

What is Biomedical and Health Informatics? (6/7)

Copyright, 2011
William Hersh, MD
Professor and Chair
Department of Medical Informatics & Clinical Epidemiology
Oregon Health & Science University
Portland, OR, USA
Email: hersh@ohsu.edu
Web: www.billhersh.info
Blog: informaticsprofessor.blogspot.com



1

Applications of knowledge-based information

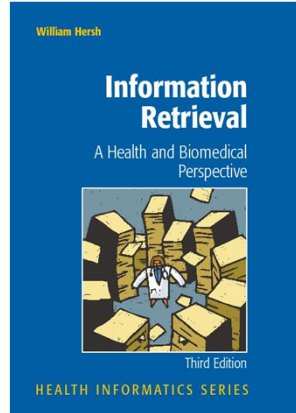
- Information retrieval (IR) systems, aka search engines, are widely available
 - Virtually all of the world's biomedical literature is now available electronically, but significant challenges still exist to its use (Hersh, 2008)
- Evidence-based medicine (EBM) is a proper framework for finding and applying knowledge in clinical care (Straus, 2005; Guyatt, 2008; Guyatt, 2008)
 - Although it must be applied properly in the larger context of healthcare (Haynes, 2002; Cohen, 2004)



2

Information retrieval – Hersh, 2009

- Focuses on indexing and retrieval of knowledge-based information
- Historically centered on text in documents, but increasingly associated with multimedia and even patient-specific information
- www.irbook.info



3

Search systems are popular icons

The image contains two overlapping screenshots of web browsers. The top-left screenshot shows the PubMed website in a Mozilla Firefox browser window. The URL is 'http://www.ncbi.nlm.nih.gov/pubmed/'. The page has a search bar and various navigation links. The bottom-right screenshot shows the Google homepage in a Mozilla Firefox browser window. The URL is 'http://www.google.com/'. The page features the Google logo, a search bar, and the text 'I'm Feeling Lucky'.

pubmed.gov

www.google.com

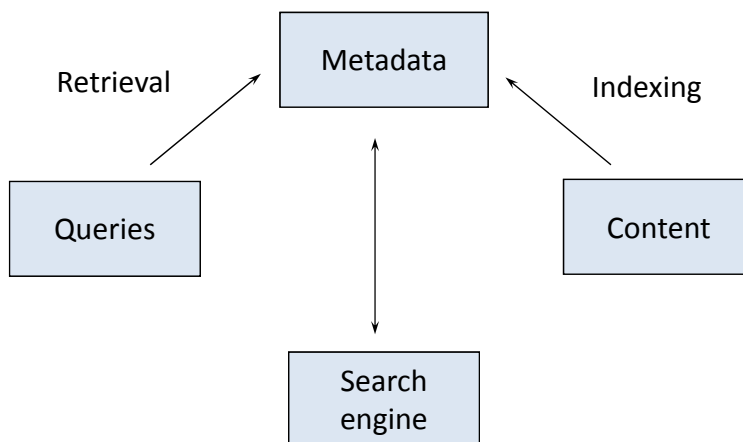
But new problems have emerged

JIM BREMAN



5

Overview of an IR system



6

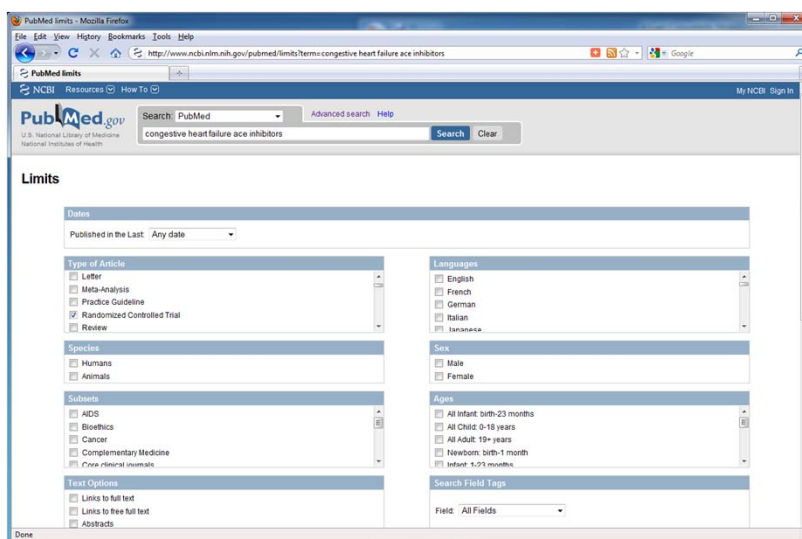
The intellectual tasks of IR

- Indexing
 - Assigning metadata to content items
 - Can assign
 - Subjects (terms) – words, phrases from controlled vocabulary, e.g., Medical Subject Headings (MeSH)
 - Attributes – e.g., author, source, publication type
- Retrieval
 - Most common approaches are
 - Boolean – use of AND, OR, NOT
 - Natural language – words common to query and content, usually ranked by “relevance,” e.g., word counts, links (Google), etc.

