

## 2<sup>ο</sup> ΕΞΑΜΗΝΟ:

### ΠΕΡΙΓΡΑΦΜΑ ΜΑΘΗΜΑΤΟΣ «Εξόρυξη Μεγάλων Δεδομένων»

#### 1. ΓΕΝΙΚΑ

<b>ΣΧΟΛΗ</b>	ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΤΗΛΕΠΙΚΟΙΝΩΝΙΩΝ		
<b>ΤΜΗΜΑ</b>	ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΤΗΛΕΠΙΚΟΙΝΩΝΙΩΝ		
<b>ΕΠΙΠΕΔΟ ΣΠΟΥΔΩΝ</b>	Μεταπτυχιακό		
<b>ΚΩΔΙΚΟΣ ΜΑΘΗΜΑΤΟΣ</b>	208	<b>ΕΞΑΜΗΝΟ ΣΠΟΥΔΩΝ</b>	2ο
<b>ΤΙΤΛΟΣ ΜΑΘΗΜΑΤΟΣ</b>	Εξόρυξη Μεγάλων Δεδομένων		
<b>ΑΥΤΟΤΕΛΕΙΣ ΔΙΔΑΚΤΙΚΕΣ ΔΡΑΣΤΗΡΙΟΤΗΤΕΣ</b>	<b>ΕΒΔΟΜΑΔΙΑΙΕΣ ΩΡΕΣ ΔΙΔΑΣΚΑΛΙΑΣ</b>	<b>ΠΙΣΤΩΤΙΚΕΣ ΜΟΝΑΔΕΣ</b>	
Διαλέξεις	3	6	
Εργαστηριακές Ασκήσεις			
<b>ΤΥΠΟΣ ΜΑΘΗΜΑΤΟΣ</b>	Ανάπτυξης δεξιοτήτων		
<b>ΠΡΟΑΠΑΙΤΟΥΜΕΝΑ ΜΑΘΗΜΑΤΑ:</b>			
<b>ΓΛΩΣΣΑ ΔΙΔΑΣΚΑΛΙΑΣ και ΕΞΕΤΑΣΕΩΝ:</b>	Ελληνική		
<b>ΤΟ ΜΑΘΗΜΑ ΠΡΟΣΦΕΡΕΤΑΙ ΣΕ ΦΟΙΤΗΤΕΣ ERASMUS</b>	ΝΑΙ (Στην Αγγλική)		
<b>ΗΛΕΚΤΡΟΝΙΚΗ ΣΕΛΙΔΑ ΜΑΘΗΜΑΤΟΣ (URL)</b>	<a href="https://ecourse.uoi.gr/course/view.php?id=1950">https://ecourse.uoi.gr/course/view.php?id=1950</a>		

#### 2. ΜΑΘΗΣΙΑΚΑ ΑΠΟΤΕΛΕΣΜΑΤΑ

##### Μαθησιακά Αποτελέσματα

Μετά την επιτυχή ολοκλήρωση του μαθήματος, οι φοιτητές θα είναι σε θέση:

- Να εκτελεί μια διερευνητική αριθμητική και γραφική ανάλυση ενός συνόλου δεδομένων μεγάλου όγκου.
- Να εφαρμόζει διαφορετικούς τύπους αλγορίθμων εξόρυξης δεδομένων σύμφωνα με τους διαφορετικούς τύπους προβλημάτων.
- Να κατανοεί τους υπολογιστικούς περιορισμούς των διαφόρων αλγορίθμων, όταν εφαρμόζονται σε δεδομένα μεγάλου όγκου.
- Να εφαρμόζει λύσεις σε προβλήματα πραγματικών εφαρμογών που παράγουν δεδομένα μεγάλου όγκου.

##### Γενικές Ικανότητες

- Αναζήτηση, ανάλυση και σύνθεση δεδομένων και πληροφοριών, με τη χρήση και των απαραίτητων τεχνολογιών.
- Αυτόνομη εργασία
- Προαγωγή της ελεύθερης, δημιουργικής και επαγωγικής σκέψης
- Αναλυτική και συνθετική σκέψη
- Κριτική σκέψη
- Επίλυση προβλημάτων
- Παραγωγή νέων ερευνητικών ιδεών
- Αναζήτηση, ανάλυση και σύνθεση πληροφοριών, με τη χρήση και των απαραίτητων τεχνολογιών

