Kode/ Nama Rumpun Ilmu : 458 / Teknik Informatika

# LAPORAN PENELITIAN

# HIBAH INTERNAL STIKI

# SISTEM TEMU BALIK INFORMASI DENGAN METODE K-MEANS

**TIM PENELITI :**
**Daniel Rudiaman S., S.T, M.Kom (NIDN: 0722037101)**
**Anita, S.Kom, M.T (NIDN: 0707077201)**

**SEKOLAH TINGGI INFORMATIKA & KOMPUTER INDONESIA**
**AGUSTUS, 2016**

# HIBAH INTERNAL STIKI

# SISTEM TEMU BALIK INFORMASI DENGAN METODE K-MEANS

**TIM PENELITI :**
**Daniel Rudiaman S., S.T, M.Kom (NIDN: 0722037101)**
**Anita, S.Kom, M.T (NIDN: 0707077201)**

**SEKOLAH TINGGI INFORMATIKA & KOMPUTER INDONESIA**
**AGUSTUS, 2016**

## HALAMAN PENGESAHAN
## PENELITIAN HIBAH INTERNAL

**Judul Penelitian** : **Sistem Temu Balik Informasi dengan Metode K-Means**

**Ketua Peneliti** :
**a. Nama Lengkap** : **Daniel Rudiaman S., S.T, M.Kom**
**a. NIDN** : **0722037101**
**b. Jabatan Fungsional** : **Asisten Ahli**
**c. Program studi** : **Teknik Informatika**
**d. Nomor HP** : **081334289205**
**e. Alamat Surel (e-mail)** : **daniel223@stiki.ac.id**

**Anggota Peneliti (1)**
**f. Nama Lengkap** : **Anita, S.Kom,MT**
**b. NIDN** : **0707077201**
**c. Perguruan Tinggi** : **Sekolah Tinggi Informatika & Komputer Indonesia (STIKI)**

**Biaya Penelitian** : **Rp. 1.500.000,-**

**Malang, 25 Agustus 2016**

**Mengetahui,**
**Ka.Prodi TI**

**Daniel Rudiaman S., S.T, M.Kom**
**NIDN. 0722037101**

**Ketua Peneliti,**

**Daniel Rudiaman S., S.T, M.Kom**
**NIDN. 0722037101**

**Menyetujui**
**Ketua LPPM**

**Subari, M.Kom**
**NIDN. 0702027201**

# DAFTAR ISI

# RINGKASAN

Daniel Rudiaman S., ST, M.Kom dan Anita, S.Kom, MT, "**Sistem Temu Balik Informasi dengan Metode K-Means**", Hibah Internal STIKI, 2015. Teknik Informatika, Sekolah Tinggi Informatika & Komputer Indonesia (STIKI) Malang.

Keyword : Sistem Temu Balik Informasi, Clustering, K-Means

Sistem Temu Balik Informasi *(Information Retrieval System)* adalah sistem yang secara otomatis melakukan pencarian atau penemuan kembali informasi yang relevan terhadap kebutuhan pengguna. Kebutuhan pengguna, diekspresikan dalam *query*, menjadi input bagi sistem dan selanjutnya sistem akan mencari dan menampilkan dokumen yang relevan dengan *query* tersebut.

Pada saat menggunakan sistem temu balik informasi, pengguna ingin mendapat sekumpulan informasi yang relevan dengan query yang diberikannya pada sistem. Penggunaan metode clustering pada sistem temu balik informasi akan memungkinkan sistem memberikan sekumpulan informasi yang relevan dengan kebutuhan pengguna. Dokumen yang ada pada cluster yang sama memiliki kemiripan dalam hal relevansi nya terhadap kebutuhan informasi. Apabila salah satu dokumen yang ada pada suatu cluster relevan dengan query yang diberikan pengguna, maka besar kemungkinan dokumen lain yang ada pada cluster tersebut juga relevan bagi pengguna.

K-Means adalah salah satu metode clustering yang dapat diterapkan pada Sistem Temu Balik Informasi. Pada penelitian ini, akan dirancang model penggunaan metode K-Means untuk melakukan clustering informasi pada sebuah Sistem Temu Balik Informasi.

# BAB I PENDAHULUAN

## 1.1. Latar Belakang

Sistem Temu Balik Informasi (*Information Retrieval System*) adalah sistem yang secara otomatis melakukan pencarian atau penemuan kembali informasi yang relevan terhadap kebutuhan pengguna. Kebutuhan pengguna, diekspresikan dalam *query*, menjadi input bagi sistem dan selanjutnya sistem mencari dan menampilkan dokumen yang relevan dengan query tersebut.

Pada saat menggunakan sistem proses temu balik informasi, pengguna ingin mendapat sekumpulan informasi yang relevan dengan *query* yang diberikannya pada sistem. Pada umumnya hasil pencarian akan ditampilkan berupa daftar hasil pencarian. Pengguna akan menelusuri tiap hasil pencarian untuk mendapatkan informasi yang sesuai dengan kebutuhannya. Penggunaan metode *clustering* pada sistem temu balik informasi akan memungkinkan sistem memberikan sekumpulan informasi yang relevan dengan kebutuhan pengguna. Sistem *clustering* akan mengelompokkan informasi-informasi yang mirip pada *cluster* yang sama. Dokumen yang ada pada *cluster* yang sama memiliki kemiripan dalam hal relevansi nya terhadap kebutuhan informasi. Pada saat penguna melakukan proses *query*, apabila salah satu dokumen yang ada pada suatu cluster relevan dengan *query* yang diberikan pengguna, maka besar kemungkinan bahwa dokumen lain yang ada pada *cluster* tersebut juga relevan bagi pengguna.

K-Means adalah salah satu metode clustering yang dapat diterapkan pada Sistem Temu Balik Informasi. Metode ini meminimalkan rata-rata kuadrat jarak Euclidean tiap dokumen

dari pusat cluster. Pada penelitian ini akan dirancang model penggunaan metode K-Means untuk melakukan *clustering* informasi pada sebuah Sistem Temu Balik Informasi.

