Deconstructing databases: Becoming a better (re)searcher

iSkills Workshop Series
Inforum

Nalini K. Singh
January 2017
What is a database?

- A structured collection of digital data/information
- Information is stored in the system in a structured way, allowing users to search for it

- Some structure lingo:
  - A database is made up of **RECORDS** – each discreet item (e.g. a book, an article) is a RECORD in the database
  - Each record is made up of discreet **FIELDS** of information
  - Some common fields are illustrated below.
  - Fields are usually named with **FIELD LABELS**, e.g. title, subject, authors, etc.
  - This information can be thought of as the **DESCRIPTIVE METADATA**, i.e., data about the particular item that each database record describes, whether it is a book, an article, etc. so that users can discover it
  - **USER INTERFACE**: point of contact between searcher and the information in the system.

**RECORD OF A BOOK FROM THE U OF T LIBRARY CATALOGUE**

<table>
<thead>
<tr>
<th>Holdings</th>
<th>Details</th>
<th>Subjects</th>
<th>A Look Inside</th>
<th>MARC View</th>
</tr>
</thead>
<tbody>
<tr>
<td>Author</td>
<td>Perelman, Michael, 1957-</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Title</td>
<td>The power of trees: the reforesting of the soul / Michael Perelman</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Imprint</td>
<td>Dallas : Spring Publications, c1994</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ISBN</td>
<td>086214962X (pbk. : acid-free paper)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Catalogue key: $36013
Includes bibliographical references (p. [222]-254) and index.
Find other items about the same subject:
- **Subject**: Nature—Psychological aspects
- **Subject**: Trees—Psychological aspects

**RECORD OF A JOURNAL ARTICLE FROM HUMANITIES ABSTRACTS DATABASE**

**Title**: Our Past and Present Selves.

- **Authors**: Allen, Brooke¹
- **Source**: Hudson Review; Winter 2012, Vol. 64 Issue 4, p712-716, 9p
- **Document Type**: Book Review
- **Subjects**: Spadea, Patricia Meyer; Emotions (Psychology); Reading; Nonfiction
- **Abstract**: The article reviews the book "On Rereading" by Patricia Meyer Spadea.
- **Author Affiliations**: ¹Swarthmore College
- **ISSN**: 0018702X
- **Accession Number**: 70244376

- **Author field**
- **Title field**:
- **Imprint field**: Place & year of publication, publisher name
- **ISBN field**
- **Subject fields**: Nature—Psychological aspects
  - Trees—Psychological aspects

**TITLE field** in this record is set apart from the other information, and does not have a field label.
Research databases @ U of T: Getting there

A) Individually listed by title in the catalogue
B) Some are listed in a small subset of popular databases
C) Grouped by categories in Subjects A–Z – this is the fastest way!

A) CATALOGUE –
Databases listed individually in the catalogue
- Pick the CATALOGUE page.
- Enter database title in double quotes E.g. "alternative press index"
- Choose TITLE search.
- Follow the link in catalogue record to get to the database

B) POPULAR DATABASES
- A much shorter list of popular databases.
- Also lists important interdisciplinary databases such as Scopus and Web of Science

C) SUBJECTS A–Z
- This is a list of hundreds of databases organized by subject area.
- This page also offers lists of REFERENCE TITLES and PRIMARY SOURCES by title
Different kinds of databases

Knowing about differences in scholarly research databases can help you find more, and more relevant results, with less time spent searching.

Databases can be categorized in a number of ways. Here are some. These are NOT mutually exclusive categories.

**Subject-specific databases**
- Databases of articles from journals & magazines covering a specific subject area
- Smaller in size because they cover the journal literature of a smaller area of study
- Each article is usually described in terms of SUBJECTS (tags, descriptors, subject terms)
- Subjects usually are drawn from a CONTROLLED VOCABULARY
- An area of study may have one or databases relating to that area.
  
  **EXAMPLES:** *International Political Science Abstracts; Communications Abstracts; Medline*

**Interdisciplinary databases**
- Databases of articles from journals & magazines covering ALL areas of studies
- Much larger in size because they cover the journal literature from many different areas
- Usually lacking a controlled vocabulary, relying on author-supplied subject terms.
- May be split up into smaller chunks: e.g. Web of Science is made up of 3 smaller databases.
  
