

Group work on Medical IR

Georgios Peikos, Wojciech Kusa, Annisa Maulida Ningtyas



Agenda

Information Extraction/Expansion with ruled-based approaches

Information Extraction/Expansion with LMs (mainly NER models)

Information Extraction/Expansion based on statistics/similarities

Information Extraction/Expansion with LLMs

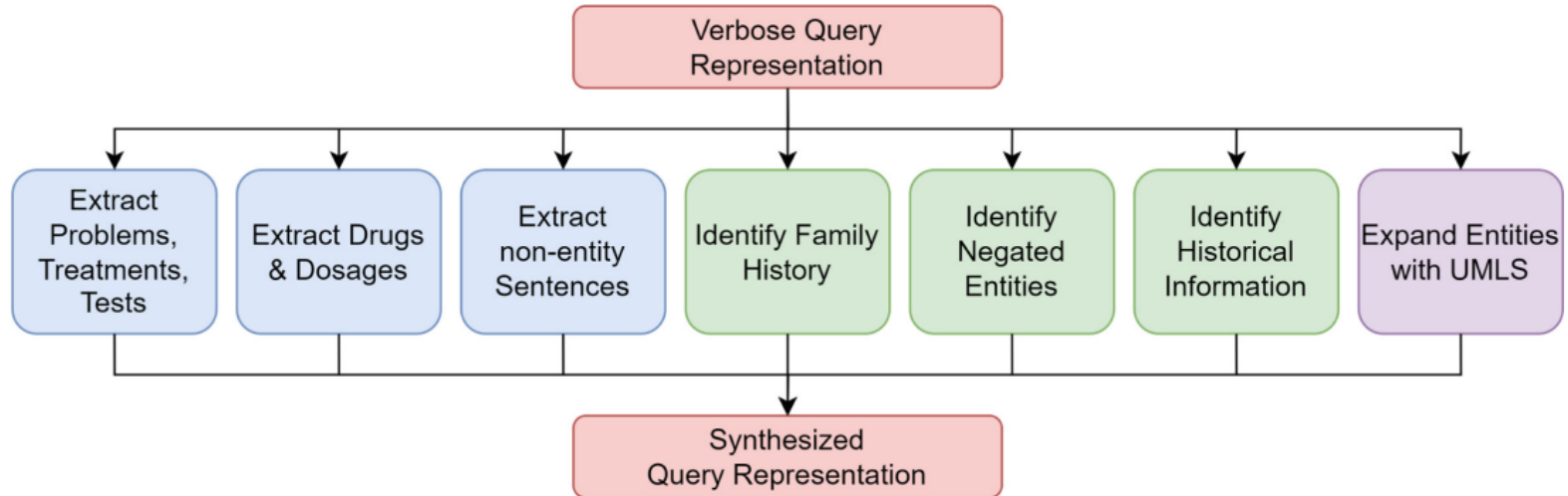


Synthetic patient's case

past medical history / current medical conditions / family description / unrelated

The patient is a 55-year-old man who was recently diagnosed with Parkinson's disease. He is complaining of slowness of movement and tremors. His disease is ranked as mild, Hoehn-Yahr Stage I. His past medical history is significant for hypertension and hypercholesterolemia. He lives with his wife. They have three children. He used to be active with gardening before his diagnosis. He complains of shaking and slow movement. He had difficulty entering through a door, as he was frozen and needed guidance to step in. His handwriting is getting smaller. He is offered Levodopa and Trihexyphenidyl. He is an alert and cooperative man who does not have any signs of dementia. He does not smoke or use any illicit drugs.

General Information Extraction





Agenda

Information Extraction/Expansion with ruled-based approaches

Information Extraction/Expansion with LMs (mainly NER models)

Information Extraction/Expansion based on statistics/similarities

Information Extraction/Expansion with LLMs



What is the purpose of extracting specific pieces of information?

Mainly to improve the quality of the first stage retrieval

Clearer representation of the information need – lower the chances to have topic drift

Task characteristics and task-based relevance

Due to the requirements of the search, we can omit some information that might lead to excluded or not relevant trials.

For example, if we retrieve relying on a patient's family history, e.g., father with diabetes, it is possible that we retrieve not relevant trials.



How can we extract these pieces of information?

Custom rules-based approaches

Spacy

Stanza

SciSpacy, medSpacy

ConText algorithm

Pre-trained Name Entity Recognition models trained and fine-tuned on medical data

Large Language models (in-context learning)



Agenda

Information Extraction/Expansion with ruled-based approaches

Information Extraction/Expansion with LMs (mainly NER models)

Information Extraction/Expansion based on statistics/similarities

Information Extraction/Expansion with LLMs



Extracting informative pieces of information?

