

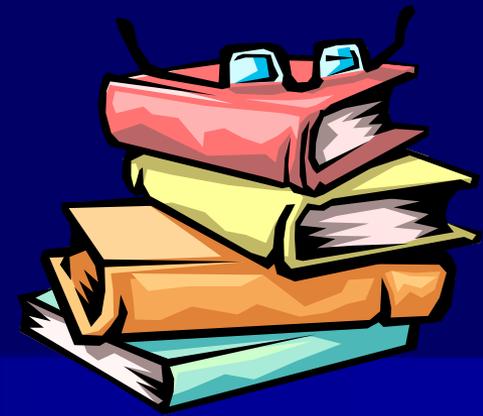
Internet Fundamentals

Lecture-13

WWW Search Engines



Outline



- Introduction
- Directories, Search Engines, Metasearch Engines
- Search Fundamentals
- Search Strategies
- How does a search engine work?
- Searching Tips
- Your site's ranking?
- Summary

Introduction

- You have probably been using search engines, but perhaps may not be as effectively as possible.
- A lot of information is available on-line, but not all of them is completely accurate.
- The web-page addresses are constantly changing, it may be only available for a short time.

Search Engine History

- In 1990, before the WWW, Alan Emtage created Archie, the first search tool for finding files on FTP sites.
- In 1993, Veronica is developed. Followed by Jughead, Wandex, ...
- In 1994, Galaxy, WebCrawler, Yahoo! and Lycos debuted.
- In 1995 and afterwards, Excite, Infoseek, Alta Vista, MetaCrawler, ...
- Next generation: specialized hybrids

Directories

- A Web Directory or Web Guide is a hierarchical representation of hyperlinks.
- The top level is typically a wide range of very general topics.
- Each topic contains hyperlinks of more specialized sub-topics.
- Very easy to use.

Hierarchical Representation

The image displays four sequential browser screenshots illustrating a hierarchical navigation path through Google's directory structure. Red arrows and circles highlight the progression from a general category to a specific sub-category and then to a list of related resources.

- Screenshot 1: Google Directory - Microsoft**
 - Left sidebar: **Arts** (circled in red), Movies, Music, Television, Business, Computers, Games, Health.
 - Main content: **Arts** (circled in red), Categories: Animation (22177), Antiques (871), Architecture (3337), **Art History** (735) (circled in red), Bodyart (1124), Celebrities (27837), Classical Studies, Comics (5104), Crafts (6086), Dance (4751), Design (2451), Digital (281), Directories (431).
- Screenshot 2: Google Directory - Art History - Micro**
 - Path: [Arts](#) > Art History
 - Categories: Aesthetics (36), Architecture (2321), Artists (1937), **Chats and Forums** (7) (circled in red), **Directories** (15) (circled in red), Graphic Design (13), Journals, Movement, Museums, Online Co, Organizati, Photograp.
 - Related Categories: Reference > Museums > Museum I, Society > History > By Topic > Cult.
 - Web Pages: **Viewing in Google**
 - [ArtMuseum.net](#) - http://www.artm... internet-based art museum experien background information not normally
 - [Mark Harden's Artchive](#) - http:// An art archive with thousands of ima names and art styles. Covers every
 - [Resources in Art History for Gi](#) - This newsletter lists grants, fellowsh
- Screenshot 3: Google Directory - Arts > Art History > Directories - Microsoft Internet Explorer**
 - Path: [Arts](#) > [Art History](#) > Directories
 - Related Category: [Arts > Directories](#) (431)
 - Web Pages: **Viewing in Google PageRank order**
 - [Artcyclopedia](#) - http://www.artcyclopedia.com/ Guide to museum-quality art on the Internet. Search hundreds of art museum sites for exhibits and artists.
 - [Art History Resources on the Web](#) - http://witcombe.sbc.edu/ARTHLinks.html In-depth hyperlinked lists divided by period and maintained by Chris Witcombe, Professor of Art History at Sweet Briar College, Virginia.
 - [ArtHistory Network](#) - http://www.arthistory.net/ Art history, archaeology and architecture resources on the web.
 - [Digital Library for the Decorative Arts and Material Culture](#) - http://decorativearts.library.wisc.edu Collects electronic resources for study and research, with a focus on Early America.
 - [Copious Art Links](#) - http://rampages.onramp.net/~dnewman/artlinks.htm Group of internet art links maintained by Brookhaven College.
 - [The Art History Research Centre](#) - http://art-history.concordia.ca/AHRC/index.htm Tool to facilitate art historical research as well as other arts research on the internet.
 - [Art History Resources](#) - http://www.asu.edu/lib/hayden/ref/hum/art/arthist.html Web sites and publications in print that are concerned with various aspects of art history, collected by the Arizona State University libraries.

Popular Directories

- AOL anywhere – search.aol.com
- CNET Search.com – www.search.com
- Excite – www.excite.com
- E-Wild life – www.ewildlife.com
- Lycos – www.lycos.com
- Yahoo! – www.yahoo.com
- Google – www.google.com

Some figures

AOL Search: Home - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Back Forward Stop Home Search Favorites Media Print Copy Paste >> Google

Links Yahoo Mail CSC1720 Old homepage Dictionary RTHK Address http://s

AOL Anywhere

Main My AOL Mail People Search Shop Channels

Send Her Valentine's Day Roses
Order Now to Guarantee Delivery by 2/14 1-800-flowers.com click here

Search Search Main | Options | Help

Categories:

- Arts & Entertainment**
Celebrities, Music, Movies, TV...
- Autos**
Buyers Guides, Makes, Repair...
- Business**
Industries, Investing, Real Estate...
- Computers & Internet**
Graphics, Hardware, Software...
- Games**
Web, Video, Card Games...
- Health**
Fitness, Illnesses, Nutrition...
- Home & Garden**
Families, Gardening, Cooking...
- Kids & Teens**
Arts, Games, School, Teens...
- Recreation**
Humor, Pets, Outdoors...
- Reference**
Dictionaries, Colleges, Maps...
- Regional**
US, Canada, UK, Europe...
- Science & Technology**
Biology, Environment, Astronomy...
- Shopping**
Auctions, Clothing, Books, Gifts...
- Society & Culture**
Religion, Relationships, Genealogy...
- Sports**
Basketball, Football, WWF, NASCAR...
- Travel**
Cruises, Destinations, Reservations...

