Attention for Multi-Ontology Concept Recognition - Presentation

Pigott-Dix, Lorcán

Version: Postprint (Verlagsversion)/Postprint (Publisher Version)

Typ/Type: Kongressschrift/Conference Proceeding

Jahr/year: 2023

Quelle/Source: https://repository.publisso.de/resource/frl:6440374

Schlagwörter/Keywords: Named Entity Recognition, Ontology, Deep Learning

Zitationsvorschlag/ Suggested Citation:

Pigott-Dix, Lorcán (2023): Attention for Multi-Ontology Concept Recognition - Presentation. International SWAT4HCLS Conference 2023. DOI: 10.4126/FRL01-006440374

Attention for Multi-Ontology Concept Recognition

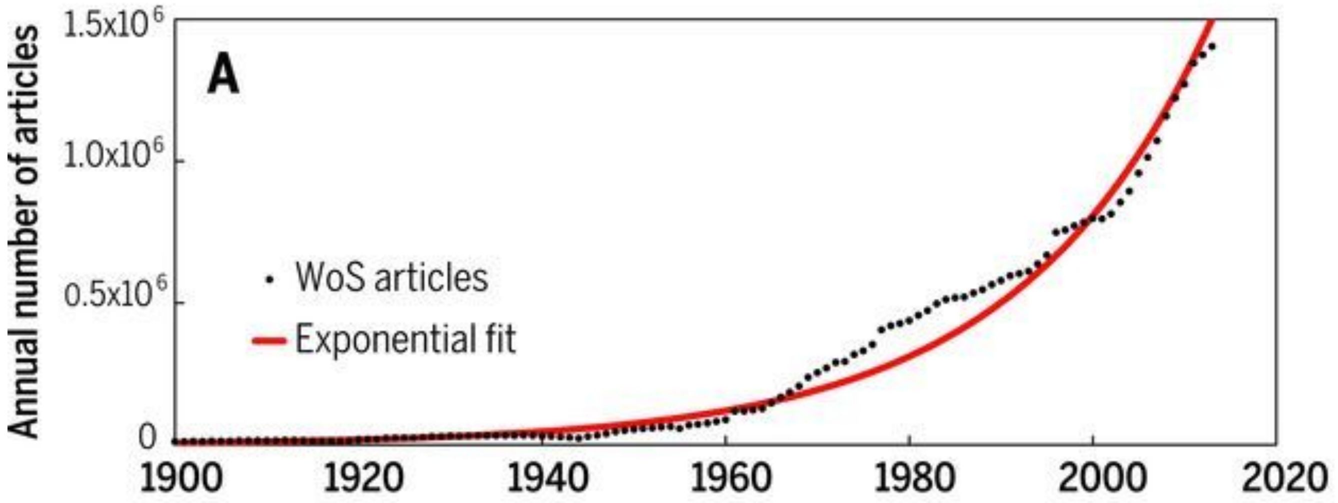
Lorcán Pigott-Dix

SWAT4HCLS 2023 - Basel

Earlham Institute

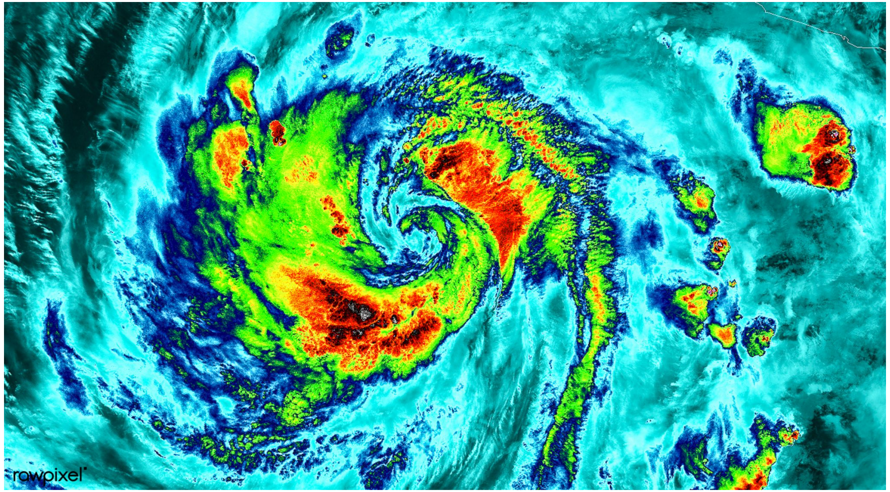
Decoding Living Systems

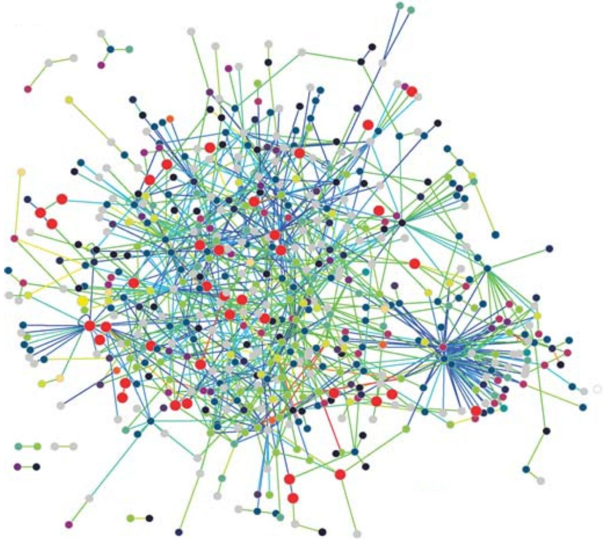
The scale of scientific output is growing exponentially...



Annual production of scientific articles indexed in the Web of Science database (from Fortunato *et al.* 2018).

… and it is increasingly complex





False-colour image of Hurricane Blanca taken using a Visible Infrared Imaging Radiometer Suite (from NASA/NOAA/UW-CIMSS/rawpixel 2015), and the protein interaction network of *T. pallidum* (from Haüser *et al.* 2008).

Ontologies describe domains of knowledge for machine agents

OBO Foundry

human phenotype ontology

Envo

CELL ONTOLOGY

GENEONTOLOGY
Unifying Biology

Current annotation is largely manual

ELIXIR-UK Launches new FAIR Data Stewardship Training Fellowship – apply now!

RDA
RESEARCH DATA ALLIANCE

GO FAIR

FAIRSFAIR
Fostering Fair Data Practices in Europe

We need to stop doing science

We need tools to scale data annotation

Aims of this work

- Develop a pipeline for training a multi-domain ontology annotation tool with minimal supervision.

- Improve performance of existing methodologies by utilising attention.

Leveraging the expertise codified in an ontology

```
[Term]
id: HP:0009226
```

An extract from the Open Biomedical Ontologies (OBO) format Human Phenotype Ontology (HPO).

Leveraging the expertise codified in an ontology

```
[Term]
id: HP:0009226
name: Short proximal phalanx of the 5th finger
```

An extract from the Open Biomedical Ontologies (OBO) format Human Phenotype Ontology (HPO).

Leveraging the expertise codified in an ontology

```
[Term]
id: HP:0009226
name: Short proximal phalanx of the 5th finger
def: "Hypoplastic/small proximal phalanx of the fifth finger." [HPO:skoehler]
synonym: "Hypoplastic/small proximal phalanx of the 5th finger" EXACT []
synonym: "Short innermost little finger bone" EXACT [ORCID:0000-0001-5208-3432]
synonym: "Short innermost pinkie finger bone" EXACT layperson [ORCID:0000-0001-5208-3432]
synonym: "Short innermost pinky finger bone" EXACT layperson [ORCID:0000-0001-5208-3432]
synonym: "Short proximal phalanx of the fifth finger" EXACT layperson []
```

An extract from the Open Biomedical Ontologies (OBO) format Human Phenotype Ontology (HPO).

Leveraging the expertise codified in an ontology

```
[Term]
id: HP:0009226
name: Short proximal phalanx of the 5th finger
def: "Hypoplastic/small proximal phalanx of the fifth finger." [HPO:skoehler]
synonym: "Hypoplastic/small proximal phalanx of the 5th finger" EXACT []
synonym: "Short innermost little finger bone" EXACT [ORCID:0000-0001-5208-3432]
synonym: "Short innermost pinkie finger bone" EXACT layperson [ORCID:0000-0001-5208-3432]
synonym: "Short innermost pinky finger bone" EXACT layperson [ORCID:0000-0001-5208-3432]
synonym: "Short proximal phalanx of the fifth finger" EXACT layperson []
xref: UMLS:C4021509
is_a: HP:0009192 ! Aplasia/Hypoplasia of the proximal phalanx of the 5th finger
is_a: HP:0009237 ! Short 5th finger
is_a: HP:0010241 ! Short proximal phalanx of finger
```

An extract from the Open Biomedical Ontologies (OBO) format Human Phenotype Ontology (HPO).

