Intermediate Searching
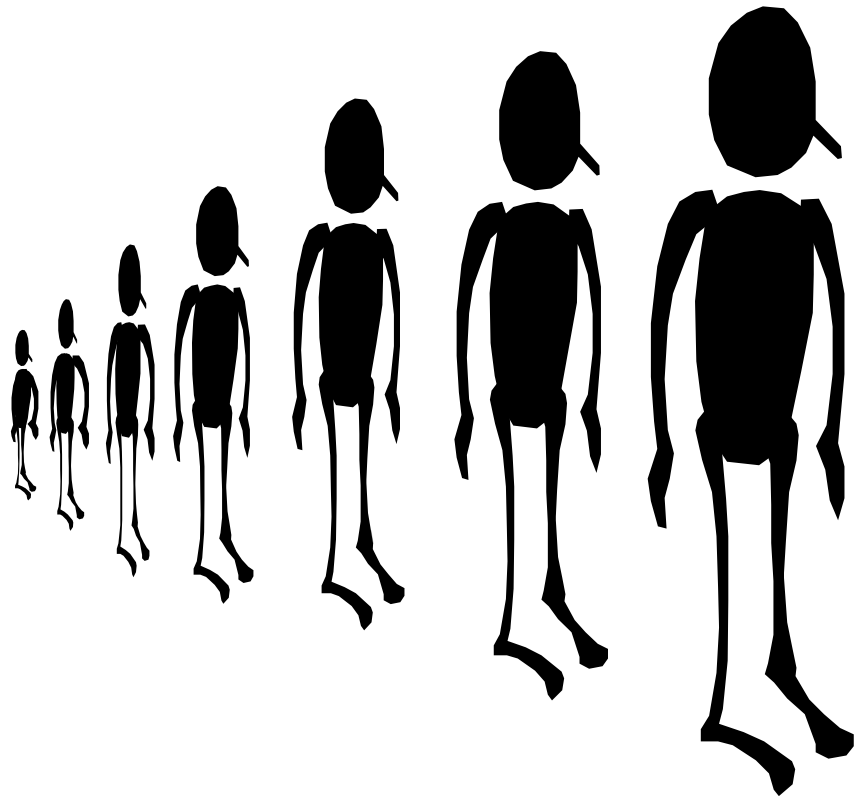
David King

IT/Web Project Manager

Kansas City Public Library

Intermediate Searching

- Recall/Precision?
- Relevance Ranking
- Boolean
- Field Searching
- Finding Images
- Meta Search Engines
- Invisible Web

Recall vs. Precision

❑High Recall

➢Getting thousands or millions of hits
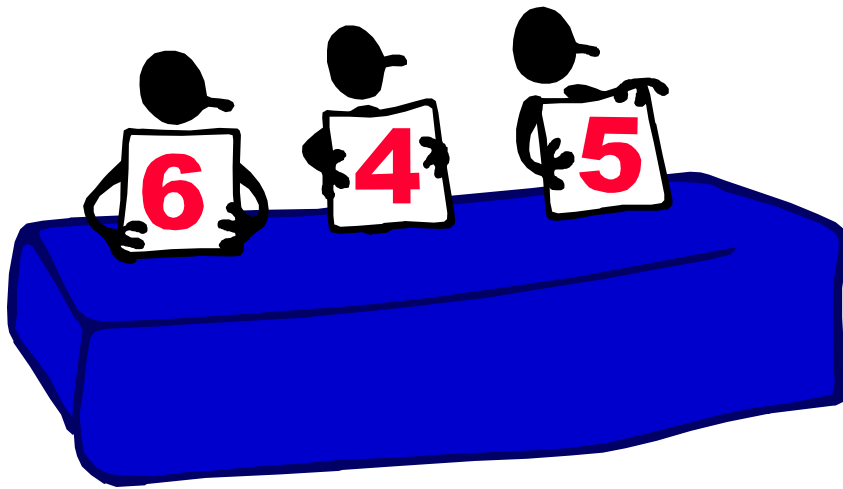
➢Not usually desirable

➢Scares some people!

❑High Precision

➢Retrieved items are all highly relevant

➢Do this by fine-tuning the search

WARNING!!!

❑ Not all advanced search techniques are supported by all search engines!

