SISTEM PENCARIAN SKRIPSI
BERBASIS *INFORMATION RETRIEVAL DI FASTIKOM UNSIQ*

Nur Hasanah [a]
[a] Fakultas Teknik dan Ilmu Komputer Universtas Sains Al Qur'an
[a] E-mail: nurh.unsiq@gmail.com

INFO ARTIKEL

Riwayat Artikel :
Diterima : 27 Desember 2016
Disetujui : 31 Desember 2016

Kata Kunci :
Information Retrival, pencarian

ABSTRAK

Pada saat ini perpusakan Fastikom UNSIQ mempunyai jumlah dokumen yang tersedia sangat besar, pencarian secara manual dapat dilakukan dengan membaca setiap dokumen pada koleksi dokumen untuk mendapatkan dokumen yang tepat dan sesuai kebutuhan. Namun, pencarian seperti itu membutuhkan waktu yang lama jika jumlah dokumen sangat banyak. Salah satu solusi untuk mengatasi masalah tersebut adalah dengan menggunakan *Information Retrieval System*. Proses dalam *Information Retrieval* dapat digambarkan sebagai sebuah proses untuk mendapatkan *relevant documents* dari *collection documents* yang ada melalui pencarian *query* yang diinputkan *user*.

Information Retrieval (IR) merupakan suatu metode untuk menemukan kembali data tidak terstruktur yang tersimpan pada sekumpulan dokumen, kemudian menyediakan informasi mengenai subyek yang dibutuhkan. (Bunyamin, 2005) Seiring peningkatan jumlah mahasiswa, maka semakin banyak penelitian dengan judul yang hampir mirip, karena susahnya melakukan penelitian dengan metode yang baru maka salah satu cara mudah untuk mengantisipasi kesamaan judul penelitian adalah dengan menggunakan sebuah sistem, yaitu *Information Retrival* (IR).

ARTICLE INFO

Riwayat Artikel :
Received : December 27, 2016
Accepted : December 31, 2016

Key words*:*
Information Retrival, searching

ABSTRACT

At this time perpusakan Fastikom UNSIQ number of documents available has very large, manual search can be done by reading each document in the collection of documents to get the proper documentation and appropriate. However, the search as it takes a long time if the number of documents very much. One solution to overcome this problem is by using Information Retrieval System. Processes in Information Retrieval can be described as a process to obtain relevant documents from the collection of existing documents through search queries entered by users.

Information Retrieval (IR) is a method to find the back of unstructured data stored in a set of documents, and then presents information on the subject is needed. (Bunyamin, 2005) As the increase in the number of students, the more research with a title that is almost similar, as difficult to do research with the new method it is one easy way to anticipate the similarity title of the research is to use a system, namely Information retrieval (IR) ,

1. PENDAHULUAN

Pada saat ini perpustakaan Fastikom UNSIQ mempunyai jumlah dokumen yang tersedia sangat besar, pencarian secara manual dapat dilakukan dengan membaca setiap dokumen pada koleksi dokumen untuk mendapatkan dokumen yang tepat dan sesuai kebutuhan. Namun, pencarian seperti itu membutuhkan waktu yang lama jika jumlah dokumen sangat banyak. Salah satu solusi untuk mengatasi masalah tersebut adalah dengan menggunakan *Information Retrieval System*. Proses dalam *Information Retrieval* dapat digambarkan sebagai sebuah proses untuk mendapatkan *relevant documents* dari *collectiondocuments* yang ada melalui pencarian *query* yang diinputkan *user*.

Information Retrieval (IR) merupakan suatu metode untuk menemukan kembali data tidak terstruktur yang tersimpan pada sekumpulan dokumen, kemudian menyediakan informasi mengenai subyek yang dibutuhkan. (Bunyamin, 2005) Tujuan dari sistem IR ini adalah memenuhi kebutuhan informasi pengguna dengan mendapatkan semua dokumen yang relevan dengan kebutuhan pengguna dan pada waktu yang sama mendapatkan sesedikit mungkin dokumen yang tak relevan.