#### 3. ΠΕΡΙΕΧΟΜΕΝΟ ΜΑΘΗΜΑΤΟΣ

Ο σκοπός του μαθήματος είναι η παρουσίαση στους φοιτητές προηγμένων τεχνικών ανάλυσης και επεξεργασίας δεδομένων καθώς και τεχνικών παρουσίασης και απεικόνισης δεδομένων καθώς επίσης και της εφαρμογής τους σε μεγάλα δεδομένα (Big Data). Η παρουσίαση εργαλείων για την δημιουργία παράλληλων αλγορίθμων όπως το Map Reduce καθώς και η μελέτη καταναμημένων συστημάτων αρχείων. Η κατανόηση της αναζήτησης ομοιότητας καθώς και αλγορίθμων και τεχνικών δεδομένων για την επεξεργασία γρήγορων ροών δεδομένων. Τα κύρια μέρη του μαθήματος περιλαμβάνουν τη διερευνητική ανάλυση δεδομένων, την επιβλεπόμενη και μη επιβλεπόμενη μάθηση, την ανάλυση ροών δεδομένων και εφαρμογές αυτών σε πραγματικά προβλήματα της βιομηχανίας. Το μάθημα θέτει τα βασικά θεμέλια αυτών των περιοχών, αλλά καλύπτει επίσης θέματα αιχμής, Εύρεση προτύπων στην εξόρυξη δεδομένων, βασικές τεχνικές εξόρυξης δεδομένων και εύρεσης συχνών προτύπων και συσχετίσεων, μέτρα αξιολόγησης προτύπων, αναζήτηση Ομοιότητας, Minhashing & κατακερματισμός με διατήρηση τονικότητας, Εξόρυξη πολυδιάστατων προτύπων, Εξόρυξη ακολουθιακών προτύπων, Βασικές τεχνικές οπτικοποίησης δεδομένων, σχεδιασμός και κατασκευή απεικονίσεων, Οπτικοποίηση μη δομημένων πληροφοριών, τεχνολογία μηχανών αναζήτησης, αλγόριθμος Page-Rank.

Εβδομ.	Τίτλος Ενότητας	Βιβλιογραφία	e-class
1	<b>Εξόρυξη Δεδομένων:</b> Εισαγωγή, Τι είναι εξόρυξη δεδομένων, στατιστικά όρια της εξόρυξης δεδομένων.	[1,2]	<a href="https://ecourse.uoi.gr/course/view.php?id=1950">https://ecourse.uoi.gr/course/view.php?id=1950</a>
2	<b>Συστήματα Αρχείων Μεγάλης Κλίμακας και Map-Reduce:</b> Καταναμημένα συστήματα αρχείων, η μεθοδολογία MapReduce, Αλγόριθμοι που χρησιμοποιούν το MapReduce, επεκτάσεις, Αλγοριθμική πολυπλοκότητα.	[1,2]	<a href="https://ecourse.uoi.gr/course/view.php?id=1950">https://ecourse.uoi.gr/course/view.php?id=1950</a>
3	<b>Αναζήτηση Όμοιων Αντικειμένων:</b> μετρικές απόστασης, αναζήτηση κοντινότερου γείτονα, θρυμματοποίηση εγγράφων.	[1,2]	<a href="https://ecourse.uoi.gr/course/view.php?id=1950">https://ecourse.uoi.gr/course/view.php?id=1950</a>
4	<b>Αναζήτηση Όμοιων Αντικειμένων:</b> συνόψεις που διατηρούν την ομοιότητα, κατακερματισμός, οικογένειες συναρτήσεων LSH, μέθοδοι για υψηλό βαθμό ομοιότητας.	[1,2]	<a href="https://ecourse.uoi.gr/course/view.php?id=1950">https://ecourse.uoi.gr/course/view.php?id=1950</a>
5	<b>Εξόρυξη από ροές δεδομένων:</b> μοντέλο δεδομένων ροής, δειγματοληψία σε ροή δεδομένων, φιλτράρισμα ροής, μέτρηση διακριτών στοιχείων ροής.	[1,2]	<a href="https://ecourse.uoi.gr/course/view.php?id=1950">https://ecourse.uoi.gr/course/view.php?id=1950</a>

6	<b>Ανάλυση Συνδέσμων:</b> ο αλγόριθμος PageRank, αποδοτικός υπολογισμός του PageRank, PageRank με βάση το θέμα.	[1,2]	<a href="https://ecourse.uoi.gr/course/view.php?id=1950">https://ecourse.uoi.gr/course/view.php?id=1950</a>
7	<b>Συχνά στοιχειοσύνολα:</b> μοντέλο καλαθιού αγορών, καλάθια αγορών και αλγόριθμος A-priori, διαχείριση μεγάλων συνόλων δεδομένων στην κύρια μνήμη.	[1,2]	<a href="https://ecourse.uoi.gr/course/view.php?id=1950">https://ecourse.uoi.gr/course/view.php?id=1950</a>
8	<b>Ομαδοποίηση:</b> Ιεραρχική Ομαδοποίηση, Αλγόριθμος κ-μέσων, αλγόριθμος Cure, ομαδοποίηση σε μη ευκλείδειους χώρους.	[1,2]	<a href="https://ecourse.uoi.gr/course/view.php?id=1950">https://ecourse.uoi.gr/course/view.php?id=1950</a>
9	<b>Ομαδοποίηση:</b> ομαδοποίηση σε μη ευκλείδειους χώρους, ομαδοποίηση για ροές και παραλληλισμό.	[1,2]	<a href="https://ecourse.uoi.gr/course/view.php?id=1950">https://ecourse.uoi.gr/course/view.php?id=1950</a>
10	<b>Διαφήμιση στο διαδίκτυο:</b> θέματα, διαδικτυακής διαφήμισης, αλγόριθμος άμεσης επεξεργασίας, το πρόβλημα του ταιριάσματος.	[1,2]	<a href="https://ecourse.uoi.gr/course/view.php?id=1950">https://ecourse.uoi.gr/course/view.php?id=1950</a>
11	<b>Συστήματα συστάσεων:</b> Μοντέλο συστάσεων, συστάσεις βασισμένες στο περιεχόμενο, συνεργατική διήθηση,	[1,2]	<a href="https://ecourse.uoi.gr/course/view.php?id=1950">https://ecourse.uoi.gr/course/view.php?id=1950</a>
12	<b>Τεχνικές μείωσης διάστασης:</b> αποσύνθεση UV, λάθος μέσης τετραγωνικής ρίζας, υπολογιστικά θέματα.	[1,2]	<a href="https://ecourse.uoi.gr/course/view.php?id=1950">https://ecourse.uoi.gr/course/view.php?id=1950</a>
13	<b>Παρουσιάσεις Εργασιών φοιτητών.</b>		