  **EXAMPLES:** *Web of Science, Scopus*

**Databases focused on special document types, or types of info**
- Not all databases focus on journal articles relating to topics.
- Below is a short list of databases that focus on other types of content

<table>
<thead>
<tr>
<th>Databases</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>DISSERTATIONS</td>
<td>Proquest Dissertations &amp; Theses Full Text</td>
</tr>
<tr>
<td>BOOK REVIEWS</td>
<td>Book Review Index Online</td>
</tr>
<tr>
<td>NEWSPAPER ARTICLES</td>
<td>Proquest Newsstand</td>
</tr>
<tr>
<td>DIRECTORIES</td>
<td><em>Urich’s Periodical Database</em></td>
</tr>
<tr>
<td></td>
<td>Information about 300,000+ serial publications (journals, magazines, newsletters, etc.) published worldwide</td>
</tr>
<tr>
<td>LIBRARY CATALOGUE</td>
<td>An online library catalogue is a database that lists the books, journals, and other holdings in a particular library or library system e.g. U of T Libraries catalogue</td>
</tr>
<tr>
<td>COUNTRY FOCUS</td>
<td><em>Canadian Periodical Index (CPI.Q)</em></td>
</tr>
</tbody>
</table>
Subscription vs. free (open access) databases

- Most article databases and other research databases are subscription-based products, published by an indexing service, and sold to libraries and other institutions.
- Databases may be sold by large information companies such as Proquest, Ebsco, Elsevier, Thomson-Reuters, etc., or by smaller companies, or non-profit organizations, e.g. JSTOR.
- There are some freely available databases, which are part of the Open Access movement.

EXAMPLES OF OPEN ACCESS DATABASES

9,487 journals, 6,715 of which are searchable at article level. Contains 2.4 million articles!

Provides both a directory listing of the academic research repositories available worldwide, and a customized Google search to search the contents of ALL listed repositories!

Mendeley – http://www.mendeley.com/research-papers/
Crowdsourced catalogue of research papers (many from journals)

Full-text vs non-full-text databases

Full-text databases

- These databases are often journal repositories. A repository's mandate is usually to house full-text documents.
  EXAMPLES: JSTOR; Scholars Portal Journals

- Some full-text databases may originate from one organization/publisher
  EXAMPLE: ACM Digital Library (full-text of all ACM journals, magazines, conferences, etc.)

Combination

These databases have some full-text content, a citations-only content. A subscribing library may link from citation-only content to full-text content coming from another supplier/vendor.

EXAMPLES: Art Full Text; Library Literature & Information Science Full Text

Citations only

- These may be large interdisciplinary databases, with NO full text content. There may be links to full-text articles coming from another vendor.
  EXAMPLES: Scopus; Web of Science

- They may also be smaller subject-specific databases, with only citations and abstracts.
  EXAMPLE: Library & Information Science Abstracts (LISA)
Different ways to search databases

