

MENYARING INFORMASI DI DUNIA WEB

Teknik Temu Balik Informasi (Information Retrieval) dan Analisa Kapabilitas Pencarian Search Engine Google (www.google.com)

Oleh:
Miyarso Dwi Ajie
Indonesia University of Education

Pendahuluan

Membayangkan Internet sebagai sekedar jaringan komputer adalah tidak tepat. Jaringan komputer hanyalah medium yang membawa informasi. Daya guna Internet terletak pada informasi itu sendiri, bukan pada jaringan komputer. Fokus yang ditekankan dalam penelitian ini adalah Internet sebagai sumber daya informasi bukan sebagai jaringan komputer.

Nilai yang ditawarkan oleh Internet dapat dikiasakan sebagai sistem jalan raya dengan transportasi berkecepatan tinggi, yang memperpendek waktu perjalanan. Atau dapat juga diibaratkan sebuah perpustakaan yang dapat di kunjungi setiap saat, dengan kelengkapan buku dan sumber informasi yang tak terbatas.

Bentuk informasi yang dikemas dalam *World Wide Web* (Web) adalah bentuk penyajian informasi yang paling populer saat ini di Internet. *Web* menawarkan bentuk sajian informasi yang menarik, variatif, dinamis dan interaktif. Di dunia web informasi disajikan dalam beragam bentuk, baik itu berupa teks, gambar, grafik, animasi, suara dan sebagainya. Kebanyakan apa yang kita lakukan dalam dunia Internet adalah mengakses suatu situs web.

Web mampu merangkul semua bentuk informasi yang ada sekarang ini, baik itu informasi berbentuk visual (teks, gambar diam maupun dalam bentuk animasi), dan juga informasi berbentuk audio serta audiovisual.

Apa yang dapat ditemukan dalam dunia web? Jawaban pendek untuk pertanyaan ini adalah "hampir semua". Puluhan juta dokumen saat ini tersedia di dunia web, meliputi rentang topik yang beragam. Beberapa diantaranya akan sangat berguna bagi akademisi adalah informasi mengenai:

- ✓ **Teaching Resources:** *lecture notes, interactive tutorials, discussion articles*
- ✓ **Research:** *project description, databases, conference proceeding, grants* (bantuan dana penelitian)
- ✓ **Computing:** *kumpulan piranti lunak, tutorials, manuals, discussions*
- ✓ **Government:** *political parties, United Nations, European Union*
- ✓ **Environment:** *Conservation, climate change, wildlife, pollution*
- ✓ **Career:** *vacancy listing, career agencies, on-line newspapers*
- ✓ **Culture:** *art galleries, music, cinema, religion* dan masih banyak lagi.

Dan sangatlah tepat untuk menyebut bahwa *World Wide Web* sebagai sebuah perpustakaan yang ter-desentralisasi atau *decentralized library*. (<http://www.abdn.ac.uk/diss/webpack/chap1.hti>, *date access: 09/11/2002*).

Memasuki dunia *Cyber* (baca: *Web*) ini memang diibaratkan berada dalam lautan informasi, bisa jadi anda yang tengah mencari suatu informasi tertentu bisa juga tersesat didalamnya. Masalah klasik yang dihadapi dalam temu balik informasi (*Information Retrieval*) di Internet adalah

menemukan berlimpah informasi (*find too much*), atau sama sekali tidak menemukan apa-apa (*find nothing*)

Untuk menanggulangi permasalahan tersebut maka terciptalah fasilitas yang khusus digunakan untuk piranti temu balik informasi di dunia web yang disebut *Search Engine*. *Search engine* ini dibuat untuk membantu pengguna Internet dalam melakukan temu balik informasi yang diinginkannya.

Saat ini banyak sekali bermunculan *search engine* di Internet, baik itu *search engine* yang menyediakan jasanya secara gratis maupun dengan sistem pembayaran setiap kali penggunaan (*pay per-click*). Diantara sekian *search engine* yang menyediakan jasanya secara gratis sebut saja misalnya, **Alltheweb** (www.alltheweb.com), **Alta Vista** (www.altavista.com), **Google** (google.com), **Naver** (www.naver.co.id), **AOL Search** (www.search.aol.com), **Lycos** (www.lycos.com) dan **MSN** (www.msn.search.com) dan masih banyak lagi. Pengertian dari *search engine* itu sendiri kurang lebih sebagai berikut,

“A program that searches documents for specified keywords and returns a list of the documents where the keywords were found” (www.webopedia.com, Kata pencarian: *search engine*, 22/11/01)

Jadi secara umum tugas *Search Engine* ini adalah melakukan temu balik informasi yang sesuai dengan *query* (kata kunci yang akan dicari) yang di ketikan, kemudian *Search Engine* mulai mencari pangkalan data (*database*), situs *web* dan juga *search engine* lainnya, yang berhubungan dengan kata pencarian, dan menyajikannya dalam bentuk list (daftar *hyperlink* dari halaman dokumen-dokumen atau *web page* yang berhubungan dengan kata pencarian)

Secara teknis *Search Engine* bekerja dengan mengirimkan *spider* (suatu program) untuk menangkap sebanyak mungkin halaman-halaman dokumen yang mempunyai konten sesuai dengan kata pencarian yang dimasukan. Kemudian setelah itu, sebuah program lainnya yang disebut *indexer*, membaca dokumen-dokumen (halaman web) yang telah ditemukan tersebut dan menciptakan suatu indeks berdasarkan kata-kata yang mengisi dari tiap dokumen. Tiap-tiap *Search Engine* menggunakan sistem algoritma yang sesuai untuk menciptakan indikasi seperti diatas, idealnya hanya hasil yang sesuai dengan kata kunci yang akan ditampilkan dari tiap pertanyaan.

Hasil dari kecenderungan yang di gambarkan diatas dalam ilmu perpustakaan adalah bahwa saat ini pustakawan memiliki tantangan untuk dapat mengakses informasi yang lebih dalam dengan kesatuan bentuk yang lebih luas. Para pengguna telah lebih canggih (pintar) dan mengharapkan penggunaan dari teknologi informasi untuk mendapatkan hasil yang lebih cepat, dan pencarian informasi yang lebih efisien.

Teknologi harus dipandang oleh pustakawan sebagai cara untuk meningkatkan layanan kepada publik (pengguna) dan membantu kerja mereka menjadi lebih efisien. Dengan cara otomasi, perpustakaan dapat menyediakan layanan katalog yang mutakhir dan mudah ditelusur.

