

**ΣΥΣΤΗΜΑ ΑΥΤΟΜΑΤΗΣ ΕΞΑΓΩΓΗΣ ΑΛΛΗΛΕΠΙΔΡΑΣΕΩΝ MICRORNA-ΓΟΝΙΔΙΩΝ ΑΠΟ
ΣΥΝΟΔΕΥΤΙΚΟ ΥΛΙΚΟ ΕΠΙΣΤΗΜΟΝΙΚΩΝ ΔΗΜΟΣΙΕΥΣΕΩΝ**

ΠΛΗΡΟΦΟΡΙΕΣ: Θανάσης Βεργούλης, 2106875423, vergoulis@imis.athena-innovation.gr, Ηλίας Κανέλλος, kanellos@dblab.ece.ntua.gr και Θεόδωρος Δαλαμάγκας, dalamag@imis.athena-innovation.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 μοντέλα που μπορούν να χρησιμοποιηθούν για την αυτόματη εξαγωγή αλληλεπιδράσεων μορίων miRNA με γονίδια χρησιμοποιώντας το κείμενο επιστημονικών δημοσιεύσεων. Όμως, η συγκεκριμένη μεθοδολογία αποτυγχάνει να εντοπίσει ένα μεγάλο μέρος των αλληλεπιδράσεων οι οποίες περιέχονται σε συνοδευτικό υλικό των εργασιών όπως είναι πίνακες και εικόνες. Σκοπός της παρούσας εργασίας είναι η υλοποίηση ενός συστήματος που θα επιδιώκει την εξαγωγή σημαντικού μέρους των προαναφερθεισών αλληλεπιδράσεων.

ΑΤΟΜΑ: 1

ΠΛΑΤΦΟΡΜΑ ΕΡΓΑΣΙΑΣ: Python, Java, 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.

Σε προηγούμενη εργασία έχουμε αναπτύξει ένα μοντέλο για την αυτόματη εξαγωγή γνώσης από κείμενα επιστημονικών δημοσιεύσεων. Η εργασία αυτή στηρίζεται σε τεχνικές επεξεργασίας φυσικής γλώσσας, επιτυγχάνει τον εντοπισμό μεγάλου μέρους των αλληλεπιδράσεων που αναφέρονται σε προτάσεις του κειμένου, όμως αποτυγχάνει να εντοπίσει όσες αλληλεπιδράσεις περιέχονται μέσα σε πίνακες, εικόνες ή άλλο συνοδευτικό υλικό των δημοσιεύσεων. Σκοπός είναι να αναπτυχθεί ένα σύστημα που θα μπορεί να εντοπίσει όσο το δυνατόν περισσότερες από τις αλληλεπιδράσεις που περιέχονται σε συνοδευτικό υλικό δημοσιεύσεων.

Η παρούσα εργασία περιλαμβάνει:

- Διερεύνηση των δυνατοτήτων που υπάρχουν για αυτόματη εξαγωγή γνώσης από εικόνες και πίνακες που περιέχονται σε επιστημονικές δημοσιεύσεις για βιομόρια microRNA (πχ χρήση αλγορίθμων OCR ή αλγορίθμων για parsing πινάκων κτλ),
- Υλοποίηση αλγορίθμων που χρησιμοποιούν τα αποτελέσματα της προηγούμενης διερεύνησης προκειμένου να εξάγουν γνώση σχετική με τα βιομόρια microRNA

ΣΧΕΤΙΚΟ ΥΛΙΚΟ:

- Σύστημα αυτόματης εξαγωγής αλληλεπιδράσεων μεταξύ μορίων microRNA και γονιδίων από επιστημονικές βάσεις δεδομένων (διπλωματική εργασία Ροδοθέας-Μυρσίνης Τσουπίδη):
<http://www.dbnet.ece.ntua.gr/pubs/uploads/DIPL-2014-6.pdf>

**ΥΛΟΠΟΙΗΣΗ ΛΕΙΤΟΥΡΓΙΩΝ ΕΞΑΤΟΜΙΚΕΥΣΗΣ ΣΤΗ ΠΛΑΤΦΟΡΜΑ ΒΙΟΠΛΗΡΟΦΟΡΙΚΗΣ
DIANA TOOLS**

ΠΛΗΡΟΦΟΡΙΕΣ: Θανάσης Βεργούλης, 2106875423, vergoulis@imis.athena-innovation.gr, Γιώργος Γιαννόπουλος, giann@imis.athena-innovation.gr και Θοδωρής Δαλαμάγκας, dalamag@imis.athena-innovation.gr

ΠΕΡΙΛΗΨΗ: Η πλατφόρμα βιοπληροφορικής DIANA TOOLS αποτελείται από ένα σύνολο εργαλείων που στοχεύουν στην υποβοήθηση των βιοεπιστημόνων που ασχολούνται με την έρευνα γύρω από τα μόρια microRNA. Μέσα από τα εργαλεία της οι βιοεπιστήμονες μπορούν να εντοπίζουν βιβλιογραφία που τους ενδιαφέρει, να ενημερώνονται για τα γονίδια και τα μεταβολικά μονοπάτια που επηρεάζονται από τη λειτουργία ενός συγκεκριμένου μορίου microRNA κτλ.

Η καταγραφή του ιστορικού αναζητήσεων ενός χρήστη, τα αποτελέσματα τα οποία επιλέγει να εξετάσει και, γενικότερα, η καταγραφή του τρόπου με τον οποίο αλληλεπιδρά με τη διεπαφή των εργαλείων μπορούν να οδηγήσουν στη δημιουργία ενός προφίλ των ερευνητικών ενδιαφερόντων και των προτιμήσεών του. Αυτό το προφίλ, όπως και το προφίλ άλλων ερευνητών που παρουσιάζουν παρόμοια συμπεριφορά, μπορούν να χρησιμοποιηθούν για να παρέχονται στο βιοεπιστήμονα εξατομικευμένες υπηρεσίες όπως προτάσεις για ενδιαφέρουσα βιβλιογραφία, για πειράματα που θα μπορούσε να πραγματοποιήσει, για γονίδια που μπορεί να σχετίζονται με τις ασθένειες που μελετά κτλ.

ΑΤΟΜΑ: 1

ΠΛΑΤΦΟΡΜΑ ΕΡΓΑΣΙΑΣ: PHP, MySQL, Java.

ΣΥΝΤΟΜΗ ΠΕΡΙΓΡΑΦΗ: Στόχος της παρούσας εργασίας είναι η ανάπτυξη ενός συστήματος καταγραφής της δραστηριότητας των χρηστών των εργαλείων της πλατφόρμας DIANA TOOLS και η υλοποίηση λειτουργιών εξατομίκευσης που θα εκμεταλλεύονται τα καταγραφόμενα δεδομένα. Ανάμεσα στα δεδομένα που θα καταγράφονται θα πρέπει να είναι το ιστορικό αναζητήσεων των χρηστών, τα αποτελέσματα που επισκέπτονται οι χρήστες μετά από κάθε αναζήτηση, οι εξωτερικοί υπερσύνδεσμοι που ακολουθούν κτλ. Με βάση τα προφίλ χρηστών που θα κατασκευαστούν θα χρησιμοποιηθούν καθιερωμένοι αλγόριθμοι εξατομίκευσης προκειμένου να παρέχονται προτάσεις βιβλιογραφίας και εξατομικευμένες λίστες αποτελεσμάτων με βάση τα ερευνητικά ενδιαφέροντα του κάθε χρήστη.