SEARCH AT amazon.com
Get the ICGQ200 and fall in

Search.com - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Back Forward Stop Home Search Favorites Media Print Copy Paste >> Google

Links Yahoo Mail CSC1720 Old homepage Dictionary RTHK Address http://www.search.com

Search.com

advertisement

Small Business Savings. See our latest offer running now. [Click for details.](#)

Easy as **DELL**

Quicklinks: [Breaking News](#) · [Jobs](#) · [Maps](#) · [Yellow Pages](#) · [Shopping](#) · [People](#) · [more...](#)

Start your metasearch Go! [Advanced Help](#)

Valentine's Day: [gifts](#) · [flowers](#) · [jewelry](#) · [chocolates](#) · [gift baskets](#)
Shopping: [cell phones](#) · [tax software](#) · [computers](#) · [airline tickets](#) · [window blinds](#)
Finance: [online tax filing](#) · [debt consolidation](#) · [credit card debt](#) · [car insurance](#)

Specialize Your Search

Metasearch more than 800 specialized engines from around the Web:

- Business & Money**
[Jobs](#), [Stock discussions](#), [Symbol lookup](#), [Stock quotes](#), [more...](#)
- Downloads**
[Shareware](#), [MP3s](#), [Images](#), [Linux](#), [more...](#)
- Entertainment**
[Movies](#), [Celebrities](#), [Television](#), [Horoscopes](#), [more...](#)
- Health**
[Drugs](#), [Alternative medicine](#), [Journals](#), [Nutrition](#), [more...](#)
- Music**
[MP3s](#), [Tours and tickets](#), [Net radio stations](#), [Prices](#), [more...](#)
- People**
[E-mail addresses](#), [Phone numbers](#), [Instant messenger lookup](#), [more...](#)
- Shopping**
[Reviews](#), [Prices](#), [Auctions](#), [Automotive](#), [more...](#)
- Travel**
[Destination guides](#), [Discount airfares](#), [Weather](#), [Maps](#), [more...](#)

February

New and Improved

- [Breaking News](#)
- [Domain Names](#)
- [Yellow Pages](#)
- [How-to](#)

Search Results

- [Submit Your Site](#)
- [Top 100 Sites](#)
- [Snoop Report](#)
- [Site Maps](#)
- [Advanced Search](#)
- [Search.com](#)
- [Download](#)

Subscribe your

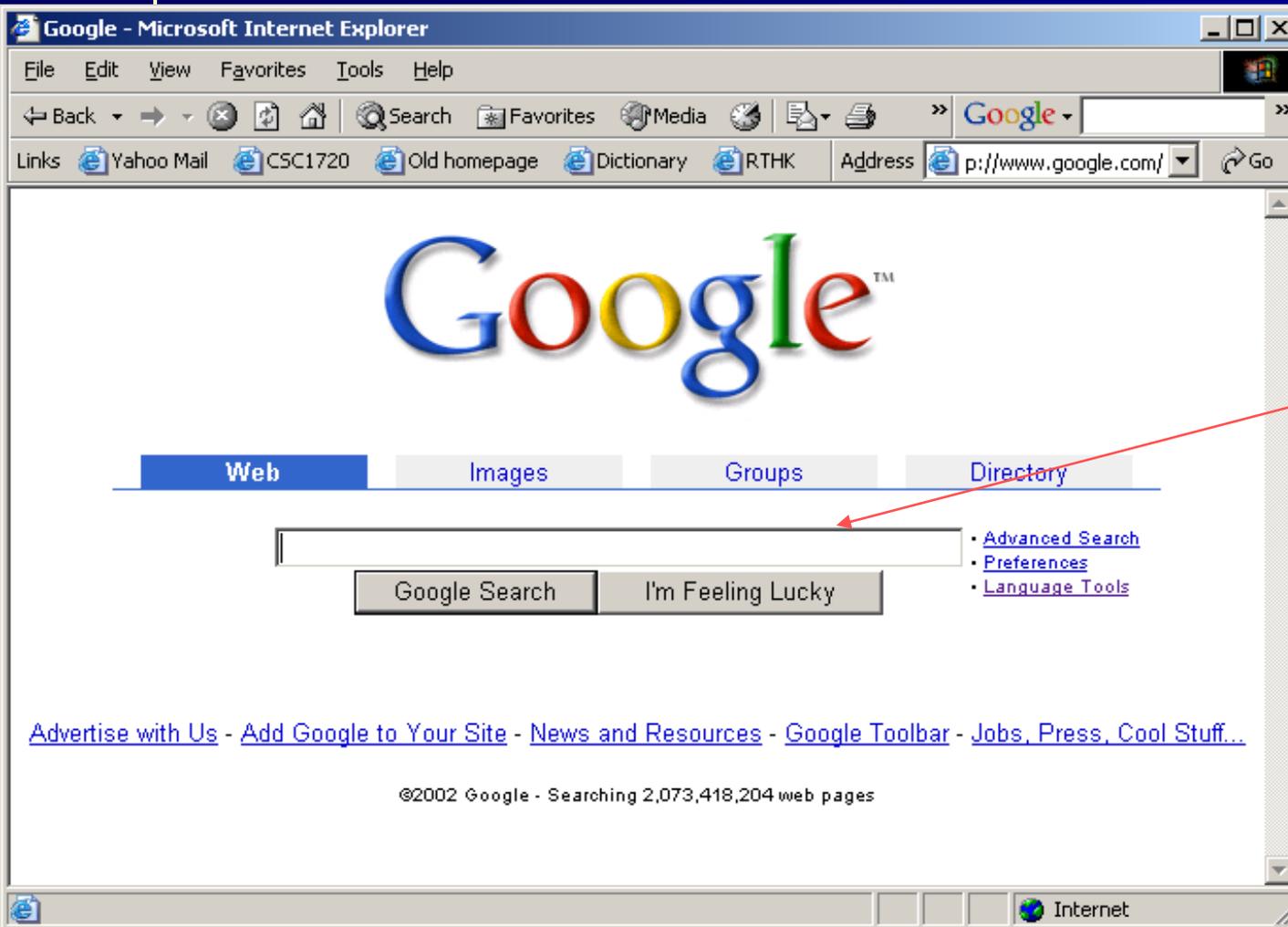
Search Google

- [Add search engines](#)
- [Customize](#)
- [Music](#)
- [Computing](#)

Search Engines

- A search engine is a computer program that does the following:
 - Allows user to submit a query that consists of a word / phase
 - Searches the database
 - Returns a list of suitable URLs which match your query.
 - Allows user to revise and resubmit.

Where to submit Query?



Submit your Query

Popular Search Engines

- AOL anywhere – search.aol.com
- AltaVista – altavista.digital.com
- Excite – www.excite.com
- HotBot – www.hotpot.com
- Magellan – www.mckinley.com
- Google – www.google.com

Metasearch Engines

- A metasearch or all-in-one search engine performs a search by the use of more than one other search engine to complete the search job.
- The duplicate retrievals are eliminated.
- The results are ranked according to how well they match with the query.
- Advantage:
 - A single query can access lot of search engines.
- Disadvantage:
 - A high noise-to-signal ratio, lot of matches will not be suitable for you.

Popular Metasearch Engines

- Metasearch – www.metasearch.com
- Metacrawler – www.metacrawler.com
- MetaFind – www.metafind.com
- Dogpile – www.dogpile.com

Some Figures

Metasearch - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Back Forward Stop Home Search Favorites Media Print Mail News RSS Feeds Google

Links Yahoo Mail CSC1720 Old homepage Dictionary RTHK Address rd+of+the+ring&l=50&s=a&w=w&o=v

METASEARCH

Yahoo!

Search 50

Find matches containing All of these words Any of these words

Match Only as complete words Within other words also



Search

Search the Web Newsgroups

Output: Verbose Standard Terse



Search 25 hits

Find matches containing of these words.

