

**PENERAPAN METODE SIMPLE TERM WEIGHTS DALAM
PENCARIAN ARTIKEL DI WWW.CSCC-JOGJA.COM**

Skripsi



Oleh
Alril Ambanaga
22064113

Program Studi Teknik Informatika Fakultas Teknologi Informasi

Universitas Kristen Duta Wacana

2012

**PENERAPAN METODE SIMPLE TERM WEIGHTS DALAM
PENCARIAN ARTIKEL DI WWW.CSCC-JOGJA.COM**

Skripsi



Diajukan kepada Program Studi Teknik Informatika Fakultas Teknologi Informasi
Universitas Kristen Duta Wacana

Sebagai Salah Satu Syarat dalam Memperoleh Gelar
Sarjana Komputer

Disusun oleh

ALRIL AMBANAGA

22064113

Program Studi Teknik Informatika Fakultas Teknologi Informasi
Universitas Kristen Duta Wacana

2012

PERNYATAAN KEASLIAN SKRIPSI

Saya menyatakan dengan sesungguhnya bahwa skripsi dengan judul:

PENERAPAN METODE SIMPLE TERM WEIGHTS DALAM PENCARIAN ARTIKEL DI WWW.CSCC-JOGJA.COM

yang saya kerjakan untuk melengkapi sebagian persyaratan menjadi Sarjana Komputer pada pendidikan Sarjana Program Studi Teknik Informatika Fakultas Teknologi Informasi Universitas Kristen Duta Wacana, bukan merupakan tiruan atau duplikasi dari skripsi keserjanaan di lingkungan Universitas Kristen Duta Wacana maupun di Perguruan Tinggi atau instansi manapun, kecuali bagian yang sumber informasinya dicantumkan sebagaimana mestinya.

Jika dikemudian hari didapati bahwa hasil skripsi ini adalah hasil plagiasi atau tiruan dari skripsi lain, saya bersedia dikenai sanksi yakni pencabutan gelar keserjanaan saya.

Yogyakarta, 7 Agustus 2012



ALRIL AMBANAGA
22064113

HALAMAN PERSETUJUAN

Judul Skripsi : PENERAPAN METODE SIMPLE TERM WEIGHTS
DALAM PENCARIAN ARTIKEL DI
WWW.CSCC-JOGJA.COM

Nama Mahasiswa : ALRIL AMBANAGA

N I M : 22064113

Matakuliah : Skripsi (Tugas Akhir)

Kode : TIW276

Semester : Genap

Tahun Akademik : 2011/2012



UKDWN

Telah diperiksa dan disetujui di
Yogyakarta,
Pada tanggal 7 Agustus 2012

Dosen Pembimbing I


Budi Susanto, SKom.,M.T.

Dosen Pembimbing II


Dra. Widi Hapsari, M.T.

HALAMAN PENGESAHAN

PENERAPAN METODE SIMPLE TERM WEIGHTS DALAM PENCARIAN ARTIKEL DI WWW.CSCC-JOGJA.COM

Oleh: ALRIL AMBANAGA / 22064113

Dipertahankan di depan Dewan Penguji Skripsi
Program Studi Teknik Informatika Fakultas Teknologi Informasi
Universitas Kristen Duta Wacana - Yogyakarta
Dan dinyatakan diterima untuk memenuhi salah satu syarat memperoleh gelar
Sarjana Komputer
pada tanggal 2 Agustus 2012

Yogyakarta, 7 Agustus 2012
Mengesahkan,

Dewan Penguji:

1. Budi Susanto, SKom.,M.T.
2. Dra. Widi Hapsari, M.T.
3. Aditya Wikan Mahastama, S.Kom
4. Yuan Lukito, S.Kom



Dekan

(Drs. Wimmie Handiwidjojo, MIT.)

Ketua Program Studi

(Nugroho Agus Haryono, M.Si)

UCAPAN TERIMA KASIH

Puji dan syukur penulis panjatkan ke hadirat Tuhan Yang Maha Esa yang telah melimpahkan rahmat, anugerah dan karunia-Nya, sehingga penulis dapat menyelesaikan Skripsi ini dengan judul Penerapan Metode Simple Term Weights Dalam Pencarian Artikel di www.cscj-jogja.com dengan baik dan tepat waktu.

