Beyond Searching- Guerrilla Tactics to Get the Information You Want & Know What You're Getting

Session # 1991 Room C-110 Anthony A. Luscre-Director of Technology, Mogadore Local Schools

S A W	R A T.
- Select Search Engine - Advanced Page - Wording	- Result Pages - Advertisements - Trim # of pages found
Q U I C K,	**G O !**
- Quality - URL (Address) - Indexes &Databases - Current - Know Who & Why	- Gleaning - Observe Copyrights - Learn About Search Engines

I. Select Search Engine (SE)

A. Best Match for Search Type
B. Major- Google, Yahoo, All The Web, etc.
C. Meta Search Tools Multiple SEs
D. Specialized Search Tools
E. Indexes & Databases

II. Advanced Page

A. Use the "Advanced Searching" page of SE
B. Automatically incorporates Boolean Logic in search by entering in variety of text boxes
C. No knowledge of Boolean Logic required .
D. Greatly increases successful searching
E. Place links on your web site to go directly to advanced search pages.

III. Wording

A. Choose the wording of your query carefully
B. Know which words are ignored by your SE ("stop words" or "trivial text" i.e. where, how, a, the, etc.)
C. Know how to exclude words
D. Try a variety of words & word orders
E. Use words that are unique to the item you are searching for-- author's name, specific state or city, dates or specialized terms
F. Use words that experts in that field would use

IV. Results Page

A. Just because an item is listed first does not mean it is the best resource !
B. What factors influence site ranking?
C. Recognize advertisements & paid placement
D. Develop a strategy for reading listed pages' descriptions to choose best candidates
E. Learn how to read URL (address) of a site

V. Advertising

A. Distinguish between advertisements and valid search results & "pay to place", "pay-to-click", "click revenue" and how vendors bid on key words

_____
_____
_____
_____
_____
_____
_____
_____
_____
_____
_____
_____
_____
_____
_____
_____
_____
_____
_____
_____
_____
_____To find out how much a key word or phrase costs today: visit overture.com, search for word & click on the link on right-top corner- "view advertiser's maximum bids".

Pay per Click-- each time you click on a link from the search engine to the high bidding company, that company pays the search engine that amount of money!

© 2004 Copyright- Anthony A. Luscre, Mogadore Local Schools all rights reserved. Please contact author for use of original materials,

Beyond Searching- Guerrilla Tactics to Get the Information You Want & Know What You're Getting

Session # 1991 Room C-110 Anthony A. Luscre-Director of Technology, Mogadore Local Schools

VI. Trim # of Pages Found

A. Use key words found in the listings of your search results to refine your search
B. If many of the sites contain a word you do not want go back and exclude that word and re-search
C. If many of the sites found are "too generic" choose word(s) to narrow search
D. If found sites have all the right words but in the wrong order go back and enter them in "specific phrase" box, then re-search

VII. Quality of Pages Found

A. Visit a few of the sites on your results page that show promise, based on their descriptions, URL, dates, etc.
B. Evaluate if you have found valid information that you were searching for
C. Decide if you need to refine your search
D. Check pages for links to other pages that might provide more relevant information

VIII. URLs of Sites

A. URLs/Addresses
   1. "...Uniform Resource Locator, the global address of documents & other resources on the World Wide Web." – from Webopedia.com
   2. Information about site owner based on .com, .edu, .gov, etc. can be helpful but is not always consistent.
   3. Caution Signs- Specific characters (@, ~, #, etc.) can cause an address to not always be what it appears to be.
   4. "Recursive Web Searching Techniques"

B. URLs (Characters)
   1. @ = DANGER ! – web browsers ignore all characters proceeding an @ in an address.
   2. ~ = Personal– addresses with a ~ (tilde) usually indicate that it is a privately maintain page even if the address denotes a larger institution (.com corporation, .edu university or .gov government) and thus usually contain personal opinions
   3. # = Named Anchors-- allow the page designer to direct you to specific spots within a page.

© 2004 Copyright- Anthony A. Luscre, Mogadore Local Schools all rights reserved. Please contact author for use of original materials,

Beyond Searching- Guerrilla Tactics to Get the Information You Want & Know What You're Getting

Session # 1991 Room C-110 Anthony A. Luscre-Director of Technology, Mogadore Local Schools

C. Recursive Searching (RS)

1. From Webster's definition of recursion "...the determination of a succession of elements... by operation on one or more preceding elements..."
2. Technique of RS Involves
   i. using information you have already found
   ii. a bit of guessing
   iii. a little knowledge of directory structure
   iv. observations of the habits of web designers/authors
   v. time to try a variety of possibilities
3. Allows you to find related pages, more info from the same source or different formats of the same document.
4. "Climbing the Tree" allows you to maneuver through a web site to other pages that might be relevant. Climb up by removing sections of address from the right of "/ ". Climb down by adding or changing characters to the right.
5. Changing endings of URL (.pdf, .html, .doc) can often provide same information in different document format
6. Substituting characters that match pattern (I.e. /list2003.html in place of /list2002.html)
7. Using portions of a URL in your search
8. Use **Ctrl-F** to find word(s) within the page you are currently viewing (**Ctrl-G** to repeat)
9. Utilize built-in searching tools and site maps to find information within web sites.
10. Find an authoritative site on a subject (national organization, trade group, governmental entity, etc.) then use information, links, indexes, etc. within their site to find more information