## 1.2. Tujuan

Tujuan yang ingin dicapai dalam penelitian ini adalah menghasilkan sebuah model sistem temu balik informasi yang dapat memberikan sekumpulan informasi yang relevan dengan pencarian yang dilakukan pengguna.

## 1.3. Hipotesis

Berdasarkan analisa awal yang dilakukan pada penelitian ini diberikan hipotesa awal bahwa sistem temu balik informasi dengan metode K-Means akan dapat memberikan hasil pencarian yang lebih relevan dengan kebutuhan pengguna.

## 1.4. Ruang lingkup

Ruang lingkup penelitian dibatasi pada pembuatan sebuah model sistem temu balik informasi dengan metode K-Means.

### 1.4.1. Fitur-fitur perangkat lunak

Secara rinci fitur-fitur yang disediakan pada perangkat lunak adalah sebagai berikut :

- Sistem input dokumen
- Sistem pencarian dengan menggunakan query biasa
- Sistem pencarian dengan menggunakan clustering K-Means

### 1.4.2. Batasan pembuatan perangkat lunak

Batasan-batasan dalam pembuatan perangkat lunak adalah :

a. Perangkat lunak yang dibangun merupakan model dan belum diterapkan pada lingkungan yang sebenarnya
b. Penelitian tidak mengerjakan aspek perangkat keras dan jaringan komputer

### 1.4.3. Target uji coba perangkat lunak

Uji coba perangkat lunak dilakukan secara bertahap, yaitu uji modul, uji fitur dan uji sistem. Uji coba dilakukan dengan memasukkan data sample dan menguji hasilnya, apakah memberikan hasil sesuai dengan rancangan.

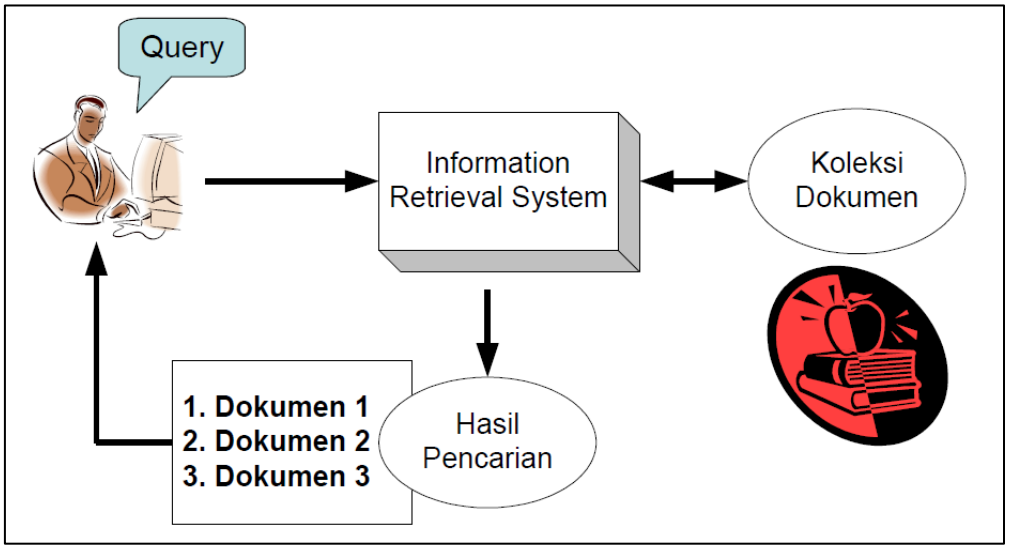
## 1.5. Luaran yang Ingin Dicapai

Target luaran yang diharapkan dengan adanya penelitian ini adalah terbangunnya model sistem temu balik informasi yang dapat digunakan memberikan hasil pencarian yang relevan dengan kebutuhan pengguna.

# BAB II TINJAUAN PUSTAKA

## 2.1 Pengertian Sistem Temu Balik Informasi

Sistem Temu Balik Informasi (*Information Retrieval System*) digunakan untuk menemukan kembali (*retrieve*) informasi-informasi yang relevan terhadap kebutuhan pengguna dari suatu kumpulan informasi secara otomatis.




**Gambar 2.1 Sistem Temu Balik Informasi**

Salah satu aplikasi umum dari Sistem Temu Balik Informasi adalah *search engine* atau mesin pencarian yang terdapat pada jaringan internet. Pengguna dapat mencari halaman-halaman web yang dibutuhkannya melalui *search engine*. Contoh lain dari Sistem Temu Balik Informasi adalah sistem informasi perpustakaan.

Sistem Temu Balik Informasi terutama berhubungan dengan pencarian informasi yang isinya tidak memiliki struktur. Demikian pula ekspresi kebutuhan pengguna yang disebut *query*, juga tidak memiliki struktur. Hal ini yang membedakan Sistem Temu Balik Informasi dengan sistem basis data. Dokumen adalah contoh informasi yang tidak terstruktur. Isi dari suatu dokumen sangat tergantung pada pembuat dokumen tersebut.

Proses yang terjadi di dalam *Information Retrieval System* terdiri dari 2 bagian utama, yaitu *Indexing* (*Text Preprocessing*) dan *Searching (Similarity Measure/Vector Space Model).*

## 2.2 Text preprocessing

*Text preprocessing* atau sering disebut juga proses *indexing,* merupakan tahapan awal pada proses merepresentasikan koleksi dokumen ke dalam bentuk tertentu untuk memudahkan dan mempercepat proses pencarian dan penemuan kembali dokumen yang relevan. Pembangunan index dari koleksi dokumen merupakan tugas pokok pada tahapan *preprocessing* di dalam IR. Kualitas index mempengaruhi efektifitas dan efisiensi sistem IR.

Index dokumen adalah himpunan term yang menunjukkan isi atau topik yang dikandung oleh dokumen. Index akan membedakan suatu dokumen dari dokumen lain yang berada di dalam koleksi. Proses *indexing* harus melibatkan konsep *linguistic processing* yang bertujuan mengekstrak *term-term* penting dari dokumen yang direpresentasikan sebagai *bag-of-words*. Ekstraksi term dibagi menjadi beberapa proses di antaranya : *case folding, tokenization, filtering , dan stemming.*

## 2.3 Case Folding dan Tokenization

*Case folding* tahap di mana merubah semua kata pada dokumen menjadi huruf kecil. Sedangkan Tokenization adalah tugas memisahkan deretan kata di dalam kalimat, paragraf atau halaman menjadi token atau potongan kata tunggal atau termmed word. Tahapan ini juga menghilangkan karakter-karakter tertentu seperti tanda baca.