Leveraging the expertise codified in an ontology

```
[Term]
id: HP:0009226
name: Short proximal phalanx of the 5th finger
def: "Hypoplastic/small proximal phalanx of the fifth finger." [HPO:skoehler]
synonym: "Hypoplastic/small proximal phalanx of the 5th finger" EXACT []
synonym: "Short innermost little finger bone" EXACT [ORCID:0000-0001-5208-3432]
synonym: "Short innermost pinkie finger bone" EXACT layperson [ORCID:0000-0001-5208-3432]
synonym: "Short innermost pinky finger bone" EXACT layperson [ORCID:0000-0001-5208-3432]
synonym: "Short proximal phalanx of the fifth finger" EXACT layperson []
xref: UMLS:C4021509
is_a: HP:0009192 ! Aplasia/Hypoplasia of the proximal phalanx of the 5th finger
is_a: HP:0009237 ! Short 5th finger
is_a: HP:0010241 ! Short proximal phalanx of finger
created_by: doelkens
creation_date: 2009-01-05T06:01:34Z
```

Figure 3: An extract from the Open Biomedical Ontologies (OBO) format Human Phenotype Ontology (HPO).

Neural Concept Recogniser (NCR)

JMIR Medical Informatics | Journal Information | Browse Journal

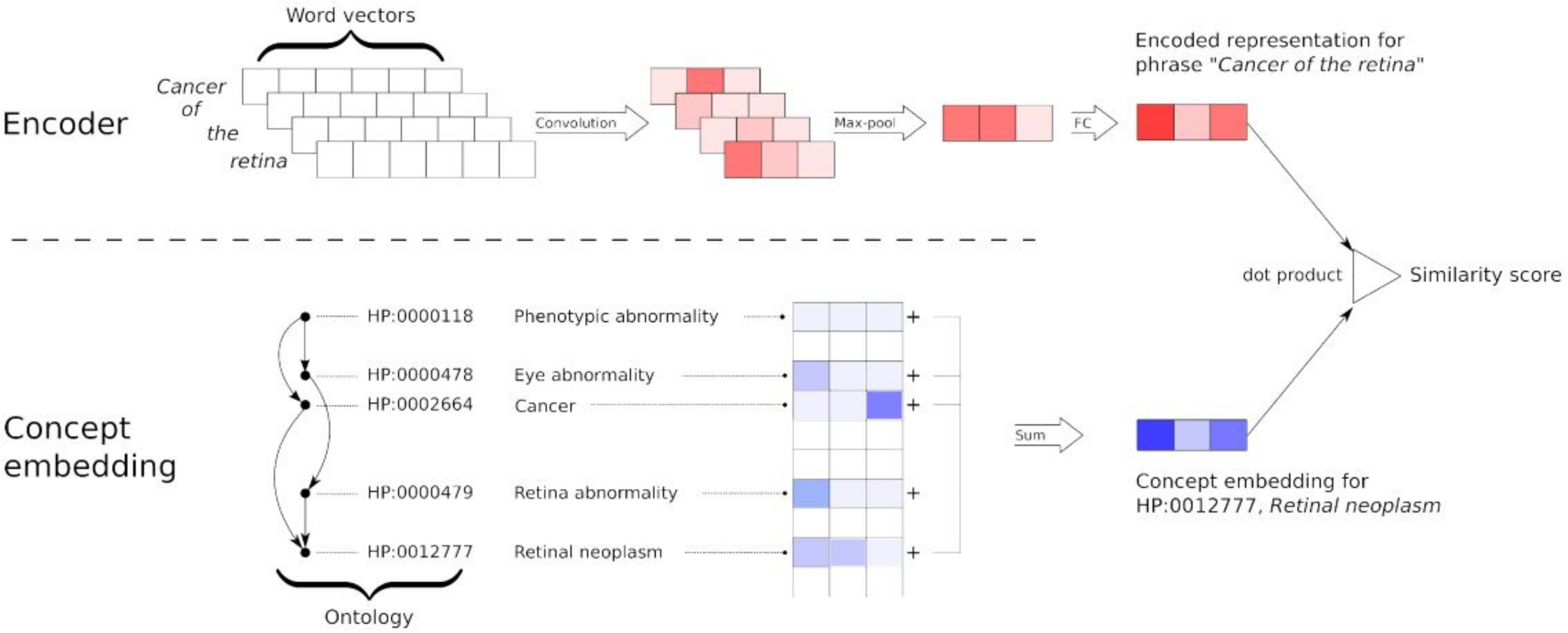
Published on 10.5.2019 in Vol 7 , No 2 (2019) :Apr-Jun

Preprints (earlier versions) of this paper are available at https://preprints.jmir.org/preprint/12596, first published October 24, 2018.

Identifying Clinical Terms in Medical Text Using Ontology-Guided Machine Learning

Aryan Arbabi [1,2]; David R Adams [3]; Sanja Fidler [1]; Michael Brudno [1,2]

Neural Concept Recogniser (NCR)



Overview of the NCR model (from Arbabi *et al.* 2019)

PhenoTagger

JOURNAL ARTICLE

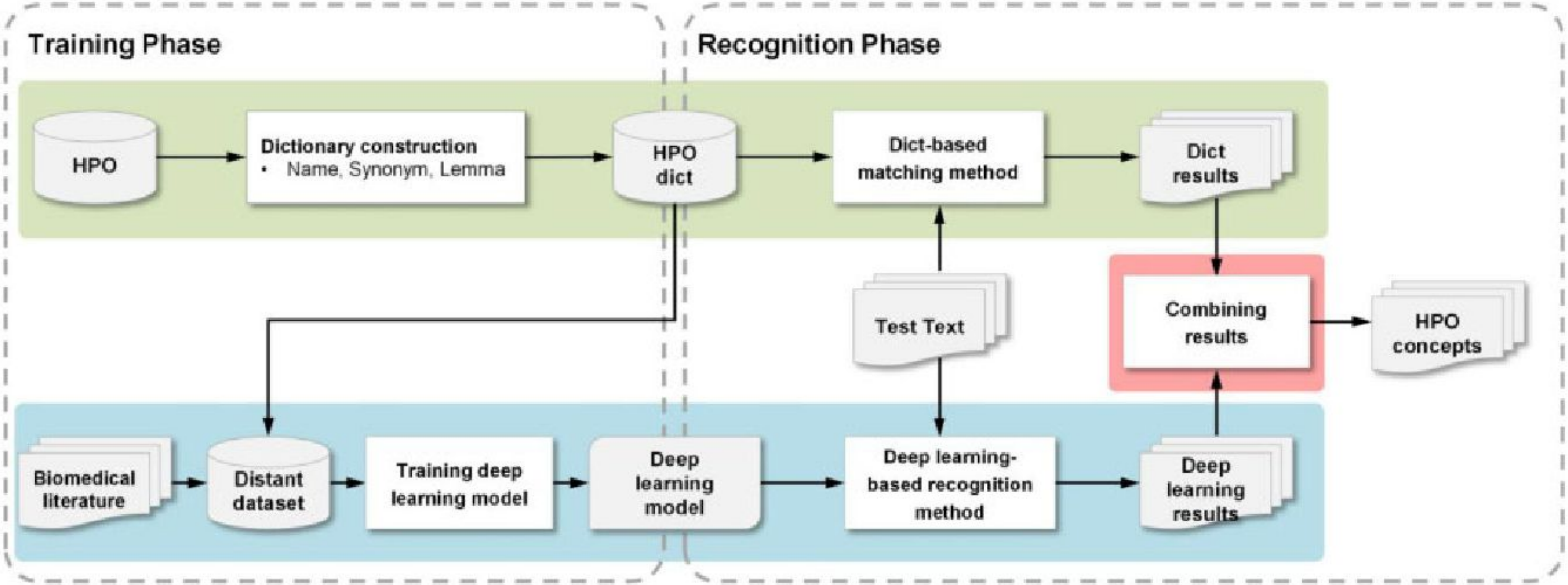
PhenoTagger: a hybrid method for phenotype concept recognition using human phenotype ontology

Ling Luo, Shankai Yan, Po-Ting Lai, Daniel Veltri, Andrew Oler, Sandhya Xirasagar, Rajarshi Ghosh, Morgan Similuk, Peter N Robinson, Zhiyong Lu

Bioinformatics, Volume 37, Issue 13, 1 July 2021, Pages 1884–1890, https://doi.org/10.1093/bioinformatics/btab019

Published: 20 January 2021 **Article history** ▾

PhenoTagger



Overview of the PhenoTagger model (from Luo *et al.* 2021)

Limitations

- Both of these methods are for a single ontology.
- PhenoTagger needs a corpus of relevant text.

