ΥΠΟΥΡΓΕΙΟ ΠΑΙΔΕΙΑΣ ΔΙΑ ΒΙΟΥ ΜΑΘΗΣΗΣ ΚΑΙ ΘΡΗΣΚΕΥΜΑΤΩΝ
ΓΕΝΙΚΗ ΓΡΑΜΜΑΤΕΙΑ ΕΡΕΥΝΑΣ & ΤΕΧΝΟΛΟΓΙΑΣ
ΕΡΕΥΝΗΤΙΚΟ ΚΕΝΤΡΟ «Αθηνά»

ΙΝΣΤΙΤΟΥΤΟ ΠΛΗΡΟΦΟΡΙΑΚΩΝ ΣΥΣΤΗΜΑΤΩΝ (Ι.Π.ΣΥ.)

Διπλωματικές Εργασίες Ακαδ. Έτους 2013-2014

Περιεχόμενα

ΕΙΣΑΓΩΓΗ

Το **Ινστιτούτο Πληροφοριακών Συστημάτων (ΙΠΣΥ - http://www.imis.athena-innovation.gr/)** είναι το νεότερο ινστιτούτο του **Ερευνητικού Κέντρου Καινοτομίας στις Τεχνολογίες της Πληροφορίας, των Επικοινωνιών και της Γνώσης "Αθηνά" (http://www.athena-innovation.gr/)**. Ξεκίνησε την λειτουργία του τον Μάρτιο του 2007, με διευθυντή τον Τίμο Σελλή, Καθηγητή της Σχολής Ηλεκτρολόγων Μηχ. και Μηχ. Υπολογιστών Ε.Μ.Π.

Το ΙΠΣΥ διεξάγει έρευνα και συμμετέχει σε αναπτυξιακά έργα στον τομέα της διαχείρισης και επεξεργασίας δεδομένων σε πληροφοριακά συστήματα μεγάλης κλίμακας. Οι περιοχές ερευνητικής δραστηριότητας του ΙΠΣΥ περιλαμβάνουν:

- Συστήματα διαχείρισης γεω-πληροφορίας και υπηρεσίες εντοπισμού κινούμενων αντικειμένων.
- Διαχείριση Δεδομένων σε Μεγάλης Κλίμακας Κατανεμημένα Περιβάλλοντα (Data Web).
- Επιστημονικές βάσεις δεδομένων Ιστού και διαχρονικές βάσεις δεδομένων.
- Προχωρημένες τεχνικές αναζήτησης πληροφορίας στον Ιστό με έμφαση στην προσωποποίηση δεδομένων στο προφίλ του χρήστη.
- Προστασία ιδιωτικότητας δεδομένων Ιστού.

Στα πλαίσια των περιοχών αυτών, το ΙΠΣΥ ανακοινώνει μια σειρά από διπλωματικές εργασίες. Οι εργασίες θα εκπονηθούν στο ΙΠΣΥ σε συνεργασία με το **Εργαστήριο Συστημάτων Βάσεων Γνώσεων και Δεδομένων της Σχολής Ηλεκτρολόγων Μηχ. και Μηχ. Υπολογιστών Ε.Μ.Π (ΕΣΒΓΔ - http://web.dbnet.ntua.gr/en/home.html)**.

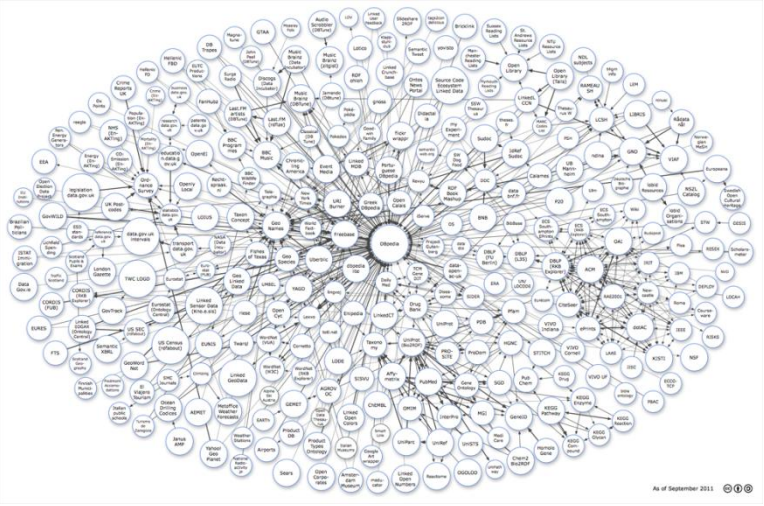
ΔΙΑΣΥΝΔΕΔΕΜΕΝΑ ΔΕΔΟΜΕΝΑ (LINKED DATA)
ΕΙΣΑΓΩΓΗ

Η εξέλιξη και ευρεία χρήση του *διαδικτύου* (*Internet*) ως μέσου διάθεσης και διακίνησης μεγάλου όγκου δομημένης πληροφορίας και γνώσης έχει οδηγήσει στην ανάπτυξη του *Σημασιολογικού Ιστού* (*Semantic Web*) και ειδικότερα των διασυνδεδεμένων (ανοιχτών) δεδομένων (Linked –Open- Data). Τα Linked Data αποτελούν ένα σύνολο από τεχνικές και εργαλεία που στοχεύουν στην δημοσίευση, κοινή αναπαράσταση, σύνδεση και διαλειτουργικότητα δεδομένων προερχόμενων από ετερογενή και κλειστά πληροφοριακά συστήματα. Τα Linked Data συνήθως αναπαρίστανται με το *RDF* (*Resource Description Framework*) μοντέλο και η *SPARQL* (*Simple Protocol and RDF Query Language*) έχει καθιερωθεί ως η γλώσσα εφαρμογής επερωτήσεων σε τέτοιου είδους δεδομένα. Η περιγραφή των δεδομένων και κατ' επέκταση η σύνταξη SPARQL ερωτημάτων σε αυτά γίνεται με ανοιχτά vocabularies όπως είναι το RDFS, το FOAF και η OWL, κτλ. Τα Linked Data αποτελούν μια δημοφιλή μέθοδο που υιοθετείται από ετερογενείς παρόχους δεδομένων, όπως π.χ., δημόσιοι φορείς, ηλεκτρονικές βάσεις δεδομένων, ψηφιακές βιβλιοθήκες, χάρτες (linkedgeodata.org) και εγκυκλοπαίδειες (π.χ.. DBPedia), κτλ.

Τα Linked Data χρησιμοποιούν απλές τεχνολογίες ιστού (URIs και links μεταξύ τους) για την αναπαράσταση και τη διασύνδεση αντικειμένων και εννοιών στο Web. Βασίζονται σε 4 βασικές αρχές.

- *Χρήση URIs* για τον προσδιορισμό αντικειμένων στο Web. URIs ανατίθενται όχι μόνο σε έγγραφα και web σελίδες (π.χ., html, pdf), αλλά και σε αντικείμενα του πραγματικού κόσμου, όπως είναι πρόσωπα, τοποθεσίες, έννοιες, κτλ.
- *Χρήση HTTP URIs* έτσι ώστε η αναφορά και πρόσβαση σε κάθε αντικείμενο να γίνεται από ανθρώπους και από υπολογιστές.
- *Χρήση ανοιχτών προτύπων* (RDF\XML) για την αναπαράσταση των δεδομένων που περιέχονται σε κάθε URI.
- Δημιουργία συνδέσεων (links) από τα δεδομένα που περιέχονται σε κάθε URI προς άλλα σχετικά URIs, για την δημιουργία συσχετίσεων στο web.

Η δημοσίευση και χρήση δεδομένων στον Ιστό με τη μορφή Διασυνδεδεμένων δεδομένων (Linked Data) αποτελεί την βασική τεχνολογία η οποία εξελίσσει σήμερα το Web από ένα δίκτυο διασυνδεδεμένων εγγράφων (Web of documents) σε ένα δίκτυο διασυνδεδεμένων δεδομένων (Web of Data), όπως φαίνεται και στην εικόνα[1], η οποία απεικονίζει τα Linked Datasets που είναι δημοσιευμένα σήμερα στο web και το δίκτυο συνδέσεων που σχηματίζουν.



[1] http://richard.cyganiak.de/2007/10/lod/

Ο Tim Berners-Lee, ο εφευρέτης του Παγκόσμιου Ιστού και των Συνδεδεμένων Δεδομένων, πρότεινε σχηματικά μια κατηγοριοποίηση[2] ***5 κανόνων*** για τα δεδομένα που διατίθενται στον ιστό και ειδικότερα για τα LOD:

★	Η πληροφορία θα πρέπει να διατίθεται στο Web (ανεξαρτήτου format) με άδεια (πρόσβασης.
★★	Η πληροφορία θα πρέπει να διατίθεται στο Web με δομημένο τρόπο (π.χ., δεδο Excel αντί για την εικόνα ενός πίνακα).
★★★	Θα πρέπει να γίνεται χρήση ανοιχτών και μη εμπορικών προτύπων για τη μορφοποί δεδομένων (π.χ., CSV αντί για Excel).
★★★★	Θα πρέπει να γίνεται χρήση URIs για τον προσδιορισμό αντικειμένων, έτσι ώστε δυνατή η αναφορά και οι δεικτοδότηση τους από άλλους χρήστες.
★★★★★	Θα πρέπει να δημιουργούνται links προς άλλα δεδομένα που περιγράφουν πληροφορία με στόχο την προσθήκη σημασιολογίας.

Η διαθεσιμότητα δομημένης πληροφορίας στον ιστό ανοίγει νέες δυνατότητες για την ανάπτυξη σύγχρονων διαδικτυακών εφαρμογών και οικοσυστημάτων τα οποία θα μπορούν να προσπελαύνουν το LOD δίκτυο και θα παρέχουν δεδομένα προστιθέμενης αξίας στους χρήστες. Οι παραπάνω τεχνολογίες προωθούν την ιδέα μια «ενοποιημένης» βάσης δεδομένων (global database), ενός ενιαίου δικτύου δεδομένων που θα παρέχει ομοιόμορφη μορφή αναπαράστασης, αναζήτησης και πρόσβασης σε δομημένη πληροφορία.

