**COURSE OUTLINE**

1. **GENERAL**

|  |  |
| --- | --- |
| **SCHOOL** | Humanities |
| **ACADEMIC UNIT** | Mediterranean Studies |
| **LEVEL OF STUDIES** | undergraduate |
| **COURSE CODE** | ΓΥE-42 | **SEMESTER** | 8 |
| **COURSE TITLE** | Topics in Natural Language Processing |
| **INDEPENDENT TEACHING ACTIVITIES** *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* | **WEEKLY TEACHING HOURS** | **CREDITS** |
|  |  |  |
|  |  |  |
|  |  |  |
| *Add rows if necessary. The organisation of teaching and the teaching methods used are described in detail at (d).* | 3 | 5 |
| **COURSE TYPE***general background, special background, specialised general knowledge, skills development* | special background, skills development |
| **PREREQUISITE COURSES:** | - |
| **LANGUAGE OF INSTRUCTION and EXAMINATIONS:** | Greek. In case of ERASMUS students: English |
| **IS THE COURSE OFFERED TO ERASMUS STUDENTS** | Yes |
| **COURSE WEBSITE (URL)** | <http://dms.aegean.gr/en/undergraduate-studies/program-studies-2017-18/> |

1. **LEARNING OUTCOMES**

|  |
| --- |
| **Learning outcomes** |
| *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.**Consult Appendix A* * *Description of the level of learning outcomes for each qualifications cycle, according to the Qualifications Framework of the European Higher Education Area*
* *Descriptors for Levels 6, 7 & 8 of the European Qualifications Framework for Lifelong Learning and Appendix B*
* *Guidelines for writing Learning Outcomes*
 |
| After the successful completion of the course, students should be able to:* understand the suitability of Linux for Natural Language Processing
* know about the basics of the Linux environment and why it is used for corpus processing by humanities and social scientists
* understand programming and how to use basic Linux commands
* implement/program for the extraction of word and collocation frequency lists from corpora
* know the use of the *grep/egrep/fgrep* commands for Regular Expressions use
* prepare a corpus for processing
* apply research questions on a corpus using *grep/egrep/fgrep*
* check output and make appropriate corrections
* deal with issues and problems in a formal, algorithmic way
 |
| **General Competences**  |
| *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?* |
| *Search for, analysis and synthesis of data and information, with the use of the necessary technology* *Adapting to new situations* *Decision-making* *Working independently* *Team work**Working in an international environment* *Working in an interdisciplinary environment* *Production of new research ideas*  | *Project planning and management* *Respect for difference and multiculturalism* *Respect for the natural environment* *Showing social, professional and ethical responsibility and sensitivity to gender issues* *Criticism and self-criticism* *Production of free, creative and inductive thinking**……**Others…**…….* |
| * Search for, analysis and synthesis of data and information, with the use of the necessary technology
* Adapting to new situations
* Decision-making
* Working independently
* Team work
* Working in an interdisciplinary environment
* Production of new research ideas
 |

1. **SYLLABUS**

|  |
| --- |
| * Corpus processing for Humanities and Social Sciences purposes
* Off-the-shelf tools
* What programming for natural language processing is
* Why Linux for Humanities and Social Sciences
* The Linux environment
* The terminal
* Basic Linux commands
* Programming for the extraction of collocation lists
* grep/fgrep/egrep for regular expressions
* Forming research questions for Humanities and Social Sciences purposes
* Applying grep/fgrep/egrep to language research questions
* Checking constructed tools for errors
* Saving results for future reference
* Assessing results
 |

1. **TEACHING and LEARNING METHODS - EVALUATION**

|  |  |
| --- | --- |
| **DELIVERY***Face-to-face, Distance learning, etc.* | Face-to-face |
| **USE OF INFORMATION AND COMMUNICATIONS TECHNOLOGY** *Use of ICT in teaching, laboratory education, communication with students* | Use of ICT in communication with students |
| **TEACHING METHODS***The manner and methods of teaching are described in detail.**Lectures, 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.**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* |

|  |  |
| --- | --- |
| ***Activity*** | ***Semester workload*** |
| Lectures | 39 hours (1.56 ECTS) |
| Personal study | 33 hours (1.32 ECTS) |
| Assignment | 50 hours (2 ECTS) |
| End of semester exam | 3 hours (0.12 ECTS) |
| Course total  | *125 hours (5 ECTS)* |

 |
| **STUDENT PERFORMANCE EVALUATION***Description of the evaluation procedure**Language of evaluation, methods 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**Specifically-defined evaluation criteria are given, and if and where they are accessible to students.* | *Language of evaluation:* Greek. In case of ERASMUS students: English*Method of evaluation:* 1. end of semester exam using ICT (50%)
2. independent/team assignment (50%)
 |

