
Perancangan Sistem Temu Kembali Informasi dengan Metode Savoy

Wilianto¹⁾

STMIK IBBI

Jl. Sei Deli No. 18 Medan, Telp. 061-4567111 Fax. 061-4527548

e-mail: wilianto@gmail.com¹⁾

ABSTRAK

Sistem Temu Kembali Informasi (*Information Retrieval System*) digunakan untuk menemukan kembali informasi - informasi yang relevan terhadap kebutuhan pengguna dari suatu kumpulan informasi secara otomatis. Salah satu hal yang perlu diingat adalah bahwa informasi yang diproses terkandung dalam sebuah dokumen yang bersifat tekstual. Pengguna sistem temu kembali informasi sangat bervariasi dengan kebutuhan informasi. Akan tetapi, yang menjadi masalah adalah bagaimana teknik mendapatkan informasi tersebut, teknik pengindeksan istilah, teknik pembobotan dokumen dan bagaimana memilih dokumen yang paling relevan sesuai dengan kueri yang diberikan. Tujuan dari penelitian ini adalah untuk menentukan indeks dan bobot dari suatu *file* yang diperlukan dalam proses pengindeksan, mencari dan menentukan *file* yang sesuai dengan *query* yang diberikan serta merancang sebuah aplikasi pencarian *file* dengan metode savoy. Metode penelitian yang digunakan adalah metode *Waterfall*. Dalam penelitian ini yang dijelaskan adalah pengindeksan berdasarkan istilah indeks. Pemberian bobot dokumen dilakukan untuk mengetahui tingkat relevansi dokumen sesuai kueri yang diberikan dan akan dilakukan secara frekwensi dan rumus savoy. Kemudian, dokumen yang telah diindeks akan ditentukan bobotnya, untuk mendapatkan informasi (dokumen) yang dibutuhkan dilakukan teknik boolean peringkat.

Kata Kunci : sistem, temu kembali, metode Savoy

ABSTRACT

Information Retrieval System is used to recover information - information that is relevant to the needs of the users of an automated information collection. One thing to remember is that the processed information contained in a document that is textual. Users of information retrieval system varies with information requirements. However, the problem is how to obtain technical information, term indexing techniques, document weighting techniques and how to choose the most relevant documents in accordance with the given query. The purpose of this study was to determine the weights of the index and a file that is needed in the process of indexing, searching and specify the file that corresponds to a given query and design a file search application with savoy method. The method used is the Waterfall method. Described in this study are based on index terms indexing. Document weighting is performed to determine the appropriate level of relevance of a document given a query and will be done in the frequency and savoy formula. Then, the document that has been indexed will be determined its weight, to get the information (documents) required boolean technique of ranking.

Keywords : system, retrieval, methods of Savoy

1. Pendahuluan

Pencarian (*searching*) merupakan pekerjaan yang sering kita lakukan dalam kehidupan sehari – hari. Dalam *text editor*, kita sering melakukan pekerjaan mencari kata, atau mencari kata1 dan menggantikannya dengan kata2, atau mencari kata tertentu dan menghitung frekuensi kemunculan kata tersebut dalam *file*. *Windows explorer* menggunakan prinsip pencarian dalam melakukan pekerjaan pencarian kata tersebut. Kecepatan perubahan dan penambahan informasi menyebabkan dibutuhkannya suatu sistem yang dapat mengakses dan menyediakan berbagai informasi tersebut. Perbedaan mesin penelusur yang satu dengan yang lain sangat tergantung pada teknik temu kembali informasi dan teknik pengindeksan yang dipakai.

Sistem Temu Kembali dirancang untuk menemukan *file* atau informasi yang diperlukan oleh masyarakat pengguna. Sistem Temu Kembali bertujuan untuk menjembatani kebutuhan informasi pengguna dengan sumber informasi yang tersedia. Sistem Temu Kembali merupakan sistem yang berfungsi untuk menemukan informasi yang relevan dengan kebutuhan pemakai. Dalam hal ini, temu kembali informasi berkaitan dengan representasi, penyimpanan, dan akses terhadap representasi *file*. *File* yang ditemukan tidak dapat dipastikan apakah relevan dengan kebutuhan informasi pengguna yang dinyatakan dalam *query*. Selain itu, pengguna Sistem Temu Kembali sangat bervariasi dengan kebutuhan informasi yang berbeda-beda. Misalkan ada sebuah kumpulan *file* dan seorang *user* yang

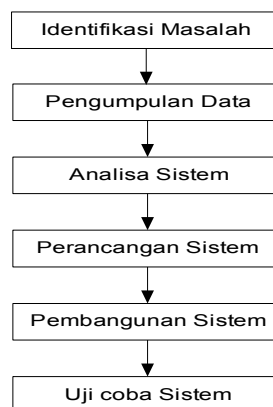
memformulasikan sebuah pertanyaan (*request* atau *query*). Jawaban dari pertanyaan tersebut adalah sekumpulan *file* yang relevan dan membuang *file* yang tidak relevan.

