Automatically Tagging Articles from PubMed with MeSH Terms

25TH MAY 2022

WEIZHONG TANG

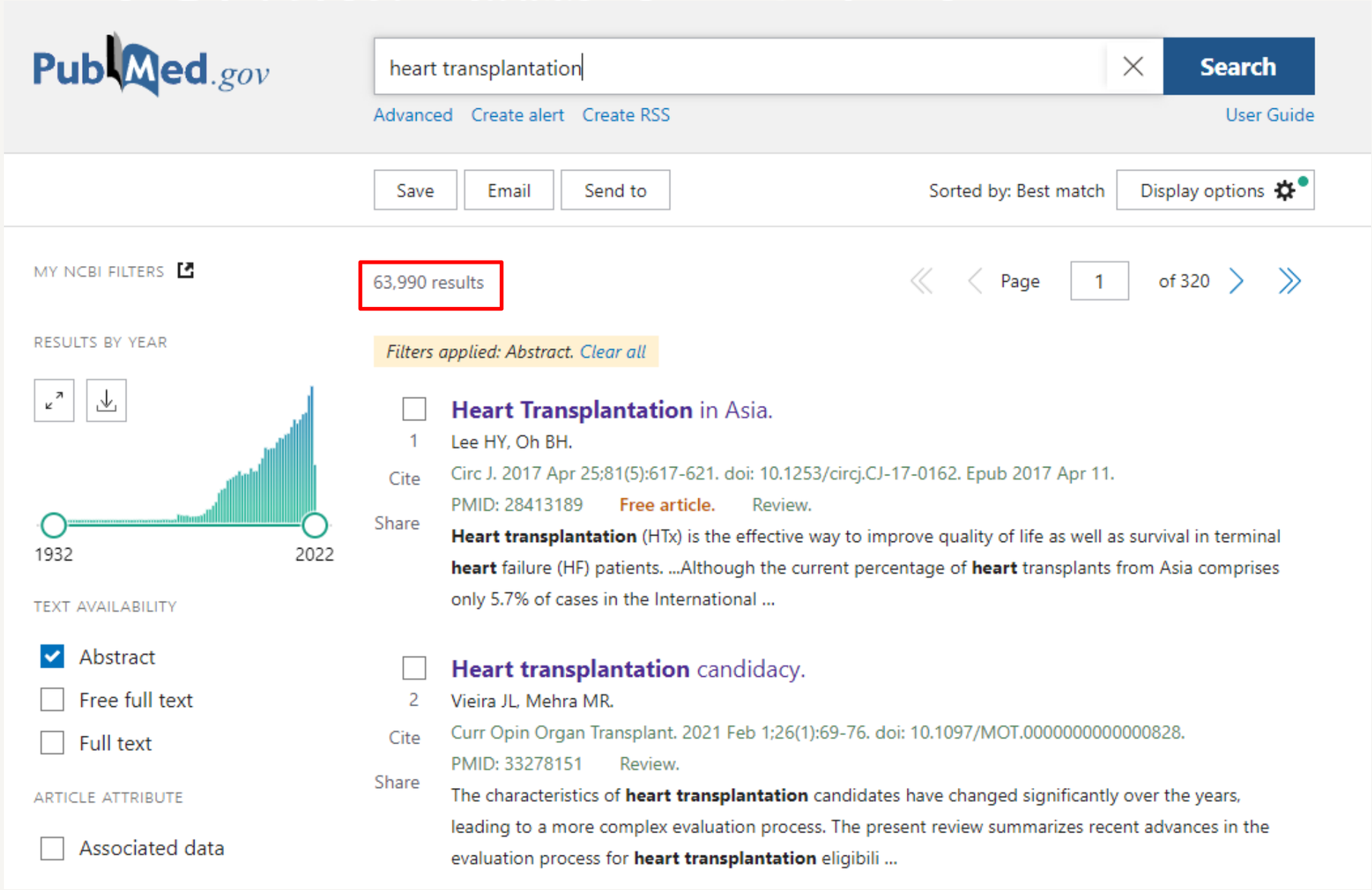
Overview

- Background
- Purpose
- Experiment
- Results
- Conclusion

Background

What is PubMed?

