JURUSAN TEKNIK INFORMATIKA

FAKULTAS TEKNOLOGI INFORMASI

**INSTITUT TEKNOLOGI SEPULUH NOPEMBER**

**SURABAYA**

**USULAN TUGAS AKHIR**

1. **IDENTITAS PENGUSUL**

**NAMA : Indra Lukmana**

**NRP : 5106 100 801**

**DOSEN WALI : Yudhi Purwananto, S.Kom. M.Kom.**

1. **JUDUL TUGAS AKHIR**

***”Implementasi Perangkingan Dokumen Berbahasa Arab Berdasarkan Susunan Posisi Kata dari Query”***

1. **ABSTRAK**

Hasil akhir dari suatu pencarian dokumen biasanya berupa nilai relevansi antara suatu kumpulan dokumen dengan suatu query. Berdasarkan relevansi tersebut di bentuk suatu sistem perangkingan. Sistem ini biasanya merepresentasikan tingkatan keterkaitan antara suatu query dengan dokumen, semakin tinggi nilai relevansi dari suatu dokumen maka akan semakin tinggi pula rangking pada hasil pencarian. Penentuan nilai relevansi pada sistem temu kembali informasi yang berdasarkan pada vector space biasanya menggunakan metode cosine similarity. penentuan nilai ini biasanya hanya memperhatikan perhitungan relevansi kata perkata dari query dengan dokumen. Tugas akhir ini mencoba untuk melakukan perhitungan nilai relevansi antara dokumen dengan query dengan memperhatikan posisi kata dari query, semisal apakah kata - kata dari query itu nanti membentuk suatu frase atau tidak.

1. **PENDAHULUAN**
   1. **Latar Belakang Masalah**

Peningkatan rangking hasil pencarian dari suatu dokumen setelah pencarian pertama telah di teliti secara mendalam dalam temu kembali informasi. Berdasarkan [1] dua cabang penelitian utama pada topik ini adalah: *query expansion* (QE)dan *automatic dokumen reranking*.

Pada QE sejumlah hasil pencarian dengan rangking tertinggi di asumsikan merupakan hasil yang relevan, dan dari hasil tersebut dilakukan perluasan *query.* Setelah itu dilakukan pencarian kembali dengan query baru yang diperluas. Dari pencarian ini di peroleh hasil pencarian yang relevan dengan query baru yang telah diperluas.

Pada *automatic dokumen reranking* peningkatan rangking dokumen dilakukan pada saat pencarian awal sehingga dokumen relevan yang di peroleh lebih tinggi. Dari proses ini hasil yang didapat dapat di gunakan oleh QE untuk lebih meningkatkan hasil pencarian yang relevan.

Dalam dokumen berbahasa arab kemunculan suatu rangkaian kata yang terjadi lebih dari satu kali biasanya kemungkinan rangkaian kata tersebut membentuk frase yang bermakna dari pada bahasa lain [4], seperti bahasa inggris atau bahasa indonesia, sebagai contoh rangakian kata bahasa arab adalah: كتب الفقة (kitab fiqih),غرفة الجلس (ruang duduk), dan اصحب الكهف(penghuni goa).

Di dalam bahasa arab banyak banyak kata-kata yang memiliki makna intrinsik tersendiri, sebagai contoh kata كتب dapat memiliki makna sesuai dengan harokat (diakritik dalam bahasa arab), kata tersebut bisa menjadi *fi’il ma’lum* (kata kerja aktif) كَتَبَ (menulis), bisa menjadi *fi’il majhul* (kata kerja pasif) كُتِبَ (ditulis) , dan dapat juga menjadi *isim* كِتَبٌ (kitab/buku). Pada saat kata كتب menjadi *fi’il* (kata kerja) sebenarnya ada *fa’il* (subyek) yang tersimpan (*mudhmar)* sehingga artinya yang sebenarnya dia (laki-laki) menulis/ditulis.

