ΓΕΝΙΚΗ ΓΡΑΜΜΑΤΕΙΑ ΕΡΕΥΝΑΣ & ΤΕΧΝΟΛΟΓΙΑΣ
ΕΡΕΥΝΗΤΙΚΟ ΚΕΝΤΡΟ «Αθηνά»

«Ι.Π.ΣΥ.Π.»
ΙΝΣΤΙΤΟΥΤΟ ΠΛΗΡΟΦΟΡΙΑΚΩΝ ΣΥΣΤΗΜΑΤΩΝ & ΠΡΟΣΟΜΟΙΩΣΗΣ

Διπλωματικές Εργασίες Ακαδ. Έτους 2009-2010

Περιεχόμενα

ΕΙΣΑΓΩΓΗ **2**

ΕΚΠΑΙΔΕΥΤΙΚΗ ΠΛΑΤΦΟΡΜΑ ΙΣΤΟΡΙΑΣ ΣΤΗΡΙΖΟΜΕΝΗ ΣΤΗ ΣΥΝΕΙΣΦΟΡΑ ΕΚΠΑΙΔΕΥΤΙΚΗΣ ΚΟΙΝΟΤΗΤΑΣ **3**

LIVE GEOBLOGGING **5**

GEOCROWD EXPLORER **7**

ΕΝΙΑΙΟ ΠΛΑΙΣΙΟ ΓΕΩΚΩΔΙΚΟΠΟΙΗΣΗΣ **9**

ΑΞΙΟΛΟΓΗΣΗ ΑΛΓΟΡΙΘΜΩΝ ΣΥΝΤΟΜΟΤΕΡΗΣ ΔΙΑΔΡΟΜΗΣ **11**

ΜΕΤΑΔΟΣΗ ΚΑΙ ΕΚΤΙΜΗΣΗ ΘΕΣΗΣ ΚΙΝΗΤΩΝ ΣΥΣΚΕΥΩΝ **12**

ΚΑΤΑΝΟΜΗ ΚΑΙ RANKING ΜΕ ΒΑΣΗ ΧΩΡΙΚΕΣ ΣΥΝΤΕΤΑΓΜΕΝΕΣ **14**

ΜΕΤΑΤΡΟΠΗ EAD ΜΕΤΑΔΕΔΟΜΕΝΩΝ ΣΕ ΜΟΡΦΗ ΟΝΤΟΛΟΓΙΑΣ CIDOC CRM **15**

ΔΙΑΧΕΙΡΙΣΗ ΣΥΣΧΕΤΙΖΟΜΕΝΩΝ ΑΛΛΑΓΩΝ ΣΤΟΝ ΙΣΤΟ **17**

"PAPER-MAP" **18**

ΥΠΗΡΕΣΙΕΣ ΑΝΑΖΗΤΗΣΗΣ ΣΤΟΝ ΙΣΤΟ ΜΕ ΧΡΗΣΗ ΕΡΓΑΛΕΙΩΝ ΔΙΑΧΕΙΡΙΣΗΣ ΣΚΕΨΗΣ **20**

ΕΞΑΤΟΜΙΚΕΥΜΕΝΗ ΑΝΑΖΗΤΗΣΗ ΜΕ ΧΡΗΣΗ ΕΡΓΑΛΕΙΩΝ ΔΙΑΧΕΙΡΙΣΗΣ ΣΚΕΨΗΣ **22**

ΠΡΟΗΓΜΕΝΗ ΕΦΑΡΜΟΓΗ ΙΣΤΟΥ ΔΙΑΧΕΙΡΙΣΗΣ ΔΕΔΟΜΕΝΩΝ ΒΙΟΕΠΙΣΤΗΜΩΝ **24**

ΕΥΦΥΕΣ ΣΥΣΤΗΜΑ ΔΙΑΧΕΙΡΙΣΗΣ ΡΟΩΝ ΔΕΔΟΜΕΝΩΝ ATOM/ RSS **26**

ΜΕΛΕΤΗ ΚΑΙ ΥΛΟΠΟΙΗΣΗ ΑΛΓΟΡΙΘΜΩΝ ΓΙΑ ΒΙΟΛΟΓΙΚΕΣ ΕΦΑΡΜΟΓΕΣ ΣΕ MAPREDUCE ΠΕΡΙΒΑΛΛΟΝ **28**

ΟΠΤΙΚΟΠΟΙΗΣΗ ΕΞΑΡΤΗΣΕΩΝ ΣΕ ΣΥΣΤΗΜΑΤΑ ΒΑΣΕΩΝ ΔΕΔΟΜΕΝΩΝ **30**

ΕΞΑΓΩΓΗ SQL ΕΡΩΤΗΜΑΤΩΝ ΑΠΟ ΑΡΧΕΙΑ ΠΗΓΑΙΟΥ ΚΩΔΙΚΑ **31**

ΣΧΕΔΙΑΣΜΟΣ ΚΑΙ ΑΝΑΠΤΥΞΗ ΕΝΟΣ TASK-BASED WIKI **32**

ΑΝΑΛΥΣΗ ΚΑΙ ΣΧΕΔΙΑΣΗ ΕΝΟΣ ΔΙΑΔΙΚΤΥΑΚΟΥ ΣΥΝΤΑΚΤΗ ΟΝΤΟΛΟΓΙΩΝ **34**

ΕΙΣΑΓΩΓΗ

Το **Ινστιτούτο Πληροφοριακών Συστημάτων και Προσομοίωσης** (ΙΠΣΥΠ - http://www.ipsyp.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 ΙΠΣΥΠ διεξάγει έρευνα και συμμετέχει σε αναπτυξιακά έργα στον τομέα της διαχείρισης και επεξεργασίας δεδομένων σε πληροφοριακά συστήματα μεγάλης κλίμακας. Οι περιοχές ερευνητικής δραστηριότητας του ΙΠΣΥΠ περιλαμβάνουν:

- Συστήματα διαχείρισης γεω-πληροφορίας και υπηρεσίες εντοπισμού κινούμενων αντικειμένων.
- 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 ΙΠΣΥΠ ανακοινώνει μια σειρά από διπλωματικές εργασίες. Οι εργασίες θα εκπονηθούν στο ΙΠΣΥΠ σε συνεργασία με το **Εργαστήριο Συστημάτων Βάσεων Γνώσεων και Δεδομένων της Σχολής Ηλεκτρολόγων Μηχ. και Μηχ. Υπολογιστών Ε.Μ.Π** (ΕΣΒΓΔ - http://www.dblab.ece.ntua.gr).

ΕΚΠΑΙΔΕΥΤΙΚΗ ΠΛΑΤΦΟΡΜΑ ΙΣΤΟΡΙΑΣ ΣΤΗΡΙΖΟΜΕΝΗ ΣΤΗ ΣΥΝΕΙΣΦΟΡΑ ΕΚΠΑΙΔΕΥΤΙΚΗΣ ΚΟΙΝΟΤΗΤΑΣ

ΠΛΗΡΟΦΟΡΙΕΣ: Dieter Pfoser, 210 6990522 (204), pfoser@imis.athena-innovation.gr

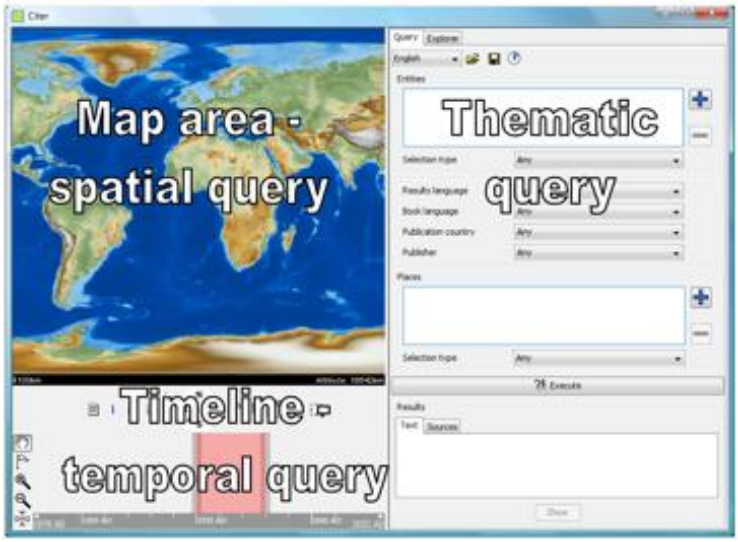
ΠΕΡΙΛΗΨΗ: Δημιουργία μιας εκπαιδευτικής πλατφόρμας για τη διδασκαλία της ιστορίας. Πρόσφατα το έργο CITER, που είχε ως κύριο στόχο την ενσωμάτωση πολυγλωσσικού ιστορικού περιεχόμενου από διάφορες χώρες, δημιούργησε μια βάση δεδομένων ιστορικού περιεχομένου με δυνατότητα αναζήτησης, που στηρίζεται σε χωροχρονικά και θεματικά μεταδεδομένα, καθώς και δυνατότητες ανάκτησης περιεχομένου, χρησιμοποιώντας εργαλεία αυτόματης μετάφρασης. Στόχος της παρούσης εργασίας, είναι η δημιουργία μια βάσης δεδομένων με ιστορικό περιεχόμενο, με υλικό που θα προέρχεται από τη συνεισφορά χρηστών και θα προορίζεται για χρήση μέσω learning objects από τη πλατφόρμα Moodle, με κύριο ζητούμενο την επέκταση και εμπλουτισμό της υπάρχουσας βάσης δεδομένων του έργου CITER και τη δημιουργία μιας εκπαιδευτικής πλατφόρμας για την ιστορία που θα επιτρέπει την ανταλλαγή μαθησιακού περιεχομένου που δημιουργείται από καθηγητές ή μαθητές.

ΑΤΟΜΑ: 1

ΠΛΑΤΦΟΡΜΑ ΕΡΓΑΣΙΑΣ: Javascript, AJAX, Ruby-on-Rails, PostgreSQL.

ΣΥΝΤΟΜΗ ΠΕΡΙΓΡΑΦΗ: Κύριος στόχος η δημιουργία μιας online διασύνδεσης που:

- Συνεργάζεται με την υπάρχουσα ιστορική βάση δεδομένων του έργου CITER
- Επιτρέπει την προσθήκη νέου περιεχομένου, είτε ως μεταδεδομένα, είτε ως νέο περιεχόμενο και εκπαιδευτικά αντικείμενα



Διεπαφή ερωτήσεων

Ειδικότερα, στην διπλωματική εργασία προβλέπεται να καλυφθούν τα εξής ζητήματα:

- Σχεδιασμός και ανάπτυξη μιας απλής online εφαρμογής που θα παρέχει τις δυνατότητες αναζήτησης, όπως φαίνεται στο σχήμα 1, χρησιμοποιώντας χάρτες (Google maps mashup), χρονολογίες (timeline) και τα υπάρχοντα θεματικά μεταδεδομένα (ταξινομημένα ήδη σε κατηγορίες)

- Σχεδιασμός και ανάπτυξη ενός μοντέλου βάσης δεδομένων που θα επιτρέπει την ενσωμάτωση νέου περιεχομένου στο ήδη υπάρχον περιεχόμενο.
- Σχεδιασμός και ανάπτυξη μιας απλής online εφαρμογής που θα επιτρέπει την προσθήκη και παρουσίαση του νέου περιεχομένου, σε σχέση με το ήδη υπάρχον περιεχόμενο. Αυτό θα συμπεριλαμβάνει από απλές σημειώσεις μέχρι εκπαιδευτικά αντικείμενα. Συνεπώς, η εφαρμογή θα δανείζεται στοιχεία από τα Wikis αλλά κυρίως από το Moodle (εκπαιδευτική πλατφόρμα ανοιχτού λογισμικού)

ΣΧΕΤΙΚΟ ΥΛΙΚΟ:

- CITER Web page: http://citer.cti.gr
- Moodle - Learning Management System: http://moodle.org/
- Timeline component: http://www.simile-widgets.org/timeline/
- Google Maps API: http://code.google.com/apis/maps/

LIVE GEOBLOGGING

ΠΛΗΡΟΦΟΡΙΕΣ: Dieter Pfoser, 210 6990522 (204), pfoser@imis.athena-innovation.gr

ΠΕΡΙΛΗΨΗ: Ο όρος Geoblogging χρησιμοποιείται για να περιγράψει μια γεωγραφικά ενήμερη εκδοχή ενός ιστολογίου που επιτρέπει την προσθήκη περιεχομένου που έχει γεωγραφική διάσταση και εμπεριέχει γεωγραφική πληροφορία. Η εφαρμογή GEOCROWD (www.geocrowd.org) είναι ένα γεω-ιστολόγιο που αναπτύχθηκε από το ΙΠΣΥΠ και χρησιμοποιεί τη γεωγραφική πληροφορία ως μέσο χωροχρονικής περιγραφής, είτε μιλάμε για ένα ταξίδι, μια απογευματινή βόλτα στη γειτονιά, την αναζήτηση μιας καλής καφετέριας, ένα ταξίδι με το ποδήλατο ή απλά μια πεζοπορία. Στόχος ήταν η δημιουργία μια απλής πλατφόρμας, που θα επιτρέπει στον χρήστη να διηγηθεί την ιστορία ενός ταξιδιού, χρησιμοποιώντας το περιεχόμενο που συνέλεξε στη διάρκειά του.
Το περιεχόμενο της παρούσης εργασίας είναι να επεκτείνουμε την ήδη υπάρχουσα στατική, Web εφαρμογή και σε κινητές συσκευές, όπως το iPhone ή το Android. Κύριος στόχος η εξερεύνηση των κατάλληλων δυνατοτήτων που παρέχουν στον χρήστη οι συγκεκριμένες συσκευές και η δημιουργία μιας πρότυπης εφαρμογής.

