



ΑΡΙΣΤΟΤΕΛΕΙΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΕΣΣΑΛΟΝΙΚΗΣ
ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ



ΕΡΓΑΣΤΗΡΙΟ ΓΛΩΣΣΩΝ ΠΡΟΓΡΑΜΜΑΤΙΣΜΟΥ &
ΤΕΧΝΟΛΟΓΙΑΣ ΛΟΓΙΣΜΙΚΟΥ



Διπλωματικές 2007-2008
του κ. Ι. Βλαχάβα

1. Ανάπτυξη και συγκριτική μελέτη μεθοδολογιών μάθησης από ρεύματα δεδομένων

Οι τελευταίες εξελίξεις στις τεχνολογίες αισθητήρων, αποθήκευσης, επεξεργασίας και δικτύων επέτρεψε τη αυτοματοποιημένη καταγραφή διάφορων τύπων δεδομένων. Η ανακάλυψη γνώσης από τέτοια ρεύματα δεδομένων (data streams) παρουσιάζει ιδιαίτερο ενδιαφέρον αφού ο όγκος και η ταχύτητα άφιξης των δεδομένων θέτουν νέες απαιτήσεις και περιορισμούς για τα μέχρι στιγμής προτεινόμενα συστήματα μάθησης. Στη διπλωματική αυτή, θα υλοποιηθούν διαθέσιμες μεθοδολογίες κατηγοριοποίησης δεδομένων από ρεύματα δεδομένων καθώς και μία συγκριτική μελέτη αυτών σε πραγματικά και τεχνητά σύνολα δεδομένων (τα οποία θα είναι επίσης διαθέσιμα). Θα γίνει χρήση της βιβλιοθήκης μηχανικής μάθησης Weka.

Απαιτήσεις: Προγραμματισμός σε γλώσσα Java

2. Περίληψη σε on-line Κριτικές Προϊόντων

Στα περισσότερα δημοφιλή ηλεκτρονικά καταστήματα (e-shops) του Παγκόσμιου Ιστού (π.χ. www.newegg.com, www.pricegrabber.com), οι παρουσιάσεις των διαθέσιμων προϊόντων συνοδεύονται από κριτικές και σχόλια χρηστών που αγόρασαν το προϊόν. Πολλές φορές ο μεγάλος αριθμός κριτικών δυσχεραίνει την αξιοποίησή τους από υποψήφιους αγοραστές. Σκοπός της διπλωματικής είναι η αυτόματη δημιουργία περιλήψεων με βάση αυτές τις κριτικές. Η περίληψη θα πρέπει να περιλαμβάνει, κατά το δυνατό, τμήματα από τα "σημαντικότερα" θετικά ή αρνητικά σχόλια. Η συγκέντρωση των κριτικών για την περαιτέρω επεξεργασία θα γίνεται με σύστημα εξαγωγής περιεχομένου από ιστοσελίδες που έχει αναπτυχθεί στο εργαστήριο. Η δημιουργία περιλήψεων θα γίνεται με εφαρμογή αλγορίθμου από υπάρχουσα βιβλιοθήκη αλγορίθμων επεξεργασίας φυσικής γλώσσας, σε γλώσσα προγραμματισμού Perl. Η επιτυχής ολοκλήρωση της διπλωματικής μπορεί να οδηγήσει και σε δημοσίευση σε σχετικό συνέδριο.

Γνωστική Περιοχή: web mining, summarization, web content extraction, e-shopping

Απαιτήσεις: Καλές γενικές γνώσεις προγραμματισμού και Αγγλικών – διάθεση για έρευνα – η γνώση γλώσσας Perl αποτελεί πλεονέκτημα

3. Μηχανή Συγκριτικών Αγορών

Τα περισσότερα ηλεκτρονικά καταστήματα (e-shops) χαρακτηρίζονται από ανύπαρκτους ή ελλιπείς μηχανισμούς σύγκρισης προϊόντων, που να επιτρέπουν στον πελάτη να επιλέξει το κατάλληλο προϊόν. Σκοπός της διπλωματικής είναι η δημιουργία ευέλικτης web εφαρμογής για την κατασκευή συγκριτικών πινάκων (comparison tables) με προδιαγραφές προϊόντων. Η database με τις προδιαγραφές των προϊόντων θα δοθεί.

