

**PERANCANGAN STBI PANDUAN PENGGUNAAN
LAYANAN HOSTING DENGAN PEMBOBOTAN
TF-IDF (STUDI KASUS: PT HOSTINGAN AWAN
INDONESIA)
PROPOSAL TUGAS AKHIR**

Oleh:

Yoki Renaldo Sinaga

33118010127

Disusun untuk pengajuan proposal Tugas Akhir Program Diploma III



**PROGRAM STUDI TEKNIK INFORMATIKA
JURUSAN TEKNIK INFORMATIKA
POLITEKNIK NEGERI BATAM
BATAM**

2020

HALAMAN PENGESAHAN PROPOSAL

**PERANCANGAN STBI PANDUAN PENGGUNAAN
LAYANAN HOSTING DENGAN PEMBOBOTAN
TF-IDF (STUDI KASUS: PT HOSTINGAN AWAN
INDONESIA)**

Oleh:

Yoki Renaldo Sinaga

3311801027

Proposal ini telah dikonsultasikan dengan dosen pembimbing
sebagai persyaratan untuk melaksanakan sidang proposal
pada

**PROGRAM DIPLOMA III
PROGRAM STUDI TEKNIK INFORMATIKA
JURUSAN TEKNIK INFORMATIKA
POLITEKNIK NEGERI BATAM**

Batam, November 2020

Disetujui oleh:

Pembimbing

Rina Yulius, S.Pd., M.Eng.
NIK. 118199

Abstrak

Sistem temu balik informasi adalah sistem yang digunakan untuk menemukan informasi yang relevan dengan kebutuhan dari pengguna secara otomatis, berdasarkan kesesuaian dengan *query* (masukan berupa ekspresi kebutuhan informasi oleh pengguna) dari suatu koleksi informasi. Salah satu contoh kasus, terkait dengan STBI yang terjadi di PT Hosting Awan Indonesia. Dalam kasus ini, pada layanan panduan hosting, belum menyediakan tempat pencarian informasi pada halaman panduan, yang menyebabkan pelanggan bingung dan kesulitan dalam mencari informasi panduan hosting. Berdasarkan hal tersebut, penulis akan melakukan tugas akhir ini untuk merancang sistem temu balik informasi untuk layanan panduan hosting dengan metode pembobotan TF-IDF. Selain dari itu, penelitian ini bertujuan untuk memudahkan pelanggan dalam mencari dan memperoleh informasi secara efektif dan efisien. Hasil dari penelitian ini diharapkan menjadi salah satu solusi untuk PT Hosting Awan Indonesia.

Keyword : Informasi, Layanan, Otomatis, Sistem.

1. Latar Belakang

Teknologi saat ini sangat berkembang pesat. Banyak teknologi yang diciptakan untuk memudahkan manusia. Salah satu pengembangan dari teknologi adalah Sistem Informasi.

Setiap perkembangan sistem informasi pasti menimbulkan beberapa dampak. Dampak positif dari perkembangan ini adalah kemudahan dalam mendapatkan informasi khususnya melalui internet, begitu juga kebutuhan pelanggan dalam memperoleh informasi dan mencari informasi harus efektif dan efisien. Pemilihan atau penemuan kembali informasi ini tidak mungkin dilakukan secara manual, karena kumpulan informasi yang sangat banyak dan beragam tentunya. Perkembangan ini juga menimbulkan dampak negatif bagi masyarakat awam yang kurang mengerti terhadap teknologi informasi ini. Hal ini juga terjadi pada PT Hosting Awan Indonesia, dikarenakan banyaknya keluhan pelanggan yang kesulitan dan bingung untuk memperoleh informasi panduan penggunaan layanan hosting.

Menurut Mandala (2002) menyatakan bahwa suatu sistem otomatis diperlukan untuk membantu pengguna dalam menemukan informasi. Berdasarkan Pernyataan tersebut, ada beberapa peneliti yang meneliti tentang sistem otomatis dengan metode sistem yang beda yaitu *Chatbot* dan *Livechat*. Hormansyah (2018) menyatakan bahwa *Chatbot* kurang optimasi dalam mencari jawaban, dan juga harus ditingkatkan untuk lebih memahami maksud pertanyaan dari si pengguna. Menurut Fidri (2019) pelayanan dari Petugas *Live chat* perlu memperhatikan mengenai respon pengguna, agar pengguna tidak menunggu terlalu lama tanggapan dari petugas.