ΑΤΟΜΑ: 1-2

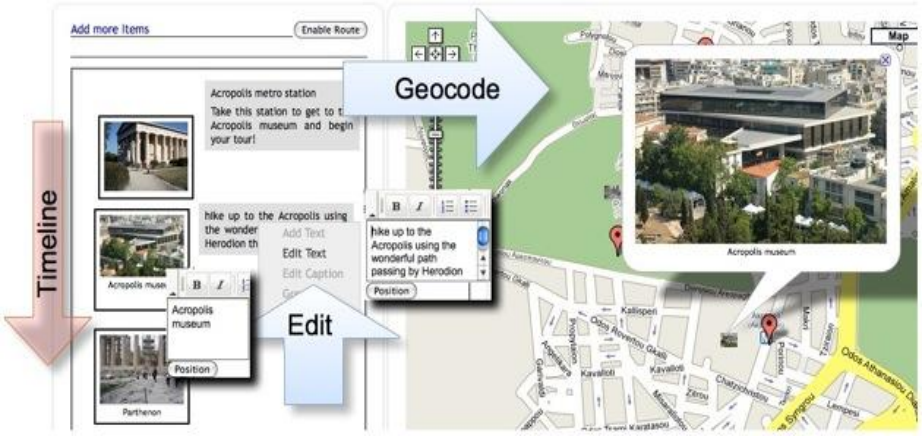
ΠΛΑΤΦΟΡΜΑ ΕΡΓΑΣΙΑΣ: [C#, iPhone API]/[Java, Android API].

ΣΥΝΤΟΜΗ ΠΕΡΙΓΡΑΦΗ: Αντικείμενο της διπλωματικής είναι η ανάπτυξη μια πλατφόρμας πελάτη που θα παρέχει:

(i) Μια απλή και κομψή διασύνδεση χρήστη που θα επιτρέπει στους χρήστες να αποθηκεύουν κείμενο και πολυμεσικό υλικό που σχετίζεται με τη θέση τους, συνδέοντας έτσι τοποθεσίες του πραγματικού κόσμου με κείμενο, εικόνες, video και ήχους.

(ii) Τη δυνατότητα μεταφόρτωσης δεδομένων (upload) στην Web υπηρεσία GEOCROWD.

Η δημιουργία ενός Γεω-ιστολογίου πρέπει να είναι το ίδιο εύκολη με τη σύνταξη ενός email!



Ειδικότερα, στη διπλωματική εργασία προβλέπεται να καλυφθούν τα εξής ζητήματα:

- Σχεδιασμός και ανάπτυξη μια πλατφόρμας – πελάτη για το Android API (Google mobile phone api)
- Η εφαρμογή θα πρέπει να περιέχει τις παρακάτω λειτουργικές υπομονάδες:
    - Διασύνδεση με τις περιφερειακές μονάδες του τηλεφώνου, όπως η GPS συσκευή (για την ανίχνευση της θέσης), η ενσωματωμένη κάμερα (για προσθήκη video και εικόνας) και του μικροφώνου (για αποθήκευση ήχων).

  - Απλό συντάκτη κειμένου (text editor) για την προσθήκη περιγραφικών σημειώσεων
- Δυνατότητα για τη μεταφόρτωση δεδομένων (upload) σε κάποια Web υπηρεσία
  - GPS δεδομένα σε διάφορες μορφές (gpx)
  - Πολυμεσικό υλικό (εικόνες, video, ήχοι και σημειώσεις κειμένου)
- Σχεδιασμός και υλοποίηση ενός συστήματος αρχείων ή τοπικής βάσης δεδομένων στη φορητή συσκευή για την αποθήκευση των τοποθεσιών και του σχετιζόμενου πολυμεσικού υλικού.

ΣΧΕΤΙΚΟ ΥΛΙΚΟ:

- GEOCROWD geoblogging platform http://www.geocrowd.org
- Android API: http://code.google.com/android/
- iPhone API: http://developer.apple.com/iphone/
- Collaborative mapping efforts e.g., http://www.openstreetmap.org/, http://www.openrouteservice.org/, http://platial.com/.

GEOCROWD EXPLORER

ΠΛΗΡΟΦΟΡΙΕΣ: Dieter Pfoser, 210 6990522 (204), pfoser@imis.athena-innovation.gr

ΠΕΡΙΛΗΨΗ: Στόχος της παρούσης εργασίας είναι η δημιουργία μιας «εκτεταμένης πραγματικότητας» - *augmented reality (AR)* εφαρμογής που επιτρέπει την κατανάλωση γεω-ιστολογικού περιεχομένου σε κινητές συσκευές.
Ο όρος Geoblogging χρησιμοποιείται για να περιγράψει μια γεωγραφικά ενήμερη εκδοχή ενός ιστολογίου, που επιτρέπει την προσθήκη περιεχομένου που έχει γεωγραφική διάσταση και εμπεριέχει γεωγραφική πληροφορία. Η εφαρμογή GEOCROWD (www.geocrowd.org) είναι ένα γεω-ιστολόγιο που αναπτύχθηκε από το ΙΠΣΥΠ και χρησιμοποιεί την γεωγραφική πληροφορία ως μέσο χωροχρονικής περιγραφής, είτε μιλάμε για ένα ταξίδι, μια απογευματινή βόλτα στη γειτονιά, την αναζήτηση μιας καλής καφετέριας, ένα ταξίδι με το ποδήλατο ή απλά μια πεζοπορία.
Στόχος ήταν η δημιουργία μια απλής πλατφόρμας, που θα επιτρέπει στον χρήστη να διηγηθεί την ιστορία ενός ταξιδιού, χρησιμοποιώντας το περιεχόμενο που συνέλεξε στη διάρκειά του. Χρησιμοποιώντας «εκτεταμένη πραγματικότητα» - *augmented reality*, παρέχεται η δυνατότητα στον χρήστη να ξαναζήσει τα ταξίδια αυτά, παρέχοντας έτσι μια κοινή πλατφόρμα ανταλλαγής ταξιδιωτικών εμπειριών. Μια τέτοια εφαρμογή μπορεί να αποτελέσει πρότυπο για τους μελλοντικούς ταξιδιωτικούς οδηγούς.

ΑΤΟΜΑ: 1-2

ΠΛΑΤΦΟΡΜΑ ΕΡΓΑΣΙΑΣ: [C#, iPhone API]/[Java, Android API].

ΣΥΝΤΟΜΗ ΠΕΡΙΓΡΑΦΗ: Ο όρος Augmented Reality (AR) χρησιμοποιείται για την περιγραφή μιας ζωντανής (άμεσης ή έμμεσης) άποψης του φυσικού κόσμου, τα στοιχεία του οποίου εμπλουτίζονται από ψηφιακές εικόνες που παράγονται από υπολογιστή, δημιουργώντας έτσι μια μικτή πραγματικότητα.
Στόχος της παρούσης εργασίας είναι η ανάπτυξη ενός AR λογισμικού για κινητές συσκευές (iPhone ή Android) που θα αλληλεπιδρά με την online υπηρεσία GEOCROWD. Συγκεκριμένα:

- Πρέπει να μελετηθούν οι υπάρχουσες AR εφαρμογές και τα αντίστοιχα παραδείγματα αλληλεπίδρασής τους με το χρήστη, προκείμενου να εντοπίσουμε τα πλεονεκτήματα και τις αδυναμίες τους. Χαρακτηριστικές εφαρμογές: LAYAR (http://layar.com/) ή Wikitude (http://www.wikitude.org/)
- Να εντοπίσουμε υπάρχοντα frameworks που θα μπορούσαν να χρησιμοποιηθούν για τους στόχους μας.

Παράδειγμα AR (Wikitude)

- Να σχεδιαστεί και να αναπτυχθεί ένας AR πελάτης, είτε για iPhone, είτε για το Android)
- Να σχεδιαστεί και να αναπτυχθεί μια διασύνδεση για την online υπηρεσία GEOCROWD, προκειμένου να έχουμε πρόσβαση στα δεδομένα.

ΣΧΕΤΙΚΟ ΥΛΙΚΟ:

- GEOCROWD geoblogging platform http://www.geocrowd.org
- Android API: http://code.google.com/android/, iPhone API: http://developer.apple.com/iphone/
- LAYAR, http://www.layar.com, Wikitude http://www.wikitude.org

ΕΝΙΑΙΟ ΠΛΑΙΣΙΟ ΓΕΩΚΩΔΙΚΟΠΟΙΗΣΗΣ

ΠΛΗΡΟΦΟΡΙΕΣ: Dieter Pfoser, 210 6990522 (204), pfoser@imis.athena-innovation.gr

ΠΕΡΙΛΗΨΗ: Πολλοί διαφορετικοί τύποι μεταδεδομένων μπορούν να χρησιμοποιηθούν για την οργάνωση, ευρετηριοποίηση, αναζήτηση και πλοήγηση σε ιστοσελίδες, όπως λέξεις-κλειδιά, θεματικές κατηγορίες, γεωγραφική θέση, χρόνος. Ένας σημαντικός, καθότι μονοσήμαντος, τέτοιος τύπος είναι η γεωγραφική θέση - ένα ζεύγος συντεταγμένων έχει μια και μόνη ερμηνεία.
Γεωκωδικοποίηση (geocoding) είναι η διαδικασία αναγνώρισης γεωγραφικής πληροφορίας σε κείμενο και η σύζευξη της γεωγραφικής αυτής πληροφορίας με γεωγραφικές συντεταγμένες. Ένα απλό παράδειγμα είναι η μετατροπή μια διεύθυνσης σε γεωγραφικές συντεταγμένες, ενώ πιο πολύπλοκο παράδειγμα αποτελεί η γεωκωδικοποίηση όρων, όπως «Πλάκα», «Αθήνα» κοκ.
Το αντικείμενο της παρούσης εργασίας είναι να αναπτυχθεί ένα πλαίσιο γεωκωδικοποίησης (geocoding framework) γενικού ενδιαφέροντος που θα μπορεί να χρησιμοποιηθεί για πληθώρα χωρών και διαλέκτων. Η εργασία αυτή στηρίζεται σε προγενέστερη εργασία, στην οποία γινόταν γεωκωδικοποίηση Ελληνικών κειμένων, καθώς και στη χρήση διαφόρων υπαρχόντων πλαισίων γεωκωδικοποίησης ανοιχτού λογισμικού.

ΑΤΟΜΑ: 1-2

ΠΛΑΤΦΟΡΜΑ ΕΡΓΑΣΙΑΣ: Java, Javascript, PostgreSQL

ΣΥΝΤΟΜΗ ΠΕΡΙΓΡΑΦΗ: Αντικείμενο της εργασίας είναι η δημιουργία ενός ενιαίου πλαισίου γεωκωδικοποίησης, κατάλληλο για οποιαδήποτε περίπτωση, είτε σε επίπεδο γλωσσών, είτε σε επίπεδο γεωγραφικής πληροφορίας (διεύθυνσης, τηλεφωνικοί αριθμοί, τοπωνύμια), είτε σε επίπεδο κειμένου (online ή όχι).
Ο γεωκωδικοποιητής θα πρέπει να παραμετροποιείται για οποιαδήποτε γλώσσα, ορίζοντας γραμματικές που απεικονίζουν τις ιδιαιτερότητες της κάθε γλώσσας. Παράλληλα, για να είναι αποτελεσματικός, σε κάθε περίπτωση, (συνδυασμός γλώσσας και περιεχομένου) θα πρέπει να υπάρχει επαρκής διαθέσιμη γεωγραφική πληροφορία, με μορφή πινάκων συντεταγμένων, τηλεφωνικών καταλόγων και λεξικό γεωγραφικών όρων (Geogazetteers). Η δημιουργία τέτοιων πόρων δεν αποτελεί στόχο της διπλωματικής. Για την περίπτωση της Ελληνικής γλώσσας άλλωστε, υπάρχουν τέτοιες πηγές. Επίσης, σημαντικός παράγοντας στην αξιολόγηση ενός γεωκωδικοποιητή αποτελεί η ακρίβεια των αποτελεσμάτων. Επομένως, θα πρέπει να επιτρέπεται η κατά προσέγγιση γεωκωδικοποίηση, αφού είτε το κείμενο δεν παρέχει αρκετή πληροφορία, είτε οι διαθέσιμοι πόροι δεν είναι πάντα επαρκείς. Συνεπώς, ενώ θα πρέπει το εργαλείο να εντοπίζει στο κείμενο όσο το δυνατόν περισσότερη γεωγραφική πληροφορία, στα αποτελέσματά του θα πρέπει να παρέχεται πληροφορία σχετικά με την πιθανότητα λάθους του (μέτρηση αβεβαιότητας).

Τα συγκεκριμένα καθήκοντα της παρούσης εργασίας είναι:

- Η αξιολόγηση των υπαρχόντων πλαισίων γεωκωδικοποίησης – είτε μιλάμε για ανοιχτό λογισμικό είτε για προϊόντα προηγούμενων εργασιών.
- Σχεδιασμός και ανάπτυξη ενός πλαισίου γεωκωδικοποίησης ανοιχτού λογισμικού στο οποίο θα ποσοστικοποιείται η πιθανότητα αβεβαιότητας των αποτελεσμάτων
- Ανάπτυξη ενδεικτικών γραμματικών για Αγγλικά και Ελληνικά.
- Αξιοποίηση των διαθέσιμων πηγών γεωγραφικής πληροφορίας

- Αξιολόγηση και σύγκριση με τα υπάρχοντα πλαίσια

ΣΧΕΤΙΚΟ ΥΛΙΚΟ:

- JGeocoder - http://jgeocoder.sourceforge.net/
- Geoparsing web pages, A. Angel, Diploma Thesis - http://www.dbnet.ece.ntua.gr/pubs/details.php?id=1427
- Google maps geocoding - http://code.google.com/apis/maps/documentation/services.html#Geocoding

ΑΞΙΟΛΟΓΗΣΗ ΑΛΓΟΡΙΘΜΩΝ ΣΥΝΤΟΜΟΤΕΡΗΣ ΔΙΑΔΡΟΜΗΣ

ΠΛΗΡΟΦΟΡΙΕΣ: Dieter Pfoser, 210 6990522 (204), pfoser@imis.athena-innovation.gr

ΠΕΡΙΛΗΨΗ: Ο υπολογισμός της συντομότερης διαδρομής, αποτελεί τον πιο σημαντικό αλγόριθμο σε μια υπηρεσία δρομολόγησης και πλοήγησης. Παρόλο που πολλοί αλγόριθμοι έχουν προταθεί, δεν υπάρχει καθαρή εικόνα για την απόδοσή τους.
Στόχος της παρούσης εργασίας είναι: (α) Η ανάπτυξη μια ενιαίας πλατφόρμας που θα μπορεί να χρησιμοποιηθεί για την αξιολόγηση αλγόριθμων συντομότερης διαδρομής (β) Να χρησιμοποιηθεί αυτή η πλατφόρμα ως μέσο αξιολόγησης αρκετών διαθέσιμων αλγόριθμων (γεγονός που προϋποθέτει τη μεταφορά τους στην συγκεκριμένη πλατφόρμα).

