



ΑΡΙΣΤΟΤΕΛΕΙΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΕΣΣΑΛΟΝΙΚΗΣ  
ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ



ΕΡΓΑΣΤΗΡΙΟ ΓΛΩΣΣΩΝ ΠΡΟΓΡΑΜΜΑΤΙΣΜΟΥ &  
ΤΕΧΝΟΛΟΓΙΑΣ ΛΟΓΙΣΜΙΚΟΥ



**Πτυχιακές 2015-2016**

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

## **Θέματα κ. Ι. Βλαχάβα**

### **1. Ανάλυση γνωσιακών δεδομένων**

Η γνωσιακή επιστήμη είναι ένα διεπιστημονικό αντικείμενο το οποίο αντλεί γνώσεις και ερευνητική μεθοδολογία από τις γνωστικές νευροεπιστήμες, τη γνωστική ψυχολογία, την τεχνητή νοημοσύνη, τη γλωσσολογία και τη φιλοσοφία του νου.

Τα σημαντικότερα μοντέλα (αναπαράσταση + διαδικασίες) της Γνωσιακής Επιστήμης σήμερα είναι η Λογική, οι Κανόνες, οι Έννοιες, οι Αναλογίες, οι Εικόνες και οι συνδέσεις.

Η γνωστική ψυχολογία θεωρεί ότι για να μελετηθεί η μάθηση πρέπει να μελετήσουμε αναλυτικά τον μετασχηματισμό της πληροφορίας σε γνώση, δηλαδή τις φάσεις της γνωστικής επεξεργασίας πληροφοριών, και ρίχνει το βάρος της στη μελέτη των συγκεκριμένων γνωστικών λειτουργιών. Για να μετατραπεί η πληροφορία σε γνώση, μεσολαβούν όχι μόνον οι αισθήσεις, οι οποίες φιλτράρουν τα εξωτερικά ερεθίσματα, αλλά και κάποιες διαδικασίες επεξεργασίας οι οποίες μετασχηματίζουν τα δεδομένα σε γνώση.

Συνεισφέρει σημαντικά στο διεπιστημονικό πεδίο της γνωσιακής επιστήμης ενώ μοιράζεται το ίδιο αντικείμενο μελέτης, υπό διαφορετική σκοπιά, με την τεχνητή νοημοσύνη.

Σκοπός της παρούσας πτυχιακής είναι η μελέτη συσχέτισης των αλγορίθμων μηχανικής μάθησης και της γνωσιακής μάθησης, η μελέτη των διαφόρων ειδών μάθησης και η πειραματική αξιολόγηση και ανάλυση των μεθόδων με χρήση ηλεκτρονικών ερωτηματολογίων.

### **2. Βιοπληροφορική – Ανάπτυξη Εφαρμογής Ανάλυσης Δεδομένων Πληθυσμιακής Γενετικής**

Την τελευταία δεκαετία η πρόοδος της βιοτεχνολογίας έχει οδηγήσει στην εκθετική αύξηση των βιολογικών δεδομένων και έχει οδηγήσει την επιστήμη της βιολογίας στην περιοχή των Big Data. Η ανακάλυψη γνώσης σε βάσεις δεδομένων είναι περισσότερο απο ποτέ αναγκαία και σημαντική καθώς παίζει και θα συνεχίσει να παίζει πρωταγωνιστικό ρόλο στις επιστημονικές ανακαλύψεις στην περιοχή. Τα υπάρχοντα εργαλεία και οι τεχνικές της βιοπληροφορικής αδυνατούν να διαχειριστούν τα νέα δεδομένα. Επομένως είναι σημαντικό να υλοποιηθούν νέες προσεγγίσεις που λαμβάνουν υπόψη την τωρινή φύση των δεδομένων (διαστασιμότητα, μέγεθος κ.τ.λ.)

Σκοπός της παρούσας πτυχιακής είναι η ανάπτυξη εφαρμογής που θα περιέχει αλγορίθμους για την ανάλυση γενετικών δεδομένων από την περιοχή της γενετικής πληθυσμών. Για την εφαρμογή θα χρησιμοποιηθούν η JAVA και JAVA FX

Απαιτούμενα προσόντα:

- Καλή γνώση JAVA
- Ενδιαφέρον για βιολογικά θέματα. ΔΕΝ απαιτούνται γνώσεις βιολογίας.

