Health Search

From Consumers to Clinicians

Slides available at

https://ielab.io/russir2018-health-search-tutorial/

Guido Zuccon
Queensland University of Technology

@guidozuc

QUT

ie
lab

Outline

- Dealing with the **semantic gap**: exploiting the semantics of medical language
    - concept based search & inference, query expansion, learning to rank
- Dealing with the nuances of **medical language**
    - negation, family history, understandability
- Understanding and aiding **query formulation**
    - query variations, query reformulation, query clarification, query suggestion, query intent, query difficulty, task-based solutions

Dealing with the semantic gap

Exploiting semantics of medical language

- What are medical concepts, where are they defined
- Why use concepts
- Why concepts and terms

Medical concepts

- Medical concepts are defined in domain knowledge resource
- Capture the key aspects of the domain or some specific sub-domain
- Relationships between concepts capture associations

Implicit VS Explicit Semantics

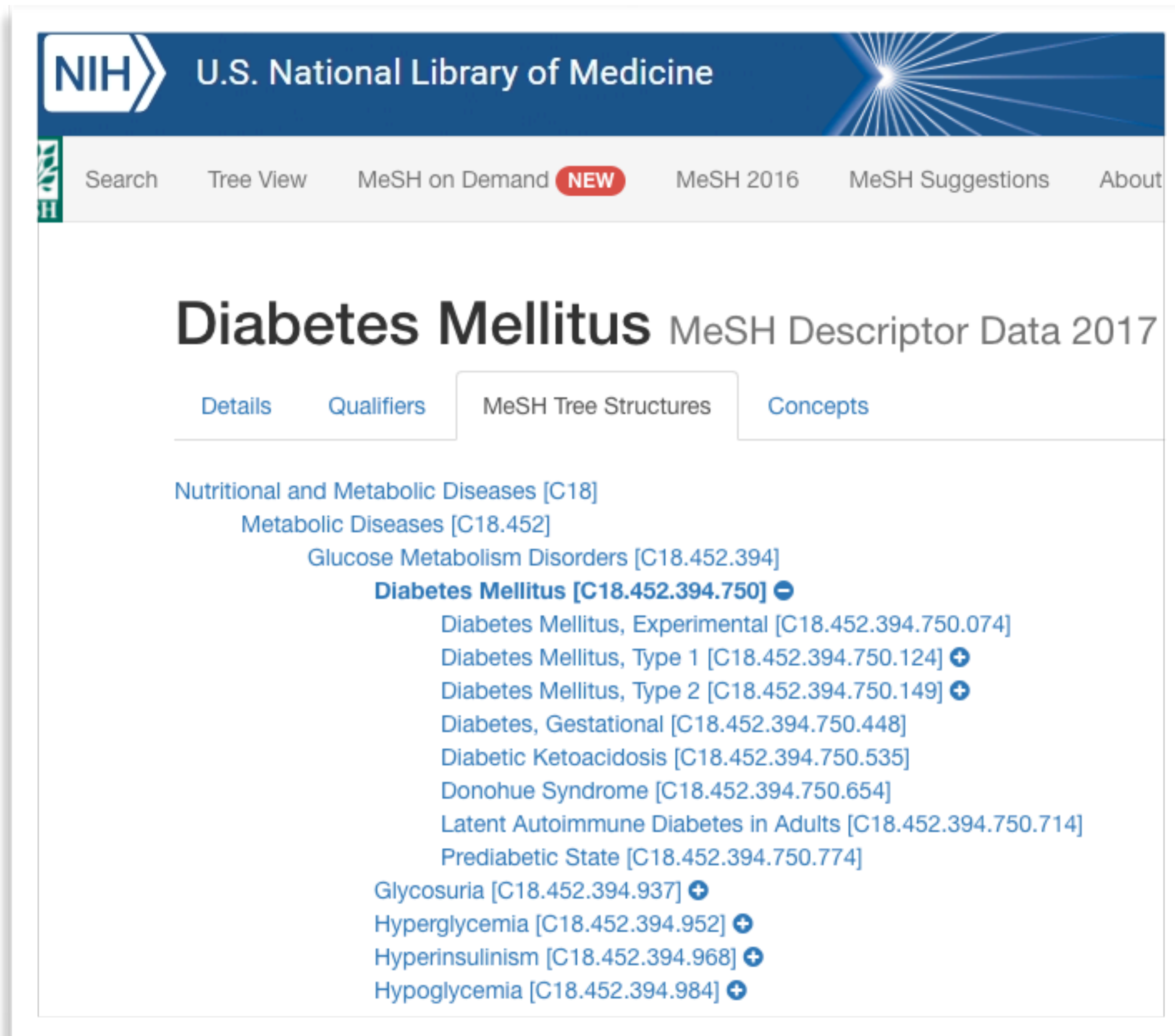
- Explicit semantics: structured human representation of knowledge and its concepts
    - e.g., medical terminologies
- Implicit Semantics: draw representation of words/concepts from data
    - e.g., distributional/latent semantic models

Key Medical Terminologies

Medical Subject Headings (MeSH)

Controlled vocabulary for **indexing journal articles**

Mainly used by researchers and clinicians searching the literature.



SNOMED CT

Formal medical ontology: ~500,000 concepts ~3,000,000 relationships

Bec … data.

Add … ven …

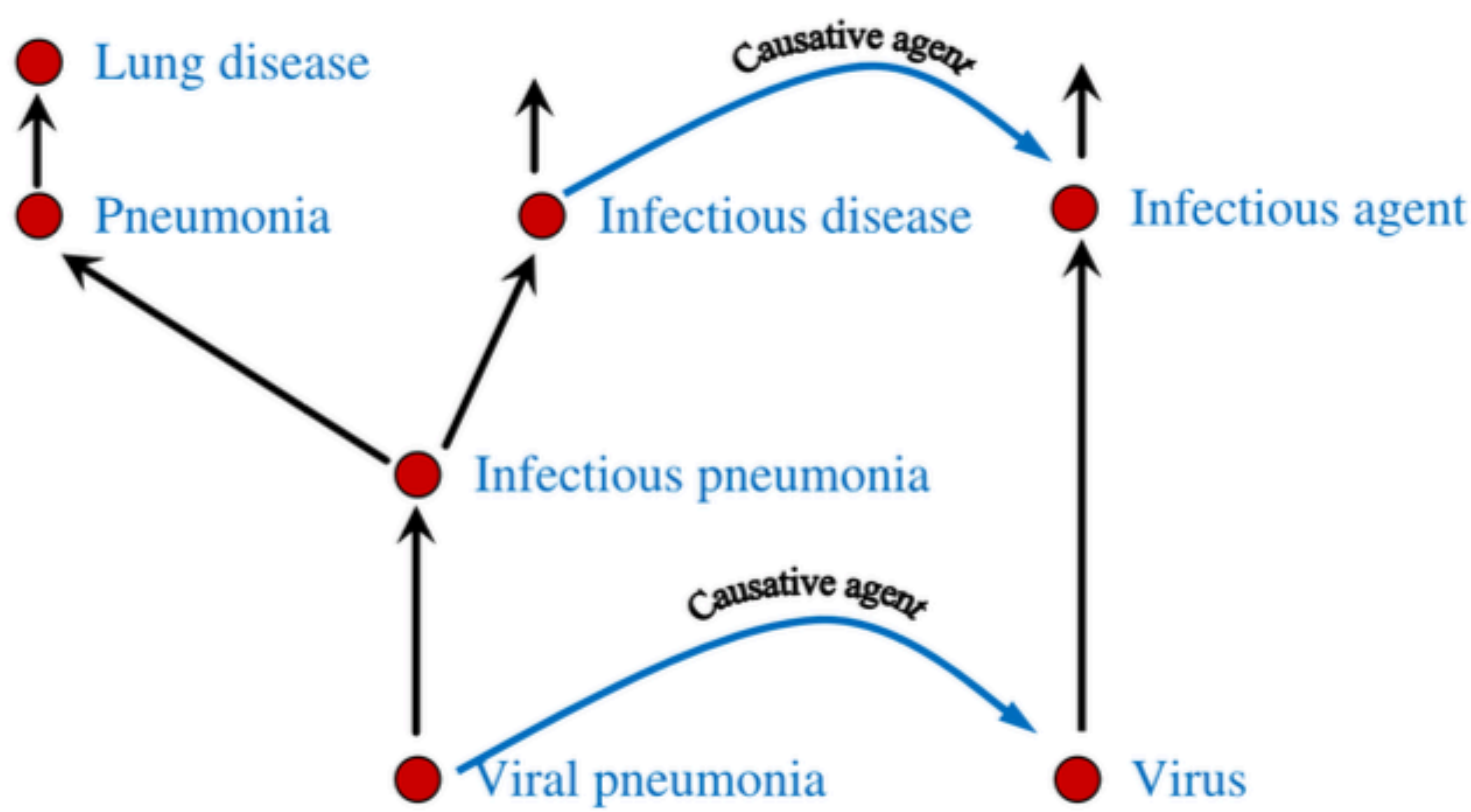


Figure 3.3: Concept hierarchy for *Viral pneumonia.*

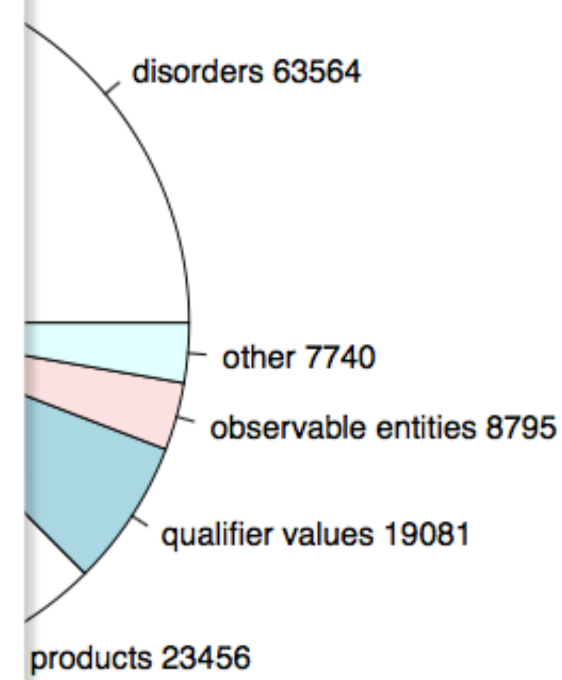


Figure 3.2: Breakdown of concept categories in the SNOMED CT ontology.

