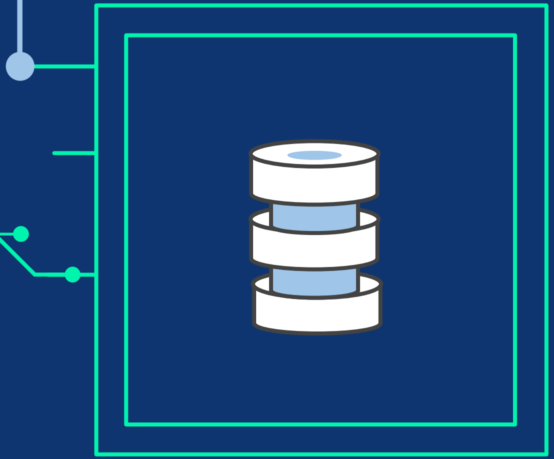
DeepOntoNLP Workshop
ESWC 2021

SUPPORTING ONTOLOGY MAINTENANCE WITH CONTEXTUAL WORD EMBEDDINGS AND MAXIMUM MEAN DISCREPANCY

Authors: Natasha Shroff
Dr. Pierre-Yves Vandenbussche
Dr. Véronique Moore
Prof. Paul Groth



ABOUT THE PROJECT

The project was managed and supported by **Dr. Véronique Moore**, the experimentations were guided by **Dr. Pierre-Yves Vandenbussche** from **Elsevier.** Research was carried out by **Natasha Shroff** (UvA).
The research was supervised by **Prof. Paul Groth** from the **University of Amsterdam**. We received weekly feedback from Elsevier NLP.

CONTENT

01

INTRODUCTION

Study background & research problem

02

EXPERIMENTS

Research methods & experiments

03

DISCUSSION & CONCLUSION

Future work & Q&A

01

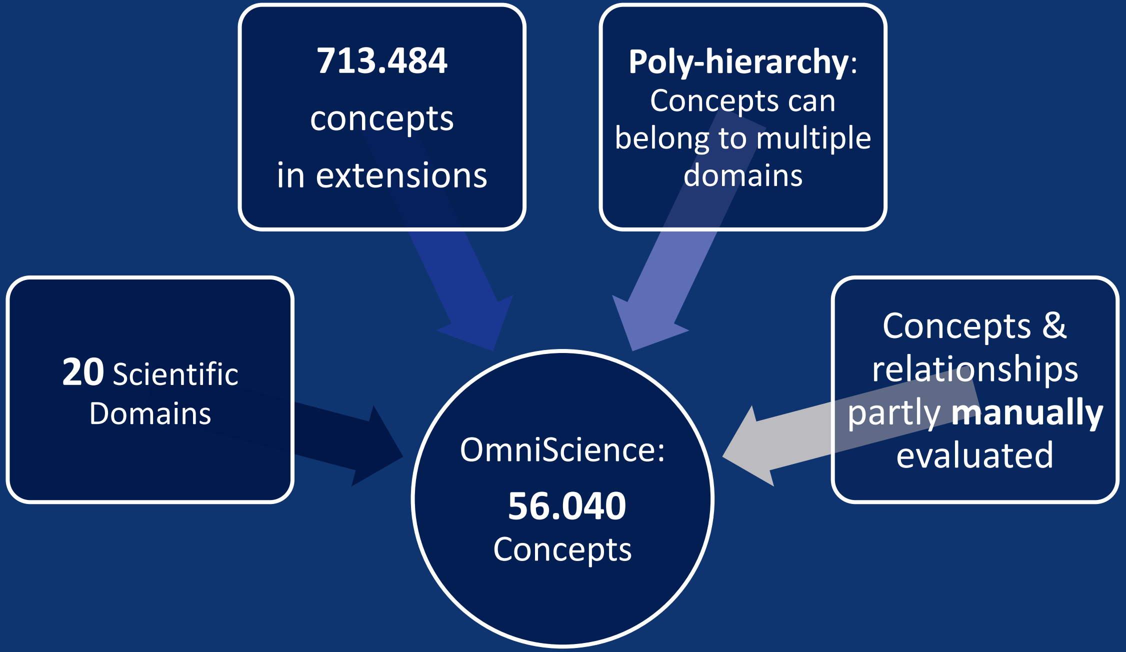
INTRODUCTION

RESEARCH PROBLEM

- Ontologies need to be frequently maintained
- Current maintenance tools cannot fully offer insights into the polysemy of a concept
- Tools are not able to accurately indicate if two similar concepts should be merged
- Curators struggle to get the best possible and unambiguous representation of their domains of interest.

