

Subject card

Subject name and code	Computer processing of documents- lecture, PG_00150266						
Field of study	English Studies						
Date of commencement of studies	October 2025		Academic year of realisation of subject		2026/2027		
Education level	Master's studies		Subject group		Obligatory subject group in the field of study 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		1.0		
Learning profile	academic		Assessment form		credit		
Conducting unit	Institute of English and American Studies -> Faculty of Languages -> Rector						
Name and surname of lecturer (lecturers)	Subject supervisor		mgr Dagmara Solska				
	Teachers						
Lesson types	Lesson type	Lecture	Tutorial	Laboratory	Project	Seminar	SUM
	Number of study hours	15.0	0.0	0.0	0.0	0.0	15
	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	15		0.0		9.0	24
Subject objectives	The aim of the course is to apply basic knowledge of computational methods used in document processing to perform practical tasks such as document representation, document comparison, information retrieval based on query matching, document classification, and the use of digital linguistic resources.						

Learning outcomes	Course outcome	Subject outcome	Method of verification
	[FAMU2_U10] Can independently plan and implement their own lifelong learning and guide others in this area within the framework of English philology and their chosen field of professional activity.	Can independently plan and implement their own lifelong learning in the field of professional activities of an English language processing specialist.	[SU2] presentation/project/paper/report
	[FAMU2_K02] Is prepared to recognise the importance of knowledge and skills in English Studies in solving cognitive and practical problems and to seek the advice of a supervisor in their chosen place of work in the event of difficulty in solving problems on their own.	Is willing to recognise the importance of knowledge and skills in computer-based document processing used in natural language processing in solving cognitive and practical problems and to consult others when having difficulty solving problems independently.	[SK2] presentation/project/paper/report
	[FAMU2_K01] Is ready to critically evaluate the extent of their knowledge and skills, in particular in the field of English-language linguistics and literature and English.	Is prepared to critically evaluate the extent of his/her knowledge and skills in the field of computer document processing used in natural language processing.	[SK2] presentation/project/paper/report
	[FAMU2_W12] Knows and understands the main development trends in literature, in particular with regard to research into English-language literature.	Knows and understands selected issues in the field of information sciences, necessary as a context for the professional activities of an English natural language processing specialist.	[SW2] presentation/project/paper/report
	[FAMU2_U04] Can select and apply information and communication techniques (ICT) during the acquisition and processing of information for research and professional purposes within the framework of English Studies.	Is able to select and apply information and communication technology (ICT) when acquiring and processing information needed in the professional activities of an English language processing specialist.	[SU2] presentation/project/paper/report
Subject contents	• Representation of textual documents • Character encoding, metadata, and the structure of digital documents • Introduction to digital linguistic resources (e.g. corpora, dictionaries, WordNet) • Data extraction from documents: headers, paragraphs, tables, structural elements • Web scraping: acquiring data from websites (using <i>BeautifulSoup</i>, <i>requests</i>) • Document comparison methods: lexical similarity, distance measures, duplicate detection • Document retrieval based on user queries: full-text search, filtering mechanisms • Text document classification: basics of rule-based and statistical approaches • Content processing and analysis using Python-based tools • Final project: independent development of a solution for document processing or analysis using the tools and methods covered in the course		
Prerequisites and co-requisites	The choice of the PJN specialisation. Credit for the course Tools of Modern Computer Science.		

Assessment methods and criteria	Subject passing criteria	Passing threshold	Percentage of the final grade
	student project	51.0%	100.0%
Recommended reading	Basic literature	Ryan, M. (2021). <i>Web Scraping with Python</i> . Packt Publishing. Eisenstein, J. (2022). <i>Introduction to Natural Language Processing</i> . The MIT Press.	
	Supplementary literature	D. Jurafsky, J. H. Martin - <i>Speech and Language Processing: An Introduction to Natural Language Processing, Speech Recognition and Computational Linguistics</i> . 2nd edition. Prentice-Hall. 2009.	
	eResources addresses		
Example issues/ example questions/ tasks being completed	• How can data be extracted from websites and social media platforms? (web scraping, APIs) • What are the differences between a documents structure and its content? • How can documents containing specific phrases or keywords be searched and filtered? • What metrics are used to compare documents? • How can documents be represented in a vector space (TF-IDF, embeddings)? • How can documents be classified based on their content (rule-based methods, statistical models)? • How can tools like spaCy and transformers be used in document analysis? • How can one create custom document collections for linguistic analysis? • Final project: design and implement an original document processing pipeline (e.g., analysis of Reddit posts, news articles, emails, or forum threads).		
Work placement	Not applicable		

Document generated electronically. Does not require a seal or signature.