SEARCHING and LOCATING Information

Lecture Four
explain the basic differences between the
- Deep Web and the
- Open Web

determine what type of search tool to use for which type of source
- databases
- search engines
- library catalogs
- digital libraries
demonstrate a basic understanding of database:
- structures
- the types of information contained in them
- subject scope
- etc.
describe how search engines work

write an effective search statement using various techniques
- demonstrate the use of Boolean operators AND and OR
- demonstrate truncation of search terms
- demonstrate both subject and keyword searching and describe the difference between the two
- identify and use limit features to refine a search
apply various search statements and different terminology using different search tools
Sources of information

- People
  - Local Sources
  - Writing Letters and E-mail
  - Reading Personal Papers
  - Attending Lectures
Sources of information

- Government Documents
  - Local - city
  - State
  - Federal
Sources of information

- Internet
  - Subject Directories
    - Commercial Portals
      - About.com
      - Go.network
      - Lycos
      - Yahoo!
    - Academic Directories
      - Librarians' Index to the Internet. http://www.ipl.org/
      - WWW Virtual Library - http://vlib.org/
      - INFOMINE - http://infomine.ucr.edu/
      - Internet public Library - http://www.ipl.org/
  - Surveys & Experiments
  - Tv, Radio, etc.
  - Library
Two types of search tools:
1. Directly provide access to the information source
2. Provide information to help find a source (i.e. citation)
Types of Search Tools – contd.

- Library Catalogs
  - to identify books and other library materials
- Databases
  - to identify journal articles
- Web Search Engines
  - to identify Websites
Open Web vs. Deep Web

- Open Web is also known as the ‘Surface Web’ or ‘Visible Web’.
  - commonly used Search engines (Google, Yahoo) provide access to it
  - Freely available to the public

- **BUT…..**
"Approximately 80% of the information on the Web belongs to the 'invisible Web'."


- Library databases are part of the Deep Web
  - subscription based or password protected
## Open Web vs. Deep Web

<table>
<thead>
<tr>
<th>OPEN WEB</th>
<th>DEEP WEB</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sources not reviewed or evaluated</td>
<td>Sources and search tools reviewed and recommended by experts</td>
</tr>
<tr>
<td>Identity and credentials of author usually not available</td>
<td>Authors are usually experts who are identified</td>
</tr>
<tr>
<td>Open to everyone</td>
<td>Resources intended for specific academic community</td>
</tr>
<tr>
<td>Little organization, limited searching features and may not be updated</td>
<td>Better organized, updated with advanced features</td>
</tr>
<tr>
<td>Web sites not always stable or archived</td>
<td>Permanent part of collection</td>
</tr>
</tbody>
</table>
Library Catalogs

- Used to identify and locate books (print or electronic)
  - Also provides access to videos, microforms, government publications and most journals (not articles)
Library Catalogs

- **WorldCat**
  - Tool for finding materials in the catalogs of thousands of worldwide libraries.
  - Interlibrary Loan service allows you to get any item from any library

- **WorldCat Local**
  - Combines both WorldCat and all resources in ODU Libraries.
    - ODU items at top of list
    - Also lists libraries with selected item by distance from ODU
Google books

- Tool for finding books online.
  - Not all online books will provide full text
  - Should not be the sole source for finding books
What is a database?

- “a comprehensive collection of related data organized for convenient access, generally in a computer”
- Online catalogs are databases
- Search tool for identifying journal and magazines, books, book chapters, dissertations and other sources
- Can be multidisciplinary or subject – specific
Looking for journal articles?

*Academic Search Complete*
- covers all topics
- Articles
- Limit to scholarly articles
- Many available in full text
- Includes materials back to 1887
Electronic Journal Collections

- Subset of Periodicals databases
  - Provide full text articles from a particular publisher
    - Project Muse
      - Humanities and Social Sciences
    - BioOne
      - American Institute of Biological Sciences
    - ScienceDirect
      - All journals published by Elsevier (science and health science)
Google Scholar

- Good starting point for a multidisciplinary database
- Includes materials from
  - Academic publishers
  - Professional societies
  - Preprint repositories
  - Universities
  - Other scholarly organizations
Google Scholar

- Since results will be huge, learn to narrow a search and evaluate the results carefully.
- Google Scholar can be linked to the full-text publications in ODU’s journal collection.
- Some issues with Google Scholar:
  - Heavier in sciences and not as good for humanities and social sciences
  - Capabilities for searching, limiting, sorting and printing not as good as ODU’s databases
  - Not everything is available in full text
  - VIDEO
For finding journal articles primarily
- Leads to other resources like:
  - Books, book chapters, dissertations, conference proceedings etc.
- Most topics fall into the following categories:
  - Sciences
  - Social Sciences
  - Humanities
ODU libraries break down these disciplines further based on the subjects taught here.

