Medical Information Retrieval

MOSIG2 Lecture

Lorraine Goeuriot[1]

[1] Laboratoire d'Informatique de Grenoble (LIG)
Université Grenoble Alpes
France
lorraine.goeuriot@univ-grenoble-alpes.fr

November 21st, 2023

Outline

Objectives

1. Introduce tasks, users and resources in the medical domain
2. Present state-of-the art models and techniques in medical information retrieval and off-the-shelf tools
3. Provide a list of open-source datasets
4. Summarize challenges and research opportunities

Slides, links and references are online:
`https://github.com/lorraine-goeuriot/MOSIG-lecture-Medical-IR`

Outline

What is IR?

The IR view

Information Retrieval (Mooers:1951)

"Information retrieval is the name for the process or method whereby a prospective user of information is able to convert his **need for information** into an actual list of citations to **documents** in storage containing information **useful** to him... Information retrieval is crucial to documentation and organization of knowledge"

Information Retrieval [Salton and McGill, 1986]

"Information retrieval systems are designed to help analyze and describe the **items** stored in a file, to organize them and search among them, and finally to retrieve them **in response** to a **user's query**. Designing and using a retrieval system involves four major activities: information analysis, information organization and **search**, **query formulation**, and i**nformation retrieval** and dissemination"

- Main keywords, phrases: naïve definitions
  - *Information item*: a single unit of (textual) information (document, blog, tweet, e-mail, medical visit report, etc.)
  - *Information need*: what the user seeks for
  - *Query*: explicit formulation of the user's information need
  - *Retrieve useful information*: select information items that are relevant to the query

WHAT IS IR?

THE WELL KNOWN SEARCH ENGINES

- When we talk about information retrieval, we think about web search engines...

What is IR?

The well known search engines

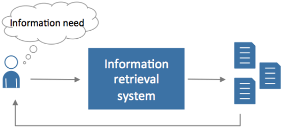
- When we talk about information retrieval, we think about web search engines...

- ...but there also a plenty of other search systems
  - Search in digital libraries
  - Search in entreprise corpora
  - Search for medical patient records
  - Search for legal texts

What is IR?

The Anatomy of an IR process

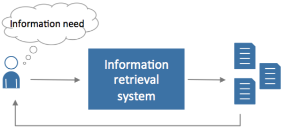
- Look at the IR process from the user side



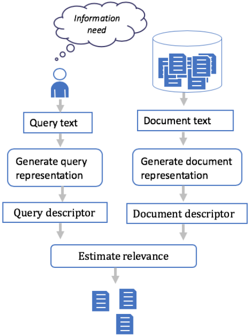
What is IR?

The Anatomy of an IR process

- Look at the IR process from the user side



- Look at the IR process from the system side



What is IR?

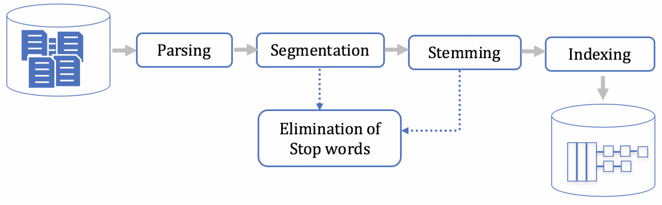
Document representation

- The document
  - Refers to a unit of information searchable by the user
  - Has a syntax and a semantics, specified by the author

What is IR?

Document representation

- The document
  - Refers to a unit of information searchable by the user
  - Has a syntax and a semantics, specified by the author
- Document preprocessing: logical view of a document from full text to a set of index terms
  - Lexical analysis of the text (parsing and segmentation): converting stream of chars into stream of words
  - Elimination of stopwords: words that appear too frequently
  - Stemming of the remaining words: reduce to stems after removing prefixes/suffixes, plural, gerund forms, ...
  - Selection of index terms or keywords: use useful words as index terms (all words, nouns, ...)
  - (Construction of term categorization structures like thesaurus, terminologies, ontologies): list of important words or concepts in a domain, related words, concepts...



What is IR?

Relevance estimation: how?

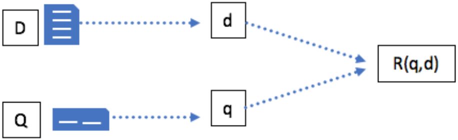
- Relevance is a core concept in IR [Borlund, 2003]
  - Relevance is subjective (user-dependent): adequacy of the document to answer the query
  - Multiple dimensions lead to multiple types of relevance: topic, novelty, understandability, reliability
    - Topical relevance: the major type of relevance addressed in IR based on the "aboutness" criteria
  - Dynamic: user's perception of relevance evolves through the search episodes

What is IR?

Relevance estimation: how?

- Relevance is a core concept in IR [Borlund, 2003]
  - Relevance is subjective (user-dependent): adequacy of the document to answer the query
  - Multiple dimensions lead to multiple types of relevance: topic, novelty, understandability, reliability
    - Topical relevance: the major type of relevance addressed in IR based on the "aboutness" criteria
  - Dynamic: user's perception of relevance evolves through the search episodes

- IR models: a framework for relevance estimation
  - Algorithmic relevance: assign scores to documents with regard to a given query based on content matching
  - Process in two main stages
    1. Design a logical (formal) framework for representing documents and queries
    2. Define a ranking function that measures the topical similarities between queries and documents and orders the documents according to this measure



What is IR?

The semantic gap issue

She takes **just like a woman**, yeah she does. She makes love **just like a woman**, yeah she does And she aches **just like a woman** But she breaks **just** like a little girl

Just like a woman

One of many Dylan songs with an unclear subject. It's often thought tobe about fellow folk-singer Joan Baez, with whom Dylan had arelationship. Edie Sedgwick, an actress affiliated with Andy Warhol, isalso thought to have inspired the song.

- Representing documents and queries:
  - Understand broad language: what's behind the surface of strings, bag of words?
  - **Semantic representation rather than string/lexical representation**
    - Disambiguation of **entities**, **concepts** and roles
    - Reasoning and inference of relations
- Relevance ranking:
  - Understand broad relevance: what's behind the surface of matching?
  - Semantic matching rather than string matching
  - Relevance matching vs. semantic matching [Guo et al., 2016]

What is IR?

IR models

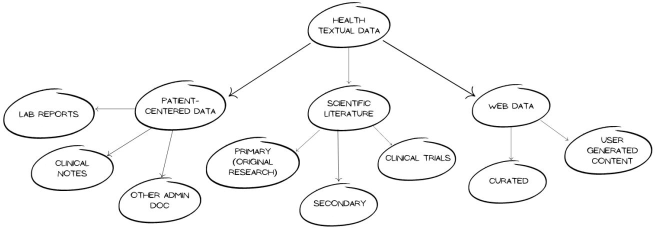
- Design retrieval models: a long standing research in IR
- Major IR models: vectorial model [Salton et al., 1975], probabilistic model [Jones et al., 2000], language model [Ponte and Croft, 1998]

Fang and Zhai [2005]: Axiomatic approach
He and Ounis [2005]: TF in BM25 and DFR
Metzler and Croft [2004]: LM and PIN's
Robertson [2005]: Event Spaces
Robertson [2004]: Understanding IDF
Sparck-Jones et al. [2003]: LM and Relevance
Croft and Lafferty [2003], Lafferty and Zhai [2003]: LM book
Zaragoza et al. [2003]: Bayesian extension to LM
Bruza and Song [2003]: probabilistic dependencies in LM
Amati and van Rijsbergen [2002]: DFR
Lavrenko and Croft [2001]: Relevance-based LM
Hiemstra [2000]: TF-IDF and LM
Sparck-Jones et al. [2000]: probabilistic model: status
Ponte and Croft [1998]: LM
Brin and Page [1998a], Kleinberg [1999]: Pagerank and Hits
Robertson et al. [1994], Singhal et al. [1996]:
Pivoted Document Length Normalization
Wong and Yao [1995]: $P(d \rightarrow q)$
Robertson and Walker [1994], Robertson et al. [1995]: 2-Poisson, BM25
Church and Gale [1995], Margulis [1992]: Poisson
Fuhr [1992b]: Probabilistic Models in IR
Turtle and Croft [1991, 1990]: PIN's
Fuhr [1989]: Models for Probabilistic Indexing
Cooper [1991, 1988, 1994]: Beyond Boole, Probability Theory in IR: An Encumbranc
Deerwester et al. [1990], Dumais et al. [1988]: Latent semantic indexing
van Rijsbergen [1986, 1989]: $P(d \rightarrow q)$
Bookstein [1980], Salton et al. [1983]: Fuzzy, extended Boolean
Croft and Harper [1979]: BIR without relevance
Robertson and Sparck-Jones [1976]: BIR
Salton [1971], Salton et al. [1975]: VSM, TF-IDF
Rocchio [1971]: Relevance feedback
Maron and Kuhns [1960]: On Relevance, Probabilistic Indexing, and IR

Outline

Medical Information

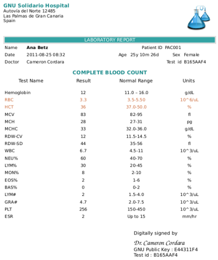
A Classification of Textual Health Information [Hersh, 2010]



- **Patient-specific information**: applies to individual patients. Tells healthcare providers, administrators and researchers about the health and disease of a patient.
  - Structured: laboratory results, vital signs
  - Narrative: history and physical, progress notes, radiology report
- **Knowledge-based information**: has been derived and organized from observational or experimental research. Usually provided in books, journals or *computerized media*.
  - Primary: original research (in journals, books, reports, etc.)
  - Secondary: summaries of research (in review articles, books, practice guidelines, etc.)
- Web data:
  - Curated websites: medical institutions, magazines, specialized websites, Wikipedia
  - User-generated-content: discussion forums, Facebook, Twitter, PatientsLikeMe

Medical Information

Structured Patient-centered Data



GNU Solidario Hospital
Autovía del Norte 12485
Las Palmas de Gran Canaria
Spain

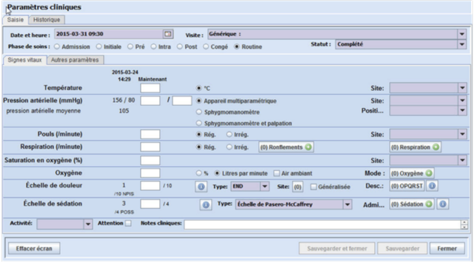
LABORATORY REPORT

Name Ana Betz — Patient ID PAC001
Date 2011-08-25 08:32 — Age 25y 10m 26d — Sex Female
Doctor Cameron Cordara — Test id B165AAF4

COMPLETE BLOOD COUNT

Test Name	Result	Normal Range	Units
Hemoglobin	12	11.0 - 16.0	g/dL
RBC	3.3	3.5-5.50	10^6/uL
HCT	36	37.0-50.0	%
MCV	83	82-95	fl
MCH	28	27-31	pg
MCHC	33	32.0-36.0	g/dL
RDW-CV	12	11.5-14.5	%
RDW-SD	44	35-56	fl
WBC	6.7	4.5-11	10^3/uL
NEU%	60	40-70	%
LYM%	30	20-45	%
MON%	8	2-10	%
EOS%	2	1-6	%
BAS%	0	0-2	%
LYM#	2	1.5-4.0	10^3/uL
GRA#	4.7	2.0-7.5	10^3/uL
PLT	256	150-450	10^3/uL
ESR	2	Up to 15	mm/hr

Digitally signed by
Dr. Cameron Cordara
GNU Public Key : E44311F4
Test id : B165AAF4



HEMATOLOGIE

NUMERATION ET FORMULE LEUCOCYTAIRE
(automate LH Beckman Coulter)

	Résultats	Normes
HEMATIES	5.320.000 /mm3	4500000 à 6500000
HEMOGLOBINE.	16,5 g/100 ml	13,0 à 17,0
HEMATOCRITE.	48,7 %	40,0 à 54,0
LEUCOCYTES	6.000 /mm3	4.000 à 10.000
PLAQUETTES	226.000 /mm3	150.000 à 500.000

VITESSE DE SEDIMENTATION
(technique Test1 BCL ALIFAX)

1ère heure	8 mm	inf. à 10

CHIMIE DU SANG

	Résultats	Normes
aspect du sérum.	normal	
GLYCEMIE . . . . (technique enzymatique à l'hexokinase)	1,08 g/l 5,99 mmol/l	0,70 à 1,05 3,88 à 5,83
UREE . . . . (technique enzymatique à l'uréase/GLDH)	0,33 g/l 5,48 mmol/l	0,10 à 0,50 1,66 à 8,30
CREATININEMIE. . . . . (technique de Jaffé compensée sans déprotéinisation)	10 mg/l 88 umol/l	7 à 12 62 à 106

EXPLORATION LIPIDIQUE

CHOLESTEROL TOTAL. . . . . . . (technique colorimétrique enzymatique CHOD-PAP)	2,86 g/l 7,38 mmol/l	inf. à 2,00 inf. à 5,16
H.D.L. . . . . (technique colorimétrique enzymatique CHOD-PAP directe)	0,53 g/l 1,37 mmol/l	0,40 à 0,80 1,03 à 2,06
TRIGLYCERIDES. . . . . (technique colorimétrique enzymatique GPO-PAP)	1,54 g/l 1,76 mmol/l	inf. à 2,00 inf. à 2,28
LDL CHOLESTEROL. . . . . (calcul selon la formule de Friedewald)	2,02 g/l 5,23 mmol/l	inf. à 1,60 inf. à 4,14
PROTEINE C-REACTIVE. . . . . . (technique immunoturbidimétrique)	inf à 3 mg/l	inf. à 5,0

Medical Information

Narrative Patient-centered Data

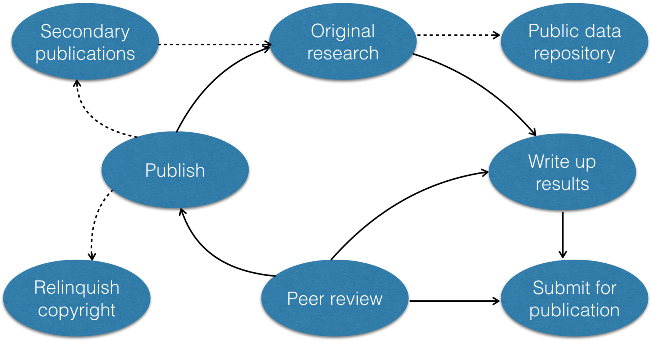
```
Admission Date: [**2015-03-17**]     Discharge Date: [**2015-03-24**]
Date of Birth: [**1974-10-03**]      Sex: F
Service: Neurosurgery
HISTORY OF PRESENT ILLNESS: The patient is a 40-year-old female with complaints
of headache and dizziness. In [**2015-01-14**], the patient had headache with
neck stiffness and was unable to walk for 45 minutes. [...]
PAST MEDICAL HISTORY: Hypothyroidism.
ALLERGIES: Penicillin and Bactrim which causes a rash.
MEDICATIONS: Levoxyl 1.75 mg.
PHYSICAL EXAMINATION: On physical examination, her blood pressure was 104/73,
pulse 79. In general, she was a woman in no acute distress. HEENT: Nonicteric.
Pupils are equal, round, and reactive to light. Extraocular movements are full.
[...]

On postoperative day #1, the patient was taken to arteriogram, where she
underwent a cerebral angiogram to evaluate clipping of the aneurysm. []
DISCHARGE MEDICATIONS:
1. Hydromorphone 2-6 mg po q4h prn.
2. Synthroid 175 mcg po q day.[...]

CONDITION ON DISCHARGE: Stable.

FOLLOW-UP INSTRUCTIONS: She will follow up in 10 days for staple removal with
Dr. [**Last Name (STitle) 570**].

(End of Report)
```

Discharge summary extracted from the MIMIC II dataset
`https://physionet.org/mimic2/`.

Medical Information

Primary Knowledge-based Documents

- Contain reports of research results: discoveries, observations, description of related work and position of the report, conclusions.
- Has never been published before
- Published in books, journals or conference proceedings
- Usually a small number of documents have the highest impact



Medical Information

Primary Knowledge-based Documents

Cyberchondria: Studies of the Escalation of Medical Concerns in Web Search

RYEN W. WHITE and ERIC HORVITZ
Microsoft Research

The World Wide Web provides an abundant source of medical information. This information can assist people who are not healthcare professionals to better understand health and illness, and to provide them with feasible explanations for symptoms. However, the Web has the potential to increase the anxieties of people who have little or no medical training, especially when Web search is employed as a diagnostic procedure. We use the term *cyberchondria* to refer to the unfounded *escalation* of concerns about common symptomatology, based on the review of search results and literature on the Web. We performed a large-scale, longitudinal, log-based study of how people search for medical information online, supported by a survey of 515 individuals' health-related search experiences. We focused on the extent to which common, likely innocuous symptoms can escalate into the review of content on serious, rare conditions that are linked to the common symptoms. Our results show that Web search engines have the potential to escalate medical concerns. We show that escalation is associated with the amount and distribution of medical content viewed by users, the presence of escalatory terminology in pages visited, and a user's predisposition to escalate versus to seek more reasonable explanations for ailments. We also demonstrate the persistence of postsession anxiety following escalations and the effect that such anxieties can have on interrupting user's activities across multiple sessions. Our findings underscore the potential costs and challenges of cyberchondria and suggest actionable design implications that hold opportunity for improving the search and navigation experience for people turning to the Web to interpret common symptoms.

Categories and Subject Descriptors: H.3.3 [**Information Storage and Retrieval**]: Information Search and Retrieval—*Search process; query formulation*

General Terms: Human Factors, Experimentation

Additional Key Words and Phrases: Cyberchondria

ACM Reference Format:

White, R. W. and Horvitz, E. 2009. Cyberchondria: Studies of the escalation of medical concerns in Web search. ACM Trans. Inf. Syst. 27, 4, Article 23 (November 2009), 37
DOI = 10.1145/1629096.1629101 http://doi.acm.org/10.1145/1629096.1629101

1. INTRODUCTION

The World Wide Web has the potential to provide valuable medical information to people, where Web sites such as WebMD (http://www.webmd.com) and MSN

Authors' addresses: R. W. White and E. Horvitz, Microsoft Research, One Microsoft Way, Redmond, WA 98052; email: {ryenw, horvitz}@microsoft.com.

DOI 10.1145/1629096.1629101 http://doi.acm.org/10.1145/1629096.1629101

ACM Transactions on Information Systems, Vol. 27, No. 4, Article 23, Publication date: November 2009.

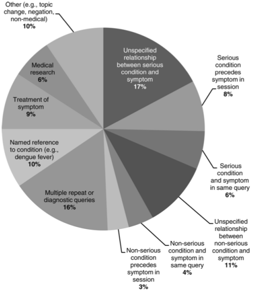


Fig. 1. Distribution of labels assigned to set of hand-labeled no-change sessions.

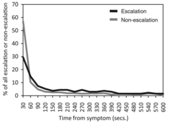


Fig. 2. Temporal distance from initial input of symptom (within session)

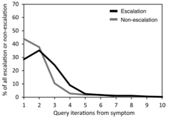


Fig. 3. Query distance from initial input of symptom (within session).

REFERENCES

Agichtein, E., Brill, E., and Dumais, S. 2006. Improving Web search ranking by incorporating user behavior information. In *Proceedings of the ACM SIGIR Conference on Research and Development in Information Retrieval*. 19–26.

American Psychiatric Association. 1994. *Diagnostic and Statistical Manual of Mental Disorders* 4th Ed. APA, Washington, DC.

Asmundson, J. G., Taylor, S., and Cox, B. J. 2001. *Health Anxiety: Clinical and Research Perspectives on Hypochondriasis and Related Conditions*. Wiley.

Ayers, S. and Kronenfeld, J. 2007. Chronic illness and health-seeking information on the Internet. *Health 11*, 3, 327–347.

