

PENCARIAN DOKUMEN TEKS ARSIP SURAT DENGAN METODE INDEXING DAN QUERY

Heriyanto

Jurusan Teknik Informatika UPN "Veteran" Yogyakarta
Jl. Babarsari no 2 Tambakbayan 55281 Yogyakarta Telp (0274)-485323
e-mail : heriyanto.yanto@ugm.ac.id, mr_heriyanto_skom@yahoo.com

Abstrak

Pengarsipan surat menyurat di dalam aktifitas baik dalam arsip surat sendiri maupun dalam lingkungan kantor sangatlah penting. Pada saat-saat tertentu terkendala pada dimana suatu saat surat yang ingin dicari yang ada tulisan kata-kata tertentu atau kalimat ada pada surat-surat apa saja, kapan surat itu dibuat dan di file yang mana berkas tersebut pernah di simpan. Berkas surat yang sedikit belum menjadi masalah namun setelah data surat yang sudah dibuat bertumpuk-tumpuk bahkan sampai bertahun-tahun dengan jumlah yang banyak menjadi bingung. Permasalahan tersebut maka dilakukan dengan proses penemuan kembali dengan indeks agar pada saat-saat tertentu data surat dapat ditemukan dengan baik. Metode indeks prinsipnya mengambil indeks yang ada pada buku. Buku yang tebal mempunyai indeks langsung dapat menemukan di dalam buku tersebut ada kata atau kalimat apa saya dapat ditemukan. Pada pengarsipan surat maka dapat dilakukan indeks yang sudah tersimpan dengan penyimpanan, dan dapat namun terlebih dahulu harus diolah ketahapan awal pemecahan kata dalam suatu teks kalimat tersebut dilakukan indeks dan sekaligus tersimpan dalam database dengan tabel penyimpanan.

Kata kunci : Pengarsipan, Indeks, Teks, Surat, Berkas.

1. PENDAHULUAN

Pada saat ini berkas semakin hari semakin banyak dan menumpuk beriringan dengan data-data terutama surat atau pengarsipan. Penyimpanan file sudah dilakukan dengan baik tetapi yang menjadi kendala adalah file tersebut berisi apa saja. Berawal dari kesulitan bagaimana menemukan kata-kata yang apa saja yang tersimpan di dalam file tersebut maka perlu dilakukan pengolahan berkas teks. Berkas teks yang berisi arsip-arsip surat perlu dilakukan pengolahan yang dapat diindeks agar memudahkan di dalam proses pencarian atau penemuan kembali berkas-berkas yang kita inginkan.

2. TINJAUAN PUSTAKA

2.1 Indexing Retrieval

Teknik yang dipakai untuk dokumen teks banyak digunakan dengan teknik IR (*Indexing* dan *Retrieval*). Teknik IR sangatlah penting di dalam informasi multimedia manajemen sistem untuk dua alasan yaitu :

1. Keberadaannya, sebagian besar dokumen teks banyak digunakan di dalam organisasi seperti perpustakaan.
Teks adalah sangat penting sumber informasi untuk organisasi. Untuk pengefisienan penggunaan penyimpanan informasi dalam naskah sangat diperlukan sistem IR.
2. Teks dapat digunakan untuk kebutuhan media lain seperti Audio, Gambar dan Video. (menurut Buku Multimedia Database Management System, Guojun Lu Hal 73) Ada 4 penggunaan komponen model IR (*Indexing Retrieval*) yaitu : 1) *Exact Match*, 2) *Vector Space*, 3) *Probabilistic*, 4) *Cluster-based*

2.2 IR (Indexing Retrieval)

IR menurut Lu (1999;75-76) melihat bahwa tujuan dari suatu sistem IR adalah untuk mendapat kembali materi relevan dari suatu database dokumen sebagai jawaban atas *query* pemakai. Kebanyakan dari sistem IR yang komersil saat sekarang dapat digolongkan pada sistem IR boolean atau sistem pencarian *text-pattern*. *Query* pencarian *text-pattern* adalah *string* atau ungkapan *reguler*.

