

MENGINDEKS BIBLIOGRAFI DENGAN MENGGUNAKAN KATAKUNCI SECARA INTERAKTIF BERASASKAN KOMPUTER MIKRO

Oleh: Harihodin bin Selamat
Institut Sains Komputer,
Universiti Teknologi Malaysia

Sinopsis

Setiap bibliografi yang disimpan di dalam pengkalan data diindeks dengan menggunakan satu atau lebih kata kunci yang dipilih oleh pengguna. Pencarian dokumen adalah berasaskan rumusan carian dalam bentuk satu atau gabungan beberapa katakunci yang disatukan dengan menggunakan operator boolean untuk mengujudkan kaedah carian.

Rumusan carian diinput melalui terminal secara interaktif dan diproses dengan kaedah logik boolean dan teori set. Sistem adalah diharapkan untuk mendapatkan bibliografi yang memenuhi syarat rumusan carian tersebut. Kemudian disusuli dengan tindakbalas di antara pengguna dengan komputer untuk menawarkan peluang carian yang lebih berkesan dan seterusnya mempercepatkan proses carian.

Katakunci : Bibliografi, Indeks, Komputer, Katakunci, Mikro, Dokumen.

Pengenalan

Kebanyakan pengetahuan manusia ditulis dalam bentuk buku, jurnal, majalah dan berbagai cara lain, kemudian dikumpul dan disimpan dalam perpustakaan. Masalah yang mungkin timbul dalam hal ini ialah pencarian semula dokumen atau rujukan dokumen yang ingin dicari oleh seseorang pengguna. Dalam dunia moden, jumlah dan kadar penambahan maklumat amatlah besar, di mana hari demi hari perkara ini menimbulkan masalah untuk mendapatkan kembali dokumen yang diperlukan oleh pengguna. Dengan wujudnya masalah ini, ia telah mendesak para pustakawan dan saintis untuk mencuba menyimpan dokumen atau rujukan dokumen secara berkompuser agar pencarian maklumat dapat dilaksanakan dengan lebih mudah dan efektif.

Maklumat mesti disimpan dalam komputer sekiranya ia adalah untuk tujuan didapati kembali oleh pengguna. Masalahnya ialah bagaimana untuk menyusun data tersebut di dalam komputer supaya pencarian maklumat yang dikehendaki boleh didapati dengan segera dan tepat. Kajian ini menghasilkan satu pakej yang dipanggil KIBS (Keyword Indexing of Bibliographic System) untuk mengindeks bibliografi atau dokumen dengan menggunakan katakunci-katakunci secara interaktif yang berasaskan komputer mikro. Pakej ini direkabentuk dan dibina menggunakan Dbase 111 Plus.

Para pustakawan, penyelidik, pelajar dan berbagai golongan masyarakat mencari dokumen untuk mendapatkan sesuatu maklumat yang diperlukan bagi sesuatu tujuan. Kebanyakan sesebuah organisasi atau perpustakaan, menyimpan koleksi dokumen mengikut subjek, nombor akses, judul dan pengarang sebagai rujukan dengan menggunakan kad-kad indeks.

Masalah yang timbul apabila menggunakan cara ini ialah kad indeks tidak boleh menghubungkan di antara satu subjek dengan subjek lain atau satu indeks dengan indeks lain yang berkaitan secara terus. Lebih-lebih lagi, tiap-tiap kad yang digunakan sebagai indeks kepada sesuatu dokumen mesti diberi nombor yang sempurna dan perlu difailkan dengan rapi, dan penyalinan semula maklumat berlaku pada kad yang lain sekiranya ia digunakan untuk dokumen yang sama. Penggunaan indeks untuk dokumen di bawah beberapa subjek memerlukan kerja menaip yang berulang, fail menjadi banyak dan besar. Masalah mendapatkan kembali maklumat ada dibincangkan dari segi teori disusuli dengan analisa proses rujukan¹.

Dalam usaha untuk menampung masalah-masalah ini, sistem mengindeks katakunci bibliografi telah dibangunkan sebagai satu eksperimen untuk menyimpan, meminda, mencari bibliografi yang dihajatkan daripada pengkalan data.