Pada sistem temu kembali, kata-kata bahasa arab yang memiliki diakritik biasanya pada tahap prapemrosesan dirubah menjadi bentuk kata paling dasarnya. Pada saat kata-kata yang memiliki diakritik atau imbuhan dirubah menjadi kata dasarnya, kata tersebut kehilangan makna intrinsik awal yang dimaksud. Makna kata sebelum dilakukan normalisasi dapat dipertahankan dengan merangkaikan kata tersebut dengan kata sebelum dan sesudahnya. Sebagai contoh: كَتَبَ احمد (ahmad menulis) dan كِتَبٌ الفقة (kitab fiqih), setelah tahap prapemrosesan di peroleh hasil كتب, احمد, dan فقة .Pada hasil tersebut makna kata كتب tidak dapat diketahui secra pasti karena telah ternormalisasi. Dengan merangkaikan kata كتب dengan kata yang mengikutinya dapat dilihat perbedaan makna seperti pada contoh pertama dan kedua. Sehingga diasumsikan rangkaian kata yang muncul pada dokumen berbahasa arab memiliki andil yang lebih besar dalam memberikan makna pada dokumen dari pada kata yang berdiri sendiri.

Tugas akhir ini menggunakan metode *automatic dokumen reranking* untuk melakukan perbaikan hasil pencarian. Dengan menitik beratkan pada dokumen-dokumen yang memiliki rangkaian kata berdasarkan query. Dengan penitik beratan ini diharapkan perangkingan dokumen dapat lebih relevan.

* 1. **Perumusan Masalah**

Permasalahan yang diangkat dalam pembuatan tugas akhir ini adalah bagaimana menerapkanperangkingan hasil pencarian berdasarkan posisi kata dari *query* dalam proses pencarian dokumen berbahasa arab?

* 1. **Batasan Masalah**

Dari permasalahan yang telah disebutkan di atas, maka batasan-batasan dalam tugas akhir ini adalah:

1. Dokumen yang digunakan sebagai data diambil dari website <http://www.shamela.ws/old_site/index.php>. kitab – kitab tersebut berjumlah 63 kitab yang diambil dari kategori fiqih syafi’i
2. Proses-proses prapemrosesan dan pencarian mengembangkan dari *lucene library* dan [3]
   1. **Tujuan dan Manfaat Tugas Akhir**

Tujuan tugas akhir ini adalah membuat aplikasi pencarian yang mengimplementasikan perangkingan berdasarkan posisi kata dari *query*. Dengan menerapkan metode perangkingan tersebut akan meningkatkan kualitas hasil pencarian dengan memberikan hasil yang lebih relevan.

1. **KAJIAN PUSTAKA**

Tugas akhir ini dikembangkan dengan di dasarkan pada paper [1], paper ini menjelaskan tentang perangkingan dari suatu hasil pencarian. Metode yang digunakan pada paper ini menitik beratkan penggunaan *keyterm* yaitu suatu kata atau rangkaian kata yang dikembangkan dari *query* sebagai penentu definisi suatu dokumen. *keyterm* lebih diperhatikan karena biasanya memiliki makna yang lebih lengkap dari pada hanya satu kata yang berdiri sendiri

Dalam metode ini nanti akan dilakukan dua tahap pencarian yaitu: **pencarian pertama dan kedua**. Pada **pencarian pertama** dilakukan pencarian dengan basis *vector space* [3], tahapan-tahapannya adalah prapemrosesan, pembentukan *index*, dan pencarian. prapemrosesan berkaitan erat dengan penerapan proses pembentukan *index* pada suatu sistem temu kembali informasi. Suatu skema prapemrosesan biasanya tidak hanya berurusan dengan bagaimana merepresentasikan suatu dokumen teks, tapi juga bagaimana menghasilkan solusi yang efektif dan efisien. Prapemrosesan juga sering diterapkan pada dokumen pencarian. Berikut adalah salah satu contoh skema prapemrosesan otomatis dokumen teks berbahasa Arab:

Identifikasikan kata per kata dalam suatu kumpulan dokumen teks.

Normalisasi teks Arab, pembuangan diakritik, dan perubahan huruf *hijaiyah* ke bentuk normal. Misalnya, perubahan huruf *hamzah* (أ dan إ) ke bentuk *alief* (ا).

Lakukan proses *filtering* dan *stopping* (menghilagkan kata-kata *stopword*) dengan berbasis pada daftar *stopword* yang ditentukan sebelumnya.

Gunakan *stemmer* untuk mereduksi jumlah *term* (kata) yang ada.

Untuk setiap *term* yang tersisa, dilakukan proses pembobotan *term* dengan metode TF-IDF.

 (1)

wij : menandakan bobot atau seberapa penting suatu term j pada dokumen i

tfij : merupakan frekuensi term j pada dokumen i

N : merupakan jumlah total dokumen yang diproses

dfj : adalah jumlah dokumen yang memiliki term j didalamnya.

Representasikan setiap dokumen dengan berdasarkan pada *term* dan nilai pembobotannya.

