

Subject card

Subject name and code	Data Analytics Classes, PG_00188625						
Field of study	English Studies						
Date of commencement of studies	October 2027		Academic year of realisation of subject			2028/2029	
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							
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		1.0		9.0	25
Subject objectives	The aim of the course is to develop students practical skills in analysing linguistic data by applying methods drawn from corpus linguistics, computational linguistics, and natural language processing. The classes are designed to enable students to systematically collect, preprocess, and interpret English-language textual data with the support of contemporary information and communication technologies. The course further aims to foster the ability to critically assess the output of automated analytical tools, to employ empirical evidence in linguistic inquiry, and to formulate well-grounded conclusions based on both quantitative and qualitative approaches. An additional objective is to promote responsible and ethically informed use of digital linguistic resources in academic and professional contexts relevant to the work of an English philologist.						

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.	Is able to independently plan and pursue lifelong learning and to guide others in this process within the domain of English philology and in relevant professional contexts, particularly in relation to data analytics applicable to computational linguistics, corpus linguistics, and natural language processing.	[SU2] presentation/project/paper/report
	[FAMU2_U02] Can apply knowledge in the field of linguistics and literature in solving practical problems in the field of the chosen specialty.	Is able to apply knowledge from linguistics and literary studies to solving practical problems within the chosen specialisation, particularly in the context of data analytics relevant to computational linguistics, corpus linguistics, and natural language processing.	[SU2] 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 technologies (ICT) when gathering and processing information for research and professional purposes within English philology, particularly in the context of data analytics relevant to computational linguistics, corpus linguistics, and natural language processing.	[SU2] presentation/project/paper/report
	[FAMU2_W14] Knows and understands selected issues connected to various disciplines constituting auxiliary and related sciences of English Philology, necessary as a context for research in the field of English linguistics and English-language literatures or in the professional activity of the English philologist.	Understands the fundamental dilemmas of contemporary civilisation in relation to issues of intercultural linguistic and literary communication, as well as the role of modern technologies in the work of the English philologist, particularly in the context of data analytics relevant to computational linguistics, corpus linguistics, and natural language processing.	[SW2] 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.	Understands selected issues within complementary and related disciplines of English studies, as well as aspects of the culture and civilisation of English-speaking countries, which provide essential context for research in English linguistics and anglophone literatures or for the professional practice of an English philologist, particularly in the context of data analytics relevant to computational linguistics, corpus linguistics, and natural language processing.	[SW2] presentation/project/paper/report
	[FAMU2_U05] Can use and present orally and in writing knowledge in the field of auxiliary and related sciences, which is a context for English linguistic and literary studies and supports the professional work of an English philologist.	Is able to employ and present, both orally and in writing, knowledge from complementary and related fields that offer contextual grounding for linguistic and literary research in English studies and support the professional work of the English philologist, particularly in the context of data analytics relevant to computational linguistics, corpus linguistics, and natural language processing.	[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 prepared to acknowledge the importance of knowledge and skills acquired within English studies in addressing cognitive and practical problems, and to seek guidance from an academic supervisor or workplace mentor when difficulties arise in resolving issues independently, particularly in the context of data analytics relevant to computational linguistics, corpus linguistics, and natural language processing.	[SK2] presentation/project/paper/report

Subject contents	Introduction to data analytics in linguistic research The scope of data analytics in contemporary linguistic and philological studies; the role of empirical data in research on the English language; relationships between qualitative and quantitative approaches. Sources and types of linguistic data Varieties of data used in linguistic and cultural research, including literary texts, online sources, user-generated content, dialogue datasets, corpora, and specialised collections; principles of ethically and legally responsible data acquisition. Data quality and representativeness The concept of representative data; issues of bias, balance, and sampling; criteria for assessing the suitability of data for linguistic or cultural research. Fundamentals of data preparation for analysis Data cleaning, selection, and evaluation; segmentation and tokenisation; basic text normalisation; challenges associated with different types of sources (literary, media-based, conversational, online). Techniques for exploring and analysing textual data Introduction to quantitative methods (frequency, distributions, co-occurrence patterns); basics of qualitative, data-informed analysis; strategies for interpreting quantitative findings in linguistic and cultural studies. Foundations of computational and corpus-based approaches in data analytics The role of annotation, linguistic categories, and metadata; the use of automated tools in data processing; critical evaluation of outputs generated by computational models. Practical exploration of textual data Working with basic tools for data acquisition and analysis; examples of analyses performed on small and medium-sized datasets; interpreting results. Visualising and presenting analytical results Preparing basic data visualisations; principles of effective presentation; formulating linguistically and culturally grounded conclusions based on textual data and metadata.		
Prerequisites and co-requisites			
Assessment methods and criteria	Subject passing criteria	Passing threshold	Percentage of the final grade
	student project	51.0%	100.0%
Recommended reading	Basic literature	1. Gutman, Alex J., & Goldmeier, Jordan (2021). <i>Becoming a Data Head: How to Think, Speak, and Understand Data Science, Statistics, and Machine Learning</i>. Wiley. 2. Claus O. Wilke (2019). <i>Fundamentals of Data Visualization: A Primer on Making Informative and Compelling Figures</i>. O'Reilly Media. 3. Lewandowska-Tomaszczyk, Barbara (2005). <i>Podstawy językoznawstwa korpusowego</i>. Wydawnictwo Uniwersytetu Łódzkiego.	
	Supplementary literature	Adam Pawłowski (red.) (2023). <i>Od Gutenberga do Zuckerberga. Wstęp do humanistyki cyfrowej</i>. Kraków: Universitas. Downey, A. B. (2025). ThinkStats:ExploratoryDataAnalysis.	

	eResources addresses	
Example issues/ example questions/ tasks being completed	1. What does it mean for linguistic data to be representative? Provide examples of factors that influence representativeness. 2. Explain the difference between raw data and processed data. 3. Identify typical sources of textual data and evaluate their usefulness for linguistic research. 4. Discuss example problems related to bias in linguistic data. 5. Present the stages involved in preparing textual data for analysis. 6. Interpret a simple frequency list or a set of word co-occurrence data. 7. Discuss the differences between qualitative and quantitative methods and explain how they can complement each other. 8. Conduct a short analysis of a selected dataset (e.g., data from Reddit) - clean the data, describe your assumptions, and present your conclusions. 9. Evaluate the challenges and limitations of automated tools used for text analysis.	
Work placement	Not applicable	

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