COURSE OUTLINE

DATA SCIENCE FOR HUMANITIES: DATA EXTRACTION, CURATION AND ANALYSIS

1. GENERAL

SCHOOL	CLASSICS AND HUMANITIES		
DEPARTMENT/UPS	HUMANITIES / DIGITAL APPLICATIONS IN ARTS AND CULTURE		
LEVEL OF STUDIES	UNDERGRADUATE – LEVEL 6		
COURSE CODE	**XXXXX**	**SEMESTER**	3RD
COURSE TITLE	DATA SCIENCE FOR HUMANITIES: DATA EXTRACTION, CURATION AND ANALYSIS		

TEACHING ACTIVITIES *If the ECTS Credits are distributed in distinct parts of the course e.g. lectures, labs etc. If the ECTS Credits are awarded to the whole course, then please indicate the teaching hours per week and the corresponding ECTS Credits.*	**TEACHING HOURS PER WEEK**	**ECTS CREDITS**
	3	6
Please, add lines if necessary. Teaching methods and organization of the course are described in section 4.		

COURSE TYPE *Background, General Knowledge, Scientific Area, Skill Development*	BACKGROUND
PREREQUISITES:	NO
TEACHING & EXAMINATION LANGUAGE:	GREEK
COURSE OFFERED TO ERASMUS STUDENTS:	YES
COURSE URL:	https://eclass.duth.gr/courses/XXXXXX/

2. LEARNING OUTCOMES

Learning Outcomes

Please describe the learning outcomes of the course: Knowledge, skills and abilities acquired after the successful completion of the course.

Upon successful completion of the course, participants will be able to:

- Recognize and extract data from public databases and websites.
- Use Python libraries for analyzing and processing image, text, and digitized document data.
- Understand the nature and functioning of data collections in the humanities.
- Apply tools for web scraping, API usage, and database management.
- Organize and document humanities data (data curation).
- Apply preprocessing techniques for cleaning and formatting text and image data.
- Use Python libraries for preparing data for analysis or machine learning models.
- Address data issues such as missing data, outliers, and inappropriate formats.
- Analyze text and image data and produce quantitative and qualitative analyses.
- Engage in discussions on the ethical aspects of data analysis in the humanities.

General Skills

Name the desirable general skills upon successful completion of the module

Search, analysis and synthesis of data and information, ICT Use
Adaptation to new situations
Decision making
Autonomous work
Teamwork
Working in an international environment
Working in an interdisciplinary environment
Production of new research ideas
Project design and management
Equity and Inclusion
Respect for the natural environment
Sustainability
Demonstration of social, professional and moral responsibility and sensitivity to gender issues
Critical thinking
Promoting free, creative and inductive reasoning

- Search, analysis and synthesis of data and information, ICT Use
- Decision making
- Autonomous work

- Working in an international environment
- Working in an interdisciplinary environment
- Project design and management
- Production of new research ideas

3. COURSE CONTENT

1. **Introduction to data science and its application in the humanities.**
2. **Data and databases in the humanities:** Structured and unstructured data in the humanities, sources of data in the humanities.
3. **Data retrieval and management from databases** (e.g., digitized libraries, museums).
4. **Web scraping techniques for extracting data from websites** (e.g., archives, digital museums). Use of APIs to obtain data from online platforms. Python tools (e.g., BeautifulSoup, Scrapy, Requests, Tweepy).
5. **Data curation and preprocessing:**
    - Curation and organization of data for analysis. Ensuring data quality: cleaning, transforming, and formatting data.
    - Preprocessing: text and image preprocessing techniques.
6. **Data curation and preprocessing:** Introduction to pandas and practical applications.
7. **Text analysis for the humanities:** Techniques for extracting and analyzing data from historical and literary archives. Introduction to topic modeling and sentiment analysis in texts.
8. **Data analysis for literary research:** Data sources for literary research (digitized texts, ancient texts, digital archives). Application of natural language processing (NLP) to literary text analysis (nltk, spaCy, gensim).
9. **Image analysis and digitized artifacts in the humanities fields** (art, archaeology, historical documents). Application of image processing methods to archaeological artifacts and artworks. Analysis of image data collections.
10. **Introduction to computer vision:** Machine learning and cultural heritage images for tasks like pattern recognition, classification, and clustering of images with Python (e.g., TensorFlow, Keras, OpenCV).
11. **Data visualization in humanities datasets.**
12. **Ethical issues and challenges in data analysis in the humanities.**
13. **Case studies.**