Το ΙΠΣΥ έχει ήδη αναπτύξει μια έντονη ερευνητική δραστηριότητα στη περιοχή αυτή και στοχεύει να διευρύνει τόσο ερευνητικά όσο και τεχνολογικά θέματα που σχετίζονται με τη δημοσίευση και διάθεση διαφόρων τύπων δεδομένων με την μορφή LOD, την αναζήτηση και την εφαρμογή επερωτήσεων σε LOD, την καταγραφή της εξέλιξής τους και τη διατήρησή τους στο χρόνο και τέλος την οπτική αναπαράσταση και πλοήγηση σε LOD. Στα πλαίσια αυτά, για φέτος θα δοθούν ένα σύνολο από διπλωματικές που αντιστοιχούν στα παραπάνω θέματα.

Γενικά για LOD.

http://linkeddata.org/

http://www.w3.org/DesignIssues/LinkedData.html

[2] http://5stardata.info/

ΑΝΑΚΤΗΣΗ ΚΑΙ ΜΕΤΑΤΡΟΠΗ ΠΛΗΡΟΦΟΡΙΩΝ ΚΟΙΝΩΝΙΚΩΝ ΔΙΚΤΥΩΝ ΣΕ ΔΙΑΣΥΝΔΕΔΕΜΕΝΑ ΔΕΔΟΜΕΝΑ

ΠΛΗΡΟΦΟΡΙΕΣ: Γ. Αλεξίου, Μ. Μεϊμάρης, Γ. Παπαστεφανάτος, 210 6875446
{galexiou,m.meimaris,gpapas}@imis.athena-innovation.gr

ΠΕΡΙΛΗΨΗ: Η διαχείριση πόρων σε συνεργατικά περιβάλλοντα εντός του διαδικτύου μπορεί να επωφεληθεί από τη χρήση τεχνολογιών Σημασιολογικού Ιστού (Semantic Web) και Διασυνδεδεμένων Δεδομένων (Linked Data) ώστε οι πόροι να αποθηκεύονται και να προβάλλονται με έξυπνες και πολυδιάστατες μεθόδους. Στην παρούσα εργασία ο στόχος είναι η ανάκτηση προσωπικών δεδομένων χρηστών από κοινωνικά δίκτυα, η μετατροπή τους σε πόρους του Σημασιολογικού Ιστού (Semantic Web) και ο περαιτέρω εμπλουτισμός τους μέσω της τεχνολογίας των Διασυνδεδεμένων Δεδομένων (Linked Data).

ΑΤΟΜΑ: 1.

ΠΛΑΤΦΟΡΜΑ ΕΡΓΑΣΙΑΣ: PHP, JavaScript, Java, SPARQL .

ΣΥΝΤΟΜΗ ΠΕΡΙΓΡΑΦΗ: Στα πλαίσια των δραστηριοτήτων του ΙΠΣΥ γύρω από τα διασυνδεδεμένα δεδομένα, αναπτύσσεται ένα συνεργατικό περιβάλλον διαχείρισης πόρων (π.χ. αρχεία) βασισμένο εξολοκλήρου σε τεχνολογίες Linked Data. Στη συγκεκριμένη εργασία ο στόχος είναι η ανάπτυξη web εφαρμογής για τον εμπλουτισμό του προφίλ ενός χρήστη του συνεργατικού περιβάλλοντος με πληροφορία προερχόμενη από τα κοινωνικά δίκτυα. Πιο συγκεκριμένα, θα πρέπει να γίνει α) διασύνδεση της πλατφόρμας με τα APIs των κυριοτέρων κοινωνικών δικτύων (Facebook, twitter, linkedin) β) ανάκτηση των προσωπικών πληροφοριών που θα επιλέξει ο εκάστοτε χρήστης γ) μετατροπή αυτών σε πόρους του Σημασιολογικού Ιστού (Semantic Web) και τέλος δ) επεξεργασία αυτών και εμπλουτισμός των μεταδεδομένων τους με την τεχνολογία των Διασυνδεδεμένων Δεδομένων (Linked Data).

Στα καθήκοντα της ανάληψης της συγκεκριμένης διπλωματικής εργασίας εντάσσονται:

- Ανάπτυξη λογισμικού για τη διασύνδεση των APIs και της πλατφόρμας
- Ανάπτυξη λογισμικού για την μετατροπή των δεδομένων σε πόρους του Σ.Ι. (RDFize)
- Σχεδιασμός και ανάπτυξη οντολογίας και χρήση υφισταμένων οντολογιών
- Ανάπτυξη λογισμικού για τη διασύνδεση των πόρων και τον περαιτέρω εμπλουτισμό των μεταδεδομένων τους

ΣΧΕΤΙΚΟ ΥΛΙΚΟ:

1. http://developers.facebook.com/,
2. https://dev.twitter.com/ ,
3. http://developer.linkedin.com/apis

ΑΝΑΠΤΥΞΗ MOBILE ΕΦΑΡΜΟΓΗΣ ΓΙΑ ΤΗ ΜΕΤΑΤΡΟΠΗ ΓΕΩΧΩΡΙΚΗΣ ΚΑΙ ΧΡΟΝΙΚΗΣ ΠΛΗΡΟΦΟΡΙΑΣ ΣΕ ΔΙΑΣΥΝΔΕΔΕΜΕΝΑ ΔΕΔΟΜΕΝΑ

ΠΛΗΡΟΦΟΡΙΕΣ: Γ. Αλεξίου, Μ. Μεϊμάρης, Γ. Παπαστεφανάτος, 210 6875446
{galexiou,m.meimaris,gpapas}@imis.athena-innovation.gr

ΠΕΡΙΛΗΨΗ: Η διαχείριση πόρων σε συνεργατικά περιβάλλοντα εντός του διαδικτύου μπορεί να επωφεληθεί από τη χρήση τεχνολογιών Σημασιολογικού Ιστού (Semantic Web) και Διασυνδεδεμένων Δεδομένων (Linked Data) ώστε οι πόροι να αποθηκεύονται και να προβάλλονται με έξυπνες και πολυδιάστατες μεθόδους. Στην παρούσα εργασία, στόχος είναι η δημιουργία mobile εφαρμογής για την ανάρτηση πόρων σε πλατφόρμα συνεργατικού περιβάλλοντος (η οποία αναπτύσσεται από το ΙΠΣΥ) και ο αυτόματος εμπλουτισμός τους με γεωχωρική και χρονική πληροφορία από την κινητή συσκευή. Ο εμπλουτισμός θα βασιστεί σε τεχνολογίες διασυνδεδεμένων δεδομένων.

ΑΤΟΜΑ: 1.

ΠΛΑΤΦΟΡΜΑ ΕΡΓΑΣΙΑΣ: PHP, JavaScript, Java, ANDROID SDK, SPARQL .

ΣΥΝΤΟΜΗ ΠΕΡΙΓΡΑΦΗ: Στα πλαίσια των δραστηριοτήτων του ΙΠΣΥ γύρω από τα διασυνδεδεμένα δεδομένα, αναπτύσσεται ένα συνεργατικό περιβάλλον διαχείρισης πόρων (π.χ. έγγραφα, φωτογραφίες, κτλ) βασισμένο εξολοκλήρου σε τεχνολογίες Linked Data. Στη συγκεκριμένη εργασία ο στόχος είναι η ανάπτυξη mobile εφαρμογής (είτε web interface είτε native mobile application (Android SDK)) για τον εμπλουτισμό των πόρων που αναρτά ένας χρήστης στο συνεργατικό περιβάλλον με γεωχωρικά και χρονικά δεδομένα τα οποία ανακτώνται δυναμικά από την κινητή συσκευή του χρήστη. Πιο συγκεκριμένα, θα πρέπει να υλοποιηθεί εφαρμογή που να δίνει α) δυνατότητα ανάρτησης πόρων (πχ. φωτογραφίες) από κινητές συσκευές (android smartphones) στην πλατφόρμα συνεργατικού περιβάλλοντος β) ανάκτηση των γεωχωρικών και χρονικών δεδομένων και συσχετισμός τους με τον εκάστοτε πόρο γ) μετατροπή αυτών σε πόρους του Σημασιολογικού Ιστού (Semantic Web) δ) επεξεργασία αυτών και εμπλουτισμός των μεταδεδομένων τους με την τεχνολογία των Διασυνδεδεμένων Δεδομένων (Linked Data) και ε) ανάπτυξη της εφαρμογής (σε περιβάλλον android) που θα συνδυάζει τις παραπάνω λειτουργικότητες.

Στα καθήκοντα της ανάληψης της συγκεκριμένης διπλωματικής εργασίας εντάσσονται:

- Ανάπτυξη λογισμικού για τη διασύνδεση της πλατφόρμας με το android SDK
- Ανάπτυξη λογισμικού για την μετατροπή των δεδομένων σε πόρους του Σ.Ι. (RDFize)
- Σχεδιασμός και ανάπτυξη οντολογίας και χρήση υφισταμένων οντολογιών για την αναπαράσταση και την εύκολη διαχείριση γεωχωρικών και χρονικών δεδομένων
- Ανάπτυξη λογισμικού για τη διασύνδεση των πόρων και τον περαιτέρω εμπλουτισμό των μεταδεδομένων τους

ΣΧΕΤΙΚΟ ΥΛΙΚΟ:

1. http://www.w3.org/2005/Incubator/geo/XGR-geo-ont-20071023/ , http://www.w3.org/TR/owl-time/
2. http://developer.android.com/index.html
3. http://dev.w3.org/geo/api/spec-source.html

ΕΝΣΩΜΑΤΩΣΗ ΑΝΟΙΚΤΩΝ ΓΕΩΓΡΑΦΙΚΩΝ ΔΕΔΟΜΕΝΩΝ ΣΤΟΝ ΣΗΜΑΣΙΟΛΟΓΙΚΟ ΙΣΤΟ

ΠΛΗΡΟΦΟΡΙΕΣ: Κώστας Πατρούμπας, 210 772 1446, kpatro@dblab.ece.ntua.gr

ΠΕΡΙΛΗΨΗ: Στόχος της εργασίας είναι η επέκταση της υπάρχουσας εφαρμογής *TripleGeo* με πρόσθετες δυνατότητες, ώστε να ανταποκρίνεται στις απαιτήσεις μετατροπής μεγάλης κλίμακας και ποικιλίας ανοικτών γεωγραφικών δεδομένων για την ενσωμάτωσή τους στον Σημασιολογικό Ιστό.

