

**Πτυχιακές και Διπλωματικές Εργασίες για τους προπτυχιακούς και μεταπτυχιακούς φοιτητές του Τμήματος Πληροφορικής και Τηλεπικοινωνιών**

**Ακαδημαϊκό Έτος 2010-2011 (και όλα τα επόμενα έτη εφόσον αυτό το αρχείο υπάρχει ακόμα στην ιστοσελίδα μου!)**

**Αναπληρωτής Καθηγητής Μανόλης Κουμπάρκης**  
<http://www.di.uoa.gr/~koubarak>  
**Τμήμα Πληροφορικής και Τηλεπικοινωνιών**  
**ΕΚΠΑ**

**Σημειώσεις για τους φοιτητές:**

- Τα θέματα είναι κοινά για διπλωματικές και πτυχιακές, αλλά οι απαιτήσεις διαφέρουν ανάλογα με το θέμα και το επίπεδο σπουδών του φοιτητή/φοιτήτριας (προπτυχιακό/μεταπτυχιακό). Γενικά, στην περίπτωση των πτυχιακών (προπτυχιακοί φοιτητές), η έμφαση θα είναι κυρίως στην κατανόηση και υλοποίηση αλγορίθμων που έχουν αναπτυχθεί από την ομάδα μου ή άλλους ερευνητές, και είναι δημοσιευμένοι στη βιβλιογραφία. Στην περίπτωση των διπλωματικών (μεταπτυχιακοί φοιτητές), θα δίνεται έμφαση και στην επιστημονική συνεισφορά και πρωτοτυπία της εργασίας.
- Όλες οι πτυχιακές μπορούν να ανατεθούν και σε ομάδα φοιτητών με κατάλληλη τροποποίηση των απαιτήσεων. Οι διπλωματικές είναι μόνο για ένα φοιτητή ή φοιτήτρια.
- Όπου αναφέρονται σχετικές περιοχές/μαθήματα, η πληροφορία είναι ενδεικτική. Στην πράξη κάποια μαθήματα θα αποδειχθούν πολύ χρήσιμα για την εργασία σας και κάποια άλλα λιγότερο. Όμως, αν στα μαθήματα που αναφέρω έχετε πάρει χαμηλούς βαθμούς (ή δεν σας ενδιαφέρουν), μάλλον η αντίστοιχη εργασία δεν είναι για σας!
- Όλες οι εργασίες επικεντρώνονται σε θέματα που είναι στην αιχμή της έρευνας και της τεχνολογίας στις αντίστοιχες θεματικές περιοχές. Συνεπώς, οι πτυχιακές είναι κατάλληλες σαν πρώτο βήμα για την μελλοντική εκπόνηση διπλωματικής, και οι διπλωματικές σαν πρώτο βήμα για την μελλοντική εκπόνηση διδακτορικού σε σχετικό αντικείμενο.
- Αν σας ενδιαφέρει κάποιο θέμα, δείτε τη σχετική βιβλιογραφία και στείλτε μου e-mail για να συναντηθούμε και να το συζητήσουμε. Αν δεν μπορείτε να εντοπίσετε κάποια από τις δημοσιεύσεις που αναφέρω παρακάτω, μη διστάσετε να μου στείλετε e-mail.

Τα παρακάτω θέματα έχουν γραφτεί με τη βοήθεια των παρακάτω συνεργατών μου: Σταύρος Βάσσος, Ζωή Καούδη, Μάνος Καρπαθιωτάκης, Κωστής Κυζηράκος, Ίριδα Μηλιαράκη και Μπάμπης Νικολάου. Η επίβλεψη των εργασιών θα είναι από κοινού με τους συνεργάτες μου.

## Προτεινόμενες Θεματικές Περιοχές:

### 1. Ιδεατά παρατηρητήρια της Γης (virtual earth observatories) βασισμένα σε τηλεπισκοπικές πολυφασματικές εικόνες

Με τη σημερινή αλματώδη ανάπτυξη των δορυφόρων και της επιστήμης της τηλεπισκόπησης, έχει δημιουργηθεί ένας τρομακτικός όγκος δεδομένων (terabytes) που βρίσκουν καθημερινά χρήση σε περιβαλλοντολογικές εφαρμογές, στην οικολογία, στην παρατήρηση των ωκεανών, στην αρχαιολογία, στην εθνική ασφάλεια, κλπ.

Οι διάφοροι διεθνείς και εθνικοί οργανισμοί διαστήματος (NASA, ESA, DLR, κλπ.) έχουν αναπτύξει ιστοχώρους τους οποίους μπορεί να χρησιμοποιήσει κανείς για να εντοπίσει δορυφορικές εικόνες που τον ενδιαφέρουν, και στην συνέχεια να τις παραγγείλει από τον αντίστοιχο οργανισμό. Ένας τέτοιος ιστοχώρος προσφέρει απλά μια μηχανή αναζήτησης που επιτρέπει την προσπέλαση σε μεταδεδομένα του αρχείου του οργανισμού χρησιμοποιώντας τα εξής κριτήρια:

- Το γεωγραφική περιοχή που καλύπτει η εικόνα.
- Το χρόνο λήψης της εικόνας.
- Ειδικά κριτήρια που οργανώνονται σ' ένα ιεραρχικό μενού και περιγράφουν τις διαθέσιμες εικόνες.

Δείτε για παράδειγμα τον ιστοχώρο EOWEB της DLR (<http://eoweb.dlr.de:8080/servlets/template/welcome/entryPage.vm>).

Στο έργο “TELEIOS: A Virtual Observatory Infrastructure for Earth Observation”, το οποίο ξεκίνησε πρόσφατα, συνεργαζόμαστε με την DLR (Γερμανική Υπηρεσία Διαστήματος) για την ανάπτυξη κατάλληλων τεχνολογιών, ώστε ιστοχώροι, όπως ο EOWEB, να μπορούν να μεταβληθούν σε ιδεατά παρατηρητήρια της Γης που θα προσφέρουν πιο πλούσιες δυνατότητες αναζήτησης πολυφασματικών εικόνων, και θα κάνουν πιο εύκολη την ανάπτυξη σχετικών εφαρμογών.