Τα καθήκοντα της εργασίας είναι τα παρακάτω:

- Σχεδιασμός και ανάπτυξη συστήματος καταγραφής της δραστηριότητας των χρηστών της πλατφόρμας DIANA TOOLS.
- Υλοποίηση υπηρεσίας που χρησιμοποιεί το προφίλ του χρήστη για την παροχή προτάσεων ενδιαφέρουσας βιβλιογραφίας και εξατομικευμένων λιστών αποτελεσμάτων.

ΣΧΕΤΙΚΟ ΥΛΙΚΟ:

- Πλατφόρμα εργαλείων DIANA TOOLS: <http://diana.imis.athena-innovation.gr/DianaTools/index.php>
- Μέθοδοι εξατομίκευσης αναζήτησης:
 - ο T. Joachims. Optimizing search engines using clickthrough data. ACM SIGKDD, 2002.

- K. Sugiyama, K. Hatano, and M. Yoshikawa. Adaptive web search based on user profile constructed without any effort from users. WWW, 2004.
- Z. Dou, R. Song and J.-R. Wen. A Large-scale Evaluation and Analysis of Personalized Search Strategies. WWW, 2007.
- Z. Dou, R. Song, J.-R. Wen, and X. Yuan. Evaluating the Effectiveness of Personalized Web Search. IEEE TKDE, 2008.

**ΜΕΛΕΤΗ ΚΑΙ ΕΠΕΚΤΑΣΗ ΑΛΓΟΡΙΘΜΩΝ ΑΝΑΓΝΩΡΙΣΗΣ ΜΟΤΙΒΩΝ ΣΕ ΧΡΟΝΟΣΕΙΡΕΣ
ΚΑΤΑΝΑΛΩΣΗΣ ΝΕΡΟΥ**

ΠΛΗΡΟΦΟΡΙΕΣ: Γιώργος Γιαννόπουλος, giann [at] imis.athena-innovation.gr

Παντελής Χρόνης, pchronis [at] imis.athena-innovation.gr

Σπύρος Αθανασίου, sathan [at] imis.athena-innovation.gr

ΠΕΡΙΛΗΨΗ: Στόχος της διπλωματικής είναι η υλοποίηση αλγορίθμων αναγνώρισης μοτίβων σε χρονοσειρές και η επέκταση/προσαρμογή τους στο πρόβλημα της ανάλυσης δεδομένων κατανάλωσης νερού. Η ανάπτυξη της αντίστοιχης εφαρμογής θα πραγματοποιηθεί για κινητές συσκευές, σύμφωνα με τις απαιτήσεις του προβλήματος, όπως περιγράφεται στη συνέχεια.

ΑΤΟΜΑ: 1

ΠΛΑΤΦΟΡΜΑ ΕΡΓΑΣΙΑΣ: Java, Javascript, Objective-C, Cordova, Flot

ΣΥΝΤΟΜΗ ΠΕΡΙΓΡΑΦΗ:

Μία αρκετά διαδεδομένη κατηγορία δεδομένων είναι οι χρονοσειρές. Η χρονοσειρά είναι μία πλειάδα τιμών κάθε μία από τις οποίες αντιστοιχεί σε κάποια χρονική στιγμή. Χρονοσειρές μπορεί να παράγονται από αισθητήρες ή από τη λειτουργία διαφόρων συστημάτων και συνήθως οι τιμές λαμβάνονται σε ίσα διαστήματα. Παράδειγμα τέτοιων δεδομένων είναι οι τιμές ενός αισθητήρα που μετράει την κατανάλωση νερού σε μία βρύση, τα δεδομένα νομισματικής ισοτιμίας ευρώ/δολαρίου ή η ποσότητα βροχόπτωσης σε μία περιοχή. Οι χρονοσειρές, ως τύπος δεδομένων, έχουν κάποιες ιδιαιτερότητες, όπως η υψηλή συσχέτιση διαδοχικών δειγμάτων και η μεγάλη διαστατικότητα.

Η συγκεκριμένη διπλωματική επικεντρώνεται στην ανάλυση χρονοσειρών που προκύπτουν από συλλογή δεδομένων κατανάλωσης νερού. Η συλλογή των δεδομένων γίνεται από ειδικούς μετρητές [1], οι οποίοι προσαρμόζονται σε μία βρύση και μπορούν να καταγράφουν ποσότητα καταναλωμένου νερού και θερμοκρασία, ανά συνεχόμενες χρονικές στιγμές. Η μετρούμενη κατανάλωση εξαρτάται από διαφορετικούς παράγοντες, όπως η εποχή, η μέρα της εβδομάδας, η θερμοκρασία περιβάλλοντος, δημογραφικά στοιχεία των καταναλωτών, κ.α. Στόχος είναι η ανάλυση των συγκεκριμένων χρονοσειρών και η δημιουργία μοντέλων που προβλέπουν τη μελλοντική κατανάλωση, με βάση τις τιμές των προαναφερθέντων παραγόντων.

Στόχος της διπλωματικής είναι η ανάπτυξη βασικών αλγορίθμων αναγνώρισης μοτίβων (pattern recognition) ([2]) και αναγνώρισης ασυνήθιστων γεγονότων (event detection) ([3]) κατανάλωσης σε χρονοσειρές. Θα υλοποιηθεί/επεκταθεί, καταρχήν, αλγόριθμος γραμμικής παλινδρόμησης (linear regression) ([4]), γύρω από τον οποίο θα μελετηθούν διάφορες τεχνικές για αναγνώριση γεγονότων και μοτίβων και θα υλοποιηθεί ένα υποσύνολο από αυτές.

Σκοπός των παραπάνω είναι η ανάπτυξη ενός πλαισίου που θα επιτρέπει στον καταναλωτή να μαθαίνει πληροφορίες για την κατανάλωσή του, τόσο σε πραγματικό χρόνο, όσο και αθροιστικά δεδομένα (πρόβλεψη κατανάλωσης, μοτίβα-είδη κατανάλωσης και ασυνήθιστες καταναλώσεις) ύστερα από ανάλυση των δεδομένων. Λόγω της πρόβλεψης των συσκευών-μετρητών για ασύρματη επικοινωνία μέσω Bluetooth, ιδανικότερες συσκευές τελικού χρήστη αποτελούν οι κινητές συσκευές (smartphones, tablets). Για αυτό το λόγο, η ανάπτυξη των αλγορίθμων και της εφαρμογής θα γίνει για κινητά τηλέφωνα και για τα λογισμικά Android ή/και iOS. Οι τεχνολογίες που θα χρησιμοποιηθούν είναι (ανάλογα με τις γνώσεις του υποψήφιου

και την κατεύθυνση της διπλωματικής): Java, Objective-C, Javascript και οι βιβλιοθήκες Cordova [5] και Flot [6].

Η διπλωματική θα πραγματοποιηθεί στα πλαίσια του ερευνητικού έργου DAIAD το οποίο είναι ένα χρηματοδοτούμενο από την ΕΕ ερευνητικό έργο, διάρκειας 3,5 ετών που αφορά στη διαχείριση μεγάλων δεδομένων νερού. Συνοπτικά, το DAIAD εστιάζει στην αποδοτική συλλογή, αποθήκευση, ανάλυση και εκμετάλλευση (σε επίπεδο πρόβλεψης κατανάλωσης, αναγνώρισης ασυνήθιστων καταναλώσεων και εξατομικευμένων συστάσεων κατανάλωσης) δεδομένων κατανάλωσης νερού. Τα ερευνητικά του αποτελέσματα δύνανται να χρησιμοποιηθούν τόσο από τελικούς χρήστες, όσο και από παρόχους ύδρευσης και ερευνητές μοντέλων κατανάλωσης νερού.