## 2.4 Filtering

*Stop-word* didefinisikan sebagai *term* yang tidak berhubungan (*irrelevant*) dengan subyek utama dari *database* meskipun kata tersebut sering kali hadir di dalam dokumen. Berikut ini adalah contoh *stop words* dalam bahasa inggris :*a, an, the, this, that, these, those, her, his, its, my, our, their, your, all, few, many, several, some, every, for, and, nor, bit, or, yet, so, also, after, although, if, unless, because, on, beneath, over, of, during, beside,* dan *etc*. Contoh *stop words* dalam bahasa Indonesia : yang, juga, dari, dia, kami, kamu, aku, saya, ini, itu, atau, dan, tersebut, pada, dengan, adalah, yaitu, ke, tak, tidak, di, pada, jika, maka, ada, pun, lain, saja, hanya, namun, seperti, kemudian, dam lain-lain.

## 2.5 Stemming

*Stemming* adalah proses untuk memecahkan setiap varian-varian suatu kata menjadi kata dasar. *Stem* (akar kata) adalah bagian dari kata yang tersisa setelah dihilangkan imbuhannya (awalan dan akhiran), contohnya kata *connect* adalah stem dari *connected*, *connecting*, *connection*, dan *connections*. Metode *stemming* memerlukan input berupa *term* yang terdapat dalam dokumen. Sedangkan outputnya berupa *stem*. Sebagai contoh dalam bahasa indonesia kata bersama, kebersamaan, menyamai, akan distem ke root wordnya yaitu "sama".

Algoritma *stemming* untuk bahasa yang satu berbeda dengan algoritma stemming untuk bahasa lainnya. Sebagai contoh bahasa Inggris memiliki morfologi yang berbeda dengan bahasa Indonesia sehingga algoritma *stemming* untuk kedua bahasa tersebut juga berbeda. Proses *stemming* pada teks berbahasa Indonesia lebih rumit/kompleks karena terdapat variasi imbuhan yang harus dibuang untuk mendapatkan *root word* (kata dasar) dari sebuah kata. Pada umumnya kata dasar pada bahasa Indonesia terdiri dari kombinasi: *Prefiks 1 + Prefiks 2 + Kata dasar + Sufiks 3 + Sufiks 2 + Sufiks 1*.

Ada beberapa algoritma untuk melakukan stemming contohnya porter dan ‘nazief dan adriani’. Salah satu algoritma yang paling akurat untuk stemming bahasa Indonesia adalah algoritma nazief dan adriani. Proses *stemming* dalam bahasa Indonesia ini lebih kompleks dari algoritma lain, karena terdapat berbagai macam variasi serta kombinasi imbuhan yang harus dihapus untuk mendapatkan kata dasar.

## 2.6 Term Weighting TF.iDf(Pembobotan)

Metode TF-IDF merupakan metode pembobotan *term* yang banyak digunakan sebagai metode pembanding terhadap metode pembobotan baru. Pada metode ini, perhitungan bobot *term t* dalam sebuah dokumen dilakukan dengan mengalikan nilai *Term Frequency* dengan *Inverse Document Frequency.*

Pada *Term Frequency* (TF), terdapat beberapa jenis formula yang dapat digunakan yaitu:

1. tf biner (*binery tf*), hanya memperhatikan apakah suatu kata ada atau tidak dalam dokumen, jika ada diberi nilai satu, jika tidak diberi nilai nol.
2. tf murni (*raw tf*), nilai tf diberikan berdasarkan jumlah kemunculan suatu kata di dokumen. Contohnya, jika muncul lima kali maka kata tersebut akan bernilai lima.
3. tf logaritmik, hal ini untuk menghindari dominasi dokumen yang mengandung sedikit kata dalam *query*, namun mempunyai frekuensi yang tinggi.

$$tf = 1 + log\ (tf)$$

4. tf normalisasi, menggunakan perbandingan antara frekuensi sebuah kata dengan jumlah keseluruhan kata pada dokumen.
5. *Inverse Document Frequency* (idf) dihitung dengan menggunakan formula

$$idf_j = log\ (D / df_j)$$

di mana

D adalah jumlah semua dokumen dalam koleksi

$df_j$ adalah jumlah dokumen yang mengandung term $t_j$

Jenis formula yang akan digunakan untuk perhitungan *term frequency* (TF) yaitu tf murni (*raw tf*). Dengan demikian rumus umum untuk TF-IDF adalah penggabungan dari formula perhitungan *raw tf* dengan formula *idf* dengan cara mengalikan nilai *term frequency (*TF*)* dengan nilai *inverse document frequency* (IDF).

$$w_{ij} = tf_{ij} \times idf_j$$
$$w_{ij} = tf_{ij} \times log\ (D / df_j)$$

Keterangan :

$w_{ij}$ adalah bobot term $t_j$ terhadap dokumen $d_i$

$tf_{ij}$ adalah jumlah kemunculan term $t_j$ dalam dokumen $d_i$

D adalah jumlah semua dokumen yang ada

$df_j$ adalah jumlah dokumen yang mengandung term $t_j$ (minimal ada satu kata yaitu term $t_j$)

## 2.7 Vector Space Model

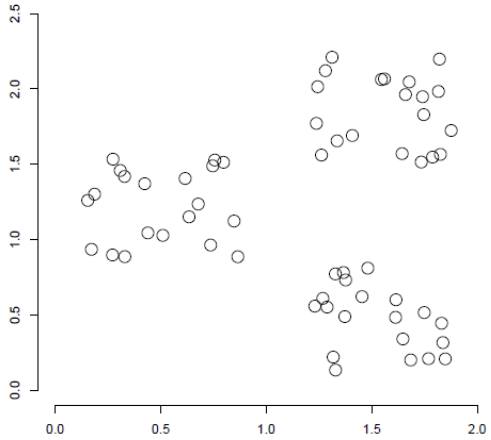
Vektor Space Model adalah model sistem temu balik informasi yang mengibaratkan masing-masing query dan dokumen sebagai sebuah vektor n-dimensi. Tiap dimensi pada vektor tersebut diwakili oleh satu term. Term yang digunakan biasanya berpatokan kepada term yang ada pada query atau keyword, sehingga term yang ada pada dokumen tetapi tidak ada pada query biasanya diabaikan.