Τα θέματα που μας ενδιαφέρουν στο έργο TELEIOS είναι, μεταξύ άλλων, τα εξής:

- i. Ψηφιακή επεξεργασία τηλεπισκοπικών πολυφασματικών εικόνων με τελικό σκοπό την εξαγωγή χαρακτηριστικών τα οποία θα μας επιτρέψουν την ανάπτυξη οντολογιών για την αποτελεσματική περιγραφή και ανάκτηση πολυφασματικών εικόνων που προέρχονται από τον δορυφόρο TerraSAR-X.
- ii. Βελτίωση υπάρχοντων τεχνικών για την ανίχνευση-παρακολούθηση πυρκαγιών και τη χαρτογράφηση καμένων εκτάσεων με χρήση δορυφορικών εικόνων και δεδομένων GIS.
- iii. Τεχνικές των βάσεων επιστημονικών δεδομένων (scientific databases, array data models and query languages) για την επεξεργασία δεδομένων πολυφασματικών εικόνων όπως αυτές χρησιμοποιούνται στα (i) και (ii). Υλοποίηση των τεχνικών αυτών στο σύστημα MonetDB (<http://monetdb.cwi.nl/>).
- iv. Υλοποίηση κατάλληλων επεκτάσεων του μοντέλου stRDF και της γλώσσας επερωτήσεων stSPARQL (<http://cgi.di.uoa.gr/~koubarak/publications/stSPARQL.pdf>) και χρήση τους στην αποτελεσματική επισημείωση και ανάκτηση πολυφασματικών εικόνων.
- v. Υλοποίηση των επεκτάσεων των stRDF/stSPARQL χρησιμοποιώντας το σύστημα MonetDB.

Στα πλαίσια του έργου αυτού υπάρχουν πτυχιακές, διπλωματικές (και διδακτορικά!) που επικεντρώνονται στα παραπάνω θέματα.

Σχετικά μαθήματα: Τεχνητή Νοημοσύνη, Βάσεις Δεδομένων, Τεχνολογίες Γνώσεων. Η γνώση τεχνικών επεξεργασίας εικόνων είναι επιθυμητή σε κάποια από τα σχετικά θέματα.

## 2. Δίκτυα όλου του κόσμου ενωθείτε!<sup>1</sup>

Τα δίκτυα αισθητήρων (sensor networks) έχουν γίνει ευρέως διαθέσιμα τα τελευταία χρόνια και χρησιμοποιούνται με επιτυχία σε διάφορους τομείς (π.χ., περιβάλλον, γεωργία, υγεία, μεταφορές, δημόσια ασφάλεια κλπ.). Στο άμεσο μέλλον θα έχουμε στη διάθεση μας διάφορα δίκτυα αισθητήρων τα οποία έχουν αναπτυχθεί από ανεξάρτητες οντότητες και τα οποία θα μας παρέχουν ένα ανεξάντλητο πλούτο δεδομένων. Η ενοποίηση αυτών των δικτύων με υπάρχουσες πηγές πληροφορίας ή υπηρεσίες διαθέσιμες στο Διαδίκτυο και τον Παγκόσμιο Ιστό θα μας δώσει την δυνατότητα να αναπτύξουμε πρωτοποριακές εφαρμογές που είναι πολύ δύσκολο ή αδύνατο να υλοποιηθούν σήμερα. Αυτό είναι το όραμα της περιοχής του **Διαδικτύου Αισθητήρων (Sensor Internet)** ή του **Παγκόσμιου Ιστού Αισθητήρων (Sensor Web)** ή του **Πλέγματος Αισθητήρων (Sensor Grid)** όπως έχουν ονομαστεί οι σχετικές πρόσφατες ερευνητικές προσπάθειες. Δείτε για παράδειγμα τα συστήματα IrisNet (<http://www.intel-iris.net/>), HourGlass (<http://www.eecs.harvard.edu/~syrah/hourglass/index.shtml>) or Global Sensor Networks (<http://lsirpeople.epfl.ch/salehi/papers/LSIR-REPORT-2006-001.pdf>).

Πολλοί ερευνητές πιστεύουν ότι μπορούμε να ανταποκριθούμε αποτελεσματικά στην ερευνητική πρόκληση που περιγράφεται παραπάνω με το πάντρεμα των τεχνολογιών των Δικτύων Αισθητήρων και του Σημασιολογικού Ιστού (Semantic Web), με τελικό στόχο την ανάπτυξη ενός **Σημασιολογικού Ιστού Αισθητήρων (Semantic Sensor Web)**. Η βασική υπόθεση εδώ είναι ότι οι τεχνολογίες που έχουν αναπτυχθεί τα τελευταία χρόνια για τον Σημασιολογικό Ιστό (technologies for managing metadata, ontologies, rules, etc.) μπορούν να χρησιμοποιηθούν με επιτυχία και στο Σημασιολογικό Ιστό Αισθητήρων για την επισημείωση (annotation) αισθητήρων, δικτύων αισθητήρων, δεδομένων και υπηρεσιών ώστε να έχουμε μεγαλύτερη διαλειτουργικότητα μεταξύ δικτύων αισθητήρων και άλλων εφαρμογών, και για να διευκολύνουμε τη σύνθεση νέων υπηρεσιών και την εύκολη ανάπτυξη mash-ups κλπ. Ένα πρόσφατο άρθρο για την περιοχή του Semantic Sensor Web είναι το εξής <http://knoesis1.wright.edu/library/resource.php?id=00311>.