Untuk proses pencarian berdasarkan [3] dari bobot TF-IDF dokumen di tentukansimilaritas kosinus dengan *query*. Hasil pencarian ini adalah dokumen yang memiliki nilai relevansi paling tinggi dengan query sejumlah K.

Dengan dasar pencarian pertama ini dimisalkan: terdapat dua puluh dokumen, setelah dilakukan pencarian dengan *query* “batik Indonesia” diperoleh lima dokumen sebagai berikut:

|  |  |  |  |
| --- | --- | --- | --- |
| Dokumen | Rank | Cosine | Isi Dokumen |
| d1 | 3 | 0.121 | Oleh karenanya, pihaknya juga telah mensosialisasikan batik indonesia kepada ratusan pengrajin batik di berbagai sentra kerajinan batik di Indonesia |
| d2 | 5 | 0.092 | Batik Indonesia secara resmi diakui UNESCO dengan dimasukkan ke dalam Daftar Representatif sebagai Budaya Tak-benda Warisan Manusia |
| d3 | 1 | 0.282 | Motif Batik Indonesia dengan dukungan Online dan berbagai jenis motif dan gambar batik dari pekalongan, batik ndonesia menjadi warisan budaya. |
| d4 | 2 | 0.179 | Batik Markets adalah toko batik online menjual kerajinan Batik Pekalongan – Indonesia dan menerima pemesanan seragam batik sekolah,perkantoran,organisasi |
| d5 | 4 | 0.117 | Batik indonesia sangat berkaitan erat dengan kerajaan Majapahit dan penyebaran agama Islam di Jawa. Oleh karena itu batik indonesia harus dilestarikan |

Tabel 5.1 Hasil pencarian pertama

Pada **pencarian kedua** menitikberatkan penggunaan *keyterm*, *keyterm* ditentukan dengan melihat posisi *query* pada dokumen, setelah itu ditentukan bobot dari setiap *keyterm* yang ditemukan, *keyterm* yang telah ditentukan bobotnya ini disebut *querytem.* Dari *queryterm* ini ditentukan nilai normalisasi yang akan digunakan pada hasil relevansi **pencarian pertama***.* Dari pembobotan tersebut diperoleh rangking dokumen berdasar nilai relevansi yang ditingkatkan.

Langkah – langkah yang dilakukan pada proses perangkingan **pencarian kedua** dilakukan dalam tahapan sebagai berikut:

1. Ekstraksi *keyterm* yang mendefinisikan dokumen

Dalam ekstraksi *keyterm* digunakan mekanisme *seeding and expansion*, yaitu menentukan suatu kata sebagai inti *keyterm* yang lalu dapat dikembangkan. Mekanisme ini terdiri dari dua fase, yaitu: **penempatan *seed*** dan **penentuan *keyterm***.

**Penempatan *seed*** adalah menentukan perkiraan posisi kata pada dokumen yang sama dengan *query* yang telah di lakukan tahapan prapemrosesan. Kata dalam dokumen yang sama dengan *query* ini disebut ***seed*.**

**Penentuan *keyterm*** menentukan rangkaian kata yang melingkupi *seed* yang membentuk *keyterm*. Penentuan *keyterm* dilakukan dengan beberapa ketentuan yaitu:

* + - 1. Sebuah *keyterm* harus terdiri dari paling sedikit satu *seed*
      2. Sebuah *keyterm* muncul paling sedikit L kali pada dokumen, dengan L adalah parameter jumlah dari sebuah *keyterm* paling sedikit muncul dan L > 1
      3. *Keyterm* tidak *m*asuk dalam suatu rangkaian kata yang memenuhi 1 dan 2

Dengan permisalan dari table 5.1, dengan *query* “batik Indonesia” dan parameter L=2 ditemukan *keyterm* sebagai berikut:

|  |  |  |
| --- | --- | --- |
| Keyterm | Dokumen | frekuensi |
| batik | d1;d3;d4 | 3 |
| indonesia | d1;d4 | 2 |
| batik indonesia | d2;d3;d5 | 3 |

Tabel 5.2 Hasil penentuan *keyterm*

1. Pembobotan *queryterm*

Dari *keyterm* yang didapat dilakukan pembobotan *queryterm* pembobotan ini di tentukan bebarapa faktor yaitu:

* + - 1. Distribusi relative: semakin sering suatu *keyterm* muncul pada hasil pencarian pertama relatif dengan seluruh koleksi, maka *keyterm* cendrung semakin penting.
      2. Term length: semakin panjang *querytem* maka kontribusi pada presisi semakin besar.
      3. Posisi rangking dokumen yang terdapat *queryterm*.