Dengan adanya perkembangan Internet ini (khususnya perkembangan Web), layanan perpustakaan akan mendobrak keterbatasan ruang dan waktu serta birokrasi. Namun untuk menuju kepada kondisi itu pustakawan harus mempersiapkan diri dengan memahami beberapa hal teknis dan praktis mengenai karakteristik persiapan layanan Internet. Pustakawan perlu berkonsentrasi pada kandungan atau isi informasi disbanding infrastruktur Internet yang telah dirintis dan ditanggapi oleh profesi atau pihak lain. (Campbel, 1998: 4)

Di internet kita seolah berada dalam lautan informasi, dan hal itu tidak akan berarti apa-apa jika kita tidak mengetahui bagaimana cara mencarinya, hal ini merupakan tugas pokok sebagai seorang *information professional* untuk mengantisipasi permasalahan tersebut, seperti yang ditegaskan oleh Richard Saul Wurman dalam *Information Anxiety*:

“You don’t have to know everything, You just need to know how to find it”
(www.sunflower.com/~madfinch/pro/map/readings.htm, date access: 12/03/02)

Alat Bantu Temu Balik Informasi di Dunia Web

Sampai saat alat bantu temu balik yang menjadi standar di dunia web (baca; Internet) dapat digolongkan menjadi dua macam, yaitu yang berbentuk Direktori dan *Search Engine*.

A. Direktori

“Directories are collections of information organized by subject. Most directories are organized hierarchically and allow the user to “drill down” to lower detailed levels” (Archer, 1996)

(<http://www.business.mcmaster.ca/msis/profs/head/papers/memos/userhist.htm>
date access: 10/11/02)

Direktori adalah suatu alat bantu temu balik informasi yang merupakan suatu kumpulan atau koleksi dari berbagai informasi yang diorganisasi berdasarkan subyek. Pada umumnya direktori web dibentuk dan dikelola oleh manusia dibandingkan dengan menggunakan software *-robot* atau *spider-* yang sudah terotomasi seperti halnya pada *Search Engine*. Keterlibatan peran “manusia” ini membuat direktori berbeda dengan *Search Engine*. Pangkalan data direktori lebih kecil dibanding dengan pangkalan data *Search Engine*, tetapi cenderung memberikan hasil yang relevan. Dalam beberapa kasus, sumber-sumber informasi yang ada dalam suatu direktori juga dinilai dan diperiksa. Direktori khusus diorientasikan kepada topik-topik tertentu. Umumnya direktori menyediakan suatu pendekatan yang terfokus dan terorganisasi untuk mengetahui sumber informasi daripada *search engine*.

Secara umum direktori web menyertakan fitur pilihan untuk melakukan pencarian menggunakan kata kunci berupa kata atau frase. Fitur ini dimaksudkan untuk membantu pengguna untuk berpindah secara cepat ke bagian direktori yang lebih sesuai, sehingga tidak perlu mencari secara porsif besar di dunia web. Dibandingkan dengan Direktori, *search engine* lebih berguna dalam membantu pengguna yang mengharapkan untuk mencari informasi tertentu dari jutaan situs-situs web untuk suatu kata atau frase pencarian.

B. Search Engine

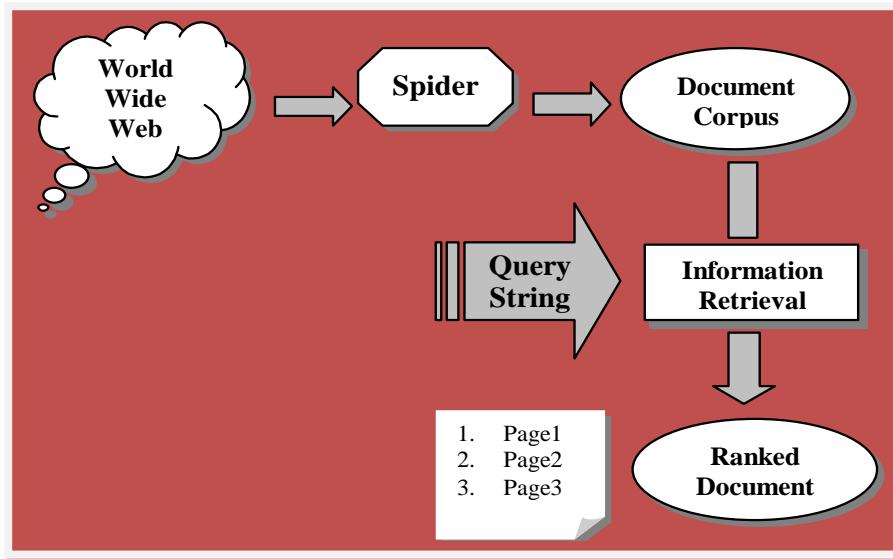
“A program that searches documents for specified keywords and returns a list of the documents where the keywords were found”

(Webopedia.com; query: *search engine*).

Search engine menyediakan layanan dan kontrol lebih kepada pengguna dalam melakukan suatu pencarian informasi. Secara garis besar tugas *Search engine* ini adalah melakukan temu balik informasi yang sesuai dengan kata pencarian yang diketikkan pada kotak pencarian. Kemudian *Search engine* mulai mencari dokumen-dokumen (dapat berupa *database*, *homepage* atau juga *Search engine* lainnya) yang berhubungan dengan kata pencarian, dan menyajikannya dalam bentuk daftar *hits* dari halaman dokumen-dokumen yang berhubungan dengan kata pencarian.

Secara teknis *Search engine* bekerja dengan mengirimkan *Spider* (atau juga dinamakan robot -- sebuah program--) untuk ‘menangkap’ sebanyak mungkin dokumen-dokumen. Kemudian setelah itu, sebuah program lainnya yang disebut *Indexer*, membaca dokumen-dokumen yang telah ditemukan tersebut dan menciptakan suatu indeks berdasarkan kata-kata yang mengisi dari tiap dokumen. Tiap-tiap *Search engine* menggunakan sistem algoritma yang sesuai untuk menciptakan indikasi seperti di atas, idealnya hanya hasil yang sesuai dengan kata kunci yang akan ditampilkan dari tiap pertanyaan. Berikut ilustrasi sistem pencarian dunia web:

Gbr1. Web Search system



Namun satu yang harus disadari bahwa setiap *search engine* mempunyai prinsip kerja yang berbeda. Beberapa *search engine* (seperti Hotbot, www.hotbot.com) bekerja dengan menggunakan program komputer untuk menyusun dan menggali dari list atau daftar situ-situs web, dan ada juga *search engine* yang bekerja dengan mencari *Search engine* lainnya atau yang dikenal sebagai "*meta-Search Engine*"

Search engine menggunakan indeks untuk menangkap terma (*term*) dari kata pencarian (*query*) yang diketikan. Ini berarti semakin banyak data yang di-indeks, semakin tinggi *Recall* (rasio perolehan). Meng-indeks setiap kata atau kata-kata yang paling sering digunakan dalam suatu dokumen, dapat memberikan kesempatan untuk menemukan informasi yang relevan dengan kata pencarian. Semakin besar indeks, semakin besar kemungkinan mendapatkan dokumen yang sesuai dengan kata pencarian. Dan dengan perkembangan *Web* yang semakin pesat, semakin sering indeks suatu *search engine* di perbarui (*update*), semakin relevan jumlah informasi yang ditelusur. Dan kehadiran *SearchEngine* ini menjadi alat bantu yang sangat penting dalam menelusur informasi di dunia web.