Baker, L., Wagner, T. H., Singer, S., and Bundorf, M. K. 2003. Use of the Internet and e-mail for health care information. *J. Amer. Med. Assoc. 289*, 18, 2401–2406.

Barsky, A. J. and Ahern, D. K. 2004. Cognitive behavioral therapy for hypochondriasis. *J. Amer. Med. Assoc. 291*, 12, 1464–1470.

Barsky, A. J. and Klerman, G. L. 1983. Overview: Hypochondriasis, bodily complaints, and somatic styles. *Amer. J. Psychiatry 140*, 273–283.

Benigeri, M. and Pluye, P. 2003. Shortcomings of health-related information on the Internet. *Health Promot. Int. 18*, 4, 381–387.

Berland, G. K., Elliott, M. N., Morales, L. S., Algazy, J. I., Kravitz, R. L., Broder, M. S., Kanouse, D. E., Muñoz, J. A., Puyol, J.-A., Marielena, L., Watkins, K. E., Yang, H., and McGlynn, E. A.

Medical Information

Primary Knowledge-based Documents

Clinical trials

Clinical trials are research studies that test a medical, surgical, or behavioral intervention in people. These trials are the primary way that researchers determine if a new form of treatment or prevention, such as a new drug, diet, or medical device (for example, a pacemaker), is safe and effective in people.
Source: `https://www.nia.nih.gov/health/what-are-clinical-trials-and-studies`

Study Overview

Brief Summary:
The purpose of this study is to evaluate whether erythropoietin can help limit the damage to the heart in patients with acute heart attacks.

Detailed Description:
REVEAL is a randomized, double-blinded, placebo-controlled, parallel phase II clinical study that will evaluate the effects of erythropoietin administration on infarct size, left ventricular remodeling and circulating endothelial progenitor cells in patients with large myocardial infarctions (MI). The study will be conducted in two phases: a dose-escalation safety phase and a single dose efficacy phase. Eligible patients who present to the hospital with an acute ST-elevation MI and who agree to participate in this study will be randomly assigned to receive a single infusion of study medication consisting either of erythropoietin or placebo. The size of the infarction and the dimensions of the heart will be assessed by cardiac magnetic resonance imaging (MRI) within 2-6 days of the infusion of the study medication, and again approximately 3 months later.

— Show less

OFFICIAL TITLE
Effects of Erythropoietin on Infarct Size and Left Ventricular Remodeling in Survivors of Large **Myocardial Infarctions**

CONDITIONS
Acute ST Elevation Myocardial Infarction

INTERVENTION / TREATMENT
Drug: Epoetin alfa

STUDY START
2005-09

PRIMARY COMPLETION (ACTUAL)
2009-07

STUDY COMPLETION (ACTUAL)
2011-01

ENROLLMENT (ACTUAL)
223

STUDY TYPE
Interventional

PHASE
Phase 2

OTHER STUDY ID NUMBERS
999905255
05-AG-N255 (Other Identifier) (OTHER: NIH)

Eligibility Criteria

DESCRIPTION

- INCLUSION CRITERIA:

Age greater than 21 years

Acute ST-elevation myocardial infarction

Referral for primary or rescue angioplasty

Revascularization procedure within 8 hours from the onset of ischemic symptoms

TIMI (Thrombolysis in myocardial infarction) flow grade 0 or 1 in the culprit coronary artery at the beginning of coronary angiography

Successful revascularization of infarct-related artery

EXCLUSION CRITERIA:

Clinical indication for erythropoietin

STEMI (ST-elevation myocardial infarction) due to occlusion of a branch vessel

Any history of prior MI, PCI (Percutaneous coronary intervention), CABG (Coronary artery bypass graft), cardiomyopathy, myocarditis, or CHF (congestive heart failure)

Hypersensitivity to human albumin, mammalian cell-derived products, or erythropoietin

Hematocrit greater than 42% in men or greater than 40% in women at the time of study drug administration

Uncontrolled hypertension at the time of study drug administration

AGES ELIGIBLE FOR STUDY
21 Years and older (Adult, Older Adult)

SEXES ELIGIBLE FOR STUDY
All

ACCEPTS HEALTHY VOLUNTEERS
No

Medical Information

Secondary Knowledge-based Documents

- All medical professionals are not researchers: primary resources need to be rephrased, summarized, synthetized
- Summary and reviews of primary resources are published in scientific journals
- Quality issue: the editorial process is not the same for secundary than primary resources
- Other category: clinical practice guidelines (many publications, very little control)

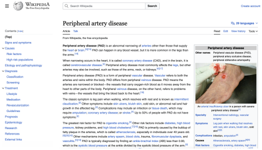
Specific case: Systematic Reviews and Meta-Analysis

- Fragmentation of the scientific literature → difficult to identify all the relevant papers on a topic
- In particular with clinical trials, large amount of publications on a similar condition or treatment
- Systematic reviews tackle a precise question, and describe the complete set of related work and factual approaches
- Meta-analysis compare results at the systematic review scale
- Topics: treatment (63%), causality and security (29%), diagnosis (4,4%), prognosis (2.1%) [Montori et al., 2004]

Medical Information

Curated Websites

1. Health portals
   - E.g. WebMD, Mayoclinic, MedlinePlus
   - Held by public/private institutions or companies
   - Varying quality and trustworthiness
2. Collaborative knowledge bases
   - E.g. Wikipedia, Radiopaedia, wikiDoc...
   - The literature provides strong evidence to position Wikipedia as a prominent health information resource for the public, patients, students, and practitioners seeking health information online [Smith, 2020]
   - 155,000 health articles using 950,000 citations to sources and which collectively received 4.8 billion pageviews in 2013 (all languages included)
3. Newspapers and magazines



Medical Information

User Generated Content

Collaborative writing websites allow users to edit collaboratively documents. It can have some sort of editorial control. It includes:

- **Wikis** such as wikipedia (collective writing and control of the content)

[Blackman, 2006] showed that information contained on wikipedia wasn't erroneous (comparison on 42 topics with the Britannica Encyclopaedia)

- **Blogs**: discussion or informational website published on the Web consisting of discrete, often informal diary-style text entries ("posts").

Not logged in Talk Contributions Create account Log i

WIKIPEDIA
The Free Encyclopedia

Main page
Contents
Featured content
Current events
Random article
Donate to Wikipedia
Wikipedia store

Interaction
Help
About Wikipedia
Community portal
Recent changes
Contact page

Tools
What links here
Related changes
Upload file
Special pages
Permanent link
Page information
Wikidata item

Article Talk Read Edit View history Search Wikipedia

Liver

From Wikipedia, the free encyclopedia

For other uses, see Liver (disambiguation).

The **liver**, an organ only found in vertebrates, detoxifies various metabolites, synthesizes proteins, and produces biochemicals necessary for digestion.[2][3][4] In humans, it is located in the right upper quadrant of the abdomen, below the diaphragm. Its other roles in metabolism include the regulation of glycogen storage, decomposition of red blood cells and the production of hormones.[4]

The liver is an accessory digestive gland that produces bile, an alkaline compound which helps the breakdown of fat. Bile aids in digestion via the emulsification of lipids. The gallbladder, a small pouch that sits just under the liver, stores bile produced by the liver.[5] The liver's highly specialized tissue consisting of mostly hepatocytes regulates a wide variety of high-volume biochemical reactions, including the synthesis and breakdown of small and complex molecules, many of which are necessary for normal vital functions.[6] Estimates regarding the organ's total number of functions vary, but textbooks generally cite it being around 500.[7]

Terminology related to the liver often starts in *hepat-* from ἡπατο-, the Greek word for liver.[8]

There is currently no way to compensate for the absence of liver function in the long term, although liver dialysis techniques can be used in the short term. Artificial livers are yet to be developed to promote long-term replacement in the absence of the liver. As of 2017,[9] liver transplantation is the only option for complete liver failure.

Contents [hide]

Liver

liver
gall bladder
spleen
stomach
colon
small intestine
appendix
bladder

The human liver is located in the upper right abdomen

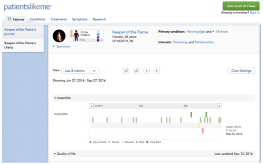
Medical Information

User Generated Content

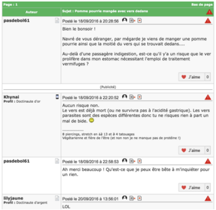
Health topics can be covered on all types of social media:

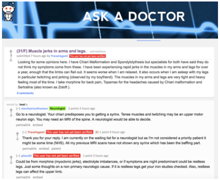
- General social media such as facebook, twitter:

- Medical social media such as PatientsLikeMe:



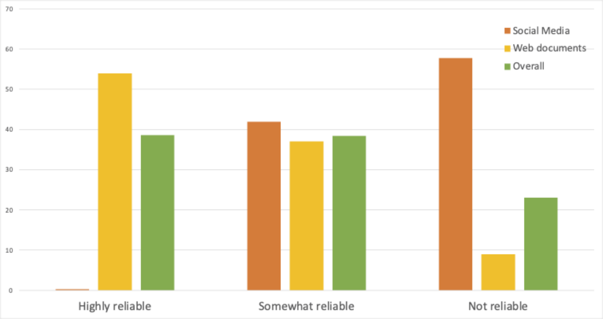
- Discussion forums: where all kinds of users (patients, doctors, students, nurses...) can discuss health topics





Medical Information

Trustworthiness of Health Information



- Within CLEF eHealth, the documents were manually assessed based on their topical relevance, readability and reliability [Goeuriot et al., 2021]
- 55 layuser queries on a broad range of topics
- 250 documents were assessed per query
- Documents in the pool come from CommonCrawl (71%) or social media Twitter and Reddit (29%)
- These results are in line with other studies such as [Scullard et al., 2010]

Medical Information

Certification

How can the quality of health information online be guaranteed?

The organization Health On the Net (HON) certifies the quality and validity of medical websites.
HON manually certifies website according to the following principles:

- Principle 1 : Authority - Give qualifications of authors
- Principle 2 : Complementarity - Information to support, not replace
- Principle 3 : Confidentiality - Respect the privacy of site users
- Principle 4 : Attribution - Cite the sources and dates of medical information
- Principle 5 : Justifiability - Justification of claims / balanced and objective claims
- Principle 6 : Transparency - Accessibility, provide valid contact details
- Principle 7 : Financial disclosure - Provide details of funding
- Principle 8 : Advertising - Clearly distinguish advertising from editorial content

https://www.hon.ch/HONcode/Guidelines/guidelines.html

Medical Information

Medical Textual Data - Summary

Data type	Characteristics	Challenges
Patient-centered Data		
Structured	Description of patients signs Contains measures in some structured form and sometimes free text	Privacy Numerical data Multiple ambiguous signs Availability
Narrative	Communication between practitioners Record of an event in the patient care Form and content Free text	Privacy Negation and uncertainty Noisy writing Many ambiguities Temporal aspects Availability
Knowledge-based Data		
Primary	Communication between experts Partially or fully available text Lightly structured - highly standardized High quality	References and links Contains tables and figures Quantity Availability
Secondary	Reliable information Good quality	Broad range of formats and content (and quality)
Websites		
Curated	Wide range of topics covered All levels of readability Available data	Broad range of formats, content Lower quality
User-generated Content	Anyone can write and publish anything	No quality control Potentially unreliable data Potentially noisy data

Outline

Medical Search

Users and Tasks

Information needs [Hersh, 2010]:

- *Retrospective information needs*:
  - The need for help in solving a certain problem or making a decision
  - The need for background information on a topic
- *Current awareness information needs*:
  - The need to keep up with information in a given subject area

General classification of search queries from [Broder, 2002]:

- Navigational
- Transactional
- Informational

Some of the tasks well explored in IR:

1. Researchers
   - Cohort studies
   - Clinical trials search
   - Systematic reviews
   - Literature search
2. Clinicians
   - Evidence-based medicine
   - Precision medicine
   - Clinical trials search
3. General public
   - Information about a condition
   - Symptom check
   - Treatments
   - Advice and support

Medical Search Queries

Clinical Queries

Clinicians queries:

- Relatively short [Palotti et al., 2016a]
- Not systematically pursued: 51% pursued, and for 78% of them some answers were found [Del Fiol et al., 2014]

Evidence-based Medicine: For a given patient case, clinicians might need evidence in the biomedical literature to support a decision regarding: a diagnosis, a treatment, a test, etc. [Roberts et al., 2015b]:

Topic 1 – Diagnosis
Description: A 44 yo male is brought to the emergency room after multiple bouts of vomiting that has a "coffee ground" appearance. His heart rate is 135 bpm and blood pressure is 70/40 mmHg. Physical exam findings include decreased mental status and cool extremities. He receives a rapid infusion of crystalloid solution followed by packed red blood cell transfusion and is admitted to the ICU for further care.
Summary: A 44-year-old man with coffee-ground emesis, tachycardia, hypoxia, hypotension and cool, clammy extremities.

Precision Medicine: Adapt a clinical decision to a specific patient, upon genetic, environmental, and lifestyle choices (e.g. oncology) [Roberts et al., 2017]

Disease: Liposarcoma **Variant:** CDK4 Amplification **Demographic:** 38-year-old male **Other:** GERD	**Disease:** Colon Cancer **Variant:** KRAS (G13D), BRAF (V600E) **Demographic:** 52-year-old male **Other:** Type II Diabetes, Hypertension

Medical Search Queries

Clinical Queries

Analysis of search queries in an EHR search utility [Natarajan et al., 2010]

- **Navigational queries (14.5%)**: were mostly aiming at retrieving a specific EHR (e.g. using the record number)
- **Transactional queries (0.4%)**: were representing an action (e.g. adding a new note)
- **Information queries (85.1%)**: the most frequent, especially among clinicians and researchers.

Top 5 semantic types of searches

Semantic type	%	Semantic type	%
Laboratory or test result	29.2	Pharmacologic substance	7.5
Disease or syndrome	21.7	Diagnostic procedure	6.2
Body part, organ or organ component	8.1		

Top 10 most frequent queries

Query	%	Query	%
class	9.8	nephrogenic	1.8
nyha	4.5	hysterectomy	1.5
hodgkins	2.9	cva	1.1
iii	2.4	ef	1.0
iv	2.3	hf	0.9

- Very short queries (1.2 term(s) on average in the corpus)
- Many acronyms (*NYHA*) and abbreviations (*tach* for tachycardia)
- Ambiguous (*class*)

Medical Search Queries

PICO Queries

Designed to answer Evidence-based Medicine problems, PICO stands for:

- **P**atient / **P**roblem / **P**opulation
- **I**ntervention
- **C**omparison / **C**ontrol
- **O**utcome

The formulation of a focused clinical question containing well-articulated PICO elements is widely believed to be **the key to efficiently finding high-quality evidence** and also **the key to evidence-based decisions** [Huang et al., 2006].

Example (from [Boudin et al., 2010]):

`"children with pain and fever how does paracetamol compared with ibuprofen affect levels of pain and fever?`

→

Patient/Problem: `children/pain and fever`
Intervention: `paracetamol`
Comparison: `ibuprofen`
Outcome: `levels of pain and fever`

Medical Search Queries

Clinical Search Queries (Genomics)

[Hersh and Voorhees, 2009] categorized clinical queries into seveval Generic Topic Types:

Generic Topic Type	Example Topic
Find articles describing standard methods or protocols for doing some sort of experiment or procedure	Method or protocol: GST fusion protein expression in Sf9 insect cells
Find articles describing the role of a gene involved in a given disease	Gene: DRD4 Disease: alcoholism
Find articles describing the role of a gene in a specific biological process	Gene: Insulin receptor gene Biological process: Signaling tumorigenesis
Find articles describing interactions (e.g. promote, suppress, inhibit, etc.) between two or more genes in the function of an organ or in a disease	Genes: HMG and HMGB1 Disease: Hepatitis
Find articles describing one or more mutations of a given gene and its biological impact	Gene with mutation: Ret Biological impact: Thyroid function

Medical Search Queries

Systematic review Queries

- Systematic reviews use boolean queries on specific databases such as the Cochrane library to retrieve all the possible relevant documents on a topic.
- Example (topic extracted from CLEF eHealth Technologically assisted reviews task [Kanoulas et al., 2017]):

```
Topic: CD009551
Title: Polymerase chain reaction blood tests for the diagnosis of
       invasive aspergillosis in immunocompromised people

Query:
exp Aspergillosis/
exp Pulmonary Aspergillosis/
exp Aspergillus/
(aspergillosis or aspergillus or aspergilloma or "A.fumigatus" or
"A. flavus" or "A. clavatus" or "A. terreus" or "A. niger").ti,ab.
or/1-4
exp Nucleic Acid Amplification Techniques/
pcr.ti,ab.
"polymerase chain reaction*".ti,ab.
or/6-8
5 and 9
exp Animals/ not Humans/
10 not 11

Pmid's:
    25815649
    26065322
    ...
```

Medical Search Queries

Layperson queries

Particularities and challenges [Zhang et al., 2012]

- *Conceptual level*: layperson have their own understandings and hypotheses about a particular condition.
- *Terminological level*: layperson's vocabulary doesn't match medical terminologies
- *Lexical level*: queries contain mispelling, partial words, etc.
- Short text (on average less than 3 words), ambiguous

fluid in leg

puffy sore calf

swollen legs

- Circumlocution: "turning around" - when many words are used to describe what could be say with fewer, more precise words [Stanton et al., 2014a]
- Commercial search engine often answer poorly such queries [Zuccon et al., 2015]

[Cartright et al., 2011a]

Topics covered:

- Symptom
- Cause
- Remedy

Types of queries:

- Evidence-directed
- Hypothesis-directed:
  - Diagnosis intent
  - Informational intent

Outline

Semantic Resources

Definitions [Hersh, 2010, Bast et al., 2016]

A concept

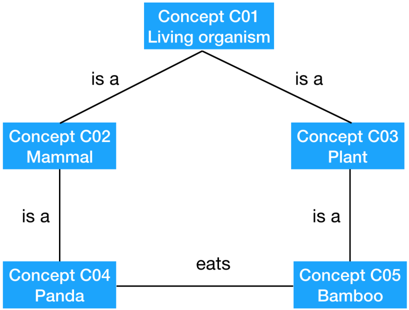
Idea or object that occurs in the world (e.g. *the condition under which human blood pressure is elevated*

A term

String of one or more words that represents a concept (e.g. *hypertension* or *high blood pressure*)

A relationship

Link between 2 concepts (e.g. the *liver* is an *organ*) or terms (e.g. *hypertension* and *high blood pressure* are synonyms)



A Knowledge Base is a collection of records in a database, which typically refers to some kind of knowledge about the world. Records are triples `(subject, predicate, object)`.

- [Bast et al., 2016] calls a *knowledge base* every collection of entities following an ontology.
- A **knowledge-base can be thought as a graph** where entities are the nodes and the relationships are the edges.

SEMANTIC RESOURCES

DEFINITIONS [HERSH, 2010, BAST ET AL., 2016]

A controlled vocabulary or Terminology

Contains a list of terms that are the canonical representation of the concepts.

An ontology

Is a formal description of the concepts in a domain of discourse. Contains a *hierarchy of concepts* with various *relationships* between them. Concepts have *attributes* and *facets*.

A thesaurus

Contains concepts, terms, and relationships between them:

- *Hierarchical* links between broader and narrower concepts
- *Synonymy* and antonymy
- *Related* somehow otherwise related

A Knowledge Base is a collection of records in a database, which typically refers to some kind of knowledge about the world. Records are triples `(subject, predicate, object)`.

- [Bast et al., 2016] calls a *knowledge base* every collection of entities following an ontology.
- A **knowledge-base can be thought as a graph** where entities are the nodes and the relationships are the edges.

