



ΑΡΙΣΤΟΤΕΛΕΙΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΕΣΣΑΛΟΝΙΚΗΣ  
ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ



ΕΡΓΑΣΤΗΡΙΟ ΓΛΩΣΣΩΝ ΠΡΟΓΡΑΜΜΑΤΙΣΜΟΥ &  
ΤΕΧΝΟΛΟΓΙΑΣ ΛΟΓΙΣΜΙΚΟΥ



**Πτυχιακές 2013-2014**

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

## **Θέματα κ. Ι. Βλαχάβα**

### **1. Ανάπτυξη επέκτασης (add-on) για Εξόρυξη Γνώσης από δεδομένα ηλεκτρονικού καταστήματος (1 ή 2 άτομα)**

Σκοπός της πτυχιακής εργασίας είναι η ανάπτυξη επέκτασης (add-on) για την εφαρμογή ηλεκτρονικού καταστήματος ανοιχτού κώδικα Magento με σκοπό, την εξόρυξη γνώσης (data mining) από τα δεδομένα συναλλαγών των πελατών του ηλ. καταστήματος. Το συγκεκριμένο πρόσθετο θα παρέχει στους διαχειριστές ενός ηλ. καταστήματος Magento τη δυνατότητα να εξάγουν ομαδοποιήσεις (clusters) των πελατών τους βάσει της αξίας των συναλλαγών τους. Παρέχεται πλήρης καθοδήγηση στους αλγορίθμους Μηχ. Μάθησης και τις τεχνολογίες που θα χρησιμοποιηθούν. Η επιτυχής υλοποίηση ενός πραγματικού πρόσθετου για το ηλ. κατάστημα Magento θα αποτελέσει σημαντική επαγγελματική εμπειρία για τον φοιτητή. Επίσης, το πρόσθετο μπορεί να κατοχυρωθεί και στο αντιστοίχο Magento Store.

#### **Απαιτούμενα Προσόντα:**

Γνώση HTML, PHP και SQL

**Επικοινωνία:** Ανέστης Φαχαντίδης, [afa@csd.auth.gr](mailto:afa@csd.auth.gr)

#### **Σχετικό υλικό:**

<http://www.magentocommerce.com/>

<http://www.magentocommerce.com/magento-connect/>

<http://www.rstudio.com/shiny/>

### **2. Ενισχυτική Μάθηση για την ευφυή παροχή διαφήμισης Google**

Οι διαφημίσεις Google (Adwords) κατέχουν σήμερα το 53.1% της παγκόσμιας αγοράς διαδικτυακής διαφήμισης. Σκοπός της πτυχιακής εργασίας είναι η ανάπτυξη εφαρμογής που θα διαχειρίζεται με βέλτιστο τρόπο τις παραμέτρους μιας συγκεκριμένης διαφήμισης Google AdWords με σκοπό τη μέγιστη προβολή της. Η εφαρμογή θα χρησιμοποιεί το Google AdWords API και θα υλοποιεί αλγόριθμο Ενισχυτικής Μάθησης για να πετύχει τη μάθηση μιας βέλτιστης πολιτικής προβολής της διαφήμισης.

Επιτυχή αποτελέσματα από την παραπάνω υλοποίηση ενδέχεται να οδηγήσουν σε υποβολή εργασίας σε συνέδριο ή περιοδικό του χώρου και θα αποτελέσει σημαντική επαγγελματική εμπειρία για τον φοιτητή.

#### **Απαιτούμενα Προσόντα:**

Γνώσεις προγραμματισμού σε μία από τις διαδεδομένες γλώσσες PHP, Java, Python, .NET

**Επικοινωνία:** Ανέστης Φαχαντίδης, [afa@csd.auth.gr](mailto:afa@csd.auth.gr)

**Σχετικό υλικό:**

<http://en.wikipedia.org/wiki/AdWords>

<https://developers.google.com/adwords/api/index>

<http://webdocs.cs.ualberta.ca/~sutton/book/ebook/node7.html>

### 3. Κατηγοριοποίηση εικόνων σε έξυπνες συσκευές

Η πτυχιακή αφορά την δημιουργία εφαρμογής που θα επιτρέπει την αυτόματη αναγνώριση του περιεχομένου φωτογραφιών που λαμβάνονται από έξυπνες συσκευές με σκοπό την (ημι-)αυτόματη οργάνωση των εικόνων με βάση το περιεχόμενό τους. Η πτυχιακή μπορεί να βασιστεί σε υπάρχουσα εφαρμογή για συσκευές Android.