ΑΤΟΜΑ: 1

ΠΛΑΤΦΟΡΜΑ ΕΡΓΑΣΙΑΣ: Java, RDF, REST API.

ΣΥΝΤΟΜΗ ΠΕΡΙΓΡΑΦΗ: Χάρη στο μοντέλο λειτουργίας του *Σημασιολογικού Ιστού* (Semantic Web), έχουν αναπτυχθεί τεχνικές και εργαλεία για την διασύνδεση ανοικτών δεδομένων που αφορούν μεν την ίδια οντότητα, αλλά τα οποία πιθανόν προέρχονται από πολλαπλές, ανεξάρτητες και μάλλον ετερογενείς πηγές. Είναι χαρακτηριστικό ότι αυτού του είδους οι πληροφορίες συνήθως εμπεριέχουν (ρητώς ή εμμέσως) μία *γεωγραφική* πτυχή. Λ.χ. όταν ένας χρήστης ψάχνει για κάποια ταινία στο Διαδίκτυο ή απ' το κινητό του, η εφαρμογή θα μπορούσε να επιστρέφει όχι μόνο φωτογραφίες, κριτικές ή σχόλια για την ταινία, αλλά επίσης τις ώρες προβολής της σε κοντινούς κινηματογράφους μαζί με την τοποθεσία τους πάνω σε χάρτη.

Για να διευκολυνθεί η ενσωμάτωση τέτοιων *διασυνδεδεμένων γεωγραφικών δεδομένων* (*linked geospatial data*) στον Σημασιολογικό Ιστό, το ΙΠΣΥ έχει ξεκινήσει την ανάπτυξη της πλατφόρμας ***TripleGeo*** σε *ανοικτό κώδικα*. Στην τρέχουσα έκδοσή της, η εφαρμογή επιτρέπει την άντληση γεωγραφικών και περιγραφικών στοιχείων από υπάρχουσες πηγές (λ.χ. αρχεία ή βάσεις δεδομένων) και την εξαγωγή τους σε διάφορες μορφές κατάλληλες για τήρηση σε *αποθετήρια* (RDF stores) στο Διαδίκτυο.

Στα πλαίσια της διπλωματικής, θα επιδιωχθεί ο εμπλουτισμός της TripleGeo με πρόσθετες λειτουργίες επεξεργασίας, καθώς και με δυνατότητες χειρισμού μεγάλου πλήθους δεδομένων. Ενδεικτικά προτείνονται:

- Ανάπτυξη web interface και επέκταση των υπαρχουσών επιλογών εισόδου/εξόδου, λ.χ. επιτρέποντας ανάκτηση γεωγραφικών δεδομένων από αρχεία KML, GML κ.ά., καθώς και απ' ευθείας εισαγωγής των αποτελεσμάτων σε συγκεκριμένο αποθετήριο (λ.χ. Virtuoso RDF store).
- Υλοποίηση ενός περιβάλλοντος RESTful API, ώστε να είναι δυνατή η κλήση της εφαρμογής για ανοικτά γεωγραφικά δεδομένα προσβάσιμα μόνο μέσω Διαδικτύου.
- Σχεδιασμός και υλοποίηση μηχανισμού παράλληλης επεξεργασίας, λ.χ. με επιμερισμό του όγκου των στοιχείων για εκτέλεση από ανεξάρτητες εικονικές μηχανές (Virtual Machines), προκειμένου να αντιμετωπίζονται κλιμακούμενοι όγκοι δεδομένων (λ.χ. όλο το οδικό δίκτυο της Ευρώπης).
- Εφαρμογή της πλατφόρμας για την μετατροπή γεωγραφικών δεδομένων από ανοικτές πηγές (λ.χ. OpenStreetMap) και δυνατότητα εκθέσεώς τους στο Διαδίκτυο μέσω SPARQL endpoints.
- Μετρήσεις επιδόσεων της επεξεργασίας για διάφορους όγκους γεωγραφικών δεδομένων.

ΣΧΕΤΙΚΟ ΥΛΙΚΟ:

- TripleGeo: https://web.imis.athena-innovation.gr/redmine/projects/geoknow_public/wiki/TripleGeo
- REST API tutorial: http://rest.elkstein.org/2008/02/what-is-rest.html
- OpenStreetMap (OSM): http://www.openstreetmap.org/
- SPARQL protocol: http://www.w3.org/TR/rdf-sparql-protocol/

ΔΗΜΙΟΥΡΓΙΑ ΥΠΟΔΟΜΗΣ ΚΑΤΑ INSPIRE ΓΙΑ ΔΙΑΣΥΝΔΕΔΕΜΕΝΑ ΓΕΩΓΡΑΦΙΚΑ ΔΕΔΟΜΕΝΑ

ΠΛΗΡΟΦΟΡΙΕΣ: Κώστας Πατρούμπας, 210 772 1446, kpatro@dblab.ece.ntua.gr

ΠΕΡΙΛΗΨΗ: Στόχος της διπλωματικής είναι να σχεδιασθούν και να υλοποιηθούν διαδικτυακές υπηρεσίες που θα επιτρέπουν την διαμόρφωση και ανάρτηση ανοικτών, διασυνδεδεμένων γεωγραφικών δεδομένων και μεταδεδομένων σύμφωνα με την κοινοτική οδηγία INSPIRE.

ΑΤΟΜΑ: 1.

ΠΛΑΤΦΟΡΜΑ ΕΡΓΑΣΙΑΣ: Java, JavaScript, RDF, (πιθανόν και OpenLayers).

ΣΥΝΤΟΜΗ ΠΕΡΙΓΡΑΦΗ: Ήδη σε ισχύ από το 2007, η κοινοτική οδηγία INSPIRE αποσκοπεί στην συγκρότηση κοινής υποδομής χωρικών δεδομένων για όλα τα κράτη-μέλη της Ευρωπαϊκής Ένωσης. Θεσπίζοντας σειρά κανόνων για την τήρηση, επεξεργασία και ανταλλαγή γεωγραφικών δεδομένων μεταξύ οργανισμών, κυβερνητικών φορέων κλπ., επιδιώκει να διευκολύνει την λήψη αποφάσεων σε διάφορους τομείς (περιβάλλον, μεταφορές, οικιστική ανάπτυξη κ.ά.). Εξάλλου, ολοένα μεγαλύτερος όγκος και ευρύτερη ποικιλία γεωγραφικών δεδομένων συνεισφέρονται ή επικαιροποιούνται απ' τους ίδιους τους χρήστες, όπως συμβαίνει λ.χ. στους χάρτες του OpenStreetMap (OSM). Επομένως, θα είχε μεγάλη σημασία αν τέτοιας κλίμακας γεωγραφική πληροφορία ήταν προσβάσιμη μέσω κατάλληλων υπηρεσιών καταλόγου που θα παρείχαν *μεταδεδομένα* (metadata) σχετικά με την κλίμακα, την τελευταία ενημέρωση, την χωρική κάλυψη κ.ά., καθώς επίσης και τα ίδια τα δεδομένα σε μορφή συμβατή με τις προδιαγραφές INSPIRE.

Η διπλωματική αποσκοπεί στον σχεδιασμό και την υλοποίηση τέτοιων υπηρεσιών που θ' αξιοποιούν επίσης τεχνολογίες του Σημασιολογικού Ιστού, ώστε τα γεωγραφικά δεδομένα και μεταδεδομένα να είναι προσβάσιμα ως RDF μέσω SPARQL endpoints. Κατά την εκπόνηση της εργασίας θα επιχειρηθούν τα εξής:

- Διαμόρφωση των μεταδεδομένων κατά INSPIRE με κατάλληλο RDF λεξιλόγιο (λ.χ., VoID). Κατ' αυτόν τον τρόπο θα υπάρχουν έτοιμα μεταδεδομένα (ή διαθέσιμα δυναμικά μέσω service) για γεωγραφικά δεδομένα σε μορφή RDF σύμφωνα με τις προδιαγραφές INSPIRE.
- Υλοποίηση διαδικτυακών υπηρεσιών καταλόγου σύμφωνα με τις απαιτήσεις του INSPIRE (INSPIRE Discovery Services). Οι υπηρεσίες αυτές (λ.χ. discovery service metadata, discovery service for queries) θα πρέπει να είναι διαθέσιμες ως middleware μέσω SPARQL, ακολουθώντας το πρότυπο GeoSPARQL για γεωγραφικά RDF δεδομένα.
- Δοκιμαστική μετατροπή δεδομένων κατά INSPIRE σε διασυνδεδεμένα γεωγραφικά δεδομένα σύμφωνα με το πρότυπο GeoSPARQL. Το σχήμα δεδομένων και οι αντιστοιχήσεις (mappings) που θα χρησιμοποιηθούν για τον σκοπό αυτό, θα μπορούσαν να γενικευθούν για τέτοιου είδους μετατροπές.
- Αυτοματοποιημένη μετατροπή δεδομένων (λ.χ. οδικό δίκτυο OSM) από/προς μορφή INSPIRE.

ΣΧΕΤΙΚΟ ΥΛΙΚΟ:

- INSPIRE: http://inspire.jrc.ec.europa.eu/ και INSPIRE portal: http://inspire-geoportal.ec.europa.eu/
- OpenStreetMap (OSM): http://www.openstreetmap.org/
- GeoSPARQL standard: https://portal.opengeospatial.org/files/?artifact_id=47664
- VoID Vocabulary: http://www.w3.org/TR/void/

ΜΕΛΕΤΗ ΚΑΙ ΕΠΕΚΤΑΣΗ ΑΛΓΟΡΙΘΜΩΝ ΣΥΓΧΩΝΕΥΣΗΣ ΟΝΤΟΤΗΤΩΝ ΣΕ ΣΗΜΑΣΙΟΛΟΓΙΚΑ ΔΕΔΟΜΕΝΑ ΜΕ ΓΕΩΧΩΡΙΚΗ ΠΛΗΡΟΦΟΡΙΑ

ΠΛΗΡΟΦΟΡΙΕΣ: Γιώργος Γιαννόπουλος, giann [at] dblab.ece.ntua.gr
Δημήτρης Σκούτας, dskoutas [at] imis.athena-innovation.gr

ΠΕΡΙΛΗΨΗ: Στόχος της διπλωματικής είναι η μελέτη τεχνικών, αλγορίθμων και εργαλείων για συγχώνευση οντοτήτων σε σημασιολογικά δεδομένα και η επέκτασή τους ή ανάπτυξη νέων τεχνικών για συγχώνευση βασιζόμενη τόσο στη σημασιολογική, όσο και στη γεωχωρική πληροφορία των δεδομένων.