Links:

- <http://www.nature.com/nature/journal/v498/n7453/full/498255a.html>
- Βιοπληροφορική: <http://en.wikipedia.org/wiki/Bioinformatics>
- Γενετική Πληθυσμών: [http://en.wikipedia.org/wiki/Population\\_genetics](http://en.wikipedia.org/wiki/Population_genetics)

Περισσότερες πληροφορίες: [ikavak@csd.auth.gr](mailto:ikavak@csd.auth.gr)

## **Θέματα κ. Ν. Βασιλειάδη**

### **3. Βιώσιμη κινητικότητα και έξυπνα δίκτυα ηλεκτροδότησης**

Ένας μεγάλος αριθμός επιστημόνων, κυβερνήσεων, επιχειρήσεων και μη κυβερνητικών οργανώσεων έχουν αντιληφθεί την ανάγκη για το πέρασμα σε μια νέα εποχή όσον αφορά τον τρόπο που θα γίνονται οι προσωπικές μας μετακινήσεις. Αυτός θα πρέπει να είναι φιλικός προς το περιβάλλον αλλά και προς τον οδηγό, και προς αυτήν την κατεύθυνση η προώθηση και η χρήση ηλεκτρικών οχημάτων είναι καθοριστικής σημασίας. 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. Αυτό γιατί, σε αντίθεση με τα συμβατικά οχήματα, η φόρτιση των ηλεκτρικών οχημάτων διαρκεί έως και 8 ώρες και η ανάγκη για ηλεκτρικό ρεύμα είναι υψηλή. Γίνεται φανερό, λοιπόν, ότι η δυνατότητα για αποτελεσματική και γρήγορη φόρτιση αυτών είναι εκ των ων ουκ άνευ ώστε να γίνουν ελκυστικά προς τους αγοραστές. Επιπρόσθετα, προκειμένου τα ηλεκτρικά οχήματα να μην είναι μόνο κατ' όνομα, αλλά και επί της ουσίας φιλικά προς το περιβάλλον, η ηλεκτρική ενέργεια που θα χρησιμοποιείται για την φόρτισή τους θα πρέπει να προέρχεται από ανανεώσιμες πηγές. Επιπλέον, τα ηλεκτρικά οχήματα θα πρέπει να συμμετέχουν στο δίκτυο όχι μόνο ως καταναλωτές, αλλά και ως παραγωγοί προσφέροντας την περισευούμενη τους ενέργεια σε σημεία και σε χρονικές στιγμές που η ζήτηση είναι αυξημένη.

Η παρακάτω ερευνητική εργασία παρέχει αρκετές πληροφορίες για (ερευνητικά) ζητήματα που σχετίζονται με τα ηλεκτρικά οχήματα:

[http://lps.csd.auth.gr/publications/ITS\\_rigas.pdf](http://lps.csd.auth.gr/publications/ITS_rigas.pdf)

Σε αυτό το πλαίσιο εμπίπτουν τα δύο ακόλουθα θέματα πτυχιακών εργασιών (είναι 2 ξεχωριστές πτυχιακές που θα δοθούν σε έναν φοιτητή η κάθε μια):

#### **A: Ανταλλαγή ενέργειας μεταξύ ηλεκτρικών οχημάτων με βάση ένα ομότιμο δίκτυο επικοινωνίας:**

Σκοπός αυτής της πτυχιακής είναι η μελέτη της διαδικασίας μεταφοράς ηλεκτρικής ενέργειας μεταξύ δύο ή περισσότερων ηλεκτρικών οχημάτων. Ποιο συγκεκριμένα, θα πρέπει να αναπτυχθεί αλγόριθμος βελτιστοποίησης που σκοπό θα έχει την αποτελεσματική μεταφορά ενέργειας μεταξύ οχημάτων στα πρότυπα των ομότιμων δικτύων (peer to peer networks). Η επικοινωνία μεταξύ οχημάτων (Vehicle-to-Vehicle V2V) με σκοπό την μεταφορά ενέργειας αποτελεί μετεξέλιξη της επικοινωνίας “όχημα προς δίκτυο” (Vehicle-to-Grid V2G) και είναι ένας από τους πιο καινοτόμους ερευνητικούς τομείς που σχετίζονται με τα ηλεκτρικά οχήματα και τα έξυπνα δίκτυα ηλεκτροδότησης.

