

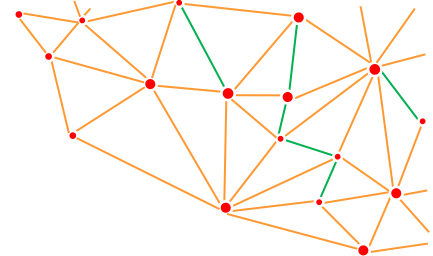
Current State of BTB-WordNet: Overview of Semantic Relations

Zara Kancheva, Kiril Simov, Petya Osenova

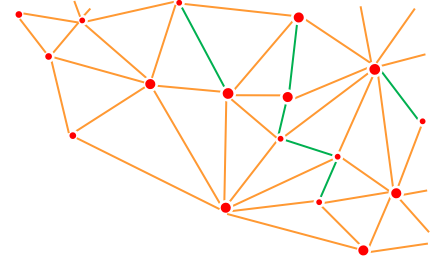
Institute of Information and Communication Technologies, BAS

Plan of the Talk

- Introduction
- Bulgarian Wordnet BTB-WN: A brief overview
- Relations in wordnets
- Relations in BTB-WN
- Conclusions and Future Work



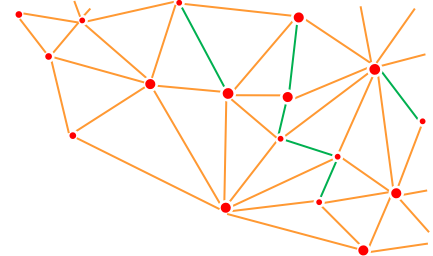
Introduction



- *What is a wordnet?*
 - lexico-semantic resource
 - senses are organized in **synsets** (synonym sets)
 - linked together in a hierarchical way by **lexical** and **semantic relations**
 - **synonymy** has a central role
- *What distinguishes wordnets from traditional dictionaries and other lexical databases?*
 - network of lexical and semantic relations between concepts (synsets)
 - **hierarchy** of senses
 - applicability/suitability for various **NLP** tasks



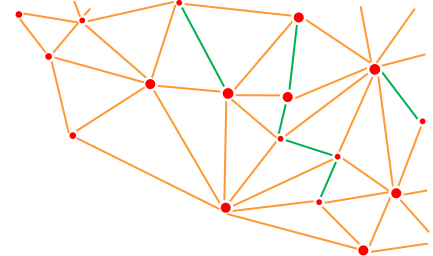
Related Works



- A non exhaustive list of literature about wordnets and wordnet relations

- **Princeton WordNet (PWN)**: Miller, G. A., Beckwith, R., Fellbaum, C., Gross, D., Miller, K. J.: *Introduction to WordNet: An On-line Lexical Database*. *International Journal of Lexicography* 3(4), 235–244 (1990)
- **Open English WordNet (OEW)**: McCrae, J. P., Rademaker, A., Bond, F., Rudnicka, E., Fellbaum, C.: *English WordNet 2019 – An Open-Source WordNet for English*. In: *Proceedings of the 10th Global WordNet Conference – GWC 2019*, pp. 245–252. Global Wordnet Association, Wroclaw (2019)
- **DanNet**: Pedersen, B. S., Nimb, S., Asmussen, J., Sørensen, N. H., Trap-Jensen, L., Lorentzen, H. .: *DanNet – the challenge of compiling a WordNet for Danish by reusing a monolingual dictionary*. *Lang Resources and Evaluation* 43, 269–299. (2009)
- **EuroWordNet**: Vossen, P.: *Introduction to EuroWordNet*. *Computers and the Humanities* 32(2/3), *EuroWordNet: A Multilingual Database with Lexical Semantic Networks*, 73–89 (1998)
- **BalkaNet**: Tufis, D., Cristea, D., Stamou, S.: *BalkaNet: Aims, Methods, Results and Perspectives. A General Overview*. *Romanian Journal on Information Science and Technology*, Dan Tufis (ed.), *Special Issue on BalkaNet*, 7(1-2), 7–41. Romanian Academy (2004)
- **plWordNet**: Rudnicka, E., Witkowski, W., Piasecki, M.: *A (Non)-Perfect Match: Mapping plWordNet onto Princeton WordNet*. In: *Proceedings of the 11th Global Wordnet Conference*, pp. 137–146. Global Wordnet Association, University of South Africa (UNISA) (2021)
- **Global WordNet**: Bond, F., Vossen, P., McCrae, J., Fellbaum, C.: *CILI: the Collaborative Interlingual Index*. In: *Proceedings of the 8th Global WordNet Conference (GWC2016)*, pp. 50–57. Global Wordnet Association, Bucharest (2016)
- **RoWordNet**: Mititelu, V. B.: *Adding Morpho-semantic Relations to the Romanian Wordnet*. In: *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC'12)*, pages 2596–2601. European Language Resources Association (ELRA), Istanbul, Turkey (2012)