Medical Knowledge Resources

Existing Medical thesauri (in English)

1. Medical Subject Headings (MeSH)

- Created by the National Library of Medicine *to index medical documents*
- 28,000 descriptors (concepts) with over 90,000 entry terms
- 3 types of relationships: hierarchical, synonymous, related

Hypertension MeSH Descriptor Data 2018

Details | Qualifiers | MeSH Tree Structures | Concepts

MeSH Heading
Hypertension
Tree Number(s)
C14.907.489
Unique ID
D006973
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(s)
Blood Pressure, High
NLM Classification #
WG 340
See Also
Antihypertensive Agents
Vascular Resistance
Date Established
1966/01/01
Date of Entry
1999/01/01
Revision Date
2010/06/25

The 16 trees in MeSH

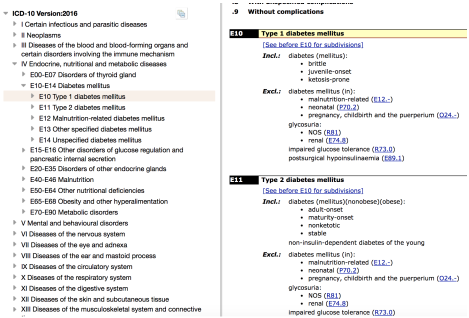
1. Anatomy
2. Organisms
3. Diseases
4. Chemicals and Drugs
5. Analytical, Diagnostic and Therapeutic Techniques and Equipment
6. Psychiatry and Psychology
7. Biological Sciences
8. Natural Sciences
9. Anthropology, Education, Sociology and Social Phenomena
10. Technology, Industry, Agriculture
11. Humanities
12. Information Science
13. Named Groups
14. Health care
15. Publication Characteristics
16. Geographicals

MEDICAL KNOWLEDGE RESOURCES

EXISTING MEDICAL THESAURI (IN ENGLISH)

2 International Classification of Medicine (ICD)

- International statistical classification of diseases and health problems
- Coded medical classification including a wide variety of signs, symptoms, trauma, etc.
- Published by the WHO
- Internationaly used to register morbidity and causes and morbidity



ICD-10 Version:2016
- I Certain infectious and parasitic diseases
- II Neoplasms
- III Diseases of the blood and blood-forming organs and certain disorders involving the immune mechanism
- IV Endocrine, nutritional and metabolic diseases
  - E00-E07 Disorders of thyroid gland
  - E10-E14 Diabetes mellitus
    - E10 Type 1 diabetes mellitus
    - E11 Type 2 diabetes mellitus
    - E12 Malnutrition-related diabetes mellitus
    - E13 Other specified diabetes mellitus
    - E14 Unspecified diabetes mellitus
  - E15-E16 Other disorders of glucose regulation and pancreatic internal secretion
  - E20-E35 Disorders of other endocrine glands
  - E40-E46 Malnutrition
  - E50-E64 Other nutritional deficiencies
  - E65-E68 Obesity and other hyperalimentation
  - E70-E90 Metabolic disorders
- V Mental and behavioural disorders
- VI Diseases of the nervous system
- VII Diseases of the eye and adnexa
- VIII Diseases of the ear and mastoid process
- IX Diseases of the circulatory system
- X Diseases of the respiratory system
- XI Diseases of the digestive system
- XII Diseases of the skin and subcutaneous tissue
- XIII Diseases of the musculoskeletal system and connective

.9 Without complications

E10 Type 1 diabetes mellitus

[See before E10 for subdivisions]

Incl.: diabetes (mellitus):
- brittle
- juvenile-onset
- ketosis-prone

Excl.: diabetes mellitus (in):
- malnutrition-related (E12.-)
- neonatal (P70.2)
- pregnancy, childbirth and the puerperium (O24.-)

glycosuria:
- NOS (R81)
- renal (E74.8)

impaired glucose tolerance (R73.0)
postsurgical hypoinsulinaemia (E89.1)

E11 Type 2 diabetes mellitus

[See before E10 for subdivisions]

Incl.: diabetes (mellitus)(nonobese)(obese):
- adult-onset
- maturity-onset
- nonketotic
- stable

non-insulin-dependent diabetes of the young

Excl.: diabetes mellitus (in):
- malnutrition-related (E12.-)
- neonatal (P70.2)
- pregnancy, childbirth and the puerperium (O24.-)

glycosuria:
- NOS (R81)
- renal (E74.8)

impaired glucose tolerance (R73.0)

ICD Classification

1 Certain infectious and parasitic diseases
2 Neoplasms
3 Diseases of the blood and blood-forming organs and certain disorders involving the immune mechanism
4 Endocrine, nutritional and metabolic diseases
5 Mental and behavioural disorders
6 Diseases of the nervous system
7 Diseases of the eye and adnexa
8 Diseases of the ear and mastoid process
9 Diseases of the circulatory system
10 Diseases of the respiratory system
11 Diseases of the digestive system
12 Diseases of the skin and subcutaneous tissue
13 Diseases of the musculoskeletal system and connective tissue
14 Diseases of the genitourinary system
15 Pregnancy, childbirth and the puerperium
16 Certain conditions originating in the perinatal period
17 Congenital malformations, deformations and chromosomal abnormalities
18 Symptoms, signs and abnormal clinical and laboratory findings, not elsewhere classified
19 ...

Medical Knowledge Resources

Existing Medical thesauri (in English)

3. Systematized Nomenclature of Medicine (SNOMED): thesaurus designed to process clinical data
4. Cumulative Index to Nursing and Allied Health Literature (CINAHL): classical medical concepts + domain-specific ones
5. EMTREE: European MeSH, used to index EMBASE
6. PsycINFO: psychology and psychiatry thesaurus
7. Gene Ontology: description of biomolecular biology (molecular functions, biological processes, cellular components) - designed to structure the knowledge rather than index content
8. National Cancer Institute (NCI) thesaurus: knowledge model enabling cross-disciplinary communication and collaboration

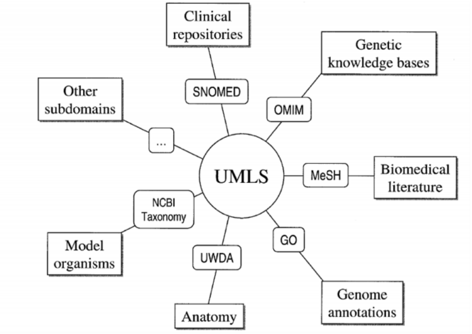
Many thesauri are also available in many well-endowed languages.

Medical Knowledge Resources

Existing Medical thesauri (in English)

The Unified Medical Language System (UMLS)

- Purpose: provide a mecanism to link existing medical thesaurus and controlled vocabularies
- Initiated in 1986 and maintained by the National Library of Medicine
- Contains: a metathesaurus, a semantic network, NLP tools
- Gathers more than 100 thesauri/vocabulary



Bodenreider, O. (2004) The Unified Medical Language System (UMLS): integrating biomedical terminology. *Nucleic Acids Research, 32, D267-D270.*

MEDICAL KNOWLEDGE RESOURCES

EXISTING MEDICAL THESAURI (IN ENGLISH)

The Unified Medical Language System (UMLS)

- All the entries corresponding to the same thing are registered as a single concept
- To each concept correspond several terms that represent an expression of the concept
- Each concept is described with: CUI, semantic type, definitions, synonyms, relations
- The information from all the thesauri containing the concept is stored

Concept (CUI)	Term (LUI)	String (SUI)	Atom (AUI)
C0004238 Atrial fibrillation (preferred) Atrial fibrillations Auricular fibrillation Auricular fibrillations	**L0004238** Atrial fibrillation (preferred) Atrial fibrillations	**S0016668** Atrial fibrillation (preferred)	**A0027665** Atrial fibrillation (from MSH)
			A0027667 Atrial fibrillation (from PSY)
		S0016669 Atrial fibrillations (plural variant)	**A0027668** Atrial fibrillations (from MSH)
	L0004327 Auricular fibrillation Auricular fibrillations (synonyms)	**S0016899** Auricular fibrillation (preferred)	**A0027930** Auricular fibrillation (from PSY)
		S0016900 Auricular fibrillations (plural variant)	**A0027932** Auricular fibrillations (from MSH)

Medical Knowledge Resources

Existing Medical thesauri (in English)

The Unified Medical Language System (UMLS)

Basic View | Report View | Raw View

- Concept: [C0004238] Atrial Fibrillation
- Semantic Type
- Definition
- Synonyms (96)
- Relations (1672) REL | RELA | RSAB| String | CUI

Medical Knowledge Resources

Existing Medical thesauri (in English)

The Unified Medical Language System (UMLS)

Basic View | Report View | Raw View

⊟ Concept: [C0004238] Atrial Fibrillation

DA Date 1990-09-30 05:00:00.000000000

MR Major Revision Date 2017-09-14 06:00:00.000000000

ST Status R

⊟ Semantic Type

Disease or Syndrome [T047]

⊞ Definition

⊞ Synonyms (96)

⊞ Relations (1672) REL | RELA | RSAB| String | CUI

Medical Knowledge Resources

Existing Medical thesauri (in English)

The Unified Medical Language System (UMLS)

Basic View | Report View | Raw View

⊞ Concept: [C0004238] Atrial Fibrillation

⊞ Semantic Type

⊟ Definition

CHV/null - rapid tremor and shake of upper chambers of the heart

CSP/null - disorder of cardiac rhythm characterized by rapid, irregular atrial impulses and ineffective atrial contractions.

HPO/null - An atrial arrhythmia characterized by disorganized atrial activity without discrete P waves on the surface EKG, but instead by an undulating baseline or more sharply circumscribed atrial deflections of varying amplitude an frequency ranging from 350 to 600 per minute. [HPO:probinson]

MEDLINEPLUS/null -

An arrhythmia is a problem with the speed or rhythm of the heartbeat. Atrial fibrillation (AF) is the most common type of arrhythmia. The cause is a disorder in the heart's electrical system.

Often, people who have AF may not even feel symptoms. But you may feel

- Palpitations -- an abnormal rapid heartbeat
- Shortness of breath
- Weakness or difficulty exercising
- Chest pain
- Dizziness or fainting
- Fatigue
- Confusion

AF can lead to an increased risk of stroke. In many patients, it can also cause chest pain, heart attack, or heart failure.

Doctors diagnose AF using family and medical history, a physical exam, and a test called an electrocardiogram (EKG), which looks at the electrical waves your heart makes. Treatments include medicines and procedures to restore normal rhythm.

Medical Knowledge Resources

Existing Medical thesauri (in English)

The Unified Medical Language System (UMLS)

Basic View | Report View | Raw View

- Concept: [C0004238] Atrial Fibrillation
- Semantic Type
- Definition
- Synonyms (96)
  - ACFA (arythmie complète par fibrillation auriculaire)
  - AF
  - AF - Atrial fibrillation
  - AFib
  - ATRIAL FIBRILLATION
  - ATRIJ, FIBRILACIJA
  - AURICULAR FIBRILLATION
  - AURICULAR, FIBRILACION
  - Afib
  - Atrial Fibrillation
  - Atrial Fibrillation [Disease/Finding]
  - Atrial Fibrillations
  - Atrial fibrillation
  - Atrial fibrillation (disorder)
  - Atrieflimmer
  - Atriumfibrillatie
  - Auricular Fibrillation
  - Auricular Fibrillations

Medical Knowledge Resources

Existing Medical thesauri (in English)

The Unified Medical Language System (UMLS)

Basic View | Report View | Raw View

⊞ Concept: [C0004238] Atrial Fibrillation
⊞ Semantic Type
⊞ Definition
⊞ Synonyms (96)
⊟ Relations (1672) REL | RELA | RSAB| String | CUI

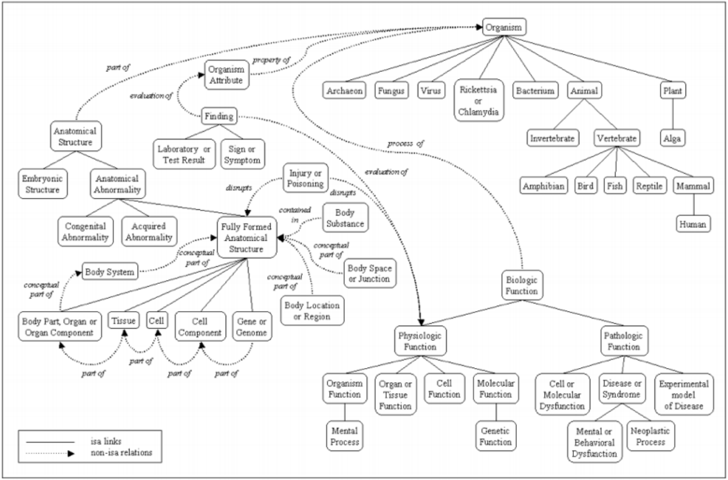
[: 1 - 10 : »]

AQ | | MSH | In Blood | C0005768
AQ | | MSH | In Cerebrospinal Fluid | C0007807
AQ | | MSH | chemically induced | C0007994
AQ | | MSH | Taxonomic | C0008903
AQ | | MSH | Congenital MeSH qualifier | C0009678
AQ | | MSH | nutritional management | C0012160
AQ | | MSH | pharmacotherapeutic | C0013217
AQ | | MSH | Economic | C0013557
AQ | | MSH | embryologic | C0013943
AQ | | MSH | enzymology | C0014445

Medical Knowledge Resources

Existing Medical thesauri (in English)

The Unified Medical Language System (UMLS)



http://genome.tugraz.at/MedicalInformatics/WinterSemester2011Holzinger/3_LV444_152_STRUCTURED_DATA_HOLZINGER_2011.pdf

Semantic Annotation

Semantic annotation consists in linking documents to knowledge bases by identifying:

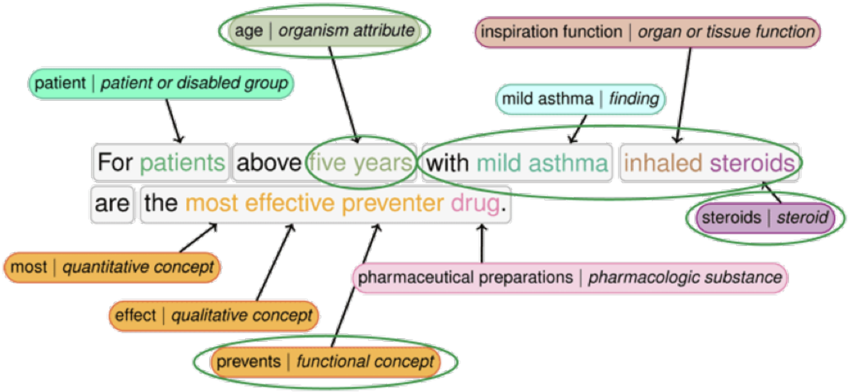
- Entities/concepts (in the document)
- Relationships (implicit or explicit):
  - Between entities: *HK1 involved in glycolytic process*
  - Between an entity and the document:
    - MeSH entities for indexing documents on MEDLINE: *PMID:3207429 is indexed with* `Glucose/metabolism` *and* `Hexokinase/genetics`
    - The ICD code in a medical report

Semantic annotation can be done:

- Automatically
  - Named Entity Recognition and normalization
  - Automatic indexing
- Manually
  - Data curation, manual labelling
  - Manual indexing (majority on MEDLINE)
  - Codes added to patients reports when billing
- Derived from other annotations
  - Using links between concepts and entities

Semantic Annotation

Annotated sentence:



http://ieg.ifs.tuwien.ac.at/~gschwand/mapface/project_page/img/corrections.gif

Existing Tools

Access to UMLS

- Requires a license (takes up to a few days - free)
- Gives access to the metathesaurus, along with other resources and tools created by the NLM
- `https://www.nlm.nih.gov/research/umls/index.html`

Existing tools for medical named-entity recognition:

Metamap : designed to annotate biomedical literature, interactive, API or batch versions available [Aronson and Lang, 2010]

QuickUMLS : designed to annotate medical text *faster* than other tools [Soldaini and Goharian, 2016]

cTakes : UIMA-based system designed to annotate EHRs [Savova et al., 2010]

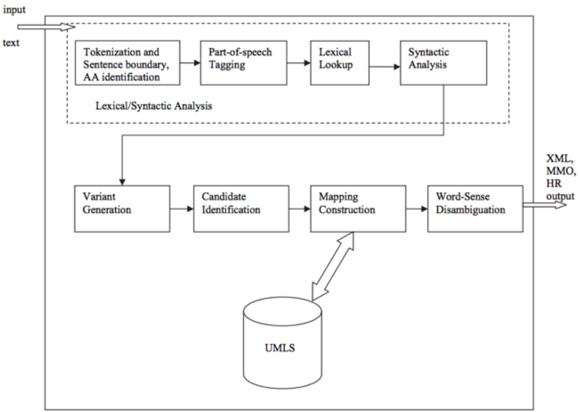
SemRep : designed to extract semantic relations from biomedical text (as 3-parts propositions) [Rindflesch and Fiszman, 2003]

MedCAT : Self-supervised NER tool trained on EHR [Kraljevic et al., 2021]

MedSpacy, SciSpacy : python-based medical and biomedical NLP toolkits

METAMAP

Figure 1 MetaMap system diagram. HR, human readable; MMO, MetaMap machine output; UMLS, unified medical language system.



Extracted from [Aronson and Lang, 2010]

Semantic Annotation

MetaMap

Input Text:

```
The female patient suffers from TIA and high blood pressure
```

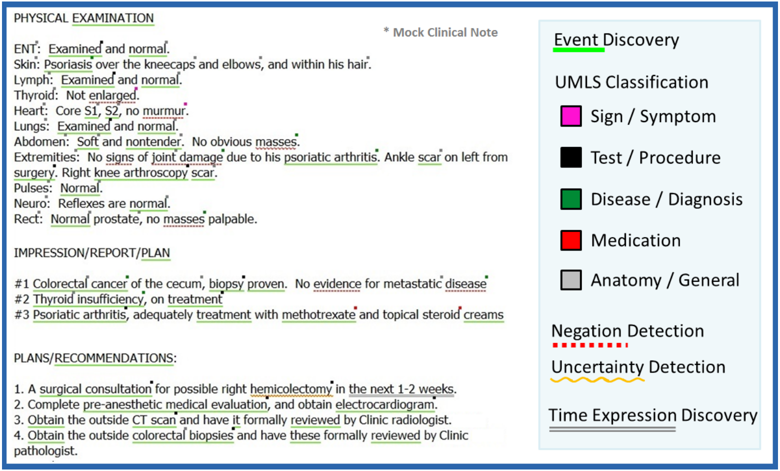
Results:

```
Processing 00000000.tx.1: The female patient suffers from TIA and high blood pressure

Phrase: The female patient
>>>>> Phrase
female patient
<<<<< Phrase
>>>>> Mappings
Meta Mapping (1000):
  1000   C0150905:female patient (patient is female) [Organism Attribute]
<<<<< Mappings

Phrase: suffers from TIA
>>>>> Phrase
suffers from tia
<<<<< Phrase
>>>>> Mappings
Meta Mapping (746):
   790   C0683278:suffers (Mental Suffering) [Mental or Behavioral Dysfunction]
   790   C0007787:TIA (Transient Ischemic Attack) [Disease or Syndrome]
Meta Mapping (746):
   790   C0683278:suffers (Mental Suffering) [Mental or Behavioral Dysfunction]
   790   C0917805:TIA (Transient Cerebral Ischemia) [Disease or Syndrome]
Meta Mapping (746):
   790   C0683278:suffers (Mental Suffering) [Mental or Behavioral Dysfunction]
   790   C1054154:tia (Tacca leontopetaloides) [Plant]
<<<<< Mappings

Phrase: and
>>>>> Phrase
<<<<< Phrase

Phrase: high blood pressure
>>>>> Phrase
high blood pressure
<<<<< Phrase
>>>>> Mappings
Meta Mapping (1000):
  1000   C0020538:Blood Pressure, High (Hypertensive disease) [Disease or Syndrome]
Meta Mapping (1000):
  1000   C2926615:High blood pressure (Ever told by doctor or nurse that you have high blood pressure:Finding:Point in time:^Patient:Ordinal) [Clinical Attribute]
<<<<< Mappings
```

Screenshot from the MetaMap interactive tool

CTAKES

Apache cTakes is a UIMA-based natural language processing system for extraction of information from electronic medical record clinical free-text:



Exctracted from https://ctakes.apache.org/

QuickUMLS

- A faster option to identify UMLS concepts from unstructured documents
- Gives comparable accuracy on several datasets

Method		*Prec*	*Rec*	*F-1*	*ms/doc*
MetaMap		0.49*	0.48*	0.48*	19,295*
cTAKES		**0.71**	0.53*	0.61	3,852*
QuickUMLS	$\alpha = 0.6$	0.50*	**0.75**	0.60	1,594*
	$\alpha = 0.7$	0.60*	0.66*	**0.63**	680*
	$\alpha = 0.8$	0.63*	0.60*	0.61	332*
	$\alpha = 0.9$	0.64*	0.56*	0.60	193*
	$\alpha = 1.0$	0.67*	0.54*	0.60	**143**

Table 1: Results for the i2b2 dataset. cTAKES outperforms QuickUMLS in precision, but QuickUMLS has better recall. QuickUMLS is 2.5 to 135 times faster than MetaMap or cTAKES. * indicates statistically significant differences from best value (Welch's t-test, $p < 0.05$).

