

8.4b Retrieval

10x10 – Introduction to Biomedical Informatics & Artificial Intelligence
William Hersh, MD, FACMI, FAMIA, FIAHSI
Copyright 2026, All Rights Reserved



Natural language retrieval

User enters natural language words without Boolean operators

- Output usually ranked based on number of words common to query and content items (non-Web) or number of links to items (Web)
- This is implicitly an OR, although some systems (e.g., Web search engines) apply an AND

Usually used in conjunction with weighted indexing (Salton, 1991)

Natural language retrieval approach

User enters free-text query

If indexing applied stop list or stemming, must be applied to query words as well

Content items scored based on weight of words common to query and content item

- Sums $TF*IDF$ weights for all words that occur in both query and content item
- Content items may be “normalized” to account for length
- Some more complex methods achieve better performance, e.g., BM25 (Robertson, 1994)

List sorted and rank list presented to user

This approach allows other features

Relevance feedback

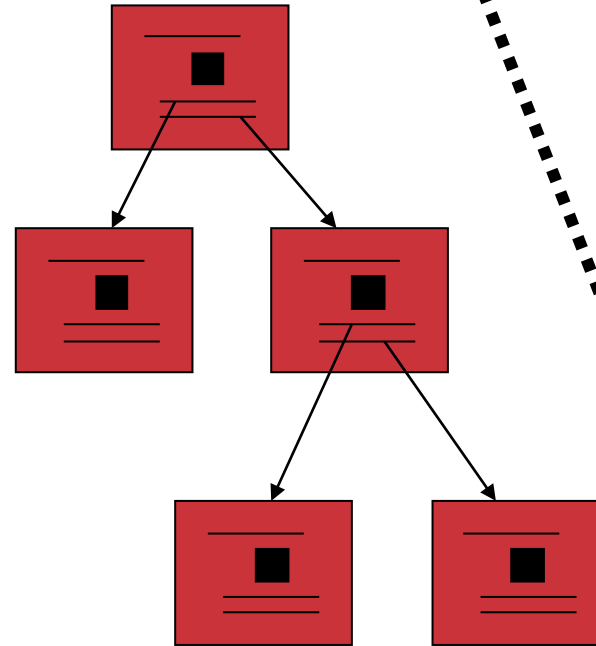
- Allows system to “find me more documents like these ones”
- After user designates relevant content items (documents), query modified
 - New words from relevant content items added
 - Query words not in relevant content items down-weighted
- Used in PubMed Related Articles feature (Fiorini, 2018)

Query expansion

- Relevance feedback without designation of relevant content items, i.e., top-ranking content items assumed to be relevant

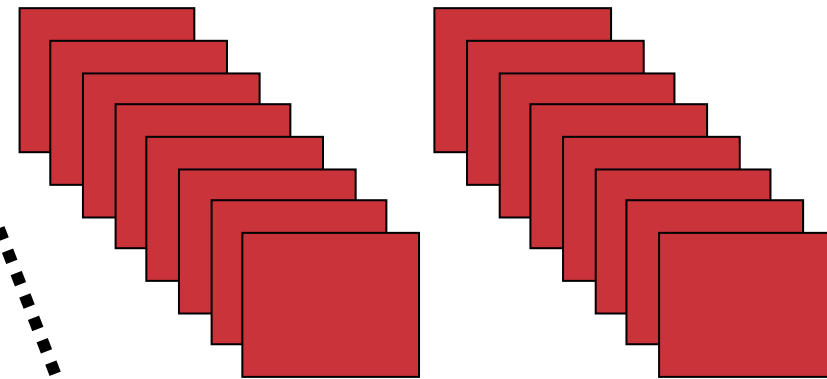
Web searching

The visible Web



Searching the Web, e.g., Google, Bing, Health Finder, etc.

The invisible or deep Web



Searching on the Web, e.g., bibliographic databases, textbooks, etc.

