Domain Specific IR

Lecture 2 of 5: Medical Information Search

Allan Hanbury

hanbury@ifs.tuwien.ac.at

Russian Summer School on Information Retrieval

August 6-11, 2012 Yaroslavl, Russian Federation

Health and Medical Information

Patient-specific information


Narrative reports


Structured data


Images


Radiology images

EHR/EMR

PACS


-omics information

HealthVault

Personal Health Records (PHR)

Knowledge-based information


Journals


Books


Practice guidelines

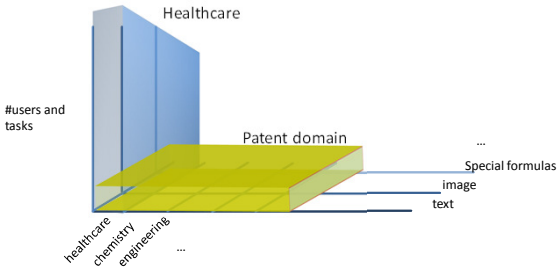

Taxonomies, vocabularies, ontologies, ...


Language resources


Websites, Web 2.0

- Primary – original research (in journals, books, reports, etc.)
- Secondary – summaries of research (in review articles, books, practice guidelines, etc.)

Patent and healthcare domains



Contents

- Introduction
- Medical Domain:
  - End users and tasks
  - Documents to be indexed
  - Search process refinements
- Future Challenges

Potential End-Users of Health Information

- Physicians
- Specialists
- Nurses
- Medical Students
- Biomedical researchers
- Lay-people (general public)
- ...

Physician Information Needs

- Unrecognized Needs
- Recognized Needs
- Pursued Needs
- Satisfied Needs

Unrecognized Needs

- Lack of awareness of the need
- Don't know that new information is available

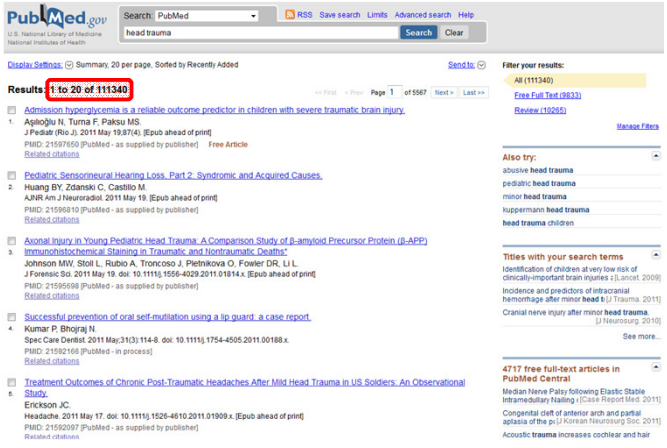
Recognized Needs

- Physicians recognise that they have an unmet information need
- Numbers from various studies:
  - Average of 2 unmet needs for every 3 patients (0.66 per patient) [CU85]
  - 1.4 questions per patient [OF91]
- Questions of type:
  - What is the cause of symptom X?
  - What is the dose of drug X?
  - How should I manage disease or finding X?
  - 69 in total [EO99]

Pursued Needs

- Physicians decided against pursuing answers for a majority of the unmet needs (from many studies)
- Most important reasons for not pursuing an answer [EO05]
  - Doubted existence of relevant information – 25%
  - Readily available consultation leading to referral rather than pursuit – 22%
  - Lack of time to pursue – 19%
  - Not important enough to pursue answer – 15%
  - Uncertain where to look for answer – 8%