Pengguna dapat menemukan informasi yang relevan dengan membaca seluruh dokumen yang ada pada tempat penyimpanannya, menyimpan dokumen-dokumen yang relevan, membuang dokumen yang tidak relevan, dan mengurutkan dokumen-dokumen yang sesuai dengan keperluannya. Hal tersebut merupakan sistem IR yang sempurna, tetapi solusi ini tidak praktis dan efisien. Dikarenakan pengguna tidak memiliki banyak waktu untuk membaca seluruh dokumen satu per satu dari sekian banyak dokumen yang ada.

Agar diperoleh sistem yang dapat mengambil dokumen yang relevan dan mengurutkan pada urutan teratas dibutuhkan suatu metode *matching* suatu *query* dengan koleksi dokumen menggunakan metode yang optimal dalam menentukan nilai *similarity score* suatu dokumen. Pada jenjang pendidikan diploma dan jenjang yang lebih tinggi, mahasiswa diwajibkan untuk menyusun karya ilmiah, baik berupa jurnal penelitian, tugas akhir, skripsi, tesis, disertasi dan lain sebagainya. Seiring dengan kesadaran masyarakat akan pentingnya pendidikan, maka berdampak pada semakin meningkatnya jumlah mahasiswa pada suatu perguruan tinggi yang berpengaruh juga terhadap banyaknya karya ilmiah yang dihasilkan oleh mahasiswa.

Apabila pada sebuah perguruan tinggi meluluskan 500 mahasiswa tiap tahun, maka pada kurun waktu 10 tahun sudah terkumpul 5000 judul penelitian yang berbeda (unik). Seiring peningkatan jumlah mahasiswa, maka semakin banyak penelitian dengan judul yang hampir mirip, karena susahnya melakukan penelitian dengan metode yang baru.

Salah satu cara mudah untuk mengantisipasi kesamaan judul penelitian adalah dengan menggunakan sebuah sistem, dimana apabila dilakukan secara manual sangat tidak memungkinkan bagi seorang mahasiswa untuk mengecek lebih dari 5000 judul karya ilmiah satu per satu.

2. LANDASAN TEORI

2.1. *Information Retrieval* (IR)

Information Retrieval System atau Sistem Temu Balik Informasi merupakan bagian dari *computer science* tentang pengambilan informasi dari dokumen-dokumen yang didasarkan pada isi dan konteks dari dokumen-dokumen itu sendiri. Menurut Gerald J. Kowalski di dalam bukunya *"Information Storage and Retrieval Systems Theory and Implementation*", sistem temu balik informasi adalah suatu sistem yang mampu melakukan penyimpanan, pencarian, dan pemeliharaan informasi. Informasi dalam konteks ini dapat terdiri dari teks (termasuk data numerik dan tanggal), gambar, audio, video, dan objek multimedia lainnya. Sistem temu balik informasi sebagai suatu sistem yang berfungsi untuk menemukan informs yang relevan dengan

kebutuhan pemakai, yang merupakan salah satu tipe istem informasi. (Mulyadin, 2014).

Prinsip kerja *Information Retrieval System* jika ada sebuah kumpulan dokumen dan seorang user yang memformulasikan sebuah pertanyaan (*request* atau *query*). Jawaban dari pertanyaan tersebut adalah sekumpulan dokumen yang relevan dan membuang dokumen yang tidak relevan (Salton, 1989).

information retrieval (IRS) merupakan suatu sistem yang menemukan informasi yang sesuai dengan kebutuhan *user* dari kumpulan informasi secara otomatis. Aplikasi *Information Retrieval System* sudah digunakan dalam banyak bidang seperti dikedokteran, perusahaan dan lain sebagainya. Salah satu aplikasi dari *Information Retrieval System* adalah mesin pencari yang dapat diterapkan diberbagai bidang. Pada mesin pencari dengan *Information Retrieval System user* dapat memasukkan *query* yang bebas dalam arti kata *query* yang sesuai dengan bahasa manusia dan sistem dapat menemukan dokuen yang sesuai dengan *query* yang ditulis oleh *user. (*Amin, 2005).

