GDAŃSK UNIVERSITY OF TECHNOLOGY

Subject card

Subject name and code	Natural Language Processing, PG_00048268		
Field of study	Informatics		
Date of commencement of studies	February 2024	Academic year of realisation of subject	2024/2025
Education level	second-cycle studies	Subject group	Optional subject group Subject group related to scientific research in the field of study
Mode of study	Full-time studies	Mode of delivery	at the university
Year of study	2	Language of instruction	Polish
Semester of study	3	ECTS credits	2.0
Learning profile	general academic profile	Assessment form	assessment
Conducting unit	Department of Intelligent Interactive Systems -> Faculty of Electronics, Telecommunications and Informatics		
Name and surname of lecturer (lecturers)	Subject supervisor	dr hab. inż. Jan Daciuk	
	Teachers	dr hab. inż. Jan Daciuk	

Lesson types and methods of instruction	Lesson type	Lecture	Tutorial	Laboratory	Project	Seminar	SUM
	Number of study hours	15.0	0.0	0.0	15.0	0.0	30
	E-learning hours included: 0.0						

Learning activity and number of study hours	Learning activity	Participation in didactic classes included in study plan	Participation in consultation hours	Self-study	SUM
	Number of study hours	30	4.0	16.0	50

Subject objectives	The purpose of this subject is to familiarize students with basic techniques used in natural language processing and to to prepare them for work on development of applications for that domain.

Learning outcomes			
	Course outcome	Subject outcome	Method of verification
	[K7_U42] can solve engineering and research problems including design, assessment and maintenance of information systems and applications, using experimental methods and management techniques	Can solve engineering and research problems n design, evaluation, and maintenance of systems and application in natural language processing domain using experimental methods and management techniques.	[SU1] Assessment of task fulfilment [SU2] Assessment of ability to analyse information [SU3] Assessment of ability to use knowledge gained from the subject [SU4] Assessment of ability to use methods and tools
	[K7_U06] can analyse the operation of components, circuits and systems related to the field of study; measure their parameters; examine technical specifications; interpret obtained results and draw conclusions	Can analyze functioning of natural language processing systems and can draw conclusions from that analysis.	[SU2] Assessment of ability to analyse information [SU4] Assessment of ability to use methods and tools
	[K7_W03] Knows and understands, to an increased extent, the construction and operating principles of components and systems related to the field of study, including theories, methods and complex relationships between them and selected specific issues - appropriate for the curriculum.	Knows and understands principles and methods for building dictionaries, morfological analysis and synthesis, part-of-speech tagging, parsing and semantic analysis.	[SW1] Assessment of factual knowledge
	[K7_W41] Knows and understands, to an increased extent, the standards, production methods, life cycle and development trends of software as well as information systems and applications.	Knows and profoundly understands standards, development methods, life cycle, and development trends of natural language processing software.	[SW1] Assessment of factual knowledge
	[K7_U04] can apply knowledge of programming methods and techniques as well as select and apply appropriate programming methods and tools in computer software development or programming devices or controllers using microprocessors or programmable elements or systems specific to the field of study, making assessment and critical analysis of the prepared software as well as a synthesis and creative interpretation of information presented with it	Can use acquired knowledge to develop natural language processing software.	[SU1] Assessment of task fulfilment [SU4] Assessment of ability to use methods and tools

Subject contents

Lecture:

1. Introduction, segmentation.
2. Morphology: concatenation.
3. Morphology: spelling rules.
4. Morphological analysis, lexicon as an automaton.
5. Incremental construction of lexicons, perfect hashing.
6. Spelling correction with automata.
7. Spelling correction: noisy channel model.
8. Document retrieval.
9. Tagging.
10. Syntax description methods.
11. Parsing with unification.
12. Representation of meaning.
13. Syntax-driven semantic analysis.
14. Dialogue.
15. Word sense disambiguation.

Project:

1. Linux/Unix tools for text processing.
2. Morphology.
3. Segmentation.
4. Document retrieval.
5. Syntax and semantics.

Prerequisites and co-requisites

Programming skills (mainly arbitrary scripting languages), understanding Prolog programs.

<table>
<tr><td>Assessment methods and criteria</td><td colspan="2"><table>
<tr><th>Subject passing criteria</th><th>Passing threshold</th><th>Percentage of the final grade</th></tr>
<tr><td>project</td><td>50.0%</td><td>50.0%</td></tr>
<tr><td>exam</td><td>50.0%</td><td>50.0%</td></tr>
</table></td></tr>
<tr><td rowspan="3">Recommended reading</td><td>Basic literature</td><td>1. Daniel Jurafsky, James Martin, Speech and Language Processing. An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition, Second Edition, Prentice Hall, 2008.
2. Christopher D. Manning, Hinrich Schütze, Foundations of Statistical Natural Language Processing, MIT Press, 2000.
3. Emmanuel Roche, Yves Schabes, Finite-State Language Processing, MIT Press, 1997.
4. Computational Linguistics journal and proceedidngs of ACL (Association for Computational Linguistics) conferences. Available from http://acl.ldc.upenn.edu/ – ACL Anthology.</td></tr>
<tr><td>Supplementary literature</td><td>1. Zygmunt Saloni, Włodzimierz Gruszczyński, ´Marcin Woliński, Robert Wołosz, Słownik gramatyczny języka polskiego. Podstawy teoretyczne. Instrukcja użytkowania, Wiedza Powszechna, 2007.
2. Stanisław Mędak, Słownik form koniugacyjnych czasowników polskich, Universitas, Kraków, 2004.
3. Stanisław Mędak, Słownik odmiany rzeczowników polskich, Universitas, Kraków, 2003.</td></tr>
<tr><td>eResources addresses</td><td>Adresy na platformie eNauczanie:</td></tr>
<tr><td>Example issues/ example questions/ tasks being completed</td><td colspan="2"></td></tr>
<tr><td>Work placement</td><td colspan="2">Not applicable</td></tr>
</table>