ΑΤΟΜΑ: 1-2

ΠΛΑΤΦΟΡΜΑ ΕΡΓΑΣΙΑΣ: Java, PostgreSQL.

ΣΥΝΤΟΜΗ ΠΕΡΙΓΡΑΦΗ: Στόχος της παρούσης εργασίας είναι η σύγκριση αλγορίθμων συντομότερης διαδρομής (κάποιοι από τους οποίους αναπτύχθηκαν από το ΙΠΣΥΠ) σε επίπεδο α) χρόνου εκτέλεσης β) ποιότητα αποτελεσμάτων, δηλαδή εάν όντως δίνουν τη βέλτιστη συντομότερη διαδρομή. Σημαντικό κομμάτι της εργασίας αποτελεί η δημιουργία ενός κανονικοποιημένου οδικού δικτύου από τα δεδομένα του Open Street Maps για την Ευρώπη και η κατάλληλη προσαρμογή του προκειμένου να χρησιμοποιηθεί σε αλγόριθμους δρομολόγησης.

Τα καθήκοντα της εργασίας είναι τα παρακάτω:

- Ορισμός των μετρήσιμων χαρακτηριστικών της απόδοσης των αλγόριθμων, δηλαδή ποια δεδομένα χαρακτηρίζουν αποτελεσματικά την απόδοση ενός αλγόριθμου συντομότερης διαδρομής
- Σχεδιασμός και ανάλυση ενός συστήματος μέτρησης απόδοσης (benchmark system), που θα περιλαμβάνει και τον αλγόριθμο Dijkstra συντομότερης διαδρομής, ο οποίος θα χρησιμοποιηθεί ως σημείο αναφοράς για την απόδοση των υπόλοιπων αλγόριθμων.
- Η δημιουργία ενός οδικού δικτύου που θα χρησιμοποιηθεί για την αξιολόγηση και θα στηρίζεται στο open street map
- Μεταφορά και ενσωμάτωση του A* αλγόριθμου, καθώς και άλλων αλγόριθμων συντομότερης διαδρομής (που θα επιλεγούν) στη δική μας πλατφόρμα
- Μέτρηση της απόδοσης (Benchmarking) των αλγόριθμων

ΣΧΕΤΙΚΟ ΥΛΙΚΟ:

- Benchmarking - http://en.wikipedia.org/wiki/Benchmarking
- Shortest-path computation and Dijkstra algorithm - http://en.wikipedia.org/wiki/Dijkstra%27s_algorithm
- Openstreetmap - http://www.openstreetmap.org/

ΜΕΤΑΔΟΣΗ ΚΑΙ ΕΚΤΙΜΗΣΗ ΘΕΣΗΣ ΚΙΝΗΤΩΝ ΣΥΣΚΕΥΩΝ

ΠΛΗΡΟΦΟΡΙΕΣ: Γιάννης Σταύρακας, 210 6990522 (103), yannis@imis.athena-innovation.gr

ΠΕΡΙΛΗΨΗ: Στόχος της διπλωματικής είναι να μελετηθούν εναλλακτικές μέθοδοι μετάδοσης και εκτίμησης θέσης και προσανατολισμού για φορητές συσκευές, που θα εκμεταλλεύονται την προθυμία κοντινών συσκευών να μοιραστούν τη δική τους θέση και προσανατολισμό.

ΑΤΟΜΑ: 1.

ΠΛΑΤΦΟΡΜΑ ΕΡΓΑΣΙΑΣ: Java για mobiles.

ΣΥΝΤΟΜΗ ΠΕΡΙΓΡΑΦΗ: Υπάρχουν περιπτώσεις όπου η τεχνολογία GPS είναι ανεπαρκής (π.χ. φαράγγια ψηλών κτιρίων - "urban canyons", ανεύρεση θέσης σε εσωτερικούς χώρους), ή και μη διαθέσιμη (π.χ. χρήστες με συμβατικές συσκευές). Έτσι, υπάρχει ενδιαφέρον για εναλλακτικές τεχνικές ανεύρεσης θέσης, με κύριο παράδειγμα την προσέγγιση "wireless positioning" (http://en.wikipedia.org/wiki/Skyhook_Wireless), που χρησιμοποιήθηκε από το iPhone.

Η διπλωματική βασίζεται στη ιδέα ότι οι φορητές συσκευές που γνωρίζουν τη θέση τους, μπορούν να την εκπέμπουν ώστε να τη χρησιμοποιούν οι κοντινές συσκευές που δεν έχουν δυνατότητα ανεύρεσης θέσης. Θα σχηματίζονται έτσι συνεχώς μεταβαλλόμενα P2P δίκτυα, για μετάδοση συντεταγμένων σε κοντινές αποστάσεις μέσω Bluetooth.

Σε ένα τέτοιο δίκτυο έχει ιδιαίτερη σημασία η εκτίμηση της ακρίβειας των συντεταγμένων. Για παράδειγμα, συντεταγμένες που έχουν ληφθεί πολύ πρόσφατα και κατευθείαν από συσκευή με GPS είναι πιθανότερο να έχουν μεγαλύτερη ακρίβεια από άλλες που έχουν ληφθεί νωρίτερα και που έχουν μεταφερθεί από συσκευή χωρίς δυνατότητα ανεύρεσης θέσης. Επίσης σημαντικός είναι ο προσεγγιστικός υπολογισμός της θέσης μιας συσκευής με βάση τις πρόσφατες συντεταγμένες που έχουν ληφθεί, και το ρυθμό αλλαγής τους.

Με παρόμοιο τρόπο μπορεί να εκτιμηθεί σε κάποιες περιπτώσεις και ο προσανατολισμός (προς τα πού είναι στραμμένη η συσκευή). Για παράδειγμα, αν όλοι οι άνθρωποι γύρω μου κοιτούν προς την ίδια κατεύθυνση είναι πιθανό ότι και εγώ κοιτάζω προς τα εκεί. Αν μια συσκευή γνωρίζει τον προσανατολισμό των άλλων κοντινών συσκευών είναι πιθανό να μπορεί να βγάλει κάποιο συμπέρασμα για το δικό της προσανατολισμό.

Στα πλαίσια της διπλωματικής θα σχεδιαστούν αλγόριθμοι και θα αναπτυχθεί κατάλληλο λογισμικό για να μελετηθούν τα εξής:

- Πόσο πρακτική είναι η χρήση Bluetooth για τη μετάδοση πληροφορίας μεταξύ κινητών; Σε τι αποστάσεις; Με τι ταχύτητες κίνησης;
- Πώς μπορώ να αξιοποιήσω τη γνώση της θέσης κοντινών μου συσκευών για να εκτιμήσω τη δική μου; Πόσο καλά μπορώ να τη προσεγγίσω; Αν ήδη έχω μια προσέγγιση της θέσης μου από GPS (ή wireless positioning κτλ.) μπορώ να τη βελτιώσω; Το ίδιο για τον προσανατολισμό.
- Πόσο εφικτή / αποδοτική είναι μια υλοποίηση των παραπάνω τοπικά στη φορητή συσκευή σε σύγκριση με μια «συγκεντρωτική» (centralized) υλοποίηση όπου οι υπολογισμοί θα γίνονταν σε έναν εξυπηρετητή;

Ο στόχος είναι να υλοποιηθεί και να δοκιμαστεί μια εφαρμογή για κινητά που θα μοιράζεται (εκπομπή-λήψη) και θα (επαν-)εκτιμά με «έξυπνο» τρόπο τη θέση του κινητού (λαβαίνοντας υπόψη τα παραπάνω) χρησιμοποιώντας Bluetooth.

ΣΧΕΤΙΚΟ ΥΛΙΚΟ:

1. Jeffrey Hightower, Anthony LaMarca, Ian E. Smith. “Practical Lessons from Place Lab”, IEEE Pervasive 2006.
2. Timothy Sohn, William G. Griswold, James Scott, Anthony LaMarca, Yatin Chawathe, Ian Smith, Mike Y. Chen. “Experiences with Place Lab: an Open Source Toolkit for Location-Aware Computing”, In Proc. 28th International Conference on Software Engineering (ICSE), 2006.
3. Paramvir Bahl, Venkata N. Padmanabhan. “RADAR: An In-Building RF-based User Location and Tracking System”, In Proc. IEEE Infocom, 2000.
4. Mobile Ad-hoc Networks. http://en.wikipedia.org/wiki/Mobile_ad_hoc_network.

ΚΑΤΑΝΟΜΗ ΚΑΙ RANKING ΜΕ ΒΑΣΗ ΧΩΡΙΚΕΣ ΣΥΝΤΕΤΑΓΜΕΝΕΣ

ΠΛΗΡΟΦΟΡΙΕΣ: Γιάννης Σταύρακας, 210 6990522 (103), yannis@imis.athena-innovation.gr

ΠΕΡΙΛΗΨΗ: Ο στόχος της διπλωματικής είναι η ανάπτυξη μιας εφαρμογής για την ομαδοποίηση διευθύνσεων με βάση (α) τις χωρικές συντεταγμένες στις οποίες αντιστοιχούν, και (β) μια σειρά από περιορισμούς που καθορίζονται από τις παραμέτρους του εκάστοτε προβλήματος.

ΑΤΟΜΑ: 1.

ΠΛΑΤΦΟΡΜΑ ΕΡΓΑΣΙΑΣ: Java (ή αντίστοιχο), προαιρετικά Google Maps (ή αντίστοιχο).

ΣΥΝΤΟΜΗ ΠΕΡΙΓΡΑΦΗ: Η διπλωματική βασίζεται σε ένα σενάριο χρήσης που προέρχεται από ένα επιμέρους πρόβλημα της ΕΣΥΕ (Εθνική Στατιστική Υπηρεσία Ελλάδας). Η ΕΣΥΕ διεξάγει τακτικά έρευνες στο πληθυσμό της χώρας για συγκέντρωση δεδομένων σχετικά με διάφορα θέματα (πχ. απογραφή πληθυσμού, χρήση γης, καλλιέργειες, κτηνοτροφία, κτλ). Σε κάθε έρευνα, ένα σύνολο ερευνόμενων μοιράζεται σε έναν αριθμό ερευνητών, οι οποίοι τους επισκέπτονται και συμπληρώνουν ερωτηματολόγια με βάση τις απαντήσεις τους. Κάθε ερευνητής μπορεί να αναλάβει ένα συγκεκριμένο αριθμό ερευνόμενων, και έχει προτίμηση σε κάποιες περιοχές. Το πρόβλημα είναι (α) η καλύτερη δυνατή κατανομή ερευνόμενων σε ερευνητές, και (β) η κατάταξη των ερευνόμενων για κάθε ερευνητή ώστε να προτείνεται μια σειρά επίσκεψης.

Στα πλαίσια της διπλωματικής θα αναπτυχθεί εφαρμογή που:

- θα δέχεται σαν είσοδο ένα κατάλογο ερευνόμενων με τις διευθύνσεις τους, ένα κατάλογο ερευνητών, και μια σειρά περιορισμών (πχ. αριθμός ερευνόμενων για κάθε ερευνητή, περιοχές προτίμησης ερευνητή).
- θα βρίσκει τις χωρικές συντεταγμένες των ερευνόμενων.
- θα κατανέμει τους ερευνόμενους σε ερευνητές.
- για κάθε ερευνητή θα κατατάσσει τους ερευνόμενους σύμφωνα με μια προτεινόμενη σειρά επίσκεψης.

Η διπλωματική συνδυάζει την υλοποίηση μιας ολοκληρωμένης εφαρμογής (σχεδιασμός, υλοποίηση σε Java, διαπροσωπεία με Google Maps) με τη μελέτη ερευνητικών εργασιών για ανεύρεση κατάλληλων τεχνολογικών λύσεων και αλγορίθμων (προβλήματα βελτιστοποίησης, ερωτήματα ανεύρεσης κοντινότερου γείτονα, clustering, ranking).

ΣΧΕΤΙΚΟ ΥΛΙΚΟ:

1. ΕΣΥΕ. http://www.statistics.gr/portal/page/portal/ESYE.
2. Travelling salesman problem. http://en.wikipedia.org/wiki/Travelling_salesman_problem.
3. Nearest neighbor search. http://en.wikipedia.org/wiki/Nearest_neighbor_search.

ΜΕΤΑΤΡΟΠΗ EAD ΜΕΤΑΔΕΔΟΜΕΝΩΝ ΣΕ ΜΟΡΦΗ ΟΝΤΟΛΟΓΙΑΣ CIDOC CRM

ΠΛΗΡΟΦΟΡΙΕΣ: Χρήστος Παπαθεοδώρου, 210 6990522 (106), c.papatheodorou@dcu.gr
Γιάννης Σταύρακας, 210 6990522 (103), yannis@imis.athena-innovation.gr

ΠΕΡΙΛΗΨΗ: Ο στόχος της διπλωματικής είναι η ανάπτυξη μιας εφαρμογής για την μετατροπή EAD XML εγγράφων σε σημασιολογικά ισοδύναμες περιγραφικές βασισμένες στην οντολογία CIDOC CRM.

ΑΤΟΜΑ: 1.

ΠΛΑΤΦΟΡΜΑ ΕΡΓΑΣΙΑΣ: Οντολογίες, XML τεχνολογίες, Java.

ΣΥΝΤΟΜΗ ΠΕΡΙΓΡΑΦΗ: Οι οργανισμοί (δημόσιοι και ιδιωτικοί) παράγουν πληροφορία η οποία τεκμηριώνει τη λειτουργία τους και οργανώνεται συστηματικά με τη μορφή αρχείων (archives). Τα αρχεία αποτελούνται από υλικό διαφορετικών μορφοτύπων (έντυπα/ηλεκτρονικά έγγραφα, φωτογραφίες, ηχητικό, πολυμεσικό υλικό κ.α.) και είναι φορείς πληροφορίας που αποδεικνύει/τεκμηριώνει τις δραστηριότητες οργανισμών. Η διατήρηση της αρχειακής πληροφορία είναι σημαντική τόσο για την εύρυθμη και αποδοτική λειτουργία των οργανισμών, όσο και για την ιστορική έρευνα και τη διατήρηση της πολιτιστικής κληρονομιάς.