Ένα πρόβλημα που μας ενδιαφέρει ιδιαίτερα σ' αυτή την περιοχή είναι η επέκταση του μοντέλου RDF (<http://www.w3.org/RDF/>) με τις έννοιες του χώρου και χρόνου ώστε να μπορεί να χρησιμοποιηθεί για αναπαράσταση χωρικών και χρονικών μεταδεδομένων (που είναι πολύ χρήσιμα για περιγραφή αισθητήρων). Πρόσφατα η ομάδα μου πρότεινε το σχετικό μοντέλο stRDF και τη γλώσσα ερωτήσεων stSPARQL τα οποία τώρα υλοποιούνται στο σύστημα Strabon (<http://cgi.di.uoa.gr/~koubarak/publications/stSPARQL.pdf>). Το μοντέλο stRDF είναι πολύ χρήσιμο στο Σημασιολογικό Ιστό Αισθητήρων ώστε να μπορούμε να αναπαραστήσουμε τη θέση ενός αισθητήρα, το χρόνο κατά τον οποίο είναι ενεργός, την τροχιά ενός κινούμενου αισθητήρα κλπ.

---

<sup>1</sup> Από την ομιλία του Παναγιώτη Τσακαλίδη «Σύγκλιση Τηλεπικοινωνιών και Πληροφορικής: Συνδέοντας το Φυσικό Κόσμο με Ασύρματα Δίκτυα Αισθητήρων» (<http://www.ics.forth.gr/~tsakalid/ppub.html>).

Η περιοχή αυτή θα απασχολήσει την ομάδα μου για τα επόμενα 2-3 χρόνια και υπάρχουν πτυχιακές/διπλωματικές (και διδακτορικές διατριβές!) στα ακόλουθα θέματα:

- Ανάπτυξη οντολογιών για την περιγραφή αισθητήρων, δικτύων αισθητήρων και σχετικών υπηρεσιών ή εφαρμογών.
- Τεχνικές υλοποίησης για το σύστημα Strabon (ανάπτυξη και υλοποίηση κατάλληλων δομών δεδομένων για το μοντέλο stRDF, ανάπτυξη τεχνικών για την αποτίμηση και τη βελτιστοποίηση επερωτήσεων της γλώσσας stSPARQL, κλπ.).
- Υπηρεσιοκεντρικές διεπαφές (service-oriented interfaces) για το σύστημα Strabon με σκοπό τη χρήση του ως registry στην υπηρεσιοκεντρική αρχιτεκτονική που σχεδιάζεται από το έργο SensorGrid4Env στο οποίο συμμετέχουμε (<http://www.sensorgrid4env.eu/>).
- Υποστήριξη συνεχών επερωτήσεων stSPARQL στο σύστημα Strabon.
- Επέκταση του συστήματος ομοτίμων κόμβων Atlas (δείτε <http://atlas.di.uoa.gr/> και άλλα σχετικά θέματα παρακάτω) ώστε να υποστηρίζει το μοντέλο stRDF και τη γλώσσα επερωτήσεων stSPARQL.

Σχετικά μαθήματα: Τεχνολογίες Διαδικτύου, Τεχνολογίες Γνώσεων, Βάσεις Δεδομένων

### 3. Περιβάλλοντα αισθητήρων όπως το SenseWeb/SensorMap

Το SenseWeb/SensorMap (<http://research.microsoft.com/en-us/projects/senseweb/>) είναι ένα ερευνητικό portal της Microsoft Research το οποίο επιτρέπει στους χρήστες του να κάνουν επερωτήσεις σε δεδομένα πραγματικού χρόνου που προέρχονται από αισθητήρες χρησιμοποιώντας διεπαφές χάρτη όπως Windows Live Local (<http://maps.live.com/>) ή Google Maps (<http://maps.google.com/>). Οι χρήστες μπορούν εύκολα να δημοσιεύουν δεδομένα πραγματικού χρόνου στο portal χρησιμοποιώντας Web service interfaces.

Υπάρχουν διάφορα ενδιαφέροντα προβλήματα που σχετίζονται με το παραπάνω portal και μπορούν να αποτελέσουν το θέμα πτυχιακών ή διπλωματικών:

- Ανάπτυξη περιβαλλοντολογικών εφαρμογών στο SenseWeb/SensorMap με έμφαση στον Ελληνικό χώρο (πτυχιακές).
- Επανασχεδιασμός/επεκτάσεις του SenseWeb/SensorMap με σημασιολογική πληροφορία ώστε να έχουμε ένα Semantic Sensor Web όπως το οραματίζονται οι συγγραφείς του άρθρου [http://en.wikipedia.org/wiki/Semantic\\_Sensor\\_Web](http://en.wikipedia.org/wiki/Semantic_Sensor_Web) (διπλωματικές και διδακτορικά).

Σχετικά μαθήματα: Τεχνολογίες Διαδικτύου, Τεχνολογίες Γνώσεων, Βάσεις Δεδομένων

### 4. Διασυνδεδεμένα δεδομένα

Το «κίνημα» των διασυνδεδεμένων δεδομένων (linked data, <http://linkeddata.org/>) έχει σκοπό να συμβάλει στην ενοποίηση δεδομένων που είναι αυτή τη στιγμή ανοικτά και διαθέσιμα στον Παγκόσμιο Ιστό ώστε να πολλαπλασιαστεί η ανάπτυξη εφαρμογών που βασίζονται σ' αυτά τα δεδομένα. Οι τεχνολογίες αυτές έχουν γίνει ιδιαίτερα δημοφιλείς τα τελευταία χρόνια και βρίσκονται σε σύμπνοια με σχετικές δράσεις ανοικτής διακυβέρνησης που έχουν αναλάβει κυβερνήσεις διαφόρων χωρών, ειδικότερα των ΗΠΑ (<http://www.data.gov/>) και του Ηνωμένου

Βασιλείου (<http://data.gov.uk/>).<sup>2</sup> Στο υπάρχον «σύννεφο» που σχηματίζεται από τις πηγές ανοικτών δεδομένων και τις διασυνδέσεις τους (<http://richard.cyganiak.de/2007/10/lod/>) μπορεί να βρει κανείς δεδομένα από την Wikipedia, τους NY Times (<http://data.nytimes.com/>), τη μουσική που παίζει το BBC (<http://www.bbc.co.uk/music>), γεωγραφικά δεδομένα, κλπ. Με τα ανοικτά διασυνδεδεμένα δεδομένα, ο Ιστός των Δεδομένων (Web of data) γίνεται πραγματικότητα, και μας επιτρέπει την έκφραση επερωτήσεων ή την ανάπτυξη εφαρμογών που δεν ήταν δυνατές στο παρελθόν. Για παράδειγμα, μια επερώτηση στο σύννεφο των ανοικτών διασυνδεδεμένων γεγονότων μπορεί να είναι η εξής: Ποιο είναι το Αγγλικό συγκρότημα του οποίου όλα τα μέλη γεννήθηκαν στο Μάντσεστερ, και το οποίο ήταν πρώτο στην λίστα των “most played artists on BBC” την εβδομάδα 20-26 Σεπτεμβρίου 2010 και την ίδια περίοδο “made the headlines” στους NY Times για δύο συνεχόμενες ημέρες.

