

Subject card

Subject name and code	Computer processing of documents- lecture, PG_00150280						
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	0.0	15.0	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 the application of IT methods for document processing to carry out practical tasks: document representation, methods of comparing documents, searching for documents matching a query, document classification, using digital language 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.	The student can independently plan and implement their own lifelong learning in computer-based document processing, essential for practical use in natural language processing	[SU2] 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.	The student knows and understands selected issues in the field of computer-based document processing necessary for practical analysis.	[SW2] 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.	The student is ready to critically assess the scope of their own knowledge and skills in computer-based document processing necessary for practical analysis.	[SK2] 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.	The student is ready to recognize the importance of knowledge and skills in computer-based document processing for solving practical problems, and to seek guidance from an academic teacher when encountering difficulties in solving problems independently.	[SK2] 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.	The student is able to select and apply information and communication technologies (ICT) techniques for computer-based document processing necessary for practical applications.	[SU2] presentation/project/paper/report
Subject contents	- Representation of text documents - Character encoding, metadata, and the structure of digital documents - Introduction to digital linguistic resources (e.g. corpora, dictionaries, WordNet) - Data extraction from documents: headings, paragraphs, tables, structural elements - Web scraping: retrieving data from websites (using BeautifulSoup, requests) - Document comparison methods: lexical similarity, distance metrics, duplicate detection - Retrieving documents relevant to user queries (full-text search, filters) - Text document classification: fundamentals of rule-based and statistical approaches - Processing and analysing document content using Python-based tools - Final project: independent development of a solution for document processing or analysis using the methods and tools covered in the course		
Prerequisites and co-requisites	Selection of the NLP specialization. Completion of the course Tools of Contemporary Computer Science.		
Assessment methods and criteria	Subject passing criteria	Passing threshold	Percentage of the final grade
	project	51.0%	100.0%

Recommended reading	Basic literature	1. Patel, A. A., & Arasanipalai, A. U. (2022). <i>Applied Natural Language Processing in the Enterprise</i>. Apress. 2. Ryan, M. (2021). <i>Modern Web Scraping with Python</i>. Packt Publishing. 3. Eisenstein, J. (2022). <i>Natural Language Processing</i>. The MIT Press.
	Supplementary literature	1. Kowsari, K., & Brown, D. E. (2023). <i>Text Analytics with Python: A Practitioner's Guide to Natural Language Processing</i>. Springer. 2. Berg, D. (2022). <i>Python Web Scraping Cookbook</i> (2nd ed.). Packt Publishing. 3. Bird, S. (2023). <i>Natural Language Processing with Python and spaCy</i>. O'Reilly. 4. Munz, P. (2021). <i>Intelligent Document Processing with NLP</i>. O'Reilly. 5. Rao, D., & McMahan, B. (2022). <i>Natural Language Processing with PyTorch</i>. O'Reilly.
	eResources addresses	
Example issues/ example questions/ tasks being completed	• How to retrieve data from websites and social media platforms? (web scraping, API) • What is the difference between a documents structure and its content? • How to search and filter documents containing specific phrases or keywords? • What metrics can be used to compare documents? • How to represent documents in a vector space (TF-IDF, embeddings)? • How to classify documents based on their content (rules, statistical models)? • How to use tools such as spaCy and transformers in document analysis? • How to create your own document collections for linguistic analysis? • Final project: design and implementation of an individual document processing pipeline (e.g. analysis of Reddit posts, news articles, emails, discussion forums).	
Work placement	Not applicable	

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