

Khai Thác Search Engine Cho Nhu Cầu Học Tập và Nghiên Cứu - bài I

vietsciences-Võ Quang Nhân

ngày 01 tháng 07 năm 2004

Các bạn thân mến,

Trong đời sống ngày nay, một phương tiện không thể thiếu cho việc học tập và nghiên cứu khoa học là biết tận dụng khả năng của tin học. Trong đó, việc sử dụng Internet một cách hiệu quả có thể góp phần giúp các bạn đẩy nhanh tiến độ của các đề án cũng như nó là một công cụ không thể thiếu cho việc tự trao đổi, cập nhật hoá kiến thức và nghiên cứu. Loạt bài này được viết ra nhằm đáp ứng phần nào việc giúp các bạn có thêm kĩ năng để sử dụng đúng mức phương tiện Internet (Hi hi! 'chat' thường không có ích gì ngoại trừ dùng để tìm bạn 4 chiều!), đặc biệt là các máy truy tìm dữ liệu (search engine). Bài giảng sẽ bao gồm nhiều phần, trong phần kết chúng tôi sẽ trình lại tất cả tài liệu tham khảo chính mà tác giả đã dùng để viết nên loạt bài này. Các tài liệu này sẽ rất hữu ích nếu như các bạn có thì giờ và trình độ Anh ngữ để nghị hãy đọc thêm chúng để có dịp hiểu chi tiết thêm những chi tiết mà chúng tôi không đề cập đến.

1. Search engine là gì: - Mục đích của phần này là trình bày kiến trúc căn bản của các search engine. Có nắm bắt được nguyên lý hoạt động của các search engine thì việc sử dụng sẽ có nhiều hiệu quả và nhanh chóng hơn.

.1.1 Khái quát về search engine

Search engine (tạm dịch - máy truy tìm) -nguyên thủy- là một công cụ phần mềm nhằm tìm ra các trang trên mạng dựa vào các thông tin mà nó có. Dữ liệu thông tin của search engine thực chất là một loại cơ sở dữ liệu (database) cực lớn. Công cụ này tìm các tài liệu dựa trên các từ khoá (keyword) và trả về một danh mục của các trang có chứa từ khoá.



Trang Google.com

1.1.1 Spider:

Cơ sở dữ liệu của các search engine được cập nhật hoá bởi các chương trình đặc biệt thường gọi là "robot", "spider" hay "Webcrawler". Các chương trình này sẽ tự động dò tìm và phân tích từ những trang có sẵn trong cơ sở dữ liệu để kiểm tra các nối kết (links) từ các trang và trở lại bổ xung dữ liệu cho các search engine sau khi phân tích. Các chương trình này cũng sẽ báo cáo về các liên kết đã bị đào thải. Từ khoá mà bạn bỏ vào là để cho search engine kiểm tra trong bảng chỉ số của nó. Kết quả đúng nhất sẽ được xếp ở thứ tự đầu tiên. Trang nổi tiếng nhất là dùng nguyên tắc này là www.google.com

1.1.2 Meta-search engine:

Trên đây là các loại search engine "nguyên thủy". Ngày nay, người ta còn lợi dụng các search engine sẵn có để thiết kế thành một loại search engine mới còn gọi là meta-search engine (tạm dịch máy truy tìm ảo). Nguyên tắc của loại search engine này rất đơn giản. Nó không có cơ sở dữ liệu. Khi hoạt động thì nó sẽ gửi từ khoá đến các search engine khác Một cách đồng loạt và nhận về tất cả các kết quả tìm được (Ui chao! sao giống ... mấy con kí sinh trùng quá !). Và nhiệm vụ của nó chỉ là phân tích hay sắp xếp lại các tài liệu tìm được cho thân chủ. Cái hay của loại meta-search engine là lợi dụng cơ sở dữ liệu của các search engine nguyên thủy để tìm ra nhiều kết quả hơn. Nhưng vì loại meta-search engine này chỉ tồn tại nếu có các search engine nguyên thủy nên gọi là meta- (tiếp đầu ngữ meta thường dịch là 'siêu hình' hay 'ảo'). Điển hình

là <http://www.metacrawler.com/>



Một phần trang www.metacrawler.com

Ngoài hai loại search engine này người ta còn đề cập đến

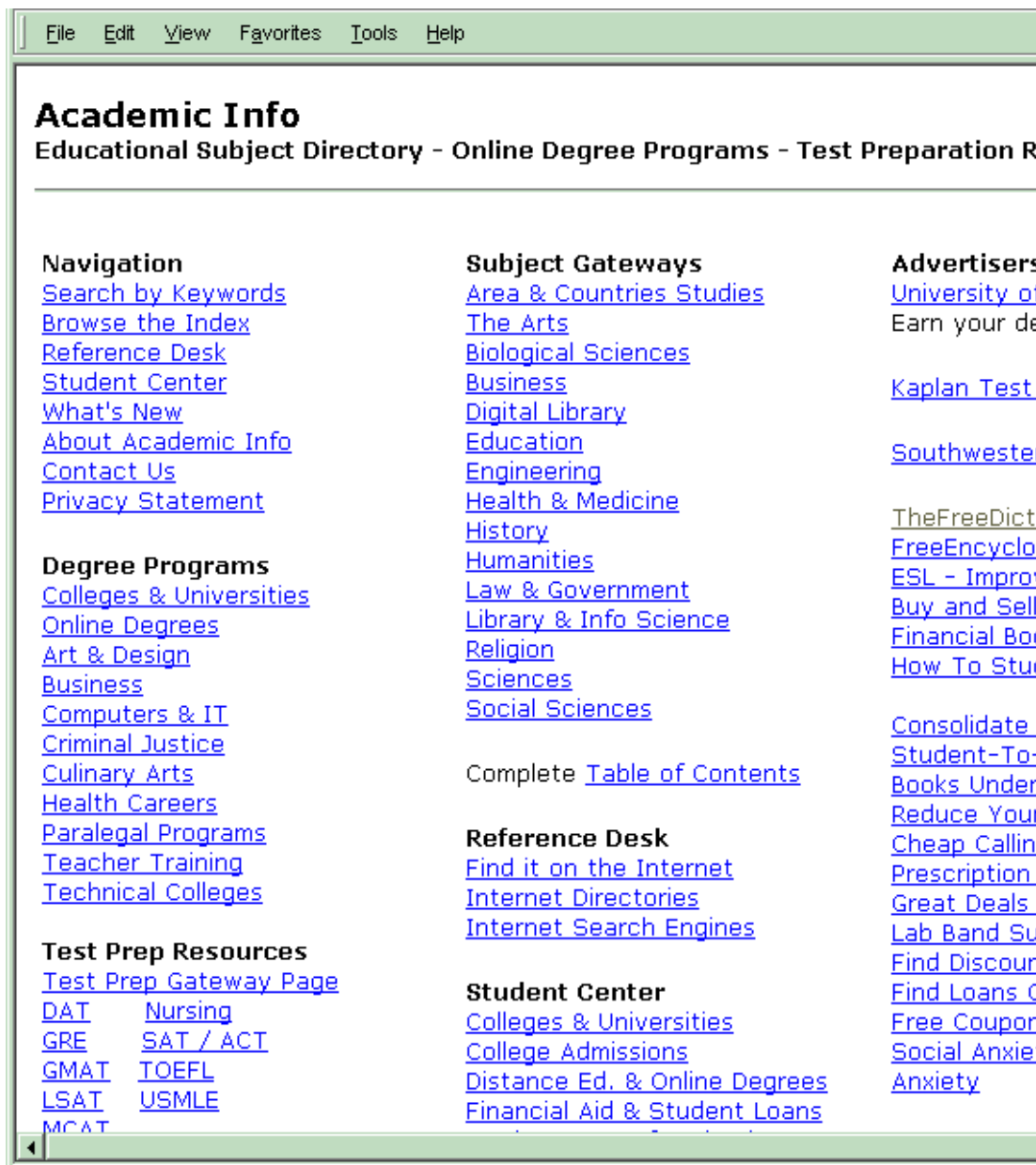
1.1.3 Thư mục đối tượng (Subject Directories):

Còn gọi là máy truy tìm theo phân lớp (hierarchical search engine) - search engine kiểu này sẽ phân lớp sẵn các đối tượng vào các thư mục và người dùng sẽ lựa rẽ nhánh từ từ cho đến khi tìm ra các trang WEB mà mình muốn. Kiểu này dễ cho người tuy cập nhưng có điểm yếu là nó không thể bao gồm hết mọi chủ đề mà mình muốn kiếm ra. Hơn nữa, sự phân loại đôi khi không được đầy đủ và chính xác. (Ngoài ra, nó còn chứa cả đồng trang 'shopping' chỉ phù hợp cho ... phái nữ hì hì). Điển hình của loại này là www.yahoo.com

Chúng ta sẽ không bàn sâu loại thư mục đối tượng.

1.1.4 Các cơ sở dữ liệu đặc biệt:

Đặc điểm của loại này là dữ liệu kiếm ra không thực sự có được địa chỉ trang WEB cụ thể qua các search engine; dữ liệu này tồn tại trong các cơ sở dữ liệu của một computer ở đâu đó mà các trang WEB được phép sử dụng. Đặc biệt, các trang WEB nghiên cứu của các ĐH hay học viện như <http://lii.org/> , <http://www.academicinfo.net/> , và <http://infomine.ucr.edu/> (Xin nhớ cho đây là cơ sở dữ liệu của các cơ quan nghiên cứu đó nha!)



Một Phần của Trang Academicinfo.net

Ngoài ra, lợi dụng google hay yahoo ta cũng có thể tìm ra các tài liệu trong các cơ sở dữ liệu bằng cách thêm vào từ khoá chữ: "database" - Nghĩa là: thay vì dùng từ khoá "*helium laser*" thì viết thành "*helium laser database*".

Các trang WEB mô tả như trên gọi là invisible WEB (tạm dịch 'vô kiến WEB')

1.2 Đặc điểm và phân loại search engine:

1.2.1 Các đặc điểm:

Như vậy, tùy theo keyword mà có thể tìm số trang liên hệ ít hay nhiều. Trong nhiều trường hợp, số trang tìm ra có đến hàng triệu trang thì vấn đề đặt ra cho một search engine là làm sao xếp thứ hạng các trang để người dùng có nhiều cơ hội tìm ra dữ liệu mình muốn nhất?