Method		*Prec*	*Rec*	*F-1*	*ms/doc*
MetaMap		0.71*	0.53*	0.61*	15,935
cTAKES		**0.89**	0.55*	0.68*	3,765*
QuickUMLS	$\alpha = 0.6$	0.68*	**0.77**	**0.72**	1,536*
	$\alpha = 0.7$	0.78*	0.67*	**0.72**	646*
	$\alpha = 0.8$	0.83*	0.61*	0.70†	340*
	$\alpha = 0.9$	0.85*	0.57*	0.68*	174*
	$\alpha = 1.0$	0.87*	0.55*	0.67*	**141**

Table 2: Results for the THYME corpus. cTAKES achieves the best precision, QuickUMLS the best recall and substantially better throughput than MetaMap or cTAKES. * indicates statistically significant differences from best value (Welch's t-test, $p < 0.05$).† indicates statistical significance from α= 0.6 but not α= 0.7.

SUMMARY

- Many types of data :
  - Patient-specific data
  - Knowledge-based data
  - User-generated data
- Various information needs: search for information, for cohorts, for evidence...
- On the solution side:
  - Some structured knowledge bases: UMLS and all its constituents (and their versions in languages other than English)
  - Several concept annotation tools: MetaMap, cTakes, QuickUMLS...
  - IR models
  - LLMs

Outline

Why is medical IR important?

A largely web-driven activity

- Search engines and social media are popular tools for seeking and sharing information about a range of health conditions
  [De Choudhury et al., 2014, White and Horvitz, 2014]. PewInternet, October 2013:

The % of adult internet users who have looked online in the last 12 months for information about...

55%	Specific disease or medical problem
43	Certain medical treatment or procedure
27	How to lose weight or how to control your weight
25	Health insurance, including private insurance, Medicare or Medicaid
19	Food safety or recalls
16	Drug safety or recalls
16	A drug you saw advertised
15	Medical test results
14	Caring for an aging relative or friend
12	Pregnancy and childbirth
11	How to reduce your health care costs
20	Any other health issue
72	**at least one of the above topics**

- Existence of behavioral patterns suggesting a strong relationship between search behavior and health care [White and Horvitz, 2010, White and Horvitz, 2014]

What makes medical search challenging?

Overview

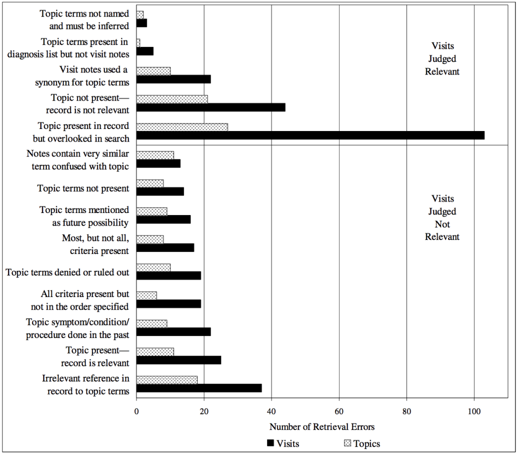
		Lexical representation	Lexical Matching	Result appraisal
Text (Queries/Documents)	High occurrence of lexical variants	■	■	■
	Ambiguity of language	■	■	■
	Significant presence of negation	■	■	■
	Time factors	■	■	■
Search task	Domain-specific task	□	■	■
User	Variability in levels of expertise	■	□	■
	Difficulty of understanding medical language	□	□	■
	Cognitive bias	□	□	■

- Lexical representation [Tamine et al., 2015, Stanton et al., 2014b, Limsopatham et al., 2013, Edinger et al., 2012, Chapman et al., 2001]
- Lexical matching [Edinger et al., 2012, Dinh and Tamine, 2012]
- Result appraisal [Koopman and Zuccon, 2014, Tamine and Chouquet, 2017, Roberts et al., 2015a, W. White and Horvitz, 2009, White and Horvitz, 2013, Cartright et al., 2011b, Palotti et al., 2016b]

What makes medical search challenging?

Lexical representation and matching issues

- Failure analysis from the TREC Medical Records Track [Edinger et al., 2012]
  - Same task across queries: retrieve cohorts of patients fitting criteria similar to those specified for participation in clinical studies
  - Same user's profile: domain-expert (physician)
  - Main results: **both precision errors and recall errors were due to bad lexical representations and lexical mismatches**



What makes medical search challenging?

Domain-specific tasks: need of task-dependent matching?

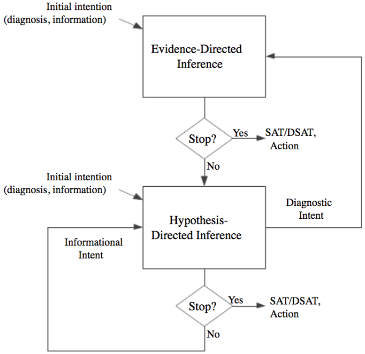
Study of the peculiarities of queries across medical tasks [Tamine et al., 2015]

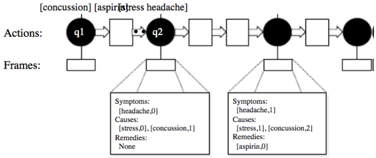
- Study of 4 search tasks:
  - (T1) Retrieve from biomedical scientific literature relevant references that address biology or related protein products (eg., '*Arginin vassopressin*')
  - (T2) Retrieve from medical scientific literature, relevant documents that fit with a patient profile (eg., '*Adult respiratory distress syndroms*')
  - (T3) Identify cohorts in clinical studies for comparative effectiveness research (eg., *Retrieve relevant medical cases including images for differential diagnosis*)
  - (T4) Identify relevant references that deal with typical relations between an entity and a medical process (eg., *What is the role of gene gamma-aminobutyric acid receptors in the process of inhibitory synaptic transmission?*)
- Same user's profile: domain-expert (physician)
- Main results:
  - Queries vary significantly across tasks: have multiple topical facets expressed using different levels of specifity w.r.t. medical terminology
  - Query performance significantly vary across tasks: however, the shorter and less specific the query is, the more difficult it is regardless of the task

What makes medical search challenging?

Domain-specific tasks: need of task-dependent matching?

- Diagnosis, a common medical search setting, is a highly complex task [Cartright et al., 2011b]
- Identifying two medical foci in search sessions corresponding to two iterative and interactive phases:
  - Evidence-directed: findings are merged to build a list of potential explanatory diagnoses ranked by likelihood
  - Hypothesis-directed: list of diagnoses used to guide collection of additional evidence to validate candidate hypotheses



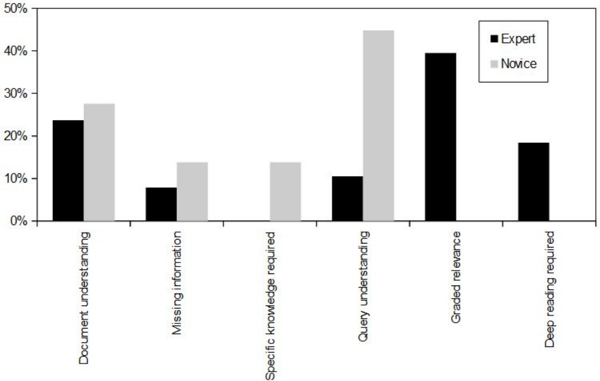


What makes medical search challenging?

Result appraisal: what does make it difficult with respect to user's expertise?

Study the factors of the relevance assessment task difficulty [Tamine and Chouquet, 2017]

- Qualitative analysis of relevance assessments for equivalent information needs
- Different user's profiles: domain-expert (physician), domain-novice (laypeople)
- Main results:
  - Levels of relevance agreement are low for both experts and novices
  - Better level of relevance agreement among experts than novices
  - More than third of the assessors found the relevance assessment task difficult but the underlying reasons different among experts vs. novices



What makes medical search challenging?

User cognitive bias - Cybercondria

- *Cybercondria: Study the potential of web search to escalate user's medical concerns* [White and Horvitz, 2013, W. White and Horvitz, 2009]
  - Qualitative and quantitative analysis of users' medical search logs and surveys
  - Observe the escalation within session: severity increase in the search terms of evolving medical search sessions
  - Main results:
    - Using Web search to perform diagnosis is a common user activity (more than 24%) while general-purpose ranking functions are based on the presence of lexical query symptoms, not designed for diagnosis inference
    - Common users consider that the system ranks the potential explanatory diagnoses by likelihood
    - Out of 8000 real sessions (extracted from Bing logs), 30% contained an escalation of concerns, 25% contained an inverse escalation

Table I. Probability of Mention of Cause Given Symptom

Symptom	Cause	Web Crawl	Web Search	Domain Search
headache	caffeine withdrawal	.29	.26	.25
	tension	.68	.48	.75
	brain tumor	.03	.26	.00
muscle twitches	benign fasciculation	.53	.12	.34
	muscle strain	.40	.38	.66
	ALS	.07	.50	.00
chest pain	indigestion	.28	.35	.38
	heartburn	.57	.28	.52
	heart attack	.15	.37	.10

Summary of issues

- Semantic gap
  - Vocabulary mismatch between experts and novices
  - Vocabulary mismatch between the query and documents
  - Interpretation of vocabulary in context: negation, lexical variants, time, task peculiarities
- Result appraisal
  - Understand medical language
  - Ability to interpret the results, make accurate inference, assess the credibility
  - Cognitive bias

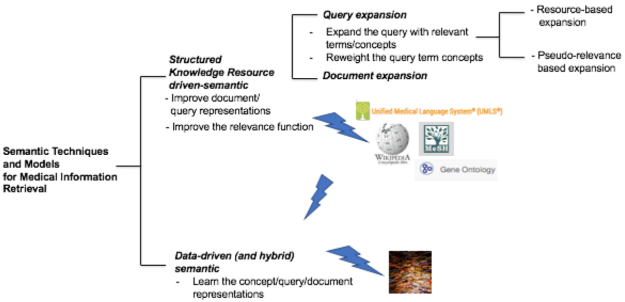
Outline

Outline

ROADMAP

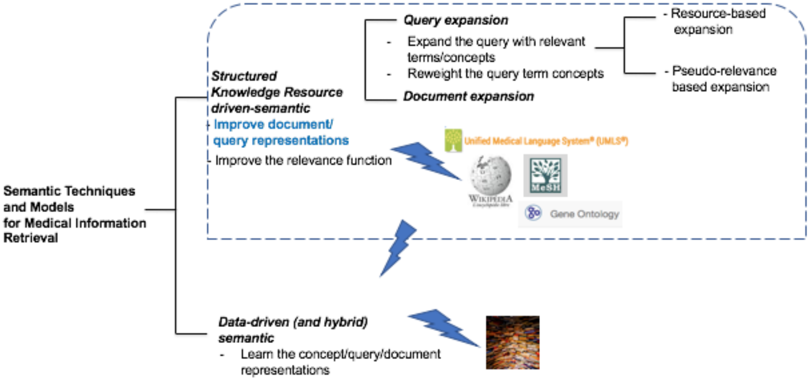
- Traditional IR
  - Q: bag of words
  - D: bag of words
  - RSV(Q,D): Alignment of Q and D

- Semantic (medical) IR
  - Q:
    - Bag of words
    - Bag of words **and concepts/entities**
    - **Embeddings**
  - D:
    - Bag of words
    - Bag of words **and concepts/entities**
    - **Embeddings**
  - RSV(Q,D): **Semantic inference**



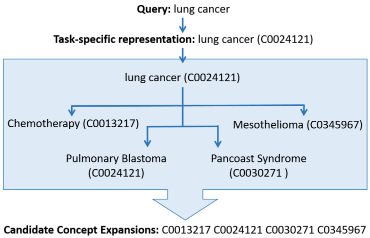
OUTLINE

Roadmap


Semantic Techniques and Models for Medical Information Retrieval
Structured Knowledge Resource driven-semantic
Improve document/ query representations
Improve the relevance function
Query expansion
- Expand the query with relevant terms/concepts
- Reweight the query term concepts
Document expansion
- Resource-based expansion
- Pseudo-relevance based expansion
Unified Medical Language System® (UMLS®)
Wikipedia
MeSH
Gene Ontology
Data-driven (and hybrid) semantic
- Learn the concept/query/document representations

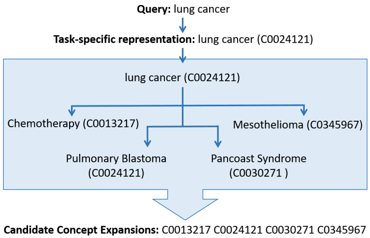
QUERY/DOCUMENT EXPANSION

- Query/document expansion
  - Enhance the Query/Document using:
    - evidence from related words/terms in semantic resources;
    - relevance feedback signals

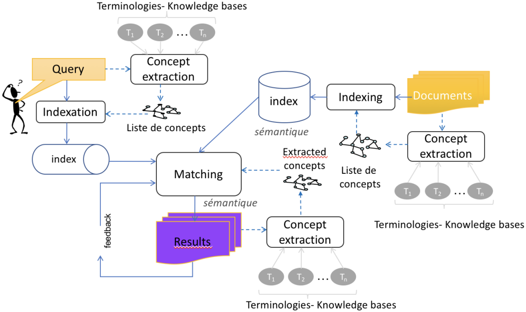


Query/document expansion

- Query/document expansion
  - Enhance the Query/Document using:
    - evidence from related words/terms in semantic resources;
    - relevance feedback signals



  - The full framework



QUERY/DOCUMENT EXPANSION

- Main impacting factors: [Dinh et al., 2013, Zuccon and Koopman, 2018]
  - Which knowledge-base to use (specialized vs. generic) and how many?
  - Which context to use (global vs. local)?
  - How to select candidate expansion terms and (how to inject them in a retrieval model) ?

QUERY/DOCUMENT EXPANSION

- Main impacting factors: [Dinh et al., 2013, Zuccon and Koopman, 2018]
  - Which knowledge-base to use (specialized vs. generic) and how many?
  - Which context to use (global vs. local)?
  - How to select candidate expansion terms and (how to inject them in a retrieval model) ?
- Resulting techniques
  - LSMo: Local context, Specialized Mono-Resource [Soldaini et al., 2017, Sondhi et al., 2012]
  - GSMo: Global context, Specialized Mono-Resource [Martinez et al., 2014, Znaidi et al., 2016, Demner-Fushman et al., 2006, Shen et al., 2014, Zhu et al., 2013]
  - GSMu: Global context, Specialized Multiple-Resources [Ando et al., 2005, Huang et al., 2005]
  - LGSMo: Local and Global contexts, Specialized Mono-Resource [Wang and Akella, 2015, Znaidi et al., 2015, Znaidi et al., 2016, Xu et al., 2019]
  - GSGMu: Global context, Specialized General Multiple-Resources [Soldaini et al., 2016]
  - LGSM : Local and Global contexts, Specialized Multiple-Resources [Limsopatham et al., 2013, Dinh and Tamine, 2012, Oh and Jung, 2015, Zhu and Carterette, 2012, Balaneshinkordan and Kotov, 2019, Ai et al., 2006]

		LSMo	GSMo	LGSMo	GSGMu	LGSMu
Context	Local (Pseudo-relevance)	■	□	■	□	■
	Global (Resource)	□	■	■	■	■
Knowledge Base	Specialized	■	■	■	■	■
	General	□	□	□	■	□
	Mono-resource	■	■	■	□	□
	Multiple-resources	□	□	□	■	■

QUERY/DOCUMENT EXPANSION

LOCAL CONTEXT, ONE SPECIALIZED RESOURCE [SONDHI ET AL., 2012]

- **Goal: retrieve medical literature relevant to case queries using medical thesauri and physician feedback**
- **Context:** Top N retrieved documents
- **Knowledge-Base**: MeSH thesaurus

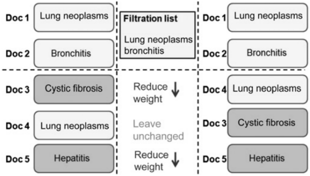
- **Key steps**
  - Map query words to UMLS semantic types and assign weights $c'(w, Q) = c(w, Q)$ if w belong to a relevant type eg., disease, syndrome, body, etc.
  - **Top-N based MeSH feedback**: identify a list of potential diagnoses from N top documents and then rerank the documents w.r.t absence of potential diseases
  - **Distribution-based MeSH feedback**: For each MeSH term, identify all the documents indexed with it, pick the M highest scoring MeSH terms as candidate term expansion
  - Expand the query and then perform a pseudo-relevance feedback based model (PRF) [Zhai and Lafferty, 2001]

Query/document expansion

Local context, one specialized resource [Sondhi et al., 2012]

- **Goal: retrieve medical literature relevant to case queries using medical thesauri and physician feedback**
- **Context:** Top N retrieved documents
- **Knowledge-Base**: MeSH thesaurus

- **Key steps**
  - Map query words to UMLS semantic types and assign weights $c^{'}(w, Q) = c(w, Q)$ if w belong to a relevant type eg., disease, syndrome, body, etc.
  - **Top-N based MeSH feedback**: identify a list of potential diagnoses from N top documents and then rerank the documents w.r.t absence of potential diseases
  - **Distribution-based MeSH feedback**: For each MeSH term, identify all the documents indexed with it, pick the M highest scoring MeSH terms as candidate term expansion
  - Expand the query and then perform a pseudo-relevance feedback based model (PRF) [Zhai and Lafferty, 2001]



- **Main results/findings**
  - Slight improvements (more than 6%) over the baseline for the Distribution-based MeSH feedback while the top N based Mesh feedback is worse than the baseline using small datasets (19 queries, 5585 documents)
  - Difficulty in recovering new treatments and rare alternative diseases

Query/document expansion

Local context, one general resource [Soldaini et al., 2017]

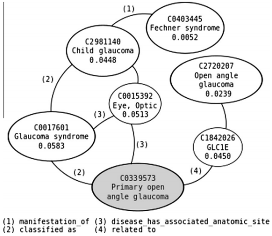
- **Goal:** Query reformulation for clinical decision support
- **Context:** Top N retrieved documents
- **Knowledge-Base**: Wikipedia

- **Key steps: Health Terms Pseudo Relevance Feedback HTPRF**
  - Retrieve the N Top documents w.r.t query Q
  - For each term from the top N documents, compute a score
    $s_j = log_{10}(10 + w_j)\, w_j = \alpha * tf(t_j, Q) + \frac{\beta}{N} \sum_1^N (tf(t_j, D_i) * idf(t_j))$
  - Select the top M terms with the highest score as the candidate expansion terms
  - For each candidate term expansion, compute the likelihood of being health-related. Compute the odds ratio as the proportion of health-related Wikipdia (W_H) documents including term t_i
    $OR(t_i) = \frac{n(t_i, W_H)}{n(t_i, W)}$
  - Consider the top M ranked terms with score $OR(t_j) > \sigma$
  - Expand the query and perform a pseudo-relevance feedback based model
- **Main results/findings**
  - Mapping the wikipedia terms to UMLS semantic types revealed that 75% are present in the UMLS: 32% are symptoms, 20.3% are treatments, 18% are a diagnosis procedure or test, 17.1% are diseases
  - The HTPRF parameters do not significantly impact the results
  - Precision oriented with slight improvement (+3.6%) over state-of the best systems in TREC CDS 2014-TREC CDS 2015

Query/document expansion

Global context, one specialized resource [Martinez et al., 2014]

- **Goal**: Searching EHRs
- **Context**: Concepts and relations between concepts
- **Knowledge-Base**: UMLS thesaurus
- **Key steps**
  - Map query words to UMLS semantic types
  - Identify the initial sub-graph based concept including query concepts and related UMLS concepts
  - Assign an uniform probability to the concepts in the sub-graph and then run the Page Rank algorithm
  - Rank the concepts using the Page Rank score
  - Expand the query with the N concepts having the highest PageRank Score
  - Perform a basic retrieval model (eg., TF-IDF, BM25)
- **Main results/findings**
  - Experiments on TREC medical records 2011-2012 show significant improvements (+30% in average)
  - Expansion terms are those related to the query with either taxonomic (eg., synonyms) and not taxonomic (eg., disease has associated anatomic site).
  - Useful expansion in the case of a cohort retrieval task.