Benchmarks

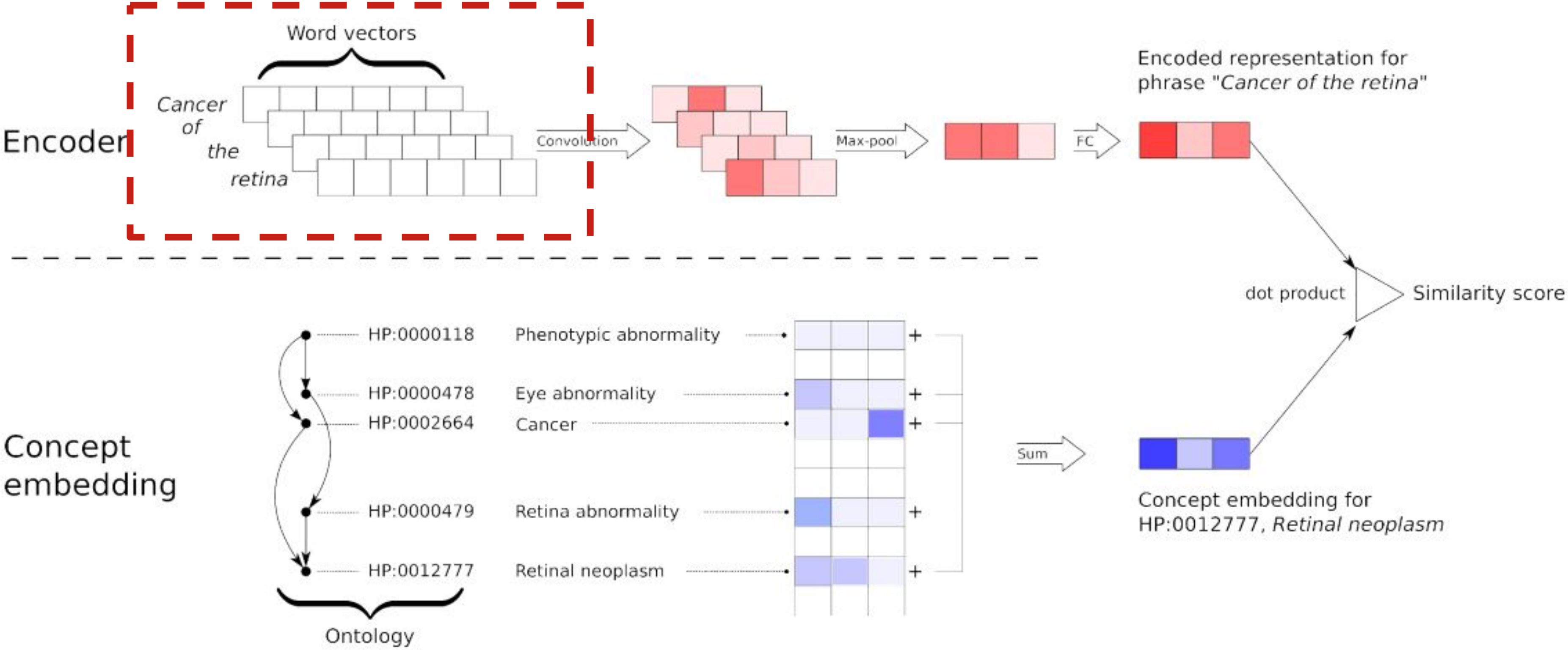
Table 1: The best scores for each model against a Gold-Standard dataset of 228 PubMed abstracts, expertly annotated with HPO terms, as reported in their respective publications, for non-overlapping concepts.

Method	Micro / %			Macro / %		
	Precision	Recall	F-score	Precision	Recall	F-score
NCR (Arbabi *et al.* 2019)	**80.3**	62.4	70.2	**80.5**	68.2	73.9
PhenoTagger (Luo *et al.* 2021) (With concept overlap)	78.9	**72.2**	**75.4**	77.4	**74.0**	**75.7**

Adapting the NCR model

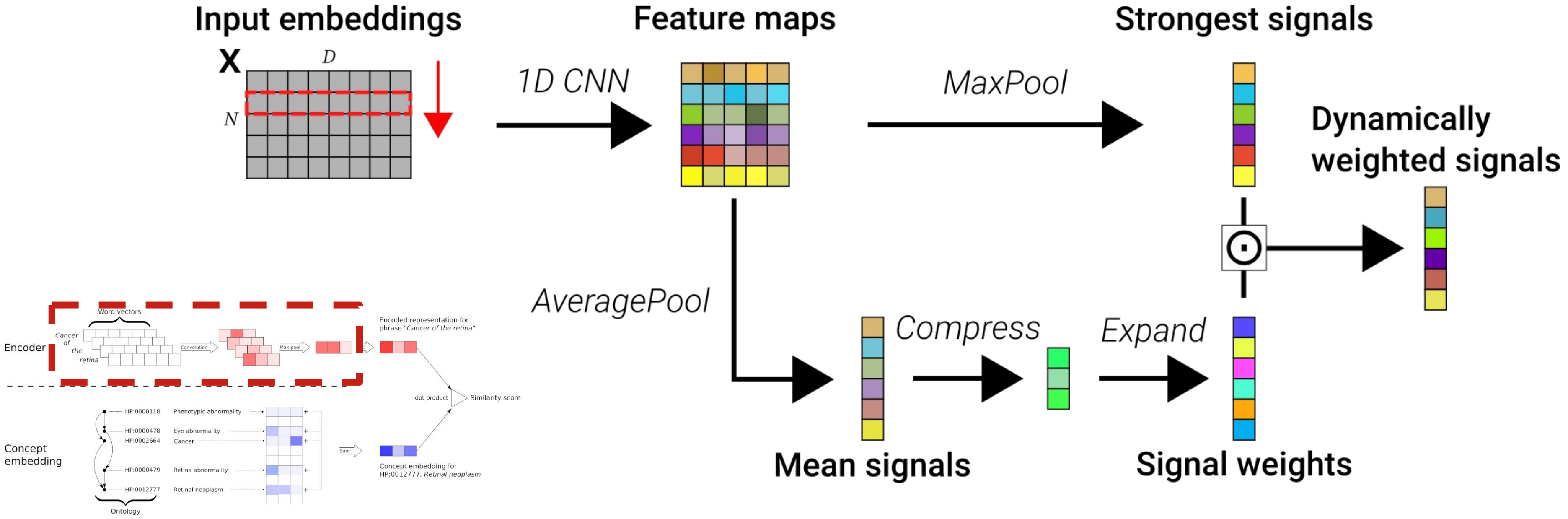
- Integrating ontologies from multiple domains
- Testing alternate architectures

NCR + ELMo Embeddings



Overview of the NCR model (from Arbabi *et al.* 2019)

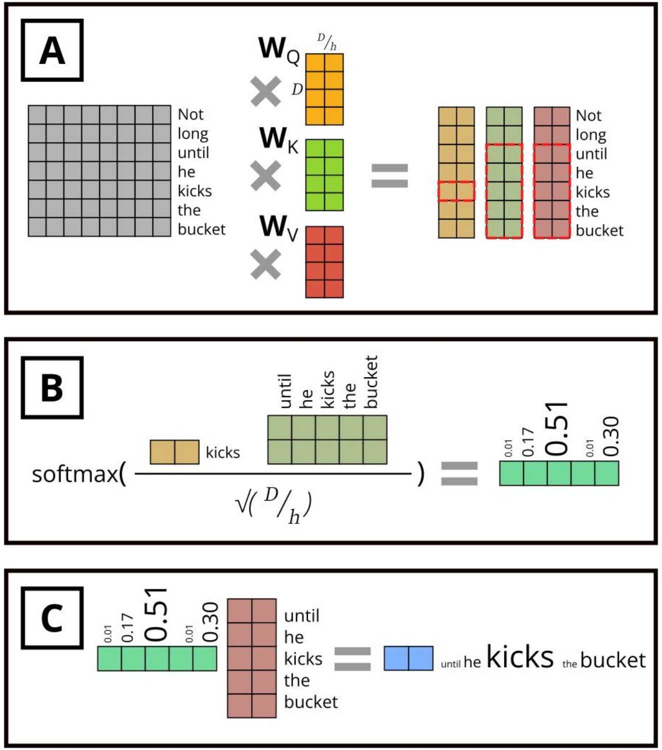
Squeeze-and-Excite (SAE) - using attention to improve CNN



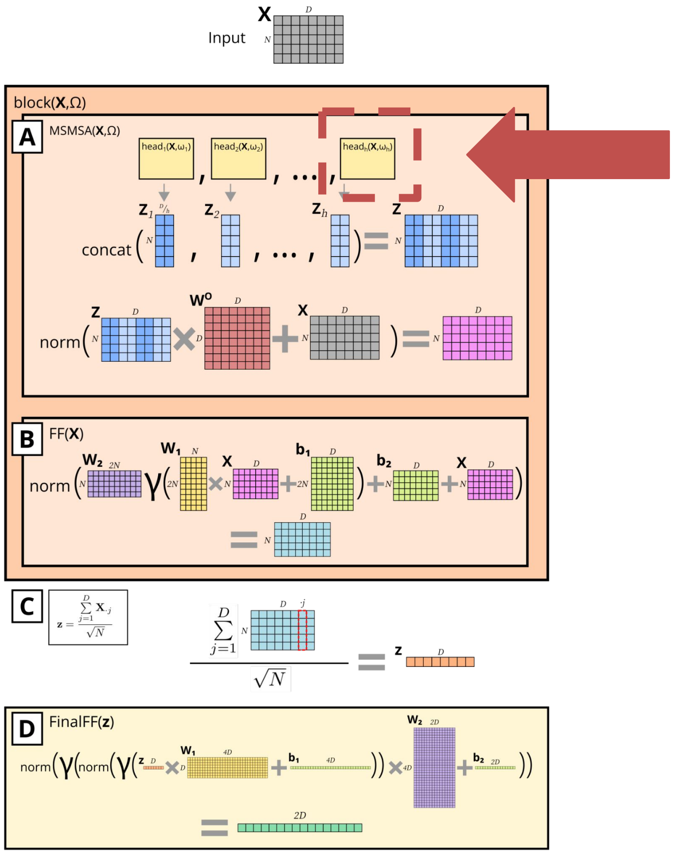
Schematic of the Squeeze-and-Excite mechanism used to augment the CNN.

