



## Προπτυχιακές και μεταπτυχιακές εργασίες Σεπτέμβριος 2009

Το Εργαστήριο Τεχνολογίας Γνώσεων και Λογισμικού του ΕΚΕΦΕ «Δημόκριτος» προσφέρει τις παρακάτω εργασίες, οι οποίες είναι χωρισμένες σε θεματικές ενότητες, σύμφωνα με τις Δράσεις του Εργαστηρίου. Οι περιγραφές των εργασιών συμπεριλαμβάνουν τα ονόματα των ερευνητών που είναι υπεύθυνοι για την επίβλεψη και την εσωτερική αξιολόγησή τους. Όλες οι εργασίες θα εκπονηθούν σε συνεργασία με μέλη ΔΕΠ ΑΕΙ και ΤΕΙ.

### **ΔΡΑΣΗ 1. Συστήματα Εξαγωγής Πληροφορίας και Εξόρυξης Γνώσης**

Στόχος είναι η έρευνα και ανάπτυξη σε θέματα φιλτραρίσματος και εξαγωγής πληροφορίας από βάσεις κειμένων (information filtering and extraction) καθώς και σε θέματα εξόρυξης γνώσης από βάσεις κειμένων ή δεδομένων (text mining or data mining). Για την επίτευξη του στόχου αυτού αξιοποιούνται μέθοδοι, τεχνικές και εργαλεία από τις τεχνολογίες Επεξεργασίας Φυσικής Γλώσσας, Εξόρυξης Γνώσης, και Μηχανικής Μάθησης.

#### **1.1 Μηχανική μάθηση ταξινομητών από ταξινομημένα και μη ταξινομημένα δεδομένα εκπαίδευσης σε περιπτώσεις ανισο-κατανομημένων κατηγοριών**

**Επίπεδο:** Προπτυχιακό ή μεταπτυχιακό.

**Απαιτούμενα προσόντα:** Γνώσεις προγραμματισμού, κατά προτίμηση σε Java.

**Επιθυμητά προσόντα:** Γνώσεις μηχανικής μάθησης.

**Περιγραφή:** Δύο βασικά προβλήματα στην εκπαίδευση ταξινομητών, ιδιαίτερα σε εφαρμογές ταξινόμησης περιεχομένου, είναι (α) η ακραία ανισοκατανομή των κατηγοριών, συνήθως εις βάρος της ενδιαφέρουσας κατηγορίας, και (β) η ύπαρξη λίγων κατανομημένων παραδειγμάτων και πολλών μη-κατανομημένων. Για την αντιμετώπιση αυτού του προβλήματος έχει αναπτυχθεί στο εργαστήριο μέθοδος μηχανικής μάθησης που εστιάζει στην κατηγορία που μας ενδιαφέρει και προσπαθεί να εκτιμήσει την κατηγορία των μη-κατανομημένων παραδειγμάτων. Στα πλαίσια αυτής της εργασίας θα υλοποιηθούν και θα αξιολογηθούν βελτιώσεις της υπάρχουσας μεθόδου, σε δεδομένα ιατρικής διάγνωσης και ταξινόμησης περιεχομένου.

**Επιβλέπων Ερευνητής:**

Γιώργος Παλιούρας, τηλ. 21-06503158, email: [paliourg@iit.demokritos.gr](mailto:paliourg@iit.demokritos.gr)

#### **1.2 Μηχανική μάθηση ακολουθιακών ταξινομητών με βάση απλούς ταξινομητές δεδομένων**

**Επίπεδο:** Προπτυχιακό ή μεταπτυχιακό.

**Απαιτούμενα προσόντα:** Γνώσεις προγραμματισμού.

**Επιθυμητά προσόντα:** Γνώσεις μηχανικής μάθησης.

**Περιγραφή:** Η ταξινόμηση ακολουθιακών δεδομένων, όπως κείμενα, ηχητικά σήματα, χρονοσειρές, έχει αντιμετωπιστεί κυρίως με ειδικά παραγωγικά μοντέλα (generative models), όπως τα Μαρκοβιανά, στη βιβλιογραφία. Πρόσφατα έχουν όμως προταθεί και μέθοδοι που μαθαίνουν μοντέλα δεσμευμένων πιθανοτήτων, όπως κάνουν και οι απλοί ταξινομητές, π.χ. Conditional Random Fields (CRFs). Στα πλαίσια της εργασίας θα μελετηθεί η δυνατότητα γενίκευσης αυτών των μοντέλων, ώστε να ενσωματώνουν γνωστές μεθόδους μάθησης ταξινομητών. Στόχος της εργασίας είναι η υλοποίηση μίας πρώτης τέτοιας μεθόδου και η σύγκρισή της με τα CRFs.

**Επιβλέπων Ερευνητής:**

Γιώργος Παλιούρας, τηλ. 21-06503158, email: [paliourg@iit.demokritos.gr](mailto:paliourg@iit.demokritos.gr)

### **1.3 Σύστημα ενημέρωσης και ανάλυσης δικτύου βιβλιογραφικών αναφορών από τον Ιστό με τεχνικές εξαγωγής πληροφορίας και ταιριάσματος όμοιων εγγραφών**

**Επίπεδο:** Προπτυχιακό ή μεταπτυχιακό.

**Απαιτούμενα προσόντα:** Γνώσεις προγραμματισμού, κατά προτίμηση σε Java και διαχείρισης βάσεων δεδομένων.

**Επιθυμητά προσόντα:** Γνώσεις μηχανικής μάθησης ή/και γλωσσικής τεχνολογίας ή/και διαδικτυακού προγραμματισμού.

**Περιγραφή:** Τα κοινωνικά δίκτυα (social networks) είναι μία από τις πιο ενδιαφέρουσες εφαρμογές για την εξαγωγή πληροφορίας και την εξόρυξη γνώσης. Το δίκτυο των βιβλιογραφικών αναφορών είναι ένα τυπικό παράδειγμα κοινωνικού δικτύου με ενδιαφέροντα χαρακτηριστικά. Στο εργαστήριο έχει αναπτυχθεί σύστημα ενημέρωσης ενός τέτοιου δικτύου με εξαγωγή πληροφορίας από το διαδίκτυο. Στόχος της εργασίας είναι η βελτίωση και επέκταση του συστήματος με νέες μεθόδους εξαγωγής πληροφορίας και ανάλυσης του δικτύου που προκύπτει. Η εξαγωγή πληροφορίας εφαρμόζεται στα αποτελέσματα γνωστών βιβλιογραφικών βάσεων, καθώς και μηχανών αναζήτησης. Η ανάλυση του κοινωνικού δικτύου γίνεται με μεθόδους εξόρυξης γνώσης. Σημαντικό ρόλο στη διαδικασία παίζει το ταίριασμα όμοιων εγγραφών (record linkage), που ανακτώνται από διαφορετικές πηγές.

**Επιβλέπων Ερευνητής:**

Γιώργος Παλιούρας, τηλ. 21-06503158, email: [paliourg@iit.demokritos.gr](mailto:paliourg@iit.demokritos.gr)

### **1.4 Επέκταση πύλης ενημέρωσης με μεθόδους αυτόματης συντήρησης των κανόνων εξαγωγής πληροφορίας από τις ειδησεογραφικές πηγές**

**Επίπεδο:** Προπτυχιακό ή μεταπτυχιακό.

**Απαιτούμενα προσόντα:** Γνώσεις προγραμματισμού, κατά προτίμηση σε Java.

**Επιθυμητά προσόντα:** Γνώσεις μηχανικής μάθησης ή/και γλωσσικής τεχνολογίας ή/και διαδικτυακού προγραμματισμού.

**Περιγραφή:** Στο Εργαστήριο έχει αναπτυχθεί μία εξατομικευμένη πύλη ενημέρωσης από τον Παγκόσμιο Ιστό η οποία συλλέγει ειδήσεις από διάφορες πηγές και τις παρέχει εξατομικευμένα στους χρήστες της. Για το σκοπό αυτό, η πύλη χρησιμοποιεί κανόνες εξαγωγής της πληροφορίας από τις πηγές. Οι κανόνες αυτοί υλοποιούνται ως κανονικές εκφράσεις, οι οποίες εφαρμόζονται στο κείμενο των ιστοσελίδων και εξάγουν την ενδιαφέρουσα πληροφορία, π.χ. τον τίτλο της είδησης. Ένα από τα βασικά προβλήματα είναι η συντήρηση αυτών των κανόνων για πολλές πηγές οι οποίες μεταβάλλουν τον τρόπο παρουσίασης της πληροφορίας στον χρόνο. Στόχος της εργασίας είναι η ανάπτυξη μεθόδων αυτόματου εντοπισμού αυτών των αλλαγών και ενημέρωσης των κανόνων εξαγωγής πληροφορίας με μεθόδους μηχανικής μάθησης.