Pada *Information Retrieval System* terdapat beberapa metode yang digunakan dalam *searching* salah satunya adalah dengan merepresentasikan proses *searching* menggunakan model ruang vektor. Model ruang vektor dibuat berdasarkan pemikiran bahwa isi dari dokumen ditentukan oleh kata-kata yang digunakan dalam dokumen tersebut. Model ini menentukan kemiripan (*similarity)* antara dokumen dengan *query* dengan cara merepresentasikan dokumen dan *query* masing-masing ke dalam bentuk vektor. Tiap kata yang ditemukan pada dokumen dan *query* diberi bobot dan disimpan sebagai salah satu elemen vektor.

Kemiripan antar dokumen didefinisikan berdasarkan representasi *bag-of-words* dan dikonversi ke suatu model ruang vektor (*Vector Space Model*, VSM).

## 2.8 Clustering pada Sistem Temu Balik Informasi

Clustering adalah bentuk paling umum dari pembelajaran tanpa pengawasan. Tidak ada pengawasan berarti bahwa tidak ada ahli yang telah mengelompokkan dokumen ke dalam kelas-kelas. Dalam clustering, distribusi dan kemiripan dari data yang akan menentukan keanggotaan cluster. Contoh sederhana adalah seperti digambarkan pada Gambar 2.2




**Gambar 2.2 Pengelompokan Data ke dalam Cluster**

## 2.9 K-Means

K-Means adalah salah satu metode clustering yang dapat diterapkan pada Sistem Temu Balik Informasi. Metode ini meminimalkan rata-rata kuadrat jarak Euclidean tiap dokumen dari pusat cluster, di mana pusat cluster dihitung dengan menggunakan persamaan 2.1 berikut.

$$\vec{\mu}(\omega) = \frac{1}{|\omega|} \sum_{\vec{x} \in \omega} \vec{x} \quad \text{.....(2.1)}$$

# BAB III METODE PENELITIAN

## 3.1. Metodologi penelitian

Metodologi penelitian yang digunakan adalah sebagai berikut :

a. Lokasi penelitian

Lokasi penelitian dilakukan di Sekolah Tinggi Informatika & Komputer Indonesia

b. Alat dan bahan

Alat yang digunakan adalah seperangkat komputer untuk melakukan dokumentasi, analisa dan pemodelan. Sedangkan bahan yang digunakan adalah sekumpulan dokumen yang akan dikelompokkan ke dalam cluster.

c. Pengumpulan data dan informasi

Pengumpulan data dilakukan dengan cara memilih sekumpulan dokumen dari dokumen yang ada di internet yang nantinya isinya akan diinputkan pada sistem yang dibuat.

d. Analisa data

Berdasarkan data yang terkumpul dilakukan analisa untuk menentukan data yang layak untuk dijadikan data yang akan digunakan pada pengujian sistem.

e. Prosedur penelitian

Penelitian dilakukan dengan menggunakan prosedur sebagai berikut :

- Planning,

  Pada tahapan ini dilakukan perencanaan terhadap tahapan pengerjaan sistem
- Analisa dilakukan untuk mencari sebab akibat dari permasalahan yang timbul dan menentukan alternatif pemecahan masalah
- Perancangan dibuat berdasarkan hasil analisa yang telah dilakukan yaitu dengan memodelkan permasalahan sehingga dokumen dapat dikelompokkan menjadi cluster oleh algoritma clustering K-Means.

- Konstruksi, yaitu pembuatan prototype berdasarkan rancangan.
- Testing, yaitu pengujian pada prototype yang dibuat.
- Implementasi, yaitu membuat sistem clustering dengan metode K-Means berbasis web yang dapat digunakan pada kondisi real.


