ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΙΓΑΙΟΥ
ΣΧΟΛΗ ΑΝΘΡΩΠΙΣΤΙΚΩΝ ΕΠΙΣΤΗΜΩΝ
Τμήμα Μεσογειακών Σπουδών

ΠΡΟΓΡΑΜΜΑ
ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ
Γλωσσολογία Νοτιοανατολικής Μεσογείου

ΠΕΡΙΓΡΑΦΗ ΜΑΘΗΜΑΤΟΣ (Syllabus)

ΤΙΤΛΟΣ ΜΑΘΗΜΑΤΟΣ:	Επεξεργασία και Εφαρμογές Σωμάτων Κειμένων
ΕΞΑΜΗΝΟ:	$2^{ο}$ - Υποχρεωτικό
ΔΙΔΑΚΤΙΚΕΣ ΜΟΝΑΔΕΣ (ECTS):	6
ΥΠΕΥΘΥΝΗ ΜΑΘΗΜΑΤΟΣ:	Κατερίνα Θ. Φραντζή
ΔΙΔΑΣΚΟΝΤΕΣ/ΟΥΣΕΣ:	Κατερίνα Θ. Φραντζή, Βάλια Κορδώνη, Σαράντος Καπιδάκης, Σοφία Στάμου, Στέλλα Μαρκαντωνάτου, Αναστασία Κατρανίδου, Κωνσταντίνα Βαδάση
ΣΤΟΧΟΙ ΜΑΘΗΜΑΤΟΣ:	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.
ΠΡΟΣΔΟΚΩΜΕΝΑ ΑΠΟΤΕΛΕΣΜΑΤΑ:	Με την ολοκλήρωση του μαθήματος οι φοιτητές και φοιτήτριες 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 [Δ1] • γνωρίζουν τις Κανονικές εκφράσεις και την εφαρμογή τους στη Γλωσσική Τεχνολογία [Δ2] • γνωρίζουν για την εφαρμογή των ΣΚ στην Αυτόματη Μετάφραση [Δ3] • γνωρίζουν για τα ΣΚ ως Μεγάλα Δεδομένα [Δ4] • γνωρίζουν για εφαρμογές ΣΚ στις Γλώσσες για Ειδικούς Σκοπούς [Δ5] • γνωρίζουν για τα Σώματα Κειμένων και τα Λεξικά για τον Άνθρωπο και τη Μηχανή [Δ6] • μπορούν να παράγουν κώδικα στο λειτουργικό σύστημα Linux για Επεξεργασία ΣΚ [Δ7, Δ8] • γνωρίζουν για την Επεξεργασία Σωμάτων Κειμένων στο Διαδίκτυο [Δ9] • γνωρίζουν για Εφαρμογές Γλωσσικής Τεχνολογίας για τη Διαχείριση Δυναμικών Δεδομένων Μεγάλου Όγκου [Δ10]
ΕΝΟΤΗΤΕΣ:	**ΔΙΑΛΕΞΗ 1 –Σώματα Κειμένων** [Διά ζώσης] [Κατερίνα Φραντζή] **ΔΙΑΛΕΞΗ 2 – Κανονικές Εκφράσεις** [Διά ζώσης] [Κατερίνα Φραντζή]