ΣΧΕΤΙΚΕΣ ΠΛΗΡΟΦΟΡΙΕΣ:

- [1] Amphiro a1 device. Available at: <http://amphiro.com/products/a1/>
- [2] Pattern Recognition in Time Series. Available at: <http://www.cs.gmu.edu/~jessica/publications/astronomy11.pdf>
- [3] Tutorial on Event Detection, KDD 2009. <http://www.cs.cmu.edu/~neill/papers/eventdetection.pdf>
- [4] Linear Regression. Available at: <http://stattrek.com/regression/linear-regression.aspx>
- [5] Apache Cordova. Available at: <http://cordova.apache.org/>
- [6] Flot: Attractive JavaScript plotting for jQuery. Available at: <http://www.flotcharts.org/>

ΜΕΛΕΤΗ ΚΑΙ ΕΠΕΚΤΑΣΗ ΑΛΓΟΡΙΘΜΩΝ ΜΗΧΑΝΙΚΗΣ ΜΑΘΗΣΗΣ ΓΙΑ ΔΕΔΟΜΕΝΑ ΧΡΟΝΟΣΕΙΡΩΝ ΣΕ MAP-REDUCE

ΠΛΗΡΟΦΟΡΙΕΣ: Γιώργος Γιαννόπουλος, giann [at] imis.athena-innovation.gr

Παντελής Χρόνης, pchronis [at] imis.athena-innovation.gr

Σπύρος Αθανασίου, sathan [at] imis.athena-innovation.gr

ΠΕΡΙΛΗΨΗ: Στόχος της διπλωματικής είναι η μελέτη αλγορίθμων μηχανικής μάθησης για δεδομένα χρονοσειρών και η επέκταση/προσαρμογή τους στο μοντέλο Map-Reduce για επεξεργασία μεγάλων δεδομένων. Οι αλγόριθμοι θα προσαρμοστούν στο σενάριο όπου τα δεδομένα αποτελούνται από χρονοσειρές κατανάλωσης νερού από μεγάλο αριθμό χρηστών.

ΑΤΟΜΑ: 1

ΠΛΑΤΦΟΡΜΑ ΕΡΓΑΣΙΑΣ: Java, Hadoop, Apache Flink, Cassandra

ΣΥΝΤΟΜΗ ΠΕΡΙΓΡΑΦΗ:

Μία αρκετά διαδεδομένη κατηγορία δεδομένων είναι οι χρονοσειρές. Η χρονοσειρά είναι μία πλειάδα τιμών κάθε μία από τις οποίες αντιστοιχεί σε κάποια χρονική στιγμή. Χρονοσειρές μπορεί να παράγονται από αισθητήρες ή από τη λειτουργία διαφόρων συστημάτων και συνήθως οι τιμές λαμβάνονται σε ίσα διαστήματα. Παράδειγμα τέτοιων δεδομένων είναι οι τιμές ενός αισθητήρα που μετράει την κατανάλωση νερού σε μία βρύση, τα δεδομένα νομισματικής ισοτιμίας ευρώ/δολαρίου ή η ποσότητα βροχόπτωσης σε μία περιοχή. Οι χρονοσειρές, ως τύπος δεδομένων, έχουν κάποιες ιδιαιτερότητες, όπως η υψηλή συσχέτιση διαδοχικών δειγμάτων και η μεγάλη διαστατικότητα.

Η συγκεκριμένη διπλωματική επικεντρώνεται στην ανάλυση χρονοσειρών που προκύπτουν από συλλογή δεδομένων κατανάλωσης νερού. Η συλλογή των δεδομένων γίνεται από ειδικούς μετρητές [1], οι οποίοι προσαρμόζονται σε μία βρύση και μπορούν να καταγράφουν ποσότητα καταναλωμένου νερού και θερμοκρασία, ανά συνεχόμενες χρονικές στιγμές. Η μετρούμενη κατανάλωση εξαρτάται από διαφορετικούς παράγοντες, όπως η εποχή, η μέρα της εβδομάδας, η θερμοκρασία περιβάλλοντος, δημογραφικά στοιχεία των καταναλωτών, κ.α. Στόχος είναι η ανάλυση των συγκεκριμένων χρονοσειρών και η δημιουργία μοντέλων που προβλέπουν τη μελλοντική κατανάλωση, με βάση τις τιμές των προαναφερθέντων παραγόντων.

Στόχος της διπλωματικής είναι η μελέτη αλγορίθμων μηχανικής μάθησης για χρονοσειρές ([2]) και η επιλογή ενός υποσυνόλου τους για υλοποίηση/επέκταση στο μοντέλο Map-Reduce [3] για κατανομημένη επεξεργασία μαζικών δεδομένων. Το σύνολο των αλγορίθμων που θα υλοποιηθούν θα περιέχει έναν τουλάχιστον αλγόριθμο συσταδοποίησης χρονοσειρών και έναν αλγόριθμο αναγνώρισης προτύπων. Οι αλγόριθμοι θα λαμβάνουν υπόψη, πέρα από τα δεδομένα χρονοσειρών και μεταδεδομένα όπως οι προαναφερθέντες παράγοντες που επηρεάζουν την κατανάλωση και θα πρέπει να είναι *παραμετροποιήσιμοι* ως προς αυτούς τους παράγοντες αλλά και ως προς τη διακριτότητα (granularity) των δεδομένων.

Η διπλωματική θα δώσει έμφαση στην αποδοτική υλοποίηση των δύο (τουλάχιστον) αλγορίθμων και στην κλιμακωσιμότητά τους σε μεγάλους όγκους δεδομένων. Για αυτό το λόγο, στα πλαίσια της διπλωματικής, θα επεκταθεί επίσης ένας υπάρχων synthetic data generator με σκοπό την παραγωγή μαζικών χρονοσειρών που προσεγγίσουν τις κατανομές πραγματικών δεδομένων κατανάλωσης νερού.

Η διπλωματική θα πραγματοποιηθεί στα πλαίσια του ερευνητικού έργου DAIAD το οποίο είναι ένα χρηματοδοτούμενο από την ΕΕ ερευνητικό έργο, διάρκειας 3,5 ετών που αφορά στη διαχείριση μεγάλων δεδομένων νερού. Συνοπτικά, το DAIAD εστιάζει στην αποδοτική συλλογή, αποθήκευση, ανάλυση και εκμετάλλευση (σε επίπεδο πρόβλεψης κατανάλωσης, αναγνώρισης ασυνήθιστων καταναλώσεων και εξατομικευμένων συστάσεων κατανάλωσης) δεδομένων κατανάλωσης νερού. Τα ερευνητικά του αποτελέσματα δύνανται να χρησιμοποιηθούν τόσο από τελικούς χρήστες, όσο και από παρόχους ύδρευσης και ερευνητές μοντέλων κατανάλωσης νερού.

