

Aplikasi Mesin Pencarian Alat Elektronik Berbasis Web Menggunakan Metode Vector Space Model

Putri Tarakanita Wijaya Ningtyas, Fitriansyah, Dikky Praseptian M.

Abstraksi— Penelitian ini bertujuan untuk membuat sistem pencarian alat elektronik Menggunakan Metode Vector Space Model pada PT. Eka Daya Tarakan. Proses pencarian Alat Elektronik masih dilakukan dengan cara mencatat ke dalam sistem komputerisasi yang ada pada PT. Eka Daya Tarakan. Metode Vector Space Model yang dibahas menggunakan konsep ruang vektor. Masukan dari pengguna dan kumpulan dokumen diterjemahkan menjadi vektor-vektor. Kemudian vektor-vektor tersebut dikenakan operasi perkalian titik dan hasilnya menjadi acuan dalam menentukan relevansi masukan pengguna (*query*) terhadap kumpulan dokumen.

Vector Space Model solusi atas permasalahan yang dihadapi jika menggunakan algoritma TF/IDF. Karena pada algoritma TF/IDF terdapat kemungkinan antar dokumen memiliki bobot yang sama, sehingga ambigu untuk diurutkan. Pada sistem ini menggunakan adalah satu kriteria yaitu kata kunci untuk metode Vector Space Model yang telah diberikan bobot penilaian kesamaan. Nilai yang paling banyak adalah hasil yang muncul di barisan paling atas. Hasil penelitian yang dilakukan membuktikan bahwa metode sukses diterapkan kedalam sistem pencarian alat elektronik pada PT. Eka Daya Tarakan.

Kata Kunci—Pencarian Alat Elektronik, Vector Space Model

I. PENDAHULUAN

Belakangan ini semakin populer penggunaan data mining untuk menggali informasi dari data mentah. Salah satu bagian dari data mining yang cukup menarik adalah algoritma *stemming*. Algoritma *stemming* adalah suatu teknik atau cara memenggal kata imbuhan, baik berupa awalan maupun akhiran sehingga membentuk kata dasar.

Pada analisa temu kembali (*Information Retrieval*) informasi imbuhan merupakan bagian dari informasi yang tidak bermakna, seperti halnya *stop word*. Sehingga imbuhan harus dihilangkan untuk mempercepat proses query. Proses *stemming* dapat dilakukan dengan 2 cara yaitu dengan menggunakan kamus dan menggunakan aturan imbuhan. Untuk mendapatkan tingkat kebenaran hasil yang tinggi biasanya digunakan kamus seperti yang diperkenalkan oleh Arifin dan Setiono. Ada beberapa metode pencarian untuk melengkapi proses Arifin dan Setiono, metode disebut dengan algoritma *Vector Space Model*.

Algoritma *Vector Space Model* yang dibahas menggunakan konsep ruang vektor. Masukan dari pengguna dan kumpulan dokumen diterjemahkan menjadi vektor-vektor. Kemudian

vektor-vektor tersebut dikenakan operasi perkalian titik dan hasilnya menjadi acuan dalam menentukan relevansi masukan pengguna (*query*) terhadap kumpulan dokumen.

Salah satu instansi yang membutuhkan mesin *stemming* dengan algoritma *Vector Space Model* adalah PT. Eka Daya Tarakan. Berdasarkan kondisi diatas, maka penulis tertarik untuk membuat sebuah aplikasi pencarian menggunakan algoritma *stemming* Arifin dan Setiono serta metode *vector space model* pada PT. Eka Daya Tarakan. Salah satu teknologi yang dapat memudahkan pembeli untuk mencari alat elektronik dimanapun ia berada adalah melalui web. Oleh karena itu, pada penelitian ini dibuat aplikasi berbasis web.

II. TINJAUAN PUSTAKA

Aplikasi merupakan suatu program komputer yang berfungsi untuk melakukan tugas khusus atau menjalankan suatu perintah seperti membuat dokumen, manipulasi foto, membuat laporan dan lain-lain.