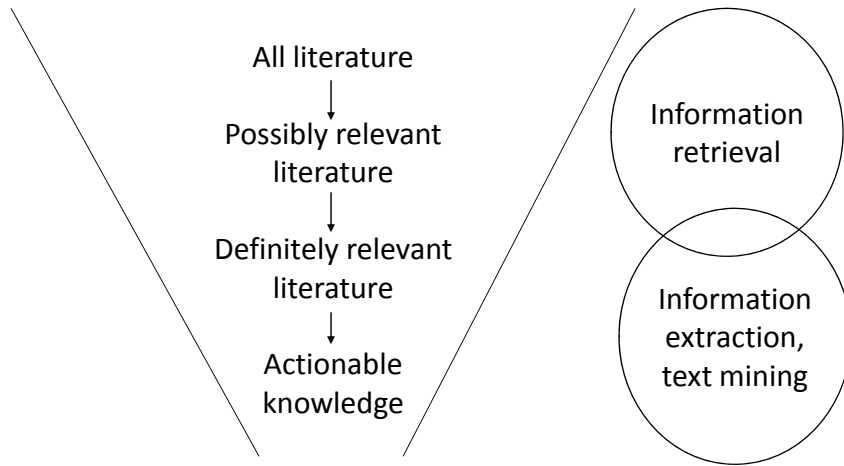


7

Some IR systems have advanced features, e.g., Pubmed limits, clinical queries, etc.



Growing role of IR and related areas in knowledge discovery

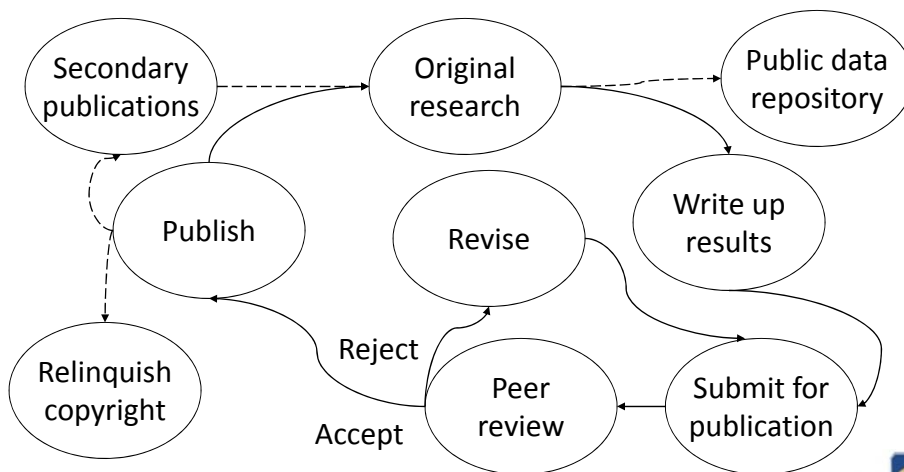


(Hersh, 2009)

9



The life-cycle of knowledge-based information – many roles for informatics



10



Some major IR challenges

- “Open access” publishing, especially to taxpayer-funded research (Björk, 2009)
 - NIH Public Access Policy aims to disseminate taxpayer-funded research better (publicaccess.nih.gov)
 - Includes deposit of papers or final submitted manuscripts into PubMed Central (pubmedcentral.gov)
 - Publishers’ alternative approach is DC Principles for Free Access to Science (www.dcpinciples.org)
- Linkage across “silos” and other applications, e.g.,
 - Digital Object Identifier (DOI) – www.doi.org
 - CrossRef project – www.crossref.org

11



IR challenges (cont.)

- What are best approaches for users to enter queries and for systems to provide output, e.g.,
 - Google works well for Web searches, but is it best for biomedicine and health?
 - What about the needs of biomedical researchers, who collect increasing amounts of data and whose research touches on many other areas?
 - What are best approaches for consumers with different levels of education, medical knowledge, etc. (Fox, 2011)?