**Επιβλέποντες Ερευνητές:**

Γιώργος Παλιούρας, τηλ. 21-06503158, email: [paliourg@iit.demokritos.gr](mailto:paliourg@iit.demokritos.gr)

Δημήτριος Βογιατζής, τηλ. 21-06503194, email: [dimitrv@iit.demokritos.gr](mailto:dimitrv@iit.demokritos.gr)

### **1.5 Μέθοδοι μείωσης μεγέθους γράφου ν-γραμμάτων με βάση τη χρησιμότητα**

**Επίπεδο:** Μεταπτυχιακό.

**Απαιτούμενα προσόντα:** Γνώσεις Τεχνητής Νοημοσύνης.

**Επιθυμητά προσόντα:** Γνώσεις Μηχανικής Μάθησης, εξοικείωση με γράφους και αλγορίθμους σε γράφους.

**Περιγραφή:** Στο Εργαστήριο έχουμε αναπτύξει μία μέθοδο αναπαράστασης κειμένων με γράφους ν-γραμμάτων. Η αναπαράσταση αυτή, αν και πλούσια σε πληροφορία, έχει την αδυναμία του μεγάλου μεγέθους, που αυξάνει την πολυπλοκότητα εφαρμογής αλγορίθμων πάνω της. Σκοπός της εργασίας είναι να μελετηθούν μέθοδοι μείωσης του μεγέθους του γράφου, αντίστοιχες αυτών που χρησιμοποιούνται σε διανυσματικές αναπαραστάσεις κειμένων (π.χ. PageRank, PCA), με παράλληλη διατήρηση της ωφέλιμης πληροφορίας.

**Επιβλέποντες Ερευνητές:**

Γιώργος Παλιούρας, τηλ. 21-06503158, email: [paliourg@iit.demokritos.gr](mailto:paliourg@iit.demokritos.gr)

Γιώργος Γιαννακόπουλος, email: [ggianna@disi.unitn.it](mailto:ggianna@disi.unitn.it)

**1.6 Μάθηση κανόνων ταύτισης όμοιων εγγραφών**

**Επίπεδο:** Προπτυχιακό ή μεταπτυχιακό.

**Απαιτούμενα προσόντα:** Γνώσεις προγραμματισμού, κατά προτίμηση σε Java.

**Επιθυμητά προσόντα:** Γνώσεις μηχανικής μάθησης ή/και γλωσσικής τεχνολογίας.

**Περιγραφή:** Στο εργαστήριο έχει υλοποιηθεί πειραματική πλατφόρμα σύγκρισης όμοιων εγγραφών σε βάσεις. Η πλατφόρμα περιλαμβάνει ποικιλία μετρικών απόστασης μεταξύ λέξεων (π.χ. απόσταση Levenshtein), κάποιες από τις οποίες βασίζονται σε μορφολογικά και άλλες σε σημασιολογικά χαρακτηριστικά. Στόχος της εργασίας είναι η ανάπτυξη μεθόδων μάθησης για τον συνδυασμό των αποτελεσμάτων σύγκρισης των διαφόρων πεδίων μίας εγγραφής με στόχο την λήψη της τελικής απόφασης σχετικά με το αν δύο εγγραφές είναι όμοιες ή όχι. Συνήθεις τομείς εφαρμογής αυτών των μεθόδων είναι η αποσαφήνιση νοήματος λέξεων (word sense disambiguation), η ταυτοποίηση υπηρεσιών νοημόνων πρακτόρων και η ταύτιση όμοιων εγγραφών σε συστήματα ολοκλήρωσης πληροφορίας (information integration) στον Ιστό. Επίσης, θα δοκιμαστεί μία νέα οικογένεια μεθόδων ασαφούς ταύτισης συμβολοσειρών (fuzzy string matching) βασισμένη σε γράφους ν-γραμμμάτων.

**Επιβλέποντες Ερευνητές:**

Γιώργος Παλιούρας, τηλ. 21-06503158, email: [paliourg@iit.demokritos.gr](mailto:paliourg@iit.demokritos.gr)

Γιώργος Γιαννακόπουλος, email: [ggianna@disi.unitn.it](mailto:ggianna@disi.unitn.it)

**1.7 Αυτόματη δημιουργία επερωτήσεων για την ανάκτηση παραδειγμάτων από τον Ιστό, με στόχο την εκπαίδευση ταξινομητών ιστοσελίδων ακολουθώντας διαδικασία ενεργής μάθησης**

**Επίπεδο:** Προπτυχιακό ή μεταπτυχιακό.

**Απαιτούμενα προσόντα:** Πολύ καλές γνώσεις προγραμματισμού, κατά προτίμηση σε Java.

**Επιθυμητά προσόντα:** Γνώσεις μηχανικής μάθησης ή/και γλωσσικής τεχνολογίας.

**Περιγραφή:** Η δημιουργία αποδοτικών ταξινομητών κειμένου με χρήση μηχανικής μάθησης, στηρίζεται στη σωστή επιλογή δεδομένων εκπαίδευσης, δηλαδή κειμένων. Στο Εργαστήριο έχει αναπτυχθεί μέθοδος δημιουργίας και εμπλουτισμού των δεδομένων εκπαίδευσης για ταξινομητές εγγράφων, ανακτώντας κατάλληλα παραδείγματα από τον Ιστό. Η μέθοδος αυτή ακολουθεί τη διαδικασία ενεργής μάθησης (active learning) σε συνδυασμό με τον σχηματισμό επερωτήσεων για μηχανές αναζήτησης. Στόχος της εργασίας είναι η αυτοματοποίηση της δημιουργίας των επερωτήσεων προς τις μηχανές αναζήτησης και η αξιολόγηση διάφορων προσεγγίσεων σε αυτό το πρόβλημα.

**Επιβλέποντες Ερευνητές:**

Γιώργος Παλιούρας, τηλ. 21-06503158, email: [paliourg@iit.demokritos.gr](mailto:paliourg@iit.demokritos.gr)

Κώστας Χανδρινός, email: [chandrinos@gmail.com](mailto:chandrinos@gmail.com)

**1.8 Μάθηση μοντέλου πρόβλεψης της δημοτικότητας μίας κριτικής**

**Επίπεδο:** Προπτυχιακό ή μεταπτυχιακό.

**Απαιτούμενα προσόντα:** Γνώσεις προγραμματισμού.

**Επιθυμητά προσόντα:** Γνώσεις μηχανικής μάθησης ή/και γλωσσικής τεχνολογίας.

**Περιγραφή:** Όλο και πιο συχνά στο διαδίκτυο οι χρήστες αξιολογούν προϊόντα και δημοσιεύουν την άποψή τους. Επιπλέον, η άποψή τους αξιολογείται από άλλους χρήστες. Αντικείμενο αυτής της εργασίας είναι η μάθηση των χαρακτηριστικών και κανόνων που προκαλούν τη δημοτικότητα μία δημόσιας κριτικής και ενός κριτή αντίστοιχα. Για το σκοπό αυτό, θα χρησιμοποιηθούν χαρακτηριστικά που σχετίζονται τόσο με το περιεχόμενο της κριτικής (εξαγωγή πληροφορίας από κείμενο) όσο και με τις συνθήκες στις οποίες δημοσιεύθηκε. Ως περίπτωση μελέτης θα εξετασθούν οι κριτικές ταινιών που υπάρχουν σε μεγάλους σχετικούς ιστοτόπους.

**Επιβλέποντες Ερευνητές:**

Γιώργος Παλιούρας, τηλ. 21-06503158, email: [paliourg@iit.demokritos.gr](mailto:paliourg@iit.demokritos.gr)

Κώστας Χανδρινός, email: [chandrinos@gmail.com](mailto:chandrinos@gmail.com)

**1.9 Κατηγοριοποίηση ιστολογίων (blogs) χρησιμοποιώντας τεχνικές ημι-επιβλεπόμενης μάθησης**

**Επίπεδο:** Προπτυχιακό ή μεταπτυχιακό.

**Απαιτούμενα προσόντα:** Γνώσεις προγραμματισμού.

**Επιθυμητά προσόντα:** Γνώσεις μηχανικής μάθησης ή/και ανάκτησης πληροφορίας.

**Περιγραφή:** Οι τεχνικές της Μηχανικής Μάθησης έχουν χρησιμοποιηθεί για διάφορες εφαρμογές πρόσβασης στην πληροφορία, όπως η κατηγοριοποίηση, η ομαδοποίηση και η εξόρυξη πληροφοριών. Η απόκτηση των επισημειωμένων δεδομένων που είναι αναγκαία για να εφαρμοστούν οι τεχνικές επιβλεπόμενης μάθησης, αποτελεί μια σημαντική πρόκληση για αυτές τις εφαρμογές, ειδικά για πολύ μεγάλες συλλογές δεδομένων. Μία από τις προσεγγίσεις για την αντιμετώπιση αυτού του προβλήματος είναι η ημι-επιβλεπόμενη μάθηση. Η προσέγγιση αυτή αποσκοπεί στην μείωση του κόστους της επισημείωσης των δεδομένων λαμβάνοντας υπ' όψη τόσο τα επισημειωμένα όσο και τα μη-επισημειωμένα διαθέσιμα δεδομένα. Στόχος της παρούσας εργασίας είναι η εφαρμογή και επέκταση τεχνικών ημι-επιβλεπόμενης μάθησης σε συλλογές ιστολογίων (blogs). Η πειραματική αξιολόγηση θα γίνει σε δεδομένα που συλλέγονται στο Εργαστήριο.