Το Encoded Archival Description (EAD) (Library of Congress, 2002) είναι ένα διεθνές πρότυπο αρχειακής περιγραφής, το οποίο εκφράζεται ως XML DTD και ως XML Schema. Η υλοποίησή του επιτρέπει τη δημιουργία ηλεκτρονικών εργαλείων έρευνας (*finding aids*), τα οποία ενσωματώνουν την περιγραφή των αρχείων (*archives*). Αναλυτικότερα, τα εργαλεία έρευνας είναι τα πληροφοριακά εργαλεία που περιλαμβάνουν μεταδεδομένα για την ταύτιση, διαχείριση και ερμηνεία ενός αρχείου και, συγχρόνως, αναλύουν το περιβάλλον δημιουργίας του. Το συγκεκριμένο XML σχήμα μεταδεδομένων υλοποιείται στις ψηφιακές βιβλιοθήκες τόσο σε εθνικό (Γενικά Αρχεία του Κράτους, Δημοτικό Κέντρο Ιστορίας και Τεκμηρίωσης Βόλου κτλ) όσο και διεθνές επίπεδο (Library of Congress, California Digital Library, Access to Archives (National Archives – UK) etc).

Στα πλαίσια μίας ευρύτερης τάσης για ολοκλήρωση δεδομένων και μεταδεδομένων (*data and metadata integration*) ενός συγκεκριμένου θεματικού τομέα, αξιοποιούνται τα τελευταία χρόνια οι *οντολογίες* ως σχήματα διαμεσολάβησης πηγών (*mediated schema*). Πιο συγκεκριμένα, η οντολογία CIDOC CRM (ICOM/CIDOC and CIDOC CRM SIG, 2008) αξιοποιείται ως σχήμα διαμεσολάβησης πηγών μεταδεδομένων του τομέα πολιτιστικής κληρονομιάς (*αρχεία, βιβλιοθήκες μουσεία*). Αξίζει να σημειωθεί ότι η συγκεκριμένη οντολογία εκφράζεται σε σύνταξη RDF, OWL ενώ παράλληλα έχουν γίνει ερευνητικές προσπάθειες αποθήκευσης δεδομένων βασισμένων στην οντολογία σε σχεσιακές βάσεις δεδομένων. Με στόχο τη σημασιολογική ολοκλήρωση πηγών μέσω της συγκεκριμένης οντολογίας έχουν οριστεί σημασιολογικές συσχετίσεις (*semantic mappings*) ανάμεσα σε διάφορα σχήματα μεταδεδομένων, όπως το EAD, το Dublin Core κτλ, και την οντολογία CIDOC CRM (Stasinopoulou et al., 2007).

Στόχος της προτεινόμενης εργασίας είναι η ανάπτυξη εργαλείου για την μετατροπή EAD XML εγγράφων σε σημασιολογικά ισοδύναμες περιγραφικές βασισμένες στην οντολογία CIDOC CRM. Το εργαλείο αυτό είναι δυνατό να αναπτυχθεί σε XSLT (ή σε κάποια γλώσσα προγραμματισμού, π.χ. JAVA, στον βαθμό που θα εκτιμηθεί ότι είναι καταλληλότερη για την ανάπτυξη ενός τέτοιου εργαλείου).

Οι σημασιολογική αντιστοίχιση των ανάμεσα στα πεδία του EAD και τις κλάσσεις/ιδιότητες της οντολογίας CIDOC CRM θα βασιστεί στην εργασία [3] ενώ θα υπάρξει συνεργασία με τα μέλη της συγγραφικής ομάδας της εργασίας αυτής.

ΣΧΕΤΙΚΟ ΥΛΙΚΟ:

1. Library of Congress (2002). EAD version 2002 official site. Retrieved from http://www.loc.gov/ead/index.html
2. ICOM/CIDOC and CIDOC CRM SIG (2008). Current Official Version of the CIDOC Conceptual Reference Model (Nick Crofts, Martin Doerr, Tony Gill, Stephen Stead, Matthew Stiff (editors)). Retrieved from http://cidoc.ics.forth.gr/docs/cidoc_crm_version_4.2.5.pdf
3. Stasinopoulou, T., Bountouri, L., Kakali, C., Lourdi, Irene, Papatheodorou, C., Doerr, M., and Gergatsoulis, M. (2007). Ontology-based metadata integration in the cultural heritage domain. In Dion Hoe-Lian Goh and Tru Hoang Cao and Ingeborg Torvik Solvberg and Edie Rasmussen (eds.), *Proceedings of the Asian Digital Libraries - Looking Back 10 Years and Forging New Frontiers. 10th International Conference on Asian Digital Libraries, ICADL 2007, Lecture Notes in Computer Science, Vol. 4822* (pp. 165-175). Heidelberg: Springer.

ΔΙΑΧΕΙΡΙΣΗ ΣΥΣΧΕΤΙΖΟΜΕΝΩΝ ΑΛΛΑΓΩΝ ΣΤΟΝ ΙΣΤΟ

ΠΛΗΡΟΦΟΡΙΕΣ: Γιάννης Σταύρακας, 210 6990522 (103), yannis@imis.athena-innovation.gr
Γιώργος Παπαστεφανάτος, 210 6990522 (101), gpapas@imis.athena-innovation.gr

ΠΕΡΙΛΗΨΗ: Ο στόχος της διπλωματικής είναι η ανάπτυξη μιας εφαρμογής που θα καταγράφει, θα παρουσιάζει και θα διαχειρίζεται αλλαγές σε ιστοσελίδες, καθώς και αλληλεξαρτήσεις μεταξύ των αλλαγών αυτών.

ΑΤΟΜΑ: 1.

ΠΛΑΤΦΟΡΜΑ ΕΡΓΑΣΙΑΣ: Java (ή αντίστοιχο).

ΣΥΝΤΟΜΗ ΠΕΡΙΓΡΑΦΗ: Το ζήτημα της αναπαράστασης και αποθήκευσης των αλλαγών σε κείμενα και γενικά σε δεδομένα έχει μελετηθεί πολύ, όπως και η ανακατασκευή προηγούμενων εκδόσεων ή στιγμιότυπων της πληροφορίας. Το πρόβλημα που αναδεικνύει ο Ιστός είναι ότι σε πολλές περιπτώσεις υπάρχουν δεδομένα που, ενώ εξελίσσονται ανεξάρτητα, οι αλλαγές που υφίστανται συσχετίζονται (συνήθως με σχέση αιτίας-αποτελέσματος). Χαρακτηριστικό παράδειγμα αποτελούν οι βιολογικές βάσεις δεδομένων που συντηρούνται από διαφορετικές ομάδες, όμως μια αλλαγή σε μια από αυτές συχνά επιφέρει αλλαγές σε άλλες βάσεις. Έτσι, χρειάζεται κανείς να αντιμετωπίσει το παλαιό πρόβλημα του ιστορικού των δεδομένων, στο νέο πλαίσιο των συσχετιζόμενων, αλληλεξαρτώμενων αλλά και ταυτόχρονα ανεξάρτητων databanks.

Ο στόχος της διπλωματικής είναι να επιδείξει μια προσέγγιση όπου η *αλλαγή* σαν έννοια θα βρίσκεται στο επίκεντρο της λύσης. Θα αναπτυχθεί εφαρμογή που:

- Θα διαχειρίζεται την εξέλιξη ενός συγκεκριμένου web εγγράφου. Όταν χρειαστεί να γίνουν αλλαγές στο έγγραφο αυτό, ο χρήστης θα βασίζεται στην συγκεκριμένη εφαρμογή, η οποία θα τις καταγράφει με κάποιον από τους ήδη γνωστούς από τη βιβλιογραφία τρόπους (στην απλούστερη περίπτωση αποθηκεύοντας ολόκληρα τα διαδοχικά έγγραφα, και δίνοντας ένα όνομα στην ομάδα των αλλαγών που οδηγούν σε κάθε νέα έκδοση).
- Θα παραχωρεί αυτόματα ένα web address σε κάθε τέτοια ομάδα αλλαγών, ώστε να μπορεί κάποιος να αναφερθεί σε μια αλλαγή όπως σε ένα οποιοδήποτε αντικείμενο στον ιστό.
- Θα υποστηρίζει τη διασύνδεση μεταξύ αλλαγών ώστε να μπορεί να καταγράψει κάποιος ποιές αλλαγές έχουν οδηγήσει σε ποιες άλλες αλλαγές (πιθανόν μεταξύ διαφορετικών εγγράφων σε άλλα websites).
- Θα αποκρίνεται σε μια συγκεκριμένη web διεύθυνση (που θα σχετίζεται και με το εν λόγω έγγραφο), δίνοντας μια "κάτοψη" όλων των (ομάδων) αλλαγών που έχουν λάβει χώρα στο έγγραφο αυτό, καθώς και των διασυνδέσεών τους με άλλες αλλαγές (στο ίδιο ή σε άλλα έγγραφα).
- Θα παρουσιάζει οποιοδήποτε προηγούμενο στιγμιότυπο του εγγράφου. Κάθε στιγμιότυπο θα έχει δική του web διεύθυνση, που θα παραχωρείται αυτόματα από την εφαρμογή.
- Θα επιτρέπει την περιήγηση στον Ιστό δια μέσου των συσχετίσεων των αλλαγών, μέχρι το επίπεδο των στιγμιότυπων των αντίστοιχων εγγράφων.
- Θα υποστηρίζει διάδοση ειδοποιήσεων όταν συμβαίνει μια αλλαγή που μπορεί να προκαλέσει άλλες αλλαγές.

ΣΧΕΤΙΚΟ ΥΛΙΚΟ:

1. A comparative study of XML diff tools. Gregory Cobena, Talel Abdessalem, Yassine Hinnach. INRIA Futurs - Projet GEMO, 2004.

"PAPER-MAP"

ΠΛΗΡΟΦΟΡΙΕΣ: Γιάννης Σταύρακας, 210 6990522 (103), yannis@imis.athena-innovation.gr

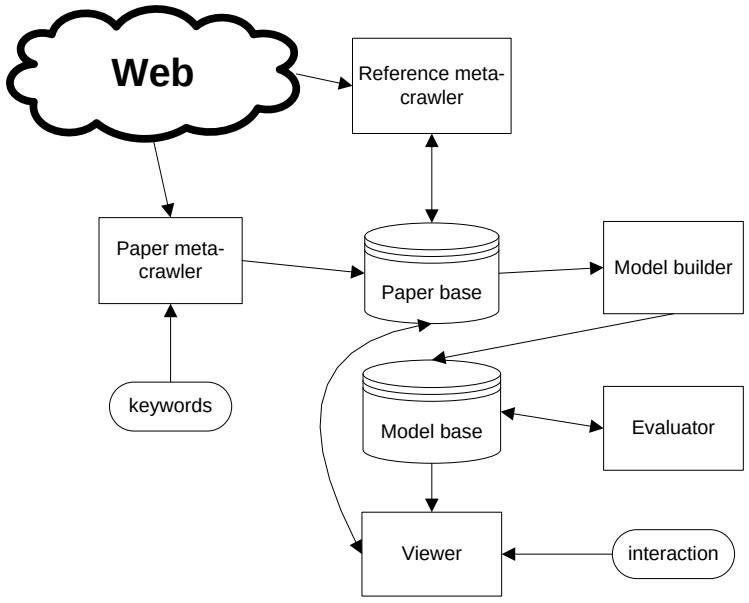
ΠΕΡΙΛΗΨΗ: Στόχος είναι η ανάπτυξη συστήματος που (α) θα ανακτά από το διαδίκτυο εργασίες για μια ερευνητική περιοχή, (β) θα μοντελοποιεί τους συσχετισμούς (πχ. αναφορές) μεταξύ τους, θα κατατάσσει τις εργασίες με σειρά σπουδαιότητας και θα προτείνει τη σειρά με την οποία πρέπει να μελετηθούν, και (γ) θα παρουσιάζει τις εργασίες και τους συσχετισμούς τους με τρόπο που να διευκολύνει τον ερευνητή να οργανώσει την μελέτη του.

ΑΤΟΜΑ: 1-2.

ΠΛΑΤΦΟΡΜΑ ΕΡΓΑΣΙΑΣ: Java (ή αντίστοιχο).

ΣΥΝΤΟΜΗ ΠΕΡΙΓΡΑΦΗ: Για την μελέτη μιας ερευνητικής περιοχής ο ερευνητής αρχίζει συνήθως από κάποιες εργασίες που ήδη γνωρίζει ή που βρίσκει στο διαδίκτυο. Από αυτές αρχικά επιλέγει surveys, που δίνουν μια συνολική εικόνα της δουλειάς στην περιοχή και περιέχουν πολλές παραπομπές σε σχετικές εργασίες. Στη συνέχεια επιλέγει εργασίες που έχουν δημοσιευτεί σε καλά συνέδρια από γνωστούς ερευνητές και που αναφέρονται από πολλές άλλες εργασίες του χώρου. Για να καθορίσει τη σειρά με την οποία θα διαβάσει τις εργασίες παίζει ρόλο η ημερομηνία δημοσίευσης, καθώς και οι εξαρτήσεις από / αναφορές σε άλλες εργασίες. Αυτές τις διαδικασίες θέλουμε να αυτοματοποιήσουμε, δίνοντας την ευκολία της άμεσης ανεύρεσης και κατάταξης εργασιών αλλά και συνεδρίων καθώς και συγγραφέων για μια θεματική περιοχή.

Ο τρόπος κατάταξης θα βασιστεί στη διαπίστωση ότι οι εργασίες μπορούν να αναπαρασταθούν σαν κόμβοι ενός γράφου όπου οι ακμές δηλώνουν αναφορές (references / citations). Στη συνέχεια θα εφαρμοστεί παραλλαγή του αλγόριθμου HITS του Google για να βρεθούν καλά hubs (δηλαδή καλά survey papers) και καλά authorities (καλές εργασίες). Η παραλλαγή θα παίρνει υπόψη / θα βγάζει συμπεράσματα για τη σπουδαιότητα των συνεδρίων και των συγγραφέων για τη συγκεκριμένη θεματική περιοχή.
Η αρχιτεκτονική απεικονίζεται στο σχήμα που ακολουθεί.



