

Chapter **Three**: Information retrieval & EBM

Information retrieval & EBM

After completing this chapter, you should know the answers to these questions:

- What types of online content are available and useful to health care practitioners, researchers, and consumers?
- What are the major components of the information retrieval process?
- What are the major approaches to retrieval of knowledge-based biomedical information?

Definition for Information retrieval (IR)

- Information retrieval (IR) is the field concerned with the acquisition, organization, and searching of knowledge-based information (Hersh, 2003).
- It means finding material (usually documents) of an unstructured nature (usually text) that satisfies an information need from within large collections (usually stored on computers).
- Information Retrieval System is a capable of storing, maintaining from a system and retrieving of information. This information may be any form that includes **audio, video and text**.

Definition for Information retrieval (IR)

- Information Retrieval System is mainly focus on **electronic searching** and retrieving of documents.
- Information Retrieval (IR) is devoted to finding **relevant documents**, not finding simple matches to patterns.

Cont...

- Modern information retrieval systems deal with storage, organization and access to text, as well as multimedia information resources.
- So, information retrieval is collectively defined as a **“science of search”** or a process, method and procedure used to select or recall recorded and/or indexed information from files of data.

Major functions of an IRS:

- To **identify** sources of information relevant to the areas of interest of the target users
- To analyze the contents of the sources (documents)
- To represent the contents of the analyzed sources for matching with the users' queries
- To match the search statement with the stored database
- To retrieve the information that is relevant

Big Issue In IR:

1. Precision and Recall:

- To measure effectiveness of IRS, two ratios are used: **precision** and **recall**.
 - **Precision:**
- It is the ratio of the number of **relevant documents retrieved** to the total number **retrieved**. Precision provides an indication of the quality of the answer set.
- However, this does not consider the total number of relevant documents. A system might have good precision by retrieving ten documents and finding that nine are relevant (a 0.9 precision), but the total number of relevant documents also matters.

PRECISION AND RECALL:

➤ Recall

- **Recall:** considers the total number of relevant documents; it is the ratio of the **number of relevant documents retrieved** to the total number of documents in the collection that are believed to be relevant.

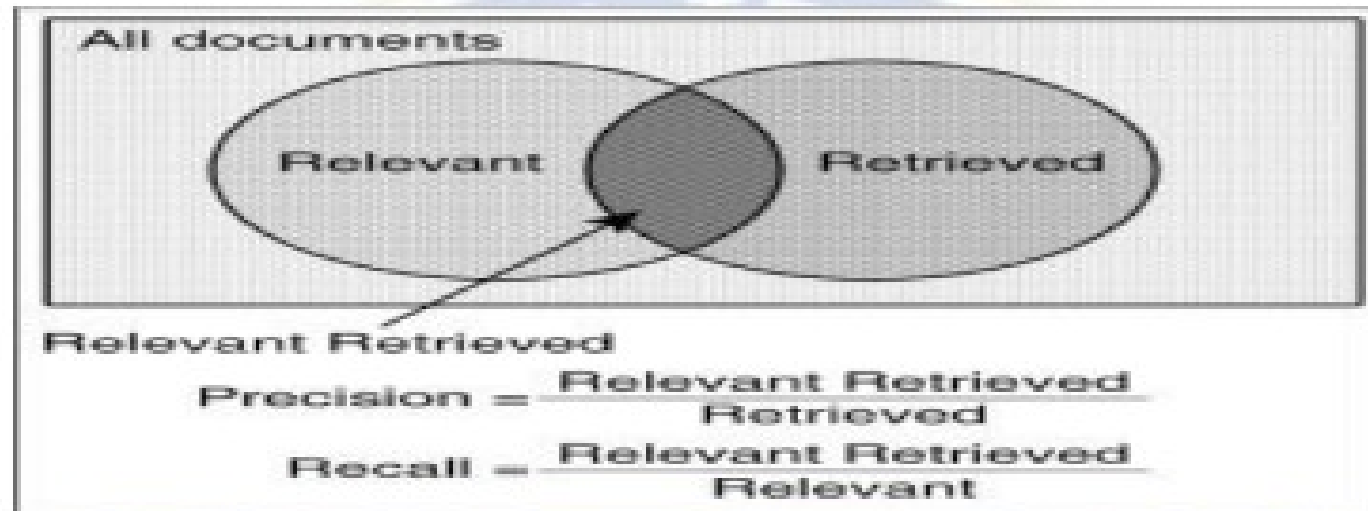


Fig: Precision And Recall

Big Issue In IR

2. Evaluation

It is necessary early on to develop **evaluation measures** and experimental procedures for acquiring quality data.

