**JURUSAN TEKNIK INFORMATIKA**

**FAKULTAS TEKNOLOGI INFORMASI**

**INSTITUT TEKNOLOGI SEPULUH NOPEMBER**

**USULAN TUGAS AKHIR**

# **IDENTITAS PENGUSUL**

Nama :**Arlisa Yuliawati**

NRP :**5107 100 090**

Dosen Wali :Sarwosri, S.Kom, M.T

1. **JUDUL TUGAS AKHIR**

***Implementasi Peringkasan Otomatis pada Dokumen Terstruktur dengan Metode Faktorisasi Matriks Nonnegatif***

1. **LATAR BELAKANG**

*Summary* atau ringkasan adalah sebuah teks yang dihasilkan dari satu atau lebih teks, yang mengandung informasi dari teks sumber dan panjangnya tidak lebih dari setengah teks sumber. Peringkasan teks otomatis (*Automatic Text Summarization* atau *ATS*) adalah pembuatan ringkasan dari sebuah teks secara otomatis dengan memanfaatkan aplikasi yang dijalankan pada komputer.

*Portable Document Format* (PDF) merupakan suatu dokumen yang dapat digunakan untuk merepresentasikan dokumen dua dimensi yang meliputi teks, huruf, citra dan grafik vektor dua dimensi. Jenis dokumen ini memiliki 4 bagian dalam strukturnya, yaitu *header*, *body area/object*, *cross reference table*, dan *trailer*. Dimana bagian isi dari suatu dokumen PDF pada umumnya terletak pada bagian *body area*.

Metode peringkasan dokumen menggunakan LSA (*Latent Semantic Analysis*) menghasilkan ringkasan dokumen yang kurang bermakna karena menggunakan vektor tunggal sebagai dasar seleksi kalimat dan komponen vektor tunggal tersebut bisa memiliki nilai negatif. Oleh karena itu diperkenalkan metode baru dengan menggunakan NMF (*Non-negative Matrix Factorization*) dalam menghasilkan ringkasan dokumen yang generic, yang menggunakan batasan nonnegatif dalam komponen vektornya sehingga menghasilkan ringkasan dokumen yang lebih bermakna karena lebih mirip dengan proses kognisi yang dilakukan manusia.

1. **RUMUSAN MASALAH**

Rumusan masalah yang diangkat dalam Tugas Akhir ini dapat dipaparkan sebagai berikut:

1. Bagaimana mengenali bagian-bagian dalam suatu dokumen PDF yang akan diringkas.
2. Bagaimana implementasi metode Non-negative Factorization dalam menghasilkan ringkasan dokumen yang lebih bermakna.
3. Bagaimana mendapatkan topik penting dari suatu dokumen untuk dijadikan ringkasan dokumen yang sifatnya generik.
4. **BATASAN MASALAH**

Dalam penerapan metode NMF untuk peringkasan dokumen secara generik ini, terdapat beberapa batasan dalam proses pengerjaannya. Peringkasan dokumen ini dikemas dalam sebuah aplikasi *dekstop* yang mengambil masukan berupa dokumen PDF yang telah diubah ke dalam bentuk dokumen biner. Jenis dokumen PDF yang digunakan adalah dokumen terstruktur yang memiliki struktur hirarki di dalamnya, sebagai contoh adalah dokumen jurnal hasil riset. Pemilihan dokumennya sendiri dibatasi pada dokumen dengan struktur paragraf yang terdiri atas satu kolom saja dan menggunakan bahasa inggris. Peringkasan atas dokumen tersebut dilakukan hanya pada bagian *Introduction*  dan *Conclusion* dengan menggunakan metode *Non-negative matrix Factorization* (NMF)*.*

1. **TUJUAN TUGAS AKHIR**

Sistem peringkasan dokumen otomatis pada dokumen PDF ini dibangun untuk mempermudah pemahaman garis besar isi dari suatu dokumen jurnal. Dengan mengambil topik-topik penting dari dalam dokumen.

1. **MANFAAT TUGAS AKHIR**

Dengan adanya sistem peringkasan dokumen secara otomatis untuk dokumen jurnal PDF ini diharapkan pengguna dapat menghemat waktu dalam proses membaca suatu dokumen jurnal yang memiliki banyak bagian. Dimana pada umumnya inti suatu dokumen jurnal dapat diambil dari bagian awal dan pada bagian kesimpulannya.