Cách thức xếp thứ hạng (ranking):

Hầu hết các search engine xếp hạng ưu tiên theo kiểu mà người design nghĩ rằng trang tìm thấy là phù hợp với từ khoá và các điều kiện cho thêm của người dùng máy (chẳng hạn như máy bên Pháp thì trang tiếng Pháp sẽ có ưu tiên hơn và tùy theo cách cho từ khoá đủ "hẹp" trong thị trường tìm kiếm). Mặc dù vậy, khi chưa quen dùng chúng ta sẽ "tìm" được cả đống những cái liên kết (links) không vừa ý và nhiều lúc ...chúng hoàn toàn vô dụng. Lý do chính có thể là do người dùng đã không hiểu rõ kĩ thuật tìm kiếm của máy và ngược lại cũng có thể là do kĩ thuật hiện tại của máy truy tìm chưa đủ sức để "hiểu" người dùng muốn gì! (bởi vậy mới phải tả rõ hơn cho 'nàng' search engine hiểu rằng 'chàng' muốn gì)

Đa số cách xếp thứ hạng là dựa vào "số lần xuất hiện" (còn gọi là *tần xuất*) của từ khoá trong 1 trang. Nghĩa là trang có xuất hiện từ khoá nhiều nhất sẽ được hiển thị trước tiên khi search engine trả lời. (Người thiết kế search engine cho rằng tài liệu nào lập lại nhiều lần cái từ khoá thì chắc nó sẽ hữu dụng nên được xếp ngay ở hàng đầu tiên của trang trả lời!).

Bởi vậy nếu bạn tra tìm trang WEB mà lại dùng từ khoá chung chung hoặc từ có nhiều nghĩa thì hầu như bạn sẽ không tìm được những trang chứa cái mình muốn. Trong trường hợp này hoặc bạn phải tìm bộ từ khoá khác hơn/chi tiết hơn hoặc dùng những kĩ thuật hỗ trợ từ search engine mà sẽ đề cập trong các phần sau của bài viết.

Có một số search engine ngoài việc xác định *tần xuất* của từ khoá chúng còn dựa vào khả năng xuất hiện sớm của từ khoá này trong một hồ sơ như là tìm thấy trong tựa đề (title), trong đầu trang (header),... Hoặc giả, họ xác định sự quan trọng của trang WEB bằng cách ... đếm số trang có liên kết từ các trang khác tới nó (Một trang WEB có hay có quan trọng thì tui mới cho 'link' và 'reference' trên trang WEB của tui chớ sao !)

Cũng vì lí do này, đa số các search engine lớn đều có thêm chức năng là "advanced" hay "advanced search" để giúp bạn tìm ra cái bạn cần dễ hơn.

1.2.2 Phân loại search engine:

Phân loại theo kiến trúc hoạt động:

Theo phân định của Berkeley thì có 3 loại công cụ tìm kiếm (search tool): các search engine tồn tại dựa trên chương trình spider, các search engine theo thư mục đối tượng, và invisible WEB

Phân loại theo chức năng:

Theo cách phân loại này thì tùy theo đối tượng tìm kiếm mà có: Kiểm địa

chỉ trang WEB, kiểm địa chỉ e-mail, hay kiểm tin tức về 1 người (<http://www.nwbuildnet.com/nwbn/fouremail.html>) hay một tổ chức (<http://www.freeality.com/findrev.htm>), kiểm việc làm, kiểm bản đồ (www.mapquest.com) ...

Chúng ta sẽ không đề cập sâu hơn trong phần này. (xem như là bài tập vậy mà ... Có muốn đi tìm người yêu cũ thì tự lo kiểm lấy chứ sao nhờ tui!).

1.2.3 Các Search Engine nên biết:

- www.google.com Được xem là search engine kiểu spider tốt nhất (4 lần thắng do Search Engine Watch lựa chọn) Bằng google, bạn có thể dùng để tìm ảnh, các bản luận đang có trên Usenet, và newsgroups (tạm dịch -- các nhóm tin tức). Nó cũng cung cấp chức năng kiểm lỗi chính tả, tra cứu các từ điển (hầu hết các bài viết của tác giả đều mượn trang này để tra từ điển) tìm giá chứng khoán, bản đồ đi đường, số điện thoại và ..vv
Google đầu tiên là một đề án từ ĐH Stanford của các sinh viên Larry Page và Sergey Brin lúc đó gọi là BackRub. Năm 1998, tên này đổi thành Google và dự án đã được tung ra thị trường trở thành một công ty tư nhân tên Google cho đến nay.
- www.yahoo.com Xuất hiện năm 1994 ban đầu trang này thiết kế theo phương pháp thư mục. Cho đến tháng 11 năm 2002 yahoo đã thêm vào chức năng spider lấy từ Google (để tăng sức cạnh tranh) cho đến tháng 2 2004. Hiện nay yahoo đã tự có kĩ thuật tìm kiếm độc lập.
Đây là trang "số 1" cho những người thích mua sắm vì cách phân loại theo đối tượng sẽ dễ cho người tiêu dùng tìm đến sản phẩm cần thiết.
Kĩ thuật spider của yahoo là sự kết nối kĩ thuật của Altavista, AllTheWeb, và của Inktomi (một đề án search engine phát khởi từ UC Berkeley)
- www.askjeeves.com Khởi sự từ năm 1998 đến 1999 thì trở thành search engine có "ngôn ngữ tự nhiên": Nó cho phép bạn đặt câu hỏi (dĩ nhiên tiếng Anh hoặc các thứ tiếng mà nó hỗ trợ) và trả lời bạn tất cả những gì dường như có thể là đúng. Thật ra, đằng sau bức màn "kĩ thuật cao" này là công ty đã có khoảng một trăm chuyên viên có nhiệm vụ khảo sát các ghi nhận khi tìm kiếm và từ đó tìm ra những gì dường như là những trang đúng nhất cho sự tìm kiếm đó. Ask Jeeves còn hỗ trợ thêm vào đó một chức năng gọi là "smart search" (tạm dịch dò kiếm tinh khôn) chức năng này cung cấp thêm một sự phân loại tối thiểu về loại đối tượng mà bạn muốn tìm (hình ảnh, phim, nhạc,) bằng cách chọn các biểu tượng và do đó cho ra kết quả chính xác hơn nhiều.
- www.vivisimo.com -- meta-search engine
- www.surfwax.com -- meta-search engine

- www.copernic.com -- meta-search engine
- www.metacrawler.com - meta-search engine
- <http://lii.org/> -- adcademic search database
- <http://www.academicinfo.net/> -- adcademic search database
- <http://infomine.ucr.edu/> -- adcademic search database

Trong bài kì tới chúng ta sẽ khảo sát sâu hơn về các chức năng nâng cao và các kĩ năng cần thiết để tận dụng các search engine.

©Vietssciences-Võ Quang Nhân Ngày 01 tháng 07 năm 2004

Khai Thác Search Engine Cho Nhu Cầu Học Tập và Nghiên Cứu - bài II

vietsciences-Võ Quang Nhân ngày 19 tháng 07 năm 2004

2. Các hỗ trợ nâng cao và các phép toán thông dụng trong máy truy tìm: Bài giảng phần này nói nhiều về các đặc điểm chuyên biệt của các search engine dựa trên các đặc điểm của chúng mà các bạn có thể có nhiều phương pháp để tìm, tra cứu tài liệu một cách hữu hiệu hơn. Chúng tôi sẽ mô tả chi tiết các khả năng chuyên biệt của các thiết bị truy tìm

2.1 Các Phép Toán đơn giản: các phép này dùng để nâng cao hiệu quả của việc tìm kiếm. Trong hầu hết các trang truy tìm, việc dùng các dấu hỗ trợ +, - và ngoặc kép đều thuận tiện.

2.1.1 Dùng dấu '+': Khi bạn muốn tìm các trang có mặt tất cả các chữ mà bạn muốn không theo thứ tự thì hãy viết nối các chữ này với nhau bởi dấu + (và nhớ chừa khoảng trống giữa các chữ)

Thí dụ muốn tìm trang nói về cách thức viết Linux scripts bạn có thể điền vào ô tìm bộ từ khoá:

+Linux +script +tutor



Hình1: tự học Linux script bằng ... search engine

Làm vậy thì chỉ có những trang nào có đủ các chữ đã nêu mới được tìm ra. Bằng cách này bạn có thể lọc bớt được một số lớn các trang không cần tìm

2.1.2 Dùng dấu '-': Nhiều lúc bạn muốn loại bỏ bớt các trang có một (hay nhiều) chữ mà bạn không muốn có thì dùng dấu - trong trường hợp này

Chẳng hạn tìm tin tức về các loại xe dùng kỹ thuật lai mới chưa có bán trên thị trường nhưng không muốn các trang bán xe lọt vào thì có thể thử từ khoá:

car +hibrid -sale -Prius -Insight

(Prius and Insight là hai kiểu xe hibrid đã có bán trên thị trường của Toyota và Honda)

2.1.3 Dùng ngoặc kép " ": để tìm cụm từ trong nguyên văn

Nhiều khi bạn muốn tìm bài viết nguyên văn của một câu nói, tên của một người hay một bộ cụm từ thì có thể để tất cả vào trong ngoặc kép. Phương pháp này rất tiện lợi cho việc tìm kiếm những trang đặc biệt (thí dụ hãy thử đánh nguyên văn một câu thơ "Quả cau nho nhỏ miếng trầu ôi" vào trong google tìm xem ai đã viết câu này? Hè hè! như zây mà đi thi thả thơ là tui thắng chắc hết 99% còn một

phần là ... xui hông ai đang bài trên In-tờ-nét)



Hình2: 0.14 giây đã tìm ra tác giả của câu "quả cau nho nhỏ"!

Để tìm ra các tài liệu viết về nhà khoa học Von Neumann thì gõ nguyên tên "**von neumann**"

Chú ý: Phương pháp này rất hiệu quả nhưng nếu bạn đánh sai chính tả thì ... kể như "bán lá giống" (hãy thử đánh tìm chữ "**Quã cau**" xem tui nói có đúng hông!)

2.2 Các phép toán Boolean

Các mệnh lệnh truy tìm dùng đại số boolean đã được phát triển từ nhiều năm. Mặc dù vậy, chúng đã khó xử dụng so với những trình độ người trung bình. Hầu hết các search engine đều có chấp nhận các mệnh lệnh kiểu này. Tuy nhiên, hầu hết các trường hợp thì bạn có thể tìm ra những gì cần mà không phải xài tới chúng

Các phép toán thông dụng thường được các search engine hỗ trợ là **OR**, **AND**, **NOT** và **NEAR**

Lưu ý: khi dùng mệnh lệnh Boolean thì các toán tử phải viết bằng chữ hoa.