PubMed is a repository of medical articles with a free search engine accessing primarily the MEDLINE database of references and abstracts on life sciences and biomedical topics.



Background

You will see the overview of the article and the most important part we are going to discuss is the MeSH terms

Review > Circ J. 2017 Apr 25;81(5):617-621. doi: 10.1253/circj.CJ-17-0162. Epub 2017 Apr 11.

Heart Transplantation in Asia

Hae-Young Lee [1], Byung-Hee Oh [1]

Affiliations + expand

PMID: 28413189 DOI: 10.1253/circj.CJ-17-0162

Free article

Abstract

Heart transplantation (HTx) is the effective way to improve quality of life as well as survival in terminal heart failure (HF) patients. Since the first heart transplant in 1968 in Japan and in earnest in 1987 at Taiwan, HTx has been continuously increasing in Asia. Although the current percentage of heart transplants from Asia comprises only 5.7% of cases in the International Society of Heart and Lung Transplantation (ISHLT) registry, the values were under-reported and soon will be greatly increased. HTx in Asia shows comparable with or even better results compared with ISHLT registry data. Several endemic infections, including type B hepatitis, tuberculosis, and cytomegalovirus, are unique aspects of HTx in Asia, and need special attention in transplant care. Although cardiac allograft vasculopathy (CAV) is considered as a leading cause of death after HTx globally, multiple observations suggest less prevalence and benign nature of CAV among Asian populations. Although there are many obstacles such as religion, social taboo or legal process, Asian countries will keep overcoming obstacles and broaden the field of HTx.

Keywords: Asia; Heart failure; Heart transplantation.