Two of the measures used, **Precision and recall**, are still popular

Big Issue In IR

3. Information needs

The users of a search engine are the ultimate judges of quality. This has led to numerous studies on how people interact with search engines and, in particular, to the development of techniques to help people express their information needs.

Components of a Search Engine

- **A search engine** is the practical application of information retrieval techniques to large-scale text collections.
- **A web search engine** is the obvious example
 - These days we frequently think first of web search, but there are many other cases:
 - E-mail search
 - Searching through your laptop
 - Legal information retrieval

IR vs. databases: Structured vs unstructured data

- Structured data tends to refer to information in “tables”

Employee	Manager	Salary
Smith	Jones	50000
Chang	Smith	60000
Ivy	Smith	50000

Typically allows numerical range and exact match (for text) queries, e.g.,

Salary < 60000 AND Manager = Smith.

Unstructured data

- Typically refers to free text
- Allows
 - Keyword queries including operators
 - More sophisticated “concept” queries

How to find information on the Web?

- You can find information by two basic means.
- **Search by Topic** and **Search by keywords**.
- Some search services offer both methods, others only one.
- Yahoo offers both.

➤ Search by Topic

You can navigate through topic lists

➤ Search by keywords





A service of the U.S. National Library of Medicine
and the National Institutes of Health

Subscription based: **PubMed-MEDLINE** is the premier North American biomedical index (millions of articles from 5,000 journals)

- Ways to access for free:
 - Check library, they may subscribe
 - Do a one-time trial subscription

What is MEDLINE ?

MEDLINE is an acronym for Medical Literature Online.

MEDLINE is the National Library of Medicine's premier bibliographic database covering the fields of medicine, nursing, dentistry, veterinary medicine, the health care system, and the preclinical sciences.

MEDLINE currently contains over 16 million references dating back to 1949. Coverage is worldwide, but most records (about 90%) are from English-language sources or have English abstracts.

Open access databases

- What makes it different:
 - ❖ Publishers and journals submit their articles for free
 - ❖ Access is free and unrestricted
 - ❖ Allows people from all over the world to read and download the full text of article

Open access databases

- PubMed Central
- BioMed Central
- Public Library of Science (PLOS)
- Google scholar
- CDC publications
- AFENET publications
- HINARI

PubMed Central (PMC)
www.pubmedcentral.nih.gov/

- Part of the U.S. National Institutes of Health free digital archive of **biomedical** and **life sciences** journal literature
- Electronic archive of full-text journal articles, with free access to its contents
- Over half a million articles, most of which have a corresponding entry in PubMed



**PubMed
Central**

A free archive of life sciences journals

[About PMC](#)

[Journal List](#)

[For Publishers](#)

[Utilities](#)

PubMed Central (PMC) is the U.S. National Institutes of Health (NIH) free digital archive of biomedical and life sciences journal literature.

Find Articles

[Advanced search](#)

Browse PMC journals: [\[A-B\]](#) [\[C-H\]](#) [\[I-M\]](#) [\[N-S\]](#) [\[T-Z\]](#) [\[Full List\]](#) [\[New Journals\]](#)

Receive notice of new journals and other major updates to PMC: join the [PMC News mail list](#) or subscribe to the PMC News [RSS feed](#) .

All the articles in PMC are free (sometimes on a delayed basis). Some journals go beyond free, to [Open Access](#). Find out what that means.

PMC's [utilities](#) include an OAI service that provides XML of the full-text of some articles, functions for scripting PMC searches and linking to specific PMC articles from your site, and more ...

Looking for a modern journal article DTD? Take a look at NLM's [Journal Publishing XML DTD and schema](#).

It's about preservation and access: [digitizing the complete run of back issues](#) of many of the journals in PMC.

The [PMC journal list](#) comprises journals that deposit material in PMC on a routine basis and generally make all their published articles available here. Find out how to [include your journal](#) in PMC.

PMC also has the [author manuscripts](#) of articles published by NIH-funded researchers in various non-PMC journals. Increasing free access to these articles is the goal of the [NIH Public Access](#) policy. Similar manuscripts from researchers funded by the Wellcome Trust are available in PMC as well.

Eligible researchers should use the [NIH Manuscript Submission](#) system to deposit manuscripts.

Get [answers](#) to other questions about PubMed Central.