Απαραίτητες γνώσεις: Java, Τεχνητή Νοημοσύνη, επιχειρησιακή έρευνα/ βελτιστοποίηση (επιθυμητή) – Γνώσεις δικτύων **δεν** απαιτούνται.

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

## B: Επέκταση και έλεγχος βιβλιοθήκης διαχείρισης λειτουργιών ηλεκτρικών οχημάτων:

Σκοπός αυτής της πτυχιακής είναι η επέκταση μιας υπάρχουσας βιβλιοθήκης διαχείρισης της φόρτισης και κίνησης ηλεκτρικών οχημάτων. Επιπλέον, στόχος είναι ο λεπτομερής έλεγχος και η αποσφαλμάτωση της βιβλιοθήκης μέσω της εκτενούς χρήσης της σε μια σειρά σεναρίων και της προσομοίωσης ενός συνόλου αλγορίθμων διαχείρισης της φόρτισης ηλεκτρικών οχημάτων.

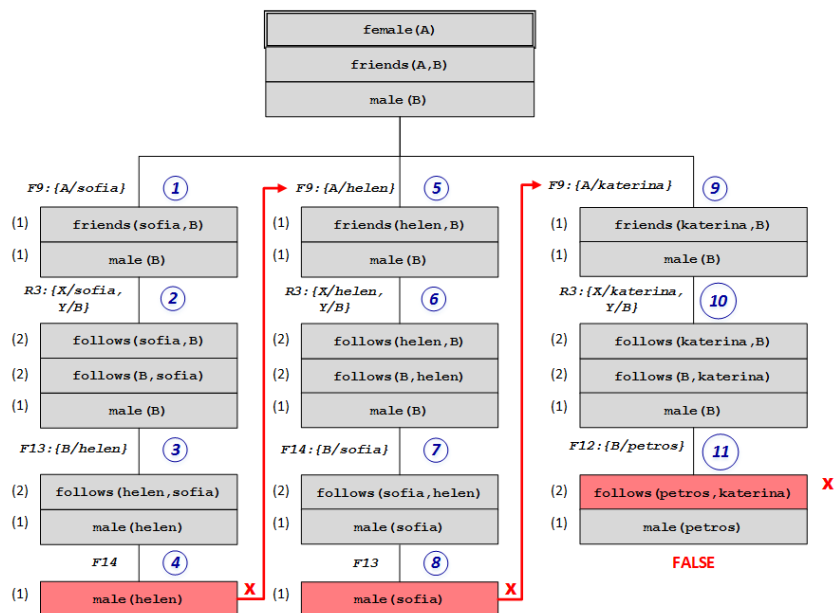
Απαραίτητες γνώσεις: Java (πολύ καλή), τεχνολογία λογισμικού (επιθυμητή)

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

### 4. Οπτική αναπαράσταση δένδρου εκτέλεσης λογικών προγραμμάτων

Η Prolog είναι η πλέον σημαντική γλώσσα λογικού προγραμματισμού και αποτελεί την βάση της κατανόησης του δηλωτικού προγραμματισμού με την χρήση λογικών κανόνων. Παρόλο που ο λογικός προγραμματισμός αποτελεί μία παλαιότερη προσπάθεια στην βελτίωση της εμπειρίας και της αποδοτικότητας στην ανάπτυξη προγραμμάτων, εντούτοις τα τελευταία χρόνια έχει επανέλθει στο προσκήνιο διαμέσου του λεγόμενου Σημασιολογικού Ιστού (Semantic Web) και των Ανοιχτών Διασυνδεδεμένων Δεδομένων (Linked Open Data), καθώς όλες οι γλώσσες αναπαράστασης και επεξεργασίας των σημασιολογικών δεδομένων βασίζεται σε μικρότερο ή μεγαλύτερο βαθμό στην λογική και στον λογικό προγραμματισμό

