Introduction to Searching Bibliographic Databases

Jennifer Lyon, M.S., M.L.I.S.
Clinical Research Librarian
Health Science Center Libraries
University of Florida

A **database** is an organized collection of data.

- Bibliographic databases are searched through many of the same techniques as any other database
- Focus will be on PubMed, but these techniques apply to ANY literature database

Examples of Records and Fields

Record #	Author	Title	Publisher	Date of Publication
1	Jones KM	Infectious Diseases	Wiley	2001
2	Smith BR	Medicine	Cambridge Univ. Press	2004
3	Johnson AS	Cancer	Elsevier	2005
4	Bradley PL	Asthma	Synergy	2003

Examples of Records and Fields

Record #	Author	Title	Publisher	Date of Publication
1	Jones KM	Infectious Diseases	Wiley	2001
2	Smith BR	Medicine	Cambridge Univ. Press	2004
3	Johnson AS	Cancer	Elsevier	2005
4	Bradley PL	Asthma	Synergy	2003

RED=RECORD BLUE=FIELD PURPLE = ONE PIECE OF DATA

Searching a Database

Different search interfaces do the same things in slightly different ways

Good search interfaces should provide

- Ability to search for a specific item
- Ability to search for related items to a known item
- Ability to search in a specific field or fields
- Ability to combine search terms using Boolean Logic
- Ability to retrieve search results in a useful way

Step-by-Step Procedure

- Asking and Parsing the Question
- Understanding Boolean Logic
- Understanding Database Structure
- Understanding Field Searching
- Understanding Controlled Vocabulary
  - Controlled Vocabulary vs Keyword Searching
- Specialty Features
- Putting It All Together

Asking the Question: PICO

- Patient, Population or Problem
  - What are the characteristics of the patient or population?
  - What is the condition or disease you are interested in?
- Intervention or exposure
  - What do you want to do with this patient (e.g. treat, diagnose, observe)?
- Comparison
  - What is the alternative to the intervention (e.g. placebo, different drug, surgery)?
- Outcome
  - What are the relevant outcomes (e.g. morbidity, death, complications)?

http://healthlinks.washington.edu/ebp/pico.html

Parsing the Question

- What are the main concepts in your question?

Sample question:

Does nutrition therapy improve decubitus (pressure) ulcer healing in an elderly patient?

Concepts:

Nutrition therapy

Decubitus/pressure ulcers

Treatment efficacy

Ulcer healing

Elderly patients

Boolean Logic

A British mathematician named George Boole (1815-1864) developed an algebraic system of logic that is now widely used in computer and electronic systems including database searching.

While Boole's algebraic system can be complex, a very simple form of Boolean Logic is used for searching most bibliographic databases.

Boolean Operators

Standard Boolean Logic for database searching uses 3 relationships among search terms.

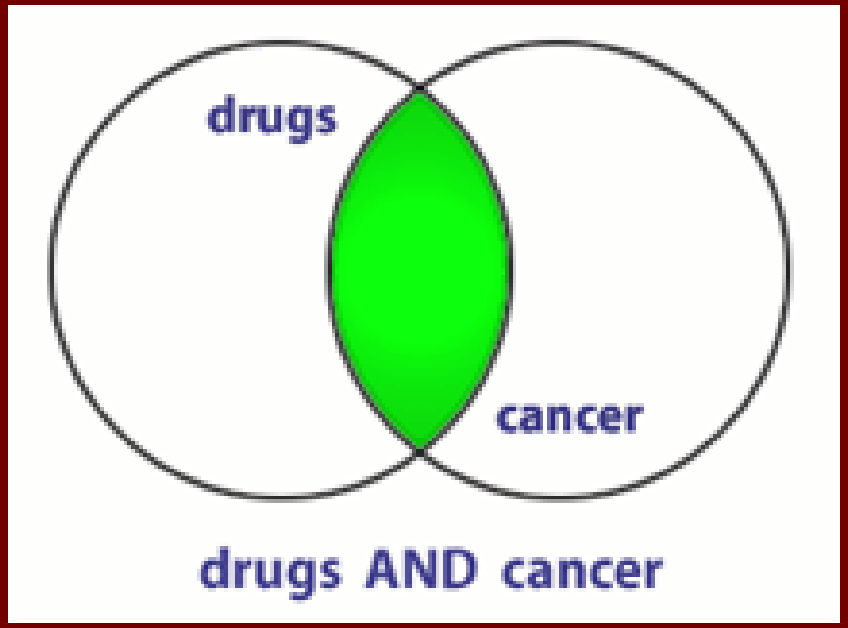
AND

OR

NOT

It is both simple and powerful.