**Επιβλέποντες Ερευνητές:**

Αναστασία Κριθαρά, email: [akrithara@iit.demokritos.gr](mailto:akrithara@iit.demokritos.gr)

Βαγγέλης Καρκαλέτσης τηλ. 21-06503197 email: [vangelis@iit.demokritos.gr](mailto:vangelis@iit.demokritos.gr)

**1.10 Σύστημα φιλτραρίσματος ανεπιθύμητων ηλεκτρονικών μηνυμάτων για ομάδες χρηστών με χρήση μηχανικής μάθησης**

**Επίπεδο:** Προπτυχιακό ή μεταπτυχιακό.

**Απαιτούμενα προσόντα:** Γνώσεις προγραμματισμού, κατά προτίμηση σε Java.

**Επιθυμητά προσόντα:** Γνώσεις μηχανικής μάθησης.

**Περιγραφή:** Η ανεπιθύμητη ηλεκτρονική αλληλογραφία (spam filtering) έχει γίνει ένα από τα σημαντικότερα προβλήματα που σχετίζονται με τη χρήση του Διαδικτύου. Η εκπαίδευση ταξινομητών των μηνυμάτων (φίλτρων) με χρήση μηχανικής μάθησης έχει οδηγήσει πρόσφατα σε μερική αντιμετώπιση του προβλήματος. Στο Εργαστήριο έχει αναπτυχθεί σύστημα εκπαίδευσης φίλτρων για ανεπιθύμητη αλληλογραφία, το οποίο χρησιμοποιεί για την εκπαίδευση ένα σώμα ανεπιθύμητων μηνυμάτων και τα κανονικά μηνύματα που λαμβάνει ένας χρήστης ή μία ομάδα χρηστών. Στόχος της εργασίας είναι η βελτίωση της μεθόδου κατασκευής φίλτρων για ομάδες χρηστών, π.χ. τα μέλη του Εργαστηρίου. Η νέα μέθοδος θα εφαρμόζει μία νέα προσέγγιση στον τρόπο που λαμβάνονται υπόψη τα φίλτρα «συναδέλφων» από το ατομικό φίλτρο του κάθε χρήστη.

**Επιβλέποντες Ερευνητές:**

Γιώργος Παλιούρας, τηλ. 21-06503158, email: [paliourg@iit.demokritos.gr](mailto:paliourg@iit.demokritos.gr)

Άρης Κοσμόπουλος, τηλ. 21-06503216, email: [akosmo@iit.demokritos.gr](mailto:akosmo@iit.demokritos.gr)

**1.11 Μέθοδος επαυξητικής και ενεργής μάθησης για φίλτρα ανεπιθύμητων ηλεκτρονικών μηνυμάτων**

**Επίπεδο:** Προπτυχιακό ή μεταπτυχιακό.

**Απαιτούμενα προσόντα:** Γνώσεις προγραμματισμού, κατά προτίμηση σε Java.

**Επιθυμητά προσόντα:** Γνώσεις μηχανικής μάθησης.

**Περιγραφή:** Τα εκπαιδευόμενα φίλτρα ανεπιθύμητης ηλεκτρονικής αλληλογραφίας (spam filtering) χρειάζεται να βελτιώνονται συνεχώς κατά τη χρήση τους, προσαρμοζόμενα στις ιδιαιτερότητες του χρήστη

αλλά και στη συνεχώς μεταβαλλόμενη φύση των ανεπιθύμητων μηνυμάτων. Γι' αυτό το λόγο απαιτείται η ανάπτυξη μεθόδων επαυξητικής μάθησης (incremental learning), οι οποίες ταυτόχρονα δεν θα απαιτούν σημαντική επίβλεψη από τον χρήστη, ώστε να μην τον/την επιβαρύνουν. Γι' αυτό το σκοπό μελετούνται στη βιβλιογραφία μέθοδοι ενεργής μάθησης (active learning). Το Σύστημα φιλτραρίσματος ανεπιθύμητης αλληλογραφίας που έχει αναπτυχθεί στο Εργαστήριο επεκτάθηκε πρόσφατα με μία μέθοδο ενεργής μάθησης. Στα πλαίσια της εργασίας, η μέθοδος αυτή θα βελτιωθεί και θα συνδυαστεί με δυνατότητες επαυξητικής μάθησης. Η βελτιωμένη έκδοση του συστήματος θα ελεγχθεί σε δεδομένα αξιολόγησης που έχουν συλλεχθεί στο Εργαστήριο, αλλά και αυτά που έχουν χρησιμοποιηθεί για διαγωνισμούς σε διεθνή συνέδρια.

#### **Επιβλέποντες Ερευνητές:**

Γιώργος Παλιούρας, τηλ. 21-06503158, email: [paliourg@iit.demokritos.gr](mailto:paliourg@iit.demokritos.gr)

Άρης Κοσμοπούλος, τηλ. 21-06503216, email: [akosmo@iit.demokritos.gr](mailto:akosmo@iit.demokritos.gr)

### **1.12 Εξαγωγή πληροφορίας από έγγραφα μέσω επαγωγικής εξαγωγής γραμματικών ανεξάρτητων από συμφραζόμενα (context-free grammars)**

**Επίπεδο:** Προπτυχιακό ή μεταπτυχιακό.

**Απαιτούμενα προσόντα:** Γνώσεις προγραμματισμού, κατά προτίμηση σε C/C++, Tcl.

**Επιθυμητά προσόντα:** Γνώσεις μηχανικής μάθησης ή/και γλωσσικής τεχνολογίας.

**Περιγραφή:** Το Εργαστήριο έχει αναπτύξει έναν αλγόριθμο επαγωγικής εξαγωγής γραμματικών ανεξάρτητων από συμφραζόμενα. Σκοπός της εργασίας αυτής είναι η αξιολόγηση του αλγορίθμου αυτού στην εργασία της εξαγωγής πληροφορίας από κείμενα της Ελληνικής και της Αγγλικής γλώσσας καθώς και η σύγκριτική αξιολόγηση του αλγορίθμου με έτοιμα συστήματα που βασίζονται σε άλλες τεχνικές μηχανικής μάθησης.

#### **Επιβλέποντες Ερευνητές:**

Γεώργιος Πετάσης, email: [petasis@iit.demokritos.gr](mailto:petasis@iit.demokritos.gr)

Γιώργος Παλιούρας, τηλ. 21-06503158, email: [paliourg@iit.demokritos.gr](mailto:paliourg@iit.demokritos.gr)

### **1.13 Ανάπτυξη συντακτικού αναλυτή γραμματικών ανεξάρτητων από συμφραζόμενα (context-free grammars), σε συνδυασμό με μεθόδους μάθησης τέτοιων γραμματικών**

**Επίπεδο:** Προπτυχιακό ή μεταπτυχιακό.

**Απαιτούμενα προσόντα:** Γνώσεις προγραμματισμού, κατά προτίμηση σε C/C++, Tcl.

**Επιθυμητά προσόντα:** Γνώσεις μηχανικής μάθησης ή/και γλωσσικής τεχνολογίας.

**Περιγραφή:** Το Εργαστήριο έχει αναπτύξει έναν αλγόριθμο επαγωγικής εξαγωγής γραμματικών ανεξάρτητων από συμφραζόμενα. Κατά την εφαρμογή του αλγορίθμου αυτού σε διάφορα προβλήματα, προκύπτουν σημαντικές δυσκολίες εξαιτίας της ακαταλληλότητας των υπάρχοντων συντακτικών αναλυτών ή/και της υπολογιστικής τους πολυπλοκότητας. Σκοπός της εργασίας αυτής είναι η ανάπτυξη ενός συντακτικού αναλυτή προσαρμοσμένου στις ανάγκες των μεθόδων μάθησης γραμματικών και ο συνδυασμός των δύο μεθόδων σε ένα κοινό σύστημα. Το σύστημα αυτό θα αξιολογηθεί σε εφαρμογές επεξεργασίας φυσικής γλώσσας που μελετώνται στο εργαστήριο.