αντίστοιχα. Έτσι πλέον η εκμάθηση της γλώσσας Prolog αποτελεί ένα ισχυρό γνωστικό θεμέλιο σε αυτόν που θέλει να εμβαθύνει απρόσκοπτα στις αναδυόμενες τεχνολογίες του Σημασιολογικού Ιστού. Καθώς η ροή εκτέλεσης της γλώσσας Prolog διαφέρει αρκετά σε σχέση με τις υπόλοιπες γλώσσες διαδικαστικού προγραμματισμού, είναι συνηθισμένο να διδάσκεται με την βοήθεια εποπτικών αναπαραστάσεων και συγκεκριμένα με την χρήση του δένδρου εκτέλεσης. Στην πτυχιακή αυτή θα υλοποιηθεί μια συγκεκριμένη μέθοδος κατασκευής δένδρων εκτέλεσης, δείγμα της οποίας εμφανίζεται στο σχήμα.



Εργαλεία που θα χρησιμοποιηθούν: Prolog (για την ανάλυση του λογικού προγράμματος), C/C++/Java/... (οποιαδήποτε διαδικαστική γλώσσα, για την κατασκευή του γραφικού μέρους)

Περισσότερες πληροφορίες: Ν. Βασιλειάδης ([nbassili@csd.auth.gr](mailto:nbassili@csd.auth.gr))

## 5. Μετατροπή των ανοιχτών δεδομένων του ΑΠΘ σε Linked Open Data

Το ΑΠΘ είναι το 1ο ελληνικό πανεπιστήμιο που ανταποκρίθηκε στην πρόσκληση συγκρότησης "Μητρώου Ανοιχτών Δεδομένων του Δημοσίου", δημοσιεύοντας επτά σύνολα δεδομένων σε ανοικτό και μηχαναγνώσιμο μορφότυπο. Δείτε την σχετική ανακοίνωση εδώ: <https://www.auth.gr/news/articles/19195>. Τα ανοικτά δεδομένα του ΑΠΘ περιλαμβάνουν πληροφορίες για Διδάσκοντες, Πρόσωπα, Προγράμματα Σπουδών, Μαθήματα και πληροφορίες σχετικά με τις Ακαδημαϊκές και Διοικητικές Μονάδες του πανεπιστημίου. Είναι διαθέσιμα μέσω του [www.data.gov.gr](http://www.data.gov.gr), με αναλυτικές περιγραφές και παραδείγματα χρήσης, σε αυτή τη διεύθυνση: <http://www.data.gov.gr/organization/aristotle-university-of-thessaloniki>. Η μορφή υπό την οποία είναι διαθέσιμα τα δεδομένα είναι σε JSON. Μία πολύ ενεργή τάση στον χώρο των ανοιχτών δεδομένων είναι αυτά να διατίθενται υπό την μορφή των ανοιχτών διασυνδεδεμένων δεδομένων (Linked Open Data – LOD), σύμφωνα με την οποία τα δεδομένα όχι μόνο είναι ανοιχτά αλλά και διασυνδεδεμένα με άλλα δεδομένα του ίδιου ή άλλου δικτυακού τύπου. Η μορφή των δεδομένων είναι RDF, που αποτελεί το πρότυπο αναπαράστασης πληροφορίας / γνώσης στον λεγόμενο Σημασιολογικό Ιστό (Semantic Web), ο οποίος αποτελεί το επόμενο μεγάλο βήμα εξέλιξης στο web (το λεγόμενο web 3.0) 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 (γλώσσες RDF Schema ή OWL). Τα Linked Open Data είναι μία μέθοδος 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. Χαρακτηριστικό παράδειγμα LOD είναι η DBpedia, η οποία δημοσιεύει σε μορφή RDF την δομημένη πληροφορία που υπάρχει στα λήμματα της Wikipedia. Περισσότερες πληροφορίες μπορούν να βρεθούν στους ακόλουθους δεσμούς:

<http://www.w3.org/standards/semanticweb/>  
<http://linkeddata.org/>

<http://lpis.csd.auth.gr/mtpx/sw/index.htm>  
<http://wiki.dbpedia.org/>

Αντικείμενο της παρούσας πτυχιακής είναι η μετατροπή και δημοσίευση των ανοιχτών δεδομένων του ΑΠΘ σε Linked Open Data, καθώς και η διασύνδεσή τους με άλλα σύνολα δεδομένων LOD. Η εργασία περιλαμβάνει προγραμματισμό στην γλώσσα Java καθώς και εγκατάσταση/συντήρηση συστήματος βάσης δεδομένων RDF. Επίσης, θα απαιτηθεί και χρήση ή/και τροποποίηση κάποιων οντολογιών.