2.2.1 Phép **OR**

Lệnh này cho phép *tìm những trang WEB nào có mặt 1 trong các thành tố (hay còn gọi là toán tử của phép toán OR) của bộ từ khoá.*

Thí dụ để tìm các bài viết về Nguyễn Trãi trong cả tiếng việt và tiếng nước ngoài thì có thể dùng bộ từ khoá

"Nguyễn Trãi" OR "Nguyen Trai"

Xa hơn nữa, một số các search engines sẽ dùng phép toán OR như là phép toán mặc định (nghĩa là nếu tui gõ **"Nguyễn Trãi" "Nguyen Trai"** thì kết quả. Chẳng hạn trường hợp của Altavista thì sẽ tìm ra ngay cả những bài viết ... không dấu)

Lưu ý: AOL search engine không làm tốt khả năng truy tìm khi dùng toán tử OR. Trong khi google sẽ không hoạt động để tìm những cụm từ khác nhau bởi phép toán OR

Các trang có thể dùng OR là: AltaVista, AOL Search, Excite, Google, Inktomi (HotBot, MSN), Ask Jeeves, Lycos, Northern Light, HotBot, và Gigablast

2.2.2 Phép **AND**

Phép toán AND *nhằm yêu cầu search engine truy tìm các trang có sự hiện diện của tất cả thành tố.* Thí dụ

"space craft" AND "health"

sẽ truy tìm các trang có chữ health và chữ "space craft"

Một số trang truy tìm sẽ dùng AND như là mặc định (trong đó có google). Bạn cũng có thể thay thế bằng cách dùng dấu + trong một số trường hợp nào search engine không có chức năng boolean

Các trang có thể dùng AND là: AltaVista, AOL Search, Excite, Inktomi (HotBot, MSN) Northern Light, Yahoo, và Gigablast

2.2.3 Phép **NOT**:

Phép này hoàn toàn tương tự như cách dùng dấu -. Nghĩa là, *sự truy tìm sẽ loại bỏ những trang có thành tố đi cùng với phép toán NOT.*

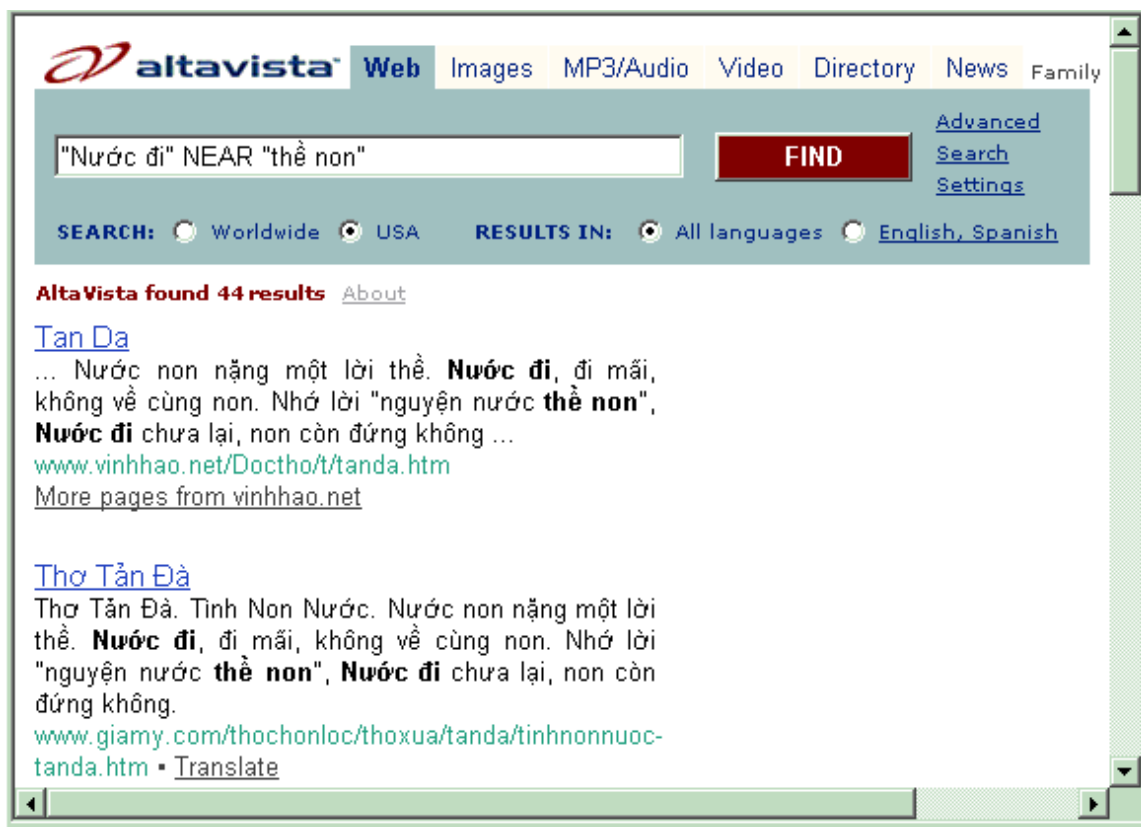
Các trang có thể dùng NOT là AOL Search, Excite, Inktomi (HotBot, MSN), Northern Light và Gigablast

2.2.4 Phép **NEAR**

Dùng để *truy tìm những trang WEB có các thành tố của từ khoá nằm gần nhau.* Phép toán này rất có lợi để tìm ra những trang có một cụm từ, một khái niệm, một định nghĩa hay một lời phát biểu mà bạn không nhớ hết được nguyên văn. Ví dụ:

Tìm lại nguyên văn câu thơ và tác giả bằng bộ từ khoá:

"Nước đi" NEAR "thề non"



Hình 3: Không nhớ hết câu, vẫn tìm ra cụ Tấn qua Altavista

Các trang cho dùng NEAR là AltaVista (10 words), AOL Search (specify number).

2.2.5 Chẻ nhánh bằng phép ()

Dùng ngoặc đơn cho phép ta tìm nhiều kết hợp phức tạp. Thí dụ:

bootable AND (CD OR CDROM OR CD-ROM) AND (howto or instruction)

(không chơi mấy cái ví dụ... thơ với thần nữa vì sợ bạn đọc là phái nữ chê tui 'yêu sắc' chỉ biết có ba bài thơ cổ thì chết)

altavista Web Images MP3/Audio Video Directory News Far

bootable AND (CD OR CDROM OR CD-ROM) AND (hc **FIND** [Advanced S](#)
[Settings](#)

SEARCH: Worldwide USA RESULTS IN: All languages [English, Spanish](#)

AltaVista found 12,100 results [About](#)

[Step-by-step Construction of a **Bootable** Rescue **CD-ROM**](#)

Caution: These instructions are not to be used without understanding them! You must be the root user for many of these operations and that means you can really mess up your system if you are careless or hasty. ... The Bootdisk-**HOWTO**, **CD-Writing HOWTO**, man mkisofs ... El Torito" style of **bootable CD-ROM** contains an implanted image ... 1024 Feb 17 23:27 **cdrom** drwxr-xr-x 2 root ...
ourworld.compuserve.com/homepages/KanjiFlash/BD-CDROM.htm
[More pages from ourworld.compuserve.com](#)

[The Linux **CD-ROM HOWTO**](#)

The Linux **CD-ROM HOWTO**. Jeff Tranter. tranter@pobox.com. v1.17, 18 July 2001. Revision History. Revision 1.17. 2001-07-18. Revised by: jjt. Merged in some questions from the no longer maintained ATAPI/IDE **CD-ROM** FAQ by Mathew Kirsch. ... Polish: <http://www.jtz.org.pl/Html/CDROM-HOWTO.pl.html> ... from a **CD-ROM instruction** manual. Don't Panic ...
www.briandowney.net/linux/docs/linuxdoc/CDROM-HOWTO.html
[More pages from briandowney.net](#)

Hình4: Tìm phương pháp làm CD tự khởi động bằng bộ từ khoá có ()

Hỗ trợ cho kiểu phân nhánh bằng ngoặc đơn là AltaVista, AOL Search, Excite, Inktomi (MSN), Northern Light

Các lưu ý:

- Trong mọi trường hợp thì từ khoá sai chính tả sẽ không thể có hiệu quả
- Cách tốt nhất là dùng chữ in hoa cho các phép toán
- Các trang tìm kiếm của Hotbot hay MSN thì bạn phải chuyển sang chọn chức năng "Boolean phrase" khi dùng các phép toán Boolean
- Trang Lycos (www.lycos.com) có thông báo là họ trợ giúp các phép toán boolean nhưng trong thực tế, chức năng này không hoạt động đúng và hiệu quả

2.3 Các hỗ trợ nâng cao khác:

Ngoài ra, nhiều search engine còn hỗ trợ thêm các từ khoá qui ước. Khi dùng các từ khoá này thì search engine sẽ chuyên biệt hoá các trang WEB, truy tìm theo ý nghĩa qui ước mà từ khoá biểu tượng. Với các hỗ trợ này bạn có thể kiểm soát được các loại trang nào mà bạn muốn truy tìm

Lưu ý: Đằng sau các từ khoá qui ước đều có dấu hai chấm ':'. Khi viết bộ từ khoá có các từ khoá qui ước này thì tốt nhất là viết từ tìm kiếm ngay liền sau dấu ':' và không chừa khoảng trống (space) nào (thí dụ ta viết từ khoá

link:<http://cnn.com>

thay vì viết

link: <http://cnn.com>)

2.3.1 Các từ khoá *host:*, *site:*, *url.host:*, và *domain:* Dùng để thu hẹp các trang WEB tìm ra

2.3.1.1 Từ khoá *host:* của Alavista(www.altavista.com): Từ khoá này sẽ chỉ cho phép search engine tìm các trang nằm trong một WEB server (host) mà bạn muốn tìm. Thí dụ

host:mars.jpl.nasa.gov mars saturn

Bộ từ khoá trên chỉ tìm trong mars.jpl.nasa.gov tất cả các trang có chứa chữ mars và chữ saturn

Kết hợp với các lệnh khác bạn có thể tìm theo cách chuyên biệt:

"carbon nanotech" -host:www.technologyreview.com

Lệnh trên cho phép tìm tất cả các trang nào có chứa cụm từ "carbon nanotech" ngoại trừ các trang xuất sứ từ www.technologyreview.com (Bởi vì muốn đọc bài từ trang này thì phải trả tiền đăng kí nên tui ...bảo thủ)

Hay là:

"heart transplan" +host:edu

mệnh lệnh trên sẽ tuy tìm các trang có đuôi gốc là .edu (vì tui biết rằng các trang có tận cùng edu thường là các trang của trường đại học)

Như là 1 bài tập, bạn hãy tìm tất cả các trang xuất phát từ vương quốc Anh viết về cloning (Hà hà! Anh là một trong những nước dẫn đầu về kĩ thuật cloning)

2.3.1.2 Từ khoá tương tự *site:*, *url.host:*, và *domain:*

Làm chức năng hoàn toàn tương tự với từ khoá **host:** của Altavista thì:

- Từ khoá **site:** được dùng bởi Excite, Google (Netscape, Yahoo)
- Từ khoá **url.host:** dùng bởi AllTheWeb
- Từ khoá **domain:** dùng bởi Inktomi (HotBot, iWon, LookSmart)

2.3.2 Các từ khoá *title:*, *intitle:*, và *allintitle:* Dùng để tìm các trang có tựa đề chứa từ khoá. Tùy theo loại search engine mà chúng ta dùng từ khoá khác nhau. Điều này tiện lợi cho việc tìm kiếm tựa đề (hay các bài viết có chủ đề tập trung) các bài viết. thông thường các bài nghiên cứu được đăng tải sẽ có tựa đề rất rõ

ràng. Do đó, nếu dùng cách này cũng có thể tìm ra nhiều tài liệu. Ngoài ra, dùng cách này bạn có thể tìm lại được các trang mà bạn chỉ nhớ một hay một phần tựa đề của nó.

