Data Mining from Clinical Research Records for Healthcare

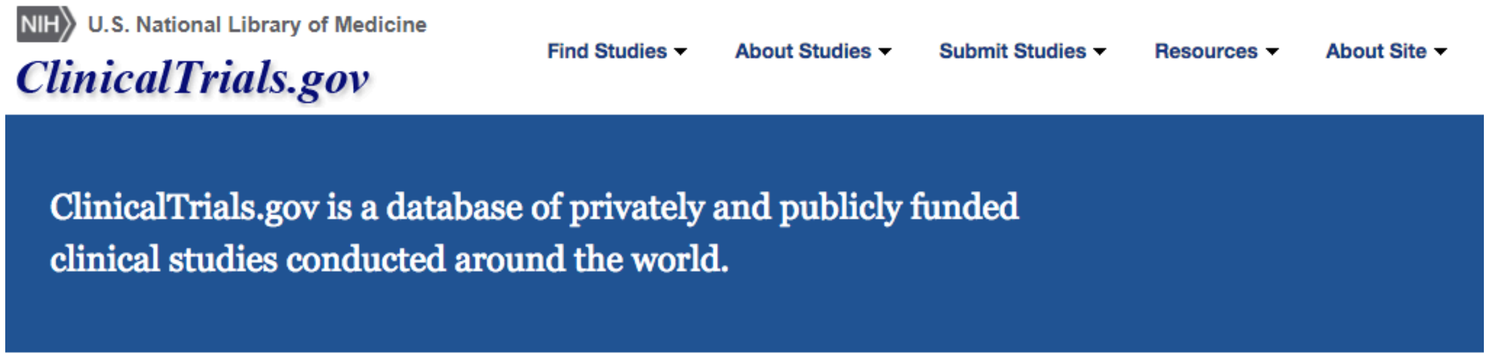
Jie Yang

Ph.D. Student at Department of Biomedical Engineering, Columbia University

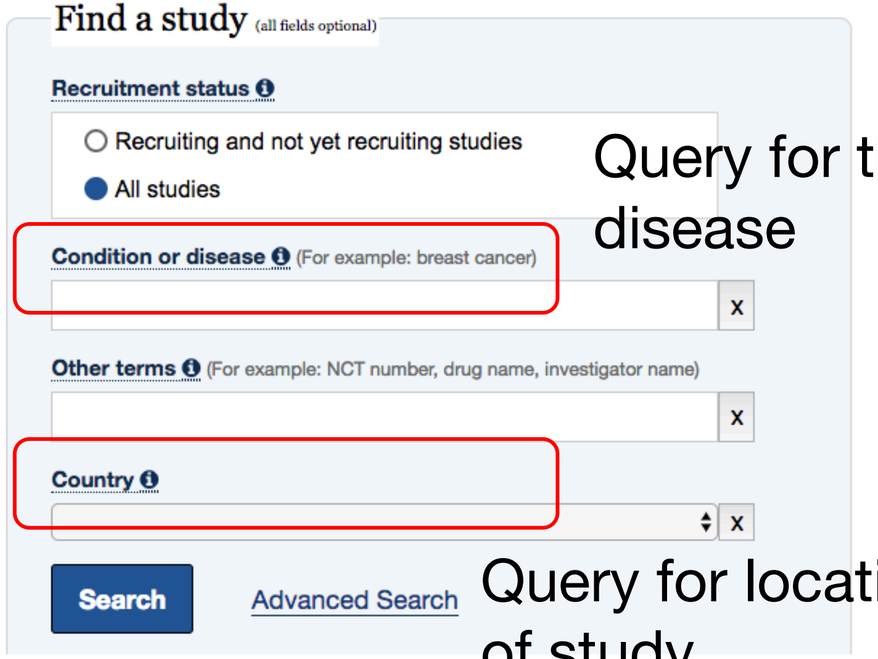
Research Area: Machine Learning for Medical Image Classification / Segmentation for Healthcare