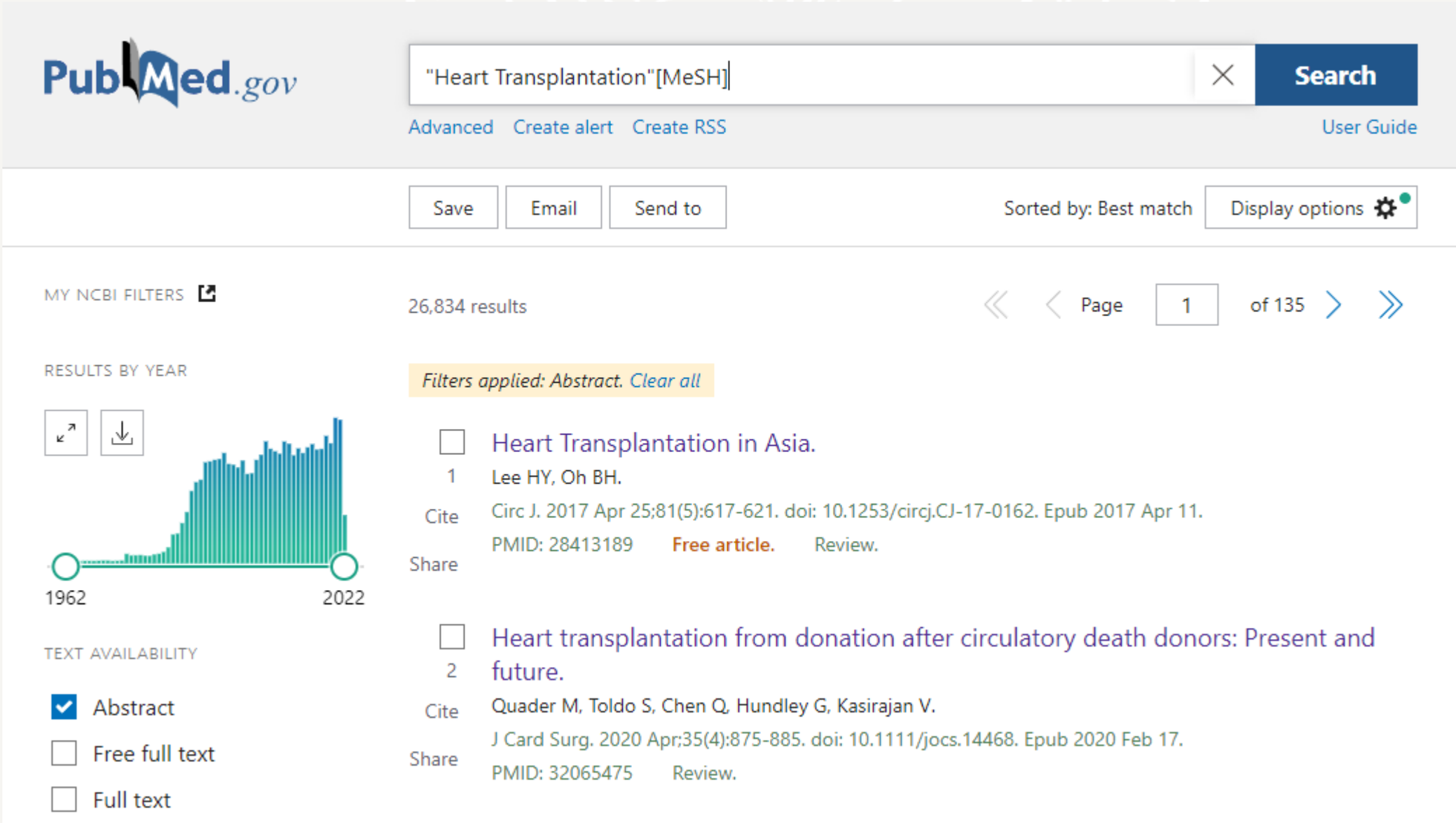
MeSH terms

- Asia
- Cardiovascular Diseases / etiology
- Endemic Diseases
- Heart Failure / complications
- Heart Failure / therapy*
- Heart Transplantation / adverse effects
- Heart Transplantation / methods
- Heart Transplantation / mortality
- Heart Transplantation / trends*
- Humans
- Japan
- Registries
- Taiwan

Background

What is the MeSH term?

Medical Subject Headings (**MeSH**) is a comprehensive controlled vocabulary for the purpose of indexing journal articles and books in the life sciences.



Background

Why do we need to search with MeSH?[1]

By using MeSH terms in your search, the various **synonyms** of a term are automatically included in the search query.

Example:

Heart transplantation

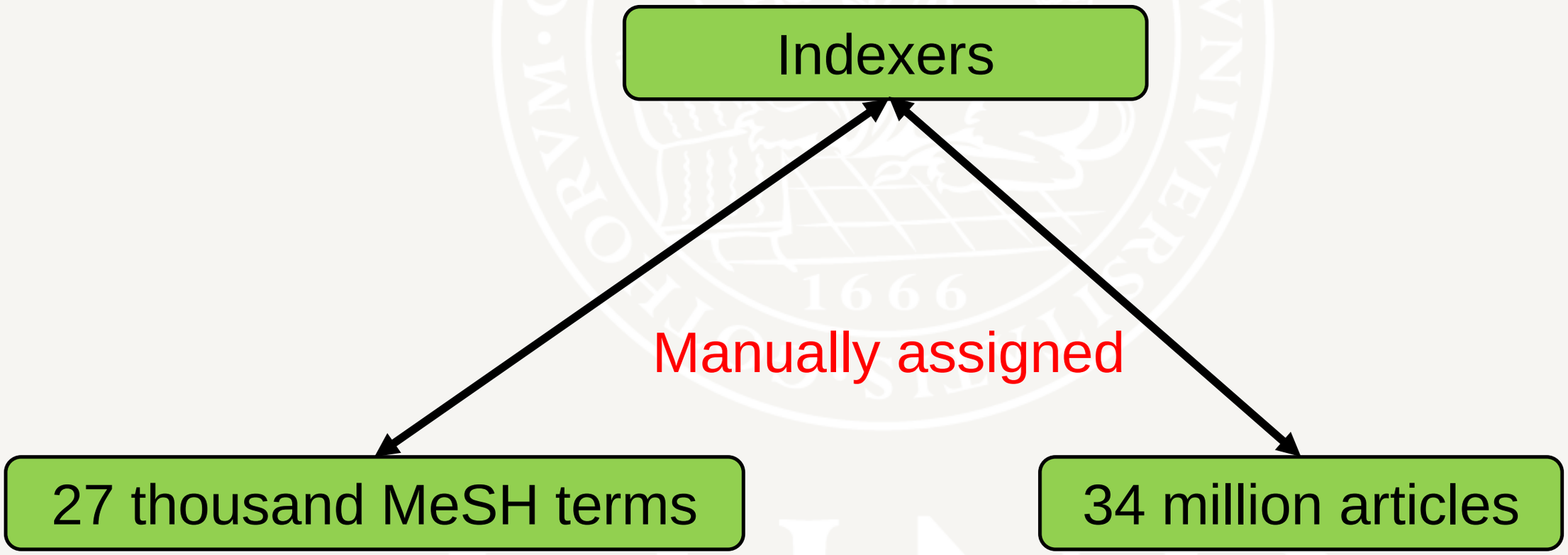
Problems:

1. You probably won't get results for the synonyms: heart/lung transplantation, organ transplant, transplanting, surgical operation, graft etc.
2. You probably will get too many irrelevant results.

[1]https://libguides.ru.nl/PubMedEN/MeSH#:~:text=MeSH%20(Medical%20Subject%20Headings)%20are,(National%20Library%20of%20Medicine).

Background

What's the problem we face to?

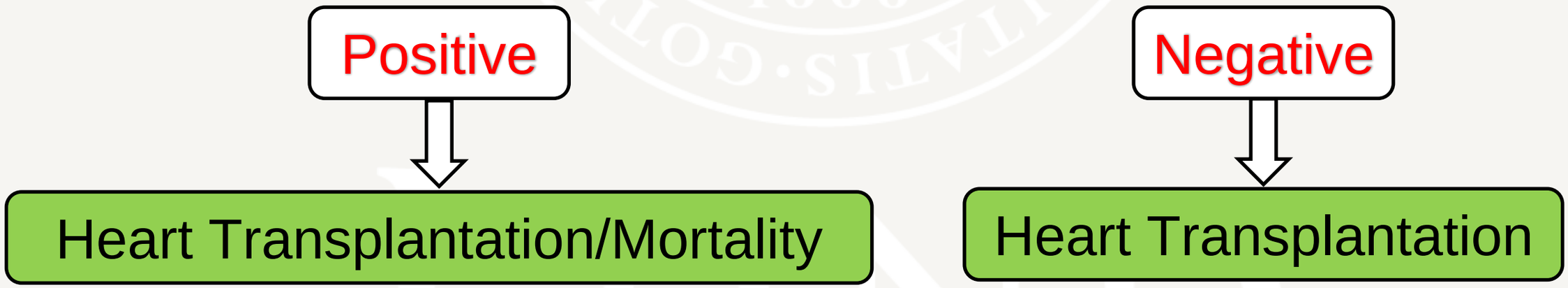


Costly and inevitable to make mistakes

Purpose

Construct a model that can automatically tag articles based on abstracts with more proper MeSH terms on PubMed