Pembobotan dari *queryterm* dilakukan dengan persamaan:

(2)

(3)

K : Jumlah dokumen hasil pencarian pertama

t : Query *keyterm*

di : Dokumen hasil pencarian dengan rangking *i*

C : koleksi dokumen

R : Jumlah dokumen pada seluruh koleksi C

df(t,C) : Jumlah dokumen yang mengandung *t* dalam C

df(t,di) : nilai normalisasi frekuensi *t* pada dokumen dalam hasil pencarian pertama

|t| : panjang dari *keyterm*

f(i) : pembobotan frekuensi berdasarkan rangking i

dari keyterm yang didapat pada table 5.2 di tentukan nilai normalisasi frekuensi dokumen (dengan i adalah rangking dokumen)

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| *Queryterm* | frekuensi | | | | | df(t,di) | | | | |
| d1 | d2 | d3 | d4 | d5 | d1 | d2 | d3 | d4 | d5 |
| batik | 3 | 0 | 1 | 4 | 0 | 0 | 1 | 0 | 0 | 1 |
| indonesia | 1 | 0 | 0 | 1 | 0 | 0 | 1 | 1 | 0 | 1 |
| batik indonesia | 0 | 1 | 2 | 0 | 2 | 1 | 0 | 0 | 1 | 0 |

Tabel 5.3 nilai normalisasi df(t,di)

Setelah itu dilakukan perhitungan bobot *queryte*rm, dengan:

A1 =

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| *Querytem* | A1 | K | df(t,C) | R |  | Wq |
| batik | 1,154 | 5 | 3 | 20 | 1 | 1,240 |
| indonesia | 1,731 | 5 | 2 | 20 | 1 | 1,861 |
| batik indonesia | 1,5 | 5 | 3 | 20 | 1,414 | 2 |

Tabel 5.4 Bobot dari *queryterm*

1. Perangkingan dokumen

Untuk menghindari suatu *queryterm* yang memiliki makna sama atau membahas suatu aspek yang sama, di tentukan korelasi *keyterm* untuk mengukur keberhubungan suatu set *queryterm.* Suatu set *queryterm* berkorelasi jika kebersamaan kemunculannya dalam top K (dengan K adalah jumlah dokumen ditemu-kembali hasi pencarian awal) relatif tinggi. Dengan kata lain jika suatu set *queryterm* membahas konsep yang independen maka kebersamaan kemunculannya rendah. Korelasi *queryterm* di definisikan sebagai berikut:

(3)

S : Jumlah dokumen hasil pencarian pertama

ti & tj : *queryterm*

Dengan berdasarkan table 5.2 ditentukan nilai korelasi untuk setiap term, dengan:

B1 = jumlah dokumen yang mengandung ti dan tj

B2 = jumlah dokumen yang mengandung tj

|  |  |  |  |
| --- | --- | --- | --- |
| ti|tj | B1 | B2 | P(ti|tj) |
| batik | indonesia | 2 | 2 | 1 |
| batik | batik indonesia | 1 | 3 | 0.33 |
| indonesia | batik indonesia | 0 | 3 | 0 |
| indonesia | batik | 2 | 3 | 0.66 |
| batik indonesia | batik | 1 | 3 | 0.33 |
| batik indonesia | indonesia | 0 | 2 | 0 |

Tabel 5.5 Nilai korelasi antar term

Untuk perangkingan dokumen hasil pencarian dilakukan perhitungan nilai penyesuaian untuk relevansi hasil pencarian pertama yang menitik beratkan pada *queryterm*. Pembobotan dokumen ini dilakukan dengan mencari *queryterm* yang muncul pada suatu dokumen, *queryterm* ini di urutkan secara menurun berdasarkan bobotnya. *Quertyterm* pertama mendapat bobotnya secara penuh, lalu untuk *queryterm* selanjutnya bila suatu *queryterm* memiliki korelasi tinggi maka bobotnya terkurangi secara proporsional. Berdasarkan hal-hal tersebut definisi pembobotan baru adalah sebagai berikut:

(4)

w(t1) : Bobot *queryterm* terbesar

w(ti) : Bobot *queryterm* ke-i

P(*ti* | *tj*) : Nilai korelasi *keyterm* ke-i

Dengan berdasarkan table 5.5 maka nilai penyesuai relevansi baru, dengan:

C1 = t1 min(1-P(ti|t1))