Selectivity:

Output: Verbose Standard Terse

[MetaCrawler®] - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Back Forward Stop Home Search Favorites Media Print Mail News RSS Feeds Google

Links Yahoo Mail CSC1720 Old homepage Dictionary RTHK Address .metacrawler.com/index_metafind.html

Show your true colors. Change your theme here.

Check Email Power Search Customize Put MetaCrawler on Your Site Tools & Tips

Search The Web

any all phrase

Results from: Yellow Pages White Pages Classifieds

Browse the Web's Best Directory

- computing
- entertainment
- lifestyles
- personal
- shopping
- sports
- work & money
- more...

Popular Searches

- [Background Checks](#)
- [Casinos Online](#)
- [Digital Hotels](#)
- [Insurance Quotes](#)
- [Long Distance Mortgages](#)
- [Online Education Refinancing](#)
- [Travel](#)
- [More Searches...](#)

Resources

- [Maps](#)
- [Personals](#)
- [People Search](#)
- [Hot Airfare Deals](#)
- [Photo Personals](#)
- [Domain Names](#)
- [Class Reunion](#)
- [More Resources...](#)

Shopping Tools

- [Go Shopping](#)
- [Product Demos](#)
- [SuperBowl Betting](#)
- [Shop @ eBay.ca](#)
- [Mortgage Quotes](#)

APPLY TODAY for a Credit Card!

VISA MC

Tell a Friend About MetaCrawler Make Us Your Start Page MetaSpy MiniCrawler

DISCOVER THE WAY WE POWER: [Wireless](#) | [Broadband](#) | [Web](#) powered by:

About InfoSpace	Investor Relations	Customer Feedback	World Sites:	Make this your start page
Partner With Us	Press Center	Privacy Policy	United Kingdom	
Advertising	Job Openings	Canada	Canada	

Traded on NASDAQ: INSP © 1996-2000 InfoSpace, Inc. All Rights Reserved. [Terms of Use](#) [MC History](#)
[Sign up for free e-mail](#) [MC Press Center](#) [MC Support](#)

White Pages / Yellow Pages

- White pages allows user to lookup information about individuals.
- We can use white page to track down the telephone numbers, email address.
- People can abuse white pages
- Some people think that white pages are an invasion of their privacy.
- **Yellow pages** contain information about businesses.

Popular White Pages & Yellow Pages

- Bigfoot – www.bigfoot.com
- Yahoo! People Search – people.yahoo.com
- WhoWhere – www.whowhere.com

- Yahoo! Yellow Page – yp.yahoo.com
- SuperPages – www.superpages.com

Some Figures – White Pages

Yahoo! People Search - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Back Forward Stop Home Search Favorites Media Print Mail News RSS Feeds

Links Yahoo Mail CSC1720 Old homepage Dictionary RTHK Address http://people.yahoo

Welcome, Guest User

Yahoo! People Search

Advanced Search People Locate

Merchant Spotlight

Go on a first date that never ends. **YAHOO! Personals** Where millions of singles meet. [click here](#)

Telephone Search

First Name: Last Name (required):

City/Town: State:

Email Search

First Name: Last Name:

[Advanced](#)

Yahoo! Resources

- Yellow Pages
- Address Book
- Yahoo! Mail
- Yahoo! Calendar
- Yahoo! Personals
- Yahoo! Wallet
- Yahoo! Bill Pay
- Yahoo! Travel
- Camp Yahoo!

Can't Find Them on Yahoo!?

powered by US

Search 1000's of Public Databases with one click! Find Addresses, Property Records, Court Records and much, much more...

First: Last:

Street: City:

State: Zip:

Copyright © 2001 Yahoo! Incorporated. All rights reserved.
[Privacy Policy](#) - [Terms of Service](#) - [Help](#)

Welcome to WhoWhere?! - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Back Forward Stop Home Search Favorites Media Print Mail News RSS Feeds

Links Yahoo Mail CSC1720 Old homepage Dictionary RTHK Address http://www.whowhere.lycos.com Go

WhoWhere? ...the way to find people on the web

First Name: Last Name: Search Type:

[Advanced Email Search](#) | [Advanced Phone Search](#) | [Ancestor Search](#) | [Public Records](#)

Your Guide to People on the Web

Find Me

[Add & Update Listing](#)

More People Searches

- Homepages
- WhoWhere? Members
- Celebrities
- Ancestors
- 3D People Finder
- Yellow Pages

[Maps & Directions](#)

Sponsors

Find a Job
Buy & Sell Cars
Find it at eBay!
Personals
FREE Psychic Reading!
Find Old Classmates!
The Best Local Deals
FREE Business Cards!
Find Low Airfares!
FIND INSTANT INFO NOW
Find an Apartment

Background Checks and Public Records

Powered by Online Detective

- DMV Records
- Background Checks
- Public Records
- State Agencies
- Social Security
- Criminal Records
- Military Records
- Driving records

LYCOS PEOPLE GUIDES

People in Entertainment

[Filmmakers](#), [Musicians](#), [Athletes](#)

People and Money

[Investors](#), [Brokers](#), [Shoppers](#), [Executives](#)

People Like You

[Women](#), [Kids](#), [Family](#), [Travelers](#)

People to Help You

[Realtors](#), [Career Advisors](#), [Employers](#)

Other Guides to the Web

[Games](#), [Autos](#), [Computers](#), [Health](#), [Companies](#)

THE LYCOS NETWORK

[Angelfire](#) | [Gamesville](#) | [HotBot](#) | [HotWired](#) | [HTML Gear](#) | [Lycos](#) | [Lycos Shop](#) | [Lycos Mail](#) | [Matchmaker](#) | [My Lycos](#) | [Suck.com](#) | [Sonique](#) | [Tripod](#) | [Webmonkey](#) | [Wired News](#)

What's New

[Get the credit you deserve with a Getsmart Visa!](#)

[Create Your Own Digital Videos!](#)

[Upload and share photos!](#)

Internet

Some Figures – Yellow Pages

Yahoo! Yellow Pages - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Back Forward Stop Search Favorites Media Print Mail News RSS Feeds Google yellow page

Links Yahoo Mail CSC1720 Old homepage Dictionary RTHK Address http://yp.yahoo.com/

Welcome, Guest User [Create My Favorite Locations](#)

Yahoo! Yellow Pages

[Change Location](#)

Feature your business in Yahoo! Yellow Pages

Featured Sponsors

click here for **Jobs**

YAHOO! Careers

[click here for Jobs!](#)

Business Solutions

?/b> [Feature your business](#) in the Yahoo! Yellow Pages.

?/b> [Enhanced web site placement](#) in the Yahoo! Directory

?/b> [Establish an online presence](#) with Yahoo!

• [Yahoo! Express](#) - Suggest your site.