1. *Paper meta-crawler*. Ξεκινώντας από μια σειρά keywords, ψάχνει συγκεκριμένα sites στον ιστό (πχ Google scholar) για εργασίες της αντίστοιχης θεματικής περιοχής. Τα αποτελέσματα ολοκληρώνονται (απάλειψη διπλοτύπων), αναλύονται λεξικά, αναγνωρίζονται οι οντότητες (πχ συγγραφείς, τίτλοι, κτλ) και η πληροφορία αποθηκεύεται στο Paper base. Η πληροφορία αυτή συνιστά το root set.
2. *Reference meta-crawler*. Επαυξάνει το root set ως εξής. Για κάθε εργασία του root set ψάχνει τον ιστό για εργασίες που την αναφέρουν ή που αναφέρονται από αυτή. Τα αποτελέσματα περνούν όπως πριν από ολοκλήρωση, ανάλυση και αναγνώριση οντοτήτων και αποθηκεύονται στο paper base. Η πληροφορία αυτή (μαζί με το root set) συνιστά το base set.
3. *Model builder*. Χτίζεται το μοντέλο που αναπαριστά πρωτίστως τις αναφορές μεταξύ εργασιών. Βάση του μοντέλου αποτελεί ένας κατευθυνόμενος γράφος. Η αναπαράσταση του μοντέλου αποθηκεύεται στο Model base.
4. *Evaluator*. Συμπληρώνει το μοντέλο με στοιχεία που επιτρέπουν τη ταξινόμηση των εργασιών με διάφορα κριτήρια. Τα στοιχεία προκύπτουν από αλγορίθμους όπως ο HITS, προσαρμοσμένος ώστε να παίρνει υπόψη εκτός από καλά hubs (survey papers) και authorities με βάση τις εισερχόμενες και εξερχόμενες ακμές, την ποιότητα συνεδρίων και συγγραφέων, τις ημερομηνίες, κτλ.
5. *Viewer*. Παρουσιάζει το μοντέλο με διάφορους τρόπους, και επιτρέπει μια σειρά από λειτουργίες, όπως: αλλαγή του base set με τον αποκλεισμό ή την εισροή εργασιών, model annotation, navigation, export, prioritize, etc.

Η έκταση της διπλωματικής θα προσαρμοστεί στον αριθμό των ατόμων (ένας ή δύο) που θα ασχοληθούν, μετά από συνεννόηση.

ΣΧΕΤΙΚΟ ΥΛΙΚΟ:

1. HITS algorithm. http://en.wikipedia.org/wiki/HITS_algorithm.

ΥΠΗΡΕΣΙΕΣ ΑΝΑΖΗΤΗΣΗΣ ΣΤΟΝ ΙΣΤΟ ΜΕ ΧΡΗΣΗ ΕΡΓΑΛΕΙΩΝ ΔΙΑΧΕΙΡΙΣΗΣ ΣΚΕΨΗΣ

ΠΛΗΡΟΦΟΡΙΕΣ: Θοδωρής Δαλαμάγκας, 210 6990522 (102), dalamag@imis.athena-innovation.gr
Γιώργος Γιαννόπουλος, 210 6990522 (102), giann@imis.athena-innovation.gr

ΠΕΡΙΛΗΨΗ: Στόχος της διπλωματικής εργασίας είναι η ανάπτυξη μιας εφαρμογής μετα-αναζήτησης πληροφορίας στον Ιστό που θα βασίζεται στην οργάνωση της αναζήτησης μέσω μοντέλων διαχείρισης σκέψης (mindmaps).

ΑΤΟΜΑ: 1-2

ΠΛΑΤΦΟΡΜΑ ΕΡΓΑΣΙΑΣ: Freemind/Java

ΠΕΡΙΓΡΑΦΗ: Στα πλαίσια των δραστηριοτήτων του ΙΠΣΥΠ έχει αναπτυχθεί μια εφαρμογή που υποστηρίζει **προηγμένες τεχνικές αναζήτησης στον Ιστό**. Βασικό καινοτόμο στοιχείο της εφαρμογής είναι η χρήση ενός μοντέλου ερώτησης που βασίζεται σε αφαιρετικές περιγραφές και συσχετίσεις εννοιών προς αναζήτηση, οργανωμένες σε χάρτες σκέψης **(mindmaps)** (δείτε σχετικά http://freemind.sourceforge.net/wiki/index.php/Main_Page).

Η εφαρμογή έχει χτιστεί πάνω από το εργαλείο χαρτών σκέψης Freemind (http://freemind.sourceforge.net/) και προσφέρει ένα σύνολο υπηρεσιών, σε ένα ενδιάμεσο επίπεδο μεταξύ χρήστη και μηχανών αναζήτησης, που εκτελούν τις παρακάτω λειτουργίες:

1. **Αναζήτηση πληροφορίας** στον Ιστό για περιεχόμενο σχετικό με τους χάρτες σκέψης και οργάνωση των αποτελεσμάτων σε χάρτες σκέψης. Το σύστημα έχει επικεντρωθεί στην **(μετα-) αναζήτηση** κειμένων επιστημονικών συνεδρίων και περιοδικών από **ειδικευμένες πηγές αναζήτησης**, όπως το DBLP (http://www.informatik.uni-trier.de/~ley/db/), Citeseerx (http://citeseerx.ist.psu.edu/) και Google scholar (http://scholar.google.com/).
2. **Εμπλουτισμός** αποτελεσμάτων αναζήτησης. Το σύστημα χρησιμοποιεί υλικό από **γενικευμένες πηγές αναζήτησης** (π.χ. Google, Yahoo, κ.λ.π.) για να βρει πληροφορία που δεν έχει βρεθεί από τη Λειτουργία 1. και είναι απαραίτητη για το χρήστη. Παραδείγματα: η περίληψη (abstract) μιας δημοσίευσης, το pdf αρχείο της (π.χ. αν υπάρχει στην προσωπική ιστοσελίδα των συγγραφέων), κ.λ.π.
3. **Οργάνωση αποτελεσμάτων σε όψεις (facets).** Το σύστημα οργανώνει τα αποτελέσματα σε ομάδες, ανάλογα με την τιμή κάποιας ιδιότητάς τους (π.χ. δημοσιεύσεις ανά έτος, ανά συγγραφέα) αλλά και ανά θεματική περιοχή (π.χ. δημοσιεύσεις σχετικές με Βάσεις Δεδομένων, με Ανάκτηση Πληροφορίας κ.λ.π.).

Η διπλωματική εργασία έχει ως στόχο την προσθήκη δύο νέων λειτουργικών δυνατοτήτων στην εφαρμογή:

1. **Κόμβοι Actors στους χάρτες σκέψεις**. Η διπλωματική εργασία θα επεκτείνει το μοντέλο χαρτών σκέψης με την προσθήκη μιας ειδικής κατηγορίας κόμβων που θα επιτελούν εξειδικευμένες λειτουργίες αναζήτησης. Οι κόμβοι αυτοί θα εξάγουν λέξεις-κλειδιά από τους άλλους κόμβους με τους οποίους σχετίζονται στο χάρτη και θα τις χρησιμοποιούν για να δρομολογούν τις αναζητήσεις. Παραδείγματα:
    a. κόμβοι AbstractCollector: θα δρομολογούν αναζήτηση περιλήψεων δημοσιεύσεων (abstracts)
    b. κόμβοι WikipediaCollector: θα δρομολογούν αναζήτηση υλικού από την εγκυκλοπαίδεια Wikipedia

2. **Αυτόματη ενσωμάτωση νέων μηχανών αναζήτησης**. Η διπλωματική εργασία θα υλοποιήσει μια υπηρεσία αυτόματης ενσωμάτωσης νέας μηχανής αναζήτησης με χρήση τεχνικών **screen scrapping/wrapping**. Οι τεχνικές αυτές βασίζονται στη χρήση παραδείγματος εκτέλεσης. Ο χρήστης δηλαδή (φάση εκπαίδευσης) (α) εκτελεί μια αρχική αναζήτηση σε μια νέα μηχανή, (β) «μαρκάρει» το είδος της πληροφορίας που τον ενδιαφέρει και (γ) της δίνει κάποιο χαρακτηριστικό, και έπειτα (φάση εκτέλεσης) το σύστημα μαθαίνει για κάθε νέα αναζήτηση (α) να «καλεί» τη μηχανή αναζήτησης, (β) να επεξεργάζεται την ιστοσελίδα του αποτελέσματος και (γ) να εξάγει την πληροφορία αυτόματα. Παράδειγμα τέτοιου εργαλείου στο οποίο θα βασιστεί η εφαρμογή είναι το Dapper (http://www.dapper.net/open/).

Όλες οι λειτουργίες θα υλοποιηθούν πάνω από το σύστημα που έχει αναπτυχθεί στο ΙΠΣΥΠ.

ΣΧΕΤΙΚΟ ΥΛΙΚΟ:

1. http://www.dbnet.ece.ntua.gr/pubs/details.php?id=1550&clang=1
2. http://www.dbnet.ece.ntua.gr/pubs/details.php?id=1537&clang=1
3. http://web.imis.athena-innovation.gr/~dalamag/mm/

ΕΞΑΤΟΜΙΚΕΥΜΕΝΗ ΑΝΑΖΗΤΗΣΗ ΜΕ ΧΡΗΣΗ ΕΡΓΑΛΕΙΩΝ ΔΙΑΧΕΙΡΙΣΗΣ ΣΚΕΨΗΣ

ΠΛΗΡΟΦΟΡΙΕΣ: Θοδωρής Δαλαμάγκας, 210 6990522 (102), dalamag@imis.athena-innovation.gr
Γιώργος Γιαννόπουλος, 210 6990522 (102), giann@imis.athena-innovation.gr

ΠΕΡΙΛΗΨΗ: Στόχος της διπλωματικής εργασίας είναι η ανάπτυξη μιας εφαρμογής εξατομικευμένης αναζήτησης πληροφορίας στον Ιστό που θα βασίζεται στην οργάνωση της αναζήτησης μέσω μοντέλων διαχείρισης σκέψης (mindmaps).

ΑΤΟΜΑ: 1-2

ΠΛΑΤΦΟΡΜΑ ΕΡΓΑΣΙΑΣ: Freemind/Java

ΠΕΡΙΓΡΑΦΗ: Στα πλαίσια των δραστηριοτήτων του ΙΠΣΥΠ έχει αναπτυχθεί μια εφαρμογή που υποστηρίζει **προηγμένες τεχνικές αναζήτησης στον Ιστό**. Βασικό καινοτόμο στοιχείο της εφαρμογής είναι η χρήση ενός μοντέλου ερώτησης που βασίζεται σε αφαιρετικές περιγραφές και συσχετίσεις εννοιών προς αναζήτηση, οργανωμένες σε χάρτες σκέψης **(mindmaps)** (δείτε σχετικά http://freemind.sourceforge.net/wiki/index.php/Main_Page).

Η εφαρμογή έχει χτιστεί πάνω από το εργαλείο χαρτών σκέψης Freemind (http://freemind.sourceforge.net/) και προσφέρει ένα σύνολο υπηρεσιών, σε ένα ενδιάμεσο επίπεδο μεταξύ χρήστη και μηχανών αναζήτησης, που εκτελούν τις παρακάτω λειτουργίες:

1. **Αναζήτηση πληροφορίας** στον Ιστό για περιεχόμενο σχετικό με τους χάρτες σκέψης και οργάνωση των αποτελεσμάτων σε χάρτες σκέψης. Το σύστημα έχει επικεντρωθεί στην **(μετα-) αναζήτηση** κειμένων επιστημονικών συνεδρίων και περιοδικών από **ειδικευμένες πηγές αναζήτησης**, όπως το DBLP (http://www.informatik.uni-trier.de/~ley/db/), Citeseerx (http://citeseerx.ist.psu.edu/) και Google scholar (http://scholar.google.com/).
2. **Εμπλουτισμός** αποτελεσμάτων αναζήτησης. Το σύστημα χρησιμοποιεί υλικό από **γενικευμένες πηγές αναζήτησης** (π.χ. Google, Yahoo, κ.λ.π.) για να βρει πληροφορία που δεν έχει βρεθεί από τη Λειτουργία 1. και είναι απαραίτητη για το χρήστη. Παραδείγματα: η περίληψη (abstract) μιας δημοσίευσης, το pdf αρχείο της (π.χ. αν υπάρχει στην προσωπική ιστοσελίδα των συγγραφέων), κ.λ.π.
3. **Οργάνωση αποτελεσμάτων σε όψεις (facets).** Το σύστημα οργανώνει τα αποτελέσματα σε ομάδες, ανάλογα με την τιμή κάποιας ιδιότητάς τους (π.χ. δημοσιεύσεις ανά έτος, ανά συγγραφέα) αλλά και ανά θεματική περιοχή (π.χ. δημοσιεύσεις σχετικές με Βάσεις Δεδομένων, με Ανάκτηση Πληροφορίας κ.λ.π.).

Η διπλωματική εργασία έχει ως στόχο τη βελτίωση και τον εμπλουτισμό του μηχανισμού αναζήτησης της εφαρμογής. Συγκεκριμένα θα προστεθούν οι παρακάτω μηχανισμοί:

1. **Εξατομικευμένη αναζήτηση με βάση το ιστορικό του χρήστη**. Η διπλωματική εργασία θα δίνει τη δυνατότητα προσαρμογής των αναζητήσεων του χρήστη με βάση προηγούμενες αναζητήσεις του. Συγκεκριμένα, θα λαμβάνει υπόψιν τα ερωτήματα που έκανε στο παραλθόν ο χρήστης και τα αποτελέσματα του κάθε ερωτήματος που επέλεξε να επεξεργαστεί/αποθηκεύσει στο χάρτη σκέψης. Χρησιμοποιώντας αυτήν την πληροφορία, θα δημιουργεί πολλαπλά προφίλ του χρήστη, που θα

αντιστοιχούν σε θεματικές κατηγορίες που τον ενδιαφέρουν. Στη συνέχεια, θα εκμεταλλεύεται αυτά τα προφίλ, ώστε να να φιλτράρει ή να επεκτείνει τα αποτελέσματα κάθε νέας αναζήτησης.

