

Introduction to Search Engines: Six Lessons for Successful Searching

David King
Web Project Manager
Kansas City Public Library

Overview

- Lesson 1: What is the Web?
- Lesson 2: Search Directories
- Lesson 3: Search Engines
- Lesson 4: Specialized Search Engines
- Lesson 5: Evaluate what you find!
- Lesson 6: What to Use?

Lesson 1: What is the Web?

- Definition of the web
- Definition of a graphical browser
- Why is the web Popular?
- Who owns the Internet?



Lesson 1: What is the Web?

- Strengths
- weaknesses
- Why can't I find what I want?



Definition of the Web:

- Interconnected network of computer networks – hence, the “web.”
- A bunch of computer files hooked together
- Each page has a unique address (URL)

Definition of a Graphical Browser:

- Computer program that requests information from other computers
- Graphical User Interface, or GUI's – grab pictures, text, etc.

Why is the Web Popular?

- Cheap computers
- Fast, reliable Internet connections – for graphics
- Commercial/corporate world jumped in
- It's fun – you don't have to read it like a book.

Who owns the Internet?

- No one.
- No one is in charge of policing it, either.
- Self publishing medium

Strengths:

- Consumer information
- Entertainment
- Government information
- Recent research
- Online communities
- News and current events

Weaknesses, or “The Web Is Not:”

- Comprehensive or all-inclusive in coverage
- A substitute for peer-reviewed articles (compliments rather than substitutes)
- Always reliable
- Censored

Why can't I find what I want?

- Searcher error:
 - Incorrect spelling/typos: kriseler/crysler
 - Poorly described or conceptualized topic: something on cars/automotive industry and government regulations
 - Query is too general: aids

Why can't I find what I want?

- Query is too specific: aids symptom triggered by eating an olive while sitting in a cafeteria in Munich.
- Search syntax is wrong or does not perform the way you expect: use HELP on Search Engine sites to guide your search syntax

Beyond Your Control

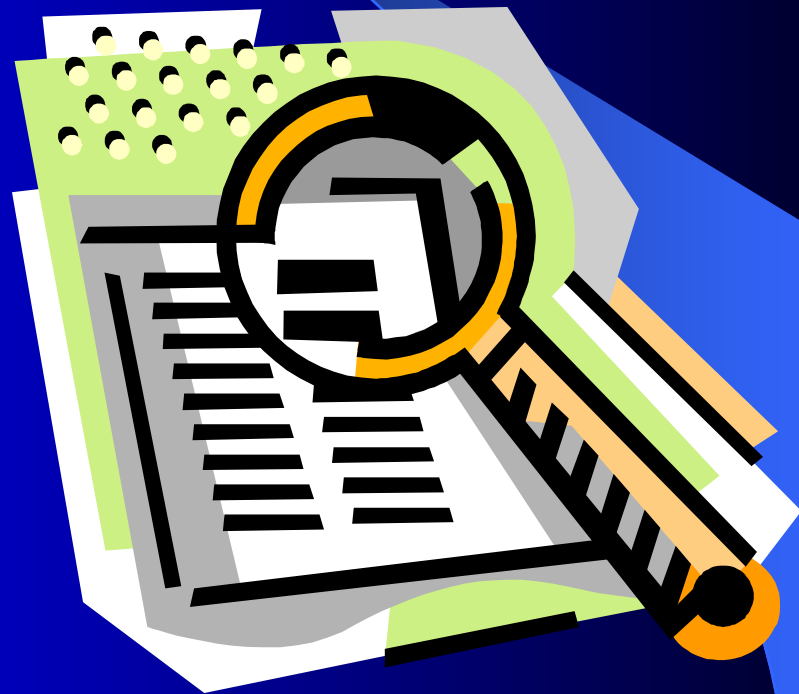
- Remember – you're searching a search engine's database, rather than the web itself
- Searching only 1/3 of the web, at best

Beyond Your Control

- Hits match search terms, but in the wrong way
 - more than one meaning of a word
 - same words are found, but in different order
 - Could be using wrong search tool (FindLaw vs. Yahoo)

Lesson 2: Search Directories

- What is a Web Directory?
- Difference between directories and search engines
- Examples of directories



What is a Web Directory?

- A listing of subjects organized into categories
 - Usually arranged hierarchically
 - Example – looking for info on guitars:
entertainment/music/instruments/stringed
instruments/guitar
- Differences between directories and engines?