Sistem Temu Kembali pada dasarnya dibagi dalam dua komponen utama yaitu sistem pengindeksan (*indexing*) yang menghasilkan basis data sistem dan temu kembali yang merupakan gabungan dari *user interface* dan *look-up-table*. *File* sebagai objek data dalam Sistem Temu Kembali merupakan sumber informasi. *File* biasanya dinyatakan dalam bentuk indeks atau kata kunci. Kata kunci dapat diekstrak secara langsung dari teks *file* atau ditentukan secara khusus oleh spesialis subjek dalam proses pengindeksan yang pada dasarnya terdiri dari proses analisis dan representasi *file*. Pengindeksan dilakukan dengan menggunakan sistem pengindeksan tertentu, yaitu himpunan kosa kata yang dapat dijadikan sebagai bahasa indeks sehingga diperoleh informasi yang terorganisasi. Sementara itu, pencarian diawali dengan adanya kebutuhan informasi pengguna. Dalam hal ini Sistem Temu Kembali berfungsi untuk menganalisis pertanyaan (*query*) pengguna yang merupakan representasi dari kebutuhan informasi untuk mendapatkan pernyataan - pernyataan pencarian yang tepat. Proses pencarian yang terdapat dalam *Microsoft Windows* mampu mendaftarkan semua *file* yang berhubungan dengan kata (topik) yang diinginkan, namun *Microsoft Windows* tidak mampu mengurutkan hasil pencarian *file* menurut hubungan atau keterkaitan *file* tersebut dengan kata (topik) yang dicari. Untuk menyelesaikan permasalahan tersebut, maka dapat diterapkan sebuah sistem temu kembali informasi.

Berbagai metode *information retrieval system* yang dapat diterapkan mencakup metode *Boolean* sederhana, *Boolean* peringkat dan *Extended Boolean* (Model Savoy). Kelemahan dari metode *Boolean* sederhana adalah model *Boolean* tidak menggunakan peringkat dokumen yang terambil, sedangkan kelemahan dari metode *Boolean* peringkat adalah perhitungan bobot yang kurang akurat karena hanya menggunakan operasi minimal dan maksimal saja. Untuk menyelesaikan permasalahan yang dihadapi oleh kedua metode diatas, maka dapat diterapkan metode Savoy (Nama lengkap Savoy adalah Profesor Jacques Savoy). Dimana kita harus menentukan terlebih dahulu jumlah kata dan *file* yang ada. Kemudian akan dihitung bobot dari suatu *file* untuk mengetahui tingkat relevansi dari suatu *file*. Setelah itu, untuk mendapatkan informasi (*file*) yang dibutuhkan akan dilakukan teknik *extended boolean* sesuai P-Norm. Dimana teknik ini akan mengurutkan *file* yang kita inginkan dan nilai dari P-Norm lah yang akan menentukan diurutkan keberapakah *file* tersebut berada. Pada teknik pembobotan ini, bobot istilah telah dinormalisasi, dengan mempertimbangkan frekwensi kemunculan istilah di satu *file*, tetapi juga memperhatikan frekwensi terbesar pada suatu istilah-istilah lain di *file* yang sama. Selain itu, teknik ini juga memperhitungkan jumlah *file* yang mengandung istilah yang bersangkutan dan jumlah keseluruhan *file*. Hal ini berguna untuk mengetahui posisi relatif bobot istilah bersangkutan pada suatu *file* dibandingkan dengan *file-file* lain yang memiliki istilah yang sama. Ini berarti, jika sebuah istilah mempunyai frekwensi kemunculan yang sama pada dua *file*, belum tentu mempunyai bobot yang sama.

2. Metode Penelitian

Dalam melakukan perancangan sistem, diperlukan sebuah kerangka kerja dalam bentuk skema metode penelitian yang dijadikan sebagai panduan mengenai langkah-langkah yang harus dikerjakan. Adapun skema metode penelitian yang digunakan dapat dilihat pada Gambar 1.