Sehingga dari beberapa kekurangan dari metode tersebut, penulis ingin menggunakan sistem otomatis dengan metode sistem temu balik informasi. Sistem temu balik informasi (*information retrieval system*) merupakan sistem yang digunakan untuk menemukan informasi yang relevan dengan kebutuhan dari penggunanya secara otomatis, berdasarkan kesesuaian dengan *query* (masukan berupa ekspresi kebutuhan informasi oleh pengguna) dari suatu koleksi informasi.

Sistem temu balik informasi ini dipadukan dengan metode kombinasi TF-IDF. TF-IDF (*term frequency – inverse document frequency*) merupakan algoritma yang menggabungkan dua konsep frekuensi kemunculan *term* dalam sebuah dokumen atau artikel dan inversi frekuensi dokumen yang mengandung kata tersebut. Jumlah dokumen yang frekuensinya banyak dapat ditemukan kembali dan yang dianggap relevan secara sekaligus. Nantinya STBI yang menggunakan metode TF-IDF ini menjadi solusi bagi PT Hosting Awan Indonesia untuk memudahkan pelanggan dalam memperoleh informasi panduan layanan hosting.

2. Rumusan Masalah

Dari latar belakang diatas maka dapat diambil rumusan masalah sebagai berikut:

- a. Bagaimana sistem temu balik informasi dengan metode kombinasi TF-IDF ini dirancang?
- b. Bagaimana cara pengguna memperoleh informasi layanan panduan hostingan.id secara efektif dan efisien?

3. Batasan Masalah

- a. Sistem ini hanya digunakan di PT Hosting Awan Indonesia.
- b. Sistem ini dirancang berbasis web.
- c. Informasi yang digunakan adalah informasi yang tersedia di halaman panduan hostingan.id.
- d. Bahasa yang digunakan adalah PHP.
- e. Sistem untuk pencarian rekomendasi artikel panduan hosting dengan *keywords*.

4. Tujuan

Tujuan dari Tugas Akhir ini adalah Menghasilkan Sistem Temu balik Informasi dengan metode pembobotan kombinasi TF-IDF untuk memudahkan pelanggan dalam memperoleh informasi di halaman layanan panduan hostingan.id.

5. Manfaat

Dengan dirancangnya sistem temu balik informasi dengan metode pembobotan kombinasi TF-IDF ini, diharapkan sistem bermanfaat bagi admin

dan pengguna dalam memperoleh dan mencari informasi secara efisien dan efektif pada halaman panduan hostingan.id.

6. Tinjauan Pustaka/Landasan Teori

6.1 Penelitian Terdahulu

Sebagai bahan pertimbangan akan dicantumkan detail perbandingan penelitian yang pada Tabel 1.

Tabel 1. Penelitian Terdahulu

Perbandingan	Penelitian Sebelumnya	Penelitian Sebelumnya	Tugas Akhir
Judul	Pemanfaata Sistem Temu Kembali Informasi dalam Pencarian Dokumen Menggunakan <i>Metode Vector Space Model</i>	Perancangan Sistem Temu Balik Informasi Dengan Metode Pembobotan Kombinasi TF-IDF Untuk Pencarian Dokumen Berbahasa Indonesia	Perancangan STBI Panduan Penggunaan Layanan Hosting Dengan Pembobotan TF-IDF (Studi Kasus: PT Hostingan Awan Indonesia)
Bahasa Pemrograman	PHP	JSP	PHP
Metode	Metode <i>Vector Space Retrieval Model</i>	Metode TF-IDF	Metode TF-IDF
Pencarian	Dokumen	Dokumen	Artikel