The PubMed problem

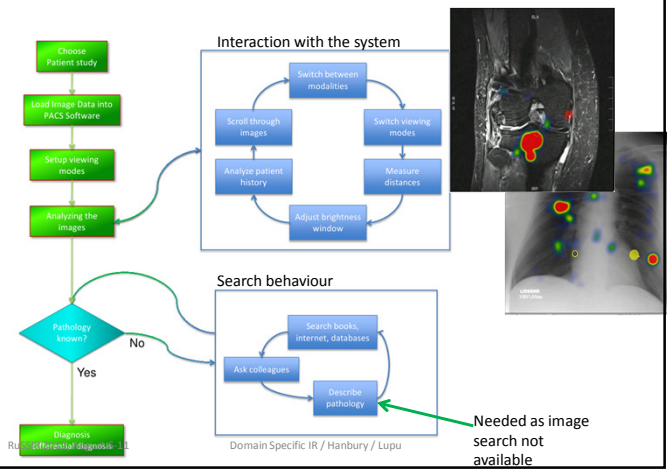


- Difficulties identified:
  - Time:
    - Physicians search on average for less than 5 minutes, and seldom search for more than 10 minutes [HSV08].
    - The time taken to answer questions using MEDLINE averages 30 minutes [HH98], and the information found is often scattered over multiple articles, making PubMed searching MEDLINE impractical for intensive clinical use [HSV08]
  - Query language:
    - Physicians tend to make simple queries, containing 2 to 3 terms on average [HSV08b], resulting in long lists of results (Boolean model of PubMed)
  - Language:
    - Dutch-speaking physicians observed in the study [HSV08b] may have used erroneous English terms, resulting in poorer returned results

Satisfied Needs

- The finding of relevant information could be improving as Internet affinity become more widespread
- Investigation of implicit search, starting automatically from an EHR
- Potential increase of mobile search

Common Radiology workflow



Other Groups

- Have different
  - Needs
  - Search behaviours
  - …

Eye-tracking

- http://youtu.be/YWo1Cx3jdOo

Consumer Health Searchers

- Non-professionals can access large amount of health information on the Internet
- 61% of American Adults seek out health advice online
- Around a third of those surveyed admitted that they changed their thinking about how they should treat a condition based on what they found online (Pew Internet and American Life Project, June 2009)

Patients searching...

- The Internet is changing the doctor-patient relationship
- Want empowered patients but no Cyberchondria
  - But can they access information of high quality?

Contents

- Introduction
- Medical Domain:
  - End users and tasks
  - Documents to be indexed
  - Search process refinements
- Future Challenges

Classification of Health Information

- Patient-specific information
  - Structured – laboratory results, vital signs
  - Narrative – history and physical, progress note, radiology report
- Knowledge-based information
  - Primary – original research (in journals, books, reports, etc.)
  - Secondary – summaries of research (in review articles, books, practice guidelines, etc.)

From Hersh

Classification of Health and Biomedical Information Content

1. Bibliographic
   - Literature reference databases
     - MEDLINE
     - Scopus
     - ...
   - Web catalogues and feeds
     - List of medical resources on the internet
     - E.g. http://www.tripdatabase.com
   - Specialized registries
     - E.g. Catalogue of U.S. Government Publications, National Guidelines Clearinghouse

From Hersh

2. Full text
   - Periodicals/Journals
   - Books and reports
   - Web collections
     - US Government: NIH, NCI
     - Clinical practice guidelines
     - Wikis
   - Evidence-based medicine (EBM) resources
     - Clinical care should be guided by the best scientific evidence

Properties of primary literature

- Growth
- Obsolescence
  - But also problem of pre-Internet literature
- Fragmentation
- Linkage and Citations
- Propagation

From Hersh

Common primary literature

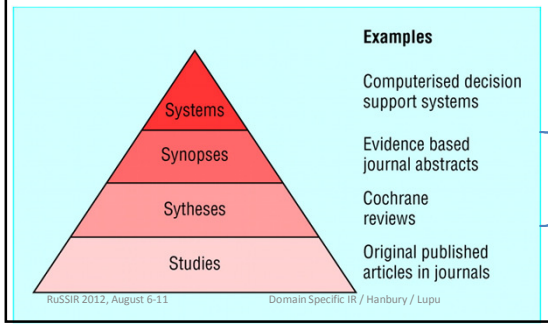
- Results of Randomized Controlled Trials
  - Potential bias as only those with positive outcomes tend to be published
  - But this is changing with clinical trial registers
- Reports of individual interesting cases

