

Θέμα 1: Εξάγοντας χάρτες από σύννεφα δεδομένων

Θεωρητικό υπόβαθρο: Οι τεχνολογίες εντοπισμού θέσης που συναντώνται σε εφαρμογές πραγματικού χρόνου πλοήγησης ή διαχείρισης στόλου, όχι μόνο επιτρέπουν την παροχή μιας σημαντικής και συστηματικής πηγής δεδομένων παρακολούθησης που δίνει τη δυνατότητα να αντλήσει κανείς πληροφορία σχετικά με μεταφορικά δίκτυα, αλλά εισάγουν επίσης ένα πλήθος ανοιχτών ερευνητικών προκλήσεων σε σχέση με την αξιοποίηση τέτοιων πλούσιων δεδομένων. Στη βιβλιογραφία έχουν προταθεί ορισμένες μέθοδοι, ως μέθοδοι αυτόματης παραγωγής χαρτών, αλλά δεν έχουν μελετηθεί επαρκώς τα προβλήματα που απορρέουν από μεγάλης κλίμακας δεδομένα και γενικευμένης γεωμετρίας μεταφορικά δίκτυα.

Στόχος της διπλωματικής εργασίας είναι: (α) Η ανάπτυξη αλγορίθμου αυτόματης παραγωγής χάρτη και εξαγωγής γνώσης για δίκτυα εναέριας κυκλοφορίας (β) Η αξιοποίηση αυτού του αλγορίθμου πάνω από ένα μεγάλο όγκο δεδομένων σε παράλληλο και κατακεντημένο περιβάλλον.

Τεχνικές λεπτομέρειες: Οι βασικές τεχνικές απαιτήσεις περιλαμβάνουν τη γνώση/εξοικείωση ανάπτυξης εφαρμογών σε Java και προαιρετικά σε MapReduce και QGIS. Κατά την εργασία αυτή θα αναπτυχθεί αλγόριθμος εξαγωγής γνώσης από τροχιές κινούμενων αεροπλάνων με αποτελεσματικό και βέλτιστο τρόπο σε επίπεδο α) χρόνου εκτέλεσης β) ποιότητας αποτελεσμάτων, δηλαδή εάν όντως εξαγάγονται σημαντικά χωρο-χρονικά στοιχεία της πορείας των αεροπλάνων.

Τα καθήκοντα της εργασίας είναι τα παρακάτω:

- Ορισμός των μετρήσιμων χαρακτηριστικών της απόδοσης των αλγορίθμων, δηλαδή τι είναι αυτό που χαρακτηρίζει αποτελεσματικά την απόδοση και την αποτελεσματικότητα ενός αλγορίθμου αυτόματης εξαγωγής χάρτη.
- Συλλογή, ανάλυση, σχεδιασμός και αξιοποίηση μεγάλου όγκου χωρο-χρονικών δεδομένων για την εξόρυξη γεωμετρικών και επιπρόσθετων εξαγόμενων χαρακτηριστικών από τροχιές αεροπλάνων.
- Η δημιουργία ενός εναέριου δικτύου που θα χρησιμοποιηθεί για την αξιολόγηση του αλγορίθμου.

Βιβλιογραφία:

- Map construction: <http://www.mapconstruction.org/>
- Trajectory data mining: <http://research.microsoft.com/apps/pubs/?id=241453>
- Computing with Spatial Trajectories: <https://goo.gl/9rriOM>
- MapReduce: <https://en.wikipedia.org/wiki/MapReduce>
- Apache Flink: <https://flink.apache.org/>

Θέμα 2: Εξάγοντας χάρτες από έναν ωκεανό δεδομένων

Θεωρητικό υπόβαθρο: Οι τεχνολογίες εντοπισμού θέσης που συναντώνται σε εφαρμογές πραγματικού χρόνου πλοήγησης ή διαχείρισης στόλου, όχι μόνο επιτρέπουν την παροχή μιας σημαντικής και συστηματικής πηγής δεδομένων παρακολούθησης που δίνει τη δυνατότητα να αντλήσει κανείς πληροφορία σχετικά με μεταφορικά δίκτυα, αλλά εισάγουν επίσης ένα πλήθος ανοιχτών ερευνητικών προκλήσεων σε σχέση με την αξιοποίηση τέτοιων πλούσιων δεδομένων. Στη βιβλιογραφία έχουν προταθεί ορισμένες μέθοδοι, ως μέθοδοι αυτόματης παραγωγής χαρτών, αλλά δεν έχουν μελετηθεί επαρκώς τα προβλήματα που απορρέουν από μεγάλης κλίμακας δεδομένα και γενικευμένης γεωμετρίας μεταφορικά δίκτυα.