Single field search, or a search that covers multiple fields

**SINGLE FIELD**: e.g. Author OR Title OR Abstract | **MULTIPLE FIELD SEARCHES**

<table>
<thead>
<tr>
<th>Searchable fields: Ebsco</th>
<th>Searchable fields: Proquest</th>
<th>Searchable fields: Scopus</th>
</tr>
</thead>
<tbody>
<tr>
<td>Select a Field (optional)</td>
<td>All fields + text</td>
<td>All fields</td>
</tr>
<tr>
<td>TX All Text Fields</td>
<td>All fields (no full text) = ALL*</td>
<td>Article title, Abstract, Keywords</td>
</tr>
<tr>
<td>AU Author Personal</td>
<td>Abstract = AB*</td>
<td></td>
</tr>
<tr>
<td>CA Author Corporate</td>
<td>Author = AU*</td>
<td></td>
</tr>
<tr>
<td>RS Responsibility Statement</td>
<td>Document text = FT*</td>
<td></td>
</tr>
<tr>
<td>PB Publisher</td>
<td>Document title = TI*</td>
<td></td>
</tr>
<tr>
<td>PD Physical description</td>
<td>Publication title = PUB*</td>
<td></td>
</tr>
<tr>
<td>SU Subject</td>
<td>Subject heading (all) = SU*</td>
<td></td>
</tr>
<tr>
<td>AB Abstract</td>
<td>Company/organization = ORG*</td>
<td></td>
</tr>
<tr>
<td>LA Language</td>
<td>Location = LOC*</td>
<td></td>
</tr>
<tr>
<td>IS ISSN</td>
<td>Person = PER*</td>
<td></td>
</tr>
<tr>
<td>IB ISSN</td>
<td>More options</td>
<td></td>
</tr>
<tr>
<td>NT Notes</td>
<td>ISSN = ISSN*</td>
<td></td>
</tr>
<tr>
<td>SO Journal Name</td>
<td>Section = SEC*</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Tag = TAG*</td>
<td></td>
</tr>
</tbody>
</table>

**About KEYWORD field**
- KEYWORD may mean different things in different databases
- Sometimes it means a MULTI-FIELD SEARCH that searches the article title, author name(s), journal name, subjects, and abstract.
- Sometimes it is a SINGLE-FIELD SEARCH that searches author-supplied subject tags.
- Read the database help info to find out exactly what you’re searching.

**Boolean searching**
- Use of Boolean operators, AND, OR, NOT to combine search terms in a logical way.
- AND = narrow a search
- OR = expand a search
- Some databases offer a single search box where a Boolean search may be entered
- Boolean searches can also be run in an ADVANCED SEARCH – see next section

**EXAMPLE**
(internet OR web) AND (young adults OR teen* OR children) AND (public libraries OR school libraries)
Advanced searching

- Can combine multiple search terms from different fields with different Boolean operators
- Usually get 3 rows of search boxes
- Usually has the ability to add more search rows

A BOOLEAN SEARCH EXPRESSED OVER 3 ROWS OF AN ADVANCED SEARCH SCREEN

<table>
<thead>
<tr>
<th>Basic Search</th>
<th>Advanced Search</th>
</tr>
</thead>
<tbody>
<tr>
<td>FIND</td>
<td></td>
</tr>
<tr>
<td>internet or web</td>
<td>Keyword</td>
</tr>
<tr>
<td>and</td>
<td>Keyword</td>
</tr>
<tr>
<td>young adults or teenagers or children</td>
<td>and Subject(s)</td>
</tr>
<tr>
<td>and</td>
<td>Subject(s)</td>
</tr>
<tr>
<td>public libraries or school libraries</td>
<td></td>
</tr>
</tbody>
</table>

Searching multiple databases selected by you!

- Databases sold by a particular vendor e.g. Proquest, Ebsco, Ovid, may be searched individually, or you can select some, or even all, to search at once.
- The search interface is the same for all databases in the group since it is the same vendor.

PARTIAL LIST OF EBSCO DATABASES SEARCHABLE AT UTL
**Federated search of many, many databases**

A simultaneous search of multiple databases over multiple search platforms, e.g. Proquest databases plus Ebsco databases plus databases from other vendors.

- Search results are then presented in a uniform way.
- The search at UTL is a federated search at UTL.
- The search engine is called Summon.
- Summon searches the contents of about 1000 different databases subscribed to by UTL.

**Cited reference (citation) searching**

- This search lets you find out which other articles have cited a specific article.
- You can also look up an individual author and generate a list of articles that have cited that individual author’s articles.
- *Web of Science (WOS)*, and *Scopus* are databases that allow this kind of searching.
- Other more subject-focused databases may offer this feature as well, but *WOS* and *Scopus* are the most comprehensive.*

![WEB OF SCIENCE](image)

- Cited reference searching is useful for:
  - Finding more articles on the topic of interest
  - Seeing how ideas change over time
  - Seeing the impact of a particular article, or of the body of work of individual researchers & scholars