#### **Επιβλέποντες Ερευνητές:**

Γεώργιος Πετάσης, email: [petasis@iit.demokritos.gr](mailto:petasis@iit.demokritos.gr)

Γιώργος Παλιούρας, τηλ. 21-06503158, email: [paliourg@iit.demokritos.gr](mailto:paliourg@iit.demokritos.gr)

### **1.14 Εφαρμογή τεχνικών αλλαγής θέματος σε έγγραφα (κείμενα, ιστοσελίδες)**

**Επίπεδο:** Προπτυχιακό ή μεταπτυχιακό.

**Απαιτούμενα προσόντα:** Πολύ καλές γνώσεις προγραμματισμού, κατά προτίμηση σε C/C++, Tcl.

**Επιθυμητά προσόντα:** Γνώσεις μηχανικής μάθησης ή/και γλωσσικής τεχνολογίας, οικειότητα με τεχνικές Τεχνητής Νοημοσύνης.

**Περιγραφή:** Οι τεχνικές αλλαγής θέματος (topic change identification) βοηθούν στην εύρεση των

διαφορετικών θεμάτων τα οποία εμφανίζονται μέσα σε ένα κείμενο. Η εν λόγω εύρεση είναι δυνατό να βασίζεται τόσο στην οπτική αναπαράσταση της εκάστοτε ιστοσελίδας όσο και σε επιπλέον πληροφορία που σχετίζεται με το περιεχόμενο αυτής. Τμήματα κειμένου τα οποία αναφέρονται στο ίδιο θέμα είναι πιθανό να αναφέρονται στις ίδιες οντότητες (named entities) ή/και έννοιες (concepts). Μια τέτοια πληροφορία μπορεί να αξιοποιηθεί κατά τη διαδικασία της εξαγωγής πληροφορίας (information extraction), όπου μας ενδιαφέρει η εύρεση και εξαγωγή των σχέσεων μεταξύ οντοτήτων οι οποίες εμφανίζονται στα κείμενα. Στόχος της εργασίας είναι η εξέταση και συνδυασμός τμηματοποίησης κειμένων βάση της οπτικής τους αναπαράστασης σε συνδυασμό με τεχνικές εξαγωγής σχέσεων από κείμενα και ιστοσελίδες βάσει του περιεχομένου τους με όσο το δυνατό αυτόματο τρόπο.

**Επιβλέποντες Ερευνητές:**

Παυλίνα Φράγκου, email: [fragou@iit.demokritos.gr](mailto:fragou@iit.demokritos.gr)

Βαγγέλης Καρκαλέτσης τηλ. 21-06503197 email: [vangelis@iit.demokritos.gr](mailto:vangelis@iit.demokritos.gr)

**1.15 Συνδυασμός μετα-δεδομένων που περιγράφουν το περιεχόμενο ιστο-σελίδων**

**Επίπεδο:** Προπτυχιακό ή μεταπτυχιακό.

**Απαιτούμενα προσόντα:** Πολύ καλές γνώσεις προγραμματισμού, κατά προτίμηση σε Java.

**Επιθυμητά προσόντα:** Γνώσεις τεχνολογιών διαδικτύου, γλωσσικής τεχνολογίας.

**Περιγραφή:** Με την εκρηκτική ανάπτυξη του Παγκόσμιου Ιστού πρακτικά ο κάθε χρήστης έχει τη δυνατότητα να αναρτήσει εκεί περιεχόμενο με βάση τις γνώσεις και τα ενδιαφέροντά του. Η επισημείωση (annotation) αυτού του περιεχομένου με επιπλέον δεδομένα (μετα-δεδομένα), όπως για παράδειγμα, για τον συγγραφέα του περιεχομένου, τη θεματική κατηγορία του περιεχομένου, τις κατηγορίες χρηστών στις οποίες απευθύνεται, κ.ά., μπορεί να διευκολύνει σημαντικά την αναζήτηση πληροφορίας, και αποτέλεσε το κίνητρο για την προώθηση της τεχνολογίας του σημασιολογικού Ιστού (semantic web). Τα μετα-δεδομένα αυτά όμως μπορούν να αξιοποιηθούν μόνο αν έχουν δημιουργηθεί σύμφωνα με ένα συγκεκριμένο φορμαλισμό αναγνωρίσιμο από μηχανές αναζήτησης. Για το σκοπό αυτό το World Wide Web Consortium (W3C) δημιούργησε πρόσφατα την ομάδα εργασίας POWDER με σκοπό τον ορισμό ενός κοινού πρωτοκόλλου για την περιγραφή των πόρων του Ιστού (web resources). Σε αυτή την ομάδα εργασίας συμμετέχει ενεργά και το Εργαστήριο. Μέσω του κοινού αυτού φορμαλισμού διαφορετικοί χρήστες θα είναι σε θέση να περιγράφουν με μετα-δεδομένα το περιεχόμενο της ίδιας ιστο-σελίδας (π.χ. εκφράζοντας απόψεις για το περιεχόμενο μιας ιστο-σελίδας με θέματα υγείας). Στόχος αυτής της εργασίας είναι η δημιουργία μιας εφαρμογής που θα συνδυάζει μετα-δεδομένα από διαφορετικούς χρήστες έτσι ώστε να παρουσιάζει μια ολοκληρωμένη εικόνα στον χρήστη που ενδιαφέρεται να μάθει τις απόψεις των άλλων χρηστών για το ίδιο περιεχόμενο.

**Επιβλέποντες Ερευνητές:**

Βαγγέλης Καρκαλέτσης τηλ. 21-06503197 email: [vangelis@iit.demokritos.gr](mailto:vangelis@iit.demokritos.gr)

Κώστας Σταματάκης, email: [kstam@iit.demokritos.gr](mailto:kstam@iit.demokritos.gr)

**1.16 Αναγνώριση γραμματικών φαινομένων από κείμενα με χρήση γράφων**

**Επίπεδο:** Προπτυχιακό ή μεταπτυχιακό.

**Απαιτούμενα προσόντα:** Γνώσεις Τεχνητής Νοημοσύνης.

**Επιθυμητά προσόντα:** Γνώσεις Μηχανικής Μάθησης, εξοικείωση με αλγορίθμους σε γράφους.

**Περιγραφή:** Επειδή η γλώσσα προϋπήρξε της γραμματικής και του συντακτικού, μέχρι σήμερα οι περισσότερες αναπαραστάσεις γραμματικής είναι προσεγγίσεις της πραγματικότητας. Στα πλαίσια της εργασίας αυτής μελετάμε το πώς γραμματικά φαινόμενα αναπαρίστανται σε γράφους ν-γραμμάτων διαφορετικών μηκών. Επίσης, ελέγχουμε την εντοπιστικότητα γραμματικών φαινομένων με μελέτη του αντιστοίχου γράφου.

**Επιβλέποντες Ερευνητές:**

Βαγγέλης Καρκαλέτσης τηλ. 21-06503197 email: [vangelis@iit.demokritos.gr](mailto:vangelis@iit.demokritos.gr)

Γιώργος Γιαννακόπουλος, email: [ggianna@disi.unitn.it](mailto:ggianna@disi.unitn.it)

### **1.17 Αναπαράσταση και αναγνώριση θέματος με χρήση γράφου ν-γραμμάτων**

**Επίπεδο:** Προπτυχιακό ή μεταπτυχιακό.

**Απαιτούμενα προσόντα:** Γνώσεις Τεχνητής Νοημοσύνης.

**Επιθυμητά προσόντα:** Γνώσεις Μηχανικής Μάθησης, εξοικείωση με αλγορίθμους σε γράφους.

**Περιγραφή:** Στην αναπαράσταση και ανάλυση κειμένου είναι ενδιαφέρον να εντοπίσει κανείς τα θέματα του κειμένου (topics), πού αυτά εντοπίζονται στο κείμενο (topic detection) και ποια η αναπαράστασή τους. Επίσης, είναι χρήσιμο να εντοπίζονται τα σημεία αλλαγής θέματος (topic shift). Όλα αυτά βοηθούν στην ταξινόμηση κειμένων βάσει θέματος. Στόχος της εργασίας είναι να μελετηθεί πώς μπορούν να αναπαρασταθούν θέματα με χρήση γράφου ν-γραμμάτων και πώς εντοπίζονται οι αλλαγές θέματος σε ένα κείμενο με βάση την ίδια αναπαράσταση.

**Επιβλέποντες Ερευνητές:**

Βαγγέλης Καρκαλέτσης τηλ. 21-06503197 email: [vangelis@iit.demokritos.gr](mailto:vangelis@iit.demokritos.gr)

Γιώργος Γιαννακόπουλος, email: [ggianna@disi.unitn.it](mailto:ggianna@disi.unitn.it)

### **1.18 Μη-εποπτευόμενη εξόρυξη πληροφορίας από κείμενα σε διαφορετικά επίπεδα ανάλυσης, με χρήση γράφων**

**Επίπεδο:** Προπτυχιακό ή μεταπτυχιακό.

**Απαιτούμενα προσόντα:** Γνώσεις Τεχνητής Νοημοσύνης. Γνώσεις αντικειμενοστραφούς προγραμματισμού.

**Επιθυμητά προσόντα:** Γνώσεις Μηχανικής Μάθησης, εξοικείωση με γράφο  $\psi$  και αλγορίθμους σε γράφους. Γνώσεις νευρωνικών δικτύων.