ICD

International Statistical Classification of Diseases and Related Health Problems (ICD)

Diagnosis classification from World Health Organisation

Used extensively in **billing**

International Statistical Classification of Diseases and Related Health Problems 10th Revision

Chapter	Blocks	Title
I	A00–B99	Certain infectious and parasitic diseases
II	C00–D48	Neoplasms
III	D50–D89	Diseases of the blood and blood-forming organs and certain disorders involving the immune mechanism
IV	E00–E90	Endocrine, nutritional and metabolic diseases
V	F00–F99	Mental and behavioural disorders
VI	G00–G99	Diseases of the nervous system
VII	H00–H59	Diseases of the eye and adnexa
VIII	H60–H95	Diseases of the ear and mastoid process
IX	I00–I99	Diseases of the circulatory system
X	J00–J99	Diseases of the respiratory system
XI	K00–K93	Diseases of the digestive system
XII	L00–L99	Diseases of the skin and subcutaneous tissue
	M00–	Diseases of the musculoskeletal system

Unified Medical Language System (UMLS)

- UMLS is a compendium of many controlled vocabularies in the biomedical sciences
- **Combined many terminologies under one umbrella**
- UMLS concept grouped into higher level semantic types
  - Concept: *Myocardial Infarction* [C0027051] of type *Disease or Syndrome* [T047]
  - https://uts.nlm.nih.gov//metathesaurus.html

An important note

- These resources contain information that can help characterise medical language
    - Synonyms of a term
    - Relationship between terms/concepts
- Rarely do these resources contain information that directly answers questions like
    - What is the drug of choice for condition x?
    - What is the cause of symptom x?
    - What test is indicated in situation x?
    - How should I treat condition x (not limited to drug treatment)?
    - How should I manage condition x (not specifying diagnostic or therapeutic)?
    - What is the cause of physical finding x?
    - What is the cause of test finding x?
    - Can drug x cause (adverse) finding y?
    - Could this patient have condition x?
- That is, they **do not directly resolve the clinical questions** presented in [Ely et al., 2000] taxonomy
- They capture truisms/**universal facts**, not subjective knowledge/things that could change over time

Convert Terms to Concepts

(aka Concept Mapping)

"metastatic breast cancer"

"metastatic"
"breast"
"cancer"

Concept Id:
60278488
(Breast Cancer
Metastatic)

Term Encapsulation

"human immunodeficiency virus"
"T-lymphotropic virus"
"HIV"
"AIDS"

86406008
(*Human immunodeficiency virus infection*)

Conflating Term-variants

"esophageal reflux"

235595009 Gastroesophageal reflux
196600005 Acid reflux or oesophagitis
47268002 Reflux
249496004 Esophageal reflux finding

Concept Expansion

[Aronson&Lang, 2010]

Concept extraction/mapping tools

- **Metamap** — National Library of Medicine [Aronson&Lang, 2010]
    - Extensive configuration option; but: default options tuned for biomedical literature, not necessarily websites or clinical text
    - Can be slow and unstable
- **QuickUMLS** [Soldaini&Goharian, 2016]
    - Modern computationally efficient mapper
    - Shown in the hands-on session
- **SemRep** — to extract relations between concepts [Rindflesch&Fiszman, 2003]
    - <subject, object, relation> from 27.9M PubMed articles stored into SemMedDB: https://skr3.nlm.nih.gov/SemMedDB/
- Others exist: cTakes [Savova et al., 2010], Ontoserver [McBride et al., 2012], etc.

Concept Mapping as an IR problem

"…the patient had $\underbrace{\text{headaches}}$ and was home…"

Issue the query "headaches" to IR system

Select top ranking concept

Ranked list of concepts

25064002
162307009
162308004
…

CONCEPT ID:	25064002
FULLY SPECIFIED NAME:	Headache (finding)
SYNONYMS:	HA - Headache Headache Cephalalgia Head pain Pain in head Cephalodynia Cephalgia
PREFERRED TERMS:	GB English : Headache US English : Headache

System	RR	S@1	S@5	S@10
Metamap	0.3015	0.2032	0.4354	0.5941
Ontoserver	0.6315	0.5323	0.7576	0.8111
TF-IDF	0.3959*	0.2967*	0.5069*	0.5920
BM25	0.3925*	0.2953*	0.5048*	0.5852
JMLM	0.3691*	0.2747*	0.4766	0.5714
DLM	0.2914	0.1848	0.4059	0.5227*

(when retrieval methods are able to generate at least one mapping)

[Mirhosseini et al., 2014]

Practical - part 1

- In this hands-on session, we will:
  1. Take a collection of clinical trials, annotate them with medical concepts, producing documents with both term and concept representation.

- In part 2, we will use these results to:
  2. Index these documents in Elasticsearch with multi term/concepts fields.
  3. Search Elaticsearch with either term or concept, demonstrating semantic search capabilities.
  4. Play a bit more (maybe)

- Instructions: https://ielab.io/russir2018-health-search-tutorial/hands-on/

Implicit Medical Concept Representations: Word Embeddings

- [Pyysalo et al., 2013]: word2vec and random indexing on very large corpus of biomedical scientific literature. http://bio.nlplab.org
- [De Vine et al., 2014]: word2vec on medical journal abstracts (embedding for UMLS)
  - Learns embedding of a concept, from co-occurrence with concepts
- [Zuccon et al., 2015, b]: word2vec on TREC Medical Records Track. http://zuccon.net/ntlm.html
- [Choi et al., 2016]: word2vec on medical claims (embedding for ICD), clinical narratives (embedding for UMLS) https://github.com/clinicalml/embeddings
- [Beam et al., 2018]: cui2vec (variation of word2vec) on 60M insurance claims + 20M health records + 1.7M full text biomedical articles. https://figshare.com/s/00d69861786cd0156d81
- Nuances of medical word embeddings:
  - [Chiu et al., 2016]: bigger corpora do not necessarily produce better biomedical word embeddings

Concept-based IR

Two types for Concept-based Retrieval

- Concept Augmented Term-based Retrieval
  e.g. [Ravindran&Gauch, 2004]
  - Maintain the original term representation of documents.
  - Use a concept-based approach to improve the query representation.
- Pure Concept-based Retrieval
  - Map the terms in documents to higher-level concepts
  - Retrieval is then done in ‘concept space’ rather than ‘term space’
  - SAPHIRE system [Hersh&Hickam, 1995]
  - Language modelling concepts [Meij et al., 2010]

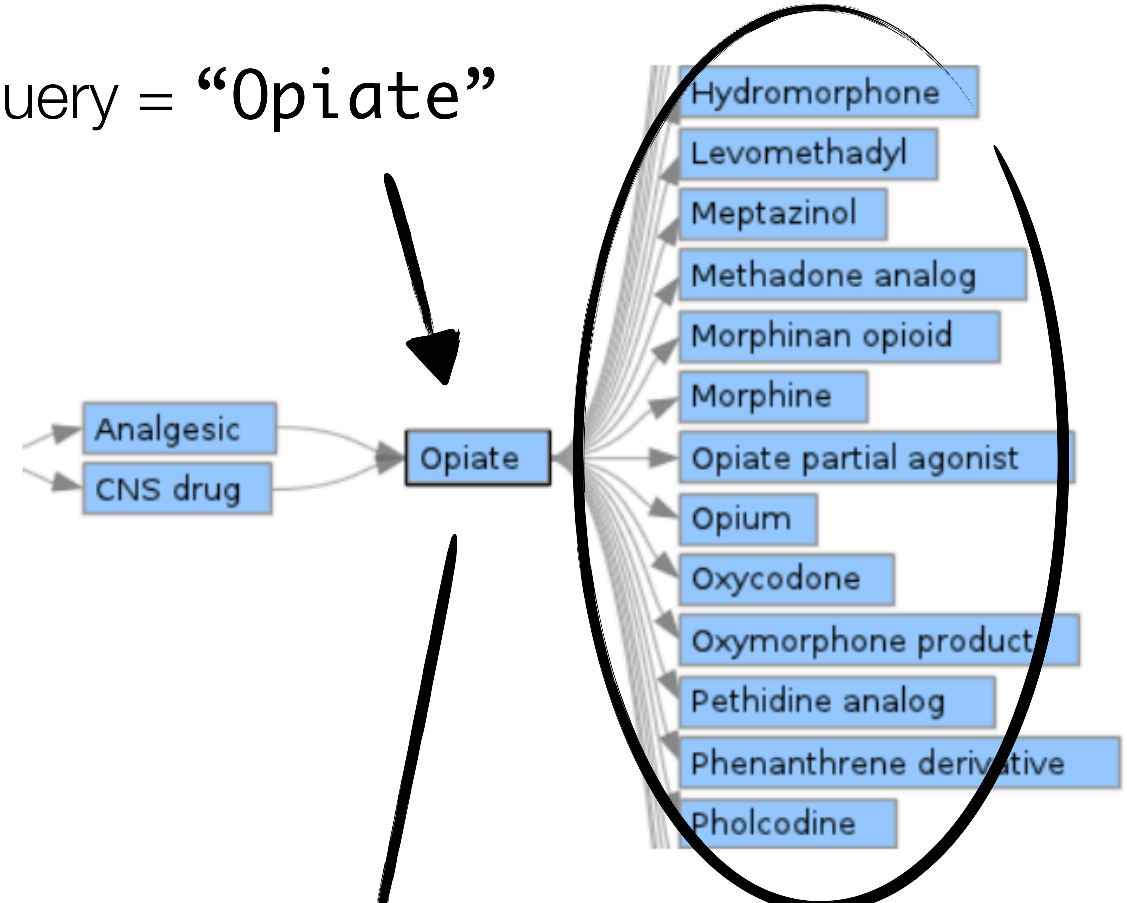
Combining Text and Concept Representations

[Limsopatham et al., 2013c]: learning framework that combines bag-of-words and bag-of-concepts representations on per-query basis

1. Linear combination model for merging scores from the two representations
2. Features: QPPs for both representations
3. Regression to infer model parameters (Gradient Boosted Regression Trees)

Exploiting concept hierarchies

Query = “Opiate”



$$RSV(d|q) = \underbrace{\sum_{c_i \in q} w(c_i, d)}_{\text{Base query concept}} + \underbrace{\sum_{c_j \prec c_i ; c_i \in q} \delta(w(c_j, d))}_{\text{Subsumed query concepts}}$$