2.3.2.1 Từ khoá *title*: Có thể dùng trong các trang WEB: [AltaVista](#), [AllTheWeb](#), [Inktomi \(MSN và HotBot\)](#). Để tìm các trang có tựa đề chứa 1 từ hay 1 cụm từ đặc biệt (trường hợp cụm từ thì bạn hãy để trong ngoặc kép) Thí dụ:

title: "Mars Landing"

sẽ giúp tìm các trang nói về Hoả tinh

Trong trường hợp bạn muốn tìm nhiều hơn một chữ nằm trong cùng 1 tựa đề bài thì có thể dùng hai lần từ khoá title. Thí dụ:

title Mars title: water

2.3.2.1 Từ khoá *intitle*: và *allintitle*: Cho phép dùng trong các trang: [google](#) và [teoma](#).

- Từ khoá **intitle**: tương tự như từ khoá title của altavista để tìm một từ hay 1 cụm từ để trong ngoặc kép. Ví dụ để so sánh với 'title':
intitle:"mars landing"
- Từ khoá **allintitle**: được dùng để tìm tựa các bài mà có chứa nhiều hơn 1 từ mà bạn muốn tìm. Ví dụ để so sánh với việc dùng nhiều lần chữ 'title':
allintitle:Mars watwer

The screenshot shows a Google search interface. At the top, the Google logo is on the left, and navigation links for 'Web', 'Images', 'Groups', 'News', 'Froogle', and 'more »' are on the right. Below the logo is a search bar containing the text 'allintitle: mars water' and a 'Search' button. The search results are displayed under the heading 'Web'. The first result is titled 'The Case of the Missing Mars Water' and includes a snippet about the debate on Mars water, with a link to 'science.nasa.gov/headlines/y2001/ast05jan_1.htm'. The second result is titled 'NASA announces discovery of evidence of water on Mars' and includes a snippet about the discovery, with a link to 'www.space.com/scienceastronomy/solarsystem/mars_water_000620.html'. The third result is titled 'Water On Mars Could Sustain Human Colonies' and includes a snippet about an advertisement, with a link to 'www.space.com/scienceastronomy/solarsystem/mars_futurist_000622.html'. The fourth result is titled 'BBC NEWS | Science/Nature | Mars rocks once 'water drenched'' and includes a snippet about the discovery, with a link to 'news.bbc.co.uk/2/hi/science/nature/3524275.stm'.

Hình 5: Dùng *allintitle*: tìm trang có tựa bao gồm chữ Mars và water trong google

2.3.3 Các từ khoá *inurl*, *allinurl*, *originurl*, và *u*: Các từ khoá này dùng để tìm những địa chỉ trang WEB mà có chứa c'ac chữ cần tìm. (Rõ ràng là việc này có lợi cho những ai thích mò tới các địa chỉ khác nhau...nhưng có cùng 1 tên)

2.3.3.1 *inurl*: và *allinurl*: trong google:

- Cũng như trên muốn tìm địa chỉ các trang WEB có một chữ đặc biệt thì dùng *inurl*. Thí dụ:
inurl:nasa
- Nếu bạn tuy tìm một địa chỉ có nhiều hơn 1 chữ thì dùng *allinurl*. Thí dụ:
allinurl:vietnam thetholucbat

Lưu ý: trang google chỉ có thể tìm ra nếu như bạn gõ nguyên 1 bộ phận của từ (nghĩa là nếu trong ví dụ trên bạn gõ thành ***allinurl vietnam lucbat*** thì bạn sẽ thất bại trong việc tìm kiếm

Ngoài ra, trong google bạn có thể lạm dụng từ khoá này để tìm các trang có đuôi

file riêng biệt. chẳng hạn như:

mars rover inurl:pdf

sẽ giúp tìm các trang có dạng đuôi tệp là pdf mà các trang này chứa từ khoá mars và chữ rover

2.3.3.2: *originurl*, *u*: và *url*:

- Việc tìm địa chỉ trong Inktomi (AOL, GoTo, HotBot) sẽ thông qua từ khoá **originurl**:
- Trong Yahoo thì dùng từ khoá **u**:
- Trong excite (www.excite.com) dùng **url**:

2.3.4 Từ khoá *Link*: và *linkdomain*: Dùng để tìm các trang có dòng liên kết tới trang mà được ghi trong từ khoá.

- Từ khoá link dùng được trong [google](#) và trong [yahoo](#). Tuy nhiên, khi dùng yahoo bạn phải cho đủ tên trang WEB với tiền tố **http://**. Thí dụ: từ khoá **link:vietsciences.free.fr** sẽ tìm được một số trang bởi google. Trong khi đó, nếu đánh **link:http://vietsciences.free.fr** thì sẽ tìm ra nhiều kết quả hơn trong yahoo
- Từ khoá **linkdomain**: được dùng để tìm các liên kết nếu dùng [MSN](#).

2.3.5 từ khoá *filetype*: Khi cần tìm các tư liệu nằm dưới các dạng tệp khác nhau thì từ khoá filetype: sẽ giúp đỡ ít nhiều. Tuy nhiên, tùy theo trang mà chúng ta truy kiếm sẽ có các giới hạn khác nhau.

- Trang [google](#): sẽ hỗ trợ truy tìm các kiểu tệp: PDF, Word (.doc), Excel (.xls), PowerPoint (.ppt), và Rich Text Format (.rtf) cũng như là PostScript (.ps), Text (.txt), HTML (.htm hay .html), WordPerfect (.wpd), và các file extensions khác ...
ví dụ: **laser filetype:pdf** sẽ giúp tìm các trang kiểu đuôi .pdf
- Trang [yahoo](#) cho phép tìm HTML (htm hay html), PDF, Excel (.xls), PowerPoint (.ppt), Word (.doc), RSS/XML (.xml), and Text Format (.txt)
- Trang [MSN](#) chỉ chuyên biệt tìm các loại tệp: HTML, PDF, PowerPoint, Word, or Excel

Lưu ý: Dưới "con mắt" của các máy truy tìm thì các tệp có đuôi .htm khác với các tệp có đuôi .html. Đó đó, nếu muốn tìm một cách chắc chắn tất cả các tệp dạng HTML thì nên tìm làm hai lần một riêng cho htm và một cho html

2.3.6: Dùng các loại *kí tự phòng định* (wildcard character): Các dấu này được hiểu tương tự như khi ta dùng lệnh có dấu phòng định trong DOS, Windows, hay Linux

- Dấu *****: dùng thay cho dãy không kể độ dài các kí tự (từ zero trở lên). hỗ trợ dấu này trong các từ khoá có [AltaVista](#), [Inktomi \(iWon\)](#), [Northern Light](#),

Gigablast, google, Yahoo, MSN ... Thí dụ:

chnology +laser

- Dấu **?**: Dùng thay cho một kí tự bất kì. Hỗ trợ kiểu tìm này có [AOL Search](#), [Inktomi \(iWon\)](#)

2.3.7 Dùng kí tự ~: Đặc biệt trong [google](#) có một cách để tìm không những các trang có chứa từ khoá mà còn tìm các trang có chứa chữ đồng nghĩa với từ khoá (trong tiếng Anh). Ví dụ:

~food facts

sẽ giúp truy tìm các dữ liệu có chữ 'food facts' và các chữ tương đương như 'nutrition facts' ,...

Trong bài tới chúng ta sẽ tìm hiểu về các chế độ nâng cao của các trang truy tìm và vài bảng tổng kết các từ khoá qui ước

© <http://vietsciences.free.fr> - Võ Quang Nhân Ngày 19 tháng 07 năm 2004

Khai Thác Search Engine Cho Nhu Cầu Học Tập và Nghiên Cứu - bài III

vietsciences-Võ Quang Nhân

ngày 03 tháng 08 năm 2004

3. Chế Độ Nâng Cao Của Các Trang Truy Tìm

Ngoài chế độ tìm kiếm thông thường hầu hết các máy truy tìm đều hỗ trợ chức năng nâng cao mà dòng liên kết của nó thường viết bởi cụm từ "*Advanced search*" hay đơn giản là "*Advanced*"



hình1a: Bấm lên dòng 'Advanced Search' cho chế độ nâng cao

Advanced Web Search

can use the options on this page to create a very specific search. Just fill in the ; you need for your current search.

Show results with

all of these words	<input type="text"/>	any part
the exact phrase	<input type="text"/>	any part
any of these words	<input type="text"/>	any part
none of these words	<input type="text"/>	any part

Tip: Use these options to look for an exact phrase or to exclude pages cor also limit your search to certain parts of pages.

Updated

Site/Domain

Any domain
 Only **.com** domains Only **.edu** domains
 Only **.gov** domains Only **.org** domains

only search in this domain/site:

Tip: You can search for results in a specific website (e.g. yahoo.com) or top .org, .gov).

File Format Only find results that are:

Hình1b: một phần hình trong chế độ nâng cao của Yahoo

Trong chế độ này thì sự tìm kiếm được hướng dẫn chi tiết hơn (nhưng cũng không có nghĩa là dễ tìm ra tất cả những gì mình muốn)

Đặc điểm chung của các chế độ nâng cao là:

- Giao diện được thêm vào nhiều ô trống có dòng hướng dẫn để tiện người dùng điền vào (như điền vào các mẫu đơn xin ... thi đại học). Không nhất thiết phải điền hết tất cả các ô trống nhưng các ô này được điền một cách đúng cách và nhiều chi tiết thì càng giúp cho máy truy tìm làm việc hiệu quả hơn
- Đa số các search engine cho phép lọc lựa (filter) lại các thông tin như là: lựa chọn ngôn ngữ (rất tiếc là chỉ có cái máy truy tìm nào thềm cho truy tìm riêng các trang ... Việt ngữ!!!), lựa chọn ngày tháng của trang WEB, lựa chọn các kiểu tệp (file type) và ngay cả lọc lựa các nội dung ... chỉ dành riêng cho người lớn

- Có thể xử dụng các phép toán đã nêu trong bài 2 để nâng cao phép truy tìm. Tuy nhiên, bạn nên cẩn thận vì có thể các kết quả sẽ chịu ảnh hưởng

4. Vài đặc thù của các search engines thông dụng. Mỗi search engine có thể cung cấp thêm một số phương tiện để giảm thiểu khó khăn của việc truy tìm nếu như bạn đã gặp (lỡ gõ) phải từ khoá có quá nhiều trang tìm được -- Có nhiều khi số trang tìm được lên đến cả triệu trang! (vậy là từ khoá này quá chung chung). Sau đây là vài công cụ hữu ích.