3. METODE PENELITIAN

Metodologi yang akan digunakan pada penelitian ini antara lain ; Studi Literatur dan SDLC (*System Development Life Cycle*) yang meliputi tahap *Analysis, Design, Implementation, Testing* dan *Maintenance*, (Pressman, 2002). Metodologi ini dirinci sebagai berikut :

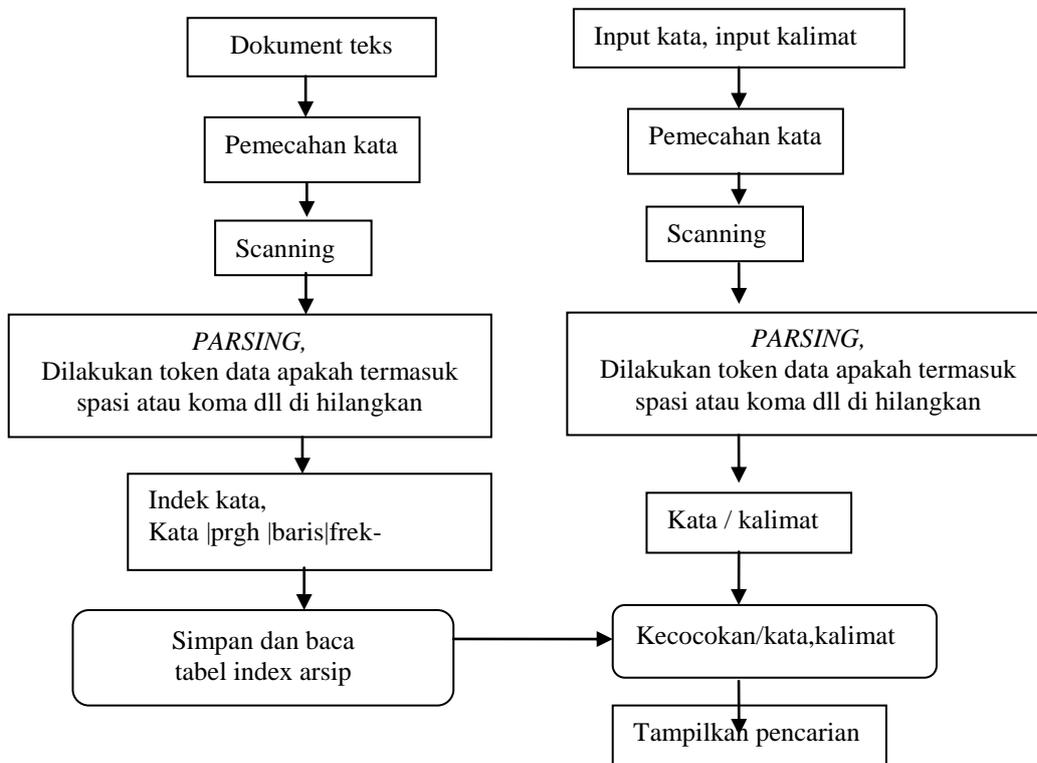
1. Studi Literatur
Tahap ini merupakan tahap pengumpulan informasi dan literatur yang diperlukan untuk pembuatan sistem. Adapun informasi dan literatur yang dipergunakan diantaranya diagnosis penyakit gigi dan sistem pakar.
2. Analisis dan perancangan
Pada tahap ini dilakukan analisis serta desain yang diperlukan dalam membuat sistem, diantaranya akuisisi pengetahuan, representasi pengetahuan, mekanisme inferensi, perancangan DFD, perancangan basisdata dan perancangan user interface
3. Implementasi
Pada tahap ini, rancangan sistem yang telah dibuat akan diimplementasikan menggunakan Borland Delphi 6 dan *Microsoft Access* serta program pendukung.
4. Uji coba dan evaluasi

Pada tahap ini, akan dilakukan uji coba dan evaluasi terhadap sistem serta akan dilakukan perbaikan-perbaikan yang diperlukan.