Στόχος της διπλωματικής εργασίας είναι: (α) Η ανάπτυξη αλγορίθμου αυτόματης παραγωγής χάρτη και εξαγωγής γνώσης για θαλάσσια δίκτυα (β) Η αξιοποίηση αυτού του αλγορίθμου πάνω από ένα μεγάλο όγκο δεδομένων σε παράλληλο και κατανεμημένο περιβάλλον.

Τεχνικές λεπτομέρειες: Οι βασικές τεχνικές απαιτήσεις περιλαμβάνουν τη γνώση/εξοικείωση ανάπτυξης εφαρμογών σε Java και προαιρετικά σε MapReduce και QGIS. Κατά την εργασία αυτή θα αναπτυχθεί αλγόριθμος εξαγωγής γνώσης από τροχιές κινούμενων πλοίων με αποτελεσματικό και βέλτιστο τρόπο σε επίπεδο α) χρόνου εκτέλεσης β) ποιότητας αποτελεσμάτων, δηλαδή εάν όντως ανακαλύπτουν σημαντικά χωρο-χρονικά στοιχεία της πορείας των πλοίων.

Τα καθήκοντα της εργασίας είναι τα παρακάτω:

- Ορισμός των μετρήσιμων χαρακτηριστικών της απόδοσης των αλγορίθμων, δηλαδή τι είναι αυτό που χαρακτηρίζει αποτελεσματικά την απόδοση και την αποτελεσματικότητα ενός αλγορίθμου αυτόματης εξαγωγής χάρτη.
- Ανάλυση, σχεδιασμός και αξιοποίηση μεγάλου όγκου χωρο-χρονικών δεδομένων για την εξόρυξη γεωμετρικών και επιπρόσθετων εξαγόμενων χαρακτηριστικών από τροχιές πλοίων.
- Η δημιουργία ενός θαλάσσιου δικτύου που θα χρησιμοποιηθεί για την αξιολόγηση και θα στηρίζεται στο OpenStreetMap ή συμπληρωματικές πηγές.

Βιβλιογραφία:

- Map construction: <http://www.mapconstruction.org/>
- Trajectory data mining: <http://research.microsoft.com/apps/pubs/?id=241453>
- Computing with Spatial Trajectories: <https://goo.gl/9rriOM>
- MapReduce: <https://en.wikipedia.org/wiki/MapReduce>
- Apache Flink: <https://flink.apache.org/>

Θέμα 3: Πρόβλεψη προτύπων κίνησης από σύννεφα δεδομένων

Θεωρητικό υπόβαθρο: Οι τεχνολογίες εντοπισμού θέσης που συναντώνται σε εφαρμογές πραγματικού χρόνου πλοήγησης ή διαχείρισης στόλου, όχι μόνο επιτρέπουν την παροχή μιας σημαντικής και συστηματικής πηγής δεδομένων παρακολούθησης που δίνει τη δυνατότητα να αντλήσει κανείς πληροφορία σχετικά με μεταφορικά δίκτυα, αλλά εισάγουν επίσης ένα πλήθος ανοιχτών ερευνητικών προκλήσεων σε σχέση με την αξιοποίηση τέτοιων πλούσιων δεδομένων. Στη βιβλιογραφία απουσιάζουν οι μέθοδοι βράχου/μάκρο-πρόθεσμης πρόβλεψης προτύπων κίνησης από δεδομένα μεγάλης κλίμακας σε εναέρια δίκτυα κυκλοφορίας.