#### 4. ΔΙΔΑΚΤΙΚΕΣ ΚΑΙ ΜΑΘΗΣΙΑΚΕΣ ΜΕΘΟΔΟΙ - ΑΞΙΟΛΟΓΗΣΗ

ΤΡΟΠΟΣ ΠΑΡΑΔΟΣΗΣ.	Πρόσωπο με πρόσωπο
<b>ΧΡΗΣΗ ΤΕΧΝΟΛΟΓΙΩΝ ΠΛΗΡΟΦΟΡΙΑΣ ΚΑΙ ΕΠΙΚΟΙΝΩΝΙΩΝ</b>	<input checked="" type="checkbox"/> Χρήση ηλεκτρονικών παρουσιάσεων, αναρτημένων στο e-class.
	<input checked="" type="checkbox"/> Χρήση λογισμικού στον υπολογιστή κατά τη διάλεξη.
	<input checked="" type="checkbox"/> Χρήση εξειδικευμένου λογισμικού
	<input checked="" type="checkbox"/> Διάθεση εκπαιδευτικού υλικού μέσω e-class.
	<input type="checkbox"/> Διαχείριση εργασιών/ασκήσεων μέσω δικτυακού τόπου.
	<input checked="" type="checkbox"/> Επικοινωνία με φοιτητές μέσω e-mail.
	<input checked="" type="checkbox"/> Ηλεκτρονικός χώρος συνομιλιών διδάσκοντος και

		φοιτητών.	
ΟΡΓΑΝΩΣΗ ΔΙΔΑΣΚΑΛΙΑΣ	Δραστηριότητα	Φόρτος Εργασίας Εξαμήνου	
	Διαλέξεις	39 ώρες	
	Εργαστηριακές Ασκήσεις	-	
	Βιβλιογραφική Εργασία	61 ώρες	
	Υλοποίηση Project	-	
	Μη καθοδηγούμενη μελέτη	70 ώρες	
	<b>Σύνολο Μαθήματος</b>	<b>170 ώρες</b>	
	ΑΞΙΟΛΟΓΗΣΗ ΦΟΙΤΗΤΩΝ	<p>Η τελική βαθμολογία για το μάθημα προκύπτει με 50% από την τελική εξέταση και 50% από τον βαθμό του project.</p> <p>Οι φοιτητές αναλαμβάνουν να παρουσιάσουν ατομικά μια συναφή εργασία (paper) από κάποιο έγκριτο περιοδικό ή συνέδριο που είναι συναφές με το αντικείμενο του μαθήματος (50%).</p> <p>Τα κριτήρια αξιολόγησης γνωστοποιούνται στους φοιτητές στην πρώτη διάλεξη, τα οποία και αναφέρονται ρητά στο syllabus του μαθήματος, το οποίο είναι και διαθέσιμο στο e-class.</p>	

## 5. ΣΥΝΙΣΤΩΜΕΝΗ-ΒΙΒΛΙΟΓΡΑΦΙΑ

- Προτεινόμενη Βιβλιογραφία:

1. Jure Leskovec, Amand Rajaraman, Jeffrey David Ullman "Mining of Massive Datasets" 2nd Edition Cambridge University Press 2014
2. Anand Rajaraman and Jeffrey David Ullman "Εξόρυξη από Μεγάλα Σύνολα Δεδομένων» Εκδόσεις Νέων Τεχνολογιών 2013
3. Tan Pang - Ning, Steinbach Michael, Kumar Vipin, Introduction to Data Mining.
4. Εισαγωγή στην Εξόρυξη δεδομένων, ΕΚΔΟΣΕΙΣ Α. ΤΖΙΟΛΑ & ΥΙΟΙ Α.Ε.
5. Jiawei Han, Micheline Kamber and Jian Pei Data Mining: Concepts and Techniques. Morgan Kaufmann 3rd edition 2011
6. Nikhil Buduma, Fundamentals of Deep Learning, O'Reily