**Περιγραφή:** Ο τομέας της μη-εποπτευόμενης εξόρυξης γνώσης εδώ και χρόνια στοχεύει στην εξαγωγή προφανών και μη-προφανών κομματιών πληροφορίας (όρων, οντοτήτων, σχέσεων και συσχετίσεων), με σκοπό την αποτελεσματικότερη ανάλυση του πολύ μεγάλου όγκου πληροφορίας που υπάρχει διαθέσιμος στις σύγχρονες πηγές. Στα πλαίσια αυτής της εργασίας, στόχος είναι να χρησιμοποιηθεί αναπαράσταση γράφου για να μοντελοποιηθεί δεδομένα κειμένων που παραδοσιακά αναπαρίστανται ως διανύσματα και να μελετηθούν αλγόριθμοι που θα μπορέσουν να χρησιμοποιήσουν την αναπαράσταση αυτή για αποτελεσματική μη-εποπτευόμενη εξαγωγή γνώσης.

**Επιβλέποντες Ερευνητές:**

Γιώργος Παλιούρας, τηλ. 21-06503158, email: [paliourg@iit.demokritos.gr](mailto:paliourg@iit.demokritos.gr)

Γιώργος Γιαννακόπουλος, email: [ggianna@disi.unitn.it](mailto:ggianna@disi.unitn.it)

### **1.19 Αναγνώριση σύνθετων γλωσσικών φαινομένων ως ιεραρχικές δομές στοιχειωδών γράφων**

**Επίπεδο:** Μεταπτυχιακό.

**Απαιτούμενα προσόντα:** Γνώσεις Τεχνητής Νοημοσύνης. Γνώσεις αντικειμενοστραφούς προγραμματισμού.

**Επιθυμητά προσόντα:** Γνώσεις Μηχανικής Μάθησης, εξοικείωση με γράφους και αλγορίθμους σε γράφους. Εξοικείωση με νευρωνικά δίκτυα.

**Περιγραφή:** Η αναγνώριση του γραπτού λόγου σύμφωνα με τις νευρεπιστήμες στηρίζεται σε ασαφή (fuzzy) αναγνώριση σε διαφορετικά επίπεδα (γράμμα, λέξη, πρόταση, κτλ.). Σκοπός της εργασίας είναι να διατυπωθεί και μελετηθεί μία διαδικασία ανάλυσης κειμένων, που να μοντελοποιεί το κείμενο ως ιεραρχία γράφων από το γράμμα έως το κείμενο, με σκοπό την ανάθεση εννοιών σε διαφορετικά επίπεδα ανάλυσης. Θα ελεγχθούν ως μέθοδοι αναπαράστασης γράφοι, ομαδοποιημένοι με αυτο-οργανωνόμενα νευρωνικά δίκτυα.

**Επιβλέποντες Ερευνητές:**

Γιώργος Παλιούρας, τηλ. 21-06503158, email: [paliourg@iit.demokritos.gr](mailto:paliourg@iit.demokritos.gr)

Γιώργος Γιαννακόπουλος, email: [ggianna@disi.unitn.it](mailto:ggianna@disi.unitn.it)

## 1.20 Επέκταση ερωτήσεων με χρήση εξελιγμένων τεχνικών σημασιολογικής ανάλυσης ερωτημάτων

**Επίπεδο:** Προπτυχιακό ή μεταπτυχιακό.

**Απαιτούμενα προσόντα:** Γνώσεις Τεχνητής Νοημοσύνης.

**Επιθυμητά προσόντα:** Γνώσεις Μηχανικής Μάθησης, εξοικείωση με γράφους και αλγόριθμους σε γράφους.

**Περιγραφή:** Στα πλαίσια σύγχρονων εργαλείων ανάκτησης πληροφορίας, όπως οι μηχανές αναζήτησης, είναι συχνό ο χρήστης να υποβοηθείται στις αναζητήσεις του με χρήση συνωνύμων λέξεων με αυτές που βάζει στο ερώτημα. Συχνά όμως η χρήση λέξεων σε αλλαγμένη μορφή ή η έλλειψη γλωσσολογικών εργαλείων δυσκολεύει τον εντοπισμό συνωνύμων. Στα πλαίσια της εργασίας, προσπαθούμε να δούμε πώς μπορούμε να χρησιμοποιήσουμε μία δομή που έχει αναπτυχθεί στα πλαίσια του εργαστηρίου που ονομάζεται «εννοιολογικό ευρετήριο», ώστε να αποφευχθούν τα παραπάνω προβλήματα. Το εννοιολογικό ευρετήριο μπορεί να συνδυάσει πληροφορίες τμήματος λέξεων για να αποδώσει έννοιες σε φράσεις ή και προτάσεις.

**Επιβλέποντες Ερευνητές:**

Βαγγέλης Καρκαλέτσης τηλ. 21-06503197 email: [vangelis@iit.demokritos.gr](mailto:vangelis@iit.demokritos.gr)

Γιώργος Γιαννακόπουλος, email: [ggianna@disi.unitn.it](mailto:ggianna@disi.unitn.it)

## **ΔΡΑΣΗ 2. Φιλικά Συστήματα Πληροφορικής**

Στόχος είναι η έρευνα και ανάπτυξη στις παρακάτω περιοχές:

- ανάλυση της χρήσης υπηρεσιών παροχής πληροφορίας (π.χ. Web-sites, news-filtering services, digital libraries), με στόχο τη διάθεση πληροφορίας στο χρήστη σύμφωνα με τις ανάγκες και τα ενδιαφέροντά του.
- διαλογικά συστήματα (dialogue systems), δηλαδή συστήματα που επιτρέπουν την επικοινωνία του χρήστη με Η/Υ ή άλλες συσκευές, με χρήση φυσικής γλώσσας (προφορικοί ή γραπτοί διάλογοι).

Για την επίτευξη του στόχου αυτού αξιοποιούνται μέθοδοι, τεχνικές και εργαλεία από τις τεχνολογίες Επεξεργασίας Φυσικής Γλώσσας, Μοντελοποίησης Χρηστών, Λογισμικού και Μηχανικής Μάθησης.

### 2.1 Τεχνικές εξόρυξης γνώσης για την μοντελοποίηση χρηστών

**Επίπεδο:** Προπτυχιακό ή μεταπτυχιακό.

**Απαιτούμενα προσόντα:** Πολύ καλές γνώσεις προγραμματισμού, κατά προτίμηση σε Java.

**Επιθυμητά προσόντα:** Γνώσεις εξόρυξης γνώσης ή/και αλληλεπίδρασης ανθρώπου-μηχανής.

**Περιγραφή:** Οι τεχνικές εξόρυξης γνώσης δίνουν τη δυνατότητα αυτόματης κατασκευής και δυναμικής ενημέρωσης μοντέλων χρηστών, τα οποία αποτελούν την βάση για την εξατομίκευση των Διαδυκτυακών υπηρεσιών. Στο Εργαστήριο έχει αναπτυχθεί ένας εξυπηρετητής εξατομίκευσης (personalization server) γενικής χρήσης. Σκοπός της εργασίας είναι η επέκταση των δυνατοτήτων του εξυπηρετητή εξατομίκευσης με νέες μεθόδους εξόρυξης γνώσης από δεδομένα χρήσης του Παγκόσμιου Ιστού.

**Επιβλέποντες Ερευνητές:**

Γιώργος Παλιούρας, τηλ. 21-06503158, email: [paliourg@iit.demokritos.gr](mailto:paliourg@iit.demokritos.gr)

Δημήτριος Βογιατζής, τηλ. 21-06503194, email: [dimitrv@iit.demokritos.gr](mailto:dimitrv@iit.demokritos.gr)

### 2.2 Εκμάθηση γραμματικών από δεδομένα χρήσης του Παγκόσμιου Ιστού, με τη βοήθεια καταλόγων του Ιστού

**Επίπεδο:** Προπτυχιακό ή μεταπτυχιακό.

**Απαιτούμενα προσόντα:** Γνώσεις προγραμματισμού, κατά προτίμηση σε Java.

**Επιθυμητά προσόντα:** Γνώσεις εξόρυξης γνώσης ή/και αλληλεπίδρασης ανθρώπου-μηχανής.

**Περιγραφή:** Η πλοήγηση ενός χρήστη σε έναν ιστιακό τόπο ή γενικότερα στον Ιστό καταγράφεται ως ακολουθία ερωτήσεων από τους εξυπηρετητές που παρέχουν το περιεχόμενο. Αυτή η ακολουθία είναι ιδιαίτερα χρήσιμη για τη μοντελοποίηση της συμπεριφοράς των χρηστών στον Ιστό. Στο Εργαστήριο έχει

αναπτυχθεί μία μέθοδος που μοντελοποιεί τη συμπεριφορά των χρηστών με γραμματικές, τις οποίες μαθαίνει από τα δεδομένα που συλλέγουν οι εξυπηρετητές. Στα πλαίσια της εργασίας θα βελτιωθεί αυτή η μέθοδος χρησιμοποιώντας καταλόγους του Ιστού (Web directories) ως κατηγοριοποίηση του περιεχομένου που υπάρχει στον Ιστό και το οποίο ανακτούν οι χρήστες.