6.2 Sistem Temu Balik Informasi

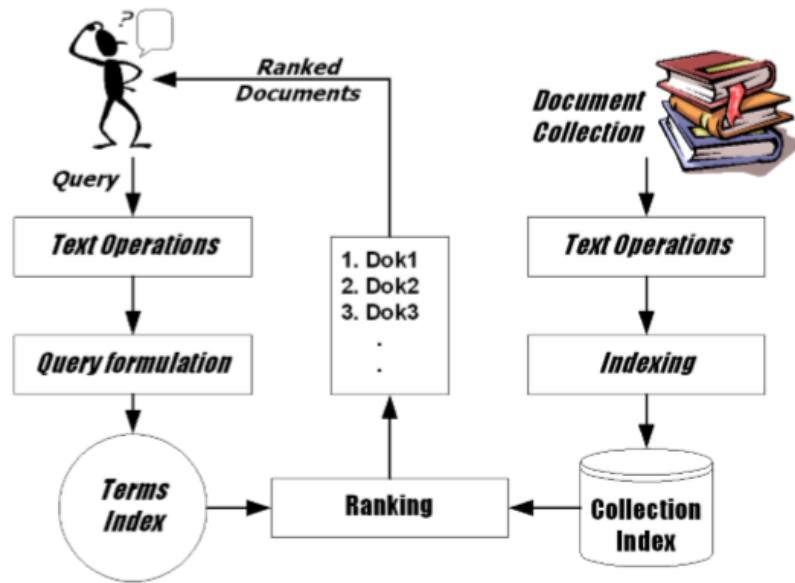
Sistem temu balik informasi (*information retrieval system*) digunakan untuk menemukan kembali (*retrieve*) informasi-informasi yang relevan terhadap kebutuhan pengguna dari suatu kumpulan informasi secara otomatis, Mandala (2004).



Gambar 6.2.1 Ilustrasi sistem temu balik informasi

Sistem temu balik informasi terdiri dari tiga komponen utama, yaitu masukan (*input*), pemroses (*processor*), dan keluaran (*output*). Pengguna adalah pemilik kebutuhan akan informasi, yang kemudian menerjemahkan kebutuhannya menjadi sebuah *query*. *Input* harus berupa *keywords* yang tepat dari artikel yang dicari sehingga *query* dapat diolah oleh pemroses. Pemroses (*processor*) bertugas menstrukturkan informasi dalam bentuk yang tepat, misalnya dengan pengindeksan dan klasifikasi serta melakukan proses temu balik, yaitu dengan menjalankan suatu strategi pencarian sebagai respon dari *query*. *Output* adalah keluaran yang diberikan oleh pemroses, biasanya berbentuk informasi tentang suatu dokumen ataupun artikel yang di cari.

Sebagai suatu sistem, sistem temu balik informasi memiliki beberapa bagian yang membangun sistem secara keseluruhan. Gambaran bagian-bagian yang terdapat pada suatu sistem temu balik informasi digambarkan pada Gambar 6.2.2



Gambar 6.2.2 Bagian-bagian sistem temu balik informasi (Mandala, 2002)

Gambar 6.2.2 memperlihatkan bahwa terdapat dua proses operasi pada sistem temu balik informasi. Alur pertama dimulai dari koleksi dokumen dan alur kedua dimulai dari *query* pengguna. Alur pertama yaitu pemrosesan terhadap koleksi dokumen menjadi basis data indeks tidak tergantung pada alur kedua. Sedangkan alur kedua tergantung dari keberadaan basis data indeks yang dihasilkan pada alur pertama (Mandala, 2002).

Bagian-bagian dari sistem temu balik informasi menurut gambar 6.2.2 meliputi :

- Text Operations* (operasi terhadap teks) yang meliputi pemilihan kata-kata dalam *query* (*term selection*) *query* menjadi *terms index* (indeks dari kata-kata).
- Query formulation* (formulasi terhadap *query*) yaitu memberi bobot pada indeks kata-kata *query*.
- Ranking* (perangkingan), mencari dokumen atau artikel yang relevan terhadap *query* dan mengurutkan dokumen artikel tersebut berdasarkan kesesuaiannya dengan *query*.

- d. *Indexing* (pengindeksan), membangun basis data indeks dari koleksi dokumen atau artikel. Dilakukan terlebih dahulu sebelum pencarian dokumen dilakukan.

