

## Τίτλος Μαθήματος: Θέματα Επεξεργασίας Φυσικής Γλώσσας

- **Κωδικός μαθήματος:** ΓΥΕ-42
- **Τύπος μαθήματος:** Κατ' επιλογήν υποχρεωτικό
- **Επίπεδο μαθήματος (Προπτυχιακό/Μεταπτυχιακό):** Προπτυχιακό
- **Έτος σπουδών:** 4
- **Εξάμηνο σπουδών:** 8
- **Αριθμός πιστωτικών μονάδων (ECTS):** 5
- **Αριθμός διδακτικών μονάδων:** 3
- **Όνομα διδάσκοντος/ διδασκόντων :** Αικατερίνη Φραντζή
- **Περιεχόμενο μαθήματος:**

Στο μάθημα οι φοιτητές/ήτριες μαθαίνουν να εκφράζουν ερευνητικά ερωτήματα που σχετίζονται με τη γλώσσα σε σώματα κειμένων και με τη χρήση κανονικών εκφράσεων τις οποίες διδάχτηκαν στο μάθημα της Υπολογιστικής Γλωσσολογίας. Τίθενται μορφολογικά, συντακτικά και σημασιολογικά ερωτήματα σε επισημειωμένα ή μη σώματα κειμένων. Το μάθημα γίνεται εξολοκλήρου σε υπολογιστικό περιβάλλον στο εργαστήριο και με χρήση του λειτουργικού συστήματος Linux. Το Linux δίνει τη δυνατότητα σε φοιτητές και φοιτήτριες, οι οποίοι δεν προέρχονται απαραίτητα από τις θετικές επιστήμες, να εισαχθούν με ομαλό τρόπο στην υπολογιστική επεξεργασία της φυσικής γλώσσας. Οι φοιτητές/ήτριες μαθαίνουν να συντάσσουν μόνοι τους τα ερωτήματά τους, χωρίς τη χρήση έτοιμων εργαλείων, ώστε να έχουν ελευθερία στη μελέτη και έρευνα τους, η οποία δεν περιορίζεται από έτοιμες παροχές. Δομούν και αποθηκεύουν τα αποτελέσματά τους για τη μελλοντική ποιοτική επεξεργασία τους.

- **Επιδιωκόμενα μαθησιακά αποτελέσματα:**

Με την επιτυχή ολοκλήρωση αυτού του μαθήματος ο φοιτητής/η φοιτήτρια θα πρέπει να:

- Γνωρίζει τι είναι το λειτουργικό σύστημα Linux και την καταλληλότητά του για την επεξεργασία της φυσικής γλώσσας.
- Γνωρίζει το βασικό περιβάλλον του λειτουργικού συστήματος Linux.
- Γνωρίζει τι είναι το «τερματικό» και να χρησιμοποιεί βασικές εντολές του Linux.
- Μπορεί να προγραμματίσει σε περιβάλλον Linux για την Παραγωγή Λίστας Συχνοτήτων Συμφράσεων από ΣΚ.
- Γνωρίζει τη χρήση των εντολών grep/egrep/fgrep.
- Γνωρίζει τη σύνταξη των εντολών grep/egrep/fgrep για τις κανονικές εκφράσεις.
- Μπορεί να προετοιμάσει ένα σώμα κειμένων για επεξεργασία.
- Μπορεί να θέσει ερευνητικά ερωτήματα συντάσσοντας κανονικές εκφράσεις με τις εντολές grep/egrep/fgrep.

- Μπορεί να εφαρμόσει τις κανονικές εκφράσεις μέσω των grep/egrep/fgrep στο σώμα κειμένων για μελέτη βασιζόμενη σε συγκεκριμένα ερευνητικά ερωτήματα.
- Μπορεί να ελέγχει τα αποτελέσματα και να κάνει τις κατάλληλες διορθώσεις.

Με το μάθημα «Θέματα Επεξεργασίας Φυσικής Γλώσσας» ο/η φοιτητής/ήτρια ενός Θεωρητικού Πανεπιστημιακού Τμήματος αποκτά ικανότητα να διαχειρίζεται θέματα και προβλήματα με τυπικό και αλγοριθμικό τρόπο, προσόν που αποκτάται με την παρακολούθηση μαθημάτων τα οποία σχετίζονται με τις Θετικές Επιστήμες.

Επιπλέον, με το πέρας του μαθήματος, ο/η φοιτητής/ήτρια φοιτήτρια Πανεπιστημιακής Σχολής Ανθρωπιστικών Επιστημών, αποκτά γνώση των πρώτων βημάτων στον υπολογιστικό προγραμματισμό.

- **Συνιστώμενη βιβλιογραφία προς μελέτη:**

- α) **Εγχειρίδια του μαθήματος:**

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.

- β) **Συμπληρωματική βιβλιογραφία:**

Γούτσος, Δ., Μ. Σηφριανού & Α. Θεοδορακοπούλου. 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.

- **Διδακτικές και μαθησιακές μέθοδοι:**

Τα μαθήματα γίνεται με διαλέξεις κατά τις οποίες οι φοιτητές/ήτριες έχουν ενεργή συμμετοχή σε θέματα των οποίων οι τρόποι αντιμετώπισης συζητούνται κατά τη διάρκεια των μαθημάτων.

- **Μέθοδοι αξιολόγησης/ βαθμολόγησης:**

Ο τελικός βαθμός είναι αποτέλεσμα των παρακάτω:

- 1 Εξέταση στο τέλος του εξαμήνου.
- 2 Εργασία (παράδοση στο τέλος του εξαμήνου).

- **Γλώσσα διδασκαλίας:** Ελληνική/ Αγγλική

- **Τρόπος παράδοσης μαθήματος (με φυσική παρουσία, εξ αποστάσεως):** Η παράδοση των μαθημάτων γίνεται με φυσική παρουσία.