Γνωστική Περιοχή: comparison e-shopping

Απαιτήσεις: Καλή γνώση HTML, Java Script και ενός από τα PHP/mysql ή ASP/SQL-Server

4. Ανάπτυξη αλγορίθμων ταξινόμησης δεδομένων πολλαπλών ετικετών

Η ταξινόμηση δεδομένων πολλαπλών ετικετών (multi-label classification) είναι ένα πρόβλημα που παρουσιάζει μεγάλο ενδιαφέρον, επειδή βρίσκει εφαρμογή στην ανάλυση δεδομένων κειμένου (ιστοσελίδες, άρθρα σε blogs, κ.α.) αλλά και στην ανάλυση βιολογικών δεδομένων. Για παράδειγμα ένα άρθρο που αφορά ένα νέο νομοσχέδιο για την παιδεία μπορεί να ταξινομηθεί τόσο στην κατηγορία «πολιτική» όσο και στην κατηγορία «παιδεία», ίσως ακόμα και στην κατηγορία «νομικά». Στόχος της διπλωματικής είναι ο εμπλουτισμός του λογισμικού MULAN (MUlti-LAbel classificatiON), το οποίο έχει αναπτυχθεί σε γλώσσα Java από την ομάδα Μηχανικής Μάθησης και Ανακάλυψης Γνώσης (mlkd.csd.auth.gr/multilabel.html), με έναν τουλάχιστο νέο αλγόριθμο από τη διεθνή βιβλιογραφία για την ταξινόμηση δεδομένων πολλαπλών ετικετών.

Απαιτήσεις: Προγραμματισμός σε γλώσσα Java

5. Μάθηση σε Συστήματα Πολλαπλών Πρακτόρων (ΣΠΠ)

Η παρούσα διπλωματική έχει ερευνητικό χαρακτήρα και σκοπός της είναι η μελέτη νέων μεθόδων Ενισχυτικής Μάθησης (EM) σε Συστήματα Πολλαπλών Πρακτόρων (ΣΠΠ). Η EM είναι μία τεχνική μάθησης μέσω της αλληλεπίδρασης ενός συστήματος με το περιβάλλον του. Η μάθηση σε ΣΠΠ έχει κερδίσει μεγάλο ερευνητικό ενδιαφέρον καθώς θέτει ενδιαφέροντα προβλήματα. Η EM είναι κατάλληλη τεχνική για επίλυση αυτών των προβλημάτων. Το περιβάλλον το οποίο θα χρησιμοποιηθεί είναι το Predator-Prey στο οποίο ένας ή περισσότεροι θηρευτές προσπαθούν να πιάσουν ένα θήραμα. Το πακέτο του περιβάλλοντος βρίσκεται στο URL: <http://staff.science.uva.nl/~jellekok/software/>.

Απαιτήσεις: Προγραμματισμός σε γλώσσα C++ ή Java

6. Ανακάλυψη Γνώσης από Δεδομένα Καταγραφής του Αριθμού Αντιγράφων DNA

Όλοι οι οργανισμοί περιέχουν το γενετικό τους υλικό (DNA), που καθορίζει τα χαρακτηριστικά των ίδιων και των απογόνων τους. Κάποια τμήματα του DNA μπορεί να εμφανίζονται σε πολλά αντίγραφα. Ο αριθμός των αντιγράφων αυτών μπορεί να διαφέρει από άτομο σε άτομο. Αυτές οι διαφορές συνήθως είναι υπεύθυνες για τη διαφορετικότητα των ατόμων αλλά και για την εκδήλωση ασθενειών. Τα τελευταία χρόνια έχουν γίνει αρκετά πειράματα για τη μελέτη του φαινομένου αυτού και έχουν προκύψει πολλά δεδομένα. Σκοπός της διπλωματικής εργασίας είναι η χρήση τεχνικών μηχανικής μάθησης και ανακάλυψης γνώσης για την ανάλυση τέτοιων δεδομένων. Συγκεκριμένα, η εργασία περιλαμβάνει τα εξής: α) Επισκόπηση της σχετικής βιβλιογραφίας και σύγκριση των γνωστών μεθόδων που έχουν εφαρμοστεί για τη μελέτη του συγκεκριμένου προβλήματος, β) Αναζήτηση και συγκέντρωση δεδομένων από βιολογικές βάσεις δεδομένων που υπάρχουν διαθέσιμες στο διαδίκτυο, γ) Χρήση έτοιμων βιβλιοθηκών μηχανικής μάθησης και ανακάλυψης γνώσης με σκοπό τη μελέτη και σύγκριση διάφορων αλγορίθμων. Τα αποτελέσματα της εργασίας είναι πιθανό να δημοσιευτούν σε συνέδριο της σχετικής περιοχής.

Απαιτήσεις: Προγραμματισμός σε γλώσσα C/C++ ή Java

7. Έρευνα και Συγκριτική Μελέτη Συστημάτων Σχεδιασμού

Η διπλωματική εργασία περιλαμβάνει την συλλογή και την κατηγοριοποίηση των σύγχρονων συστημάτων σχεδιασμού, που έλαβαν μέρος στους διεθνείς διαγωνισμούς σχεδιασμού και την ποιοτική τους σύγκριση με βάση τα χαρακτηριστικά και τις δυνατότητες τους. Σε δεύτερη φάση θα γίνει ποσοτική σύγκριση χρησιμοποιώντας μια δεξαμενή προβλημάτων από διάφορα πεδία που χρησιμοποιήθηκαν ως πλατφόρμα δοκιμών στους διαγωνισμούς. Ο φοιτητής που θα αναλάβει την εργασία θα κληθεί να ερευνήσει το Διαδίκτυο για ορισμένα Συστήματα Σχεδιασμού Ενεργειών καθώς και για πεδία/προβλήματα. Η διπλωματική εργασία πιθανώς να οδηγήσει σε δημοσίευση σε συνέδριο του χώρου. Απαιτείται γνώση του λειτουργικού συστήματος UNIX (SOLARIS ή LINUX) καθώς και απλές γνώσεις προγραμματισμού σε C, C++ ή Java.