**Gambar 3.1. Prosedur Penelitian**

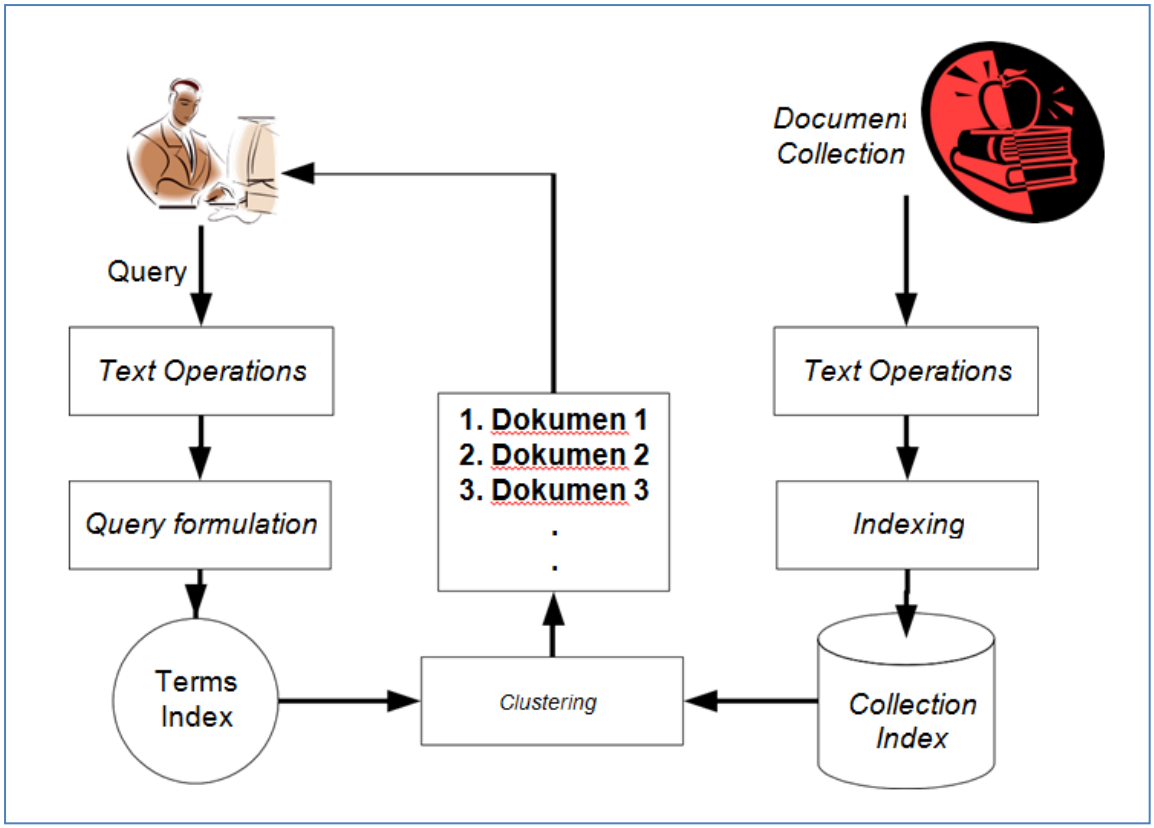
# BAB IV PERANCANGAN DAN PEMBAHASAN

Pada saat menggunakan sistem proses temu balik informasi, pengguna ingin mendapat sekumpulan informasi yang relevan dengan *query* yang diberikannya pada sistem. Pada umumnya hasil pencarian akan ditampilkan berupa daftar hasil pencarian. Pengguna akan menelusuri tiap hasil pencarian untuk mendapatkan informasi yang sesuai dengan kebutuhannya. Penggunaan metode *clustering* pada sistem temu balik informasi akan memungkinkan sistem memberikan sekumpulan informasi yang relevan dengan kebutuhan pengguna. Sistem *clustering* akan mengelompokkan informasi-informasi yang mirip pada *cluster* yang sama.

K-Means adalah salah satu metode clustering yang dapat diterapkan pada Sistem Temu Balik Informasi. Metode ini meminimalkan rata-rata kuadrat jarak Euclidean tiap dokumen dari pusat cluster. Pada penelitian ini akan dirancang model penggunaan metode K-Means untuk melakukan *clustering* informasi pada sebuah Sistem Temu Balik Informasi.

## 4.1 Pemodelan Sistem Temu Balik Informasi

Pemodelan sistem yang akan dibuat diberikan pada gambar 4.1 berikut.




**Gambar 4.1 Pemodelan Sistem**

## 4.2 Text Operation

*Text operation,* merepresentasikan koleksi dokumen ke dalam bentuk tertentu untuk memudahkan dan mempercepat proses pencarian dan penemuan kembali dokumen yang relevan. Proses ini melibatkan konsep *linguistic processing* yang bertujuan mengekstrak *term-term* penting dari dokumen yang direpresentasikan sebagai *bag-of-words*. Ekstraksi term dibagi menjadi beberapa proses di antaranya : *case folding, tokenization, filtering , dan stemming.*

*Case folding* merubah semua kata pada dokumen menjadi huruf kecil. Sedangkan Tokenization adalah tugas memisahkan deretan kata di dalam kalimat, paragraf atau halaman menjadi token atau potongan kata tunggal atau termmed word. Tahapan ini juga menghilangkan karakter-karakter tertentu seperti tanda baca.

*Filtering* bertujuan menghilangkan s*top-word* yang dianggap sebagai *term* yang tidak berhubungan (*irrelevant*) dengan subyek utama meskipun kata tersebut sering kali hadir di dalam dokumen.

*Stemming* berfungsi memecahkan setiap varian-varian suatu kata menjadi kata dasar. *Stem* (akar kata) adalah bagian dari kata yang tersisa setelah dihilangkan imbuhannya (awalan dan akhiran), contohnya kata *connect* adalah stem dari *connected*, *connecting*, *connection*, dan *connections*. Metode *stemming* memerlukan input berupa *term* yang terdapat dalam dokumen. Sedangkan outputnya berupa *stem*. Sebagai contoh dalam bahasa indonesia kata bersama, kebersamaan, menyamai, akan distem ke root wordnya yaitu "sama".

### 4.3 Term Weighting TF.iDf(Pembobotan)

*Inverse Document Frequency* (idf) dihitung dengan menggunakan formula

$$idf_j = log\ (D / df_j)$$

di mana

D adalah jumlah semua dokumen dalam koleksi

$df_j$ adalah jumlah dokumen yang mengandung term $t_j$

Jenis formula yang akan digunakan untuk perhitungan *term frequency* (TF) yaitu tf murni (*raw tf*). Dengan demikian rumus umum untuk TF-IDF adalah penggabungan dari formula perhitungan *raw tf* dengan formula *idf* dengan cara mengalikan nilai *term frequency (*TF*)* dengan nilai *inverse document frequency* (IDF).

$$w_{ij} = tf_{ij} \times idf_j$$
$$w_{ij} = tf_{ij} \times log\,(D/df_j)$$

Keterangan :

$w_{ij}$ adalah bobot term $t_j$ terhadap dokumen $d_i$

$tf_{ij}$ adalah jumlah kemunculan term $t_j$ dalam dokumen $d_i$

D adalah jumlah semua dokumen yang ada

$df_j$ adalah jumlah dokumen yang mengandung term $t_j$ (minimal ada satu kata yaitu term $t_j$)

## 4.4 Vector Space Model

Masing-masing query dan dokumen akan direpresentasikan sebagai sebuah vektor n-dimensi.