Pencarian atau *searching* di dalam ilmu komputer merupakan kegiatan untuk mendapatkan suatu bentuk informasi yang terdapat di dalam berkas. Pada sub bab ini akan dibahas mengenai tahapan pencarian mulai dari case folding, tokenisasi, Stop Word Removal, *stemming*, pembobotan *tf/idf*, pembobotan query dengan kalimat menggunakan *vector space model* yang nantinya diperoleh hasil pencarian. Hasil pencarian diperoleh dari nilai cosine >0 dan diurutkan dari nilai maksimal ke nilai minimal.

Case folding adalah tahapan proses mengubah semua huruf dalam teks dokumen menjadi huruf kecil, serta menghilangkan karakter selain a-z. Case folding adalah proses pemrosesan teks dimana semua teks diubah ke dalam *case* yang sama dan pada kasus ini teks diubah representasinya ke dalam huruf kecil semua. Model kesalahan *orthographic* model, akan dikoreksi dengan cara mengubah semua huruf menjadi lowercase atau huruf kecil. Berikut contoh penerapan case folding pada peringkasan sistem. Contoh proses case folding dapat dilihat pada Tabel I.

Tokenisasi adalah proses pemotongan atau memisahkan deretan kata di dalam kalimat, paragraf atau halaman menjadi token atau potongan kata tunggal dengan *scan* kalimat dengan pemisah (delimiter) *white space* (spasi, tab, dan *newline*). Contoh proses tokenisasi dapat dilihat pada tabel II.

TABEL I
 CASE FOLDING

Doc	Deskripsi
1	mesin cuci sharp es t65mw sharp twin tub washer est65mw merupakan mesin cuci berkapasitas cuci 6 kg mesin cuci ini didesain dengan penggunaan listrik yang rendah sebesar 199 watt desain bodi yang kuat dan menarik kemampuan dengan pulsator yang besar mampu menghasilkan air yang kuat sehingga hasil cucian lebih bersih ukuran sangat bagus harga 1450000 kita juga dapat menghemat daya listrik karena mesin cuci ini menggunakan watt yang tidak begitu besar dan juga mesin cuci ini harganya tidak begitu mahal karena mesin cuci ini ekonomis
2	sharp aquos 32 in lc32le100m televisi led yang punya kemampuan berkualitas yang bisa dinikmati pencari televisi sharp memberikan kualitas gambar yang begitu nyata dengan resolusi grafis sangat memukau sharp memberikan kualitas suara yang jernih pula, mau diputar di volume kecil sekalipun masih terdengar jelas nah soal harga sharp aquos memiliki harga yang relatif terjangkau kisaran 18 jutaan aja untuk kelas tv led harganya sangat terjangkau kan padahal ukuran layarnya cukup besar lho 32 inch

TABEL II
 TOKENISASI

Doc	Token	Filter	Word
D1	Mesin	1	D1
	Cuci	1	D1
D1	Sharp	1	D1
	Es	1	D1
	t65mw	1	D1
	Sharp	1	D1
	Twin	1	D1
	Tub	1	D1
	washer	1	D1
	Es	1	D1
	t65mw	1	D1
	merupakan	0	D1
	Mesin	1	D1
	Cuci	1	D1
	berkapasitas	1	D1
	Cuci	1	D1
	6	0	D1
	Kg	1	D1
	Mesin	1	D1
	Cuci	1	D1
	Ini	0	D1
	didesain	1	D1
...
...
D2	Lho	1	D1
	32	0	D1
	Inch	1	D1

Stopword adalah kata-kata yang sering kali muncul dalam dokumen namun artinya tidak deskriptif dan tidak memiliki keterkaitan dengan tema tertentu. Didalam bahasa Indonesia stopword dapat disebut sebagai kata tidak penting, misalnya “di”, “oleh”, “pada”, “sebuah”, “karena” dan lain sebagainya.¹ Contoh proses filtering pada tabel III.

Stemming adalah proses mengembalikan kata ke dalam bentuk dasarnya dengan cara memotong imbuhan pada kata tersebut, baik itu berupa awalan, akhiran, maupun sisipan [1]. Dalam penelitian ini stemming yang digunakan adalah

stemming Arifin dan Setiono. Contoh proses stemming pada tabel IV.