C2 = t2 min(1-P(ti|t2))

C3 = t3 min(1-P(ti|t3))

C4 = C1+C2+C3

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Dokumen | C1 | C2 | C3 | C4 | Weightnew |
| d1 | 0.413 | 0 | 0 | 0.413 | 2.274 |
| d2 | 0 | 0 | 1.333 | 1.333 | 3.333 |
| d3 | 0.413 | 0 | 1.333 | 1.746 | 3.746 |
| d4 | 0.413 | 0 | 0 | 0.413 | 2.274 |
| d5 | 0 | 0 | 1.333 | 1.333 | 3.333 |

Tabel 5.6 nilai penyesuai relevansi

Dari bobot dokumen baru dengan nilai similaritas awal dokumen ditentukan skor rangking baru. Dari dari skor baru ini di tentukan urutan baru dari M dokumen. Skor baru ini ditentukan dengan:

(5)

Weightnew : bobot baru berdasarkan (4)

Scoreold : nilai similaritas pencarian awal, variabel ini di peroleh dari pencarian pertama

Dari table 5.6 dan 5.1 ditentukan nilai relevansi pencarian yang ditekankan pada pembobotan keyterm, dengan:

R1 = rangking pencarian pertama

R2 = rangking pencarian kedua

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Dokumen | Weightnew | Scoreold | Scorenew | R2 | R1 |
| d1 | 3,234 | 0,282 | 1,194 | 1 | 3 |
| d2 | 4,892 | 0,179 | 1,058 | 2 | 5 |
| d3 | 5,584 | 0,121 | 0,798 | 3 | 1 |
| d4 | 3,234 | 0,117 | 0,499 | 5 | 2 |
| d5 | 4,892 | 0,092 | 0,546 | 4 | 4 |

Berdasarkan langkah langkah tersebut berikut adalah gambaran alur tugas akhir ini.



1. **METODE PENELITIAN**

Pengerjaan Tugas Akhir ini terdiri atas beberapa tahapan, yaitu :

* 1. **Studi Literatur**

Studi literatur dilakukan untuk mencari pemahaman mengenai hal-hal yang berkaitan dengan proses perangkingan pada dokumen berbahasa arab. Terdapat beberapa hal yang perlu dipahami pada proses ini, antara lain:

*Prapemrosesan* dan *pencarian*

Pada tahap ini perlu pemahaman terhadap *prapemrosesan* dan *pencarian* yang dilakukan pada *lucene library* dan [3].

Ekstraksi kata

Pada tahap ini dicari pemahaman pada penentuan *keyterm* yang mendefinisikan dokumen.

Pembobotan query berdasarkan distribusi kata

Pada tahap ini dicari pemahaman pembentukan bobot *queryterm* yaitu *keyterm* yang ditemukan dalam *query*.

Perangkingan hasil pencarian

Pada tahap ini dicari pemahaman bagaimana penentuan korelasi *queryterm*, pembobotan baru, dan skor dokumen berdasarkan hasil pembobotan baru. Lalu dari hasil penentuan tersebut dilakukan perangkingan dokumen

* 1. **Perancangan Algoritma Perangkingan Dokumen**

Berdasarkan kajian pustaka prosedur dari pengurutan kembali dokumen adalah sebagai berikut:

Dengan *q* sebagai *query*, *K* adalah jumlah hasil pencarian pertama,

* + - 1. Ekstrak *keyterm*  yang muncul pada *q* dari setiap dokumen hasil pencarian pertama
      2. Menghitung bobot *queryterm* dari *keyterm*
      3. Pengurutan kembali top M dokumen
  1. Memperoleh korelasi *queryterm* dari hasil pencarian pertama; menentukan korelas*i keyterm*
  2. Untuk setiap dokumen d pada M dokumen, menghitung nilai rangking skor yang baru
  3. Mengurutkan top M dokumen berdasar skor
  4. **Implementasi Sistem**

Pada tahap pengimplementasian ini aplikasi sudah mulai dibuat secara menyeluruh berdasar pada desain sistem dan tampilan perangkat lunak. Bahasa yang digunakan untuk membangun aplikasi ini adalah JAVA dengan menggunakan GUI JAVA dan database yang digunakan adalah MySQL. Pada proses pembentukan index dan pencarian menggabungkan *lucene library* dan [3]. Aplikasi ini dibangun sebagai aplikasi berbasis desktop*.*

*Lucene library* yang digunakan pada proses *pembentukan index* dan *pencarian* adalah library pencarian *open source* yang di implementasikan dalam java. Library ini memiliki terdiri dari fungsi-fungsi java yang biasa digunakan untuk proses pencarian. Library ini banyak dipergunakan pada aplikasi-aplikasi besar, seperti: Wikipedia, IDE eclipse, dll.

