

ΠΕΡΙΓΡΑΜΜΑ ΜΑΘΗΜΑΤΟΣ

1. ΓΕΝΙΚΑ

ΣΧΟΛΗ	ΠΟΛΥΤΕΧΝΙΚΗ, ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΑΤΡΩΝ		
ΤΜΗΜΑ	ΜΗΧΑΝΙΚΩΝ Η/Υ ΚΑΙ ΠΛΗΡΟΦΟΡΙΚΗΣ		
ΕΠΙΠΕΔΟ ΣΠΟΥΔΩΝ	ΠΡΟΠΤΥΧΙΑΚΟ, ΕΠΙΛΟΓΗΣ ΚΟΡΜΟΥ		
ΚΩΔΙΚΟΣ ΜΑΘΗΜΑΤΟΣ	CEID_NE5597	ΕΞΑΜΗΝΟ ΣΠΟΥΔΩΝ	ΧΕΙΜΕΡΙΝΟ
ΤΙΤΛΟΣ ΜΑΘΗΜΑΤΟΣ	ΑΝΑΚΤΗΣΗ ΠΛΗΡΟΦΟΡΙΑΣ		
ΑΥΤΟΤΕΛΕΙΣ ΔΙΔΑΚΤΙΚΕΣ ΔΡΑΣΤΗΡΙΟΤΗΤΕΣ σε περίπτωση που οι πιστωτικές μονάδες απονέμονται σε διακριτά μέρη του μαθήματος π.χ. Διαλέξεις, Εργαστηριακές Ασκήσεις κ.λπ. Αν οι πιστωτικές μονάδες απονέμονται ενιαία για το σύνολο του μαθήματος αναγράψτε τις εβδομαδιαίες ώρες διδασκαλίας και το σύνολο των πιστωτικών μονάδων		ΕΒΔΟΜΑΔΙΑΙΕΣ ΩΡΕΣ ΔΙΔΑΣΚΑΛΙΑΣ	ΠΙΣΤΩΤΙΚΕΣ ΜΟΝΑΔΕΣ
Διαλέξεις		2	2
Φροντιστήριο		1	1
Εργαστηριακές Ασκήσεις		2	2
Προσθέστε σειρές αν χρειαστεί. Η οργάνωση διδασκαλίας και οι διδακτικές μέθοδοι που χρησιμοποιούνται περιγράφονται αναλυτικά στο 4.			
ΤΥΠΟΣ ΜΑΘΗΜΑΤΟΣ <i>Υποβάθρου , Γενικών Γνώσεων, Επιστημονικής Περιοχής, Ανάπτυξης Δεξιοτήτων</i>	Ανάπτυξης Δεξιοτήτων		
ΠΡΟΑΠΑΙΤΟΥΜΕΝΑ ΜΑΘΗΜΑΤΑ:	Δεν υπάρχουν προαπαιτούμενα. Συνιστάται η κατοχή ενός βασικού συνόλου γνώσεων όπως παρέχεται από τα μαθήματα Δομές Δεδομένων (CEID_23Y210) , Αλγόριθμοι (CEID_23Y205) , και Βάσεις Δεδομένων (CEID_24Y334).		
ΓΛΩΣΣΑ ΔΙΔΑΣΚΑΛΙΑΣ και ΕΞΕΤΑΣΕΩΝ:	Ελληνικά. Μπορούν όμως να γίνουν και επιπλέον παραδόσεις στην αγγλική γλώσσα στην περίπτωση που αλλοδαποί φοιτητές παρακολουθούν το πρόγραμμα.		
ΤΟ ΜΑΘΗΜΑ ΠΡΟΣΦΕΡΕΤΑΙ ΣΕ ΦΟΙΤΗΤΕΣ ERASMUS	ΝΑΙ		
ΗΛΕΚΤΡΟΝΙΚΗ ΣΕΛΙΔΑ ΜΑΘΗΜΑΤΟΣ (URL)	https://eclass.upatras.gr/courses/CEID1037/		

