Domain Specific Applications

Sophia Althammer

sophia.althammer@tuwien.ac.at
/sophiaalthammer

Today

Domain Specific Applications

1 Medical domain

- Systematic reviews, clinical decision support
- TREC Covid, TREC biomedical track, BioASQ evaluation campaigns, CLEF eHealth

2 Legal domain

- Prior case retrieval, eDiscovery
- COLIEE, FIRE AILA evaluation campaigns

What is domain specific information retrieval?

- So far in this lecture mainly web search and extractive QA considered

- BUT also information retrieval applications in **other domains like the medical, legal, patent or academic domain**

What is domain specific information retrieval?

- Search in databases of specific domains

Medical domain

Legal domain

Scientific domain

What is domain specific information retrieval?

- Specific users

Medical domain

Doctors, medical researcher

Legal domain

Legals, paralegals

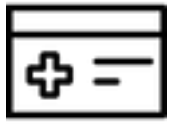
Scientific domain

Academics, researcher

What is domain specific information retrieval?

- Specific tasks

Medical domain



Systematic reviews, clinical decision support

Legal domain

Prior case retrieval, eDiscovery

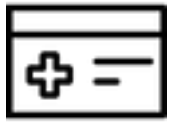
Scientific domain

Prior literature review, surveys

What is domain specific information retrieval?

- Tasks characteristics different to web search

Medical domain



Long exhaustive search process, transparent results

Legal domain

Long search sessions, multiple queries

Scientific domain

Specific domain language, long search sessions

Medical domain

Tasks, evaluation campaigns and solutions in the medical domain

Different textual health information

- Patient-specific information is common to health practitioners working daily with patients
  - Searching for this information tell doctors about the patient's health and disease

- Knowledge-based information from experiments, summaries, and observations

Patient-specific information	**Knowledge-based information**
Structured: - lab results, vital signs	Primary: - original research (e.g. in journals, books, reports)
Narrative: - progress note, radiology report	Secondary: - summaries of research (e.g. in systematic reviews, practice guidelines)

Different textual health information

Patient-specific information	**Knowledge-based information**
Structured: - lab results, vital signs	Primary: - original research (e.g. in journals, books, reports)
Narrative: - progress note, radiology report	Secondary: - summaries of research (e.g. in systematic reviews, practice guidelines)