Consumer Yellow Pages

Search by Name or Category (e.g. Hilton, Hotel or Hilton Hotel)

- **Automotive**
[Dealers, Motorcycles, Repair, ...](#)
- **Home and Garden**
[Housewares, Nurseries, Util](#)
- **Community**
[Government, Libraries, Religion, ...](#)
- **Legal and Financial**
[Banks, Insurance, Law Firm](#)
- **Computers and Internet**
[Computer Stores, Internet Services, ...](#)
- **Other Shopping and Services**
[Pets, Books, Flowers, Cellul](#)
- **Education and Instruction**
[Adult Education, Colleges, K-12, ...](#)
- **Personal Care**
[Drug Stores, Hair Care, Nail Salons, ...](#)
- **Entertainment and Arts**
[Nightclubs, Movies, Music, ...](#)
- **Real Estate**
[Agents, Rental Agencies, ...](#)
- **Food and Dining**
[Catering, Grocers, Restaurants, ...](#)
- **Recreation and Sport**
[Golf, Hobbies, Sporting Good](#)
- **Health and Medicine**
[Hospitals, Doctors, Fitness, ...](#)
- **Travel and Transport**
[Hotels, Taxis, Travel Agents](#)

Copyright © 2002 Yahoo! Inc. All rights reserved. [Help](#) - [Privacy Policy](#) - [Terms of Service](#)

SuperPages.com: Yellow Pages - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Back Forward Stop Search Favorites Media Print Mail News RSS Feeds Google yellow page

Links Yahoo Mail CSC1720 Old homepage Dictionary RTHK Address es.com/sform.phtml?SRC=&STYPE=S Go

verizon

SUPERPAGES.COM

BETTER PAGES FOR BETTER DECISIONS

english | spanish

HOME PAGE MY DIRECTORY YELLOW PAGES PEOPLE PAGES SUPER TOPICS CONSUMER GUIDE CITY PAGES ABOUT & HEL

YELLOW PAGES

Home → [Yellow Pages](#) → Simple Search

U.S. YELLOW PAGES

SIMPLE SEARCH

1 **Category** ([Browse](#)) OR **Business Name**

2 **City** (optional) **State** (required)

3 [Search Tips](#)

Detailed Search

Search using almost any combination of business name, category, street address, city, state, ZIP, or phone number.

Search By Distance

Search for businesses within a certain distance from a location you specify or that you saved in your Directory.

Nationwide Searches

Search Nationwide by:

- [Business Name](#)
- [Telephone Number](#)

My Directory

Quickstart

Category Browser

Apparel & Discount Stores Clothing, Accessories, Department Stores, ...	Government & Municipal Services Government Offices, Libraries, Police, ...
Art, Music, Books, & Photography Art Galleries, Art Supplies, Videos, ...	Home & Office Appliances, Gardening, Pets, Office Supplies, ...
Automotive Dealers, Rentals, Parts, Repair, ...	Media & Advertising Marketing Consultants, Printers

Comparison

Directory	Search Engine
A directory allows you to explore and get what you want eventually.	A search engine brings you to the exact page on the words or phrases you are looking for.
Use a directory to find cooking-related websites.	Use a search engine to find a specific recipe, by providing the name of the ingredients.
Use a directory to find travel guides in a country.	Use a search engine to find the transport trains schedule in South Africa.

Search Fundamentals

- Example: www.yahoo.com
- **Header:** Yahoo Logo and some advertising.
- **Information bar:** contains other hyperlinks.
- **Search form area:** consists a form which allows you to type a query.
- **Directory area:** a large number of categories, channels.
- **Yahoo Links:** Link to other yahoo sites.
- **Footer:** contains information about yahoo, copyright and a disclaimer.

Search Fundamentals

The screenshot shows the Yahoo! homepage in Microsoft Internet Explorer. The browser title is "Yahoo! - Microsoft Internet Explorer". The page features a navigation bar with icons for Finance, Messenger, Check Email, What's New, Personalize, and Help. Below this is a search bar with the text "Search" and a link to "advanced search".

The main content area is divided into several sections:

- 2002 Olympic Games:** Hockey, Figure Skating, Speed Skating.
- LOSE 10 POUNDS:** Tax Center tips, forms, online filing.
- Tax Center:** Online Filing with Turbo Tax, Federal Forms, State Forms, Calendar.
- Shop:** Auctions, Autos, Classifieds, Shopping, Travel, Yellow Pgs, Maps, Media, Finance/Quotes, News, Sports, Weather, Connect, Careers, Chat, Clubs, GeoCities, Greetings, Mail, Members, Messenger, Mobile, Personals, People Search, Photos, Personal, Addr Book, Briefcase, Calendar, My Yahoo!, PayDirect, Fun, Games, Kids, Movies, Music, Radio, TV, more...
- Make a Connection with Yahoo! Personals:** A form to find a match based on gender (Man/Woman), city, state, or ZIP.
- In the News:**
 - Italy thwarts cyanide terror plot
 - Ex-aid to Mullah Omar captured
 - Millions of Muslims mass for Hajj
 - ISS astronauts take spacewalk
 - 'X-Files' finale to star Duchovny
 - 2002 Winter Olympic Games
- Marketplace:**
 - Y! Exclusive - digital cam, web cam, camcorder all in one, only \$49.99 - Save 43%
 - Sony 27" WEGA TV - Only \$16/month, free in-home delivery
 - FDNY Brotherhood Auction - Bid on signed photos from the book
 - Olympic Gear - Official US Team Store
- Broadcast Events:**
 - Jodie Foster's Panic Room sneak peek
 - 8pm ET Alabama vs. Arkansas
 - 8pm Oklahoma St vs. Texas
- Inside Yahoo!:**
 - Play free Fantasy Auto Racing
 - GeoCities - build your own web site
 - Y! Games - pool, literati, spades, chess, dominoes, euchre...
 - Astrology - what's your sign?
 - Make Yahoo! your home page

The right sidebar contains a "powered by COMPAQ" logo and a "Local Yahoos" section with links to various regional sites (Europe, Asia Pacific, Americas, U.S. Cities). Below this is a "More Yahoos" section with links to various services like Buzz Index, Education, Health, Outdoors, Pets, Real Estate, Yahoo!igans!, Entertainment, Horoscopes, Broadcast, Games, Movies, Music, Radio, Tickets, TV, more, Finance, Banking, Bill Pay, Money Manager, Insurance, Loans, Taxes, FinanceVision, more, Local, Autos, Careers, Classifieds, Events, Lodging, Maps, Yellow Pages, more, News, Top Stories, Business, Entertainment, Lottery, Sports, Technology, Weather, Publishing, Briefcase, Clubs, Experts, Invites, Photos, Home Pages, Message Boards, Small Business, Domain Registration, Small Biz Center, Store Building, Web Hosting, Access Yahoo! via: Pagers, PDAs, Web-enabled Phones and Voice (1-800-My-Yahoo!).

At the bottom, there is a "Make Yahoo! your home page" link, a "How to Suggest a Site" link, and a "Privacy Policy" link. The copyright notice reads: "Copyright © 2002 Yahoo! Inc. All rights reserved. Privacy Policy".

Search Terminology

- **Search Tool:** Any mean to locating information on the Internet.
- **Query:** Information typed into the form on the search engine.
- **Query syntax:** Rules for constructing a valid query.
- **Query semantics:** Rules for defining the meaning of a query.
- **Hit/Match:** A URL that the search engine returns for a specific query.
- **Relevancy score:** A value that indicates the quality of the URL (match close to the query 1 to 100).