[Zuccon et al., 2012]

Semantic Inference for IR

Concept-based retrieval that exploits ontology relationships

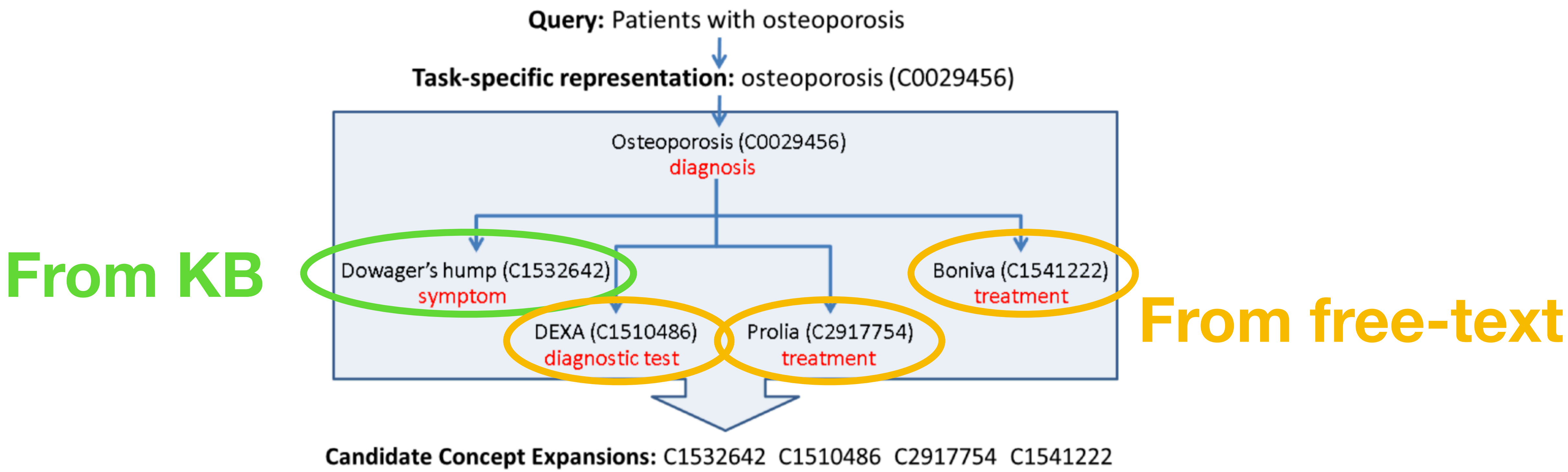
- Inferring conceptual relationships [Limsopatham et al., 2013]
- Information Retrieval as Semantic Inference [Koopman et al., 2016]
  - both: expand queries by inferring additional conceptual relationships from KB, but in different ways
  - [Limsopatham et al., 2013] also infers relationships
    - from collection of medical free-text, and
    - via PRF

> "This is a 62-year-old gentleman who has Type 1 DM and is on hemodialysis. He is currently taking Avapro"

- Hemodialysis ✔
- DM? Diabetes mellitus?
- Avapro? Hypertension!

Inferring conceptual relationships
[Limsopatham et al., 2013]

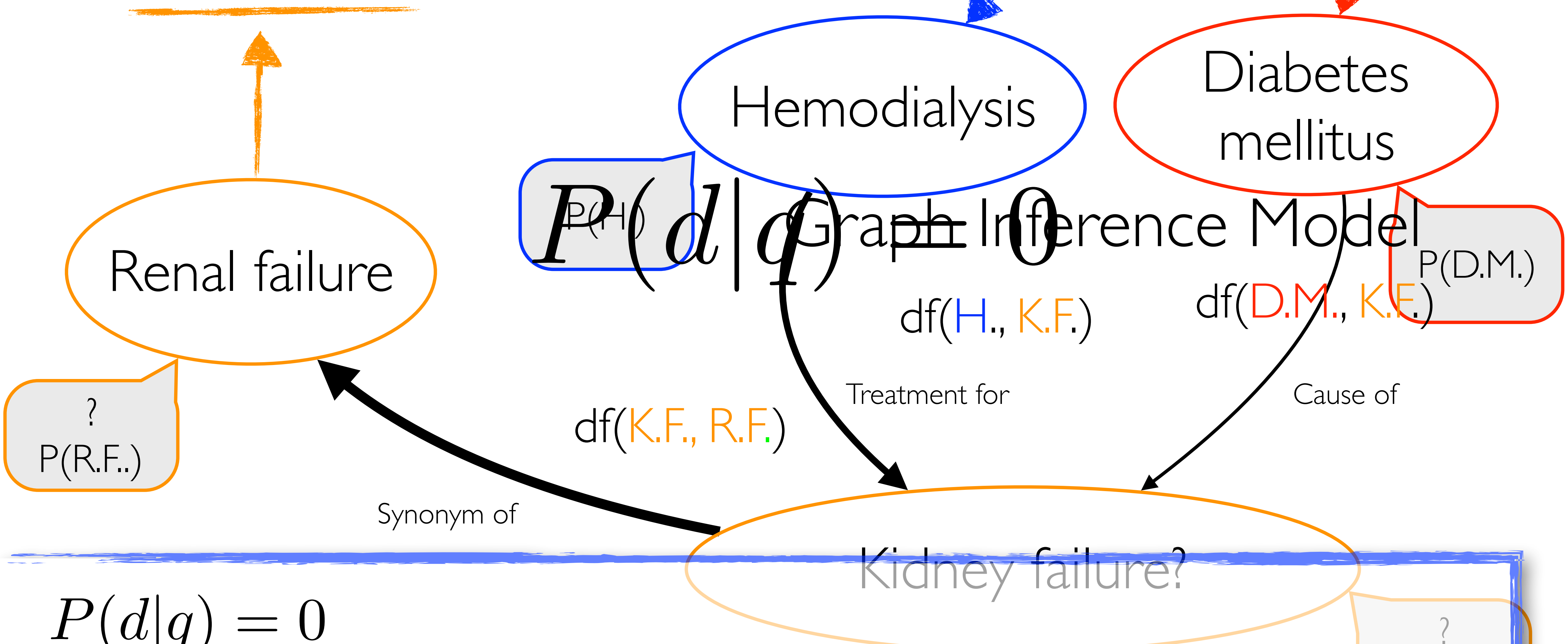
- For KB: use semantic relationships of concepts to represent the relationships between concepts.
- For free-text: MetaMap to identify concepts from the free-text, then infer relationships by co-occurence/association rules



[Koopman et al., 2016]

q "Patients with diabetes and renal failure"

d "This is a 62-year-old gentleman who has history of Type I DM and is on hemodialysis."



$$P(d|q) = 0$$

$$P(d \rightarrow q) \approx P(D.M.) * df(D.M., K.F.) + P(H.) * df(H., K.F.)$$

Practical - part 2

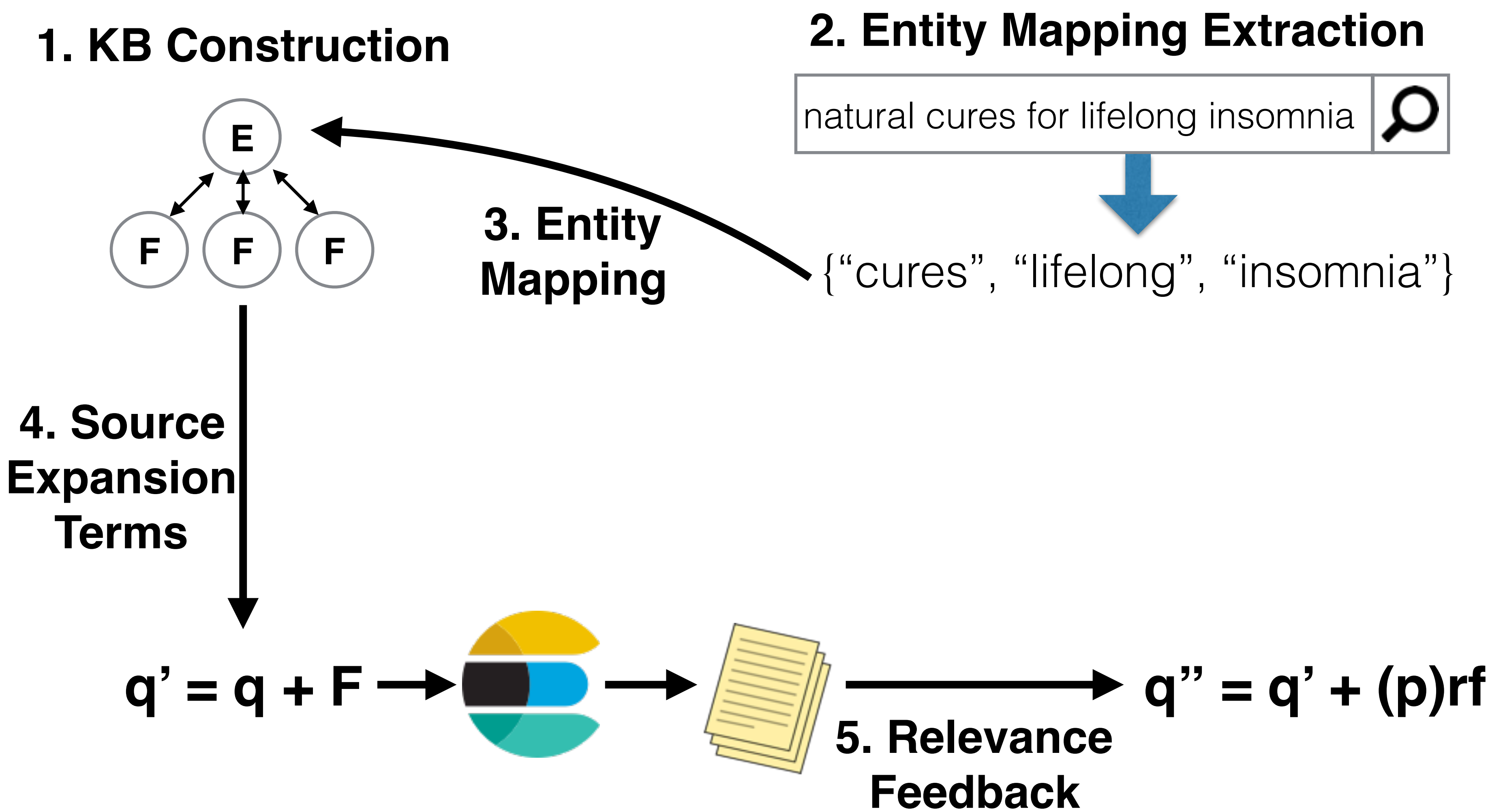
- Let's resume from where we left in part 1, and let's do:

1. Index these documents in Elasticsearch with multi term/concepts fields.
2. Search Elaticsearch with either term or concept, demonstrating semantic search capabilities.
3. Play a bit more (maybe)

- Instructions: https://ielab.io/russir2018-health-search-tutorial/hands-on/

Choices in KB Query Expansion

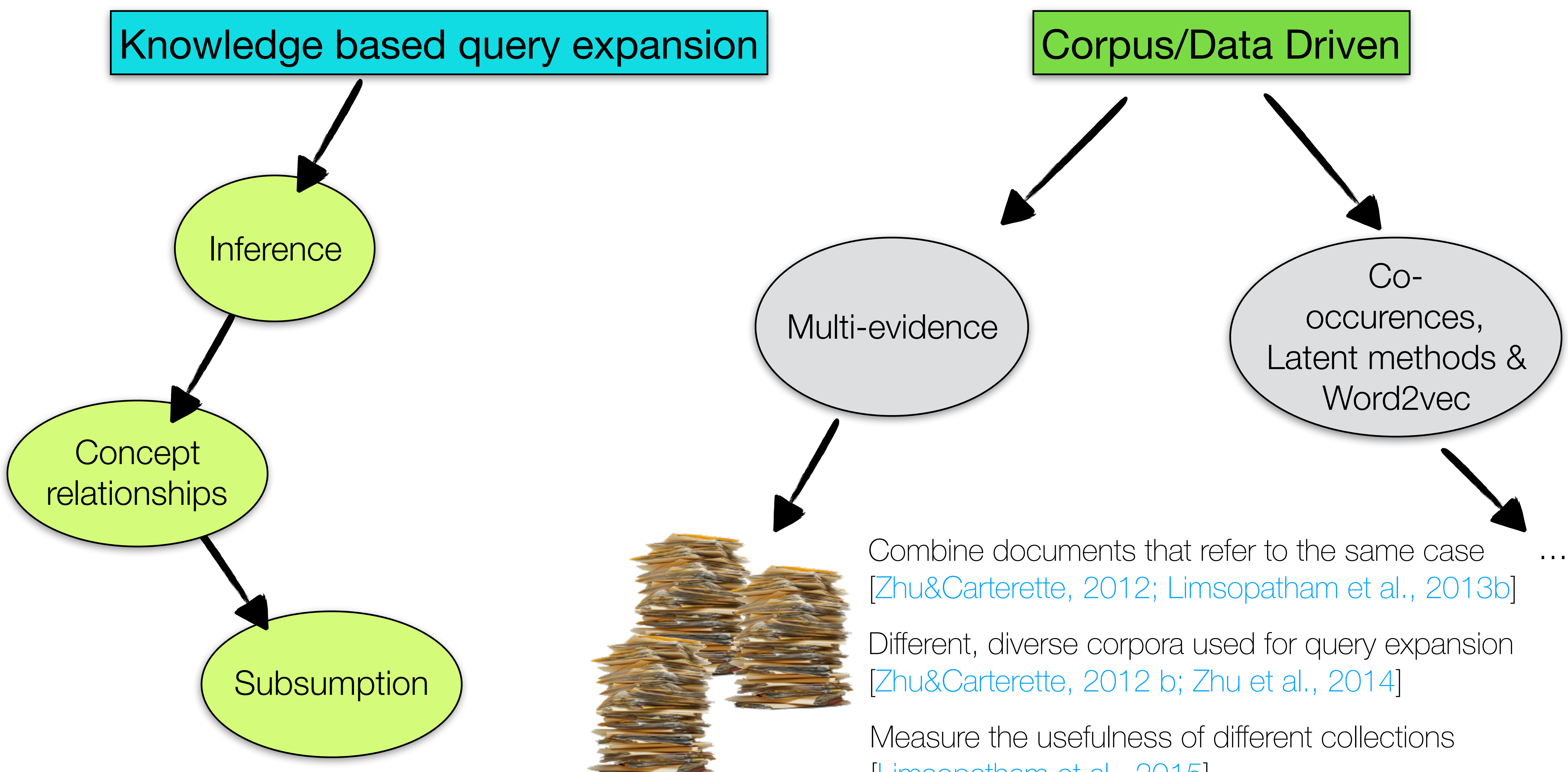
- Many other approaches to do inference over KB data
- [Jimmy et al., 2018] consider the Entity Query Feature Expansion model [Dalton et al., 2014] and the influence settings choices have



Choices in KB Query Expansion Findings for CHS

- For CHS, EQFE based on **UMLS** is more effective than based on Wikipedia.
  - Choice 1: Index **all** UMLS **concepts**
  - Choice 2: Use **all uni-, bi-, and tri-grams** of the original **queries**
  - Choice 3: **Map** mentions to **UMLS aliases**
  - Choice 4: Source **expansion** from the **UMLS title**
  - Choice 5: Add **relevance feedback** terms

Knowledge based vs data-driven Query Expansion



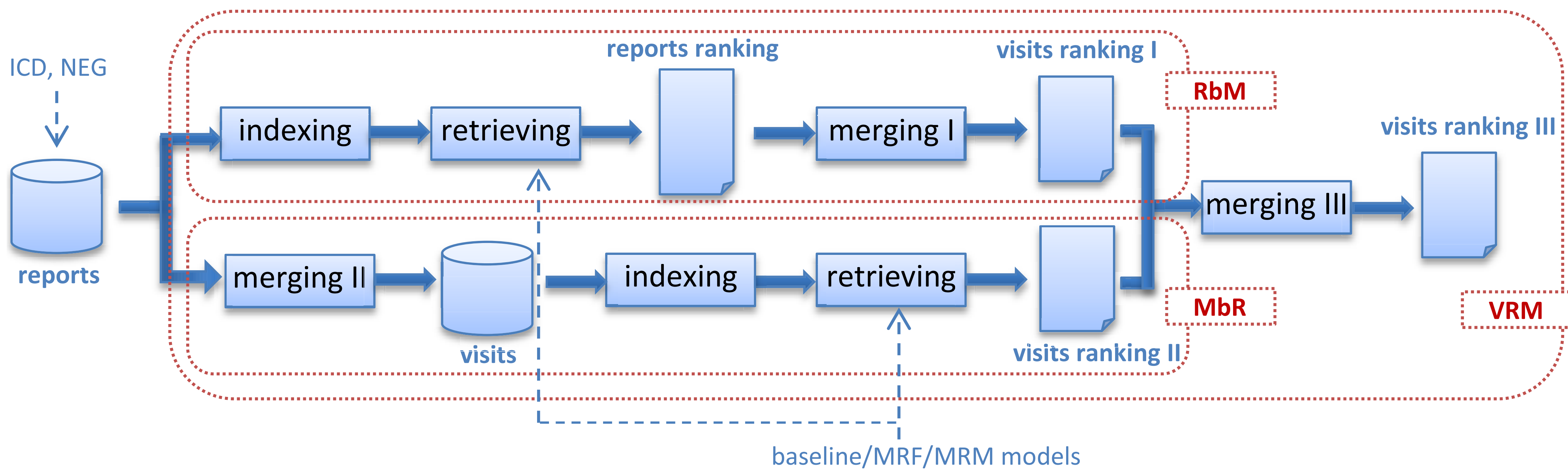
Combine documents that refer to the same case [Zhu&Carterette, 2012; Limsopatham et al., 2013b]

Different, diverse corpora used for query expansion [Zhu&Carterette, 2012 b; Zhu et al., 2014]

Measure the usefulness of different collections [Limsopatham et al., 2015]

Combine multiple-evidences in the collection that refer to the same case

[Zhu&Carterette, 2012]



- Ranking generated for each document, individually
- Ranking generated for an aggregated case

} Fused into new ranking

- Online possible in situations where multiple documents are available for one case (e.g. with health records, where case=patient)

Adaptively Combine (or not) Records of a Case

[Limsopatham et al., 2013b]

- Choose between:
    1. Combine records for a patient, then rank patient
    2. Rank records, then identify patients based on relevance of records ranking
- Classifier to **learn to select which ranking approach** to use, depending on query
- Features: query difficulty measures (QPPs), number of medical concepts in query

Different, diverse corpora used for query expansion

[Zhu et al., 2014]

- Mixture of relevance models to **combine evidence from different collections** to derive **query expansions**
  - Collections: Mayo Clinic health records (39M), TREC Genomics (166K), ClueWeb09B (44M), TREC Medical Records (100K)
- Findings:
  - Access to **large clinical corpus** significantly improves query expansion
  - The more difficult the query, the more it benefit expansion benefits from auxiliary collections
  - “use **all available data**" is **sub-optimal**: value in collection curation

Measure the usefulness of different collections

[Limsopatham et al., 2015]

- Automatically decide which collection to use for query expansion evidence
  - 14 different document collections, from domain-specific (e.g. MEDLINE abstracts) to generic (e.g. blogs and webpages)
    - But they are **not all useful**, and **not to the same extent** to generate query expansion terms
- Techniques based on resource selection and learning to rank

Co-occurences, Latent Methods & Word2vec

- (Co-occurence of) concepts as a **graph** -> application of **link analysis** methods [Koopman et al., 2012; Martinez et al., 2014]
- **Explicit** and **latent concepts** [Balaneshin-kordan&Kotov, 2016]
- Word **embeddings** and concept embeddings [Zuccon et al., 2015, b; Nguyen et al., 2017]

Co-occurence Graphs, Semantic Graphs and Page Rank

- [Koopman et al., 2012]:
  1. Build concept **graph** from **document concepts** as they **co-occur** in document
  2. Run Pagerank
  3. Use Pagerank scores as additional weights for retrieval
- [Martinez et al., 2014]:
  1. Build concept **graph** from **query concepts** and **related concepts** in **UMLS**
  2. Run Pagerank
  3. Rank concepts using page rank scores; select top K concepts as query expansion
- Analysis shows expansion terms selected by Pagerank: taxonomic (eg., synonyms) and not taxonomic (eg., disease has associated anatomic site).