Περισσότερες πληροφορίες: Ν. Βασιλειάδης ([nbassili@csd.auth.gr](mailto:nbassili@csd.auth.gr))

## 6. Ανάπτυξη browser plugin για εκτέλεση κανόνων στον browser

Τα τελευταία χρόνια ο Σημασιολογικός Ιστός (Semantic Web) και τα Ανοιχτά Διασυνδεδεμένα Δεδομένα (Linked Open Data) γνωρίζουν πολύ μεγάλη ανάπτυξη καθώς ρίχνουν το κόστος δημοσίευσης, διαμοιρασμού και επανάχρησης των μεγάλων συνόλων δεδομένων που προκύπτουν από κυβερνητικές υπηρεσίες ή επιστημονικά πειράματα. Η μορφή αυτών των δεδομένων είναι RDF, μία μορφή αναπαράστασης γνώσης που βασίζεται στην λογική. Επομένως ο πιο φυσιολογικός τρόπος επεξεργασίας τους είναι μέσω γλωσσών προγραμματισμού που

βασίζονται στην λογική ή στους κανόνες επεξεργασίας IF-THEN. Καθώς όμως μεγαλώνει η ανάγκη επεξεργασίας των δεδομένων αυτών, μεγαλώνει και ο υπολογιστικός φόρτος των servers που τα φιλοξενούν. Σκοπός της διπλωματικής αυτής είναι να διερευνήσει την δυνατότητα να αναπτυχθεί σύστημα εκτέλεσης κανόνων στον client (δηλαδή browser), υπό την μορφή ενός browser plugin, και όχι στον server, προκειμένου η επεξεργασία των δεδομένων να γίνεται κατανεμημένα. Η διπλωματική δεν θέτει περιορισμό στις τεχνολογίες που θα χρησιμοποιηθούν για την γλώσσα κανόνων, για την ανάπτυξη του plugin, ή για τον browser για τον οποίο θα αναπτυχθεί το plugin. Επιστάται η προσοχή ότι πρόκειται για ένα εξαιρετικά καινοτόμο και δύσκολο θέμα.

## **Θέματα κ. Γ. Τσουμάκα**

### **7. Ανάλυση Συναισθήματος Επιστημονικών Αναφορών**

Μια επιστημονική εργασία περιλαμβάνει συνήθως ένα σύνολο βιβλιογραφικών αναφορών προς προηγούμενες εργασίες, τις οποίες αναφέρει εντός τους κειμένου της. Στόχος αυτής της πτυχιακής εργασίας είναι η ανάπτυξη ενός συστήματος, το οποίο αναλύει το κείμενο γύρω από την αναφορά (το λεγόμενο context), προκειμένου να συμπεράνει αν η αναφορά περιέχει θετικό, ουδέτερο ή αρνητικό συναίσθημα. Αυτό μπορεί να γίνει με δύο κυρίως τρόπους, είτε με χρήση λεξικού συναισθηματικών λέξεων και εργαλείων όπως το SentiWordNet<sup>1</sup>, είτε με την εκπαίδευση ενός μοντέλου ταξινόμησης με χρήση μιας συλλογής από τέτοια κείμενα με γνωστό συναίσθημα (μπορεί να κατασκευαστεί στο πλαίσιο της πτυχιακής). Απώτερος σκοπός είναι προώθηση του συστήματος ως μια νέα υπηρεσία προς όφελος επιστημόνων και ερευνητών αλλά και οργανισμών αξιολόγησης ερευνητών. Ήδη το κοινωνικό δίκτυο ResearchGate, επιτρέπει σε έναν ερευνητή να βλέπει το context των αναφορών σε εργασίες του.