Pattern Matching Queries

- It is also called Fuzzy Query.
- You can enter “ungrammatical sentences”, “incomplete sentence fragments”, “disjoint phrases”, “nonsense words”.
- The search engine gets a collection of keywords.
- Required keyword: Mark with “+” before the keyword.
- Prohibited keyword: Mark with “-” before the keyword.

Pattern Matching Queries

Google Search: "+Chocolate mousse" +servings - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Back Forward Stop Home Search Favorites Media Print Mail News Groups RSS Feeds Settings Profile

Address 22+%2Bservings Go

Links Yahoo Mail CSC1720 Old homepage Dictionary RTHK Google "ocolate mousse" +servings Search Web

Google™ Advanced Search Preferences Language Tools Search Tips

+ "Chocolate mousse" +servings Google Search

Web Images Groups Directory

Searched the web for + "Chocolate mousse" +servings Results 1 - 10 of about 3,920. Search took 0.12 seconds.

[Passover Chocolate Mousse Pie](#)
... Passover **Chocolate Mousse** Pie Submitted by: Sara, ,
Fresh Flowers direct from the grower. Makes 8 **servings**. ...
www.holidayrecipe.com/AZ/PassoverChocolateMoussePie.asp - 33k - [Cached](#) - [Similar pages](#)

[Chocolate Mousse Cake II](#)
... to **servings** About Recipe Scaling. Find Similar Recipes 9 inch springform pan
www.cakerecipe.com/AZ/ChocolateMousseCakell.asp - 35k - [Cached](#) - [Similar pages](#)

[Chocolate Mousse Cake IV](#)
... Scale this Recipe to **servings** About Recipe Scaling. Find Similar Recipes 9 inch round pan
www.cakerecipe.com/az/ChcltMssCkIV.asp - 33k - [Cached](#) - [Similar pages](#)
[\[More results from www.cakerecipe.com \]](#)

[Welcome to Dessert Gallery](#)
... a thick layer of fudge topped with white **chocolate mousse**. Cookies 'N Cream Cheesecake (12-14 **servings**) ... \$38.50 [pic] Creamy filling with chunks of Oreos in ...
www.dessertgallery.com/menu_dessert.htm - 21k - [Cached](#) - [Similar pages](#)

[Hershey's \(R\) Chocolate Mousse](#)
... with cocoa, whipped with gelatin and chilled for a cool **chocolate mousse**. " , ... Makes 2 cups Scale this Recipe to **servings** About Recipe Scaling. ...
dessert.allrecipes.com/AZ/HrshysChcltMss.asp - 32k - [Cached](#) - [Similar pages](#)

[Blender Chocolate Mousse](#)

Internet

Boolean Queries

- A Boolean Query is a query that consists keywords but with logical operators (AND, OR, NOT).
- **X AND Y** – will return URLs that contain both X and Y.
- **X OR Y** – will return URLs that contain either X or Y.
- **X AND NOT Y** – will return URLs that contain X and do not contain Y.
- Symbol: **AND** - **&**, **OR** - **|**, **NOT** - **!**, **NEAR** - **~**

Boolean Queries

- **AND is used for narrowing a query**
 - If you know that your target documents will contain a group of keywords, list them using the AND operator
- **OR is used for broadening a query**
 - If you can think of related words for a topic, list them using the OR operator
- **NOT is used to redirect a query**
 - If you find that a keyword or phrase is leading irrelevant hits, then represent it in your query as **AND NOT *keyword***

Boolean Queries

Google Search: (Hackers OR Hacker) AND CNN - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Back Forward Stop Home Search Favorites Media Print Mail News Groups People Address http:// Go

Links Yahoo Mail CSC1720 Old homepage Dictionary RTHK Google rankings-college

Google™ (Hackers OR Hacker) AND CNN Google Search

Advanced Search Preferences Language Tools Search Tips

The "AND" operator is unnecessary -- we include all search terms by default. [\[details\]](#)

Web Images Groups Directory

Searched the web for (Hackers OR Hacker) AND CNN. Results 1 - 10 of about 74,800. Search took 0.05 seconds.

CNN In-Depth Specials - Hackers
... underground'. Friday night in the mall with the 'hacker underground' Full story. ... Recent stories on **hackers**: Search, **CNN.com**, www.cnn.com/TECH/specials/hackers/ - 18k - [Cached](#) - [Similar pages](#)

www.cnn.com/2002/TECH/ptech/01/04/hacking.home.computers.ap/
[Similar pages](#)
[[More results from www.cnn.com](#)]

Inforeading - Hacker Articles: CNN
... university users' passwords - November 6, 1998; **CNN** - NY Times Web page back online after **hacker** attack - September 13, 1998; **Hackers** break into NY Times Web ... www.inforeading.com/archive/hacker_articles/cnn.php3 - 47k - [Cached](#) - [Similar pages](#)

Google Directory - News > Online Archives > CNN.com > 1996 ...
... Web Pages. Viewing in Google PageRank order View in alphabetical order. **Hackers** vandalize CIA home page - <http://www.cnn.com/TECH/9609/19/cia.hacker/> [**CNN.com**]. ... directory.google.com/Top/News/Online_Archives/CNN.com/1996/September/Technology/ - 30k - [Cached](#) - [Similar pages](#)

Hackers: Re: CNN - The Rare Glitch Project The legendary ...
Re: **CNN** - The Rare Glitch Project The legendary search for a stable version of Wind. ... www.seds.org/UASEDS/mail/hackers/0592.html - 5k - [Cached](#) - [Similar pages](#)

Done Internet

Using Wildcards

- Wildcards are useful for retrieving variations of a word
- For example, art* will search for art, artwork, artist, artistry, and so forth
- An excellent way to broaden a search
- Different wildcard characters are used by different search engines
- The most common characters are: *, #, and ?

Advanced Search Options

AltaVista Advanced Search Cheat Sheet - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Back Forward Stop Home Search Favorites Media Print Mail News RSS Feeds

Links Yahoo Mail CSC1720 Old homepage » Google

Advanced Search Cheat Sheet

AND	Finds documents containing all of the specified words or phrases. Peanut AND butter finds documents with both the word peanut and word butter.
OR	Finds documents containing at least one of the specified words or phrases. Peanut OR butter finds documents containing either peanut or butter. The found documents could contain both items, but not necessarily.
AND NOT	Excludes documents containing the specified word or phrase. Peanut AND NOT butter finds documents with peanut but not containing butter. NOT must be used with another operator, like AND. AltaVista does not accept 'peanut NOT butter', instead, specify peanut AND NOT butter .
NEAR	Finds documents containing both specified words or phrases with words of each other. Peanut NEAR butter would find documents containing peanut butter, but probably not any other kind of butter.
()	Use parentheses to group complex Boolean phrases. For example, (peanut AND butter) AND (jelly OR jam) finds documents with words 'peanut butter and jelly' or 'peanut butter and jam' or both.
anchor:text	Finds pages that contain the specified word or phrase in the text of a hyperlink. anchor:job +programming would find pages with the word job in a link and with the word programming in the content of the page. Do not put a space before or after the colon. You must repeat the keyword to search for more than one word or phrase; for example anchor:job OR anchor:career to find pages with anchors containing either the word job or the word career.

