

ΕΙΔΙΚΑ ΘΕΜΑΤΑ ΒΑΣΕΩΝ ΔΕΔΟΜΕΝΩΝ

ΠΕΡΙΓΡΑΜΜΑ ΜΑΘΗΜΑΤΟΣ

1. ΓΕΝΙΚΑ

ΣΧΟΛΗ	ΘΕΤΙΚΩΝ ΕΠΙΣΤΗΜΩΝ		
ΤΜΗΜΑ	ΠΛΗΡΟΦΟΡΙΚΗΣ		
ΕΠΙΠΕΔΟ ΣΠΟΥΔΩΝ	ΠΡΟΠΤΥΧΙΑΚΟ		
ΚΩΔΙΚΟΣ ΜΑΘΗΜΑΤΟΣ	704ΕΔΕΕ	ΕΞΑΜΗΝΟ ΣΠΟΥΔΩΝ	7 ^ο
ΤΙΤΛΟΣ ΜΑΘΗΜΑΤΟΣ	ΕΙΔΙΚΑ ΘΕΜΑΤΑ ΒΑΣΕΩΝ ΔΕΔΟΜΕΝΩΝ		
ΑΥΤΟΤΕΛΕΙΣ ΔΙΔΑΚΤΙΚΕΣ ΔΡΑΣΤΗΡΙΟΤΗΤΕΣ <i>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</i>		ΕΒΔΟΜΑΔΙΑΙΕΣ ΩΡΕΣ ΔΙΔΑΣΚΑΛΙΑΣ	ΠΙΣΤΩΤΙΚΕΣ ΜΟΝΑΔΕΣ
Διαλέξεις		2	5
Φροντιστηριακές Ασκήσεις		1	
<i>Προσθέστε σειρές αν χρειαστεί. Η οργάνωση διδασκαλίας και οι διδακτικές μέθοδοι που χρησιμοποιούνται περιγράφονται αναλυτικά στο 4.</i>			
ΤΥΠΟΣ ΜΑΘΗΜΑΤΟΣ <i>Υποβάθρου, Γενικών Γνώσεων, Επιστημονικής Περιοχής, Ανάπτυξης Δεξιοτήτων</i>	Επιστημονικής Περιοχής		
ΠΡΟΑΠΑΙΤΟΥΜΕΝΑ ΜΑΘΗΜΑΤΑ:	-		
ΓΛΩΣΣΑ ΔΙΔΑΣΚΑΛΙΑΣ και ΕΞΕΤΑΣΕΩΝ:	Ελληνική		
ΤΟ ΜΑΘΗΜΑ ΠΡΟΣΦΕΡΕΤΑΙ ΣΕ ΦΟΙΤΗΤΕΣ ERASMUS	ΟΧΙ		
ΗΛΕΚΤΡΟΝΙΚΗ ΣΕΛΙΔΑ ΜΑΘΗΜΑΤΟΣ (URL)	-		

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.

Συμβουλευτείτε το Παράρτημα Α

- 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 μάθημα θα δοθεί έμφαση σε νέες τεχνολογίες Βάσεων Δεδομένων που αναφέρονται και στο πλαίσιο του όρου "Big Data". Τα "Big Data" δημιουργούν σημαντικότερες προκλήσεις στην αποτελεσματική επεξεργασία, αποθήκευση και μεταφορά τους. Το μάθημα αποσκοπεί στο να παρουσιάσει στους φοιτητές ορισμένες σύγχρονες τεχνικές, συστήματα και πλατφόρμες για αποδοτική διαχείριση και ανάλυση δεδομένων μεγάλης κλίμακας.

Επειδή το θέμα των Big Data είναι ευρύτατο θα δοθεί κυρίως έμφαση σε τεχνολογίες που αναπτύσσονται στα πλαίσια του open-source συστήματος Apache Spark. Το Spark αποτελεί ένα σύγχρονο και αποτελεσματικό σύστημα για καταμετρημένη επεξεργασία των Big Data και στα πλαίσιά του είναι δυνατή η μελέτη πλήθους προβλημάτων που συσχετίζονται με Big Data και αποτελεσματικών τρόπων αντιμετώπισής τους. Στο πλαίσιο των Big Data θα γίνει και εισαγωγή της νέας κατεύθυνσης του machine learning που αφορά το deep learning. Με το μάθημα αυτό, οι φοιτητές αναμένεται ότι θα αποκτήσουν σημαντικές τεχνικές δεξιότητες σε ότι αφορά την διαχείριση δεδομένων μεγάλης κλίμακας και θα εξοικειωθούν με αλγορίθμους και μεθόδους ανάλυσής τους.