2. **Εξατομικευμένη αναζήτηση με βάση το context του χρήστη**. Η διαδικασία εξατομίκευσης της αναζήτησης θα είναι παρόμοια με παραπάνω, τα προφίλ, όμως, του χρήστη θα δημιουργούνται με διαφορετικό τρόπο. Συγκεκριμένα, ο χρήστης θα μπορεί να συνδέει κόμβους του χάρτη σκέψης με συγκεκριμένη βιβλιογραφία που έχει διαθέσιμη στον υπολογιστή του. Με αυτόν τον τρόπο, ο εκάστοτε κόμβος θα χαρακτηρίζεται θεματικά με βάση το περιεχόμενο της συνδεδεμένης με αυτόν βιβλιογραφίας, εξειδικεύοντας ταυτόχρονα τυχόν αναζητήσεις που σχετίζονται με αυτόν.

Όλες οι λειτουργίες θα υλοποιηθούν πάνω από το σύστημα που έχει αναπτυχθεί στο ΙΠΣΥΠ.

ΣΧΕΤΙΚΟ ΥΛΙΚΟ:

1. http://www.dbnet.ece.ntua.gr/pubs/details.php?id=1550&clang=1
2. http://www.dbnet.ece.ntua.gr/pubs/details.php?id=1537&clang=1
3. http://web.imis.athena-innovation.gr/~dalamag/mm/

ΠΡΟΗΓΜΕΝΗ ΕΦΑΡΜΟΓΗ ΙΣΤΟΥ ΔΙΑΧΕΙΡΙΣΗΣ ΔΕΔΟΜΕΝΩΝ ΒΙΟΕΠΙΣΤΗΜΩΝ

ΠΛΗΡΟΦΟΡΙΕΣ: Θοδωρής Δαλαμάγκας, 210 6990522 (102), dalamag@imis.athena-innovation.gr
Θανάσης Βεργούλης, 210 6990522 (102), vergoulis@imis.athena-innovation.gr

ΠΕΡΙΛΗΨΗ: Στόχος της διπλωματικής εργασίας είναι η ανάπτυξη νέων υπηρεσιών για εφαρμογές Ιστού που υποστηρίζουν ερευνητικές διεργασίες για ερευνητές βιοεπιστημών. Οι συγκεκριμένες εφαρμογές δίνουν δυνατότητες αποθήκευσης, διαχείρισης και αναζήτησης βιολογικών δεδομένων, πειραμάτων και σχετιζόμενων πληροφοριών από βάσεις βιομορίων και ψηφιακές βιβλιοθήκες επιστημονικών δημοσιεύσεων.

ΑΤΟΜΑ: 1-2

ΠΛΑΤΦΟΡΜΑ ΕΡΓΑΣΙΑΣ: Ανοικτή

ΣΥΝΤΟΜΗ ΠΕΡΙΓΡΑΦΗ: Τα μόρια μικρο-RNA (micro-RNA) είναι βιομόρια πολύ μικρού μήκους, τα οποία προσκολλώνται επάνω σε μόρια «αγγελιοφόρου» RNA (mRNA) εμποδίζοντας την παραγωγή πρωτεϊνών. Για την έρευνα στις βιοεπιστήμες, είναι χρήσιμο σε κάποιον ερευνητή να γνωρίζει σε ποια mRNA προσκολλάται ένα συγκεκριμένο μόριο μικρο-RNA. Τα σημεία προσκόλλησης αποκαλούνται στόχοι (targets). Η πληροφορία αυτή είναι πολύτιμη για την εξαγωγή συμπερασμάτων σχετικά με τα αίτια ασθενειών. Επειδή ο ρόλος των μικρο-RNA ανακαλύφθηκε μόλις το 2001, η περιοχή είναι την περίοδο αυτή πολύ δημοφιλής στην ερευνητική κοινότητα των βιοεπιστημών.
Το **DIANA microT** (http://diana.cslab.ece.ntua.gr/microT/) και το **DIANA mirGen** (http://diana.cslab.ece.ntua.gr/mirGen/) είναι δύο εφαρμογές Ιστού, που έχουν αναπτυχθεί από το ΙΠΣΥΠ σε συνεργασία με το Ερευνητικό Κέντρο Βιοϊατρικών Επιστημών "Αλέξανδρος Φλέμινγκ". Η εφαρμογή DIANA microT προσφέρει στον ερευνητή τη δυνατότητα να αναζητήσει στόχους (targets) για διάφορα μικρο-RNA και σχετικές πληροφορίες. Οι στόχοι έχουν προβλεφθεί με βάση υπολογιστικές τεχνικές της ερευνητικής ομάδας του "Αλέξανδρος Φλέμινγκ" και στη συνέχεια έχουν αποθηκευτεί σε μια σχεσιακή βάση δεδομένων. Η εφαρμογή DIANA mirGen προσφέρει στον ερευνητή τη δυνατότητα να αναζητήσει πληροφορίες σχετικές με την έκφραση των διαφόρων μικρο-RNA.
Η διπλωματική εργασία στοχεύει στη σχεδίαση και ανάπτυξη νέων υπηρεσιών για τις προηγούμενες εφαρμογές, όπως για παράδειγμα:

- **Αναζήτηση και απεικόνιση πειραματικών αποτελεσμάτων για στόχους (targets) που έχουν υπολογιστεί με βάση διάφορες μεθόδους.** Κάθε στόχος (target) Σ που έχει υπολογιστεί από την ομάδα του «Αλέξανδρος Φλέμινγκ» αποτελεί μια πρόβλεψη και συνεπώς συνοδεύεται από κάποιο βαθμό εμπιστοσύνης β. Ο βαθμός αυτός εκφράζει το πόσο σίγουρη είναι η πρόβλεψη με βάση τη συγκεκριμένη μέθοδο. Εκτός από τα αποτελέσματα των μεθόδων της ομάδας του «Αλέξανδρος Φλέμινγκ» υπάρχουν και άλλες ομάδες που καταλήγουν σε στόχους με βάση δικές τους υπολογιστικές μεθόδους. Όταν κάποιος αναζητά στόχους (π.χ. για κάποιο συγκεκριμένο μικρο-RNA) κάθε μέθοδος θα επιστρέψει από μια λίστα με ζεύγη (Σ, β). Ένας στόχος Σ μπορεί να είναι στην $1^{η}$ θέση σε μια λίστα, στην $3^{η}$ θέση σε μία άλλη, κ.λ.π. Επίσης, ένας στόχος Σ1 μιας λίστας μπορεί να αντιστοιχεί στους στόχους Σ2 και Σ3 σε μια άλλη λίστα. Σκοπός είναι να υλοποιηθούν αλγόριθμοι που θα παράγουν μια ενιαία λίστα αποτελεσμάτων (fused rank list) η οποία θα λαμβάνει υπ' όψη (α) τις σχετικές θέσεις των στόχων στις επιμέρους λίστες και (β) τις αντιστοιχίσεις που αναφέρθηκαν.

- **Αναζήτηση δημοσιεύσεων**. Στις παρούσες εφαρμογές προσφέρεται και η δυνατότητα εύρεσης σχετικών με τα αποτελέσματα μιας αναζήτησης δημοσιεύσεων. Οι δημοσιεύσεις αντλούνται από το PubMed (http://www.ncbi.nlm.nih.gov/pubmed/). Η εργασία θα διευρύνει αυτές τις δυνατότητες και θα προτείνει τη χρήση γνώσης που θα προκύπτει από μιας μορφής οντολογία για τη βελτίωση της ακρίβειας στην αναζήτηση σχετικών δημοσιεύσεων.

Οι επεκτάσεις είναι ενδεικτικές γιατί η εφαρμογή έχει πλήθος από ενδιαφέρουσες λειτουργικές απαιτήσεις. Ο καθορισμός τους στα πλαίσια της διπλωματικής εξαρτάται και από τη φαντασία και τη δημιουργικότητά σας.

ΕΥΦΥΕΣ ΣΥΣΤΗΜΑ ΔΙΑΧΕΙΡΙΣΗΣ ΡΟΩΝ ΔΕΔΟΜΕΝΩΝ ATOM/ RSS

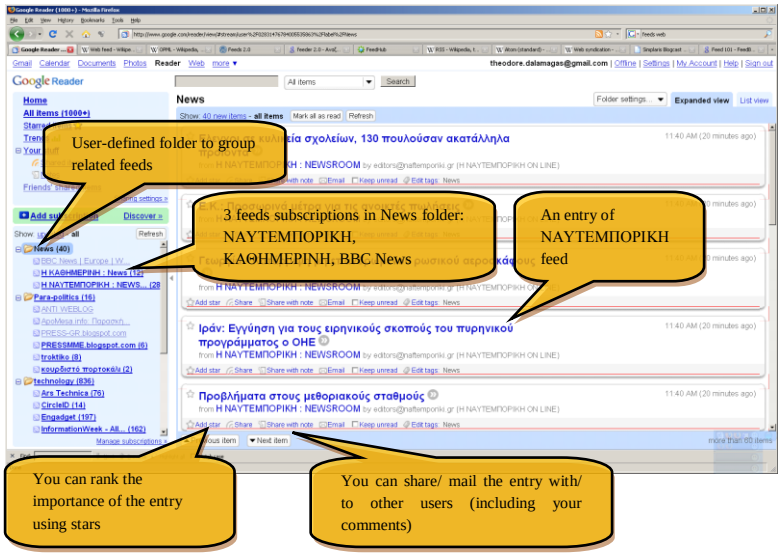
ΠΛΗΡΟΦΟΡΙΕΣ: Θοδωρής Δαλαμάγκας, 210 6990522 (102), dalamag@imis.athena-innovation.gr
Θανάσης Βεργούλης, 210 6990522 (102), vergoulis@imis.athena-innovation.gr

ΠΕΡΙΛΗΨΗ: Στόχος της διπλωματικής εργασίας είναι η ανάπτυξη ενός συστήματος διαχείρισης ροών δεδομένων RSS/Atom (intelligent Web RSS/Atom feeder) που θα δίνει έμφαση στη διαχείριση περιεχομένου με μορφή ιστοριών.

ΑΤΟΜΑ: 1-2

ΠΛΑΤΦΟΡΜΑ ΕΡΓΑΣΙΑΣ: Ανοικτή

ΣΥΝΤΟΜΗ ΠΕΡΙΓΡΑΦΗ: Οι ηλεκτρονικές ροές δεδομένων (web feeds) είναι μια ευρέως διαδεδομένη μέθοδος για τη διάθεση πληροφορίας στον Ιστό με συχνά ανανεούμενο ψηφιακό περιεχόμενο (π.χ. εγγραφές ιστολογίων (blogs), πρωτοσέλιδα ειδήσεων κτλ). Μια ροή ειδήσεων περιλαμβάνει συνήθως αντικείμενα με τα εξής στοιχεία: (α) τίτλο, (β) σύντομη περιγραφή, (γ) μεταδεδομένα και (δ) σύνδεση (http link) στο πλήρες περιεχόμενο. Το RSS και το Atom είναι πρότυπα αναπαράστασης ενός web feed.



Οι χρήστες προκειμένου να λαμβάνουν το περιεχόμενο ενός feed χρησιμοποιούν ένα συσσωρευτή Web feed (Web feed aggregator/reader) με τον οποίο και εγγράφονται (subscribe) στην RSS υπηρεσία. Δείτε για παράδειγμα τον συσσωρευτή Google Reader στη διπλανή εικόνα.

Ο συσσωρευτής παρακολουθεί τα εγγεγραμμένα σε αυτόν feeds για το πότε έχουν νέο περιεχόμενο και ενημερώνει σχετικά το χρήστη ώστε να το δει.

Τα τελευταία χρόνια έχουν εμφανιστεί στην αγορά συσσωρευτές Web feed νέας γενιάς. Οι συσσωρευτές αυτοί δεν περιορίζονται απλώς στην εμφάνιση των περιεχομένων του feed, αλλά επιχειρούν επίσης να τα βαθμολογήσουν (ranking) με βάση τις προτιμήσεις του χρήστη. Οι προτιμήσεις του χρήστη μπορούν να εξαχθούν (α) άμεσα, π.χ. ρωτώντας το χρήστη εκ των προτέρων για το ποια είναι τα ενδιαφέροντά του και ποια άρθρα του αρέσουν ή (β) έμμεσα, π.χ. καταγράφοντας τις επιλογές του και ποια άρθρα έχει διαβάσει στο παρελθόν. Παράδειγμα τέτοιων συστημάτων είναι το FreeHub (http://www.freehub.com).

Η διπλωματική εργασία στοχεύει στη σχεδίαση και ανάπτυξη ενός προηγμένου συσσωρευτή Web feed με τα ακόλουθα χαρακτηριστικά:

- **Monitor Alert.** Ο χρήστης θα μπορεί να ορίζει ένα θέμα με κάποιες λέξεις-κλειδιά και το σύστημα θα ξεχωρίζει τα αντικείμενα της ροής που σχετίζονται με το θέμα ώστε να ενημερώνει το χρήστη άμεσα.

- Έξυπνος χειρισμός **διπλοτύπων.** Το περιεχόμενο που φτάνει στον συσσωρευτή από δύο διαφορετικές πηγές, αλλά αφορά το ίδιο θέμα, θα ομαδοποιείται ώστε να μειωθεί ο όγκος της πληροφορίας που φτάνει στο χρήστη.
- **Διαχείριση περιεχομένου με μορφή ιστοριών**. Για κάθε νέο αντικείμενο ροής Α, το σύστημα θα μπορεί να ανιχνεύσει παλαιότερα αντικείμενα ροής σχετικά με τα Α και να τα παρουσιάζει ως μια ιστορία, δηλαδή μια χρονική αλληλουχία γεγονότων. Οι ιστορίες θα διαχειρίζονται ως αυτόνομες οντότητες. Θα υπάρχουν δηλαδή μηχανισμοί:
    - αποθήκευσης ιστοριών,
    - ανανέωσης ιστοριών,
    - αναζήτησης παρόμοιων ιστοριών και
    - αναζήτησης κομβικών σημείων των ιστοριών.