Στην περιοχή των ανοικτών διασυνδεδεμένων δεδομένων υπάρχουν πτυχιακές και διπλωματικές στις εξής περιοχές:

- Ανάπτυξη πιλοτικών εφαρμογών (π.χ., περιβαλλοντολογικών εφαρμογών) στην Ελλάδα που βασίζονται σε ανοικτά διασυνδεδεμένα δεδομένα και τεχνολογίες του Σημασιολογικού Ιστού (Semantic Web). Παραδείγματα τέτοιων εφαρμογών στο εξωτερικό υπάρχουν στις ιστοσελίδες <http://www.data.gov/semantic/index> και <http://data.gov.uk/apps>.
- Εργαλεία και τεχνικές υλοποίησης για ανοικτά διασυνδεδεμένα δεδομένα (ειδικότερα γεωγραφικά διασυνδεδεμένα δεδομένα, δείτε το άρθρο <http://www.informatik.uni-leipzig.de/~auer/publication/linkedgeodata.pdf>).

Σχετικά μαθήματα: Τεχνολογίες Γνώσεων, Τεχνητή Νοημοσύνη, Βάσεις Δεδομένων

##### **5. Λογικές περιγραφών (description logics) για χρονικό και χωρικό λογισμό και σχετικά προγράμματα συλλογιστικής (reasoners)**

Τα πιο δημοφιλή προγράμματα συλλογιστικής για λογικές περιγραφών όπως ο Pellet (<http://clarkparsia.com/pellet/>), ο Hermit (<http://hermit-reasoner.com/>) και άλλα (<http://www.cs.man.ac.uk/~sattler/reasoners.html>) δεν υποστηρίζουν χρονικό και χωρικό λογισμό. Οι πτυχιακές και διπλωματικές που είναι διαθέσιμες σ’ αυτή την περιοχή αποσκοπούν στην επέκταση τέτοιων προγραμμάτων συλλογιστικής με μεθόδους χωρικού και χρονικού λογισμού βασισμένους σε προβλήματα ικανοποίησης περιορισμών. Για παράδειγμα, δείτε το πρόσφατο άρθρο <http://clarkparsia.com/pellet/spatial/>.

Σχετικά μαθήματα: Τεχνολογίες Γνώσεων, Τεχνητή Νοημοσύνη

---

<sup>2</sup> Η αντίστοιχη Ελληνική πρωτοβουλία ανοικτής διακυβέρνησης (<http://www.opengov.gr/home/>) δεν είναι τόσο μπροστά σ’ αυτό τον τομέα αυτή τη στιγμή (20 Σεπτεμβρίου 2010) αλλά φαντάζομαι θα προχωρήσει πολύ γρήγορα, αν κρίνουμε από τον πρόσφατο ιστοχώρο <http://geodata.gov.gr/geodata/>.

## 6. Αποδοτικές υλοποιήσεις αλγορίθμων για χωρικό και χρονικό λογισμό βασισμένο σε προβλήματα ικανοποίησης περιορισμών (CSPs)

Πολλά προβλήματα χωρικού και χρονικού λογισμού μπορούν να εκφραστούν σαν προβλήματα ικανοποίησης περιορισμών (δείτε το σχετικό άρθρο <http://cgi.di.uoa.gr/~koubarak/publications/2006/koubarakis-cp-handbook.pdf> που αναφέρεται σε προβλήματα χρονικού λογισμού ή το άρθρο <http://users.cecs.anu.edu.au/~jrenz/papers/renz-nebel-los.pdf> που αναφέρεται σε προβλήματα χωρικού λογισμού). Οι πτυχιακές και διπλωματικές εργασίες που είναι διαθέσιμες σ' αυτή την περιοχή αποσκοπούν στην αποδοτική υλοποίηση αλγορίθμων ικανοποίησης χρονικών ή χωρικών περιορισμών όπως αυτοί που αναφέρονται στα προηγούμενα άρθρα. Για παράδειγμα, δείτε την υλοποίηση μερικών γνωστών αλγορίθμων χωρικού λογισμού στο <http://users.rsise.anu.edu.au/~jrenz/software/rcc8-csp-solving-jair.tar.gz>.

## 7. Protégé plug-ins

Το Protégé είναι ένα γραφικό περιβάλλον για την διαχείριση οντολογιών (<http://protege.stanford.edu/>). Οι προτεινόμενες πτυχιακές ή διπλωματικές σ' αυτή την περιοχή αποσκοπούν στην ανάπτυξη κατάλληλων plug-ins για το Protégé (<http://protege.stanford.edu/doc/dev.html#plugins>) τα οποία προσφέρουν δυνατότητες που θεωρούμε χρήσιμες από την εμπειρία μας με το Protégé αλλά οι οποίες δεν έχουν αναπτυχθεί ακόμα. Ένα τέτοιο plug-in μπορεί να είναι ένα “Querying plug-in” που να προσφέρει την δυνατότητα σ' ένα χρήστη να γράψει μια πολύπλοκη επερώτηση (εκφρασμένη σε μια γλώσσα επερωτήσεων όπως την SPARQL) η οποία να απαντιέται με βάση την ενεργή οντολογία OWL.