**Επιθυμητά προσόντα:** Άριστη γνώση Java, γνώσεις Μηχανικής Μάθησης. Θα εκτιμηθεί ιδιαίτερα η γνώση Android SDK.

**Επικοινωνία:** Ελευθέριος Σπυρομήτρος-Ξιούφης, [espyromi@csd.auth.gr](mailto:espyromi@csd.auth.gr)

### 4. Πρόβλεψη καιρού

Η πτυχιακή θα μελετήσει τη δυνατότητα πρόβλεψης μετεωρολογικών δεικτών όπως θερμοκρασία, υγρασία, πιθανότητα βροχής, κλπ. με χρήση ιστορικών μετεωρολογικών δεδομένων αλλά και προγνώσεων από online πηγές (π.χ. [www.meteo.gr](http://www.meteo.gr), [www.weather.com](http://www.weather.com)). Για το σκοπό αυτό θα χρησιμοποιηθούν/επεκταθούν τεχνικές Μηχανικής Μάθησης που έχουν αναπτυχθεί από μέλη της ομάδας MLKD. Στα πλαίσια της πτυχιακής θα πραγματοποιηθεί συλλογή δεδομένων και θα διερευνηθούν κατάλληλοι τρόποι συνδυασμού και αναπαράστασής τους με σκοπό τη δημιουργία μοντέλων με υψηλή ακρίβεια πρόβλεψης.

**Επιθυμητά προσόντα:** Άριστη γνώση Java, γνώσεις Μηχανικής Μάθησης.

**Επικοινωνία:** Ελευθέριος Σπυρομήτρος-Ξιούφης, [espyromi@csd.auth.gr](mailto:espyromi@csd.auth.gr)

### 5. Βιοπληροφορική - Ανάλυση Δεδομένων Πληθυσμιακής Γενετικής

Τα τελευταία χρόνια με την πρόοδο της βιοτεχνολογίας έχει προκύψει ένας μεγάλος όγκος βιολογικών δεδομένων. Τα δεδομένα αυτά αναλύονται από ένα ευρύ φάσμα αλγορίθμων για την ανακάλυψη πολύτιμης βιολογικής γνώσης. Αυτό είναι το πεδίο της Βιοπληροφορικής. Σκοπός της παρούσας διπλωματικής είναι η ανάπτυξη και η τροποποίηση υπάρχοντων αλγορίθμων μηχανικής μάθησης για την ανάλυση γενετικών δεδομένων από την περιοχή της γενετικής πληθυσμών.

**Απαιτούμενα προσόντα:** Καλή γνώση προγραμματισμού σε JAVA. Ενδιαφέρον για βιολογικά θέματα. ΔΕΝ απαιτούνται γνώσεις βιολογίας.

**Σχετικό υλικό:**

*Βιοπληροφορική:* <http://en.wikipedia.org/wiki/Bioinformatics>

*Γενετική Πληθυσμών:* [http://en.wikipedia.org/wiki/Population\\_genetics](http://en.wikipedia.org/wiki/Population_genetics)

**Επικοινωνία:** Ιωάννης Καβακιώτης, [ikavak@csd.auth.gr](mailto:ikavak@csd.auth.gr)

## Θέματα κ. Ν. Βασιλειάδη

### **6. Δημοσίευση Πληροφοριών Δικτυακού Τύπου ως Διασυνδεδεμένα Δεδομένα στον Σημασιολογικό Ιστό (1 ή 2 άτομα)**

Ο Σημασιολογικός Ιστός (Semantic Web) αποτελεί το επόμενο μεγάλο βήμα εξέλιξης στο web με σκοπό την μετατροπή της αδόμητης πληροφορίας του περιεχομένου του web, σε ημιδομημένα ή δομημένα δεδομένα, τα οποία να μπορούν να γίνουν αυτόματα κατανοητά και επεξεργάσιμα από διαδικτυακές εφαρμογές, όπως είναι οι διαδικτυακές υπηρεσίες και οι ευφυείς πράκτορες. Ο Σημασιολογικός Ιστός περιλαμβάνει τη σαφή αναπαράσταση του νοήματος των πληροφοριών, των εγγράφων και των υπηρεσιών, με τη χρήση μεταδεδομένων (μοντέλο RDF) και των οντολογιών που περιγράφουν τη σημασία των μεταδεδομένων (γλώσσα OWL). Ο όρος linked data (διασυνδεδεμένα δεδομένα) περιγράφει μία μέθοδο δημοσιοποίησης δομημένων δεδομένων στο Σημασιολογικό Ιστό ώστε να είναι αλληλένδετα και να γίνουν πιο χρήσιμα. Η μέθοδος αυτή στηρίζεται στις γνωστές τεχνολογίες του Ιστού, όπως HTTP και URIs, αλλά αντί να τις χρησιμοποιεί για να εξυπηρετεί ιστοσελίδες για τους ανθρώπινους αναγνώστες, τις επεκτείνει ώστε να ανταλλάσσουν πληροφορίες με τρόπο που να μπορούν να διαβαστούν αυτόματα από τους υπολογιστές. Αυτό επιτρέπει δεδομένα από διαφορετικές πηγές να συνδέονται και να μπορούν να αναζητηθούν.