TABEL III
 STOP WORD REMOVAL

Doc	Token	Filter	Word
D1	Mesin	1	D1
	Cuci	1	D1
	Sharp	1	D1
	Es	1	D1
	t65mw	1	D1
	Sharp	1	D1
	Twin	1	D1
	Tub	1	D1
	Washer	1	D1
	Es	1	D1
	t65mw	1	D1
	merupakan	0	D1
	Mesin	1	D1
	Cuci	1	D1
	berkapasitas	1	D1
	Cuci	1	D1
	Kg	1	D1
	Mesin	1	D1
	Cuci	1	D1
	Ini	0	D1
	didesain	1	D1
...
...
D2	Lho	1	D1
	32	0	D1
	Inch	1	D1

TABEL IV
 STEMMING

Term	Steam	Word
Mesin	Mesin	D1
Cuci	Cuci	D1
Sharp	Sharp	D1
Es	Es	D1
t65mw	t65mw	D1
Sharp	Sharp	D1
Twin	Twin	D1
Tub	Tub	D1
Washer	Washer	D1
Es	Es	D1
t65mw	t65mw	D1
Mesin	Mesin	D1
...
...
Lho	Lho	D2
Inch	Inch	D2

Vector space model adalah salah satu model yang digunakan untuk mengukur derajat kemiripan antara suatu dokumen dengan suatu query.

$$\text{Cosine} \rightarrow \text{sim}(dj, q) = \frac{\vec{a}_j \cdot \vec{q}}{|\vec{a}_j| \cdot |\vec{q}|} = \frac{\sum_{i=1}^t (W_{ij} \cdot W_{iq})}{\sqrt{\sum_{i=1}^t W_{ij}^2 \cdot \sum_{i=1}^t W_{iq}^2}} \quad (1)$$

Keterangan :

t = term dalam Kalimat

D = Dokumen

Q = Kata Kunci

Wij = bobot term i dalam blok ij

$$W_{iq} = \text{bobot term } i \text{ dalam blok } i_q$$

$$\text{Cosine}(D_i) = \frac{\text{sum}(kk^2 \cdot D_i)}{[\text{sqrt}(kk^2) * \text{sqrt}(D_i^2)]}$$

Vector space model merupakan pemodelan untuk menghitung nilai cosinus dari dua vektor W dari dokumen dan W dari kata kunci, contoh pemodelan untuk perhitungan vector space model pada contoh kasus berikut :

- D1 = mesin cuci SHARP ES T65MW.SHARP - TWIN TUB WASHER EST65MW merupakan mesin cuci berkapasitas cuci 6 kg.
- D2 = SHARPAQUOS 32 in. LC-32LE100M. Televisi LED yang punya kemampuan berkualitas yang bisa dinikmati pencari televisi.

Jika terdapat query (Q) : “Mesin cuci”.

Langkah 1. Praproses terhadap semua (n=3) dokumen yang terlibat, yaitu Q, D1, dan D2.

Hasil praproses (tokenisasi, case folding, stop word, dan stemming) yaitu :

- D1 = mesin cuci sharp es t65mw sharp twin tub washer est65mw merupakan mesin cuci kapasitas cuci 6 kg
- D2 = sharp aquos 32 in lc 32le100m. televisi led mampu berkualitas bisa nikmat cari televisi

Selanjutnya hitung nilai W untuk masing-masing dokumen menggunakan algoritma TF/IDF.

kk^2 adalah kata kunci yang dikuadratkan, diambil dari nilai w yang kemudian dikuadratkan. $D1^2$ dan $D2^2$ merupakan dokumen 1 sampai dengan dokumen yang dikuadratkan, diambil dari nilai w. Selanjutnya hitung nilai *cosinus* sudut antara vektor kata kunci dengan tiap dokumen dengan menggunakan rumus pada persamaan 2.2.

$$\text{Cosine}(D1) = \frac{\text{sum}(kk^2 \cdot D1)}{[\text{sqrt}(kk^2) * \text{sqrt}(D1^2)]}$$

$$= \frac{7,449}{[1,114 * 6,685]}$$

$$= 1,002$$

$$\text{Cosine}(D2) = \frac{\text{sum}(kk^2 \cdot D2)}{[\text{sqrt}(kk^2) * \text{sqrt}(D2^2)]}$$

$$= \frac{0}{[0 * 0]}$$

$$= 0$$

Sesuai hasil perhitungan diatas maka nilai cosinus setiap dokumen telah didapat, seperti tabel V. Nilai tersebut diranking

TABEL V

TABEL HASIL AKHIR PERHITUNGAN VECTOR SPACE MODEL

	D1	D2
Cosine	1,002	0
	#1	#2

Dari hasil diatas, dapat diketahui bahwa urutan tingkat similitas dokumen tertinggi adalah D1, kemudian disusul D2.

