4. Frekuensi Kemunculan konsep Pada Dokumen

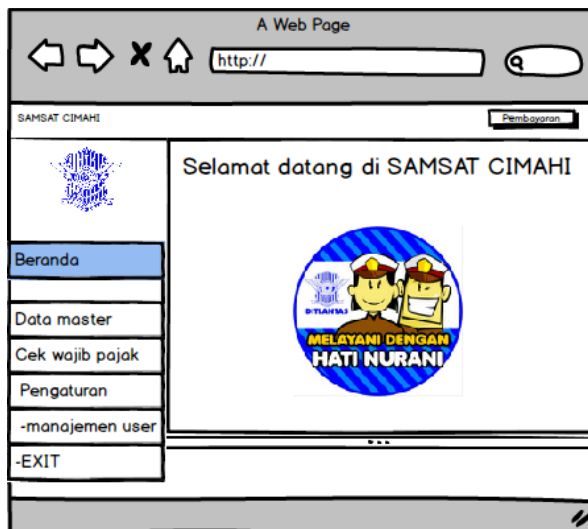
Kata	CF			DF
	D1	D2	D3	
maranata	1	1	0	2
kompetensi	1	1	0	2
jalan	1	0	0	1
penentuan	1	0	0	1
desa	1	0	0	1

Tabel 3.5. Perhitungan Pembobotan dengan CF-IDF

Kata	CF			DF	D/df	idf	W		
	D1	D2	D3				D1	D2	D3
Sukamaju	1	1	0	2	2,5	0,40	0,4	0,4	0
printis	1	1	0	2	2,5	0,40	0,4	0,4	0
malaka	1	0	0	1	5,0	0,70	0,7	0	0
jalan	1	0	0	1	5,0	0,70	0,7	0	0
du	1	0	0	1	5,0	0,70	0,7	0	0
unjani	1	0	0	1	5,0	0,70	0,7	0	0
menggunakan	1	1	0	3	1,7	0,22	0,22	0,22	0

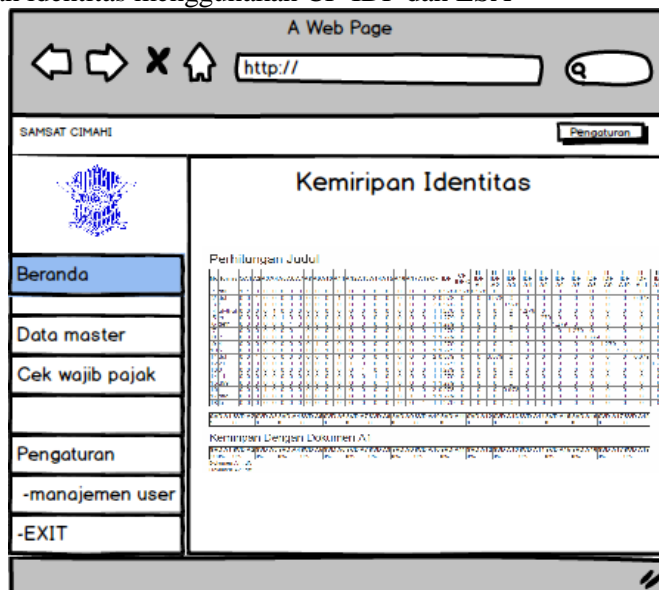


Gambar 3.1. Use Case Diagram



Gambar 3. 2 Beranda

Perhitungan Latent Semantic Analysis (LSA) adalah halaman yang digunakan untuk proses pengecekan kemiripan identitas menggunakan CF-IDF dan LSA



Gambar 3. 3 Perhitungan Latent Semantic Analysis (LSA).

5. IMPLEMENTASI DAN PENGUJIAN SISTEM

5.1. Implementasi Halaman Gambar kemiripan dokumen

Implementasi halaman gambar kemiripan dokumen ini merupakan halaman untuk menampilkan menu gambar kemiripan dokumen data dapat dilihat pada Gambar 4. 1 Halaman kemiripan dokumen.

Kemiripan dokumen

Perhitungan Judul

No	Term	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
1	ber	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	berf	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	berh	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4	berh	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
5	ber	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6	berh	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
7	ber	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
8	berh	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
9	ber	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
10	berh	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
11	ber	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
12	berh	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
13	ber	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
14	berh	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
15	ber	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
16	berh	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
17	ber	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
18	berh	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
19	ber	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
20	berh	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
21	ber	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
22	berh	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Kemiripan Dengan Dokumen A1

100%	95%	90%	85%	80%	75%	70%	65%	60%	55%	50%	45%	40%	35%	30%	25%	20%	15%	10%	5%	0%
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

- Dokumen A1: ber
- Dokumen A2: berf
- Dokumen A3: berh
- Dokumen A4: berh
- Dokumen A5: ber
- Dokumen A6: berh
- Dokumen A7: ber
- Dokumen A8: berh
- Dokumen A9: ber
- Dokumen A10: berh
- Dokumen A11: ber
- Dokumen A12: berh
- Dokumen A13: ber
- Dokumen A14: berh
- Dokumen A15: ber
- Dokumen A16: berh
- Dokumen A17: ber
- Dokumen A18: berh
- Dokumen A19: ber
- Dokumen A20: berh
- Dokumen A21: ber
- Dokumen A22: berh

Gambar 4. 1 Halaman kemiripan dokumen

DAFTAR PUSTAKA

Aji, R. B., Baiza, A., & Firdaus, Y. (17-18 JUNI 2011). Automatic Essay Grading System Menggunakan Metode Latent Semantic Analysis. *seminar nasional aplikasi teknologi informasi 2011 (SNATI 211)*. Yogyakarta.

Indonesia, K. K. (2016, febuari rabu). *Darurat Infrastruktur Jalan*. Retrieved from <http://www.kemenkeu.go.id/>: <http://www.kemenkeu.go.id/en/node/44722>

Kebumen, P. K. (2016, april kamis). http://kependudukan.kebumenkab.go.id/?page_id=48. Retrieved from <http://kependudukan.kebumenkab.go.id/>: http://kependudukan.kebumenkab.go.id/?page_id=48

purwitasari, d., sahputra, d., yuniar, e., yuhana, u. l., & siahaan, d. (2011). jurnal teknologi informasi. *SISTEM PEMBANGKIT ANOTASI PADA ARTIKEL bergambar dengan pendekatan kontekstual*, 9(1), 21-28.

sari, y. a., & puspanigrum, e. y. (19 januari 2013). PencarianSemantikDokumenBeritaMenggunakanEssentialDimension of latent semantic indexing dengan memakai reduksi fitur dokumen frequency dan information Gain thresholding. *seminar nasional teknologi informasi dan multimedia 2013*. yogyakarta.

Yustiana, D. (2015). Penilaian Otomatis Terhadap Jawaban Esai Pada Soal Berbahsa Indonesia Menggunakan Latent Semantic Analysis (LSA). *SEMINAR NASIONAL "INOVASI DALAM DESAIN DAN TEKNOLOGI" IdeaTech 2015*. Surabaya.