Queries with highest improvement for PageRank, together with the learnt expansion terms and the Bpref increase.

Query	TREC version	Expansion terms	Bpref increase
Hospitalized patients treated for methicillin-resistant Staphylococcus aureus (MRSA) endocarditis	2011	MRSA elsewhere/NOS Personal history of poliomyelitis Personal history of other infectious and parasitic disease	0.931
Patients with Primary Open Angle Glaucoma (POAG)	2012	Eye, Eyeball, Globe, Ocular… Glaucoma syndrome Open cleft glaucoma GLC1E	0.742
Patients with adult respiratory distress syndrome	2012	Immunology Taxonomy Metabolism Historical aspects	0.722

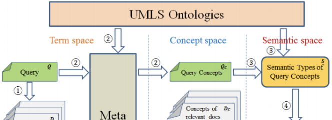
Query/document expansion

Global and local contexts, one specialized resource [Wang and Akella, 2015]

- **Goal:** Adhoc medical information retrieval
- **Context:** top N retrieved documents
- **Knowledge-base:** UMLS

- **Key steps**
  - Map query Q and pseudo-relevant document words to UMLS semantic types
  - Build the document model using the distribution of semantic type s in the set of relevant document D_C: $p(c \mid D_C) = \sum_s p(s \mid D_C) P(c \mid s, D_C)$
  - Compute $P(c \mid s, D_C)$ by the maximum likelihood estimation of c in this semantic type, and then smoothed by the collection background model
  - Computed $p(c \mid R_c)$ with R_c the unknown concept relevance model by estimating the joint probability of observing the concept c together with query concepts

    $$p(c \mid R_C) = p(c \mid q_{C1} \ldots q_{Ck}) = \frac{p(c, q_{C1} \cdots q_{Ck})}{(q_{C1} \cdots q_{Ck})}$$
  - Rank the concepts using the $p(c \mid R_c)$ score
  - Expand the query with the N concepts having the highest score $p(c \mid Q_c') = (1 - \lambda) * p(c \mid D_c) + \lambda * p(c \mid R_c)$



Query/document expansion

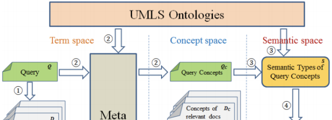
Global and local contexts, one specialized resource [Wang and Akella, 2015]

- **Goal:** Adhoc medical information retrieval
- **Context:** top N retrieved documents
- **Knowledge-base:** UMLS

- **Key steps**
  - Map query Q and pseudo-relevant document words to UMLS semantic types
  - Build the document model using the distribution of semantic type s in the set of relevant document D_C: $p(c \mid D_C) = \sum_s p(s \mid D_C) P(c \mid s, D_C)$
  - Compute $P(c \mid s, D_C)$ by the maximum likelihood estimation of c in this semantic type, and then smoothed by the collection background model
  - Computed $p(c \mid R_c)$ with R_c the unknown concept relevance model by estimating the joint probability of observing the concept c together with query concepts

    $$p(c \mid R_C) = p(c \mid q_{C1} \ldots q_{Ck}) = \frac{p(c, q_{C1} \cdots q_{Ck})}{(q_{C1} \cdots q_{Ck})}$$
  - Rank the concepts using the $p(c \mid R_c)$ score
  - Expand the query with the N concepts having the highest score $p(c \mid Q'_c) = (1 - \lambda) * p(c \mid D_c) + \lambda * p(c \mid R_c)$

- **Main results/findings**



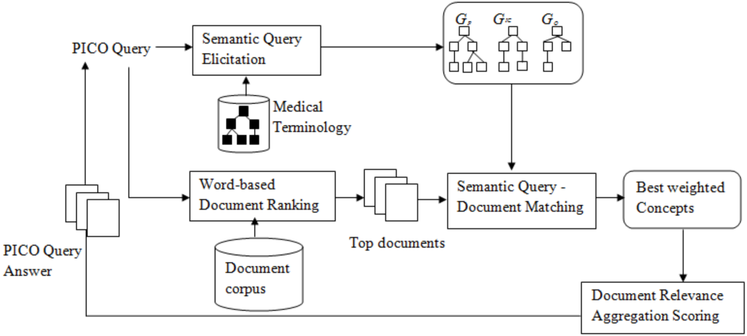
a dense space for
expected to produce

cularly for long
nt evidence to infer

QUERY/DOCUMENT EXPANSION

GLOBAL AND LOCAL CONTEXTS, ONE SPECIALIZED RESOURCE [ZNAIDI ET AL., 2016]

- **Goal:** answer PICO queries for Evidence Based Medicine
- **Context:** Concepts and relations between concepts, Top N retrieved documents
- **Knowledge-Base**: UMLS thesaurus



Query/document expansion

Global and local contexts, one specialized resource [Znaidi et al., 2016]

- **Goal:** answer PICO queries for Evidence Based Medicine
- **Context:** Concepts and relations between concepts, Top N retrieved documents
- **Knowledge-Base**: UMLS thesaurus

- **Key steps: tailored for PICO queries**
  - Map each PICO facet of query Q to UMLS
  - Identify the UMLS query concepts for each facet
  - Build a concept-based tree for each facet
  - Build the candidate query expansion terms using a recursive propagation algorithm from active concept *c* through sub-concepts *csub*
    - Compute $score(c) = +level(csub) * score(s)$
  - Expand the query with the N concepts having the highest score
  - Apply a prioritized-scoring based retrieval

$$RSV_{PICO}(Q,d) = \lambda_P * RSV_P(Q_P,d) + \lambda_{IC} * RSV_{IC}(Q_{IC},d) + +\lambda_O * RSV_O(Q_O,d)$$

$$\lambda_{IC} = 1, \lambda_P = \lambda_{IC} * RSV(Q^c_{IC},d), \lambda_O = \lambda_P * RSV(Q^c_P,d)$$

  - Combine the PICO score with a classic score (eg., BM25)

- **Mains findings**
  - Significant improvements (+30%) on a standard PICO collection
  - Slight but non significant improvement in comparison to PICO tailored models
  - The model performs better for relatively short queries (still be long in comparison to classic queries)
  - For long queries, it is more likely that the document matches the query according to different facets but with misleading interpretations of the search intent.

Query/document expansion

Global and local contexts, multiple specialized resource [Dinh and Tamine, 2012]

- **Goal:** document retrieval in the biomedical domain
- **Context:** Concepts and relations between concepts, Top N retrieved documents
- **Knowledge-Base**: MeSH, SNOMED, GO, ICD10 thesaurus,

- **Key steps**
  - Document expansion (indexing level)
    - Map each document to N resources and extract N lists of M top ranked candidate concepts
    - Apply a fusion strategy to select the best expansion concepts C
    - Expand the document with concepts in C

Stage 1 Conceptual Document Indexing

Input: Collection C, Terminologies T
Output: Index I

```
for all document D in C do
    # Mono-terminology extraction
    for all terminology T_i in T do
        R(D, T_i) ← extract(D, T_i);
    end for
    # Concept fusion
    R(D, T) ← ∪_{i=1}^{n} R(D, T_i);
    # Document expansion
    D' ← expand(D, R(D, T));
    # Document indexing
    I ← addIndex(D');
end for
return I;
```

Category	Technique	score(c_j, D)
Rank-based	CombRank	$\sum_{i=1}^{n}(\|R(D,T_i)\| - r_{ji}^{D})$
	CombRCP	$\sum_{i=1}^{n} 1/r_{ji}^{D}$
Score-based	CombSUM	$\sum_{i=1}^{n} w_{ji}^{D}$
	CombMIN	$\min\{w_{ji}^{D}, \forall i = \overline{1..n}\}$
	CombMAX	$\max\{w_{ji}^{D}, \forall i = \overline{1..n}\}$
	CombMED	$median\{w_{ji}^{D}, \forall i = \overline{1..n}\}$
	CombANZ	$\sum_{i=1}^{n} w_{ji}^{D} \div \|\{c_j \in R(D,T)\}\|$
	CombMNZ	$\sum_{i=1}^{n} w_{ji}^{D} \times \|\{c_j \in R(D,T)\}\|$

QUERY/DOCUMENT EXPANSION

GLOBAL AND LOCAL CONTEXTS, MULTIPLE SPECIALIZED RESOURCE [DINH AND TAMINE, 2012]

- **Goal:** document retrieval in the biomedical domain
- **Context:** Concepts and relations between concepts, Top N retrieved documents
- **Knowledge-Base**: MeSH, SNOMED, GO, ICD10 thesaurus,

- **Key steps**
  - Document expansion (indexing level)
    - Map each document to N resources and extract N lists of M top ranked candidate concepts
    - Apply a fusion strategy to select the best expansion concepts *C*
    - Expand the document with concepts in *C*
  - Query expansion
    - Perform an initial search with query
    - Build a document relevance model; use the Bose-Einstein statistics [Amati, 2003] to weight terms in the expanded query q^e derived from the original query Q:

      $$weight(t \in Q^e) = tfq_n + \beta * \frac{\text{Info}_{\text{Bo1}}}{\text{MaxInfo}}$$
  - Perform a post- retrieval with the expanded query

Stage 1 Conceptual Document Indexing

Input: Collection C, Terminologies T
Output: Index I

for all document D in C **do**
  # Mono-terminology extraction
  for all terminology T_i in T **do**
    $R(D, T_i) \leftarrow extract(D, T_i)$;
  end for
  # Concept fusion
  $R(D, T) \leftarrow \cup_{i=1}^{n} R(D, T_i)$;
  # Document expansion
  $D' \leftarrow expand(D, R(D, T))$;
  # Document indexing
  $I \leftarrow addIndex(D')$;
end for
return I;

Category	Technique	score(c_j, D)
Rank-based	CombRank	$\sum_{i=1}^{n}(\lVert R(D, T_i)\rVert - r_{ji}^{D})$
	CombRCP	$\sum_{i=1}^{n} 1/r_{ji}^{D}$
Score-based	CombSUM	$\sum_{i=1}^{n} w_{ji}^{D}$
	CombMIN	$\min\{w_{ji}^{D}, \forall i = \overline{1..n}\}$
	CombMAX	$\max\{w_{ji}^{D}, \forall i = \overline{1..n}\}$
	CombMED	$median\{w_{ji}^{D}, \forall i = \overline{1..n}\}$
	CombANZ	$\sum_{i=1}^{n} w_{ji}^{D} \div \lVert\{c_j \in R(D, T)\}\rVert$
	CombMNZ	$\sum_{i=1}^{n} w_{ji}^{D} \times \lVert\{c_j \in R(D, T)\}\rVert$

Query/document expansion

Global and local contexts, multiple specialized resource [Dinh and Tamine, 2012]

- **Context:** Concepts and relations between concepts, Top N retrieved documents
- **Knowledge-Base**: MeSH, SNOMED, GO, ICD10 thesaurus

- **Main results/findings**
  - Multi-terminology based retrieval significantly better than mono-terminology-based retrieval but with varying and moderate improvement rates according to the fusion strategy
  - Need raises toward the weighting of each terminology according to query specificities

	TREC Genomics 2004		*TREC Genomics 2005*	
Run	MAP	Δ (%)	MAP	Δ (%)
Median	0.2074		0.2173	
	Mono-terminology indexing and retrieval			
MeSH	0.4412†††	(+112.73)	0.2639	(+21.45)
SNOMED	0.4222†††	(+103.57)	0.2630	(+21.03)
ICD-10	0.4138†††	(+99.52)	0.2592	(+19.28)
GO	0.4408†††	(+112.54)	0.2536	(+16.71)
	Multi-terminology indexing and retrieval			
CombANZ	0.4435†††	(+113.84)	0.2647	(+20.89)
CombMAX	0.4387†††	(+111.52)	0.2684†	(+23.52)
CombMED	0.4459†††	(+115.00)	0.2683†	(+23.47)
CombMIN	0.4440†††	(+114.08)	**0.2685**†	(+23.56)
CombMNZ	**0.4529**†††	(+118.37)	0.2593	(+19.33)
CombRank	0.4407†††	(+112.49)	0.2594	(+19.37)
CombRCP	0.4371†††	(+110.75)	0.2601	(+19.70)
CombSUM	0.4470†††	(+115.53)	0.2601	(+19.70)

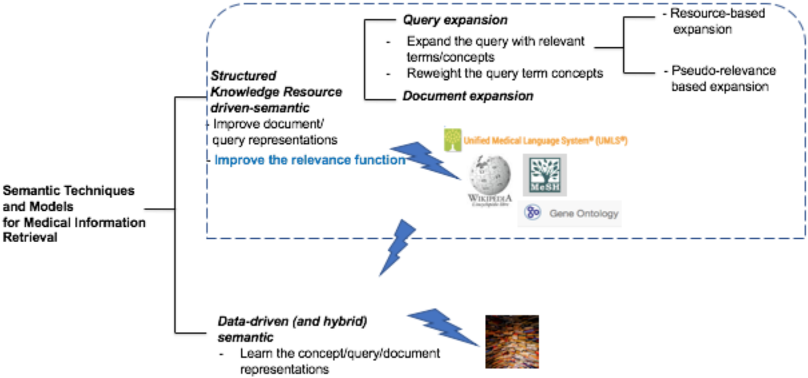
Significant changes at $p \leq 0.05, 0.01$ and 0.001 are denoted $^{\dagger}, ^{\dagger\dagger}$ and †††.

Query/document expansion

- A large body of state-of-the art work in medical IR rely on query expansion techniques
  - Combine statistical concepts and semantic concepts
  - Exploit generally evidence from multiple resources according to the polyrepresentation view
  - Use generally a local relevance model

- Robustness of query expansion techniques in the medical domain is questionable
  - They do not systematically lead to significant and important improvements over baselines
  - Several impacting factors on performance:
    - The knowledge bases used: general vs. specific
    - The semantic relations exploited: taxonomic vs. non taxonomic
    - The task at hand (collection): searching for cases, scientific literature, etc.

Roadmap



Semantic Techniques and Models for Medical Information Retrieval
Structured Knowledge Resource driven-semantic
Improve document/ query representations
Improve the relevance function
Query expansion
- Expand the query with relevant terms/concepts
- Reweight the query term concepts
Document expansion
- Resource-based expansion
- Pseudo-relevance based expansion
Unified Medical Language System® (UMLS®)
WIKIPEDIA
MeSH
Gene Ontology
Data-driven (and hybrid) semantic
- Learn the concept/query/document representations

Relevance ranking

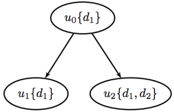
- How to incorporate semantics in the document relevance estimation?
  - Ranking as a semantic inference
    [Goodwin and Harabagiu, 2016, Koopman et al., 2016, Cao et al., 2011]
  - Ranking as learning the discriminant relevant (semantic) features
    [Balaneshin-kordan and Kotov, 2016, Xiong and Callan, 2015, Soldaini and Goharian, 2017]

Relevance ranking

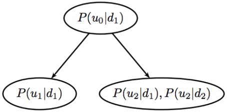
Ranking as a semantic inference: a graph-based approach [Koopman et al., 2016]

- **Goal:** Adhoc medical search
- **Key model components**
  - Graph-based representation of the documents
  - Document ranking as an inference process over related concepts in the graph
  - Knowledge resources with directed relationships between concepts
  - Different types of relationships
- **Key inference rationale:** tune the inference mechanism according to semantic gap issues: lexical mismatch, granularity mismatch, conceptual mismatch
  - Lexical mismatch (eg., *hypertension* vs. *high blood pressure*): association and deductive inference
  - Granularity mismatch (eg., *antipsychotic* and *Diazepman*): introduce uncertainty in the taxonomic (hierarchical eg., IS A) relationships
  - Conceptual mismatch (eg., *treatments* $\rightarrow$ *disease*): deductive inference and logical deduction

- The Graph-based corpus representation



(a) Basic node-document representation.



(b) Representation with initial probabilities assigned to node.

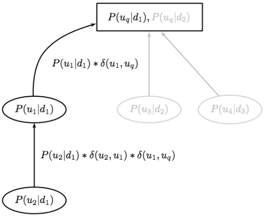
Relevance ranking

Ranking as a semantic inference: a graph-based approach [Koopman et al., 2016]

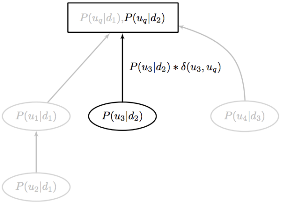
- **The retrieval model**
  - Strength of the association between two information nodes: compute recursively over the graph:

$$\sigma(u,u') = \begin{cases} 1 \; if \; u = u' \\ \sigma_0(u,u') \; if \; uRu' \\ argmax_{u_i \in U : uRu_i} \sigma(u,u_i) \times \sigma(u_i,u'), otherwise \end{cases} \quad (1)$$

  - Relevance of document-query $RSV(d,q) = \prod_{u_q \in q} \prod_{u_d \in d} p(u_d \mid d)\sigma(u_d,u_q)$
- **Main results/findings**
  - Effective improvement of queries suffering from the conceptual implication problem
  - Degradation for 'simple' queries do not requiring inference. Inference highlighted general irrelevant concepts



(a) Retrieval process for document d_1.



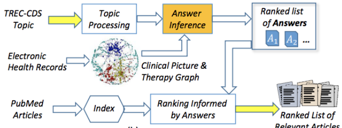
(b) Retrieval process for document d_2.