4.1 Tìm thêm dữ liệu theo liên hệ (related search): Sau khi search, máy truy tìm có thể giúp bạn đào sâu sự truy tìm bằng cách cung cấp các bộ từ khoá chi tiết hơn dưới dạng các dòng link. Các link này thu nhỏ thị trường tìm kiếm (nếu số trang tìm theo từ khoá của bạn là quá lớn). **Thực ra, nếu bạn biết cách chọn từ khoá cho "xít xao" thì sẽ không phải mượn đến chức năng này.** Thí dụ như bạn tìm các trang viết về sao hoả mà dùng từ khoá mars thì kể như ... một vì số lượng trang tìm ra sẽ vô cùng lớn. Do đó, bạn có thể bấm thêm vào các chữ liên hệ để search engine thu nhỏ thị trường tìm kiếm lại

những trang có chức năng để tìm dữ liệu theo liên hệ là: AltaVista (*Related Searches*), AllTheWeb (*Narrow your search*) , Excite(*Refine Your Results*), HotBot, Lycos, MSN, Yahoo (*also try*)



Hình2: Thu hẹp bằng cách lựa bấm vào một link thích hợp

4.2 Chống hội tụ (cluster): trong lúc tìm kiếm nếu search engine thấy rằng các trang tìm ra có thể ở chung một trang nguồn thì chỉ chọn ra 1 trang đại diện còn các trang khác sẽ được xếp gọn vào và bạn có thể khai triển rộng ra thành nhiều trang nếu muốn ... như vậy sự tìm kiếm sẽ dễ hơn và ít bị "nhiều loạn" thông tin của cùng một trang



Hình3: Các hỗ trợ của google

Những trang có chức năng này là: AltaVista(*Additional relevant pages from this site*), AllTheWeb(*more hits from*), Excite, Google(*More results from*), HotBot, MSN, Northern Light

4.3 Trang Tương Tự (similar) và trang có chính tả gần giống

Trong một số trường hợp thì chức năng tìm các trang tương tự như các trang mà search engine tìm ra có thể hữu dụng. Để dùng chức năng này bạn chỉ việc bấm vào cái liên kết tương ứng. Riêng trong google thì còn có link gợi ý thêm cho việc truy tìm bằng chữ "Did you mean ..."

Các trang có chức năng tìm trang tương tự là: AltaVista (*similar pages, related pages*), AOL Search (*Show me more like this*), Google (*did you mean, similar pages*)

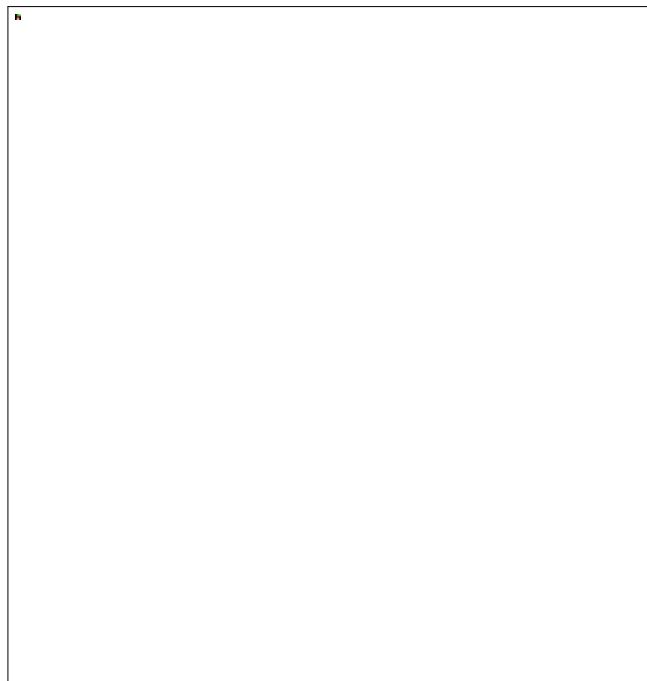
4.4 Trang có từ nối dài (stemming): Đây là khả năng của một số máy truy tìm cho phép tìm những trang có chữ mà có chưa từ khoá bên trong nó.

Thí dụ, thay vì chỉ tìm chữ tutor trang sẽ tìm luôn các chữ tutors, tutorial, ...

AOL Search (*mặc định*), Direct Hit, HotBot(*Enable Word Stemming*)

4.5 Dùng chức năng tồn trữ (cache) của google:

Có nhiều trường hợp trang WEB đã bị xoá sổ không thể hiển thị được nữa nhưng thông tin của nó trong cơ sở dữ liệu của máy tìm vẫn còn mà chúng ta có thể cần dùng đến. Trong thời gian các tin tức này chưa bị xoá khỏi search engine thì bạn có thể đọc được nó nhờ vào chức năng đặc biệt này của google



Hình4: Thông tin cũ tồn đọng trong 'cache'

Ngoài ra, trong google, bạn có thể dùng từ khoá qui ước 'cache' để đọc trực tiếp các thông tin này. (Hà hà! Trong vài trường hợp bạn có thể dùng kĩ thuật này để lôi ra các thông tin đã bị di chuyển đi nơi khác của các ... hackers!)

Ví dụ:

cache:cnn.com

4.6 Chuyển Dịch (translation)

Một số trang cho khả năng dịch lại trang tìm thấy sang thứ tiếng mà người dùng đang có (thường thì chỉ có dịch được ra Anh Pháp Đức hay Tây Ban Nha thôi! Nhưng cũng trặc trẹo lắm) Như vậy, bạn có thể nắm được một phần thông tin có thể dịch ra.

Các trang có chức năng chuyển dịch là: [AltaVista \(Translate\)](#), [Google\(Translate this page\)](#)

5.7 Địa chỉ các search engine thông dụng:

5.7.1 Trang kiểu Spider

- Google: www.google.com
- Ask Jeeves www.ask.com/
- Teoma: www.teoma.com/
- Altavista: www.altavista.com/
- Excite: www.excite.com/
- Gigablast: www.gigablast.com/
- lycos: www.lycos.co.uk/
- Nothern Light www.nlsearch.com

5.7.2 Kiểu Meta:

- Dog Pile www.dogpile.com tìm trên LookSmart, GoTo.com, Thunderstone, Yahoo!, Open Directory, About.com, Direct Hit, Lycos, and AltaVista.
- Meta Find www.metafind.com tìm trên Meta Find Excite, AltaVista, Infoseek, and WebCrawler
- Meta crawlwr www.metacrawler.com tìm trên Lycos, WebCrawler, Infoseek, Excite, Thunderstone, AltaVista, GoTo, và Yahoo

5.7.3 Kiểu Directory:

- Yahoo: www.yahoo.com
- Librarians's Index to the Internet: lii.org/
- Infomine: infomine.ucr.edu/
- Academic Info: www.academicinfo.net
- About: www.about.com

5.8 Phần mềm search engine: Một số nhà sản xuất có viết phần mềm để hỗ trợ cho việc tìm kiếm thông tin trên Internet các phần mềm này có thể được bán (từ 25.00 USD trở lên) hay cho miễn phí tiêu biểu là:

- copernic: www.copernic.com/
- webFerret www.ferretsoft.com
- Seeker www.bluesquirrel.com/store/

Kì tới: Dùng search engine theo lời khuyên của các chuyên gia và phân tích kỹ năng

Khai Thác Search Engine Cho Nhu Cầu Học Tập và Nghiên Cứu - bài IV

vietsciences-Võ Quang Nhân ngày 05 tháng 08 năm 2004

6. Sử dụng đúng chỗ search engine theo lời khuyên của nhà chuyên môn:

Mỗi một máy truy tìm sẽ có điểm mạnh và yếu riêng. Tuỳ theo mục đích, nội dung, và cấp độ của thông tin cần thiết mà bạn nên lựa chọn trang truy tìm cho thích hợp. Sau đây là một số hướng dẫn (và đánh giá) chung. Với một sự tìm kiếm sơ xài không quá chuyên sâu thì rõ ràng việc lựa chọn máy truy tìm không cần thiết; bạn có thể chọn loại nào tiện lợi và quen dùng

6.1 Những tìm kiếm dữ liệu thông thường

Thế nào gọi là thông thường? Thực sự không có biên giới nào về các cấp độ sâu hay nông của thông tin. Người ta thường có xu hướng muốn nhận về càng nhiều chi tiết về 1 sự việc càng tốt. (hà hà nhiều khi có nhiều quá thì lại không biết cái nào cần cái nào không cái nào hay cái nào dở thì tổ ...mệt thêm) Đôi khi giá phải trả cho các chi tiết này lại không xứng với công sức bỏ ra (thời gian và đầu tư) thì cũng là điều đáng suy nghĩ. Theo chúng tôi, đối với cấp độ học vấn thông thường (từ Đại Học và tốt nghiệp đại học trở xuống) thì việc tìm kiếm dữ liệu có thể chỉ cần *loay hoay* trên các máy truy tìm thông dụng (như google, altavista, và yahoo) là đủ cho nhu cầu

6.1.1 Kỹ Năng lựa chọn bộ từ khoá và yêu cầu tối thiểu :

Nếu sử dụng các máy truy tìm thì việc đầu tiên là gõ vào ô tìm kiếm bộ từ khoá. Việc này tưởng chừng như dễ dàng (như ăn cơm với ...mắm zậy mà) nhưng thực tế nhiều khi không đơn giản tí nào. Nhiều khi gõ vào một chữ mà số trang tìm ra đến mấy ngàn thì làm sao biết trang nào là trang cần mở ra? (mở hết thì ... overload mà mắt trang mở ngẫu nhiên lại ... quá đồm) Ngược lại, có lúc gõ vào lại nhận được quá ít hay không có một thông tin nào! Có nhiều nguyên do. Khách quan như là hạn chế của khả năng truy tìm của search engine, số người đăng bài về 'đề tài' này quá ít hay hoàn toàn không có (nhưng cả hai nguyên do này đều đáng ngờ ..hồng lẽ cả trăm triệu trang WEB không có trang nào ...hợp nhãn sao). Nếu loại bỏ các nguyên nhân khách quan, thì sở dĩ thất bại là vì người dùng thiếu kỹ năng viết xuống từ khoá thích hợp hoặc không đáp ứng được một số điều kiện tối thiểu

6.1.2 Những yêu cầu thiết yếu:

- **Vốn sinh ngữ:** Nếu chỉ bó buộc thông tin trong Việt ngữ thì chắc chắn bạn sẽ bị trở ngại lớn. (Các thống kê cho biết số dân Đại Cồ Việt có WEB site riêng thì ... dưới 1 triệu/ 80 triệu dân). Đa số các trang WEB sẽ thông tin bằng ngôn ngữ riêng và hoạ hoàng trình bày thêm 1 ngoại ngữ thường lại là tiếng Anh. (hi hi nói chơi thôi -- nếu muốn tìm tin tức về khủng bố thì học tiếng ..Ả Rập là chắc như bấp ...rang). Như vậy, việc có vốn ngoại ngữ (ít nhất là đọc hiểu) là không thể tránh khỏi. Tuỳ theo mục đích mà dùng ngoại ngữ nào sẽ thích hợp hơn. Tuy nhiên, trong mọi hoàn cảnh, Anh ngữ thường là yêu cầu lớn nhất. (Dẫu sao thì việc biết càng nhiều ngoại ngữ sẽ càng có lợi cho việc tìm kiếm thông tin.) Ngoại trừ tìm tài liệu chuyên biệt về 1 loại ngôn ngữ, khi dùng Anh ngữ để gõ từ khoá thì sắc xuất tìm ra sẽ lớn nhất. Mặc dù vậy, khi dùng từ chung chung thì số lượng trang tìm được sẽ có thể quá nhiều và do đó bạn phải có khả năng để tìm một từ (hay cụm từ) khác chuyên biệt hơn nhằm thu hẹp lại số trang tìm ra. Ngoài ra việc gõ sai chính tả cũng gây mất thì giờ không kém.