© 2004 Copyright- Anthony A. Luscre, Mogadore Local Schools all rights reserved. Please contact author for use of original materials,

Beyond Searching- Guerrilla Tactics to Get the Information You Want & Know What You're Getting

Session # 1991 Room C-110 Anthony A. Luscre-Director of Technology, Mogadore Local Schools

IX. Indexes & Databases

A. Many items are not found by SEs because they are contained in dynamic or sequestered web pages
B. Dynamic pages are actually created on the fly from a database (php, asp, sql, etc.) of information when the users selects specific information (i.e. AnyWho.com)
C. Sequestered pages
D. Behind a firewall
E. Requires password access
F. HTML coding in a page to cause search robots & spiders to ignore pages on that site (robots.txt)
G. I & D- often have more human involvement in compiling, maintaining, sorting and annotating
H. I & D- some may require a subscription, registration or user fee
I. Examples include Lexus/Nexus, SIRS, and NARA (National Archives & Records Administration.)
J. Your friendly local librarian, always a great resource when searching for information, can point you to good online I & Ds.

X. Current?

A. When was the page created/modified
B. Ctrl-I provides information on the page being viewed
C. Many SEs allow you to specify limiting dates for search result from their advanced search pages

XI. Know Who & Why

A. Who owns the site and why do they want you to read it ?
B. Why is a web page there? Know Who & Why
C. What kind of document is it
   1. Basis of document
   2. Author's credentials
   3. References,
   4. Citations, etc.
D. "Hierarchy of Reliability" as measured by ownership, author, type of document, etc.
E. Finding the owner of a site or page-
F. Read the URL
G. Which sites link to the site? Using a SE type-- link:http://www.xyz.com
H. Who owns the URL- "whois" search at Network Solutions

© 2004 Copyright- Anthony A. Luscre, Mogadore Local Schools all rights reserved. Please contact author for use of original materials,

Beyond Searching- Guerrilla Tactics to Get the Information You Want & Know What You're Getting

Session # 1991 Room C-110 Anthony A. Luscre-Director of Technology, Mogadore Local Schools

XII. **Gleaning-** Techniques for gathering only the specific material (text and/or graphics) you need from a web page

(Warning- Indiscriminate or improper use of the following techniques could result in violations of Copyright Laws and/or Plagiarism! The author does not condone either activity!)

A. Gleaning (Printing)
   1. Print Preview, then print only desired pages
   2. Print Screen, Paste into Image editing software

B. Often the best way to save information from a web site is to instead save a link to that page (bookmark or favorite)
   1. Pros & Cons of Bookmarks vs. Gleaning -http://www.mogadore.summit.k12.oh.us/library/gleaning.html#book

C. Gleaning (Saving)
   1. Save Image- right click mouse on image, menu appears, choose "save image as"
   2. Save an entire page or site
      i. "Save as" function in explorer
      ii. Opening a web page in a graphical web editor, then saving
   3. Alternative- Use screen recording software (Camtasia-Windows, Snapz-Mac, etc.) of the page and its links

D. Gleaning (Getting Text)
   1. Copy & Paste to into a word processor (wp)
   2. Often needs reformation to remove paragraph returns at the end of each line (use find & replace feature of wp) or other formatting
   3. Text vs. Graphics- not everything that looks like text is really text. Many titles, headers, selection "buttons", etc. are actually graphics

E. Spreadsheets & Databases
   1. Copy & then use "Paste Special" when bringing data into spreadsheets
   2. Often data will be in one cell per row. To separate into multiple columns use parsing or "text to table" functions
   3. If data is separated improperly, not in the needed format, use concatenation function

© 2004 Copyright- Anthony A. Luscre, Mogadore Local Schools all rights reserved. Please contact author for use of original materials,

Beyond Searching- Guerrilla Tactics to Get the Information You Want & Know What You're Getting

Session # 1991 Room C-110 Anthony A. Luscre-Director of Technology, Mogadore Local Schools

XIII. Observe Copyright

A. Copyright Information Links from Stanford University-- http://fairuse.stanford.edu/
B. Electronic format lends itself to very easy duplication, distribution and resulting copyright violations and/or plagiarism
C. Citations for web research
D. Digital Millennium Copyright Act
E. Plagiarism- Web can enable plagiarism both individualized or as a business ("internet paper mills.") Web can also detect and deter plagiarism with sites such as Turn It In .com.

XIV. ! "...and then..."

A. Evaluate your results & the effectiveness of your search by trying alternative search engines and recursive searching
B. Visit web sites on searching and search engines
C. Visit web sites on evaluating the information you find
D. Try something different– Web Brain.Com

Resources & Contact Information

Download this presentation & see links to other resources and view information on my previous OSN presentations at:
http://www.mogadore.summit.k12.oh.us/OSN_conf.html

Anthony A. Luscre
mo_luscre@neonet.k12.oh. Us
Director of Technology, Mogadore Local Schools
Visit our district's web portal-
http://www.mogadore.summit.k12.oh.us/portal_pages.html

© 2004 Copyright- Anthony A. Luscre, Mogadore Local Schools all rights reserved. Please contact author for use of original materials,