Relevance Ranking

- What is relevance ranking?
- How does relevance ranking help my search?

What is Relevance Ranking?

❑ Results of your search are "ranked" according to:

❑ Uniqueness of terms

❑ Proximity of terms to each other

❑ Number of different terms present

❑ Number of times the term is present

How does Relevance Ranking help my Search?

❑Search engine attempts to place the "most important" results of your search at the top of the results list

❑How should this info be used?

➢Don't let the "2 million records found" numbers fool you!

➢Examine the first 50 items on results list

➢If these aren't accurate – re-do your search

Exercise

- Do a simple search in Google
- Compare screen one results with screen 10-20 results
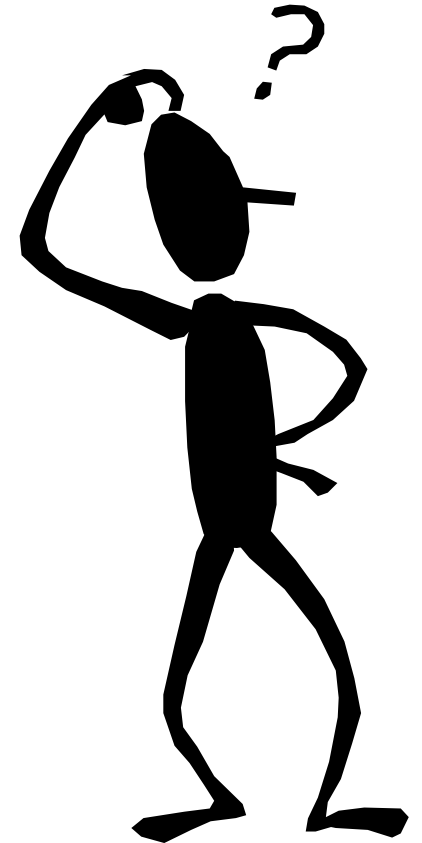- Which results seem more relevant?

Boolean Searching

❑Three types of Boolean Searches:
  ➢Full Boolean
  ➢Simple Boolean
  ➢Menu-driven Boolean
❑All can be very confusing.

Full Boolean

- AND, OR, NOT, NEAR, ADJ
- Sometimes found on Advanced Search pages
- Usually needs to be in all caps

Boolean Operators

❑AND – libraries AND technology

➢Results include both terms

❑OR – libraries OR technology

➢Results include one or both terms

❑NOT – libraries NOT technology

➢Results exclude specified term

Proximity Operators

❑ NEAR – one word appears close to the second word

❑ ADJ – Both terms appear next to each other

➢ You could find "Power Search"

➢ You could find "Search Power"

Simple Boolean

- AKA "Implied Boolean."
- Can usually do this type of search on most Simple Search option pages.
- +, -, ""
- + = AND searching
- - = NOT searching
- "" = phrase searching

Simple Boolean

- +libraries +technology = libraries AND technology
- +libraries -technology = libraries NOT technology
- "technology in libraries" = technology in libraries (the phrase)

Menu-Driven Boolean

❑ Pull-down menus offering menu choices:
- ➢ all the words = AND
- ➢ any of the words = OR
- ➢ exact phrase = AND, ADJ

