



# Тесла



# ICIST2024

International Conference on Information Society and Technologies  
Kopaonik, Serbia 10-13 March, 2024

## Semantic Textual Similarity of Courses based on Text Embeddings

Oliyera Kitanović, Aleksandra Tomašević,  
Mihailo Škorić, Ranka Stanković and Ljiljana Kolonja  
Serbia, University of Belgrade: Faculty of Mining and Geology, Serbia



Фонд за науку  
Републике Србије

# Research questions

- **RQ1.** Can a text vectorization approach, based on the course descriptions, be useful to identify courses with high semantic similarity?
- **RQ2.** To what extent can similarities between professor's specialty be calculated based on course vectors?
- **RQ3.** How does course similarity detection differ when using vectorization based on the course description compared to using the course title and description?





# Evaluation

	Professor ID	Course ID	Symilar1	Eval1	Symilar2	Eval2	Symilar3	Eval3	Symilar4	Eval4	Symilar5	Eval5
1	540	8382	[8597, 90]	90	[8535, 81]		[8457, 78]		[8680, 76]		[8526, 76]	
2	540	8597	[8382, 90]	90	[8535, 81]		[8538, 78]		[8680, 78]		[8145, 78]	
3	4060	8322	[8323, 99]	99	[8139, 94]		[8745, 92]		[8376, 89]		[8338, 89]	
4	388	8513	[8483, 91]		[8546, 89]	5	[8471, 88]	3	[8514, 88]	3	[8201, 88]	1
5	388	8546	[8483, 90]		[8513, 89]	5	[8514, 89]		[8491, 88]		[8399, 88]	
6	168	8589	[8655,100]	100	[8463, 90]		[8525, 88]		[8587, 86]		[8542,86]	
7	168	8655	[8589,100]	100	[8463, 89]		[8525, 87]		[8587, 86]		[8542,86]	
8	1784	8768	[8766, 99]	99	[8769,99]	99	[8767, 98]	99	[8722, 86]		[8729, 84]	
9	1600	8652	[8510, 94]	95	[8509, 89]		[8649, 87]		[8625, 87]		[8288, 86]	

Rows 4 and 5 => calculated similarity <> manual evaluation

Rows 6 and 7 => complete match

Row 3 => "metallic" <> "non-metallic".

Row 8 => "flora"/"fauna"/"animals", "organism"/"microorganism"

Row 9 => same general topics, different wording

Thank you for your  
attention!

olivera.kitanovic@rgf.bg.ac.rs

<https://tesla.rgf.bg.ac.rs/>

A research conducted with the support of  
the Science Fund of the Republic of  
Serbia, project number: 7276.

**Text Embeddings -  
Serbian Language Applications –**

**TESLA**

**ТЕСЛА**