AND



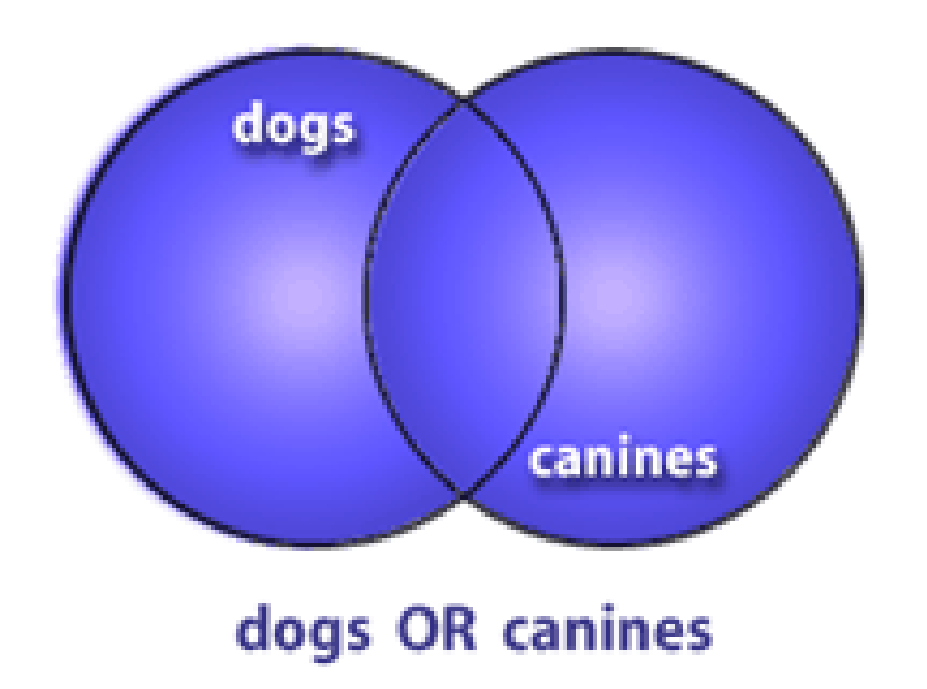
BOTH terms included in any results.

If a record has only one of the two terms, it will not be retrieved.

If the record has neither term, it will not be retrieved.

What does this do to the amount of records retrieved?

