

Serbian Wordnet for Biomedical Sciences

Sanja Antonic, University library University of Belgrade, Serbia

Cvetana Krstev, Faculty of Philology, University of Belgrade, Serbia

Wordnet is a lexical semantic network having as nodes the sets of synonymous word (synsets) which are linked by various semantic relations. The first Wordnet built was the so-called Princeton Wordnet consisting today of approximately 140,000 synsets. Due to its remarkable size and successful inclusion in various computer-based applications it is considered as a de facto standard. In this paper we present the new electronic information resource Serbian Wordnet (SWN) and its initial development. Serbian Wordnet is based on the same principles as Princeton Wordnet (PWN), which means that it has not only inherited its basic structure but also the additional information associated to PWN. The development of Serbian Wordnet started with Balkanet project which was funded by European Commission from (2001-2004). The main goal of the project was development of wordnets for the Balkan languages: Bulgarian, Greek, Romanian, Serbian, Turkish and Czech and their alignment with PWN. The development of Serbian Wordnet continues after Balkanet with its enhancement in various domains. We describe how various traditional and Internet resources are being used for the enhancement of SWN in domain of biomedicine and how we see the possible practical usage of SWN in information retrieval this domain.



Table 1. Distribution of literals per synsets for (Serbian Wordnet - SWN):

Part of speech (PoS)	sunsets	percentage	literals	1/s	1	2	3	max
nouns	9964	79.81%	17164	1.72	4823	3710	1101	10
verbs	1864	14.93%	3487	1.87	778	700	287	11
adjectives	639	5.12%	807	1.26	501	113	22	8
adverbs	18	0.14%	18	1.00	1			1
total	12485	100%	21476	1.72	6120	4523	1320	

Table 2. Status of the Serbian Wordnet

	The end of Balkanet project 2001	September 2007
Synsets	8050	12485
Nouns	5919	9964
Verbs	1803	1864
Adjectives	324	639
Adverbs	13	18
Literals	13295	21476
The medium length of synsets	~ 1.625	1.72
The average number of senses per literal	~ 1.21	1.14

Table 3.

	The end of Balkanet project 2004	September 2007
Base/Concepts		
BC1	1219	1219
BC2	3469	3469
BC3	1369	1953
Domain specific synsets	305	453
Law	103	176
Politics	101	125
Economy	101	152
Balkan specific synsets	117	535
Serbian specific synsets	206	156