

ΙΔΡΥΜΑ	ΕΘΝΙΚΟ ΚΑΙ ΚΑΠΟΔΙΣΤΡΙΑΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ									
ΣΧΟΛΗ	ΘΕΤΙΚΩΝ ΕΠΙΣΤΗΜΩΝ									
ΤΜΗΜΑ	ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΤΗΛΕΠΙΚΟΙΝΩΝΙΩΝ									
ΕΠΙΠΕΔΟ ΣΠΟΥΔΩΝ	ΜΕΤΑΠΤΥΧΙΑΚΟ - ΔΠΜΣ ΓΛΩΣΣΙΚΗ ΤΕΧΝΟΛΟΓΙΑ									
ΤΙΤΛΟΣ ΜΑΘΗΜΑΤΟΣ	Ποσοτική Ανάλυση Γλωσσικών Δεδομένων									
ΚΩΔΙΚΟΣ ΜΑΘΗΜΑΤΟΣ	M905	Εξάμηνο	2	ECTS	6					
ΩΡΕΣ ΔΙΔΑΣΚΑΛΙΑΣ	ΘΕΩΡ.	39	ΦΡΟΝΤ.	ΕΡΓΑΣΤ.						
ΤΥΠΟΣ ΜΑΘΗΜΑΤΟΣ	<p>Επιλέξτε ένα από τα ακόλουθα και διαγράψτε τα υπόλοιπα Βασικό (B) / Επιλογής (E)</p> <table border="1"> <tr> <td>B</td> <td>E</td> </tr> <tr> <td></td> <td>X</td> </tr> </table>						B	E		X
B	E									
	X									
ΗΛΕΚΤΡΟΝΙΚΗ ΣΕΛΙΔΑ ΜΑΘΗΜΑΤΟΣ (URL)										
ΠΡΟΑΠΑΙΤΟΥΜΕΝΑ ΜΑΘΗΜΑΤΑ	ΟΧΙ									
ΓΛΩΣΣΑ ΔΙΔΑΣΚΑΛΙΑΣ και ΕΞΕΤΑΣΕΩΝ:	ΕΛΛΗΝΙΚΑ [Θα διατεθούν σημειώσεις και στα ΑΓΓΛΙΚΑ]									
ΤΟ ΜΑΘΗΜΑ ΠΡΟΣΦΕΡΕΤΑΙ ΣΕ ΦΟΙΤΗΤΕΣ ERASMUS	ΟΧΙ									

ΠΕΡΙΕΧΟΜΕΝΟ ΜΑΘΗΜΑΤΟΣ

Συμπληρώστε (λεκτική περιγραφή) το περιεχόμενο του μαθήματος αναφέροντας τη δομή και τα θέματα που καλύπτονται.

Το μάθημα διαρθρώνεται σε τέσσερις ενότητες:

(α) Δημιουργία και επεξεργασία κειμένων: Παρουσιάζονται γνωστά διαθέσιμα σώματα κειμένων, και περιγράφονται βασικά στάδια δημιουργίας (π.χ. web crawling) και προ-επεξεργασίας σωμάτων κειμένων (π.χ. sentence splitting, tokenization, normalization).

(β) Αναπαράσταση κειμενικών συλλογών: Παρουσιάζονται βασικές μέθοδοι μοντελοποίησης κειμενικών συλλογών (π.χ. Bag Of Words, TF-IDF, Dictionaries) και παραδείγματα οπτικοποίησης δεδομένων.

(γ) Περιγραφική Στατιστική: Περιγράφονται βασικές έννοιες (π.χ. μέση τιμή, διάμεση τιμή, επικρατούσα τιμή, εύρος, διασπορά, μεταβλητότητα) και παρουσιάζονται γνωστές κατανομές (π.χ. Διωνυμική,, Κανονική, Poisson) συνοδευόμενες με απλές εφαρμογές τους στην περιγραφή σωμάτων κειμένων.

