

## **CUSTOMER INFORMATION GATHERING MENGGUNAKAN METODE TEMU KEMBALI INFORMASI DENGAN MODEL RUANG VEKTOR**

**Muhammad Erwin Ashari Hariyono<sup>1</sup> dan Wahyudi<sup>2</sup>**

*Laboratorium Pemrograman dan Informatika Teori*

*Jurusan Teknik Informatika, Fakultas Teknologi Industri, Universitas Islam Indonesia*

*Kampus Terpadu UII, Jl. Kaliurang Km 14.5 Yogyakarta*

*E-mail: meah@fti.uii.ac.id<sup>1</sup>, yudi\_alcampari@yahoo.com<sup>2</sup>*

### **ABSTRAK**

Suatu perusahaan sangat memerlukan saran dan masukan dari para pelanggan (*customer*) yang mempergunakan jasa dan layanan yang diberikan oleh perusahaan tersebut. Kesan dan saran tersebut akan digunakan perusahaan untuk meninjau dan memperbaiki kinerja dan proses pelayanan yang selama ini telah dilakukan demi kepentingan perusahaan untuk lebih maju lagi kedepannya. Terkadang saran dari pelanggan tersebut kurang mendapat respon dan tanggapan dari perusahaan. Saran telah diberikan tetapi untuk waktu yang lama perusahaan tersebut tidak secara baik meninjau. Hal ini terjadi karena proses yang dilakukan secara manual dan terkadang penggolongan terhadap kesan dan kritik terhadap sub bidang proses perusahaan tersebut sangat sulit untuk diimplementasikan. Hal ini dikarenakan sistem peninjau harus mencari satu persatu saran tersebut secara manual, walaupun telah menggunakan media komputer untuk merekam informasi tersebut. Model Vektor sangat tepat digunakan untuk mencari dokumen masukan saran pelanggan (*customer gathering information*) terutama pada penentuan kemiripan *query* dan penggolongan kritik dan saran tersebut. Model yang dikembangkan dan diusulkan telah diterapkan dan di implementasikan pada salah satu perusahaan multinasional di Indonesia yang oleh penulis tidak disebutkan dalam paper ini.

**Kata Kunci:** *customer information gathering, model ruang vektor*

### **1. PENDAHULUAN**

Manusia tidak pernah luput dan lepas dari namanya permasalahan pencarian, dimanapun dan kapanpun selalu ada model pencarian yang secara mutlak diperlukan. Dalam komputer, pencarian juga menjadi hal yang sangat penting di dalamnya, dimana seluruh aktifitas IT selalu berkaitan dengan hal tersebut. Bidang jaringan komputer, seperti pencarian alamat DNS yang telah tersimpan dalam *cache*, *traffic sniffing*, dan lain-lain. Bidang basis data, yang tentunya hampir 90 % melibatkan pencarian didalam prosesnya, bidang grafik, seperti *rendering*, *edge detection* dan lain-lain yang secara tidak langsung juga melibatkan proses pencarian didalamnya. Konsep pencarian terutama di dalam SI (baca: Sistem Informasi) telah banyak mempergunakan konsep basis data didalamnya. Basis data merupakan konsep penyimpanan informasi yang terpilah menjadi data-data kecil (*field*) yang tentunya berguna untuk mempercepat model pencarian. Hal ini sangat baik sekali terutama untuk kebutuhan pencarian yang sangat cepat. Hanya saja permasalahannya bahwa konsep ini adalah konsep yang kaku. Penulis mengungkapkan hal ini karena menurut penulis kelemahan dalam basis data adalah pembuatan *query* dan pembentukan relasi antar *query* yang bersifat *Artificial* (Buatan) sehingga pengguna sistem merasa seolah-olah terbatas dengan aturan-aturan yang telah dibuat, tanpa memiliki kewenangan untuk melakukan modifikasi *query* sendiri.

### **2. TEMU KEMBALI INFORMASI**

*Information Retrieval* adalah suatu sistem yang menemukan (*retrieve*) informasi yang sesuai dengan kebutuhan pengguna dari kumpulan informasi secara otomatis. Salah satu aplikasi dari *Information Retrieval* adalah mesin pencari yang dapat diterapkan diberbagai bidang. Pada mesin pencari dengan *information retrieval* pengguna dapat memasukkan *query* yang bebas dalam arti kata *query* yang sesuai dengan bahasa manusia dan Sistem dapat menemukan dokumen yang sesuai dengan *query* yang ditulis oleh user.

