

## Module/Course Title: Topics in Natural Language Processing

- **Code number:**
- **Level of Module/Course (under-/postgraduate):**
- **Type of Module/Course:**
- **Year of Study**
- **Semester**
- **Number of ects allocated:**
- **Number of teaching units:**
- **Name of lecturer / lecturers :**
- **Content outline:**

Students learn how to apply Regular Expressions to Corpora in order to express and study research questions on various language and linguistic issues. Research questions may be on morphological, syntactic, semantic, discourse analysis issues and more. Students learn how to apply regular expressions in both raw and part-of-speech tagged corpora. They learn how to build their own tools rather than use ready off-shelf software. They structure and save their products for future use. Lectures take place in a computing lab under Linux environment, which is a suitable environment for the introduction of students, even those without a computational background, to language processing.

- **Learning outcomes:**

After the successful completeness of the course, students should:

- Know what Linux is and why it is suitable for Natural Language Processing.
- Know the basics of the Linux environment.
- Know what a terminal is and how to use basic Linux commands.
- Be able to program in the Linux environment in order to extract word and collocation frequency lists from corpora.
- Know the use of the *grep/egrep/fgrep* commands for Regular Expressions use.
- Be able to prepare a corpus for processing.
- Be able to express research questions on a corpus using regular expressions to *grep/egrep/fgrep*.
- Be able to check the output and make the appropriate corrections.

The course “Natural Language Processing Issues” gives students of a theoretical background the ability to deal with issues and problems in a formal, algorithmic way, an ability acquired with the study of mathematics/computing subjects.

Moreover, the students are introduced to the first steps of computer programming.

- **Prerequisites:** students should have already been introduced to Regular Expressions (taught at the “Computational Linguistics” course)
- **Recommended Reading:**

**a) Basic Textbooks:**

Μαρκόπουλος, Γ. Α. 2006. Ζητήματα Υπολογιστικής Γλωσσολογίας. Γ. Α. Μαρκόπουλος.

Μικρός, Γ. 2016. Υπολογιστική Υφολογία. Ελληνικά Ακαδημαϊκά Ηλεκτρονικά Συγγράμματα και Βοηθήματα - Αποθετήριο "Κάλλιπος"

Σαριδάκης, Ι. 2010. Σώματα Κειμένων και μετάφραση - Θεωρία και Εφαρμογές. Αθήνα : Παπαζήσης.

Τάντος, Αλ. 2016. Υπολογιστική Γλωσσολογία. Ελληνικά Ακαδημαϊκά Ηλεκτρονικά Συγγράμματα και Βοηθήματα - Αποθετήριο "Κάλλιπος".

Φραντζή, Κ. 2012. Εισαγωγή στην Επεξεργασία Σωμάτων Κειμένων. Αθήνα: Ίων.

**b) Additional References:**

Γούτσος, Δ., Μ. Σηφριανού & Α. Θεοδωρακοπούλου. Η ελληνική ως ξένη γλώσσα. Αθήνα: Πατάκης, 2005.

Παναρέτου, Ε. 2009. Νομικός Λόγος - Γλώσσα & Δομή των Νόμων. Παπαζήση.

Dougherty, D. 1990. sed & awk. Sebastopol, CA: O'Reilly & Associates, Inc.

Friedl, J. E.F. 2006. Mastering regular Expressions. Sebastopol, CA: O'Reilly Media, Inc.

Oakes, M. P. 1998. Statistics for corpus linguistics. Edinburgh University Press.

Watt, A. 2005. Beginning Regular Expressions. Indianapolis, IN: Wiley Publishing, Inc.

Welsh, M., Dalheimer, M. K. & L. Kaufman. Ο Οδηγός του LINUX. Αθήνα: Κλειδάριθμος.

Young, S. & G. Bloothoof 1997. Corpus-based methods in language and speech processing, editors:. Kluwer Academics.

- **Learning Activities and Teaching Methods:** interactive lectures
- **Assessment/Grading Methods:**
  1. end-of-semester exams
  2. assignment
- **Language of Instruction:** Greek/English
- **Mode of delivery (face-to-face, distance learning):** face-to-face