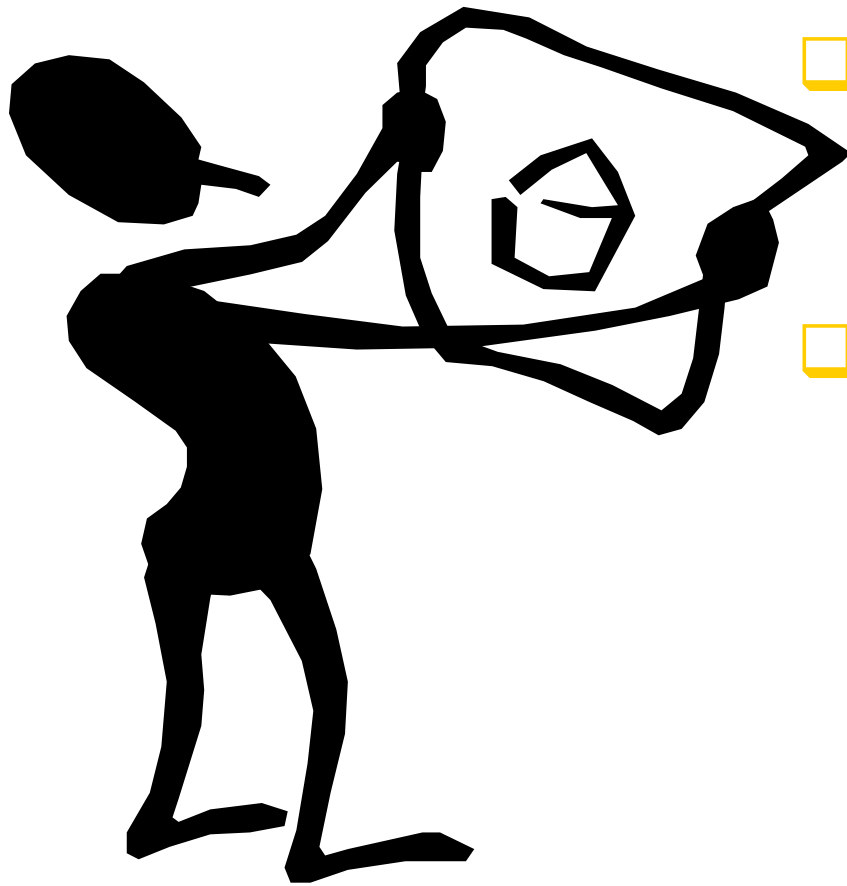❑ Easy way to do Boolean

❑ Limited by available options

Exercise

- ❑ Do a simple search
- ❑ Do a simple Boolean search
- ❑ Are the results similar?
- ❑ Which search gave more results?

Field Searching

- Special keywords allow searching specific parts of a web page
- **Cache:search word**
  - Finds the cached page
  - **cache:www.kclibrary.org web** will show the cached content of the page.

Field Searching

- **site:www.something.com**
  - Finds pages within a specific domain
  - **bridger site:www.kclibrary.org** will find pages about Jim Bridger within www.kclibrary.org
  - **help site:com** will find pages about help within .com urls.

Field Searching

❑ **Related:www.kcstar.com**
- ➢ lists webpages that are "similar" to a specified webpage.
- ➢ related:www.kcstar.com lists webpages that are similar to the Star's homepage.
- ➢ Also accessible by clicking on the "Similar Pages" link

❑ **Link:www.google.com**
- ➢ list webpages that have links to the specified webpage.
- ➢ link:www.google.com lists webpages that point to the Google homepage.

Field Searching

info:www.google.com

➢ finds info that Google has about a webpage

➢ info:www.google.com finds info about the Google homepage.

➢ Also accessible by typing the web page url directly into a Google search box.

Field Searching

□ **Stocks:**

➢ Google treats the rest of the query terms as stock ticker symbols, and will link to a page showing stock information for those symbols

➢ stock: intc yhoo will show information about Intel and Yahoo

➢ also available if you search just on the stock symbols (e.g. [intc yhoo]) and then click on the "Show stock quotes" link on the results page.

Field Searching

- **allintitle:**
  - Finds all search words in the title of a webpage
  - **allintitle: google search** returns documents that have both "google" and "search" in the title.
- **intitle:**
  - finds documents containing that word in the title.
  - **intitle:google search** will return documents that mention the word "google" in their title, and mention the word "search" anywhere in the document
- **intitle:google intitle:search** is the same as **allintitle: google search**

Field Searching

❑ **allinurl:**
  ➢ Finds all search words in a URL
  ➢ **allinurl: google search** returns documents that have both "google" and "search" in the url.