(δ) Επαγωγική Στατιστική: Περιγράφονται βασικές μέθοδοι εξαγωγής συμπερασμάτων (π.χ. έλεγχος υποθέσεων, t test, ANOVA) συνοδευόμενες με απλές εφαρμογές τους στην ποσοτική ανάλυση κειμένων.

(ε) Αναπαράσταση κειμένων ως διανύσματα: Περιγράφεται η αναπαράσταση δεδομένων σε n διαστάσεις και παρουσιάζονται βασικές έννοιες συσχέτισης χαρακτηριστικών (π.χ. correlation, similarity, regression).

Παρουσιάζονται απλά παραδείγματα εφαρμογών χρησιμοποιώντας απλά scripts της Python.

ΜΑΘΗΣΙΑΚΑ ΑΠΟΤΕΛΕΣΜΑΤΑ

Περιγράψτε τους στόχους ή/και τα μαθησιακά αποτελέσματα του μαθήματος. Μπορείτε να αναφερθείτε στις επιμέρους κατηγορίες των μαθησιακών αποτελεσμάτων σε επίπεδο γνώσεων, δεξιοτήτων (νοητικών, πρακτικών) και ικανοτήτων.

Συμβουλευτείτε το Παράρτημα «Περίληπτικός Οδηγός συγγραφής Μαθησιακών Αποτελεσμάτων» για τη συγγραφή των μαθησιακών αποτελεσμάτων .

Διδακτικοί-Μαθησιακοί Στόχοι:

- Η απόκτηση γνώσεων για την προ-επεξεργασία κειμενικών συλλογών, την αναπαράστασή τους σε μορφή κατάλληλη για την ανάλυσή τους και τελικά την εξαγωγή συμπερασμάτων.

Οι φοιτητές που θα παρακολουθήσουν με επιτυχία το μάθημα θα μπορούν να:

- εφαρμόζουν βασικές μεθόδους επεξεργασίας και αναπαράστασης κειμενικών συλλογών,
- κατανοούν τη χρήση στατιστικών εννοιών για την περιγραφή κειμενικών συλλογών,
- εφαρμόζουν τεχνικές επαγωγικής στατιστικής για την εξαγωγή συμπερασμάτων κατά την εξέταση και σύγκριση κειμένων/συλλογών κειμένων,
- κατανοούν και εφαρμόζουν μεθόδους εξέτασης της ομοιότητας/συσχέτισης δεδομένων