4. HASIL DAN PEMBAHASAN

Berkas surat yang sedikit kemudian lama-lama menjadi sangat banyak dan menjadi rumit manakala mencari tema atau kalimat atau kata yang pernah dibuat terdapat di file, paragraf dan baris ke berapa surat surat yang sudah begitu banyak. Permasalahan tersebut maka dilakukan dengan proses penemuan kembali dengan indeks agar pada saat-saat tertentu data surat dapat ditemukan dengan baik dengan langkah dokumen teks atau surat tersebut dapat dianalisa sebagai berikut : 1) Data teks terdiri dari beberapa kata, 2) Data kata-kata ada di beberapa paragraph 3) Data paragraph terdiri atas beberapa baris, 4) Data baris pada kata terdiri atas kemunculan / jumlah kemunculan atau frekwensi, 5) Data kata, paragraf, baris dan frekwensi disimpan dalam tabel

Model perancangan dapat dilihat pada gambar 1 sebagai berikut :



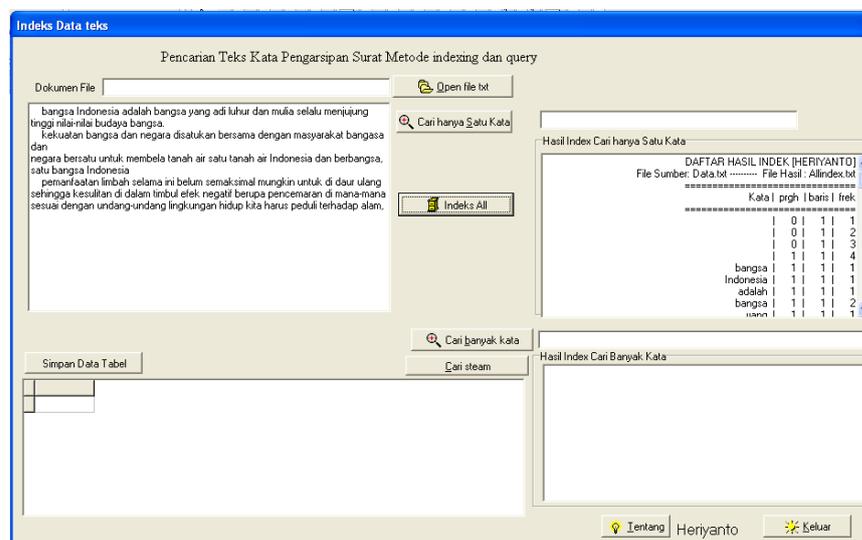
Gambar 1. Pencarian dokumen teks arsip surat

Setelah proses melakukan scanning, parsing, token kemudian dilakukan indek dan hasil indek disimpan dalam tabel data indek arsip seperti pada tabel 1 di bawah ini:

Tabel 1. Data tabel indek arsip

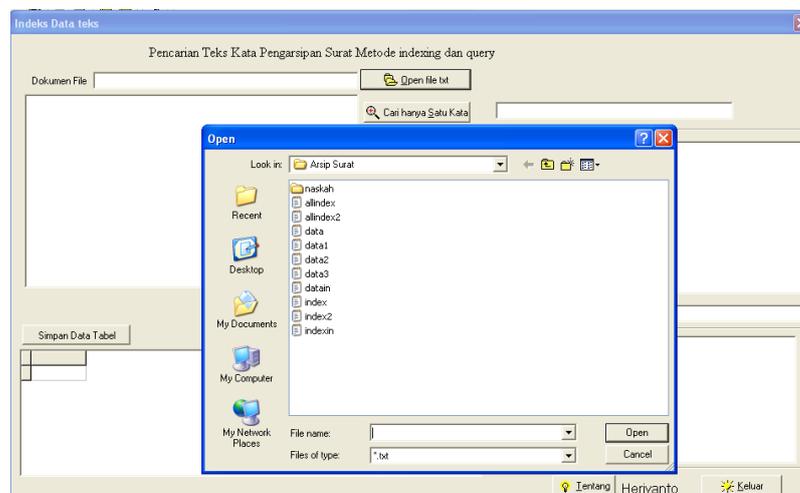
Name Fields	Type Data	Size	Constraint
Nourut	Number	10	Primary Key
NamaFile	Text	50	
Kata	Number	10	
Paragraph	Number	10	
Baris	Number	10	
Frekwensi	Number	10	

Adapun hasil desain dan rancangan terdiri dari : input data teks surat, dengan input data kalimat atau kata yang dicari seperti pada gambar 1 menu tampilan program rancangan dibawah ini :



