



ΑΡΙΣΤΟΤΕΛΕΙΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΕΣΣΑΛΟΝΙΚΗΣ
ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ



ΕΡΓΑΣΤΗΡΙΟ ΓΛΩΣΣΩΝ ΠΡΟΓΡΑΜΜΑΤΙΣΜΟΥ &
ΤΕΧΝΟΛΟΓΙΑΣ ΛΟΓΙΣΜΙΚΟΥ



Πτυχιακές 2011-2012

των κ.κ. Βλαχάβα, Βασιλειάδη, Βράκα και Τσουμάκα

Διαθέσιμα Θέματα μετά την πρώτη ανάθεση

Θέμα κ. Ι. Βλαχάβα

6. Ανάπτυξη εφαρμογών επίδειξης (demo) για βασικά θέματα ή εφαρμογές Τεχνητής Νοημοσύνης

Ανάπτυξη εφαρμογής για την παραμετρική επίλυση ενός προβλήματος και την γραφική απεικόνιση των αποτελεσμάτων, με χρήση τεχνικών επίλυσης προβλημάτων (όπως αλγορίθμων αναζήτησης ή γενετικών αλγορίθμων ή νευρωνικών δικτύων, κλπ).

Απαιτούμενα προσόντα: Γνώσεις προγραμματισμού σε Java, C/C++

Θέματα κ. Ν. Βασιλειάδη

7. Γραφικό περιβάλλον συγγραφής κανόνων για την διαχείριση Ευφυών Πρακτόρων

Η πρόοδος στον τομέα των ευφυών πρακτόρων είναι σημαντική, ωστόσο ακόμη υπάρχει σημαντική έλλειψη εργαλείων. Το γεγονός αυτό σε συνδυασμό με την ολοένα αυξανόμενη ανάγκη για ευέλικτη δημιουργία και εποπτεία των πρακτόρων που δραστηριοποιούνται στα πολύ-πρακτορικά συστήματα, οδήγησε ένα κομμάτι της έρευνας να στραφεί στην χρήση κανόνων για τον έλεγχο της συμπεριφοράς τους. Στόχος της εργασίας είναι να μελετηθεί μια γλώσσα κανόνων και να υλοποιηθεί (σε JAVA) ένα γραφικό περιβάλλον συγγραφής κανόνων για αυτή, το οποίο θα υποβοηθά την δημιουργία και διαχείριση της συμπεριφοράς των ευφυών πρακτόρων. Παραδείγματα γλωσσών κανόνων που μπορούν να χρησιμοποιηθούν είναι οι Jess (Clips-like rule language) και/ή Prova (Prolog-like rule language), οι οποίες είναι ισχυρές εκφραστικά γλώσσες κανόνων.

Εργαλεία / Τεχνολογίες που θα χρησιμοποιηθούν: JAVA, υποσύνολο κάποιας γλώσσας κανόνων. (η γνώση ευφυών πρακτόρων είναι επιθυμητή αλλά όχι προαπαιτούμενη)

8. Διαχείριση εφοδιαστικής αλυσίδας μέσω Ευφυών Πρακτόρων

Η διαχείριση της εφοδιαστικής αλυσίδας ασχολείται με τον προγραμματισμό και τον συντονισμό των δραστηριοτήτων σε όλη την εφοδιαστική αλυσίδα, από την προμήθεια πρώτων υλών μέχρι την παράδοση προϊόντων. Στη σημερινή παγκόσμια οικονομία, η αποτελεσματική διαχείριση της εφοδιαστικής αλυσίδας είναι ζωτικής σημασίας για την ανταγωνιστικότητα των επιχειρήσεων μεταποίησης. Επιπλέον τεχνολογικά, η πρόοδος της έρευνας στον τομέα των ευφυών πρακτόρων τους έχει καταστήσει ως μια από τις πλέον υποσχόμενες τεχνολογίες για την υλοποίηση πολύπλοκων συστημάτων με πολλά αλληλεπιδρώντα μέρη. Στόχος της εργασίας

είναι η μελέτη του προβλήματος της διαχείρισης της εφοδιαστικής αλυσίδας και η υλοποίηση ενός σεναρίου προμήθειας όπου οι συμμετέχοντες πράκτορες θα πρέπει να εξισορροπήσουν τους στόχους τους ώστε αφενός να μεγιστοποιούν το κέρδους τους και αφετέρου να καλύπτουν τις απαιτήσεις των πελατών τους. Οι προδιαγραφές του σεναρίου θα υιοθετηθούν από τον διαγωνισμό TAC-SCM Procurement Challenge (SCM-PC): http://cmieux.isr.cs.cmu.edu/scm_pc/. Το SCM-PC αποτελεί τμήμα του TAC SCM Game που διοργανώνεται κάθε χρόνο από το Trading Agent Competition (TAC - <http://www.sics.se/tac/page.php?id=13>), ένα διεθνές φόρουμ με στόχο την προώθηση της έρευνας για το πρόβλημα των συναλλαγών μεταξύ των πρακτόρων.