Gambar 1. Skema Metode Penelitian

Adapun uraian dari langkah-langkah penelitian adalah sebagai berikut.

1. Identifikasi Masalah. Pada bagian ini dikemukakan masalah yang ada di dalam pencarian informasi sehingga dibutuhkan metode Savoy untuk alternatif pencarian informasi yang cukup baik.
2. Pengumpulan Data. Pengumpulan data dilakukan dengan cara membaca literatur-literatur yang berhubungan dengan cara kerja dari Metode Savoy.
3. Analisa Sistem. Analisa sistem berhubungan dengan upaya mencari spesifikasi dari rancangan yang akan dibangun berikut dengan cara kerja dari metode Savoy.
4. Perancangan Sistem. Pada bagian ini akan dirancang aplikasi berdasarkan pada data-data yang sudah terkumpul dan hasil analisa sistem.
5. Pembangunan Sistem. Setelah terdapat skema rancangan dari perangkat lunak yang akan dibangun, maka pada bagian ini proses pembangunan sistem akan dilaksanakan berdasarkan pada hasil rancangan yang sudah ada.
6. Uji coba sistem. Bilamana sistem sudah selesai dibangun, maka pada langkah selanjutnya adalah dilakukan uji coba sistem untuk didapatkan gambaran mengenai hasil kerja sistem.

Teknik Pembobotan Savoy

Teknik pembobotan berdasarkan rumus Savoy (1993), ini merepresentasikan bobot (W_{ik}) dari setiap istilah k , dimana $k=1, 2, \dots, t$; dalam sebuah dokumen i adalah:

$$W_{ik} = tf_{ik} * \log \left[\frac{n}{df_k} \right] = tf_{ik} * idf_k ; \dots\dots\dots 1)$$

Keterangan :

- a. W_{ik} adalah bobot dari istilah k dalam dokumen i
- b. tf_{ik} adalah frekwensi dari istilah k dalam dokumen i .
- c. n adalah jumlah dokumen dalam kumpulan dokumen.
- d. df_k adalah jumlah dokumen yang mengandung istilah k .

Rumus di atas terdiri dari dua komponen yaitu tf_{ik} dan idf_k , dimana kedua komponen ini dinormalisasi dalam rentang $[0, 1]$, menjadi :

$$W_{ik} = ntf_{ik} * nidf_k$$

dimana :

$$ntf_{ik} = \frac{tf_{ik}}{\text{Max}_j tf_{ij}} \dots\dots\dots 2) \text{ dan}$$

$$nidf_k = \frac{\log \left[\frac{n}{df_k} \right]}{\log(n)} \dots\dots\dots 3)$$

dimana $\text{Max}_j tf_{ij}$ adalah frekwensi istilah terbesar pada satu dokumen. Pada teknik pembobotan ini, bobot istilah telah dinormalisasi, dengan mempertimbangkan frekwensi kemunculan istilah di satu dokumen, tetapi juga memperhatikan frekwensi terbesar pada suatu istilah-istilah lain di dokumen yang sama. Selain itu, teknik ini juga memperhitungkan jumlah dokumen yang mengandung istilah yang bersangkutan dan jumlah keseluruhan dokumen. Hal ini berguna untuk mengetahui posisi relatif bobot istilah bersangkutan pada suatu dokumen dibandingkan dengan dokumen-dokumen lain yang memiliki istilah yang sama. Ini berarti, jika sebuah istilah mempunyai frekwensi kemunculan yang sama pada dua dokumen, belum tentu mempunyai bobot yang sama.