Explicit and Latent Concepts

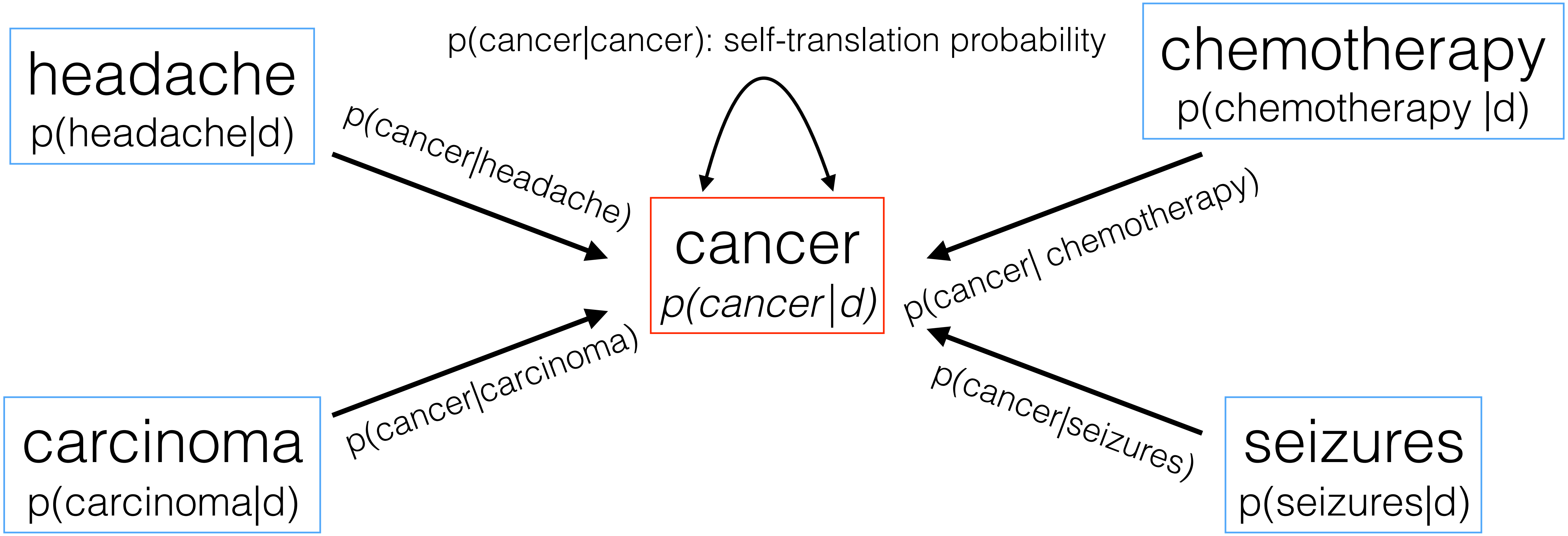
- [Balaneshin-kordan&Kotov, 2016]: different **concept types/ sources** (KBs, PRF) should have **different weights**
- Builds upon Markov Random Field retrieval [Metzler&Croft, 2005]
- Different **features** for different **semantic types** + **topical features** of KB graphs, and **statistics** of concepts in collection
- Learns optimal query concept weight using multivariate optimisation
- Base approach (without optimisation) best system at TREC CDS 2015

Word Embeddings and Concept Embeddings: Neural Translation LM

[Zuccon et al., 2015, b]

$$p_t(w|d) = \sum_{u \in d} \boxed{p_t(w|u)} p(u|d)$$

use Word Embeddings for computing this



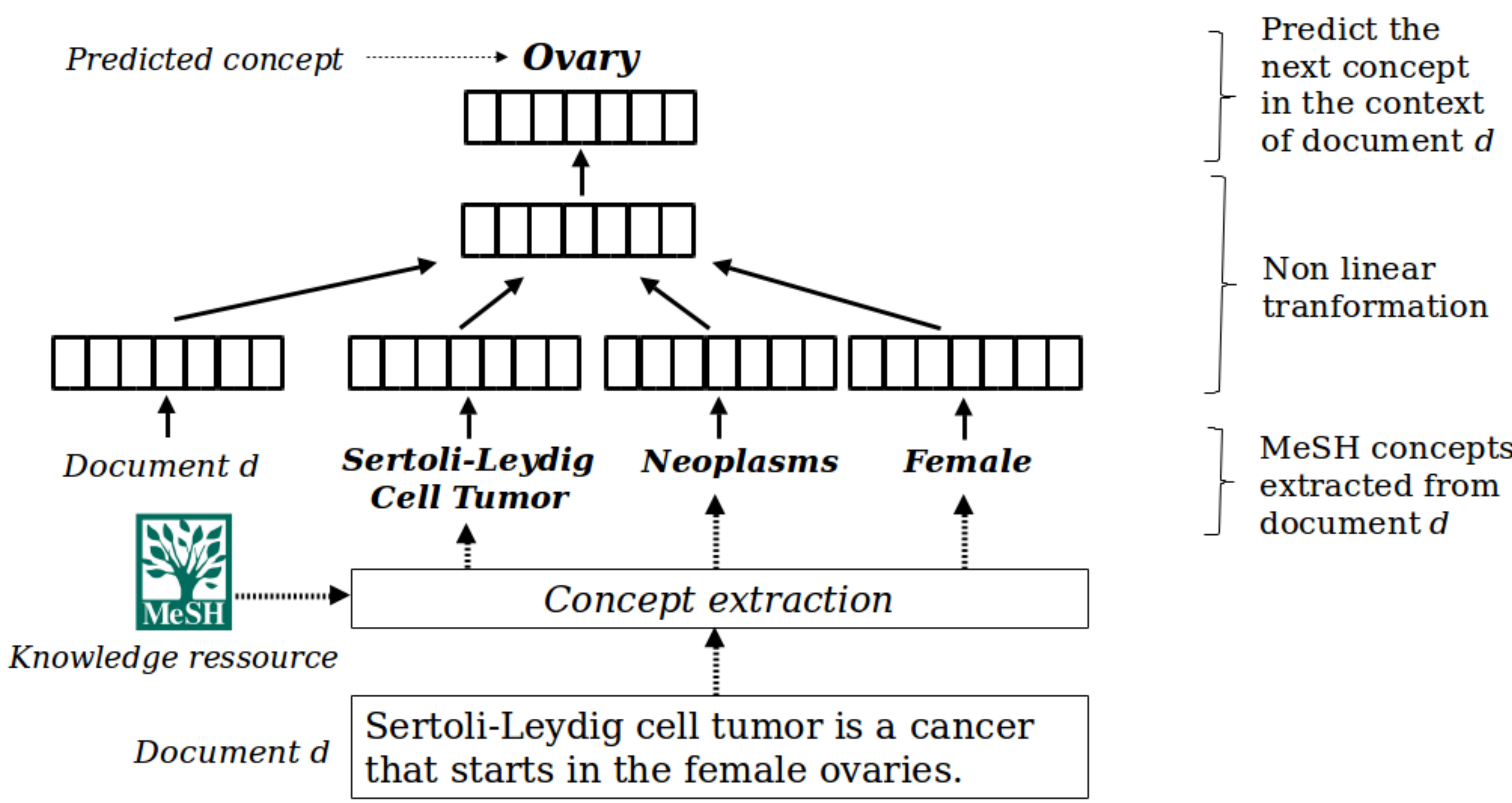
Constraining word embeddings by prior knowledge

- [Liu et al., 2016]: learn concept embeddings **constrained by relations** in KB (UMLS)
- Results in a modified CBOW
- Use word embeddings to **re-rank** results: **interpolate** original relevance score with similarity based on embeddings
- Experiments only limited to synonym relations & single-word concepts

Concept-Driven Medical Document Embeddings

[Nguyen et al., 2017]: optimises document representation for medical content

- Uses neural-based approach (akin to *doc2vec*) to create embedding that captures latent relations from concepts and terms in text.
- Uses embedding to identify top documents
- Extract top words and concepts from top documents to produce expansions



Learning to Rank

[Soldaini&Goharian, 2017]: compares 5 LTR in CHS context:

- LTR: logistic regression, random forests, LambdaMART, AdaRank, ListNet
- Features: statistical (36 features), statistical health (9), UMLS (26), latent semantic analysis (2), word embeddings (4).
- **LambdaMART** performed best; **all features** required

Dealing with the nuances of medical language

Negation & Family History

`“denies fever”`
`“no fracture”`

`“mother had breast cancer”`

NegEx/ConText [Harkema et al., 2009]:
Algorithm for extracting negated content

- Negated content best handled by:
  - **Not removing** negated content (as is commonly done)
  - **Indexing** positive, negated & family history content **separately** [Limsopatham et al., 2012]
  - **Weighting** content **separately** [Koopman & Zuccon, 2014]

PICO

- PICO: framework for formulating clinical questions
  P: Patient/Problem (P) (e.g., males aged 20-50)
  I: Intervention (e.g., weight loss drug)
  C: Comparison (e.g., controlled exercise regime)
  O: Outcome (e.g., weight loss)

- Exploiting PICO elements in IR:
  - Language modelling based **content weighting** [Boudin et al., 2010]
    - Tagging PICO elements for IR - "I" & "P" elements most beneficial for retrieval
  - **Field retrieval** based on PICO [Scells et al., 2017b]
    - promising, but needs method to predict which keywords require PICO annotations