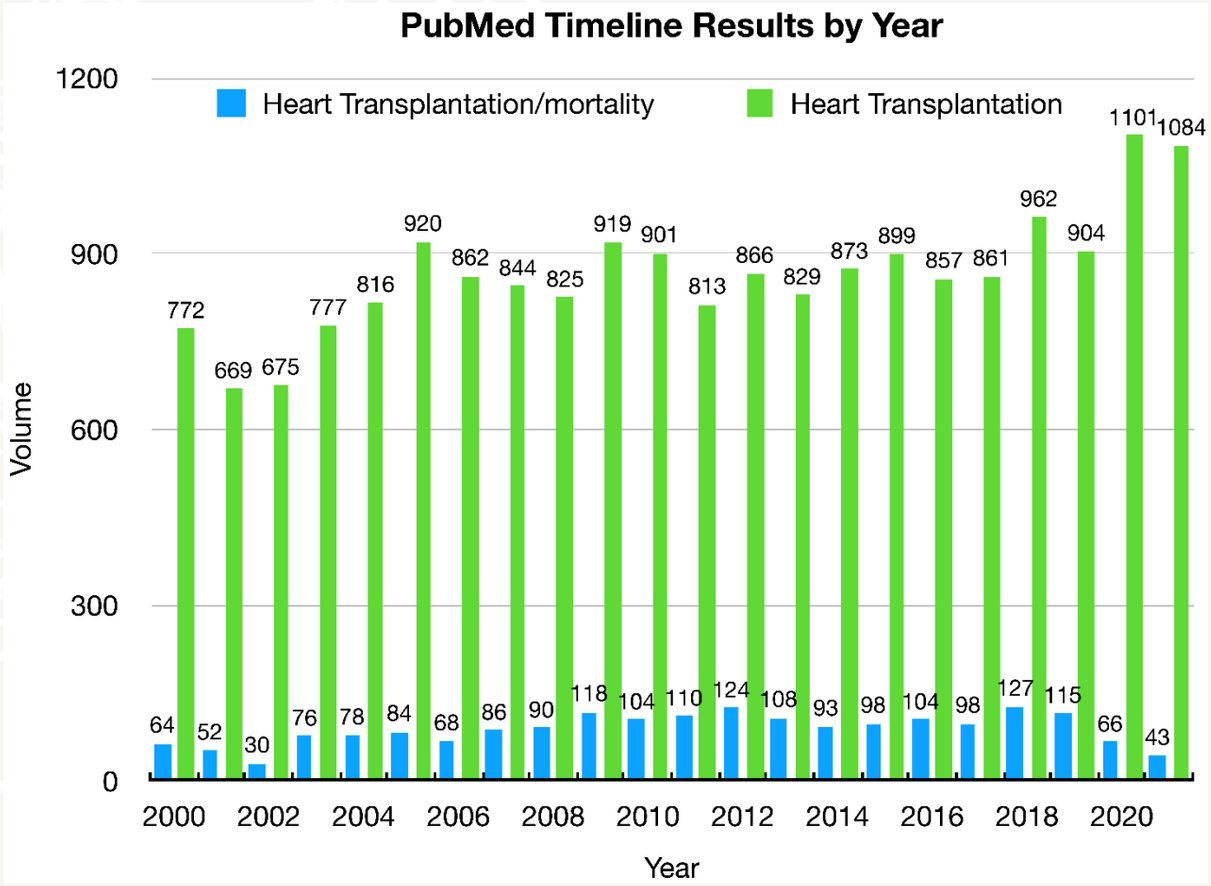
After a small experiment, we decided to focus on two terms



Essentially, the positives are the subclass of the negatives. We wonder how many of the negatives are supposed to be tagged as positive

Experiment-Dataset Construction

- Get PubMed data via FTP
- Build a database on Sqlite3
- Index them by the PMID
- Extract relevant abstracts
- Data cleaning
- Data splitting



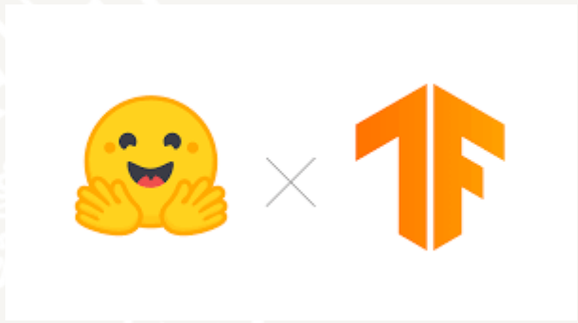
Type	MeSH Term	Qualifier
Positive	Heart Transplantation	Mortality
Negative	Heart Transplantation	