Dunia teknologi sangatlah berkembang pesat di masa ini dan mendatang, banyak sekali alat elektronik yang diciptakan sehingga dapat memanjakan para peminatnya. PT. Eka Daya Tarakan merupakan salah satu perusahaan yang bergerak dibidang alat elektronik yang terletak di kota tarakan tepatnya pada JL. Yos Sudarso, No. 51 RT. 008 77114 Tarakan, dengan kegiatan usaha yang terdiri dari alat elektronik. Saat ini PT. Eka Daya Tarakan mempekerjakan sekitar 20 karyawan di tarakan,

beberapa diantaranya adalah karyawan milik masing-masing produsen alat elektronik diantaranya sharp, samsung, panasonic lg dan produsen lain yang bekerjasama pada PT. Eka Daya Tarakan. Alat elektronik pada PT. Eka Daya Tarakan sangatlah mempunyai banyak varian yaitu televisi, kulkas, pendingin ruangan, mesin cuci dan masih banyak lagi alat elektronik lainnya. PT. Eka Daya Tarakan ini sendiri masih menggunakan sistem konvensional yaitu pelanggan harus datang membeli dengan mendatangi PT. Eka Daya Tarakan dan membeli barang yang diinginkan, kondisi ini membuat pemasaran dan penjualan terhambat oleh jarak dan waktu. Selain itu promosi pada perusahaan tersebut belum sepenuhnya dilakukan dengan baik karena hanya dilakukan dengan menginformasikan langsung kepada pembeli yang datang ke toko atau kepada teman atau kerabat dari pemilik dan pegawai toko, sehingga menyebabkan konsumen yang berada di tempat yang jauh kesulitan dalam informasi barang yang tersedia di toko.

III. ANALISA DAN DESAIN SISTEM

A. Penerapan Metode

Pada penelitian ini metode yang digunakan untuk melakukan pencarian alat elektronik adalah *vector space model*. Proses pencarian diawali dengan pengguna memasukkan kata kunci (*query*) sesuai dengan yang diinginkan pengguna, sehingga pengguna dengan mudah mendapatkan informasi dalam pencarian alat elektronik.

Pengembangan perangkat lunak dengan menggunakan *vector space model* meliputi beberapa tahapan. Di dalam penelitian ini, pengembangan sistem temu-kembali informasi ini hanya akan dibahas tahap pemodelan dan pembangunan perangkat lunak yang meliputi analisis, perancangan, pembuatan serta pengujian perangkat lunak. Pada bab ini akan dibahas mengenai analisis dan perancangan pengembangan sistem temu-kembali informasi.

Analisis sistem temu-kembali yang akan dibangun merupakan bagian dari sistem pencarian data. Akan tetapi, pada pembahasan analisis dan perancangan serta implementasi dan pengujian pada bab selanjutnya akan lebih diutamakan kepada sistem temu-kembali informasi saja. Untuk lebih jelasnya mengenai sistem pencarian data secara garis besar sistem penyimpanan data akan melakukan proses pencarian file dokumen berbasis teks yang dimasukan oleh *user* pada media pencarian. Dokumen teks akan dikelompokkan menjadi kelompok dokumen sesuai dengan jenis/karakteristiknya. Perluasan dapat dilakukan pada vektor *query*, vektor dokumen, atau pada kedua vektor tersebut. Hubungan antara kata pada *database*, dokumen dan kata kunci.

Perhitungan bobot *query relevance* merupakan bobot hasil perbandingan kemiripan (similaritas) antara *query* yang dimasukkan oleh user terhadap keseluruhan kalimat. Perhitungan bobot *query relevance* yaitu menghitung bobot kemiripan antara *query* dengan kalimat dalam dokumen, dengan menghitung cosinus sudut dari dua vektor yaitu W, bobot dari tiap kalimat dan W (bobot) *query*.