Tiga komponen utama dalam sistem temu balik adalah masukan (*input*), proses (*processor*)dan keluaran (*output*). Menurut maning et al (2009) mengakatakn bahwa Information Retrival (IR) adalah menemukan bahan (biasanya dokumen) yang bersifat tidak trstruktur (biasanya teks) yang memenuhi kebutuhan informasi dari dalam koleksi besar (bisanya disimpan di komputer)

2.2. Model *Information Retrieval*

Model IR adalah model yang digunakan untuk melakukan pencocokan antara term-term (kata) dari *query* dengan *term-term* dalam *document collection* (*folder file*), model yang terdapat dalam IR terbagi dalam 3 model besar, yaitu (J.Kowalski Gerald, 2000):

a. *Set-theoritic models,* model merepresentasikan dokumen sebagai himpunan kata atau *frase*. Contoh model ini ialah *Standard Boolean model* dan *Extended Boolean model.*
b. *Algebraic model,* model merepresentasikan dokumen dan *query* sebagai vektor *similarity* antara vektor dokumen dan vektor *query* yang direpresentasikan sebagai sebuah nilai skalar. Contoh model ini ialah *Vektor Space Model* (model ruang vektor), *Latent Semantic Indexing* (LSI) dan *Generalized Vector Space Model* (GVSM).
c. *Probabilistic model,* model memperlakukan proses pengambilan dokumen sebagai sebuah *probabilistic inference.* Contoh model ini ialah penerapan teorema *bayes* dalam model probabilistik.

2.1.2. Arsitektur *Information Retrieval System*

10 dalam aplikasi teori probabilitas. Teori probabilitas dapat digunakan untuk menghitung suatu ukuran relevansi antara sebuah *query* dan sebuah dokumen.

Prinsip dari pemodelan probabilistik ini adalah dengan melakukan estimasi bobot suatu kata berdasarkan prinsip seberapa sering kata tersebut muncul atau tidak muncul baik dalam dokumen-dokumen yang relevan maupun dalam dokumen-dokumen yang tidak relevan. Model probabilistik akan memberikan nilai probilitas pada tiap kata yang menjadi komponen dalam suatu *query*, dan kemudian menggunakan bukti-bukti tersebut untuk menghitung probabilitas akhir bahwa suatu dokumen relevan dengan suatu *query*.

3. METODOLOGI PENELITIAN

3.1. Data Penelitian

Data penelitian yang digunakan dalam penelitian ini adalah judul skripsi atau tugas akhir mahasiswa UNSIQ terutama program studi Teknik Informatika dan Manajemen Informatika. Penulis hanya mengambil judul kripsi Teknik Informatika dan Manajemen Informatika dimaksudkan agar sistem yang akan dibuat dapat bekerja secara optimal untuk mencari kemiripan judul skripsi dari satu rumpun atau jurusan.

3.2. Metode Pengumpulan Data

Metode pengumpulan data yang digunakan dalam penelitian ini adalah sebagai berikut:

3.2.1. Sumber Data

Data yang diperlukan untuk membuat aplikasi diambil dari:

a. Data Primer, yaitu data yang diperoleh langsung dari sumbernya, baik melalui wawancara ataupun observasi dengan pihak-pihak terkait. Pada penelitian ini, yang dijadikan sebagai data utama penelitian yaitu judul skripsi mahasiswa Teknik Informatika dan Manajemen Informatika yang berjumlah 20 judul yang diperoleh dari perpustakaan.