ΑΤΟΜΑ: 1

ΠΛΑΤΦΟΡΜΑ ΕΡΓΑΣΙΑΣ: Java

ΣΥΝΤΟΜΗ ΠΕΡΙΓΡΑΦΗ: Η ευρεία εξάπλωση και χρήση του διαδικτύου, η διάθεση και διακίνηση μεγάλου όγκου πληροφορίας μέσω αυτού, σε συνδυασμό με την ανάπτυξη πληροφοριακών συστημάτων, τεχνολογιών και προτύπων βασισμένων σε διαφορετικές ανάγκες και ιδιαιτερότητες, έχουν σαν αποτέλεσμα την εμφάνιση *ετερογένειας* (*heterogeneity*) και τον περιορισμό των δυνατοτήτων του σημερινού *παγκόσμιου ιστού* (*WWW*). Τα παραπάνω καλείται να αντιμετωπίσει ο *Σημασιολογικός Ιστός* (*Semantic Web*) [1], ο οποίος αποτελεί τη μεγαλύτερη προσπάθεια αυτόματης ενοποίησης συστημάτων, με σκοπό να συνεργάζονται διαλειτουργικά σε παγκόσμιο επίπεδο. Στον *Σημασιολογικό Ιστό*, τα δεδομένα ακολουθούν το *RDF* (*Resource Description Framework*) [2],[3] μοντέλο, με κυρίαρχη γλώσσα ερωτήσεων, την *SPARQL* (*Simple Protocol and RDF Query Language*) [4].
Από την άλλη πλευρά, η διαχείριση και εκμετάλλευση μεγάλου όγκου γεωχωρικών δεδομένων είναι υψηλής σημασίας τόσο στη βιομηχανία (π.χ. τουριστικά πρακτορεία που αναλύουν τις τάσεις των πελατών τους) όσο και στην κοινωνική ζωή (π.χ. εκμετάλλευση γεωγραφικής πληροφορίας στα κοινωνικά δίκτυα για διασύνδεση χρηστών, διαφημίσεις, κ.α.).
Παρόλο που στις δύο παραπάνω περιοχές (ειδικά στα γεωχωρικά δεδομένα) έχουν γίνει σημαντικά βήματα προόδου, δεν ισχύει το ίδιο για τη διαχείριση δεδομένων που συνδυάζουν ιδιότητες και από τα δύο πεδία, δηλαδή για σημασιολογικά, γεωχωρικά δεδομένα.
Η συγκεκριμένη διπλωματική θα επικεντρωθεί στο πρόβλημα της συγχώνευσης δεδομένων (data fusion), το οποίο συνίσταται στα εξής στάδια: (α) αναζήτηση σημασιολογικών οντοτήτων οι οποίες ενδέχεται να έχουν κωδικοποιηθεί με διαφορετικό τρόπο, αλλά αντιστοιχούν σε κοινή οντότητα και (β) δημιουργία συσχετίσεων μεταξύ των οντοτήτων και μετασχηματισμός των μεταδεδομένων τους, έτσι ώστε να είναι δυνατή η αυτόματη αντιστοίχιση τους και η εύρεση του συνόλου των μεταδεδομένων τους σε περίπτωση αναζήτησης των οντοτήτων.
Στα πλαίσια της διπλωματικής θα μελετηθούν διάφορες τεχνικές σημασιολογικής αντιστοίχισης και συγχώνευσης δεδομένων [5] καθώς και αντίστοιχα εργαλεία [6], [7] και, στη συνέχεια, θα αναπτυχθούν αλγόριθμοι οι οποίοι θα εκμεταλλεύονται, πέρα από τα σημασιολογικά μεταδεδομένα, και τη γεωχωρική πληροφορία των οντοτήτων για αποδοτική συγχώνευση των δεδομένων. Επιπλέον, θα σχεδιαστούν benchmarks, πάνω σε κατάλληλα σύνολα σημασιολογικών γεωχωρικών δεδομένων [8], για την αξιολόγηση και σύγκριση των διαφόρων εξεταζόμενων ή προς υλοποίηση μεθόδων.

Η διπλωματική θα πραγματοποιηθεί στα πλαίσια του ερευνητικού έργου GeoKnow το οποίο είναι ένα τριετές, χρηματοδοτούμενο από την ΕΕ ερευνητικό έργο, που αφορά στο γεωχωρικό Σημασιολογικό Ιστό. Συνοπτικά, το GeoKnow καταπιάνεται με τη διασύνδεση, διαχείριση, ποιότητα, συνάθροιση, οπτικοποίηση και δημιουργία

γεωχωρικών διαδικτυακών δεδομένων. Τα ερευνητικά μας αποτελέσματα θα εφαρμοστούν στις περιοχές των εφοδιαστικών αλυσίδων και των ταξιδιωτικών εταιρειών.

ΣΧΕΤΙΚΕΣ ΠΛΗΡΟΦΟΡΙΕΣ:

[1] Tim Berners-Lee et al. *The Semantic Web*, Scientific American, May 17, 2001, Available at: *www.dblab.ece.ntua.gr/~bikakis/SW.pdf*
[2] Resource Description Framework (RDF), *http://www.w3schools.com/rdf/rdf_intro.asp*
[3] RDF Primer, *http://www.w3.org/TR/rdf-syntax/*
[4] Simple Protocol and RDF Query Language (SPARQL), *http://www.slideshare.net/olafhartig/an-introduction-to-sparql*
[5] Bernstein et al. Generic Schema Matching, Ten Years Later, VLDB'11 http://www.sigmod.org/publications/sigmod-record/0906/publications/1003/p41.survey.drosou.pdf
[6] Interlinking tools, http://stack.lod2.eu/
[7] Fusion tools, http://sieve.wbsg.de/
[8] Datasets, http://linkedgeodata.org

ΕΡΓΑΛΕΙΑ ΓΙΑ ΤΗ ΣΥΓΧΩΝΕΥΣΗ ΔΕΔΟΜΕΝΩΝ ΜΕ ΣΗΜΑΣΙΟΛΟΓΙΚΑ ΚΑΙ ΓΕΩΧΩΡΙΚΑ ΚΡΙΤΗΡΙΑ

ΠΛΗΡΟΦΟΡΙΕΣ: Γιώργος Γιαννόπουλος, giann [at] dblab.ece.ntua.gr
Δημήτρης Σκούτας, dskoutas [at] imis.athena-innovation.gr

ΠΕΡΙΛΗΨΗ: Στόχος της διπλωματικής είναι η σχεδίαση και υλοποίηση ενός εργαλείου που θα διευκολύνει την αντιστοίχιση και συγχώνευση γεωχωρικών οντοτήτων με τεχνικές που συνδυάζουν σημασιολογική και γεωχωρική πληροφορία.

ΑΤΟΜΑ: 1

ΠΛΑΤΦΟΡΜΑ ΕΡΓΑΣΙΑΣ: Java, PostGIS, OpenStreetMap

ΣΥΝΤΟΜΗ ΠΕΡΙΓΡΑΦΗ: Το ανοιχτό μοντέλο λειτουργίας του Παγκόσμιου Ιστού συνεπάγεται ότι συνήθως μπορεί κανείς να αντλήσει πληροφορία για μία οντότητα από μια πληθώρα διαφορετικών, ετερογενών και ανεξάρτητων μεταξύ τους πηγών. Οι πηγές αυτές συχνά χρησιμοποιούν διαφορετικά αναγνωριστικά για να αναφερθούν στην ίδια οντότητα, διαφορετικούς τρόπους αναπαράστασης, και παρέχουν διαφορετικές πληροφορίες που ενδέχεται να είναι συμπληρωματικές, επικαλυπτόμενες ή και αντιφατικές. Για το λόγο αυτό, ένα μεγάλο μέρος της έρευνας έχει επικεντρωθεί στην ανάπτυξη αλγορίθμων και τεχνικών για την αντιστοίχιση και συγχώνευση τέτοιων δεδομένων [9].
Η συνήθης πρακτική είναι η κατασκευή εργαλείων τα οποία με τη βοήθεια ευριστικών αλγορίθμων προσπαθούν να εντοπίσουν πιθανές αντιστοιχίσεις μεταξύ οντοτήτων, και κατόπιν μέσω μια γραφικής διεπαφής επιτρέπουν στον χρήστη να δει τις προτεινόμενες αντιστοιχίσεις και να τις επιβεβαιώσει, απορρίψει ή συμπληρώσει. Συνεπώς είναι σημαντικό το γραφικό περιβάλλον να είναι εύχρηστο, κατανοητό και σχεδιασμένο με τρόπο που να υποστηρίζει και να διευκολύνει τη διαδικασία, ώστε να μειώνεται ο απαιτούμενος χρόνος και προσπάθεια, να αποφεύγονται λάθη, και να βελτιώνεται η ποιότητα και πληρότητα των αποτελεσμάτων.
Ωστόσο τα εργαλεία αυτά συνήθως δεν λαμβάνουν κάποια ειδική μέριμνα για τη διαχείριση γεωχωρικών δεδομένων, όπως π.χ. τη δυνατότητα απεικόνισής τους σε χάρτη ή την αξιοποίηση της χωρικής διάστασης στην ανεύρεση αντιστοιχίσεων. Το κενό αυτό είναι σημαντικό, καθώς τα γεωχωρικά δεδομένα αποτελούν μια πολύ μεγάλη και σημαντική κατηγορία δεδομένων, με πληθώρα πρακτικών εφαρμογών, και προέρχονται συχνά από ετερογενείς πηγές οπότε είναι υπαρκτή η ανάγκη για αντιστοίχιση και συγχώνευσή τους.
Σκοπός λοιπόν της διπλωματικής είναι να μελετηθούν διάφορα υπάρχοντα εργαλεία για διασύνδεση δεδομένων (π.χ. [10][11]), προκειμένου να σχεδιαστεί και να υλοποιηθεί ένα εργαλείο (είτε ως επέκταση υπάρχοντος είτε ανεξάρτητο) που θα είναι προσανατολισμένο και προσαρμοσμένο σε γεωχωρικά δεδομένα. Αυτό θα αφορά τόσο τις τεχνικές που θα ενσωματώνει όσο και τη δομή, σχεδίαση και λειτουργικότητα που θα προσφέρει το γραφικό του περιβάλλον.