Pemilihan Katakunci Sebagai Indeks

Tujuan mengindeks bibliografi adalah untuk membolehkan ia didapatkan kembali apabila perlu. Dalam sistem ini katakunci yang dipilih haruslah perkataan yang melambangkan isi kandungan bibliografi, pengarang, penerbit, subjek bibliografi dan sebagainya yang mempunyai kaitan rapat atau bermakna kepada sesuatu bibliografi yang ingin disimpan. Pokoknya katakunci-katakunci tersebut mestilah yang bersifat penerangan dan membawa erti ke atas bibliografi tersebut. Misalan, sekiranya sesuatu dokumen itu menerangkan tentang kaedah mengorganisasi fail berkomputer, dalam kontek ini 'fail' dan 'komputer' boleh digunakan sebagai katakunci di samping katakunci yang lain; asalkan pengguna boleh membayangkan dan mengaitkan katakunci dengan bibliografi yang ingin dicari. Dalam kajian ini bilangan katakunci yang boleh digunakan untuk mengindeks sesuatu bibliografi dihadkan sebanyak sepuluh katakunci. Walau bagaimanapun Sistem KIBS ini boleh menerima lebih dari sepuluh dengan mengubah bilangan katakunci yang dibenarkan untuk mengindeks sesuatu bibliografi dalam model yang berkenaan. Tujuan utama adalah supaya pengguna boleh mendapatkan kembali bibliografi yang dikehendaki dengan mengaitkan katakunci-katakunci yang bermakna di samping membentuk formula carian berupa Logik Boolean agar carian bibliografi dapat dilaksanakan oleh sistem dengan lebih berkesan. Formula carian boleh dibentuk menggunakan cara carian tunggal atau berganda yang menggunakan beberapa katakunci yang dicantumkan dengan menggunakan operator Boolean.

Operator Boolean

Operator Boolean akan digunakan sekiranya formula carian menggunakan lebih dari satu katakunci untuk menambahkan ketepatan carian. Boolean operator tersebut adalah 'AND' 'OR' dan 'NOT' di mana dalam sistem KIBS ini operator '*' mewakili 'AND', '+' mewakili 'OR', dan '-' mewakili 'NOT'. Adalah lebih baik sekiranya pengguna tahu dan memahami logik Boolean dan operasi set² sebelum menggunakan sistem ini untuk menghasilkan carian yang berkesan dan tepat. Boolean operator ini digunakan bagi mencantumkan katakunci-katakunci untuk membentuk formula carian Boolean yang lebih kompleks. Operator-operator ini tidak mempunyai sebarang keutamaan dan formula carian adalah diproses dari kiri ke kanan.

Kaedah Carian

Terdapat beberapa kaedah Carian bibliografi yang boleh digunakan seperti 'browsing', 'best-fit matching', carian bersyarat dan kaedah 'deductive'³. Kaedah carian yang digunakan dalam program ini ialah carian bersyarat dengan teknik sistem Boolean dan teori Set. Ia dijalankan secara interaktif supaya kaedah perhubungan antara mesin dan manusia dapat memberikan carian yang lebih berkesan.

Bagi mengarahkan komputer untuk melaksanakan carian pengkalan data, ia memerlukan cara yang formal untuk permohonan carian (search request). Di sini katakunci-katakunci yang berkaitan dengan bibliografi yang ingin dicari akan digunakan di dalam pembentukan formula carian dengan menggunakan kaedah logik Boolean dan operasi Set. Setelah pemrosesan formula carian tersebut tamat, komputer akan menghasilkan satu senarai nombor bibliografi yang memenuhi syarat seperti yang ditetapkan dalam formula carian itu. Berikut ini satu proses operasi ke atas formula carian adalah dibincangkan.

Sebagai contoh, katakunci-katakunci yang ingin digunakan dalam carian adalah diinput melalui terminal dan sistem akan memberikan kod bagi tiap-tiap katakunci tersebut seperti berikut:

Kod	Katakunci
a	- Merekabentuk
b	- Martin J.
c	- Dokumentasi
d	- 1982
e	- Dates C. J.
f	- Struktur
g	- Analisa Sistem

Andaian yang dibuat di sini adalah semua katakunci yang diinput berada di dalam fail kamus katakunci dalam pengkalan data. Untuk makluman, sekiranya mana-mana katakunci yang diinput oleh pengguna tiada dalam fail kamus, maka sistem akan memberitahu pengguna yang katakunci tersebut tiada dalam pengkalan data.

Dalam pembentukan formula carian, hanya kod bagi katakunci digunakan dalam formula rumusan carian untuk mengurangkan kesilapan menaip seperti di bawah.

$$f * g * a - c * b + e * d ,$$

di mana formula carian tersebut bermaksud:—

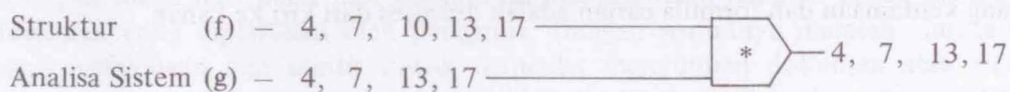
Cari dokumen yang membincangkan tentang analisa sistem dan merekabentuk secara berstruktur tetapi bukan tentang dokumentasi yang dikarang oleh Martin J. atau Dates C. J. yang diterbitkan dalam tahun 1982.