6.3 *Term Frequency - Inverse Document Frequency (TF-IDF)*

Metode TF-IDF merupakan metode untuk menghitung bobot setiap kata yang paling umum digunakan pada *information retrieval*. Fitri (2013) menyatakan bahwa Metode TF-IDF ini menggabungkan dua konsep yaitu frekuensi kemunculan sebuah kata di dalam sebuah dokumen dan *inverse* frekuensi dokumen yang mengandung kata tersebut. Menurut Ma'arif (2015), metode ini juga terkenal efisien, mudah dan memiliki hasil yang akurat.

Frekuensi dokumen yang mengandung *keywords* menunjukkan seberapa umum kata tersebut. Sehingga bobot hubungan antara sebuah kata dan sebuah dokumen akan tinggi apabila frekuensi kata tersebut tinggi di dalam dokumen dan frekuensi keseluruhan dokumen yang mengandung kata tersebut yang rendah pada kumpulan dokumen.

Rumus yang digunakan pada *Term Frequency (tf)*, yaitu nilai TF diberikan berdasarkan frekuensi kemunculan suatu kata dokumen atau artikel. Berdasarkan penelitian Manning (2009) *Inverse Document Frequency (idf)* dihitung dengan menggunakan rumus sebagai berikut :

$$IDF = \log\left(\frac{D}{DF}\right)$$

Dimana,

D = Jumlah dokumen atau artikel

DF = Dokumen frekuensi

Manning(2009) meneliti bahwa Rumus umum untuk TF-IDF adalah gabungan dari rumus perhitungan *raw tf* dengan rumus *idf* dengan cara mengalikan nilai *term frequency* (tf) dengan nilai *inverse document frequency* (idf).

$$W_{dt} = tf_{dt} * IDF_t$$

Dimana,

d : dokumen ke-d.

t : kata ke-t dari kata kunci.

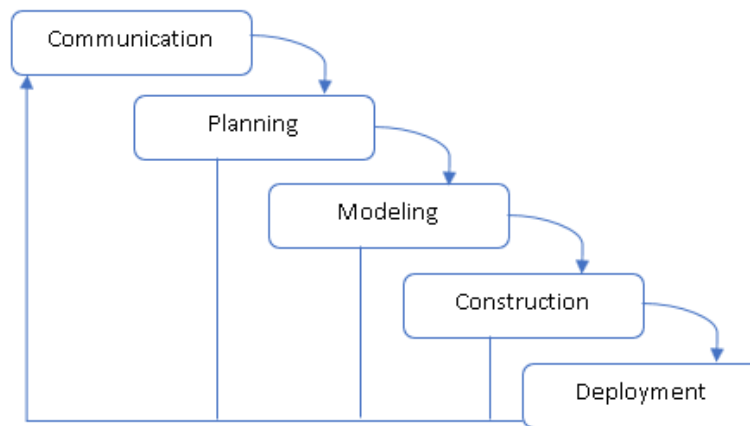
W : bobot dokumen ke-d terhadap kata ke-t.

tf : banyaknya kata yang dicari pada sebuah dokumen.

IDF : nilai idf yang di dapatkan.

7. Metode Penyelesaian Masalah

Menurut Pressman (2010), metode *Waterfall* disebut juga dengan *classic life cycle*, artinya suatu siklus klasik dari pembangunan suatu sistem. Langkah-langkahnya adalah sebagai berikut:



Gambar 1. Metode Waterfall

a. *Communication*

Proses *Communication* adalah tahap awal dari perancangan suatu sistem. Pada tahap ini dilakukan analisis kebutuhan sistem, dan mempelajari teori-teori yang berhubungan dengan sistem yang akan dirancang.

b. Planning

Proses *Planning* adalah tahap kedua dari perancangan suatu sistem. Pada tahap ini dilakukan perkiraan kebutuhan sistem, penjadwalan, dan jalur yang akan di tempuh.

c. Modeling

Proses *Modeling* adalah tahap ketiga dari perancangan suatu sistem. Pada tahap ini dilakukan analisis rancangan basis data, dan desain tampilan.

d. Construction

Proses *Construction* merupakan tahap penerjemahan desain menjadi bahasa pemrograman yang dipahami sistem.

e. Deployment

Proses *Deployment* adalah tahap terakhir dari perancangan sistem. Setelah rancangan sistem dibuat dan tidak ada *bug* lagi dalam sistem tersebut, maka sistem dapat dikenalkan ke pengguna. Tujuannya adalah untuk mendapatkan *feedback* dan saran.