Penulisan Skripsi ini merupakan kelengkapan dan pemenuhan dari salah satu syarat dalam memperoleh gelar Sarjana Komputer. Selain itu bertujuan melatih mahasiswa untuk dapat menghasilkan suatu karya yang dapat dipertanggungjawabkan secara ilmiah, sehingga dapat bermanfaat bagi penggunanya.

Dalam menyelesaikan pembuatan program dan laporan Skripsi ini, penulis telah banyak menerima bimbingan, saran dan masukan dari berbagai pihak, baik secara langsung maupun secara tidak langsung. Untuk itu dengan segala kerendahan hati, pada kesempatan ini penulis menyampaikan ucapan terima kasih kepada:

1. **Bpk. Budi Susanto, S.Kom., M.T.** selaku dosen pembimbing I yang telah memberikan bimbingan dan saran untuk penyusunan program kepada penulis, juga kepada
2. **Ibu Dra. Widi Hapsari, M.T.** selaku dosen pembimbing II atas bimbingan, petunjuk dan masukan yang diberikan selama pengerjaan skripsi ini.
3. Kedua orangtuaku dan seluruh keluargaku yang telah memberikan semangat dan dorongan untuk penyelesaian skripsi ini.
4. Sahabatku Susilawati, S.Pd.K. yang selalu memberikan dorongan, motivasi dan mengingatkan penulis untuk menyelesaikan laporan skripsi ini.
5. Ibu Sarlin Mataheru selaku Gembala Sidang CornerStone Community Church (CSCC) Yogyakarta, untuk arahan dan motivasi bagi penulis dalam melakukan penelitian pada website www.cscj-jogja.com.

6. Semua teman-teman pelayanan yang ada di gereja CSCC Yogyakarta atas dukungannya.

Penulis menyadari bahwa program dan laporan Skripsi ini masih jauh dari sempurna. Oleh karena itu, penulis sangat mengharapkan kritikan dan saran yang membangun dari pembaca sekalian. Sehingga suatu saat penulis dapat memberikan karya yang lebih baik lagi.

Akhir kata penulis ingin meminta maaf bila ada kesalahan baik dalam penyusunan laporan ini dan semoga program dan laporan Skripsi ini dapat berguna bagi kitasemua.

© UKDWN

INTISARI

PENERAPAN METODE SIMPLE TERM WEIGHTS DALAM PENCARIAN ARTIKEL DI WWW.CSCC-JOGJA.COM

Pengelompokan berdasarkan kategori pada suatu artikel tertentu sangat penting dalam sebuah website yang memiliki artikel yang banyak. Pengunjung *website* bisa mencari artikel yang ingin dibaca berdasarkan kategori tersebut. Setelah memilih salah satu kategori, pengunjung bisa memilih berdasarkan judul artikel sebagai perwakilan dari isi artikel yang ingin mereka dibaca. Namun jika judul artikel dalam kategori tersebut sudah sangat banyak, atau kategori artikel pun sudah sangat banyak, pengunjung akan kesulitan dalam mencari artikel yang ingin dibaca. Untuk mengatasi masalah tersebut perlu dibuat sistem pencarian artikel yang membantu pengunjung dalam mencari artikel yang ingin mereka baca.

Untuk mempermudah pengguna dalam mencari artikel, penulis membuat sistem pencari artikel dengan menggunakan *Simple Term Weights model*. Model ini akan melakukan pembobotan pada setiap kata yang ada pada isi artikel kemudian dilakukan estimasi apakah dokumen tersebut relevan atau tidak. Setiap kata yang ada pada artikel atau dokumen, akan disaring terlebih dahulu untuk membuang kata-kata yang tidak penting yang tergolong dalam *stopwords*. Dasar dari proses estimasi adalah bobot dari setiap kata atau *term* yang dihitung nilai probabilitasnya. Penghitungan nilai probabilitas ini terjadi ketika pengguna sistem atau pengunjung website memberikan *query*. *Query* ini disaring terlebih dahulu untuk membuang kata-kata yang termasuk *stopwords* baru kemudian dilakukan proses estimasi berdasarkan dari *term* yang ada pada *query* tersebut. Dari proses estimasi ini akan didapatkan dokumen yang mungkin relevan yang kemudian dihitung nilai relevansinya. Selanjutnya dokumen yang telah diberi nilai relevansi akan ditampilkan kepengunjung website dengan urutan berdasarkan nilai relevansi terbesar sampai dengan nilai terkecil. Dokumen yang ditampilkan

ke pengunjung web inilah yang merupakan hasil dari proses pencarian sesuai kata kunci atau *query* yang diberikan.