3.2.2. Teknik Pengumpulan Data

Pada bagian ini, penulis melakukan pengumpulan data-data yang terkait langsung sesuai dengan kebutuhan untuk membangun aplikasi temu kembali (*Information Retrieval*). Teknik pengumpulan data dalam penelitian ini adalah sebagai berikut:

a. Observasi

Pengumpulan data dengan melakukan pengamatan langsung terhadap data yang diperlukan. Langkah observasi ini dilakukan dengan mengumpulkan data judul skripsi atau tugas akhir mahasiswa Teknik Informatika dan Manajemen Informtika UNSIQ serta langkah-langkah membandingkan kemiripan antar judul skripsi atau tugas akhir oleh dosen sehingga dipastikan tidak ada judul skripsi atau tugas akhir yang sama.

b. *Interview*

Pengumpulan data dengan melakukan wawancara atau tanya jawab langsung dengan seorang yang dianggap berkompeten. Dalam tahap ini, penulis mewawancarai berberapa dosen pembimbing skripsi Teknik Informatika dan Manajemen Informatika mengenai tata cara membandingkan judul skripsi atau tugas akhir Teknik Informatika dan Manajemen Informatika UNSIQ.

c. Literatur

Pengumpulan data dengan melakukan studi pustaka mencakup buku-buku teks, diktat, makalah, artikel di media cetak dan internet dan buku petunjuk teknis terpadu yang berkaitan dengan objek penelitian. Penulis melakukan studi literatur terutama di perpustakaan UNSIQ dan sumber-sumber lain yang sebagian besar diperoleh dari berbagai jurnal penelitian yang tersedia pada perpustakaan digital beberapa perguruan tinggi di Indonesia.

4. HASIL DAN PEMBAHASAN

4.1. Perancangan Sistem

4.1.1. Kebutuhan Masukan

Kebutuhan masukan dari aplikasi sistem pencarian arsip skripsi berbasis *information retrieval* menggunakan metode probabilistic yaitu berupa abstrak dari skripsi berjumlah 20 buah sebagai sampel.

4.1.2. Kebutuhan Proses

a. Membaca data abstrak

Proses pertama pada *information retrieval* ini yaitu membaca data abstrak skripsi yang telah dimasukkan ke dalam database sebelumnya. Penulis memilih data abstrak yang digunakan sebagai objek pencarian daripada judul skripsi

dikarenakan abstrak dari skripsi lebih mewakili keseluruhan isi dan juga wilayah pencariannya lebih akurat. Data abstrak skripsi dibaca kemudian dibuat menjadi koleksi menggunakan *array*.

b. Menghilangkan Tanda Baca

Koleksi abstrak skripsi yang diperoleh dari pembacaan data, masing-masing dihilangkan tanda bacanya. Tanda baca pada sistem IR tidak termasuk ke dalam objek yang diperhitungkan, sehingga harus dihilangkan. Diantara tanda baca yang dihilangkan yaitu titik (.), koma (,), garis miring (/), kurang buka dan kurung tutup, simbol-simbol seperti operator matematika (kali, bagi, persen, dan lain sebagainya). Hasil dari proses penghilangan tanda baca yaitu koleksi abstrak yang siap diproses dan hanya terdiri dari alfabet (a-z) serta spasi.

d. Konversi menjadi huruf kecil

Proses pencarian dengan model IR membutuhkan objek yang seragam, sehingga diperlukan perubahan dari kalimat menjadi huruf kecil semua. Hal ini tentunya sangat berpengaruh, mengingat kata Ekonomi dan ekonomi pada beberapa bahasa pemrograman merupakan kata yang berbeda disebabkan perbedaan sturktur karakter. Kata Ekonomi yang pertama huruf E nya kapital, sedangkan kata ekonomi yang kedua huruf nya kecil semua.

e. *Tokenizing*

Koleksi abstrak skripsi yang telah dibersihkan dipecah (dibagi) berdasarkan spasi dan disimpan ke dalam *array*. *Array* ini yang selanjutnya akan dioperasikan untuk proses-proses sistem IR berikutnya.

f. Stopwords Filtering

Setelah data abstrak skripsi dijadikan token (per kata), langkah selanjutnya adalah proses pengecekan adanya kata umum pada abstrak tersebut. Apabila ditemui kata umum (misalnya yang, dari, dan, dengan, dan sebagainya) maka akan dihapus. Hal ini tentunya akan meningkatkan akurasi dari pencarian dimana objek yang dicari benar-benar berbeda antara abstrak yang satu dengan lainnya.