Εργαλεία / Τεχνολογίες που θα χρησιμοποιηθούν: JAVA. (η γνώση ευφυών πρακτόρων είναι επιθυμητή αλλά όχι προαπαιτούμενη)

11. Σημαιολογική Διαχείριση Επαφών σε SmartPhones

Σκοπός της διπλωματικής εργασίας είναι να αναπτυχθεί μια εφαρμογή για smartphones η οποία θα κάνει χρήση τεχνολογιών του Σημαιολογικού Ιστού για τη διαχείριση των επαφών των χρηστών. Πιο συγκεκριμένα, η εφαρμογή θα πρέπει να δίνει τη δυνατότητα στο χρήστη να αποθηκεύει μεταδεδομένα για τις επαφές του, καθώς και να αναζητά επαφές σύμφωνα με την οντολογία FOAF. Τα μεταδεδομένα θα αποθηκεύονται σε μια σημαιολογική βάση με την οποία θα επικοινωνεί η εφαρμογή μέσω δικτύου τόσο για την προσθήκη νέων μεταδεδομένων όσο και για την αναζήτηση. Το άτομο που θα αναλάβει τη διπλωματική θα πρέπει να έχει κάποια εμπειρία στην ανάπτυξη εφαρμογών για κινητά (κατά προτίμηση για Android) και καλείται να εξοικειωθεί με μεθοδολογίες αποθήκευσης και ανάκτησης δεδομένων (ερωτήματα SPARQL) από σημαιολογικές βάσεις.

Εργαλεία / Τεχνολογίες που θα χρησιμοποιηθούν: Java, περιβάλλον ανάπτυξης mobile apps (κατά προτίμηση για Android), RDF/OWL, SPARQL, εγκατάσταση και διαχείριση σημαιολογικής βάσης (π.χ. OWLIM)

Θέματα κ. Γ. Τσουμάκα

16. Βελτίωση Συστήματος Διαχείρισης Διαδικτυακής Φήμης

Ο όρος διαδικτυακή φήμη αφορά τη δημόσια εικόνα μιας επιχείρησης, ενός προϊόντος, ή ενός προσώπου (π.χ. πολιτικού), όπως αυτή αποτιμάται από δημοσιεύματα στον παγκόσμιο ιστό (σε κοινωνικά μέσα, blogs, forums, κ.α.). Λόγω του σημαντικού ρόλου που παίζει πλέον στην καθημερινότητα μας ο παγκόσμιος ιστός, οι δημοσιεύσεις σε αυτόν μπορούν να επηρεάσουν καταλυτικά τη φήμη επιχειρήσεων και ανθρώπων. Επομένως, η σύγχρονη επιχείρηση και το σύγχρονο δημόσιο πρόσωπο, θα πρέπει να παρακολουθεί την διαδικτυακή κοινή γνώμη, ώστε να μπορεί να επέμβει άμεσα και κατάλληλα σε περίπτωση κακόβουλων δημοσιευμάτων ή ακόμα και πραγματικών ατυχών συμβάντων. Στόχος της πτυχιακής αυτής είναι η βελτίωση ενός λογισμικού που έχει αναπτυχθεί σε προηγούμενη πτυχιακή για την αυτόματη παρακολούθηση σε βάθος χρόνου των δημοσιευμάτων του παγκόσμιου ιστού για λέξεις κλειδιά που αφορούν μια επιχείρηση, ένα προϊόν ή ένα πρόσωπο και την αυτόματη κατάταξη τους σε θετικά/αρνητικά με τεχνικές μηχανικής μάθησης.

Επιθυμητά προσόντα: Γνώση προγραμματισμού σε Java και τεχνολογιών ανάπτυξης δυναμικών εφαρμογών παγκόσμιου ιστού (JSP, MySQL, JavaScript). Ενδιαφέρον για την περιοχή της μηχανικής μάθησης.