Penulis memperkenalkan konsep sistem temu kembali (*Information Retrieval System*) dan sistem basis informasi (*Information base system*) yang konsepnya hampir sama dengan basis data, yaitu menyimpan informasi didalamnya (bukan ekstraksi data). Konsep ini tentunya memiliki kebebasan dalam berekspresi baik dalam penyimpanannya maupun dalam *query* yang diberikan. Paper ini penulis tidak akan mengutarakan lebih jauh tentang konsep yang penulis ajukan (*Information base*) karena akan penulis berikan lebih lanjut pada penelitian penulis berikutnya. Sistem temu kembali informasi melakukan proses-proses tertentu untuk mendapatkan hasil yang baik, beberapa proses yang telah kita kenal adalah *string gathering* atau pemilahan kata-kata dalam sekumpulan *sentence*, penghilangan kata-kata yang memiliki frekuensi kemunculan sangat tinggi (*stoplist*) seperti *in*, *are*, *go*, *is*, *were*, *without*, *break*, dan lain-lain, proses *stemming*, yaitu menghilangkan *suffix*, *prefix* dan

*infix* dari suatu kata, sehingga menjadi kata dasar, yang sangat baik untuk normalisasi dan penyamaan arti kata nantinya. Proses pemberian bobot dan pembuatan index kata.

### 3. METODOLOGI PENELITIAN

*Customer gathering* yang dicoba di angkat dalam penelitian ini menerapkan konsep pengolahan bahasa alami dalam pencarian dokumen yang sesuai dengan yang diinginkan pengguna (pegawai bank, toko, pemerintahan dan masih banyak lagi). Pengguna layanan (*customer*) memberikan masukan berupa data-data yang diberikan oleh perusahaan (gambar 1). Setelah data-data tersebut terkumpul, maka *sistem customer gathering* akan membuat suatu daftar *parsing* terhadap total dokumen pelanggan yang masuk ke sistem. Proses *parsing* kemudian dilanjutkan dengan pembentukan *inverted file* dan *clustering* terhadap kelompok data-data berdasarkan kritik atau masukan tertentu.

Jika mencari dokumen dalam *inverted file*, pengguna memasukkan *query* yang terdiri dari kumpulan kata-kata yang akan dihubungkan secara otomatis menggunakan struktur data yang telah dibuat penulis. *Query* ini dipecah menjadi beberapa kata dan dicari pada *inverted file* tanpa melalui proses stemmed. Kemudian setiap dokumen yang ditemukan dihitung nilai kesamaannya dengan query yang dimasukkan. Nilai ini disebut dengan *similarity coefficient (SC)*. Setelah itu dokumen diurutkan mengecil berdasarkan nilai SC. Hasilnya ditampilkan kepada pengguna.

Idf dapat ditentukan dengan menggunakan rumus:

$$idf_j = \log \left[ \frac{d}{df_j} \right]$$

dengan ketentuan:

Idf<sub>j</sub> = *inverse document frequency* t<sub>j</sub>

d = total dokumen

df<sub>j</sub> = jumlah dokumen yang mengandung istilah t<sub>j</sub>

$$SC(Q, D_i) = \frac{\sum_{j=1}^l (w_{qj} * d_{ij})}{\sqrt{\sum_{j=1}^l (w_{qj})^2 \sum_{j=1}^l (d_{ij})^2}}$$

dimana:

w<sub>qj</sub> = bobot istilah j pada *query* q = frek<sub>qj</sub> \* idf<sub>j</sub>

d<sub>ij</sub> = bobot istilah j pada dokumen i = tf<sub>ij</sub> \* idf<sub>j</sub>

tf<sub>ij</sub> = *term frequency* = kemunculan istilah t<sub>j</sub> pada dokumen D<sub>i</sub>

untuk medapat persamaan di atas didapat dari model ruang vektor. Vektor adalah himpunan semua garis berarah yang besar dan arahnya sama. Contoh vektor adalah gaya, percepatan, kecepatan.

