

Τίτλος Μαθήματος: Επεξεργασία και Εφαρμογές Σωμάτων Κειμένων

- Κωδικός μαθήματος:
- Τύπος μαθήματος:
- Επίπεδο μαθήματος (Προπτυχιακό/Μεταπτυχιακό):
- Έτος σπουδών:
- Εξάμηνο σπουδών:
- Αριθμός πιστωτικών μονάδων (ECTS):
- Αριθμός διδακτικών μονάδων:
- Όνομα διδάσκοντος / διδασκόντων :
- Περιεχόμενο μαθήματος:

Τα τελευταία χρόνια, τα Σώματα Κειμένων, μεγάλες συλλογές πραγματικού γραπτού ή προφορικού λόγου σε ηλεκτρονική μορφή, έχουν σημαντική συμβολή στη μελέτη τομέων της γλωσσολογίας, όπως στη μορφολογία, σύνταξη, σημασιολογία, πραγματολογία, ανάλυση λόγου, λεξικογραφία, ορολογία, κειμενογλωσσολογία, υφογλωσσολογία, κοινωνιογλωσσολογία, ιστορική και συγκριτική γλωσσολογία, εφαρμοσμένη γλωσσολογία, διαλεκτολογία, κ.ά. Λόγω της αξιοπιστίας, ακρίβειας και ταχύτητας που προσφέρει η επεξεργασία της γλώσσας με τη χρήση του υπολογιστή σε πραγματικό γλωσσικό υλικό (ειδικά μεγάλου μεγέθους του οποίου η επεξεργασία θα ήταν αδύνατη αλλιώς), οι εφαρμογές της Επεξεργασίας Σωμάτων Κειμένων απλώνονται σε επιστημονικά πεδία κι εκτός της Γλωσσολογίας, όπως στην Εκπαίδευση, Ψυχολογία, Φιλολογία, Επικοινωνία Πολιτική, Δικαστική Γλωσσολογία, στις Επιστήμες των Μέσων Μαζικής Ενημέρωσης, κ.ά. Η γλώσσα μπορεί να μελετηθεί πλέον από την πραγματική της χρήση και όχι με υποθέσεις, εκτιμήσεις και προκαταλήψεις.

Ενότητες μαθήματος:

Σώματα Κειμένων (ορισμός, είδη, επισημείωση, κανονικοποίηση, κατασκευή)

Κανονικές Εκφράσεις

Αυτόματα Πεπερασμένων Καταστάσεων

Επεξεργασία Σωμάτων Κειμένων στο Διαδίκτυο

Επεξεργασία Σωμάτων Κειμένων σε Υπολογιστικό Περιβάλλον Linux

Εφαρμογές Σωμάτων Κειμένων σε ποικίλα αντικείμενα

- **Επιδιωκόμενα μαθησιακά αποτελέσματα:**

Στόχος του μαθήματος είναι αφενός η παρουσίαση των Σωμάτων Κειμένων και η εφαρμογή τους σε γλωσσολογικά και άλλα γνωστικά αντικείμενα και αφετέρου η εισαγωγή των φοιτητριών και φοιτητών στην ανάπτυξη λογισμικού σε κατάλληλα για την επεξεργασία τους υπολογιστικά περιβάλλοντα.

Με την ολοκλήρωση του μαθήματος οι φοιτητές/ήτριες θα:

- γνωρίζουν τι είναι τα Σώματα Κειμένων (ΣΚ), τα είδη ΣΚ, τι είναι η επισημείωση ΣΚ, τι είναι και πώς γίνεται η κανονικοποίηση ΣΚ, θέματα κατασκευής ΣΚ
- γνωρίζουν τις Κανονικές εκφράσεις και την εφαρμογή τους στη Γλωσσική Τεχνολογία
- γνωρίζουν τα Αυτόματα Πεπερασμένων Καταστάσεων και την εφαρμογή τους στη Γλωσσική Τεχνολογία
- γνωρίζουν για την Επεξεργασία Σωμάτων Κειμένων στο Διαδίκτυο
- γνωρίζουν για εφαρμογές ΣΚ σε ποικίλα αντικείμενα

Με το μάθημα ο 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.