8. Deskripsi Umum Sistem

Pada sistem yang akan dirancang nanti, dapat digunakan oleh admin maupun pengguna. Sistem akan ada di halaman panduan pada web hostingan.id. Pengguna melakukan pencarian terkait panduan pada web menggunakan Kotak pencarian yang ada. Lalu sistem melakukan pembobotan *keywords* yang diinginkan pengguna dengan metode pembobotan kata TF-IDF.

9. Rencana Pelaksanaan

a. Perencanaan dan pencarian referensi

Tahap pencarian informasi, referensi dan pembuatan rencana. Dalam tahapan ini, penyusun akan mempelajari kebutuhan dan informasi terkait sistem dan metode yang digunakan.

b. Perancangan aplikasi

Pembuatan desain tampilan antarmuka ,perancangan skema sistem yang akan dirancang. Rancangan skema berupa diagram DCD dan UML dari sistem yang akan dibangun.

c. Implementasi / *coding*

Tahap perancangan sistem dengan cara mengubah kebutuhan dan rancangan sistem menjadi sebuah program dengan bahasa pemrograman PHP. Tahap ini sangat bergantung pada tahapan sebelumnya.

10. Scheduling

No	Kegiatan	Bulan dan Minggu ke-																															
		Agustus		September				Oktober				November				Desember				Januari				Februari				Maret				April	
		3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2				
1	Communication																																
2	Planning																																
3	Modeling																																
4	Construction																																
5	Deployment																																