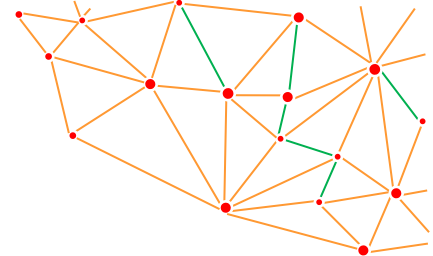
BTB-WN: Overview



The Bulgarian BulTreeBank WordNet (BTB-WN) has been created in several steps:

- manual translation of English synsets from *Core WordNet* – a subset of PWN into Bulgarian
- identification of senses used in Bulgarian Treebank *BulTreeBank*
- sense extension, which includes two activities: **a)** detection of the missing senses of processed lemmas in BulTreeBank and adding them to the BTB-WN, and **b)** a semi-automatic extraction of information from the *Bulgarian Wiktionary* mapped to synsets from PWN and then manually checked
- some older versions of BTB-WN are already part of the *Open Multilingual WordNet project*, the latest version will be added too

BTB-WN: Overview(1)



- employing the *expand model* where a monolingual wordnet is created by transferring and translating synsets and structure from an existing wordnet ~ *PWN*
- then moved to the *merge model* where a monolingual wordnet is created by building the synsets and the structure from scratch and later mapped via interlingual relations to other wordnets ~ *OEW*
- in accordance with its mappings BTB-WN uses the relation inventory of the *PWN* and *OEW*
- *equivalent-to* relation has a central role - aligns BTB-WN with *OEW*

BTB-WN: Overview (2)