1. **ATTACHED BIBLIOGRAPHY**

|  |
| --- |
| *Suggested bibliography:*Γούτσος, Δ., Σηφιανού, M. & Θεοδωρακοπούλου, A. (2005) Η ελληνική ως ξένη γλώσσα. Αθήνα: Πατάκης. Γούτσος, Διονύσης (2015) Εισαγωγή στη Γλωσσολογία Σωμάτων Κειμένων.  Ελληνικά Ακαδημαϊκά Ηλεκτρονικά Συγγράμματα και Βοηθήματα. Αθήνα: Εθνικό Μετσόβιο Πολυτεχνείο, Σύνδεσμος Ελληνικών Ακαδημαϊκών Βιβλιοθηκών, Αποθετήριο "Κάλλιπος", www.kallipos.gr.Μαρκόπουλος, Γ. Α. (2006) Ζητήματα Υπολογιστικής Γλωσσολογίας. Γ. Α. Μαρκόπουλος.Μικρός, Γ. Κ. (2009) Η Ποσοτική Ανάλυση της Κοινωνιογλωσσολογικής Ποικιλίας – Θεωρητικές και Μεθοδολογικές Προσεγγίσεις. Αθήνα: Μεταίχμιο.Μικρός, Γ. (2016) Υπολογιστική Υφολογία.  Ελληνικά Ακαδημαϊκά Ηλεκτρονικά Συγγράμματα και Βοηθήματα - Αποθετήριο "Κάλλιπος"Σαριδάκης, Ι. (2010) Σώματα Κειμένων και μετάφραση – Θεωρία και Εφαρμογές. Αθήνα: Παπαζήση.Παναρέτου, Ε. (2009) Νομικός Λόγος – Γλώσσα & Δομή των Νόμων. Αθήνα: Παπαζήση.Τάντος, Α. (2016) Υπολογιστική Γλωσσολογία.  Ελληνικά Ακαδημαϊκά Ηλεκτρονικά Συγγράμματα και Βοηθήματα - Αποθετήριο "Κάλλιπος".Φραντζή, Κ. (2012) Εισαγωγή στην Επεξεργασία Σωμάτων Κειμένων. Αθήνα: Ίων.Baker, P. (2009) Contemporary Corpus Linguistics (Contemporary Studies in Linguistics). Continuum International Publishing Group Ltd.Biber, D., Conrad, S. & Randi Reppen, R. (1998) Corpus linguistics: investigating language structure and use. Cambridge University Press.Cheng, Winnie (2011) Exploring Corpus Linguistics: Language in Action (Routledge Introductions to Applied Linguistics). Routledge.Dougherty, D. (1990) sed & awk. Sebastopol, CA: O’Reilly & Associates, Ιnc.Friedl, J. E.F. (2006) Mastering regular Expressions. Sebastopol, CA: O’Reilly Media, Inc.Hyland, Ken (2013) Corpus Applications in Applied Linguistics. Bloomsbury Academic.Kennedy, G. (1998) An Introduction to Corpus Linguistics. Longman. McEnery, A., Xiao, R. & Tono, Y. 2006. Corpus-based Language Studies: an advanced resource book. Routledge, 2006.McEnery, A. (2011) Corpus Linguistics: Method, Theory and Practice (Cambridge Textbooks in Linguistics). Cambridge University PressOakes, M. P. (1998) Statistics for corpus linguistics. Edinburgh University Press.O’Keeffe, A. (2012) The Routledge handbook of corpus linguistics (Routledge Handbooks in Applied Linguistics). Routledge.Stubbs M. (1996) Text and Corpus Analysis: Computer Assisted Studies of Language and Institutions (Language in Society). John Wiley & Sons.Watt, A. (2005) Beginning Regular Expressions. Indianapolis, IN: Wiley Publishing, Inc.Welsh, M., Dalheimer, M. K. & L. Kaufman (2011) Ο Οδηγός του LINUX. Αθήνα: Κλειδάριθμος.Wolfgang, T, (2007) Corpus Linguistics: A Short Introduction. Continuum.*Related academic journals:** International Journal of Corpus Linguistics (IJCL) https://benjamins.com/#catalog/journals/ijcl
* Corpora http://www.euppublishing.com/loi/cor
* Corpus Linguistics and Linguistic Theory (CLLT)
* Corpus Pragmatics [http://www.springer.com/education+%26+language/](http://www.springer.com/education%2B%26%2Blanguage/) linguistics/journal/41701
* Language Resources and Evaluation https://link.springer.com/journal/10579
 |