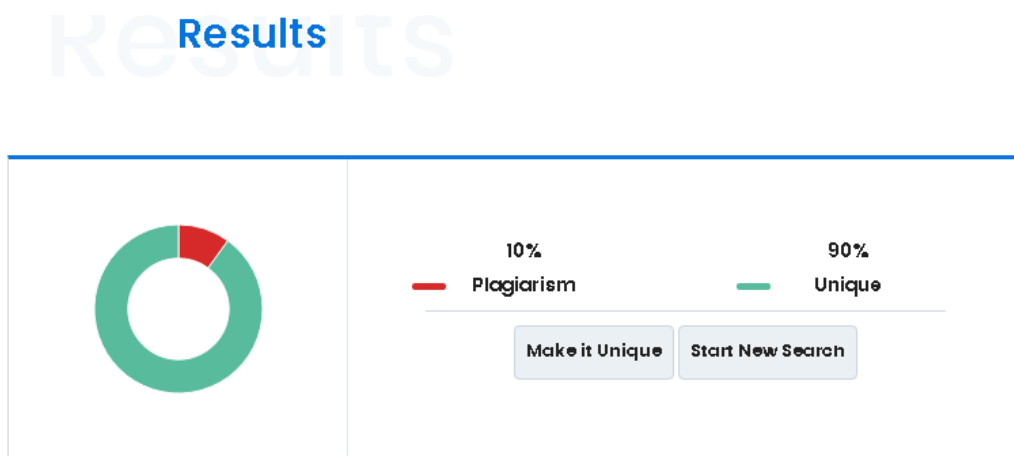
Tabel 1. Scheduling

11. Daftar Pustaka

- Mandala, R., dan Setiawan, H. 2002. Peningkatan Performansi Sistem Temu Kembali Informasi dengan Perluasan *Query* secara Otomatis, Institut Teknologi Bandung, Indonesia.
- Hormansyah, Dhebys Suryani dan Utama, Yoga Putera. 2018. Aplikasi *Chatbot* Berbasis Web Pada Sistem Informasi Layanan Publik Kesehatan Di Malang Dengan Menggunakan Metode TF-IDF, Politeknik Negeri Malang, Indonesia.
- Fidri, Imansyah Fadhil, dan Tasrif, Elfi. 2019. Analisis Kepuasan Pengguna Live Chat BPS Sumatera Barat Menggunakan Model Delone Mclean, Universitas Negeri Padang, Indonesia.
- Fitri, Meisya. 2013. Perancangan Sistem Temu Balik Informasi Dengan Metode Pembobotan Kombinasi TF-IDF Untuk Pencarian Dokumen Berbahasa Indonesia, Universitas Tanjungpura, Indonesia.
- Sanjaya, Ferry. 2017. Pemanfaatan Sistem Temu Kembali Informasi dalam Pencarian Dokumen Menggunakan Metode *Vector Space Model*. Journal of Information and Technology Volume 05 Nomor 02, STIKI, Indonesia.
- Mandala, R. 2004. Bahan Kuliah Sistem Temu Balik Informasi. Institut Teknologi Bandung, Indonesia.
- Ma'arif, Abdul Aziz. 2015. Penerapan Algoritma Tf-Idf Untuk Pencarian Karya Ilmiah. Jurnal. Jurusan Teknik Informatika. Universitas Dian Nuswantoro, Indonesia.
- Manning, D. Christopher, Raghavan, P. & Schütze H. 2009. An Introduction to Information Retrieval. Cambridge University Press
- Pressman, Roger S. 2010. *Software Engineering A Practitioner's Approach, Seventh Edition* .New York : Mc Graw Hill

12. Hasil Pengecekan Plagiarisme

Berikut ini adalah hasil pengecekan plagiarisme melalui duplichecker.com.



Gambar 3. Hasil Pengecekan Plagiarisme

13. Lampiran Artikel

Lampiran 1

Pemanfaatan Sistem Temu Kembali Informasi dalam Pencarian Dokumen Menggunakan Metode *Vector Space Model*

Ferry Sanjaya

Program Studi Teknik Informatika, Sekolah Tinggi Informatika & Komputer Indonesia (STIKI)
Malang
Email: ferrysanjaya99@gmail.com

ABSTRAK

Seiring banyaknya dokumen menyebabkan berkembangnya teknologi, penyimpanan dokumen dalam bentuk file semakin banyak digunakan. Akan tetapi, file-file dokumen tersebut akan terus bertambah setiap harinya dan untuk mencari informasi dari isi file-file tersebut akan menjadi sulit. Untuk itu perlu diterapkannya metode ilmu pencarian yang dikenal dengan temu kembali informasi (*Information Retrieval*). Salah satunya metode dalam sistem temu kembali adalah *Vector Space Model*. Pada metode VSM sebelum melakukan pencarian dokumen akan dilakukan *indexing* dengan memecah isi teks dari dokumen-dokumen tersebut menjadi *index term*. *Index term* ini yang akan digunakan untuk proses pencarian. Proses pembentukan *index term* dari teks yang terdapat di dalam dokumen akan melalui beberapa tahapan yaitu *parsing*, *text preprocessing*, penghitungan bobot, dan juga pengukuran kesamaan (*Similarity Measure*).

Kata Kunci: *pencarian dokumen, information retrieval, vector space model*

1. PENDAHULUAN

Perkembangan jaman hingga saat ini sangat berdampak besar terhadap kehidupan sekarang. Salah satu hal yang berubah adalah cara menggunakan data sebagai informasi pada era *big data* sekarang ini. Dengan semakin banyaknya jumlah dokumen yang beredar saat ini menimbulkan sebuah masalah untuk melakukan pencarian yang diinginkan dengan cepat dan akurat baik pada media online melalui internet (*search engine, e-library*) ataupun offline (sistem penyimpanan pada komputer). meskipun saat ini beberapa teknologi sudah digunakan

sebuah dokumen ke sebuah *query* didasarkan pada kesamaan di antara vektor dokumen dan vektor *query*. Dalam merepresentasikan dalam sebuah vektor dibutuhkan bobot *term* dari dokumen ataupun *query*. *Term* dapat berupa kata, frase, atau unit hasil *indexing* lain dalam suatu dokumen sebagai gambaran dari isi setiap dokumen tersebut.