- **Προαπαιτούμενα:**
- **Συνιστώμενη βιβλιογραφία προς μελέτη:**

α) Εγχειρίδια του μαθήματος:

Γούτσος, Διονύσης και Γεωργία Φραγκάκη (2015) Εισαγωγή στη Γλωσσολογία Σωμάτων Κειμένων. Ελληνικά Ακαδημαϊκά Ηλεκτρονικά Συγγράμματα και Βοηθήματα - Αποθετήριο "Κάλλιπος".

Μικρός, Γεώργιος (2009) Ποσοτική Ανάλυση της Κοινωνιογλωσσολογικής Ποικιλίας - Θεωρητικές και Μεθοδολογικές Προσεγγίσεις. Αθήνα : Μεταίχμιο.

Φραντζή, Κατερίνα Θ. (2012) Εισαγωγή στην επεξεργασία σωμάτων κειμένων. Αθήνα: Ίων.

β) Συμπληρωματική βιβλιογραφία:

Gavrilidou, M., Karagiannis, G., Markantonatou, S. Piperidis, S. & G. Stainhaouer (2000) Proceedings of the Second International Conference on Language Resources and Evaluation, Athens, Greece.

Καπιδάκης, Σαράντος (2008) Χειρισμός πνευματικών δικαιωμάτων σε ψηφιακούς πόρους. Πρακτικά Συνεδρίου «Αρχεία, βιβλιοθήκες και δίκαιο στην κοινωνία της πληροφορίας», Αθήνα 2-3 Φεβρουαρίου 2006. Εθνική Βιβλιοθήκη της Ελλάδος, 301-307.

Κατσογιάννου, Μαριάννα & Ελένη Ευθυμίου (2004) Ελληνική Ορολογία: Έρευνα και εφαρμογές. Αθήνα: Καστανιώτης.

Kordoni, Valia, Carlos Ramisch & Aline Villavicencio (2013) Proceedings of the 9th Workshop on Multiword Expressions, NAACL HLT 2013, Atlanta, Georgia, USA, June, 2013.

Kyriacopoulou, Panagiota (2005) Analyse automatique des textes écrits: le cas du grec moderne. Θεσσαλονίκη: University Studio Press.-

Markantonatou, S., Sofianopoulos, S., Spilioti, V., Tambouratzis, G., Vasileiou, M., Yannoutsou, O. (2005) Monolingual Corpus-based MT using Chunks. In Proceedings of the Workshop on EBMT, held within the MT Summit X, Phuket, Thailand, 91-98.

Μαρκόπουλος, Γεώργιος (2006) Ζητήματα Υπολογιστικής Γλωσσολογίας: Prolog και μορφολογική ανάλυση. Αθήνα: περιοδικό Παρουσία, παράρτημα 69.

Σαριδάκης, Ιωάννης Ε. (2010) Σώματα κειμένων και μετάφραση: θεωρία και εφαρμογές, Γλώσσα και πολιτισμός. Αθήνα: Παπαζήση.

Σταματάτος, Ευστάθιος (2000) Στατιστική Αναγνώριση Είδους Κειμένου και Συγγραφέα σε Νεοελληνικά Κείμενα χωρίς Περιορισμούς, Διδακτορική Διατριβή, Τμήμα Ηλεκτρολόγων Μηχανικών και Τεχνολογίας Υπολογιστών, Πανεπιστήμιο Πατρών.
Σταύρου, Μελίτα & Μαρία Τζεβελεκου (2000) Η μηχανική μετάφραση και η ελληνική γλώσσα. Αθήνα: Καστανιώτη.

- **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:**

Εργασία (παράδοση κατά την εξεταστική περίοδο)

- **Γλώσσα διδασκαλίας:** Ελληνική

- **Τρόπος παράδοσης μαθήματος (με φυσική παρουσία, εξ αποστάσεως):** δια ζώσης/μέσω διαδικτύου (σύγχρονα)