Clinical decision support

Systematic Reviews

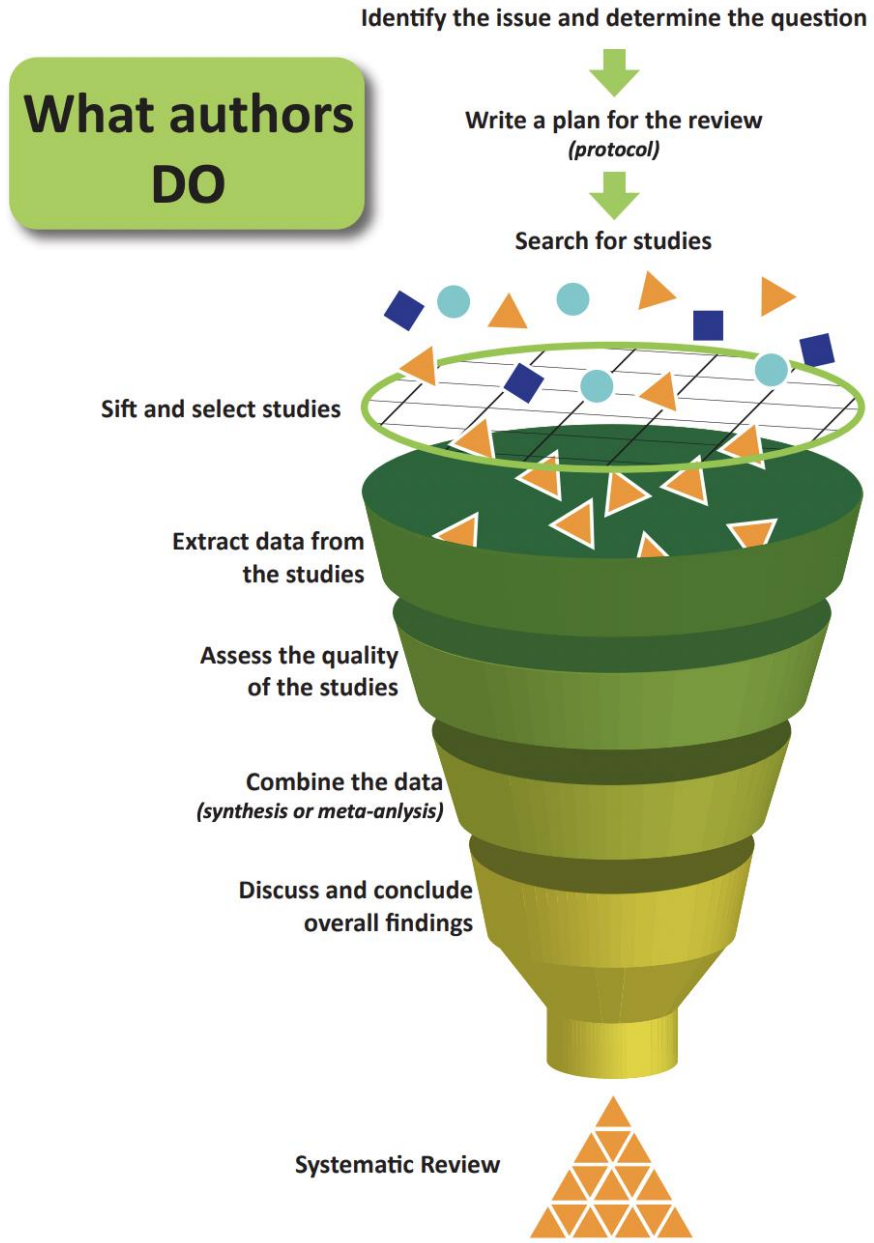
Systematic reviews

- Secondary study that summarises all available data fitting pre-specied criteria to answer specific research questions
  - Uses rigorous scientific methods to minimise bias and generate solid conclusions from which health practitioners can make decisions
  - To verify all empirical evidence, researchers in the search step must find all publications relevant to the research question

- Resource-intensive process – relies mostly on human labour
  - Exponential number of publications

Systematic reviews

- Hierarchical approach of conducting a systematic review[1]
  - define search strategy, refine research question
  - Retrieval: Boolean search queries
  - Appraisal: screening on a abstract and document level
  - Synthesis



1 Cochrane Handbook for Systematic Reviews of Interventions, 2nd Edition, Julian Higgins

Systematic reviews

- Legal requirements for a review
  - Explainable
  - Transparent
  - Reliable
  - Reproducible
- Therefore traditional Boolean search queries and lexical matching methods for their IR systems preferred so far
- But we can change that!

Clinical decision support

- Provides clinicians, staff, patients with knowledge and consider patient-specific information from the electronic health record (EHR)
- Generate case specific advice for treatment of a patient
- Link health observations with health knowledge

