Bachelor Project Report

Ivan Cvetanovic

The list of things implemented into the project currently:

- Input of a single or multiple words.
- English Translation of the input as a whole and word by word in the table.
- Random Sentence generation by using 12 fairy tales on bajke.rs
- Morphological Analysis: Lemma, Type, Number, Person, Case and Gender.
- Dependency Parsing.
- Named Entity Recognition.

Visualization:

Text Analyzer for Serbian Language
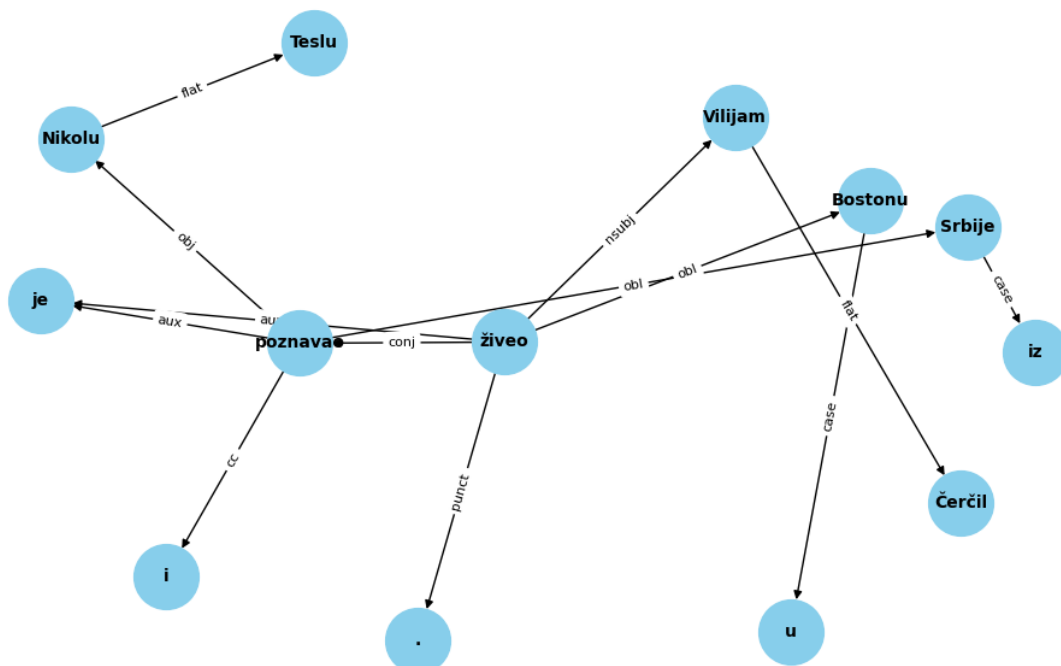
Enter something…   Submit Random

Original Input:
Vilijam Čerčil je živeo u Bostonu i poznavao je Nikolu Teslu iz Srbije.
Translated Input:
William Churchill lived in Boston and knew Nikola Tesla from Serbia.

Original Word	Translation	Lemma (Base Form)	Word Type	Number	Person	Case	Gender	Head	Dependency Relation	Named Entity
Vilijam	William	Vilijam	PROPN	Sing	/	Nom	Masc	4	nsubj	B-PER
Čerčil	Churchill	Čerčil	PROPN	Sing	/	Nom	Masc	1	flat	I-PER
je	be	biti	AUX	Sing	3	/	/	4	aux	/
živeo	live	živeti	VERB	Sing	/	/	Masc	0	root	/
u	in	u	ADP	/	/	Loc	/	6	case	/
Bostonu	Boston	Boston	PROPN	Sing	/	Dat	Masc	4	obl	B-LOC
i	and	i	CCONJ	/	/	/	/	8	cc	/
poznavao	know	poznavati	VERB	Sing	/	/	Masc	4	conj	/
je	be	biti	AUX	Sing	3	/	/	8	aux	/
Nikolu	Nikola	Nikola	PROPN	Sing	/	Acc	Masc	8	obj	B-PER
Teslu	teslo	teslu	NOUN	Sing	/	Acc	Masc	10	flat	I-PER
iz	from	iz	ADP	/	/	Gen	/	13	case	/
Srbije	Serbia	Srbija	PROPN	Sing	/	Gen	Fem	8	obl	B-LOC
.	.	.	PUNCT	/	/	/	/	4	punct	/

Known Issues:

- Sometimes the translation of single word entries is not correct because the context is unknown. Example: the word "da" is always translated to "yeah" but that is not the only possible translation.
- Sometimes the server crashes because mathplotlib is not used in the main thread.
- The visualization can sometimes overlap in some cases where sentences are really long.