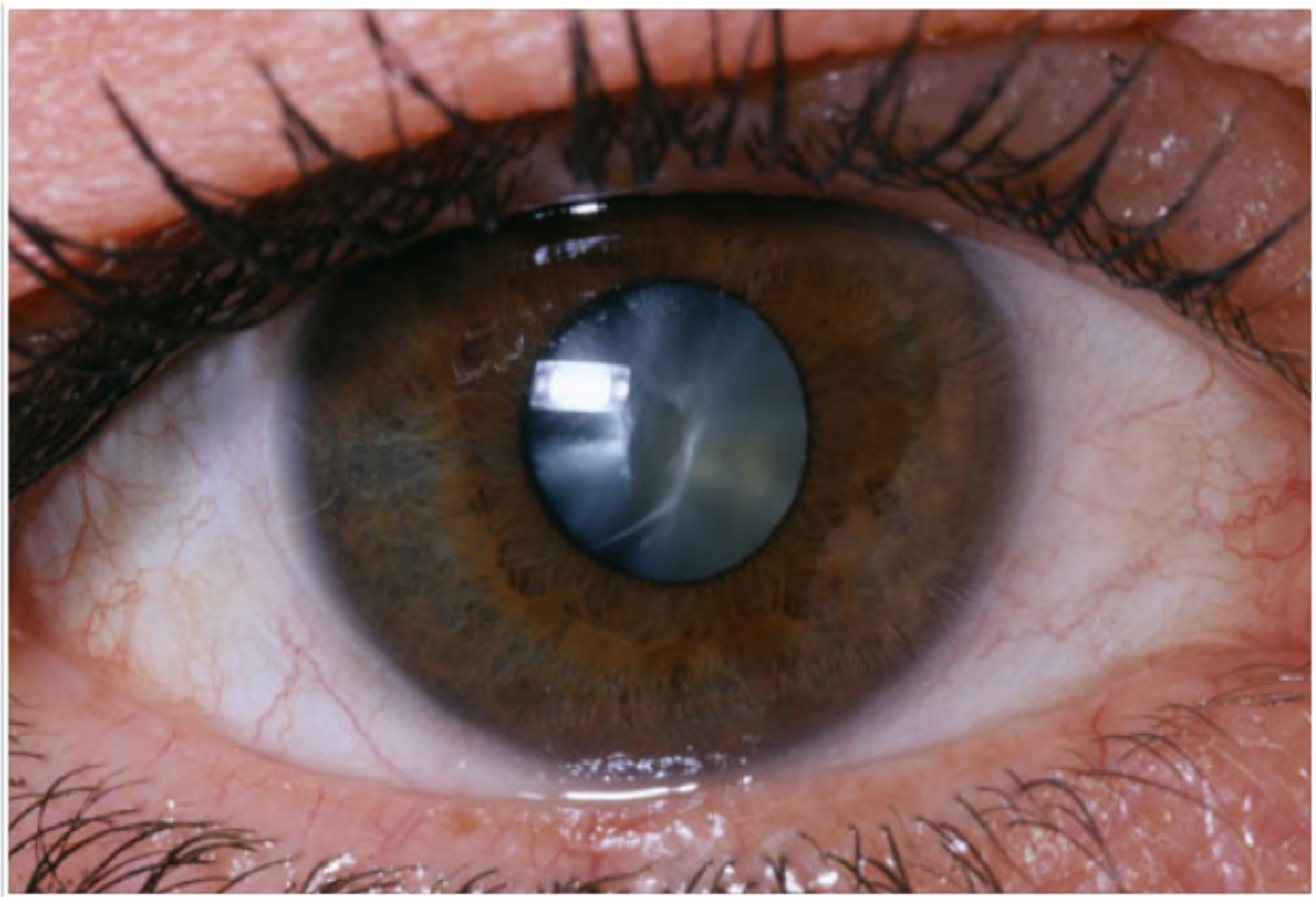
RobotReviewer [Marshall et al., 2015]: Algorithm for extracting PICO elements from free-text

Readability & Understandability

- **Laypeople** do not necessarily **understand medical** documents that clinicians would understand
- Need to retrieve documents that are both understandable and relevant
- [Palotti et al., 2016 b]: LTR with two sets of features:
    - Estimate relevance: standard IR features
    - Estimate understandability: features based on **readability measures** and **medical lexical aspects**

Understanding and aiding query formulation

What would **search** for?



Enter your search terms at http://chs.ielab.webfactional.com/

“Circumlocutory” queries

Symptom Group	Crowdsourced Circumlocutory Queries
alopecia	baldness in multiple spots, circular bald spots, loss of hair on scalp in an inch width round
angular cheilitis	broken lips, dry cracked lips, lip sores, sores around mouth
edema	fluid in leg, puffy sore calf, swollen legs
exophthalmos	bulging eye, eye balls coming out, swollen eye, swollen eye balls
hematoma	hand turned dark blue, neck hematoma, large purple bruise on arm
jaundice	yellow eyes, eye illness, white part of the eye turned green
psoriasis	red dry skin, dry irritated skin on scalp, silvery-white scalp + inner ear
urticaria	hives all over body, skin rash on chest, extreme red rash on arm

[Stanton et al., 2014]

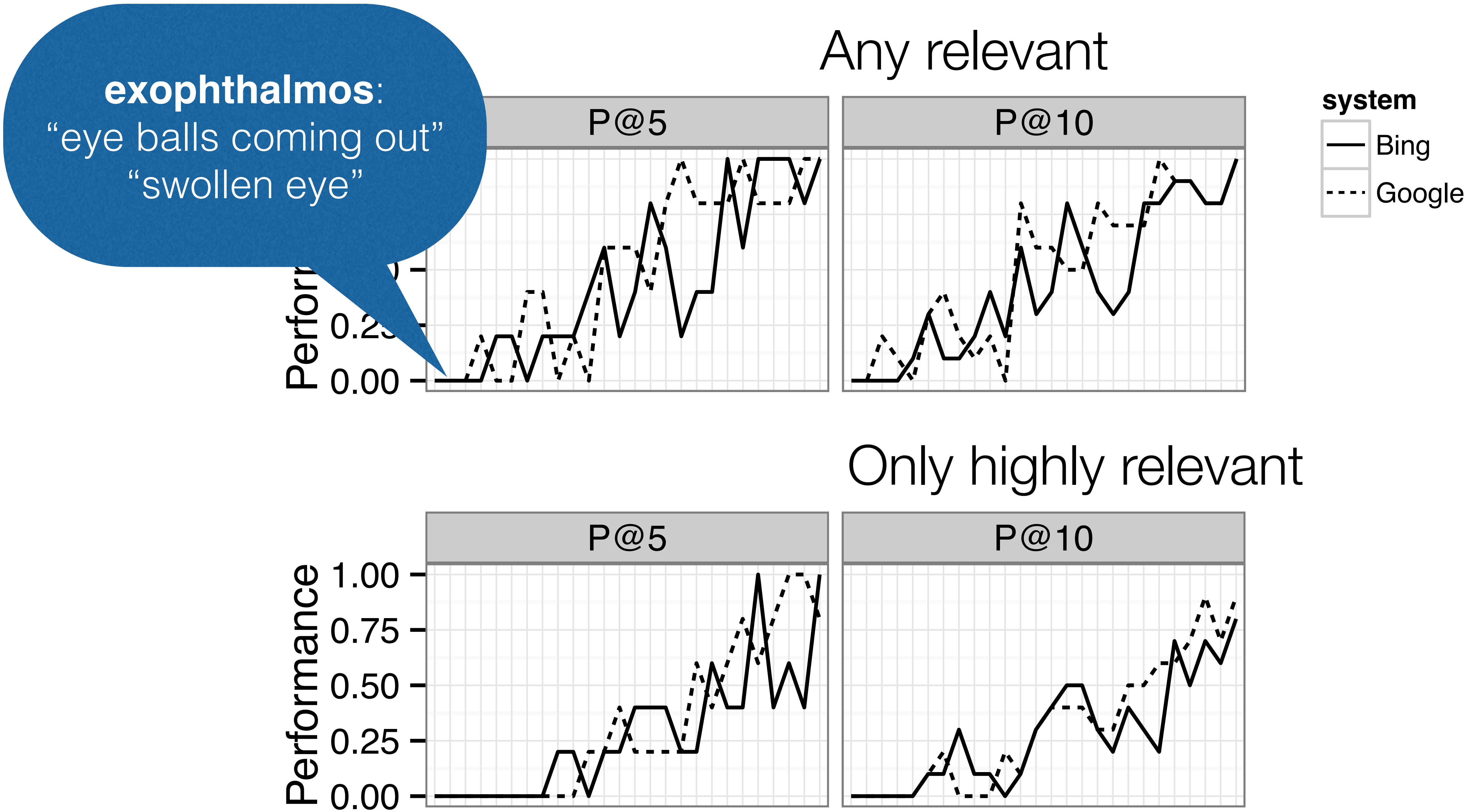
How effective are Google & Bing at Health Search?

System	ndcg@1		ndcg@5		ndcg@10		P@5		P@10	
	Rel	Hrel	Rel	Hrel	Rel	Hrel	Rel	Hrel	Rel	Hrel
Bing	.3846	.2308	.3812	.2654	.3802	.2764	.4385	.2769	.4308	.2769
Google	.3846	.3077	.4242	.3142	.4252	.3138	.5000	.3154	.4923	.3115

Effectiveness of two widely used commercial search engines when prompted with (circumlocutory) medical queries aimed at self-diagnosis purposes. Results are averaged over 26 queries. P@k stands for precision at rank k; ndcg@k stands for normalised discounted cumulative gain at rank k.

[Zuccon et al., 2015]