Namun sayangnya, kebanyakan dari mereka yang telah terbiasa berinteraksi di dunia maya (*web*) rata-rata baru mempergunakan 10-15% dari "kekuatan" Internet yang sebenarnya. Hal ini disebabkan karena banyaknya pengguna Internet belum mengetahui teknik menelusur informasi yang tepat dan efektif, sehingga pengguna internet dapat memperoleh informasi, maupun pengetahuan yang relevan dengan apa yang dibutuhkannya, terutama dalam usahanya untuk mengembangkan diri dan mempermudah pekerjaan atau aktifitas sehari-hari (Indrajit, 2002: 1).

Di Internet (*web*), terdapat ratusan bahkan ribuan *Search engine* yang dapat diakses secara cuma-cuma. Karena sifatnya sebagai pintu gerbang bagi para pengguna sebelum memasuki situs (*website*) yang diinginkan.

Teknik Menyaring Informasi & Analisa Kapabilitas Pencarian Google

Untuk menghindari adanya limpahan informasi yang tidak relevan maka perlu dilakukan filterisasi informasi (*information filtering*) atau dengan pembatasan pencarian (limit). Tiap search engine menawarkan berbagai fitur dalam pembatasan pencarian, namun dalam penelitian ini dilakukan dengan yang paling sederhana yaitu cara tiap kata pencarian akan ditambah dengan penggunaan boolean operator (**AND**, **OR**, dan **NOT**) atau dapat berupa simbol matematika seperti (+), (-), dan batasan pencarian umum lainnya seperti penggunaan double *quote* atau tanda petik (" "),

Nesting (tanda kurung) atau fitur batasan pencarian lainnya yang dimiliki oleh masing-masing search engine.

a. Boolean Searching

Boolean Searching merupakan suatu cara untuk menggabungkan kata pencarian dengan menggunakan beberapa "konektor" agar memperoleh hasil termu balik informasi yang diinginkan. *Boolean Searching* tersebut diantaranya:

1. **AND** digunakan untuk kata pencarian lebih dari satu. Contoh: library AND librarian. Maka *search engine* akan mencari berbagai dokumen yang mengandung kata "library" dan "librarian".
2. **OR** sama artinya dengan ATAU dalam Bahasa Indonesia, digunakan jika ingin mencari dokumen yang mengandung salah satu kata pencarian. Contoh: library OR librarian, maka *search engine* akan mencari beragam dokumen yang mengandung kata "library" atau "librarian" atau gabungan keduanya. Pada beberapa *search engine* (misalnya Google), simbol ini bekerja dengan cara yang sama dengan spasi.
3. **NOT** digunakan bila ingin membatasi pencarian dengan menghilangkan salah satu kata pencarian. Misalnya, library AND Asia NOT Vietnam, maka *search engine* akan mencari berbagai dokumen yang mengandung kata library dan asia kecuali Vietnam.
4. **()** Tanda kurung ganda atau **Nest** ini berfungsi sama dengan tanda kurung dalam ilmu matematika, yaitu melakukan isolasi terhadap sebuah perhitungan yang harus didahulukan, contoh: **management AND (archive OR record)**, kata pencarian tersebut akan memiliki makna agar *search engine* terlebih dahulu mencari dokumen yang ada dalam tanda kurung yaitu "archive" atau "record" atau gabungan atas keduanya dan hasil termuan tersebut harus pula mengandung kata "management".
5. **[+]** Simbol plus ini sama dengan simbol **AND**, namun simbol ini harus ditempatkan didepan kata pencarian. Contoh: untuk kata pencarian: +library+librarian, maka hal ini ekuivalen dengan:library AND librarian.
6. **[-]** Simbol minus ini berfungsi sama dengan operator NOT, namun ditempatkan didepan kata yang ingin dihilangkan, contoh: library AND (asia or afrika)-indonesia, maka *search engine* akan mencari dokumen yang mengandung kata library kemudian diikuti oleh asia atau afrika kecuali Indonesia.

b. Case Sensitive

Secara umum, *search engine* akan mencocokkan penggunaan huruf kapital (*uppercase*), huruf kecil (*lowercase*) sebagai kata pencarian yang sama. Beberapa *search engine* mempunyai kemampuan untuk mencocokkan *case* yang tepat. Memasukan kata pencarian dalam *lowercase* biasanya akan menemukan semua *cases*. Dalam suatu *Case Sensitive Search Engine*, memasukan beberapa huruf kapital dalam kata pencarian akan mencari (memanggil) *case* yang cocok. Contoh: memasukan kata pencarian **opac** maka *search engine* akan mencari kata **opac**, **Opac**, **OpAC**, dan varian lainnya sedangkan memasukan kata pencarian **OpAC** *search engine* akan mencari dokumen sesuai kata pencarian yaitu **OpAC**

c. Databases

Sebuah pangkalan data adalah suatu koleksi informasi tertentu yang diorganisasikan dalam sebuah bentuk struktur tertentu. Beberapa alat bantu (*tools*) *search engine* menyediakan *results* (hasil pencarian) lebih dari satu pangkalan data. Sebagai contoh, Yahoo selain mencari direktori *database* sendiri juga menggunakan pangkalan data milik *search engine* Inktomi

d. Default Operation (Konfigurasi Normal)

Default merupakan reaksi *setting* atau operasi tertentu yang secara otomatis diambil oleh *hardware* atau *software* tertentu. Jika pengguna tidak memberikan perintah yang spesifik dan dapat diidentifikasi oleh *search engine*. Sebagai contoh jika pengguna memasukkan kata pencarian *knowledge management* tanpa *Boolean searching* pada *search engine* Google, maka kata tersebut dapat diproses sebagai: *knowledge AND management*, *knowledge OR management*, atau "*knowledge management*" sebagai sebuah frase secara otomatis.

e. Display

Adalah ulasan informasi hasil temu balik yang ditampilkan dalam setiap pencarian dan memberikan gambaran mengenai layanan alternatif jika tersedia. Ini meliputi jumlah dokumen yang dapat ditampilkan dalam tiap pencarian dan jika ada batas jumlah dokumen yang bisa ditampilkan.

