



INFORMATION RETRIEVAL AND RECOMMENDER SYSTEMS

Laboratory 5 – Hands-on IR Experiments on
Clinical Trials Retrieval

Georgios Peikos

georgios.peikos@unimib.it

Building U14, DISCo

University of Milan-Bicocca

Viale Sarca 336, 20126 Milan

Agenda

Clinical Trials Retrieval

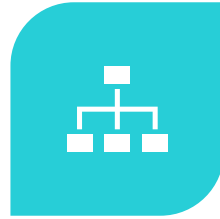
- Task Presentation and Discussion
- Teams of three and develop IR systems
 - Indexing
 - Retrieval
 - Evaluation
- Best team in terms of $P(\text{rel}=2)@10$ gets 1 point in the exam



Format of this LAB



20-30 MINUTE
PRESENTATION OF THE
TOPIC.



YOU WILL HAVE **1 HOUR
AND 30 MINUTES** TO WORK
IN TEAMS.

BRAINSTORM IDEAS, RUN
BASELINES, CREATE YOUR
IDEA.



AT THE **45-MINUTE MARK**
OF THE TEAM SESSION
EACH TEAM WILL WRITE
DOWN THE BEST
P(R=2)@10 VALUE THEY
ACHIEVED.



AT THE END OF THE
SESSION, WE **WILL REVIEW
THE RESULTS** AND PICK THE
WINNING TEAM.

Ideas to work on



CHOICE OF INDEXING
STRATEGY



QUERY PROCESSING AND
FORMULATION



RANKING ALGORITHM
TUNING: USE A BASELINE
RANKING (E.G., BM25)
THEN ADJUST ITS
PARAMETERS (K1, B) OR
INCORPORATE FIELD
WEIGHTING (FOR EXAMPLE
GIVE TITLE HIGHER
WEIGHT).



FILTERING AND RESULT
POST-PROCESSING

Introduction to Clinical Trials Retrieval

Modified slides from the Working Group on Medical IR, presented in the 14th ESSIR by
Georgios Peikos, Wojciech Kusa, Annisa Maulida Ningtyas



Topics

Introduction to clinical trials retrieval

Users

Task formulation

Queries

Documents

Notion of Relevance – retrieval goal

Your task and materials





Clinical trials & Clinical trials retrieval

Clinical trials

Experiments conducted in the development of new medical treatments, drugs or devices

Task Description

A user must allocate patients to clinical trials (find eligible participants)



Users

Laypeople (e.g., patients)

Employees in health organizations

Researchers

Medical Experts, e.g., physicians



Formulation of the retrieval task

Retrieval Formulation

Given a **synthetic patient's case** as a query, retrieve eligible clinical trials from a collection of clinical trials



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.



A clinical trial document

Structured Documents

Title, Description, Condition, Eligibility Criteria (important for this task)

Eligibility Criteria

Mention a trial's inclusion and exclusion criteria, along with its demographic information (gender, age)

Inclusion Criteria:

- Male or female adults ages 18-40 or of 65 and or older at the time of enrollment
- Eligible to receive Fludax® (MF59Flu) or Fluzone® (HDFlu) if age 65 or older
- No history of anaphylactic reaction to gelatin, neomycin, or other vaccine component
- Not pregnant
- No immunosuppression or immunodeficiency
- No acute illness at time of vaccination
- Determined by medical history and clinical judgment to be eligible for the study, by being generally healthy, with no autoimmune or immunosuppressive conditions and having stable current medical conditions (subjects with preexisting stable disease, defined as disease not requiring significant change in therapy or hospitalization for worsening disease 12 weeks before receipt of study vaccine, will be eligible. A change in dose or therapy within a category (e.g., change from one nonsteroidal anti-inflammatory drug to another) is allowed. A change to a new therapy category (e.g., surgery or addition of a new pharmacological class) is only allowed if it is not caused by worsening disease. A change to a new therapy category caused by worsening disease is considered significant and therefore ineligible for enrollment.
- Patients with diabetes mellitus are eligible for inclusion if they have had a hemoglobin A1c measurement of <8.0 within the past 6 months prior to enrollment. These hemoglobin A1c measurements are recommended at least twice yearly by the American Diabetes Association (ADA), and the target levels here are representative of the goals of the ADA. These hemoglobin A1c levels will ensure that these participants have good glycemic control. (American Diabetes Association. American Diabetes Association Position Statement: Standards of Medical Care in Diabetes- 2015. Diabetes Care 2015;38(Suppl. 1): S1-S94)
- Able to follow study procedures in the opinion of the investigator
- Expected to be available for the duration of the study
- Weighs >110 lbs

Exclusion Criteria:

- Known or suspected immunodeficiency or receiving treatment with immunosuppressive therapy including cytotoxic agents or systemic corticosteroids (e.g., for cancer, HIV, or autoimmune disease). If systemic corticosteroids have been administered short term for treatment of an acute illness, subjects will be included if corticosteroid therapy (inhaled, intranasal, and intra-articular corticosteroid therapy is permitted) has been discontinued for at least 30 days.
- Serious chronic medical conditions including metastatic malignancy, severe chronic obstructive pulmonary disease requiring supplemental oxygen, end-stage renal disease with or without dialysis, clinically unstable cardiac disease, or any other disorder that, in the investigator's opinion, precludes the subject from participating in the study. Diabetic patients will be excluded if they do not have a hemoglobin A1c measurement within the past 6 months or if they had a hemoglobin A1c measurement of an A1c >8.0



What makes a clinical trial relevant for a given patient (query)?

Highly Relevant Trials (eligible trial)

Those for which the patient covers as many inclusion criteria as possible (i.e., all) and, at the same time, none of the exclusion criteria

Relevant Trials (excluded trials)

The patient is excluded from the clinical trial, but the condition of the query matches the patient's condition

Not Relevant

All the rest



Text Retrieval Conference (TREC)

Information: <https://trec.nist.gov/overview.html>

Clinical Trials Tracks: <https://www.trec-cds.org/2021.html>



Current System

<https://clinicaltrials.gov/>



Working on this task

You will be given:

- A collection of clinical trials in a .csv format for you to index
- A set of queries
- Relevance judgments

You will be asked:

By the end of this working group, each team will have to develop one end-to-end retrieval pipeline that:

1. Analysis of the collection of documents
2. Processes the queries
3. Indexes the collection
4. Retrieves eligible clinical trials