17. Μάθηση από Δεδομένα Πολλαπλών Ετικετών (3 διαφορετικά υποθέματα)

Δεδομένα πολλαπλών ετικετών (multi-label data), ονομάζουμε τα δεδομένα εκείνα τα οποία έχουν σημασθεί με μία ή παραπάνω ετικέτες (labels ή tags) από ένα πεπερασμένο σύνολο ετικετών. Τα δεδομένα αυτά διαφέρουν από τα δεδομένα που χρησιμοποιούνται στο κλασικό πρόβλημα της ταξινόμησης (classification), όπου κάθε δεδομένο ανήκει σε μία και μόνο κλάση από ένα πεπερασμένο σύνολο κλάσεων. Για παράδειγμα, πολλά τραγούδια των Scorpions θα μπορούσαν να σημασθούν τόσο με την ετικέτα «ροκ», όσο και με την ετικέτα «μπαλάντα», αλλά και με πολλές άλλες ετικέτες που θα μπορούσαν να αφορούν το συναίσθημα του κομματιού, τη γλώσσα των στίχων, τα μουσικά όργανα που χρησιμοποιήθηκαν στη σύνθεση, κ.α. Τα τελευταία χρόνια, η μάθηση από δεδομένα πολλαπλών ετικετών (multi-label learning) είναι ένα πρόβλημα που παρουσιάζει πολύ μεγάλο ενδιαφέρον, επειδή ανακύπτει σε πολλές ενδιαφέρουσες εφαρμογές όπως στην ανάλυση δεδομένων κειμένου (ιστοσελίδες, άρθρα σε blogs, κ.α.), βιολογικών δεδομένων (λειτουργία πρωτεϊνών), εικόνων (σημασιολογική σήμανση) και μουσικής (ταξινόμηση τραγουδιών σε συναισθήματα). Η αναγκαιότητα χρήσης τεχνικών μάθησης οφείλεται κυρίως στο ότι οι συλλογές δεδομένων (εικόνων, μουσικής, κειμένων, βιολογικών δεδομένων κτλ.) έχουν στις μέρες μας πολύ μεγάλο μέγεθος, και η χειροκίνητη σήμανση τους με ετικέτες έχει μεγάλο χρονικό και οικονομικό κόστος. Απαιτούνται λοιπόν τεχνικές που από ένα μικρό σύνολο δεδομένων, για το οποίο οι ετικέτες είναι γνωστές, μαθαίνουν ένα μοντέλο, το οποίο μπορεί στη συνέχεια να παράγει αυτόματα τις ετικέτες για τα υπόλοιπα δεδομένα, τα οποία δεν έχουν σημασθεί.

Η ομάδα Μηχανικής Μάθησης και Ανακάλυψης Γνώσης (<http://mlkd.csd.auth.gr>) έχει αναπτύξει σε γλώσσα Java το λογισμικό Mulan (<http://mulan.sourceforge.net>), ένα ανοιχτού κώδικα λογισμικό για μάθηση από δεδομένα πολλαπλών ετικετών, το οποίο χρησιμοποιείται διεθνώς από ερευνητές αλλά και από απλούς χρήστες εργαλείων ανάλυσης δεδομένων.

Στο παραπάνω πλαίσιο, προτείνονται οι εξής πτυχιακές εργασίες:

(α) Graphical User Interface. Το λογισμικό Mulan αυτή τη στιγμή μπορεί να χρησιμοποιηθεί μόνο ως βιβλιοθήκη, δεν παρέχει δηλαδή κάποιου είδους διεπαφής για χρήστες. Σκοπός της πτυχιακής είναι η ανάπτυξη μιας γραφικής διεπαφής, η οποία θα επιτρέπει την παρουσίαση απλών στατιστικών στοιχείων για ένα σύνολο δεδομένων πολλαπλών ετικετών, την οπτικοποίηση των συσχετίσεων των ετικετών, καθώς και την χρησιμοποίηση αλγορίθμων μάθησης είτε για αξιολόγηση της απόδοσης τους είτε για την πρόβλεψη των ετικετών νέων άγνωστων δεδομένων.

Επιθυμητά προσόντα: Πολύ καλή γνώση ανάπτυξης γραφικών διεπαφών χρήστη σε JAVA. Ενδιαφέρον για την περιοχή της μηχανικής μάθησης.

(β) *Active multi-label learning*. Λόγω του μεγάλου κόστους της χειροκίνητης σήμανσης των δεδομένων έχουν αναπτυχθεί τεχνικές, οι οποίες επιλέγουν από το σύνολο των δεδομένων δίχως σήμανση, εκείνο το οποίο αν σημασθεί χειροκίνητα, θα προσφέρει τη μέγιστη πληροφορία στον αλγόριθμο μάθησης. Οι τεχνικές αυτές ονομάζονται τεχνικές ενεργητικής μάθησης (active learning). Σκοπός της πτυχιακής είναι ο εμπλουτισμός του Mulan με τεχνικές ενεργητικής μάθησης της βιβλιογραφία και η πειραματική συγκριτική αξιολόγηση της απόδοσης τους.

Επιθυμητά προσόντα: Πολύ καλή γνώση προγραμματισμού σε JAVA. Ενδιαφέρον για την περιοχή της μηχανικής μάθησης.