g. *Stemming*

Langkah berikutnya yaitu *stemming* (mencari akar kata atau kata dasar). Proses *stemming* sendiri berdampak pada akurasi hasil pencarian, dimana apabila dengan menggunakan stemming, maka kata perjanjian dan janji akan diperoleh bentuk kata dasar yang sama yaitu janji.

h. Menghitung nilai probabilistik

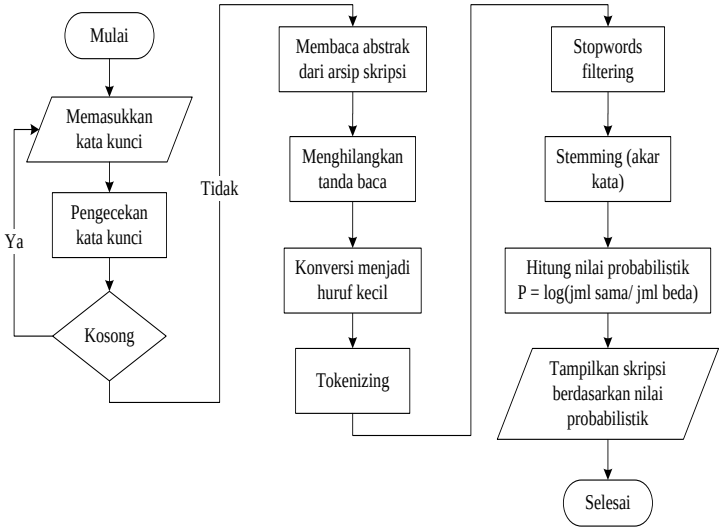
Setelah dipersiapkan pra proses (*preprocessing*) meliputi keenam langkah di atas, maka langkah inti dari sistem IR metode probabilistik yaitu menghitung probabilitas dari masing-masing dokumen, dalam hal ini abstrak skripsi untuk dicari nilai probabilitas tertingginya. Nilai probabilitas tertinggi berarti mempunyai kemiripan yang paling mirip dengan kata kunci. Rumus menghitung nilai probabilistik yaitu log(jumlah sama/jumlah beda).

4.1.3. Data Keluaran

Data keluaran yang diinginkan dari aplikasi sistem IR yang dibuat yaitu berupa daftar hasil pencarian arsip skripsi berdasarkan nilai probabilistik tertinggi hingga terendah. Hasil tersebut dapat dijadikan acuan dalam berbagai hal, misalnya mengukur kemiripan antar skripsi dan lain sebagainya .

Data lain yang dibutuhkan yaitu berupa data kata umum yang dijadikan acuan pada proses *stopwords filtering* dan juga data kata

dasar yang digunakan pada proses *stemming*. Data kata dasar diperoleh dari kata dasar pada Kamus Besar Bahasa Indonesia (KBBI). Sedangkan kata umum diperoleh dari kata-kata umum pada beberapa topik penelitian.



Gambar 4.1 *Flowchart* Proses IR Probabilistik

4.2. Aplikasi IR Metode Probabilistik

4.2.1. Koneksi *Database*

Aplikasi yang memanfaatkan basis data tentunya memiliki koneksi ke *database* yang digunakan. Pada pemrograman PHP, koneksi *database MySQL* sangat sederhana, karena modul *MySQL* sudah terintegrasi dengan modul *Apache* dalam satu lingkungan pengembangan aplikasi.

Modul koneksi ini digunakan juga pada sebagian besar modul pada aplikasi guna berkomunikasi dengan basis data.

