

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

na rok akademicki 2022/23

Promotor:	dr hab. Jozef Kapusta, prof. UP
Temat pracy magisterskiej (j. polski oraz j. angielski):	Documents Similarity based on Cluster Analysis <i>Podobieństwo dokumentów na podstawie analizy skupień</i>
Zakres i oczekiwane rezultaty pracy:	Text clustering is an important application of data mining. It is concerned with grouping similar text documents together. The aim of the thesis is to design several models for clusterization using selected clustering techniques (k-means, k-medoids, etc.). The student will extract a dataset from the website with stories or speeches. The student will create a vector representation of documents (TF-IDF, etc.) and apply clustering analysis to the data. The quality of the obtained models will be evaluated and compared.
*Aspekt naukowy, problemowy pracy:	Selection and implementation of natural language processing methods, implementation selected machine learning methods.
Literatura	<ol style="list-style-type: none"> 1. 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. 2. Bird, S., Klein, E., Loper, E. (2009). Natural Language Processing with Python - Analyzing Text with the Natural Language Toolkit. O'Reilly Media 3. Natural Language Toolkit, online: https://www.nltk.org/ 4. scikit-learn: Machine Learning in Python, online: https://scikit-learn.org/stable/ 5. Shan, C., Du, YG. (2021). A Web Service Clustering Method Based on Semantic Similarity and Multidimensional Scaling Analysis. In. Scientific Programming, 2021 6. Buatoom, U., Kongprawechnon, W., Theeramunkong, T. (2020). Document Clustering Using K-Means with Term Weighting as Similarity-Based Constraints. In. Symmetry-Basel, Vol. 12/6.
**Oprogramowanie, język programowania, środowisko systemowe:	Jupyter Notebook environment (Python)
**Środowisko uruchomieniowe:	Windows or Linux
Dodatkowe wymagania i uwagi:	english language

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

na rok akademicki 2022/23

* 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