2. ΜΑΘΗΣΙΑΚΑ ΑΠΟΤΕΛΕΣΜΑΤΑ

Μαθησιακά Αποτελέσματα Περιγράφονται τα μαθησιακά αποτελέσματα του μαθήματος οι συγκεκριμένες γνώσεις, δεξιότητες και ικανότητες καταλλήλου επιπέδου που θα αποκτήσουν οι φοιτητές μετά την επιτυχή ολοκλήρωση του μαθήματος. <i>Συμβουλευτείτε το Παράρτημα Α</i> - Περιγραφή του Επιπέδου των Μαθησιακών Αποτελεσμάτων για κάθε ένα κύκλο σπουδών σύμφωνα με Πλαίσιο Προσόντων του Ευρωπαϊκού Χώρου Ανώτατης Εκπαίδευσης - Περιγραφικοί Δείκτες Επιπέδων 6, 7 & 8 του Ευρωπαϊκού Πλαισίου Προσόντων Διά Βίου Μάθησης και Παράρτημα Β - Περιληπτικός Οδηγός συγγραφής Μαθησιακών Αποτελεσμάτων
Στο τέλος αυτού του μαθήματος ο φοιτητής θα μπορεί: - να εξηγήσει τις βασικές έννοιες αποθήκευσης και ανάκτησης πληροφορίας, και να κατανοήσει τα θέματα που είναι ειδικά για αποτελεσματική ανάκτηση πληροφορίας - να εξηγήσει τις διαφορές ανάμεσα σε εναλλακτικά μοντέλα ανάκτησης πληροφορίας και να αναλύσει γιατί ένα συγκεκριμένο μοντέλο είναι κατάλληλο για μία συγκεκριμένη εφαρμογή. - να κατανοεί τα βασικά σχεδιαστικά πίσω από την ανάπτυξη μίας αποτελεσματικής μηχανής ψαξίματος στο παγκόσμιο ιστό. - να σχεδιάζει και να υλοποιεί ένα μικρό προς μεγάλο σύστημα αποθήκευσης και ανάκτησης πληροφορίας. - να κατανοεί την έννοια της διανυσματικής αναπαράστασης λέξεων (word embeddings) και της χρήσης της σε διάφορες εφαρμογές

Στο τέλος του μαθήματος ο φοιτητής θα έχει περαιτέρω αναπτύξει τις ακόλουθες δεξιότητες:

1. ικανότητα να κατανοεί τις διάφορες έννοιες που εμπλέκονται σε συστήματα ανάκτησης πληροφορίας, και πως αυτά διασυνδέονται με την απόδοση συστημάτων που ήδη είναι σε λειτουργία.
2. ικανότητα να εφαρμόζει μεθοδολογικά αυτές τις έννοιες με σκοπό την σχεδίαση και την υλοποίηση αποτελεσματικών συστημάτων
3. ικανότητα να δουλεύει συνεργατικά, με σκοπό την επίλυση προβλημάτων που ανακύπτουν κατά την δόμηση ενός πλήρως λειτουργικού συστήματος ανάκτησης πληροφορίας

Γενικές Ικανότητες

Λαμβάνοντας υπόψη τις γενικές ικανότητες που πρέπει να έχει αποκτήσει ο πτυχιούχος (όπως αυτές αναγράφονται στο Παράρτημα Διπλώματος και παρατίθενται ακολούθως) σε ποια / ποιες από αυτές αποσκοπεί το μάθημα;

Αναζήτηση, ανάλυση και σύνθεση δεδομένων και πληροφοριών, με τη χρήση και των απαραίτητων τεχνολογιών