- Interdisciplinary topic - i.e. Criminal justice
- Comes under Social Sciences, Psychology (criminal behavior) or Education (juvenile delinquents/school)
- Check all applicable databases for the most comprehensive research
Read the annotations on the Databases page

Subject Scope:

- Subject-specific – focuses on one or just a few areas
  - Example: CINAHL covers nursing, medicine, allied health disciplines

- Broader scope – covers all science areas
  - Science Citation Index (Web of Science)
Understanding Databases

- Multidisciplinary – covers all subject areas
  - Academic Search Complete

- Material Types Covered:
  - Variety of materials (books, journal articles, reports, government publications)
    - PAIS International
  - Only journals
    - Web of Science
  - Only images
    - AP Images
Understanding Databases

- Popular or Scholarly?
- Indexing or Full text?

**Time Range:**

- Materials before 1980 may not be indexed
  - PAIS Archive dates back to 1915
  - PubMed dates back to 1950
  - ERIC (education) dates back to 1966
  - JSTOR will not include the most recent 3-5 years off journals
Geographic Orientation:
- Many databases focus on U.S. materials
- Others will include material from different parts of the world
  - Medline – article citations from biomedical journals available all over the world
  - Check language variations in searching world databases
    - Color, analyze, centre etc.
Access Restrictions

- Majority of databases are paid for through library subscription or through VIVA (the Virtual Library of Virginia) and must be restricted to ODU users.
Database Structures

- Database terminology for conducting efficient searches:
  - **Databases**
    a comprehensive collection of related data organized for convenient access
  - **Fields. Records**
    records are made up of **Fields**, such as Author, Article title, Journal title, Publication date, Page number etc.
  - **Metadata**
    data about data so resources can be more readily identified and found
Database Structures

- **Search Systems**
- ODU libraries databases are part of a larger Search System (Database Package)
  - EBSCOhost
  - CSA Illumina
  - FirstSearch
  - WilsonWeb
  - ProQuest
- Each search system has its own interface and approach to searching.
Search Engines

- Programs designed to search for information on the World Wide Web.
  - Google, WebCrawler, Ask.com, Bing, Exalead
- We are talking about the Open Web here
  - Freely available for input and viewing
- Because nobody owns the Internet, there is no “policing” and anyone can publish anything they want.
  - Good and bad material available
  - Scholarly and layman articles
  - Finding high-quality sources is difficult and so careful evaluation is essential.
- No one Search engine indexes all Web Sites.
What is a Search engine?
- a computer program that retrieves documents or files or data from a database or from a computer network (especially from the internet) [wordnetweb.princeton.edu/perl/webwn](http://wordnetweb.princeton.edu/perl/webwn)
- A web search engine is designed to search for information on the World Wide Web. The search results are usually presented in a list of results and are commonly called hits. The information may consist of web pages, images, information and other types of files. ... [en.wikipedia.org/wiki/Search_engine](http://en.wikipedia.org/wiki/Search_engine)
Top 5 Web Search engines according to searchenginewatch.com are:

- Google, Yahoo, Bing, AOL, Ask.com

Advanced Search on Google - video
How do Search Engines work?

- [http://www.lib.berkeley.edu/TeachingLib/Guides/Internet/SearchEngines.html](http://www.lib.berkeley.edu/TeachingLib/Guides/Internet/SearchEngines.html)

Terminology

- Database
- Spiders
  - Computer robot programs that build search engine databases by following the links in the pages they already have in their database
  - If a web page is never linked from any other page, spiders cannot find it.
  - Must be manually linked (submit a URL) or other pages must link to it.
Indexing

- Spiders pass the pages they find to another program that indexes - ie. Identifies, the text, links and other content in the page and stores it in the search engine database so it can be searched by keyword and any other advanced approaches.

Invisible Web

- Web pages that are excluded from most search engines by policy
  - Library catalogs and article databases
Search Engines

- Ranking results
  - Depends on the Search engine
  - By how many times the sites are linked from other sites (popularity)
    - Google uses this method
    - Wikipedia shows up in the top 5 results because it cites its own articles – many times
  - By the number of times the word appears in the page, especially in the header
  - By the amount the site pays to be listed (sponsors)
- Comparison of search results from Google, Yahoo Search, MSN and Ask.com
  - [Link to comparison](http://www.lib.odu.edu/genedinfolit/3searching/how_do_they_rank_results.html)
Search Engines

- Wikipedia
  - What’s all the fuss? Why is it not a legitimate source?