12



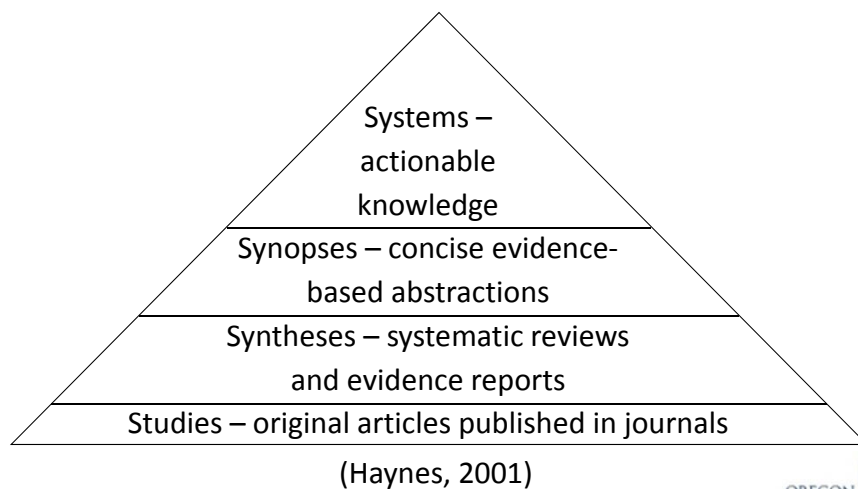
Applying EBM

- Steps of EBM
 - Phrasing a clinical question that is pertinent and answerable
 - Identifying evidence to address the question
 - Critically appraising the evidence to determine if it applies to the patient
- There are some of the criticisms of the EBM approach (Cohen, 2004)
- This has led to growing advocacy for
 - Evidence-based practice (Slawson, 2005; Dawes, 2005)
 - Practical clinical trials (Tunis, 2003; Luce, 2009)
 - Comparative effectiveness research (McGinnis, 2007; Murray, 2010)

13



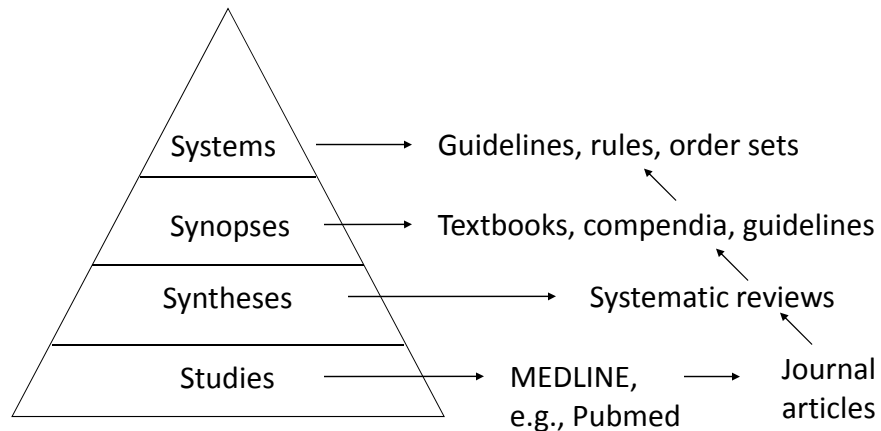
Hierarchy of knowledge-based information (evidence)



14



Where does the best evidence come from?



15

Bioinformatics

- Area of informatics focused on cellular and molecular level
 - Usually associated with genomics, proteomics, metabolomics, and other -omics
 - Genome is the collection of all genes in an organism
- Translational bioinformatics focuses on application in human health and disease (Sarkar, 2011), presaging “personalized medicine” (Collins, 2010)
- New areas emerging due to changing nature of bioscience (Gibson, 2009; www.genome.gov)
 - Gene expression microarrays (Bier, 2008)
 - Genome-wide association studies (GWAS) (Hardy, 2009)

16

Informatics challenges become more complex as science advances

- It used to be so simple – the central dogma
 - DNA → mRNA → protein
 - Genes in exons; rest is “non-coding” (used to be called “junk,” but is found to be increasingly important)
 - Growing understanding of role of microRNAs (Hanrahan, 2011)
- Increased recognition that genomic variation plays important roles in disease, consisting mainly of
 - Single-nucleotide polymorphisms (SNPs) (Lupski, 2007)
 - Copy-number variations (CNVs) (Redon, 2006; Chen, 2011)

17



Informatics challenges (cont.)

- Key focus now is genetic variation and its impact on health and disease (Hardy, 2009)
 - Constant discovery of new genetic variants associated with development of common and uncommon diseases
 - But need to be careful of spurious associations, e.g., the “incidentalome” (Kohane, 2006), and apply evidence-based approaches (Attia, 2009 – three papers)
- “High-throughput” technologies generate much data ... and information needs, e.g., need for information retrieval and text mining (Cohen, 2005; Krallinger, 2008)

18



Increased understanding of molecular basis of disease impacts bioinformatics

- And vice versa
- Case in point: cancer
 - “Hallmarks” of cancer point to cellular activities that reduce cell death while increasing invasiveness (Hanahan, 2011; Johnson, 2011)
 - Growing evidence that these changes enabled by changes in genome and its function (Cowin, 2010)
 - These changes are manifested temporally (Durinck, 2011) and more readily measured with newer sequencing technologies (Meyerson, 2010)
 - New resources, such as the Cancer Genome Atlas (Chin, 2008) and PharmGKB (Thorn, 2010), will guide research, diagnosis, and treatment

19



Melding of bioinformatics and clinical informatics: eMerge Network

- Consortium aiming to link growing number of biorepositories with sequenced DNA to data in EHR systems for “large-scale, high-throughput genetic research” (McCarty, 2011; Wilke, 2011)
 - Map the phenotype (EHR and tissue) to the genotype (DNA sequences) using informatics
- Some early work includes use of NLP for
 - Replicating finding of known gene-disease associations from research data in EHR data (Denny, 2010)
 - Discovering new associations (Denny, 2010)
 - Mapping to the phenotype, including controlled terminologies (Pathak, 2011)

20