Properties of secondary literature

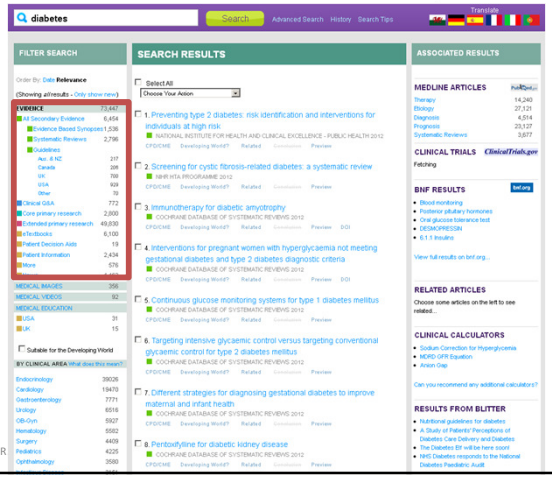
- Syntheses
- Less fragmented
- Contain older information
- Linkage and Citations
- Potentially more certain results due to combination and re-analysis of many studies
- Obsolesence

The Haynes 4S Model (EBM)

- An alternative knowledge-based information classification



TRIP Database example



3. Annotated (metadata tightly integrated)
   - Images
     - E.g. Visible Human Project
   - Videos
   - Citations
     - E.g. Science Citation Index
   - Molecular biology and –omics
     - E.g. Genomics, proteomics, ...
   - Other
     - E.g. clinicaltrials.gov, PubChem, ...

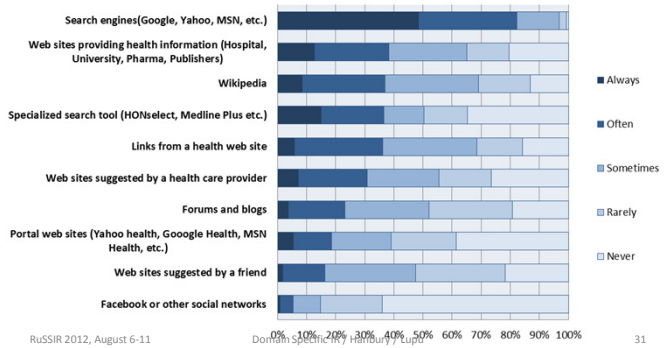
4. Aggregations
   - Consumer
     - E.g. http://www.medlineplus.gov
   - Professional
     - E.g. http://www.mdconsult.com
   - Body of knowledge
     - Has the goal of mapping all knowledge in a field
     - E.g. Health Information Management Body of Knowledge
   - Model organism databases
     - All information about an organism brought together
     - E.g. Mouse Genome Informatics, FlyBase (fruit fly), ...

Consumer Health Information

- Concerns
  - Inaccurate or out-of-date information
  - Readability
  - Trustability
  - Web 2.0 sources (forums, Wikipedia, …)
  - …

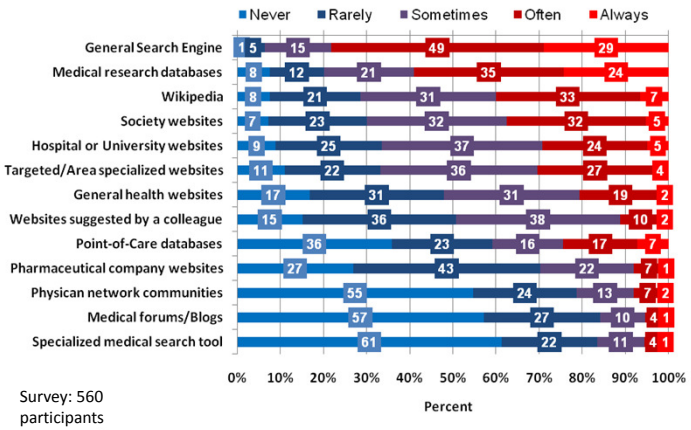
Same question for the public

How often do you use the following types of online sources to find online health information?



Where do physicians search for medical information?