Relevance ranking

Ranking as a semantic inference: a probabilistic-based approach [Goodwin and Harabagiu, 2016]

- **Key model components**
  - Clinical decision support task through question-answering
  - Document ranking as an inference process over a bayesian graph
  - Medical knowledge-bases for document annotation
  - Full-text documents (EHRs, scientific articles) for background knowledge in the retrieval phase
- **Key inference rationale**: Consider the physician's belief value (eg., hypothetical, present, absent) in the inference process
  - Build the medical knowledge graph as a probabilistic graphical model: compute the probability distribution over all the possible clinical pictures and therapies of patients CPTG
  - Given a question topic t, the set of medical concepts and their assertions derived from the question is viewed as a hidden sketch of the clinical picture and therapy described in the question $Z(t)$ through concepts A with the same type
  - Answering a medical question associated with topic t consists in performing an inference process in the CPTG to determine the medical concepts having the highest likelihood given $Z(t)$

- The architecture of medical question-answering system for clinical decision-support



(b)

Relevance ranking

Ranking as learning the concept-document weighting features [Balaneshin-kordan and Kotov, 2016]

- **Goal:** Retrieve medical literature from clinical cases
- **Key model components**
  - Document ranking as learning the optimal query concept weighting based on the intuition that different concept types have different importance
  - A set of features characterizing query concept computed on the basis of knowledge-bases (Wikipedia, UMLS) as global context and top retrieved documents as local context of the query
- **The learning process**: A multivariate optimization method to train the weights of all features as contributions to fix the concept weight w.r.t. a document
- **The retrieval model**
  - Relevance status value based on the Markov Random Field retrieval framework [Metzler and Croft, 2005]: weighted linear combination of matching scores of concepts types in a query: $sc(Q, D) = \sum_{c \in C_T} \lambda_T(c) * f_T(c, D)$
  - Importance weight of concept c computed as a linear combination of importance features: $\lambda_T(c) \sum_{n=1}^{N} w_\phi^n \phi^n$
  - Consider global features (eg., popularity concept node in the UMLS graph) and local features (eg., Number of top retrieved documents containing concept c)
  - Learn $\lambda_T(c)$ using a multivariate optimization solving problem.
- **Main results/findings**
  - Reasonable performance improvement (5-9%) over the baselines according to search accuracy
  - Exploiting evidence from multiple resources (Wikipedia and UMLS) positively affects performance

Relevance ranking

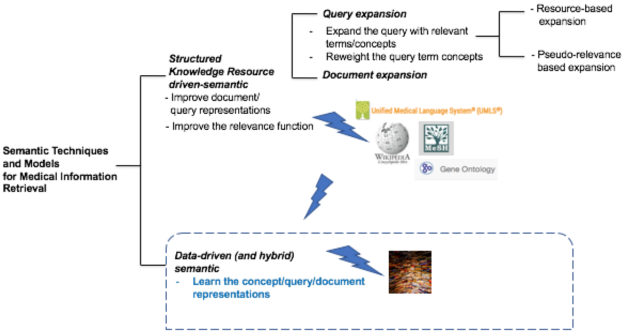
Let us discuss

- **A few work addressed the semantic search at the relevance function level**
  - Identify logical matching between words and concepts
  - Identify relevant semantic features that connect words to concepts, queries to documents

- **Findings: the general trend**
  - High-level inference yields to high computational complexity
  - The good balance between lexical matching and semantic matching is difficult to tune
  - Robustness to concept annotation quality is important

Outline

LEARNING



Semantic Techniques and Models for Medical Information Retrieval
Structured Knowledge Resource driven-semantic
- Improve document/ query representations
- Improve the relevance function
Query expansion
- Expand the query with relevant terms/concepts
- Reweight the query term concepts
Document expansion
- Resource-based expansion
- Pseudo-relevance based expansion
Unified Medical Language System® (UMLS®)
WIKIPEDIA
MeSH
Gene Ontology
Data-driven (and hybrid) semantic
- Learn the concept/query/document representations

Representation learning for medical search

Overview of early research

- What do the models learn? (Pre-BERT)
  - *Word, CUI, entity, phrase embeddings*: to learn high-level similarity between information units [De Vine et al., 2014, Limsopatham and Collier, 2016, Liu et al., 2016, Ghosh et al., 2017, Jagannatha and Yu, 2016, Cai et al., 2018, Henry et al., 2018]
  - *Document embeddings*: to improve semantic representations of texts that bridge the gap between data-driven semantic and knowledge resource driven semantic [Minarro-Gimnez et al., 2014, Nguyen et al., 2017, Loza Mencía et al., 2016, Peng et al., 2016, Banerjee et al., 2017, Nguyen et al., 2018]
  - *Medical objects embeddings*: care events/episodes, disease embeddings [Ghosh et al., 2016, Moen et al., 2015, Choi et al., 2016], patient embeddings [Baytas et al., 2017, Ni et al., 2017, Zhu et al., 2016, Stojanovic et al., 2017, Sushil et al., 2018]
- For which search tasks?
  - Relevance matching RM (eg., document retrieval, care-episode retrieval)
  - Semantic matching (eg., patient similarity)

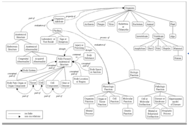
- Transformer-based IR:
  - Encoding of queries and documents [Jin et al., 2023, Ueda et al., 2021]
  - (Multiple) re-ranking phases [Pradeep et al., 2022, Biester et al., , Tahami et al.,]
  - Query/document classification [Wang et al., 2022]
  - Data augmentation/generation [Pradeep et al., 2022, Karimi, , Bondarenko et al., 2022]
  - Combinations of all the above

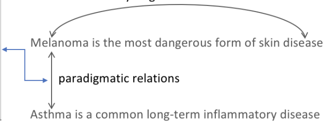
Representation learning for medical search

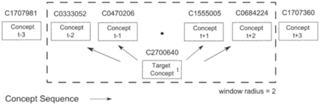
Learning Word, CUI, entity, phrase embeddings

- Different purposes yield to different objective functions
  - Learn readable concept/CUI/entity representations from raw texts: driven by paradigmatic relations provided in knowledge-bases
  - Learn concept and associated poly-senses: learn one vector representation per sense
- Different neural architectures
  - Extension of the CBOW and Skip-Gram models
  - Deep architectures (CNN, RNN, ...)
  - Transformer-based models

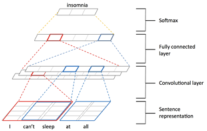
Medical thesaurus



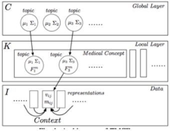




Devine et al. 2014



Limsopatham et al. 2016



Qian et et al. 2017

Representation learning for medical search

Learning Word, CUI, entity, phrase embeddings [De Vine et al., 2014, Liu et al., 2016]

- Learn word meaning representation from language models, in order to perform a concept similarity task - extension of the Skip-Gram model [De Vine et al., 2014]
  - Learn UMLS concept representations from sequences of concepts in annotated texts
  - Maximize the average log probability of the objective function $\frac{1}{2r}\sum_{i=1}^{2r}\sum_{-r\leq j\leq r} log(c_{t+j} \mid c_t)$
  - Valid representations when compared to human-assessments within a concept similarity task
  - Requires huge amount of annotated data.
  - Sensitivity to concept annotation quality?

- Extension of the CBOW model [Liu et al., 2016]
  - Learn concept representations constrained by relations established in a knowledge base
  - Maximize the log probability of the objective function $L = \sum_{i=1}^{T}(log(p(w_t \mid w_{t+k}) + \alpha \sum_{w_s:(w_t,w_s)\in R} wt(w_s \mid w_t)(logp(w_t \mid w_{t\pm k} - logp(w_s \mid w_{s\pm k})))^2)$
    $w_t(w_s \mid w_t) = \frac{f(w_s)}{\sum_{(w_t,w)\in R} f(w)}$
  - Experimental evaluation on IR tasks (query expansion) show: 1) sensitivity to model parameters and collections; 2) ability to select related words in the UMLS thesaurus; 3) slight improvement on a medical document search task

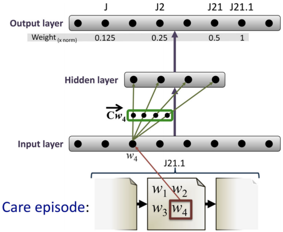
The most similar words to « heart »

CBOW		Online	
Cardiac	0.4891	Cardiac	0.5205
Synergist	0.4494	Hearts	0.5030
Hearts	0.4276	Cor	0.4939
Cardiovascular	0.4096	Synergist	0.4690
Acyanotic	0.3987	Cardiovascular	0.4156
Ouvrier	0.3934	Cerebrovascular	0.4149
Multiorgan	0.3931	Acyanotic	0.3985
Ventricular	0.3837	Ventricular	0.3979
Cardiorespiratory	0.3829	Cardiorespiratory	0.3969
Thrive	0.3766	Biventricular	0.3831

Representation learning for medical search

Learning care episodes, disease codes [Ghosh et al., 2016, Moen et al., 2015, Choi et al., 2016]

- Learning care episodes [Moen et al., 2015, Choi et al., 2016]
  - Introduce a longitudinal temporal view of medical documents (eg., care episode as successive clinical notes related to patient visits). Two types of relations:
    - Coocurrence of words: basic relation addressed in neural models
    - Sequential order of visits, codes
    - Inputs are high-level data (eg., EHR): computational complexity is questionable
    - Interpretability of the learned representations are highly required
  - Evaluation in a retrieval task [Moen et al., 2015]
    - Single care episod vectors: one vector representation per care episode based on average word vectors ; compute vector similarity between care episodes embeddings (eg., cosinus)
    - Average note vector similarity: one vector representation per care episode based on average note vectors; compute average pairwise (note) similarities, optimal pairing, sequence alignment



Representation learning for medical search

Learning care events/episodes, disease representations
[Ghosh et al., 2016, Moen et al., 2015, Choi et al., 2016]

- Learning disease: identify taxonomic and interpretable representations of diseases from raw data
- Extension of the Skip-Gram model [Choi et al., 2016]
  - **Input**: ordered sequence of medical codes extracted from patient visits (raw data of the patient visits)
  - **Output**: medical code representation, visit representation
  - **Learning objective**: two levels of learning functions unified in the same framework (yield to representations learned in a shared space)
    - Predict the medical codes given a visit: what happend in the past, what coul happen in the future?

$$min_{W_s, b_s} \frac{1}{T} \sum_{t=1}^{T} \sum_{-w \leq \leq w} -x_{t+1}^{T} log\hat{y}_t - (1 - x_{t+i}^{T} log(1 - \hat{y}_t)$$

    - Predict the medical codes given the code represenions in the same visit

$$min_{W'_c} \frac{1}{T} \sum_{t=1}^{T} \sum_{i:c_i \in V_t} \sum_{j:c_j \in V_t, j \neq i} log(c_j \mid c_i)$$

Representation learning for medical search

Learning care events/episodes, disease representations [Ghosh et al., 2016, Moen et al., 2015, Choi et al., 2016]

- Learning disease: identify taxonomic and interpretable representations of diseases from raw data
- Extension of the Skip-Gram model [Choi et al., 2016]
  - **Input**: ordered sequence of medical codes extracted from patient visits (raw data of the patient visits)
  - **Output**: medical code representation, visit representation
  - **Learning objective**: two levels of learning functions unified in the same framework (yield to representations learned in a shared space)
    - Predict the medical codes given a visit: what happend in the past, what coul happen in the future?

$$min_{W_s, b_s} \frac{1}{T} \sum_{t=1}^{T} \sum_{-w \leq \leq w} -x_{t+1}^{T} log \hat{y}_t - (1 - x_{t+i}^{T} log(1 - \hat{y}_t)$$

    - Predict the medical codes given the code represention in the same visit

$$min_{W_c'} \frac{1}{T} \sum_{t=1}^{T} \sum_{i:c_i \in V_t} \sum_{j:c_j \in V_t, j \neq i} log(c_j \mid c_i)$$

  - **Qualitative evaluation**
    - Interpretable visit representations validated by medical experts

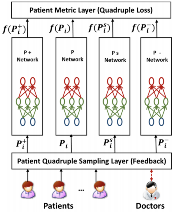
Coordinate 112	Coordinate 152	Coordinate 141
Kidney replaced by transplant (V42.0) Hb-SS disease without crisis (282.61) Heart replaced by transplant (V42.1) RBC antibody screening (P) Complications of transplanted bone marrow (996.85) Sickle-cell disease (282.60) Liver replaced by transplant (V42.7) Hb-SS disease with crisis (282.62) Prograf PO (R) Complications of transplanted heart (996.83)	X-ray, knee (P) X-ray, thoracolumbar (P) Accidents in public building (E849.6) Activities involving gymnastics (E005.2) Struck by objects/persons in sports (E917.0) Encounter for removal of sutures (V58.32) Struck by object in sports (E917.5) Unspecified fracture of ankle (824.8) Accidents occurring in place for recreation and sport (E849.4) Activities involving basketball (E007.6)	Cystic fibrosis (277.02) Intracranial injury (854.00) Persistent mental disorders (294.9) Subdural hemorrhage (432.1) Neurofibromatosis (237.71) Other conditions of brain (348.89) Conductive hearing loss (389.05) Unspecified causes of encephalitis, myelitis, encephalomyelitis (323.9) Sensorineural hearing loss (389.15) Intracerebral hemorrhage (431)
Sickle-cell disease and organ transplant	Sport-related injuries	Brain injuries and hearinh loss

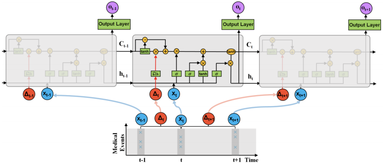
Representation learning for medical search

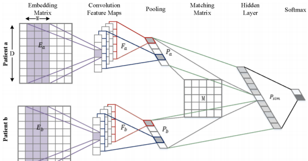
Learning patient profiles, patient similarity

[Baytas et al., 2017, Ni et al., 2017, Zhu et al., 2016, Stojanovic et al., 2017, Sushil et al., 2018]

- Two main objectives
  - Learn patient profile: input (EHR) - output (patient vector) [Baytas et al., 2017, Stojanovic et al., 2017, Sushil et al., 2018]
  - Learn end-to-end patient-patient similarity:
    - Input: EHR patient A, EHR patient B
    - Output: similarity class[Zhu et al., 2016, Ni et al., 2017]
- Input data
  - Heterogeneous patient data: demographic, medication, diagnosis codes etc.
  - Historical data: considering the sequence of medical events with irregular intervals
- Tasks
  - Predict patient mortality, primary diagnosis, length of stay, total incurred charges, ...







Representation learning for medical search

Learning patient profiles, patient similarity

[Baytas et al., 2017, Ni et al., 2017, Zhu et al., 2016, Stojanovic et al., 2017, Sushil et al., 2018]

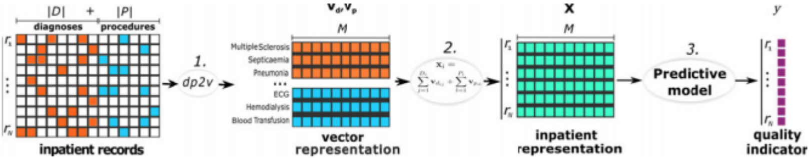
Learning patient vector from patient disease and procedure descriptions [Stojanovic et al., 2017]

- Learn the representations of diseases and procedures (dp2vec): extend the w2vec model by considering patient record as a "sentence" and diseases/procedures as "words"
- Build patient vector: sum the vector representations of in-patient diseases and procedures

Task evaluation

- Predict length of patient stay and total charges as a regression problem (vector representations used as features)
- Predict mortality as a classification problem

Neighbors of *respiratory failure*	Neighbors of *congestive heart failure*
Insertion of endotracheal tube	Insertion of implantable heart assist system
Tracheostomy toilette	Implantation of cardiac resynchronization defibrillator total system (CRT-D)
Other lavage of bronchus and trachea	Implantation of cardiac resynchronization defibrillator pulse generator (CRT-D)
Bronchoscopy through artificial stoma	Insertion of percutaneous external heart assist device
Other oxygen enrichment	Heart transplantation
Other repair and plastic operations on trachea	Excision destruction or exclusion of left atrial appendage (LAA)
Fiber-optic bronchoscopy	Aquapheresis
Infusion of vasopressor agent	Automatic implantable cardioverter-defibrillator (AICD) check
Replacement of tracheostomy tube	Noninvasive programmed electrical stimulation (NIPS)
Replacement of gastrostomy tube	Removal of lead(s) [electrode] without replacement
Complete glossectomy	Endovascular removal of obstruction from head and neck vessel(s)
Other intubation of respiratory tract	Replacement of automatic cardioverter-defibrillator lead(s) only



Representation Learning for Medical Search

Transformer-based models

- Encoding of queries and documents [Jin et al., 2023, Ueda et al., 2021]
- (Multiple) re-ranking phases [Pradeep et al., 2022, Biester et al., , Tahami et al.,]
- Query/document classification [Wang et al., 2022]
- Data augmentation/generation [Pradeep et al., 2022, Karimi, , Bondarenko et al., 2022]
- Combinations of all the above

Off-the-shelf medical models:

- Trained on text: ClinicalBERT, BioBERT, SciBERT, PubMedBERT,
- Including external knowledge: UMLSBERT, Clinical Kb-BERT
- In other languages!
- Check this amazing list: `https://mr-nlp.github.io/posts/2021/05/transformer-based-biomedical-pretrained-language-models-list/`
- (note that it is very complete but not up-to-date)

Representation Learning for Medical Search

Transformer-based models: Clinical Trials Matching [Pradeep et al., 2022]

- **Goal:** Given a patient and the patients electronic health record (EHR) as the query and a collection of actively recruiting clinical trials, return those that the patient is eligible for. (=*Clinical Matching*)
- **Context:** TREC2021 Clinical Trials Track

Their approach is two-fold:

1. Neural query synthesis (NQS): zero-shot document expansion model to generate sentence-long queries from patient descriptions:
   - doc2queryT5 model trained on MS MARCO V2 passage ranking
   - generated queries issued to BM25 + RM3 = first list of candidates
   - **+33% compared to using the EHR as the query**

Patient Description - #23: A 39-year-old man came to the clinic with cough and shortness of breath that was not relieved by his inhaler. · · ·

Query 1: causes for wheezing and shortness of breath
Query 2: what could be wrong when a chef has a cough and is short of breath all of a sudden
Query 3: how often should fluticasone be used for asthma
Query 4: what causes shortness of breath even with inhaler
⋮

Table 2: Examples of synthetic queries for the patient description in Table 1.

2. two-stage neural reranking pipeline trained on clinical trial matching data
   - neural sequence-to-sequence ranking model Med-Mono-T5
   - fine-tuned on CT dataset [Koopman and Zuccon, 2016]
   - model trained and used to find the best segments in each trial from the candidate list
   - final ranking based on the combination of segment's scores
   - **best automatic system at the TREC 2021 Clinical Trials Track**

Representation Learning for Medical Search

Transformer-based models: niomedical literature search [Ueda et al., 2021]

- **Goal:** Leverage biomedical abstract's structure to improve biomedical literature search
- Structure of an abstract: background, methods, results, conclusions
- standard cascading architecture for fine-tuning contextual language models:
  1. for each query q, retrieve initial k abstracts $a \in A$ to be re-ranked: $\mathcal{L} = log(1 + e^{-h_A(q,a^+,a^-)})$
  2. fine-tuning multiple pretrained SciBERT models concurrently: $a = (a|s \in \{B, M, R, C\})$, learn one model h_{A_s} per section
- Comparison of models on TREC PM and TREC COVID: locally fine-tuned features give consistent and significant improvements for the sum aggregator

Table 2: Retrieval effectiveness of different rankings leveraging globally (h_A) and locally ($h_{A_s} \mid s \in S = \{B, M, R, C\}$) fine-tuned SciBERT models as features. The symbols † and ‡ denote significant increases over the global model in the first row for $p < 0.05$ and $p < 0.01$, respectively.