f. Field

Field Searching menyediakan layanan pengguna untuk meramu kata pencarian, agar memperoleh hasil temu balik yang lebih sesuai. *Field searching* memberikan struktur khusus dalam menentukan jenis dokumen yang akan ditemu balik. Judul, URL (*Unit Resource Location* atau alamat situs), gambar, atau adalah beberapa hal tertentu yang harus dikandung oleh dokumen yang akan dicari.

g. Homepage

Adalah tampilan awal dari *search engine*. Hal ini sangat penting dalam menjangkau minat pengguna dalam menggunakan *search engine*. Tampilan yang sederhana dan mudah digunakan merupakan hal pertama yang diperhatikan pengguna dalam memilih *search engine*.

h. Limits

Adalah kemampuan *search engine* untuk mempersempit atau membatasi hasil temu balik (*results*) dengan menambahkan suatu *specific restriction* (batasan khusus) dalam sebuah pencarian. Misalnya, *date limit* yaitu batasan pencarian berdasarkan tanggal dan *language limits* atau batasan pencarian berdasarkan bahasa.

h. Proximity Searching

Proximity Searching merujuk pada kemampuan *search engine* untuk menspesifikasi dokumen yang akan ditemu balik (*retrieve*) dengan menggunakan kata pencarian ganda (*multiple terms*). Simbol standar yang digunakan untuk mengidentifikasi frase adalah tanda petik ganda " " mengurung frase. [" "] tanda petik ganda atau *double quotes* digunakan jika kata pencarian berbentuk frase ingin dianggap sebagai satu kesatuan, contoh: "*knowledge management*" maka *search engine* akan mencari dokumen yang mengandung kata *knowledge management* secara utuh.

i. Sorting

Kemampuan *search engine* untuk mengorganisasi hasil temu balik (*results*) dari setiap pencarian. Biasanya *search engine* mengurut hasil temu balik berdasar relevansinya dengan kata pencarian, yang ditentukan menggunakan pengurutan (*ranking*) sistem algoritma yang sesuai. Pilihan lainnya adalah untuk mengatur hasil pencarian berdasar tanggal (*date*), judul berurut secara alfabetis, atau berdasar URL (*Unit resource locator*) atau *host name*.

j. Stop word

Beberapa *search engine* mengikutkan kata-kata tertentu seperti 'the' dan lainnya yang juga memungkinkan ikutnya sejumlah atau HTML *string* yang sering dicari. Contoh: *the, a, is, of, be, I, html, com*.

k. Kekuatan

Keuntungan-keuntungan dan alasan-alasan untuk menggunakan alat bantu cari. Hal ini bersifat subjektif dan didasarkan atas opini peneliti.

I. Truncation / wild card / Stemming

Teknik pencarian ini merujuk pada kemampuan untuk mencari hanya sejumlah porsi dari suatu kata. Biasanya, sebuah simbol seperti asterisk (*) digunakan untuk mewakili sisa dari kata pencarian. Contoh; Swim*, maka *search engine* akan mencari dokumen yang mengandung varian kata swim seperti kata *swimming, swims, swimmer* dan varian lainnya.

- **Stemming** masih berhubungan dengan *truncation*, istilah ini merujuk pada kemampuan *search engine* untuk menemukan variant dari suatu kata, seperti dalam bentuk jamak (*plurals*), bentuk tunggal, bentuk lampau (*past tense*), *present tense* dan sebagainya, misalnya kata eat* maka search engine mencari dokumen yang mengandung kata *eat, eaten, ate* dan sebagainya. Beberapa *stemming* hanya meliputi bentuk jamak dan tunggal saja.
- **End Truncation**, contoh: college* akan mencari dokumen yang mengandung varian kata college seperti kata *colleges, collegium, collegial*.
- **Internal Truncation**, contoh: kata pencarian col*r maka *search engine* akan mencari dokumen dengan kata-kata *color, colour, colander* dan sebagainya.
- **Stemming**: lights akan mencari *lights, lighting, lit*.

m. Field Searching

Google menawarkan beberapa *field searches* yang berhubungan dengan URLs dalam melakukan pencarian. Pada bulan Desember 2000 suatu revisi pada bentuk layanan *Advanced search*, Google menambahkan beberapa fitur pencarian *Title* (judul) dan *URLs Field searches* (pencarian berdasarkan alamat situs).

Harus dicatat bahwa kebanyakan *field searching* tidak dapat dikombinasikan dengan kata pencarian lainnya. Dengan kata lain, sebuah kata pencarian yang diketikkan seperti: **carikata link:nama.com** hanya akan diproses hanya jika *field search* diketikkan seperti: **link:nama.com**. carikata diabaikan oleh *search engine*. Satu pengecualian adalah fitur **site:** yang tidak dapat digunakan sendiri dan harus ditambah dengan kata pencarian lainnya. Pengecualian terbaru adalah fitur field **intitle:** dan **inurl:** yang dapat dikombinasikan dengan kata pencarian lainnya. Berikut tabel fitur *field searching* yang dimiliki Google:

| No. | FIELD | KETERANGAN |
|-----|--------------------|---|
| 1 | intitle: | Akan mencari dokumen-dokumen yang mempunyai elemen judul yang sesuai dengan kata pencarian yang diketikkan. Dapat dikombinasikan dengan kata-kata pencarian lainnya. Contoh: intitle: knowledge management |
| 2 | inurl: | <i>Search engine</i> akan mencari berbagai dokumen yang mempunyai URL atau alamat situs sampai pada level terendah (nama <i>file</i>) yang sesuai dengan kata pencarian. Dapat dikombinasikan dengan kata pencarian lainnya. Contoh: inurl: searchenginewatch |
| 3 | allintitle: | <i>Search engine</i> akan mencari dokumen-dokumen yang mempunyai judul (HTML) yang sesuai dengan kata pencarian. Sebagai contoh: allintitle:knowledge management , maka <i>search engine</i> akan mencari berbagai dokumen yang mempunyai judul <i>knowledge management</i> |
| 4 | link: | <i>Search engine</i> akan mencari berbagai dokumen yang mengandung kata-kata yang sesuai dengan kata pencarian pada alamat situs (URL) yang diketikkan. Contoh: link:kompas.com/korupsi , maka <i>search engine</i> akan mencari berbagai dokumen yang mengandung kata korupsi yang merujuk pada situs <i>kompas.com</i> |