ΣΧΕΤΙΚΕΣ ΠΛΗΡΟΦΟΡΙΕΣ:

- [1] Amphiro a1 device. Available at: <http://amphiro.com/products/a1/>
- [2] Pattern Recognition in Time Series. Available at: <http://www.cs.gmu.edu/~jessica/publications/astronomy11.pdf>
- [3] MapReduce Tutorial. Available at: http://hadoop.apache.org/docs/r1.2.1/mapred_tutorial.html
- [4] Hadoop. Available at: <http://hadoop.apache.org/>
- [5] Apache Flink. Available at: <https://flink.incubator.apache.org/>
- [6] Cassandra. Available at: <http://cassandra.apache.org/>

**ΥΛΟΠΟΙΗΣΗ ΑΛΓΟΡΙΘΜΩΝ ΜΗΧΑΝΙΚΗΣ ΜΑΘΗΣΗΣ ΓΙΑ ΑΥΤΟΜΑΤΟΠΟΙΗΜΕΝΗ
ΣΥΓΧΩΝΕΥΣΗ ΣΗΜΑΣΙΟΛΟΓΙΚΩΝ ΓΕΩΧΩΡΙΚΩΝ ΔΕΔΟΜΕΝΩΝ**

*ΠΛΗΡΟΦΟΡΙΕΣ: Γιώργος Γιαννόπουλος, giann [at] dlablab.ece.ntua.gr
Δημήτρης Σκούτας, dskoutas [at] imis.athena-innovation.gr
Σπύρος Αθανασίου, sathan [at] imis.athena-innovation.gr*

ΠΕΡΙΛΗΨΗ: Στόχος της διπλωματικής είναι η υλοποίηση αλγορίθμων για αυτοματοποιημένη συγχώνευση (fusion) οντοτήτων σε σημασιολογικά δεδομένα με γεωχωρική πληροφορία. Οι αλγόριθμοι θα εκπαιδεύονται πάνω στο ιστορικό του χρήστη, δηλαδή σε προηγούμενες στρατηγικές συγχώνευσης δεδομένων που έχει επιλέξει και θα μπορούν, με βάση το εκπαιδευμένο μοντέλο, να προτείνουν στρατηγικές για νέα, προς συγχώνευση δεδομένα.

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.).

Η συγκεκριμένη διπλωματική θα επικεντρωθεί στο πρόβλημα της συγχώνευσης δεδομένων (data fusion), το οποίο συνίσταται στα εξής στάδια: (α) Επιλογή ενός ζεύγους οντοτήτων που έχουν διασυνδεθεί ως όμοιες, (β) Αντιστοίχιση των ιδιοτήτων της μιας οντότητας με τις ιδιότητες της άλλης, ώστε να βρεθούν τα σωστά ζεύγη ιδιοτήτων που περιγράφουν τα ίδια χαρακτηριστικά των οντοτήτων (γ) επιλογή, για κάθε ζεύγος ιδιοτήτων, συνδυασμού του επιμέρους περιεχομένου τους (στρατηγική συγχώνευσης), ώστε η τελική ιδιότητα να περιέχει το ορθότερο και πληρέστερο δυνατό περιεχόμενο.

Στα πλαίσια της διπλωματικής θα υλοποιηθούν αλγόριθμοι supervised learning [13] και active learning [14] οι οποίοι θα λαμβάνουν υπόψη τα ιδιαίτερα χαρακτηριστικά των δεδομένων που χειριζόμαστε και θα τα χρησιμοποιούν ως χαρακτηριστικά εκπαίδευσης (training features). Το μοντέλο θα εκπαιδεύεται είτε πάνω στα ιστορικά δεδομένα συγχώνευσης του χρήστη (supervised learning), είτε συνδυάζοντας ιστορικά δεδομένα και άμεση ανάδραση χρήστη (active learning). Η ανάπτυξη των αλγορίθμων θα γίνει πάνω στο FAGI-gis [12], εργαλείο που αναπτύσσεται στο ΠΠΣΥ στα πλαίσια του ευρωπαϊκού ερευνητικού έργου GeoKnow.

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 Σηματολογικό Ιστό. Συνοπτικά, το GeoKnow εστιάζει στη 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:

- [7] Tim Berners-Lee et al. *The Semantic Web*, Scientific American, May 17, 2001, Available at: www.dblab.ece.ntua.gr/~bikakis/SW.pdf
- [8] Resource Description Framework (RDF), http://www.w3schools.com/rdf/rdf_intro.asp
- [9] RDF Primer, <http://www.w3.org/TR/rdf-syntax/>
- [10] Simple Protocol and RDF Query Language (SPARQL), <http://www.slideshare.net/olafhartig/an-introduction-to-sparql>
- [11] 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
- [12] FAGI tool, http://svn.aksw.org/projects/GeoKnow/Public/D3.2.1_Fusing_of_geographic_features.pdf
- [13] Supervised learning. Available at: http://en.wikipedia.org/wiki/Supervised_learning
- [14] Active Learning Literature Survey. Available at: <http://burrsettles.com/pub/settles.activelearning.pdf>

ΑΝΑΛΥΣΗ ΚΑΙ ΑΝΑΖΗΤΗΣΗ ΓΕΩΧΩΡΙΚΗΣ ΠΛΗΡΟΦΟΡΙΑΣ
ΑΠΟ ΠΗΓΕΣ ΤΟΥ ΔΙΑΔΙΚΤΥΟΥ

ΠΛΗΡΟΦΟΡΙΕΣ: Δημήτρης Σκούτας, 2106875415, dskoutas@imis.athena-innovation.gr

ΠΕΡΙΛΗΨΗ: Στόχος της διπλωματικής εργασίας είναι η σχεδίαση και υλοποίηση μιας Web ή mobile εφαρμογής που θα συνδυάζει γεωχωρικά δεδομένα διαθέσιμα στον Παγκόσμιο Ιστό και τα κοινωνικά δίκτυα και θα παρέχει δυνατότητες αναζήτησης και οπτικοποίησης των αποτελεσμάτων.

ΑΤΟΜΑ: 1

ΠΛΑΤΦΟΡΜΑ ΕΡΓΑΣΙΑΣ: Java, PostGIS, Solr.

ΣΥΝΤΟΜΗ ΠΕΡΙΓΡΑΦΗ: Στις μέρες μας υπάρχει μια μεγάλη πληθώρα γεωχωρικών δεδομένων που δημιουργούνται από μεμονωμένους χρήστες ή κοινότητες χρηστών και είναι διαθέσιμα στο διαδίκτυο και στα διάφορα μέσα κοινωνικής δικτύωσης (π.χ. OpenStreetMap, Wikimapia, Flickr, Twitter). Η συνδυασμένη ανάλυση τέτοιων δεδομένων είναι ιδιαίτερα χρήσιμη για την παροχή υπηρεσιών βασισμένων στη θέση του χρήστη, καθώς και για την εξαγωγή γνώσης σχετικά με το είδος μιας τοποθεσίας ή γεγονότα που συμβαίνουν σε αυτή.

Σκοπός της παρούσας εργασίας είναι η ανάπτυξη μιας εφαρμογής για τη διαχείριση και απεικόνιση τέτοιας πληροφορίας από διαφορετικές πηγές. Συγκεκριμένα, η εφαρμογή θα περιλαμβάνει:

- συνδυασμό γεωχωρικής και κειμενικής πληροφορίας από διάφορες πηγές (π.χ. σημεία ενδιαφέροντος, φωτογραφίες, tweets)
- ταξινόμηση αποτελεσμάτων αναζήτησης και δημιουργία περιλήψεων
- οπτικοποίηση αποτελεσμάτων μέσω Web ή mobile διεπαφής με βάση την τοποθεσία του χρήστη ή περιοχές που έχει επιλέξει.