Searching the Web

Web search engines tend to use natural language search, although most allow some Boolean operators, e.g., Google,

- + before word indicates word must occur (AND), e.g., +congestive
- - before word indicates word must not occur (NOT), e.g., -congestive

Most Web search engines use implicit AND between search terms

Google has other features

[Ads](#) – matching search terms to advertising but clearly demarcating from regular search results

[Image](#) – images on pages retrieved by query

[Scholar](#) – searching of scientific papers

[Maps](#) – maps

[Earth](#) – satellite photos

[News](#) – latest news

Why does Google work so well?

Page Rank algorithm ranks pages based on number of links to them (Brin, 1998)

- Even though it has had to be “schooled” over the years (Lohr, 2011)

Default AND between search terms also helps due to large size of Web

This approach works well for Web pages but not necessarily for other types of content

Google has many other features, including [API for programmers](#)

Another feature of Google Scholar allows researchers to create profiles

William Hersh, MD, FACP, FACMI, FAMIA

[Oregon Health & Science University](#)
Verified email at ohsu.edu - [Homepage](#)
[Biomedical and health infor...](#) [information retrieval](#)

TITLE	CITED BY	YEAR
OHSUMED: An interactive retrieval evaluation and new large test collection for research W Hersh, C Buckley, TJ Leone, D Hickam SIGIR'94: Proceedings of the Seventeenth Annual International ACM-SIGIR ...	1196	1994
A survey of current work in biomedical text mining AM Cohen, WR Hersh Briefings in bioinformatics 6 (1), 57-71	1100	2005
Desiderata for controlled medical vocabularies in the twenty-first century JJ Cimino Methods of information in medicine 37 (04/05), 394-403	979	1998
Caveats for the use of operational electronic health record data in comparative effectiveness research WR Hersh, MG Weiner, PJ Embi, JR Logan, PRO Payne, EV Bernstam, ... Medical care 51, S30-S37	702	2013
Information retrieval: a health and biomedical perspective WR Hersh Springer New York	675	2003

Cited by

[VIEW ALL](#)

	All	Since 2021
Citations	27524	5833
h-index	86	37
i10-index	274	115

Year	Citations
2019	1000
2020	1100
2021	1150
2022	1050
2023	1050
2024	1200
2025	1100
2026	100

Public access

[VIEW ALL](#)

5 articles

63 articles

not available

available

Based on funding mandates

8.4b - Hersh - AMIA

9

Will Google ever be dethroned? Bing?

The image shows a side-by-side comparison of the Microsoft Bing search interface. On the left is a smaller, mobile-optimized version of the search page, and on the right is the full desktop version. Both pages show the search results for the query "ace inhibitors in congestive heart failure".

Desktop Version (Right):