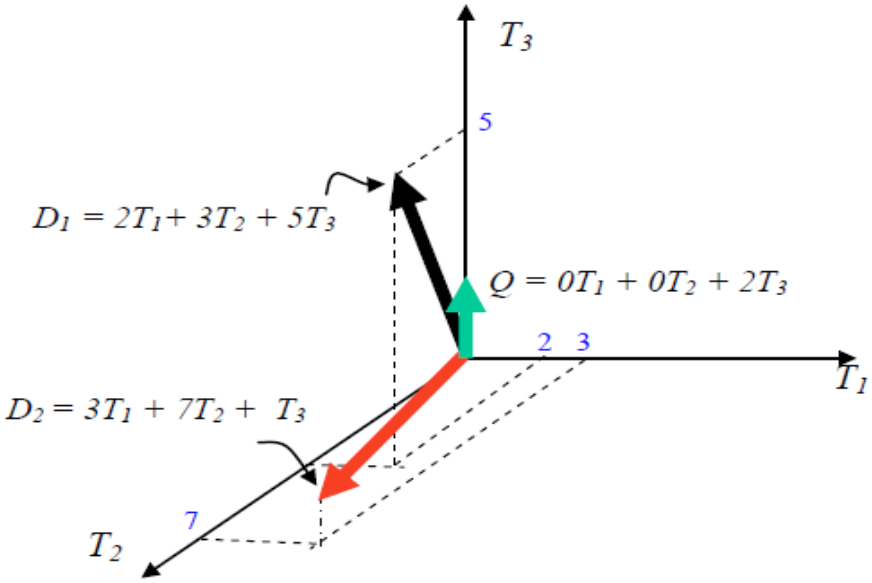
Sebagai contoh terdapat 3 buah kata ($T_1$, $T_2$ dan $T_3$), 2 buah dokumen ($D_1$ dan $D_2$) serta sebuah *query Q*. Masing-masing bernilai :

$D_1 = 2T_1+3T_2+5T_3$

$D_2 = 3T_1+7T_2+0T_3$

$Q = 0T_1+0T_2+2T$

Maka representasi grafis dari ketiga vektor ini adalah seperti pada gambar 4.2 berikut.




**Gambar 4.2 Ilustrasi VSM pada *query* dan dokumen**

## 4.5 Clustering pada Sistem Temu Balik Informasi

Setiap dokumen akan dianggap sebagai sebuah vektor pada n-dimensi. Dokumen-dokumen yang mirip akan memiliki kata-kata yang sama, dan kata-kata itu akan muncul dengan frekuensi yang tinggi dan letaknya akan berdekatan pada ruang berdimensi n. Metode clustering K-Means akan mengelompokkan dokumen-dokumen yang mirip ke dalam cluster yang sama berdasarkan kedekatannya dengan pusat cluster.

Untuk menentukan ukuran kedekatan digunakan Euclidean Distance. Jarak Euclidean antara titik p dan q adalah panjang segmen garis yang menghubungkan kedua titik tersebut. Dalam koordinat Cartesian, jika **p** = (p1, p2, ..., pn) dan **q** = (q1, q2, ..., qn) adalah dua poin di Euclidean n-space, maka jarak (d) dari p ke q, atau dari q ke p diberikan oleh rumus Pythagoras sebagai berikut:

$$\mathrm{d}(\mathbf{p}, \mathbf{q}) = \mathrm{d}(\mathbf{q}, \mathbf{p}) = \sqrt{(q_1 - p_1)^2 + (q_2 - p_2)^2 + \cdots + (q_n - p_n)^2}$$

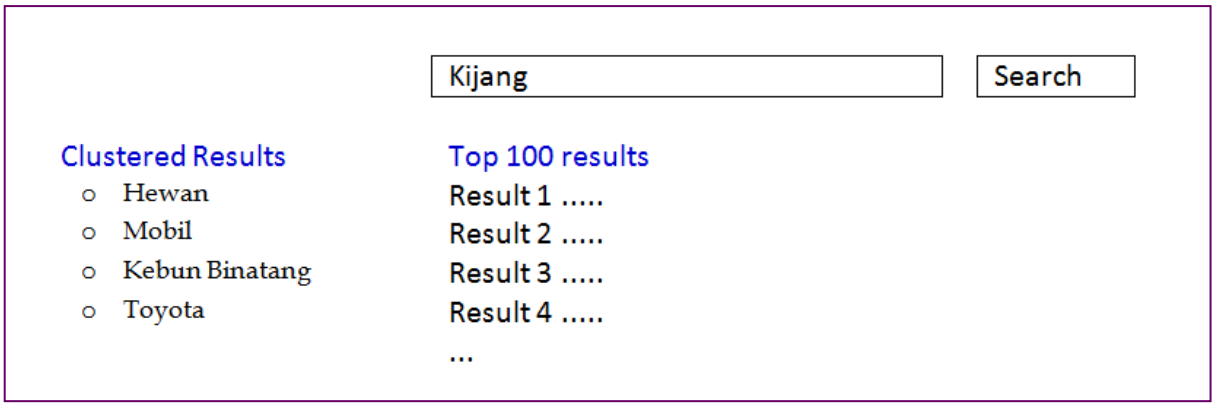
$$= \sqrt{\sum_{i=1}^{n} (q_i - p_i)^2}.$$

Penghitungan pusat cluster dapat dilakukan dengan menggunakan algoritma berikut ini.

K-MEANS($\{\vec{x}_1, \ldots, \vec{x}_N\}, K$)
1 $(\vec{s}_1, \vec{s}_2, \ldots, \vec{s}_K) \leftarrow$ SELECTRANDOMSEEDS($\{\vec{x}_1, \ldots, \vec{x}_N\}, K$)
2 **for** $k \leftarrow 1$ **to** $K$
3 **do** $\vec{\mu}_k \leftarrow \vec{s}_k$
4 **while** stopping criterion has not been met
5 **do for** $k \leftarrow 1$ **to** $K$
6 **do** $\omega_k \leftarrow \{\}$
7 **for** $n \leftarrow 1$ **to** $N$
8 **do** $j \leftarrow \arg\min_{j'} |\vec{\mu}_{j'} - \vec{x}_n|$
9 $\omega_j \leftarrow \omega_j \cup \{\vec{x}_n\}$ *(reassignment of vectors)*
10 **for** $k \leftarrow 1$ **to** $K$
11 **do** $\vec{\mu}_k \leftarrow \frac{1}{|\omega_k|} \sum_{\vec{x} \in \omega_k} \vec{x}$ *(recomputation of centroids)*
12 **return** $\{\vec{\mu}_1, \ldots, \vec{\mu}_K\}$

## 4.6 User Interface

Berikut diberikan rancangan user interface yang digunakan pada prototype sistem.