3. Hasil dan Analisis

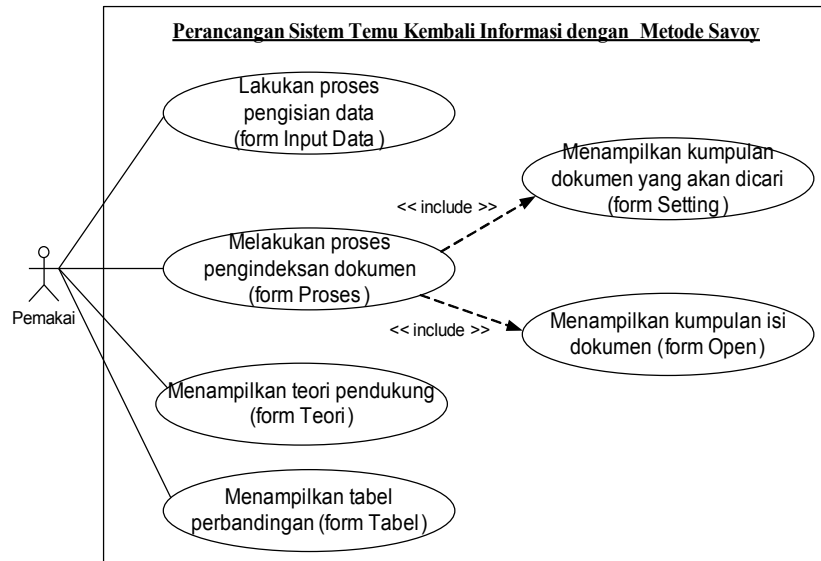
Analisis dan Perancangan

Berdasarkan analisa pada Metode Savoy untuk pencarian kembali informasi, maka diperoleh cara kerja dari metode Savoy yaitu sebagai berikut.

1. Tentukan jumlah kata dan *file* yang ada.
2. Tentukan bobot dari setiap kata atau istilah dengan menggunakan teknik *extended boolean* sesuai dengan kaidah P-Norm yang dikemukakan oleh Savoy. Adapun formulasi untuk menghitung bobot dapat dilihat pada formula 1, 2, dan 3.
3. Istilah atau informasi dengan bobot yang terbesar lah yang akan dijadikan sebagai hasil dari metode Savoy

Perancangan Diagram Use Case

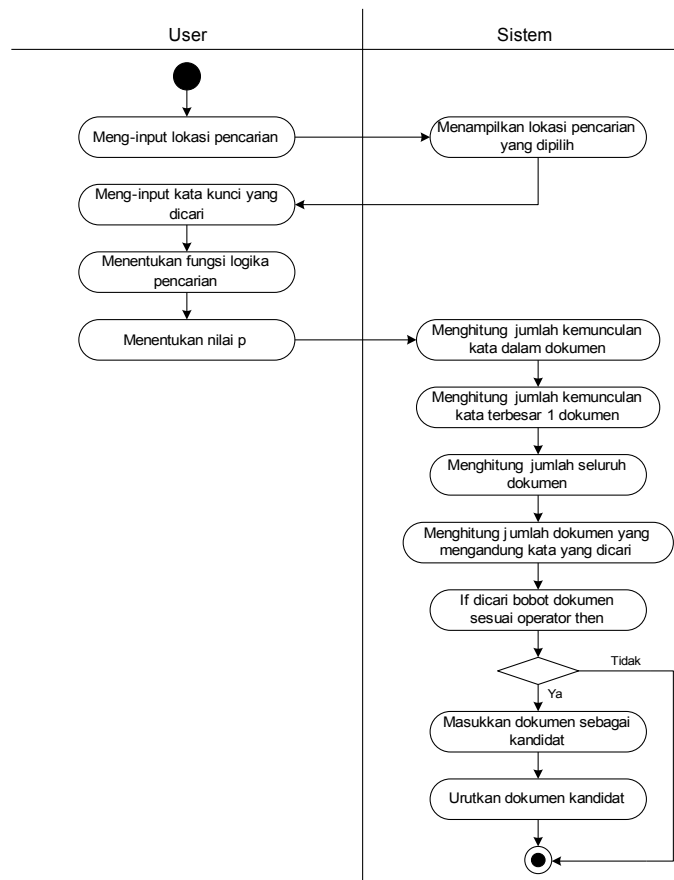
Adapun Diagram *Use Case* dari Perancangan Sistem Temu Kembali Informasi dengan Metode Savoy dapat dilihat pada Gambar 2.



Gambar 2. *Use Case Diagram*

Perancangan Activity Diagram

Adapun *Activity Diagram* dari Perancangan Sistem Temu Kembali Informasi dengan Metode Savoy dapat dilihat pada Gambar 3.



Gambar 3. *Activity Diagram*

Perancangan Form

Adapun perancangan dari *form* yang digunakan di dalam Sistem Temu Kembali Informasi Metode Savoy dapat dilihat pada Gambar 4 sampai dengan Gambar 6.

a. Rancangan *Form* Utama

Adapun rancangan *form* Utama dapat dilihat pada Gambar 4.

The screenshot shows a window titled "Perancangan Sistem Temu Kembali Informasi Metode Savoy". Inside the window, there is a single menu item displayed as "Teori Penemuan Kembali Keluar".