Dokumen-dokumen bahasa arab di masukkan pada vector space model, dari prapemrosesan ini dilakukan pencarian. Proses-proses ini dikembangangak dari *lucene library* dan [3]. Dari hasil pencarian ini dilakukan proses perangkingan berdasar metode yang dibahas dalam kajian pustaka. Dari proses perangkingan ini dihasilkan perangkingan baru hasil dokumen.

* 1. **Uji Coba dan Analisa**

Dari proses-proses ini akan dilakukan uji coba dan analisa sebagai berikut:

Analisa perubahan parameter L (jumlah minimal suatu *keyterm* muncul dalam dokumen) pada tahapan *keyterm* extraction pada presisi hasil

Perbandingan antara hasil pencarian pertama dengan hasil pencarian kedua

* 1. **Penyusunan Buku Tugas Akhir**

Pada tahap ini dilakukan untuk pembuatan laporan dari semua dasar teori dan metode yang digunakan serta hasil – hasil yang diperoleh selama pengerjaan tugas akhir. Laporan tugas akhir ini akan dibagi menjadi beberapa bab sebagai berikut:

1. Bab I, Pendahuluan, berisi latar belakang, permasalahan, tujuan, batasan permasalahan, metodologi, dan sistematika penulisan.
2. Bab II, Landasan Teori, akan dibahas dasar ilmu yang mendukung pembahasan tugas akhir ini.
3. Bab III, Desain Aplikasi.
4. Bab IV, Implementasi dari aplikasi yang telah dibuat, akan dilakukan pembuatan aplikasi yang dibangun dengan komponen-komponen yang telah ada yang sesuai dengan permasalahan dan batasannya yang telah dijabarkan pada bab pertama.
5. Bab V, Uji coba dan analisa hasil, akan dilakukan uji coba berdasarkan parameter-parameter yang ditetapkan, dan kemudian dilakukan analisa terhadap hasil uji coba tersebut.
6. Bab VI, Penutup, berisi kesimpulan yang dapat diambil dari Tugas Akhir ini beserta saran untuk pengembangan selanjutnya.
   1. **JADWAL KEGIATAN**

Tugas akhir ini diharapkan bisa dikerjakan menurut jadwal sebagai berikut:

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| No | Tahapan | Bulan | | | | | | | | | | | | | | | |
| Maret | | | | April | | | | Mei | | | | Juni | | | |
| 1 | Studi Literatur |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  | * 1. Ekstraksi Kata |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  | * 1. query berdasarkan distribusi kata |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  | * 1. Reranking hasil pencarian |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  | * 1. Evaluasi Hasil |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 2 | Perancangan Perangkat Lunak |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  | a. Perancangan Sistem |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  | b. Perancangan Antar Muka Pengguna |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 3 | Implementasi Perangkat Lunak |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 4 | Uji Coba dan Evaluasi |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 5 | Penyusunan Buku Tugas Akhir |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |

1. **DAFTAR PUSTAKA**
   * + 1. L. Yang, D. Ji and M. Leong (2007), Dokumen reranking by *keyterm* distribution and maximal marginal relevance for Chinese information retrieval. Information Processing and Management: An International Journal 43 (2), pp. 315–326
       2. M. Mitra, A. Singhal, & C. Buckley (1998), Improving Automatic Query Expansion. In Proceedings of ACM SIGIR’98.
       3. A. Najibullah (2010), Implementasi N-gram dalam Pencarian Teks Sebagai Penunjang Aplikasi Perpustakaan Kitab Berbahasa Arab. Tugas Akhir Institut Teknologi Sepuluh Nopember, FTIF Teknik Informatika.
       4. S. R. El-Beltagy, A. Rafea (2009), KP-Miner: A keyphrase extraction system for english and arabic dokumens. Infromation Systems 34 (2009) pp. 132-144.

LEMBAR PENGESAHAN

Surabaya, April 2010

Mengetahui/Menyetujui

# Dosen pembimbing I

**Diana Purwitasari, S.Kom., M.Sc.**

NIP. 197804102003122001

# Dosen pembimbing II

**Dr. Agus Zainal Arifin, S.Kom., M.Kom.**

NIP. 197208091995121001