Στην παρούσα εργασία θα χρησιμοποιηθούν τα διασυνδεδεμένα δεδομένα που αφορούν στον δικτυακό τόπο της ερευνητικής ομάδας LPIS (<http://lpis.csd.auth.gr>) και έχουν αναπτυχθεί από προηγούμενη διπλωματική εργασία, προκειμένου να ενσωματωθούν δυναμικά σε ιστοσελίδες (XHTML ή HTML5) οι οποίες θα παρέχουν την ίδια πληροφορία στον επισκέπτη με τις σημερινές ιστοσελίδες του παραπάνω δικτυακού τύπου, αλλά επιπλέον θα έχουν ενσωματωμένα επισημειώσεις μεταδεδομένων χρησιμοποιώντας το πρότυπο RDFa.

Εργαλεία / Τεχνολογίες που θα χρησιμοποιηθούν: JAVA ή .NET ή PHP. RDF, Sparql, RDFa, HTML5 ή XHTML.

### **7. Αναγνώριση Συμβάντων από Καταναμημένες Πηγές Πληροφορίας μέσω Οντολογιών. (1 ή 2 άτομα)**

**Σενάριο:** Φανταστείτε ότι είστε σε ένα συνέδριο και παρακολουθείτε την ανακοίνωση εργασιών. Επίσης, θεωρείστε ότι όλοι οι παρόντες σε ένα χώρο όπου ανακοινώνεται μια εργασία σε μια δεδομένη χρονική στιγμή έχουν ένα μηχανισμό στο κινητό τους ή σε μια άλλη συσκευή (RFID tag) που δηλώνουν σε έναν αισθητήρα εντός της αίθουσας την παρουσία τους. Οι παρουσίες αυτές (καθώς και όλες οι αλλαγές) αποτελούν συμβάντα που αποθηκεύονται προσωρινά σε μια βάση δεδομένων. Καθώς παρακολουθείτε την ανακοίνωση, μπορείτε να ανιχνεύσετε την παρουσία συγκεκριμένων ατόμων, και από τα παρεχόμενα ανοικτά αλληλένδετα δεδομένα (linked open data) από διάφορες πηγές (π.χ. DBLP, Wikipedia / DBpedia ή από FOAF δεδομένα) ή από διάφορες άλλες πηγές πληροφορίας (όπως π.χ. το πρόγραμμα του συνεδρίου) μπορείτε να ανιχνεύσετε:

1. Πόσο έχει εργαστεί ο καθένας από τους παρόντες στην επιστημονική περιοχή που εντάσσεται η τρέχουσα ανακοίνωση, και ποια είναι τα συγκεκριμένα ενδιαφέροντά του
2. Πώς τα διάφορα παρόντα πρόσωπα σχετίζονται μεταξύ τους, π.χ. με τη συ-συγγραφή εργασιών, με την εργασία σε κοινά ερευνητικά έργα ή σε κοινά ινστιτούτα / πανεπιστήμια κ.ο.κ
3. Πώς ο ομιλητής συνδέεται με καθένα από τους παρόντες
4. Ένα σύνολο συνδέσμων σχετικά με προηγούμενες εργασίες του ομιλητή και άλλες συμπληρωματικές γενικές πηγές (DBpedia)

**Στόχος της διπλωματικής:** Η ανάπτυξη ενός προτύπου του παραπάνω συστήματος που θα λειτουργεί σε περιβάλλον προσομοίωσης σε ό,τι αφορά τους πραγματικούς χώρους ενός συνεδρίου, αλλά που θα επικοινωνεί με διάφορες πηγές πληροφορίας ιστού που παρέχουν ετερογενή πληροφορία για την αναγνώριση συγκεκριμένων προτύπων (patterns).

**Τεχνολογία:** Οντολογίες και μεταδεδομένα RDF για την μοντελοποίηση (α) πηγών πληροφορίας, (β) της παρεχόμενης πληροφορίας με ομοιογενή τρόπο, και (γ) των προτύπων που μας ενδιαφέρουν να

ανιχνευθούν. Θα μελετηθούν επίσης μηχανισμοί διάσπαρτης, κατανεμημένης συλλογιστικής που θα συνδυάζουν την πληροφορία από τις διάφορες οντολογίες, καθώς και μηχανισμοί και γλώσσες ανάκλησης μεταδεδομένων από βάσεις RDF (Sparql). Η γλώσσα υλοποίησης αφήνεται ανοιχτή (π.χ. Java, Python, Prolog, ...).

**Επιβλέποντες:** Ν. Βασιλειάδης, Γ. Βούρος (Παν. Πειραιώς)

## **8. Using and Extending Thea for Linked Data applications<sup>1</sup>**

The Semantic Web vision is about creating a web of machine processable data analogous to the web of (html) documents and pages we know today. In the last 10 years since the vision was first outlined, RDF and OWL specifications were developed for publishing data and ontologies respectively on the web. Linked Data (<http://www.w3.org/DesignIssues/LinkedData.html>) is an initiative to publish and make interconnected data available on the web using open standards (RDF). There exist now a large number of interconnected data sources that adhere to the Linked Data principles (called the cloud). Thea is an OWL2 library for Prolog ([http://ceur-ws.org/Vol-529/owled2009\\_submission\\_43.pdf](http://ceur-ws.org/Vol-529/owled2009_submission_43.pdf)). It can be used to access, load, query and manipulate OWL2 ontologies as Prolog programs and fact databases.