Gambar 4. Rancangan *Form* Utama

b. Rancangan *Form Input* Data

Adapun Rancangan *Form Input* Data dapat dilihat pada Gambar 5.

The screenshot shows a window titled "Input Data dan Setting". It contains several elements: a "Lokasi File" text box with a "Browse" button to its right; two buttons, "Tambahkan ke List" and "Hapus dari List", positioned below the file location; a large empty text area labeled "Isi File"; and two buttons, "Proses" and "Keluar", at the bottom of the window.

Gambar 5. Rancangan *Form Input* Data

c. Rancangan *Form* Proses

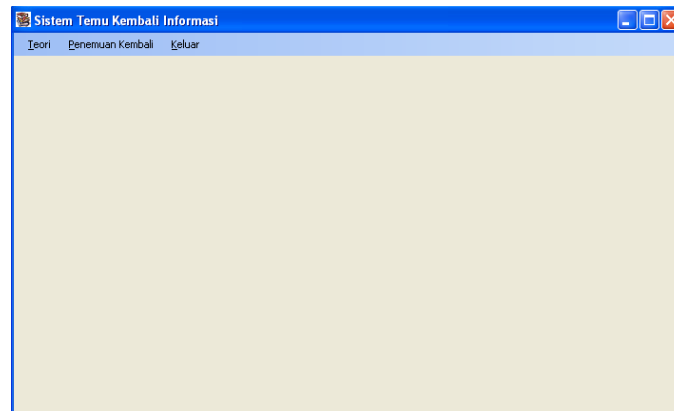
Rancangan *Form* Proses dapat dilihat pada Gambar 6.

The screenshot shows a window titled "Proses Pencarian". It contains: an "Input Kata yang Dicari" text box; a "Nilai P" text box; a "Tambahkan Kumpulan File yang Akan Dicari" button; and "Cari" and "Batal" buttons.

Gambar 6. Rancangan *Form* Proses

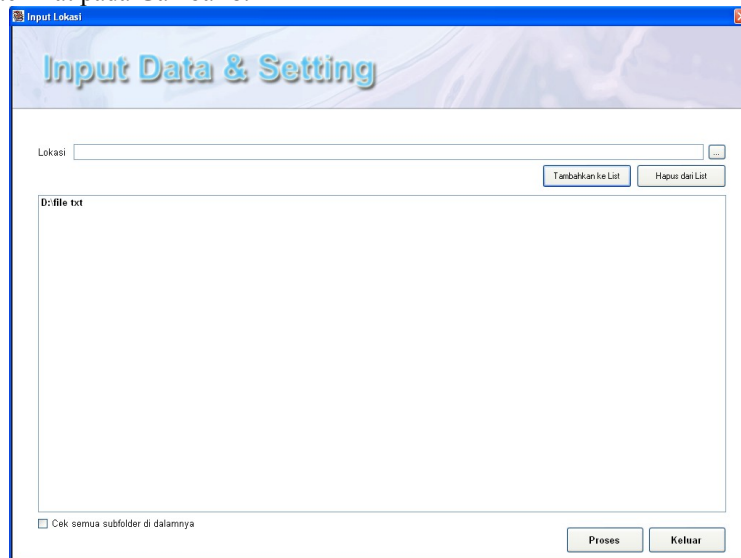
Hasil rancangan pada aplikasi ini adalah sistem yang dapat mencari kembali informasi dengan metode Savoy.

1. Saat sistem ini dijalankan, maka akan muncul tampilan menu utama seperti yang terlihat pada Gambar 7.



Gambar 7. Tampilan Menu Utama

2. Jika dipilih menu penemuan kembali maka akan ditampilkan tampilan menu *Input Data* dan *Setting* seperti yang terlihat pada Gambar 8.



Gambar 8. Tampilan *Form Input Data*

Pada form input data ini, pengguna akan mencari folder-folder mana saja yang diperlukan dan mengklik folder tersebut untuk diproses oleh sistem. Untuk mengecek secara keseluruhan isi dokumen maka pengguna harus menandai *checkbox* yang ada disudut kiri bawah form yang bertuliskan “Cek semua subfolder didalamnya” karena jika tidak menandainya, maka dokumen yang ada didalam subfolder tidak akan diproses juga. Setelah itu, pengguna mengklik button proses untuk menjalankannya.