*Web of Science* started life as a print product called *Science Citation Index* in the 1960s. It led to two other print products – *Social Science Citation Index*, & *Arts/Humanities Citation Index*. The 3 combined indexes make up the online *WOS*. Citations in *WOS* go back to the 1900s in some cases, while *Scopus* was launched in 2004, has citations going back to 1996.
Common database features & functionalities

- You will be searching many databases while doing your research.
- Search interfaces vary from database to database
- This means that across databases, there may be different:
  - Ways of presenting search results
  - Ways of describing features and functionalities that tools you can use as you search
  - Terminology used to describe the same thing (e.g. PUBLICATION TITLE vs SOURCE TITLE)
  - Types of limits and search refinement features
- However, they all mostly offer common ways of doing things
- Familiarizing yourself with common features available in most databases, and looking for them as you explore, will improve your search efficiency and effectiveness.

| A. Syntax |
| B. Limits |
| C. Facets |
| D. Field-specific indexes |
| E. Thesaurus |
| F. Changing the display |
| G. Output options |
| H. Search history |
| I. Alerts, feeds, etc. |

A) Syntax – or, the way you write the search
- Does the search require you to type in Boolean or other operators?
- If you put a string of words into a search box, does it look for ALL the words? (i.e. is the Boolean AND implied?)
- In an AUTHOR search, do you have to say LAST NAME [comma] FIRST NAME?
- Which truncation symbols are used?
- Phrase searching: quotes or no quotes?

B) Limits
Most databases offer some way to limit results, BEFORE or AFTER running your search, e.g.
- Peer-reviewed
- Full text availability
- Language
- Publication date ranges
- Document types, eg. Articles, reviews, conference paper
- Other more specialized limits might be available
C) Facets, or categorization of your results list

- These categorizations often appear to the side of your search results.
- Lets you work with your results to NARROW them down in useful ways.
- i.e. you can see smaller categorized subsets of the results you found.

- Common facets are:
  - Source type (e.g. journal, book, conference proceeding)
  - Document type (e.g. scholarly article, review)
  - Authors
  - Date
  - Subjects: particularly useful in helping you identify the subject terms used by the database, if these are supplied
  - Publication title
  - Language
  - Geographic coverage

D) Field-specific indexes

Some indexes allow you to browse special lists e.g.
- All AUTHOR names contained in the database
- All SUBJECT descriptors used in the database
- More specialized indexes may be available (e.g. list of JOURNAL TITLES)

Browsing field-specific indexes (lists) such as lists of all authors in AUTHOR field or SUBJECTS field or PUBLICATION NAME field is useful because you can:
- Get at all articles by the same author in the database despite any name variations (e.g. W. Duff, W. M. Duff, Wendy M. Duff, Wendy Duff)
- Look up subject terms used by the database to find articles on that topic
- Look up names of journals to see if their articles are included in the database.

This feature might be called INDEXES, or BROWSE, or available in a tab called OTHER.
E) A thesaurus (hierarchy of subjects) may be available.

A thesaurus is useful to search for articles on a topic because you can use it to:
- Look up subject terms used by the database
- See where the term fits into a conceptual hierarchy, so you can discover (and then search for) narrower terms, or broader terms, or related terms of that concept
- A thesaurus may also refer you from a subject NOT used to the right one.

F) Changing/customizing the display:

You can change how the search results are displayed on your screen:
- Sort results in different ways – by DATE or RELEVANCE
- Show more results: This saves you time – easier to scroll/scan through a long list of results, than to click through page after page
- Show brief info or full info – e.g. display the abstract or not
- Some interfaces let you choose precisely which fields of info to display.

G) Marking & outputting search results

- You can select (MARK) multiple items, to work with them as a group.
- The MARKED RECORDS are temporarily saved to a folder or list.
- From the MARKED LIST or folder, the group of selected items can then be printed, emailed, saved, exported to a reference manager like RefWorks, or EndNote.
Some variations on the MARKED LIST

FOLDERS (EBSCO APPROACH)

MY LIST (SCOPUS APPROACH)

SELECTED ITEMS (PROQUEST APPROACH)
H. Working with a search history

You can track searches for an individual session in a SEARCH HISTORY, from where you can:

- Re-run searches
- Combine different searches with AND, or OR
- Use this feature to search for selected topics in a selected group of journals:
  - Search 1: Journal A OR Journal B OR Journal C OR Journal D
  - Search 2: Topic 1 OR Topic 2 OR Topic 3.
- Working from the search history screen, combine Search 1 AND Search 2.