The proposed work aims a) in developing Prolog libraries that extend Thea to accessing and processing Linked data and b) building prototype applications of Linked data utilizing the extra semantics that Ontologies provide and the power of Prolog's rule-based programming paradigm. A well-known example of Linked Data is Dbpedia (<http://dbpedia.org>).

Εργαλεία / Τεχνολογίες που θα χρησιμοποιηθούν: Prolog, Thea, OWL.

## **9. Γραφικό περιβάλλον αναπαράστασης ευφυών πρακτόρων.**

Με την ανάπτυξη του Σημαιολογικού Ιστού (Semantic Web), η ανάπτυξη πολύ-πρακτορικών συστημάτων (MAS) αποτελεί σημαντικό τομέα έρευνας. Ιδιαίτερο ενδιαφέρον παρουσιάζει ο τομέας γραφικών αναπαραστάσεων. Στόχος της παρούσας εργασίας είναι η επέκταση ενός υλοποιημένου, στην γλώσσα JAVA, γραφικού περιβάλλοντος (GUI), το οποίο θα επιτρέπει γρήγορο και αποτελεσματικό έλεγχο των ευφυών πρακτόρων από τους ανθρώπους – χρήστες.

Εργαλεία / Τεχνολογίες που θα χρησιμοποιηθούν: Java.

## **10. Ανάπτυξη κατανεμημένου μοντέλου διαχείρισης εμπιστοσύνης.**

Η σημαντική ανάπτυξη πολύ-πρακτορικών συστημάτων (MAS) στο Σημαιολογικό Ιστό (Semantic Web) έχει καταστήσει αναγκαία την ανάπτυξη μηχανισμών εμπιστοσύνης (Trust) που θα διέπουν τις συναλλαγές μεταξύ των ευφυών πρακτόρων. Στόχος της εργασίας είναι η ανάπτυξη ενός νέου κατανεμημένου μοντέλου εμπιστοσύνης κατόπιν μελέτης της σχετικής βιβλιογραφίας.

Εργαλεία / Τεχνολογίες που θα χρησιμοποιηθούν: Java.

## **11. Ανάπτυξη ευφυούς πράκτορα διαχείρισης έξυπνων ηλεκτρικών δικτύων σύμφωνα με τους κανόνες του διαγωνισμού Power TAC.**

Ο οργανισμός TAC (Trading Agent Competition) θέλοντας να δώσει ώθηση στην έρευνα στον τομέα της έξυπνης διαχείρισης καθαρής ενέργειας, δημιούργησε τον διεθνή διαγωνισμό Power TAC. Το Power TAC είναι μια πλατφόρμα η οποία μοντελοποιεί μια αγορά ηλεκτρικής ενέργειας (βασισμένη σε ανανεώσιμες πηγές) στην οποία οι πελάτες συνάπτουν συμβόλαια αγοράς ηλεκτρικής ενέργειας σε λιανική με ιδιωτικές εταιρίες "brokers" και οι brokers εξυπηρετούν τους πελάτες αγοράζοντας ενέργεια από μια αγορά χονδρικής. Επιπλέον, οι πελάτες πέρα από καταναλωτές μπορεί να είναι και παραγωγοί ενέργειας (π.χ. μέσω

---

<sup>1</sup> Η εργασία αυτή θα πραγματοποιηθεί σε συνεργασία με εταιρία.

φωτοβολταϊκών στις στέγες). Στόχος των brokers (ευφυείς πράκτορες) είναι η μεγιστοποίηση του κέρδους τους μέσω της διαδικασίας αγοραπωλησίας ενέργειας.

Σκοπός αυτής της πτυχιακής είναι αρχικά η μελέτη και ανάλυση της σχετικής βιβλιογραφίας και στην συνέχεια, η μελέτη της πλατφόρμας και η ανάπτυξη ενός ευφυούς πράκτορα που να ενεργεί με βάση τους κανόνες του Power TAC. Επιτυχής εκτέλεση της πτυχιακής ενδέχεται να οδηγήσει σε συμμετοχή στον διαγωνισμό. Περισσότερες πληροφορίες σχετικά με το Power TAC μπορείτε να βρείτε στην παρακάτω διεύθυνση: <http://www.powertac.org>.

