

COURSE OUTLINE

(1) GENERAL

SCHOOL	Science and Technology		
ACADEMIC UNIT	Science and Technology		
PROGRAMME OF STUDIES	MSc in Data Science		
LEVEL OF STUDIES	Postgraduate		
COURSE CODE	DSE03	SEMESTER	2
COURSE TITLE	Natural Language Processing and Text Mining		
COURSE TYPE <i>Elective, compulsory</i>	Elective		
INSTRUCTOR(S)	Theory: Assoc. Prof. A. Papadopoulos, Dr Christos Berberidis Lab: Dr. Christos Berberidis		
INDEPENDENT TEACHING ACTIVITIES <i>if credits are awarded for separate components of the course, e.g. lectures, laboratory exercises, etc. If the credits are awarded for the whole of the course, give the weekly teaching hours and the total credits</i>	WEEKLY TEACHING HOURS	CREDITS	
	3	6	
<i>Add rows if necessary. The organisation of teaching and the teaching methods used are described in detail at (d).</i>			
TEACHING ACTIVITIES BREAKDOWN	WEEKLY HOURS		
Theory	2,6		
Lab	0,4		
<i>Add rows if necessary. The organisation of teaching and the teaching methods used are described in detail at (d).</i>			
COURSE TYPE <i>general background, special background, specialised general knowledge, skills development</i>	<i>specialised general knowledge</i>		
PREREQUISITE COURSES:	-		
LANGUAGE OF INSTRUCTION and EXAMINATIONS:	English		
IS THE COURSE OFFERED TO ERASMUS STUDENTS	Yes		
COURSE WEBSITE (URL)	https://elearn-ucips.ihu.gr/		

(2) LEARNING OUTCOMES

<p>Learning outcomes</p> <p><i>The course learning outcomes, specific knowledge, skills and competences of an appropriate level, which the students will acquire with the successful completion of the course are described.</i></p> <p><i>Consult Appendix A</i></p> <ul style="list-style-type: none"> • <i>Description of the level of learning outcomes for each qualifications cycle, according to the Qualifications Framework of the European Higher Education Area</i> • <i>Descriptors for Levels 6, 7 & 8 of the European Qualifications Framework for Lifelong Learning and Appendix B</i> • <i>Guidelines for writing Learning Outcomes</i>
<p>On completing the course, the student will be able to:</p> <ul style="list-style-type: none"> • Understand how natural language processing (NLP) draws upon other areas of computer science and data analysis.

<ul style="list-style-type: none"> ● Design and build computer systems and software for various tasks of NLP. ● Understand and implement the most important algorithms and techniques in NLP and text mining. ● Formulate models and construct computational solutions to text and speech-based processing problems. 																		
<p>General Competences</p> <p><i>Taking into consideration the general competences that the degree-holder must acquire (as these appear in the Diploma Supplement and appear below), at which of the following does the course aim?</i></p> <table border="0"> <tr> <td><i>Search for, analysis and synthesis of data and information, with the use of the necessary technology</i></td> <td><i>Project planning and management</i></td> </tr> <tr> <td><i>Adapting to new situations</i></td> <td><i>Respect for difference and multiculturalism</i></td> </tr> <tr> <td><i>Decision-making</i></td> <td><i>Respect for the natural environment</i></td> </tr> <tr> <td><i>Working independently</i></td> <td><i>Showing social, professional and ethical responsibility and sensitivity to gender issues</i></td> </tr> <tr> <td><i>Team work</i></td> <td><i>Criticism and self-criticism</i></td> </tr> <tr> <td><i>Working in an international environment</i></td> <td><i>Production of free, creative and inductive thinking</i></td> </tr> <tr> <td><i>Working in an interdisciplinary environment</i></td> <td>.....</td> </tr> <tr> <td><i>Production of new research ideas</i></td> <td><i>Others...</i></td> </tr> <tr> <td></td> <td>.....</td> </tr> </table>	<i>Search for, analysis and synthesis of data and information, with the use of the necessary technology</i>	<i>Project planning and management</i>	<i>Adapting to new situations</i>	<i>Respect for difference and multiculturalism</i>	<i>Decision-making</i>	<i>Respect for the natural environment</i>	<i>Working independently</i>	<i>Showing social, professional and ethical responsibility and sensitivity to gender issues</i>	<i>Team work</i>	<i>Criticism and self-criticism</i>	<i>Working in an international environment</i>	<i>Production of free, creative and inductive thinking</i>	<i>Working in an interdisciplinary environment</i>	<i>Production of new research ideas</i>	<i>Others...</i>	
<i>Search for, analysis and synthesis of data and information, with the use of the necessary technology</i>	<i>Project planning and management</i>																	
<i>Adapting to new situations</i>	<i>Respect for difference and multiculturalism</i>																	
<i>Decision-making</i>	<i>Respect for the natural environment</i>																	
<i>Working independently</i>	<i>Showing social, professional and ethical responsibility and sensitivity to gender issues</i>																	
<i>Team work</i>	<i>Criticism and self-criticism</i>																	
<i>Working in an international environment</i>	<i>Production of free, creative and inductive thinking</i>																	
<i>Working in an interdisciplinary environment</i>																	
<i>Production of new research ideas</i>	<i>Others...</i>																	
																	