- **Vốn chuyên môn:** Bạn không thể học toán nhân (x) nếu chưa biết toán cộng (+). Cũng tương tự, khi truy tìm một tài liệu về bất kì chuyện gì thì điều kiện là bạn phải thỏa mãn yêu cầu kiến thức tối thiểu về chuyện đó. Hiểu biết càng nhiều thì việc tìm kiếm càng hiệu quả. (Ai mà biểu tui đi tìm tài liệu về bộ lọc khói ống bô xe hơi đời mới thì kể như ... luá vì ngoài việc thay nhớt xe tui chả có tí kiến thức nào)
- **Một tí kiên trì:** Làm gì cũng vậy yêu cầu tối thiểu cũng phải kiên nhẫn. Nếu lỡ gõ vào một từ khoá mà được quá nhiều kết quả hay các kết quả không như ý thì bạn nên nghĩ lại xem từ khoá đó có chính xác không nên thêm hay bớt chữ nào. Ngược lại cũng vậy khi mà máy truy tìm không tìm ra cái gì hết thì việc trước tiên là xem lại có viết đúng chính tả không. Từ khoá điền vào có rõ và đầy đủ không (nhiều khi mình dịch từ tiếng việt ra tiếng Anh rồi gõ đại vào và nhận được toàn là những thứ không vừa ý chẳng qua là vì mình dùng sai chỗ/chữ)
- **Dùng thêm hỗ trợ của các sách tra cứu hay sách giáo khoa:** Tại sao phải có mấy thứ linh kinh này? Thật ra, để tìm ra các từ vựng chi tiết và chính xác về đề tài mà mình kiếm thì cách hay nhất là có từ điển và sách giáo khoa để xem lại ... nhiều khi bạn không thể tìm trang WEB bằng từ khoá mà bạn nghĩ ngay tới vì .. nó quá nhiều nghĩa (ambiguous) hay quá chung chung (common/generic meaning). Dùng các từ điển hay sách tra cứu là để kiểm tra chữ khác đặc biệt, cũng đi đi sát với đề tài bạn tìm nhưng lại không hàm chứa nhiều ý nghĩa khác. Ngoài ra, do các yêu cầu tìm kiếm tài liệu thì việc viết đúng chính tả dùng đúng khái niệm và các từ vựng khoa học để đặt vào bộ từ khoá là cần yếu. Như đã trình bày, các từ điển này cũng là nguồn hỗ trợ để "màiũa" lại bộ từ khoá cho bén nhọn hơn
Trên google bạn có thể dùng từ khoá mặc định **define:** để tra cứu định nghĩa của 1 khái niệm hay 1 chữ (ví dụ **define:LASER**)

6.1.3 Kỹ năng lựa chọn từ khoá:

Luôn luôn nhớ cho rằng bạn là người chủ động và thông minh còn computer chỉ là cái máy rất nhanh nhưng thật sự ngờ nghệch. **Hãy xử dụng sự khôn khéo của mình kết hợp với cái nhanh của máy chứ đừng làm ngược lại! Cần phải biết thật rõ bạn muốn đào kiếm cái gì và bao sâu**

Trong khi truy tìm các dữ liệu cho việc nghiên cứu và học tập bằng các search engine thì việc lựa chọn từ khoá rất quan trọng. Nên chú ý thêm rằng, máy truy tìm sẽ không thể tìm ra những bài viết nào chỉ có các từ đồng nghĩa với từ khoá nhưng từ này lại không có mặt trong bản thân của từ khoá mà bạn điền vào ô truy tìm.

- **Từ khoá càng chuyên sâu càng đạt kết quả cao.** Ở đây nếu bạn dùng đúng bộ từ khoá **mô tả chính xác và chi tiết** những gì mình muốn thì kết quả tìm gặp sẽ cao hơn. Nên nhớ rằng nhiều trạng từ hay liên từ thường là các từ không đóng vai trò quan trọng (trừ khi nó đứng trong 1 cụm từ trích dẫn " ") và do đó thường bị các search engine bỏ qua. (Trong Anh ngữ thì tránh dùng các chữ: a, an, the, is, and, or, of, you, me, my .. trừ khi nó đứng trong ngoặc kép của trích dẫn). Ví dụ: Khi install Redhat 7.3, một số máy compaq hiệu 1850R sẽ bị treo. Bạn muốn tìm giải pháp của chứng này trên Internet. Bạn có thể tự mình so sách kết quả việc dùng các bộ từ khoá sau:
 1. **Linux hang**
 2. **Redhat 7.3 install**
 3. **Redhat 7.3 1850R hang**
 4. **Redhat 7.3 1850R intall hang**
Nếu như nhận xét thì bạn dễ dàng tìm ra lời giải bằng cách đọc vài trang tìm thấy bởi google với một trong hai bộ từ khoá thứ 3 hay 4. (và địa chỉ của cách giải (có thể) nằm trong các trang đầu tìm kiếm) trong khi đó nếu dùng hai bộ từ khoá đầu thì hơi khó

- **"Cụm" từ khoá và tổ hợp các từ khoá:** Như trên chúng ta cũng đã thấy dùng tổ hợp từ khoá có ảnh hưởng nhiều đến việc truy tìm ... (và ngay cả đảo thứ tự các từ khoá cũng có ảnh hưởng đến sự sắp xếp thứ tự các trang tìm ra .. bạn hãy tự so sánh khi dùng bộ từ khoá 4a: hang install Redhat 7.3 1850R so với bộ từ khoá 4 xem). Trong nhiều trường hợp thì việc dùng ngoặc kép để bảo vệ nguyên văn cụm từ đi tìm sẽ có vai trò khá quan trọng (nhất là trường hợp bạn muốn tìm lại văn bản của một bài viết đặc biệt nào đó). thí dụ: Bạn muốn tìm càng nhiều càng tốt bài viết về một người tên là Nguyễn Khắc Hiếu nhưng lại không phải là "Tản Đà" bạn có thể thấy kết quả khác nhau vô cùng của hai bộ từ khoá:

Nguyễn Khắc Hiếu -Tản -Đà -phố

"Nguyễn Khắc Hiếu" -Tản -Đà -phố



hình1a: Dấu ngoặc kép giúp truy tìm chính xác hơn (chỉ còn 384 trang)



Hình 1b: Không có ngoặc Kép thì có đến hơn 2000 trang! (kể như trường hợp này là "lắm mối, tối nằm ...chơi")

Ví dụ2: **biomedical engineering" AND cancer** sẽ cho ra quá nhiều kết quả liên quan đến các chương trình học bạn có thể loại bớt chúng bằng từ khoá **"biomedical engineering" cancer -"Department of" -"School of"**

- **Biến chiểu (với các phép toán)** Ngoài các dấu + và - thì việc tận dụng các phép toán sẽ giúp quá trình tìm kiếm thông tin nhanh hơn và hiệu quả hơn. (Rất tiếc, google lại không hỗ trợ tốt các phép toán Boolean như là Altavista). Xin đơn cử 1 ví dụ là bạn muốn tìm tài liệu để tự học về Linux script (--Tạm dịch: văn lệnh của hệ thống Linux). Trong tiếng Anh các tài liệu chỉ dẫn thường được viết là tutor (n) tutorial (adj) hay là guide, howto (chữ guide thường ở mức độ sơ sài hơn). Như vậy ta có thể thử dùng bộ từ khoá (trong Altavista):

"Linux script" AND (tutor OR tutorial OR Howto OR guide)

Lưu ý: Phép toán phân nhánh trong Altavista chỉ được hỗ trợ mức thấp nhất nghĩa là bạn không thể lồng hai hay nhiều dấu ngoặc đơn vào nhau

- **Biết nhiều ngoại ngữ:** Tiếng Anh là thứ tiếng thông dụng nhất thế giới (kể đến là tiếng ..Tàu. Nhưng rất tiếc, không có mấy ai viết bài nghiên cứu khoa học bằng thứ tiếng tượng hình này ngoại trừ dân bản xứ mà xứ này thì lại còn bị giới hạn nhiều

thứ) Nếu kể về số lượng các tài liệu khoa học chuyên khảo thì bên cạnh tiếng Anh còn có Tiếng Đức, Tây Ban Nha, và Pháp. Nhiều khi tài liệu mà bạn muốn tìm lại không viết bằng tiếng Anh! Do đó, biết càng nhiều thứ tiếng thì càng có lợi. Ngay cả việc tìm tài liệu nghiên cứu về Việt Nam thì việc biết thêm Anh hay Pháp ngữ cũng vô cùng quan trọng (để biết cái nhìn khách quan của người nước ngoài). Bạn hãy thử so sánh số trang tìm được của hai từ khoá sau đây về Ampere: "Andre Marie Ampere" (Anh) và "André Marie Ampère" (Pháp) thì sẽ rõ.

- **Dùng từ tương đương hay đồng nghĩa:** Trong ví dụ về việc tìm bài học về Linux script, chúng ta cũng đã dùng kĩ thuật từ tương đương hay đồng nghĩa. Vấn đề là vì một số tác giả thích dùng chữ này mà lại không dùng chữ kia, và do đó, nếu bạn bỏ sót 1 từ đồng nghĩa **X** (hay từ tương đương **X**) thì cũng có nghĩa là bạn mất hẳn một cơ hội tìm ra bài viết của các tác giả chỉ thích dùng từ **X** này. (Ngộ nhờ bài đó lại là của 1 tác giả xuất sắc thì lại càng uổng hơn). Ngoài ra, việc dùng ACRONYM (--tạm dịch: chữ viết tắt) thêm vào bộ từ khoá của một danh từ khoa học cũng rất quan trọng. thí dụ người ra hay dùng chữ **LED** (đi-ốt phát quang) trong tài liệu hơn là chữ nguyên gốc "**light emitting diode**". Cũng như vậy cho chữ laser.
- **Dùng từ cùng chuyên khoa:** Khi bạn tìm các bài chuyên khảo về một ngành nghề nào đó thì việc hiểu viết nhiều từ vựng trong ngành đó sẽ giúp bạn rất nhiều trong việc "tinh lọc" các thông tin nhận về. Thí dụ trường hợp bạn tìm tài liệu nghiên cứu về kí sinh trùng sốt rét ác tính thì ngoài chữ *intensive malaria* bạn có thể thay bằng tên khoa học của con kí sinh trùng này (*Plasmodium Falciparum*) để thêm vào bộ từ khoá.