Your search histories may be saved to a special password-protected space, to use later.

I. Setting up alerts

ALERTS are a current awareness feature that allows you to track new results on a topic.

- First, develop a search statement that captures your topic.
- Then, save it to a password-protected space provided by the database
- When new records matching this saved search statement are added to the database, you get email notification with citations of those new records.
- This is very useful when you are researching a topic over an extended period of time.

General search tips

- Define your search carefully
  Break it down into multiple concepts, and use synonyms for each concept
  Use multiple search boxes and Boolean operators.

  - Synonyms & related: Global warming OR climate change OR greenhouse gases
  - Level of specificity: Pets vs. dogs vs. black Russian terriers
  - Word variants: Colour vs. color

- Use appropriate search statements for the database in which you are searching:

  Single words: museology
  Multiple words: academic AND libraries AND “collection development”
  internet OR web
  networks NOT computer
  Phrases: “archival finding aids”
  “enterprise resource planning systems”
  “preservation of cultural heritage”
  Truncation to search canad* gets you canada, canadian, etc.
  for word variants: wom?n gets you woman, women
  Symbols vary from database to database
  Command search language: (reference services and management) in SU
  Syntax varies from database to database
• **Try different kinds of searches:**
  o Do broad searches (e.g. ANYWHERE or ALL field) to search the whole record, OR
  o Do more precise searches, e.g.:
    ▪ AUTHOR search
    ▪ TITLE search
    ▪ SUBJECT search
  o Search for subjects in different ways:
    ▪ Choose the SUBJECT field, after finding out what subjects are valid in the database in which you are searching. SUBJECT may be called DESCRIPTOR, depending on the database
    ▪ Browse the SUBJECT INDEX, if available, to check for valid subject terms
    ▪ Use the THESAURUS if one is available.

• **Search in more than 1 database!**

  Think about all the disciplines that your topic may fit in.
  For example, management information systems articles may be found in BUSINESS & MANAGEMENT databases as well as COMPUTER SCIENCE databases

• **USE the help feature of the database.**
  o Search interfaces can change regularly and without much notice.
  o Guides like this one can quickly become outdated!
  o If you familiarize yourself with features/functions common to databases in general, you can apply this knowledge to them all.

**Hints & hazards re: database searching**

**Is your journal covered?**
Check the journal source lists (journals covered by the database) if available, to see if your journal of interest is actually included in the database you are using.

**Time periods per journal covered?**
A database may not cover a journal from when the journal was first published. It may also decide to stop covering a journal altogether.

**Time range**

**Does the database cover the particular document type you are after?**
For example, some databases may not include book reviews.
Selective indexing
- Databases do not always include all articles from all journals they cover.
- While many will include all articles in a journal, the database service may decide to omit some articles in a journal, such as editorials, reviews, articles not in the subject area, etc.

Vocabulary varies
- Each database uses its own set of subject descriptors, so always check to see that you are using valid descriptors for the database you are in.
- Terminology may differ from continent to continent (LISA uses British terminology and spelling, Library Literature uses North American terminology and spelling).
- Keeping this in mind while doing searches in multiple databases simultaneously is important!

Time lag
- There may be a time delay before new articles make it into a database as human indexers take time to process new material.
- Some databases have a faster update rate. Databases with abstracts may take longer to update than those without abstracts.

Glossary of selected terms

**Citation:** "A reference to a text, or part of a text, identifying the document in which it may be found" (Harrod’s Librarians Glossary, 9th ed. 2000). In a database, usually refers to the brief version of a record, containing enough info necessary to find the item in a library, a database, or elsewhere, e.g. TITLE, AUTHOR, SOURCE, YEAR.

**Controlled vocabulary:** “Established list of preferred terms from which a cataloguer or indexer must select when assigning subject headings or descriptors in the bibliographic record to indicate the content of a work” (Dictionary for Library and Information Science, 2004).