Satu lagi andaian yang dibuat dalam contoh ini ialah katakunci-katakunci tersebut mempunyai senarai kejadian seperti berikut:—

Katakunci	Senarai Kejadian
merekabentuk	— 2, 4, 7, 10, 12, 13, 17
Martin J	— 2, 3, 4, 10, 13, 16
Dokumentasi	— 1, 5, 7, 19, 25
1982	— 1, 3, 4, 5, 18, 26
Dates C.J.	— 2, 4, 7, 8, 17, 26
Struktur	— 4, 7, 13, 17
Analisa Sistem	— 4, 7, 10, 13, 17

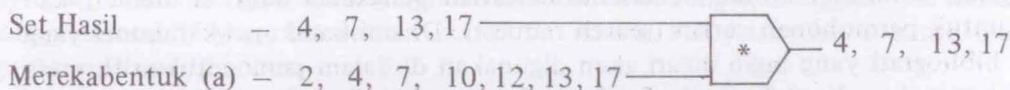
Formula carian di atas akan melibatkan 6 operasi boolean dan set seperti di bawah:—

Operasi 1 — $f * g$ yang bermaksud mengambil semua nombor kejadian yang sama berlaku di antara katakunci 'struktur' dan 'analisa sistem' di mana ia akan menghasilkan set 4, 7, 13, 17.

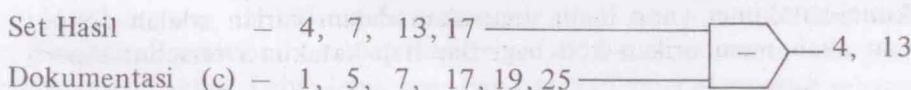


Di sini konsep teori set diperkenalkan untuk mendapatkan hasil dari operasi sesuatu operator boolean bagi dua operan (katakunci). Gambarajah 1 menggambarkan operasi set bagi operator '*' (AND).

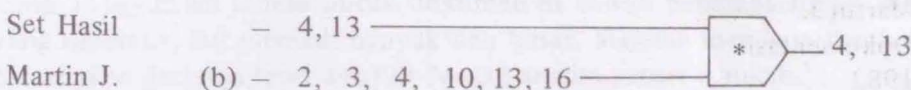
Operasi 2 — Set (4, 7, 13, 17) dari operasi 1 diproses dengan nombor kejadian bagi katakunci 'merekabentuk' dan menghasilkan set (4, 7, 13, 17), i.e.



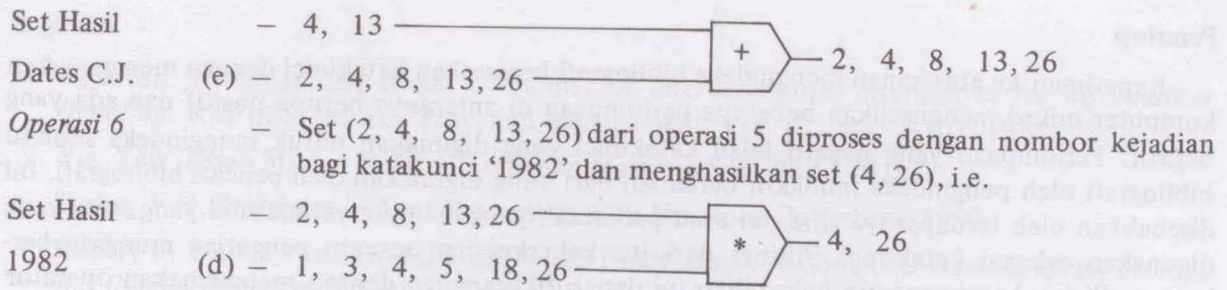
Operasi 3 — Set (4, 7, 13, 17) dari operasi 2 diproses dengan nombor kejadian bagi katakunci 'Dokumentasi' dan menghasilkan set (4, 13). Gambarajah 2 memberi gambaran operasi set bagi operator '-' (NOT).



Operasi 4 — Set (4, 13) dari operasi 3 diproses dengan nombor kejadian-kejadian bagi katakunci 'Martin J.' dan menghasilkan set (4, 13), i.e.