Στόχος της διπλωματικής εργασίας είναι: (α) Η ανάπτυξη αλγορίθμου για την παραγωγή πληροφορίας/γνώσης από δίκτυα εναέριας κυκλοφορίας και σχετίζεται με ακραίες τιμές, πρότυπα κίνησης καθώς και την πρόβλεψη μελλοντικών θέσεων βράχου/μάκρο-πρόθεσμα (β) Η αξιοποίηση αυτού του αλγορίθμου πάνω από ένα μεγάλο όγκο δεδομένων σε παράλληλο και καταναμημένο περιβάλλον.

Τεχνικές λεπτομέρειες: Οι βασικές τεχνικές απαιτήσεις περιλαμβάνουν τη γνώση/εξοικείωση ανάπτυξης εφαρμογών σε Java και προαιρετικά σε MapReduce και QGIS. Κατά την εργασία αυτή θα αναπτυχθεί αλγόριθμος πρόβλεψης μελλοντικών θέσεων και εξαγωγής γνώσης από τροχιές κινούμενων αεροπλάνων με αποτελεσματικό και βέλτιστο τρόπο σε επίπεδο α) χρόνου εκτέλεσης β) ποιότητας αποτελεσμάτων, δηλαδή εάν όντως προβλέπονται/εξαγάγονται σημαντικά χωρο-χρονικά στοιχεία της πορείας των αεροπλάνων.

Τα καθήκοντα της εργασίας είναι τα παρακάτω:

- Ορισμός των μετρήσιμων χαρακτηριστικών της απόδοσης των αλγορίθμων, δηλαδή τι είναι αυτό που χαρακτηρίζει αποτελεσματικά την απόδοση και την αποτελεσματικότητα ενός αλγόριθμου εξαγωγής γνώσης και πρόβλεψης μελλοντικών θέσεων από τροχιές αεροπλάνων.
- Συλλογή, ανάλυση, σχεδιασμός και αξιοποίηση μεγάλου όγκου χωρο-χρονικών δεδομένων για την εξόρυξη γεωμετρικών και επιπρόσθετων εξαγόμενων χαρακτηριστικών από τροχιές αεροπλάνων.

Βιβλιογραφία:

- Trajectory data mining: <http://research.microsoft.com/apps/pubs/?id=241453>
- Bill Gill, Bob Maddock. "Prediction of Optimal 4D Trajectories in the Presence of Time and Altitude Constraints."
- Aircraft trajectory prediction: <http://goo.gl/9iNiuV>
- Computing with Spatial Trajectories: <https://goo.gl/9rriOM>
- MapReduce: <https://en.wikipedia.org/wiki/MapReduce>
- Apache Flink: <https://flink.apache.org/>

Θέμα 4: Πρόβλεψη προτύπων κίνησης από έναν ωκεανό δεδομένων

Θεωρητικό υπόβαθρο: Οι τεχνολογίες εντοπισμού θέσης που συναντώνται σε εφαρμογές πραγματικού χρόνου πλοήγησης ή διαχείρισης στόλου, όχι μόνο επιτρέπουν την παροχή μιας σημαντικής και συστηματικής πηγής δεδομένων παρακολούθησης που δίνει τη δυνατότητα να αντλήσει κανείς πληροφορία σχετικά με μεταφορικά δίκτυα, αλλά εισάγουν επίσης ένα πλήθος ανοιχτών ερευνητικών προκλήσεων σε σχέση με την αξιοποίηση τέτοιων πλούσιων δεδομένων. Στη βιβλιογραφία απουσιάζουν οι μέθοδοι βράχου/μάκρο-πρόθεσμης πρόβλεψης προτύπων κίνησης από δεδομένα μεγάλης κλίμακας σε δίκτυα κυκλοφορίας πλοίων.

Στόχος της διπλωματικής εργασίας είναι: (α) Η ανάπτυξη αλγορίθμου για την παραγωγή πληροφορίας/γνώσης από δίκτυα πλοίων και σχετίζεται με ακραίες τιμές, πρότυπα κίνησης καθώς και την πρόβλεψη μελλοντικών θέσεων βράχου/μάκρο-πρόθεσμα (β) Η αξιοποίηση αυτού του αλγορίθμου πάνω από ένα μεγάλο όγκο δεδομένων σε παράλληλο και καταναμημένο περιβάλλον.

