

Προγραμματισμός & Συστήματα στον Παγκόσμιο Ιστό

Εργαστηριακή Άσκηση Ακαδημαϊκού Έτους 2019 - 2020

Θέμα: Σύστημα πληθοπορισμού (crowdsourcing) καταγραφής και αξιοποίησης της δραστηριότητας των χρηστών

Στόχος

Στόχος της παρούσας εργασίας είναι η ανάπτυξη ενός **πλήρους συστήματος συλλογής, διαχείρισης και ανάλυσης πληθοποριστικής (crowdsourced) πληροφορίας**, που αφορά χωροχρονικά (spatiotemporal) δεδομένα ανθρώπινης δραστηριότητας.

Λειτουργικές Προδιαγραφές

Η υπηρεσία Google Maps χρησιμοποιείται στα κινητά εκατομμυρίων χρηστών για να προσφέρει υπηρεσίες πλοήγησης και αναζήτησης χωρικής πληροφορίας. Για τη λειτουργία της, απαιτείται η ενεργοποίηση των λειτουργιών εύρεση τοποθεσίας (μέσω GPS, Wi-Fi ή 3/4G δικτύου). Προαιρετικά, οι ρυθμίσεις του λογαριασμού Google που διατηρεί κάποιος στο κινητό του (ανεξάρτητα από το λειτουργικό του σύστημα) επιτρέπουν την ενεργοποίηση της υπηρεσίας Location History. Με την υπηρεσία αυτή, η Google αποθηκεύει περιοδικά την τρέχουσα τοποθεσία του χρήστη στο cloud, με σκοπό να μπορεί ο χρήστης να βλέπει το ιστορικό των μετακινήσεών του αλλά και για να του προτείνονται σημεία ενδιαφέροντος κοντά στην τοποθεσία του.

Τα στοιχεία που διατηρεί η Google για οποιονδήποτε χρήστη είναι προσβάσιμα μόνο από τον ίδιο. Ο χρήστης μπορεί να μεταφορτώσει τα δεδομένα που τον αφορούν πηγαίνοντας στον ιστότοπο <https://takeout.google.com> και επιλέγοντας τα δεδομένα που αφορούν την υπηρεσία Location History. Τα δεδομένα μπορούν να μεταφορτωθούν σε μορφή JSON ή KMZ (zipped KML).

Τα δεδομένα αυτά έχουν ιδιαίτερη αξία τόσο για τον ίδιο το χρήστη (ανάλυση σε ατομικό επίπεδο) όσο και για την ευρύτερη κοινωνία (ανάλυση σε επίπεδο πληθυσμού). Για το σενάριο της τρέχουσας εργασίας, φανταζόμαστε ότι οι χρήστες θα μπορούσαν να συνεισφέρουν εθελοντικά τα δεδομένα ιστορικού τοποθεσίας τους σε ένα κοινό αποθετήριο μιας έξυπνης πόλης, και να λαμβάνουν μια ανάλυση που περιγράφει πόσο φιλικές προς το περιβάλλον είναι οι μετακινήσεις τους (μετακίνηση με τα πόδια ή με το ποδήλατο). Παράλληλα, οι διαχειριστές του συστήματος μπορούν να βλέπουν σε επίπεδο πληθυσμού τις διαδρομές που επιλέγουν οι χρήστες να κάνουν με τα πόδια ή το ποδήλατο (για να μπορέσουν να σχεδιάσουν καλύτερα τις αστικές υποδομές, π.χ. βελτιώνοντας τα πεζοδρόμια), αλλά και τις περιοχές που συγκεντρώνεται ο πληθυσμός, ώστε σε περίπτωση έκτακτης ανάγκης (π.χ. μεγάλος σεισμός) να είναι πιο αποδοτικός ο σχεδιασμός της αποστολής ασθενοφόρων και πυροσβεστικής.

Στα πλαίσια αυτά, σας έχει ανατεθεί η εκπόνηση ενός συστήματος πληθοπορισμού που μπορεί να λειτουργήσει για την πόλη της Πάτρας. Στο σύστημα υπάρχουν δύο τύποι χρηστών: Διαχειριστής και Χρήστης.

Διαχειριστής

Ο Διαχειριστής αποκτά πρόσβαση στο σύστημα με σταθερό υπολογιστή, μέσω κατάλληλου μηχανισμού username / password. Κατά την είσοδό του στο σύστημα έχει τις εξής δυνατότητες.