| | | |
|---|---------------------|--|
| 5 | allinurl: | <i>Search engine</i> akan mencari dokumen-dokumen yang mempunyai alamat atau URL yang sesuai dengan kata pencarian yang ketikan. <i>Search engine</i> hanya akan melakukan pencarian sampai sub direktori terendah saja (tidak mengikutsertakan nama file). Contoh: allinurl:digitallibrary |
| 6 | site: | <i>Search engine</i> akan mencari berbagai dokumen yang sesuai dengan kata pencarian dari situs yang ketikan. <i>Path</i> dan nama <i>file</i> tidak dapat dimasukkan. Tambahan terma pencarian harus digunakan. Contoh, notes site:notess.com , akan mencari berapa jumlah dokumen yang telah diindeks atau di daftar Google pada situs notess.com |
| 7 | allinanchor: | <i>Search engine</i> akan mencari berbagai dokumen yang mengandung kata sesuai dengan kata pencarian dimanapun pada <i>links</i> yang merujuk pada dokumen tersebut |
| 8 | related: | <i>Search engine</i> akan mencari berbagai dokumen yang mengandung isi sesuai dengan kata pencarian, dimanapun pada links yang merujuk pada dokumen tersebut |
| 9 | flink: | Fitur ini akan memanggil <i>GoogleScout</i> untuk mencari dokumen yang mirip (serupa) dalam <i>linkage patterns</i> untuk memberikan URL pada sebuah level hirarki yang sama. URL harus diketikan secara tepat. Sebagai contoh related:kompas.com dan related:www.kompas.com akan memberikan hasil temu balik yang berbeda. |

Tabel 4.1. *Field searching* Google

j. Limits

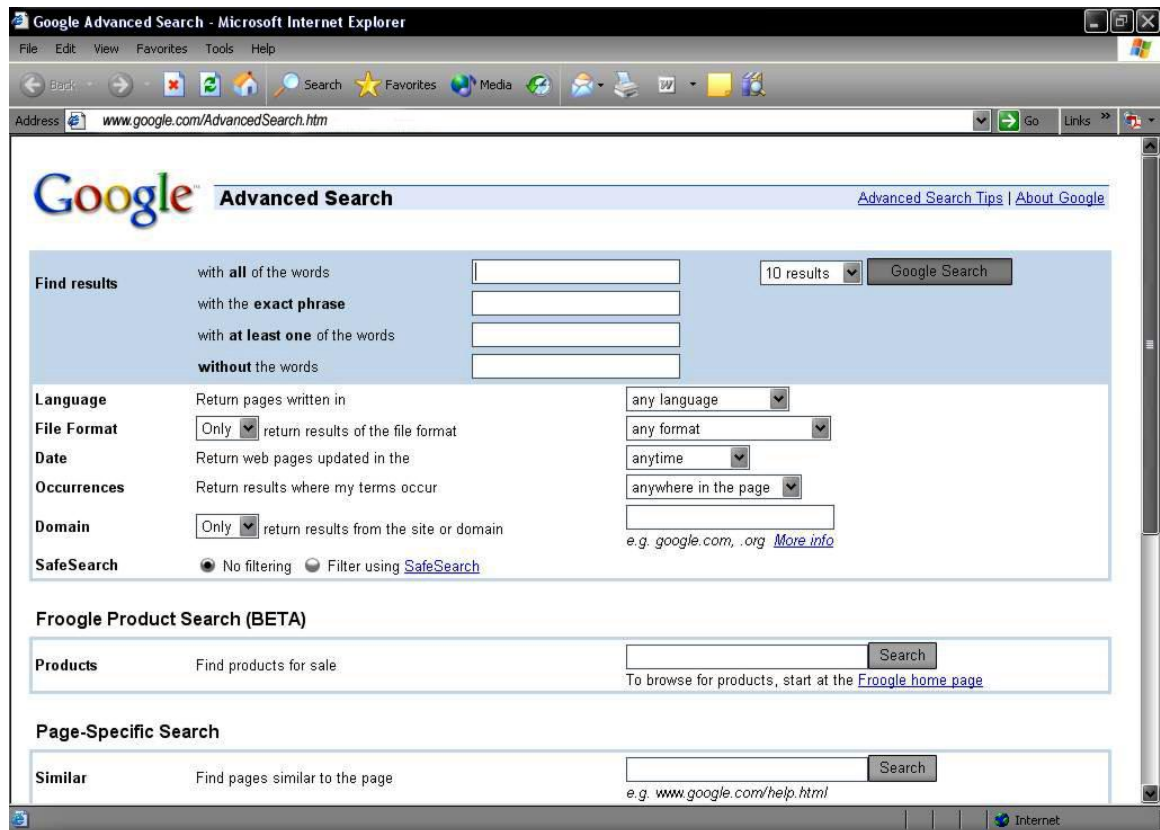
Google mempunyai fitur batasan pencarian berdasarkan bahasa, domain, tanggal (*date*), tipe file (*filetype*), dan konten untuk orang dewasa (*adult content*). Fitur batasan pencarian berdasarkan tanggal (*date limit*), diluncurkan pada bulan July 2001, dan hanya tersedia pada layanan *Advanced Search*.

Hanya tersedia tiga pilihan yang tersedia: *Past 3 Montsh*, *Past 6 Months*, atau *Past Year*. Fitur batasan pencarian berdasarkan tipe file, ditambahkan pada layanan *Advanced Search* bulan November 2001. Layanan *Advanced Search* hanya menawarkan fitur limit tipe file untuk jenis file tertentu yaitu, PDF (.pdf), Word (.doc), Excel (.xls), Power Point (.ppt), dan Rich Text Format (.rtf), dan dengan mengetikan jenis file didepan fitur limit **filetype:file**, fitur limit tipe file ini dapat juga digunakan untuk jenis file *PostScript* (.ps), Text (.txt), .html, *Word Perfect* (.wpd), dan jenis ekstensi file lainnya.

Google memperkenalkan pembatasan pencarian berdasar bahasa (*language limit*) pada bulan April 2000 dengan 11 bahasa yang kemudian dikembangkan lagi pada bulan Agustus 2000 menjadi 24 bahasa, bahasa Rusia ditambahkan pada November 2001, dan sampai saat in terdapat 29 bahasa dan mendekati 86 bahasa grafis antarmuka atau *interface language* (termasuk Bahasa Indonesia) yang tersedia pada layanan *Advanced Search* dan layanan *Language Tools*.