Τεχνικές λεπτομέρειες: Οι βασικές τεχνικές απαιτήσεις περιλαμβάνουν τη γνώση/εξοικείωση ανάπτυξης εφαρμογών σε Java και προαιρετικά σε MapReduce και QGIS. Κατά την εργασία αυτή θα αναπτυχθεί αλγόριθμος πρόβλεψης μελλοντικών θέσεων και εξαγωγής γνώσης από τροχιές κινούμενων πλοίων με αποτελεσματικό και βέλτιστο τρόπο σε επίπεδο α) χρόνου εκτέλεσης β) ποιότητας αποτελεσμάτων, δηλαδή εάν όντως προβλέπονται/εξαγάγονται σημαντικά χωρο-χρονικά στοιχεία της πορείας των αεροπλάνων.

Τα καθήκοντα της εργασίας είναι τα παρακάτω:

- Ορισμός των μετρήσιμων χαρακτηριστικών της απόδοσης των αλγορίθμων, δηλαδή τι είναι αυτό που χαρακτηρίζει αποτελεσματικά την απόδοση και την αποτελεσματικότητα ενός αλγορίθμου εξαγωγής γνώσης και πρόβλεψης μελλοντικών θέσεων από τροχιές πλοίων.
- Συλλογή, ανάλυση, σχεδιασμός και αξιοποίηση μεγάλου όγκου χωρο-χρονικών δεδομένων για την εξόρυξη γεωμετρικών και επιπρόσθετων εξαγόμενων χαρακτηριστικών από τροχιές πλοίων.

Βιβλιογραφία:

- Trajectory data mining: <http://research.microsoft.com/apps/pubs/?id=241453>
- Perera, Lokukaluge P., and Carlos Guedes Soares. "Ocean Vessel Trajectory Estimation and Prediction Based on Extended Kalman Filter."
- Dimitris Zissis, Elias K. Xidias, Dimitrios Lekkas. "Real-time vessel behavior prediction"
- Computing with Spatial Trajectories: <https://goo.gl/9rriOM>
- MapReduce: <https://en.wikipedia.org/wiki/MapReduce>
- Apache Flink: <https://flink.apache.org/>

Θέμα 5: Παρατηρώντας από μια κλειδαρότρυπα τον κόσμο

Θεωρητικό υπόβαθρο: Οι social media πλατφόρμες, όπως το Twitter, έχουν γίνει ένα πολύ δημοφιλές εργαλείο επικοινωνίας, όπου εκατομμύρια χρήστες μοιράζονται απόψεις σχετικά με διάφορες πτυχές της καθημερινής ζωής. Το Twitter έχει πάνω από 100 εκατομμύρια ενεργούς χρήστες όπου ανταλλάσσουν 500 εκατομμύρια tweets κάθε μέρα. Αυτό ο τεράστιος όγκος δεδομένων και το γεγονός ότι αυτά προσφέρονται σε πραγματικό χρόνο στους χρήστες έχει κάνει τη διαχείριση και την ανάλυσή τους ιδιαίτερα ενδιαφέρουσα. Πολλές ερευνητικές μελέτες έχουν διεξαχθεί, προκειμένου να διαπιστωθεί αν το Twitter μπορεί να δώσει πραγματικά ιδέες για το πώς οι άνθρωποι συμπεριφέρονται. Τέτοιες μελέτες έχουν επικεντρωθεί στην ανάλυση μιας ποικιλίας χωροχρονικών φαινομένων, καθώς και σε θέματα που αφορούν στην ανάλυση συναισθημάτων και των κοινωνικών αλληλεπιδράσεων.

Οι χρήστες του Twitter επιτρέπεται να δημοσιεύσουν ένα μήνυμα 140 χαρακτήρων που ονομάζεται "τιτίβισμα" (tweet). Οι χρήστες μπορούν να έχουν μια σχέση "follow" με άλλους χρήστες, και ένας χρήστης "follower" μπορεί να δει όλα τα tweets του χρήστη που ακολουθεί. Επιπλέον, τα tweets μπορεί να διαδοθούν και να γίνουν δημοφιλή, αν ένας χρήστης θέλει σε συγκεκριμένο μήνυμα να κάνει ένα "re-tweet". Έτσι ένα "τιτίβισμα" (tweet) μπορεί να διαδοθεί όταν συγκεντρώσει πολλούς "followers" μέσω του δικτύου.