Προσαρμογή σε νέες καταστάσεις

Λήψη αποφάσεων

Αυτόνομη εργασία

Ομαδική εργασία

Εργασία σε διεθνές περιβάλλον

Εργασία σε διεπιστημονικό περιβάλλον

Παράγωγή νέων ερευνητικών ιδεών

Σχεδιασμός και διαχείριση έργων

Σεβασμός στη διαφορετικότητα και στην πολυπολιτισμικότητα

Σεβασμός στο φυσικό περιβάλλον

Επίδειξη κοινωνικής, επαγγελματικής και ηθικής υπευθυνότητας και ευαισθησίας σε θέματα φύλου

Άσκηση κριτικής και αυτοκριτικής

Προαγωγή της ελεύθερης, δημιουργικής και επαγωγικής σκέψης

Αναζήτηση, ανάλυση και σύνθεση δεδομένων και πληροφοριών, με τη χρήση και των απαραίτητων τεχνολογιών

Λήψη αποφάσεων

Αυτόνομη εργασία

Ομαδική εργασία

Σχεδιασμός και διαχείριση έργων

Προαγωγή της ελεύθερης, δημιουργικής και επαγωγικής σκέψης

Παράγωγή νέων ερευνητικών ιδεών

3. ΠΕΡΙΕΧΟΜΕΝΟ ΜΑΘΗΜΑΤΟΣ

1. Εισαγωγή (μοντελοποίηση χρήστη, λογική αναπαράσταση κειμένου, διεργασία ανάκτησης)
2. Μετρικές εκτίμησης απόδοσης (recall, precision, average precision, R-precision, precision histograms, NDCG metric, harmonic median, user oriented metrics).
3. Μοντελοποίηση στην Ανάκτηση Πληροφορίας
4. Συνολοθεωρητικά Μοντέλα (boolean models, fuzzy set model, extended boolean model), αλγεβρικά μοντέλα ((vector space models, latent semantic indexing model, topic models), πιθανοτικά μοντέλα (κλασσικά και γλωσσικά μοντέλα)
5. Ανάκτηση Πληροφορίας στον Παγκόσμιο Ιστό και ιδιαιτερότητες αυτού
6. Μηχανές Ψάξιματος στον Παγκόσμιο Ιστό (crawler, indexer), HITS algorithm (Hyperlink-induced topic search), Google μηχανή αναζήτησης (η PageRank μετρική), ο SALSA αλγόριθμος, παραλλαγές στο ψάξιμο στον παγκόσμιο Ιστό.
7. Τεχνικές Μηχανικής Μάθησης και Νευρωνικά Μοντέλα στην Ανάκτηση Πληροφορίας (Learning to Rank, διανυσματική αναπαράσταση λέξεων και ενσωματώσεις όπως word2vec, CBOW, skipgram, Transformers, BERT, GPT, Μεγάλα Γλωσσικά Μοντέλα και Sparse vs Dense Αναζήτηση, Vector Search σε λ.χ. FAISS (HNSW, ...), χρήση dense ή/και sparse search σε Retrieval Augmented Generation (RAG), Search engines vs reasoning engines)
8. Δομές Δεικτοδότησης (inverted files, signature files, bitmaps).
9. Τεχνικές Αποθήκευσης σε Κατανεμημένη Ανάκτηση Πληροφορίας (Mapreduce, Apache Spark),
10. Full indexing structures in main memory (suffix trees, suffix arrays, acyclic directed graphs (DAWG) for strings), and in secondary memory (supra-suffix array, prefix B-tree, string B-tree).
11. Συμπύκνωση Κειμένων και Δομών Δεικτοδότησης
12. Εξόρυξη Γνώσης από Κείμενα και Μοντέλα Βασισμένα σε Γράφους (Graph embeddings)

4. ΔΙΔΑΚΤΙΚΕΣ και ΜΑΘΗΣΙΑΚΕΣ ΜΕΘΟΔΟΙ - ΑΞΙΟΛΟΓΗΣΗ

ΤΡΟΠΟΣ ΠΑΡΑΔΟΣΗΣ

Πρόσωπο με πρόσωπο, Εξ αποστάσεως
εκπαίδευση κ.λπ.