OR

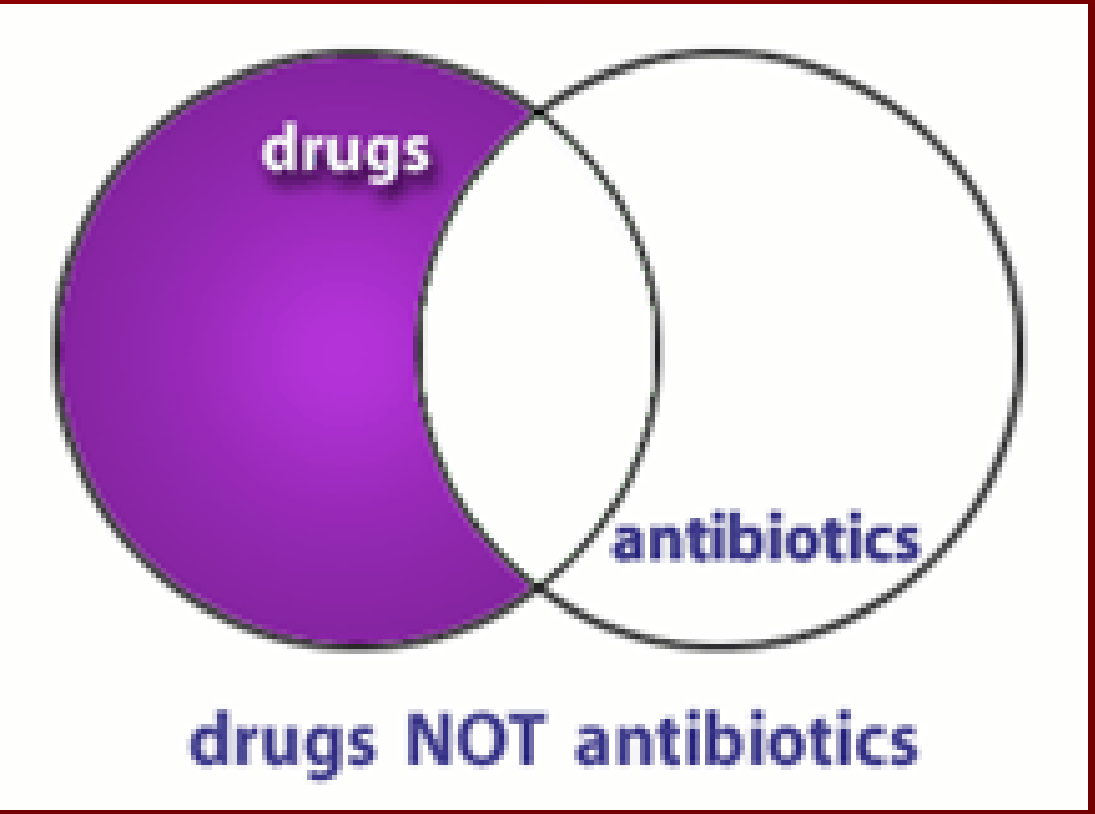


Only one (**NOT both**) of the terms are in the results

'OR' will retrieve the record if both are included.

What does OR do to the amount of records retrieved?

NOT



Excludes any results containing the term

Records containing both will not be retrieved.

What does NOT do to the amount of records retrieved?

Using OR

"OR" groupings contain terms for the same idea/concept and are usually put in parenthesis

(term OR term OR term)

where all terms are difference ways of representing the same concept

(faculty OR teachers OR professors)

(students OR learners OR pupils)

Using AND

"AND" groupings contain terms for different ideas/concepts and can combine OR groupings

Term AND (Term OR Term) where each represents a different concept

heart attack AND smoking

Diabetes AND exercise

Cancer AND (treatment OR therapy)

Using NOT

"NOT" statements are usually put last and can contain an "OR" grouping; they are often used to get rid of a common subgroup

Students NOT dental

Diabetes NOT juvenile

Putting Them Together

1. Identify the concepts (Parse the question)
2. List specific terms for each concept
3. Put the terms for each concept in an OR statements within parentheses
4. Combine OR statements with AND
5. Add any NOT statements to the end

Creating a Boolean Search

QUESTION: Is Vitamin C helpful in treating the flu?

1. Identify concepts and list terms

Concept 1 Influenza	**Concept 2 Vitamin C**	**Concept 3 Treatment**	**Concept 4 helpfulness**
Influenza	Vitamin C	Treatment	Outcome
Flu	Ascorbic acid	Therapy	Recovery
	Orange Juice	Management	Success

Step 2

2. Make your OR statements, one per concept
   - (influenza OR flu OR orthomyxovirus)
   - (vitamin C OR ascorbic acid OR ascorbate)
   - (treatment OR therapy OR management)
   - (outcome OR recovery OR success)

Steps 3 and 4

3. Put "AND" between each of the OR statements

(influenza OR flu) AND (vitamin C OR ascorbic acid OR orange juice) AND (treatment OR therapy OR management) AND (outcome OR recovery OR success)

4. Consider any NOT statements you might want to add.

Note: NOT isn't used very often

Parsing a Boolean Search

(emergency OR acute OR critical) AND (treatment OR therapy OR management OR care) AND (motor vehicle accident OR car crash) NOT (pedestrian OR walking)

What are the four concepts?

What terms are used for each concept?

Which three concepts must be included in all records found?

Which concept must not be included in any record found?

Beyond Basic Boolean

- Field Searching
- Controlled Vocabulary
  - Subject vs. Keyword Searching
- Specialty Features
  - Truncation
  - Phrase searching

Field Searching

Almost all databases will provide you with some ability to search a specific field or fields.

- Allows faster searching
- Allows more accurate searching

Not all databases may make all fields searchable.

Each search system will require a specific format.

All Field vs. Specific Field Searches

I would like to find articles by John Smith.

- Search all fields: John Smith
- Search Author Field only: John Smith

I would like to find an article published in 1997.