Evaluation campaigns and research datasets

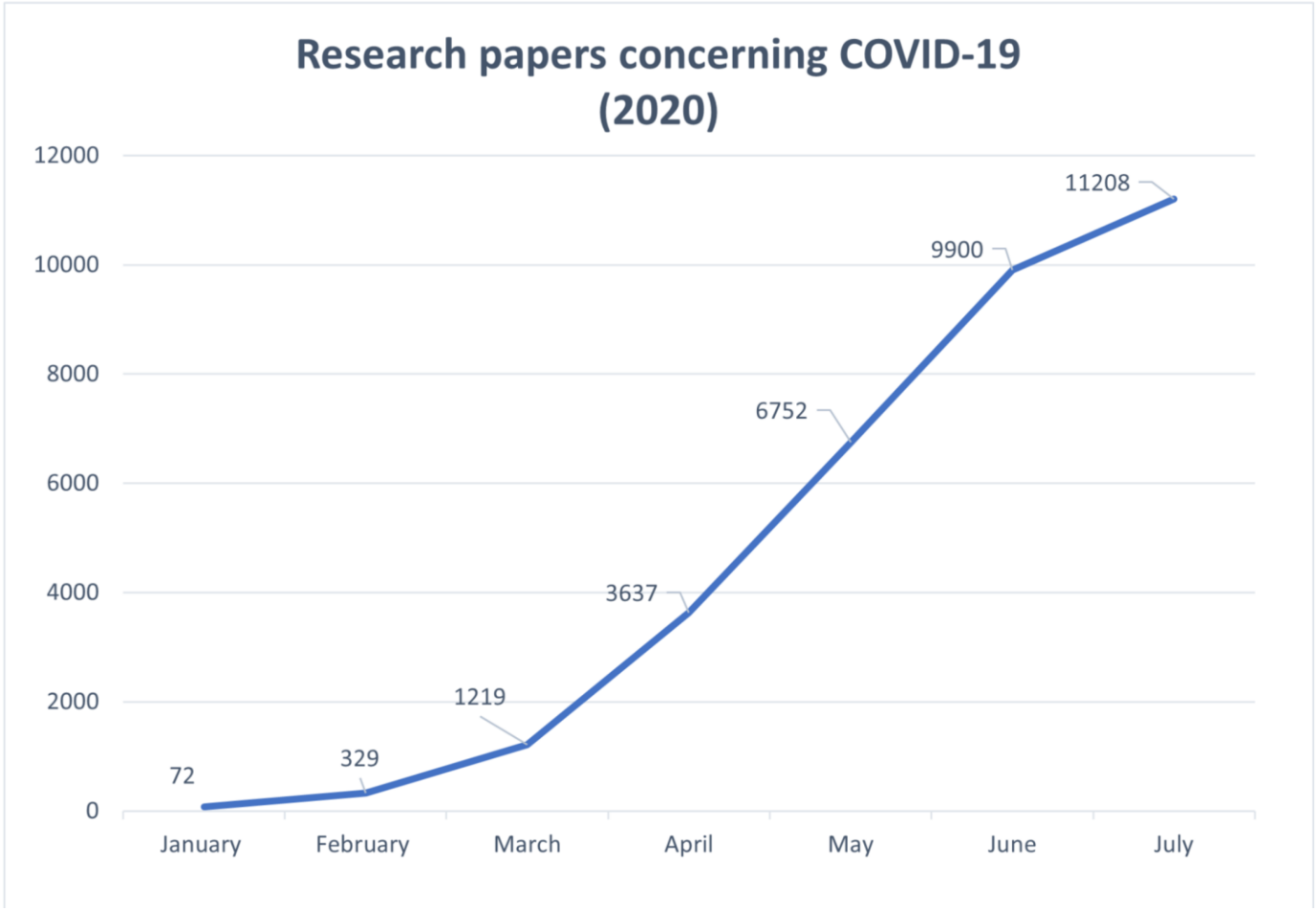
- TREC Clinical Trials[1]
  - Matches patients depending on their health record to clinical trials
- BioASQ[2]
  - Biomedical question-answering
- CLEF eHealth[3]
  - Patient centered information retrieval/consumer health search
  - Technology assisted reviews in empiricial medicine

1 TREC Biomedical Tracks, http://www.trec-cds.org/
2 BioASQ A challenge on large-scale biomedical semantic indexing and question answering, http://www.bioasq.org/
3 Clef eHealth Lab Series, https://clefehealth.imag.fr/

Evaluation campaigns: TREC Covid[1]