The Multi-Scale Self-Attention (MSSA) model limits attention

- Transformer-style encoder
- Adapted to work on low data volumes by scaling attention to local neighbourhood
- **More relevance + sharper features → greater inductive bias**

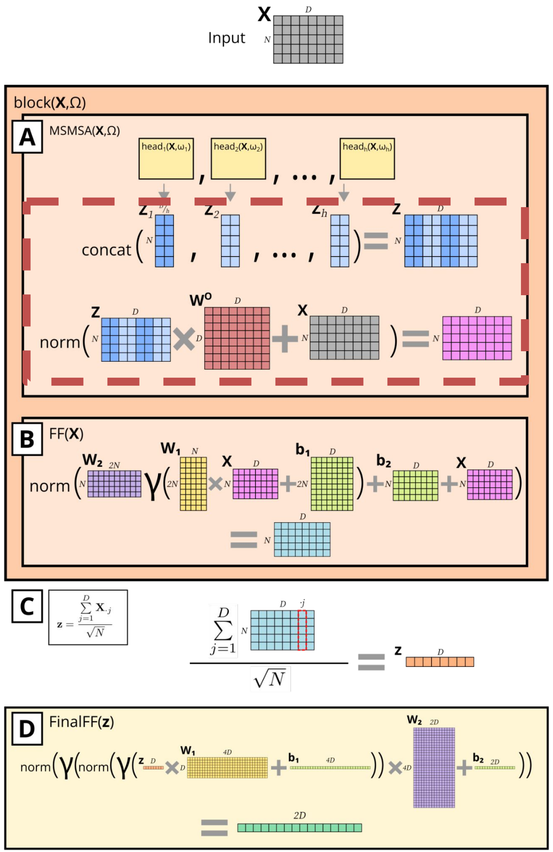


The output of the attention blocks is averaged into a single vector



Earlham Institute

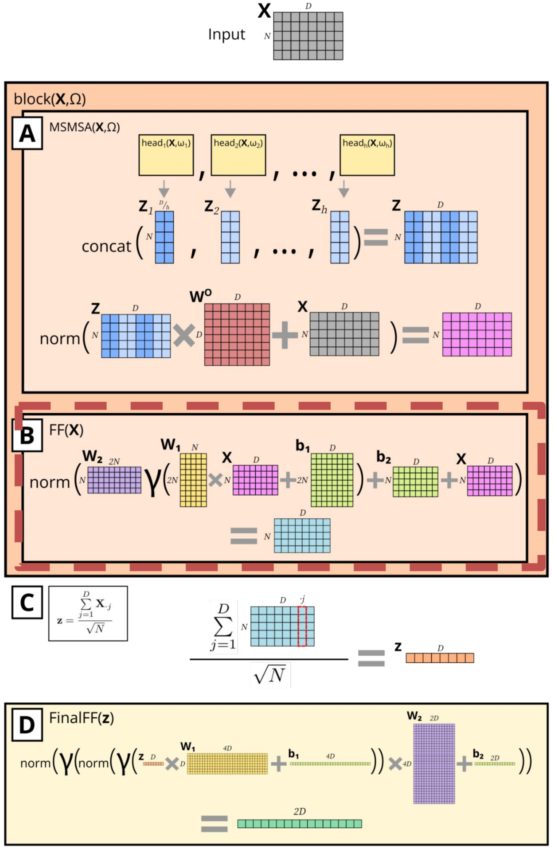
The output of the attention blocks is averaged into a single vector



www.earlham.ac.uk

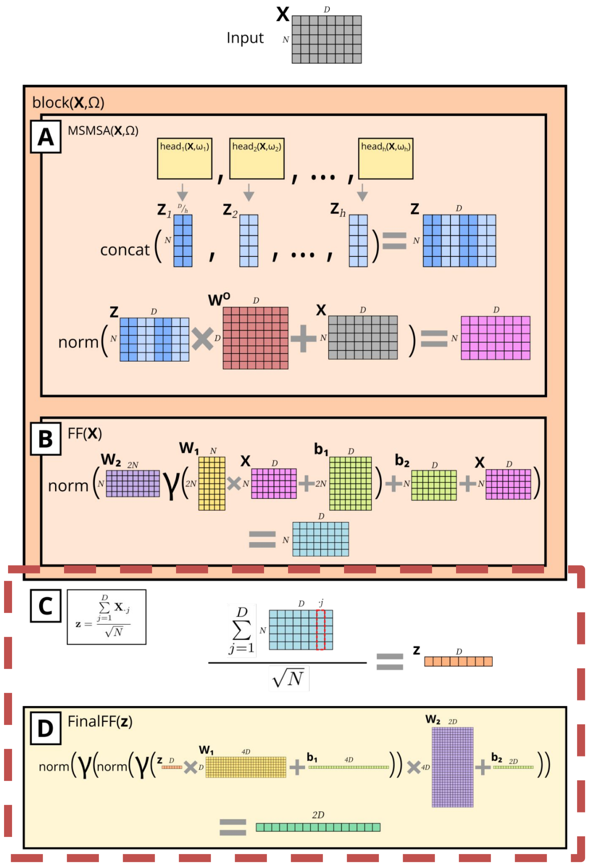
Earlham Institute

The output of the attention blocks is averaged into a single vector



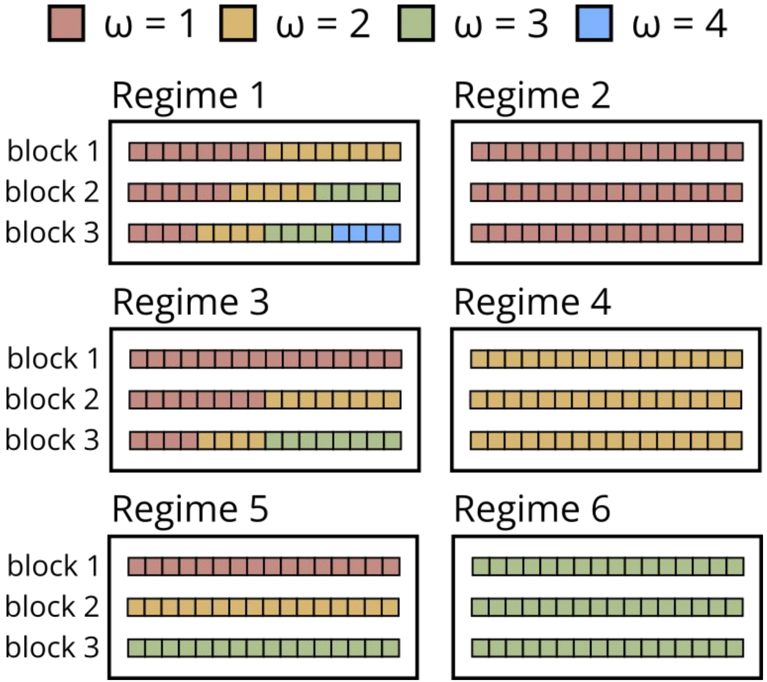
Earlham Institute

The output of the attention blocks is averaged into a single vector



Earlham Institute

Various scaling regimes were explored



Encompassing multiple domains

The different ontology combinations used to train each model.

Ontologies	Unique concepts	Training examples
Human Phenotype Ontology	16 059	35 969
Human Phenotype Ontology and Mammal Phenotype Ontology	29 370	75 298
Human Phenotype Ontology, Cell Ontology, and Ontology of Host-Pathogen Interactions	29 662	59 175

Assessing performance

- 228 PubMed Abstracts
- Expertly annotated with Human Phenotype Ontology terms

MSSA is competitive in micro metrics, and beats SOTA in macro

The PubMed benchmarking results of the best performing scaled-attention model for each ontology combination.