	#	Features	Agg	NDCG	MAP	P@10
PM19	1	$\{h_A(q,a)\}$	–	0.530	0.206	0.456
	2	$\{h_A(q,a)\} \cup \{h_A(q,a_s) \mid s \in S\}$	sum	0.526	0.209	0.462
	3	$\{h_A(q,a)\} \cup \{h_{A_s}(q,a_s) \mid s \in S\}$	sum	**0.569**‡	**0.262**‡	**0.562**‡
	4	$\{h_A(q,a)\} \cup \{h_A(q,a_s) \mid s \in S\}$	max	0.517	0.198	0.428
	5	$\{h_A(q,a)\} \cup \{h_{A_s}(q,a_s) \mid s \in S\}$	max	0.533	0.217	0.465
COVID	1	$\{h_A(q,a)\}$	–	0.467	0.160	0.402
	2	$\{h_A(q,a)\} \cup \{h_A(q,a_s) \mid s \in S\}$	sum	0.486†	0.180‡	0.484‡
	3	$\{h_A(q,a)\} \cup \{h_{A_s}(q,a_s) \mid s \in S\}$	sum	**0.492**‡	**0.192**‡	**0.516**‡
	4	$\{h_A(q,a)\} \cup \{h_A(q,a_s) \mid s \in S\}$	max	0.475	0.169	0.454
	5	$\{h_A(q,a)\} \cup \{h_{A_s}(q,a_s) \mid s \in S\}$	max	0.482†	0.175‡	0.438

Table 3: Impact in retrieval performance on PM19 when removing individual features from our summed feature aggregation. Symbols † and ‡ denote significant decreases from the baseline (first row) for $p < 0.05$ and $p < 0.01$, respectively.

Set of features	NDCG	MAP	P@10
Using all sections	**0.569**	**0.262**	**0.562**
(-) Original Abstract	0.561 (-0.008)‡	0.249 (-0.013)‡	0.532 (-0.030)‡
(-) Background	0.565 (-0.004)	0.254 (-0.008)†	0.547 (-0.015)
(-) Methods	0.564 (-0.005)†	0.252 (-0.010)‡	0.557 (-0.005)
(-) Results	0.563 (-0.006)‡	0.253 (-0.009)‡	0.551 (-0.011)
(-) Conclusions	0.564 (-0.005)‡	0.251 (-0.011)‡	0.540 (-0.022)†

Representation learning for medical search

Let us discuss

- In summary
  - Big trend towards the use of neural models in medical search
  - Learned representations reusable in a wide range of search tasks and prediction tasks
  - Background knowledge (eg., Knowledge-base, expert's assessments) driven learning increases the readability of the representations and the explicability of the learning outcomes

- Pending issues
  - What are the impacting factors? What works vs. fails in the black box?
  - Lack of sufficient amount of labeled data to learn accurate representations (eg., patient similarity, IR tasks)
  - Performance sensitivity to a large size of network parameters, hyper-parameters and models parameters
  - Transfert learning, data augmentation/generation
  - Performance variability across medical tasks

Outline

OUTLINE

Challenges in Evaluating Medical Information Retrieval

How is Information Retrieval Evaluated

Evaluating an IR system consists in checking how it satisfies an information need

What satisfies a human is different from what satisfies a system!

Two levels of evaluation can be distinguished:

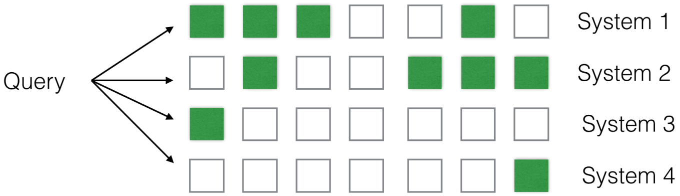
- Evaluation at the document level: does this document satisfy the information need?
- Evaluation at the system level: does this system helped in satisfying the information need? I.e. does it retrieve one or more document(s) satisfying the information need?

Efficient and effective system

- Efficiency: time and space
  - Pre-development specifications
  - Easy to measure
- Effectiveness: good results
  - What is a good result?

Challenges in Evaluating Medical Information Retrieval

Evaluation at the system level



Which system is the best one?

- Looking for the highest number of relevant documents? → Systems 1 and 2
- Looking for a single document giving the answer? → Systems 1, 2, 3 and 4
- Looking at the ranking of the documents? → Systems 1 and 3

Evaluation Paradigms

Laboratory-based: the Cranfield paradigm

- Testing and comparing search systems requires a laboratory environment that doesnt change
- Initiated by Cyril Cleverdon in the Cranfield College of Aeronautics called *the Cranfield Tests*
- Retrieval experiments conducted on test databases in a controlled setting

User studies

- Measuring user satisfaction with feedback from *real users*
- Also allows to measure a system's usability
- Mostly done in Interactive Information Retrieval
- [Kelly et al., 2009]

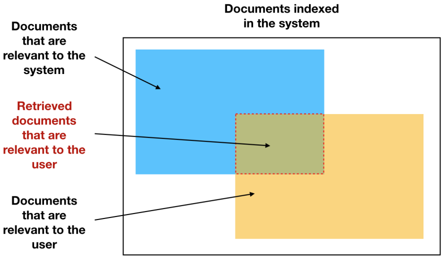
Log-based/click based

- When there is no space/time for laboratory evaluation, e.g. online systems
- A/B testing: redirecting n% of the trafic on a new version of the system
- Click analysis: inference on document relevance/usefulness

Challenges in Evaluating Medical Information Retrieval

Evaluation at the system level

It's all a matter of precision and recall...

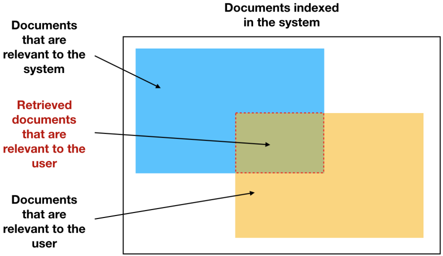


$$Precision = \frac{|P \cap R|}{|R|}, Recall = \frac{|P \cap R|}{|P|}$$

Challenges in Evaluating Medical Information Retrieval

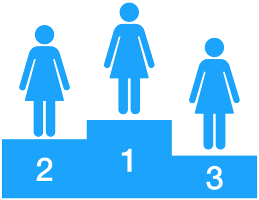
Evaluation at the system level

It's all a matter of precision and recall...



$$Precision = \frac{|P \cap R|}{|R|}, Recall = \frac{|P \cap R|}{|P|}$$

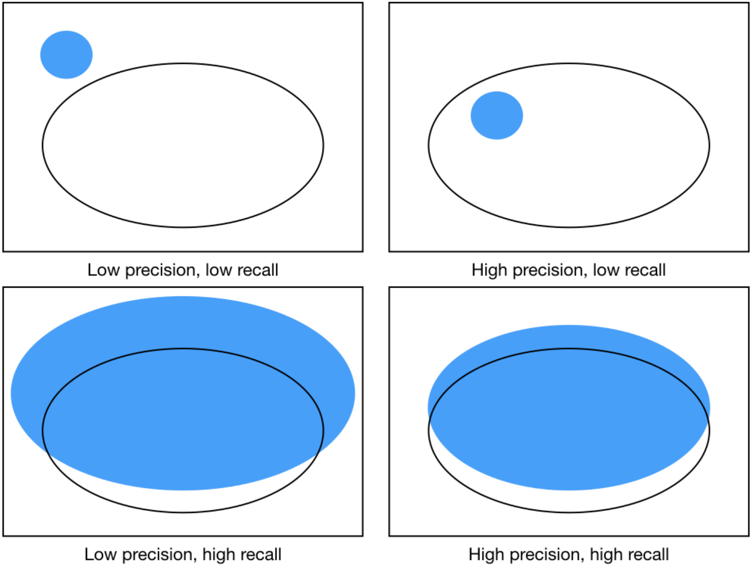
... And of rank!



Unless they are looking for the entire set of documents, nobody goes through the entire set of results.
Ranked metrics:

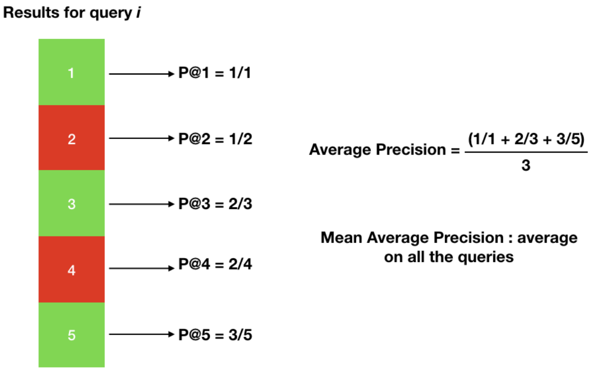
- P@N
- Mean Average Precision (MAP) [Voorhees, 1998]
- Normative Discounted Cumulation Gain [Jarvelin and Kekalainen, 2000]

Balancing precision and recall



P@N and Mean Average Precision

- Precision @r: computes the precision at a certain rank
- Mean Average Precision:
  - Average precision over all relevant documents
  - Rewards systems retrieving relevant documents quickly



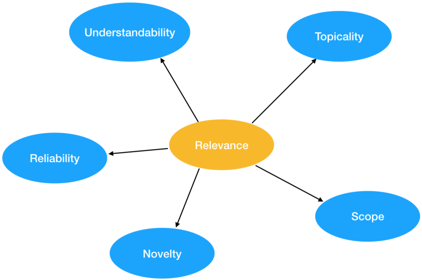
Challenges in Evaluating Medical Information Retrieval

Evaluation at the document level

In classical IR

- From the user point-of-view: a relevant document answer the initial information need
- From the system point-of-view: A relevant document covers the same topic as the query, i.e. contains the query's terms

Relevance has many other dimensions [Zhang et al., 2014]



In the medical domain:

- For patients:
  - Documents must be readable and understandable for a given user
  - The information contained in the documents should be trustworthy
- For medical professionals:
  - Documents must contain up-to-date information
  - Documents must properly cover the topic searched

Challenges in Evaluating Medical Information Retrieval

Each search task has its proper objectives:

- How should the retrieval and the ranking be implemented?
- How should the system be evaluated?

Examples:

- Physician adhoc search: priority given to the rank, P@10, the topicality, scope...
- Patient adhoc search: priority given to the rank, P@10, the topicality, understandability, readability...
- Clinical trials: priority given to the rank, the topicality, the scope, the novelty...
- Systematic reviews: priority given to the recall, the topicality, the scope...

Outline

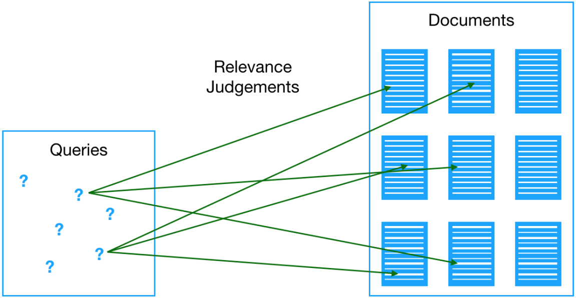
Evaluation Challenges

What is a benchmark?

- Comparing 2 search systems results on a common dataset allows to compare their effectiveness.
- These common datasets are called *benchmarks*.

An IR benchmark contains:

- A document collection that can be indexed
- A set of topics (enriched queries)
- Relevance judgements (linking queries to the relevant documents in the collection)



Evaluation Challenges

The Cranfield Paradigm

Given:

1. A test collection (T, D, R)
2. A retrieval run for the test collection : a doc-list L_t for each topic t in T

For each topic t in T

- Use a measure (e.g. P@10) to compute the quality of L_t

Combine scores:

- Mean average precision

Relevance judgement:

- For a given topic $t \in T$, a given document $d \in D$, $R(d, t)$ is the relevance score of d for topic t.
- $R(d, t)$ can be:
  - a discrete value: e.g. $\in 0, 1$ for binary assessment or $\in 0, 1, 2, 3$ for graded assessment
  - a continuous value: e.g. $\in [0, 1]$
- Assumption: if $R(d, t, u_1)$ is the judgement of assessor u_1 on topic t and document d and $R(d, t, u_2)$ the judgement of assessor u_2 on topic t and document d, $R(d, t, u_1) = R(d, t, u_2)$

Evaluation Challenges

Main Organizers

There are many forums organizing challenges and benchmarking activities, and most of them have some medical tracks:

- **Text REtrieval Conference (TREC)**: organized by the US National Institute of Standards and Technology (NIST). Provides since 1992 numerous evaluation challenges and a forum to discuss the results.
  - `http://trec.nist.gov`
- **Conference and Labs of the Evaluation Forum (CLEF)**: European (smaller) version of TREC, organized by the CLEF initiative. Organizes since 2000 evaluation challenges along with a conference.
  - `http://www.clef-initiative.eu/`
- **NTCIR**: organized by the Japanese National Institute of Informatics (NII). Provides since 1998 numerous evaluation challenges and a forum to discuss the results.
  - `http://ntcir.nii.ac.jp/`
- **Forum for Information Retrieval Evaluation (FIRE)**: South Asian counterpart, organized in India. Provides since 2008 numerous evaluation challenges and a forum to discuss the results.
  - `http://fire.irsi.res.in/`

Summary of the Benchmarking Activities/datasets

Venue	Task	Dataset	Activity
TREC	Genomics adhoc retrieval	Clinical information need Biomedical articles	Terminated
	Genomics passage retrieval	Clinical information need Biomedical articles	Terminated
	Medical records	Patient cohort search	Terminated
	Clinical decision support / Precision medicine	Case reports Biomedical articles	Ongoing
	Health Misinformation	Yes-no questions Web pages	Ongoing
	TREC COVID-19	Round-based biomedical literature Search on COVID-19 topics	Terminated
CLEF	ImageCLEF medical retrieval	Image and medical reports Collection of medical images	Terminated
	CLEF eHealth consumer search	Health information need Large web crawl	Terminated
	CLEF eHealth technological assisted reviews	Boolean queries Biomedical articles	Terminated
*	BioASQ	Annotated biomedical abstracts and QA dataset	Ongoing

IR dataset containing PICO queries: `https://github.com/boudinfl/CLIREC`
TRIP Click: Log files of a biomedical search engine [Rekabsaz et al., 2021]
Clinical trial matching test collection [Koopman and Zuccon, 2016]
The majority of these datasets are still available and can be used for research!

Outline

Conclusion

- *A large and growing body of work on Information Retrieval in the medical domain*
  - Focus on task, user profile, information need elicitation in context (time, task, user's expertise, etc.)
  - Model semantic w.r.t. polyrepresentation view: document collections, knowledge bases, users, etc.
  - Shift from lexical matching to semantic matching by considering domain-specific peculiarities
  - Understand relevance assessment facets according to task, user (laypeople vs.expert)
  - Increasing amount of tasks due to a higher collaboration with the medical community
- *Challenges ahead*
  - IR at the service of the medical community:
    - Domain-driven IR models for medical search: Are IR heuristics similar to medical search heuristics?
    - Complex tasks not solved yet
    - Effort in structuring knowledge: dynamic terminologies, languages other than English...
    - Conceptualization of medical search tasks: eg., model hypothetico-deductive approach of medical experts?
  - Neural networks for all
    - Need for big data
    - Black-box: towards explainability
  - Large-scale evaluation under privacy-constraints

Discussion

References I

Ai, Q., Yang, L., Guo, J., and Croft, W. B. (2006).
Maxmatcher: biological concept extraction using approximate dictionary lookup.
In *PRICAI'06*, pages 1145–1149.

Amati, G. (2003).
Probabilistic models for Information Retrieval based on Divergence from Randomness.
PhD thesis, University of Glasgow.

Ando, R. K., Dredze, M., and Zhang, T. (2005).
Trec 2005 genomics track experiments at ibm watson.
In *Proceedings of TREC 2005*.

Aronson, A. R. and Lang, F.-M. (2010).
An overview of metamap: historical perspective and recent advances.
Journal of the American Medical Informatics Association, 17(3):229–236.

Balaneshin-kordan, S. and Kotov, A. (2016).
Optimization method for weighting explicit and latent concepts in clinical decision support queries.
In *Proceedings of the 2016 ACM International Conference on the Theory of Information Retrieval*, ICTIR '16, pages 241–250, New York, NY, USA. ACM.

Balaneshinkordan, S. and Kotov, A. (2019).
Bayesian approach to incorporating different types of biomedical knowledge bases into information retrieval systems for clinical decision support in precision medicine.
Journal of Biomedical Informatics, 98:103238.

Banerjee, I., C. Chen, M., Lungren, M., and Rubin, D. (2017).
Radiology report annotation using intelligent word embeddings: Applied to multi-institutional chest ct cohort.
Journal of Biomedical Informatics, 77.

References II

Bast, H., Björn, B., and Haussmann, E. (2016).
Semantic search on text and knowledge bases.
Found. Trends Inf. Retr., 10(2-3):119–271.

Baytas, I. M., Xiao, C., Zhang, X., Wang, F., Jain, A. K., and Zhou, J. (2017).
Patient subtyping via time-aware lstm networks.
In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '17, pages 65–74.

Biester, L., Joopudi, V., and Dandala, B.
Ibm@ trec clinical trials track 2021.

Blackman, S. (2006).
Nature has wikipedia in its cities.
The scientist, 20(2).

Bondarenko, A., Fröbe, M., Gienapp, L., Pugachev, A., Reimer, J. H., Schlatt, F., Artemova, E., Potthast, M., Stein, B., Braslavski, P., et al. (2022).
Webis at trec 2022: Deep learning and health misinformation.
In *31st Text REtrieval Conference. TREC*, pages 500–338.

Borlund, P. (2003).
The concept of relevance in ir.
J. Am. Soc. Inf. Sci. Technol., 54(10):913–925.

Boudin, F., Nie, J.-Y., and Dawes, M. (2010).
Clinical information retrieval using document and pico structure.
In *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, HLT '10, pages 822–830, Stroudsburg, PA, USA. Association for Computational Linguistics.

References III

Broder, A. (2002).
A taxonomy of web search.
SIGIR Forum, 36(2):3–10.

Cai, X., Gao, J., Ngiam, K. Y., Ooi, B. C., Zhang, Y., and Yuan, X. (2018).
Medical concept embedding with time-aware attention.
In *IJCAI*.

Cao, Y., Liu, F., Simpson, P., Antieau, L., Bennett, A., Cimino, J. J., Ely, J., and Yu, H. (2011).
Askhermes: An online question answering system for complex clinical questions.
Journal of Biomedical Informatics, 44(2):277 – 288.

Cartright, M.-A., White, R. W., and Horvitz, E. (2011a).
Intentions and attention in exploratory health search.
In *Proceedings of the 34th International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '11, pages 65–74, New York, NY, USA. ACM.

Cartright, M.-A., White, R. W., and Horvitz, E. (2011b).
Intentions and attention in exploratory health search.
In *Proceedings of the 34th International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '11, pages 65–74.

Chapman, W., Bridewell W, H. P., GF, C., and BG, B. (2001).
Evaluation of negation phrases in narrative clinical reports.
In *Proceedings of the AMIA Symposium*, pages 105–109.

Choi, E., Bahadori, M. T., Searles, E., Coffey, C., Thompson, M., Bost, J., Tejedor-Sojo, J., and Sun, J. (2016).
Multi-layer representation learning for medical concepts.
In *Proceedings of the 22Nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '16, pages 1495–1504.

References IV

De Choudhury, M., Morris, M. R., and White, R. W. (2014).
Seeking and sharing health information online: Comparing search engines and social media.
In *Proceedings of the 32Nd Annual ACM Conference on Human Factors in Computing Systems*, CHI '14, pages 1365–1376.

De Vine, L., Zuccon, G., Koopman, B., Sitbon, L., and Bruza, P. (2014).
Medical semantic similarity with a neural language model.
In *Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management*, CIKM '14, pages 1819–1822, New York, NY, USA. ACM.