Introduction

https://clinicaltrials.gov/



5GB of text records



Query for the disease

Query for location of study

Introduction

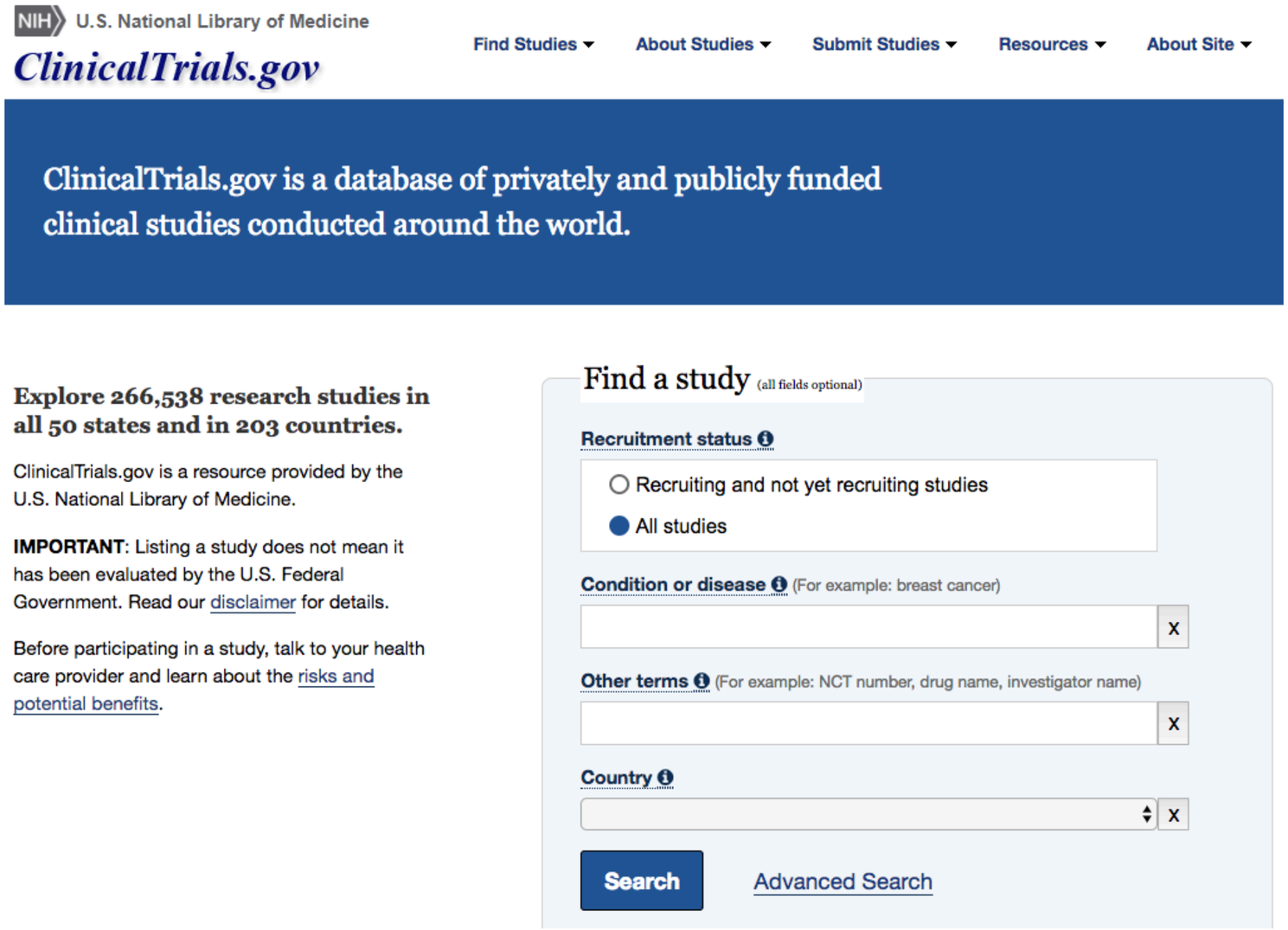
https://clinicaltrials.gov/

Rich information:

- Disease type;
- Study location;
- Proposed treatment;
- Detailed description;
- Starting date;
- Ongoing / final results; …

For patient use

For research use



Proposal

Develop data science interface for mining this dataset:

Time consuming to query for the information.

Better better summarization and visualization of the data:

- E.g. to inspect ***past*** treatments, ***trending*** treatments, and predict ***future*** plans.