Επιθυμητά προσόντα: Πολύ καλή γνώση προγραμματισμού. Ενδιαφέρον για την περιοχή της μηχανικής μάθησης.

18. Ανάπτυξη Ανοιχτής Βιβλιοθήκης Λογισμικού για Μεθόδους Κλαδέματος Ομάδας

Οι μέθοδοι ομάδας (ensemble methods) χρησιμοποιούνται για την αύξηση της ακρίβειας μοντέλων ταξινόμησης (classification) ή παλινδρόμησης (regression). Οι ομάδες μπορούν να πετύχουν μεγαλύτερη απόδοση πρόβλεψης από τα μεμονωμένα μοντέλα κυρίως μέσω της διόρθωσης των σφαλμάτων τους. Η μέθοδοι ομάδας περιλαμβάνουν τη διαδικασία παραγωγής των μοντέλων και τη διαδικασία συνδυασμού τους. Ανάμεσα στις δύο αυτές διαδικασίες μπορεί να παρεμβληθεί η φάση της κλαδέματος της ομάδας (ensemble pruning) η οποία αφορά τη μείωση του μεγέθους της. Το κλάδεμα αποτελεί μια σημαντική διαδικασία καθώς βοηθάει στη βελτίωση της απόδοσης (μειώνοντας το πλήθος των απαιτούμενων μοντέλων) και στην ακρίβεια πρόβλεψης (διατηρώντας τα μοντέλα με υψηλή ικανότητα πρόβλεψης) της ομάδας. Σκοπός της πτυχιακής είναι να αναπτυχθεί μία ανοιχτή (open source) βιβλιοθήκη λογισμικού για μεθόδους κλαδέματος ομάδας. Η βιβλιοθήκη θα περιέχει υλοποιήσεις γνωστών αλγορίθμων κλαδέματος της βιβλιογραφίας. Θα πρέπει να σημειωθεί ότι η βιβλιοθήκη θα βασιστεί στην παρακάτω βιβλιοθήκη όπου και υπάρχουν υλοποιημένες μέθοδοι ομάδας: <http://mlkd.csd.auth.gr/ensemblepruning.html>. Ένας ακόμη στόχος της διπλωματικής είναι να προστεθεί η βιβλιοθήκη στο αποθετήριο ανοιχτού κώδικα <http://mloss.org>, καθώς και η υποβολή εργασίας στο αντίστοιχο track του περιοδικού Journal of Machine Learning Research (<http://jmlr.csail.mit.edu/mloss/>).

Απαιτούμενα προσόντα: γνώσεις προγραμματισμού σε Java
Επιθυμητά προσόντα: γνώσεις μηχανικής μάθησης

Θέμα κ. Δ. Βράκα

19. Ανάπτυξη Εργαλείου Συγγραφής (Editor) αρχείων PDDL

Η πτυχιακή εργασία περιλαμβάνει την ανάπτυξη ενός εργαλείου συγγραφής (editor) για τη γλώσσα PDDL (Planning Domain Definition Language), που χρησιμοποιείται στο Σχεδιασμό Ενεργειών. Το εργαλείο θα πρέπει να δίνει στον χρήστη τη δυνατότητα συγγραφής νέων αρχείων πεδίου και προβλημάτων, προσφέροντας του ευκολίες, όπως το keyword highlighting, την αναγνώριση δομών, την αυτόματη συμπλήρωση, την επικύρωση της ορθότητας κ.α. Απαιτείται καλό υπόβαθρο γνώσεων σε TN και πολύ καλή γνώση προγραμματισμού σε JAVA.

Πρόταση Θέματος Πτυχιακής από Φοιτητές

Ο φοιτητής μπορεί να προτείνει θέμα της επιλογής του που να σχετίζεται με τις εκπαιδευτικές/ερευνητικές δραστηριότητες του Εργαστηρίου σε συνεννόηση με τον αντίστοιχο καθηγητή.

Οι ενδιαφερόμενοι καλούνται να αποστείλουν email στο lpis-ai@googlegroups.com επισυνάπτοντας: α) σύντομο βιογραφικό σημείωμα, β) αναλυτική βαθμολογία, γ) πρόσφατη φωτογραφία, και δ) λίστα με τα θέματα για τα οποία ενδιαφέρονται, ταξινομημένη σε σειρά προτίμησης (μέχρι 8 θέματα), π.χ. 8,1,16,16α,16γ,17.

Καταληκτική ημερομηνία υποβολής αιτήσεων: **1/11/2011**

Περισσότερες λεπτομέρειες για τα ερευνητικά ενδιαφέροντα των ομάδων ISKP και MLKD θα βρείτε στις διευθύνσεις <http://iskp.csd.auth.gr> και <http://mlkd.csd.auth.gr> αντίστοιχα.