Dari hasil implementasi sistem dengan menggunakan *Simple Term Weights model*, pengunjung web bisa mendapatkan dokumen yang ingin mereka baca berdasarkan pada kata kunci yang mereka berikan kesistem. Namun dalam pemberian kata kunci, terkadang kata kunci tersebut belum benar-benar mewakili isi keseluruhan artikel. Untuk itu dalam pengembangan selanjutnya penulis menyarankan untuk mengembangkan sistem agar bisa melakukan ekspansi kata kunci atau *query* agar pencarian dapat lebih optimal dan akurat.

© UKDW

DAFTAR ISI

HALAMAN JUDUL.....	ii
PERNYATAAN KEASLIAN SKRIPSI.....	iii
HALAMAN PERSETUJUAN.....	iv
HALAMAN PENGESAHAN.....	v
UCAPAN TERIMA KASIH.....	vi
INTISARI.....	viii
DAFTAR ISI.....	x
DAFTAR TABEL	xiii
DAFTAR GAMBAR	xiv
BAB 1 PENDAHULUAN	1
1.1. LatarBelakangMasalah.....	1
1.2. PerumusanMasalah	2
1.3. BatasanMasalah	2
1.4. Hipotesis.....	2
1.5. TujuanPenelitian	2
1.6. Metode/Pendekatan.....	3
1.7. SistematikaPenulisan	4
BAB 2 TINJAUAN PUSTAKA	5
2.1. LandasanTeori.....	5
2.2.1. Information Retrieval	5
2.2.2. Simple Term Weight.....	6
2.1.3. EvaluasiSistem Information Retrieval	15
BAB 3 ANALISIS DAN PERANCANGAN SISTEM	19
3.1. Analisis Data	19
3.2. BahasaPemrograman.....	19
3.3. Alat.....	20
3.4. Bahan	20
3.5. Rancangan Input	20

3.5.1. Input Dokumen	21
3.5.2. Input Query	21
3.6. Rancangan Output.....	21
3.7. Rancangan Proses	22
3.7.1. Rancangan Proses Indexing	22
3.7.1.1. Indexing DokumenBaru	22
3.7.1.2. Indexing Update Dokumen	24
3.7.1.3. Indexing Delete Dokumen	27
3.7.2. Rancangan Proses Retrieval.....	29
3.8. Rancangan Interface.....	32
3.8.1. Rancangan Form MasukanDokumen.....	32
3.8.2. Rancangan Form Query	33
3.8.3. RancanganDokumen Retrieve.....	34
3.8.4. Rancangan Isi Dokumen	35
3.9. Rancangan Database	36
BAB 4 IMPLEMENTASI DAN ANALISIS SISTEM.....	39
4.1. ImplementasiSistem	39
4.1.1. Proses Indexing.....	39
4.1.1.1. Proses Indexing Insert dokumenbaru	39
4.1.1.2. Proses Indexing Update Dokumen.....	40
4.1.1.3. Proses Indexing Delete Dokumen.....	42
4.1.1.4. Proses Indexing Menggunakan MySQL event	42
4.1.2. Proses Retrieval.....	43
4.1.2.1. Proses EstimasiDokumenRelevan.....	44
4.1.2.2. Proses PenghitunganBobotRelevansiAkhir	44
4.2. Implementasi Form Input Dokumen	45
4.3. Implementasi Form Input Query	46
4.4. ImplementasiKeluaranSistem	47
4.5. Implementasi View Document.....	48
4.6. AnalisisMetode Simple Term Weight.....	49
4.7. EvaluasiMetode Simple Term Weight	51