Dalam penentuan *term* pada suatu dokumen ataupun *query* diperlukan beberapa tahapan antara lain *filtering*, *stemming* dan *tokenizing*. Setiap *term* tentunya memiliki tingkat kepentingan yang berbeda dalam dokumen untuk itulah diperlukan *term*

Lampiran 2

PERANCANGAN SISTEM TEMU BALIK INFORMASI DENGAN METODE PEMBOBOTAN KOMBINASI TF-IDF UNTUK PENCARIAN DOKUMEN BERBAHASA INDONESIA

Meisya Fitri

Program Studi Teknik Informatika
Jurusan Teknik Elektro Fakultas Teknik Universitas Tanjungpura
meisya.fitri@yahoo.co.id

Abstrak - Information will be easily to access especially from internet that has increased the number of various of digital information. It has increased need of user about information retrieval tools, which it can describe the most relevant document based on user question. Information retrieval can support in managing and finding information effectively and efficiently. It designed to handle searching based on keyword from user. This implemented query expansion technique and term weighting method with combination of TF-IDF (*term frequency-inverse document frequency*). Query expansion is a method that increases the term in query for improving search performance. Term weighting based on combination of TF-IDF gives more weight for more important term. In this system, searching query add based on synonym of searching keywords that the result is not only document which have keywords, but also document associated with searching keyword based on

Akan tetapi, istilah-istilah yang terdapat dalam dokumen dan *query* sering memiliki banyak varian morfologik, sehingga pasangan istilah seperti “bergeser” dan “menggeser” tidak akan dianggap ekuivalen oleh sistem tanpa suatu bentuk *Natural Language Processing* (NLP).

Pembobotan dasar dilakukan dengan menghitung frekuensi kemunculan istilah dalam dokumen karena dipercaya bahwa frekuensi kemunculan istilah merupakan petunjuk sejauh mana istilah tersebut mewakili isi dokumen [1]. Pembobotan kata diharapkan dapat menemukan kembali informasi yang paling relevan dengan indeks istilah terbaik. Metode pembobotan kata berdasarkan kombinasi TF-IDF memberikan bobot lebih kepada istilah yang lebih penting. Dengan menggunakan metode pembobotan kata dapat diperoleh informasi penting dari suatu dokumen berdasarkan kata-kata yang ada dalam dokumen tersebut.

Penambahan istilah pada *query* juga diperlukan

Lampiran 3

PENERAPAN ALGORITMA TF-IDF UNTUK PENCARIAN KARYA ILMIAH

Abdul Azis Maarif

Jurusan Teknik Informatika, Fakultas Ilmu Komputer, Universitas Dian Nuswantoro
Jl. Nakula I No. 5-11 Semarang, Jl. Imam Bonjol No. 207 Semarang, 50131-Indonesia
E-mail: azismaarif@gmail.com

ABSTRAK

Sorting a scientific paper can be done easily by humans, but sorting of documents is done automatically by the computer will bring its own problems. Similarly, by measuring the level of similarity of a document with other documents, humans can easily measure whether a document has the level of similarity/similaritas with other documents. Keyword that is used in the extraction process document in the process of sorting the categories document. In order for the results of the measurement of level of similaritas documents with keywords to get optimal results then used algorithms for text mining algorithm used in the process whereby the TF-IDF (Term Frequency – Inversed Document Frequency) of the model IR (information retrieval) as a measure of the level of similaritas between documents with keywords obtained from the extraction of the text in the document. The purpose of this study was to apply the algorithm TF-IDF that can be used to find the scientific papers as the measuring level similaritas between documents with keywords obtained from the extraction of the text in the document so that it gets sorted data from that similarity (level similaritas) most high so the search papers become more efficient as the relevant information

Kata Kunci : TF-IDF, Scientific Papers, Keyword

1. Pendahuluan

1.1. Latar Belakang

Perkembangan teknologi khususnya internet sangat berperan dalam kehidupan sehari-hari. Dengan adanya internet, informasi dapat dengan mudah disebarluaskan dan diakses oleh banyak orang. Banyaknya informasi yang beredar tentu membuat kebutuhan akan informasi yang relevan semakin meningkat. Salah satu cara yang bisa digunakan untuk mendapatkan informasi yang relevan adalah dengan menggunakan sistem temu kembali

1.4. Manfaat

Adapun manfaat yang diharapkan dalam penelitian ini adalah :

Mempermudah dalam pencarian informasi karya ilmiah karena dapat mempresentasikan hasil informasi secara terurut berdasarkan kemiripan antara *query* dengan informasi yang ada pada dokumen karya ilmiah.

2. Tinjauan Pustaka

2.1. Information Retrieval