Σχετικά μαθήματα: Τεχνολογίες Γνώσεων, Τεχνητή Νοημοσύνη, Βάσεις Δεδομένων

## 8. Ερωτήσεις εκφρασμένες με λέξεις-κλειδιά για δεδομένα RDF

Τα τελευταία χρόνια, ο πιο δημοφιλής τρόπος αναζήτησης πληροφορίας είναι με τη χρήση λέξεων-κλειδίων (keywords). Αυτό οφείλεται, κατά πρώτον, στη ραγδαία εξάπλωση του Παγκόσμιου Ιστού, όπου ένας τεράστιος όγκος πληροφορίας γίνεται διαθέσιμος προς αναζήτηση σε ένα κοινό παγκόσμιας εμβέλειας, και κατά δεύτερον, στο ότι οι λέξεις-κλειδιά αποτελούν έναν εύκολο και διαισθητικό τρόπο έκφρασης της αναζήτησης ανεξαρτήτως του επιπέδου τεχνικής κατάρτισης ενός ατόμου.

Την τελευταία δεκαετία μάλιστα, αυτός ο τρόπος αναζήτησης ξεπερνά τα όρια εφαρμογής του — που μέχρι τότε ήταν, στο μεγαλύτερο ποσοστό του, στο πεδίο της Ανάκτησης Πληροφορίας (Information Retrieval, IR) και του Παγκόσμιου Ιστού — και αρχίζει να εφαρμόζεται σε πεδία, όπως αυτό των Βάσεων Δεδομένων και του Σημασιολογικού Ιστού. Χαρακτηριστικές εργασίες γι' αυτά τα δύο πεδία είναι αντίστοιχα οι εξής: <http://users.cis.fiu.edu/~vagelis/publications/IRKeyword.pdf> και [http://www.aifb.kit.edu/images/5/51/2009\\_1856\\_Tran\\_Top-k\\_Explorati\\_1.pdf](http://www.aifb.kit.edu/images/5/51/2009_1856_Tran_Top-k_Explorati_1.pdf). Αυτή η παρρέκλιση δεν έχει να κάνει μόνο με τις τάσεις που επιβάλλει η ανάπτυξη της τεχνολογίας — όπου ο Σημασιολογικός Ιστός κατέχει μία από τις υψηλότερες θέσεις —, αλλά και με την ανάγκη συγχώνευσης των δεδομένων που διαχειρίζονται αυτά τα πεδία (Παγκόσμιος Ιστός, Βάσεις Δεδομένων, Σημασιολογικός Ιστός) και επίτευξης της διαλειτουργικότητας μεταξύ των αντίστοιχων εφαρμογών τους. Είναι η περίοδος όπου το κάθε πεδίο δεν μπορεί να λειτουργεί

αυτόνομα, αλλά πρέπει να προσφέρει τη μεθοδολογία και τις τεχνολογίες του στα υπόλοιπα, με μόνο σκοπό την αποτελεσματικότερη αναζήτηση της παγκόσμιας πληροφορίας. Προς αυτή την κατεύθυνση προσανατολίζεται η έρευνα που αφορά στην αναζήτηση πληροφορίας με λέξεις-κλειδιά. Ένα άρθρο που περιγράφει το όραμα για την επίτευξη αυτού του σκοπού μπορείτε να βρείτε στο <http://delivery.acm.org/10.1145/1500000/1498784/p56-weikum.pdf?key1=1498784&key2=7551235821&coll=GUIDE&dl=GUIDE&CFID=103087606&CFTOKEN=27345600><sup>3</sup>.

Αυτή η εργασία κινείται επίσης προς αυτή την κατεύθυνση και εστιάζει κυρίως στην αναζήτηση πληροφορίας με χρήση λέξεων-κλειδιών πάνω από το μοντέλο RDF. Μία σχετική και πρόσφατη εργασία είναι: [http://www.aifb.kit.edu/images/5/51/2009\\_1856\\_Tran\\_Top-k\\_Explorati\\_1.pdf](http://www.aifb.kit.edu/images/5/51/2009_1856_Tran_Top-k_Explorati_1.pdf). Στη γενική του μορφή, το πρόβλημα της αναζήτησης με λέξεις-κλειδιά για δεδομένα που παρουσιάζουν δομή γράφου, όπως είναι τα δομημένα δεδομένα στην περιοχή των βάσεων δεδομένων και τα ημιδομημένα δεδομένα (XML, RDF) στην περιοχή του διαδικτύου και του σημασιολογικού ιστού, μπορεί να χωριστεί στα παρακάτω στάδια:

1. Ερμηνεία των λέξεων-κλειδιών βάσει του περιεχομένου των αποθηκευμένων δεδομένων και του μοντέλου δεδομένων και αντιστοίχισή τους σε στοιχεία του γράφου δεδομένων. Αυτό το στάδιο εμπλέκει ζητήματα αναζήτησης πλήρους κειμένου (full-text search), όπως επίσης και κατασκευή κατάλληλων ευρετηρίων, γνωστά από την περιοχή IR.
2. Εξερεύνηση του γράφου με αρχή τα στοιχεία του 1<sup>ου</sup> σταδίου για την εύρεση υπογράφων μέσω των οποίων τα στοιχεία αυτά συνδέονται.

Αυτό το στάδιο εμπλέκει θεωρητικά ζητήματα γράφου, κατασκευή ευρετηρίου για την ταχύτερη εξευρέυνηση του γράφου, κατασκευή συναρτήσεων κατάταξης για την βαθμολόγηση της ποιότητας των αποτελεσμάτων, όπως επίσης και κατασκευή αλγορίθμων για την εύρεση των k καλύτερων (top-k) απαντήσεων.

3. Μετάφραση των υπογράφων του 2<sup>ου</sup> σταδίου σε κατάλληλες επερωτήσεις για το μοντέλο των αποθηκευμένων δεδομένων.

Αυτό το στάδιο εμπλέκει την ανάπτυξη αλγορίθμων για την μετάφραση ενός υπογράφου σε μία κατάλληλη επερώτηση πάνω από το μοντέλο των αποθηκευμένων δεδομένων.

4. Κατασκευή απάντησης τόσο από τους υπογράφους, όσο επίσης και από τα αποτελέσματα των επερωτήσεων του 3<sup>ου</sup> σταδίου.