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

- Αυτόνομη Εργασία
- Ομαδική Εργασία

3. ΠΕΡΙΕΧΟΜΕΝΟ ΜΑΘΗΜΑΤΟΣ

- Εισαγωγή
- Βασικές έννοιες των Big Data (Volume, Variety, Velocity, Veracity, Validity and Volatility). Εφαρμογές των Big Data με έμφαση στην Βιοπληροφορική, περιπτώσεις χρήσης, ορισμοί, ανοικτά ερευνητικά θέματα, απαιτήσεις για πλατφόρμες διαχείρισης μεγάλης κλίμακας δεδομένων.
- Επεξεργασία Big Data
- Δυνατότητα κλιμάκωσης, αποδοτικότητα, ανεκτικότητα σε λάθη, προγραμματιστικές λύσεις για ανάλυση δεδομένων μεγάλης κλίμακας, MapReduce/Hadoop, Hadoop Distributed File System (HDFS), Spark core, Spark SQL, Spark Machine Learning.
- Επεξεργασία δεδομένων μεγάλης κλίμακας σε πραγματικό χρόνο
- Επεξεργασία ροών δεδομένων, επεξεργασία δεδομένων σε πραγματικό χρόνο, συστήματα διαχείρισης δεδομένων στη κύρια μνήμη, real time streaming στο Spark με DStreams και το νέο structured stream API.
- Ορισμένες αποτελεσματικές προσεγγίσεις στη διαχείριση μεγάλου όγκου δεδομένων
- Η έννοια των Resilient Distributed Data Sets του Spark και η υλοποίησή τους. Το πλαίσιο ανάπτυξης εφαρμογών με RDDs. Υψηλού επιπέδου API του Spark: DataFrames, DataSets, Spark SQL.
- Αρχιτεκτονική cluster computing στο Spark, standalone cluster management, Apache YARN, Apache Mesos, Cloud-based deployments.
- Machine Learning στο Spark, SparkMLlib και Spark ML
- Feature Extraction, Dimensionality Reduction, Principal Components Analysis, Binary and Multiclass Classification, Clustering techniques, Bayesian Inference, Text Analysis, Εισαγωγή στο πλαίσιο του deep learning και εφαρμογών του σε Big Data.

4. ΔΙΔΑΚΤΙΚΕΣ και ΜΑΘΗΣΙΑΚΕΣ ΜΕΘΟΔΟΙ - ΑΞΙΟΛΟΓΗΣΗ

ΤΡΟΠΟΣ ΠΑΡΑΔΟΣΗΣ <i>Πρόσωπο με πρόσωπο, Εξ αποστάσεως εκπαίδευση κ.λπ.</i>	Πρόσωπο με πρόσωπο	
ΧΡΗΣΗ ΤΕΧΝΟΛΟΓΙΩΝ ΠΛΗΡΟΦΟΡΙΑΣ ΚΑΙ ΕΠΙΚΟΙΝΩΝΙΩΝ <i>Χρήση Τ.Π.Ε. στη Διδασκαλία, στην Εργαστηριακή Εκπαίδευση, στην Επικοινωνία με τους φοιτητές</i>	Εξειδικευμένο open-source λογισμικό σε περιβάλλον Linux	
ΟΡΓΑΝΩΣΗ ΔΙΔΑΣΚΑΛΙΑΣ <i>Περιγράφονται αναλυτικά ο τρόπος και μέθοδοι διδασκαλίας.</i>	Δραστηριότητα	Φόρτος Εργασίας Εξαμήνου

<p>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.</p> <p>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</p>	Διαλέξεις	26 x 2= 52 ώρες
	Φροντιστηριακές Ασκήσεις που εστιάζουν στην εφαρμογή	13 x 2= 26 ώρες
	Αυτοτελής Μελέτη	45 ώρες
	Γραπτές Εξετάσεις	2 x 1 = 2 ώρες
	Σύνολο Μαθήματος (25 ώρες φόρτου εργασίας ανά πιστωτική μονάδα)	125 ώρες
<p>ΑΞΙΟΛΟΓΗΣΗ ΦΟΙΤΗΤΩΝ</p> <p>Περιγραφή της διαδικασίας αξιολόγησης Γλώσσα Αξιολόγησης, Μέθοδοι αξιολόγησης, Διαμορφωτική ή Συμπερασματική, Δοκιμασία Πολλαπλής Επιλογής, Ερωτήσεις Σύντομης Απάντησης, Ερωτήσεις Ανάπτυξης Δοκιμίων, Επίλυση Προβλημάτων, Γραπτή Εργασία, Εκθεση / Αναφορά, Προφορική Εξέταση, Δημόσια Παρουσίαση, Εργαστηριακή Εργασία, Κλινική Εξέταση Ασθενούς, Καλλιτεχνική Ερμηνεία, Άλλη / Άλλες</p> <p>Αναφέρονται ρητά προσδιορισμένα κριτήρια αξιολόγησης και εάν και που είναι προσβάσιμα από τους φοιτητές.</p>	Γραπτή τελική εξέταση (100%)	

5. ΣΥΝΙΣΤΩΜΕΝΗ-ΒΙΒΛΙΟΓΡΑΦΙΑ

-Προτεινόμενη Βιβλιογραφία :

-Συναφή επιστημονικά περιοδικά:

- Rajkumar Buyya, Rodroigo N. Calheiros, “Big Data: Principles and Paradigms”, Morgan Kaufmann, 2016
- Jules J Berman, “Principles and Practice of Big Data: Preparing, Sharing and Analyzing Complex Information”, 2nd edition, Academic Press, 2018
- Md. Rezaul Karim, Sridhar Alla,, Scala and Spark for Big Data Analytics, Packt Publishing, 2017
- Ian Foster, Dennis B. Gannon, William Grop, Ewing Lusk, Rich Wolski, Stig Telfer, “Cloud Computing for Science and Engineering (Scientific and Engineering Computation)”, MIT Press, 1st edition, 2017
- Ian Goodfellow, Yoshua Bengio, Aaron Courville, “Deep Learning”, MIT Press, 2016
- Kai Hwang, Min Chen, “Big-Data Analytics”, Wiley, 2017
- Bill Chambers, Matel Zaharia, “Spark: The Definite Guide: Big Data Processing Made Simple”, O’Reilly, 2018