BioMed Central
www.biomedcentral.com/



- Independent publishing house providing immediate open access to peer-reviewed biomedical research
- 196 journals listed

Public Library of Science (PLOS)
www.plos.org/



Nonprofit organization of scientists and physicians
committed to making the world's scientific and
medical literature a freely available public resource.

Google scholar
<http://scholar.google.com/>



- Searches across many disciplines and sources:
 - peer-reviewed papers, theses, books, abstracts and articles, from academic publishers, professional societies, preprint repositories, universities and other scholarly organizations
- Helps identify most relevant research across world of scholarly research

CDC publications www.cdc.gov

- Emerging Infectious Diseases (www.cdc.gov/ncidod/eid/)
- Morbidity and Mortality Weekly
Report (www.cdc.gov/mmwr/)
- Preventing Chronic Disease (www.cdc.gov/pcd/)
- Other publications (www.cdc.gov/Publications/)

Health Inter-Network Access to Research Initiative



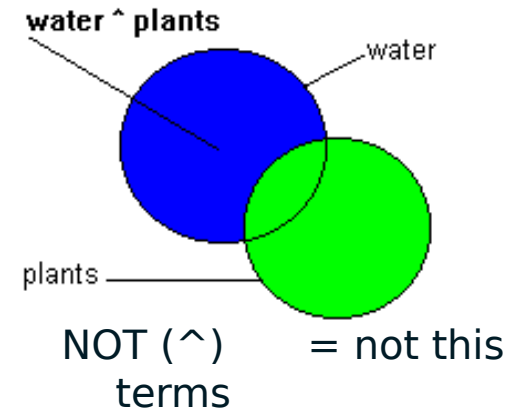
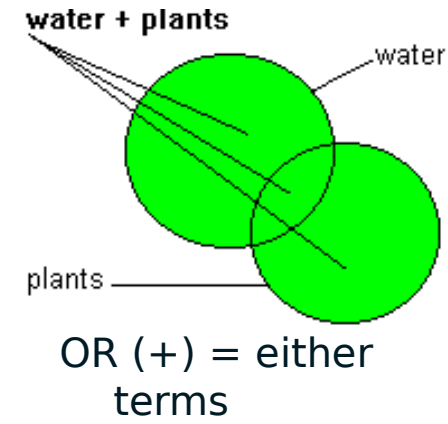
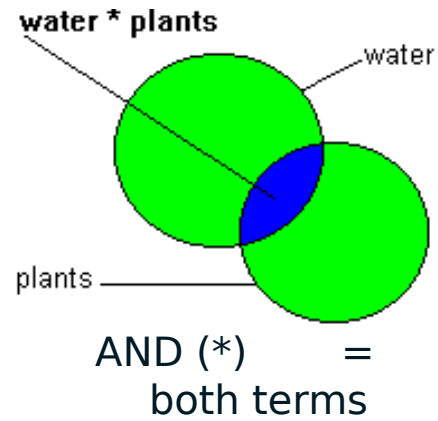
- Free database for **developing countries**
- Provided by WHO and major publishers
- Gives developing countries access to one of the world's largest collections of biomedical and health literature
- Has more than 6,000 journal titles available to health institutions in 108 countries

Learn how to enter the search into the database or search engine using Boolean searches

- **Definition:** a web search allowing the use of words like AND, OR, or NOT to limit, widen, or define search
- Using AND narrows search by combining terms
- Using OR expands search
- Using NOT narrows search by excluding terms

Boolean Operator

- AND (*) = both terms
- OR (+) = either terms
- NOT (^) = not this term



An example

- Looking for materials on **malaria** and **medication** for children < 12 months **in West Africa**
- Identify keywords (female anopheles, Plasmodium, artmesinin, insecticide-treated bednets, infants, West Africa)
- Use Boolean operators: start with (malaria AND medication)
- Filter search result....keep filtering

Example of a Boolean search

Database: PubMed Central

1. Malaria AND medication: 1,439
2. Malaria AND medication AND children: 842
3. Malaria AND medication AND infants NOT children: 27
4. Malaria AND medication AND infants NOT children
AND West Africa: 2

Tips on conducting Boolean searches

- Never use Google (too many useless links)
- Use the science databases we discussed
- Use the “OR”, ‘AND’ “NOT” to facilitate the nesting technique such as: bird flu AND avian flu OR H5N1 AND Africa

Don't forget the “snowball”
technique

To find relevant articles:

- Locate the most recent article you can find on your topic
- Find all the articles cited in that article

The End