→ Online browsing in BTB-WN:
<https://btb-wn.webclark.org/>

Clark Българска мрежа от думи - BTB-WordNet

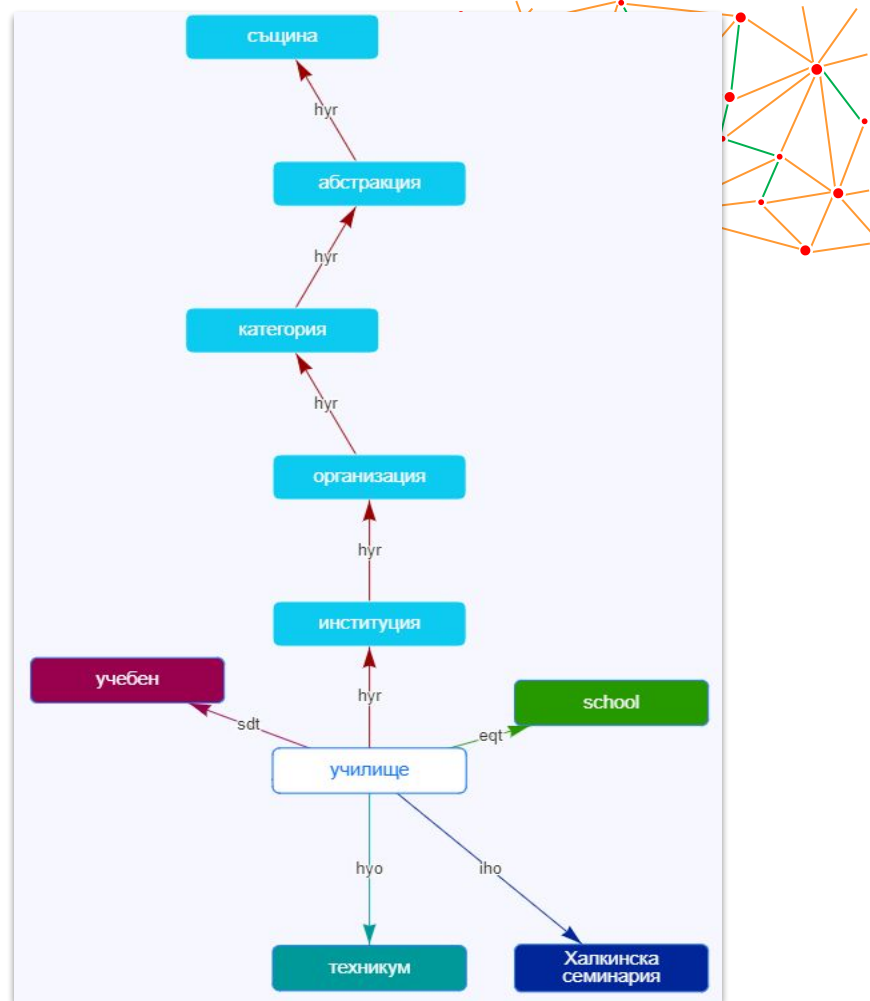
училище

търси форми

#	Лема	Част на речта	Описание
1	училище	п	Сграда, където се помещава учебно заведение, което хората посещават редовно, обикновено в своите детски и младежки години, за да придобиват знания и опит и да получават съответен документ за завършено образование. schoolhouse
2	училищен	а	
3	училищен	п	
	настоятел		

(n) училище, даквало, школо
 Период, време на учене, учебен процес в учебно заведение. **schooltime**

(n) училище, даквало, школо
 Учебно заведение, институция за обучение, образование. **school**



BTB-WN: Overview (3)

→ CLaDA-BG
Dict Editor

Lemma: училище

Synset

Part of speech	Category	Order	Definition	№	BTB
n	noun.group	1	Учебно заведение, институция за обучение, образование.	2	btbw
n	noun.artifact	2	Сграда, където се помещава учебно заведение, което хората посещават редовно, обикновено в своите детски и младежки години, за да се обучат.	1	btbw
n	noun.time	3	Период, време на учене, учебен процес в учебно заведение.	3	btbw

Lemma

50 examples

Example	Lemma	ID	Order	Part of speech
училище	училище	10241	1	n
даскало	даскало	10240	2	n
школа	школа	10242	3	n

Relations

Hypernym chain

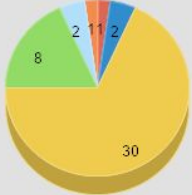
Additional information

Tickets

Open ticket

Temporary notes

Учебно заведение, институция за обучение, образование.

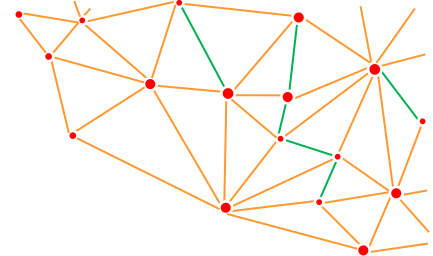


Legend:

- equivalent-to (yellow)
- hypernym (blue)
- hyponym (orange)
- instance_hyponym (green)
- mero_member (light blue)
- sem-derives-to (red)

equivalent-to

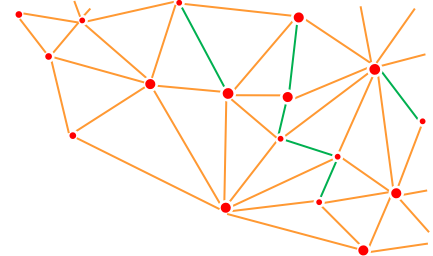
Lemma	Definition	Part of speech
school	an educational institution	n



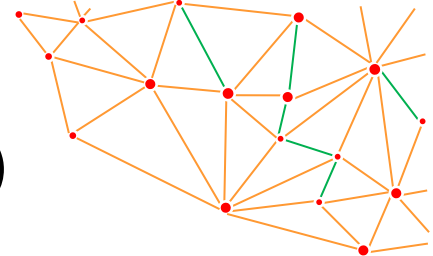
Relations in wordnets

- *Synonymy*, *hypernymy* and *hyponymy* relations could be considered mandatory
~ since *Princeton WordNet (PWN)*
- Initially in *PWN*: *synonymy*, *antonymy*, *hypernymy*, *hyponymy*, *meronymy* and *morphological relations*
- Later: *instance hypernym* and *instance hyponym*; *holonymy* and *meronymy*; *attribute*; *derivationally related form*; *domain of synset* and *member of domain*; *entailment* and *cause*; *also see*; *similar to*; *participle of verb*; *pertainym*; and *derived from adjective*
~ most wordnets

BTB-WN Relations: What is new?