Done Internet

AltaVista Advanced Search Cheat Sheet - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Back Forward Stop Home Search Favorites Media Print Mail News RSS Feeds

Links Yahoo Mail CSC1720 Old homepage » Google

applet:classname	Finds pages that contain a specified Java applet. Use applet:morph to find pages using applets called morph.
domain:domainname	Finds pages within the specified domain. Use domain:uk to find pages from the United Kingdom, or use domain:com to find pages from commercial sites.
host:hostname	Finds pages on a specific computer. The search host:www.shopping.com would find pages on the Shopping.com computer, and host:dilbert.unitedmedia.com would find pages on the computer called dilbert at unitedmedia.com.
image:filename	Finds pages with images having a specific filename. Use image:beaches to find pages with images called beaches.
like:URLtext	Finds pages similar to or related to the specified URL. For example, like:www.abebooks.com finds Web sites that sell used and rare books, similar to the www.abebooks site. like:sfpl.lib.ca.us/ finds public and university library sites. like:http://www.indiax.com/ finds sites about culture on the Indian subcontinent.
link:URLtext	Finds pages with a link to a page with the specified URL text. Use link:www.myway.com to find all pages linking to myway.com.
text:text	Finds pages that contain the specified text in any part of the page other than an image tag, link, or URL. The search text:graduation would find all pages with the term graduation in them.
title:text	Finds pages that contain the specified word or phrase in the page title (which appears in the title bar of most browsers). The search title:sunset would find pages with sunset in the title.
url:text	Finds pages with a specific word or phrase in the URL. Use url:garden to find all pages on all servers that have the word garden anywhere in the host name, path, or filename.

Done Internet

Break Time – 10 minutes



Search Strategies

- You should find a search engine that meets the following conditions:
 - A user-friendly interface
 - Easy-to-understand documentation
 - Convenient to access
 - A large indexed database
 - Assigning good relevancy scores.
- Learn the syntax of this particular search engine, but not several different engines.

Search Generalization

- Too few hits?
 - Needs to generalize your search query.
- Pattern matching query: eliminate one of the more specific keywords of the query.
- Boolean query: remove the keywords with **AND** operator, or delete the **NOT** item, or use the **OR** operator.
- Use a directory or metasearch engine if still cannot locate the matched URL.

Search Specialization

- Too many hits?
 - Needs to specialize your search query.
- Pattern matching query: add more keywords.
- Boolean query: use **AND** with other keyword, or add **NOT** operator to excluded some unwanted pages.
- Try capitalizing proper nouns or names.
- Use a directory to locate your information.

Sample Searches

- Queries about Kayaking in Alaska
- Example: Using **infoseek**

Query:	No. of Hits
alaska	176,954
Alaska	176,064
+”Prince William Sound” +Alaska	778
+kayak +”Prince William Sound” +Alaska	44
+kayaking +”Prince William Sound” +Alaska	60
+kayaking +”Prince William Sound” +Alaska +rental	20

How does it work?

- **User Interface** – Allows you to type a query and displays the results.
- **Searcher** – The engine searches the database for matching your query.
- **Evaluator** – The engine assigns scores to the retrieved information.
- **Gatherer** – The component that travels the WEB, and collects information.
- **Indexer** – The engine that categorizes the data collected by the gatherer.

User Interface

- Provides a mechanism for a user to submit queries to the search engine.
- Uses forms, very user friendly.
- The user interface displays the search results in a convenient way.
- A summary of each matched page is shown.

Searcher



- It is a program that uses the search engine's database to locate the matches for a specific query.
- The database of a search engine holds extremely large indexed pages.
- A highly efficient search algorithm is necessary.
- Computer Scientists have spent years to develop the searching and sorting methods.
- You can refer to computer books.

Evaluator

- The searcher returns a set of URLs that match your query.
- Not all of the hits equally match your query.
- More references to the page, the ranking of the page will be higher.
- How the relevancy score is calculated?
 - Varies from one engine to another one.
 - The number of times of the word appears?
 - The query words appear in the title?
 - The query words appear in the META tag?

Link Popularity

LinkPopularity.com
THE FREE LINK POPULARITY SERVICE

What is link popularity?
Link popularity is the total number of web sites that link to your site.

Why is it important?
Because good link popularity can dramatically increase traffic to your web site. Well placed links are an excellent source of consistent and targeted traffic. And due to recent developments, they can even generate

Helpful Hints:
Your web pages are optimized for search engines but they still rank low and you can't figure out why?
Search here for higher ranking web sites and see if the answer is LinkPopularity.

reference

Gatherer

- It is a program that traverses the Web and gathers information about the Web documents.
- It runs at a short and regular intervals.
- It returns information and will be indexed to the database.
- Alternate names: Bot, Crawler, Robot, Spider and Worm.

Spiderlist

■ <http://www.spiderhunter.com/>

Spider Hunter: Learn to write cloaking scripts and track spiders - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Address <http://www.spiderhunter.com/spiderlist/index.cfm?FuseAction=Results&Key>

 **Spider Hunter.com**

SimpleApps.net
Co-branded applications on your own domain name

 Calendar
 Briefcase
 Contacts
 Email

[Compare Scripts](#)
[Ethics](#)
[Free Scripts](#)
[Links](#)
[Non-Spiders](#)
[Pseudocode](#)
[Spider Scope](#)
[Spiders](#)
[Spider List](#)
[Spider Trap](#)
[Tools](#)
[Tutorial](#)

[Help Support this site](#)

[Donations](#)

Search at Interrogative



Results

Spider IP	Spider DNS Name	Search Engine	Spider User Agent
209.185.253.175	crawler.googlebot.com	Google	Googlebot/2.1 (+http://googlebot.com/bot.html)
209.185.108.147	a10.google.com	Google	Googlebot/1.0 (googlebot(at)googlebot.com)
209.185.108.155	a18.google.com	Google	Googlebot/1.0 (googlebot(at)googlebot.com)
209.185.108.140	a3.google.com	Google	Googlebot/1.0 (googlebot(at)googlebot.com)
209.185.108.141	a4.google.com	Google	Googlebot/1.0 (googlebot(at)googlebot.com)
209.185.108.143	a6.google.com	Google	Googlebot/1.0 (googlebot(at)googlebot.com)
209.185.108.145	a8.google.com	Google	Googlebot/1.0 (googlebot(at)googlebot.com)
209.185.108.148	a11.google.com	Google	Googlebot/1.0 (googlebot(at)googlebot.com)
209.185.108.150	a13.google.com	Google	Googlebot/1.0 (googlebot(at)googlebot.com)
209.185.108.152	a15.google.com	Google	Googlebot/1.0 (googlebot(at)googlebot.com)
209.185.108.154	a17.google.com	Google	Googlebot/1.0 (googlebot(at)googlebot.com)
209.185.108.157	c3.google.com	Google	Googlebot/1.0 (googlebot(at)googlebot.com)
209.185.108.159	a19.google.com	Google	Googlebot/1.0 (googlebot(at)googlebot.com)
209.185.108.161	c4.google.com	Google	Googlebot/1.0 (googlebot(at)googlebot.com)
209.185.108.164	a23.google.com	Google	Googlebot/1.0 (googlebot(at)googlebot.com)
209.185.108.165	a24.google.com	Google	Googlebot/1.0 (googlebot(at)googlebot.com)

Internet

Indexer

- It organizes the data by creating a set of keys or an index.
- Indexes need to be rebuilt frequently.
- E.g. Libraries – Author, Title, ISBN, etc...
- In order to ensure the returned URL is not out of date.
- The search engine is very complex and needs to break down into different components.