❑ **inurl:**
  ➢ finds documents containing that word in the url.
  ➢ **inurl:google search** finds documents with the word "google" in the url, and with the word "search" anywhere in the document

❑ **inurl:google inurl:search** is the same as **allinurl: google search**

Field Searching

❑ Google's not the only one with specific field searches

❑ Yahoo and ask.com also have specialized search features
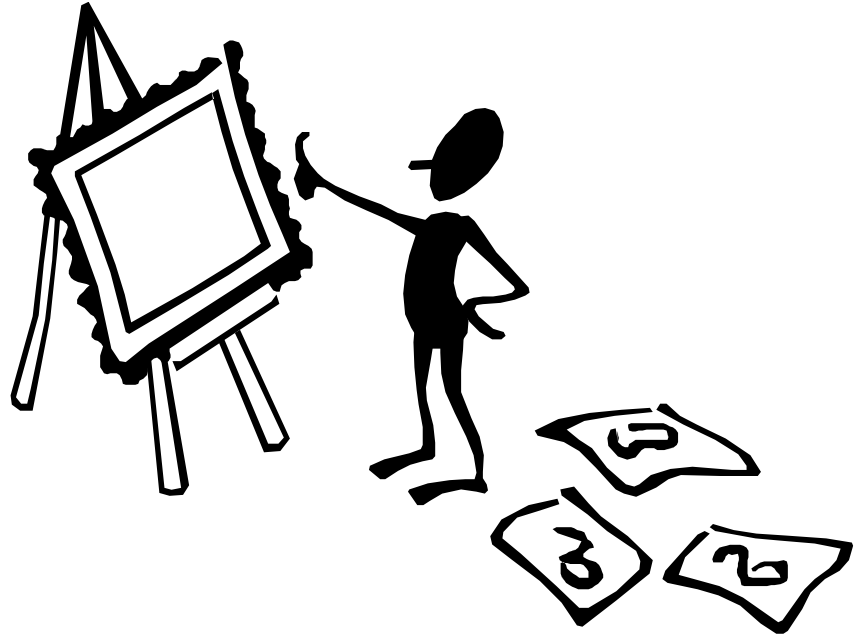
❑ Always check the Help pages!

Exercise

❑ Search for KCMLIN using inurl:text

❑ Search for KCMLIN using link:text

❑ Search for KCMLIN using info:text

Images

- ❑ Why are images hard to find?
- ❑ Where to go for images

Why are images hard to find?

❑ We search for images with text-based keywords

❑ We hope these words have been added to the images, in a file name or in descriptive terms on the same page

Where to go for images

- Google (www.google.com)
- Yahoo (www.yahoo.com)
- Ditto (ditto.com)
- Flickr (flickr.com)
- Other similar search engines
- Ebay.com!!!

Exercise

- ❏ Find a photo of a famous person
- ❏ Find a drawing of an apple

What is a meta-search engine?

- Searches other search engines at the same time
- Meta Search engines do not maintain a database

Selection

❑Most search several major search engines at once

❑May also search directories, specialized databases, or news wires

❑Ability to customize search is important!

How queries are submitted

- ❏ Varies with each meta engine
- ❏ Some submit simultaneously
- ❏ Some submit one or several at a time

Big Problem with Meta Search Engines

- Each search engine has different search features
- Different advanced features
- Advanced features tend to be filtered out
- Usually, simple Boolean (+,-) will work

Manipulation of results

- ❑ Some allow display options for results
- ❑ Options are relevance, site, source
- ❑ Most don't offer this feature

When to use them

- Want to see top hits from several databases at once
- You want to search a variety of sources simultaneously

Some Meta Search Engines

- Dogpile, metacrawler, webcrawler (all same company)
- Search.com (www.search.com)
- A new-fangled one - HotBot (www.hotbot.com)

Exercise

- Perform a search in Google
- Perform same search using a meta search engine
- Compare results

Invisible Web

❑ 16% of websites found using search engines!

❑ Where are the other 84%???

➤ Not linked

➤ Not important

➤ Database-driven

➤ Great, but not yet discovered by you!

Specialized Search Engines?

❑ What are they?

❑ Where do you find them?
- ➢ http://lii.org/
- ➢ http://infomine.ucr.edu/
- ➢ http://www.invisible-web.net/
- ➢ http://www.completeplanet.com/

Exercise

- Perform a search in Google
- Perform same search using a specialized search engine
- Compare results

Blogs!!!

- Web Log = blog
- AKA news feeds, RSS,
- Can be a college student's diary
- Can be up-to-date news from the BBC or Reuters

Blog Search Engines

- http://www.feedster.com/
- http://www.syndic8.com/
- http://www.topix.net/
- http://www.daypop.com/
- http://www.newsisfree.com/
- http://www.blogdigger.com/