Berikut tampilan *Advanced Search* Google:

Gbr.4.1 Tampilan *Advanced Search* Google



k. Stop Words

Google mengabaikan kata-kata sandang (*stop words*). Namun Google dapat melakukan pencarian terhadap kata-kata seperti *'the', 'of, 'and* dan *'or* dengan menambahkan symbol **+** didepannya. Pada bulan Maret 2000 kata stop words *'the'* tidak dapat dicari walaupun dengan menggunakan tanda **+**. Pastikan selalu menempatkan tanda **+** didepan stop words. Jika tanda **+** didepan kata bukan stop words dalam kata pencarian yang sama, semua tanda **+** akan diabaikan. Stop words yang ada dala sebuah frase pencarian tidak lagi memerlukan tanda **+** karena akan secara otomatis dibaca sebagai *boolean operator* "AND".

l. Sorting

Hasil temu balik (*result*) diurut secara relevan didasarkan atas analisis teknologi *PageRank* milik Google, ditetapkan berdasarkan links dari dokumen lain dengan penekanan yang besar pada situs-situs *authoritative* (berwenang). Dokumen juga di-*cluster* berdasarkan situs (Cluster adalah suatu layanan dimana sering kali dijumpai suatu keadaan hasil temu balik berasal dari sebuah situs yang sama, dan tentu saja bukan seperti yang dikehendaki oleh pengguna. Setiap *search engine* biasanya telah memiliki fasilitas otomatis untuk mencegah hal ini terjadi)

Hanya dua dokumen yang berasal dari situs yang sama yang akan ditampilkan, dengan ditandai masuknya paragraf (*indent*) dibandingkan dengan situs lainnya (memperlihatkan bahwa link tersebut sama atau merupakan bagian dari link yang ditampilkan di atasnya), sedangkan sisanya akan tersedia melalui link [*More results from...*]. Jika pencarian menemukan kurang dari 1.000 results ketika di-cluster dengan dua dokumen tiap situs dan jika dimajukan pada dokumen terakhir, setelah dokumen terakhir akan muncul pesan seperti:

In order to show you the most relevant results, we have omitted some entries very similar to the 63 already displayed. If you like, you can repeat the search with the omitted results included.

Meng-klik opsi "repeat the search" akan menghasilkan lebih banyak dokumen, beberapa diantaranya mendekati atau hampir sama persis dari dokumen-dokumen yang telah ditemukan sebelumnya.

m. Display

Google akan menampilkan judul, URL, sebuah laporan singkat (ekstraksi) yang mendekati kata pencarian, ukuran file, dan jumlah hits, sebuah link merujuk pada *cached copy* pada tiap hasil pencariannya. *Cache copy* adalah berasal dari Index Google dan mungkin lebih lama usianya dari versi sekarang yang tersedia di dunia web. *Cached Copy* akan menampilkan hasil penelusuran dengan kata pencarian yang sejenis namun dalam kurun waktu yang lampau. Jika lebih dari satu macam kata pencarian maka, setiap kata pencarian tersebut akan memiliki warna yang berbeda.

Hasil temu balik akan ditampilkan secara *default* (dalam konfigurasi normal) 10 hits per tampilan, namun pengguna dapat menkostum menjadi 20, 30, 50 atau 100 per tampilan dengan menggunakan layanan *preferences*.

Pada bulan April 2000, Google mulai meng-*cluster* hasil temu balik (results) berdasarkan situs (**Cluster** adalah suatu layanan dimana sering kali dijumpai suatu keadaan hasil temu balik berasal dari sebuah situs yang sama (duplikasi), dan tentu saja hal ini tidak diinginkan oleh pengguna. Setiap *search engine* biasanya telah memiliki fasilitas otomatis untuk mencegah hal ini terjadi). Sebelumnya hits dari situs yang sama akan dicantumkan dibawah yang pertama, dan mulai April 2000 juga, hanya dua *hits* dari situs yang sama akan ditampilkan (situs kedua akan dimasukan dalam paragraph situs pertamanya) dan sisanya dapat diakses melalui link [*More results from hostname*].

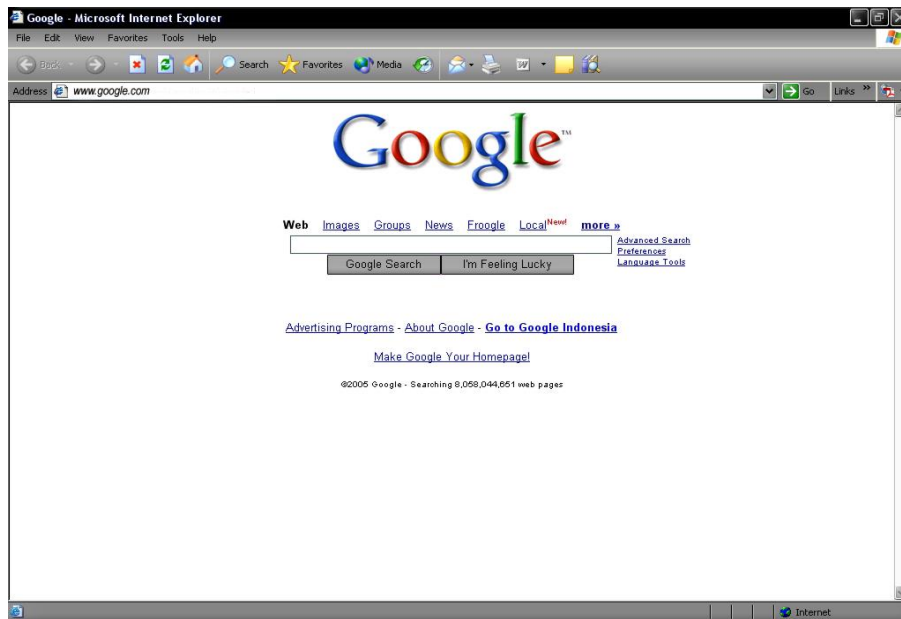
Dengan tujuan untuk file-file non-HTML pada tahun 20001, Google menambahkan dua catatan dalam display untuk mengidentifikasi file-file non-HTML. Sebelum judul dari baris pertama tampilan hasil temu balik, [PDF] atau [PS] atau [XLS] digunakan untuk menyatakan ulang format file yang berbeda. Dan beberapa, pada baris kedua daftar tampilan akan muncul File Format: PDF/Adobe Acrobat – view as text.