- Search all fields: 1997
- Search Publication Date Field: 1997

Why waste time searching for a date in the author field or an author in the volume field?

Formats for Field Searching

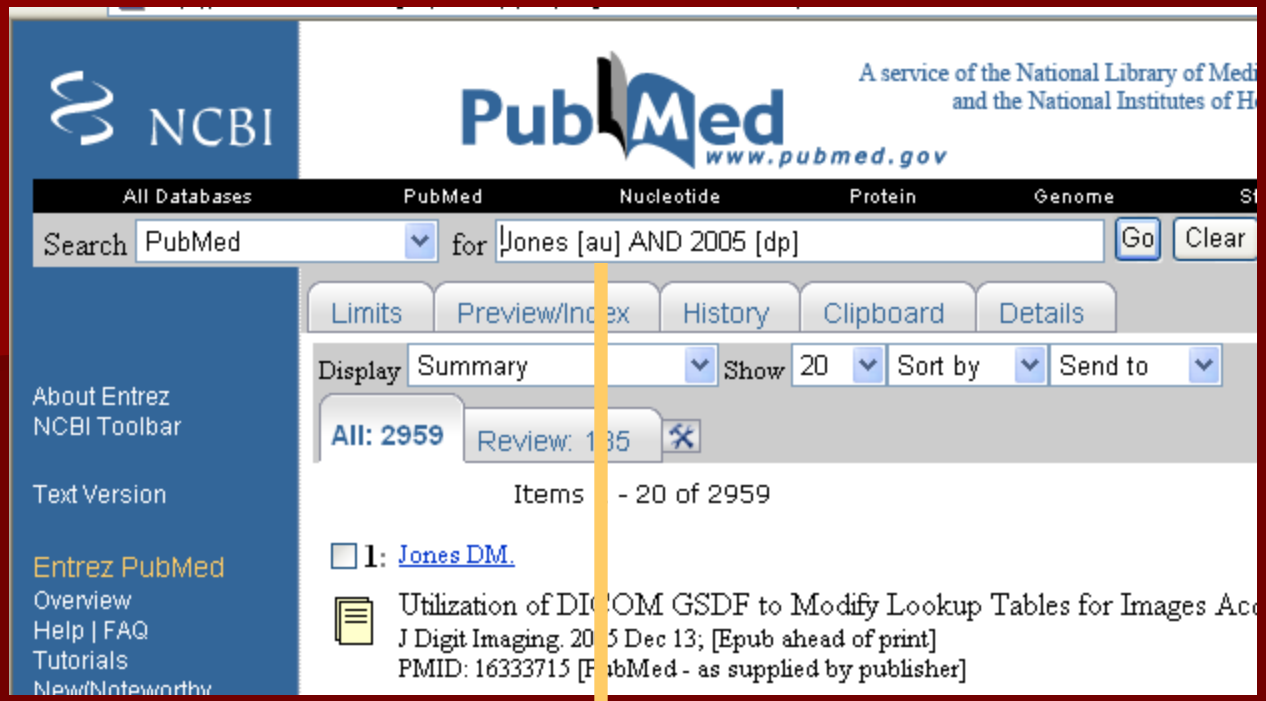
Different databases provide different formats for specifying fields.

Most use field names or nicknames

- Field ‘tags’ OR ‘labels’

which may follow a period or be placed in brackets or parentheses.

Some databases offer forms or drop-down menus.



PubMed

- Field tags go in [] and follow term
- Field tags can be used within Boolean queries

PubMed Field Tags

http://www.ncbi.nlm.nih.gov/books/bv.fcgi?rid=helppubmed.section.pubmedhelp.Search_Field_Descrip