- With the development of BTB-WN the need of more relations was identified ~ *problem solving oriented relations*
- Novel relations introduced in BTB-WN: *near-equivalent-to*, several *sem-derived-to* and *sem-derived-from* relations, and *predecessor* and *successor*
- Total inventory of **36** relations in version 4.0



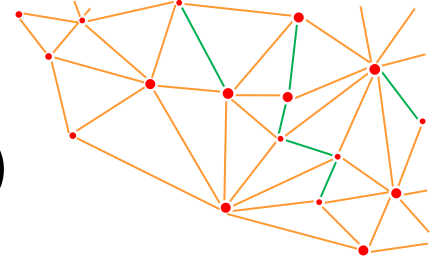
BTB-WN Relations: What is new? (1)

- *Near-equivalent-to* relation
 - nouns, verbs, adjectives and adverbs
 - very similar, but not completely matching Bulgarian and English senses
 - links head and satellite adjectives
 - links noun synsets for men and women, that are not connected in OEW

личен (*Which refers only to a certain person, originates from one person, belongs to one person.*)

equivalent-to - ‘personal’ (*concerning or affecting a particular person or his or her private life and personality*)

near-equivalent-to - ‘individual, private’ (*concerning one person exclusively*)



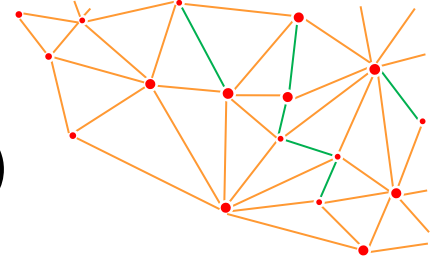
BTB-WN Relations: What is new? (2)

- *Sem-derived-to* and *sem-derived-from* relations
 - nouns and adjectives; adjectives and adverbs; verbs and adjectives; nouns, adverbs and prepositions
 - link semantically and derivationally related concepts
 - not strictly following the derivational paradigm, but making semantic connections explicit

ЛИСТО, **ЛИСТ** 'leaf' (plant organ)

sem-derives-to

ЛИСТЕН (which is related with leaf or leaves)
ЛИСТОВИДЕН (which has form of a leaf)
ТРИЛИСТЕН (which has three leaves)
ДРЕБНОЛИСТЕН (which has small leaves)



BTB-WN Relations: What is new? (3)

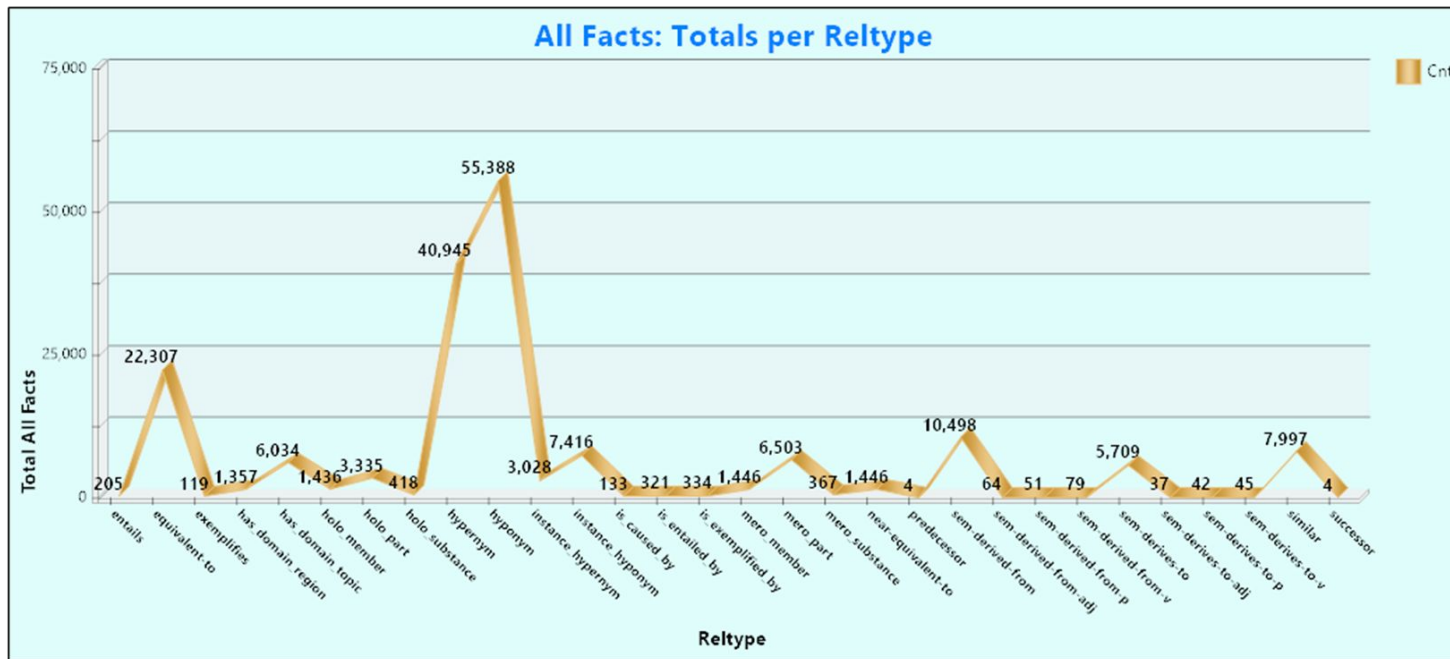
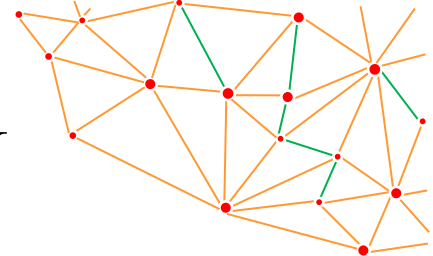
- *Predecessor* and *successor* relations
 - named entities
 - connect a concept and its previous word form/name (*predecessor*) and a concept and its subsequent, new word form/name (*successor*)
 - if the previous and the current names of a country/city/organization, etc. are presented as members of the same synset, they would have a **false** notion of equivalence