Ontology	Scale regime	Blocks	Threshold	Micro Precision	Recall	F-score	Macro Precision	Recall	F-score
HPO	4	2	0.45	75.04	71.63	73.30	78.16	74.08	76.06
+ MPO	3	2	0.4	75.00	65.00	69.64	76.89	68.05	72.20
+ CLO + OHPI	2	1	0.5	80.37	68.58	74.01	83.21	70.72	76.46

NCR (Arbabi *et al.* 2019)	80.3	62.4	70.2	80.5	68.2	73.9
PhenoTagger (Luo *et al.* 2021)	78.9	72.2	75.4	77.4	74.0	75.7

MSSA beats SOTA in macro f-score …

The PubMed benchmarking results of the best performing scaled-attention model for each ontology combination.

Ontology	Scale regime	Blocks	Threshold	Micro Precision	Recall	F-score	Macro Precision	Recall	F-score
HPO	4	2	0.45	75.04	71.63	73.30	78.16	74.08	76.06
+ MPO	3	2	0.4	75.00	65.00	69.64	76.89	68.05	72.20
+ CLO + OHPI	2	1	0.5	80.37	68.58	74.01	83.21	70.72	76.46

NCR (Arbabi *et al.* 2019)	80.3	62.4	70.2	80.5	68.2	73.9
PhenoTagger (Luo *et al.* 2021)	78.9	72.2	75.4	77.4	74.0	75.7

… and is competitive in micro metrics

The PubMed benchmarking results of the best performing scaled-attention model for each ontology combination.

Ontology	Scale regime	Blocks	Threshold	Micro Precision	Micro Recall	Micro F-score	Macro Precision	Macro Recall	Macro F-score
HPO	4	2	0.45	75.04	71.63	73.30	78.16	74.08	76.06
+ MPO	3	2	0.4	75.00	65.00	69.64	76.89	68.05	72.20
+ CLO + OHPI	2	1	0.5	80.37	68.58	74.01	83.21	70.72	76.46
NCR (Arbabi *et al.* 2019)				80.3	62.4	70.2	80.5	68.2	73.9
PhenoTagger (Luo *et al.* 2021)				78.9	72.2	75.4	77.4	74.0	75.7

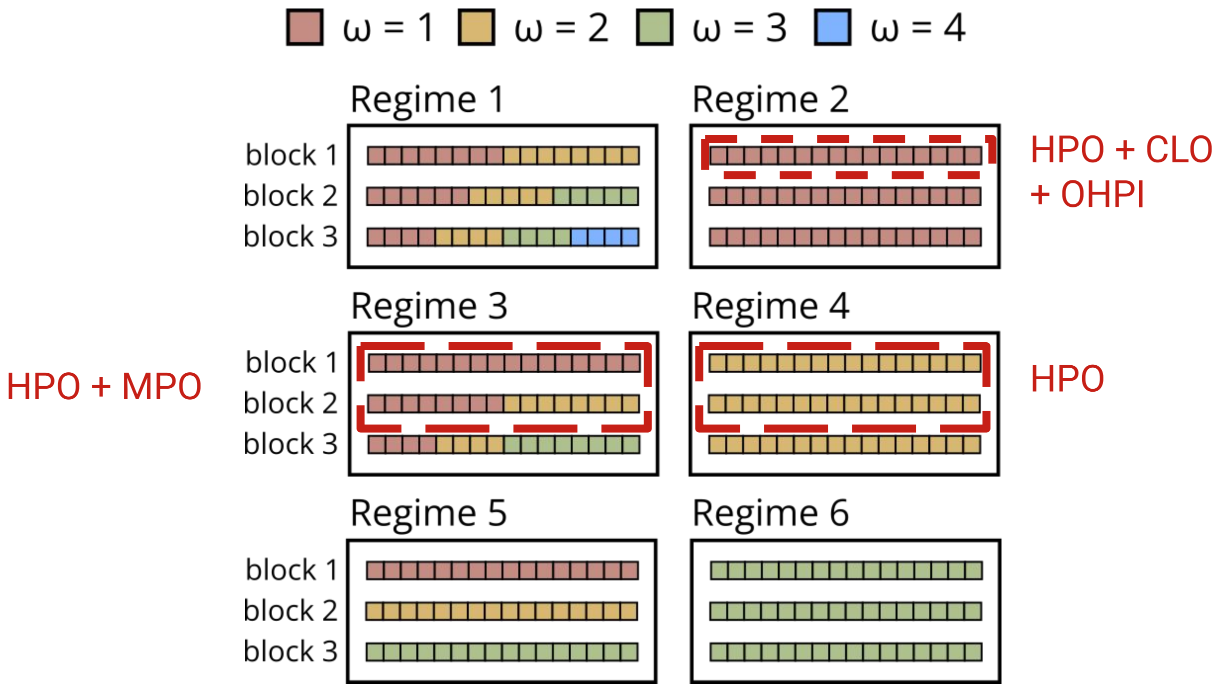
Those trained with HPO and MPO do not perform as well

The PubMed benchmarking results of the best performing scaled-attention model for each ontology combination.

Ontology	Scale regime	Blocks	Threshold	Micro Precision	Recall	F-score	Macro Precision	Recall	F-score
HPO	4	2	0.45	75.04	71.63	73.30	78.16	74.08	76.06
+ MPO	3	2	0.4	75.00	65.00	69.64	76.89	68.05	72.20
+ CLO + OHPI	2	1	0.5	80.37	68.58	74.01	83.21	70.72	76.46

NCR (Arbabi *et al.* 2019)	80.3	62.4	70.2	80.5	68.2	73.9
PhenoTagger (Luo *et al.* 2021)	78.9	72.2	75.4	77.4	74.0	75.7

The narrower the attention the better



Earlham Institute

Sorry for the horrible table

The results of the CNN-based models.

Ontology	Model	Filters	Threshold	Micro Precision	Micro Recall	Micro F-score	Macro Precision	Macro Recall	Macro F-score
HPO	NCR	1024	0.5	78.72	**72.75**	75.62	82.16	74.78	78.29
	SAE	1024	0.55	77.63	72.60	75.03	80.46	**75.04**	77.65
		1536	0.55	80.81	70.89	75.53	83.07	73.49	77.99
		2048	0.7	80.25	70.81	75.24	82.81	73.65	77.96
+ MPO	NCR	1024	0.5	74.12	66.12	69.89	77.38	70.75	73.91
	SAE	1024	0.85	82.28	52.20	63.87	84.28	57.09	68.07
		1536	0.5	76.17	65.23	70.28	78.68	68.90	73.46
		2048	0.5	74.62	61.73	67.56	76.78	66.19	71.10
+ CLO + OHPI	NCR	1024	0.75	**84.04**	64.71	73.12	86.01	66.61	75.08
	SAE	1024	0.6	83.05	70.07	**76.01**	85.13	72.82	**78.50**
		1536	0.5	80.55	69.69	74.73	82.18	72.07	76.79
		2048	0.65	82.94	65.15	72.98	**86.12**	66.96	75.34
NCR (Arbabi *et al.* 2019)				80.3	62.4	70.2	80.5	68.2	73.9
PhenoTagger (Luo *et al.* 2021)				78.9	72.2	75.4	77.4	74.0	75.7

SAE + diverse domain ontologies = Best Performance

The results of the CNN-based models.

Ontology	Model	Filters	Threshold	Micro Precision	Micro Recall	Micro F-score	Macro Precision	Macro Recall	Macro F-score
HPO	NCR	1024	0.5	78.72	**72.75**	75.62	82.16	74.78	78.29
	SAE	1024	0.55	77.63	72.60	75.03	80.46	**75.04**	77.65
		1536	0.55	80.81	70.89	75.53	83.07	73.49	77.99
		2048	0.7	80.25	70.81	75.24	82.81	73.65	77.96
+ MPO	NCR	1024	0.5	74.12	66.12	69.89	77.38	70.75	73.91
	SAE	1024	0.85	82.28	52.20	63.87	84.28	57.09	68.07
		1536	0.5	76.17	65.23	70.28	78.68	68.90	73.46
		2048	0.5	74.62	61.73	67.56	76.78	66.19	71.10
+ CLO + OHPI	NCR	1024	0.75	**84.04**	64.71	73.12	86.01	66.61	75.08
	SAE	1024	0.6	83.05	70.07	**76.01**	85.13	72.82	**78.50**
		1536	0.5	80.55	69.69	74.73	82.18	72.07	76.79
		2048	0.65	82.94	65.15	72.98	**86.12**	66.96	75.34
NCR (Arbabi *et al.* 2019)				80.3	62.4	70.2	80.5	68.2	73.9
PhenoTagger (Luo *et al.* 2021)				78.9	72.2	75.4	77.4	74.0	75.7

Better embeddings alone improve the NCR ...

The results of the CNN-based models.