Figure 4.1 Frequency of use of online resources ?*



Survey: 560 participants

*Question asked: *"How often do you use the following types of online resources to find online medical information?"**

A Google Example

Search Engines

- About 70% of the top websites with information on oral cancers gathered by Google and Yahoo searches had serious deficiencies [LC09]
  - web sites failed to attribute authorship, cite sources and report conflicts of interest.
- On the first page of results, "lawyers were the most common sponsors of websites retrieved by the terms cerebral palsy (52%), birth trauma (48%), and shoulder dystocia (43%)" [KCB08]

Wikipedia

- Wikipedia articles appear in the top 10 results for more than 70% of medical queries in four different search engines tested in [LV09]
- Whereas Wikipedia medical articles have been found to be accurate, they are also often incomplete.
  - E.g. a study on drug information comparing Wikipedia to the Medscape Drug Reference [CPK08] found that "no factual errors were found in Wikipedia, whereas 4 answers in Medscape conflicted with the answer key." However, "Wikipedia was able to answer significantly fewer drug information questions (40.0%) compared with MDR (82.5%)."
  - An advantage of Wikipedia was that "there was a marked improvement in Wikipedia over time, as current entries were superior to those 90 days prior."

Codes of Conduct

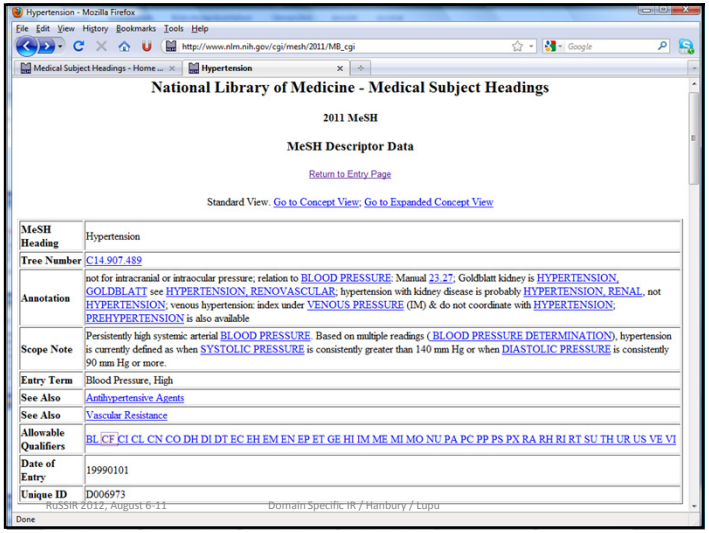
- Various criteria for the quality of health web pages have been put forward.
- E.g. Health on the Net is an NGO that certifies health web pages satisfying the HONcode Principles
  - http://www.healthonnet.org
- Semi-automatic certification
- Have a search engine that searches certified pages

HONcode principles

1. Authoritative
   - Indicate the qualifications of the authors
2. Complementarity
   - Information should support, not replace, the doctor-patient relationship
3. Privacy
   - Respect the privacy and confidentiality of personal data submitted to the site by the visitor
4. Attribution
   - Cite the source(s) of published information, date and medical and health pages
5. Justifiability
   - Site must back up claims relating to benefits and performance
6. Transparency
   - Accessible presentation, accurate email contact
7. Financial disclosure
   - Identify funding sources
8. Advertising policy
   - Clearly distinguish advertising from editorial content

Contents

- Introduction
- Medical Domain:
  - End users and tasks
  - Documents to be indexed
  - Search process refinements
- Future Challenges


Hypertension - Mozilla Firefox

National Library of Medicine - Medical Subject Headings

2011 MeSH

MeSH Descriptor Data

Return to Entry Page

Standard View. Go to Concept View; Go to Expanded Concept View

MeSH Heading	Hypertension
Tree Number	C14.907.489
Annotation	not for intracranial or intraocular pressure; relation to BLOOD PRESSURE: Manual 23.27; Goldblatt kidney is HYPERTENSION, GOLDBLATT see HYPERTENSION, RENOVASCULAR; hypertension with kidney disease is probably HYPERTENSION, RENAL, not HYPERTENSION; venous hypertension: index under VENOUS PRESSURE (IM) & do not coordinate with HYPERTENSION; PREHYPERTENSION is also available
Scope Note	Persistently high systemic arterial BLOOD PRESSURE. Based on multiple readings (BLOOD PRESSURE DETERMINATION), hypertension is currently defined as when SYSTOLIC PRESSURE is consistently greater than 140 mm Hg or when DIASTOLIC PRESSURE is consistently 90 mm Hg or more.
Entry Term	Blood Pressure, High
See Also	Antihypertensive Agents
See Also	Vascular Resistance
Allowable Qualifiers	BL CF CI CL CN CO DH DI DT EC EH EM EN EP ET GE HI IM ME MI MO NU PA PC PP PS PX RA RH RI RT SU TH UR US VE VI
Date of Entry	19990101
Unique ID	D006973