Vektor  $\vec{AB}$  mempunyai besar  $|\vec{AB}|$  dan arahnya

$\vec{AB}$ . Vektor  $\vec{AB}$  dapat dinyatakan dalam bentuk pasangan bilangan  $\left( \frac{p}{q} \right)$  yang mana:

p menyatakan pergeseran sejajar sumbu X dan q menyatakan pergeseran sejajar sumbu Y.

berdasarkan hukum pythagoras maka besar  $\left( \frac{p}{q} \right)$

adalah  $|\vec{AB}| = \sqrt{p^2 + q^2}$  jadi besar  $|\vec{AB}| = \left( \frac{p}{q} \right)$

adalah  $\sqrt{p^2 + q^2}$

model ruang vektor dengan normalisasi Euclidean dapat dilihat pada persamaan dibawah ini:

$$r = \frac{(Q, D)}{\|Q\| \|D\|} = \cos(\theta)$$

keterangan:

Q = *query*

D = *document*

θ = sudut antara garis horizontal dengan vektor

$\vec{AB}$

Sebagai contoh, misalkan ada sebuah *query* (Q) dan sebuah koleksi dokumen yang terdiri dari 3 buah dokumen (D<sub>1</sub>, D<sub>2</sub>, D<sub>3</sub>).

Q = “mutu pelayanan ditingkatkan”

D<sub>1</sub> = “pelayanan sudah cukup bagus”

D<sub>2</sub> = “ mutu agak kurang tetapi pelayanan sudah ditingkatkan”

D<sub>3</sub> = “ mutu pada pelayanan sudah meningkat” (1)

Jika sebuah istilah muncul pada sebuah dokumen, maka idf =  $\log \left[ \frac{3}{1} \right] = 0,477$  jika sebuah

istilah muncul pada 2 buah dokumen, maka idf =  $\log \left[ \frac{3}{2} \right] = 0,176$ . Jika sebuah istilah muncul pada semua

dokumen maka idf =  $\log \left[ \frac{3}{3} \right] = 0$ . Jika nilai W<sub>qj</sub> sama (2)

dengan idf<sub>j</sub> maka perhitungan SC untuk setiap dokumen adalah sebagai berikut:

SC (Q, D<sub>1</sub>) =

(0.176)(0.176)+(0.477)(0)+(0.176)(0) = 0.031

SC (Q, D<sub>2</sub>) =

(0.176)(0)+(0.477)(2\*0.477)+(0.176)(0.176) =

0.48

SC (Q, D<sub>3</sub>) =

(0.176)(0.176)+(0.477)(0)+(0.176)(0.176) =

0.062

Proses perankingan akan menghasilkan dokumen dengan urutan D<sub>2</sub>, D<sub>3</sub>, D<sub>1</sub>.

#### 4. IMPLEMENTASI

Metodologi penelitian telah dirancang untuk membentuk perangkat lunak *customer gathering*, pada bab berikut akan diberikan uraian singkat implementasi perangkat lunak yang telah dibuat. Tidak semua *interface* perangkat lunak ditampilkan dalam makalah ini, dikarenakan ada lebih dari 15 modul, sehingga menurut penulis hanya ditampilkan yang berhubungan dengan permasalahan di atas. Penelitian yang dilakukan adalah membuat suatu mesin pencari untuk mengetahui informasi yang diberikan oleh konsumen terhadap perusahaan. Pertama kali konsumen diminta untuk mengisi form seperti gambar 1 di bawah ini.

Gambar 1. Masukan data pelanggan

setelah itu maka data yang disikan akan disimpan dalam sistem basis data seperti pada gambar 2.