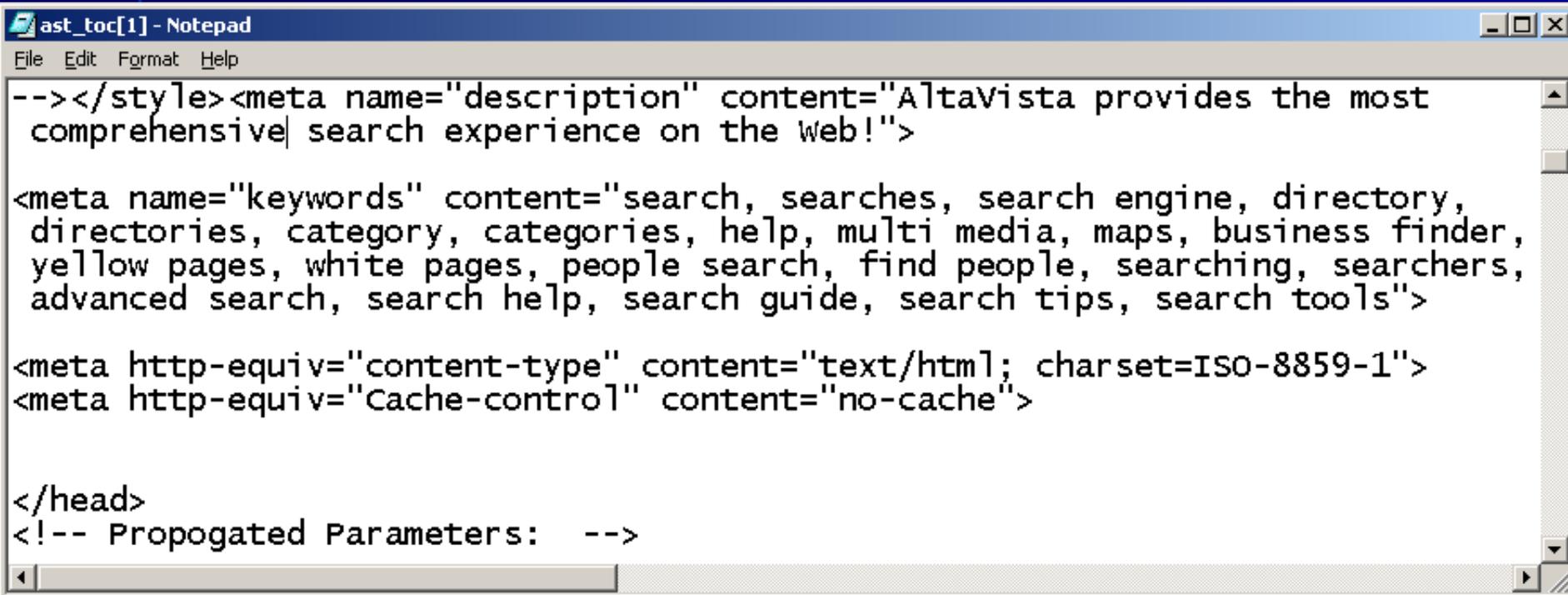
Case Study - AltaVista

- Sending out Crawlers (robot programs) that capture information from the web and bring them back.
- The main crawler – “Scooter” simultaneously send out HTTP requests like blind users on the Web.
- Store all these information to the indexing engine.
- Scooter’s cousins help to remove “dead” links.
- A typical day, Scooter will visit over 10 million pages.
- Web pages with no links referencing will never be found.
- You can also submit your URL to AltaVista.

Case Study - AltaVista

- **METAtags** – special keywords embedded in the headers of the webpage.
- **Full-text index** – Every word on every page is also included during searching.
- AltaVista is using Full-text indexing.

METAtag Example

A screenshot of a Notepad window titled "ast_toc[1] - Notepad". The window contains HTML code for meta tags. The code is as follows:

```
--></style><meta name="description" content="AltaVista provides the most  
comprehensive search experience on the Web!">  
  
<meta name="keywords" content="search, searches, search engine, directory,  
directories, category, categories, help, multi media, maps, business finder,  
yellow pages, white pages, people search, find people, searching, searchers,  
advanced search, search help, search guide, search tips, search tools">  
  
<meta http-equiv="content-type" content="text/html; charset=ISO-8859-1">  
<meta http-equiv="Cache-control" content="no-cache">  
  
</head>  
<!-- Propogated Parameters: -->
```

Case Study - AltaVista

- Limit a search to a domain
- E.g. searching “edu” domain
- +domain:edu +”molecular biophysics”
- The above query would only search for molecular biophysics at educational institutions.
- Here is a list of Top-level Internet Domains

Searching Tips

■ Be natural

- Is cell phone harmful?
- Ask the search engine : “Cell phone” AND harmful

■ Capitalize

- Always use lowercase
- star will search “Star, STAR, stAr, ...”
- Type “Star” unless you really want to search “Star”.

Searching Tips

- **Use uncommon keywords**
 - The more specific results will return to you.
 - Think a valid and uncommon keyword.
- **Require words**
 - Add a “+” before the keyword.
 - It will be in every match.
- **Exclude words**
 - Use “-” before the keyword.
 - In what situation should we use?

Searching Tips

■ Correct Spelling

- Beware of the differences between English and American spellings (Color, Colour) → (color OR colour)

■ Stop words

- Ignore the most common words “the, is, ...”
- “searching the web” and the search engine will ignore “the web”.
- Add more relevant keyword.

Searching Tips

■ Use wildcards

- Use “*” in some search engines.
- “funk*” → funk, funky, funkiest, ...

■ Solve dead links

- If the search engine returns <http://www.hit.com/a/b/c.html> which is a dead link.
- You can try <http://www.hit.com/a/b/>
- Or <http://www.hit.com/a/> ...

Factors affect your site's Ranking

- Keyword prominence
- Keyword frequency
- Keyword weight
- Keyword proximity
- Keyword placement
- Click popularity & Stickiness

Keyword Prominence

- How early in a web site do the keywords first appear?
 - The first element in HTML is the title tag
 - What happen if your title is:
 - This is my homepage
 - Welcome to my company's homepage
- Include the keywords in head, Meta tag, early in the body, ...

Keyword Frequency

- Search engine may determine your site's popularity by checking how frequently the keyword or phrase appears on the page.
- What is the problem if you put too many same keywords into one single page?

Keyword Weight

- It is also called keyword density
- Measure by comparing the number of keywords appearing on the web page with the total number of words on the page.
- In most case, we try not to exceed a keyword weight of 3 to 10 percent.