ΣΧΕΤΙΚΟ ΥΛΙΚΟ:

http://portal.acm.org/citation.cfm?id=1558334.1558343

ΜΕΛΕΤΗ ΚΑΙ ΥΛΟΠΟΙΗΣΗ ΑΛΓΟΡΙΘΜΩΝ ΓΙΑ ΒΙΟΛΟΓΙΚΕΣ ΕΦΑΡΜΟΓΕΣ ΣΕ MAPREDUCE ΠΕΡΙΒΑΛΛΟΝ

ΠΛΗΡΟΦΟΡΙΕΣ: Θοδωρής Δαλαμάγκας, 210 6990522 (102), dalamag@imis.athena-innovation.gr
Θανάσης Βεργούλης, 210 6990522 (102), vergoulis@imis.athena-innovation.gr

ΠΕΡΙΛΗΨΗ: Στόχος της διπλωματικής εργασίας είναι αφενός η μελέτη της δουλειάς που έχει γίνει σχετικά με τη χρήση του MapReduce προγραμματιστικού μοντέλου για την επίλυση προβλημάτων που σχετίζονται με βιολογικά (κυρίως ακολουθιακά) δεδομένα, αφετέρου η υλοποίηση MapReduce αλγορίθμων για τέτοιου είδους προβλήματα.

ΑΤΟΜΑ: 1

ΠΛΑΤΦΟΡΜΑ ΕΡΓΑΣΙΑΣ: Hadoop (Java)

ΣΥΝΤΟΜΗ ΠΕΡΙΓΡΑΦΗ: Στον πυρήνα της έρευνας στις Βιοεπιστήμες βρίσκονται αλγόριθμοι αναζήτησης και επεξεργασίας ακολουθιακών δεδομένων (π.χ. ο αλγόριθμος BLAST). Μια ακολουθία είναι απλώς μια σειρά από σύμβολα, τα οποία επιλέγονται από ένα συγκεκριμένο αλφάβητο. Τα βασικά βιολογικά μόρια (π.χ. DNA, RNA, πρωτεΐνες), μπορούν να αναπαρασταθούν με τη μορφή ακολουθιών. Στη συνέχεια εφαρμόζονται επάνω τους αλγόριθμοι σαν τον BLAST και, από τα αποτελέσματα που παράγονται, προκύπτουν χρήσιμα βιολογικά συμπεράσματα.
Οι προαναφερθέντες αλγόριθμοι είναι συνήθως απαιτητικοί σε χρόνο υπολογισμού και μνήμη. Καθώς τα δεδομένα γίνονται όλο και περισσότερα (π.χ. Personal Genome Project), η κατανομή του υπολογισμού σε πολλούς επεξεργαστές φαίνεται ως μονόδρομος για την γρήγορη παραγωγή αποτελεσμάτων. Η κατανομή των δεδομένων σε πολλούς σκληρούς δίσκους φαίνεται επίσης αναπόφευκτη, καθώς τα μεγέθη των δίσκων είναι ανεπαρκή.
Το MapReduce είναι ένα προγραμματιστικό μοντέλο το οποίο έχει κυριαρχήσει τα τελευταία χρόνια στο πεδίο του κατανεμημένου υπολογισμού. Έχει γίνει ιδιαίτερα δημοφιλές γιατί απλοποιεί τον προγραμματισμό σε τέτοιο περιβάλλον και γιατί υπάρχουν ακόμα και ανοιχτού κώδικα υλοποιήσεις του (π.χ. το Hadoop), οι οποίες λειτουργούν καλά ακόμα και για μεγέθη εισόδου της τάξης των PBs.
Ως αποτέλεσμα, τα τελευταία χρόνια έχουν εμφανιστεί δουλειές που προσπαθούν να κάνουν χρήση του MapReduce προγραμματιστικού μοντέλου για την αναζήτηση και επεξεργασία ακολουθιακών δεδομένων. Συνήθως τέτοιου τύπου αλγόριθμοι αποτελούν μεταγραφές (άλλοτε απλές, άλλοτε λιγότερο απλές) υπαρχόντων αλγορίθμων έτσι ώστε να λειτουργούν σε MapReduce.

Στα πλαίσια της παρούσας διπλωματικής:

- Θα μελετηθεί η βιβλιογραφία προκειμένου να αποτυπωθούν οι τελευταίες εξελίξεις στο πεδίο των MapReduce αλγορίθμων για βιολογικά δεδομένα.
- Θα υλοποιηθούν MapReduce αλγόριθμοι για βιολογικά προβλήματα και θα πραγματοποιηθούν αντίστοιχα πειράματα.

ΣΧΕΤΙΚΟ ΥΛΙΚΟ:

1. http://labs.google.com/papers/mapreduce.html -- Η εργασία που εισήγαγε το MapReduce μοντέλο.
2. http://hadoop.apache.org/ -- Η ανοιχτού κώδικα υλοποίηση του MapReduce.

3. http://www-math.mit.edu/~lippert/18.417/papers/altschuletal1990.pdf -- Ο BLAST αλγόριθμος για βιολογικά δεδομένα.
4. http://www.cs.umd.edu/Grad/scholarlypapers/papers/MichaelSchatz.pdf -- Ένας MapReduce αλγόριθμος που λύνει το ίδιο πρόβλημα με τον BLAST.
5. http://www.personalgenomes.org/ -- Το Personal Genome Project που οδηγεί σε αύξηση των βιολογικών δεδομένων.

ΟΠΤΙΚΟΠΟΙΗΣΗ ΕΞΑΡΤΗΣΕΩΝ ΣΕ ΣΥΣΤΗΜΑΤΑ ΒΑΣΕΩΝ ΔΕΔΟΜΕΝΩΝ

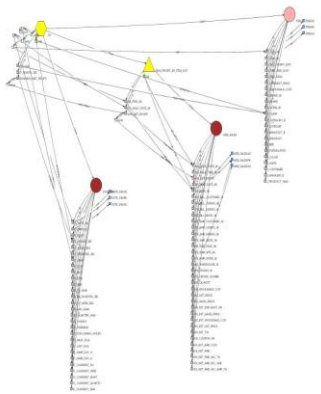
ΠΛΗΡΟΦΟΡΙΕΣ: Γιώργος Παπαστεφανάτος, 210 6990522 (101), gpapas@imis.athena-innovation.gr

ΠΕΡΙΛΗΨΗ: Ο στόχος της διπλωματικής είναι η μελέτη και υλοποίηση τεχνικών και αλγορίθμων οπτικοποίησης για την αναπαράσταση εξαρτήσεων σε συστήματα βάσεων δεδομένων. Οι τεχνικές θα μελετηθούν και θα ενσωματωθούν στο υπάρχον σύστημα «ΕΚΑΤΑΙΟΣ» (http://www.cs.uoi.gr/~pvassil/projects/hecataeus/), το οποίο χρησιμοποιείται για την αναπαράσταση και εξέλιξη πληροφοριακών συστημάτων.

ΑΤΟΜΑ: 1

ΠΛΑΤΦΟΡΜΑ ΕΡΓΑΣΙΑΣ: Java

ΣΥΝΤΟΜΗ ΠΕΡΙΓΡΑΦΗ: Τα σύγχρονα πληροφοριακά συστήματα και ειδικότερα τα συστήματα βάσεων δεδομένων αποτελούν σύνθετα περιβάλλοντα με πολύμορφες εξαρτήσεις μεταξύ των διαφόρων τμημάτων τους. Τυπικά, η δομή μιας σχεσιακής βάσης δεδομένων αποτελείται από ένα σύνολο πινάκων και περιορισμών, όψεων (views) και αποθηκευμένων συναρτήσεων που προσπελαύνουν τους πίνακες αυτούς, triggers, κτλ. Αντίστοιχα, μια σειρά από δομικές μονάδες εκτός του σχήματος της βάσης δεδομένων, όπως είναι, π.χ., Web σελίδες ή φόρμες εισαγωγής δεδομένων κ.α., προσπελαύνουν το σχήμα της βάσης δημιουργώντας με αυτόν τον τρόπο ένα οικοσύστημα με επιπλέον εξαρτήσεις. Η αναπαράσταση και οπτικοποίηση των εξαρτήσεων αυτών με έναν ενιαίο και ομοιόμορφο τρόπο παίζουν σημαντικό ρόλο στη διαχείριση ολόκληρου του οικοσυστήματος και ειδικότερα σε ότι αφορά την εξέλιξή του. Στο υπάρχον σύστημα «ΕΚΑΤΑΙΟΣ», έχουμε ήδη προχωρήσει στην μοντελοποίηση και οπτικοποίηση εξαρτήσεων χρησιμοποιώντας γράφους ως τεχνική μοντελοποίησης, όπως φαίνεται στο παρακάτω σχήμα.



Οι γράφοι αποτελούν ιδανικά εργαλεία αναπαράστασης εξαρτήσεων σε μικρή κλίμακα (π.χ. της τάξης 10 αντικειμένων) αλλά παρουσιάζουν μειονεκτήματα σε σχέση με τη χρηστικότητα τους όταν ο αριθμός των αντικειμένων και εξαρτήσεων γίνεται αρκετά μεγάλος. Σε αυτές τις περιπτώσεις αναζητούνται κατάλληλες τεχνικές τμηματοποίησης και ομαδοποίησης του γράφου που θα επιτρέπουν την αποδοτική και χρηστική απεικόνιση των εξαρτήσεων.

Στα πλαίσια της διπλωματικής θα μελετηθούν και θα υλοποιηθούν (α) τεχνικές αναπαράστασης και οπτικοποίησης εξαρτήσεων με δυνατότητα αλληλεπίδρασης του χρήστη, (β) αλγόριθμοι μετασχηματισμού γράφων, (γ) μετρικές αξιολόγησης του σχεδιασμού του συστήματος που θα βασίζονται σε ιδιότητες του γράφου.

ΕΞΑΓΩΓΗ SQL ΕΡΩΤΗΜΑΤΩΝ ΑΠΟ ΑΡΧΕΙΑ ΠΗΓΑΙΟΥ ΚΩΔΙΚΑ

ΠΛΗΡΟΦΟΡΙΕΣ: Γιώργος Παπαστεφανάτος, 210 6990522 (101), gpapas@imis.athena-innovation.gr

ΠΕΡΙΛΗΨΗ: Ο στόχος της διπλωματικής είναι η υλοποίηση τεχνικών εξαγωγής SQL ερωτημάτων από αρχεία πηγαίου κώδικα (π.χ. Java, PL\SQL, κτλ.) με στόχο την αναπαράσταση των εξαρτήσεων τους με αντικείμενα της βάσης δεδομένων που προσπελαύνουν. Οι τεχνικές θα μελετηθούν και θα ενσωματωθούν στο υπάρχον σύστημα «ΕΚΑΤΑΙΟΣ» (http://www.cs.uoi.gr/~pvassil/projects/hecataeus/) το οποίο χρησιμοποιείται για την αναπαράσταση και εξέλιξη πληροφοριακών συστημάτων.

ΑΤΟΜΑ: 1

ΠΛΑΤΦΟΡΜΑ ΕΡΓΑΣΙΑΣ: Java, Oracle

ΣΥΝΤΟΜΗ ΠΕΡΙΓΡΑΦΗ: Τα σύγχρονα πληροφοριακά συστήματα και ειδικότερα τα συστήματα βάσεων δεδομένων αποτελούν σύνθετα περιβάλλοντα με πολύμορφες εξαρτήσεις μεταξύ των διαφόρων τμημάτων τους. Τυπικά, η δομή μιας σχεσιακής βάσης δεδομένων αποτελείται από ένα σύνολο πινάκων και περιορισμών, όψεων (views) και αποθηκευμένων συναρτήσεων που προσπελαύνουν τους πίνακες αυτούς, triggers, κτλ. Αντίστοιχα, μια σειρά από δομικές μονάδες εκτός του σχήματος της βάσης δεδομένων, όπως είναι, π.χ., Web σελίδες ή φόρμες εισαγωγής δεδομένων κ.α., προσπελαύνουν το σχήμα της βάσης δημιουργώντας με αυτόν τον τρόπο ένα οικοσύστημα με επιπλέον εξαρτήσεις. Η ανίχνευση και μοντελοποίηση των εξαρτήσεων αυτών παίζουν σημαντικό ρόλο στην διαχείριση ολόκληρου του οικοσυστήματος και ειδικότερα σε ότι αφορά την εξέλιξή του. Δυστυχώς, η πληθώρα των εξαρτήσεων βρίσκεται «κρυμμένη» υπό μορφή SQL ερωτημάτων σε πηγαίο κώδικα (Web σελίδες, αντικείμενα στο Data Access Layer, JDBC connections, αποθηκευμένων συναρτήσεων στη βάση, κτλ) από όπου θα πρέπει να ανιχνευθούν και εξαχθούν οι διάφορες συσχετίσεις τους με το σχήμα της βάσης που προσπελαύνουν. Στο υπάρχον σύστημα «ΕΚΑΤΑΙΟΣ» (http://www.cs.uoi.gr/~pvassil/projects/hecataeus/), έχουμε ήδη προχωρήσει στην μοντελοποίηση απλών SQL ερωτημάτων.

Στα πλαίσια της διπλωματικής θα μελετηθούν και θα υλοποιηθούν (α) τεχνικές ανίχνευσης και εξαγωγής SQL ερωτημάτων μέσα από αρχεία πηγαίου κώδικα (Java και PL\SQL) (β) repository για την αποθήκευση, ευρετηρίαση και διαχείριση SQL ερωτημάτων από αρχεία πηγαίου κώδικα (ένα είδος extended dictionary).

ΣΧΕΔΙΑΣΜΟΣ ΚΑΙ ΑΝΑΠΤΥΞΗ ΕΝΟΣ TASK-BASED WIKI

ΠΛΗΡΟΦΟΡΙΕΣ: Τάσος Αρβανίτης, 210 6990522 (205), anarv@imis.athena-innovation.gr

ΠΕΡΙΛΗΨΗ: Στόχος της διπλωματικής εργασίας είναι ο σχεδιασμός και ανάπτυξη ενός wiki portal που θα δίνει στους χρήστες τη δυνατότητα να οργανώσουν τις δραστηριότητές τους γύρω από tasks (εργασίες).