Dalam pembobotan *query relevance* dengan menggunakan *vector space model*, dokumen dan *query* direpresentasikan dalam bentuk vector. Untuk setiap kata diberi bobot

menggunakan skema tf atau tf-idf. Kesamaan antara dokumen dan que atau $sim(d,q)=1$ jika $d=q$, sebaliknya $sim(d,q)=0$ jika d dan q tidak memiliki kesamaan.

Pada sekumpulan data elektronik sebanyak 50 data elektronik pada database akan diambil 1 data yang akan digunakan sebagai contoh untuk perhitungan query relevance menggunakan vector space model. Query yang diberikan pada penelitian ini berupa kata kunci dari spesifikasi. Adapun contoh dokumen tersebut ditunjukkan pada table VI. Similarity diukur berdasarkan sudut cosinus dari vector query (cosine measure).

TABEL VI
 CASE FOLDING

Doc	Deskripsi
1	mesin cuci sharp es t65mw sharp twin tub washer est65mw merupakan mesin cuci berkapasitas cuci 6 kg mesin cuci ini didesain dengan penggunaan listrik yang rendah sebesar 199 watt desain bodi yang kuat dan menarik kemampuan dengan pulsator yang besar mampu menghasilkan air yang kuat sehingga hasil cucian lebih bersih ukuran sangat bagus harga 1450000 kita juga dapat menghemat daya listrik karena mesin cuci ini menggunakan watt yang tidak begitu besar dan juga mesin cuci ini harganya tidak begitu mahal karena mesin cuci ini ekonomis
2	sharp aquos 32 in lc32le100m televisi led yang punya kemampuan berkualitas yang bisa dinikmati pencari televisi sharp memberikan kualitas gambar yang begitu nyata dengan resolusi grafis sangat memukau sharp memberikan kualitas suara yang jernih pula, mau diputar di volume kecil sekalipun masih terdengar jelas nah soal harga sharp aquos memiliki harga yang relatif terjangkau kisaran 18 jutaan aja untuk kelas tv led harganya sangat terjangkau kan padahal ukuran layarnya cukup besar lho 32 inch

Tahapan-tahapan praproses tersebut sama dengan tahapan praproses data untuk pencarian data elektronik pada bab 2. Hasil tokenisasi, penghapusan stopword dan stemming dokumen pada tabel VI untuk membentuk sebuah matrik antara term dengan masing-masing dokumen, ditunjukkan pada tabel VII

TABEL VII
 MATRIKS TERM DATA ELEKTRONIK

Term	D	kk	D1	D2
Mesin	D1	1	8	0
Cuci	D1	1	9	0
Sharp	D1	1	2	4
Es	D1	0	2	0
t65mw	D1	0	2	0
Twin	D1	0	1	0
..
Lho	D2	0	0	1
Inch	D2	0	0	1

Selanjutnya menghitung bobot W untuk masing-masing dokumen. Terdapat perhitungan 0 pada Tabel VIII. Pada tabel VIII perhitungan nilai W dengan menggunakan tf-idf yang dijumlahkan hanya query yang berupa kata kunci pada data elektronik, contoh perhitungan tf-idf pada tabel 3.3 kata kunci pada data elektronik yang digunakan untuk contoh perhitungan yaitu “mesin dan cuci” diambil dari kode 1 dan 2.

TABEL VIII
 HASIL PERHITUNGAN TF-IDF

Term	D	k k	D 1	D 2	d f	D/ df	IDF	kk	D1	D2
mesin	D1	1	8	0	1	2	1,11	1,1	8,91	0,00
Cuci	D1	1	9	0	1	2	1,11	1,1	10,0	0,00
sharp	D1	1	2	4	2	1	1,00	1	2,00	4,00
Es	D1	0	2	0	1	2	1,11	0	2,23	0,00
t65m w	D1	0	2	0	1	2	1,11	0	2,23	0,00
Twin	D1	0	1	0	1	2	1,11	0	1,11	0,00
Tub	D1	0	1	0	1	2	1,11	0	1,11	0,00
wash er	D1	0	1	0	1	2	1,11	0	1,11	0,00
kapas itas	D1	0	1	0	1	2	1,11	0	1,11	0,00
Kg	D1	0	1	0	1	2	1,11	0	1,11	0,00
desai n	D1	0	2	0	1	2	1,11	0	2,23	0,00
Guna	D1	0	1	0	1	2	1,11	0	1,11	0,00
..
..
Lho	D2	0	0	1	1	2	1,11	0	0,00	1,11
Inch	D2	0	0	1	1	2	1,11	0	0,00	1,11