## 4.7 Pembahasan

Dengan menggunakan pemodelan seperti diuraikan di atas diharapkan sistem akan dapat mengelompokkan dokumen-dokumen yang mirip ke dalam cluster yang sama. Pada proses clustering dengan metode K-Means ada beberapa poin yang harus diperhatikan yang akan dapat menentukan hasil clustering, yaitu:

- Pemilihan seed dilakukan secara random, hal ini dapat menyebabkan penentuan cluster tidak optimal.
- Penentuan K dilakukan secara coba-coba, hal ini juga dapat menyebabkan penentuan cluster tidak optimal.

# BAB V KESIMPULAN DAN SARAN

## 5.1 Kesimpulan

Dari perancangan pemodelan sistem temu balik ini dapat ditarik beberapa kesimpulan sebagai berikut:

1. Penggunaan clustering pada dokumen akan dapat mempermudah proses pencarian, karena jika sebuah query dekat dengan salah satu dokumen, maka semua dokumen yang ada pada cluster yang sama juga diasumsikan relevan dengan query tersebut dan akan ditampilkan juga.
2. Dengan rancangan user interface yang dirancang, maka selain menampilkan semua dokumen yang dianggap relevan, sistem juga dapat menampilkannya dalam kelompok dokumen atau cluster.

## 5.2 Saran

Hasil penelitian ini baru berupa model dari sistem temu balik informasi dengan menggunakan metode K-Means. Untuk mendapatkan sistem yang dapat digunakan pada kondisi real model di atas masih perlu diimplementasikan dalam sebuah sistem berbasis web.

# DAFTAR PUSTAKA

Christopher D. Manning, Prabhakar Raghavan, Hinrich Schütz, 2009,"An Introduction to Information Retrieval", Chambridge UP

C. J. van RIJSBERGEN B.Sc., Ph.D., M.B.C.S. , "Information Retrieval" Department of Computing Science University of Glasgow

Hendra Bunyamin, 2005, "Information Retrieval System dengan Metode Latent SemanticIndexing", Program Studi Rekayasa Perangkat Lunak ITB

**Lampiran 1 : Justifikasi Anggaran Penelitian**

| 1. Honorarium | | | |
|---|---|---|---|
| Material | kuantitas | Harga satuan(Rp.) | Biaya per Tahun (Rp.) |
| Honorarium Ketua Peneliti | 1 | 725.000 | 500.000 |
| Honorarium Anggota Peneliti | 1 | 500.000 | 400.000 |
| | | **Sub Total (Rp.)** | **900.000** |
| 1. Bahan Habis pakai | | | |
| Material | kuantitas | Harga satuan(Rp.) | Biaya per Tahun (Rp.) |
| Fotocopy dan Penjilidan | 4 | 15.000 | 60.000 |
| ATK | 1 | 48.300 | 48.300 |
| | | **Sub Total (Rp.)** | **108.300** |
| 2. Publikasi | | | |
| Material | kuantitas | Harga satuan(Rp.) | Biaya per Tahun (Rp.) |
| Publikasi | 1 | 500.000 | 500.000 |
| | | **Sub Total (Rp.)** | **500.000** |
| | | **Total** | **1.508.300** |

**Lampiran 2: Biodata Ketua/ Anggota Tim Peneliti/Pelaksana**

**Ketua Peneliti :**

A. Identitas Diri

| | | |
|---|---|---|
| 1. | Nama Lengkap (dengan gelar) | Daniel Rudiaman Sijabat, S.T., M. Kom. |
| 2. | Jenis Kelamin | Laki-laki |
| 3. | Jabatan Fungsional | Asisten Ahli |
| 4. | NIP/NIK/Identitas lainnya | 010052 |
| 5. | NIDN | 0722037101 |
| 6. | Tempat, Tanggal Lahir | Kabanjahe, 22 Maret 1971 |
| 7. | Email | daniel223@stiki.ac.id |
| 8. | Nomor Telepon/HP | 081334289205 |
| 9. | Alamat Kantor | Jl. Tidar No. 100 Malang |
| 10. | Nomor Telepon/Faks | 0341-560823 /0341-562525 |
| 11. | Lulusan yang telah dihasilkan | 84 orang |
| 12. | Mata Kuliah yang diampu | Jaringan Komputer |
| | | Keamanan Jaringan |
| | | Artificial Intelligence |
| | | Organisasi dan Arsitektur Komputer |

B. Riwayat Pendidikan

| | S-1 | S-2 | S-3 |
|---|---|---|---|
| Nama Perguruan Tinggi | Institut Teknologi Bandung | STTS | |
| Bidang Ilmu | Teknik Elektro | Teknologi Informasi | |
| Tahun Masuk-Lulus | 1989 - 1997 | 2004-2006 | |
| Judul Skripsi/Tesis/Disertasi | Jaringan Syaraf Tiruan untuk Pengenalan Sinyal Sonar | Pengenalan Wajah dengan Menggunakan Principal Component Analysis dan Radial Basis Function Network | |
| Nama Pembimbing/Promotor | Dr. Ir. Adang Suwandi | Ir. Endang Setyati, M.T. | |

C. Pengalaman Penelitian dalam 5 tahun terakhir

| No | Tahun | Judul Penelitian | Pendanaan | |
|---|---|---|---|---|
| | | | Sumber | Jml Juta (Rp) |
| 1 | 2006 | Pengenalan Wajah dengan Menggunakan Principal Component Analysis dan Radial Basis Function Network | Mandiri | 3,00 |
| 2 | 2014 | Sistem Penunjang Keputusan Untuk Menentukan Golongan Masyarakat Dengan Metode Simple Additive Weighting | STIKI | 1,00 |
| 3 | 2015 | Sistem Informasi Kredit Poin | Mandiri | 1,00 |

| | | | | |
|---|---|---|---|---|
| | | Mahasiswa<br>Guna Mendukung Pembuatan Surat Keterangan Pendamping Ijazah | | |