Ontology	Model	Filters	Threshold	Micro Precision	Micro Recall	Micro F-score	Macro Precision	Macro Recall	Macro F-score
HPO	NCR	1024	0.5	78.72	**72.75**	75.62	82.16	74.78	78.29
	SAE	1024	0.55	77.63	72.60	75.03	80.46	**75.04**	77.65
		1536	0.55	80.81	70.89	75.53	83.07	73.49	77.99
		2048	0.7	80.25	70.81	75.24	82.81	73.65	77.96
+ MPO	NCR	1024	0.5	74.12	66.12	69.89	77.38	70.75	73.91
	SAE	1024	0.85	82.28	52.20	63.87	84.28	57.09	68.07
		1536	0.5	76.17	65.23	70.28	78.68	68.90	73.46
		2048	0.5	74.62	61.73	67.56	76.78	66.19	71.10
+ CLO + OHPI	NCR	1024	0.75	**84.04**	64.71	73.12	86.01	66.61	75.08
	SAE	1024	0.6	83.05	70.07	**76.01**	85.13	72.82	**78.50**
		1536	0.5	80.55	69.69	74.73	82.18	72.07	76.79
		2048	0.65	82.94	65.15	72.98	**86.12**	66.96	75.34
NCR (Arbabi *et al.* 2019)				80.3	62.4	70.2	80.5	68.2	73.9
PhenoTagger (Luo *et al.* 2021)				78.9	72.2	75.4	77.4	74.0	75.7

… but NCR performance declines with additional ontologies

The results of the CNN-based models.

Ontology	Model	Filters	Threshold	Micro Precision	Micro Recall	Micro F-score	Macro Precision	Macro Recall	Macro F-score
HPO	NCR	1024	0.5	78.72	**72.75**	75.62	82.16	74.78	78.29
	SAE	1024	0.55	77.63	72.60	75.03	80.46	**75.04**	77.65
		1536	0.55	80.81	70.89	75.53	83.07	73.49	77.99
		2048	0.7	80.25	70.81	75.24	82.81	73.65	77.96
+ MPO	NCR	1024	0.5	74.12	66.12	69.89	77.38	70.75	73.91
	SAE	1024	0.85	82.28	52.20	63.87	84.28	57.09	68.07
		1536	0.5	76.17	65.23	70.28	78.68	68.90	73.46
		2048	0.5	74.62	61.73	67.56	76.78	66.19	71.10
+ CLO + OHPI	NCR	1024	0.75	**84.04**	64.71	73.12	86.01	66.61	75.08
	SAE	1024	0.6	83.05	70.07	**76.01**	85.13	72.82	**78.50**
		1536	0.5	80.55	69.69	74.73	82.18	72.07	76.79
		2048	0.65	82.94	65.15	72.98	**86.12**	66.96	75.34

NCR (Arbabi *et al.* 2019)	80.3	62.4	70.2	80.5	68.2	73.9
PhenoTagger (Luo *et al.* 2021)	78.9	72.2	75.4	77.4	74.0	75.7

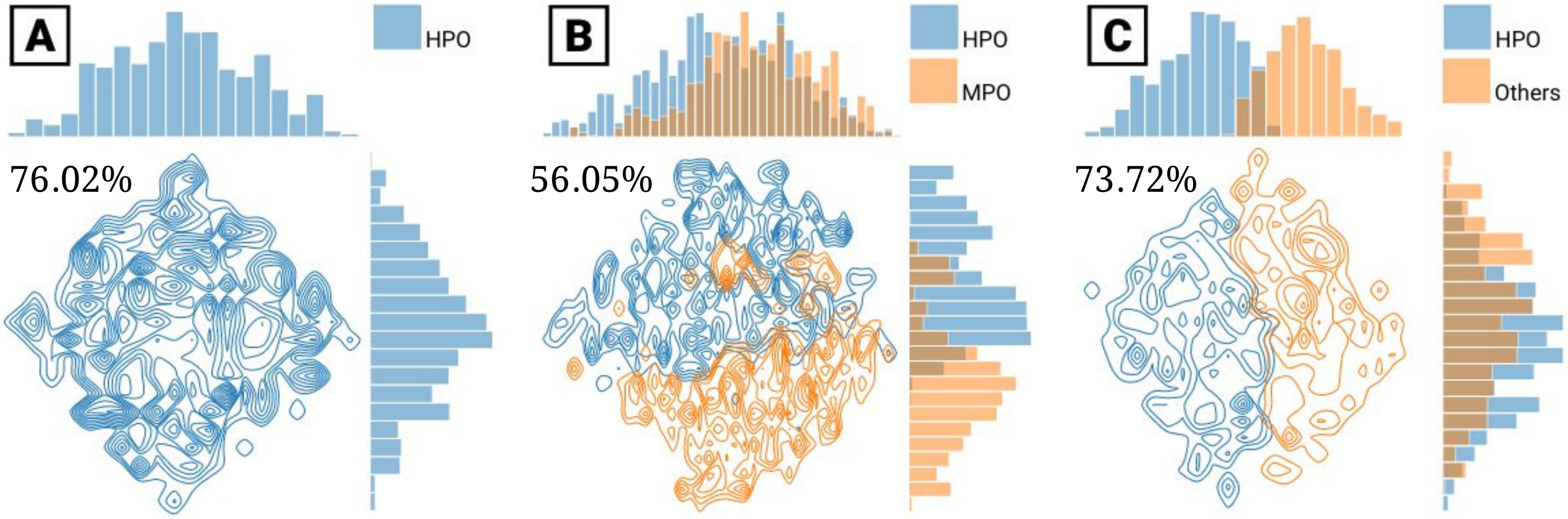
Again … combination of similar domain ontologies doesn't perform as well

The results of the CNN-based models.

Ontology	Model	Filters	Threshold	Micro Precision	Micro Recall	Micro F-score	Macro Precision	Macro Recall	Macro F-score
HPO	NCR	1024	0.5	78.72	**72.75**	75.62	82.16	74.78	78.29
	SAE	1024	0.55	77.63	72.60	75.03	80.46	**75.04**	77.65
		1536	0.55	80.81	70.89	75.53	83.07	73.49	77.99
		2048	0.7	80.25	70.81	75.24	82.81	73.65	77.96
+ MPO	NCR	1024	0.5	74.12	66.12	69.89	77.38	70.75	73.91
	SAE	1024	0.85	82.28	52.20	63.87	84.28	57.09	68.07
		1536	0.5	76.17	65.23	70.28	78.68	68.90	73.46
		2048	0.5	74.62	61.73	67.56	76.78	66.19	71.10
+ CLO + OHPI	NCR	1024	0.75	**84.04**	64.71	73.12	86.01	66.61	75.08
	SAE	1024	0.6	83.05	70.07	**76.01**	85.13	72.82	**78.50**
		1536	0.5	80.55	69.69	74.73	82.18	72.07	76.79
		2048	0.65	82.94	65.15	72.98	**86.12**	66.96	75.34

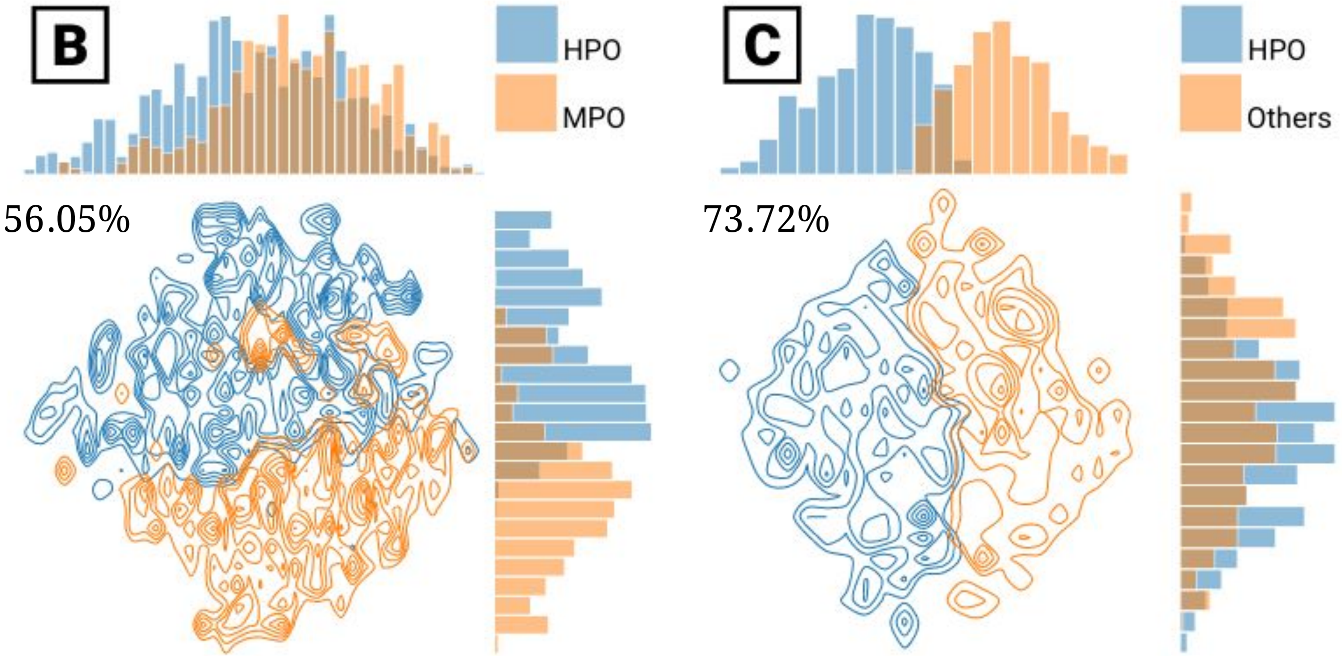
NCR (Arbabi *et al.* 2019)	80.3	62.4	70.2	80.5	68.2	73.9
PhenoTagger (Luo *et al.* 2021)	78.9	72.2	75.4	77.4	74.0	75.7