Keyword Density

Webjectives keyword density analyzer - Microsoft Internet Explorer

File Edit View Favorites Tools » Address <http://www.webjectives.com/keyword> Go

Back Forward Stop Home Search Favorites Links [CSC1720 Introduction to Internet](#)

- home
- about webjectives
- evaluations
- development
- promotion
- merchant accounts
- position exact
- request a proposal
- resources
- clients
- press
- career opportunities
- contact webjectives

keyword density analyzer

Webjectives Keyword Density Analyzer Version 2.0

Use this tool to analyze the pages of your competition to determine the optimum and search word or phrase (without quotes). Press the analysis.

URL: (eg. <http://www.webjectives.com>)

keyword / phrase:

Perform partial keyword analysis

Case sensitive (treat "test" and "Test" as different words)

Done

[reference](#)

[reference](#)

Keyword Density Analyzer - Microsoft Internet Explorer

File Edit View Favorites Tools » Address <http://www.keyworddensity.com/> Go

Back Forward Stop Home Search Favorites Links [CSC1720 Introduction to Internet](#) [RTHK - TV](#)

Keyword Density Analyzer

keyword density and word depth calculator

Fast free analysis of any web page

Enter URL-1:

Enter URL-2:

Enter Keyword or Keyphrase:

Raw Data Comparison: **Customized Analysis:** **Case Sensitive:**

Find popular [keywords for your website](#)

[Internet Marketing Coach](#)

Personalized Guidance & Advice for Motivated Webmasters

[Internet Marketing](#)

High search engine ranking requires your...

Done

Keyword Proximity

- The placement of keywords on a web page in relation to each other is measured in “Keyword Proximity”.
- “Home loans” will outrank a citation about “home mortgage loans”.
- E.g.
 - *Smith Brothers Inc has been selling **puppy food** for over 50 years.*
 - *Smith Brothers Inc has been selling **food** for your **puppies** for over 50 years.*

Keyword Placement

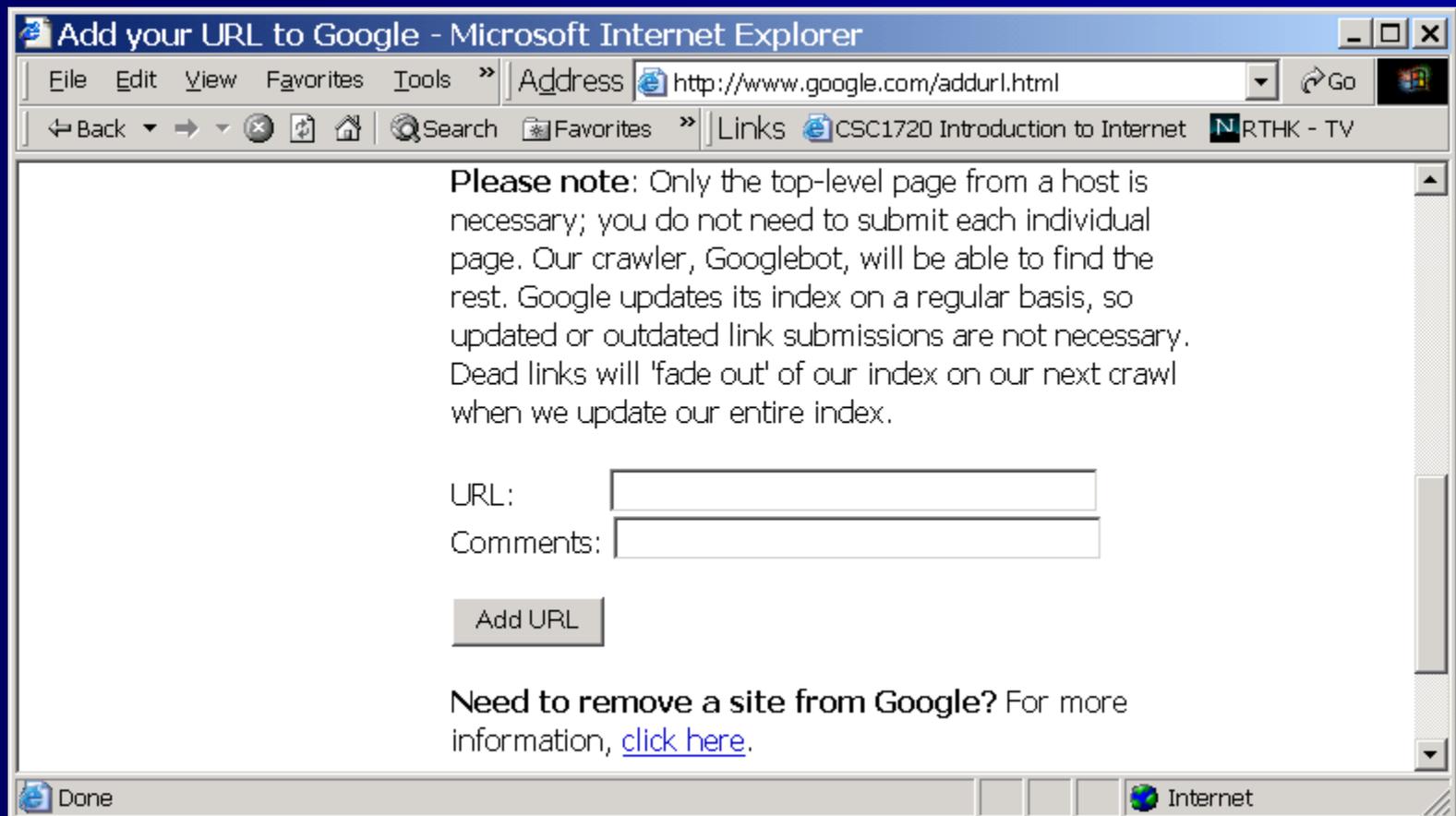
- Search engines favor web sites that contain keywords in:
 - The title tag
 - The keyword META tag
 - The headline tag <H1> ...
 - The first 25 words of body
 - Hyperlinks
 - Image <ALT> tags
 - Text near the end of the document

Click popularity & Stickiness

- **Click popularity** is a measure of the number of clicks received by each site in a search engine's results page.
- **Stickiness** is a measure of the amount of time a user spends at a site. It's calculated according to the time that elapses between each of the user's clicks on the search engine's results page.
- Reference: <http://www.directhit.com/>

Submit your site to search engines

- Google – 5 pages/day, Excite – 25 pages/week



Please note: Only the top-level page from a host is necessary; you do not need to submit each individual page. Our crawler, Googlebot, will be able to find the rest. Google updates its index on a regular basis, so updated or outdated link submissions are not necessary. Dead links will 'fade out' of our index on our next crawl when we update our entire index.

URL:

Comments:

Need to remove a site from Google? For more information, [click here](#).

Done Internet 2003.

Summary

- Use different resources to find/search different kinds of information.
- Use successive query refinement to achieve effective search queries.
- Think carefully for the keywords typed in the search engine.
- Use Boolean queries when you need combinations of keywords.
- Think carefully when you create your own homepage, can it be easily indexed by search engines?

References

- searchenginewatch.com
- Information retrieval
- Search Engine Positioning – Fredrick Marckini (Wordware Publishing Inc.)

- The End.
- Thank you for your patience!