<ul style="list-style-type: none"> ● Search for, analysis and synthesis of data and information, with the use of the necessary technology ● Teamwork ● Production of free, creative, and inductive thinking 																		

(3) SYLLABUS

<p>The course introduces fundamental concepts and tools of Machine Learning. The student is exposed to the necessary mathematical/algorithmic background and coding with the Python programming language. The topics covered include:</p> <ul style="list-style-type: none"> ● Introduction to natural language processing and its challenges. ● Syntax and language modeling ● Information extraction from text ● Text classification and clustering. ● Sentiment analysis. ● Language models
--

(4) TEACHING and LEARNING METHODS - EVALUATION

DELIVERY <i>Face-to-face, Distance learning, etc.</i>	Hybrid: Face to face and synchronous distance learning	
USE OF INFORMATION AND COMMUNICATIONS TECHNOLOGY <i>Use of ICT in teaching, laboratory education, communication with students</i>	<p>Use of ICT in Teaching</p> <p>During the educational process, various machine learning and programming tools are used, along with the material available at the e-learning platform.</p> <p>The hybrid teaching method involves synchronous learning with the support of the videoconferencing tool Zoom.</p> <p>Students are taught a variety of tools related to the course content and material.</p> <p>Use of ICT in Communication with students</p> <ul style="list-style-type: none"> ● The course material (slides, scientific articles, exercises, etc.) is posted on the course page at the e-learn platform (Moodle). ● Use of Moodle Forums announcements. ● Live video meetings via Zoom/Teams. ● Contact via email. 	
TEACHING METHODS <i>The manner and methods of teaching are described in detail.</i>	<i>Activity</i>	<i>Semester workload</i>
	Lectures	26 hrs.

<p>Lectures, recitation, seminars, laboratory practice, fieldwork, study and analysis of bibliography, tutorials, placements, clinical practice, art workshop, interactive teaching, educational visits, project, essay writing, artistic creativity, etc.</p> <p>The student's study hours for each learning activity are given as well as the hours of non-directed study according to the principles of the ECTS</p>	Lab	4 hrs.
	Project	40 hrs.
	Exams	3 hrs.
	Non-Directed Study	77 hrs.
	Course total	150 hrs.
COURSE MATERIAL ARRANGEMENT	Theory/Recitation	
	Introduction to NLP	3 hrs.
	Information Extraction from Text	6 hrs.
	Word embeddings	4 hrs.
	Text classification and clustering	2 hrs.
	Text mining	2 hrs.
	Data streams	3 hrs.
	Dimensionality reduction	3 hrs.
	Topic Analysis	3 hrs.
	Lab	
	Language Models	4 hrs.
STUDENT PERFORMANCE EVALUATION	Language of Evaluation: English	
	<p><i>Description of the evaluation procedure</i></p> <p>Language of evaluation, summative or conclusive, multiple choice questionnaires, short-answer questions, open-ended questions, problem solving, written work, essay/report, oral examination, public presentation, laboratory work, clinical examination of patient, art interpretation, other</p> <p>Specifically-defined evaluation criteria are given, and if and where they are accessible to students</p>	
STUDENT OBLIGATIONS	<p>Evaluation Procedure:</p> <ul style="list-style-type: none"> ● Written Exams (70%). Methods of evaluation: <ul style="list-style-type: none"> ○ Open-ended questions ○ Problem solving ○ Multiple choice questions (on lab material) ● Group project (30%): <ul style="list-style-type: none"> ○ Training and evaluation of various ML models ○ The students should achieve a passing grade to participate in the written exams. <p>The evaluation procedure is announced to the students during the first lecture and is also accessible at the e-learn platform throughout the entire semester.</p>	
	<ul style="list-style-type: none"> ● Compulsory attendance of lectures ● Compulsory attendance of labs ● Compulsory participation in the exams ● Compulsory delivery of project 	

(5) ATTACHED BIBLIOGRAPHY

1. Manning C., Schütze H. (1999), Foundations of statistical natural language processing, MIT Press.

2. Jurafsky D., Martin J. (2008), *Speech and language processing*, Prentice Hall, 2nd edition.
3. Bird S., Klein E., Loper E. (2009), *Natural language processing with Python: analyzing text with the Natural Language Toolkit*, O'Reilly.