

:: STUDIA II STOPNIA ::

Zgłoszenie tematu pracy dyplomowej na rok akademicki 2019/2020

Promotor:	dr hab. Jozef Kapusta, prof. UP
Temat pracy magisterskiej (j. polski, j.angielski):	Metody przetwarzania języka naturalnego dla języka polskiego <i>Natural Language Processing Methods for the Polish Language</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.
Aspekt naukowy, problemowy, innowacyjny pracy:	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.
*Oprogramowanie, język programowania, środowisko systemowe:	Jupyter Notebook environment (Python)
*Środowisko uruchomieniowe	Windows or Linux
Dodatkowe wymagania i uwagi:	The diploma thesis will be written in English.
*Literatura:	<ol style="list-style-type: none">1. Steven Bird, Ewan Klein, and Edward Loper: Natural Language Processing with Python - Analyzing Text with the Natural Language Toolkit. O'Reilly Media, 2009.2. Benjamin Bengfort, Tony Ojeda, Rebecca Bilbro: Applied Text Analysis with Python: Enabling Language - Aware Data Products with Machine Learning, O'Reilly Media, 2018, 332p.3. UDPipe, online: http://lindat.mff.cuni.cz/services/udpipe/info.php4. Natural Language Toolkit, online: https://www.nltk.org/

*pola opcjonalne