Πρόσωπο με πρόσωπο

ΧΡΗΣΗ ΤΕΧΝΟΛΟΓΙΩΝ ΠΛΗΡΟΦΟΡΙΑΣ ΚΑΙ ΕΠΙΚΟΙΝΩΝΙΩΝ Χρήση Τ.Π.Ε. στη Διδασκαλία, στην Εργαστηριακή Εκπαίδευση, στην Επικοινωνία με τους φοιτητές	Γίνεται χρήση τεχνολογιών πληροφορίας και επικοινωνιών στην επικοινωνία με τους φοιτητές. Χρησιμοποιείται e_class, e_mail.																			
ΟΡΓΑΝΩΣΗ ΔΙΔΑΣΚΑΛΙΑΣ Περιγράφονται αναλυτικά ο τρόπος και μέθοδοι διδασκαλίας. Διαλέξεις, Σεμινάρια, Εργαστηριακή Άσκηση, Άσκηση Πεδίου, Μελέτη & ανάλυση βιβλιογραφίας, Φροντιστήριο, Πρακτική (Τοποθέτηση), Κλινική Άσκηση, Καλλιτεχνικό Εργαστήριο, Διαδραστική διδασκαλία, Εκπαιδευτικές επισκέψεις, Εκπόνηση μελέτης (project), Συγγραφή εργασίας / εργασιών, Καλλιτεχνική δημιουργία, κ.λπ. Αναγράφονται 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	<table><tr><th>Δραστηριότητα</th><th>Φόρτος Εργασίας Εξαμήνου</th></tr><tr><td>Διαλέξεις</td><td>60 ώρες εργασίας</td></tr><tr><td>Φροντιστήριο</td><td>30 ώρες εργασίας</td></tr><tr><td>Εργαστηριακές Ασκήσεις</td><td>60 ώρες εργασίας</td></tr><tr><td>Σύνολο Μαθήματος</td><td>150 ώρες εργασίας</td></tr><tr><td></td><td></td></tr><tr><td></td><td></td></tr><tr><td></td><td></td></tr><tr><td>Σύνολο Μαθήματος (25-30 ώρες φόρτου εργασίας ανά πιστωτική μονάδα)</td><td>150</td></tr></table>	Δραστηριότητα	Φόρτος Εργασίας Εξαμήνου	Διαλέξεις	60 ώρες εργασίας	Φροντιστήριο	30 ώρες εργασίας	Εργαστηριακές Ασκήσεις	60 ώρες εργασίας	Σύνολο Μαθήματος	150 ώρες εργασίας							Σύνολο Μαθήματος (25-30 ώρες φόρτου εργασίας ανά πιστωτική μονάδα)	150	
Δραστηριότητα	Φόρτος Εργασίας Εξαμήνου																			
Διαλέξεις	60 ώρες εργασίας																			
Φροντιστήριο	30 ώρες εργασίας																			
Εργαστηριακές Ασκήσεις	60 ώρες εργασίας																			
Σύνολο Μαθήματος	150 ώρες εργασίας																			
Σύνολο Μαθήματος (25-30 ώρες φόρτου εργασίας ανά πιστωτική μονάδα)	150																			
ΑΞΙΟΛΟΓΗΣΗ ΦΟΙΤΗΤΩΝ Περιγραφή της διαδικασίας αξιολόγησης Γλώσσα Αξιολόγησης, Μέθοδοι αξιολόγησης, Διαμορφωτική ή Συμπερασματική, Δοκιμασία Πολλαπλής Επιλογής, Ερωτήσεις Σύντομης Απάντησης, Ερωτήσεις Ανάπτυξης Δοκιμίων, Επίλυση Προβλημάτων, Γραπτή Εργασία, Έκθεση / Αναφορά, Προφορική Εξέταση, Δημόσια Παρουσίαση, Εργαστηριακή Εργασία, Κλινική Εξέταση Ασθενούς, Καλλιτεχνική Ερμηνεία, Άλλη / Άλλες Αναφέρονται ρητά προσδιορισμένα κριτήρια αξιολόγησης και εάν και που είναι προσβάσιμα από τους φοιτητές:	(1) στην παράδοση και στην παρουσίαση μιας εργασίας από ομάδες 1-2 ατόμων, που συνίσταται στη κριτική παρουσίαση, υλοποίηση και ανάλυση ενός συνόλου επιστημονικών εργασιών που πραγματοποιείται ένα από τα ζητήματα του μαθήματος. Εναλλακτικά οι φοιτητές μπορούν να εμπλακούν στην εις βάθος υλοποίηση τμημάτων ενός ολοκληρωμένου συστήματος κατάταξης ανάκτησης (45% του συνολικού βαθμού, μόνο όταν ο φοιτητής στην τελική εξέταση εξασφαλίσει πάνω από ένα ελάχιστο βαθμό), (2) σε γραπτή εξέταση (55% του συνολικού βαθμού)																			