Performance per query

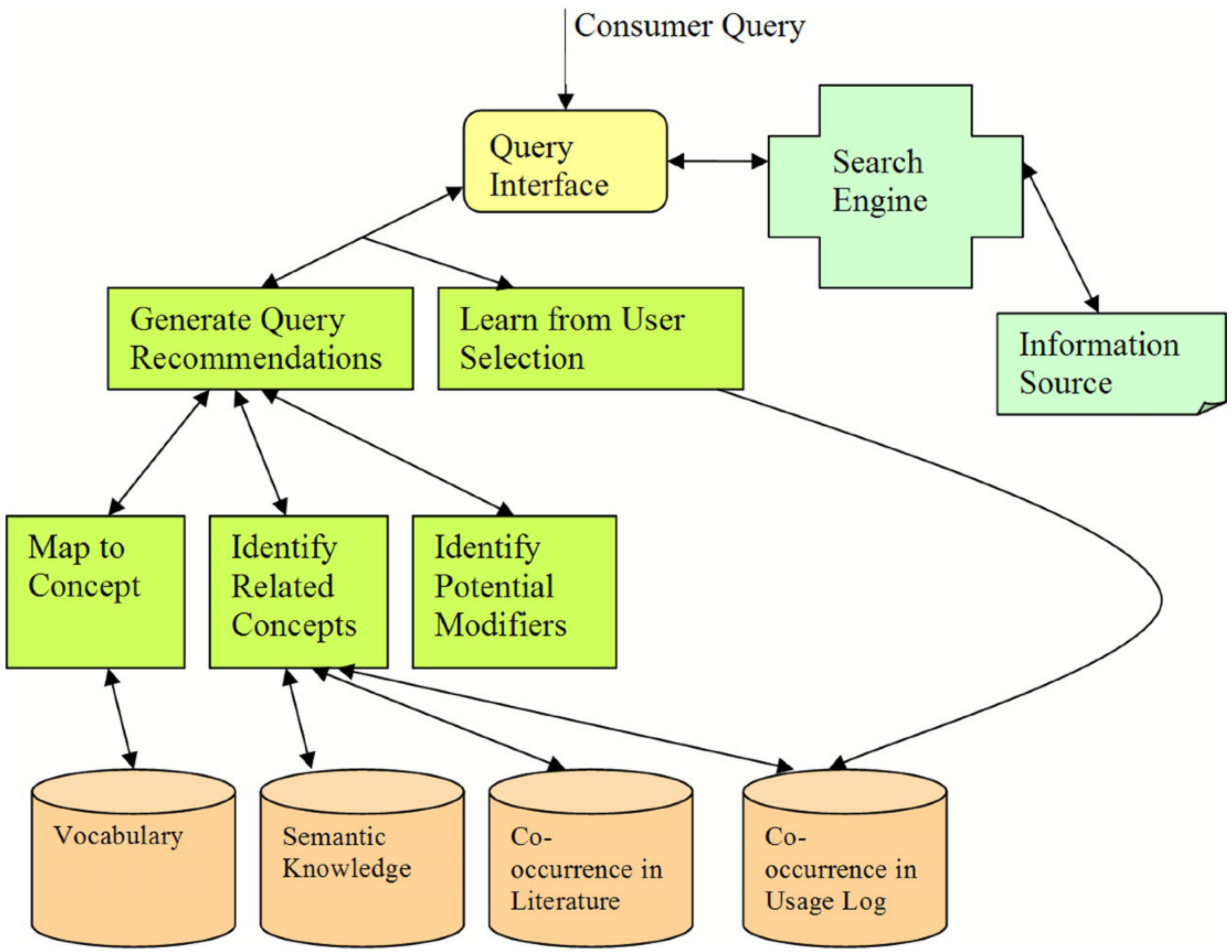


[Zuccon et al., 2015]

Query Recommendation

[Zeng et al, 2006]: recommend queries based on UMLS and query log (CHS task)

- Leads to higher user satisfaction and query success rate



Query Reformulation

[Soldaini et al., 2015]: compares the effectiveness of 7 query reformulation techniques (CDS task)

1. **UMLS Concepts Selection** (MMselect): remove all terms with no mapping to any UMLS concepts
2. **Health-related terms selection** (HT): compute ratio of associated Wikipedia page P being health-related over being not-health-related. Retain only query terms with ratio ≥ 2.
3. **Query Quality Predictors** (QQP): use QPPs as features of SVMrank to select query terms.
4. **Faster QQP**: rank sub-queries using MI and retains the top 50. In addition to QQP features, add features: UMLS concepts found, UMLS sem-types found, HT ratio, and MeSH found.

Query Reformulation

[Soldaini et al., 2015]: compares the effectiveness of 7 query reformulation techniques (CDS task)

5. **UMLS Concepts Extraction** (MMexpand): append the preferred terms UMLS query concepts to expand original query

6. **Pseudo Relevance Feedback** (PRF): weight terms in top 10 initial results, rank and add top 20 terms not in original query.

7. **Health Terms PRF** (HT-PRF): as PRF, but candidate expansion terms filtered health term ratio

    - This is empirically identified as the best technique

    - The HT component in general seems effective

Query Reformulation with deep learning

[Soldaini et al., 2017]: considers short clinical notes as queries (CDS task)

1. Generate candidate terms using PRF
2. Train supervised neural network to predict Weight Relevance Ratio (WRR) of candidate terms: importance of term in relevant documents
3. For representations it uses word embeddings, statistical features over multiple collections, syntactical and semantical features

- The neural network approach and HT-PRF perform similarly

Query Clarification

[Soldaini et al., 2016]: **add** the most appropriate **expert expression** to queries submitted by users

- Acquire expert expressions **from 3 KBs**: behavioral (logs), MedSyn, and DBpedia
- Select expression with the highest probabilities of appearing in health-related Wikipedia pages, using logistic regression classifier
- Finding through user study evaluation (CHS task):
  - Expressions from all 3 KBs **improve rate of correct answers** (behavioural KB best)
  - number of **correct** answers significantly increases when users clicked **HON-certified websites**

Query Reduction

- [Koopman et al., 2017 c]: reduce verbose clinical queries (health records, CDS task) using generic & domain-specific methods
- Reduce to only UMLS Medical Concepts & Tasked UMLS
- Combined model UMLS + IDF-r (proportion of top-ranked IDF terms retained)
- Comparison vs human-generated queries: human generated queries significantly more effective
  - per-query parameter learning promising
  - automated reduction handicapped in that they only use terms from narrative

Query Reduction

[Soldaini et al., 2017 b]: use convolutional neural networks (CNN) to reduce queries (CDS task)

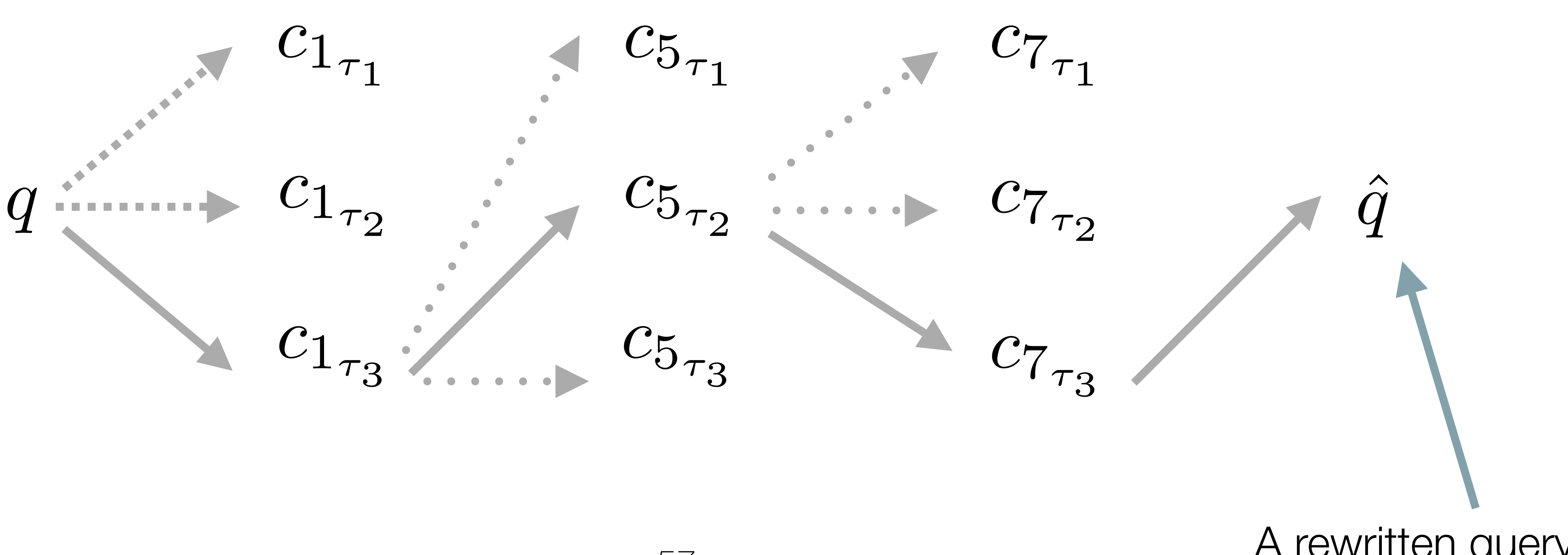
- Queries are short clinical notes
- CNN is used to estimate the importance of each query term
- Given a query, a relevant document and a non-relevant document:

1. Use CNN to determine weights terms in query
2. Use term weights to score relevant and non-relevant documents
3. Back-propagate a loss if non-relevant document is scored higher than relevant document

Query Rewriting

[Scells&Zuccon, 2018]: through a chain of transformation, generates better (Boolean) queries (for systematic reviews compilation)

- Defines set of transformations: mostly syntactic transformations
- Selects transformations based on: heuristics, classifier, learning to rank
- Large gains possible by transforming queries



Query Difficulty

- [Boudin et al., 2012]: predictor that **exploits MeSH structure** to ascertain how difficult queries are — estimates query variability and specificity

$$MeSH\text{-}QD(Q,\mathcal{T}) = \sum_{t \in Q} \overbrace{\frac{df(t)}{\sum_{t' \in V(t)} df(t')} \cdot ln\Big(1 + \frac{N}{df(t)}\Big)}^{\text{term variability}} \cdot \overbrace{\frac{\text{depth}(t)}{\text{length}(t)}}^{\text{term generality}}$$

  - V(t): set of alternative expressions of the concept t; depth/length in MeSH
  - **coverage** of thesaurus & concept **mapping** influence quality
- [Scells et al., 2018]: **standard predictors** for QPP and QVPP (V=variation) in systematic reviews compilation
  - Predictors **not suited** to the domain-specific nature of the task
  - Identifying best performing **variations hard task**

Task based retrieval

- Research on how clinicians' query shows a set of standard query types [Ely et al., 2000]
- Can be simplified to three clinical tasks:

  i. searching for **diagnoses** given a list of symptoms;

  ii. searching for relevant **tests** given a patient's situation

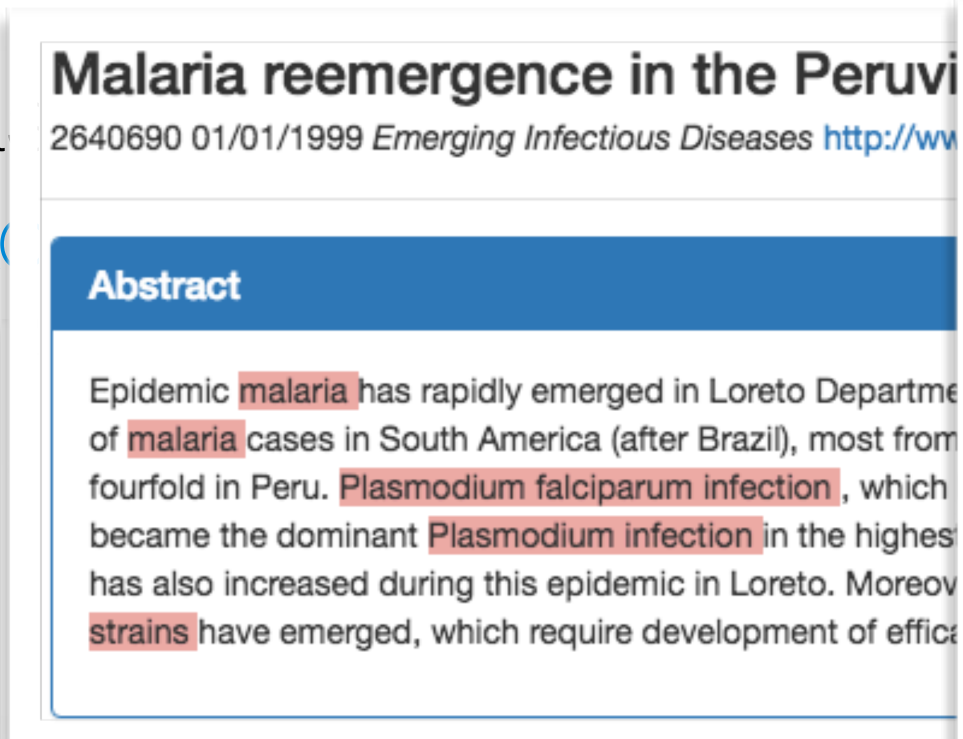
  iii. searching for effective **treatments** given a particular condition.

