

<b>Promotor:</b>	<b>dr hab. Jozef Kapusta, prof. UP</b>
Temat pracy dyplomowej (j. polski, j. angielski):	<i>Information Extraction from Unstructured Texts</i> <i>Ekstrakcja informacji z tekstów nieustrukturyzowanych</i>
Zakres pracy i oczekiwane rezultaty praktyczne:	<p>Process of information extraction is used to extract useful information from unstructured or semi-structured data. Information extraction obtains structured data or knowledge from an unstructured text by identifying references to named entities well as stated relationships between such entities. The main area which is helpful for information extraction from text is natural language processing. There are methods of part-of-speech identification which use syntactical and morphological analysis of natural language.</p> <p>The aim of the thesis is implemented methods of syntactical and morphological analysis to the Polish language. Based on these methods the student will find some simple relationships from the text. There are some web-services and libraries for syntactical and morphological analysis. The student will develop a tool for simplification (reduction) of text information from news articles. The tool (based on web-services and libraries) will find the base elements of part-of-speech (nouns, verbs, etc.) and elements from sentence construction (subjects, predicates, etc.). These elements will be used for the creation of simplification of text information from articles.</p>
Aspekt inżynierski*:	Implementation methods of syntactical and morphological analysis to the Polish language, development a tool for reduction of text information from news articles.
Wymagane oprogramowanie/języki programowania**:	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</li> </ul>

Zgłoszenie tematu **INŻYNIERSKIEJ** pracy dyplomowej

**STUDIA I STOPNIA** rok akademicki 2019/20

---

	<p>with Python: Enabling Language - Aware Data Products with Machine Learning, O'Reilly Media, 332 p.</p> <ul style="list-style-type: none"><li>• 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>• Natural Language Toolkit, online: <a href="https://www.nltk.org/">https://www.nltk.org/</a></li></ul>
--	---

**\*należy uzasadnić/wskazać, czy praca spełnia wymagania inżynierskie**

**\*\*pola opcjonalne**