1. **Απεικόνιση κατάστασης ΒΔ (dashboard).** Ο διαχειριστής βλέπει σε μία σελίδα κατάλληλη πληροφορία, σε πίνακες ή/και γραφήματα τα οποία απεικονίζουν:

- a. Την κατανομή των δραστηριοτήτων των χρηστών (ποσοστό εγγραφών ανά τύπο δραστηριότητας)
 - b. Την κατανομή του πλήθους εγγραφών ανά χρήστη
 - c. Την κατανομή του πλήθους εγγραφών ανά μήνα
 - d. Την κατανομή του πλήθους εγγραφών ανά ημέρα της εβδομάδας
 - e. Την κατανομή του πλήθους εγγραφών ανά ώρα
 - f. Την κατανομή του πλήθους εγγραφών ανά έτος
2. **Απεικόνιση στοιχείων σε χάρτη.** Επιλέγοντας έτος, μήνα (Ιαν.-Δεκ.), ημέρα (Δευτέρα - Κυριακή), ώρα (0-23) και δραστηριότητα, εμφανίζονται σε οθόνη χάρτη με μορφή heatmap οι πληροφορίες τοποθεσίας των χρηστών. Ο διαχειριστής μπορεί να επιλέξει εύρος τιμών από τις παραπάνω επιλογές, ήτοι έτη (π.χ. «2017-2019»), ημέρα (π.χ. «Σάββατο – Κυριακή»), ώρα (π.χ. «20:00 – 23:00»), δραστηριότητα με πολλαπλή επιλογή (π.χ. «WALKING και STILL»), ή για κάποιο κριτήριο να επιλέξει «όλα».
3. **Διαγραφή δεδομένων.** Επιλέγοντας αυτή τη λειτουργία και κατόπιν επιβεβαίωσης, το σύστημα διαγράφει όλα τα δεδομένα στη βάση.
4. **Εξαγωγή δεδομένων.** Αφού έχει επιλέξει κάποια κριτήρια ερωτήματος για απεικόνιση σε χάρτη, ο διαχειριστής μπορεί να κάνει export τα σχετικά δεδομένα που έχουν επιστραφεί, σε μορφή CSV (comma-separated values), XML ή JSON, για να τα μεταφορτώσει σε τοπικό υπολογιστή. Κάθε εγγραφή περιέχει όλα τα πεδία που αναφέρονται στον πίνακα του παραρτήματος, συν μια επιπλέον στήλη ΣΤΟ ΤΕΛΟΣ που είναι το userid του χρήστη στον οποίο ανήκουν τα δεδομένα.

Χρήστης

Ο χρήστης συνδέεται στο σύστημα μέσω σταθερού υπολογιστή ή smartphone, και έχει τις εξής δυνατότητες:

1. **Εγγραφή στο σύστημα.** Ο χρήστης εγγράφεται και αποκτά πρόσβαση στο σύστημα επιλέγοντας κάποιο username & password της αρεσκείας του, και παρέχοντας το email του. Το password πρέπει να είναι τουλάχιστον 8 χαρακτήρες και να περιέχει τουλάχιστον ένα κεφαλαίο γράμμα, ένα αριθμό και κάποιο σύμβολο (π.χ. # \$ * & @). Το σύστημα δημιουργεί ένα μοναδικό αναγνωριστικό (user id) για τον κάθε εγγεγραμμένο χρήστη με τη χρήση 2-way encryption, χρησιμοποιώντας το email του χρήστη και κλειδί το password του χρήστη. Το password του χρήστη θα πρέπει να αποθηκεύεται στη συνέχεια στη ΒΔ ως hashed τιμή (π.χ. MD5).
2. **Απεικόνιση στοιχείων χρήστη.** Αμέσως μετά την είσοδο στο σύστημα, εμφανίζονται
 - a. Το score οικολογικής μετακίνησης του χρήστη (ποσοστό τοποθεσιών με δραστηριότητα σώματος σε σχέση με όλες τις δραστηριότητες μετακίνησης) για τον τρέχοντα μήνα. Για παράδειγμα, αν έχει 100 τοποθεσίες με δραστηριότητα σώματος και 50 με όχημα, το σκορ του είναι $100/150 = 66\%$. Επίσης εμφανίζεται γράφημα με το score του χρήστη για τους τελευταίους 12 μήνες.
 - b. η περίοδος που καλύπτουν οι εγγραφές του χρήστη,
 - c. η ημερομηνία τελευταίου upload που έκανε ο χρήστης,
 - d. Ένα leaderboard των top 3 χρηστών με συντετμημένο όνομα (π.χ. «Ανδρέας Κ.») για τον τελευταίο μήνα. Το leaderboard αφορά τους χρήστες με την πιο οικολογική μετακίνηση. Στο leaderboard περιέχεται επίσης και ο χρήστης με την κατάταξή του (εκτός των top 3).
3. **Ανάλυση στοιχείων χρήστη.** Ο χρήστης επιλέγει έτος και μήνα (Ιαν.-Δεκ.). Ο χρήστης μπορεί να επιλέξει εύρος τιμών από τις παραπάνω επιλογές, π.χ. έτη «2017-2019», μήνες «Απρίλιος-Σεπτέμβριος». Τα στοιχεία εμφανίζονται με τη μορφή πινάκων ή/και γραφημάτων και αφορούν:

- a. Το ποσοστό εγγραφών ανά είδος δραστηριότητας
- b. Την ώρα της ημέρας με τις περισσότερες εγγραφές ανά είδος δραστηριότητας
- c. Την ημέρα της εβδομάδας με τις περισσότερες εγγραφές ανά είδος δραστηριότητας

Επίσης, εμφανίζεται χάρτης με heatmap που δείχνει τις τοποθεσίες του συγκεκριμένου μόνο χρήστη και για την επιλεγμένη περίοδο.

4. **Upload δεδομένων.** Ο χρήστης μπορεί να επιλέξει κάποιο αρχείο δεδομένων για να το ανεβάσει τη ΒΔ.
- a. Το σύστημα «κόβει» αυτόματα (δεν εισάγει) δεδομένα τα οποία δεν αφορούν την πόλη της Πάτρας (>10km από τις συντεταγμένες 38.230462,21.753150).
 - b. Για την προστασία της ιδιωτικότητάς του, ο χρήστης μπορεί να μην επιθυμεί να ανεβάσει δεδομένα από συγκεκριμένες περιοχές, συνεπώς μπορεί να επιλέξει, πριν την υποβολή, και με τη χρήση οθόνης χάρτη, μια ή περισσότερες περιοχές για τις οποίες δεν επιθυμεί να υποβληθούν στοιχεία (ορίζοντας π.χ. παραλληλόγραμμα περιοχές με click-and-drag).

Περιορισμοί

- 1. Ομάδες 3 (τριών) το πολύ ατόμων.
- 2. Ελεγχόμενη πρόσβαση στα υποσυστήματα που απαιτούν σύνδεση/ αποσύνδεση.
- 3. Οι τεχνολογίες που θα χρησιμοποιηθούν θα είναι από τις διδαχθείσες στο μάθημα. Μπορείτε όμως να χρησιμοποιήσετε **επιπλέον** όποια άλλη τεχνολογία κρίνετε απαραίτητο αρκεί να είναι open-source.
- 4. Η εμφάνιση και η λειτουργικότητα της εφαρμογής αξιολογείται.

Παραδοτέα

- 1. Συνοπτική αναφορά που θα περιλαμβάνει τον σχεδιασμό της βάσης (ER, σχέσεις πινάκων) και τον κωδικό data upload της ομάδας σας.
- 2. Τον πηγαίο κώδικα και ένα export της ΒΔ

Παράρτημα

Δεδομένα Google Location History

Τα πεδία δεδομένων για κάθε εγγραφή περιγράφονται στον παρακάτω πίνακα.

Πεδίο	Περιγραφή
heading	Τρέχουσα κατεύθυνση του χρήστη
activity.type	Ανιχνευόμενη δραστηριότητα του χρήστη
activity.confidence	Βαθμός εμπιστοσύνης ανιχνευόμενης δραστηριότητας (0 – 100)
activity.timestampMs	Timestamp ανιχνευόμενης δραστηριότητας
verticalAccuracy	Ακρίβεια μέτρησης υψόμετρου
velocity	Τρέχουσα ταχύτητα του χρήστη
accuracy	Ακρίβεια τοποθεσίας χρήστη (<800 είναι υψηλή ακρίβεια και >5000 χαμηλή).
longitudeE7	Γεωγραφικό μήκος του χρήστη (διαίρεση με 10^{-7} για να μετατραπεί σε δεκαδικές συντεταγμένες)
latitudeE7	Γεωγραφικό πλάτος του χρήστη (διαίρεση με 10^{-7} για να μετατραπεί σε δεκαδικές συντεταγμένες)
altitude	Υψόμετρο του χρήστη
timestampMs	Timestamp καταγραφής τοποθεσίας

Για κάθε αποθηκευμένη τοποθεσία του χρήστη, διατηρούνται κατ' ελάχιστο οι συντεταγμένες του χρήστη, η ημερομηνία (timestamp) και η ακρίβεια των συντεταγμένων. Για ορισμένες τοποθεσίες, ανάλογα με το αν είναι διαθέσιμη και η σχετική πληροφορία, μπορούν να υπάρχουν επίσης η τρέχουσα ταχύτητα και κατεύθυνση του χρήστη, το υψόμετρο και ακρίβεια υψόμετρου, καθώς και η ανιχνευόμενη δραστηριότητα του χρήστη από τους διάφορους αισθητήρες του κινητού (και ο σχετικός βαθμός εμπιστοσύνης). Η δραστηριότητα του χρήστη μπορεί να είναι κάποια από αυτές που αναφέρονται στο σύνδεσμο <https://developers.google.com/android/reference/com/google/android/gms/location/DetectedActivity#constant-summary>