- Wikipedia’s Disclaimers
  ...The structure of the project allows anyone with an Internet connection to alter its content. Please be advised that nothing found here has necessarily been reviewed by people with the expertise required to provide you with complete, accurate or reliable information. That is not to say that you will not find valuable and accurate information in Wikipedia; much of the time you will. However, Wikipedia cannot guarantee the validity of the information found here. The content of any given article may recently have been changed, vandalized or altered by someone whose opinion does not correspond with the state of knowledge in the relevant fields.

A meta-search engine is a search tool that sends user requests to several other search engines and/or databases and aggregates the results into a single list or displays them according to their source. ...

en.wikipedia.org/wiki/Meta_search_engine

- Dogpile
- Ask
- Metacrawler
Subject Directories

- Compiled and maintained by humans
- Evaluated websites are included in the directory
- Useful for finding general information about a subject
  - Yahoo directory
  - Broad category (Health) to narrower focus (Nursing).
  - Keyword searches
- Other search engines provide Subject Directories
Scholarly Focus

- These search engines use people (not spiders and robots) to evaluate and gather Web sites.

 Examples:

- Infomine: a directory of scholarly search engines.
- Scout Report Archives: extensive list of carefully selected Web sites and mailing lists by subject with critical annotations
Search Engines

- **Academic Info**: Browse or search for Internet resources that are directed toward an academic audience.
- **Internet Public Library**: Use "Resources by Subject" to find many quality Internet resources on your topic.
Other Search Tools

- For Government Documents
  - Catalogs, databases search engines


Catalog of US Government Publications (online): for publications 1976-present

See "Find Government Resources" on the library Web site.
Other Search Tools

- For Digital Objects/ Collections
  - Search tools under development
  - Major collections accessible through Internet search engines and directories
    - e.g. Yahoo’s Subject Directory - "Digital Libraries > Projects and Collections"
  - Library catalog and WorldCat
For Digital Objects/Collections

**OAlster** (accessible from the library's Databases page) is a specialized database that:
- contains digital resources from open archive collections
- represents multidisciplinary resources from more than 1000 contributors worldwide.
- contain a digital object link allowing users access to the object in a single click.


**Library and Archival Exhibitions on the Web** "This site features links to online exhibitions that have been created by libraries, archives, and historical societies, as well as to museum online exhibitions with a significant focus on library and archival materials."

Many of the "**Subject Guides**" on the Library's Web site will link to various Web Resources that contain digital collections.
For Finding Archival Collections

**ArchiveGrid** is a database that allows you to search for historical documents, personal papers, and family histories held in archives around the world. Thousands of libraries, museums, and archives have contributed collection descriptions to ArchiveGrid.

**Repositories of Primary Sources** -- "a listing of over 5000 websites describing holdings of manuscripts, archives, rare books, historical photographs, and other primary sources for the research scholar"

**Virginia Heritage** is a union database of finding aids to archival and manuscripts collections in twenty-seven Virginia repositories.
Identify Concepts and Keywords

- e.g. Does nonverbal communication play a role in social aggression among teenage girls?
- Searching via natural language is not a successful way to search databases.
- Better to use strings of characters (meaning not useful)
- Articles (a, an, the) and prepositions (in, to, for etc) not used as search terms.
Identify Concepts and Keywords

“does it play a role” not a good string to use in a search because there are so many other ways to express them.

Some words are already implicit – “psychology” in a psychology database and therefore unnecessary.

So, you need to be flexible. The best way to search is to break down your topic into searchable components and include useful synonyms:
Identify Concepts and Keywords

Suggestions for finding keywords and synonyms:
- Use dictionary.com -- both the dictionary and the thesaurus portion.
- Use a graphical dictionary such as: http://www.visuwords.com/ Just type a keyword in the search box, and Visuwords will provide synonyms, definitions, and associations with other words and concepts. You can then click on those other words to expand the "tree" or "neural network." Try it -- it's fun!
- In a database, you can use the thesaurus (more later).
- As you explore your topic, jot down the keywords used in your readings, whether it be an encyclopedia or a journal article.
Search Techniques

- **Identify Concepts and Keywords**
  - Remember that you are searching for concepts: the terminology and words you use to describe those concepts can vary and may change as you search.