D. Pengalaman Pengabdian kepada masyarakat dalam 5 tahun terakhir

| No | Tahun | Judul Pengabdian kepada masyarakat | Pendanaan | |
|---|---|---|---|---|
| | | | Sumber | Jml Juta (Rp) |
| 1 | 2012 | Penguji Ujian Kompetensi Jaringan Komputer di SMKN 2 Blitar | SMKN 2 Blitar | 3,00 |
| 2 | 2014 | Pembentukan POSDAYA di kelurahan Karang Besuki dan Pisang Candi kota Malang | DAMANDIRI | 4,00 |

E. Publikasi artikel dalam jurnal dalam 5 tahun terakhir

| No | Judul Artikel Ilmiah | Nama Jurnal | Volume/nomor/tahun |
|---|---|---|---|
| 1 | Pengenalan Wajah dengan Menggunakan Principal Component Analysis dan Radial Basis Function Network | Dinamika Teknologi | Volume 1, No. 2, tahun 2007 |
| | Sistem Penunjang Keputusan Untuk Menentukan Golongan Masyarakat Dengan Metode Simple Additive Weighting | Proceeding IC-Itechs ISSN 2356-4407 | Tahun 2014 |

**Anggota Peneliti :**

A. Identitas Diri

| | | |
|---|---|---|
| 1. | Nama Lengkap (dengan gelar) | **Anita, S.Kom, MT** |
| 2. | Jenis Kelamin | Perempuan |
| 3. | Jabatan Fungsional | Asisten Ahli |
| 4. | NIP/NIK/Identitas lainnya | 010034 |
| 5. | NIDN | 0707077201 |
| 6. | Tempat, Tanggal Lahir | Banyuwangi, 7 Juli 1972 |
| 7. | Email | ant@stiki.ac.id |
| 8. | Nomor Telepon/HP | 08125259973 |
| 9. | Alamat Kantor | Jl. Tidar No. 100 Malang |
| 10. | Nomor Telepon/Faks | 0341-560823 / 0341-562525 |
| 11. | Lulusan yang telah dihasilkan | 90 orang |
| 12. | Mata Kuliah yang diampu | Algoritma & Struktur Data I |
| | | Analisa Sistem Informasi |
| | | Perancangan Sistem Informasi |

B. Riwayat Pendidikan

| | S-1 | S-2 | S-3 |
|---|---|---|---|
| Nama Perguruan Tinggi | Sekolah Tinggi Informatika & Komputer Indonesia | Universitas Brawijaya | |
| Bidang Ilmu | Teknik Informatika & Komputer | Teknik Elektro ( Sistem Komunikasi dan Informatika) | |
| Tahun Masuk-Lulus | 1991-1996 | 2008-2010 | |
| Judul Skripsi/Tesis/Disertasi | Desain Sistem Informasi administrasi keuangan pada Sekolah Tinggi Informatika & Komputer Indonesia | Sistem penunjang keputusan untuk mencapai kesehatan optimal berdasarkan pola makan dan gaya hidup dengan menggunakan tabel keputusan dan | |

| | | | |
|---|---|---|---|
| | | group teknologi | |
| Nama Pembimbing/Promotor | Alm. Ir. Heru Budiono, M.Sc | Ir. Purnomo Budi Santoso, M.Sc., Ph.D dan Ir. Heru Nurwarsito, M.Kom | |

C. Pengalaman Penelitian dalam 5 tahun terakhir

| No | Tahun | Judul Penelitian | Pendanaan | |
|---|---|---|---|---|
| | | | Sumber | Jml Juta Rp) |
| 1 | 2010 | Sistem Penunjang Keputusan Penilaian Pengajuan Pembayaran Bertempo pada UD Mitra Sejati | Mandiri | 2,75 |
| 2 | 2012 | Sistem Penunjang Keputusan Untuk Memprediksi Penyakit Degeneratif Yang Akan Diderita Berdasarkan Pola Makan Dan Gaya Hidup | Mandiri | 3,00 |
| 3 | 2013 | Sistem Informasi RT/RW sebagai media komunikasi warga berbasis web | Penelitian Dosen Pemula | 12,00 |
| 4 | 2014 | Sistem Penunjang Keputusan Untuk Menentukan Golongan Masyarakat Dengan Metode Simple Additive Weighting | STIKI | 1,00 |

D. Pengalaman Pengabdian kepada masyarakat dalam 5 tahun terakhir

| No | Tahun | Judul Pengabdian kepada masyarakat | Pendanaan | |
|---|---|---|---|---|
| | | | Sumber | Jml Juta Rp) |
| 1 | 2012 | Peningkatan Proses Pembelajaran Siswa dalam Pemanfaatan Teknologi Informasi di SMAN 1 SumberPucung Malang | Seven Management | - |
| 2 | 2013 | Peningkatan Proses Pembelajaran Siswa dalam Pemanfaatan Teknologi Informasi di SMAN 1 SumberPucung Malang | Seven Management | - |
| 3 | 2014 | Pembentukan POSDAYA di kelurahan Karang Besuki dan Pisang Candi kota Malang | DAMANDIRI | |

E. Publikasi artikel dalam jurnal dalam 5 tahun terakhir

| No | Judul Artikel Ilmiah | Nama Jurnal | Volume/nomor/tahun |
|---|---|---|---|
| 1 | Sistem Penunjang Keputusan Penilaian Pengajuan Pembayaran Bertempo pada UD Mitra Sejati | Dinamika Dotcom, Jurnal pengembangan manajemen informatika dan komputer ISSN.2086-2652 | Volume 1, No. 2, tahun 2010 |
| 2 | Sistem Penunjang Keputusan Untuk Memprediksi Penyakit Degeneratif Yang Akan Diderita Berdasarkan Pola Makan Dan Gaya Hidup | SMATIKA Jurnal ISSN 2087-0256 | Volume 02, Nomor 01 tahun 2012 |
| 3 | Sistem informasi RT/RW sebagai media komunikasi warga berbasis web | SMATIKA Jurnal ISSN 2087-0256 | Volume 04, Nomor 01 tahun 2014 |
| 4 | Sistem Penunjang Keputusan Untuk Menentukan Golongan Masyarakat Dengan Metode Simple Additive Weighting | Proceeding IC-Itechs ISSN 2356-4407 | Tahun 2014 |