ONTOLOGY

- Large-scale ontology partly maintained manually
- Ontology contains many ambiguous synonyms that need to be organized (merged or not merged)



RESEARCH QUESTION

How can contextual word embeddings be leveraged to advance the automation of ontology maintenance?

02

METHOD & EXPERIMENTS

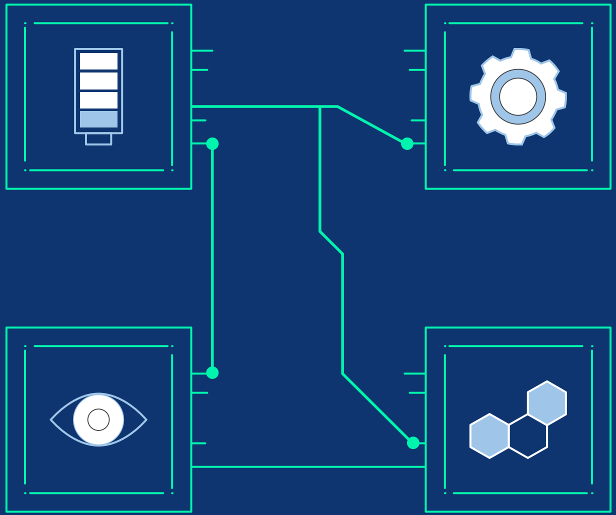
METHODOLOGY

1. EXTRACTING CONCEPTS

Large corpus of scientific articles

2. CREATING EMBEDDINGS

Using Sci-BERT (AllenAI, 2019)



3. USING EMBEDDINGS

MMD score per set of two synonyms

4. EVALUATING SCORES

Larger evaluation of MMD score applicability

9,315,365

Sentences were extracted from the corpus of scientific articles

MAXIMUM MEAN DISCREPANCY

- MMD is based on probability measures in the **R**eproducing **K**ernel **H**ilbert **S**pace
- Measures **difference between distributions** through distance of mean embeddings

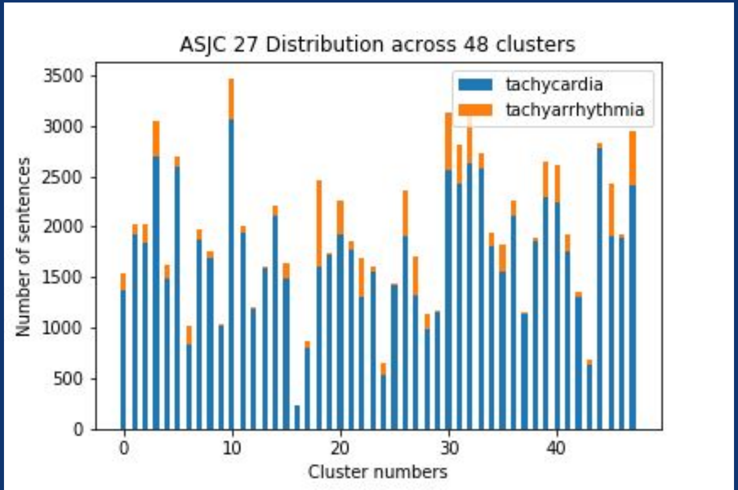
- MMD returns score between 0 & 1:
  - 0 == equal distributions
  - 1 == separate distributions

- Score > 1 : Sample size probably contains less than 20 sentences

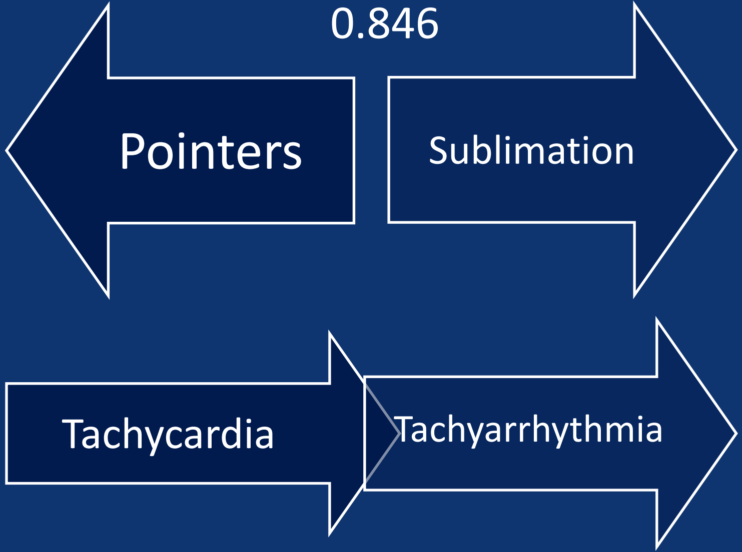
- We used sample sizes of 1000 embeddings per synonym occurrence