Απαραίτητες γνώσεις: Java, τεχνητή νοημοσύνη

(Το συγκεκριμένο θέμα μπορεί να δοθεί και σε περισσότερους του ενός φοιτητών)

Περισσότερες πληροφορίες: Μανώλης Ρήγας, [erigas@csd.auth.gr](mailto:erigas@csd.auth.gr)

## **12. Δημιουργία βιβλιοθήκης και διεπαφής χρήστη για υπάρχον λογισμικό προσομοίωσης κίνησης και φόρτισης ηλεκτρικών οχημάτων σε περιβάλλον πόλης.**

Ένας μεγάλος αριθμός επιστημόνων, κυβερνήσεων, επιχειρήσεων και μη κυβερνητικών οργανώσεων έχοντας αντιληφθεί την ανάγκη για πέρασμα σε μια νέα εποχή σε ότι έχει να κάνει με τις προσωπικές μας μετακινήσεις, η οποία θα είναι φιλική προς το περιβάλλον αλλά και προς τον οδηγό, υποστηρίζει σθεναρά την ανάπτυξη και προώθηση ηλεκτρικών οχημάτων. Η είσοδος αυτών των οχημάτων στην αγορά θα μπορούσε να στεφθεί με επιτυχία μόνο αν υπήρχε η δυνατότητα ευφυούς και δυναμικής διαχείρισης της διαδικασίας φόρτισης αυτών. Τούτο γιατί, σε αντίθεση με τα συμβατικά οχήματα, η φόρτιση των ηλεκτρικών οχημάτων διαρκεί έως και 8 ώρες και η ανάγκη για ηλεκτρικό ρεύμα είναι μεγάλη. Γίνεται φανερό, λοιπόν, ότι η δυνατότητα για αποτελεσματική και όσο το δυνατόν γρηγορότερη φόρτιση αυτών είναι εκ των ων ουκ άνευ ώστε να γίνουν ελκυστικά προς τους αγοραστές.

Σκοπός αυτής της πτυχιακής είναι η δημιουργία βιβλιοθήκης σε Java που θα βασίζεται πάνω σε ένα ήδη υλοποιημένο λογισμικό το οποίο προσομοιώνει την κίνηση και την φόρτιση ηλεκτρικών οχημάτων σε περιβάλλον πόλης. Επιπλέον της βιβλιοθήκης, θα πρέπει να υλοποιηθεί και μια διεπαφή χρήστη (user interface) για την εύκολη και αποτελεσματική διαχείριση και παρακολούθηση της προσομοίωσης.

Απαραίτητες γνώσεις: Java (πολύ καλή), τεχνολογία λογισμικού(επιθυμητή)

Περισσότερες πληροφορίες: Μανώλης Ρήγας, [erigas@csd.auth.gr](mailto:erigas@csd.auth.gr)

## **13. Web & Android Εφαρμογή Επαυξημένης Πραγματικότητας (Augmented Reality) σε Ασύρματο Δίκτυο Αισθητήρων**

Στόχος της εργασίας είναι η ανάπτυξη μιας εφαρμογής επαυξημένης πραγματικότητας (augmented reality) για τηλέφωνα Android και για Web πελάτες, στα πρότυπα των και ιδεών των Internet Of Things και Sensor Web. Ένα ασύρματο δίκτυο αισθητήρων, εγκατεστημένο στο Διεθνές Πανεπιστήμιο Ελλάδος (IHU), μετρά την κατανάλωση ρεύματος από διάφορες συσκευές, τη θερμοκρασία, την υγρασία, τη φωτεινότητα και τη συγκέντρωση CO<sub>2</sub> στον αέρα, σε διάφορα σημεία. Ήδη ανεπτυγμένο λογισμικό\* βασισμένο σε Υπηρεσίες Ιστού (Web Services) μας δίνει τη δυνατότητα απομακρυσμένου χειρισμού των δεδομένων. Η εφαρμογή επαυξημένης πραγματικότητας σε Android, θα κάνει χρήση hardware και βιβλιοθηκών ώστε να προβάλλει πάνω στην εφαρμογή κάμερας, πληροφορίες και διαγράμματα των μετρήσεων, μετά από ανάγνωση QR Codes για την αναγνώριση των σημείων. Η εφαρμογή web θα δίνει τη δυνατότητα απευθείας πρόσβασης σε αυτά τα δεδομένα και διαγράμματα για πελάτες web.

Απαιτείται καλή γνώση προγραμματισμού, ιδανικά σε Java

Πληροφορίες: Σταυρόπουλος Θάνος, Γραφείο 1, Ημιώροφος, [athstavr@csd.auth.gr](mailto:athstavr@csd.auth.gr)

\* Υπάρχουσες εφαρμογές <http://lpis.csd.auth.gr/people/thanosgstavr/applications.html>

## **Θέματα κ. Γ. Τσουμάκα**

### **14. Μάθηση από Δεδομένα Πολλαπλών Ετικετών (πολλά θέματα)**

Δεδομένα πολλαπλών ετικετών (multi-label data), ονομάζουμε τα δεδομένα εκείνα τα οποία έχουν σημειωθεί 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 σημειωθούν 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 σημειωθεί.