Directories

- Compiled by people
- Browsable
- Looking for quality
- Search using controlled vocabulary
- Use when not sure of topic
- Use when beginning a search

Search Engines

- Automated compilers
- Searchable
- Looking for quantity
- Search using keywords
- Use when topic is known

Examples of directories:

- Yahoo (dir.yahoo.com)
- dmoz (www.dmoz.org)
- Librarian's Index to the Internet
(www.iii.org)
- LookSmart (www.looksmart.com)

Assignments:

- Browse a directory for a specific topic
- See how many pages are indexed for that particular topic

Lesson 3: Search Engines



- What's a search engine?
- Relevance Ranking

What's a search engine?

- Bot, spider, intelligent agent, crawler, robot worm
- A computer program that looks for websites.
- Remembers URLs

What's a search engine?

- Remembers words on pages and their position within the text.
- When searching – you search the index – NOT the web.
- Some are very large – 8 billion web pages.

Relevance Ranking

- Algorithm that calculates which pages appear first in search results.
- Calculated by noting where the term occurs within the page and assigning this position a weight or level of importance
- Terms appearing in title, summary, key paragraph positions, or appearing several times within the text usually carry more weight

Each search engine:

- Has a different ranking scheme
- Frequently changes their ranking scheme
- Keeps their ranking methodology a secret – it's a business, after all

Examples of Search Engines:

- Google (www.google.com)
- Yahoo! (www.yahoo.com)
- Ask Jeeves (www.ask.com)

Exercises

- Search for a specific topic in 2-3 different search engines
- Compare results of searches

Lesson 4: Specialized Search Engines

- What's a specialized search Engine?
- Why use these?



What's a Specialized Search Engine?

- Focuses on a specific subject, a geographic region, or a certain type of computer file format
- Index fewer web pages
- Weed out off-topic information
- On-target pages will more likely be found at the top of one's search

More “Human Interaction.”

- Many specialized search engines hire subject specialists who gather, rank, and annotate each link.
- Entries are weeded in order to be subject-specific, and those weeded entries are winnowed even further so that only “useful information” is left.

Why use these?

- Saves Time
 - focuses on a niche subject
 - there's a good chance you'll find what you need in less time, since all hits found will focus on your topic
- Pre-weeded Hits
 - manually choose entries
 - you'll be saved the frustration of finding a page created by a 6-year-old

Why Use These?

- Unique Entries
 - gather useful sites from user submissions
 - Some of these entries won't be found on major search engines
- Annotations
 - Many specialized search engines annotate and, in some cases, rank each site

Examples:

- Blogs: Feedster (www.feedster.com)
- Legal: FindLaw (findlaw.com)
- Where to find more: CompletePlanet

Lesson 5: Evaluate what you find!



- Why Evaluate?
- What to look for

Why evaluate?

- The web is self-publishing
- Anyone can web-publish
- No such thing as the Internet Police
- No such thing as the Internet Editor

What to look for?

- Who sponsored or created the site?
 - Helps find bias (NRA on gun control)
 - Authenticity/Credibility – health on Mayo Clinic site
- Domains:
 - edu – educational
 - com – commercial
 - gov – government
 - org – organization

What to look for?

- Individual's home page – often has a tilde (~)
- Author/sponsor in footer/header of page
- Contact information
- Why is the site there?
 - Is it PR? Is it a web catalog?
 - Look in “about the website”
- Intended audience

What to look for?

- Check language – scholarly or consumer?
- Revision dates
 - Frequent updates are generally good.
 - Depends on the info presented.

What to look for?

- Good navigation
- Grammar/spelling
- Professional-looking rather than annoying color schemes, etc.

Exercises

- Look at 2-3 web sites
- Evaluate using above criteria

Lesson 6: What to Use?

- Depends on many things
- How much time you have
- What you need
- What you know



Search Directories

- Use when browsing
- Use to get a feel for a subject
- Use when you're not quite sure what you're looking for

Major search engines

- Use when you know what you're looking for
- Use for more hits/pages
- Use when you have useful keywords

Specialized Search Engines/Directories

- Use for “only professional/consumer” searches
- Use when you need something fast

All of them

- Use more than one search engine
- Pick the best one for the task
- Still might have to surf a little