RuSSIR 2012, August 6-11 Domain Specific IR / Hanbury / Lupu

Medical Vocabularies (1)

- Many such vocabularies available
- Medical Subject Headings (MeSH)
  - Produced by the NLM
  - Used to manually index MEDLINE entries
  - Contains 23.000 headings (concepts)
  - Contains the following relationships:
    - Hierarchical: organised into 16 trees
    - Synonymous: entry terms are synonyms of headings (e.g. plurality, word order, hyphenation)
    - Related: terms that may be useful for searchers to add to their searches

Medical Vocabularies (2)

- SNOMED CT: patient-specific information
  - 283,000 Active concept codes

CONCEPTID	CONCEPTSTATUS	FULLYSPECIFIEDNAME	CTV3ID	SNOMEDID
210566005	0	Open wound of hand with tendon involvement (disorder)	S922.	DD-3317D
210567001	0	Complete division extensor tendon hand (disorder)	S9220	DF-008E6
210568006	0	Complete division flexor tendon hand (disorder)	S9221	DF-008E7
210569003	0	Partial division extensor tendon hand (disorder)	S9222	DF-008E8
210570002	0	Partial division flexor tendon hand (disorder)	S9223	DF-008E9
210571003	0	Degloving injury of hand (disorder)	S923.	DD-30125
210572005	0	Degloving injury hand, palmar (disorder)	S9230	DD-30126
210573000	0	Degloving injury hand, dorsum (disorder)	S9231	DD-30127
210574006	0	Severe multi tissue damage hand (disorder)	S924.	DD-00414
210575007	0	Massive multi tissue damage hand (disorder)	S925.	DD-00415
210576008	6	Open wound of hand, excluding fingers, NOS (disorder)	S92z.	DD-33163
210577004	4	Open wound: [finger(s) or of thumb] or [fingernail] or [nail] or [thumbnail]	S93..	R-F5944
210578009	6	Open wound of finger or thumb without mention of complication (disorder)	S930.	DD-3317E
125653000	0	Open wound of finger (disorder)	S9300	DD-33169
210579001	0	Open wound, finger, multiple (disorder)	S9301	DD-3317F
125654006	0	Open wound of thumb (disorder)	S9302	DD-3316A
210580003	0	Open wound of finger or thumb with complication (disorder)	S931.	DD-33189
210581004	0	Open wound: [finger or thumb with tendon involvement] or [finger with tendon injury]	S932.	R-F5945

RuSSIR 2012, August 6-11 Domain Specific IR / Hanbury / Lupu

Medical Vocabularies (3)

- WHO International Classification of Diseases (ICD-10): codes diagnoses for statistics, epidemiology and billing
  - Available in 42 languages
- Current Procedural Terminology (CPT): codes procedures

Disorders of lens
(H25-H28)

H25 Senile cataract
Excl.: capsular glaucoma with pseudoexfoliation of lens (H40.1)

H25.0 Senile incipient cataract
Senile cataract:
- coronary
- cortical
- punctate

Subcapsular polar senile cataract (anterior)(posterior)
Water clefts

H25.1 Senile nuclear cataract
Cataracta brunescens
Nuclear sclerosis cataract

H25.2 Senile cataract, morgagnian type
Senile hypermature cataract

H25.8 Other senile cataract
Combined forms of senile cataract

H25.9 Senile cataract, unspecified

H26 Other cataract
Excl.: congenital cataract (Q12.0)

H26.0 Infantile, juvenile and presenile cataract

H26.1 Traumatic cataract
Use additional external cause code (Chapter XX), if desired, to identify cause.