3. Jika diklik tombol proses, maka akan ditampilkan *form* proses seperti yang terlihat pada Gambar 9.

Gambar 9. Tampilan Form Proses

Pada form proses ini, pengguna akan menginput nilai P dan kata-kata apa saja yang akan dicari serta operator-operator mana yang akan digunakan. Setelah itu, pengguna akan mengklik *button* cari dan sistem akan menjalankan prosesnya dan akan memunculkan hasil pencariannya, yaitu sistem akan memunculkan nama dokumen, lokasi dari file tersebut, nilai indeksnya serta jumlah kata yang terdapat pada dokumen tersebut yang sudah diurutkan berdasarkan nilai indeks dari yang terbesar ke yang paling kecil. Untuk melihat halaman berikutnya, pengguna dapat mengklik *button* halaman berikutnya dan jika pengguna ingin melihat secara acak halamannya maka dapat memilih untuk melihat dari *button* untuk lompat ke halaman yang diinginkan.

4. Jika diklik salah satu file, maka akan ditampilkan form open seperti yang terlihat pada Gambar 10.

Gambar 10. Tampilan Form Open

Pada form open ini, sistem akan menampilkan isi file dari dokumen yang telah dipilih oleh pengguna yang kemudian hasilnya dapat dilihat oleh pengguna. Untuk keluar dari form ini, pengguna dapat mengklik *button* keluar yang ada didalam sistem.

4. Kesimpulan

Setelah menyelesaikan penelitian ini, maka dapat ditarik beberapa kesimpulan antara lain untuk operator “AND” pada metode Savoy, semakin tinggi nilai p menyebabkan nilai bobot dokumen semakin menurun. Dengan menggunakan sistem temu kembali, maka dapat dilakukan pencarian dokumen yang relevan dengan kata kunci yang dimasukkan. Keunggulan dari program yang dibuat mencakup program dapat digunakan untuk menguji pengaruh nilai p terhadap hasil pencarian. Program mampu melakukan pencarian terhadap file berekstensi *.txt dan *.doc.

Berikut diberikan beberapa saran yang berhubungan dengan pengembangan sistem lebih lanjut yaitu untuk menemukan informasi yang akan dicari, *user* sebaiknya menggunakan rumus Savoy karena akan lebih memudahkan dan mempersingkat proses pencarian informasinya. Perlu dicari teknik - teknik lain untuk memperbaiki kinerja teknik frekwensi dan savoy. Perlu dikembangkan atau dicari alternatif lain untuk teknik *Boolean* peringkat dan *extended Boolean*.

Daftar Pustaka

- [1.] Ricardo Baeza-Yates, Berthier Ribeiro-Neto, **Modern Information Retrieval**, ACM Press, New York, 2001.
- [2.] Kadir Abdul, **Pengenalan Sistem Informasi**, Edisi Pertama, Andi Yogyakarta, 2003.
- [3.] <http://www.cs.ui.ac.id/WebKuliah/TKSI/MIK/MIK%20Bab%201%20konsep%20IRS.doc>
- [4.] <http://www.cs.ui.ac.id/WebKuliah/TKSI/MIK/IR%20Models.doc>
- [5.] <http://staff.ui.edu/internal/132137885/material/retrieval%20techniques.doc>
- [6.] <http://yb1zdx.arc.itb.ac.id/data/OWP/library-sw-hw/digitallibrary/gdl31/IDLN%20III%202002%20Meeting-Paper/Rila%20Mandala-Peningkatan%20Performasi%20Sistem%20Temu%20Kembali.doc>
- [7.] http://www.ilkom.unsri.ac.id/dosen/hartini/materi/I_PengantarSI.pdf
- [8.] <http://images.sweetgirl0409.multiply.com/attachment/0/R4MNqAoKCEIAAHNvP4A1/ANALISIS%20SISTEM%20INFORMASI.doc?nmid=76609011>
- [9.] <http://kangbudhi.wordpress.com/2007/10/18/evaluasi-bahasa-alami-vsboolean-query-sebuah-perbandingan-dari-pendayagunaan-sistem-temukembali-informasi/>
- [10.] <http://72.14.235.132/search?q=cache:dPy6Ffg4mR0J:staff.ui.edu/internal/132137885/material/retrieval%2520techniques.doc+sistem+temu+kembali+informasi&hl=id&ct=clnk&cd=6&gl=id>
- [11.] http://pkukmweb.ukm.my/~library/ind_abs.htm
- [12.] <http://www.cs.ui.ac.id/WebKuliah/textpro/PT08-MODEL.pdf>