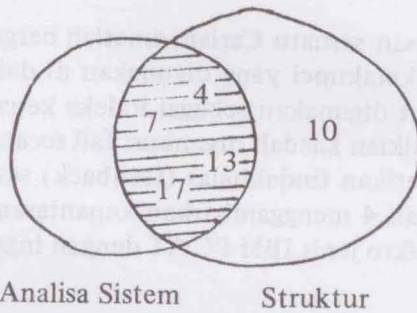


Operasi 5 — Set (4, 13) dari operasi 4 diproses dengan nombor kejadian bagi Kata Kunci 'Date C.J' dan menghasilkan set (2, 4, 8, 13, 26). Gambarajah 3 menggambarkan operasi set bagi operator '+' (OR).

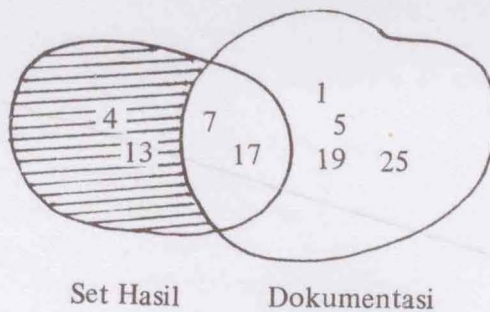


Berdasarkan kepada kejadian katakunci tersebut sebagai indeks, formula Carian telah menghasilkan satu set nombor bibliografi di mana bibliografi inilah yang memenuhi syarat Boolean setelah operasi set dilaksanakan. Sistem KIBS akan mempamerkan atau mencetak bibliografi-bibliografi tersebut mengikut kehendak pengguna.

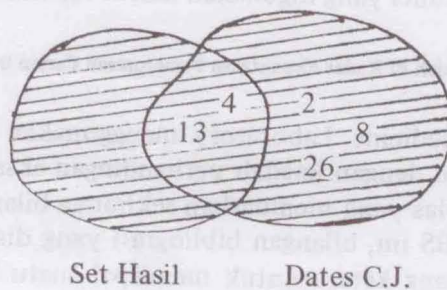
Pengguna boleh menggubah formula carian dengan menggunakan katakunci yang sama atau menginput katakunci baru untuk mendapatkan bibliografi yang lebih tepat atau membuat carian baru.



Gambarajah 1 Operasi Set Ke Atas Kejadian Katakunci Struktur * Analisa Sistem



Gambarajah 2 Operasi Set Ke Atas Set Hasil - Dokumentasi



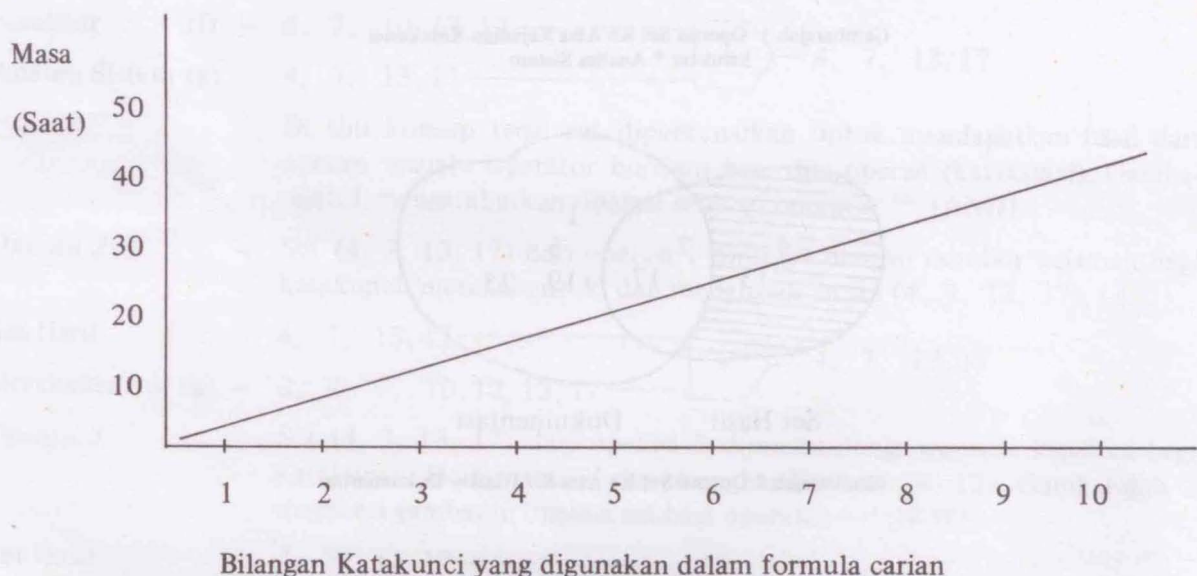
Gambarajah 3 Operasi Set Ke Atas Set Hasil + Dates C. J.