Mainly to improve the quality of the first stage retrieval

Clearer representation of the information need – lower the chances to have topic drift

Agnostic approach in terms of the semantic meaning of the extracted terms

Mainly we rely on the importance of a term or phrase within a given context



How can we extract these pieces of information?

KeyBERT

<https://github.com/MaartenGr/KeyBERT>

Kullback-Leibler divergence for Informativeness (KLI)

[Verberne, S., Sappelli, M., Hiemstra, D., & Kraaij, W. \(2016\). Evaluation and analysis of term scoring methods for term extraction. Information Retrieval Journal, 19, 510-545.](#)



Agenda

Information Extraction/Expansion with ruled-based approaches

Information Extraction/Expansion with LMs (mainly NER models)

Information Extraction/Expansion based on statistics/similarities

Information Extraction/Expansion with LLMs



Benefits of using LLMs for Information Extraction

Unity

Instead of fine-tuning and maintain several NER models, you only use one

Performance, Simplicity, Generalization

In-context learning (prompting)



Prompt Examples

Prompt Text	Model's output (example)
<p>Please identify the patient's medical condition and current treatments, including any alternative names, abbreviations, or synonyms for these terms, as well as any additional criteria that may be important for identifying clinical trials of interest. Respond with a comma-separated list of keywords that will be used for search. Do not elaborate or explain.</p> <p>Patient's medical note: { <i>Clinical Note</i> }</p>	<p>“anaplastic astrocytoma”, spine, “lower extremity weakness”, “urinary retention”, “Foley catheter”, “high-dose steroids”, hypertension, “chronic pain”, “T-L spine”, radiation, “RLE weakness”, temozolomide, “CPT-11 Weekly”, Avastin</p>

Prompt Text

Model's output (example)

Follow my instructions precisely to extract the requested information from a patient's medical clinical note. Do not explain or elaborate. Respond with exactly what I request, and reply in the requested format.

From the following clinical note, resolve all the abbreviation mentioned in the text, and then extract: the patient's, age, gender, medical problem, diagnosis, diseases, symptoms, medications, drugs, dosages, treatments, medical history, family history, lifestyle factors, lab examinations, lab results, vital signs. Add MeSH terms that are relevant to the patient's medical problem, diagnosis or disease. If you are unable to extract the information, write 'N-A'. Answer in JSON format:

```
{ "answer":{
"abbreviations": "resolved abbreviations",
"diagnosis": "", "medical_problem": "",
"diseases": "", "drug": "", "dosages": "", "symp-
toms": "", "treatments": "",
"medications": "", "medical_history": "",
"family_history": "", "lifestyle_factors": "",
"lab_examinations": "",
"lab_results": "", "vital_signs": "",
"gender": "",
"age": "", "MeSH_terms": "" }}
```

Medical note: {Clinical Note}

```
{ "answer": {
"abbreviations": "s-p: status post, RLE: right
lower extremity, MRI: magnetic resonance imaging,
t10-11: thoracic vertebrae 10 to lumbar vertebrae 1,
CPT-11: irinotecan, Q2 weeks: every two weeks",

"diagnosis": "anaplastic astrocytoma",
"medical_problem": "anaplastic astrocytoma of the
spine, severe lower extremity weakness and urinary
retention",

"diseases": "hypertension, chronic pain",
"drug": "steroids, temozolomide, irinotecan,
Avastin",
"dosages": "high-dose steroids, temozolomide 7
days on and 7 days off, CPT-11 Weekly x4 with
Avastin Q2 weeks- 2 weeks rest and repeat cycle",

"symptoms": "lower extremity weakness, right
anterior thigh numbness, urinary retention",

"treatments": "field radiation, temozolomide,
CPT-11 with Avastin",
"medications": "steroids, temozolomide, irinotecan,
Avastin",
"medical_history": "anaplastic astrocytoma of the
spine",
"family_history": "N-A", "lifestyle_factors": "N-A",
"lab_examinations": "N-A", "lab_results": "N-A",
"vital_signs": "N-A",
"gender": "male", "age": "45",

"MeSH_terms": "Astrocytoma, Spinal Cord
Neoplasms, Lower Extremity, Urinary Retention,
Steroids, Temozolomide, Irinotecan, Bevacizumab,
Radiation, Magnetic Resonance Imaging" }}
```

Able to identify and structure
patient related information



Hands - on

We will go step-by-step and introduce some approaches for information extraction, you can use them with the previous retrieval pipeline, to process the queries

Approaches: SciSpacy, KeyBERT, BERT-based LMs, LLMs and prompting.