Η διπλωματική θα πραγματοποιηθεί στα πλαίσια του ερευνητικού έργου GeoKnow το οποίο είναι ένα τριετές, χρηματοδοτούμενο από την ΕΕ ερευνητικό έργο, που αφορά στο γεωχωρικό Σημασιολογικό Ιστό. Συνοπτικά, το GeoKnow καταπιάνεται με τη διασύνδεση, διαχείριση, ποιότητα, συνάθροιση, οπτικοποίηση και δημιουργία γεωχωρικών διαδικτυακών δεδομένων. Τα ερευνητικά μας αποτελέσματα θα εφαρμοστούν στις περιοχές των εφοδιαστικών αλυσίδων και των ταξιδιωτικών εταιρειών.

ΣΧΕΤΙΚΕΣ ΠΛΗΡΟΦΟΡΙΕΣ:

[9] Bernstein et al. Generic Schema Matching, Ten Years Later, VLDB'11
http://www.sigmod.org/publications/sigmod-record/0906/publications/1003/p41.survey.drosou.pdf
[10] Interlinking tools: http://stack.lod2.eu/
[11] Fusion tools: http://sieve.wbsg.de/
[12] Datasets, http://linkedgeodata.org

ΣΥΣΤΗΜΑ ΔΙΑΧΕΙΡΙΣΗΣ ΔΙΑΧΡΟΝΙΚΩΝ ΔΕΔΟΜΕΝΩΝ ΓΙΑ ΓΟΝΙΔΙΑ ΚΑΙ ΜΟΡΙΑ MICRORNA

ΠΛΗΡΟΦΟΡΙΕΣ: Θανάσης Βεργούλης, 210 6875423, vergoulis@imis.athena-innovation.gr και Ηλίας Κανέλλος, kanellos@dblab.ece.ntua.gr

ΠΕΡΙΛΗΨΗ: Η μελέτη των *βιομορίων* που συμμετέχουν στους μηχανισμούς της ζωής (πχ DNA, πρωτεΐνες, μόρια microRNA κτλ) είναι απαραίτητη για να μπορέσουν οι ερευνητές να κατανοήσουν και να θεραπεύσουν γενετικές ασθένειες που απασχολούν την Ιατρική και τη Βιολογία τους τελευταίους αιώνες. Οι πληροφορίες που σχετίζονται με αυτά τα βιομόρια αποκαλύπτονται μέσω βιολογικών πειραμάτων και καταγράφονται σε βάσεις δεδομένων για να είναι προσβάσιμες στους ερευνητές. Όμως τα βιολογικά πειράματα χρησιμοποιούν μηχανήματα και αναλύσεις που είναι επιρρεπή σε σφάλματα. Ως αποτέλεσμα, οι προαναφερθείσες βάσεις δεδομένων οφείλουν να μεταβάλλονται συνεχώς προκειμένου να είναι ενημερωμένες με τις ακριβέστερες μετρήσεις. Για διάφορους λόγους (πχ για την συγκριτική μελέτη σχετικής βιβλιογραφίας) είναι χρήσιμη η πρόσβαση όχι μόνο στις πιο πρόσφατες εκδόσεις των πληροφοριών αλλά επίσης και στις προηγούμενες καταστάσεις τους. Οι υπάρχουσες βάσεις δεδομένων δεν διευκολύνουν την αναζήτηση αυτών των προηγούμενων καταστάσεων δημιουργώντας πρόβλημα στους ερευνητές.
Με μια προηγούμενη εργασία μας καταγράψαμε και οπτικοποιήσαμε τις σχετικές με τα μόρια microRNA πληροφορίες διαχρονικά, διευθετώντας έτσι μέρος του προβλήματος. Στόχος της παρούσας εργασίας είναι (α) η βελτίωση και επέκταση του υπάρχοντος λογισμικού οπτικοποίησης διαχρονικών δεδομένων για μόρια microRNA, (β) η μελέτη, η μοντελοποίηση, η καταγραφή και η οπτικοποίηση των πληροφοριών που σχετίζονται με τις οικογένειες των microRNA μορίων και με τα γονίδια του DNA διαχρονικά και (γ) ο εμπλουτισμός όλων των διαχρονικών δεδομένων για βιομόρια με πρόσθετες πληροφορίες (πχ στόχοι microRNA, συμμετοχή σε βιολογικές διαδικασίες κτλ).

ΑΤΟΜΑ: 1-2

ΠΛΑΤΦΟΡΜΑ ΕΡΓΑΣΙΑΣ: PHP, JavaScript (JQuery), Python, MySQL.

ΣΥΝΤΟΜΗ ΠΕΡΙΓΡΑΦΗ: Το σύνολο της γενετικής πληροφορίας ενός οργανισμού κωδικοποιείται σε ακολουθίες DNA, που ονομάζονται *γονίδια*. Το κύτταρο «διαβάζει» τη γενετική πληροφορία που κωδικοποιούν τα γονίδια και, με βάση αυτή, παράγει *πρωτεΐνες*, θέτοντας έτσι σε εφαρμογή τους μηχανισμούς της ζωής. Δυσλειτουργίες κατά την παραγωγή πρωτεϊνών μπορούν να δημιουργήσουν προβλήματα στους μηχανισμούς αυτούς. Τέτοιες δυσλειτουργίες αποτελούν την αιτία πολλών γενετικών ασθενειών. Τα μόρια microRNA, τα οποία λειτουργούν ως ρυθμιστές της παραγωγής πρωτεϊνών, υπόσχονται την παροχή ενός τρόπου αντιμετώπισης τέτοιων ασθενειών.
Λόγω όσων αναφέρθηκαν προηγουμένως, οι πληροφορίες που σχετίζονται με γονίδια και μόρια microRNA είναι ιδιαίτερα χρήσιμες στους ερευνητές. Έτσι τις τελευταίες δεκαετίες έχουν εμφανιστεί αρκετές διαδικτυακές βάσεις δεδομένων που συγκεντρώνουν τέτοιες πληροφορίες (πχ Enseml, miRBase κτλ). Τα δεδομένα που καταγράφονται σε αυτές προκύπτουν από βιολογικά πειράματα, που χρησιμοποιούν μηχανήματα και αναλύσεις επιρρεπείς σε σφάλματα. Ως αποτέλεσμα, το περιεχόμενο των βάσεων αυτών ενημερώνεται συνεχώς με βάση τις νεότερες και ακριβέστερες μετρήσεις. Για διάφορους λόγους (πχ για την συγκριτική μελέτη σχετικής βιβλιογραφίας) είναι χρήσιμη η πρόσβαση όχι μόνο στις πιο πρόσφατες εκδόσεις των πληροφοριών αλλά επίσης και στις προηγούμενες καταστάσεις τους. Όμως οι υπάρχουσες

υποδομές δεν διευκολύνουν την αναζήτηση αυτών των προηγούμενων καταστάσεων, δημιουργώντας έτσι πρόβλημα στους ερευνητές.

Στο Ινστιτούτο Πληροφοριακών ΣΥστημάτων του ΕΚ «Αθηνά» και σε συνεργασία με το ΕΚ βιοιατρικής «Αλέξανδρος Φλέμινγκ» μελετήσαμε τις βασικές πληροφορίες που σχετίζονται με τα μόρια microRNA διαχρονικά, όπως καταγράφονται από τις διάφορες εκδόσεις της βάσης δεδομένων miRBase. Επιπλέον, καταγράψαμε αυτές τις πληροφορίες σε ένα βολικό σχήμα βάσης δεδομένων και υλοποιήσαμε ένα λογισμικό οπτικοποίησής τους. Επίσης, χρησιμοποιήσαμε τη διαχρονική πληροφορία προκειμένου να βοηθήσουμε την αναζήτηση βιβλιογραφίας για τα μόρια microRNA. Με τον τρόπο αυτό διευθετήσαμε ένα πρώτο κομμάτι του προβλήματος της διαχείρισης των διαχρονικών δεδομένων για βιομόρια.

Η παρούσα εργασία περιλαμβάνει:

- Βελτίωση της διεπαφής και επέκταση της λειτουργικότητας του υπάρχοντος λογισμικού οπτικοποίησης διαχρονικών δεδομένων για μόρια microRNA
- Μελέτη, μοντελοποίηση, καταγραφή και οπτικοποίηση των διαχρονικών δεδομένων για οικογένειες microRNA
- Μελέτη, μοντελοποίηση, καταγραφή και οπτικοποίηση των διαχρονικών δεδομένων για γονίδια διαφόρων οργανισμών
- Εμπλουτισμός όλων των διαχρονικών δεδομένων για βιομόρια με επιπλέον πληροφορίες από διαδικτυακές βάσεις δεδομένων.