8. Ανάπτυξη Συστήματος Σχεδιασμού Ενεργειών σε JAVA

Η διπλωματική εργασία περιλαμβάνει την ανάπτυξη ενός απλού συστήματος σχεδιασμού ενεργειών σε γλώσσα JAVA. Το σύστημα θα πρέπει να διαβάζει το πρόβλημα από τα 2 αρχεία που το περιγράφουν σε γλώσσα PDDL. Πιο συγκεκριμένα το αρχείο με το πεδίο (domain) και το αρχείο με το στιγμιότυπο του προβλήματος (problem). Στη συνέχεια θα αποθηκεύει το πρόβλημα χρησιμοποιώντας κατάλληλες δομές δεδομένων και θα το επιλύει χρησιμοποιώντας έναν απλό τυφλό αλγόριθμο αναζήτησης (π.χ. DFS, BFS). Για το τμήμα του parsing των αρχείων μπορούν να χρησιμοποιηθούν έτοιμοι parsers που υπάρχουν διαθέσιμοι στο δίκτυο. Απαιτείται καλή γνώση προγραμματισμού σε Java.

9. Ανάπτυξη Λογισμικού Ευφυούς Αυτόνομου Συστήματος

Η διπλωματική εργασία περιλαμβάνει την ανάπτυξη λογισμικού ενός απλού ευφυούς αυτόνομου συστήματος. Το λογισμικό θα πρέπει να είναι σε θέση να κατευθύνει ένα πράκτορα λογισμικού (softbot) σε ένα δυναμικό τρισδιάστατο επιβάλλον, ώστε αυτός να μετακινείται στο χώρο αποφεύγοντας τα εμπόδια που πιθανόν να συναντά. Το λογισμικό που θα αναπτυχθεί θα ακολουθεί τις αρχές των υβριδικών αρχιτεκτονικών αυτονομίας, χρησιμοποιώντας Σχεδιασμό Ενεργειών και Τεχνικές των Πεδίων Δυναμικού. Η πλατφόρμα ανάπτυξης του συστήματος θα είναι είτε το Microsoft Robotics Studio ή το SimBad (<http://simbad.sourceforge.net/>). Απαιτείται καλή γνώση προγραμματισμού σε C# ή σε Java.

10. Επίλυση Προβλημάτων Δημιουργίας Εκπαιδευτικών Προγραμμάτων με Χρήση Σχεδιασμού Ενεργειών

Η διπλωματική εργασία περιλαμβάνει τη δημιουργία ενός αριθμού από προβλήματα σύνθεσης εκπαιδευτικού υλικού για την ανάπτυξη εκπαιδευτικών προγραμμάτων (curricula). Κάθε τέτοιο πρόβλημα θα θεωρεί ότι ο χρήστης (μαθητευόμενος) γνωρίζει ένα σύνολο από θεματικές ενότητες και επιθυμεί να μάθει κάποιες ακόμη. Για να επιτευχθεί αυτό θα πρέπει να δημιουργηθεί ένα σύστημα το οποίο θα του προτείνει μια ακολουθία από εκπαιδευτικά αντικείμενα (κείμενα, παρουσιάσεις, βίντεο κ.α.) τα οποία θα επιφέρουν το επιθυμητό αποτέλεσμα. Η διπλωματική εργασία περιλαμβάνει τη δημιουργία ενός πλήθους τέτοιων προβλημάτων με αυτόματο τρόπο, την κωδικοποίησή τους σε προβλήματα Σχεδιασμού Ενεργειών και την Επίλυση των προβλημάτων, χρησιμοποιώντας κάποιο έτοιμο σύστημα σχεδιασμού ενεργειών. Απαιτείται γνώση προγραμματισμού σε C, C++ ή Java.

Οι ενδιαφερόμενοι καλούνται να επικοινωνήσουν με τον κ. Βλαχάβα, προσκομίζοντας:

- σύντομο βιογραφικό σημείωμα,
- απλή φωτοτυπία αναλυτικής βαθμολογίας, και
- φωτοτυπία του πάσο ή της 1ης σελίδας του βιβλιαρίου σπουδών.

Περισσότερες λεπτομέρειες για τα ερευνητικά ενδιαφέροντα της ερευνητικής ομάδας Λογικού Προγραμματισμού & Ευφών Συστημάτων (LPIS): <http://lpis.csd.auth.gr>