Understanding why similar domains reduce performance



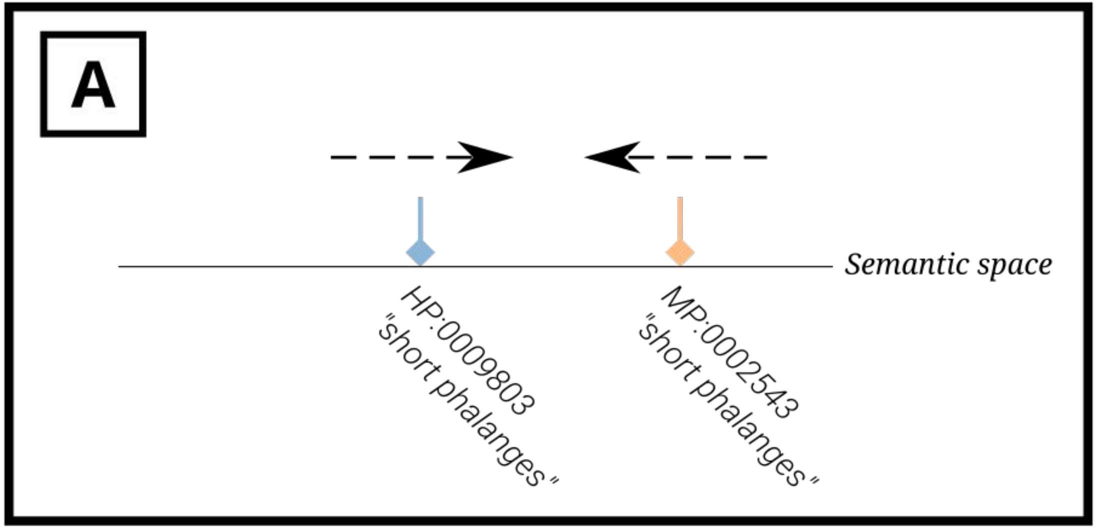
Density-contour t-SNE plots of the concept embeddings from the best-performing SAE models. [**A**] Using the HPO; [**B**] HPO and MPO; and [**C**] HPO, CLO and OHPI.

Human Phenotype and Mammal Phenotype embeddings are congested

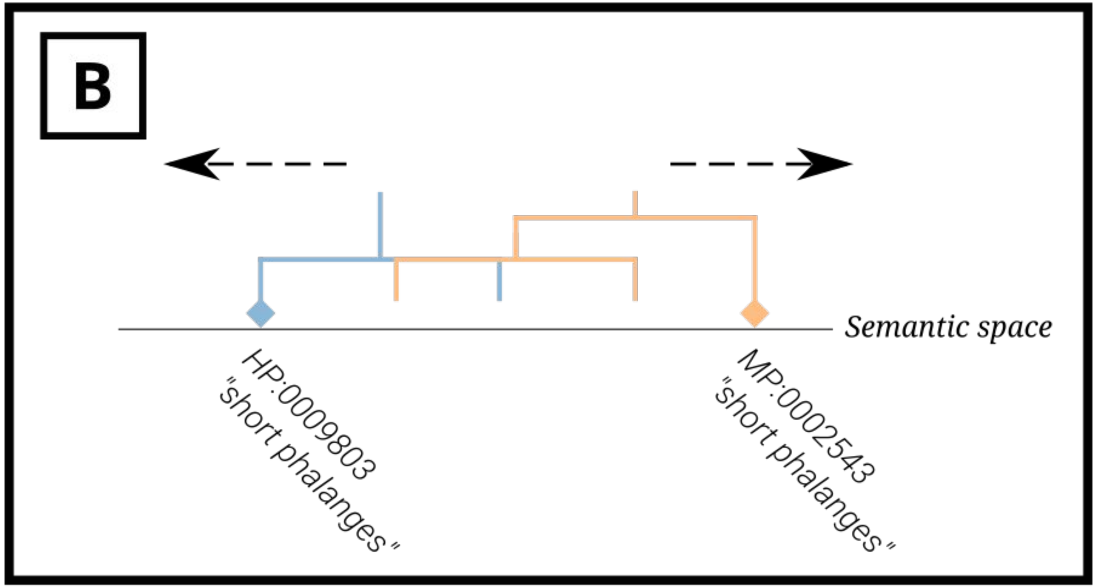


Earlham Institute

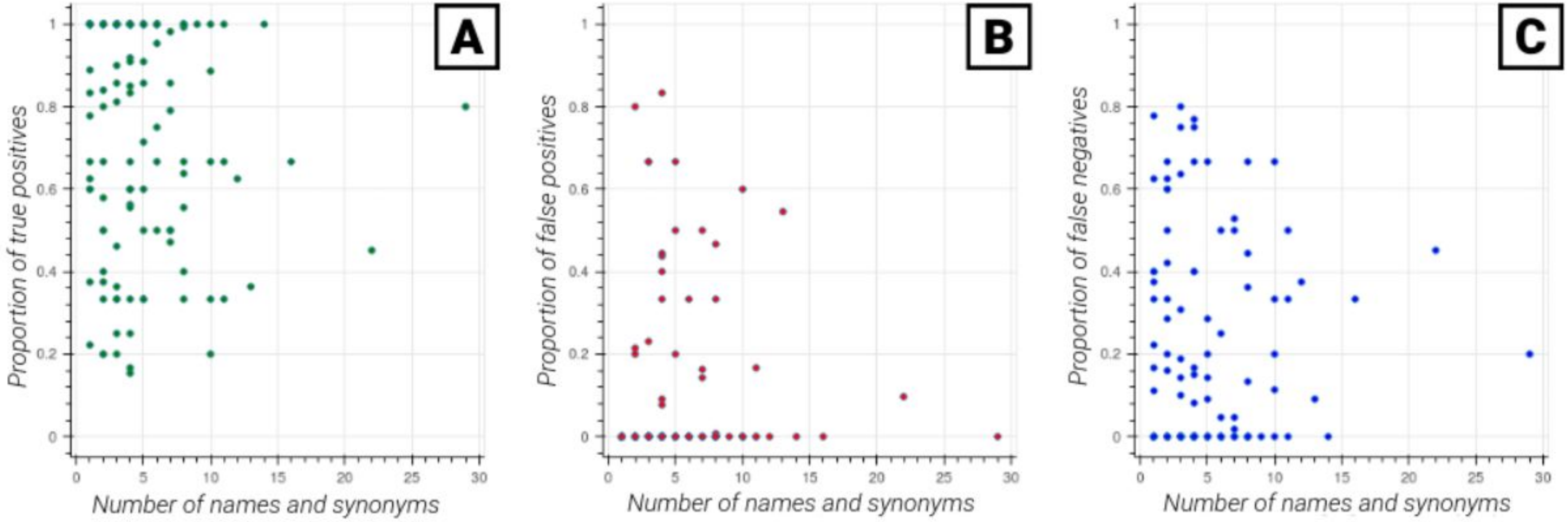
The encoder tries to produce similar encodings for similar concepts from different ontologies…



