

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

na rok akademicki 2024/25

<b>Promotor:</b>	<b>dr hab. Jozef Kapusta, prof. UKEN</b>
Temat pracy magisterskiej (j. polski oraz j. angielski):	Comparison of Methods for Name Entities Recognition Porównanie metod wyodrębniania jednostek nazw
Zakres i oczekiwane rezultaty pracy:	<p>Information extraction obtains structured data or knowledge from an unstructured text by identifying references to named entities as well as stated relationships between such entities. Named entity recognition is probably the first step towards information extraction that seeks to locate and classify named entities in text into pre-defined categories such as the names of persons, organizations, locations, expressions of times, quantities, monetary values, percentages, etc.</p> <p>In the theoretical part: The theoretical part of the thesis will summarize the methods and approaches for named entity recognition. Exploring new models and techniques for Polish and English languages is an essential part of the work.</p> <p>In the practical part: The thesis aims to compare selected libraries and approaches for named entity recognition based on performance measures. The student selects libraries for entity recognition in Python (NLTK, spaCy, etc.) and applies their methods to a selected dataset. An important part of the thesis will be a mapping table for comparing different entity tags in different libraries. The comparison will focus on the suitability of selected libraries for the Polish and English languages.</p>
*Aspekt naukowy, problemowy pracy:	implementation of a named entity recognition, implementation of selected libraries, critical evaluation of libraries possibilities
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> <li>• Zhang Y., Xiao G. (2024). Named Entity Recognition Datasets: A Classification Framework. In. International Journal of Computational Intelligence Systems, Vol. 17 / 1, art. no. 71</li> <li>• Dagdelen J., et. al. (2024). Structured information extraction from scientific text with large language models. In. Nature Communications, Vol. 15 /1, art. no. 1418</li> </ul>
, **Oprogramowanie, język programowania, środowisko systemowe:	Jupyter Notebook Environment (Python)

Zgłoszenie tematu pracy dyplomowej :: **STUDIA II STOPNIA** ::  
na rok akademicki 2024/25

---

**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