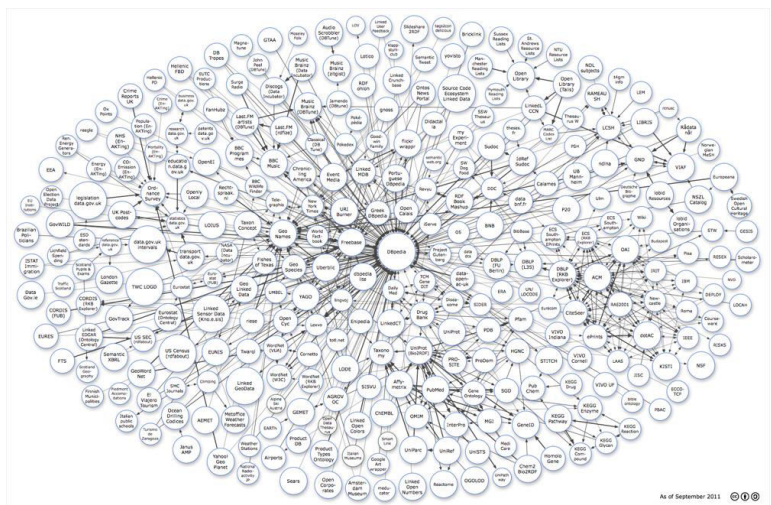
ΔΙΑΣΥΝΔΕΔΕΜΕΝΑ ΔΕΔΟΜΕΝΑ (LINKED DATA)
 ΕΙΣΑΓΩΓΗ

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 προερχόμενα 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. Αποτελεί 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 εικόνα¹, 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 κατηγοριοποίηση² **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 ανοιχτής πρόσβασης.
- ★★ Η πληροφορία θα πρέπει να διατίθεται στο Web με δομημένο τρόπο (π.χ., δεδομένα σε Excel αντί για την εικόνα ενός πίνακα).
- ★★★ Θα πρέπει να γίνεται χρήση ανοιχτών και μη εμπορικών προτύπων για τη μορφοποίηση των δεδομένων (π.χ., CSV αντί για Excel).
- ★★★★ Θα πρέπει να γίνεται χρήση URIs για τον προσδιορισμό αντικειμένων, έτσι ώστε να είναι δυνατή η αναφορά και οι δεικτοδότηση τους από άλλους χρήστες.
- ★★★★★ Θα πρέπει να δημιουργούνται links προς άλλα δεδομένα που περιγράφουν σχετική 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>

Tom Heath and Christian Bizer (2011) *Linked Data: Evolving the Web into a Global Data Space* (1st edition). Synthesis Lectures on the Semantic Web: Theory and Technology, 1:1, 1-136. Morgan & Claypool. (<http://linkeddatabook.com/editions/1.0/>)

Linked Data Glossary (<http://www.w3.org/TR/2013/NOTE-ld-glossary-20130627/>)

² <http://5stardata.info/>

**ΑΝΑΚΤΗΣΗ ΚΑΙ ΜΕΤΑΤΡΟΠΗ WEB DATA
ΣΕ ΠΟΡΟΥΣ ΔΙΑΣΥΝΔΕΔΕΜΕΝΩΝ ΔΕΔΟΜΕΝΩΝ**

ΠΛΗΡΟΦΟΡΙΕΣ: Γ. Αλεξίου, Μ. Μειμάρης, Γ. Παπαστεφανάτος,
{galexiou,m.meimaris,grapas}@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. Η εκμετάλλευση υπάρχουσας πληροφορίας μεταδεδομένων σε διαδικτυακούς πόρους σε συνδυασμό με τεχνικές διασύνδεσης δεδομένων μπορεί να οδηγήσει σε πιο πλήρεις αναπαραστάσεις πόρων, με απώτερο στόχο την επίτευξη συγκρισιμότητας μεταξύ ετερογενών αντικειμένων ή αντικειμένων από διαφορετικές πηγές. Στην παρούσα εργασία ο στόχος είναι η αυτόματη ενσωμάτωση δεδομένων που εκφράζονται με γνωστά λεξιλόγια για html injection (π.χ. microdata, json-ld) από ιστοσελίδες με παρόμοιο περιεχόμενο (π.χ. προϊόντα από ηλεκτρονικά καταστήματα) εντός της πλατφόρμας σημασιολογικού ιστού LinkZoo³ και η δημιουργία προσωπικών/συνεργατικών χώρων ενσωμάτωσης τέτοιου είδους δεδομένων, ώστε οι χρήστες να μπορούν να συγκρίνουν και να μοιράζονται προσωποποιημένες συλλογές ετερογενών αντικειμένων χωρίς περαιτέρω γνώση ή εξειδίκευση.

ΑΤΟΜΑ: 1.

ΠΛΑΤΦΟΡΜΑ ΕΡΓΑΣΙΑΣ: Java, RDF, Javascript.

ΣΥΝΤΟΜΗ ΠΕΡΙΓΡΑΦΗ: Στα πλαίσια των δραστηριοτήτων του ΠΠΣΥ γύρω από τα διασυνδεδεμένα δεδομένα, αναπτύσσεται το LinkZoo, ένα συνεργατικό περιβάλλον διαχείρισης πόρων (π.χ. αρχεία, ιστοσελίδες, επαφές) βασισμένο εξολοκλήρου σε τεχνολογίες Linked Data. Στη συγκεκριμένη εργασία ο στόχος είναι η ανάπτυξη κώδικα προς επέκταση της πλατφόρμας για την αυτόματη εξαγωγή μεταδεδομένων και ενσωμάτωση αυτών στην πλατφόρμα. Πιο συγκεκριμένα, θα πρέπει να γίνει α) περιορισμός του πεδίου εφαρμογής σε ένα πλούσιο σε μεταδεδομένα και ετερογένεια πεδίο (π.χ. ηλεκτρονικό εμπόριο) β) ανάκτηση των μεταδεδομένων από ιστοσελίδες που περιγράφουν αντικείμενα του πεδίου (π.χ. προϊόντα προς πώληση) γ) μετατροπή αυτών σε πόρους του Σημασιολογικού Ιστού (Semantic Web) και τέλος δ) εισαγωγή στην πλατφόρμα διασυνδεδεμένων δεδομένων LinkZoo.

Στα καθήκοντα της ανάληψης της συγκεκριμένης διπλωματικής εργασίας εντάσσονται:

- Ανάπτυξη λογισμικού για την αναγνώριση και εξαγωγή υπάρχοντων μεταδεδομένων σε ιστοσελίδες
- Ανάπτυξη λογισμικού για την μετατροπή των δεδομένων σε πόρους του Σ.Ι.
- Ενσωμάτωση του λογισμικού με τα API services του LinkZoo
- Ανάπτυξη λογισμικού για τη διασύνδεση των πόρων και τον περαιτέρω εμπλουτισμό των μεταδεδομένων τους.