[au] = author

[ti] = title

[tw] = textword

[tiab] = title and abstract

[mh] = medical subject heading

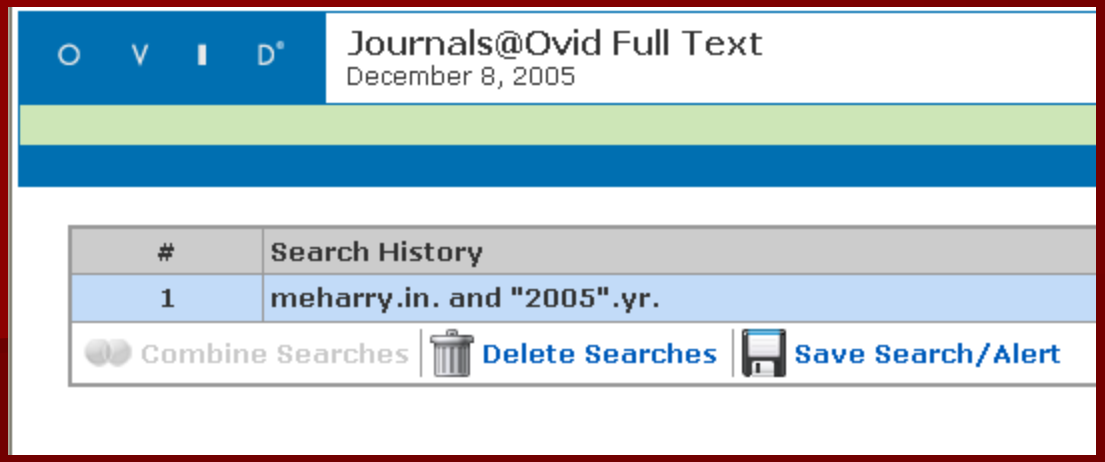
[dp] = date of publication

[la] = language

[gr] = grant number

[ta] = journal name

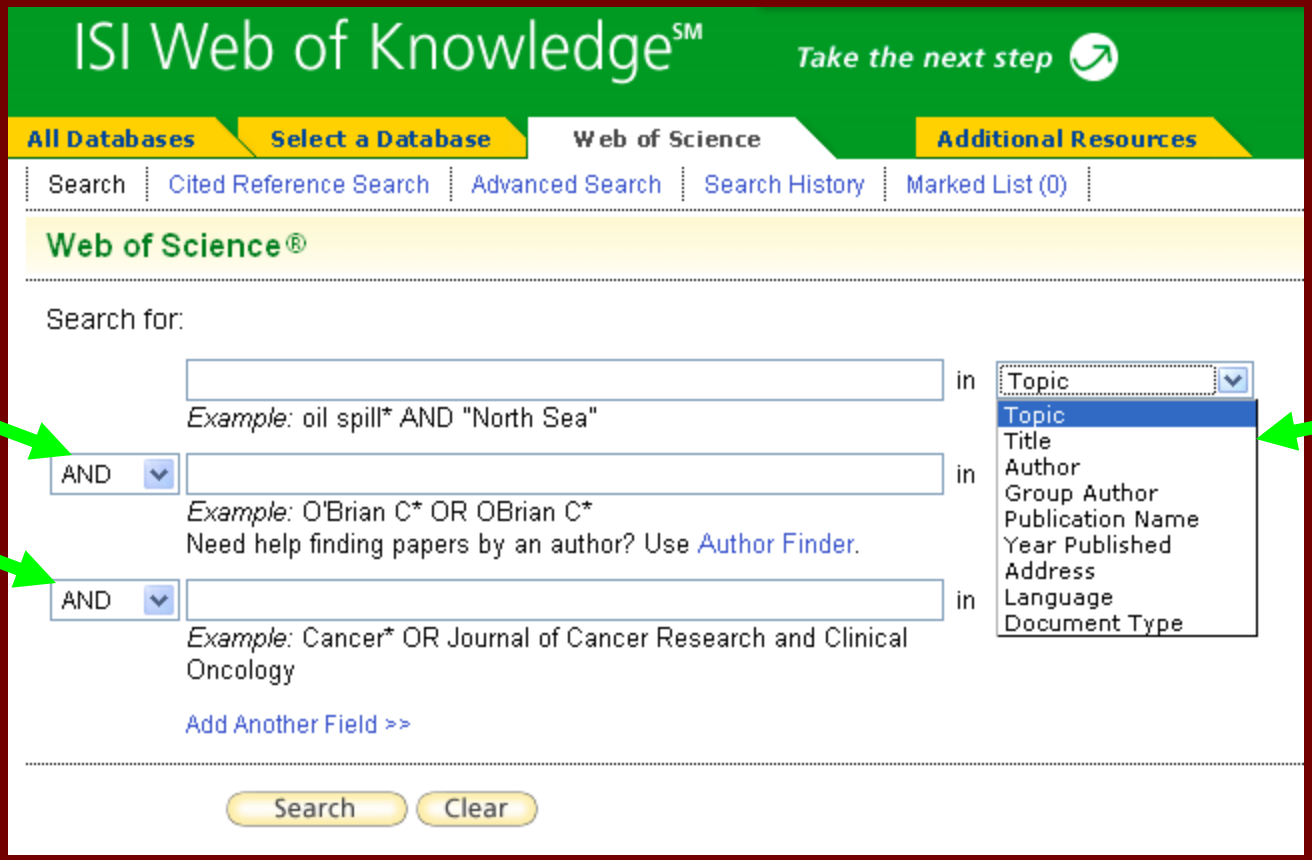
[ad] = affiliation



OVID databases require field tags to follow the term separated by a period.

Meharry.in and 2005.yr

Some databases, such as the ISI's Web of Science (Science Citation Index Expanded) provide forms to fill out.



Select Boolean Operator

Select Field from drop-down menu

Combining Field Searches

Multiple field searches can be combined using Boolean logic.

Find a 2005 article by an author named Hubble about ankle fractures.

Combine with AND:

- 2005 in date/year field
- Hubble in author field
- Ankle fractures in title field

2005[dp] AND Hubble[au] AND ankle fractures[ti]

Final Notes on Fields

- Each database provides its own specific fields
- Each database requires a specific format to designate field searching
- When searching a new database, take a moment to read the help documentation; most will provide a list of fields and how to search them.

Controlled Vocabulary

A controlled vocabulary is a set of established terms where

- every term represents a single concept
- only one term is used for that concept

Example: Many words can be used to represent the concept "people who teach."

teachers
faculty
instructors
professors
tutors
educators
lecturers, etc

In a controlled vocabulary one of these will be chosen to represent the concept

MeSH = Faculty
Faculty, Dental
Faculty, Medical
Faculty, Nursing, etc.

Another example

- How many words could you think of for the idea of "cancer"?
  - Cancer, tumor, malignancy, neoplasm, sarcoma…
- Articles in a database
  - Article one: "Breast tumors in young women"
  - Article two: "Surgery for prostrate cancer."
  - Article three: "Diagnosing Melanoma."

All three articles are about types of cancer but different terms are used in titles.