6.2 Tìm Kiếm Đề Tài có Mục tiêu Xác Định:

Trước tiên xử hãy dụng cách thức thông thường (Vì cái địa chỉ của google nó dễ nhớ ! Hi hi) thông thường. Nếu bạn vẫn muốn tìm thêm dữ liệu thì hãy thử những gì ghi ra dưới đây tùy theo mục tiêu truy tìm

6.2.1 Dữ liệu đặc biệt:

Hãy dùng đến các trang chứa cơ sở dữ liệu Nhắc lại từ bài 1 rằng có nhiều thông tin được cất giữ riêng trong các cơ sở dữ liệu mà chỉ có thể tìm ra khi chúng ta vào các trang riêng biệt để tìm chúng. Một vài trang chứa cơ sở dữ liệu như vậy là

<http://lii.org/> -- adcademic search database

<http://www.academicinfo.net/> -- adcademic search database

<http://infomine.ucr.edu/> -- adcademic search database

<http://www.invisible-web.net/> -- invisible WEB

<http://www.rdn.ac.uk/> Resouce Discovery Network

<http://turbo10.com/cgi-bin/showdeepnet.cgi> --Turbo 10

<http://www.freepint.com/gary/direct.htm> Direct Search

<http://www.noodletools.com/debbie/literacies/information/5locate/advicedepth.html> Noodle tools

<http://www.lib.berkeley.edu/Collections/acadtarg.html> Thư mục của Đại học Berkely

Trong nhiều trường hợp nếu bạn muốn truy tìm tin tức từ các cơ sở dữ liệu chỉ việc thêm vào bộ từ khoá chữ "**database**" ở vị trí cuối cùng của bộ từ khoá.

6.2.2 Dữ liệu có liên quan đến cấu trúc viết chữ và từ khoá (chính tả, đồng nghĩa, câu cú, chữ hoa hay chữ thường):

Trong trường hợp này cách tốt nhất là ... bỏ các chữ này vào các ngoặc kép thì sẽ giải quyết được vấn đề.

6.2.3 Truy tìm định nghĩa, khái niệm, tài liệu tham khảo hay dịch nghĩa:

Nếu chỉ muốn biết chính tả thì có thể dùng từ đó làm từ khoá và gõ vào google rồi đọc chức năng hỗ trợ có dòng "did you mean" Đây là là chức năng có khả năng chỉnh chính tả. Nếu để tra ý nghĩa chữ, khái niệm, hay danh từ khoa học thì dùng từ khoá qui ước **define**: trên google. (ví dụ **define:OLED**) Ngoài ra, bạn có thể lên các trang tự điển để tra cứu chẳng hạn như:

<http://www.yourdictionary.com>
<http://dictionary.reference.com/>
<http://www.reference.com/>
<http://www.britannica.com/>
<http://www.search.com/search?channel=19>
<http://www.iTools.com/research-it>

Các tài liệu tham khảo chung còn có

<http://www.odci.gov/cia/publications/factbook/index.html> : the world fact book của CIA
<http://www.guinnessworldrecords.com/> : Trang Guinness các kỉ lục thế giới
<http://portal.unesco.org>: Trang của tổ chức Văn hoá Giáo dục và Khoa Học của Liên Hợp Quốc
<http://www.noodletools.com/debbie/resources/math/stats.html> trang liên kết về các dữ liệu thống kê.

6.2.4 Tìm trên các thư viện:

Các thư viện trên WEB cũng có thể cung cấp nhiều thông tin quý báu:

<http://www.ipl.org/> Thư viện Internet công cộng
<http://www.lib.berkeley.edu/> Thư viện của ĐH Berkeley
http://web4.si.edu/sil/onlineexhibitions/oe_search2.cfm Thư viện WEB của Smithsonian Institution

Các trước tác:

<http://libweb.sonoma.edu/assistance/subject/primaryonline.html> : Nguồn link của các trước tác (primary source) Đây là một nguồn liên kết sưu tầm rất lớn các tài liệu từ các thư viện

<http://www.cyberschoolbus.un.org/infonation/info.asp> so sánh thông tin về địa lý và chính quyền các nước.

<http://www.uidaho.edu/special-collections/iil.htm> Bộ sưu tập lớn các liên kết tới trang WEB của các quốc gia

<http://www.tulane.edu/~lmliller/ArchivesResources.html> thư viện NET go

<http://vlib.org/Overview.html> Thư viện ảo

<http://www.loc.gov/rr/askalib/> Hỏi các thủ thư viện.

Thông tin về các vấn đề nóng:

<http://infodome.sdsu.edu/research/guides/hot/supersites.shtml>

<http://www.publicagenda.com/issues/issuehome.cfm>

6.2.5 Các trích dẫn câu phát biểu (Quotation):

<http://www.quoteland.com>

<http://www.quotationspage.com/search.php3>

<http://www.aphids.com/quotes/index.shtml>

<http://creativequotations.com> (Những phát biểu liên quan tới sự sáng tạo)

6.2.6 Truy tìm phương thức giải quyết:

Các bài viết về cách thức làm hay giải quyết vấn đề nào đó thường có những chữ đặc trưng như là "howto", "how to", "instruction", "do it yourself", "DYI",..

Trong nhiều trường hợp có những vấn đề về kĩ thuật, bạn có thể viết thẳng câu hỏi bằng Anh ngữ lên và xem như đó là bộ từ khoá (Dĩ nhiên là phải đúng văn phạm và chính tả).
Thí dụ: Bạn muốn tìm tài liệu để chế tạo thiết bị Laser hãy thử bộ từ khoá trên google:

How to build a laser device

Bạn sẽ nhận được kết quả rất thú vị bất ngờ.

Tương tự như vậy khi mà bạn gặp hư hỏng hay trục trặc trong quá trình xử dụng computer hay phần mềm thì cứ chép ngay nguyên văn (hay một phần lớn) cái dòng báo lỗi lên làm thành bộ từ khoá. Sau đó có thể thêm vào một số chi tiết để lọc bớt thông tin.

Thí dụ: đôi khi dùng X-window của Linux có thể bạn sẽ gặp trục trặc mà máy báo lỗi như sau: "**Could not init font path element Unix/:7100, removing from list!**". Bạn có thể dùng ngay câu báo lỗi này làm từ khoá và bỏ vào google để tìm các bài viết về cách giải quyết:

"Could not init font path element Unix/:7100"

Bạn sẽ tìm được khá nhiều bài viết về việc này. Dĩ nhiên là bạn có thể sẽ phải thêm vào bộ từ khoá vài chi tiết để tìm ra cách giải quyết thoả đáng.

Ngoài ra, các bạn có thể trực tiếp đặt câu hỏi giải quyết khó khăn trong các forum mà trang sau đây là một chỗ tập trung các links:

<http://www.forumone.com/index/>

6.2.7 Truy tìm tài liệu học vấn (MÃ NGUỒN, BÀI VỞ...):

Hãy chú ý đến những từ khoá như là **tutor, tutorial, research, database, how-to, instruction, source, source code, schematic, flow chart, circuit, theorem, proof, ...** nếu như bạn muốn tìm hiểu về một định lý hay lí thuyết nào đó thì cứ việc đánh thẳng tên (nếu có) của định lí hay lí thuyết đó làm từ khoá xem. Từ kết quả này bạn có thể lọc lại bằng cách thêm vào bộ từ khoá các phép toán (dùng dấu - hay **NOT**) . Ngay cả trường hợp chỉ nhớ đúng một phần của tên cũng có thể hữu dụng. Thí dụ bạn hãy thử đánh **Cauchy theorem** trên google xem có thể tìm được bao nhiêu định lí Cauchy khác nhau.

6.2.8 Truy tìm tin tức:

Tin tức thì có rất nhiều nhưng nổi bật là:

<http://dailynews.yahoo.com> Yahoo news

<http://news.google.com> Google news

Tin tức với bình luận và phân tích:

<http://www.rdn.ac.uk/bth/> BBC news

<http://www.usnews.com/usnews/briefings/nbhome.htm>

Trang WEB của các hãng thông tấn lớn:

www.cnn.com CNN

<http://www.ap.org/> AP

<http://www.afp.com/english/home/> AFP

<http://www.cbs.com/> CBS

<http://nbcpublish.console.net/> NBS

Trang liên kết các trang về tin tức:

<http://www.headlinespot.com>

Truy tìm các tin tức từ mọi khía cạnh (bằng nhiều ngôn ngữ):

<http://www.alltheweb.com/advanced?c=news&cs=utf-8>

6.2.9 Truy tìm các loại tệp văn bản:

Hãy dùng từ khoá qui ước **filetype:** trên www.google.com để tìm ra cá loại tệp văn bản có format riêng. Hoặc có thể dùng advanced search từ MSN:

<http://search.msn.com/advanced.aspx?FORM=HP>

6.2.10 Truy tìm các quan hệ:

Một số trường hợp bạn lại muốn tìm các trang có chỉ thêm quan hệ giữa các từ khoá thì trang search engine say đây có thể cung cấp các thông tin thú vị

<http://www.kartoo.com/flash04.php3> của trang Kart00

ví dụ: từ khoá gasoline electric car sẽ cho kết quả nhiều bản đồ liên hệ:



Hình2: Kart00

Tương tự có trang http://www.webbrain.com/html/default_win.html

6.2.11 Truy tìm các loại dữ liệu không phải văn bản (nhạc, hoạ, phim hình)

Có thể dùng trang www.Altavista.com Rồi bấm vào các đề mục (như images, MP3/Audio, video)

<http://memory.loc.gov/ammem/collections/finder.html> sưu tập các thể loại nghệ thuật

Tim hình ảnh có thể dùng:

www.google.com hay www.ask.com

<http://classroomclipart.com> Các loại hình ảnh computer

http://www.si.umich.edu/Art_History/ Thư viện nghệ thuật tranh-ảnh của ĐH Michigan

<http://www.thinker.org/fam/about/imagebase/subpage.asp?subpagekey=420> Trang thư viện hội hoạ của San Francisco

Trang có chứa nhiều tập tin âm thanh: <http://www.findsounds.com>

6.2.12 Truy tìm người, địa chỉ số phone:

- *Truy tìm thông tin về các nhân vật nổi tiếng:*
<http://www.infoplease.com/people.html>
<http://www.biography.com/search/>
Các nhân vật nổi tiếng nhưng đã chết: <http://amillionlives.com>
- *Truy tìm người thường hay các thông tin về cá nhân:*
Hãy thử gõ tên đúng chính tả tên của người cần tìm trên www.google.com
Các trang truy tìm người và thông tin cá nhân:
<http://people.yahoo.com>
<http://www.nwbuildnet.com/nwbn/fouremail.html>
<http://www.freeality.com/findrev.htm>
<http://find.intelius.com/>

