

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

na rok akademicki 2021/22

<b>Promotor:</b>	<b>dr hab. Jozef Kapusta, prof. UP</b>
Temat pracy magisterskiej (j. polski oraz j. angielski):	Comparison of Models for Generating Word Vectors <i>Porównanie modeli do generowania wektorów słów</i>
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.
**Oprogramowanie, język programowania, środowisko systemowe:	Jupyter Notebook Environment (Python)
**Środowisko uruchomieniowe:	Windows or Linux
Dodatkowe wymagania i uwagi:	English language
**Literatura:	<ul style="list-style-type: none"> <li>• Bird, S., Klein E., and Loper, E. (2009). Natural Language Processing with Python - Analyzing Text with the Natural Language Toolkit. O'Reilly Media.</li> <li>• 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.</li> <li>• Natural Language Toolkit, online: <a href="https://www.nltk.org/">https://www.nltk.org/</a></li> </ul>

\* Regulamin studiów § 35 2. Praca dyplomowa na profilu praktycznym, podobnie jak praca inżynierska, powinna mieć charakter aplikacyjny, badawczy, projektowy lub oceniający praktykę w świetle teorii.

\*\* pola opcjonalne