Table 1: Two types of articles

Dataset	Positive	Negative
Training	1000	1000
Validation	500	500
Test	246	13707

Table 2: Dataset Composition

Experiment-Model Construction



Logistic Regression Model

- Count-Vectorizer
- Tfidf-Transformer
- Logistic Regression

DistilBERT Model

- Checkpoint
- Tokenizer
- Data collator
- Transformer

Results-Model Performance

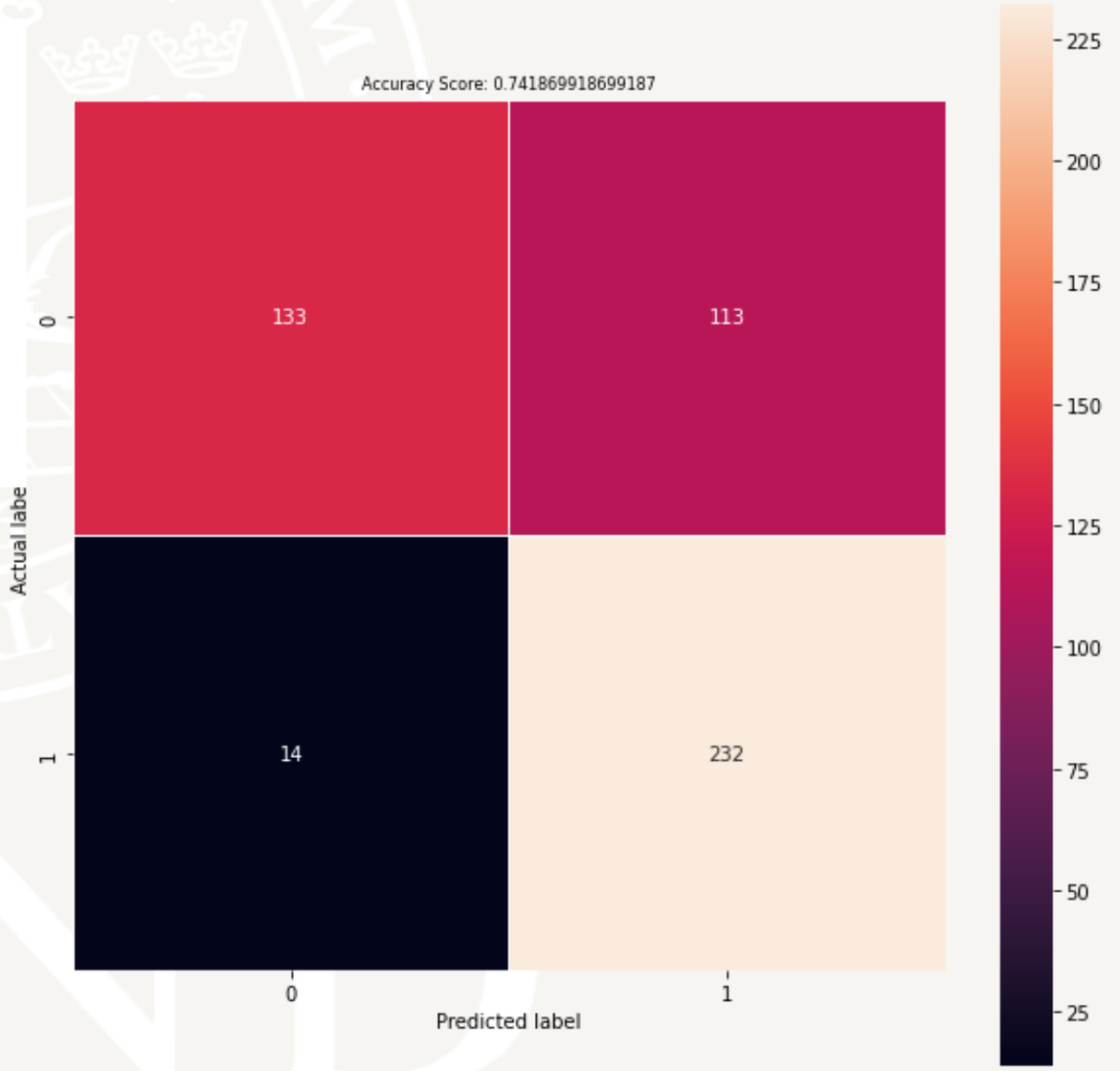
Considering PubMed as ground truth

	Logistic Regression	DistilBERT
Training dataset	**0.93**	**0.96**
Validation dataset	**0.89**	**0.91**
Test dataset	**0.44**	**0.46**