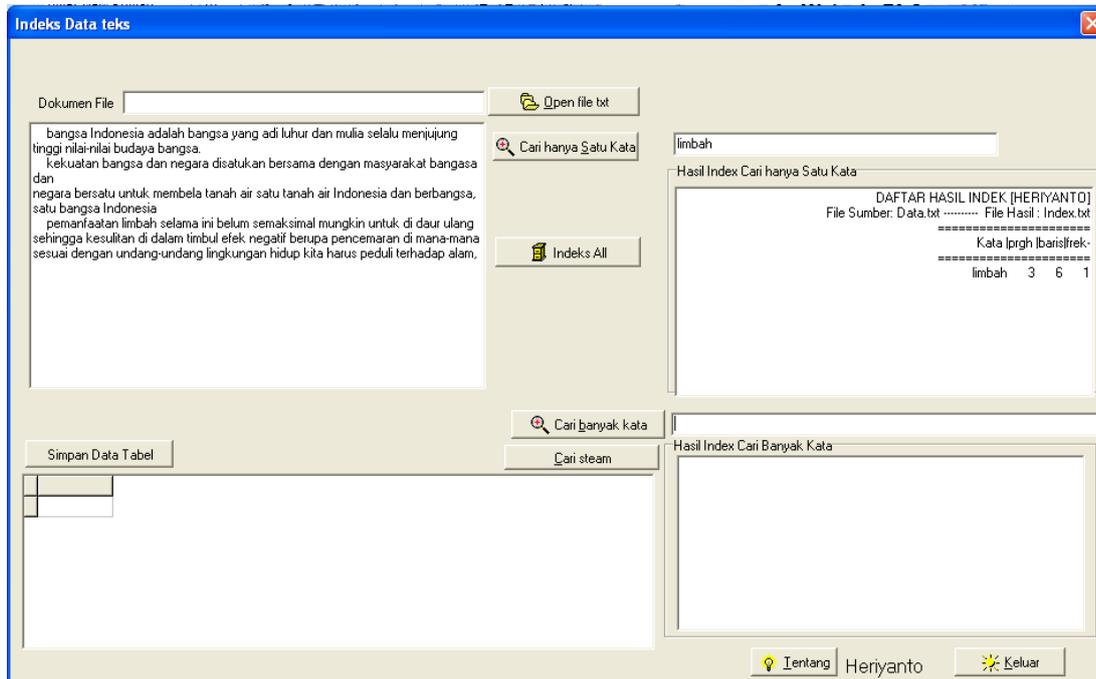
Gambar 1. Menu tampilan data pengambilan data surat dan indek

Data teks dari pengolah kata berupa surat dimasukkan dalam dokumen file .txt dan dibuka seperti pada gambar 2 dibawah ini:



