PEMANFAATAN BERKAS *STOPWORD* PADA WINISIS

B. Mustafa
mus@ipb.ac.id atau b.mustafa@eudoramail.com

Berkas ***STOPWORD*** atau **Kata-abai** dalam sistem penelusuran informasi adalah suatu daftar kata atau istilah tertentu yang disusun sedemikian rupa sehingga pada saat proses penelusuran informasi, semua kata-kata atau istilah yang terdaftar dalam berkas tersebut diabaikan oleh sistem pencari *(search enginee)*. Manfaat dari berkas kata-abai ini adalah mempercepat proses pencarian informasi dengan cara mengabaikan atau melewatkan atau tidak mencari kata/istilah yang dianggap tidak mempunyai makna dalam pencarian informasi.

Pada umumnya kata-kata atau istilah yang didaftar dalam berkas **kata-abai** ini terdiri atas kata-kata depan (atau artikel misalnya di, ke, dari, diatas, dibawah, *a*, *an*, *at*, *from*, *for*, *on*, *of*, *off*, *with*), atau kata sambung/penghubung (misalnya karena, sebab, walaupun, tetapi, *because*, *although*, *but*). Kata-kata atau istilah ini, dalam konteks penelusuran informasi suatu subjek tertentu, tidak berkaitan langsung maknanya terhadap isi (*content words*), tetapi hanya sebagai kata-kata yang mempunyai fungsi pendukung *function words*). Oleh karena itu dapat diabaikan dalam pencarian informasi, yang pada gilirannya akan mempercepat proses penelusuran.

Prinsip Kata Abai pada Winisis

Dalam program CDS/ISIS (baik versi DOS maupun versi Windows atau Winisis), fungsi **kata abai** ini dapat dimanfaatkan untuk mempercepat proses penelusuran sekaligus untuk menghemat pemakaian ruang penyimpanan data. Kalau kita menyiapkan suatu berkas yang berfungsi sebagai daftar **kata-abai** untuk suatu basisdata, maka semua kata-kata atau istilah yang terdapat pada seluruh cantuman basisdata dan didaftar di dalam berkas **kata-abai**, tidak akan diindeks oleh Winisis. Oleh karena itu akan mengurangi besarnya *byte* dari berkas indeks (XRF dan IFP). Dalam proses penelusuran, kalau kata-kata/istilah itu digunakan oleh penelusur, maka langsung diabaikan atau tidak dicari oleh program. Winisis akan langsung mencari berdasarkan kata-kata/istilah lain yang tidak terdaftar dalam berkas **kata-abai** itu.

Semua kata-kata atau istilah yang terdapat dalam berkas **kata-abai** pada suatu basisdata tertentu (misalnya basisdata BUKU), tidak akan terdapat dalam berkas kamus (*dictionary* atau berkas indeks) basisdata itu. Seperti diketahui bahwa setiap basisdata dalam Winisis kalau dibuatkan berkas teknik pengindeksannya secara automatis akan membuat berkas indeks yang biasa disebut kamus *(dictionary)*. Di dalam berkas kamus inilah akan didaftar menurut abjad semua istilah yang terdapat dapat dalam semua cantuman basisdata dan diindeks melalui teknik pengindeksan tertentu. Berkas kamus ini akan menunjukkan posisi (*posting*) pada cantuman berapa saja kata-kata atau istilah tersebut terdapat dalam basisdata. Dengan demikian jika seseorang melakukan penelusuran pada Winisis, maka kata/istilah yang digunakan dalam menelusur pertama-tama akan dicocokkan dengan daftar kamus, kemudian pada daftar kamus itu akan diperiksa nomor cantuman (MFN=*Master File Number*) yang memuat kata/istilah itu, kemudian Winisis akan menampilkan cantuman yang berkaitan satu per satu. Teknik penelusuran dan penampilan data inilah yang membuat program Winisis dapat mencari dan menampilkan cantuman secara cepat, dan tidak bergantung pada besarnya basisdata atau jumlah cantuman yang ada dalam basisdata. Pada Winisis mencari data pada basisdata dengan jumlah cantuman 100 akan sama cepatnya dengan mencari data pada basisdata dengan jumlah cantuman 100 ribu. Karena keduanya akan menempuh mekenisme yang sama. Pertama mengecek pada daftar kamus, lalu menampilkan cantuman sesuai posting MFN yang memuat kata/istilah yang dicari.

Membuat Kata-abai pada Winisis

Nama berkas **kata-abai** pada Winisis harus sesuai dengan nama masing-masing basisdata. Jadi kalau akan dibuat **kata-abai** untuk basisdata BUKU, maka nama berkas **kata-abai**nya harus BUKU.STW (STW adalah singkatan dari STOPWORDS). Lokasinya pun sebaiknya dalam satu direktori dengan berkas-berkas basisdata yang lain (berkas MST, FDT, FMT, FST dsb). Ini berarti kalau ada banyak basisdata, maka setiap basisdata perlu mempunyai berkas **kata-abai** masing-masing. Berkas **kata-abai** itu berbentuk berkas teks (*text file*), artinya dapat dibuat dengan program pengolah teks misalnya **NOTEPAD** atau **MSWINWORD**. Disarankan membuatnya dengan NOTEPAD.

Selanjutnya isi dari berkas kata-abai itu, yakni dalam contoh ini BUKU.STW, mungkin adalah sebagai berikut:

A, AN, AND, AS, AT, ATAU, BUT, DAN, DARI, DENGAN, DI, ETC, FOR, FROM, IN, INTO, IT, ITS, KARENA, KE, KEPADA, NO, NOT, OF, OFF, ON, OVER, PADA, THE, TO, TOO, UNTUK, WITH, WITHIN.

Kata/istilah ini merupakan kata depan dan kata sambung yang banyak digunakan dalam bahasa Indonesia dan bahasa Inggris. Kata depan dan kata sambung yang sering digunakan dalam bahasa lain, misalnya Peracis, Belanda atau Jerman, dapat pula ditambahkan.

Sesungguhnya cara pengetikan kata-kata/istilah diatas tidak seperti itu, melainkan harus diketik rapat ke kiri, satu baris satu kata tanpa tanda koma, diketik dengan huruf besar (kapital) dan disusun menurut abjad. Jumlah huruf maksimal 10 dan maksimal jumlah **katai abai** adalah 799 kata. Jadi cara pengetikan berkas itu sesungguhnya seperti berikut ini:

A
AN
AND
AS
AT
ATAU
BUT
DAN
DARI
dst

Kata-kata/istilah dalam berkas ini sewaktu-waktu dapat diedit (ditambah atau dikurangi) sesuai kebutuhan

Setelah berkas ini dengan nama yang sesuai dengan nama basisdata (BUKU.STW) disimpan di dalam direktori yang sama dengan berkas-berkas lain dari basisdata yang sama, dan agar dapat bekerja aktif, maka basisdata perlu diindeks secara lengkap (*FULL*). Setelah basisdata diindeks secara lengkap, kalau kita memeriksa berkas kamus, maka tidak akan menemukan lagi kata-kata/istilah yang didaftar dalam berkas **kata-abai**. Dengan demikian proses penelusuran akan lebih cepat sekaligus terdapat penghematan pemakaian ruang penyimpanan (*space*).