Del Fiol, G., Workman, T. E., and Gorman, P. N. (2014).
Clinical questions raised by clinicians at the point of care: a systematic review.
JAMA internal medicine, 174(5):710–718.

Demner-Fushman, D., Humphrey, S. M., Ide, N. C., Loane, R. F., Ruch, P., Ruiz, M. E., Smith, L. H., Tanabe, L. K., Wilbur, W. J., and Aronson, A. R. (2006).
Finding relevant passages in scientific articles: Fusion of automatic approaches vs. an interactive team effort.
In *Proceedings of TREC 2006*.

Dinh, D. and Tamine, L. (2012).
Towards a context sensitive approach to searching information based on domain specific knowledge sources.
Web Semantics: Science, Services and Agents on the World Wide Web, 12(0).

Dinh, D., Tamine, L., and Boubekeur, F. (2013).
Factors affecting the effectiveness of biomedical document indexing and retrieval based on terminologies.
Artificial Intelligence in Medicine, 57(2):155 – 167.

Edinger, T., Cohen, A., Bedrick, S., Ambert, K., and Hersh, W. (2012).
Barriers to retrieving patient information from electronic health record data: failure analysis from the trec medical records track.
AMIA ... Annual Symposium proceedings / AMIA Symposium. AMIA Symposium, 2012:180–188.

References V

Ghosh, S., Chakraborty, P., Cohn, E., Brownstein, J. S., and Ramakrishnan, N. (2016).
Characterizing diseases from unstructured text: A vocabulary driven word2vec approach.
In *Proceedings of the 25th ACM International on Conference on Information and Knowledge Management*, CIKM '16, pages 1129–1138.

Ghosh, S., Chakraborty, P., Cohn, E., Brownstein, J. S., and Ramakrishnan, N. (2017).
Joint learning of representations of medical concepts and words from ehr data.
In *Proceedings (IEEE Int Conf Bioinformatics Biomed)*, Biomed '17, pages 764–769.

Goeuriot, L., Suominen, H., Pasi, G., Bassani, E., Brew-Sam, N., Gonzalez Saez, G., Kelly, L., Mulhem, P., Seneviratne, S., Upadhyay, R., et al. (2021).
Consumer health search at clef ehealth 2021.
In *CEUR Workshop Proceedings: CLEF 2021–Conference and Labs of the Evaluation Forum, September 21–24, 2021, Bucharest, Romania*. CEUR Workshop Proceedings.

Goodwin, T. R. and Harabagiu, S. M. (2016).
Medical question answering for clinical decision support.
In *Proceedings of the 25th ACM International on Conference on Information and Knowledge Management*, CIKM '16, pages 297–306, New York, NY, USA. ACM.

Guo, J., Fan, Y., Ai, Q., and Croft, W. B. (2016).
A deep relevance matching model for ad-hoc retrieval.
In *Proceedings of the 25th ACM International on Conference on Information and Knowledge Management*, CIKM '16, pages 55–64, New York, NY, USA. ACM.

Henry, S., Cuffy, C., and McInnes, B. T. (2018).
Vector representations of multi-word terms for semantic relatedness.
Journal of Biomedical Informatics, 77:111 – 119.

Hersh, W. (2010).
Information Retrieval - A Health and Biomedical Perspective.
Springer.

References VI

Hersh, W. and Voorhees, E. (2009).
Trec genomics special issue overview.
12:1–15.

Huang, X., Lin, J., and Demner-Fushman, D. (2006).
Evaluation of pico as a knowledge representation for clinical questions.
In *AMIA Annual Symposium Proceedings*, pages 359–363.

Huang, X., Zhong, M., and Si, L. (2005).
York university at trec 2005: Genomics track.
In *Proceedings of TREC 2005*.

Jagannatha, A. and Yu, H. (2016).
Structured prediction models for RNN based sequence labeling in clinical text.
In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pages 856–865.

Jarvelin, K. and Kekalainen, J. (2000).
IR evaluation methods for retrieving highly relevant documents.
In *Proceedings of the 23rd SIGIR conference*, pages 41–48.

Jin, Q., Shin, A., and Lu, Z. (2023).
Lader: Log-augmented dense retrieval for biomedical literature search.
arXiv preprint arXiv:2304.04590.

Jones, K. S., Walker, S., and Robertson, S. E. (2000).
A probabilistic model of information retrieval: Development and comparative experiments.
Inf. Process. Manage., 36(6):779–808.

REFERENCES VII

Kanoulas, E., Li, D., Azzopardi, L., and Spijker, R. (2017).
Clef 2017 technologically assisted reviews in empirical medicine overview.
In *Proceedings of the Conference and Labs of the Evaluation Forum (CLEF) - Online working notes*.

Karimi, V. N. M. R. S.
Matching a patient from an admission note to clinical trials: Experiments with query generation and neural-ranking.

Kelly, D. et al. (2009).
Methods for evaluating interactive information retrieval systems with users.
Foundations and Trends® in Information Retrieval, 3(1–2):1–224.

Koopman, B. and Zuccon, G. (2014).
Why assessing relevance in medical ir is demanding.
In *Proceedings of the Medical IR workshop, in conjunction with the 2014 Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, Gold Coast, Australia. ACM.

Koopman, B. and Zuccon, G. (2016).
A test collection for matching patients to clinical trials.
In *Proceedings of the 39th International ACM SIGIR conference on Research and Development in Information Retrieval*, pages 669–672.

Koopman, B., Zuccon, G., Bruza, P., Sitbon, L., and Lawley, M. (2016).
Information retrieval as semantic inference: A graph inference model applied to medical search.
Inf. Retr., 19(1-2):6–37.

Kraljevic, Z., Searle, T., Shek, A., Roguski, L., Noor, K., Bean, D., Mascio, A., Zhu, L., Folarin, A. A., Roberts, A., et al. (2021).
Multi-domain clinical natural language processing with medcat: the medical concept annotation toolkit.
Artificial intelligence in medicine, 117:102083.

References VIII

Limsopatham, N. and Collier, N. (2016).
Normalising medical concepts in social media texts by learning semantic representation.
In *Proceedings of the International Conference on the Association for Computational Linguistics*, ACL '16, pages 1014–1023.

Limsopatham, N., Macdonald, C., and Ounis, I. (2013).
Learning to handle negated language in medical records search.
In *Proceedings of the 22Nd ACM International Conference on Information & Knowledge Management*, CIKM '13, pages 1431–1440, New York, NY, USA. ACM.

Liu, X. and Croft, W. B. (2004).
Cluster-based retrieval using language models.
In *Proceedings of the 27th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '04, pages 186–193, New York, NY, USA. ACM.

Liu, X., Nie, J.-Y., and Sordoni, A. (2016).
Constraining word embeddings by prior knowledge – application to medical information retrieval.
pages 155–167.

Loza Mencía, E., de Melo, G., and Nam, J. (2016).
Medical concept embeddings via labeled background corpora.
In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016)*, pages 4629–4636.

Martinez, D., Otegi, A., Soroa, A., and Agirre, E. (2014).
Improving search over electronic health records using umls-based query expansion through random walks.
Journal of Biomedical Informatics, 51:100 – 106.

Metzler, D. and Croft, W. B. (2005).
A markov random field model for term dependencies.
In *Proceedings of the 28th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '05, pages 472–479, New York, NY, USA. ACM.

References IX

Minarro-Gimnez, J., O, M.-A., and M, S. (2014).
Exploring the application of deep learning techniques on medical text corpora.
In *Studies in health technology and in informatics*, pages 205:584–8.

Moen, H., Ginter, F., Marsi, E., Peltonen, L.-M., T, S., and S., S. (2015).
Care episode retrieval: distributional semantic models for information retrieval in the clinical domain.
BMC Medical Informatics and Decision Making, 15.

Montori, V. M., Wilczynski, N. L., Morgan, D., and Haynes, R. B. (2004).
Systematic reviews: a cross-sectional study of location and citation count.
BMC Medicine, 1(2).

Natarajan, K., Stein, D., Jain, S., and Elhadad, N. (2010).
An analysis of clinical queries in an electronic health record search utility.
International Journal of Medical Informatics, 79(7):515–522.

Nguyen, G.-H., Tamine, L., Soulier, L., and Souf, N. (2017).
Learning Concept-Driven Document Embeddings for Medical Information Search.
In *Conference on Artificial Intelligence in Medicine (AIME 2017)*, Vienna, Austria.

Nguyen, G.-H., Tamine, L., Soulier, L., and Souf, N. (2018).
A Tri-Partite Neural Document Language Model for Semantic Information Retrieval.
In *ESWC 2018 - 15th European Semantic Web Conference*, volume 10843 of *Lecture Notes in Computer Science*, pages 445–461, Heraklion, Crète, Greece. Springer.

Ni, J., Liu, J., Zhang, C., Ye, D., and Ma, Z. (2017).
Fine-grained patient similarity measuring using deep metric learning.
In *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*, CIKM '17, pages 1189–1198.

References X

Oh, H.-S. and Jung, Y. (2015).
Cluster-based query expansion using external collections in medical information retrieval.
Journal of Biomedical Informatics, 58:70 – 79.

Palotti, J., Hanbury, A., Müller, H., and Kahn, C. E. (2016a).
How users search and what they search for in the medical domain: Understanding laypeople and experts through query logs.
Information Retrieval Journal, 19:189–224.

Palotti, J., Hanbury, A., Müller, H., and Kahn, Jr., C. E. (2016b).
How users search and what they search for in the medical domain.
Inf. Retr., 19(1-2):189–224.

Peng, S., You, R., Wang, H., Zhai, C., Mamitsuka, H., and Shanfeng Zhu1, . (2016).
Deepmesh: deep semantic representation for improving large-scale mesh indexing.
Bioinformatics, 32.

Ponte, J. M. and Croft, W. B. (1998).
A language modeling approach to information retrieval.
In *Proceedings of the 21st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '98, pages 275–281.

Pradeep, R., Li, Y., Wang, Y., and Lin, J. (2022).
Neural query synthesis and domain-specific ranking templates for multi-stage clinical trial matching.
In *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 2325–2330.

Rekabsaz, N., Lesota, O., Schedl, M., Brassey, J., and Eickhoff, C. (2021).
Tripclick: the log files of a large health web search engine.
In *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 2507–2513.

References XI

Rindflesch, T. C. and Fiszman, M. (2003).
The interaction of domain knowledge and linguistic structure in natural language processing: interpreting hypernymic propositions in biomedical text.
Journal of biomedical informatics, 36(6):462–477.

Roberts, K., Demner-Fushman, D., Voorhees, E. M., Hersh, W. R., Bedrick, S., Lazar, A. J., and Pant, S. (2017).
Overview of the trec 2017 precision medicine track.
In *Proceedings of the Text REtrieval Conference*.

Roberts, K., Simpson, M., Demner-Fushman, D., Voorhees, E., and Hersh, W. (2015a).
State-of-the art in biomedical literature rerieval for clinical cases: a survey of the trec 2014 cds track.
Information retrieval.

Roberts, K., Simpson, M. S., Voorhees, E., and Hersh, W. R. (2015b).
Overview of the trec 2015 clinical decision support track.
In *Proceedings of the Text REtrieval Conference*.

Salton, G. and McGill, M. J. (1986).
Introduction to Modern Information Retrieval.
McGraw-Hill, Inc., New York, NY, USA.

Salton, G., Wong, A., and Yang, C. S. (1975).
A vector space model for automatic indexing.
Commun. ACM, 18(11):613–620.

Savova, G. K., Masanz, J. J., Ogren, P. V., Zheng, J., Sohn, S., Kipper-Schuler, K. C., and Chute, C. G. (2010).
Mayo clinical text analysis and knowledge extraction system (ctakes): architecture, component evaluation and applications.
Journal of the American Medical Informatics Association, 17(5):507–513.

References XII

Scullard, P., Peacock, C., and Davies, P. (2010).
Googling children's health: reliability of medical advice on the internet.
Archives of disease in childhood, 95(8):580–582.

Shen, W., Nie, J.-Y., Liu, X., and Liui, X. (2014).
An investigation of the effectiveness of concept-based approach in medical information retrieval grium@ clef2014ehealthtask 3.
Proceedings of CLEF 2014 (Working Notes).

Smith, D. A. (2020).
Situating wikipedia as a health information resource in various contexts: A scoping review.
PloS one, 15(2):e0228786.

Soldaini, L. and Goharian, N. (2016).
Quickumls: a fast, unsupervised approach for medical concept extraction.
In *MedIR workshop, sigir*.

Soldaini, L. and Goharian, N. (2017).
Learning to rank for consumer health search: A semantic approach.
In Jose, J. M., Hauff, C., Altıngovde, I. S., Song, D., Albakour, D., Watt, S., and Tait, J., editors, *Advances in Information Retrieval*, pages 640–646, Cham. Springer International Publishing.

Soldaini, L., Yates, A., and Goharian, N. (2017).
Learning to reformulate long queries for clinical decision support.
Journal of the Association for Information Science and Technology, 68(11):2602–2619.

Soldaini, L., Yates, A., Yom-Tov, E., Frieder, O., and Goharian, N. (2016).
Enhancing web search in the medical domain via query clarification.
Inf. Retr., 19(1-2):149–173.

References XIII

Sondhi, P., Sun, J., Zhai, C., Sorrentino, R., and Kohn, M. (2012).
Leveraging medical thesauri and physician feedback for improving medical literature retrieval for case queries.
Journal of the American Medical Informatics Association: JAMIA, 19(5):851–858.

Stanton, I., Ieong, S., and Mishra, N. (2014a).
Circumlocution in diagnostic medical queries.
In *Proceedings of the 37th international ACM SIGIR conference on Research & development in information retrieval*, pages 133–142. ACM.

Stanton, I., Ieong, S., and Mishra, N. (2014b).
Circumlocution in diagnostic medical queries.
In *Proceedings of the 37th International ACM SIGIR Conference on Research & Development in Information Retrieval*, SIGIR '14, pages 133–142, New York, NY, USA. ACM.

Stojanovic, J., Gligorijevic, D., Radosavljevic, V., Djuric, N., Grbovic, M., and Obradovic, Z. (2017).
Modeling healthcare quality via compact representations of electronic health records.
IEEE/ACM Trans. Comput. Biol. Bioinformatics, 14(3):545–554.

Sushil, M., uster, S. A., Luyckx, K., and Daelemans, W. (2018).
Patient representation learning and interpretable evaluation using clinical notes.
Journal of Biomedical Informatics, 84:103 – 113.

Tahami, A. V., Zhang, D., and Smucker, M. D.
Uwaterloomds at the trec 2022 health misinformation track.

Tamine, L. and Chouquet, C. (2017).
On the impact of domain expertise on query formulation, relevance assessment and retrieval performance in clinical settings.
Inf. Process. Manage., 53(2):332–350.

References XIV

Tamine, L., Chouquet, C., and Palmer, T. (2015).
Analysis of biomedical and health queries: Lessons learned from trec and clef evaluation benchmarks.
JASIST, 66(12):2626–2642.

Ueda, A., Santos, R. L., Macdonald, C., and Ounis, I. (2021).
Structured fine-tuning of contextual embeddings for effective biomedical retrieval.
In *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 2031–2035.

Voorhees, E. (1998).
Variations in relevance judgments and the measurement of retrieval effectiveness.
In *Proceedings of the 21st SIGIR conference*, pages 315–323.

W. White, R. and Horvitz, E. (2009).
Cyberchondria: Studies of the escalation of medical concerns in web search.
ACM Trans. Inf. Syst., 27.

Wang, C. and Akella, R. (2015).
Concept-based relevance models for medical and semantic information retrieval.
In *Proceedings of the 24th ACM International on Conference on Information and Knowledge Management*, CIKM '15, pages 173–182, New York, NY, USA. ACM.

Wang, Y., Wang, S., Li, Y., and Dou, D. (2022).
Recognizing medical search query intent by few-shot learning.
In *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 502–512.

Weerkamp, W., Balog, K., and de Rijke, M. (2012).
Exploiting external collections for query expansion.
ACM Trans. Web, 6(4):18:1–18:29.

REFERENCES XV

White, R. and Horvitz, E. (2010).
Web to world: predicting transitions from self-diagnosis to the pursuit of local medical assistance in web search.
In *AMIA Annu Symp Proc*, pages 882–6.

White, R. and Horvitz, E. (2014).
From health search to healthcare: explorations of intention and utilization via query logs and user surveys.
J Am Med Inform Assoc, 21(1):4955.

White, R. W. and Horvitz, E. (2013).
Captions and biases in diagnostic search.
ACM Trans. Web, 7(4):23:1–23:28.

Xiong, C. and Callan, J. (2015).
Esdrank: Connecting query and documents through external semi-structured data.
In *Proceedings of the 24th ACM International on Conference on Information and Knowledge Management*, CIKM '15, pages 951–960, New York, NY, USA. ACM.

Xu, B., Lin, H., and Lin, Y. (2019).
Learning to refine expansion terms for biomedical information retrieval using semantic resources.
IEEE/ACM Trans. Comput. Biol. Bioinformatics, 16(3).

Zhai, C. and Lafferty, J. (2001).
Model-based feedback in the language modeling approach to information retrieval.
In *Proceedings of the Tenth International Conference on Information and Knowledge Management*, CIKM '01, pages 403–410, New York, NY, USA. ACM.

Zhang, Y., Wang, P., Heaton, A., and Winkler, H. (2012).
Health information searching behavior in medlineplus and the impact of tasks.
In *Proceedings of the 2nd ACM SIGHIT International Health Informatics Symposium*, IHI'12, pages 641–650.

References XVI

Zhang, Y., Zhang, J., Lease, M., and Gwizdka, J. (2014).
Multidimensional relevance modeling via psychometrics and crowdsourcing.
In *Proceedings of the 37th International ACM SIGIR Conference on Research & Development in Information Retrieval*, SIGIR '14, pages 435–444, New York, NY, USA. ACM.

Zhu, D. and Carterette, B. (2012).
Improving health records search using multiple query expansion collections.
In *2012 IEEE International Conference on Bioinformatics and Biomedicine*, pages 1–7.

Zhu, D., Wu, S. T., Masanz, J. J., Carterette, B., and Liu, H. (2013).
Using discharge summaries to improve information retrieval in clinical domain.
In *Proceedings of CLEF 2013 (Working Notes)*.

Zhu, Z., Yin, C., Qian, B., Cheng, Y., Wei, J., and Wang, F. (2016).
Measuring patient similarities via a deep architecture with medical concept embedding.
In *2016 IEEE 16th International Conference on Data Mining (ICDM)*, pages 749–758.

Znaidi, E., Tamine, L., and Latiri, C. (2015).
Answering pico clinical questions: A semantic graph-based approach.
In Holmes, J. H., Bellazzi, R., Sacchi, L., and Peek, N., editors, *Artificial Intelligence in Medicine*, pages 232–237, Cham. Springer International Publishing.

Znaidi, E., Tamine, L., and Latiri, C. (2016).
Aggregating semantic information nuggets for answering clinical queries.
In *Proceedings of the 31st Annual ACM Symposium on Applied Computing, Pisa, Italy, April 4-8, 2016*, pages 1041–1047.

Zuccon, G. and Koopman, B. (2018).
Choices in knowledge-base retrieval for consumer health search.
In *Proceedings of ECIR'2018*. LNCS.

References XVII

Zuccon, G., Koopman, B., and Palotti, J. (2015).
Diagnose this if you can: On the effectiveness of search engines in finding medical self-diagnosis information.
In *Advances in Information Retrieval: 37th European Conference on IR Research, ECIR 2015, Vienna, Austria, March 29-April 2, 2015. Proceedings 37*, pages 562–567. Springer.