

Zgłoszenie tematu pracy dyplomowej :: **STUDIA II STOPNIA** ::

na rok akademicki 2023/24

Promotor:	dr hab. Jozef Kapusta, prof. UP
Temat pracy magisterskiej (j. polski oraz j. angielski):	Comparison of Models for Generating Word Vectors Porównanie modeli do generowania wektorów słów
Zakres i oczekiwane rezultaty pracy:	<p>Machine learning models take vectors (arrays of numbers) as input. When working with text, the first thing you must do is come up with a strategy to convert strings to numbers (or to "vectorize" the text) before feeding it to the model. Word embeddings give us a way to use an efficient, dense representation in which similar words have a similar encoding.</p> <p>In the theoretical part: The thesis aims are summarization and comparison of the "traditional" methods for word embedding (Word2vec, Tf-Idf, GloVe). Exploring the new experimental methods for word embeddings from current scientific sources is expected.</p> <p>In the practical part: The practical part of the thesis will focus on creating text classifier and impact assessment word embedding methods for performance measures of the models (accuracy, precision, recall, f1-score).</p>
*Aspekt naukowy, problemowy pracy:	definition and implementation of a documents models, implementation methods natural language, implementation machine learning methods.
Literatura	<ul style="list-style-type: none"> • Bird, S., Klein E., and Loper, E. (2009). Natural Language Processing with Python - Analyzing Text with the Natural Language Toolkit. O'Reilly Media. • Bengfort, B., Ojeda, T., Bilbro, R. (2018). Applied Text Analysis with Python: Enabling Language - Aware Data Products with Machine Learning, O'Reilly Media, 332 p. • Natural Language Toolkit, online: https://www.nltk.org/ • Kapusta, J., Drlik, M., Munk, M. (2021). Using of N-grams From Morphological Tags for Fake News Classification. PEERJ COMPUTER SCIENCE, PeerJ, 7, e624. • Nagy, K., Kapusta, J. (2021). Improving fake news classification using dependency grammar. PLOS ONE, Public Library of Science, 16(9), e0256940 • Chen, JT., Tao, YB., Lin, H. (2018). Visual exploration and comparison of word embeddings. JOURNAL OF VISUAL LANGUAGES AND COMPUTING, 48, pp. 178-186
, **Oprogramowanie, język programowania, środowisko systemowe:	Jupyter Notebook Environment (Python)

Zgłoszenie tematu pracy dyplomowej :: **STUDIA II STOPNIA** ::
na rok akademicki 2023/24

**Środowisko uruchomieniowe:	
Dodatkowe wymagania i uwagi:	English language

UWAGA:

W polu literatura należy wskazać minimum 1 publikację z listy czasopism punktowanych wg wykazu MEiN z dnia 21 grudnia 2021 r. związaną z proponowanym tematem pracy dyplomowej.

* Regulamin studiów § 36 2. Praca dyplomowa na profilu praktycznym, podobnie jak praca inżynierska, powinna mieć charakter aplikacyjny, badawczy, projektowy lub oceniający praktykę w świetle teorii.

** pola opcjonalne