A Excluded semantic types

Abbreviation	TUI	Full semantic type name
bird	T012	Bird
dora	T056	Daily or Recreational Activity
edac	T065	Educational Activity
fish	T013	Fish
food	T168	Food
geoa	T083	Geographic Area
gora	T064	Governmental or Regulatory Activity
idcn	T078	Idea or Concept
inpr	T170	Intellectual Product
lang	T171	Language
mnob	T073	Manufactured Object
ocac	T057	Occupational Activity
ocdi	T090	Occupation or Discipline
mcha	T066	Machine Activity
orgt	T092	Organization
phob	T072	Physical Object
phpr	T067	Phenomenon or Process
prog	T097	Professional or Occupational Group
pros	T094	Professional Society
qlco	T080	Qualitative Concept
qnco	T081	Quantitative Concept
rnlw	T089	Regulation or Law
shro	T095	Self-help or Relief Organization
spco	T082	Spatial Concept
tmco	T079	Temporal Concept
vtbt	T010	Vertebrate

Table 1: List of the twenty-six semantic types that were considered non-relevant for biomedical entity linking by the authors. For a full list of all semantic types in the UMLS see: `https://lhncbc.nlm.nih.gov/ii/tools/MetaMap/documentation/SemanticTypesAndGroups.html`

B SPARQL query

```
SELECT ?concept ?conceptLabel ?cui ?article  WHERE {
     ?concept wdt:P2892 ?cui .
    ?article schema:about ?concept .
    ?article schema:isPartOf <https://nl.wikipedia.org/>.

    SERVICE wikibase:label {
       bd:serviceParam wikibase:language "nl"
    }
}
```

Listing 1: SPARQL query for retrieving all Wikidata entities that contain a UMLS CIU and where there exists an article about the entity that is part of the Dutch Wikipedia

C Hyperparameters

Hyperparameter	**Search space**	**Value**
Learning rate (2nd-phase and fine-tuning)		1×10^{-4}
Batch size		512
Weight decay		0.01
Max sequence length		25
Miner margin (λ)		0.2
Random seed		1993
Loss function		MS loss
α in MS loss		2
β in MS loss		50
ϵ in MS loss		0.5
Representation of input string		`[CLS]-token`

Table 2: List of hyperparameters used in the 2nd phase pretraining and fine-tuning steps.

D Hardware details

Hardware	Details
RAM	32GB
GPU	V100 Nvidia GPU
CPU	2vCPU @ 2.2GHz

Table 3: Hardware details that were used for training and evaluating our models on Google Colab Pro.

E Training protocol

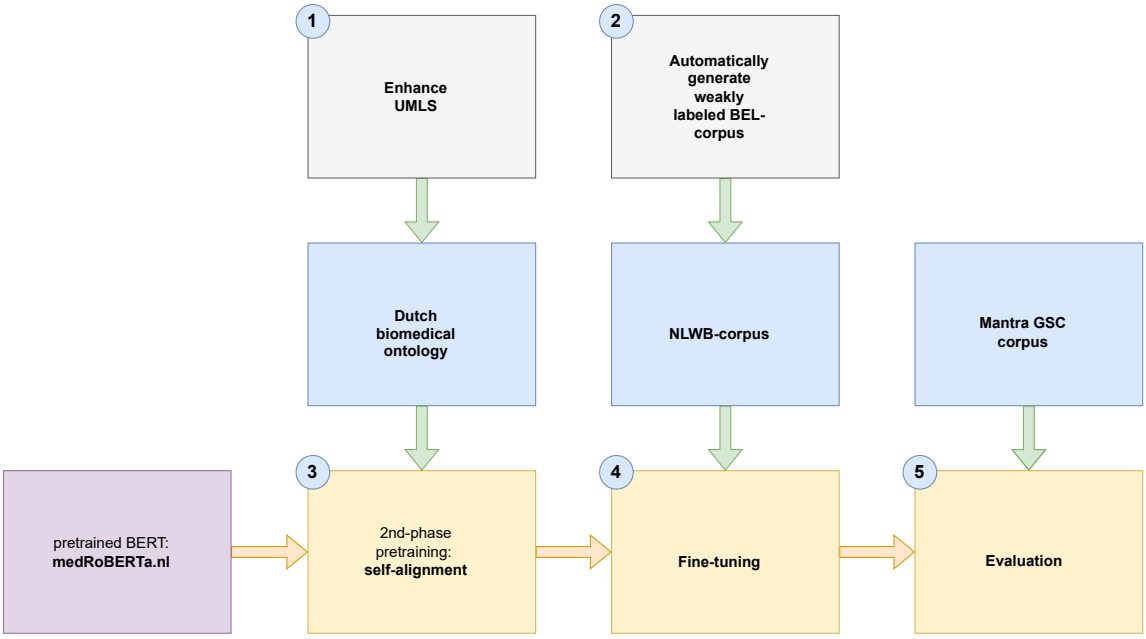

Figure 1: Overview of the training protocol.