ΣΧΕΤΙΚΟ ΥΛΙΚΟ:

1. <http://www.heppnetz.de/projects/goodrelations/>, <http://www.w3.org/TR/microdata-rdf/>
2. M. Meimaris, G. Alexiou, G. Papastefanatos: LinkZoo: A Linked Data Platform for Collaborative Management of Heterogeneous Resources. In ESWC 2014: 407-412. http://2014.eswc-conferences.org/sites/default/files/eswc2014pd_submission_74.pdf

³ <http://www.linkzoo.gr>

**ΑΝΑΚΤΗΣΗ ΚΑΙ ΜΕΤΑΤΡΟΠΗ OPEN DATASETS
ΣΕ ΠΟΡΟΥΣ ΔΙΑΣΥΝΔΕΔΕΜΕΝΩΝ ΔΕΔΟΜΕΝΩΝ**

ΠΛΗΡΟΦΟΡΙΕΣ: Γ. Αλεξίου, Μ. Μειμάρης, Γ. Παπαστεφανάτος,
{galexiou,m.meimaris,grapas}@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. Η εκμετάλλευση υπάρχουσας πληροφορίας μεταδεδομένων σε datasets από πηγές όπως το CKAN, σε συνδυασμό με τεχνικές διασύνδεσης δεδομένων μπορεί να οδηγήσει σε πιο πλήρεις καταγραφές καταλόγων datasets, με απώτερο στόχο την ανακαλυψιμότητα και την επίτευξη συγκρισιμότητας μεταξύ ετερογενών datasets. Στην παρούσα εργασία ο στόχος είναι η εξαγωγή μετδεοδμένων και η αυτόματη ενσωμάτωση datasets εντός της πλατφόρμας σημασιολογικού ιστού LinkZoo⁴ και η δημιουργία προσωπικών/συνεργατικών χώρων ενσωμάτωσης τέτοιου είδους δεδομένων.

ΑΤΟΜΑ: 1.

ΠΛΑΤΦΟΡΜΑ ΕΡΓΑΣΙΑΣ: Java, Javascript, RDF.

ΣΥΝΤΟΜΗ ΠΕΡΙΓΡΑΦΗ: Στα πλαίσια των δραστηριοτήτων του ΠΠΣΥ γύρω από τα διασυνδεδεμένα δεδομένα, αναπτύσσεται το LinkZoo, ένα συνεργατικό περιβάλλον διαχείρισης πόρων (π.χ. αρχεία, ιστοσελίδες, επαφές) βασισμένο εξολοκλήρου σε τεχνολογίες Linked Data. Στη συγκεκριμένη εργασία ο στόχος είναι η ανάπτυξη κώδικα προς επέκταση της πλατφόρμας για την αυτόματη εξαγωγή μεταδεδομένων για υπάρχοντα datasets σε πηγές όπως CKAN και ενσωμάτωση αυτών στην πλατφόρμα για εύκολη διαχείριση, ανακάλυψη και συνεργασία. Πιο συγκεκριμένα, θα πρέπει να γίνει α) αναγνώριση πιθανών πηγών καταλόγων datasets με διαθέσιμα μεταδεδομένα β) ανάκτηση των μεταδεδομένων των datasets γ) μετατροπή αυτών σε πόρους του Σημασιολογικού Ιστού (Semantic Web) και τέλος δ) εισαγωγή στην πλατφόρμα διασυνδεδεμένων δεδομένων LinkZoo.

Στα καθήκοντα της ανάληψης της συγκεκριμένης διπλωματικής εργασίας εντάσσονται:

- Αναγνώριση πιθανών πηγών datasets (π.χ. κατάλογοι CKAN) με διαθέσιμα μεταδεδομένα
- Ανάπτυξη λογισμικού για την αναγνώριση και εξαγωγή υπαρχόντων μεταδεδομένων σε datasets
- Ανάπτυξη λογισμικού για την μετατροπή των δεδομένων σε πόρους του Σ.Ι.
- Ενσωμάτωση του λογισμικού με API services του LinkZoo

ΣΧΕΤΙΚΟ ΥΛΙΚΟ:

1. M. Meimaris, G. Alexiou, G. Papastefanatos: LinkZoo: A Linked Data Platform for Collaborative Management of Heterogeneous Resources. In ESWC 2014: 407-412. http://2014.eswc-conferences.org/sites/default/files/eswc2014pd_submission_74.pdf
2. <http://ckan.org/>, <http://www.w3.org/TR/vocab-dcat/>

⁴ <http://www.linkzoo.gr>

ΑΝΑΓΝΩΡΙΣΗ ΚΑΙ ΑΝΤΙΣΤΙΧΟΙΣΗ ΙΕΡΑΡΧΙΩΝ ΣΤΟ ΣΗΜΑΣΙΟΛΟΓΙΚΟ ΙΣΤΟ

ΠΛΗΡΟΦΟΡΙΕΣ: Μ. Μειμάρης, Γ. Παπαστεφανάτος
{m.meimaris, gparas}@imis.athena-innovation.gr

ΠΕΡΙΛΗΨΗ: Οι ιεραρχικές δομές σε λεξιλόγια όρων είναι μια συνήθης πρακτική για την καταγραφή σχέσεων γονέα-παιδιού μεταξύ των όρων (π.χ. ταξινόμηση γεωγραφικών περιοχών, επαγγελμάτων, προϊόντων, κτλ) και την ταξινόμηση \ κωδικοποίηση τιμών. Τέτοιου είδους λεξιλόγια χρησιμοποιούνται στο πλαίσιο της καταγραφής πληροφοριών με πολυδιάστατο τρόπο σε διαφορετικά επίπεδα γενίκευσης/ειδίκευσης, με στόχο την εφαρμογή τεχνικών πολυδιάστατης ανάλυσης. Εντός του σημασιολογικού ιστού, αποτελεί καλή πρακτική η επαναχρησιμοποίηση κοινών λεξιλογίων όρων μεταξύ διαφορετικών πηγών, ώστε να είναι δυνατή η διαλειτουργικότητα μεταξύ τους. Παρόλα αυτά, στην πραγματικότητα είναι αρκετά σύνηθες να χρησιμοποιούνται διαφορετικά λεξιλόγια για την περιγραφή των ίδιων όρων. Σκοπός της εργασίας αυτής είναι η ανάπτυξη μεθόδων για την αναγνώριση και αντιστοίχιση ιεραρχιών από διαφορετικές πηγές που χρησιμοποιούν ίδιους όρους, στα πλαίσια πολυδιάστατων πηγών στατιστικών και άλλων δεδομένων.

ΑΤΟΜΑ: 1.

ΠΛΑΤΦΟΡΜΑ ΕΡΓΑΣΙΑΣ: Java, RDF, R/MATLAB.

ΣΥΝΤΟΜΗ ΠΕΡΙΓΡΑΦΗ: Στα πλαίσια των δραστηριοτήτων του ΠΙΣΥ γύρω από τα διασυνδεδεμένα δεδομένα, πηγές πολυδιάστατων δεδομένων μελετούνται ως προς την ανάλυση της πληροφορίας τους, την αναπαράσταση υπαρχόντων στατιστικών δεδομένων σε μορφή διασυνδεδεμένων δεδομένων και την εξαγωγή νέας γνώσης βασισμένης σε υπάρχοντα δεδομένα. Στόχος είναι η αξιοποίηση και ανάλυση της πληθώρας διασυνδεδεμένων δεδομένων που είναι δημοσιευμένα στον ιστό μέσω της εφαρμογής τεχνικών πολυδιάστατης ανάλυσης σε αυτά και γενικότερα τεχνικών data analytics. Σε αυτό το ερευνητικό πλαίσιο, αναζητούνται τρόποι, τεχνικές και μέθοδοι για την ανακάλυψη κοινών όρων, την ευθυγράμμιση των ιεραρχικών επιπέδων, καθώς και τη διασύνδεση κοινών όρων που όμως έχουν διαφορετική αναπαράσταση, περιγραφή ή αναγνωριστικό. Πιο συγκεκριμένα, θα πρέπει α) να μελετηθούν οι τρόποι με τους οποίους εμφανίζονται συνηθισμένες διαφορές σε κοινούς σημασιολογικούς όρους ιεραρχιών (π.χ. διαφορετικοί κωδικοί που περιγράφουν τις ίδιες περιοχές), β) να μελετηθούν υπάρχουσες μέθοδοι διασύνδεσης όρων, γ) να αναπτυχθεί μεθοδολογία για την υλοποίηση αυτών των μεθόδων και δ) να επεκταθεί η μεθοδολογία ώστε να επιτυγχάνει ευθυγράμμιση μεταξύ των διαφορετικών επιπέδων ιεραρχιών μεταξύ λεξιλογίων.