### **8. Εύρεση Πολιτισμικού Περιεχομένου βάσει Γεωγραφικών Συντεταγμένων**

Στόχος αυτής της πτυχιακής εργασίας είναι η ανάπτυξη μιας εφαρμογής για κινητές συσκευές (android), η οποία δεδομένου του γεωγραφικού στίγματος του χρήστη, θα του προτείνει σχετικό πολιτισμικό περιεχόμενο. Η πτυχιακή θα εστιάσει στην τεχνολογία των ανοιχτών διασυνδεδεμένων δεδομένων (linked open data), η αξιοποίηση των οποίων προσφέρει εξαιρετικές ευκαιρίες ανάπτυξης καινοτόμων και χρήσιμων για την κοινωνία εφαρμογών. Συγκεκριμένα, προτείνεται η εφαρμογή να επιτρέπει πιλοτικά στον χρήστη να ανακαλύψει ταινίες<sup>2</sup> που έχουν γυριστεί στην συγκεκριμένη (ευρύτερη) τοποθεσία<sup>3</sup>, αλλά και τραγούδια<sup>4</sup> που έχουν εμπνευστεί από συγκεκριμένες τοποθεσίες. Για τα τραγούδια θα αξιοποιηθεί πληροφορία με στίχους από συγκεκριμένο ιστότοπο<sup>5</sup>, η οποία και θα μετατραπεί στο πλαίσιο της πτυχιακής σε μορφή ανοιχτών διασυνδεδεμένων δεδομένων. Πιλοτικά η εφαρμογή θα εστιάζει στην πόλη της Θεσσαλονίκης.

### **9. Εντοπισμός Εκφοβισμού σε Κοινωνικά Μέσα**

Στόχος αυτής της πτυχιακής εργασίας είναι αρχικά η αναζήτηση ιχνών διαδικτυακού σχολικού εκφοβισμού (cyber-bullying) στην Ελληνική γλώσσα στα κοινωνικά μέσα (twitter, facebook). Στη συνέχεια τα ίχνη αυτά θα αναλυθούν με σκοπό την κατηγοριοποίηση τους σε είδη

---

<sup>1</sup> <http://sentiwordnet.isti.cnr.it/>

<sup>2</sup> <http://data.linkedmdb.org/>

<sup>3</sup> <http://linkedgeo.org/>

<sup>4</sup> <http://dbtune.org/>

<sup>5</sup> <http://www.stixoi.info/>



εκφοβισμού<sup>6</sup>, τον διαχωρισμό του θύτη και θύματος, και την εκμάθηση μοντέλων αυτόματης αναγνώρισης νέων περιπτώσεων διαδικτυακού σχολικού εκφοβισμού μέσω τεχνικών μηχανικής μάθησης. Ο διαδικτυακός σχολικός εκφοβισμός είναι ένα φαινόμενο αρκετά διαδεδομένο στη χώρα μας και με άσχημες κοινωνικές επιπτώσεις. Με τη χρήση τεχνολογιών πληροφορικής μπορούμε να βοηθήσουμε στον έγκαιρο εντοπισμό και αντιμετώπιση του. Η δουλειά που θα γίνει θα βασιστεί στην αντίστοιχη διεθνή βιβλιογραφία<sup>7</sup>.