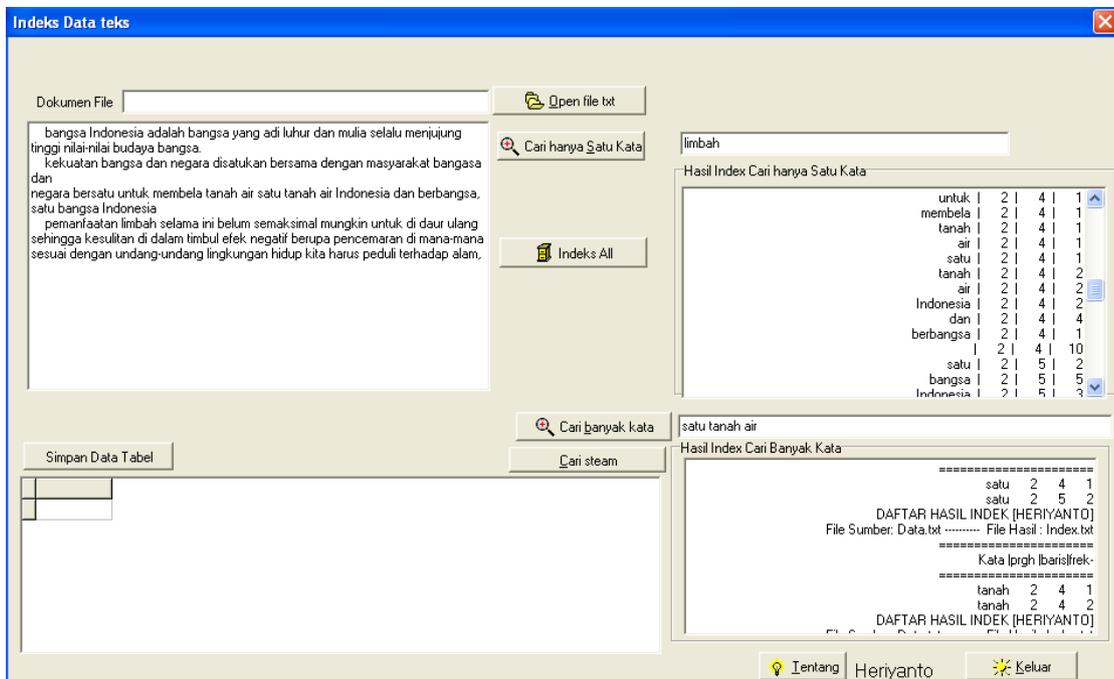
Gambar 2. Membuka file teks

Hasil pengambilan dokumen text dilakukan ke tahap indeks yang selanjutnya dapat disimpan di dalam tabel penyimpanan dan dilakukan pencarian perkata seperti pada gambar 3 dibawah ini:



Gambar 3. Pencarian indeks perkata

Proses pencarian selain perkata juga dapat dilakukan pencarian berupa kalimat seperti pada gambar 4 di bawah ini:



Gambar 4. Pencarian indeks per kalimat

Tabel 2 source code baca pemotongan berkas teks

<pre>while not.eof(berkasText))do begin jumbaris:=jumbaris+1; readln(berkastext,baris); bacaambilperkata2(baris,kata); end; {tutup file} closefile(berkasText); {menyimpan indek pada data.txt} jumbaris:=0; jumbaris3:=0; jumbaris1:=0; brs:=0; {brsa[1]:=0;} prg:=0; jumk:=0; kataada:=0; hit:=0; m:=0; {menghubungkan dengan external file} Assignfile(berkasText,OpenDialog1.filename); {buka file} reset(berkasText);</pre>	<pre>while not.eof(berkasText))do begin jumbaris1:=jumbaris1+1; readln(berkastext,baris); ambilperkata2(baris,kata,crk7); end; {tutup file} closefile(berkasText); { writeln('jumlah data record=',jumk); readln;} {i:=0;} {menampilkan pada richedit2} Assignfile(berkastext,'index2.txt'); reset(berkastext); while not (eof(berkastext)) do begin readln(berkastext,baris); RichEdit3.Lines.Add(baris); end; closefile(berkastext);</pre>
---	--

5. KESIMPULAN

Secara garis besar sistem yang dibangun untuk mengetahui kata atau kalimat dalam suatu surat yang dalam bentuk teks dapat di dilakukan proses index kata tersebut atau kalimat tersebut terdapat pada paragraf, baris, frekwensi kemunculan dan dalam file apa tersimpan di dalam database.

Hasil penyimpanan tabel disimpan secara permanen untuk dilakukan pengolahan dengan query apabila dicari data kata yang telah terindeks sesuai dengan keinginan. Pengembangan query dapat dilakukan pengolahan yang lebih lanjut dapat dilakukan pengelompokkan berdasarkan kata sehingga dapat terrangkum menjadi simpel.

DAFTAR PUSTAKA

- Candra, Ian , *Utility Komputer Multimedia*, 1999, Elex Media Komputindo, Jakarta
Heriyanto, *Identifikasi Naskah Dokumen Teks dengan Metode Indexing*, Jurnal Telematika, UPN "Veteran" Yogyakarta, 2012
Lu, Guajun, *Multimedia Database Manajemen Systems*, 1999 Artech House, Inc
Martina, Inge, *36 Jam Belajar Komputer Delphi 5.0 Database Client/Server Menggunakan Delphi*, 2000, PT. Elex Media Komputindo, Jakarta
Sanjaya Dwi, *Bertualang dengan Struktur Data di Planet Pascal*, edisi Pertama 2001, J& J Learning Yogyakarta
Silberschatz, Korth, Sudarsan, *Databases Systems Concept*, 4th ed, 2002, McGrawHill