ΣΧΕΤΙΚΟ ΥΛΙΚΟ:

1. M. Meimaris, G. Papastefanatos: Containment and Complementarity Relationships in Multidimensional Linked Open Data. In SemStats 2014 (in conjunction with ISWC'14), http://semstats2014.files.wordpress.com/2014/10/semstats2014_submission_6.pdf
2. <http://www.w3.org/TR/vocab-data-cube/>

ΕΠΕΞΕΡΓΑΣΙΑ ΚΑΙ ΔΗΜΟΣΙΕΥΣΗ ΕΞΕΛΙΣΣΟΜΕΝΩΝ ΑΝΟΙΚΤΩΝ ΒΙΟΛΟΓΙΚΩΝ
ΔΙΑΣΥΝΔΕΔΕΜΕΝΩΝ ΔΕΔΟΜΕΝΩΝ

ΠΛΗΡΟΦΟΡΙΕΣ: Θεωδωρή Δαλαμάγκας, 210 6875415, dalamag@imis.athena-innovation.gr

ΠΕΡΙΛΗΨΗ: Σκοπός της εργασίας είναι η μελέτη βιολογικών βάσεων δεδομένων και η επεξεργασία τους με σκοπό τη μετατροπή και δημοσίευσή τους σε ανοικτά διασυνδεδεμένα δεδομένα (linked data).

ΑΤΟΜΑ: 1-2

ΠΛΑΤΦΟΡΜΑ ΕΡΓΑΣΙΑΣ: Java/PHP/MySQL/D2R/Virtuoso

ΣΥΝΤΟΜΗ ΠΕΡΙΓΡΑΦΗ: Το **DIANA microT** και το **DIANA mirPath** [1] είναι, μεταξύ άλλων, δύο εφαρμογές Ιστού διεθνώς αναγνωρισμένες. Έχουν αναπτυχθεί από το ΠΣΥ σε συνεργασία με το Πανεπιστήμιο Θεσσαλίας και το Ερευνητικό Κέντρο Βιοϊατρικών Επιστημών "Αλέξανδρος Φλέμινγκ" για τη διαχείριση δεδομένων miRNA (βιομόρια πολύ μικρού μήκους που προσκολλώνται επάνω σε μόρια «αγγελιοφόρου» RNA-mRNA εμποδίζοντας την παραγωγή πρωτεϊνών).

Καθοριστικό ρόλο στη διάθεση βιολογικών δεδομένων στον Ιστό, αλλά και σύνδεσης μεταξύ τους, κατέχει το Linked Data Project (LDP) [2]. Στο LDP, σύνολα δεδομένων δημοσιεύονται και συνδέονται μεταξύ τους ακολουθώντας το μοντέλο RDF [3]. Τα δεδομένα αυτά είναι γνωστά και ως **διασυνδεδεμένα δεδομένα (linked data)**. Παράδειγμα ιστοτόπου με διασυνδεδεμένα δεδομένα βρίσκεται στο [4], όπου κανείς μπορεί να περιηγηθεί ή να διατυπώσει ερωτήσεις ειδικού τύπου (SPARQL queries) [5] σε δεδομένα miRNA.

Στις επιστημονικές βάσεις βιολογικών δεδομένων εμφανίζονται αλλαγές στις ιδιότητες και τα χαρακτηριστικά των βιομορίων. Οι αλλαγές αυτές καταγράφονται με συγκεκριμένο τρόπο σε αρχεία τα οποία και είναι διαθέσιμα στους ερευνητές σε τακτά χρονικά διαστήματα. *Παράδειγμα 1:* ένα βιομόριο μπορεί να αλλάξει ονομασία τη χρονική στιγμή T1, να αλλάξει δομή τη χρονική στιγμή T2 και τέλος να αντικατασταθεί από κάποιο άλλο τη χρονική στιγμή T3. *Παράδειγμα 2:* δύο βιομόρια μπορεί να θεωρηθεί ότι αντιπροσωπεύονται από ένα άλλο, οπότε και διαγράφονται τελείως τη χρονική στιγμή T1. Επομένως, κάθε βιομόριο έχει μια ιστορική εξέλιξη. Οι ιστορικές εξελίξεις όλων των μορίων δημιουργούν ένα γράφο ιστορικής εξέλιξης.

Η διπλωματική εργασία στοχεύει στην έκδοση ανοικτών (δηλ. ελεύθερη πρόσβαση σε όλους) διασυνδεδεμένων βιολογικών δεδομένων με τρόπο που να είναι εφικτή η περιήγηση στο γράφο ιστορικής εξέλιξης των βιομορίων miRNA. Η έκδοση θα γίνει με δύο εναλλακτικούς τρόπους, οι οποίοι και θα αποτιμηθούν ως προς την αποτελεσματικότητά τους: (α) virtual approach: θα χρησιμοποιηθούν εργαλεία που μετατρέπουν σχεσιακά δεδομένα σε RDF δεδομένα, χωρίς όμως τα τελευταία να έχουν «φυσική» αποθήκευση, και (β) materialized approach: θα χρησιμοποιηθούν εργαλεία που αποθηκεύουν αποκλειστικά RDF δεδομένα. Και στις δύο περιπτώσεις, σκοπός είναι η έκδοση χρονικά εξελισσόμενων RDF δεδομένων.

Η εργασία περιλαμβάνει τις παρακάτω δραστηριότητες:

1. Εξοικείωση με τις υπάρχουσες εφαρμογές.
2. Εξοικείωση με την πλατφόρμα έκδοσης διασυνδεδεμένων δεδομένων D2R [6] και την RDF βάση δεδομένων Virtuoso [7].
3. Σχεδιασμός μοντέλου ιστορικής εξέλιξης βιομορίων.

4. Έκδοση διασυνδεδεμένων δεδομένων στον Ιστό και με τις δύο στρατηγικές: virtual/materialized approach.
5. Αξιολόγηση στρατηγικών.

ΧΡΗΣΙΜΟ ΥΛΙΚΟ:

- [1] <http://diana.imis.athena-innovation.gr/>
- [2] <http://linkeddata.org>
- [3] <http://www.w3.org/RDF/>
- [4] <http://diwis.imis.athena-innovation.gr:8181>
- [5] <http://www.cambridgesemantics.com/semantic-university/sparql-by-example>
- [6] <http://d2rq.org/d2r-server>
- [7] <http://virtuoso.openlinksw.com>

**ΑΝΑΖΗΤΗΣΗ ΣΕ ΒΑΣΕΙΣ ΔΕΔΟΜΕΝΩΝ ΒΙΟΜΗΧΑΝΙΚΗΣ ΣΥΜΒΙΩΣΗΣ ΜΕ ΠΑΡΑΓΩΓΗ
ΥΠΟΨΗΦΙΩΝ ΕΡΩΤΗΣΕΩΝ**

ΠΛΗΡΟΦΟΡΙΕΣ: Θεωδωρή Δαλαμάγκας, 210 6875415, dalamag@imis.athena-innovation.gr

ΠΕΡΙΛΗΨΗ: Σκοπός της εργασίας είναι ο σχεδιασμός και η ανάπτυξη ενός συστήματος αναζήτησης σε βάσεις δεδομένων βιομηχανικής συμβίωσης, όπου η αρχική αναζήτηση με λέξεις κλειδιά θα οδηγεί στην παραγωγή συνόλου υποψηφίων δομημένων ερωτήσεων που θα βοηθήσουν το χρήστη να προχωρήσει στην αναζήτησή του αποδοτικότερα.