#### 10. Υποστήριξη και Αυτοματοποίηση Συστηματικών Ανασκοπήσεων (2 άτομων)

Οι συστηματικές ανασκοπήσεις<sup>8</sup> αποτελούν τον ακρογωνιαίο λίθο της τεκμηριωμένης ιατρικής<sup>9</sup>. Ωστόσο οι μελέτες αυτές δεν παράγονται αρκετά γρήγορα για να είναι χρήσιμες στην κλινική πρακτική. Το κόστος της παραγωγής, η διαθεσιμότητα της απαιτούμενης εμπειρογνομosύνης και η επικαιρότητα αναφέρονται συχνά ως σημαντικοί παράγοντες για αυτήν την καθυστέρηση. Απαιτούνται επομένως πληροφοριακά συστήματα υποστήριξης και αυτοματοποίησης των επιμέρους εργασιών που εμπλέκονται σε μια συστηματική ανασκόπηση<sup>10,11</sup>. Μια πτυχιακή σε αυτό το αντικείμενο θα μπορούσε να εστιάσει στο εξής θέμα: *Ανάπτυξη συστήματος συνεργατικής διαλογής περιλήψεων (abstracts) επιστημονικών δημοσιεύσεων που είναι σχετικές με το αντικείμενο μιας ανασκόπησης*. Το σύστημα θα πρέπει να επιτρέπει σε έναν βασικό χρήστη την εισαγωγή ενός μεγάλου αριθμού από περιλήψεις, οι οποίες μπορεί να προέρχονται από το PubMed (λίστα ids των περιλήψεων) ή από κάποιο άλλο σύστημα διαχείρισης βιβλιογραφίας (τουλάχιστον από το Mendeley). Στη συνέχεια το σύστημα θα πρέπει να επιτρέπει στον βασικό χρήστη την προσθήκη των χρηστών που θα αναλάβουν συνεργατικά τη διαλογή. Διαφορετικοί χρήστες θα πρέπει να εξετάζουν διαφορετικές περιλήψεις με τυχαία σειρά και να τις χαρακτηρίζουν ως σχετικές ή μη με το αντικείμενο μιας ανασκόπησης. Έπειτα από επαρκή αριθμό χαρακτηρισμένων περιλήψεων, το σύστημα θα πρέπει να εκπαιδεύει ένα μοντέλο δυαδικής ταξινόμησης, το οποίο να χρησιμοποιείται για την ταξινόμηση των υπόλοιπων περιλήψεων ως προς τη σχετικότητά τους ή ως προς την αβεβαιότητα του συστήματος για αυτές (active learning) αν το επιθυμεί ο βασικός χρήστης, αλλιώς οι περιλήψεις θα συνεχίζουν να εξετάζονται με τυχαία σειρά. Ουσιαστικά το θέμα αυτό έχει ως στόχο την δημιουργία ενός συστήματος σαν το abstract<sup>12</sup>. Το σύστημα θα πρέπει να αναπτυχθεί σε ένα Java web application framework όπως το Vaadin.

#### 11. 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.),

<sup>6</sup> <http://www.hfcbeatbullying.info/σχολικός-εκφοβισμός/>

<sup>7</sup> <http://research.cs.wisc.edu/bullying/data.html>

<sup>8</sup> [http://en.wikipedia.org/wiki/Systematic\\_review](http://en.wikipedia.org/wiki/Systematic_review)

<sup>9</sup> [http://en.wikipedia.org/wiki/Evidence-based\\_medicine](http://en.wikipedia.org/wiki/Evidence-based_medicine)

<sup>10</sup> <http://www.systematicreviewsjournal.com/content/3/1/74>

<sup>11</sup> <http://www.cochranelibrary.com/editorial/10.1002/14651858.ED000091>

<sup>12</sup> [http://www.cebm.brown.edu/static/papers/wallace\\_ihi\\_2011\\_preprint.pdf](http://www.cebm.brown.edu/static/papers/wallace_ihi_2011_preprint.pdf)

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.

## **Θέματα κ. Δ. Βράκα**

### **12. 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.

### **13. 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.

### **14. 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: PDDLStudio, itSimple και το VLEPPO. Στα πλαίσια της πτυχιακής 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.

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

Η πτυχιακή 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.

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

Η πτυχιακή εργασία περιλαμβάνει την μελέτη του συστήματος VLEPPO και την επέκτασή 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.

#### **17. Υλοποίηση Παιγνίου (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 "2048", στο οποίο το σύστημα θα πρέπει να έχει μια λειτουργία προβολής συστάσεων προς τον χρήστη με βάση αλγορίθμους TN.

Απαιτείται καλό υπόβαθρο γνώσεων σε TN και πολύ καλή γνώση προγραμματισμού σε JAVA.

#### **18. Υλοποίηση Αλγορίθμου Αποστολής Ενημερώσεων για την Κατάσταση Ηλεκτρικών Συσκευών σε Έξυπνα Σπίτια**

Η πτυχιακή εργασία περιλαμβάνει την ανάπτυξη ενός συστήματος το οποίο θα παρακολουθεί τη συνολική κατανάλωση ενέργειας ενός "έξυπνου" σπιτιού και θα στέλνει ενημερώσεις στο χρήστη όταν αντιλαμβάνεται "μη αναμενόμενες" συμπεριφορές συσκευών (π.χ. ο χρήστης ξέχασε ανοικτό το θερμοσίφωνο). 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, 11γ, 1, 3α, 16, 17.*

*Περισσότερες λεπτομέρειες για τα ερευνητικά ενδιαφέροντα των ομάδων ευφών 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: 15/10/2015. Ανακοίνωση αναθέσεων 20/10/2015.*