Συλλογή δεδομένων Location History

- Η κάθε ομάδα θα πρέπει να συλλέξει σε μορφή JSON από την υπηρεσία Google Takeout, ενεργοποιώντας την υπηρεσία location history σε τουλάχιστον ένα κινητό της ομάδας και συλλέγοντας δεδομένα για τουλάχιστον δύο εβδομάδες.
- Εννοείται ότι αν κάποιο μέλος της ομάδας έχει ήδη δεδομένα και από παλαιότερη περίοδο, μπορεί να τα χρησιμοποιήσει.
- Η ενεργοποίηση του location history μπορεί να γίνει από τα κινητά σας ή τον ιστότοπο <https://myaccount.google.com/data-and-personalization>
- Για να ξεκινήσετε την εργασία, σας παρέχεται ένα μικρό dataset σε σύνδεσμο που θα ανακοινωθεί στο e-class. Τα δεδομένα αφορούν ένα χρήστη οπότε θα πρέπει να τα «κόψετε» σε κομμάτια για να εξομοιώσετε την ύπαρξη πολλών χρηστών.

Κοινή χρήση δεδομένων

- Επειδή χρειάζονται πολλά δεδομένα για να μπορέσετε να αξιολογήσετε σωστά τις λειτουργίες της άσκησης, θα πρέπει όλες οι ομάδες να συνεισφέρετε δεδομένα σε ένα κοινό αποθετήριο για όλους.
- Κάθε ομάδα θα πρέπει να δημιουργήσει στο σύστημά της τουλάχιστον 5 χρήστες και για κάθε ένα από αυτούς να περάσει στο σύστημα από ένα αρχείο δεδομένων. Κατά περίπτωση, επειδή το αρχείο δεδομένων που θα επιλέξετε για κάποιο χρήστη μπορεί να είναι αρκετά αραιό, μπορείτε να προσθέσετε κι άλλα αρχεία δεδομένων στον ίδιο χρήστη.

- Μετά τη συλλογή των δικών σας δεδομένων, θα πρέπει υποχρεωτικά να κάνετε upload τα δεδομένα που συλλέξατε στο σύνδεσμο που θα ανακοινωθεί στο e-class.
- Τα δεδομένα που θα κάνετε upload θα πρέπει να είναι σε μορφή zipped JSON (όχι KML).
- Με το upload θα σας εμφανιστεί ένας μοναδικός κωδικός (πιστοποιητικό), τον οποίο θα πρέπει να καταγράψετε και να τον αναφέρετε στην αναφορά του project σας. Μετά το upload, το αρχείο σας θα είναι διαθέσιμο προς χρήση και από τις υπόλοιπες ομάδες. Οι ομάδες που δε θα συνεισφέρουν δεδομένα θα έχουν επίπτωση στο βαθμό.
- Μετά το πέρας της περιόδου συλλογής στοιχείων μπορείτε, αν επιθυμείτε, να διαγράψετε το ιστορικό τοποθεσίας που διατηρεί η Google πηγαίνοντας στο <http://www.google.com/maps/timeline> και επιλέγοντας την κατάλληλη λειτουργία από το εικονίδιο με το γρανάζι.
- Αν επιθυμείτε να ανεβάσετε δεδομένα στο κοινό αποθετήριο αλλά χωρίς κάποιες τοποθεσίες που θεωρείτε ευαίσθητες, μπορείτε να υλοποιήσετε κάποιον parser για να «κόψετε» τα δεδομένα αυτά.

Χρήση τεχνολογιών

- Θα χρησιμοποιήσετε αποκλειστικά open-source τεχνολογίες για τη ΒΔ (MySQL, PostgreSQL), τους χάρτες (Leaflet), τα γραφήματα (προτείνεται η chart.js) και τα heatmaps (<https://leafletjs.com/plugins.html#heatmaps>, προτείνεται η leaflet-heatmap.js).
- Η φόρτωση δεδομένων από τη ΒΔ θα πρέπει να γίνει χρησιμοποιώντας αποκλειστικά τεχνικές AJAX (προσοχή: όχι PHP για τη δημιουργία Javascript κώδικα).
- Τα δεδομένα που κατεβαίνουν από το takeout.google.com μπορεί να είναι αρκετά μεγάλα, ειδικά αν έχετε πολλά έτη χρήσης. Για την επεξεργασία και καταχώρησή τους στη ΒΔ καλό είναι να σκεφτείτε τη χρήση ενός streaming parser (π.χ. salsify, json-machine).