tabel	Nama	Kesan1	Kesan2	Kesan3
137	Dr. H. Hermet Lubis, MA	sebaik	ditingkatkan sistem ATM	-
138	Dr. P. P. Sudis MSc	system layanan selain penabung perlu peningkatan dan informasi	perlu sosialisasi peran daerah masih ditingkatkan lagi pelayanan	-
139	Dr. K. Sunanto Goenardi	Pelayanan terhadap nasabah baik	perlu sosialisasi peran daerah masih ditingkatkan lagi pelayanan	-
140	Dr. Kusniantardi Setya C	perlu sosialisasi peran daerah masih ditingkatkan lagi pelayanan	perlu sosialisasi peran daerah masih ditingkatkan lagi pelayanan	-
141	Dr. Tri Atmaja	perlu sosialisasi peran daerah masih ditingkatkan lagi pelayanan	perlu sosialisasi peran daerah masih ditingkatkan lagi pelayanan	-
142	Dr. Tri Ermin Fadila	- banyak kantor cabang, lokasi de	- kemudahan yang lain	-
143	Dr. Yoseph Wibowo S	kecepatan layanan	kemudahan - kemudahan yang lain	-
144	Dr. Djatmiko T Soedomo	kemudahan dalam transaksi dan kadang - kadang antrian terlalu lama	kemudahan - kemudahan yang lain	-
145	Dr. H. Siti Simamir	layanan baik	kemudahan - kemudahan yang lain	-
146	Dr. Pandjaj	Proses mendapat kredit dengan c	Jika mungkin bunga pinjaman/ tabun	Usaha memperoleh jaringan lai
147	Dr. Siti Pujiastuti	jenis - jenis kredit	ATM di Kulon Progo	perlu sosialisasi peran daerah masih ditingkatkan lagi pelayanan
148	Dr. Andar Sari	1. keramah tamahan karyawan pe	1. minimal tiap - tiap cabang harus di	1. bunga pinjaman diminimalk
149	Dr. Baq Nur Hasan Ke	kekeluargaan	ATM dan kartu kredit lebih bisa dimi	perlu sosialisasi peran daerah masih ditingkatkan lagi pelayanan
150	Dr. Enry Indiyati	saya pribadi puas dengan layan	-	-
151	Dr. Hj. Neor Djannah B	perhatianker terus apa - apa yan	- penukaran valas di tingkat Kecamatan	- pelayanan ATM
152	Dr. Isti Yurnanah	Pelayanan dan keramah tamahan	Pelayanan yang lebih prima	Disamping memberikan Vie ag
153	Dr. Neor Hidayati	- baik, kekeluargaan, ramah	- pelayanan masih kurang cepat, mo	- tambah ATM, disus kantor B
154	Dr. Pratiwi	pelayanan nasabah cukup ramah	-	- di kantor kas (prambanan) seni
155	Dr. Siti Hibana	- pelayanan yang memasyarakat	-	-
156	Dr. Siti Solihah	pelayanan prima bagi nasabah	-	-
157	Dr. Siti Suwabih	pelayannya agar ditingkatkan	-	-
158	Dr. Subati	-	-	-
159	Dr. Tri Wiryanti, SH	pelayanan cepat, staf yang cukup	rolling pegawai yang terlalu cepat	ATM diperbanyak
160	Dr. Yuryanti	- pelayanan yang cepat dan ramal	ATM diperbanyak	buatan ATM untuk semua

Gambar 2. Informasi yang tersimpan di basis data

dari basis data diatas akan dibuat suatu file berbentuk teks yang terdiri dari banyak dokumen. Tiap tiap dokumen dalam teks tersebut terdiri dari nomor dokumen nama nasabah dan kesan kesan terhadap perusahaan, seperti gambar 3. Proses pembentukan teks tersebut dilakukan pada kolom-kolom tertentu (gambar 2) yang berhubungan dengan kasus *retrieval* terhadap informasi yang akan dicari dan di *cluster*. Kolom-kolom tersebut adalah no dokumen, nama nasabah, kesan dan pesan nasabah.

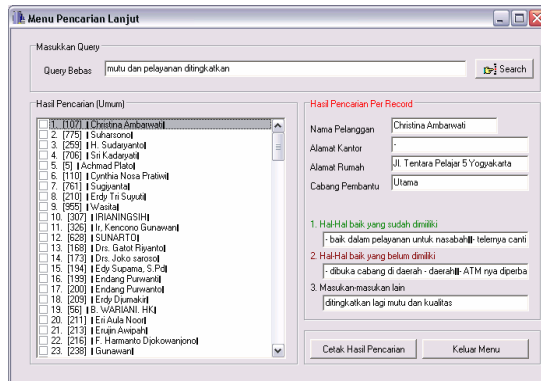
Gambar 3. Pembentukan teks eksternal untuk *pra processing* temu kembali informasi

dari dokumen di atas akan diolah menjadi *inverted file* yang isinya dapat dilihat pada gambar 4. Proses pembentukan tersebut dilakukan sebagai fungsi atomic penunjuk (pointer) antar kata dan dokumen dan kemunculan kata tersebut di dokumen teks yang telah dibentuk.

Gambar 4. *Inverted file* yang telah dibentuk menggunakan model Tf.Idf

file diatas merupakan file yang sudah dilakukan pembobotan untuk mendapatkan tf.idf nya.

setelah nilai idf nya diperoleh maka dilakukan perangkian dari yang terbesar sampai yang terkecil (*descending*). Sistem juga menyediakan form untuk mesin pencari dalam menemukan dokumen yang relevan dengan *query* yang diinginkan oleh pengguna (gambar 5).