EXAMPLE CASE

TACHYCARDIA TACHYARRHYTHMIA



MMD Score:
0.846



MMD Score:
0.098

EVALUATION

SYNONYMS	DOMAIN	MMD SCORE	SAMPLES	SAMPLE RATIO
COST BENEFIT ANALYSIS, COST BENEFIT	Economics	0.0141	743, 1553	0.743
RISK MODELING, RISK MODELLING	Economics	0.1314	238, 200	0.840
GROSS NATIONAL INCOME, GNI	Economics	0.1678	322, 1255	0.322
STANDARD DEVIATION, S.D.	Economics	0.3697	66953, 2611	1

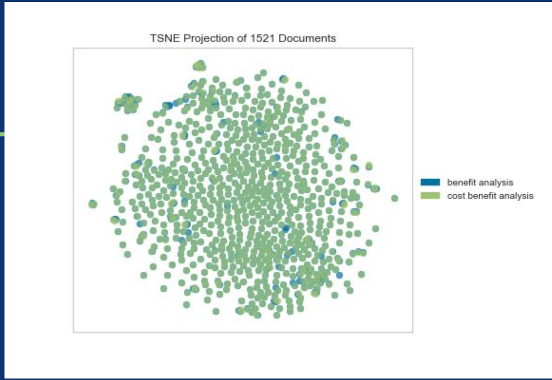
EVALUATION

MMD:
0.00046927

Sample size ratio: 0.95

TERM 1

Cost benefit analysis



TERM 2

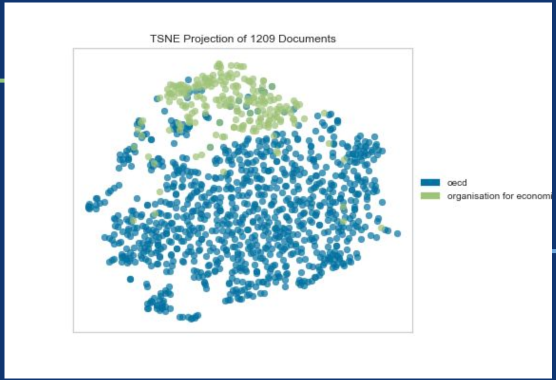
Benefit analysis

MMD:
0.5027

Sample size ratio: 0.2

TERM 1

OECD

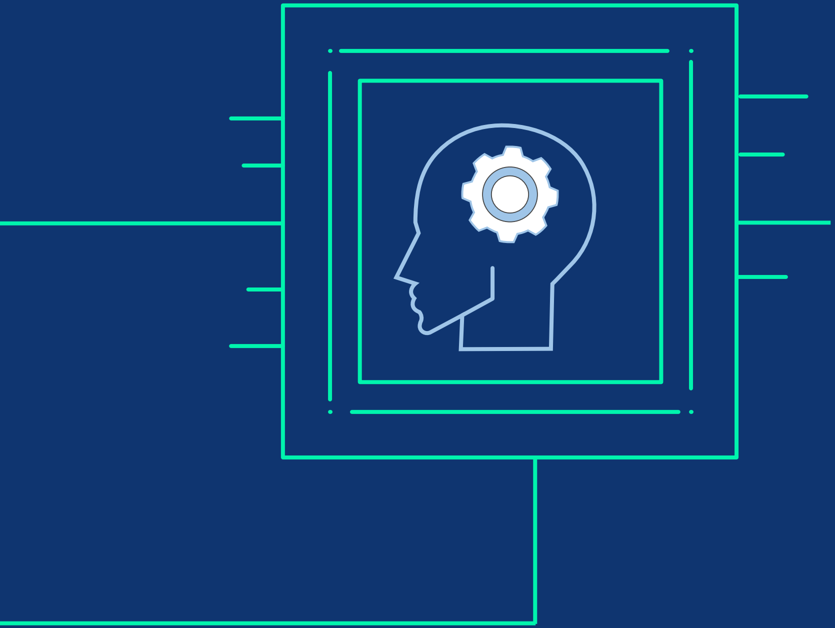


TERM 2

Organization for economic co-operation and development

03

DISCUSSION & CONCLUSION



LIMITATIONS

- Not all scores were as low as expected for pairs of synonyms -> balanced dataset is required for reliable MMD score
- Assumption that extracted occurrences are associated to synonyms that they contain
- Current pipeline is memory intensive

FUTURE WORK

ADDITIONAL SCORE EVALUATION

Test the MMD score applicability with a larger evaluation test set

ONTOLOGY CURATORS ASSESSING SCORES

Concepts with MMD scores below 0.15 to be examined by ontology curators

USE DIFFERENT EMBEDDING MODELS

E.g. contextual embedding models suited for sentences -> Sentence-BERT (Reimers, 2019)

CONCLUSION

EXTRACT SYNONYMS

Sets of two synonyms are required for the score calculation

CALCULATE MMD

Create contextual word embeddings for sets for calculation

ACCEPT OR REJECT MERGE

Support ontology curation with accept/reject suggestion per concept set

THANK YOU!

