

## HY563 - Προχωρημένα Θέματα Ανάκτησης Πληροφοριών

Διδάσκων: Γιάννης Τζιτζικας

Εξάμηνο: Εαρινό 2018-2019

### Σύντομη Περιγραφή και Περιεχόμενο Μαθήματος

Η αύξηση του όγκου και της ποικιλομορφίας των δεδομένων που παράγονται κάνει επιτακτική την ανάγκη για όλο και πιο ευέλικτους, αποτελεσματικούς και αποδοτικούς τρόπους για *εύρεση, εξερεύνηση και κατανόηση* πληροφοριών, και για αξιοποίησή τους στη διαδικασία *λήψης αποφάσεων*.

Το μάθημα αυτό επικεντρώνεται σε προχωρημένα και σύγχρονα θέματα της περιοχής της Ανάκτησης Πληροφοριών που αφορούν τις παραπάνω προκλήσεις. Το μάθημα προϋποθέτει γνώση των βασικών εννοιών και τεχνικών της Ανάκτησης Πληροφοριών (εκ τούτου αποτελεί συνέχεια του προπτυχιακού μαθήματος HY463 – Ανάκτηση Πληροφοριών).

### Περιεχόμενο

Έμφαση δίνεται στις ακόλουθες ενότητες:

- **(Α) Επεξεργασία Φυσικής Γλώσσας**
  - Υπόβαθρο και σύνδεση με γλωσσολογία, προ-επεξεργασία, μορφολογική, συντακτική, σημασιολογική, πραγματολογική ανάλυση, εργαλεία για επεξεργασία φυσικής γλώσσας
- **(Β) Συστήματα Απάντησης Ερωτήσεων (Question Answering Systems)**
  - Τυπολογία συστημάτων, τεχνικές απάντησης ερωτήσεων και διαλόγου από έγγραφα ή δεδομένα
- **(Γ) Τεχνικές Μηχανικής Μάθησης στην Ανάκτηση Πληροφοριών**
  - Περιβλήματα Λέξεων (Word Embeddings), Τεχνικές Κατάταξης μέσω Εκμάθησης (Learning to Rank)
- **(Δ) Τεχνικές για Εξόρυξη Γνώμης και Συναισθήματος, Θέματα Προκατάληψης, Ποικιλίας και Δικαιοσύνης**
  - Ανάκτηση και ανάλυση γνώμης, σχετικές δραστηριότητες αξιολόγησης, θέματα προκατάληψης, ποικιλίας, δικαιοσύνης, κάλυψης
- **(Ε) Αξιολόγηση Διαδραστικής (Interactive) Ανάκτησης Πληροφοριών**
  - Μέθοδοι, Μέτρα, Συλλογές, Στατιστική Σημαντικότητα

### Εκπαιδευτικοί Στόχοι

Μετά το πέρας του μαθήματος οι φοιτητές πρέπει να:

- γνωρίζουν τις βασικές έννοιες της επεξεργασίας φυσικής γλώσσας και να έχουν εξοικειωθεί με σχετικά εργαλεία
- έχουν κατανοήσει πως λειτουργούν τα συστήματα απάντησης ερωτήσεων (question answering systems) και να έχουν υλοποιήσει ένα σχετικό σύστημα
- έχουν κατανοήσει τεχνικές και αλγόριθμους για εξόρυξη απόψεων (opinion mining)
- έχουν κατανοήσει πως τεχνικές μηχανικής μάθησης μπορούν να αξιοποιηθούν στην Ανάκτηση πληροφοριών
- Να μπορούν να αξιολογήσουν ένα μοντέλο διαδραστικής ανάκτησης πληροφοριών

### Απαιτήσεις Μαθήματος και Τρόπος Βαθμολόγησης

Παρακολούθηση του μαθήματος απαιτεί:

- Παρακολούθηση των διαλέξεων
- Εκπόνηση δύο ή τριών σειρών ασκήσεων)
- Παρουσίαση επιστημονικών άρθρων της τρέχουσας χρονιάς (από τα συνέδρια ECIR και το SIGIR)
- [Συμμετοχή στην εργασία – project - του μαθήματος]
- Τελική γραπτή εξέταση

### Βαθμολογία (δεν έχει οριστικοποιηθεί)

- 40% Ασκήσεις
- 30% Παρουσιάσεις Άρθρων
- 30% Τελική Γραπτή Εξέταση

### Προαπαιτούμενα (για προπτυχιακούς):

HY463 – Συστήματα Ανάκτησης Πληροφοριών

### Εκπαιδευτικό Υλικό

Σχετικά βιβλία και επιστημονικά άρθρα θα αναρτηθούν στην ιστοσελίδα του μαθήματος.

## CS563 - Advanced Topics in Information Retrieval

Instructor: Yannis Tzitzikas (Γιάννης Τζιτζίκας)

Semester: Spring2018-2019

### Brief Description

The increase in the volume and diversity of the data that are being produced necessitates the need for more flexible, effective and efficient ways for finding, exploring and understanding information, as well as for exploiting that information in various decision-making processes.

This course focuses on modern and advanced topics in Information Retrieval that are related to the above challenge. The course requires knowledge of the basic concepts and techniques of Information Retrieval (hence this course can be considered as the continuation of the undergraduate course HY463 – Information Retrieval Systems).

### Contents

- **(A) Natural Language Processing**
  - Connection to Linguistics, Pre-processing, morphological, syntactical, semantic and pragmatic analysis.
- **(B) Question Answering Systems**
  - Typology of QA systems, techniques for question answering and dialogue systems over documents and data
- **(F) Machine Learning Techniques in Information Retrieval**
  - Word Embeddings, Learning to Rank
- **(Δ) Techniques for Sentiment and Opinion Mining, Bias and Fairness in Information Access**
  - Opinion Retrieval, Opinion Analysis, Related Evaluation Activities, Fairness, Diversity/Coverage/Novelty
- **(E) Evaluation of Interactive Information Retrieval**
  - Methods, metrics, collections, statistical significance

### Learning Objectives

After the end of this course the students should have understood

- the basic notions of natural language processing
- how question answering systems work
- how machine learning can be used in the context of IR
- techniques and algorithms for opinion mining

Moreover, at the end of this course the students should have the skills to:

- Carry out various basic NLP tasks
- Design and implement a QA system
- Evaluate an interactive information retrieval system.

### Student Duties and Grading

The duties and the grading system will be finalized based on the number and the background of the students that will enroll to this course. In general, the duties include:

- Attendance of lectures (almost obligatory, graded)
- Assignments (two or three).
- Presentation of two papers in the class (from ECIR or SIGIR)
- [Participation to the project ]
- Successful participation to final written examination

### Grading Formula (not finalized)

- 40% Assignments
- 30% Papers presentation
- 30% Final exam

### Prerequisites (for undergraduate students):

CS463 – Information Retrieval Systems