4.2.2. Autentikasi

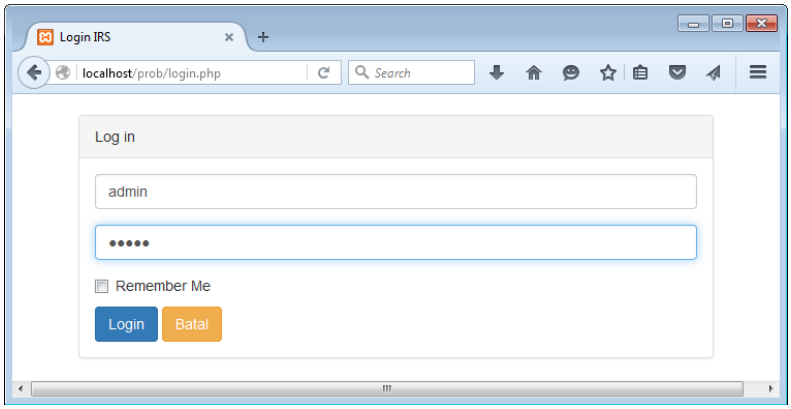
Sistem autentikasi (meliputi *login* dan *logout*) merupakan modul dasar pada sebuah sistem informasi. Latar belakang perlunya autentikasi yaitu untuk membatasi akses sistem bagi beberapa pihak, seperti administrator dan pengguna tamu. Modul-modul yang dapat diakses oleh tiap-tiap pengguna dapat dibedakan berdasarkan hak akses.

Pada aplikasi IR metode probabilistik, pengguna tamu tidak memerlukan proses autentikasi, tetapi hanya dapat mengakses modul pencarian, sedangkan pengguna sebagi *administrator* yang melalui proses autentikasi dapat mengakses seluruh fitur yang tersedia pada aplikasi.

4.2.3. Implementasi Sistem

Tahap implementasi sistem ini adalah merupakan suatu penerapan sistem yang terkomputerisasi terhadap seluruh data dan informasi dalam menenukan keterkaitan antar skripsi yang berbasis web dengan tujuan mempermudah mahasiswa untuk mengetahui keterkaitan antar skripsi tersebut sehingga diharapan bisa memerikan informasi yang tepat, cepat dan akurat.