ΑΤΟΜΑ: 1-2

ΠΛΑΤΦΟΡΜΑ ΕΡΓΑΣΙΑΣ: Java/PHP/MySQL/Lucene/D2R

ΣΥΝΤΟΜΗ ΠΕΡΙΓΡΑΦΗ: Η βιομηχανική συμβίωση (industrial symbiosis) [1] αποσκοπεί στη συστηματική αύξηση της αποδοτικότητας της ενέργειας και των υλικών που χρησιμοποιούνται στις βιομηχανικές διεργασίες με τη δημιουργία ρευμάτων σύνδεσης μεταξύ των διαφορετικών βιομηχανικών μονάδων. Αυτό επιτυγχάνεται μέσω της συνεργασίας μεταξύ διαφορετικών βιομηχανιών για τις οποίες η ύπαρξη της μιας αυξάνει τη βιωσιμότητα των συνεργαζόμενων υπολοίπων και στην οποία η απαίτηση για εξοικονόμηση πρώτων υλών, ενέργειας και προστασία του περιβάλλοντος είναι ζωτικής σημασίας. Τα **συμβιωτικά δίκτυα** είναι η πρακτική μέσω της οποίας οι εταιρείες συγκεντρώνονται προκειμένου να ανταλλάξουν υλικά, ενέργεια, νερό και υποπροϊόντα. Με τον τρόπο αυτό, όχι μόνο μειώνεται το κόστος παραγωγής αλλά προωθείται παράλληλα και η προστασία του περιβάλλοντος. Κλειδί για τη λειτουργία ενός συμβιωτικού δικτύου αποτελούν οι δυνατότητες συνεργειών οι οποίες προσφέρονται βάσει της γεωγραφικής εγγύτητας. Κύριοι τύποι συνεργειών σε συμβιωτικά δίκτυα είναι: (α) συνέργειες σε μια ενιαία αλυσίδα εφοδιασμού, (β) συνέργειες από την κοινή χρήση των υπηρεσιών κοινής ωφελείας, και (γ) συνέργειες από την τοπική χρήση των υποπροϊόντων (ενέργεια ή/και απόβλητα).

Το ΠΣΥ σε συνεργασία με το Εθνικό Μετσόβιο Πολυτεχνείο και άλλους εταίρους συμμετέχει στο ερευνητικό έργο AI4B [2], στόχος του οποίου είναι να αναπτύξει καινοτόμες υποδομές πληροφορικής για τη βελτιστοποίηση αλυσίδων εφοδιασμού βιομάζας και την ανάπτυξη συμβιωτικών δικτύων στην Περιφέρεια Θεσσαλίας. Στα πλαίσια του έργου υπάρχει η δυνατότητα εκπόνησης διπλωματικών εργασιών σε διάφορα θέματα επεξεργασίας, ανάλυσης, αναζήτησης και οπτικοποίησης περιεχομένου βάσεων δεδομένων βιομηχανικής συμβίωσης (απευθυνθείτε στις ΠΛΗΡΟΦΟΡΙΕΣ παραπάνω).

Η παρούσα διπλωματική εργασία στοχεύει στο σχεδιασμό και στην ανάπτυξη ενός συστήματος αναζήτησης σε βάση δεδομένων συμβιωτικού δικτύου βιομάζας ΒΔΒ. Η βάση δεδομένων έχει ήδη σχεδιαστεί και αναπτυχθεί [3] χρησιμοποιώντας την τεχνολογία των διασυνδεδεμένων δεδομένων (linked data). Η τεχνολογία αυτή περιγράφεται στο Linked Data Project (LDP) [4]. Στο LDP, σύνολα δεδομένων δημοσιεύονται και συνδέονται μεταξύ τους ακολουθώντας το μοντέλο RDF [5], που έχει τη μορφή γράφου. Στα δεδομένα RDF γράφων μπορεί να διατυπώνει κανείς ερωτήσεις ειδικού τύπου (SPARQL queries) [6], κατ'αναλογία με τη γλώσσα ερωτήσεων SQL σε σχεσιακές βάσεις δεδομένων.

Η αναζήτηση στην RDF βάση δεδομένων ΒΔΒ θα γίνεται αρχικά με λέξεις κλειδιά, αλλά στη συνέχεια ο χρήστης θα καθοδηγείται στην αναζήτησή του από την ημι-αυτόματη παραγωγή υποψηφίων ερωτήσεων, τις οποίες και θα μπορεί να επιλέγει ώστε να προχωρήσει η αναζήτησή του στην κατεύθυνση που επιθυμεί. Οι

ερωτήσεις αυτές θα είναι σε μορφή SPARQL, και θα παρουσιάζονται στο χρήστη με σειρά προτεραιότητας: ερωτήσεις που έχουν υψηλή θέση κατάταξης θα έχουν περισσότερη πιθανότητα να ικανοποιούν την πραγματική ανάγκη αναζήτησης του χρήστη. Η προτεραιότητα θα υπολογίζεται με βάση ένα μοντέλο κόστους που θα εφαρμόζεται κατά τη διάσχιση του RDF γράφου για την παραγωγή των υποψηφίων ερωτήσεων [7].

Η εργασία περιλαμβάνει τις παρακάτω δραστηριότητες:

1. Εξοικείωση με την υπάρχουσα υποδομή στο έργο AI4B.
2. Εξοικείωση με την πλατφόρμα έκδοσης διασυνδεδεμένων δεδομένων D2R [8].
3. Σχεδιασμός αλγορίθμου παραγωγής SPARQL ερωτήσεων από λέξεις κλειδιά.
4. Υλοποίηση συστήματος αναζήτησης και οπτικοποίησης αποτελεσμάτων.
5. Αξιολόγηση αλγορίθμων (αποδοτικότητα και ποιότητα αποτελεσμάτων).

ΧΡΗΣΙΜΟ ΥΛΙΚΟ:

[1] http://en.wikipedia.org/wiki/Industrial_symbiosis

[2] <http://www.ai4b.gr>

[3] <http://snf-533531.vm.oceanos.grnet.gr:2020>

[4] <http://linkeddata.org>

[5] <http://www.w3.org/RDF/>

[6] <http://www.cambridgesemantics.com/semantic-university/sparql-by-example>

[7] Top-k Exploration of Query Candidates for Efficient Keyword Search on Graph-Shaped (RDF) Data (<http://goo.gl/jKOzlw>)

[8] <http://d2rq.org/d2r-server>

ΕΝΣΩΜΑΤΩΣΗ ΑΝΟΙΚΤΩΝ ΓΕΩΓΡΑΦΙΚΩΝ ΔΕΔΟΜΕΝΩΝ ΣΤΟΝ ΣΗΜΑΣΙΟΛΟΓΙΚΟ ΙΣΤΟ

ΠΛΗΡΟΦΟΡΙΕΣ: Κώστας Πατρούμπας, 210 772 1446, kpatro@dblab.ece.ntua.gr και Σπύρος Αθανασίου, spathan@imis.athena-innovation.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 ΠΣΥ 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/>