ΑΤΟΜΑ: 1-2

ΠΛΑΤΦΟΡΜΑ ΕΡΓΑΣΙΑΣ: PHP, JavaScript/AJAX, PostgreSQL, Google Maps

ΣΥΝΤΟΜΗ ΠΕΡΙΓΡΑΦΗ: Είναι γεγονός ότι οι άνθρωποι συνηθίζουν να οργανώνουν τις δραστηριότητές τους γύρω από ένα σύνολο επιμέρους εργασιών (problem-solving tasks). Για παράδειγμα, έστω ότι θέλουμε να προετοιμάσουμε το ταξίδι μας σε μια ξένη χώρα. Θα χρειαστεί να συγκεντρώσουμε ένα σύνολο από πληροφορίες οι οποίες συσχετίζονται με δραστηριότητες, όπως για παράδειγμα: τι χρειάζομαι για να βγάλω διαβατήριο ή βίζα, πού πρέπει να αναζητήσω και πώς να αγοράσω εισιτήρια για το ταξίδι μου, ποια είναι τα αξιοθέατα που θα πρέπει να επισκεφτώ, τι αξίζει να αγοράσω κατά τη διαμονή μου κλπ. Για να συγκεντρώσει κανείς τις πληροφορίες αυτές μέσω του διαδικτύου θα πρέπει να καταφύγει σε αναζητήσεις βασισμένες σε λέξεις-κλειδιά (keyword-based) όπως: “Hong Kong” + “what to visit” ή “China” + “visa” + “embassy”. Αφού βρούμε τις πληροφορίες αυτές θα χρειαστεί επιπλέον να εντοπίσουμε πού βρίσκονται στον χάρτη τα αξιοθέατα που αναφέρονται, η πρεσβεία, οι ώρες λειτουργίας κλπ. Τελικά θα πρέπει να συνθέσουμε όλες τις πληροφορίες που συγκεντρώσαμε από διάφορες πηγές για να οργανώσουμε το ταξίδι μας.

Οι μηχανές αναζήτησης οργανώνουν την πληροφορία που υπάρχει διαθέσιμη στο διαδίκτυο γύρω από λέξεις-κλειδιά (keywords). Σε αρκετές περιπτώσεις αυτή η τεχνική έρχεται σε αντίθεση με τον τρόπο που ο άνθρωπος ψάχνει στο διαδίκτυο, αφού συνηθίζουμε να αναζητούμε χρήσιμες πληροφορίες (συνήθως) με κάποιο σκοπό. Για την οργάνωση ενός ταξιδιού για παράδειγμα θα ήταν περισσότερο χρήσιμο ένα ευρετήριο από το οποίο να μπορούμε να επιλέξουμε «διαβατήρια/βίζα->πρεσβείες», «διαβατήρια/βίζα->απαιτούμενα δικαιολογητικά» ή «διαβατήρια/βίζα->ραντεβού και ώρες λειτουργίας πρεσβειών» με συγκεντρωμένο όλο το σχετικό υλικό.

Η στρατηγική οργάνωσης της πληροφορίας με βάση «αυτό που θέλει να κάνει ο χρήστης» (task-based) και όχι με βάση τους όρους που περιέχονται σε μια σελίδα (keyword-based) προωθείται ιδιαίτερα από μεγάλες εταιρίες όπως η Fujitsu [1] και η NTT DoCoMo (ο μεγαλύτερος τηλεπικοινωνιακός πάροχος στην Ιαπωνία) [2], οι οποίες έχουν ήδη αναπτύξει συστήματα «εργασιοστραφούς υπολογιστικής» (Task Computing).

Τα wikis είναι αυτή τη στιγμή ο πιο διαδεδομένος τρόπος οργάνωσης πληροφορίας στο Web 2.0. Επιτρέπουν την προσθήκη και διαμόρφωση περιεχομένου από τους χρήστες με απλό και εύχρηστο τρόπο και είναι αρκετά φιλικά στον τρόπο παρουσίασης της πληροφορίας. Η δημοφιλέστερη τέτοιου τύπου εφαρμογή είναι η online εγκυκλοπαίδεια Wikipedia με πάνω από 12.000.000 λήμματα και περίπου 10.000.000 εγγεγραμμένους χρήστες. Παρόλη τη χρησιμότητά τους, τα wikis βασίζονται επίσης κατά κύριο λόγο σε αναζητήσεις με λέξεις-κλειδιά.

Σκοπός της συγκεκριμένης διπλωματικής είναι η επέκταση ενός wiki software (λογισμικό στο οποίο βασίζεται η λειτουργία ενός wiki) ώστε να μπορεί να διαχειρίζεται περιεχόμενο συσχετισμένο με εργασίες (task annotated content). Για παράδειγμα θα θέλαμε να υποστηρίζονται κάποια από τα παρακάτω:

- Υποστήριξη για επισημείωση (annotation) κειμένου με εργασίες
- Task-based αναζήτηση περιεχομένου και υπηρεσιών
- Ομαδοποίηση (ως quick links ή με κάποια ιεραρχική δομή) σελίδων που αφορούν την ίδια ή σχετικές εργασίες και εμφάνιση των πιο συχνά χρησιμοποιούμενων εργασιών στην αρχική σελίδα
- Συσχέτιση της χωρικής πληροφορίας μιας εργασίας με Google Maps (για παράδειγμα να εμφανίζονται τα προτεινόμενα αξιοθέατα ή διαδρομές στον χάρτη)
- Χρονοπρογραμματισμός (scheduling) των εργασιών. Για παράδειγμα έστω ότι μια εργασία είναι η έκδοση διαβατηρίου. Στην περίπτωση αυτή χρειάζεται να προσκομιστούν μια σειρά από δικαιολογητικά (π.χ. παράβολο χαρτοσήμου από ΔΟΥ, φωτογραφίες, αίτηση στο αστυνομικό τμήμα, φωτοτυπία της αστυνομικής ταυτότητας κ.α.). Η εφαρμογή θα πρέπει να βοηθά τον χρήστη να προγραμματίσει τη σειρά και τον χρόνο εκτέλεσης κάθε εργασίας αξιοποιώντας πληροφορία για τις αλληλεξαρτήσεις μεταξύ των εργασιών, τις αποστάσεις των υπηρεσιών που εκδίδουν τα δικαιολογητικά και τις ώρες λειτουργίας τους. Παράδειγμα μιας υπηρεσίας με παρόμοια λειτουργικότητα είναι το Planjam

Πιο αναλυτικά η εκπόνηση της διπλωματικής εργασίας περιλαμβάνει τα ακόλουθα στάδια:

- Ανάλυση απαιτήσεων από την πλευρά του χρήστη της εφαρμογής
- Σχεδιασμός της βάσης δεδομένων και του διεπιφάνειας χρήστη (user interface). Σε αυτό το σημείο θα χρησιμοποιηθεί το μοντέλο και η γλώσσα περιγραφής εργασιών που περιγράφεται αναλυτικά στο [5]
- Επέκταση του συστήματος MediaWiki ώστε να μπορεί να διαχειρίζεται περιεχόμενο συσχετισμένο με εργασίες, και ανάπτυξη της web εφαρμογής
- Εξαγωγή μιας σειράς από ενδεικτικές εργασίες από την Εθνική Πύλη Δημόσιας Διοίκησης ΕΡΜΗΣ (π.χ. αίτηση χορήγησης αντιγράφου πιστοποιητικού γέννησης, αίτηση έκδοσης διαβατηρίου, κλπ)
- Εισαγωγή δεδομένων από το ΕΡΜΗΣ (π.χ. απαιτούμενα δικαιολογητικά, τρόποι διεκπεραίωσης αιτημάτων, τοποθεσίες γραφείων ΚΕΠ και ώρες λειτουργίας κ.α.) και συσχετισμός αυτών με τις εργασίες/υπηρεσίες που έχουν εξαχθεί

ΣΧΕΤΙΚΟ ΥΛΙΚΟ:

1) http://taskcomputing.org
2) Kurakake, S., Yamazaki, K., Imai, K.: Real-World-Oriented Service Platform Based on the Task Model. IEEE Communications Magazine, September 2006
3) http://www.planjam.com/
4) http://www.mediawiki.org/wiki/MediaWiki
5) Liagouris, J.: ToDo: A Task Modeling Approach for TALOS (θα ανακοινωθεί link)

ΑΝΑΛΥΣΗ ΚΑΙ ΣΧΕΔΙΑΣΗ ΕΝΟΣ ΔΙΑΔΙΚΤΥΑΚΟΥ ΣΥΝΤΑΚΤΗ ΟΝΤΟΛΟΓΙΩΝ

ΠΛΗΡΟΦΟΡΙΕΣ: Γιάννης Λιαγούρης, 210 6990522 (205), liagos@imis.athena-innovation.gr

ΠΕΡΙΛΗΨΗ: Η διπλωματική εργασία στοχεύει στην ανάλυση των απαιτήσεων που πρέπει να ικανοποιεί ένας Συντάκτης Οντολογιών (Ontology Editor) και στην υλοποίησή του με χρήση τεχνολογιών διαδικτύου.

ΑΤΟΜΑ: 1

ΠΛΑΤΦΟΡΜΑ ΕΡΓΑΣΙΑΣ: JavaScript/XML/UML/RDFS/OWL

ΣΥΝΤΟΜΗ ΠΕΡΙΓΡΑΦΗ: Ο όρος **Οντολογία** (Ontology) χρησιμοποιείται σε μια σειρά από επιστημονικά πεδία, συνήθως για διαφορετικούς σκοπούς και κατά συνέπεια με διαφορετική σημασία. Στην περίπτωση που μας ενδιαφέρει, αυτή του **Σημασιολογικού Ιστού** (Semantic Web), οι οντολογίες έχουν εξέχοντα ρόλο επειδή ακριβώς προσφέρουν μια «κοινή αντίληψη» των δεδομένων του διαδικτύου, απαραίτητη προϋπόθεση για τον αυτόματο χειρισμό και ανάλυση των τελευταίων από «έξυπνους» πράκτορες (agents).
Πιο συγκεκριμένα, οι οντολογίες του Σημασιολογικού Ιστού αποτελούνται από δυο βασικά μέρη [1]:

- Ένα λεξιλόγιο (intensional knowledge) που περιλαμβάνει ονόματα εννοιών (concepts) και ιδιοτήτων (properties). Οι έννοιες ή αλλιώς κλάσεις (classes) εκφράζουν σύνολα ατόμων (individuals) που μοιράζονται ένα τουλάχιστον κοινό χαρακτηριστικό, ενώ οι ιδιότητες εκφράζουν συσχετίσεις μεταξύ ατόμων.
- Ένα σύνολο επιπλέον γνώσης (extensional knowledge) σχετικά με τον «κόσμο» που μοντελοποιούμε (domain of interest) το οποίο περιλαμβάνει δηλώσεις ή αλλιώς ισχυρισμούς (assertions). Οι δηλώσεις αντιστοιχίζουν άτομα σε έννοιες και ζεύγη ατόμων σε ιδιότητες. Κατ' αναλογία με την ορολογία του αντικειμενοστρεφούς μοντέλου προγραμματισμού, τα άτομα μιας οντολογίας αναφέρονται συχνά στη βιβλιογραφία και ως στιγμιότυπα (instances) και στην πράξη είναι αναγνωριστικά πόρων (resources) του παγκόσμιου ιστού πληροφοριών.

Δεδομένου του πλήθους της πληροφορίας που υπάρχει στο διαδίκτυο, ένα εξαιρετικά ενδιαφέρον πρόβλημα που παρουσιάζεται αφορά στην ανάπτυξη εργαλείων που θα παρέχουν τη δυνατότητα κατασκευής μεγάλων οντολογιών και θα ικανοποιούν ταυτόχρονα μια σειρά από άλλες απαιτήσεις όπως φιλικότητα προς το χρήστη, διαχείριση εξελισσόμενων πόρων, υποστήριξη συνεργατικού περιβάλλοντος με πολλούς χρήστες κ.λπ.

Το δημοφιλέστερο αυτή τη στιγμή εργαλείο κατασκευής οντολογιών αναπτύσσεται από το πανεπιστήμιο Stanford των ΗΠΑ και ονομάζεται Protégé Ontology Editor [2]. Παρά την ευελιξία που προσφέρει (plug-and-play architecture), λειτουργεί τοπικά ως αυτόνομη Java εφαρμογή. Ωστόσο, όπως προαναφέρθηκε, εξαιτίας του όγκου και της πολυπλοκότητας των πραγματικών οντολογιών του διαδικτύου, πολλές από τις τελευταίες είναι δύσκολο να "φορτωθούν" τοπικά σε ένα συνήθη προσωπικό υπολογιστή με περιορισμένες δυνατότητες (π.χ. 3 GB RAM).

Στην κατεύθυνση επίλυσης του παραπάνω προβλήματος, μια ενδιαφέρουσα προσέγγιση (που πρόσφατα άρχισε να μελετά και η ομάδα ανάπτυξης του Protégé [3]) είναι η χρήση τεχνικών συννέφου (cloud computing), όπου η επεξεργασία και ανάλυση μιας μεγάλης οντολογίας θα παρέχεται ως υπηρεσία

(software-as-service) από απομακρυσμένους εξυπηρετητές. Στόχος της διπλωματικής είναι ο σχεδιασμός μιας τέτοιας αρχιτεκτονικής και η υλοποίηση του αντίστοιχου εργαλείου. Επομένως, η διπλωματική θα ασχοληθεί με τα παρακάτω θέματα:

1. Μελέτη και εξοικείωση με τις επικρατέστερες γλώσσες κατασκευής οντολογιών RDFS και OWL.
2. Μελέτη και εξοικείωση με υπάρχοντα εργαλεία ανάπτυξης, συντήρησης και οπτικοποίησης οντολογιών, όπως ο Protégé Ontology Editor και το OWLSight [4].
3. Ανάλυση απαιτήσεων και σχεδιασμός της αρχιτεκτονικής ενός διαδικτυακού συντάκτη οντολογιών.
4. Υλοποίηση τεχνικών επεξεργασίας οντολογιών με έμφαση στο μοίρασμα του φόρτου εργασίας από τον πελάτη (client) στον εξυπηρετητή (server).

ΣΧΕΤΙΚΟ ΥΛΙΚΟ:

[1] http://www.ksl.stanford.edu/people/dlm/papers/ontologies-come-of-age-mit-press-(with-citation).htm
[2] http://protege.stanford.edu/
[3] Tania Tudorache, Jennifer Vendetti, Natalya F. Noy, Web-Protege: A Lightweight OWL Ontology Editor for the Web. In *The OWL Experiences and Directions Workshop (OWLED)*, Karlsruhe, 2008.
[4] http://pellet.owldl.com/ontology-browser