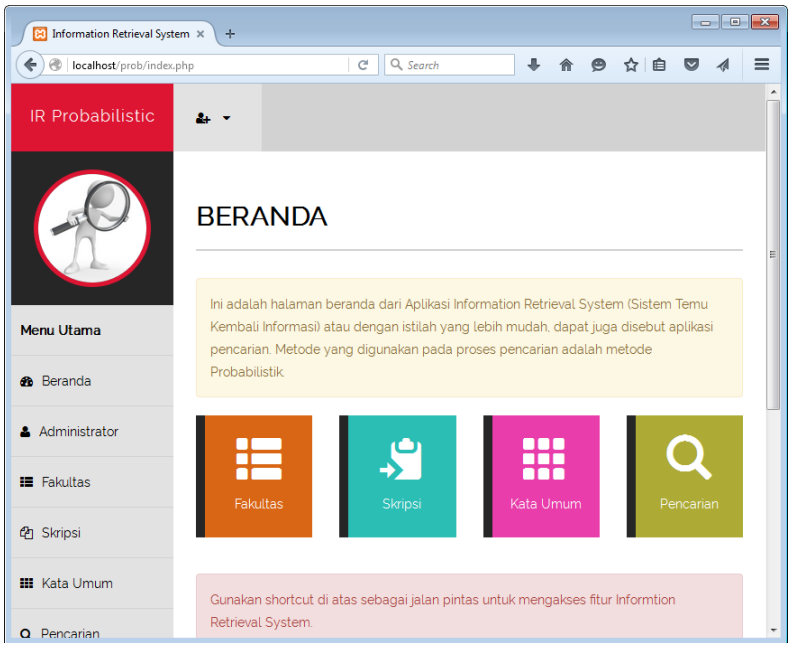
4.2.4. Tampilan Halaman Login



Gambar 4.2. Tampilan Halaman Login

4.2.5. Tampilan Halaman Beranda

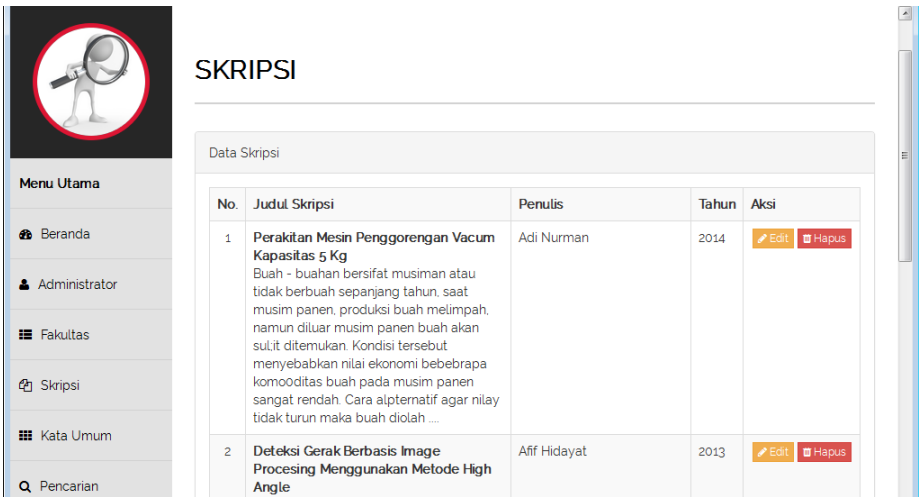
Modul beranda memuat informasi umum aplikasi, struktur menu secara lengkap dan *shortcut* (pintasan) menuju modul-modul tertentu. Tampilan halaman beranda dapat dilihat pada gambar berikut:



Gambar 4.3. Tampilan Halaman Beranda

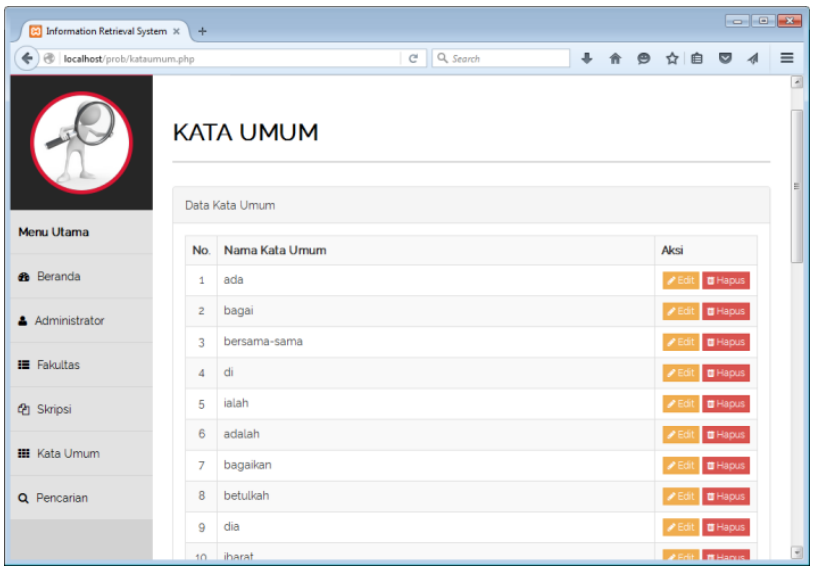
4.2.6. Tampilan Data Skripsi

Modul skripsi merupakan modul yang digunakan untuk menampilkan, menambah, mengubah dan menghapus data skripsi. Data skripsi merupakan objek utama pada sistem IR yang akan dibangun menggunakan metode *probabilistic*. Data skripsi memuat id skripsi, fakultas, judul, abstrak, penulis, dan tahun.



