Improving Case Based Search of Clinical Trials

Contributors :-

Pranjal Doshi
Udit Agrawal
Mayank Jain

What are Clinical Trials?

Clinical Trials are experiments or observations done in Clinical Research.

- They are designed to answer specific questions about medical, surgical or behavioral intervention of people.
- A clinical trial may be used to find out if a new treatment is more effective and has less harmful effects than standard treatment.
- They are conducted after receiving approval of Health Authorities.

Why and Who are searching for Clinical Trials ?

There are mainly 3 type of persons searching for clinical trials :-

1) **Doctors** : Usually, Doctors search for them in order to get latest researches about the trial and the up to date information of drugs.

2) **Trial Participants** : These are the people who have some disease and are recommended by the doctors to search for the particular trials related to it. If they don’t get the required results, they search for something relevant to the trial.

3) **Clinitians** : They are also people working in clinics searching for trials and since they have direct contact with patients so they recommend the trials to patients.

Why clinicaltrials.gov ?

- It is our data source as it contains the studies of most recent clinical trials (public or private funded) done across the world.
- Also within a month of completion of a trial, Research papers get published for that trial here.
- This is a government website maintained by NLM (US) and the most trusted of all.

What are we trying to solve?

In traditional Information Retrieval we have queries in the form of text.
E.g. Who is president of India ?

So what's different in this ??

We've unique trial Id's as input query and trial Id's which are similar to input trial are the output of the query.

Why do we need to do this ?

Because our user is naïve. He don't know much about medical terminologies. For him it is difficult to find out what query he should fire next to get similar result.

What is our noble contribution?

Previous Works

- In the Baseline paper mentioned, researchers have implemented the system using only eligibility criteria to classify the clusters.
- High threshold is taken to get a good precision. But it reduced the recall.
- We tried merging the results of eligibility criteria and brief summary together to improve quality of clusters but the approach failed miserably giving poor clusters.

Our Approach

- We have considered eligibility criteria as our primary classifier but with a reduced threshold.
- This results in a higher recall but a reduced precision.
- In order to compensate the loss of precision we are using brief summary.

Why this approach will work better?

- Lower threshold (~0.7) ensures all relevant trials are extracted since higher threshold doesn't include many trials needed by people.
- Eligibility criteria ensures that those results are relevant for the user.

Reference :- Clustering clinical trials with similar eligibility criteria features (Tianyong Hao, Alexander Rusanov)

Chrome File Edit View History Bookmarks People Window Help 51% Thu 11:32 PM

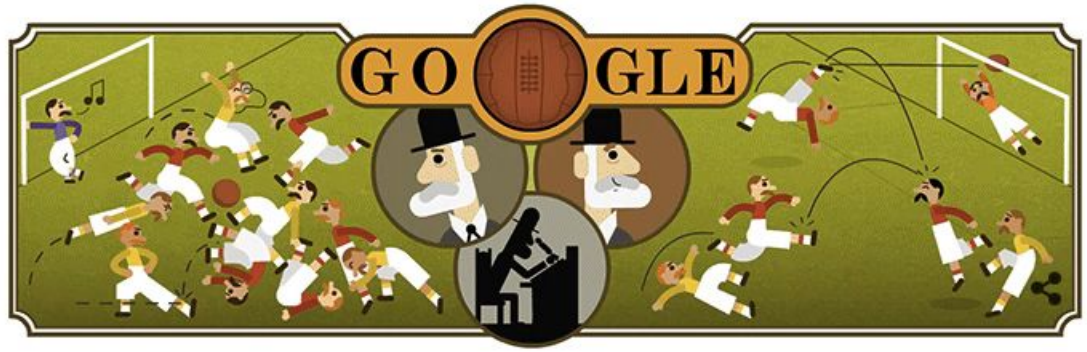
Google

mayank

Secure | https://www.google.com

Apps Bookmarks Inbox (6,274) - j.ma... online2pdf.com IR MOODLE ML MOODLE Other Bookmarks

Gmail Images m



Google Search I'm Feeling Lucky

Google offered in: हिन्दी বাংলা తెలుగు मराठी தமிழ் ગુજરાતી ಕನ್ನಡ മലയാളം ਪੰਜਾਬੀ

India

Advertising Business About Privacy Terms Settings

GOALS

- Improve Cluster Quality
- Improve Cluster Coverage
- Representation of trials as feature vectors

How are we going to achieve our goal ?

Using n-gram model

We'll be using some naïve classification to find vectors that are major deciding factors in order to understand domain specific knowledge of the model so that there is no bias in our understanding.

THANK YOU