… the Ancestry matrix keeps otherwise semantically similar concepts apart

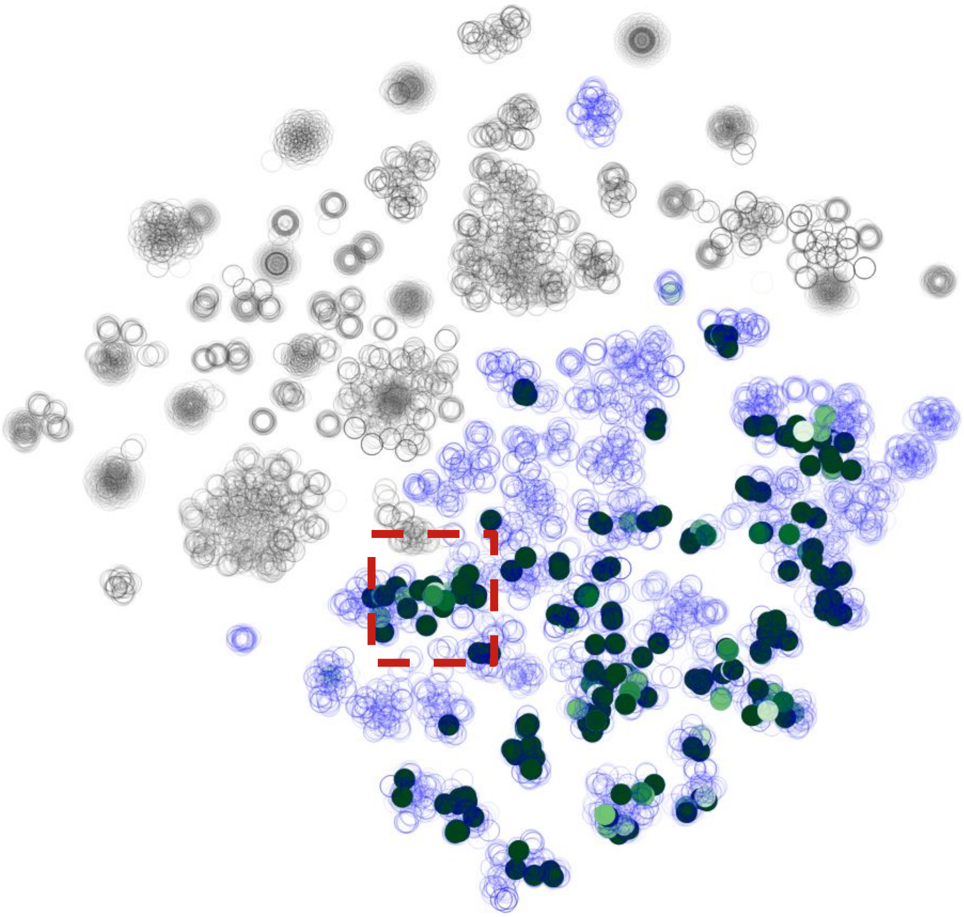


No relationship between the number of descriptions a concept has and the error rate



The proportion of [**A**] true positive; [**B**] false positive; and [**C**] false negative identifications made by the best performing model, for each ontology concept that is mentioned more than five times in the Gold-Standard benchmarking corpora.

Exploring semantic similarity and error rate



Semantic proximity appears to influence error rate

In summary

- New SOTA for Neural Dictionary methods
- More ontology data + attention → improved performance
- Domain overlap appears to hinder model performance
- Concept overlap heuristics are important
- Attention-based encoders can be adapted to low data volumes
- Available at: https://github.com/lorcanpd/adorNER

Acknowledgements

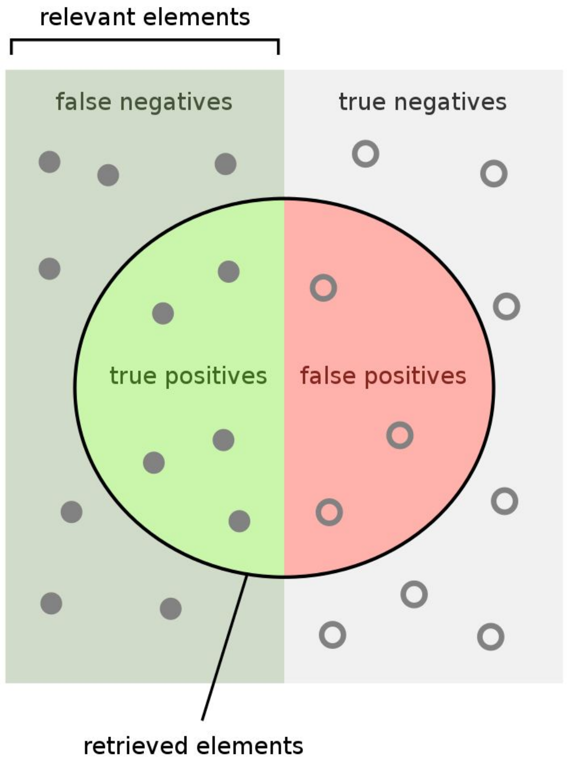
Rob Davey
Felix Shaw
Simon Tyrrell
Nicola Soranzo
Martin Ayling
Evanthia Samota
Krister-Jazz Urog
Aaliyah Providence
Daniel Olvera Cabrera
Wilfried Haerty

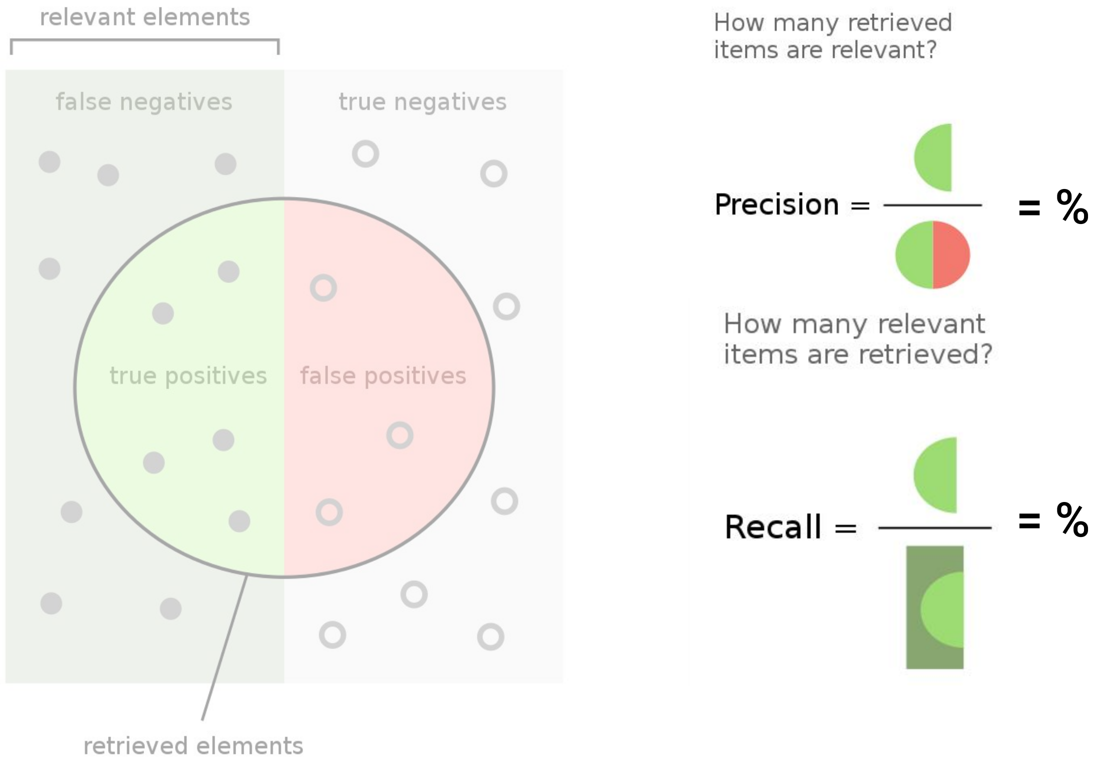
Biotechnology and Biological Sciences Research Council

This work was funded by the BBSRC as part of the Norwich Research Park Biosciences Doctoral Training Partnership, grant number BB/M011216/1, reference code 2243628

References

- S. Fortunato, *et al.* (2018) Science of science, Science 359, eaao0185.
- NASA/NOAA/UW-CIMSS / Free Public Domain Illustrations by rawpixel, “NASA-NOAA's Suomi NPP Satellite Gets Colorful Look at Hurricane Blanca. Original from NASA. Digitally enhanced by rawpixel”, (https://commons.wikimedia.org/wiki/File:NASA-NOAA's_Suomi_NPP_Satellite_Gets_Colorful_Look_at_Hurricane_Blanca._Original_from_NASA._Digitally_enhanced_by_rawpixel.jpg), https://creativecommons.org/licenses/by/2.0/legalcode
- R. Haüser, *et al*. (2008) The Binary Protein Interactome of *Treponema pallidum* – The Syphilis Spirochete. PLoS ONE 3(5): e2292.
- A. Arbabi, *et al.* (2019) ‘Identifying Clinical Terms in Medical Text Using Ontology-Guided Machine Learning’. In: JMIR Medical Informatics 7.2, e12596.
- Ling Luo, *et al.* (2021) ‘PhenoTagger: A Hybrid Method for Phenotype Concept Recognition using Human Phenotype Ontology’. In: Bioinformatics 37.13, pp. 1884–1890.


relevant elements
false negatives
true negatives
true positives
false positives
retrieved elements


relevant elements
false negatives
true negatives
true positives
false positives
retrieved elements
How many retrieved items are relevant?
Precision =
= %
How many relevant items are retrieved?
Recall =
= %