Στόχος της διπλωματικής εργασίας είναι: (α) Να μελετηθεί η συμπεριφορά των χρηστών για τα top 3 δημοφιλή θέματα σε διάστημα 2 μηνών (β) Να εξαγάγετε τα δικά σας συμπεράσματα σχετικά με τη διάρκεια/αντίκτυπο/εξέλιξη του θέματος.

Τεχνικές λεπτομέρειες: Χρησιμοποιήστε τη γλώσσα προγραμματισμού της επιλογής σας για να ανιχνεύσετε και να συλλέξετε δεδομένα από το twitter για να διερευνήσετε ποια είναι τα top 3 δημοφιλή θέματα. Θα χρησιμοποιήσετε το Twitter Streaming API (<https://dev.twitter.com/streaming/overview>). Υπάρχουν βιβλιοθήκες που περιέχουν παραμέτρους για τη χρήση του Twitter API, όπως οι ακόλουθες Twitter4j σε Java (<http://twitter4j.org/en/index.html>) και tweepy σε Python. Ένας πληρέστερος κατάλογος για όλες τις βιβλιοθήκες σε όλες τις υποστηριζόμενες γλώσσες προγραμματισμού μπορεί να βρεθεί εδώ: <https://dev.twitter.com/overview/api/twitter-libraries>.

Τα καθήκοντα της εργασίας είναι τα παρακάτω:

- Το Twitter API δέχεται ερωτήματα σε 2 μορφές (και συνδυασμό αυτών): (α) συλλέξτε tweets που ικανοποιούν μια συγκεκριμένη λίστα λέξεων-κλειδιών (β) συλλέξτε tweets που πληρούν συγκεκριμένα γεωγραφικά όρια. Πριν από την αναζήτηση των tweets θα χρειαστεί να κάνετε πρώτα μια έρευνα για να εντοπίσετε τις λέξεις-κλειδιά που θα σας βοηθήσουν να ανακτήσετε τα πιο σχετικά tweets για κάθε θέμα. Οι σχετικές ετικέτες hashtag είναι ένα καλό σημείο εκκίνησης. Στη συνέχεια, θα τρέξετε κάποιον crawler ώστε να συλλέξετε τα δεδομένα. Βεβαιωθείτε ότι συλλέγετε όλες τις πληροφορίες που επιστρέφονται από το Twitter API και όχι μόνο το κείμενο του tweet (βλέπε εδώ: <https://dev.twitter.com/overview/api/tweets>). Κατόπιν, θα χρησιμοποιήσετε κάποια βιβλιοθήκη επεξεργασίας κειμένου για να εξαγάγετε όρους, να απομακρύνετε σημεία στίξης και stopwords. Παραδείγματα βιβλιοθηκών για την επεξεργασία κειμένου είναι τα ακόλουθα: η βιβλιοθήκη NLTK σε Python (<http://www.nltk.org/>) και Apache OpenNLP για Java (<https://opennlp.apache.org/>).
- Θα χρειαστεί να κρατήσετε πληροφορία σχετικά με: tweet_id, user_id, tweet text, coordinates, created_at, place, retweet_count, favorite_count. Φυσικά μπορείτε να χρησιμοποιήσετε οποιαδήποτε άλλη πληροφορία κρίνετε απαραίτητη.
- Θα χρειαστεί να αναλύσετε τα δεδομένα και να εξαγάγετε χρήσιμη γνώση ανακαλύπτοντας πρότυπα και πληροφορίες, βάσει της συχνότητας των περιστατικών/γεγονότων/συμβάντων, του τι κάνουν και πώς αισθάνονται οι άνθρωποι εκεί που συχνάζουν αξιοποιώντας γεωχωρική πληροφορία (geotweets) και εργαλεία ανάλυσης συναισθημάτων, κτλ.