

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

na rok akademicki 2020/21

<b>Promotor:</b>	<b>dr hab. Jozef Kapusta, prof. UP</b>
Temat pracy magisterskiej (j. polski, j.angielski):	Natural Language Processing Methods for the Polish Language <i>Metody przetwarzania języka naturalnego dla języka polskiego</i>
Zakres pracy i oczekiwane rezultaty praktyczne:	Natural Language Processing is a sub-field of Artificial Intelligence that is focused on enabling computers to understand and process human languages, to get computers closer to a human-level understanding of language. The aim of the practical part is to modify and apply the basic Natural Language Processing methods for newspaper news analysis in the Polish language. Essential characteristics will be the count of sentences, count of words, average count words per sentence, "bag of words". Advanced features will be obtained from the morphological analysis, analysis of synsets, document similarities (words to vectors, TF/IDF etc.). The student will be using known libraries (NLTK, UDPipe, Polyglot, TreeTagger etc.) and will find specific solutions for the Polish language.
Aspekt naukowy, problemowy, innowacyjny pracy:	Selection and implementation of natural language processing methods, modify methods for specifics of Polish language as flective type of language.
Oprogramowanie, język programowania, środowisko systemowe:	Jupyter Notebook environment (Python)
Środowisko uruchomieniowe	Windows or Linux
Dodatkowe wymagania i uwagi:	english language
Literatura:	<ol style="list-style-type: none"> <li>1. Steven Bird, Ewan Klein, and Edward Loper: Natural Language Processing with Python - Analyzing Text with the Natural Language Toolkit. O'Reilly Media, 2009.</li> <li>2. Benjamin Bengfort, Tony Ojeda, Rebecca Bilbro: Applied Text Analysis with Python: Enabling Language - Aware Data Products with Machine Learning, O'Reilly Media, 2018, 332 p.</li> <li>3. UDPipe, online: <a href="http://lindat.mff.cuni.cz/services/udpipe/info.php">http://lindat.mff.cuni.cz/services/udpipe/info.php</a></li> <li>4. Natural Language Toolkit, online: <a href="https://www.nltk.org/">https://www.nltk.org/</a></li> </ol>