Table 7: The Macro-F1 scores

	precision	recall	f1-score	support
negative	0.90	0.54	0.68	246
positive	0.67	0.94	0.79	246
accuracy			0.74	492
macro avg	0.79	0.74	0.73	492
weighted avg	0.79	0.74	0.73	492

Macro-F1 score: 0.7309773833972695



Overfitting OR Suspicion of ground truth

Results-Model Performance

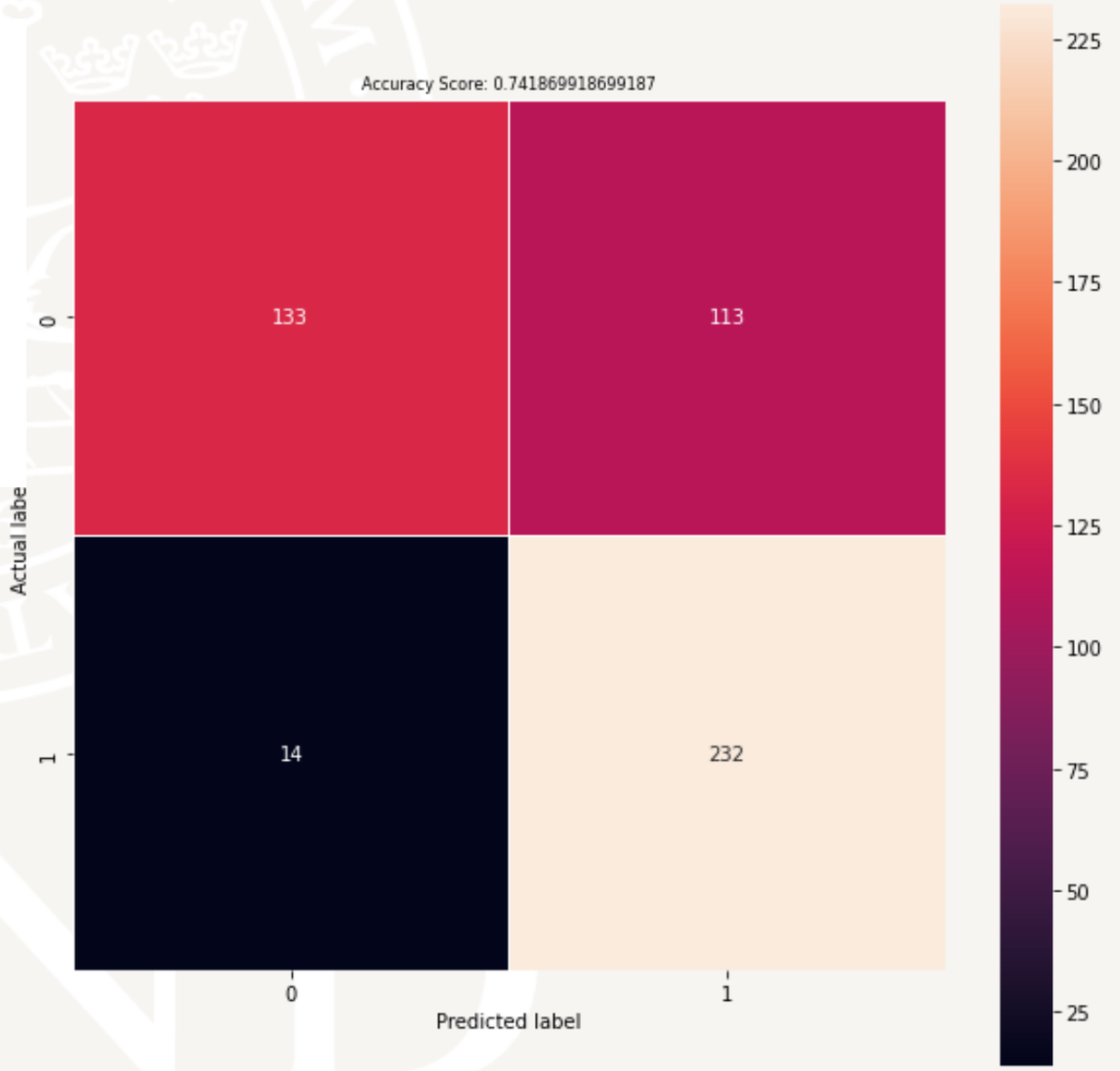
Considering PubMed as ground truth

	Logistic Regression	DistilBERT
Training dataset	**0.93**	**0.96**
Validation dataset	**0.89**	**0.91**
Test dataset	**0.44**	**0.46**

Table 7: The Macro-F1 scores

	precision	recall	f1-score	support
negative	0.90	0.54	0.68	246
positive	0.67	0.94	0.79	246
accuracy			0.74	492
macro avg	0.79	0.74	0.73	492
weighted avg	0.79	0.74	0.73	492

Macro-F1 score: 0.7309773833972695



Overfitting AND Suspicion of ground truth

Results-30 Articles Evaluation

Considering human expert Johan Nilsson cardiac surgeon's answers as ground truth

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Human Expert	x	P		P		x	x	P	P		P	P			
PubMed	P	P		P				P	P		P	P			
Logistic	P	P		P	P	P	P	P	P		P				
DistilBERT	P	P		P	P	P	P	P	P		P	P			
	16	**17**	**18**	**19**	**20**	**21**	**22**	**23**	**24**	**25**	**26**	**27**	**28**	**29**	**30**
Human Expert	P	x	P		P		P	P	P		x	P	P		P
PubMed	P	P	P	P	P							P	P		P
Logistic	P	P	P	P	P	P	P		P		P	P	P		P
DistilBERT	P	P		P	P		P		P		P	P	P		P