**Επιβλέποντες Ερευνητές:**

Δημήτρης Πιερράκος, email: [dpie@iit.demokritos.gr](mailto:dpie@iit.demokritos.gr)

Γιώργος Παλιούρας, τηλ. 21-06503158, email: [paliourg@iit.demokritos.gr](mailto:paliourg@iit.demokritos.gr)

**2.3 Βελτίωση συστημάτων συστάσεων (recommender systems) με ανακάλυψη κοινοτήτων εμπιστών χρηστών**

**Επίπεδο:** Προπτυχιακό ή μεταπτυχιακό.

**Απαιτούμενα προσόντα:** Γνώσεις προγραμματισμού, κατά προτίμηση σε Java.

**Επιθυμητά προσόντα:** Γνώσεις εξόρυξης γνώσης ή/και αλληλεπίδρασης ανθρώπου-μηχανής

**Περιγραφή:** Τα συστήματα συστάσεων (ΣΣ) χρησιμοποιούνται ευρέως (δες το Amazon) προκειμένου να βοηθήσουν τους χρήστες να βρουν χρήσιμες πληροφορίες μέσα από ένα κυκεώνα δεδομένων. Ένα τέτοιο σύστημα συνήθως προτείνει στο χρήστη κάποιες πληροφορίες που πιθανότατα τον ενδιαφέρουν. Η λειτουργία τους βασίζεται στην ιστορία αλληλεπίδρασης του χρήστη αλλά και στην ομοιότητα του χρήστη με άλλους χρήστες. Ανακύπτει όμως ένα πρόβλημα που αφορά σε καινούργιους χρήστες, για τους οποίους δεν υπάρχουν επαρκείς πληροφορίες προκειμένου να γίνουν συστάσεις. Εν μέρει αυτό το θέμα μπορεί να αντιμετωπιστεί μέσω της δημιουργίας *κοινοτήτων εμπιστών χρηστών*. Στο Εργαστήριο έχει αναπτυχθεί ένα σύστημα που προσπαθεί να εντοπίσει έμπειρους χρήστες και να χρησιμοποιήσει αυτή την πληροφορία για τη βελτίωση των συστάσεων. Στόχος της προτεινόμενης εργασίας είναι η βελτίωση αυτού του συστήματος και η επέκτασή του ώστε να προσφέρει στον χρήστη επεξήγηση του τρόπου δημιουργίας της σύστασης. Αυτό βοηθά σημαντικά στην οικοδόμηση εμπιστοσύνης από το χρήστη στο σύστημα. Η πειραματική ανάπτυξη του συστήματος θα γίνει στον Personalisation Server (PServer), που έχει επίσης αναπτυχθεί στο εργαστήριο και διευκολύνει στην αποθήκευση προφίλ των χρηστών. Μία πιθανή επέκταση της εργασίας αφορά στην υποστήριξη χρηστών με κινητές συσκευές τύπου PDA/smart phone.

**Επιβλέποντες Ερευνητές:**

Δημήτριος Βογιατζής, τηλ. 21-06503194, email: [dimitrv@iit.demokritos.gr](mailto:dimitrv@iit.demokritos.gr)

Γεώργιος Παλιούρας, τηλ. 21-06503158, email: [paliourg@iit.demokritos.gr](mailto:paliourg@iit.demokritos.gr)

**2.4 Εντοπισμός προσπαθειών χειραγώγησης σε συστήματα συστάσεων (recommender systems)**

**Επίπεδο:** Προπτυχιακό ή μεταπτυχιακό.

**Απαιτούμενα προσόντα:** Γνώσεις προγραμματισμού, κατά προτίμηση σε Java.

**Επιθυμητά προσόντα:** Γνώσεις εξόρυξης γνώσης ή/και αλληλεπίδρασης ανθρώπου-μηχανής

**Περιγραφή:** Τα συστήματα συστάσεων (ΣΣ) χρησιμοποιούνται ευρέως (δες το Amazon) προκειμένου να βοηθήσουν τους χρήστες να βρουν χρήσιμες πληροφορίες μέσα από ένα κυκεώνα δεδομένων. Ένα τέτοιο σύστημα συνήθως προτείνει στο χρήστη κάποιες πληροφορίες ή προϊόντα που πιθανότατα τον ενδιαφέρουν. Η λειτουργία τους βασίζεται στην ιστορία αλληλεπίδρασης του χρήστη αλλά και στην ομοιότητα του χρήστη με άλλους χρήστες. Σε πολλές περιπτώσεις όμως επιχειρείται η χειραγώγηση αυτών των συστημάτων, ώστε να προτείνονται συχνά συγκεκριμένες πληροφορίες ή προϊόντα. Αυτές οι προσπάθειες επηρεάζουν σημαντικά την εμπιστοσύνη των χρηστών στο σύστημα και είναι σημαντικό να εντοπίζονται. Στόχος της προτεινόμενης εργασίας είναι η ανάπτυξη μεθόδων αποκάλυψης προσπαθειών χειραγώγησης, μέσω της αναγνώρισης προτύπων στη βαθμολόγηση προϊόντων. Η πειραματική ανάπτυξη του συστήματος θα γίνει στον Personalisation Server (PServer), που έχει αναπτύξει το εργαστήριο και διευκολύνει στην αποθήκευση προφίλ των χρηστών.

**Επιβλέποντες Ερευνητές:**

Δημήτριος Βογιατζής, τηλ. 21-06503194, email: [dimitrv@iit.demokritos.gr](mailto:dimitrv@iit.demokritos.gr)

Γεώργιος Παλιούρας, τηλ. 21-06503158, email: [paliourg@iit.demokritos.gr](mailto:paliourg@iit.demokritos.gr)

**2.5 Διαχείριση διαλόγου σε ένα διαλογικό σύστημα με τρόπο που διευκολύνει την προσαρμογή σε νέες θεματικές περιοχές**

**Επίπεδο:** Προπτυχιακό ή μεταπτυχιακό.

**Απαιτούμενα προσόντα:** Γνώσεις προγραμματισμού, κατά προτίμηση σε Java.

**Επιθυμητά προσόντα:** Γνώσεις γλωσσικής τεχνολογίας ή/και διαλογικών συστημάτων.

**Περιγραφή:** Τα τελευταία χρόνια έχει σημειωθεί σημαντική πρόοδος στον τομέα των διαλογικών συστημάτων (ΔΣ - dialogue systems) με γραπτούς και προφορικούς διαλόγους. Ένα σύνηθες διαλογικό σύστημα περιλαμβάνει γλωσσικά εργαλεία για κατανόηση της απαίτησης του χρήστη (από πιο απλά που αφορούν τον εντοπισμό λέξεων-κλειδιών έως πιο σύνθετα που αναλύουν συντακτικά, σημασιολογικά και πραγματολογικά τις ερωτήσεις/διευκρινίσεις του χρήστη), εργαλεία για τη διαχείριση του διαλόγου που αξιοποιούν ένα διαλογικό μοντέλο για την αναπαράσταση και αποθήκευση των σταδίων του διαλόγου, και εργαλεία για την παραγωγή των απαντήσεων του συστήματος. Στο Εργαστήριο αναπτύχθηκε πρόσφατα η πρώτη έκδοση ενός διαλογικού συστήματος όπου το διαλογικό μοντέλο αναπαριστάται από ένα γράφο καταστάσεων. Στόχος της εργασίας είναι η επέκταση του εργαλείου διαχείρισης του διαλόγου έτσι ώστε να διευκολύνεται η προσαρμογή του διαλογικού συστήματος σε μια νέα θεματική περιοχή.

