

Semantic Relation Discovery by Using Co-occurrence Information

Stefan Schulz, Catalina Martínez Costa, Markus Kreuzthaler, Jose A. Miñarro-Giménez, Ulrich Andersen, Anders B. Jensen, Bente Maegaard

Background: MEDLINE contains high quality semantic metadata covering more than 22 million bibliographic records, by manually assigned MeSH descriptors. Can this resource be used as a “non-ontological knowledge” layer on top of the clinical ontology SNOMED CT?

Source concept	<i>Name</i>	Bipolar disorder
	<i>Type</i>	Disorder
Target concept	<i>Name</i>	Tricyclic antidepressant
	<i>Type</i>	Substance
		DT=9,CI=7,DI=5,PX=4,CO=2, EP=2,GE=2,BL=1,ET=1,PA=1, PC=1,PP=1,TH=1
	<i>Absolute co-occurrence</i>	17
	<i>Log-likelihood</i>	54.57

MeSH subheadings

qualify the source concept, e.g. DT = drug therapy, PC = prescription and control, CO = complication

Semantic Relation Discovery by Using Co-occurrence Information

Stefan Schulz, Catalina Martínez Costa, Markus Kreuzthaler, Jose A. Miñarro-Giménez, Ulrich Andersen, Anders B. Jensen, Bente Maegaard

Hypothesis: the following combination of information permits the generation of factoid Subject – Predicate – Object statements:

- MeSH co-occurrence in MEDLINE (source UMLS)
- MeSH subheading profiles (source UMLS)
- MeSH – UMLS – SNOMED CT mappings
- SNOMED CT semantic types

		Object			
		Disease	Finding	Substance	Organism
Subject	Finding	<i>sign of symptom of</i>	<i>accompanied by</i>	<i>treated by</i>	<i>affects caused by</i>
	Substance	<i>causes treats prevents metabolite of</i>	<i>causes treats prevents</i>	<i>Interacts with</i>	<i>affects produced by</i>
	Organism	<i>causes affected by</i>	<i>causes</i>	<i>sensitive to</i>	<i>interacts with</i>
	Body part	<i>possible location of</i>	<i>possible location of</i>	<i>targeted by</i>	<i>targeted by</i>

Example: A high score of the “TU” qualifier on **Substance** allows to induce the predicate “*treats*” with **Disorder** as object; a high score of the “PC” qualifier suggests “*prevents*”, accordingly

Semantic Relation Discovery by Using Co-occurrence Information

Stefan Schulz, Catalina Martínez Costa, Markus Kreuzthaler, Jose A. Miñarro-Giménez, Ulrich Andersen, Anders B. Jensen, Bente Maegaard

Results: Preliminary testing for “treats” and “prevents”. Results are promising, however requiring further refinement.

+TO=0.0	+TU=0.34	-BL=0.08	-CF=0.0	-CI=0.72	-CL=0.0	-CO=0.03
+TO=0.0	+TU=0.84	-BL=0.08	-CF=0.0	-CI=0.72	-CL=0.0	-CO=0.02
+TO=0.01	+TU=0.63	-BL=0.09	-CF=0.0	-CI=0.05	-CL=0.0	-CO=0.01
+TO=0.0	+TU=0.31	-BL=0.08	-CF=0.0	-CI=0.78	-CL=0.0	-CO=0.05
+TO=0.0	+TU=0.11	-BL=0.12	-CF=0.0	-CI=0.73	-CL=0.0	-CO=0.02
+TO=0.0	+TU=0.75	-BL=0.09	-CF=0.0	-CI=0.06	-CL=0.0	-CO=0.09
+TO=0.0	+TU=0.61	-BL=0.16	-CF=0.0	-CI=0.05	-CL=0.0	-CO=0.0
+TO=0.02	+TU=0.44	-BL=0.06	-CF=0.0	-CI=0.7	-CL=0.02	-CO=0.02
+TO=0.0	+TU=0.42	-BL=0.11	-CF=0.0	-CI=0.02	-CL=0.0	-CO=0.0
+TO=0.0	+TU=0.32	-BL=0.04	-CF=0.0	-CI=0.73	-CL=0.03	-CO=0.01
+TO=0.0	+TU=0.75	-BL=0.07	-CF=0.0	-CI=0.04	-CL=0.0	-CO=0.04
+TO=0.0	+TU=0.37	-BL=0.04	-CF=0.0	-CI=0.82	-CL=0.0	-CO=0.04

S: Coagulation factor VIIa (substance) (Therapeutic Use)
 O: Bleeding (finding)

Outlook: Publication as linked data.

Possible use cases: question answering, query expansion, decision support, knowledge discovery, background knowledge for different NLP applications