

## Διαχείριση Περιεχομένου στον Παγκόσμιο Ιστό και Γλωσσικά Εργαλεία

---

**Περιγραφή:** Ο Παγκόσμιος Ιστός αποτελεί το πλέον διαδεδομένο δίκτυο διανομής πληροφορίας για εκατομμύρια χρήστες σε όλο τον κόσμο. Όσο αυξάνεται το μέγεθος της πληροφορίας που διακινείται στον Παγκόσμιο Ιστό και όσο μεγαλώνει ο αριθμός των χρηστών που προσπελάζουν τα δεδομένα του Ιστού για την αποκομιδή πληροφορίας και την εξόρυξη γνώσης από αυτά, τόσο κλιμακώνεται η ανάγκη για αποτελεσματική διαχείριση των δεδομένων Διαδικτύου. Δεδομένου ότι η επικοινωνία του ανθρώπου με τα συστήματα διαχείρισης και διανομής της πληροφορίας του Παγκόσμιου Ιστού πραγματοποιείται με φυσικό τρόπο και κυρίως με τη χρήση φυσικής γλώσσας καθίσταται προφανής η ανάγκη για τη γλωσσική επεξεργασία της πληροφορίας του Παγκόσμιου Ιστού. Στα πλαίσια του μαθήματος θα μελετηθούν διεξοδικά οι εφαρμογές και τα εργαλεία Γλωσσικής Τεχνολογίας που έχουν υιοθετηθεί από τη διεθνή κοινότητα και έχουν ως στόχο τη βελτίωση της πρωτογενούς λειτουργίας των συστημάτων διαχείρισης δυναμικών δεδομένων κειμένου αλλά και την παροχή προηγμένων υπηρεσιών προς τους χρήστες που αλληλεπιδρούν με τα δεδομένα αυτά.

Πιο αναλυτικά, οι θεματικές ενότητες τις οποίες θα καλύψει το μάθημα συνοψίζονται στις ακόλουθες και μπορούν να ιδωθούν ως το οργανόγραμμα του μαθήματος:

### Περιεχόμενο Μαθήματος:

*Εισαγωγικές Έννοιες*

*Γλώσσα και Αναζήτηση Πληροφορίας*

*Text Mining*

*N-grams*

*Data Mining*

*Opinion Mining*

*Δυναμική Εξαγωγή Πληροφορίας από τα Δεδομένα Διαδικτύου*

*Οντολογίες*

*Question-Answering*

*Ο πολυγλωσσικός Παγκόσμιος Ιστός*

*Επισκόπηση μαθήματος και έρευνα σε εξέλιξη*

**Προτεινόμενο Σύγγραμμα:** "Υπολογιστική γλωσσολογία", Χριστίνα Αλεξανδρή, Εκδόσεις Παπασωτηρίου

### **Απαιτήσεις Μαθήματος:**

**Διαλέξεις:** Στόχος των διαλέξεων είναι η θεωρητική θεμελίωση των ζητημάτων που άπτονται της οργάνωσης, της διαχείρισης και της επεξεργασίας του περιεχομένου του Παγκόσμιου Ιστού, δίνοντας έμφαση στη χρησιμότητα και στις εφαρμογές εργαλείων Επεξεργασίας Φυσικής Γλώσσας.

**Φροντιστήριο:** Σκοπός του φροντιστηρίου είναι η πρακτική εξάσκηση σε εφαρμογές Επεξεργασίας Φυσικής Γλώσσας για την αποδοτική διαχείριση και την αποτελεσματική πρόσβαση στα δεδομένα του Παγκόσμιου Ιστού. Στο φροντιστήριο θα δίνονται διευκρινίσεις για το project του μαθήματος, το οποίο είναι υποχρεωτικό.

**Εξέταση Μαθήματος:** για την επιτυχή ολοκλήρωση του μαθήματος προβλέπεται συμμετοχή στις διαλέξεις και τα φροντιστήρια του μαθήματος, εκπόνηση ενός project και συμμετοχή σε προφορική εξέταση. Η επίδοσή σας στο project συμμετέχει κατά 80% στη διαμόρφωση του τελικού βαθμού και η επίδοσή σας στην προφορική εξέταση κατά 20%.

**Επικοινωνία:** απορίες, διευκρινίσεις και οτιδήποτε χρειαστείτε σχετικά με το μάθημα μπορούν να κοινοποιούνται στο forum του μαθήματος στο [my.ceid](http://my.ceid) ή εναλλακτικά στις διευθύνσεις [kozanid@ceid.upatras.gr](mailto:kozanid@ceid.upatras.gr) και [simaki@ceid.upatras.gr](mailto:simaki@ceid.upatras.gr). Παρακαλώ στο subject του email να χρησιμοποιείται το ακρώνυμο του μαθήματος (Web NLP) και το AM σας.

### **Διδάσκων:**

*Χριστοδουλάκης Δημήτριος*

### **Υπεύθυνοι Διδακτορικοί:**

*Κοζανίδης Λευτέρης*

*Σιμάκη Βάσω*