Table 5: Evaluation on 30 articles. "P" means tagged or classified as positive; "x" means the modification after discussion.

Results-30 Articles Evaluation

Considering human expert Johan Nilsson cardiac surgeon's answers as ground truth

PubMed	Precision	Recall	Macro-F1
Negative	0.60	0.90	0.72
Positive	0.93	0.70	0.80
Macro-F1	0.77	0.80	**0.76**
Logi_Rreg	Precision	Recall	Macro-F1
Negative	0.78	0.70	0.74
Positive	0.86	0.90	0.88
Macro-F1	0.82	0.80	**0.81**
DistilBERT	Precision	Recall	Macro-F1
Negative	0.80	0.80	0.80
Positive	0.90	0.90	0.90
Macro-F1	0.85	0.85	**0.85**

Table 6: The performances of three classifiers based on Johan's ground truth

Conclusion

PubMed	Precision	Recall	Macro-F1
Negative	0.60	0.90	0.72
Positive	0.93	0.70	0.80
Macro-F1	0.77	0.80	**0.76**
Logi_Rreg	Precision	Recall	Macro-F1
Negative	0.78	0.70	0.74
Positive	0.86	0.90	0.88
Macro-F1	0.82	0.80	**0.81**
DistilBERT	Precision	Recall	Macro-F1
Negative	0.80	0.80	0.80
Positive	0.90	0.90	0.90
Macro-F1	0.85	0.85	**0.85**

Table 6: The performances of three classifiers based on Johan's ground truth

From our experiment, the two models indeed tag these two types of articles with more proper MeSH terms

We may think that this idea can be applied to any MeSH terms as well as articles

Thank you for listening!

Thanks a lot to

Supervisor: Pierre Nugues

Others: Johan Nilsson, Marcus Klang, Dennis Medved

Question time!