Αυτό το στάδιο μπορεί να διαφοροποιηθεί περαιτέρω ως προς τον τύπο της απάντησης, όπου ενδέχεται να έχει τη μορφή οντοτήτων, γράφου, ελεύθερου κειμένου, κτλ.

Πιο συγκεκριμένα, σ' αυτή την εργασία κανείς θα μπορούσε να ασχοληθεί με κάποιο(α) από τα παρακάτω θέματα:

1. Την επέκταση του μοντέλου ερώτησης με λέξεις-κλειδιά καθώς επίσης και του μοντέλου RDF, ώστε να περιλαμβάνει δομές που να καλύπτουν χρονική και χωρική πληροφορία. Για παράδειγμα, κανείς θα μπορούσε να ρωτά ως εξής:
  - a. flat **in** athens **around** lycabettus hill, έτσι ώστε να αναζητήσει ένα διαμέρισμα στην Αθήνα γύρω από το λόφο του Λυκαβηττού.

---

<sup>3</sup> Ο σύνδεσμος είναι προσβάσιμος μόνο μέσω του δικτύου του τμήματός μας.

- b. hard-bop albums **between** 1950 **and** 1960, έτσι ώστε να αναζητήσει hard-bop jazz άλμπουμς τη δεκαετία του 50'.
  - c. athens population **before** World War II, έτσι ώστε να αναζητήσει πληροφορίες σχετικά με τον πληθυσμό της Αθήνας πριν την περίοδο του 2<sup>ου</sup> Παγκοσμίου Πολέμου.
2. Κατασκευή ειδικών ευρετηρίων για την ταχύτερη επεξεργασία των ερωτήσεων με λέξεις-κλειδιά. Αυτά τα ευρετήρια δύναται να αφορούν τόσο το μοντέλο RDF, όσο επίσης χρονική ή/και χωρική πληροφορία (δείτε τα σχετικά θέματα 1 και 2), καθώς επίσης και λέξεις κλειδιά.
  3. Ανάπτυξη αλγορίθμων εξερεύνησης γράφων, όπως περιγράφηκε προηγουμένως.
  4. Ανάπτυξη μεθολογίας για την αξιολόγηση συστημάτων που χρησιμοποιούν λέξεις-κλειδιά. Σε αυτό το θέμα υπάρχει εξίσου μεγάλο ερευνητικό ενδιαφέρον, καθώς οι τρέχουσες προσεγγίσεις αξιολόγησης είναι ελλιπείς και ανεπαρκείς. Χαρακτηριστική εργασία είναι η [http://km.aifb.kit.edu/ws/semsearch09/semse2009\\_18.pdf](http://km.aifb.kit.edu/ws/semsearch09/semse2009_18.pdf), όπως επίσης και η πρωτοβουλία INEX<sup>4</sup> πάνω σε δεδομένα XML.

Σχετικά μαθήματα: Βάσεις Δεδομένων, Τεχνολογίες Γνώσεων, Τεχνολογίες Διαδικτύου, Τεχνητή Νοημοσύνη

## 9. Cluster and Cloud Computing and the MapReduce framework

Τα τελευταία χρόνια σε πολλές επιχειρήσεις του Διαδικτύου (Google, Amazon, Microsoft, Yahoo!), η ανάπτυξη πρωτοποριακών δραστηριοτήτων βασίζεται σε μεγάλο βαθμό στη συλλογή και ανάλυση τεράστιου όγκου δεδομένων όπως σελίδες του Παγκόσμιου Ιστού, πληροφορίες σχετικές με τη συμπεριφορά των χρηστών μηχανών αναζήτησης, κλπ. Ο τεράστιος όγκος αυτών των δεδομένων επιβάλλει την αποθήκευση και την επεξεργασία του σε clusters ή clouds υπολογιστών.

Το προγραμματιστικό μοντέλο MapReduce, που προτάθηκε αρχικά από τους Dean και Ghemawat της Google (<http://labs.google.com/papers/mapreduce.html>), χρησιμοποιείται εδώ και καιρό από την Google και άλλες επιχειρήσεις για την ανάλυση δεδομένων όπως αυτά για τα οποία μιλήσαμε παραπάνω. Η πρόσφατη open-source υλοποίηση Hadoop της MapReduce (<http://hadoop.apache.org/>) προσφέρει τη δυνατότητα και σε ακαδημαϊκά και ερευνητικά ιδρύματα να πειραματιστούν με το προγραμματιστικό μοντέλο MapReduce και να το χρησιμοποιήσουν για την επίλυση διαφόρων ερευνητικών προβλημάτων.

Στα πλαίσια των προτεινόμενων διπλωματικών ή πτυχιακών αυτής της θεματικής περιοχής, μας ενδιαφέρει να μελετήσουμε τη χρήση του προγραμματιστικού μοντέλου MapReduce στα παρακάτω προβλήματα:

- Text indexing and information retrieval. Δείτε για παράδειγμα το βιβλίο <http://www.umiacs.umd.edu/~jimmylin/book.html>.
- SQL query processing. Δείτε για παράδειγμα το σύστημα HadoopDB (<http://db.cs.yale.edu/hadoopdb/hadoopdb.html>).
- Προβλήματα από την περιοχή της Τεχνητής Νοημοσύνης (π.χ., αλγόριθμοι μηχανικής μάθησης). Δείτε για παράδειγμα το άρθρο <http://www.cs.stanford.edu/people/ang/papers/nips06-mapreducemulticore.pdf>.

- Search Engines, Reasoning and Query Processing for the Semantic Web. Δείτε για παράδειγμα το άρθρο <http://portal.acm.org/citation.cfm?id=1449601>.
- Άλλα γνωστά προβλήματα που είναι κατάλληλα για το μοντέλο και θα ήταν χρήσιμο να τα λύσουμε σε μεγάλη κλίμακα.

Σχετικά μαθήματα: Παράλληλα Συστήματα, Λειτουργικά Συστήματα, Κατανεμημένα Συστήματα και μαθήματα ανάλογα με το αντικείμενο του προβλήματος που θα μελετήσουμε (π.χ., Τεχνητή Νοημοσύνη ή Βάσεις Δεδομένων ή Τεχνολογίες Γνώσεων).