1. **RINGKASAN TUGAS AKHIR**

Pada Tugas Akhir ini penulis mengusulkan untuk membuat suatu aplikasi untuk melakukan peringkasan dokumen secara otomatis. Peringkasan otomatis yang dimaksud adalah ketika pengguna memasukkan dokumen atau artikel bertipe PDF, maka sistem akan secara otomatis membuatkan ringkasan dari artikel atau dokumen tersebut. Panjang artikel yang ditampilkan adalah kurang lebih 20% dari panjang artikel aslinya.

Pengguna dapat memasukan dokumen mana yang ingin diketahui ringkasannya, kemudian sistem akan mengolah dokumen tersebut untuk kemudian dikembalikan berupa keluaran ringkasan dari artikel dari dokumen yang dimasukkan oleh pengguna.

Secara umum alur dari sistem serta proses peringkasan dokumen otomatis yang terjadi di dalam sistem sendiri dapat dijelaskan pada diagram berikut ini.

|  |  |
| --- | --- |
| Pengguna | Sistem |
|  |  |

Penjelasan Gambar:

1. Tahap Pra-pemrosesan
2. Pemecahan satu dokumen menjadi kalimat-kalimat.

Pada langkah ini, dokumen yang telah terbaca oleh sistem akan dipecah-pecah menjadi kalimat-kalimat yang menyusunnya.

1. Penghilangan *Stopwords*

Dari kalimat-kaliat individu yang telah terdata, kemudian kata-kata yang dianggap tidak penting dihilangkan. Penghilangan stopwords ini menggunakan Rijsbergen’s stopword list.

1. Penerapan proses *Stemming*

Pengambilan kata dasar dari tiap kata/*term* dalam kalimat-kalimat yang telah terdata, menggunakan metode Porter’s stemming.

1. Pembentukan representasi dokumen dalam sebuah matriks

Matriks dibangun berdasarkan *terms-by-sentences (mxn)* yang ada pada dokumen. Matriks ini berisi *term-frequency* dari *term j* pada kalimat ke-*i*.

Dimana Wgt(j,i) merupakan bobot *term j* pada kalimat ke-*i*.

* No Weight *(recall evaluation*)

* Binary Weight (*precission evaluation*)
* Modified Binary Weight (*F-measure evaluation*)

1. Tahap Ekstraksi Kalimat
2. Penerapan NMF (*Nonnegative Matrix Factorization*)

NMF diterapkan pada matriks A untuk memperoleh nilai matriks W (*Non-negative Semantic Feature*), sedangkan H (*Non-negative Sementic Variable*) diperoleh dengan persamaan berikut:

Dimana nilai W dan H terus diperbaharui hingga sampai batas treshold tertentu, dengan cara:

1. Penghitungan GRS (*Generic Relevances of Sentences*)

GRS ini dipergunakan untuk memberi semacam bobot tertentu untuk setiap kalimat yang telah terrepresentasi dalam matriks.

*Generic Relevance* dari kalimat ke-j =

Dengan

1. Ekstraksi kalimat-kalimat yang penting

Dalam tahap ekstraksi kalimat yang penting ini dilakukan dengan cara memilih sebanyak *k* kalimat yang memiliki *Generic Relevance Sentrence* tertinggi. Dimana nilai *k*  merupakan jumlah kalimat untuk *generic document summarization*.

1. **METODOLOGI**

Perencanaan tugas akhir ini terdiri dari beberapa tahap pengerjaan, yaitu :

1. Penyusunan Proposal Tugas Akhir

Tahap awal untuk memulai pengerjaan Tugas Akhir adalah penyusunan Proposal Tugas Akhir. Pada proposal ini, penulis mengajukan gagasan pembuatan peringkas dokumen otomatis dengan masukan berupa dokumen ber-ekstensi .pdf dan menghasilkan keluaran berupa ringkasan dokumen.

1. Studi Literatur

Pada tahapan ini akan dilakukan studi literatur mengenai metode yang digunakan, diantaranya :

* + - * 1. Berbagai pustaka mengenai rumusan masalah (*paper* mengenai peringkasan dokumen otomatis).
        2. Cara pembacaan isi dokumen PDF dalam platform java.
        3. Fungsi *Frobenius norm.*
        4. Metode *Nonnegative Matrix Factorization* (NMF).