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. Οπτικοποίηση ετικετών**

Στόχος της εργασίας είναι η διερεύνηση κατάλληλων τρόπων οπτικοποίησης δεδομένων πολλαπλών ετικετών. Κυρίως μας ενδιαφέρει να εστιάσουμε στην οπτικοποίηση των σχέσεων μεταξύ των ετικετών σε ένα σύνολο δεδομένων πολλαπλών ετικετών, π.χ. (α) της έντασης της συσχέτισης των ετικετών ανά δύο μέσω για παράδειγμα ενός διδιάστατου πίνακα με διαβαθμίσεις του γκρι, (β) των σχέσεων γονέας-παιδί μεταξύ των ετικετών μέσω ενός γράφου. Άλλες χρήσιμες οπτικοποιήσεις αφορούν στη συχνότητα των ετικετών, στον μέσο αριθμό ετικετών ανά έγγραφο, στον αριθμό διαφορετικών συνδυασμών ετικετών και τη συχνότητα τους. Θα πρέπει να μελετηθεί αρχικά το πεδίο της μάθησης από δεδομένα πολλαπλών ετικετών, να μελετηθούν τα διάφορα είδη σχέσεων που υπάρχουν συνήθως μεταξύ των ετικετών, να αναπτυχθεί κώδικας υπολογισμού τους, να γίνει έρευνα ως προς τον καλύτερο τρόπο οπτικοποίησης τους και τέλος να υλοποιηθούν αυτοί οι τρόποι οπτικοποίησης. Για την οπτικοποίηση προτείνεται η ανάπτυξη εφαρμογής παγκοσμίου ιστού με τη χρήση της βιβλιοθήκης JavaScript <http://d3js.org/>, η οποία επιτρέπει πολύ εντυπωσιακά αποτελέσματα.

Επιθυμητά προσόντα: Γνώσεις JavaScript. Ενδιαφέρον για την περιοχή της μηχανικής μάθησης.

## **β. Βελτίωση της υποδομής του Mulan**

Στόχος της εργασίας αυτής είναι αφενός η επέκταση του Mulan με ένα command-line interface (CLI) και αφετέρου η βελτίωση του εσωτερικού τρόπου αναπαράστασης των δεδομένων. Το CLI θα επιτρέπει την εκτέλεση πειραμάτων μάθησης, την εφαρμογή των μοντέλων που προκύπτουν για πρόβλεψη σε νέα δεδομένα, και την εμφάνιση στατιστικών στοιχείων για ένα σύνολο δεδομένων. Θα ήταν καλό να διερευνηθούν οι επιλογές (α) της χρήσης έτοιμης βιβλιοθήκης Java για κατασκευή CLI (π.χ. <http://commons.apache.org/cli>) και (β) της ανάπτυξης εξειδικευμένου interface από το μηδέν. Στη συνέχεια θα πρέπει να υλοποιηθεί το interface, κάτι που θα επιτρέψει τη χρήση του Mulan και από απλούς χρήστες που δεν γνωρίζουν τη γλώσσα Java για να το χρησιμοποιήσουν ως βιβλιοθήκη. Η εσωτερική αναπαράσταση των δεδομένων αυτή τη στιγμή βασίζεται σε ένα αντικείμενο κλάσης της βιβλιοθήκης Weka για όλες τις μεταβλητές, ενώ θα ήταν πιο αποδοτικό αν αλλάξει στο πλαίσιο της εργασίας αυτής και χρησιμοποιεί δύο αντικείμενα της κλάσης αυτής, ένα για τις μεταβλητές εισόδου και ένα για τις μεταβλητές εξόδου.

Επιθυμητά προσόντα: Άριστη γνώση Java. Ενδιαφέρον για την περιοχή της μηχανικής μάθησης.

## **γ. Παραλληλοποίηση αλγορίθμων του Mulan**

Στόχος της εργασίας αυτής είναι η επέκταση του κώδικα αλγορίθμων που υπάρχουν ήδη στη βιβλιοθήκη Mulan έτσι ώστε να επιτρέπεται η παράλληλη εκτέλεση τους με χρήση πολλαπλών νημάτων (threads) της Java. Αρκετοί αλγόριθμοι του Mulan είναι από τη φύση τους παράλληλοι, π.χ. μάθηση ενός μοντέλου ανά ετικέτα, ωστόσο έχουν υλοποιηθεί μέχρι τώρα σειριακά, και σε αυτούς θα εστιάσει η πτυχιακή. Δεν θα χρειαστεί δηλαδή να διερευνηθούν τρόποι παραλληλοποίησης των υλοποιήσεων τους.

Επιθυμητά προσόντα: Άριστη γνώση Java.