**Επιβλέπων Ερευνητής:**

Βαγγέλης Καρκαλέτσης, τηλ. 21-06503197, e-mail: [vangelis@iit.demokritos.gr](mailto:vangelis@iit.demokritos.gr)

## **2.6 Βελτίωση αναγνώρισης συναισθημάτων μέσω συνέργειας ομιλίας και κειμένου**

**Επίπεδο:** Προπτυχιακό ή μεταπτυχιακό.

**Απαιτούμενα προσόντα:** Καλές γνώσεις προγραμματισμού, κατά προτίμηση σε Matlab ή Java.

**Επιθυμητά προσόντα:** Γνώσεις Επεξεργασίας φυσικής γλώσσας, Μηχανικής μάθησης.

**Περιγραφή:**

Η επικοινωνία των ανθρώπων με τους υπολογιστές θα καθίσταται περισσότερη φυσική και ευφυής, όσο οι υπολογιστές θα μπορούν να αναγνωρίζουν κάποια ανθρώπινα συναισθήματα. Υπάρχει πάντα ένα μεγάλο ποσό πληροφορίας το οποίο οι άνθρωποι μεταδίδουν με το χρωματισμό της φωνής αλλά και άλλους τρόπους.

Τα πεδία της αναγνώρισης συναισθημάτων, με βάση το ακουστικό σήμα, έχει εξερευνηθεί σε ένα βαθμό. Στόχος της παρούσας εργασίας είναι να συνδυαστεί η επεξεργασία του σήματος με την επεξεργασία κειμένου (που αφορά στα λεχθέντα του χρήστη), ώστε να βελτιωθεί η ποιότητα αναγνώρισης συναισθημάτων. Τα πειράματα μπορούν να γίνουν με δεδομένα που έχει συλλέξει το εργαστήριο αλλά και με άλλα των οποίων επιτρέπεται η δημόσια χρήση.

**Επιβλέποντες Ερευνητές:**

Δημήτριος Βογιατζής, τηλ. 21-06503194, email: [dimitrv@iit.demokritos.gr](mailto:dimitrv@iit.demokritos.gr)

Βαγγέλης Καρκαλέτσης, τηλ. 21-06503197, email: [vangelis@iit.demokritos.gr](mailto:vangelis@iit.demokritos.gr)

## **2.7 Κατάτμηση και διαχωρισμός ομιλίας σε έξυπνα περιβάλλοντα διάχυτης νοημοσύνης**

**Επίπεδο:** Προπτυχιακό ή μεταπτυχιακό.

**Απαιτούμενα προσόντα:** Γνώσεις προγραμματισμού, κατά προτίμηση σε C/C++ και Matlab.

**Επιθυμητά προσόντα:** Γνώσεις επεξεργασίας σήματος, εικόνας, και μηχανικής μάθησης..

**Περιγραφή:** Πρόσφατα, παρατηρούνται συνεχώς αυξανόμενες ερευνητικές προσπάθειες στον τομέα της ανίχνευσης της ανθρώπινης δραστηριότητας σε περιβάλλοντα έξυπνων χώρων, κατάλληλα εξοπλισμένων με δίκτυα ακουστικών και οπτικών αισθητήρων (μικροφωνικές διατάξεις και διακριβωμένες κάμερες), όπως για παράδειγμα διαλέξεις ή συναντήσεις εντός «έξυπνων» αιθουσών (smart rooms). Ένα πρώτο απαραίτητο στοιχείο για την αποτελεσματική περιγραφή και εξαγωγή γνώσης από τα καταγεγραμμένα οπτικο-ακουστικά δεδομένα είναι η σωστή κατάτμηση / «ημερολογιοποίηση» των ομιλητών (speaker diarization), όπως και ο διαχωρισμός των σημάτων φωνής αυτών σε περίπτωση επικάλυψης. Η εργασία αυτή θα επικεντρωθεί στην ανάπτυξη ενός πρωτότυπου, νέου πλαισίου για τον αποτελεσματικό συνδυασμό της διαθέσιμης πολυ-καναλικής πληροφορίας (μικροφώνων και καμερών) για την αντιμετώπιση των δύο αυτών προβλημάτων.

**Επιβλέπων Ερευνητής:**

Γεράσιμος Ποταμιάνος, τηλ. 21-06503194, e-mail: [gpotam@iit.demokritos.gr](mailto:gpotam@iit.demokritos.gr)

## ΔΡΑΣΗ 3. Τεχνολογία Γνώσεων

Στόχος της δράσης είναι η έρευνα και ανάπτυξη σε θέματα αναπαράστασης και διαχείρισης γνώσεων.

### 3.1 Ανάπτυξη συστήματος αναγνώρισης γεγονότων από δεδομένα

**Επίπεδο:** Προπτυχιακό ή μεταπτυχιακό.

**Απαιτούμενα προσόντα:** Γνώσεις προγραμματισμού.

**Επιθυμητά προσόντα:** Γνώσεις αναπαράστασης γνώσης και συμπερασμού.

**Περιγραφή:** Τα συστήματα αναγνώρισης συμβολικών γεγονότων/συμπεριφορών από δεδομένα έχουν μελετηθεί και αναπτυχθεί σε διάφορα ερευνητικά πεδία, όπως αυτά της Τεχνητής Νοημοσύνης, των Κατανεμημένων Συστημάτων και της Τεχνολογίας Λογισμικού. Για παράδειγμα, έχουν αναπτυχθεί συστήματα που επιβλέπουν την λειτουργία κατανεμημένων συστημάτων και αναγνωρίζουν “ενδιαφέρουσες” συμπεριφορές που λαμβάνουν χώρα, όπως επιθέσεις που αποσκοπούν στην άρνηση παροχής υπηρεσιών (denial of service attacks). Στο πλαίσιο της εργασίας θα αναπτυχθεί βάση γνώσης με ορισμούς “ενδιαφεροσών” συμπεριφορών οι οποίοι θα εκφράζονται με τη χρήση ενός φορμαλισμού Τεχνητής Νοημοσύνης. Επιπλέον, θα αναπτυχθεί αλγόριθμος αναγνώρισης “ενδιαφεροσών” συμπεριφορών. Το τελικό σύστημα θα χρησιμοποιηθεί σε δεδομένα παρακολούθησης δημοσίων χώρων με video (για την αναγνώριση γεγονότων/συμπεριφορών όπως κλοπή, βανδαλισμός, κλπ.), παρακολούθησης κατανεμημένων συστημάτων (για την αναγνώριση γεγονότων όπως denial of service attack, κλπ.), ή ακουστικού σήματος που αφορούν τραγούδια της καμπούρας της φάλαινας (για την αναγνώριση τραγουδιών φάλαινας).

**Επιβλέποντες Ερευνητές:**

Αλέξανδρος Αρτίκης, τηλ. 21-06503217, email: [a.artikis@iit.demokritos.gr](mailto:a.artikis@iit.demokritos.gr)

Γιώργος Παλιούρας, τηλ. 21-06503158, email: [paliourg@iit.demokritos.gr](mailto:paliourg@iit.demokritos.gr)

### 3.2 Κατάτμηση χρονικά μεταβαλλόμενων δεδομένων (σημάτων) και αναγνώριση των γεγονότων που προκύπτουν, με χρήση χρονικών οντολογιών

**Επίπεδο:** Προπτυχιακό ή μεταπτυχιακό.

**Απαιτούμενα προσόντα:** Γνώσεις προγραμματισμού.

**Επιθυμητά προσόντα:** Γνώσεις αναπαράστασης γνώσης και επεξεργασίας σήματος.

**Περιγραφή:** Στην αναγνώριση γεγονότων από σήματα, με χρήση οντολογιών, η διαδικασία ξεκινά με την ανάλυση του σήματος. Σ’ αυτή τη φάση χρειάζεται να γίνει κατάτμηση του σήματος σε βασικές μονάδες και χαρακτηρισμός των μονάδων ως γεγονότα χαμηλού επιπέδου. Η σειρά αυτών των γεγονότων με τις χρονικές παραμέτρους τους τοποθετούνται στην οντολογία ως στιγμιότυπα κλάσεων και κατόπιν με τη βοήθεια μεθόδων συμπερασμού, αναγνωρίζονται επιπλέον γεγονότα υψηλότερου επιπέδου. Στα πλαίσια της εργασίας θα αναπτυχθούν μέθοδοι κατάτμησης σημάτων και αναγνώρισης γεγονότων. Οι μέθοδοι θα εφαρμοστούν σε ηχοδιαγράμματα (sonograms) που καταγράφουν τραγούδια της καμπούρας φάλαινας, για τα οποία έχουμε δημιουργήσει και την αντίστοιχη οντολογία.