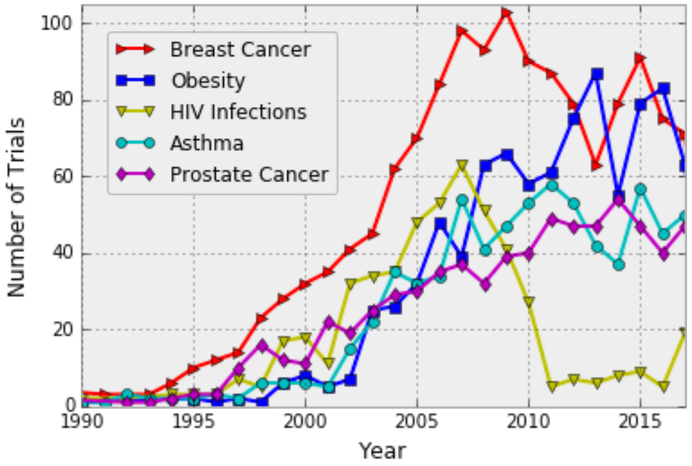
for researcher

for patient

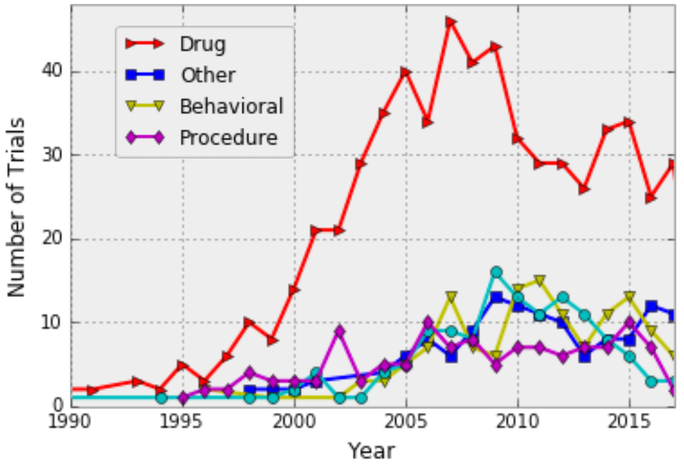
for company

Proposal

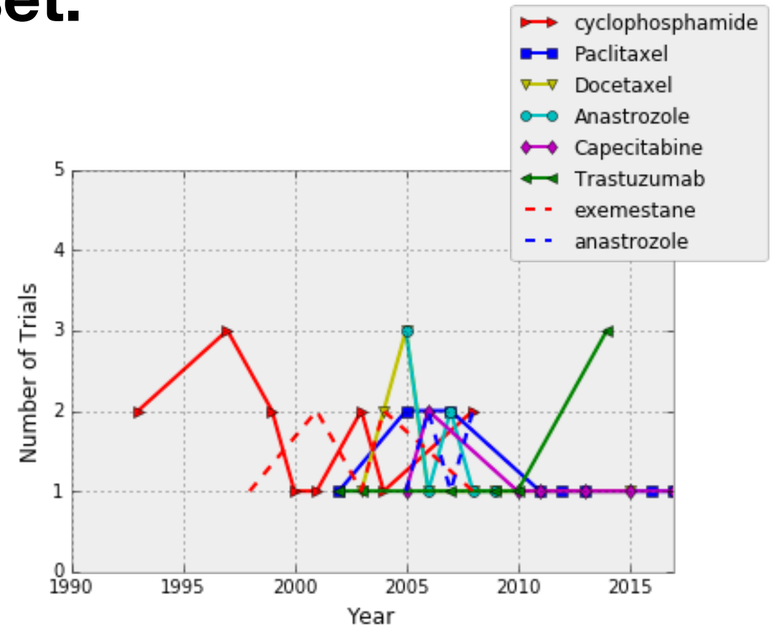
Develop data science interface for mining this dataset:



Popular Disease



Popular Treatment Type



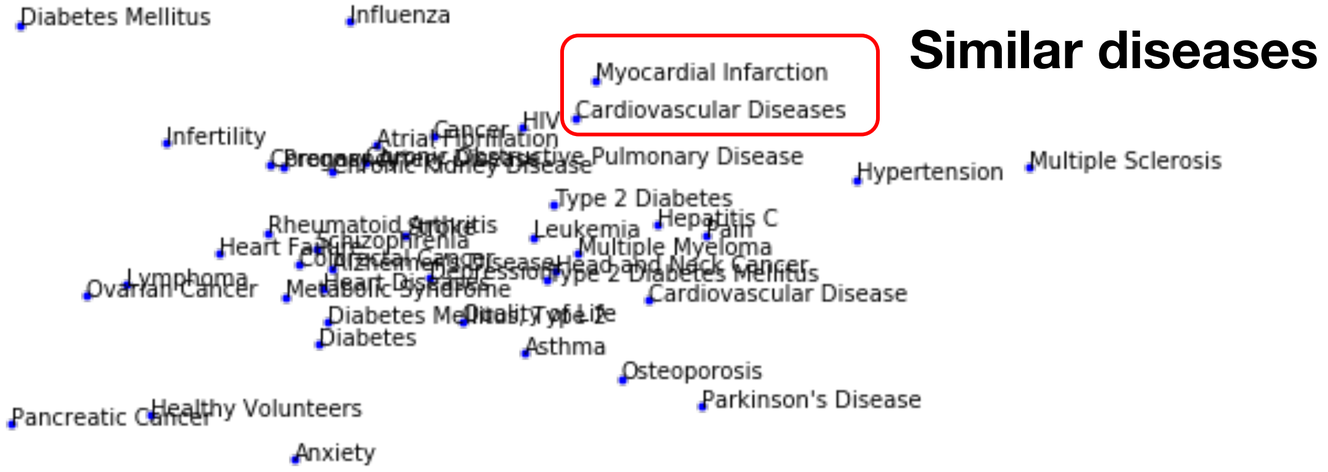
Popular Treatment Name

https://github.com/wanwanbeen/text_mining_ClinicalTrials/blob/master/demo_ClinicalTrials_explore.ipynb

Proposal

Develop data science interface for mining this dataset:

- Different study: Same / similar / correlated disease but registered with different names.
- Better better query with word embedding:



Similar diseases

https://github.com/wanwanbeen/text_mining_ClinicalTrials/blob/master/demo_ClinicalTrials_embedding.ipynb

Proposal

Future plan:

- Learn better embedding from more detailed study description
- Categorize study results for treatment query
- Additional functions: e.g. study location map, …