Perhitungan dilakukan untuk kolom D1 & D2. Diperoleh nilai bobot W untuk kata kunci dan masing-masing dokumen sebagai berikut:

$$\begin{aligned}
 KK &= 1,1+8,91+0=10,01 \\
 D1 &= 1,1+10,03+0=11,13 \\
 D2 &= 1+2,00+4,00=7
 \end{aligned}$$

Langkah berikutnya adalah menghitung akar (Sqrt) dari jumlah KK^2 dan semua dokumen yang telah dikuadratkan. Untuk nilai jumlah KK^2 dan semua dokumen yang telah dikuadratkan dapat dilihat pada tabel IX. adapun perhitungannya akan dijelaskan sebagai berikut:

$$\begin{aligned}
 \text{Sqrt}(KK^2) &= 1,2 = 1,095445 \\
 \text{Sqrt}(D1^2) &= 44,70 = 6,685806 \\
 \text{Sqrt}(D2^2) &= 0,00 = 0,00
 \end{aligned}$$

Dari hasil uji coba pertama, nilai tertinggi dimiliki oleh deskripsi produk Samsung dengan total nilai bobot 0,69.

Dari hasil uji coba kedua, nilai tertinggi dimiliki oleh deskripsi produk Panasonic dengan total nilai bobot 0.746.

Berdasarkan sistem yang telah dibuat pada form pencarian apabila data elektronik dipanggil kembali nilai bobot bisa saja berubah. Apabila kata kunci di rubah sesuai dengan keinginan user, mengacu kembali pada kriteria tertinggi dengan urutan nilai bobot tertinggi. Dari hasil uji coba penulis mendapat hasil dimana jika sewaktu-waktu nilai bobot berubah maka posisi perankingan dapat berganti dan total nilai bobot juga mengalami perubahan.

IV. KESIMPULAN

Berdasarkan langkah-langkah yang ditempuh dalam penelitian yang berjudul “Sistem Pencarian Alat Elektronik Berbasis Web Menggunakan Metode Vector Space Model” adalah:

1. Aplikasi sistem pencarian pada penulisan penelitian ini ketika dijalankan menggunakan mozilla firefox, nilai pada perhitungan vector space model belum akurat.
2. Metode yang digunakan dalam sistem pencarian dapat membantu proses penilaian, akan tetapi terdapat fitur-fitur yang masih kurang lengkap yang terdapat pada form seperti, gambar dan letak yang kurang spesifik.
3. Sistem ini menggunakan beberapa pra proses dalam satu waktu, sehingga cukup lama untuk menjalankannya.

REFERENSI

- [1] Jiwa Malem Marsya, Analisa dan Evaluasi Afiks Stemming untuk Bahasa Indonesia, FMIPA Universitas Syiah Kuala, 2011.
- [2] Abdillah, Abdul Aziz. 2013. *Implementasi vector Space Model Untuk Pencarian Dokumen*. Dari <https://www.researchgate.net> Diakses tanggal 30 April 2016 Pukul 11.08 WITA
- [3] Susanto, Budi. 2006. *Text Mining*. Dari <http://lecturer.ukdw.ac.id> Diakses tanggal 18 April 2016, Pukul 11.30 WITA.

Penulis I, Putri Tarakanita Wijaya Ningtyas, memperoleh gelar Sarjana Komputer (S.Kom), Jurusan Sistem Informasi STMIK PPKIA Tarakanita Rahmawati Tarakan, lulus tahun 2016.

Penulis II, Fitriansyah, memperoleh gelar Magister Komputer (M.Kom), Sekolah Tinggi Teknik Surabaya. Saat ini menjadi Dosen di STMIK PPKIA Tarakanita Rahmawati

Penulis III, Dikky Praseptian M., memperoleh gelar Sarjana Komputer (S.Kom), Jurusan Sistem Informasi STMIK PPKIA Tarakanita Rahmawati Tarakan, lulus tahun 2013. Saat ini menjadi Dosen di STMIK PPKIA Tarakanita Rahmawati