- In order to enhance research on IR systems to support researchers keeping track of new published trials

- Dataset:
  - Updated research corpus of Covid19 related studies
  - Updated queries about Covid19



2

1 TREC-COVID: Building a Pandemic Retrieval Test Collection, https://ir.nist.gov/covidSubmit/index.html

2 Domain-specific language model pretraining for biomedical natural language processing**,** Microsoft Blog, https://www.microsoft.com/en-us/research/blog/domain-specific-language-model-pretraining-for-biomedical-natural-language-processing/

Challenges

- Medical domain specific language
  - Domain specific language models
    - SciBERT: trained on scientific and biomedical abstracts of Semantic Scholar
    - BioBERT: trained on biomedical study abstracts
    - PubMedBERT: trained on clinical abstracts from PubMed

Challenges

- Few labelled data
  - Evaluation campaigns
  - Take relevance labels from click logs:
    - TripClick: Large corpus of medical studies (title and abstracts), relevance annotations based on click data of users
- High recall setting
  - Missing a relevance document can be catastrophic!

Neural IR approach

- Similar as in web search:

1. First stage Retrieval
   - Boolean queries or
   - Dense retrieval approaches (in next lecture)

2. Neural re-ranking
   - Domain specific language models
   - Domain specific relevance annotations

Legal domain

Tasks, evaluation campaigns and solutions in the legal domain

Different legal systems

Statute law	vs	Case law
• Statutes and legal regulatories are the primary information source		• Precedent cases are a primary source for legal evidence
• Statute retrieval is important		• Prior case retrieval of high importance
• In European countries		• In Canada, US, Australia

Prior case retrieval in case law systems

- Task should lead to prior cases which should be noticed for solving the current case
- Information source is primary literature containing previous court decisions
- Desired output of the search is a list of prior cases, sorted by relevance or temporal aspects

Prior case retrieval in case law systems

- Aspects:
  - Precision-oriented task
  - Domain specific language of court decisions
  - Hierarchy of decisions (depending from which court) needs to be considered
  - Temporal aspects also need to be considered

eDiscovery

- Discovery in legal litigation with the requirement that documents be produced as critical evidence in litigation of a case

- Subject to rules of civil procedure and agreed upon processes, often involving review for privilege and relevance before data are turned over to the requesting party

- High recall demanded

Evaluation Campaigns and research datasets

- TREC Legal[1]
  - eDiscovery for a production request
- COLIEE[2]
  - Competition on Legal Information Extraction/Entailment for case law and statute law systems for Canadian and Japanese law systems
- FIRE AILA Track[3]
  - Precedent & Statute retrieval for Indian case law system

1 TREC Legal Track, https://trec-legal.umiacs.umd.edu/
2 COLIEE Competition on Legal Information Extraction/Entailment https://sites.ualberta.ca/~rabelo/COLIEE2021/
3 FIRE AILA Track, https://sites.google.com/view/aila-2020/task-1-precedent-statute-retrieval

Challenges

- Legal domain specific language
  - Domain specific language models
    - LegalBERT: trained on legal documents from CourtListener
- Long documents
  - Relevance on paragraph-level
  - Create summaries of cases
- More precision-oriented task than systematic reviews
  - Need for relevance-based re-ranking

Neural IR approach

- Similar as in web search:

1. First stage Retrieval
   - Boolean queries or
   - Dense retrieval approaches (in next lecture)

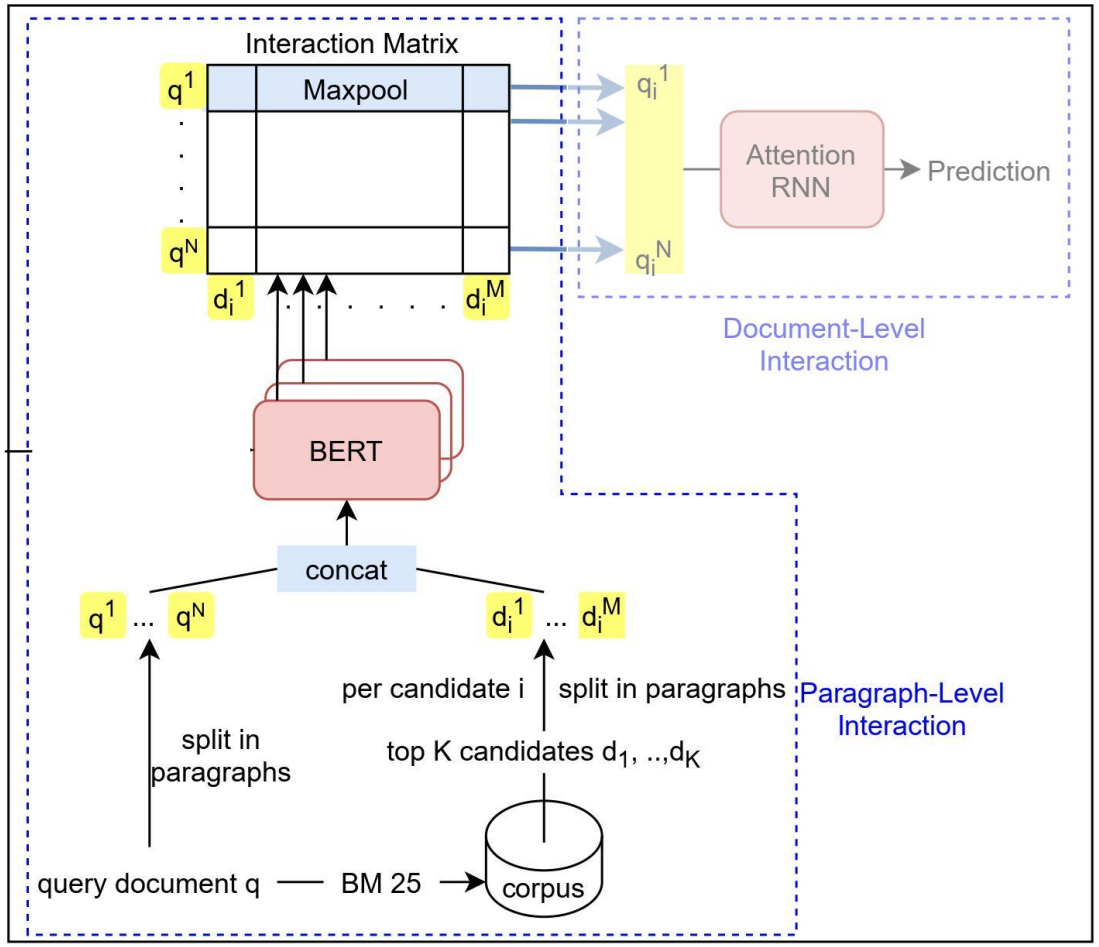
2. Neural re-ranking
   - Domain specific language models
   - Domain specific relevance annotations
   - Handling long documents with summaries or on paragraph-level by splitting up

Neural re-ranking: BERT-PLI[1] for case law retrieval

- Bert-based re-ranking of first stage retrieved candidates for prior case retrieval, re-ranking is reduced to a binary classification problem
- Handling long documents by splitting up the document in their paragraphs
- Modelling the interactions of queries and candidates paragraphs using a BERT_CAT approach by concatenating each query and candidate paragraphs