ΔΙΔΑΚΤΙΚΕΣ και ΜΑΘΗΣΙΑΚΕΣ ΜΕΘΟΔΟΙ - ΑΞΙΟΛΟΓΗΣΗ											
ΤΡΟΠΟΣ ΠΑΡΑΔΟΣΗΣ	Με τηλεδιάσκεψη										
ΧΡΗΣΗ ΤΕΧΝΟΛΟΓΙΩΝ ΠΛΗΡΟΦΟΡΙΑΣ ΚΑΙ ΕΠΙΚΟΙΝΩΝΙΩΝ	<p>Υποστήριξη μαθησιακής διαδικασίας μέσω της ηλεκτρονικής πλατφόρμας e-class:</p> <ul style="list-style-type: none"> • Παροχή υλικού • Ανακοινώσεις • Ανάθεση εργασιών • Ομάδες φοιτητών <p>Επικοινωνία μέσω ηλεκτρονικού ταχυδρομείου</p>										
ΟΡΓΑΝΩΣΗ ΔΙΔΑΣΚΑΛΙΑΣ – ΔΙΔΑΚΤΙΚΕΣ ΤΕΧΝΙΚΕΣ <i>Περιγράφονται αναλυτικά ο τρόπος και τεχνικές διδασκαλίας και αναγράφονται αναλυτικά οι ώρες μελέτης του φοιτητή για κάθε μαθησιακή δραστηριότητα καθώς και οι ώρες μη καθοδηγούμενης μελέτης ώστε ο συνολικός φόρτος εργασίας σε επίπεδο εξαμήνου να αντιστοιχεί στα standards του ECTS (25 ώρες φόρτου εργασίας ανά πιστωτική μονάδα)</i> <i>Εμπλουτισμένες Διαλέξεις,</i> <i>Online Διαλέξεις,</i> <i>Σεμινάρια,</i> <i>Φροντιστήριο,</i> <i>Εργαστήριο,</i> <i>Εργαστηριακή Άσκηση,</i> <i>Μελέτη & ανάλυση βιβλιογραφίας,</i> <i>Πρακτική</i> <i>Εκπόνηση project,</i> <i>Εκπόνηση ατομικών / ομαδικών εργασιών</i> <i>Τηλεσυνεργασία (αναφορά σε εργαλεία)</i> <i>Κλπ</i>	<p>Η θεωρία παρουσιάζεται με προβολή διαφανειών οι οποίες είναι διαθέσιμες στο eclass.</p> <table border="1"> <thead> <tr> <th>Δραστηριότητα</th> <th>Φόρτος (ώρες)</th> </tr> </thead> <tbody> <tr> <td>Διαλέξεις</td> <td>39</td> </tr> <tr> <td>Ασκήσεις</td> <td>27</td> </tr> <tr> <td>Αυτοτελής Μελέτη</td> <td>85</td> </tr> <tr> <td>Σύνολο ωρών</td> <td>150</td> </tr> </tbody> </table>	Δραστηριότητα	Φόρτος (ώρες)	Διαλέξεις	39	Ασκήσεις	27	Αυτοτελής Μελέτη	85	Σύνολο ωρών	150
Δραστηριότητα	Φόρτος (ώρες)										
Διαλέξεις	39										
Ασκήσεις	27										
Αυτοτελής Μελέτη	85										
Σύνολο ωρών	150										
ΑΞΙΟΛΟΓΗΣΗ ΦΟΙΤΗΤΩΝ <i>Περιγραφή της διαδικασίας αξιολόγησης</i> <i>Μέθοδοι αξιολόγησης (Διαμορφωτική ή/και Τελική),</i> <i>Εργαλεία Αξιολόγησης (Δοκιμασία Πολλαπλής Επιλογής, Ερωτήσεις Σύντομης Απάντησης, Ερωτήσεις Ανάπτυξης Δοκιμίων, Επίλυση Προβλημάτων, Γραπτή Εργασία, Έκθεση/Αναφορά, Προφορική Εξέταση, Δημόσια Παρουσίαση, Εργαστηριακή Εργασία, Άλλη/άλλες,)</i> <i>Παροχή ανατροφοδότησης (περιγραφική, μέσω κλίμακας διαβαθμισμένων κριτηρίων)</i>	<p>Περιγράφονται ρητά οι μέθοδοι, τα εργαλεία αξιολόγησης και η παρεχόμενη ανατροφοδότηση αποτελεσμάτων. Συμπληρώνεται αναλόγως και ο παρακάτω πίνακας.</p> <table border="1"> <thead> <tr> <th>Αξιολόγηση</th> <th>Αριθμός</th> <th>Ποσοστό</th> </tr> </thead> <tbody> <tr> <td>Ασκήσεις</td> <td>3</td> <td>60%</td> </tr> <tr> <td>Τελική εργασία</td> <td>1</td> <td>40%</td> </tr> </tbody> </table>	Αξιολόγηση	Αριθμός	Ποσοστό	Ασκήσεις	3	60%	Τελική εργασία	1	40%	
Αξιολόγηση	Αριθμός	Ποσοστό									
Ασκήσεις	3	60%									
Τελική εργασία	1	40%									

ΣΥΝΙΣΤΩΜΕΝΗ-ΒΙΒΛΙΟΓΡΑΦΙΑ

1. Statistics for Linguists by David Eddington
2. Statistics in Corpus Linguistics by Václav Brezina
3. Statistical Methods in Language and Linguistic Research
4. <http://www.nltk.org/book/>
5. <https://nlp.stanford.edu/fsnlp/>
6. Applied Text Analysis with Python: Enabling Language-Aware Data Products with Machine Learning . Benjamin Bengfort, Rebecca Bilbro, and Tony Ojeda