In a controlled vocabulary ONE word (*i.e.,* cancer) is chosen and placed in a special field, usually called a subject field.

For all three articles

- Article one: "Breast tumors in young women"
- Article two: "Surgery for prostrate cancer."
- Article three: "Diagnosing Melanoma."

The subject term (concept term) "cancer" is placed in the subject field by database indexers.

Now, Searching the database for cancer in the subject field will identify all records about the concept of cancer even if a different word for cancer is used.

- Search 'cancer-in-subject-field' finds all three articles
- How many articles would the search 'cancer-in-title-field' find?

Structure of Controlled Vocabulary

Broader Concepts

- All MeSH Categories
  - Analytical, Diagnostic and Therapeutic Techniques and Equipment Category
    - Diagnosis
      - Diagnostic Techniques and Procedures
        - **Diagnostic Imaging**

Narrower Concepts

- Image Interpretation, Computer-Assisted
  - Neuronavigation
  - Radiographic Image Interpretation, Computer-Assisted
  - Tomography, Emission-Computed +
  - Tomography, X-Ray Computed +
- Imaging, Three-Dimensional
  - Echocardiography, Three-Dimensional +
  - Holography
- Magnetic Resonance Imaging
  - Cholangiopancreatography, Magnetic Resonance
  - Diffusion Magnetic Resonance Imaging
  - Echo-Planar Imaging
  - Magnetic Resonance Angiography
  - Magnetic Resonance Imaging, Cine
  - Magnetic Resonance Imaging, Interventional

More on Controlled Vocabulary

- "Expanding" = Search includes all narrower terms beneath the searched term
  - Some databases do it automatically, others don't
- "Focusing" or "Majoring" = For a given item, some subject terms are considered the major focus; you can select to return only those articles.
  - Hip fracture[majr] = only give me articles where hip fracture is an important concept

Subject vs. Keyword Searching

Controlled vocabulary searching

- Matches terms against a specific field in the record.
- You need to consult a thesaurus (paper or online) to find out what the controlled vocabulary term is for each concept.