1. Implementasi Perangkat Lunak

Implementasi merupakan tahap untuk membangun sistem. Implementasi terdiri atas perancangan dan pembuatan sistem penringakasan dokumen otomatis dengan menggunakan metode Faktorisasi Matriks Nonnegatif.

1. Pengujian dan Evaluasi

Pada tahap ini dilakukan uji coba terhadap sistem yang telah dibuat, mengamati kinerja sistem yang baru dibuat, serta mengidentifikasi kendala yang mungkin timbul. Pengujian dilakukan dengan mencoba memasukkan dokumen-dokumen ke dalam sistem. Evaluasi dianggap berhasil jika ringkasan dokumen yang dihasilkan sudah sesuai yang diharapkan, yaitu dapat menangkap 80% maksud dari artikel.

1. Penyusunan Buku Tugas Akhir

Tahap terakhir merupakan penyusunan laporan yang memuat dokumentasi mengenai pembuatan serta hasil dari implementasi perancangan dan *prototyping* yang telah dibuat. Secara garis besar, buku laporan tugas akhir ini terdiri atas beberapa bagian yaitu:

1. Pendahuluan
   1. Latar Belakang
   2. Permasalahan
   3. Batasan Tugas Akhir
   4. Tujuan
   5. Metodologi
   6. Sistematika Penulisan
2. Tinjauan Pustaka
3. Desain dan Implementasi
4. Uji Coba dan Evaluasi
5. Kesimpulan dan Saran
6. Daftar Pustaka
7. **JADWAL KEGIATAN TUGAS AKHIR**

Tugas akhir ini diharapkan bisa dikerjakan menurut jadwal sebagai berikut:

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| No. | Kegiatan | Bulan (2010) | | | | | | | | | | | |
| Okt ‘10 | | Nov ‘10 | | Des ‘10 | | Jan ‘11 | | Feb ‘11 | | Mar ‘11 | |
| 1. | Penyusunan Proposal Tugas Akhir |  | |  |  |  |  |  |  |  |  |  |  |
| 2. | Studi Literatur |  |  |  | |  |  |  |  |  | |  |  |
| 3. | Implementasi Perangkat Lunak |  | |  |  |  |  |  |  |  |  |  |  |
| 4. | Pengujian dan Evaluasi |  | |  |  |  |  |  |  |  |  |  |  |
| 5. | Penyusunan Buku Tugas Akhir |  | |  |  |  |  |  |  |  |  |  |  |

1. **DAFTAR PUSTAKA**
2. Internet. ***Automatic Text Summarization***. Diakses tanggal 10 Oktober 2010 dari <http://www.ittelkom.ac.id/library/index.php?option=com_content&view=article&id=582:automatic-text-summarization&catid=20:informatika&Itemid=15>
3. \_\_\_\_\_\_. ***Quick Post: About the Physical and Logical Structure of PDF Files.*** Diakses tanggal 25 Oktober 2010 dari <http://blog.didierstevens.com/2008/04/09/quickpost-about-the-physical-and-logical-structure-of-pdf-files/>
4. \_\_\_\_\_\_. ***The PDF File Format***. Diakses tanggal 25 Oktober 2010 dari <http://www.prepressure.com/pdf/basics/fileformat>
5. \_\_\_\_\_\_. ***PDFBox: Extract Text From PDF***. Diakses tanggal 26 Oktober 2010 dari <http://thottingal.in/blog/2009/06/24/pdfbox-extract-text-from-pdf/>
6. \_\_\_\_\_\_. ***Portable Document Format***. Diakses tanggal 28 Oktober 2010 dari <http://id.wikipedia.org/wiki/Portable_Document_Format>
7. Ju-Hong, L., Park Sun, Chan-Min, A., dan Kim, Daeho. 2009. ***Automatic Generic Document Summarization based on Non-negative Matrix Factorization.*** Jurnal Information Processing & Management pp 20-34

**LEMBAR PENGESAHAN**

###### **Surabaya, 29 Oktober 2010**

Menyetujui,

Dosen Pembimbing I

Diana Purwitasari, S.Kom, M.Sc

NIP : 19780410 200312 2 001

Dosen Pembimbing II

Umi Laily Yuhana, S.Kom, M.Sc

NIP : 19790626 200501 2 002