**Gambar 5.** Form utama pencarian data pelanggan berdasar kritik yang telah dibentuk

Form pencarian natural pada gambar lima terdiri dari pemasukan query, hasil pencarian terhadap seluruh nasabah yang ditemukan yang sesuai dengan query yang diinginkan oleh pengguna, dan juga informasi lengkap tentang data nasabah termasuk informasi kritik dan masukan-masukan yang diinginkan terhadap perusahaan yang bersangkutan. Dari *query* yang dimasukkan maka akan dilakukan proses pembobotan terhadap dokumen yang ada. Pembobotan otomatis biasanya berdasarkan istilah yang sering muncul dalam sebuah dokumen atau dikenal dengan istilah *term frequency(tf)* dan jumlah kemunculannya dalam koleksi dokumen disebut *inverse document frequency(idf)*. Bobot suatu istilah makin kecil jika istilah tersebut muncul dalam banyak dokumen dan makin besar jika sering muncul pada suatu dokumen.

Untuk mengukur efektifitas dua rasio umum yang biasa digunakan adalah precision dan recall. Precision adalah ukuran kemampuan suatu sistem untuk menampilkan hanya dokumen relevan. Recall adalah kemampuan suatu sistem untuk menampilkan seluruh dokumen yang relevan.

$$Precision = \frac{\text{jumlah dokumen relevan yang berhasil ditemukan}}{\text{Jumlah dokumen yang ditemukan}}$$

$$Recall = \frac{\text{jumlah dokumen relevan yang berhasil ditemukan}}{\text{Jumlah dokumen relevan dalam koleksi}}$$

Nilai *recall* dan *precision* selalu berbanding terbalik, semakin tinggi nilai *recall* semakin rendah nilai *precision*, begitu juga sebaliknya semakin rendah nilai *recall* semakin tinggi nilai *precision*. *Precision* dapat dihitung pada berbagai titik *recall*. Secara umum, semakin tinggi nilai *recall* semakin banyak jumlah dokumen yang harus dicari. Pada mesin pencarian yang sempurna, hasil pencarian semuanya merupakan dokumen yang relevan atau dengan kata lain pada setiap nilai *recall*, nilai *precision* selalu 1.00. pada kenyataannya, ada dokumen yang tidak relevan juga diambil oleh mesin pencari.

## 5. PENUTUP

Model yang diusulkan pada pencarian dokumen bagi review atau peningkatan kualitas layanan dari sesuatu bidang layanan. Terkadang suatu perusahaan tidak menghiraukan mekanisme perawatan (maintenance) masukan dari pengguna. Dengan model ini maka pegawai dapat dengan mudah memarsing dan mencari data nasabah sesuai dengan bidang-bidang masukan dan kritik tertentu secara otomatis, tidak digunakan cara-cara yang manual dalam pengelompokannya. Hasil yang didapat dengan model yang diusulkan ini memiliki tingkat akurasi pengelompokan cukup baik (sekitar 89 persen untuk tingkat recall dan 60 persen untuk tingkat precisionnya). Diperkirakan faktor stemming berpengaruh terhadap hasil pembentukan kelompok data, walaupun menurut penulis hanya meningkatkan sekitar 2-5 persen saja (cukup kecil).

## DAFTAR PUSTAKA

- [FRA92] Frakes William and Ricardo Baeza-Yates; *Information Retrieval Data Structure and Algorithms*, Prentice Hall, 1992.
- [GRO98] Grossman David, and Ophir Frieder, *Information Retrieval : Algorithms and Heuristics*, Kluwer Academic Publisher, 1998.
- [MIY90] Miyamoto, Sadaki; *Fuzzy sets in Information Retrieval and cluster analysis*; Kluwer Academic Publisher; London, 1990.
- [ERW04] Erwin, Muhammad, “*Relevance feedback pada sistem temu kembali informasi menggunakan algoritma genetika*”, Thesis Magister Informatika ITB, 2004.
- [MAN99] Mandala Rila, Takenobu Takunaga, Hozumi Tanaka. “*Query expansion using heterogenous thesauri*”. “Proceeding of Information Processing and Management. 1999.
- [MAN00] Mandala Rila, Takenobu Takunaga, Hozumi Tanaka. “*The exploration and Analysis of Using Multiple Thesaurus types for Query Expansion in Information Retrieval*”. Journal of Information Processing. 2000.
- [MAN02] Mandala Rila, “*Sistem Temu-kembali informasi dengan menggunakan model probabilistik*” Jurnal Informatika, ITB, Bandung, 2002.
- [SET03] Setiawan, Kuswara. “*Paradigma Sistem Cerdas*”. BayuMedia Publishing, Malang, Jawa Timur. 2003.