### 10. Αποτίμηση ερωτήσεων εκφρασμένων σε SPARQL στο σύστημα Atlas

Η εργασία θα επικεντρωθεί στην υλοποίηση αλγορίθμων για την αποτίμηση ερωτήσεων εκφρασμένων σε SPARQL (<http://www.w3.org/TR/rdf-sparql-query/>) στο σύστημα ομοτίμων κόμβων Atlas.

Πληροφορίες για το σύστημα Atlas που υλοποιείται από την ομάδα μας τα τελευταία χρόνια μπορείτε να βρείτε στον ιστότοπο <http://atlas.di.uoa.gr>.

Επειδή το σύστημα Atlas έχει αναπτυχθεί πάνω από τον DHT (κατανεμημένο πίνακα κατακερματισμού - distributed hash table) Bamboo (<http://bamboo-dht.org/>), οι φοιτητές θα χρειαστεί πρώτα να μάθουν καλά το σύστημα Bamboo.

Σχετικά μαθήματα: Βάσεις Δεδομένων, Τεχνολογίες Γνώσεων, Δίκτυα Υπολογιστών, Κατανεμημένα Συστήματα.

### 11. Υλοποίηση κατανεμημένου συστήματος Δημοσιεύσεων/Συνδρομών (Publish/subscribe) για την XML

Η εργασία θα επικεντρωθεί στην υλοποίηση σε Java ενός κατανεμημένου συστήματος Δημοσιεύσεων/Συνδρομών (Publish/subscribe) για την XML πάνω από ένα DHT. Οι αλγόριθμοι που θα υλοποιηθούν θα είναι επεκτάσεις του συστήματος που περιγράφεται στα άρθρα <http://cgi.di.uoa.gr/~koubarak/publications/miliaraki08.pdf> και <http://cgi.di.uoa.gr/~koubarak/publications/miliaraki-73-cr.pdf>.

Για την υλοποίηση θα χρησιμοποιηθεί το DHT Pastry (<http://research.microsoft.com/~antr/Pastry/default.htm>).

Σχετικά μαθήματα: Βάσεις Δεδομένων, Δίκτυα Υπολογιστών, Κατανεμημένα Συστήματα.

### 12. Οργάνωση, διαχείριση, και ανάκτηση Φ.Ε.Κ. χρησιμοποιώντας τεχνολογίες του Σημασιολογικού Παγκόσμιου Ιστού

Αυτή η εργασία προσανατολίζεται στην οργάνωση, διαχείριση, και αποθήκευση των Φύλλων της Εφημερίδας της Κυβερνήσεως (Φ.Ε.Κ.), τα οποία παράγονται και διανέμονται από το Εθνικό Τυπογραφείο (<http://www.et.gr/>) σε καθημερινή βάση από το 1900. Πρόσφατα, σύμφωνα με το

άρθρο 7 του Ν. 3861/2010 (ΦΕΚ 112/13.7.2010, Τεύχος Α') η ηλεκτρονική πρόσβαση στα Φ.Ε.Κ. έγινε «ανοικτή» (ελεύθερη προς κάθε πολίτη και δωρεάν)<sup>5</sup>.

Το κίνητρο αυτής της εργασίας είναι η αποδοτική αναζήτηση Φ.Ε.Κ. Η κατοχή ενός Φ.Ε.Κ. είναι απαραίτητη όταν κανείς επιθυμεί να μελετήσει τις υπαγορεύσεις/αποφάσεις/τροποποιήσεις ενός συγκεκριμένου νόμου, προεδρικού διατάγματος, ή ακόμα τις εμπορικές και άλλου είδους συμφωνίες μεταξύ κρατών, κτλ. Κάθε τέτοιος νόμος, προεδρικό διάταγμα, και συμφωνία δημοσιεύεται σε ένα Φ.Ε.Κ.<sup>6</sup> Ο τρόπος με τον οποίο μπορεί να αναζητήσει κανείς ένα Φ.Ε.Κ. είναι μέσω της αντίστοιχης ηλεκτρονικής φόρμας αναζήτησης του Εθνικού Τυπογραφείου<sup>7</sup>. Για την εύρεση ενός συγκεκριμένου Φ.Ε.Κ., κανείς θα πρέπει να γνωρίζει το έτος κυκλοφορίας, το τεύχος, τον αύξοντα αριθμό Φ.Ε.Κ., και προαιρετικά λέξεις κλειδιά, και ημερομηνία κυκλοφορίας και δημοσίευσης. Ως εκ τούτου, είναι σχεδόν αδύνατο για έναν απλό πολίτη η ανάκτηση του<sup>8</sup>. Το πρόβλημα γίνεται ακόμα πιο περίπλοκο αν αναλογισθεί κανείς το συχνό φαινόμενο στην ελληνική νομοθεσία των συχνών τροποποιήσεων και των αντίστοιχων εξαρτήσεων τους από άλλους νόμους/αποφάσεις. Σε τέτοιες περιπτώσεις, είναι επιθυμητή η ανάκτηση ενός συνόλου Φ.Ε.Κ. με τη χρονολογική σειρά που επιβάλλει η εξάρτησή τους. Με άλλα λόγια, είναι επιθυμητή η καταγραφή και ανάκτηση της εξέλιξης ενός νόμου, καθώς επίσης και των εξαρτήσεών του<sup>9</sup>.

Στα πλαίσια αυτής της εργασίας, θα πρέπει να αναπτυχθεί μία οντολογία βάσει της οποίας θα οργανώνονται και θα αναπαρίστανται τα Φ.Ε.Κ. Το επίπεδο αναπαράστασης της πληροφορίας των Φ.Ε.Κ. μπορεί να φθάσει ακόμα και στο επίπεδο άρθρων ή παραγράφων. Θα μπορεί να καλύπτει προεδρικά διατάγματα, νόμους, κλπ. καθώς επίσης και τις τροποποιήσεις τους. Είναι, επίσης, απαραίτητη η ανάπτυξη πρότυπης εφαρμογής για την επίδειξη του τρόπου ανάκτησης της σχετικής πληροφορίας.