- These can be exploited in a retrieval scenario…

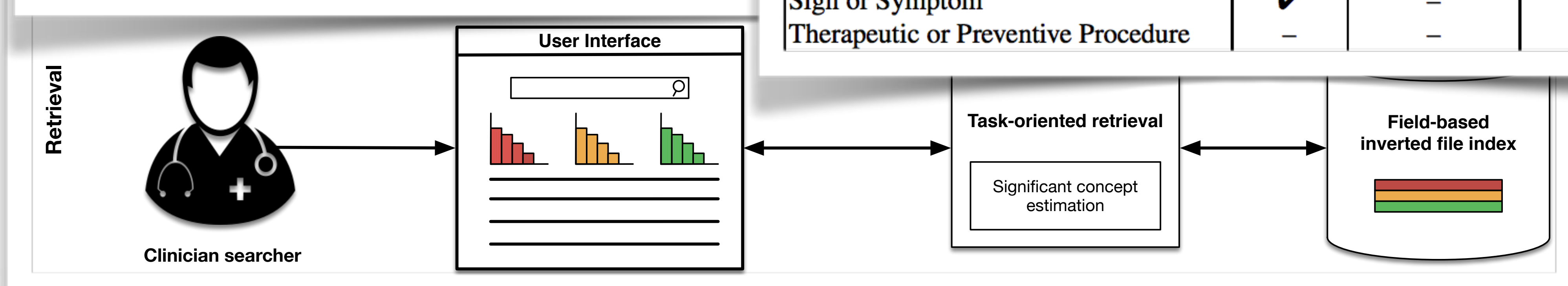
Tasked-based retrieval

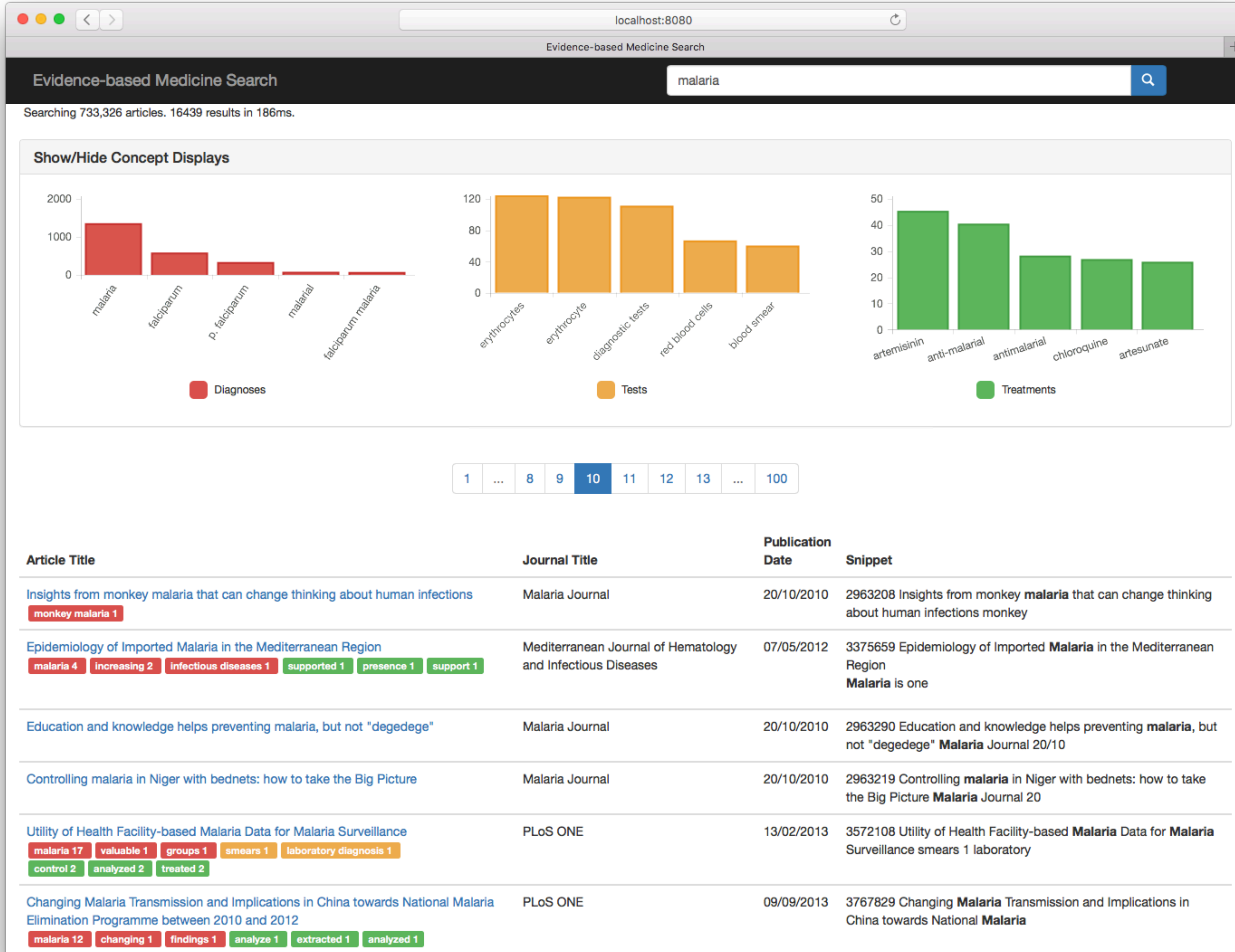
- Concept-based approach but “focusing only on medical concepts essential for the information need of a medical search task” [Limsopatham et al., 2013]

- Ta
[K



MetaMap’s Semantic Type	Aspects of the Medical Decision Criteria			
	Symptom	Diagnostic test	Diagnosis	Treatment
Body Location or Region	✔	✔	✔	✔
Body Part, Organ, or Organ Component	✔	✔	✔	✔
Clinical Drug	–	–	–	✔
Diagnostic Procedure	–	✔	–	–
Disease or Syndrome	–	–	✔	–
Finding	✔	–	–	–
Health Care Activity	–	✔	–	✔
Injury or Poisoning	✔	–	–	–
Intellectual Product	–	✔	–	✔
Medical Device	–	✔	–	✔
Mental or Behavioral Dysfunction	✔	–	✔	–
Neoplastic Process	✔	✔	✔	✔
Pathologic Function	✔	–	–	–
Pharmacologic Substance	–	–	–	✔
Sign or Symptom	✔	–	–	–
Therapeutic or Preventive Procedure	–	–	–	✔




localhost:8080
Evidence-based Medicine Search
Evidence-based Medicine Search
malaria
Searching 733,326 articles. 16439 results in 186ms.
Show/Hide Concept Displays
2000
1000
0
malaria
falciparum
p. falciparum
malarial
falciparum malaria
Diagnoses
120
80
40
0
erythrocytes
erythrocyte
diagnostic tests
red blood cells
blood smear
Tests
50
40
30
20
10
0
artemisinin
anti-malarial
antimalarial
chloroquine
artesunate
Treatments
1 ... 8 9 10 11 12 13 ... 100
Article Title
Journal Title
Publication Date
Snippet
Insights from monkey malaria that can change thinking about human infections
monkey malaria 1
Malaria Journal
20/10/2010
2963208 Insights from monkey malaria that can change thinking about human infections monkey
Epidemiology of Imported Malaria in the Mediterranean Region
malaria 4
increasing 2
infectious diseases 1
supported 1
presence 1
support 1
Mediterranean Journal of Hematology and Infectious Diseases
07/05/2012
3375659 Epidemiology of Imported Malaria in the Mediterranean Region
Malaria is one
Education and knowledge helps preventing malaria, but not "degedege"
Malaria Journal
20/10/2010
2963290 Education and knowledge helps preventing malaria, but not "degedege" Malaria Journal 20/10
Controlling malaria in Niger with bednets: how to take the Big Picture
Malaria Journal
20/10/2010
2963219 Controlling malaria in Niger with bednets: how to take the Big Picture Malaria Journal 20
Utility of Health Facility-based Malaria Data for Malaria Surveillance
malaria 17
valuable 1
groups 1
smears 1
laboratory diagnosis 1
control 2
analyzed 2
treated 2
PLoS ONE
13/02/2013
3572108 Utility of Health Facility-based Malaria Data for Malaria Surveillance smears 1 laboratory
Changing Malaria Transmission and Implications in China towards National Malaria Elimination Programme between 2010 and 2012
malaria 12
changing 1
findings 1
analyze 1
extracted 1
analyzed 1
PLoS ONE
09/09/2013
3767829 Changing Malaria Transmission and Implications in China towards National Malaria

How does a good health query look like?

- [Tamine&Chouquete, 2017] found that in health search, query quality is influenced by medical expertise
- [Koopman et al., 2017] studied the querying behaviour of 4 clinicians
  - most effective clinicians those who entered **short queries** (but retrieval models optimised for short queries)
  - most effective clinicians those who **inferred novel keywords** most likely to appear in relevant documents
  - most effective clinicians posed queries around **treatments rather than diagnoses** (but influenced by task: searching for clinical trials)