Sekitar bulan Agustus 2001, Google mulai me-*refresh* (memperbarui) indeks tertentu (khususnya yang di-*update* secara harian) lebih sering dari index database lainnya. Hal ini ditandai dengan "Fresh!" setelah URL dan size. Pada Desember 2001, tag (label) ini dirubah menjadi indexing date. Pada bulan February 2002, 3 Miliar dokumen web di perbaharui (*refresh*) hampi per harinya.

n. Homepage Google

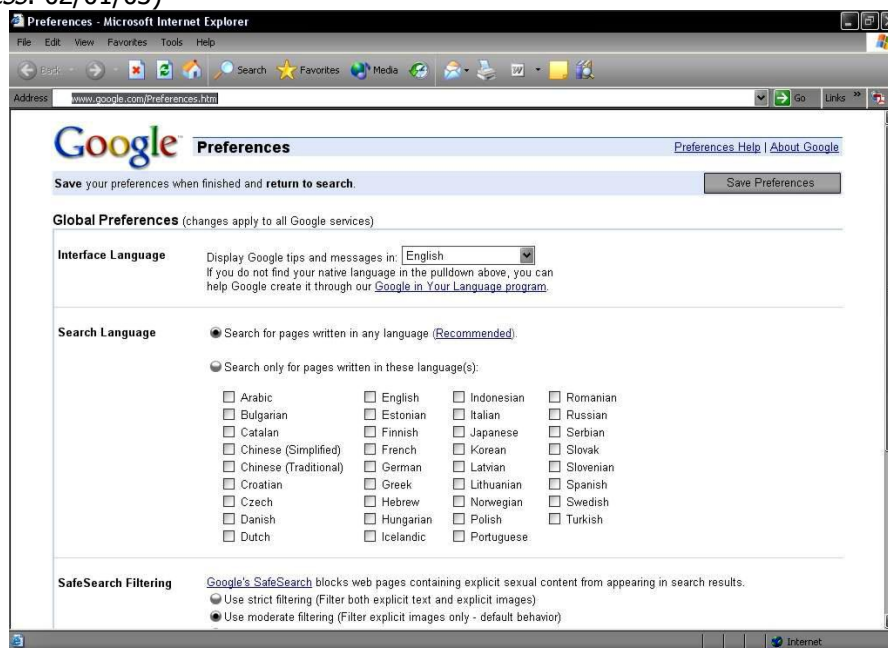
Untuk tampilan muka google kesederhanaan dalam warna dan aksesori menjadi ciri khas utama, sehingga grafis antar muka Google mudah digunakan oleh pengguna awam sekalipun. Kesederhanaan warna dan tampilan Google ini menjadikan fokus pencarian pengguna menjadi tidak kacau.

Gbr. *Homepage* Google



Secara *webdesign* pun Google cukup bagus, tulisan di halaman depan warnanya seimbang. Navigasinya lebih mudah dan cepat, pertama muncul tampilan awa Google kita langsung tahu dimana harus mengetik kata pencarian yang ingin ditemu balik, sementara pada search engine lain kita harus melihat secara keseluruhan dulu untuk mendapatkan gambaran yang lebih jelas, latar belakang yang berwarna putih justru netral dan lebih cerah dibanding latar altavista. Jadi yang ditonjolkan adalah hasil penelusurannya karena *tools search engine* kan memang fungsi utamanya untuk ini.

Tapi kelemahannya adalah tersembunyinya fitur layanan lain yang dimiliki Google, dikarenakan *link* yang tersembunyi, seperti halnya layanan *preference* (untuk mengkustomisasi hasil pencarian), *advanced search* dan *language tools*. Namun secara keseluruhan desain tampilan muka dari Google ini cukup efektif dan simpel terbukti dengan diraihnya penghargaan Pandia Award 2002 sebagai *search engine* terbaik dengan kelebihan utama dalam *User friendly interface* dan dalam menyajikan hasil temu balik yang relevan. (<http://www.pandia.com/post/016.html>, date access: 02/01/03)



Gbr. Tampilan Fitur Preference Google

Saran Bagi Pengguna *Search Engine*

- ✚ Sebagai pilihan utama dalam memilih *search engine* sebagai alat temu balik pergunakanlah Google sebagai pilihan utama, kemudian disarankan menggunakan *search engine* Alltheweb jika anda ingin melimitasi pencarian secara maksimal (bagi pengguna tingkat lanjut).
- ✚ Dalam melakukan pencarian informasi atau temu balik informasi di dunia web hendaknya pengguna:
 - ✓ Merincikan (spesifik) dalam memasukan kata pencarian
 - ✓ Gunakan bantuan *Boolean operator* (AND, NOT, dan OR) serta simbol matematis lainnya (*, tanda petik ganda (" "), tanda kurung atau (*parantheses*) guna menghindari berlimpahnya informasi hasil temu balik.
 - ✓ Pilihlah alat bantu temu balik yang tersedia, apakah direktori atau *search engine*. Direktori sangat membantu untuk memberikan hasil yang lebih relevan jika pengguna sudah mengetahui jenis informasi yang dibutuhkannya. Penggunaan *search engine* sangat bermanfaat dalam memberikan beragam pilihan terhadap informasi yang belum diketahui maupun yang sudah diketahui oleh pengguna.
 - ✓ Masalah klasik dalam temu balik informasi di dunia web adalah menemukan berlimpah informasi (*find too much information*) atau tidak menemukan sama sekali informasi (*find nothing*). Oleh karena itu disarankan untuk bereksperimen dengan kata pencarian (mencari persamaan katanya), dan usahakan gunakan lebih dari satu *search engine* untuk memberikan ragam pilihan.

[]

DAFTAR PUSTAKA

- Agung, Greogorius. 2001. *Panduan Praktis Berinternet Untuk Pemula*. Yogyakarta: Panduan.
- Buick, Joanna and Zoran Jevtic. 1997. *Mengenal Cyberspace; For Beginners*. Bandung: Mizan.
- Campbell, Jane E. 1998. *Kepustakawanan dalam abad informasi; suatu konsep Yang Usang?* Jakarta: British Council.
- Clegg, Brian. 1999. *Mining the Internet; Information Gathering and Research on The Net*. London: Kogan Page.
- Dudenev, Gavin. 2000. *The Internet and the Language Classroom; A Practical Guide for Teachers*. Cambridge University.
- Febrian, Jack. 2001. *Menggunakan Internet*. Bandung: Informatika.
- Houghton, Bernard & John Convey. 1997. *Online Information Retrieval System*. London: Clive Bingley.
- Kovacs, Diane K. 1995. *The Internet Trainer's Guide*. Van Nostrand Reinhold.
- LaQuey, Tracy. 1997. *Sahabat Internet; Pedoman bagi pemula untuk memasuki Jaringan Global*. Bandung: ITB.
- Longkutoy, John. 1983. *Pengenalan Komputer*. Jakarta: Mutiara.
- M. Rogers, Everett. 1986. *Communication Technology; The New Media in Society*. New York: The Free.
- M.Jusup, Pawit. 1995. *Pedoman Praktis Mencari Informasi*. Bandung: Rosda Karya.
- Mieadows, A.J. 1987. *The origin of information science*. London: Taylor Graham.
- Naisbitt, John. 1999. *High Tech High Touch*. Bandung: Mizan.
- Nasution, Zulkarimein. 1989. *Teknologi Komunikasi; dalam perspektif latar Belakang dan Perkembangannya*. Jakarta: FE UI.
- Piliang, Yasraf Amir. 1999. *Sebuah Dunia yang dilipat*. Bandung: Mizan.
- Paice, C.D. 1997. *Information Retrieval and the Computer*. London: MacDonald & Jane's
- Purbo, Onno dan Handoko, Yus Dwi. 1999. *Trik Mengoptimalkan Komputer; Pedoman Bagi Calon Pembeli Komputer Dan Semu Pemakai Komputer*. Jakarta: Elex Media komputindo.
- Rakhmat, Jalaludin. 1997. *Hegemoni Budaya*. Jogjakarta: Bentang Budaya.
- _____ 1992. *Teknik dan Jasa Dokumentasi*. Jakarta: Gramedia.