ΣΧΕΤΙΚΟ ΥΛΙΚΟ:

- Ensembl γονιδιακή ΒΔ: http://www.ensembl.org/index.html
- miRBase ΒΔ για microRNA: http://www.mirbase.org
- Ηλίας Κανέλλος «Αναζήτηση σε επιστημονικές βάσεις δεδομένων με βάση την ιστορική εξέλιξη των δεδομένων» (2012) διπλωματική εργασία: http://www.dbnet.ece.ntua.gr/pubs/details.php?id=1656&clang=1

ΣΥΣΤΗΜΑ ΑΥΤΟΜΑΤΗΣ ΕΞΑΓΩΓΗΣ ΠΛΗΡΟΦΟΡΙΩΝ ΓΙΑ ΤΑ ΒΙΟΜΟΡΙΑ MICRORNA ΑΠΟ ΕΠΙΣΤΗΜΟΝΙΚΕΣ ΔΗΜΟΣΙΕΥΣΕΙΣ ΣΤΙΣ ΒΙΟΕΠΙΣΤΗΜΕΣ

ΠΛΗΡΟΦΟΡΙΕΣ: Θανάσης Βεργούλης, 210 6875423, vergoulis@imis.athena-innovation.gr και Ηλίας Κανέλλος, kanellos@dblab.ece.ntua.gr

ΠΕΡΙΛΗΨΗ: Η μελέτη των *βιομορίων* που συμμετέχουν στους μηχανισμούς της ζωής (πχ DNA, πρωτεΐνες, μόρια microRNA κτλ) είναι απαραίτητη για να μπορέσουν οι ερευνητές να κατανοήσουν και να θεραπεύσουν γενετικές ασθένειες που απασχολούν την Ιατρική και τη Βιολογία τους τελευταίους αιώνες. Οι πληροφορίες που σχετίζονται με αυτά τα βιομόρια αποκαλύπτονται μέσω βιολογικών πειραμάτων ή υπολογιστικών προβλέψεων και τα αποτελέσματα δημοσιεύονται σε επιστημονικά περιοδικά και συνέδρια. Για να καταστεί η πληροφορία που καταγράφεται σε αυτές τις δημοσιεύσεις εύκολα προσβάσιμη στους βιοεπιστήμονες ανατίθεται σε επιμελητές η ανάγνωση της βιβλιογραφίας και η καταγραφή της γνώσης που εντοπίζεται με συστηματικό τρόπο. Επειδή όμως η συγκεκριμένη εργασία έχει αποδειχθεί ιδιαίτερα χρονοβόρα και επίπονη, κρίθηκε απαραίτητη η υλοποίηση εργαλείων που (α) βοηθούν τον επιμελητή να εντοπίσει γρήγορα τις δημοσιεύσεις που τον αφορούν και (β) σκιαγραφούν το είδος της πληροφορίας που καταγράφεται σε καθεμία από τις δημοσιεύσεις.
Σε προηγούμενες εργασίες έχουμε υλοποιήσει έξυπνα εργαλεία για την αυτόματη επισημείωση εγγράφων με όρους που περιγράφουν το περιεχόμενό τους (πχ Gontogle) όπως επίσης και διαδικτυακές εφαρμογές που συγκεντρώνουν πληροφορίες που έχουν εξαχθεί από επιστημονικές δημοσιεύσεις βιοεπιστημών (πχ TarBase). Στόχος της παρούσας εργασίας είναι (α) η διερεύνηση των δυνατοτήτων που υπάρχουν για αυτόματη εξαγωγή γνώσης από επιστημονικές δημοσιεύσεις για βιομόρια microRNA (πχ χρήση αλγορίθμου του Gontogle για αυτόματη επισημείωση εργασιών με κλάσεις της οντολογίας MeSH, δυνατότητα εξαγωγής συνδέσεων μεταξύ γονιδίων και μορίων microRNA με βάση το κείμενο κτλ), (β) η υλοποίηση αλγορίθμων που χρησιμοποιούν τα αποτελέσματα της προηγούμενης διερεύνησης προκειμένου να εξάγουν γνώση σχετική με τα βιομόρια microRNA και (γ) η υλοποίηση μιας διαδικτυακής διεπαφής για να μπορούν να εκτελεστούν οι αλγόριθμοι πάνω σε εργασίες που επιλέγονται από τους χρήστες.

ΑΤΟΜΑ: 1

ΠΛΑΤΦΟΡΜΑ ΕΡΓΑΣΙΑΣ: PHP, JavaScript (JQuery), Python, (Java) Lucene, MySQL.

ΣΥΝΤΟΜΗ ΠΕΡΙΓΡΑΦΗ: Το σύνολο της γενετικής πληροφορίας ενός οργανισμού κωδικοποιείται σε ακολουθίες DNA, που ονομάζονται *γονίδια*. Το κύτταρο «διαβάζει» τη γενετική πληροφορία που κωδικοποιούν τα γονίδια και, με βάση αυτή, παράγει *πρωτεΐνες*, θέτοντας έτσι σε εφαρμογή τους μηχανισμούς της ζωής. Δυσλειτουργίες κατά την παραγωγή πρωτεϊνών μπορούν να δημιουργήσουν προβλήματα στους μηχανισμούς αυτούς. Τέτοιες δυσλειτουργίες αποτελούν την αιτία πολλών γενετικών ασθενειών. Τα μόρια microRNA, τα οποία λειτουργούν ως ρυθμιστές της παραγωγής πρωτεϊνών, υπόσχονται την παροχή ενός τρόπου αντιμετώπισης τέτοιων ασθενειών.
Λόγω όσων αναφέρθηκαν προηγουμένως, οι πληροφορίες που σχετίζονται με γονίδια και μόρια microRNA είναι ιδιαίτερα χρήσιμες στους ερευνητές. Τέτοιες πληροφορίες προκύπτουν από την εκτέλεση βιολογικών πειραμάτων ή από προγνώσεις υπολογιστικών προσομοιώσεων και καταγράφονται σε επιστημονικές εργασίες που δημοσιεύονται σε περιοδικά. Θεωρητικά η δημοσίευση των εργασιών παρέχει πρόσβαση στην επιστημονική γνώση για τους ερευνητές όλου του κόσμου, όμως, στην πράξη, το πλήθος των δημοσιεύσεων είναι τόσο μεγάλο που είναι δύσκολο ένας ερευνητής να είναι ενήμερος για όλη τη σχετική βιβλιογραφία.

Για να βελτιωθεί η κατάσταση, ανατίθεται σε επιμελητές η ανάγνωση της βιβλιογραφίας και η καταγραφή της γνώσης που εντοπίζεται με τρόπο συστηματικό. Όμως και πάλι η εργασία των επιμελητών είναι χρονοβόρα και κοπιαστική. Για να διευκολυνθεί η εργασία αυτή είναι απαραίτητη η χρήση εργαλείων που βοηθούν τους επιμελητές να εντοπίσουν γρήγορα τις δημοσιεύσεις που τους αφορούν και να εξάγουν με εύκολο τρόπο την πληροφορία που περιέχεται σε αυτές τις δημοσιεύσεις.

Στο Ινστιτούτο Πληροφοριακών ΣΥστημάτων του ΕΚ «Αθηνά» και σε συνεργασία με το ΕΚ βιοϊατρικής «Αλέξανδρος Φλέμινγκ» έχουμε υλοποιήσει διαδικτυακές εφαρμογές που συγκεντρώνουν πληροφορίες που έχουν εξαχθεί από επιστημονικές δημοσιεύσεις βιοεπιστημών (πχ TarBase). Οι πληροφορίες που παρουσιάζουν οι συγκεκριμένες εφαρμογές έχουν καταγραφεί από βιοεπιστήμονες-επιμελητές μετά από εκτεταμένη μελέτη της βιβλιογραφίας. Ταυτόχρονα στο ΕΚ «Αθηνά» έχουμε αναπτύξει εργαλεία για την αυτόματη επισημείωση εγγράφων με όρους που περιγράφουν το περιεχόμενό τους (πχ Gontogle). Τα συγκεκριμένα εργαλεία, θα μπορούσαν να παραμετροποιηθούν έτσι ώστε να χρησιμοποιούνται από τους επιμελητές των εφαρμογών μας για να διευκολύνονται κατά την προσπάθεια εξαγωγής γνώσης από επιστημονικές εργασίες.
Η παρούσα εργασία περιλαμβάνει:

- Διερεύνηση των δυνατοτήτων που υπάρχουν για αυτόματη εξαγωγή γνώσης από επιστημονικές δημοσιεύσεις για βιομόρια microRNA (πχ χρήση αλγορίθμου του Gontogle για αυτόματη επισημείωση εργασιών με κλάσεις της οντολογίας MeSH, δυνατότητα εξαγωγής συνδέσεων μεταξύ γονιδίων και μορίων microRNA με βάση το κείμενο κτλ),
- Υλοποίηση αλγορίθμων που χρησιμοποιούν τα αποτελέσματα της προηγούμενης διερεύνησης προκειμένου να εξάγουν γνώση σχετική με τα βιομόρια microRNA
- Υλοποίηση μιας διαδικτυακής διεπαφής, η οποία παρέχει στον ερευνητή ως υπηρεσία τη δυνατότητα αυτόματης εξαγωγής γνώσης, ή αυτόματων επισημειώσεων σε δημοσιεύσεις της επιλογής του.

ΣΧΕΤΙΚΟ ΥΛΙΚΟ:

- Gontogle: http://www.dblab.ntua.gr/~bikakis/papers/GoNTogle@ODBASE10.pdf
- miRBase, ΒΔ που συγκεντρώνει πληροφορίες για τα microRNAs: http://www.mirbase.org
- Pubmed, ΒΔ που συγκεντρώνει τις δημοσιεύσεις στο χώρο των βιοεπιστημών: http://www.ncbi.nlm.nih.gov/pubmed

ΥΛΟΠΟΙΗΣΗ ΜΗΧΑΝΙΣΜΟΥ ΤΑΞΙΝΟΜΗΣΗΣ (RANKING) ΠΑΝΩ ΣΕ ΔΗΜΟΣΙΕΥΣΕΙΣ ΣΧΕΤΙΚΕΣ ΜΕ ΒΙΟΜΟΡΙΑ MICRORNA

ΠΛΗΡΟΦΟΡΙΕΣ: Θανάσης Βεργούλης, 210 6875423, vergoulis@imis.athena-innovation.gr και Ηλίας Κανέλλος, kanellos@dblab.ece.ntua.gr