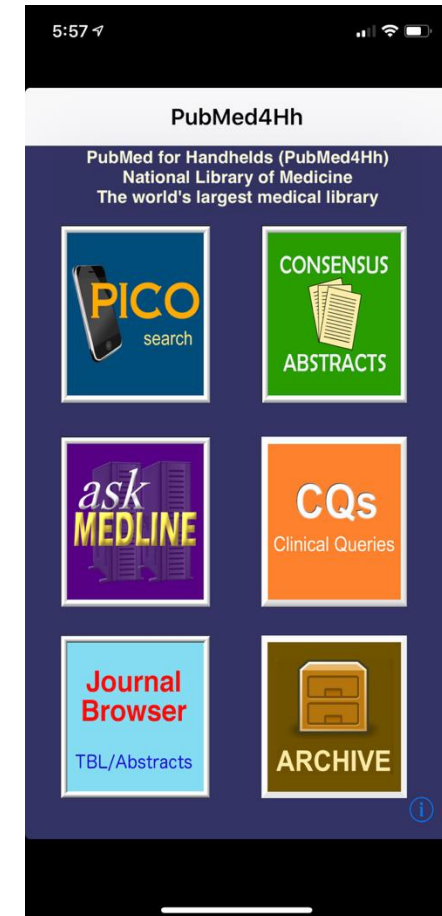
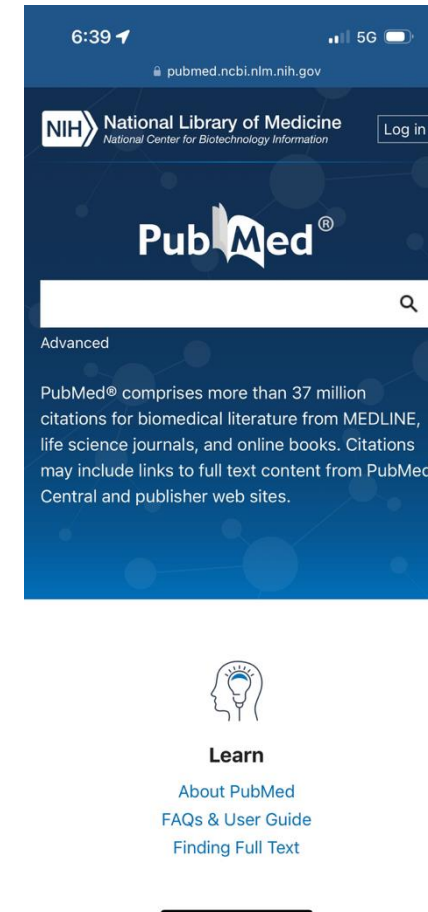
- Search Bar:** "ace inhibitors in congestive heart failure"
- Navigation:** SEARCH, COPILOT, SCHOOL, SHOPPING, IMAGES, VIDEOS, MAPS, NEWS, MORE
- Results:**
 - Oregon Health & Science University:** "Who can see school results?"
 - Heart Failure Treatment Info | Official Patient Site:** "Ad Visit The Official Site And See If You Qualify To Save On A Once-Daily Treatment Option. Learn More About A Heart Failure Treatment Option. Get More Info Here." Types: FDA Approved, Dosing, Safety Info, Doctor Discussion Guide, Savings Program.
 - Multiple Medications:** "Meet The Medisafe App And Get Help Managing Multiple Medications."
 - Your Treatment Plan:** "What To Look Out For As You Begin Your Treatment."
 - Doctor Discussion Guide:** "Discover Helpful Questions To Bring To Your Next Appointment."
 - Your Care Team:** "Meet Your Heart Failure Healthcare Team."
 - See more at heartfailure-information.com**
 - ACE inhibitors are commonly used in the treatment of congestive heart failure^{1, 2}. They have been shown to:**
 - Reduce the need for hospitalization
 - Improve survival
 - Lower the risk of heart attacks
 - Improve symptoms such as dyspnea (shortness of breath) and fatigue
 - Help the heart pump better and manage blood pressure².
 - Learn more:**
 - 1 How ACE Inhibitors Treat Heart Failure - Verywell... (verywellhealth.com)
 - 2 Medications to Help Treat Heart Failure - CL... (my.clevelandclinic.org)
 - Is there a summary? Can you show me something related? Do you know mc >**
 - Type a message...**
 - Feedback**
- Deep search:** Showing results for:
 - Heart Failure Treatment
 - ACE Inhibitors Side Effects
 - ACE Inhibitors Alternatives
- People also ask:**
 - What is the role of ACE inhibitors in the treatment of heart failure?
 - Will corticosteroids interfere with medications for congestive heart failure?
 - What is failure inhibit

Retrieval on smartphones and other mobile devices

Very popular in clinical settings, with many applications, both proprietary and free

Available in app stores

But small form factor may not be amenable to more complex searching and viewing of large documents, images, etc.



Infobuttons – direct linkage of patient-based information to knowledge

Contexts in EHR or PHR (e.g., specific diagnoses, test results, etc.) lead to generic queries that can be passed to on-line resources

The wide variety of content accessible from the Web facilitates this linkage

Leading researcher in this area has been Cimino (1996), who has developed Infobutton Manager to manage context and communications between applications (Cimino, 2006)

Now an HL7 standard and a requirement for EHR certification (Del Fiol, 2012)