4. LEARNING & TEACHING METHODS - EVALUATION

TEACHING METHOD *Face to face, Distance learning, etc.*	• Face to face • Workshops • Hands-on learning • Team work
USE OF INFORMATION & COMMUNICATIONS TECHNOLOGY (ICT) *Use of ICT in Teaching, in Laboratory Education, in Communication with students*	Use of ICT in teaching and communication with students - PPT presentations - Use of digital tools and platforms - Teaching materials, announcements and communication through the eClass platform - Study by students of supporting material relevant to the course content - Communication with students via email
TEACHING ORGANIZATION *The ways and methods of teaching are described in detail.* *Lectures, Seminars, Laboratory Exercise, Field Exercise, Bibliographic research & analysis, Tutoring, Internship (Placement), Clinical Exercise, Art Workshop, Interactive learning, Study visits, Study / creation, project, creation, project. Etc.* *The supervised and unsupervised workload per*	

Activity	*Workload/semester*
Lectures	26
Workshop	13
End of semester assignment	37
Weekly projects/tests	46
Independent study	55
Final exam	3
Total	180

activity is indicated here, so that total workload per semester complies to ECTS standards.	
STUDENT EVALUATION *Description of the evaluation process* *Assessment Language, Assessment Methods, Formative or Concluding, Multiple Choice Test, Short Answer Questions, Essay Development Questions, Problem Solving, Written Assignment, Essay / Report, Oral Exam, Presentation in audience, Laboratory Report, Clinical examination of a patient, Artistic interpretation, Other/Others* *Please indicate all relevant information about the course assessment and how students are informed*	Weekly projects: 40% Assignment (compulsory): 30% Final exam: 30%

5. SUGGESTED BIBLIOGRAPHY

Jurafsky, D. and James H. Martin. 2023. Speech and Language Processing, , Pearson Education, 3rd edition, 2023, ISBN-13: 978-0135041963.

Lesk Michael, Understanding Digital Libraries, Second Edition, Elsevier.

McGillivray, Barbara et al. 2020. The challenges and prospects of the intersection of humanities and data science: A White Paper from The Alan Turing Institute. Figshare. dx.doi.org/10.6084/m9.figshare.12732164

Schiuma Giovanni, and Daniela Carlucci. 2018. *Big Data in the Arts and Humanities: Theory and Practice*. Boca Raton: Taylor & Francis.

Schneider Gerold. 2024. *Text Analytics for Corpus Linguistics and Digital Humanities,* Bloomsbury.

Shalin Hai-Jew (ed.). 2017. *Data Analytics in Digital Humanities*. Springer Cham. https://doi.org/10.1007/978-3-319-54499-1

ANNEX OF THE COURSE OUTLINE

Alternative ways of examining a course in emergency situations

Teacher (full name):	xxxx
Contact details:	xxxx
Supervisors: (1)	YES
Evaluation methods: (2)	Weekly projects: 40% Assignment (compulsory): 30% Final exam: 30%
Implementation Instructions: (3)	Written assessments and the final examination will be conducted through eClass on a date and time that will be announced, along with the duration and content, within a reasonable period before they take place. The assignment will be submitted via eClass on a specified date.

(1) Please write YES or NO

(2) Note down the evaluation methods used by the teacher, e.g.

- *written assignment* or/and exercises
- written or oral examination with distance learning methods, provided that the integrity and reliability of the examination are ensured.

(3) In the **Implementation Instructions** section, the teacher notes down clear instructions to the students:

a) in case of **written assignment and / or exercises:** the deadline (e.g. the last week of the semester), the means of submission, the grading system, the grade percentage of the assignment in the final grade and any other necessary information.
b) in case of **oral examination with distance learning methods:** the instructions for conducting the examination (e.g. in groups of X people), the way of administration of the questions to be answered, the distance learning platforms to be used, the technical means for the implementation of the examination (microphone, camera, word processor, internet connection, communication platform), the hyperlinks for the examination, the duration of the exam, the grading system, the percentage of the oral exam in the final grade, the ways in which the inviolability and reliability of the exam are ensured and any other necessary information.
c) in case of **written examination with distance learning methods**: the way of administration of the questions to be answered, the way of submitting the answers, the duration of the exam, the grading system, the percentage of the written exam of the exam in the final grade, the ways in which the integrity and reliability of the exam are ensured and any other necessary information.
There should be an attached list with the Student Registration Numbers only of students eligible to participate in the examination.