	ΔΙΑΛΕΞΗ 3 – Σώματα Κειμένων και Αυτόματη Μετάφραση [Βιντεοδιάλεξη] [Βάλια Κορδώνη, Κατερίνα Φραντζή] **ΔΙΑΛΕΞΗ 4 Σώματα Κειμένων ως Μεγάλα Δεδομένα** [Βιντεοδιάλεξη] [Σαράντος Καπιδάκης] **ΔΙΑΛΕΞΗ 5 – Σώματα Κειμένων και Γλώσσες για Ειδικούς Σκοπούς** [Βιντεοδιάλεξη] [Κωνσταντίνα Βαδάση, Κατερίνα Φραντζή] **ΔΙΑΛΕΞΗ 6 – Σώματα Κειμένων και Λεξικά για τον Άνθρωπο και τη Μηχανή** [Διά ζώσης] [Στέλλα Μαρκαντωνάτου, Κατερίνα Φραντζή] **ΔΙΑΛΕΞΗ 7 – Επεξεργασία Σωμάτων Κειμένων σε Υπολογιστικό Περιβάλλον Linux** [Διά ζώσης] [Κατερίνα Φραντζή] **ΔΙΑΛΕΞΗ 8 – Επεξεργασία Σωμάτων Κειμένων σε Υπολογιστικό Περιβάλλον Linux** [Διά ζώσης] [Κατερίνα Φραντζή] **ΔΙΑΛΕΞΗ 9 – Επεξεργασία Σωμάτων Κειμένων στο Διαδίκτυο** [Διά ζώσης] [Αναστασία Κατρανίδου, Κατερίνα Φραντζή] **ΔΙΑΛΕΞΗ 10 – Εφαρμογές Γλωσσικής Τεχνολογίας για τη Διαχείριση Δυναμικών Δεδομένων Μεγάλου Όγκου** [Βιντεοδιάλεξη] [Σοφία Στάμου]
ΑΞΙΟΛΟΓΗΣΗ:	Δραστηριότητες : 10 μονάδες Παράδοση: Εξεταστική Περίοδος
ΒΙΒΛΙΟΓΡΑΦΙΑ:	**Ελληνική (ενδεικτικά)** Γεωργακοπούλου, Αλεξάνδρα & Διονύσιος Γούτσος (2011) Κείμενο και επικοινωνία, Ελληνική Γλώσσα: Θεωρία και διδακτική. Αθήνα: Πατάκης. Γούτσος, Διονύσης, Σηφιανού, Μαρία & Αλεξανδρα Γεωργακοπούλου (2006) Η ελληνική ως ξένη γλώσσα : από τις λέξεις στα κείμενα, Σπύρος Α. Μοσχονάς (επιμ.) Ελληνική γλώσσα : θεωρία και διδακτική. Αθήνα: Πατάκης. Γούτσος, Διονύσιος (επιμ.) (2006) Ο κόσμος των κειμένων: μελέτες αφιερωμένες στον Γεώργιο Μπαμπινιώτη. Αθήνα: Ελληνικά Γράμματα. 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. Καπιδάκης, Σαράντος (2006) Από την έντυπη στην ψηφιακή πληροφόρηση. *Πρακτικά Συνεδρίου Οι Βιβλιοθήκες χθές, σήμερα, αύριο*, 19 Μαΐου 2006 (με την ευκαιρία 50 χρόνων λειτουργίας του Ιδρύματος Ευγενίδου), Ίδρυμα Ευγενίδου, Βιβλιοθήκη. Κατσογιάννου, Μαριάννα & Ελένη Ευθυμίου (2004). *Ελληνική Ορολογία: Έρευνα και εφαρμογές*. Αθήνα: Καστανιώτης. Μαρκόπουλος, Γ. (2006) Ζητήματα Υπολογιστικής Γλωσσολογίας: Prolog και μορφολογική ανάλυση. Αθήνα: περιοδικό Παρουσία, παράρτημα 69. Μικρός, Γεώργιος Κ. (2004) Ηλεκτρονικά Σώματα κειμένων και ορολογία. Στο Κατσογιάννου, Μ. και Ε. Ευθυμίου (επιμ.) Ελληνική Ορολογία – Έρευνα και Εφαρμογές. Αθήνα: Καστανιώτης, 139-164. Μικρός, Γεώργιος Κ. (2009) Η ποσοτική ανάλυση της κοινωνιογλωσσολογικής ποικιλίας: Θεωρητικές και μεθοδολογικές προσεγγίσεις, Γλώσσα – Εκπαίδευση – Κοινωνία. Αθήνα: Μεταίχμιο. 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 Tzevelekou (2000). Η μηχανική μετάφραση και η ελληνική γλώσσα. Αθήνα: Καστανιώτη.
Φραντζή, Κατερίνα Θ. (2012) Εισαγωγή στην επεξεργασία σωμάτων κειμένων. Αθήνα: Ίων.

Ξενόγλωσση (ενδεικτικά)

Gavrilidou, M., Karagiannis, G., Markantonatou, S. Piperidis, S., Stainhaouer, G. (2000) Proceedings of the Second International Conference on Language Resources and Evaluation, Athens, Greece.
Baker, Paul, Hardie, Andrew & Tony McEnery (2006) A glossary of corpus linguistics. Edinburgh: Edinburgh University Press.
Biber, Douglas, Conrad, Susan & Randi Reppen (1998) Corpus linguistics: investigating language structure and use. Cambridge approaches to linguistics Cambridge: Cambridge University Press.
Dologlou, I., Markantonatou, S., Tambouratzis, G., Yannoutsou, O., Fourla, & Ioannou, N. (2003) Using Monolingual Corpora for Statistical Machine Translation: The METIS System. In Proceedings of the EAMT-CLAW 03: Controlled Language Translation, pages 61-68. Dublin, Ireland.
Gavrilidou, Maria (2002) The Hellenic National Corpus on-line. Revue Belge de Philologie et d' Histoire 80/2002, Societe po.
Kennedy, Graeme (1998) An introduction to corpus linguistics, Studies in language and linguistics. London; New York: Longman.
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.-
Leistyna, Pepi & Charles F. Meyer (2003) Corpus analysis: language structure and language use, Language and computers. Amsterdam: Rodopi.
Markantonatou, S., Sofianopoulos, S., Spilioti, V., Tambouratzis, G., Vasileiou, M., Yannoutsou, O. et al. (2005). Monolingual Corpus-based MT using Chunks. In Proceedings of the Workshop on EBMT, held within the MT Summit X, pages 91-98. Phuket, Thailand.
McEnery, Tony & Andrew Wilson (2001) Corpus linguistics: an introduction, Edinburgh textbooks in empirical linguistics. Edinburgh: Edinburgh University Press.
Meyer, Charles F. (2002) English corpus linguistics: an introduction, Studies in English language. Cambridge; New York: Cambridge University Press.
Sinclair, John & Ronald Carter (eds.) (2004) Trust the text: language, corpus and discourse. London; New York: Routledge.
Stubbs, Michael (1996) Text and corpus analysis: computer-assisted studies of language and culture, Language in society 23. Oxford: Blackwell.
Teubert, Wolfgang & Anna Cermakova (2007) Corpus linguistics: a short introduction. London; New York: Continuum.
Wynne, Martin (editor) (2005) Developing Linguistic Corpora: a Guide to Good Practice. Oxford: Oxbow Books. Available online from http://ahds.ac.uk/linguistic-corpora/