4.8. Kelebihan dan Kekurangan Program	62
BAB 5 KESIMPULAN DAN SARAN	64
5.1. Kesimpulan	64
5.2. Saran.....	64

© UKDW

DAFTAR TABEL

Tabel2.1.TabelProbabilitas	7
Tabel2.2.TabelKontingensidengankorelasifaktor k	9
Tabel2.3.Tabelestimasirelevansidokumen	13
Tabel2.4.Tabelbobot term	15
Tabel2.5.Tabelbobotdokumen.....	15
Tabel2.6.TabelKontingensiprecisiondanrecall.....	17
Tabel2.7.Tabelnilaiprecisiondanrecall	18
Tabel3.1.Tabel Terms	36
Tabel 3.2.Tabelestimasi_relevansi.....	36
Tabel 3.3.Tabelstopwords	37
Tabel 3.4.Tabelartikel	37
Tabel 3.5.Tabelquery_list	38
Tabel 4.1.Tabelperbandinganrelevansidokumen	50
Tabel 4.2.Tabelpengujianpertama.....	51
Tabel 4.3.Tabelinterpolasiprecisiondokumenuntuk query 1	52
Tabel 4.4.Tabelpengujiankedua	53
Tabel 4.5.Tabelinterpolasiprecisiondokumenuntuk query 2	54
Tabel 4.6.Tabelpengujianketiga.....	55
Tabel 4.7.Tabelinterpolasiprecisiondokumenuntuk query 3	56
Tabel 4.8.Tabelpengujiankeempat	57
Tabel 4.9.Tabelinterpolasiprecisiondokumenuntuk query 4	58
Tabel 4.10.Tabelpengujiankelima.....	59
Tabel 4.11.Tabelinterpolasiprecisiondokumenuntuk query 5	59
Tabel 4.12.Tabelpengujiankeenam	60
Tabel 4.13.Tabelinterpolasiprecisiondokumenuntuk query 6	61

DAFTAR GAMBAR

Gambar 3.1. Flow Chart Proses Indexing Dokumen Baru	23
Gambar 3.2. Flow Chart Proses Indexing Update Dokumen.....	25
Gambar 3.3. Flow Chart Proses Indexing Delete Dokumen.....	28
Gambar 3.4. Flow Chart Proses Retrieval	30
Gambar 3.5. Form Masukan Dokumen.....	32
Gambar 3.6. Form Masukan query	33
Gambar 3.7. Rancangan <i>interface</i> dokumen retrieval	34
Gambar 3.8. Rancangan <i>interface</i> isi dokumen	35
Gambar 3.9. Gambar ER Diagram	38
Gambar 4.1. Halaman <i>form input</i> dokumen.....	46
Gambar 4.2. Halaman <i>form input query</i>	47
Gambar 4.3. Halaman hasil retrieval	48
Gambar 4.4. Halaman lihat isi dokumen.....	49
Gambar 4.5. Grafik interpolasi precision untuk query 1	53
Gambar 4.6. Grafik interpolasi precision untuk query 2.....	55
Gambar 4.7. Grafik interpolasi precision untuk query 3	56
Gambar 4.8. Grafik interpolasi precision untuk query 4	58
Gambar 4.9. Grafik interpolasi precision untuk query 5	60
Gambar 4.10. Grafik interpolasi precision untuk query 6	62

BAB I

PENDAHULUAN

1.1. Latar Belakang Masalah

Sudah sangat umum pada suatu website jika memiliki banyak informasi. Informasi itu pun sangat beragam diantaranya adalah artikel. Pada website yang berfokus pada portal artikel, tentu memiliki sekian banyak judul artikel beserta isinya, untuk memenuhi kebutuhan pengunjung web tersebut. Dalam penyajiannya ada beragam cara seperti menampilkan judul artikel dan dikategorikan berdasarkan abjad, dikategorikan berdasarkan judul artikel yang paling banyak dikunjungi, atau dikategorikan berdasarkan artikel yang baru masuk pada *database*. Untuk pencarian artikel yang ingin dibaca jika menggunakan cara *browsing*, pengunjung harus membuka satu persatu judul yang dianggap sesuai dengan yang ingin mereka cari informasinya. Namun jika pada *database* web tersebut telah menampilkan ratusan bahkan ribuan judul artikel, tentu akan sulit bagi pengunjung untuk mencari informasi yang mereka butuhkan dari artikel yang akan dicari. Jika hanya mengandalkan *browsing* artikel akan memakan waktu dan menyulitkan pengguna karena akan menampilkan ratusan judul dan kategorinya.