5. ΣΥΝΙΣΤΩΜΕΝΗ ΒΙΒΛΙΟΓΡΑΦΙΑ

- Christofer D. Manning, Prabhakar Raghavan, Hinrich Schutze, Εισαγωγή στην Ανάκτηση Πληροφορικών 1η/2012 ΕΚΔΟΣΕΙΣ ΚΛΕΙΔΑΡΙΘΜΟΣ - Baeza-Yates Ricardo, Ribeiro-Neto Berthier, Ανάκτηση Πληροφορίας, 2η Έκδοση ΕΚΔΟΣΕΙΣ Α. ΤΖΙΟΛΑ & ΥΙΟΙ Α.Ε., 2η, 2015 - Omar Alonso and Ricardo Baeza-Yates, Information Retrieval: Advanced Topics and Techniques, December 2024, Association for Computing Machinery, New York, United States, ISBN:979-8-4007-1050-6 - Jimmy Lin, Rodrigo Nogueira, and Andrew Yates, Pretrained Transformers for Text Ranking: BERT and Beyond by University of Waterloo, University of Campinas, University of Amsterdam) Morgan & Claypool (Synthesis Lectures on Human Language Technologies, edited by Graeme Hirst, volume 53), 2021. - Dan Jurafsky and James H. Martin Speech and Language Processing (3rd ed. draft) Jan 2, 2025 draft (https://web.stanford.edu/~jurafsky/slp3/) - Mohammad Taher Pilehvar , Jose Camacho-Collados, Embeddings in Natural Language Processing Theory and Advances in Vector Representations of Meaning, Morgan Claypool, 2020 (https://sites.google.com/view/embeddings-in-nlp , https://sites.google.com/view/embeddings-in-nlp/tutorial) - Stefan Büttcher Charles L. A. Clarke, Gordon V. Cormack, Information Retrieval: Implementing and Evaluating Search Engines (MIT Press) Paperback – February 12, 2016
--

8. Langville Amy, Meyer Carl, Η μέθοδος Pagerank της Google και άλλα συστήματα κατάταξης ιστοσελίδων
GOOGLE ΚΑΙ ΑΛΛΑ ΣΥΣΤΗΜΑΤΑ ΚΑΤΑΤΑΞΗΣ ΙΣΤΟΣΕΛΙΔΩΝ ΙΔΡΥΜΑ ΤΕΧΝΟΛΟΓΙΑΣ & ΕΡΕΥΝΑΣ-
ΠΑΝΕΠΙΣΤΗΜΙΑΚΕΣ ΕΚΔΟΣΕΙΣ ΚΡΗΤΗΣ, 1Η/2010
9. Witten, A. Moffat, T. Bell, Managing Gigabytes: Compressing and Indexing Documents and Images, Morgan
Kaufmann Publishers, 1999.