## **δ. Σημασιολογική δεικτοδότηση βιοϊατρικών κειμένων**

Στην εθνική βιβλιοθήκη ιατρικής των ΗΠΑ κάθε χρόνο προστίθενται περίπου 1.000.000 περιλήψεις (abstracts) ιατρικών δημοσιεύσεων. Για την αποτελεσματικότερη ανάκτηση τους από ενδιαφερόμενους επαγγελματίες και ερευνητές της ιατρικής, η βιβλιοθήκη αναθέτει σε ειδικούς την ανάγνωση των περιλήψεων και τη σημασιολογική δεικτοδότηση τους με όρους από την ιατρική οντολογία MeSH (Medical Subject Headings), η οποία περιλαμβάνει περίπου 27.000 όρους. Για τη διευκόλυνση του έργου τους η βιβλιοθήκη έχει αναπτύξει ένα σύστημα, το οποίο αναλύει το περιεχόμενο της περίληψης και προτείνει στους ειδικούς εκείνους τους όρους που πιστεύει πως είναι κατάλληλοι για της δεικτοδότηση τους. Η ομάδα MLKD έχει αναπτύξει ένα παρόμοιο σύστημα, το οποίο σε πρόσφατο διεθνή διαγωνισμό (<http://bioasq.org>) ξεπέρασε σε ακρίβεια πρόβλεψης το σύστημα της εθνικής βιβλιοθήκης των ΗΠΑ. Το σύστημα στηρίζεται σε παράλληλη μάθηση γραμμικών μηχανών διανυσμάτων υποστήριξης για κάθε όρο της MeSH από το κείμενο των περιλήψεων. Στόχος της εργασίας αυτής είναι η βελτίωση του υπάρχοντος συστήματος και η περαιτέρω αύξηση της βελτίωσης του. Στο πλαίσιο αυτό θα διερευνηθούν τροποποιήσεις του υπάρχοντος συστήματος καθώς και υλοποιήσεις εναλλακτικών προσεγγίσεων και θα εκτελεστούν αντίστοιχα πειράματα.

Επιθυμητά προσόντα: Άριστη γνώση Java. Ενδιαφέρον για την περιοχή της μηχανικής μάθησης.

## **ε. Ανάπτυξη αλγορίθμων μάθησης πολλαπλών ετικέτων**

Στόχος της εργασίας είναι η επέκταση του Mulan με νέους αλγορίθμους της βιβλιογραφίας, η εκτέλεση πειραμάτων με στόχο την επαλήθευση των αποτελεσμάτων που παρουσιάζονται στη

βιβλιογραφία και τέλος η συγκριτική μελέτη ενός συνόλου σχετικών αλγορίθμων με σκοπό την εξαγωγή συμπερασμάτων ως προς τα πλεονεκτήματα και τα μειονεκτήματα του καθενός.

Επιθυμητά προσόντα: Άριστη γνώση Java. Ενδιαφέρον για την περιοχή της μηχανικής μάθησης.

## **Θέματα κ. Δ. Βράκα**

### **15. Ανάπτυξη Γραφικού Εργαλείου Απεικόνισης και Περιγραφής Προβλημάτων Σχεδιασμού Ενεργειών**

Η πτυχιακή εργασία περιλαμβάνει την ανάπτυξη ενός συστήματος με εύχρηστη γραφική διεπαφή το οποίο θα επιτρέπει στον χρήστη να οπτικοποιεί (visualize) και να σχεδιάζει προβλήματα σχεδιασμού ενεργειών. Το εργαλείο θα αποτελέσει επέκταση του συστήματος VLEPPO και μπορεί είτε να βασιστεί στον υπάρχοντα κώδικα (JAVA) ή να αναπτυχθεί εξολοκλήρου από την αρχή σε οποιαδήποτε γενική γλώσσα προγραμματισμού (π.χ. C++, JAVA). Απαιτείται καλό υπόβαθρο γνώσεων σε TN και πολύ καλές γνώσεις προγραμματισμού.

### **16. Ανάπτυξη Συστήματος Ελέγχου της Εγκυρότητας Προβλημάτων Σχεδιασμού Ενεργειών**

Η πτυχιακή εργασία περιλαμβάνει την ανάπτυξη ενός συστήματος το οποίο θα ελέγχει την εγκυρότητα αρχείων που περιλαμβάνουν περιγραφές προβλημάτων σχεδιασμού ενεργειών. Πιο συγκεκριμένα το σύστημα θα πρέπει να κάνει λεξικολογική και συντακτική ανάλυση των 2 αρχείων που περιγράφουν ένα πρόβλημα σχεδιασμού ενεργειών, σύμφωνα με τη γραμματική και το συντακτικό της γλώσσας PDDL. Στην συνέχεια, το σύστημα θα πρέπει να ελέγχει, χρησιμοποιώντας διάφορες τεχνικές την επιλυσιμότητα του προβλήματος και σε περίπτωση αρνητικού αποτελέσματος να προτείνει διορθώσεις στην περιγραφή που θα καταστήσουν το πρόβλημα επιλύσιμο. Απαιτείται πολύ καλό υπόβαθρο γνώσεων σε TN και καλές γνώσεις προγραμματισμού.