Untuk mengatasi permasalahan diatas, maka perlu dibangun fitur mesin pencari artikel. Yang dimaksud dengan mesin pencarian artikel ini adalah sistem *Information Retrieval*. Dalam pengembangannya ada banyak model yang bisa dipakai dalam proses *Information Retrieval* diantaranya adalah *Simple Term Weights*. *Simple Term Weights* adalah model yang digunakan untuk proses pembobotan relevansi antara masukan *query* dengan koleksi dokumen yang ada pada sistem. Dengan adanya sistem *Information Retrieval* ini diharapkan akan mempermudah dalam proses pencarian artikel di website www.cscce-jogja.com/artikel.

1.2. Perumusan Masalah

Dari permasalahan di atas dapat dibuat suatu perumusan masalah:

1. Bagaimana metode *Simple Term Weights* menentukan dokumen relevan dalam pencarian artikel?
2. Bagaimana metode *Simple Term Weights* melakukan pengurutan dokumen yang relevan?
3. Bagaimana hasil evaluasi dari hasil pencarian dokumen menggunakan metode *Simple Term Weights* terhadap masukan *query* berdasarkan pada kriteria evaluasi *precision* dan *recall*?

1.3. Batasan Masalah

Dalam Tugas Akhir ini terdapat batasan-batasan masalah sebagai berikut:

1. Pencarian hanya dilaksanakan pada dokumen yang ada pada satu database pada sistem itu sendiri.
2. Dokumen yang dicari berformat HTML.
3. Sistem yang dibangun berbasis web.
4. Tidak ada proses *stemming* terhadap *term*.
5. *Query* tidak mengenal tanda baca atau operator logika.

1.4. Hipotesis

Dugaan akhir dari penelitian yang akan dilakukan adalah sebagai berikut:

1. Sistem mampu menampilkan hasil pencarian berdasarkan tingkat relevansinya terhadap *query* yang dimasukkan.

1.5. Tujuan Penelitian

Adapun tujuan penelitian bisa dibedakan menjadi dua kategori yaitu tujuan utama dan sub tujuan. Tujuan utama penelitian ini adalah untuk menemukan cara

pencarian artikel yang efektif dengan menggunakan model Simple Term Weights. Adapun sub tujuan dari penelitian ini adalah (1) mengukur relevansi antara dokumen koleksi dengan masukan *query* dari pengguna, (2) mengukur interpolasi *precision* dan *recall* dari proses *retrieval* yang dilakukan.

1.6. Metode/Pendekatan

Dalam penyusunan Tugas Akhir ini akan digunakan langkah-langkah seperti berikut:

1. Metode pengumpulan data

Pengumpulan data dimulai dengan studi pustaka terlebih dahulu untuk mempelajari serta memperdalam pemahaman teori-teori yang mendukung penelitian ini seperti *Information Retrieval*, terlebih khusus Metode *Simple Term Weights*. Dari studi pustaka ini akan didapatkan variable-variabel yang dibutuhkan dalam pengembangan sistem, syarat-syarat yang harus dipenuhi untuk bisa menerapkan metode tersebut dan *tools* yang bisa dipakai untuk membuat sistem. Langkah selanjutnya adalah membuat daftar *stop words* yang akan dipakai. Menganalisa sistem yang akan diimplementasikan kedua metode tersebut yaitu pada bagian artikel dalam hal ini penelitian dilakukan di website dengan alamat url <http://www.csccljogja.com/artikel>.

2. Metode pengembangan sistem

Pada pengembangan sistem, hal pertama yang harus dilakukan adalah merancang aplikasi mulai dari perancangan antarmuka sistem, perancangan database dan alur kerja sistem. Dari proses perancangan sistem tersebut kemudian dilakukan pengkodean dengan menggunakan metode *Simple Term Weights*.

3. Metode Evaluasi

Metode evaluasi dilakukan menghitung nilai *precision* dan *recall* dari proses *retrieval* yang dilakukan. Nilai dari *precision* dan *recall*

kemudian akan dipakai untuk mengetahui tingkat keberhasilan dari sistem yang dibangun.