Free-text (keyword) searching

- Some concepts have many synonyms. A free-text search statement would mean "OR"ing all those terms together
- Matches terms against words anywhere in record (abstract, title, etc.).

Advantages to Controlled Vocabularies

- Using the controlled vocabulary can make your search more precise and easier.
- Increases the relevancy of results (fewer false drops)
- The indexers have already done much of the work for you.
- Searchable tree structures of terms can help you find new terms to use.

Problems with Controlled Vocabularies

- NOT all databases use a controlled vocabulary
- New concepts take time to be added
- There is often a lag phase during which the newest articles aren't indexed
- Controlled vocabularies can contain some very strange things and some concepts may not be handled well
- The controlled vocabulary must be easily searchable

Trying to understand what is and isn't in a particular controlled vocabulary can give you a big headache!

Combining Subject and Keyword Searching

To be comprehensive, it is often helpful to combine subject and keyword searching

(diabetes mellitus[mh] OR diabetes[tw])

(sickle cell anemia[mh] OR sickle cell anaemia[ti])

Don’t forget…

- Boolean logic to combine terms
- Use of other search fields in combination with subject terms

A Complex Search:

(head[mh] OR head[tw]) AND (wound and injuries[mh] OR trauma[ti] OR injury[ti]) AND 2005[dp] AND English[la]

Some Specialty Features

- Truncation
- Phrase searching
- Neighboring and other rarer Boolean operators

Truncation

What about including the singular and plural versions of words as well as other word variations?

For example: **therapy, therapies, therapeutics,**

You could combine them all in an OR relationship:

(therapy OR therapies OR therapeutics OR therapeutic)

But an easier way is by the use of truncation.

therap*

Each database handles truncation in a unique way.

The '*' and '$' are the most common wildcard symbols.

More on Truncation

Some examples:

Bacter$

Proc*

Vir?

Staph?

Be cautious when truncating!

If the word stem is too short, there may be too many possible variations and you might pick up unrelated terms.

For example, using proc* for finding procaine-like drugs will also include words like proceedings and process.

Phrase Searching

- Sometimes you want to force the database to search for a set of words in exact order

  "fever of unknown origin"

Most databases will accept a phrase in quotes.

BUT…some do not handle phrases well and will automatically break them up – usually 'AND'-ing the terms

Check how the database handles phrase searching before doing it!

Limits Options

- Many databases provide “limits” pages that make it easier for you to select common options such as language, article type, publication dates, human or animal, gender, age groups, etc.
- Each database’s limits options are unique
- Most limits can be done ‘by hand’ using field tags, but sometimes limit pages save time

PubMed Limits Page

Humans or Animals CLEAR

- [ ] Humans
- [ ] Animals

Gender CLEAR

- [ ] Male
- [ ] Female

Languages CLEAR

- [ ] English
- [ ] French
- [ ] German
- [ ] Italian
- [ ] Japanese
- [ ] Russian
- [ ] Spanish

More Languages

- [ ] Afrikaans
- [ ] Albanian

Subsets CLEAR

Journal Groups

- [ ] Core clinical journals
- [ ] Dental journals
- [ ] Nursing journals

Topics

- [ ] AIDS
- [ ] Bioethics
- [ ] Cancer
- [ ] Complementary Medicine
- [ ] History of Medicine

Type of Article CLEAR

- [ ] Clinical Trial
- [ ] Editorial
- [ ] Letter
- [ ] Meta-Analysis
- [ ] Practice Guideline
- [ ] Randomized Controlled Trial
- [ ] Review

More Publication Types

- [ ] Addresses

Ages CLEAR

- [ ] All Infant: birth-23 months
- [ ] All Child: 0-18 years
- [ ] All Adult: 19+ years
- [ ] Newborn: birth-1 month
- [ ] Infant: 1-23 months
- [ ] Preschool Child: 2-5 years
- [ ] Child: 6-12 years
- [ ] Adolescent: 13-18 years
- [ ] Adult: 19-44 years

Step-By-Step Search Construction

1. State the question
2. Identify the concepts in the question
3. For each concept, determine keywords and subject terms
4. Specify field tags after terms if needed
5. Combine terms for the same concept with “OR” in parenthesis
6. Combine “OR” statements with AND
7. Put any NOT terms at the end

Keep track of your searches, how many articles were found total, and how many you selected as relevant