Salton, Gerald & Michael J McGill. 1983. *Introduction to Modern Information Retrieval*. USA: McGraw-Hill.

Smits, Jan.1991. *Komputer: Suatu tantangan baru di bidang hukum (pengantar)*. Surabaya: Airlangga Press.

Trimo, Soejono.1992. *Pengantar Ilmu Dokumentasi*. Bandung: Rosda Karya.

Majalah Internet. Ed.15 November. 2000: 24

Situs Web:

Gillian Westera Website (Date Access: 06/11/2002):

Westera, Gillian. 2002. *Comparison of Search Engine User Interface Capabilities* [Online]. <http://lisweb.curtin.edu.au/staff/gwpersonal/compare.html>

Westera, Gillian. 2002. *Creating Your Search Strategy*. [Online]. <http://lisweb.curtin.edu.au/staff/gwpersonal/searchtut/how.html>

Westera, Gillian.2002. *Deciding Which Web Search Tool To Use*. [Online]. <http://lisweb.curtin.edu.au/staff/gwpersonal/searchtut/which.html>

Westera, Gillian. 2002. *Discussion of Results*. [Online]. <http://lisweb.curtin.edu.au/staff/gwpersonal/senginestudy/discuss.htm>

Westera, Gillian. 2002. *Glossary of Terms*. [Online]. <http://lisweb.curtin.edu.au/staff/gwpersonal/senginestudy/zgloss.htm>

Westera, Gillian. 2002. *Overview of Search Engines*. [Online]. <http://lisweb.curtin.edu.au/staff/gwpersonal/searchtut/se.htm>.

Westera, Gillian. 2002. *Using the Best Search Engine For Your World Wide Web Research*. [Online]. <http://lisweb.curtin.edu.au/staff/gwpersonal/senginestudy/zindex.htm>

Westera, Gillian. 2002. *Viewing Results and Evaluating Qualit*. [Online]. <http://lisweb.curtin.edu.au/staff/gwpersonal/searchtut/eval.htm>

Google (Date access: 06 Juli 2002)

Google Homepage. 2002. [online]. <http://www.google.com/>

Corporate Information.2002. [online]. <http://www.google.com/corporate/>

Google Corporate Information. 2002. [online]. http://www.google.com/press/overview_biz.html

Google Inc. Company Overview . 2002. <http://www.google.com/press/overview.html>

Our Customers. 2002. [online]. <http://www.google.com/press/customers.html>

Google Press Center: Fact Sheet. 2002. [online]. <http://www.google.com/press/facts.html>

Life of a Query. 2002. [online]. <http://www.google.com/press/query.html>
Google Press Center: Management.2002. [online]. <http://www.google.com/press/management.html>

Lebedev, Alexander. 17 May 1997. *Best search engines for finding scientific information in the web*. [Online]. <http://www.chem.msu.su/eng/comparison.html> [10 November 2001].

Leighton, Vernon, H. and Srivastava, J. 16 June 1997. *Precision among World Wide Web search services (Search engines): Alta Vista, Excite, Hotbot, Infoseek, and Lycos*. [Online]. <http://www.winona.msus.edu/library/webind2/webind2.htm> [Note: Link moved; URL changed 8/5/98 by ald].

Rettig, James. *Beyond cool: Analog models for reviewing digital resources*. [Online]. <http://www.onlineinc.com/onlinemag/SeptOL/rettig9.html> [30 April 1998].

Singh, Amarendra and Lidsky, David. 1996. "All-out search." *PC Magazine* 15(21): 213-249.

Tomaiuolo, Nicholas G. and Packer, Joan G. 1 December 1997. *Quantitative analysis of five WWW "search engines"*. [Online]. [Note: Broken link to <http://neal.ctstateu.edu:2001/htdocs/websearch.html> removed 12/21/98 by ald] [10 November 2002].

Webster, Kathleen and Paul, Kathryn. 1996. *Beyond surfing: Tools and techniques for searching the web*. [Online]. <http://magi.com/~mmelick/it96jan.htm> [10 November 2002].

Steve Lawrence and C. Lee Giles. 1998. *Searching the World Wide Web*. [online]. <http://www.neci.nj.nec.com/~lawrence/science98.html>. [10 November 2002]

Search Engine Showdown (Date access: 10 10 2002)

Notess, R Greg. 2002. *Search Engines Statistics: Database Change Over Time*. [online]. <http://searchengineshowdown.com/stats/change.shtml>

Notess, R Greg. 2002. *Search Engines Statistics: Database Overlap*. [online]. <http://searchengineshowdown.com/stats/overlap.shtml>

Notess, R Greg. 2002. *Search Engines Statistics: Database Relative Size*. [online]. <http://www.searchengineshowdown.com/stats/size.shtml>

Notess, R Greg. 2002. *Search Engine Features by Feature Type*. [online]. <http://www.searchengineshowdown.com/features/byfeature.shtml>

Notess, R Greg. 2002. *Search Engine Features Chart*. [online]. <http://searchengineshowdown.com/features/>

Notess, R Greg. 2002. *Search Engine Showdown Reviews*. [online]. <http://www.searchengineshowdown.com/reviews/>

Notess, R Greg. 2002. *Search Engine Showdown: Size Comparison Methodology*. [online]. <http://www.searchengineshowdown.com/stats/methodology.shtml>

Search Engine Watch (Date access: 10 02 2003)

Sullivan, Danny. 2001. *The Major Search Engines and Directories*. [online]. <http://searchenginewatch.com/links/major.html>

Sullivan, Danny.2002. *Nielsen NetRatings Search Engine Ratings*. [online].
<http://searchenginewatch.com/reports/netratings.html>

Sullivan, Danny.2001. *Nielsen//NetRatings Search Engine Ratings*. [online].
<http://searchenginewatch.com/reports/netratings.html>

Media Metrix Search Engine Ratings. [online].
<http://www.searchenginewatch.com/reports/mediamatrix.html>

Search Engine Sizes. [online]
<http://www.searchenginewatch.com/reports/sizes.html>

European Search Engine Chart. [online].
<http://www.webmasterworld.com/forum18/544.htm>

The Major Search Engines and Directories. [online].
<http://searchenginewatch.com/links/major.html>

[]