1.7. Sistematika Penulisan

Penyusunan tugas akhir ini terdiri dari lima bab: Pendahuluan, Tinjauan Pustaka, Analisis dan Perancangan Sistem, Implementasi dan Analisis Sistem, Kesimpulan dan saran dengan uraian sebagai berikut:

Bab 1: Pendahuluan, berisi tentang deskripsi umum tugas akhir yang meliputi latar belakang masalah, perumusan masalah, batasan masalah, hipotesis, tujuan penelitian, metode/pendekatan dan sistematika penulisan.

Bab 2: Tinjauan Pustaka, berisi dasar teori yang diperlukan dalam penelitian yaitu: *Information Retrieval*, *Simple Term Weights* dan metode evaluasi yaitu interpolasi *precision* dan *recall*.

Bab 3: Analisis dan Perancangan Sistem, berisi tentang arsitektur sistem, *Flow Chart*, bahan dan alat penelitian, rancangan input, rancangan proses, rancangan *interface*, dan rancangan *database*.

Bab 4: Implementasi dan Analisis Sistem, memuat hasil riset/implementasi sistem, serta analisis dari riset yang dilakukan.

Bab 5: Kesimpulan dan Saran, pada bagian ini berisi kesimpulan yang didapat dari penelitian dan saran-saran bagi penelitian akan datang yang menggunakan skripsi ini sebagai acuan.

BAB V

KESIMPULAN DAN SARAN

5.1. Kesimpulan

1. *Simple Term Weights Model* menganggap jumlah kemunculan suatu *term* dalam dokumen tidak akan mempengaruhi bobot relevansi dokumen maupun nilai relevansi *term*.
2. Pengulangan kata dalam *query* tidak akan mempengaruhi hasil *retrieval*, karena tetap dianggap sebagai satu kata.
3. Urutan kata dalam *query* tidak akan mempengaruhi nilai relevansi maupun proses *retrieval*.
4. Nilai relevansi dokumen sangat bergantung pada kehadiran seluruh *term* yang ada pada masukan *query*.

5.2. Saran

Setelah melakukan penelitian dan melakukan penulis memberikan saran untuk pengembangan sistem pencarian artikel dengan menggunakan metode *Simple Term Weights* sebagai berikut:

1. Proses input dokumen bisa dikembangkan dengan cara melakukan upload file dengan ekstensi *.txt atau *.doc
2. Untuk meningkatkan performa hasil *retrieval*, dapat diterapkan ekspansi *query* secara otomatis oleh sistem atau memberikan saran *query* pada *form input query* dengan *query* yang telah diekspansi.
3. Penambahan fitur untuk menambahkan, merubah atau menghapus daftar *stopword*.
4. Melakukan pengembangan pada proses *indexing* untuk judul dan kategori artikel dengan memberikan prioritas khusus kepada kategori kemudian pada judul artikel.

DAFTAR PUSTAKA

- Baeza-Yates, R.&Ribeiro-Neto, B. (2006).*Modern Information Retrieval*.New York: ACM Press.
- Garcia, E. (2009). *RSJ-PM Tutorial: A Tutorial on the Robertson-Sparck Jones Probabilistic Model for Information Retrieval*.
- Grossman, A. D. &Frieder, O. (2004).*Information Retrieval Algorithms and Heuristics Second edition*. Netherlands: Springer.
- Intan, R. &Defeng, A. (2006).HARD: SUBJECT-BASEDSEARCH ENGINE MENGGUNAKAN TF-IDF DAN JACCARD'S COEFFICIENT.*JURNAL TEKNIK INDUSTRI*. Surabaya: Universitas Kristen Petra Surabaya, VOL. 8, NO. 1, JUNI 2006: 61-72.
- Manning, C.D., Raghavan, P., Schütze, H. (2008). *Introduction to Information Retrieval*. Cambridge: CAMBRIDGE UNIVERSITY PRESS.
- Robertson, S. Understanding Inverse Document Frequency: On theoretical arguments for IDF. *Journal of Documentation*. London: City University, 60 no. 5, pp 503-520.
- Robertson, S.E. & Jones, K.S. (1976).Relevance Weighting of Search Terms.*Jurnal of the American Society for Information Science*.Volume 27 number 3, May-June 1976, pp 129-146.