ΠΕΡΙΛΗΨΗ: Η μελέτη των *βιομορίων* που συμμετέχουν στους μηχανισμούς της ζωής (πχ DNA, πρωτεΐνες, μόρια microRNA κτλ) είναι απαραίτητη για να μπορέσουν οι ερευνητές να κατανοήσουν και να θεραπεύσουν γενετικές ασθένειες που απασχολούν την Ιατρική και τη Βιολογία τους τελευταίους αιώνες. Οι πληροφορίες που σχετίζονται με αυτά τα βιομόρια αποκαλύπτονται μέσω βιολογικών πειραμάτων ή υπολογιστικών προβλέψεων και τα αποτελέσματα δημοσιεύονται σε επιστημονικά περιοδικά και συνέδρια. Αυτές οι δημοσιεύσεις, μαζί με μεταδεδομένα που σχετίζονται με το περιεχόμενό τους, καταγράφονται σε βάσεις δεδομένων που προσφέρουν δυνατότητες αναζήτησης μέσω διαδικτύου και διευκολύνουν τους βιοεπιστήμονες στην αναζήτηση βιβλιογραφίας σχετικά με την έρευνά τους.
Σε προηγούμενες εργασίες έχουμε κατασκευάσει το xPub μια βάση δεδομένων που καταγράφει τις επιστημονικές δημοσιεύσεις που είναι σχετικές με τα βιομόρια microRNA και μια διαδικτυακή εφαρμογή που παρέχει προηγμένες υπηρεσίες αναζήτησης πάνω στα δεδομένα που αποθηκεύονται στη συγκεκριμένη βάση. Οι εργασίες που επιστρέφονται ως αποτέλεσμα μιας αναζήτησης στο συγκεκριμένο σύστημα είναι σχετικές με το ερώτημα του χρήστη και παρουσιάζονται με βάση τη χρονολογία δημοσίευσης, από την πιο πρόσφατη προς την πιο παλιά. Στόχος της παρούσας εργασίας είναι: (α) η υλοποίηση προφίλ για τους χρήστες του xPub για καταγραφή των προτιμήσεών τους σχετικά με τις δημοσιεύσεις του ενδιαφέροντός τους, (β) η υλοποίηση ενός μηχανισμού καταγραφής της δραστηριότητας των χρηστών του xPub προκειμένου να εξάγονται συμπεράσματα για τις προτιμήσεις τους με βάση τη συμπεριφορά τους κατά τη χρήση της εφαρμογής και (γ) η υλοποίηση ενός μηχανισμού ταξινόμησης (ranking) για την παρουσίαση των αποτελεσμάτων της εφαρμογής με βάση διάφορα κριτήρια (πχ σημαντικότητα περιοδικού, σημαντικότητα εργασίας, πλήθος εμφανίσεων όρου αναζήτησης στην εργασία, δηλωμένες ή διαπιστωμένες προτιμήσεις του χρήστη κτλ).

ΑΤΟΜΑ: 1-2

ΠΛΑΤΦΟΡΜΑ ΕΡΓΑΣΙΑΣ: PHP, JavaScript (JQuery), Python, (Java) Lucene, MySQL.

ΣΥΝΤΟΜΗ ΠΕΡΙΓΡΑΦΗ: Το σύνολο της γενετικής πληροφορίας ενός οργανισμού κωδικοποιείται σε ακολουθίες DNA, που ονομάζονται *γονίδια*. Το κύτταρο «διαβάζει» τη γενετική πληροφορία που κωδικοποιούν τα γονίδια και, με βάση αυτή, παράγει *πρωτεΐνες*, θέτοντας έτσι σε εφαρμογή τους μηχανισμούς της ζωής. Δυσλειτουργίες κατά την παραγωγή πρωτεϊνών μπορούν να δημιουργήσουν προβλήματα στους μηχανισμούς αυτούς. Τέτοιες δυσλειτουργίες αποτελούν την αιτία πολλών γενετικών ασθενειών. Τα μόρια microRNA, τα οποία λειτουργούν ως ρυθμιστές της παραγωγής πρωτεϊνών, υπόσχονται την παροχή ενός τρόπου αντιμετώπισης τέτοιων ασθενειών.
Στο Ινστιτούτο Πληροφοριακών ΣΥστημάτων του ΕΚ «Αθηνά» και σε συνεργασία με το ΕΚ βιοϊατρικής «Αλέξανδρος Φλέμινγκ» έχουμε αναπτύξει το xPub, μια βάση δεδομένων για την καταγραφή επιστημονικών δημοσιεύσεων σχετικών με τα βιομόρια microRNA και μια διαδικτυακή εφαρμογή που προσφέρει προηγμένες υπηρεσίες αναζήτησης πάνω στα δεδομένα που αποθηκεύονται στη συγκεκριμένη βάση δεδομένων. Η συγκεκριμένη εφαρμογή λαμβάνει υπόψη τόσο τα διαθέσιμα μέρη των κειμένων των εργασιών όσο και μεταδεδομένα που τα συνοδεύουν (πχ πληροφορίες για το περιεχόμενό τους που έχουν

καταγραφεί είτε από τους συγγραφείς των εργασιών είτε από επιμελητές που τις έχουν μελετήσει) για να προσφέρει στους χρήστες της τον πληρέστερο κατάλογο δημοσιεύσεων που είναι σχετικές με τις αναζητήσεις τους. Στην παρούσα φάση, τα αποτελέσματα των αναζητήσεων επιστρέφονται με βάση την ημερομηνία δημοσίευσης ξεκινώντας από τις πιο πρόσφατες εργασίες και καταλήγοντας στις πιο παλιές. Όμως ο συγκεκριμένος τρόπος παρουσίασης δεν είναι ιδανικός καθώς μπορεί σημαντικές δημοσιεύσεις να κρύβονται αρκετά χαμηλά στη λίστα των αποτελεσμάτων ενώ εργασίες ήσσονος σημασίας να εμφανίζονται ως πρώτα αποτελέσματα. Είναι απαραίτητο λοιπόν να χρησιμοποιηθούν κάποια ποιοτικά κριτήρια που θα δίνουν μεγαλύτερη προτεραιότητα εμφάνισης στις πιο σημαντικές εργασίες. Επίσης, δεδομένου ότι για τον κάθε χρήστη η ιδανική σειρά αποτελεσμάτων μπορεί να διαφέρει με βάση τα ερευνητικά του ενδιαφέροντα, είναι σημαντική η διατήρηση ενός προφίλ για τους χρήστες της εφαρμογής, έτσι ώστε να λαμβάνονται υπόψη οι προτιμήσεις τους κατά την εμφάνιση των αποτελεσμάτων.
Η παρούσα εργασία περιλαμβάνει:

- Υλοποίηση προφίλ χρηστών για το xPub ώστε να καταγράφονται οι προτιμήσεις τους σχετικά με τα είδη δημοσιεύσεων που προτιμούν και τα περιοδικά που θεωρούν ως πιο αξιόπιστα.
- Υλοποίηση μηχανισμού καταγραφής της δραστηριότητας των χρηστών για το xPub (πχ καταγραφή των αναζητήσεων, των αποτελεσμάτων τα οποία επιλέγουν για να δουν περισσότερες πληροφορίες κτλ) προκειμένου να εξάγονται συμπεράσματα για τις προτιμήσεις τους.
- Υλοποίηση ενός μηχανισμού ταξινόμησης (ranking) για την παρουσίαση των αποτελεσμάτων της εφαρμογής με βάση διάφορα κριτήρια (πχ σημαντικότητα περιοδικού, σημαντικότητα εργασίας, πλήθος εμφανίσεων όρου αναζήτησης στην εργασία, δηλωμένες ή διαπιστωμένες προτιμήσεις του χρήστη κτλ) έτσι ώστε οι πιο σημαντικές εργασίες να επιστρέφονται στις πρώτες θέσεις της λίστας αποτελεσμάτων.

ΣΧΕΤΙΚΟ ΥΛΙΚΟ:

- miRBase, ΒΔ για microRNA: http://www.mirbase.org
- DIANA tools, σουίτα εργαλείων για αναζητήσεις σχετικές με microRNAs: http://diana.imis.athena-innovation.gr/DianaTools/index.php (υλοποιημένη από το ΕΚ Αθηνά)
- PubMed, ΒΔ για δημοσιεύσεις στις βιοεπιστήμες: http://www.ncbi.nlm.nih.gov/pubmed

ΑΝΑΚΤΗΣΗ ΚΑΙ ΑΝΑΛΥΣΗ ΧΡΗΣΤΩΝ ΤΟΥ TWITTER

ΠΛΗΡΟΦΟΡΙΕΣ: Γιάννης Σταύρακας, τηλ. 2106875413, yannis@imis.athena-innovation.gr
Βασίλης Πλαχούρας, τηλ. 2106875413, vassilis.plachouras@gmail.com

ΠΕΡΙΛΗΨΗ: Το αντικείμενο της διπλωματικής είναι η ανάπτυξη εφαρμογής που θα ανακτά τους χρήστες που αναρτούν tweets με κάποιο συγκεκριμένο θέμα, και θα αναλύει τις σχέσεις που έχουν μεταξύ τους καθώς και τη σχετικότητα τους με το συγκεκριμένο θέμα.

ΑΤΟΜΑ: 1-2.

ΠΛΑΤΦΟΡΜΑ ΕΡΓΑΣΙΑΣ: Java, Twitter API.

ΣΥΝΤΟΜΗ ΠΕΡΙΓΡΑΦΗ: Οι χρήστες του Twitter συνδέονται μεταξύ τους: κάθε χρήστης μπορεί να «ακολουθεί» άλλους χρήστες, των οποίων επιλέγει να διαβάζει τα μηνύματα. Αυτές οι συνδέσεις σχηματίζουν έναν κατευθυνόμενο γράφο, όπου οι χρήστες είναι οι κόμβοι. Ο στόχος της διπλωματικής είναι να βρει τους χρήστες και τις ομάδες χρηστών που είναι πιο «δραστήριοι» σε σχέση με κάποιο συγκεκριμένο θέμα. Για παράδειγμα, ας θεωρήσουμε ότι ψάχνω το Twitter για tweets σχετικά με την κρίση του ευρώ. Θα ήθελα να δω ποιοι χρήστες έχουν αναρτήσει τα περισσότερα τέτοια tweets και αν σχετίζονται μεταξύ τους σχηματίζοντας «παρέες» που συζητούν συχνά για το θέμα αυτό.

Ένας απλός τρόπος να γίνει αυτό είναι ο παρακάτω:

- Θεωρώντας ότι έχω έναν αριθμό tweets που αναφέρονται στο θέμα που με ενδιαφέρει (αυτή η δυνατότητα υπάρχει υλοποιημένη μέσω ενός campaign manager και θα δοθεί στην διπλωματική έτοιμη), ρωτάω το API του Twitter για τις σχέσεις των χρηστών που έχουν αναρτήσει τα μηνύματα αυτά.
- Με βάση αυτές τις σχέσεις φτιάχνω τον γράφο «συσχέτισης» των χρηστών. Εφαρμόζοντας διαφορές τεχνικές, προσπαθώ να ανακαλύψω (α) τους πιο δραστήριους χρήστες που επηρεάζουν τους περισσότερους άλλους χρήστες, και (β) τις πιο δραστήριες «παρέες» συνδεδεμένων χρηστών.