### **17. Συγκριτική Μελέτη και Πιλοτική Εφαρμογή Εργαλείων Σχεδιασμού Ενεργειών στην Εκπαιδευτική Διαδικασία**

Η πτυχιακή εργασία περιλαμβάνει την μελέτη ενός αριθμού γραφικών εργαλείων που χρησιμοποιούνται για την δημιουργία προβλημάτων Σχεδιασμού Ενεργειών (Planning) και την επίλυση τους. Στα εργαλεία αυτά συμπεριλαμβάνονται τα: PDDL Studio, itSimple και το VLEPPO. Στα πλαίσια της διπλωματικής θα πρέπει να γίνει μια συγκριτική μελέτη των συστημάτων αυτών (ενδεχομένως και άλλων), να καταγραφούν οι δυνατότητες, τα πλεονεκτήματα και τα μειονεκτήματα τους και να γίνει και μια πειραματική μελέτη της ευχρηστίας και των δυνατοτήτων τους στα πλαίσια του μαθήματος Τεχνητή Νοημοσύνη του 4<sup>ου</sup> εξαμήνου. Απαιτείται καλό υπόβαθρο γνώσεων σε TN και πολύ καλή γνώση Αγγλικών.

### **18. Υλοποίηση Μεθόδων Τεχνητής Νοημοσύνης για την Αποτροπή μη Εξουσιοδοτημένης Πρόσβασης σε Ευαίσθητα Δεδομένα**

Η διπλωματική εργασία περιλαμβάνει μελέτη στην περιοχή της Ελέγχου Συμπερασμάτων (Inference Controls), η οποία ασχολείται με μεθόδους αποτροπής της εξαγωγής συμπερασμάτων για προσωπικά δεδομένα ατόμων σε βάσεις δεδομένων, μέσω στατιστικών παρατηρήσεων. Πιο συγκεκριμένα στα πλαίσια της πτυχιακής εργασίας θα πρέπει να υλοποιηθούν αλγόριθμοι βελτιστοποίησης και αναζήτησης, οι οποίοι μεταβάλλουν τους πίνακες ελέγχου πρόσβασης σε βάσεις δεδομένων, ώστε να μειωθεί ή να εξαλειφθεί η δυνατότητα εξαγωγής συμπερασμάτων για δεδομένα στα οποία ο χρήστης δεν θα έπρεπε να έχει πρόσβαση. Απαιτείται καλό υπόβαθρο γνώσεων σε TN και πολύ καλή γνώση προγραμματισμού σε C/C++ ή JAVA.



## **19. Ανάπτυξη Γραφικού Εργαλείου Απεικόνισης και Συγγραφής Προβλημάτων Ιεραρχικού Σχεδιασμού Δικτύων Διεργασιών (HTN)**

Η διπλωματική εργασία περιλαμβάνει την μελέτη του συστήματος VLEPPO και την επέκταση του σε προβλήματα Ιεραρχικού Σχεδιασμού Δικτύων Διεργασιών (Hierarchical Task Network Planning). Το εργαλείο που θα αναπτυχθεί θα πρέπει να δίνει στον χρήστη τη δυνατότητα γραφικής απεικόνισης προβλημάτων σχεδιασμού (HTN), τη σύνθεση νέων προβλημάτων, χρησιμοποιώντας έτοιμα συστατικά και παλιά προβλήματα, καθώς και την επίλυση τους χρησιμοποιώντας κάποιον εξωτερικό HTN σχεδιαστή. Απαιτείται γνώση των αρχών του Σχεδιασμού Ενεργειών και καλή γνώση της γλώσσας προγραμματισμού Java.

## **20. Υλοποίηση Παιγνίου Δύο Αντιπάλων (Computer Game) με χρήση τεχνικών Τεχνητής Νοημοσύνης**

Η πτυχιακή εργασία περιλαμβάνει την ανάπτυξη ενός παιχνιδιού υπολογιστή (computer game) σε JAVA. Το παιχνίδι θα είναι το "Cold War: CIA vs KGB", λεπτομέρειες για το οποίο μπορούν να βρεθούν στο <http://www.boardgamegeek.com/boardgame/24742>

Η πτυχιακή εργασία περιλαμβάνει την ανάπτυξη εφαρμογής σε JAVA η οποία θα υλοποιεί τους κανόνες παιχνιδιού και τη δυνατότητα νέου παιχνιδιού με δύο παίκτες (ανθρώπους). Στη συνέχεια θα πρέπει να δημιουργηθεί και η δυνατότητα παιχνιδιού ενός ατόμου ενάντια στον υπολογιστή. Ο εικονικός παίκτης (υπολογιστής) θα πρέπει να επιλέγει την κίνηση του με βάση τον αλγόριθμο Alpha – Beta και κατάλληλες ευριστικές συναρτήσεις. 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 να αποστείλουν email στο group LPIS [lpis-ai@googlegroups.com](mailto: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, 13γ, 1, 16, 17.*

*Περισσότερες λεπτομέρειες για τα ερευνητικά ενδιαφέροντα των ομάδων ευφυών συστημάτων και επεξεργασίας γνώσης (ISKP) και μηχανικής μάθησης και ανακάλυψης γνώσης (MLKD) θα βρείτε στις διευθύνσεις <http://iskp.csd.auth.gr> και <http://mlkd.csd.auth.gr> αντίστοιχα.*

*Καταληκτική ημερομηνία υποβολής αιτήσεων: 18/10/2013. Ανακοίνωση αναθέσεων 23/10/2013.*