

Γλωσσική Τεχνολογία
Ακαδημαϊκό Έτος 2012-2013
Τμήμα Μηχανικών Η/Υ και Πληροφορικής

Γλωσσική Τεχνολογία

Η γλωσσική τεχνολογία ασχολείται με την ανάπτυξη υπολογιστικών μοντέλων για την επεξεργασία πληροφορίας εκφρασμένης σε φυσική γλώσσα από συστήματα πληροφορικής. Σαν φυσική γλώσσα ορίζεται το μέσο καταγραφής και ανταλλαγής πληροφορίας που χρησιμοποιείται για την επικοινωνία μεταξύ ανθρώπων (σε αντίθεση με τις τεχνητές γλώσσες που χρησιμοποιούνται για την επικοινωνία μεταξύ ανθρώπου – μηχανής). Η γλωσσική τεχνολογία αφορά τόσο την αυτόματη ανάλυση όσο και την παραγωγή γραπτών ή προφορικών εκφράσεων φυσικής γλώσσας. Οι γλωσσικές δυνατότητες που ενσωματώνονται σε υπολογιστικές εφαρμογές, οι οποίες χρησιμοποιούνται σε συστήματα πληροφορικής και επικοινωνιακής τεχνολογίας, αποτελούν σημαντικές προσθήκες, καθώς προσφέρουν νέους τρόπους επικοινωνίας που επιτρέπουν την καλύτερη συμβίωση του ανθρώπου με την τεχνολογία.

Το μάθημα στοχεύει στην επισκόπηση των διαθέσιμων τεχνικών επεξεργασίας φυσικής γλώσσας και στην μελέτη εκείνων των τεχνικών που χρησιμοποιούνται σήμερα από δημοφιλείς εφαρμογές, όπως μηχανές αναζήτησης, υπηρεσίες παγκόσμιου ιστού, εξόρυξη δεδομένων, εξαγωγή πληροφορίας, κατηγοριοποίηση εγγράφων, κ.α. 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:

1. Εισαγωγή στην γλωσσική τεχνολογία
2. Βασικές έννοιες
3. Μορφολογία
4. Συντακτική ανάλυση
5. Σημασιολογική ανάλυση
6. Ανάλυση πραγματείας
7. Ανάκτηση πληροφορίας
8. Εξαγωγή πληροφορίας
9. Σημασιολογική αποσαφήνιση
10. Αυτόματη δημιουργία περιλήψεων
11. Θεματική κατηγοριοποίηση
12. Μηχανική μετάφραση
13. Γλωσσική επεξεργασία πολυμεσικών δεδομένων

Προτεινόμενο Σύγγραμμα: «Η τεχνολογία της πληροφορίας στην επεξεργασία φυσικής γλώσσας – προβλήματα επεξεργασίας φυσικής γλώσσας», Κωνσταντίνος Τ. Φράγγος, Αναστάσιος Π. Κουτσούκος, εκδόσεις ΜΥΡΜΙΔΟΝΕΣ, 2010, ISBN: 978-960-99279-0-1.

Απαιτήσεις Μαθήματος:

Διαλέξεις: Στόχος των διαλέξεων είναι η θεωρητική θεμελίωση των ζητημάτων που άπτονται της ενσωμάτωσης γλωσσικής τεχνολογίας σε υπολογιστικές εφαρμογές, μέσω της παρουσίασης εδραιωμένων αλγορίθμων και τεχνικών επεξεργασίας που αφορούν την επεξεργασία και παραγωγή φυσικής γλώσσας.

Φροντιστήριο: Σκοπός του φροντιστηρίου είναι η πρακτική εξάσκηση σε εφαρμογές επεξεργασίας φυσικής γλώσσας. 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 70% στη διαμόρφωση του τελικού βαθμού και η επίδοσή σας στην προφορική εξέταση κατά 30%. Η βαθμολόγηση του project θα γίνει σε 2 στάδια: η πρώτη εξέταση (πρόοδος) θα γίνει σε ημερομηνία που θα ανακοινωθεί στο forum του μαθήματος και θα αφορά το 30% του βαθμού του project, ενώ η δεύτερη εξέταση θα αφορά το υπόλοιπο 70% του βαθμού του project, και θα διεξαχθεί στην τελική εξέταση του μαθήματος, μαζί με την προφορική εξέταση.

Επικοινωνία: απορίες, διευκρινίσεις και οτιδήποτε χρειαστείτε σχετικά με το μάθημα μπορούν να κοινοποιούνται στο forum του μαθήματος στο my.ceid ή εναλλακτικά στις διευθύνσεις petasis@ceid.upatras.gr και simaki@ceid.upatras.gr. Παρακαλώ στο θέμα του e-mail να χρησιμοποιείται το ακρώνυμο του μαθήματος (LT-NLP) και το ΑΜ σας.

Διδάσκοντες:

Χριστοδουλάκης Δημήτριος
Πετάσης Γεώργιος

Υπεύθυνες Μεταπτυχιακοί:

Σιμάκη Βάσω
Τζέκου Βιβή