University of national and world economy (Университет за национално и стопанство)

has a *predecessor* is a *successor* of

Karl Marx Higher Institute of Economics (previous name of the institution until 1990)

Total number of relations in BTB-WN

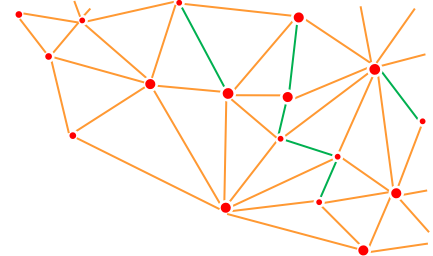


- Summary:
- *hyponymy* (55388)
- *hypernymy* (40945)
- *equivalent-to* (22307)

Distribution of Relations in BTB-WN

- Summary:
- **nouns** (137275)
- **adjectives** (23263)
- **verbs** (19600)
- **adverbs** (1022)
- **prepositions** (127)

	Pos						
Reltype	a	n	p	r	s	v	Общо
also	1153	8				30	1191
antonym			10				10
attribute	359	366				7	732
causes						99	99
domain_region	2	258				2	262
domain_topic	151	1465			13	33	1925
entails						205	205
equivalent-to	1697	15556		778	1431	2845	22307
exemplifies	1	92		17	8	1	119
has_domain_region		1357					1357
has_domain_topic		6034					6034
holo_member		1436					1436
holo_part		3335					3335
holo_substance		418					418
hypernym		34366	26		2	6551	40945
hyponym		47129	26		1	8232	55388
instance_hyponym		3028					3028
instance_hyponym		7416					7416
is_caused_by						133	133
is_entailed_by						321	321
is_exemplified_by		334					334
mero_member		1446					1446
mero_part		6503					6503
mero_substance		367					367
near-equivalent-to	349	604		94	123	276	1446
predecessor		4					4
sem-derived-from	9992					506	10498
sem-derived-from-adj				64			64
sem-derived-from-p			51				51
sem-derived-from-v	71				8		79
sem-derives-to		5708			1		5709
sem-derives-to-adj	24				13		37
sem-derives-to-p		40		2			42
sem-derives-to-v						45	45
similar	5434	1	14	54	1865	629	7997
successor		4					4
Total	19233	137275	127	1022	4030	19600	181287



Conclusion and Future Work

- BTB-WN adopts the *common wordnet relations* that originate from the *PWN*, but are also used in many wordnets for various languages
- With the development and expansion of BTB-WN the *need of more relations* was recognized, so several new relations were introduced
- They *solve different problems* and provide the concepts with *proper position in the hierarchy*, contributing for their *better interpretation*
- Relations *strictly following the derivational directions* are planned to be introduced in BTB-WN



Thank you!