Penutup

Experimen ke atas kajian mengindeks bibliografi berasaskan katakunci dengan menggunakan komputer mikro menghasilkan beberapa perjumpaan di antaranya berupa positif dan ada yang negatif. Perjumpaan yang negatif ialah katakunci yang digunakan untuk mengindeks sesuatu bibliografi oleh pengindeks mungkin berlainan dari yang digunakan oleh pencari bibliografi. Ini disebabkan oleh terdapatnya sinonim atau perkataan yang mempunyai maksud yang sama yang digunakan sebagai katakunci. Akibat dari itu keberkesanan sesuatu pencarian mungkin berkurang. Walau bagaimanapun kelemahan ini dapat dikurangkan dengan menggunakan operator '+' (OR).

Jika dilihat dari segi kebaikannya pula, pengindeks boleh menggunakan apa jua jenis katakunci untuk dijadikan indeks. Ini memberi kelonggaran pada penyediaan indeks bagi sesuatu bibliografi. Katakunci yang digunakan sebagai indeks boleh terdiri daripada pengarang, penerbit, tahun diterbitkan, subjek, isi kandungan dan sebagainya asalkan katakunci yang digunakan itu menggambarkan erti bibliografi tersebut. Penggunaan Carian Boolean dengan menggunakan operator boolean '*' (AND), '+' (OR) dan '-' (NOT) untuk mencantumkan beberapa katakunci supaya membentuk satu Formula Carian Boolean, dapat menghasilkan suatu carian yang berkesan dan memuaskan. Didapati adalah lebih baik sekiranya pencari bibliografi memahami teori set dan pengoperasian boolean sebelum menggunakan pakej ini supaya dapat menghasilkan carian yang berkesan.

Dari segi kepantasan pemerosesan sesuatu Carian, amatlah bergantung kepada jenis prosesor PC yang digunakan dan bilangan katakunci yang digunakan di dalam formula Carian dan juga bilangan sesuatu katakunci tersebut digunakan sebagai indeks kepada bibliografi yang disimpan dalam pengkalan data. Namun demikian kaedah organisasi fail secara semi-songsang yang digunakan dalam sistem ini dapat memberikan tindakbalas (feedback) sesuatu carian dengan memuaskan, i.e. dalam beberapa saat. Rajah 4 menggambarkan kepantasan pemerosesan sesuatu carian dengan menggunakan komputer mikro jenis IBM PC AT dengan ingatan 640 K.



Gambarajah 4: Kadar Kepantasan Pemerosesan Carian Boolean

Penyelidik di Computer Medicine Laboratory menggunakan komputer untuk membantu mendapatkan kembali dokumen dengan kaedah perbandingan aksara/perkataan⁴, tetapi kaedah ini tidak memberikan tindakbalas yang memuaskan sekiranya bilangan dokumen yang disimpan adalah besar. Dalam sistem KIBS ini, bilangan bibliografi yang disimpan dalam Pengkalan data ini tidak memberikan kesan yang ketara untuk mencapai suatu kepantasan yang munasabah. Namun demikian bilangan katakunci yang digunakan dalam formula carian adalah menjadi satu faktor dalam kepantasan proses carian seperti yang digambarkan dalam Gambarajah 4.

Rujukan

- 1 Kathleen T. Bibias dan Roger C. Palmer, *A Microcomputer alternatives for information Handling*: REFLES, Inform. Proc, and Management, Vol. 17., ms 93-101, 1981.
- 2 Yap Yew, *Hisab Moden Buku 4*, Longman Malaysia Sdn. Bhd., 1978.
- 3 Teskey F.N, *Principles of Text Processing*. Ellis Horwood Publishers, 1982.
- 4 Robert F. Beckley and Howard L. Bleich, *Paper Chase: A Computer-Based Reprint Storage and Retrieval System*, Comput. Biomed. Research, Vol. 10, ms 423-430, 1977.
- 5 Harold Barko, *Experiment in Book Indexsing by Computer*, Inform. Storage & Retrieval. Vol 6, ms 5-16, 1970.
- 6 Simon M Newman, *An Operative Information Retrival System Based on Relational Factors*, Inform Storage & Retrieval, Vol. 3, ms 331-350, 1967.