**Επιβλέποντες Ερευνητές:**

Γιώργος Παλιούρας, τηλ. 21-06503158, email: [paliourg@iit.demokritos.gr](mailto:paliourg@iit.demokritos.gr)

Βασίλης Γάτος, τηλ. 21-06503183, email: [bgat@iit.demokritos.gr](mailto:bgat@iit.demokritos.gr)

### 3.3 Μέθοδοι βελτίωσης χρονικών οντολογιών για αναγνώριση γεγονότων από δεδομένα

**Επίπεδο:** Προπτυχιακό ή μεταπτυχιακό.

**Απαιτούμενα προσόντα:** Γνώσεις προγραμματισμού, κατά προτίμηση σε Java.

**Επιθυμητά προσόντα:** Γνώσεις αναπαράστασης γνώσης και συμπερασμού.

**Περιγραφή:** Στην πλειοψηφία τους οι οντολογίες που χρησιμοποιούνται στην πράξη περιορίζονται σε ιεραρχικές δομές εννοιών, με ελάχιστες μη-ιεραρχικές σχέσεις μεταξύ τους. Υπάρχει όμως έντονη ανάγκη αναπαράστασης λιγότερο στατικής γνώσης, όπως χρονικές σχέσεις που χαρακτηρίζουν διαδικασίες και γεγονότα. Στα πλαίσια της εργασίας θα υλοποιηθούν υπάρχουσες μέθοδοι αναγνώρισης γεγονότων και

βελτίωσης της οντολογίας γεγονότων με βάση ένα μικρό σύνολο επισημειωμένων δεδομένων. Το τελικό σύστημα θα αξιολογηθεί σε δεδομένα ακουστικού σήματος που αφορούν τραγούδια της καμπούρας φάλαινας.

**Επιβλέποντες Ερευνητές:**

Γιώργος Παλιούρας, τηλ. 21-06503158, email: [paliourg@iit.demokritos.gr](mailto:paliourg@iit.demokritos.gr)

Αλέξανδρος Αρτίκης, τηλ. 21-06503217, email: [a.artikis@iit.demokritos.gr](mailto:a.artikis@iit.demokritos.gr)

### **3.4 Δυναμικές προδιαγραφές πολυ-πρακτορικών συστημάτων**

**Επίπεδο:** Προπτυχιακό ή μεταπτυχιακό.

**Απαιτούμενα προσόντα:** Οικειότητα με τεχνικές Τεχνητής Νοημοσύνης.

**Επιθυμητά προσόντα:** Οικειότητα με τεχνικές αναπαράστασης γνώσης και συμπερασμού.

**Περιγραφή:** Στο εργαστήριο έχουμε αναπτύξει ένα πλαίσιο για δυναμικές προδιαγραφές πολυ-πρακτορικών συστημάτων. Σε αυτό το πλαίσιο, η προδιαγραφή ενός συστήματος αναπτύσσεται πριν από την έναρξη της λειτουργίας του (design-time specification). Επιπλέον, υπάρχει η δυνατότητα αλλαγής της προδιαγραφής ενός συστήματος κατά τη διάρκεια της λειτουργίας του (run-time modification). Για παράδειγμα, μπορεί οι πράκτορες μίας δημοπρασίας να αλλάξουν κατά τη διάρκεια της δημοπρασίας τον αριθμό γύρων προσφοράς (bidding rounds), τον τύπο προσφοράς (π.χ. αλλαγή από ανοιχτές προσφορές σε εμπιστευτικές προσφορές), κλπ. Σε αυτή την εργασία θα υλοποιηθεί μία πλατφόρμα αξιολόγησης των δυναμικών προδιαγραφών ενός πολυ-πρακτορικού συστήματος. Π.χ., η πλατφόρμα αυτή θα υπολογίζει, μεταξύ άλλων:

- την “απόσταση” που έχει η προδιαγραφή μία συγκεκριμένη χρονική στιγμή από μία “ιδανική” προδιαγραφή (π.χ. μία προδιαγραφή που επιθυμεί ο σχεδιαστής του πολυ-πρακτορικού συστήματος),
- το χρονικό διάστημα που πέρασε μέχρι η προδιαγραφή ενός συστήματος να γίνει “ιδανική”,
- το χρονικό διάστημα κατά το οποίο η προδιαγραφή ενός συστήματος ήταν σε κοντινή “απόσταση” από την “ιδανική” προδιαγραφή.

Η υλοποίηση που θα πραγματοποιηθεί στα πλαίσια της εργασίας θα βασιστεί σε μία πλατφόρμα αξιολόγησης δυναμικών προδιαγραφών που αναπτύχθηκε στα πλαίσια μίας πρόσφατης πτυχιακής εργασίας.

**Επιβλέπων Ερευνητής:**

Αλέξανδρος Αρτίκης, τηλ. 21-06503217, email: [a.artikis@iit.demokritos.gr](mailto:a.artikis@iit.demokritos.gr)

### **3.5 Μεθοδολογίες δομημένης αναπαράστασης κειμένου και εφαρμογές στην τεχνολογία γνώσεων**

**Επίπεδο:** Μεταπτυχιακό.

**Απαιτούμενα προσόντα:** Γνώσεις προγραμματισμού, κατά προτίμηση σε Java.

**Επιθυμητά προσόντα:** Γνώσεις δομών δεδομένων, μηχανικής μάθησης ή/και γλωσσικής τεχνολογίας.

**Περιγραφή:** Για την αναπαράσταση κειμένων χρησιμοποιούνται συχνά δομημένες μέθοδοι, όπως ιεραρχικές ή γραφικές αναπαραστάσεις. Σκοπός της εργασίας είναι να αξιολογηθεί η χρησιμότητα των παραπάνω αναπαραστάσεων στο πλαίσιο της εξαγωγής γνώσης από κείμενα, αλλά και της χρησιμοποίησής τους για τον εμπλουτισμό μίας οντολογίας που να περιγράφει τη σημασιολογία του κειμένου.

**Επιβλέποντες Ερευνητές:**

Βαγγέλης Καρκαλέτσης, τηλ. 21-06503197, e-mail: [vangelis@iit.demokritos.gr](mailto:vangelis@iit.demokritos.gr)

Γιώργος Γιαννακόπουλος, email: [ggianna@disi.unitn.it](mailto:ggianna@disi.unitn.it)

### **3.6 Σχεδιασμός και ανάπτυξη συστημάτων υποστήριξης αποφάσεων (Decision support Systems) και χρησιμοποίησή τους για τη διάταξη των διαφόρων υποσυστημάτων και τεχνολογιών του συστήματος Υποβοηθούμενης Διαβίωσης με Ευφυείς Τεχνολογίες (Ambient Assisted Living) A2E2**

**Επίπεδο:** Προπτυχιακό ή μεταπτυχιακό.

**Απαιτούμενα προσόντα:** Γνώσεις προγραμματισμού.

**Επιθυμητά προσόντα:** Γνώσεις αναπαράστασης γνώσης, γνώσεις περιρρέουσας ευφυΐας (ambient intelligence).

**Περιγραφή:** Στόχος του A2E2 είναι να συνδυάσει ευφυείς τεχνολογίες για την στήριξη της όσο το δυνατόν ανεξάρτητης διαβίωσης των ηλικιωμένων στο σπίτι τους. Ανάμεσα στα άλλα ερευνητικά θέματα, θα εξετασθούν όλες οι απόψεις αλληλεπίδρασης μεταξύ του χρήστη και του συστήματος δίνοντας έμφαση στην παροχή κινήτρων για την ανάπτυξη και διατήρηση της καλής φυσικής κατάστασης μέσω της φυσικής εξάσκησης. Κατά τη διάρκεια του έργου, θα σχεδιαστεί και θα δημιουργηθεί ένα avatar, το οποίο θα έχει το ρόλο του προσωπικού εκπαιδευτή (personal coach).

Ο στόχος της εργασίας θα είναι η ανάπτυξη οντολογιών πεδίου (domain ontologies) σχετικών με την ευημερία των ατόμων (well being) καθώς και η δημιουργία νέων οντολογιών όπου χρειάζεται. Κατόπιν, οι οντολογίες αυτές θα χρησιμοποιηθούν για την ανανέωση των σχετικών πληροφοριών που θα λαμβάνονται από διάφορα μέρη του A2E2 (αισθητήρες, αποτελέσματα αλληλεπίδρασης χρήστη συστήματος, κτλ.) που θα γίνεται σε πραγματικό χρόνο και για τη αυτόματη λήψη αποφάσεων σχετικά με κατάλληλες ενέργειες και αντιδράσεις του προσωπικού εκπαιδευτή. Για παράδειγμα, εάν το σύστημα αντιληφθεί κάποια αλλαγή στη διάθεση του χρήστη, βασισμένο σε αλλαγές στη χροιά της φωνής του καθώς και στην έλλειψη δραστηριότητας στο περιβάλλον του σπιτιού, ο εκπαιδευτής μπορεί να κάνει ερωτήσεις και να προσφέρει λύσεις ανάλογα με το πρόβλημα.

**Επιβλέποντες Ερευνητές:**

Βαγγέλης Καρκαλέτσος, τηλ. 210-6503197, e-mail: [vangelis@iit.demokritos.gr](mailto:vangelis@iit.demokritos.gr)

Άννα Σαχινόπουλου, e-mail: [Anna.Sachinopoulou@vtt.fi](mailto:Anna.Sachinopoulou@vtt.fi)