H26.2 Complicated cataract
Cataract in chronic iridocyclitis
Cataract secondary to ocular disorders
Glaucomatous flecks (subcapsular)

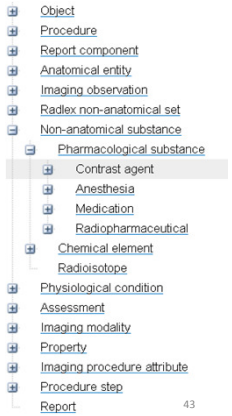
H26.3 Drug-induced cataract
Use additional external cause code (Chapter XX), if desired, to identify drug.

H26.4 After-cataract
Secondary cataract
Soemmering's ring

H26.8 Other specified cataract

Medical Vocabularies (4)

- RadLex (Radiology Lexicon)
  - Single unified source of Radiology terms
  - Links to SNOMED CT and DICOM
  - 34 446 active classes



ICD-10

Exposure to inanimate mechanical forces
(W20-W49)

See at the beginning of this chapter for the classification of the place of occurrence and activity

Excl.: assault (X85-Y09)
contact or collision with animals or persons (W50-W64)
intentional self-harm (X60-X84)

W20 Struck by thrown, projected or falling object
Incl.: cave-in without asphyxiation or suffocation
collapse of building, except on fire
falling:
- rock
- stone
- tree

Excl.: collapse of burning building (X00)
falling object in:
- cataclysm (X34-X39)
- machinery accident (W24, W28-W31)
- transport accident (V01-V99)

object set in motion by:
- explosion (W35-W40)
- firearm (W32-W34)

sports equipment (W21)

W21 Striking against or struck by sports equipment
Incl.: struck by:
- hit or thrown ball
- hockey stick or puck

W22 Striking against or struck by other objects
Incl.: walked into wall

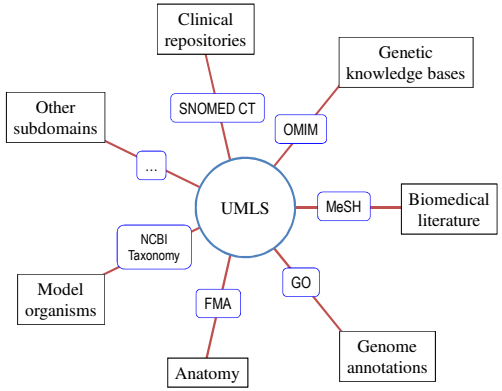
W23 Caught, crushed, jammed or pinched in or between objects
Incl.: caught, crushed, jammed or pinched:
- between:
  - moving objects
  - stationary and moving objects

  such as: folding object; sliding door and door-frame; packing crate and floor, after losing grip; washing-machine wringer
- in object

Excl.: injury caused by:
- cutting or piercing instruments (W25-W27)

Medical Vocabularies (5)

- Unified Medical Language System (UMLS)
  - Goal of providing a mechanism for linking diverse medical vocabularies
  - Metathesaurus component links more than 100 source vocabularies
  - Multilingual (non-English translations are synonyms of English translations)



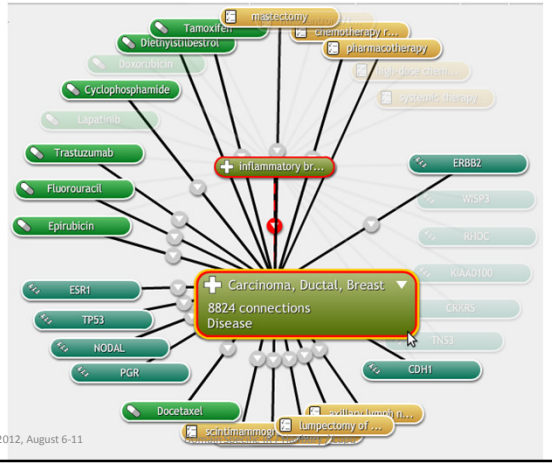
From: http://www.nlm.nih.gov/research/umls/presentations/AMIA%202007%20T12%20final.ppt