Η διπλωματική θα αναπτύξει web-based εφαρμογή που θα υλοποιεί τα παραπάνω. Η εφαρμογή θα επεκτείνει έναν campaign manager (που έχει ήδη υλοποιηθεί) και που ανακτά από το twitter τα tweets που ικανοποιούν κάποια θεματικά κριτήρια (πχ περιέχουν κάποιες λέξεις-κλειδιά). Η εφαρμογή θα:

- Σώζει την πληροφορία που ανακτά από το twitter σχετικά με τις σχέσεις μεταξύ των χρηστών.
- Παρουσιάζει τους πιο «δραστήριους» χρήστες και τις «παρέες» που έχει ανακαλύψει.
- Επιτρέπει να θύσει κάποιος κριτήρια βασισμένα στους χρήστες και τις «παρέες», που στη συνεχεία θα οδηγούν (μαζί με τα θεματικά κριτήρια λέξεων-κλειδιών) την ανάκτηση των κατάλληλων tweets στον campaign manager.

Η διπλωματική θα προσφέρει μια καλή προγραμματιστική εμπειρία σε Java, σε web-based εφαρμογές, και στο API του Twitter.

ΣΧΕΤΙΚΟ ΥΛΙΚΟ

- Twitter: https://support.twitter.com/articles/

ΑΞΙΟΛΟΓΗΣΗ ΤΕΧΝΟΛΟΓΙΩΝ NOSQL ΓΙΑ ΔΕΔΟΜΕΝΑ ΚΟΙΝΩΝΙΚΩΝ ΔΙΚΤΥΩΝ

ΠΛΗΡΟΦΟΡΙΕΣ: Γιάννης Σταύρακας, τηλ. 2106875413, yannis@imis.athena-innovation.gr
Βασίλης Πλαχούρας, τηλ. 2106875413, vassilis.plachouras@gmail.com

ΠΕΡΙΛΗΨΗ: Στο ΙΠΣΥ έχουμε αναπτύξει την πλατφόρμα TwitHoard για τη συλλογή μεγάλου όγκου θεματικά-εστιασμένων tweets, τα οποία αποθηκεύονται σε μια σχεσιακή βάση δεδομένων. Στόχος της εργασίας αυτής είναι: (α) η προσαρμογή της πλατφόρμας TwitHoard ώστε να αποθηκεύει τα συλλεγόμενα δεδομένα από το Twitter σε πλατφόρμα NoSQL, και (β) η αξιολόγηση και σύγκριση της αποδοτικότητας σχεσιακών και NoSQL τεχνολογιών για την αποθήκευση, ανάκτηση και επεξεργασία μεγάλου όγκου δεδομένων από κοινωνικά δίκτυα.

ΑΤΟΜΑ: 1-2.

ΠΛΑΤΦΟΡΜΑ ΕΡΓΑΣΙΑΣ: Java, NoSQL data stores.

ΣΥΝΤΟΜΗ ΠΕΡΙΓΡΑΦΗ: Καθώς η δραστηριότητα των χρηστών κοινωνικών δικτύων, όπως το Twitter ή το Facebook, παράγουν τεράστιο όγκο δεδομένων – καθημερινά δημιουργούνται περισσότερα από 400 εκατομμύρια tweets – η χρήση σχεσιακών βάσεων για την αποθήκευση δεδομένων δεν επιτρέπει την κλιμάκωση των συστημάτων. Στόχος της παρούσας εργασίας είναι η προσαρμογή της πλατφόρμας TwitHoard ώστε να αντικατασταθεί η σχεσιακή βάση δεδομένων στην οποία αποθηκεύονται δεδομένα από πλατφόρμες NoSQL, και η συγκριτική αξιολόγηση της απόδοσης τους. Σημαντικό κομμάτι της εργασίας αποτελεί η αξιολόγηση και χρήση διαφορετικών τεχνολογιών NoSQL.

Τα καθήκοντα της εργασίας είναι τα παρακάτω:

- Προσαρμογή των μηχανισμών αποθήκευσης δεδομένων της πλατφόρμας TwitHoard, ώστε να επιτρέπεται η χρήση διαφορετικών τεχνολογιών (σχεσιακών ή NoSQL).
- Ορισμός των μετρήσιμων χαρακτηριστικών της απόδοσης της πλατφόρμας TwitHoard σχετικά με την τεχνολογία αποθήκευσης δεδομένων.
- Σχεδιασμός και ανάλυση ενός συστήματος μέτρησης της επίδοσης που θα περιλαμβάνει τουλάχιστον την αξιολόγηση επιδόσεων για την εγγραφή, ανάκτηση και επεξεργασία δεδομένων. Το σημείο αναφοράς (baseline) θα είναι η υπάρχουσα υλοποίηση που χρησιμοποιεί τη σχεσιακή βάση δεδομένων.
- Δημιουργία ενός συνόλου δεδομένου που θα χρησιμοποιηθεί για την αξιολόγηση και θα βασίζεται σε δεδομένα που παράγει το TwitHoard.Μέτρηση της απόδοσης (Benchmarking) ενός αριθμού διαφορετικών σχεσιακών βάσεων και NoSQL τεχνολογιών.
- Μέτρηση της απόδοσης (Benchmarking) ενός αριθμού διαφορετικών σχεσιακών βάσεων και NoSQL τεχνολογιών.

ΣΧΕΤΙΚΟ ΥΛΙΚΟ

- NoSQL Databases - http://en.wikipedia.org/wiki/NoSQL
- What is NoSQL - http://www.mongodb.com/nosql
- Twitter API Documentation - https://dev.twitter.com/docs

ΔΙΑΧΕΙΡΙΣΗ ΕΞΕΛΙΣΣΟΜΕΝΩΝ ΒΙΟΛΟΓΙΚΩΝ ΔΕΔΟΜΕΝΩΝ ΣΤΟΝ ΙΣΤΟ

ΠΛΗΡΟΦΟΡΙΕΣ: Θεοδώρα Γαλάνη, theodora@imis.athena-innovation.gr
Γιάννης Σταύρακας, yannis@imis.athena-innovation.gr
Γιώργος Παπαστεφανάτος, gpapas@imis.athena-innovation.gr

ΠΕΡΙΛΗΨΗ: Στόχος της παρούσας διπλωματικής εργασίας είναι η υλοποίηση μιας εφαρμογής για τη διαχείριση εξελισσόμενων βιολογικών δεδομένων στον ιστό. Συγκεκριμένα, η εφαρμογή αυτή θα επιτρέπει τη διαχείριση βιολογικών δεδομένων με τη μορφή γράφου, την εκτέλεση απλών και σύνθετων αλλαγών πάνω σε αυτόν και την προβολή παλαιών στιγμιότυπων. Η διπλωματική έχει ερευνητικό χαρακτήρα και προγραμματιστικό περιεχόμενο.

ΑΤΟΜΑ: 1-2

ΠΛΑΤΦΟΡΜΑ ΕΡΓΑΣΙΑΣ: Java.

ΣΥΝΤΟΜΗ ΠΕΡΙΓΡΑΦΗ: Σε πολλές περιπτώσεις δεδομένα που έχουν δημοσιευτεί στον Ιστό υπόκεινται σε αλλαγές λόγω της εξέλιξης της γνώσης που έχουμε για αυτά. Για παράδειγμα, η επιστημονική κοινότητα των βιολόγων δημοσιεύει αποτελέσματα έρευνας και πειραμάτων στον ιστό, τα οποία επανεξετάζονται και συχνά αλλάζουν οδηγώντας στη δημοσίευση νέων versions. Ένα θέμα που ανακύπτει είναι το πώς οι επιστήμονες θα μπορούσαν να επανεξετάσουν τον τρόπο και τους λόγους για τους οποίους έχουν αλλάξει τα δεδομένα, και να έχουν πρόσβαση σε όλες τις versions βλέποντας την κατάσταση των δεδομένων για διάφορες χρονικές στιγμές. Για αυτό το σκοπό έχει προταθεί ο *evo-graph*, ένα μοντέλο το οποίο καταγράφει εξελισσόμενα δεδομένα ιστού και τις αλλαγές που συμβαίνουν σε αυτά.
Στόχος της παρούσας διπλωματικής εργασίας είναι να υλοποιηθεί μια εφαρμογή για βιολογικά δεδομένα, η οποία θα παρέχει τις δυνατότητες εκτέλεσης αλλαγών πάνω στα δεδομένα και προβολής παρελθοντικών στιγμιότυπων (snapshots) που ισχύουν συγκεκριμένες χρονικές στιγμές. Η υλοποίηση αυτών των λειτουργιών θα βασιστεί στο μοντέλο του *evo-graph* που έχουμε αναπτύξει. Να σημειωθεί ότι για την πραγματοποίηση της παρούσας διπλωματικής στον υποψήφιο φοιτητή θα δοθούν οι προδιαγραφές των συγκεκριμένων λειτουργιών. Ένα επιπλέον θέμα που θα εξεταστεί είναι η ανάπτυξη κατάλληλου user interface που θα επιτρέπει την οπτικοποίηση του γράφου των δεδομένων, ώστε να είναι φιλική στο χρήστη η παρουσίαση των δεδομένων και η εκτέλεση των αλλαγών.
Η διπλωματική θα δώσει μια καλή εμπειρία σε προγραμματισμό, ενώ παράλληλα έχει ερευνητική διάσταση που αφορά τη διαχείριση εξελισσόμενων δεδομένων στον ιστό, θέτοντας τις αλλαγές σαν *πρώτης τάξης πολίτες*.

ΣΧΕΤΙΚΟ ΥΛΙΚΟ:

- Yannis Stavrakas, and George Papastefanatos. Supporting Complex Changes in Evolving Interrelated Web Databanks. International Conference on Cooperative Information Systems (CoopIS 2010), 2010.