**Database:** An organized electronic, searchable collection of related information. Made up of records, which are in turn made up of fields. Article databases are the digital version of what were know as periodical indexes in the print-only era.

**Display:** How or what kind of information from a database record, or set of database records, is shown on the screen. Displays can be changed to show more or less information about the items found in a search, or to sort the items in a particular order.

**Electronic journal:** Also called e-journals or digital journals. “A journal for which the full end product is available on optical disc, over a network or in any other electronic form. Strictly, a journal in which all aspects of preparation, refereeing, assembly and distribution are carried out electronically.” (Harrod’s Librarians Glossary, 9th ed. 2000). Commonly used to describe the electronic form of a journal that is also published in paper.
Fields: the bits of discrete information which make up a database record. Each field contains a specific kind of info – author's name, article title, abstract, the subject terms.

Field-specific indexes: also called browsable indexes: Browsable lists of particular information such as all authors in a database, or all subject descriptors in a database, so that database users can search for variations in author names, see subject terms, etc.

Format: The physical type of material. For example, journals may be available in paper and/or electronic formats.

Full-text: when the full content of the item (e.g. a journal article) is available in electronic form.

Index: A resource “which specifies, (or) indicates ... the information, contents or topics of a document or group of documents.” (Harrod's Librarians Glossary, 9th ed. 2000). In the online age, the word ‘index’ has mostly been replaced by ‘database’ or ‘article database’

Periodical indexes (usually now called article databases) contain information about articles found in journals, and may be either general (covers all subject areas); or subject-specific (covers a particular field of study).

Interface: also called user interface: "the facility that enables interaction to take place between a PC and an end-user" (Harrod's Librarians Glossary, 9th ed. 2000). The interface between the database and user is where the search for the database content takes place.

Journal: A periodical, especially one containing scholarly articles and/or disseminating current information on research and development in a particular subject field. (ALA Glossary, 1983). SEE also Periodical.

Operators: special words used to combine search terms for a more complex search. Different types of operators e.g. Boolean operators (AND, OR, NOT, which are used to narrow down, expand or exclude items from a search). Other operators are: NEAR, IN.

Periodical: “A publication with a distinctive title which appears at stated or regular intervals, without prior decision as to when the last issue shall appear.” (Harrod’s Librarians Glossary, 9th ed. 2000)

Publisher: “The firm or other corporate body or the person responsible for the manufacture and distribution of a document to the public.” (ALA Glossary 1983)
Publishers are the producers of journals but not necessarily the suppliers of the journal in its electronic format.

Records: units of information which make up a database. Database size is measured by how many records it contains. Databases may contain thousands or even millions of records. 1 record = 1 unit of information (e.g. 1 article, 1 report, 1 book)
**Search terms:** words you type into the search box common to most database interfaces to look for information in the database.

**Stop words:** words of little meaning occurring too often to be useful in a search, or words that are used as operators, so cannot be used as search terms by themselves (e.g. *and, a, the, or, is, an*).

**Subject descriptors:** Also called descriptors, or subject keywords. These are words used to describe the topics discussed in an article, book, etc. Users can search a database by subject descriptors in order to find articles on a topic.

**Thesaurus:** "A compilation of groups of words, consisting of the links between words used in documents and words used as descriptors, prepared for consultation in information retrieval. They display relations within the vocabulary based on semantics, not orthography." (Harrod’s Librarians Glossary, 9th ed. 2000). A thesaurus in a database will help you pick good subject descriptors to find articles on a topic.

**Truncation symbols:** Also known as wild cards. Characters you use to shorten your search term to look for variations e.g. spelling, plurals – *librar* to find “libraries”, “library”, “librarians” or *wom?n* to find women and woman. Symbols vary in different databases.

**Vendor (Host):** “A commercial enterprise that makes available access to any number of databases via its own computer, and using a common command language.” (Harrod’s Librarians Glossary, 9th ed. 2000) Vendors (sometimes called information providers) provide e-journals articles in databases, with their own standards for interface design, and search functionalities.