Authors:
Natasha Shroff
Dr. Véronique Moore
Dr. Pierre-Yves Vandenbussche
Prof. Paul Groth

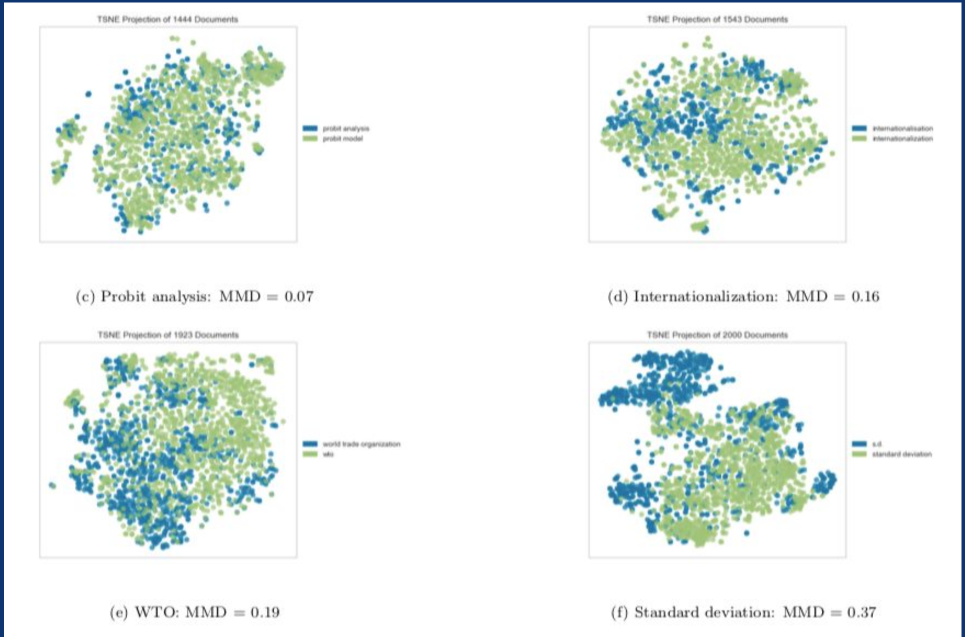
Project on GitHub:
https://github.com/curiousseikatsu/Ontology-Maintenance-with-MMD

REFERENCES

This presentation was based on our paper "Supporting Ontology Maintenance with Contextual Word Embeddings and Maximum Mean Discrepancy". All references can be found in the paper.

Q&A

T-SNE CORPUS VISUALIZATION MAPS



"One very popular method for **visualizing document similarity** is to use t-distributed stochastic neighbor embedding, t-SNE.

By decomposing high-dimensional document vectors into 2 dimensions using **probability distributions** from both the **original** dimensionality and the **decomposed dimensionality**, t-SNE is able to effectively cluster similar documents.

By decomposing to 2 […], the documents can be visualized with a scatter plot." (Yellowbricks, 2019)
https://www.scikit-yb.org/en/latest/api/text/tsne.html

MMD: HOW RELIABLE ARE THE SCORES?

We found that MMD scores with higher sample size ratios (> 0.8) were reliable, as the upsampling of the lower sample sizes to match 1000 sentences reduces the diversity of the term usage.

Why are some scores higher than others?

Because they often include acronyms or different spellings (US vs UK) that have a different context and different journals. This means that their semantic similarity might be lower because their sentences contain different geographic locations, or other location-related words.

Term 1	Term 2	MMD	Silhouette k = 2	Silhouette k = 50	Elbow method	Silhouette k = elbow
benefit analysis	cost benefit analysis	0.00046927	0.075	0.06	19	0.04
child	children	0.039332716	0.08	0.02	19	0.02
industrial structure	industry structure	0.062659373	0.07	0.03	20	0.025
probit model	probit analysis	0.0763714	0.125	0.03	11	0.049
internationalisation	internationalization	0.158332354	0.075	0.03	21	0.03
world trade organization	wto	0.195318492	0.065	0.02	12	0.03
standard deviation	s.d.	0.369657965	0.18	0.04	14	0.065
organisation for economic co-operation and development	oecd	0.502747078	0.09	0.025	24	0.3

Green MMD = Score below 0.1
Yellow MMD = Score between 0.1 and 0.2
Red MMD = Score above 0.2