Σε δεύτερη φάση, αυτή η εργασία μπορεί να συνδυαστεί με τη Θεματική Περιοχή 8 για την παροχή μιας διεπαφής αναζήτησης με χρήση λέξεων-κλειδιών.

Σχετικά μαθήματα: Τεχνολογίες Γνώσεων, Τεχνητή Νοημοσύνη, Βάσεις Δεδομένων

---

<sup>5</sup> Μπορείτε να βρείτε το σχετικό Φ.Ε.Κ., σε ηλεκτρονική μορφή, στον παρακάτω σύνδεσμο:  
<http://www.et.gr/idos-nph/search/pdfViewerForm.html?args=5C7QrtC22wGyK2xFpSwMnXdtvSoClrL8XL8tG9fmXDI5MXD0LzQTLWPU9yLzB8V68knBzLCmTXKaO6fpVZ6Lx9hLslJUqeiQaB5V0WvEtW4s52npxTZ6Iy0SgQn3RJV7wAFWXhuepyM>.

<sup>6</sup> Πληροφορίες για τα διάφορα τεύχη στα οποία διακρίνονται τα Φ.Ε.Κ. μπορείτε να βρείτε στον παρακάτω σύνδεσμο:

[http://www.et.gr/index.php?option=com\\_content&view=article&id=149&Itemid=181&lang=en](http://www.et.gr/index.php?option=com_content&view=article&id=149&Itemid=181&lang=en)

<sup>7</sup> [http://www.et.gr/index.php?option=com\\_wrapper&view=wrapper&Itemid=104&lang=en](http://www.et.gr/index.php?option=com_wrapper&view=wrapper&Itemid=104&lang=en)

<sup>8</sup> Ενδεικτικά, αναφέρεται ότι για το έτος 2000, το τεύχος Β' κυκλοφόρησε 999 Φ.Ε.Κ. από τα οποία μόνο τα 200 μπορούν να προσπελασθούν μέσω της φόρμας αναζήτησης του Εθνικού Τυπογραφείου.

<sup>9</sup> Ως πρόκληση, δοκιμάστε να βρείτε το Φ.Ε.Κ. που αντιστοιχεί στον κανονισμό λειτουργίας του προγράμματος μεταπτυχιακών σπουδών που προσφέρει το τμήμα μας. Ως μεγαλύτερη πρόκληση, βρείτε το σύνολο των Φ.Ε.Κ. που αναφέρονται στον ίδιο κανονισμό από το πρώτο έτος προσφοράς του μεταπτυχιακού προγράμματος σπουδών από το τμήμα μας. Να χρησιμοποιήσετε αποκλειστικά την ηλεκτρονική φόρμα του Εθνικού Τυπογραφείου.

### 13. Υλοποίηση και αξιολόγηση τεχνικών Τεχνητής Νοημοσύνης σε video games

Στην εργασία αυτή θα μελετηθεί η δυνατότητα εφαρμογής τεχνικών από το πεδίο της Τεχνητής Νοημοσύνης για τον έλεγχο χαρακτήρων (Non-Player Characters ή NPCs) σε εμπορικά video games.

Στα εμπορικά video games η επιλογή των ενεργειών που πραγματοποιεί ένας χαρακτήρας (π.χ. αν θα επιτεθεί ή αν θα προστατευτεί πίσω από το κοντινότερο κάλυμμα) βασίζεται κατά κανόνα σε προ-αποφασισμένες συμπεριφορές που ενεργοποιούνται ανάλογα με την τρέχουσα κατάσταση του χαρακτήρα και του παίκτη. Η αύξηση της υπολογιστικής ισχύος στις κονσόλες και τους προσωπικούς υπολογιστές έχει επιτρέψει τα τελευταία χρόνια τη χρήση κάποιων απλοποιημένων τεχνικών από το πεδίο της τεχνητής νοημοσύνης για τη δημιουργία χαρακτήρων με πιο ευέλικτες συμπεριφορές. Δείτε π.χ. Τα άρθρα:

- Three States and a Plan: The A.I. of F.E.A.R.  
[http://web.media.mit.edu/~jorkin/gdc2006\\_orkin\\_jeff\\_fear.pdf](http://web.media.mit.edu/~jorkin/gdc2006_orkin_jeff_fear.pdf)
- Symbolic Representation of Game World State: Toward Real-Time Planning in Games  
<http://web.media.mit.edu/~jorkin/WS404OrkinJ.pdf>

Στα πλαίσια της εργασίας θα υλοποιηθεί μια βιβλιοθήκη για τον έλεγχο χαρακτήρων σε video games χρησιμοποιώντας τεχνικές για την αναπαράσταση και ανανέωση της γνώσης του χαρακτήρα ως αυτόνομης οντότητας μέσα στο παιχνίδι, και στην πραγματοποίηση σχεδιασμού (planning) για τη λήψη αποφάσεων. Επιπλέον θα αναπτυχθεί ένα κατάλληλο mini-game για την πρακτική αξιολόγηση των τεχνικών για το σχεδιασμό ενεργειών σε σχέση με την χρονική απόκριση και τις απαιτήσεις ενός εμπορικού video game.

Για την υλοποίηση προτείνεται η πλατφόρμα Unity3D (<http://unity3d.com/unity/>) η οποία απλοποιεί πολλά στάδια στη διαδικασία υλοποίησης ενός video game και δίνει τη δυνατότητα να επικεντρωθεί κανείς στον κώδικα για τον έλεγχο των χαρακτήρων στο παιχνίδι. Για παράδειγμα δείτε το tutorial για τη δημιουργία ενός “first person shooter” μαζί με το συνοδευτικό project: <http://unity3d.com/support/resources/tutorials/fpstutorial>. Επίσης, δείτε περισσότερα παραδείγματα εδώ: <http://unity3d.com/support/resources/example-projects/>.

Σχετικά μαθήματα: Τεχνητή Νοημοσύνη, Τεχνητή Νοημοσύνη II (δίνεται το εαρινό εξάμηνο του ακαδημαϊκού έτους 2010-2011 για πρώτη φορά ως Ειδικά Θέματα).