6.2.13 Các báo cáo Khoa Học NSDL (National Science, Technology, Engineering, and Mathematics Education Digital Library)

Trang này của Internet Scout Project chuyên chở các báo cáo về các lãnh vực Khoa học Kì thuật và Toán Giáo dục Thư viện số của các cơ quan nghiên cứu:
<http://scout.cs.wisc.edu/index.php>

6.2.14 Thu nhận ý kiến thăm dò (kết quả):

<http://www.opinion-pages.org>
<http://www.pollingreport.com>
<http://blogdex.net>

Các đánh giá về chất lượng:

<http://www.rateitall.com>

6.2.15 Hỏi ý kiến của các chuyên gia:

<http://www.vrd.org/locator/> Dành cho các câu hỏi về giáo dục và các lớp bậc trung học của bộ giáo dục Hoa kì
<http://www.allexperts.com/index.htm> Trang WEb của các chuyên gia thiện nguyện
[http://experts.mediamap.com/\(dx3lkv553etvqq55ys1njhqb\)/search.aspx](http://experts.mediamap.com/(dx3lkv553etvqq55ys1njhqb)/search.aspx) Dùng để tìm chuyên gia về các ngành nghề
<http://www.ibiblio.org/slanews/internet/experts.html> các dòng liên kết tới các trang WEB chuyên ngành

6.2.16 Tìm Bản đồ:

<http://plasma.nationalgeographic.com/mapmachine/> Trang của National Geographic
<http://atlas.geo.cornell.edu/webmap/> Bản đồ địa lí thế giới
<http://www.mapsarea.com/> Tập trung các link về bản đồ đủ loại

6.2.17 Tải về các software:

Có lẽ là thừa nếu tôi trình bày với các bạn trong nước ở đâu để kiếm phần mềm (vừa rẻ vừa tốt). Tuy nhiên, nguồn tải các phần mềm hợp pháp lớn nhất là:
<http://www.shareware.com/> và www.download.com

6.2.18 Các liên kết khoa Học:

Ngoài các địa chỉ đã được giới thiệu, các link sau đây có thể giúp bạn có thêm các tài liệu Khoa học cũng như giáo khoa bậc ĐH
<http://www.martindalecenter.com/> Mantidale Reference Desk

<http://www.sciencegems.com/> Frank Potter's Science Gems chủ yếu cho Khoa Vật Lý
<http://explanation-guide.info/> Cho rất nhiều giải thích về các hiện tượng và hiệu ứng
<http://searchenginewatch.com/links/science.html> Có các links về các trang khoa học
<http://searchenginewatch.com/links/medical.html> Có các link về các trang y học

7. Giới thiệu một số máy truy tìm của các quốc gia khác Hoa kì:

8. Phụ Lục:

7.1: Kỹ thuật trình bày META TAG cho một trang WEB:

Như đã đề cập trong bài I, một số máy truy tìm sẽ dùng các thông tin cho trong META TAG để xếp hạng các bài viết. Nếu bạn chưa rõ cách thức xếp hạng của máy truy tìm xin [xem lại bài I](#)

Nhiều khi bạn đi tìm một tài liệu chuyên về khoa học thì lại có khá nhiều trang ... quảng cáo về chuyện hoàn toàn khác lọt vào trong danh mục tìm thấy. Đó là vì các trang (vô duyên) này đã dùng đến kỹ thuật nhồi nhét các 'từ khoá' mà họ cho là hay là quan trọng vào trong các META TAG của họ

Hiểu biết về META TAG sẽ có thể giúp bạn trong việc lọc lựa các thông tin xuất hiện trên máy truy tìm bằng cách điều chỉnh bộ từ khoá.

7.1.1 META TAG là gì:

- Về mặt mã nguồn thì bạn có thể tìm thấy META TAG nằm trong phần đầu <HEAD> của tất cả các trang WEB. Thí dụ:

<HEAD>

```
<meta http-equiv="Content-Type" content="text/html; charset=utf-8">
```

```
<meta http-equiv="Content-Language" content="vi">
```

```
<meta name="GENERATOR" content="Microsoft FrontPage 6.0">
```

```
<meta name="keywords" content="Sciences; science; khoa học; khoa học; vật lý; hóa; sinh học; gene; danh nhân; danh nhân; Tiểu sử danh nhân; thiên văn; astronomy; astronomie; physics; physique; maths; langues; languages; anh văn; english; vietnamese; vietnam; lịch sử; lịch sử; histoire; inventions; sáng chế; phát minh; giải thưởng; tặng đá bên đường; tìm hiểu; danh ngôn; Nobel, Crafoord; Fields">
```

- Các META TAG này dùng để chỉ thị thêm cho các WEB Browser (tạm dịch -- Máy truy Cập WEB) thêm một số thông tin đặc biệt về định nghĩa, mô tả, các yêu cầu đặc trưng để tải,...
- Một bộ phận quan trọng của META TAG là cung cấp thêm dữ liệu 'từ khoá' cho các máy truy tìm. Do đó, những trang nào có các META soạn thảo một cách kỹ lưỡng thì sẽ có nhiều cơ hội hơn để được các máy truy tìm đưa lên thứ tự ưu tiên trong danh mục kiếm theo từ khoá khi nó bắt gặp. Ngược lại nhiều trang WEB nhất là trang WEB của những người không chuyên môn về cách trình bày META cho WEB site thì cơ hội bài của họ được xếp cao trong danh mục tìm thấy sẽ thấp hơn và do đó khó tìm ra hơn (mặc dù thường các bài viết này có thể chứa thông tin mà bạn đang cần)
- Các nội dung của META TAG bao gồm Content (nội dung) Subcategory (phân lớp), Access level (cấp độ xử dụng) và Access Restriction Requirements (các yêu cầu giới hạn khi xử dụng), Expiry Date (thời hạn), Publisher (nhà xuất bản), Keywords (từ khoá) và Description (mô tả)

7.1.2 Đặc Điểm:

- Tùy theo cấu trúc của từng máy truy tìm mà các META TAG có thể có ảnh hưởng khác nhau đến việc xếp hạng.
- Trong thời gian gần đây thì tầm quan trọng của META TAG có suy giảm vì sự lạm dụng của các trang quảng cáo (họ nhồi nhét đủ thứ vô để đánh lừa máy truy tìm). do đó dẫn đến việc người soạn thảo máy truy tìm thay đổi phương án xây dựng các chỉ số cho cơ sở dữ liệu (hạ thấp vai trò của META TAG)
- Các thông tin trong META TAG ngoại trừ việc xử dụng cho Máy truy cập và máy truy tìm, sẽ không được hiển thị trên trang WEB
- Những từ vựng viết trong "title", "description", hay "keyword" sẽ ảnh hưởng đến cách sắp xếp của search engine.
- Không có biện pháp tuyệt hảo nào để giúp cho 1 trang WEB có thể đứng đầu vĩnh viễn trong danh mục truy tìm theo một bộ 'từ khoá' cố định. Lý do là sự thay đổi cách làm việc của máy truy tìm và sự thay đổi của META TAG trong các trang khác trở nên hiệu quả hơn trang vốn đứng đầu trong danh mục.

7.1.3 Vài cách thức để áp dụng cho trang WEB: Chúng tôi nêu thêm ra ở đây dành cho các bạn để điều chỉnh trang WEB của mình để tạo thêm cơ hội cho nhiều người tìm đến đọc:

- Bạn có thể dùng hai dạng quang trọng sau:
<META name ="description" content="(viết thật ngắn mô tả trang WEB của bạn)"
hay là:
<META name ="keywords" content="(liệt kê tất cả những từ có thể làm từ khoá *)"
(*) Mỗi 'từ khoá' được liệt kê ra sẽ phân biệt với 'từ khoá' khác bằng dấu ';' thường thì từ được liệt kê có thể là những chữ đồng nghĩa, chữ viết tắt, chữ được dịch ra các thứ tiếng khác nhau, hay các mô tả ngắn về trang WEB của bạn
- Có nhiều WEB master đã lạm dụng META TAG đưa vào rất rất nhiều từ khoá không đúng với nội dung của trang WEB để nâng cao sự phân loại. Nên nhớ rằng sớm hay muộn các máy truy tìm sẽ tìm ra và loại trang đó khỏi vị trí cần thiết. Hãy kiểm tra chắc chắn rằng các 'từ khoá' viết vào là có liên hệ trực tiếp tới bài viết
- Nhiều máy truy tìm sẽ tập trung đánh giá vào phần đầu các dòng chữ của trang WEB, nên có thể để ý trình bày cho phù hợp
- Các TAG khác như <H1>, <H2>,... cũng được máy truy tìm tính đến để phân hạng
- Tên của tập tin tương đồng với keyword cũng là nhân tố để đẩy sự xếp hạng lên cao
- Trong các ALT tag thì bạn cũng có thể điền các keyword vào để gây thêm sự chú ý của máy truy tìm.

7.2 Phụ lục về gốc (extension name) của tên miền: Trong khi tìm kiếm về các trang WEB bạn cũng có thể dựa vào tên đuôi (gốc) của miền để biết thêm đặc điểm (về ngôn ngữ chẳng hạn). Sau đây là những tên miền thường thấy. Dĩ nhiên, các tên này cũng không có tính cách bắt buộc (nghĩa là không nhất thiết phải có đúng các đặc tính :

- .com (ví dụ www.google.com) com là viết tắt của chữ comercial. Các địa chỉ loại này thường tập trung vào các loại tổ chức có thu nhập hay buôn bán hay có trao đổi thương mại (bán quảng cáo)
- .net: thường là cho các cơ quan thuộc về mạng (network) <http://www.nw.net>
- .gov: thường là các cơ quan công quyền hay chính phủ (government) <http://www.INS.gov>
- .org: Các tổ chức thường là các tổ chức không trực tiếp làm thương mại. (organization). ví dụ: <http://httpd.apache.org>
- edu: Đây là tập trung của những tổ chức chức năng giáo dục như là các trường học, ban ngành về giáo dục. Ví dụ: <http://www.lib.berkeley.edu/>
- Ngoài ra, các tên đuôi bao gồm hai chữ thì sẽ cho biết xuất xứ của trang WEB. Thí dụ: .vn (vietnam), .uk (United Kingdom), .fr (french), ... riêng các trang WEB từ nước Mỹ sẽ không có tên này.

Tài liệu tham khảo chính:

<http://www.lib.berkeley.edu/TeachingLib/Guides/Internet/>

<http://searchenginewatch.com/>

<http://www.searchengineshowdown.com>

<http://www.monash.com/spidap.html>

<http://www.noodletools.com/debbie/literacies/information/5locate/adviceengine.html>

<http://www.topwebsite.co.uk/>

<http://Vietsscience.free.fr> - Ngày 13 tháng 08 năm 2004