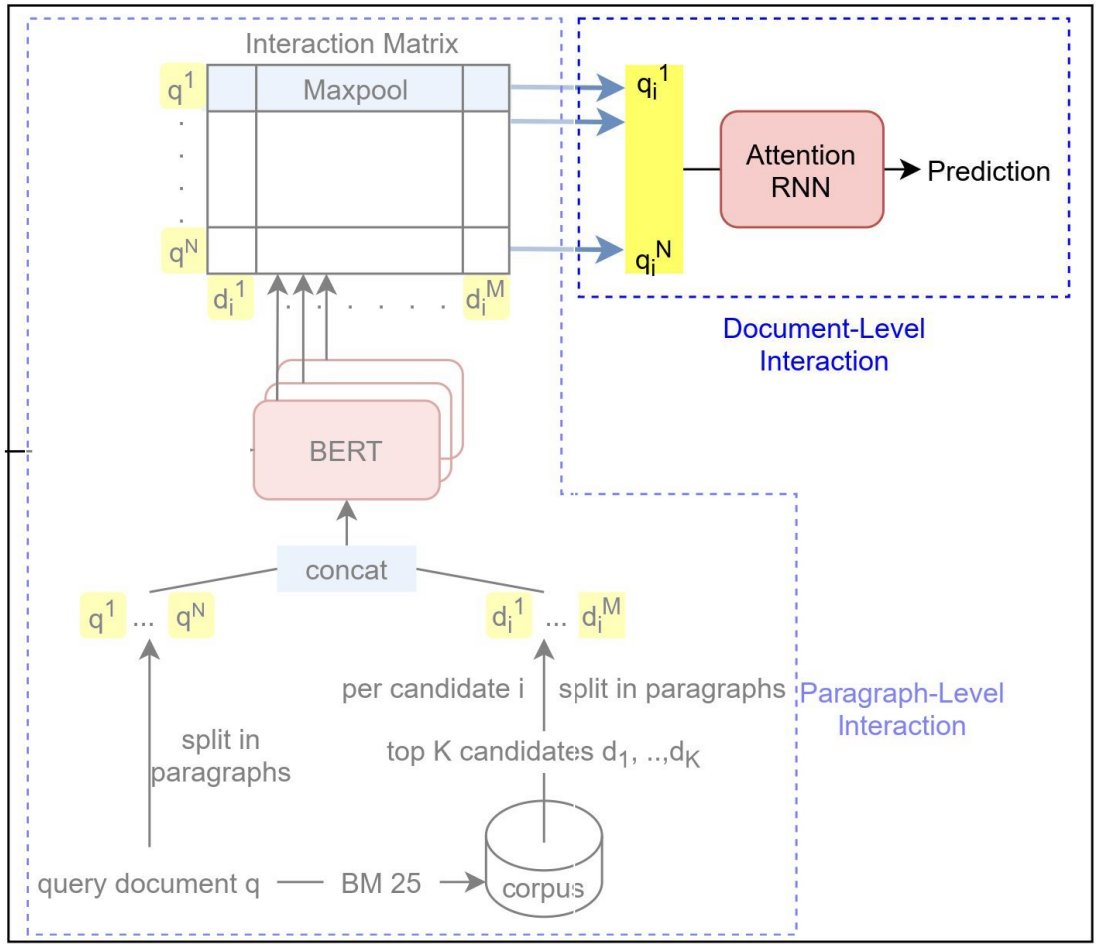
1 BERT-PLI: Modeling Paragraph-Level Interactions for Legal Case Retrieval, Yunqiu Shao

Neural re-ranking: BERT-PLI[1]



1 BERT-PLI: Modeling Paragraph-Level Interactions for Legal Case Retrieval, Yunqiu Shao

Neural re-ranking: BERT-PLI[1]



1 BERT-PLI: Modeling Paragraph-Level Interactions for Legal Case Retrieval, Yunqiu Shao

Summary: Domain specific applications

1. Not only web search, but also search in specific domains

2. Systematic reviews and clinical decision support in medical domain

3. Prior case retrieval and eDiscovery in legal domain

1. Not only web search, but also search in specific domains
2. Systematic reviews and clinical decision support in medical domain
3. Prior case retrieval and eDiscovery in legal domain

Thank You