Gambar 4.4. Tampilan Halaman Data Skripsi

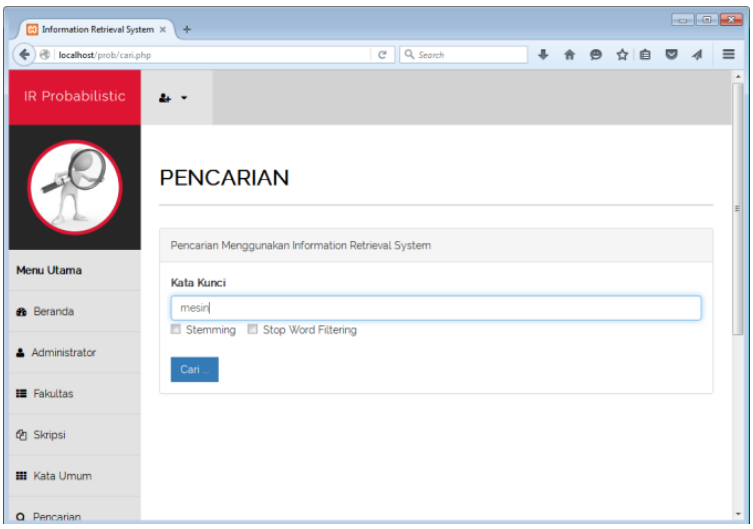
Modul kata umum merupakan modul yang digunakan untuk menampilkan, menambah, mengubah dan menghapus data kata umum. Data data umum digunakan pada proses *stopwords filtering*, yaitu proses yang pada intinya menghilangkan kata-kata yang sifatnya umum pada suatu teks.



Gambar 4.5. Tampilan Halaman Data Kata Umum

4.2.7. Tampilan Pencarian

Modul pencarian merupakan modul utama pada aplikasi IR metode probabilistik. Pada modul ini dimuat proses pengambilan data skripsi, menghilangkan tanda baca, *tokenizing*, *stopwords filtering*, *stemming*, perhitungan nilai probabilistik sampai pada menampilkan hasil pencarian. Berikut merupakan tampilan form pencarian:



Gambar 4.6. Tampilan Halaman Pencarian

Pada form pencarian, pengguna harus memasukkan kata kunci, memilih opsi menggunakan *stemming* atau tidak, dan memilih opsi penggunaan fitur *stopwords filtering*.

5. PENUTUP

5.1. Simpulan

Dari perancangan, implementasi dan pengujian aplikasi temu kembali (*information retrieval*) metode probabilistik, maka dapat diambil kesimpulan sebagai berikut:

1. Berdasarkan Hasil Perancangan, implementasi dan pengujian yang dilakukan pada aplikasi ini dapat terbangun.
2. *Stemming* dan *stopwords filtering* mempengaruhi nilai *probabilistic*, yaitu berpengaruh terhadap berkurangny anilai *probabilistic*.
3. Hasil pengujian menggunakan standar ISO 9126 diperoleh hasil rata-rata pengujian bernilai 79 sehingga dapat disimpulkan bahwa aplikasi yang dibuat layak digunakan.

5.2. Saran

Berdasarkan ujicoba, ada beberapa saran yang perlu dipertimbangkan dalam pengembangan penelitian ini, yaitu:

1. Pengembangan aplikasi sehingga dapat menerima semua jenis dokumen yang berbasis teks, misalnya .txt, .html, dan lain sebagainya.
2. Optimalisasi *query* SQL dan proses pencarian agar waktu pencarian lebih cepat.
3. *Multi language stemming*, terutama untuk kata - kata berbahasa Inggris.

6. DAFTAR PUSTAKA

Amin, Fathul dan Purwatiningtyas. 2015. Rancang Bangun *Informations Retrival System* (IRS) Bahasa Jawa Ngoko pada Palintangan Penjebar Semangad dengan Metode *Vector Space Model* (VSM). Jurnal Teknologi Informasi DINAMIKA Volume 20, No. 1

Bunyamin. 2005. *Algoritma Umum Pencarian Informasi Dalam Sistem Temu*. Institut Teknologi Bandung. Bandung

Gerald J. Kowalski. 2000. *Information Storage and Retrieval Systems: Theory and Implementation*. United States

Manning, D. Christopher, Raghavan, P. & Schütze H. 2009. *An Introduction to Information Retrieval*. Cambridge University Press

Mulyadin dan Eko Aribowo. 2014. Sistem Penentuan Keterkaitan antar skripsi berdasarkan *keyword seeking.Jurnal* Sarjana Teknik Informatika Volume 2 Nomor 1

Salton, G., 1989, *Automatic Text Processing, The Transformation, Analysis, and Retrieval of information by computer*. Addison – Wesly Publishing Company, Inc. USA.