- Classification constraint
  - Know from the labels and ontology information if a classification of organs in an image is possible
- Multilingual search
  - Map terms in many languages into the vocabulary
- Search term suggestion or disambiguation
- Example of previous two: http://www.wrapin.org

Use of Vocabularies (Domain Knowledge)

- Query Expansion
  - PubMed is an NLM search engine to search MEDLINE: http://www.pubmed.gov
  - Boolean search
  - Uses MeSH terms to expand queries
- Document annotation
  - Find occurrences of words in documents and link them to the vocabulary
  - Exopatent: http://fda.semanticannotation.com

Coreminer.com



Contents

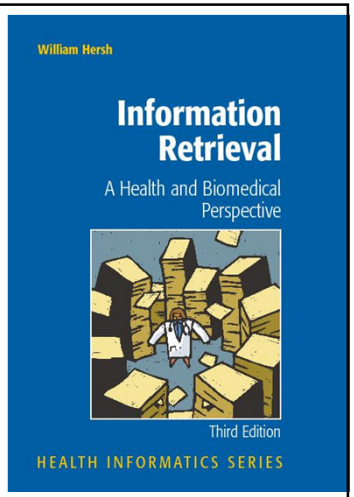
- Introduction
- Medical Domain:
  - End users and tasks
  - Documents to be indexed
  - Search process refinements
- Future Challenges

Challenges

- Exploding amount of information
- Ever increasing rate of generating new information
- Multilingual information
- Huge amounts of information stored unused in hospital archives
- Future information sources, e.g. Genome information
- ...

Reference

- William Hersh, M.D., Information Retrieval: A Health and Biomedical Perspective, Third Edition, Springer, 2009



References

[CPK08] K. A Clauson, H. H. Polen, M. N. Kamel Boulos, J. H. Dzenowagis, Scope, Completeness, and Accuracy of Drug Information in Wikipedia, The Annals of Pharmacotherapy, Volume 42, No. 12, pages 1814-1821, 2008

[CU85] D. Covell, G. Uman, et al, Information needs in office practice: are they being met? Annals of Internal Medicine, 103:596-599, 1985

[EO99] J. Ely, J. Osheroff, et al., Analysis of questions asked by family doctors regarding patient care, British Medical Journal, 319(7206):358-61, 1999

[EO05] J. Ely, J. Osheroff, Answering Physicians' Clinical Questions: Obstacles and Potential Solutions, J Am Med Inform Assoc., 12(2): 217–224, 2005.

[HH98] W. R. Hersh, D. H. Hickam, How Well Do Physicians Use Electronic Information Retrieval Systems? A Framework for Investigation and Systematic Review, Journal of the American Medical Association, 280:15, 1998

[HSV08] A Hoogendam, A. F. H. Stalenhoef, P. F de Vries Robbé, A. J. P. M. Overbeke, Answers to Questions Posed During Daily Patient Care Are More Likely to Be Answered by UpToDate Than PubMed, J Med Internet Res, Volume 10, Number 4, 2008.

[HSV08b] A. Hoogendam, A. F. H. Stalenhoef, P. F. de Vries Robbé, A. J. P. M. Overbeke, Analysis of queries sent to PubMed at the point of care: Observation of search behaviour in a medical teaching hospital, BMC Medical Informatics and Decision Making 2008, Volume 8, Number 42, 2008

[KCB08] A. J. Kamal, Y. W. Cheng, A. S. Bryant, M. E. Norton, B. L. Shaffer, A. B. Caughey, Google obstetrics: who is educating our patients?, American Journal of Obstetrics & Gynecology, Volume 198, Number 6, June 2008.

[LC09] P. López-Jornet, F. Camacho-Alonso, The quality of Internet sites providing information relating to oral cancer, Oral Oncology, 2009.

[LV09] M. R. Laurenta, T. J. Vickers, Seeking Health Information Online: Does Wikipedia Matter?, Journal of the American Medical Informatics Association, Volume 16, pages 471-479, 2009

[OF91] J. Osheroff, D. Forsythe, et al., Physicians' information needs: analysis of questions posed during clinical teaching, Annals of Internal Medicine, 114:576-581, 1991