



ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΙΓΑΙΟΥ
ΠΟΛΥΤΕΧΝΙΚΗ ΣΧΟΛΗ

ΤΜΗΜΑ ΜΗΧΑΝΙΚΩΝ ΠΛΗΡΟΦΟΡΙΑΚΩΝ ΚΑΙ ΕΠΙΚΟΙΝΩΝΙΑΚΩΝ ΣΥΣΤΗΜΑΤΩΝ

ΠΡΟΓΡΑΜΜΑ ΠΡΟΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ

**Χρονολόγιο Χαρτών Google: Μελέτη και ανάπτυξη
εφαρμογής για δεδομένα του χρονολογίου**

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

του

Σαλαμούρα Νικόλαου

Επιβλέπων : Αναπλ. Καθηγήτρια Ακριβή Βλάχου, Πανεπιστήμιο Αιγαίου

Μέλη εξεταστικής επιτροπής: Αναπλ. Καθηγητής Κωστούλας Θεόδωρος, Πανεπιστήμιο Αιγαίου
Επικ. Καθηγήτρια Διαμαντοπούλου Βασιλική, Πανεπιστήμιο Αιγαίου

Σάμος, Ιούνιος 2024

Η σελίδα αυτή είναι σκόπιμα λευκή.

Πρόλογος και ευχαριστίες

Η παρούσα Διπλωματική Εργασία εκπονήθηκε κατά το Εαρινό Εξάμηνο του ακαδημαϊκού έτους 2023-2024 στα πλαίσια του Προπτυχιακού Προγράμματος Σπουδών του τμήματος Μηχανικών Πληροφοριακών και Επικοινωνιακών Συστημάτων του Πανεπιστημίου Αιγαίου.

Το κύριο αντικείμενο της παρούσας διπλωματικής είναι η επεξεργασία και η εις βάθος ανάλυση των γεωχωρικών δεδομένων που παρέχονται από τους χρήστες, τα οποία συλλέγονται μέσω της πλατφόρμας Google και πιο συγκεκριμένα μέσω της υπηρεσίας χρονολογίου (Timeline). Η εν λόγω υπηρεσία προσφέρει μια λεπτομερή ανασκόπηση των διαδρομών που έχουν ακολουθήσει οι χρήστες στο παρελθόν. Τα συλλεχθέντα δεδομένα από αυτήν την υπηρεσία αυτή δύνανται να αξιοποιηθούν για την εξαγωγή πολύτιμων συμπερασμάτων σχετικά με τις συνήθειες και τα μοτίβα συμπεριφοράς των χρηστών, μέσω της εφαρμογής εξειδικευμένων αλγορίθμων.

Τα εξαγόμενα συμπεράσματα μπορούν εν συνεχεία να αξιοποιηθούν για την βελτίωση διαφόρων πτυχών της καθημερινότητας στον τομέα της μεταφοράς, όπως την αναβάθμιση του αστικού ιστού, των επιχειρήσεων, όπως την ανάπτυξη εξατομικευμένων διαφημιστικών προσεγγίσεων, ή και του περιβάλλοντος, όπως την παρακολούθηση εξαπλωμένων ασθενειών.

Η ολοκλήρωση της παρούσας εργασίας δεν θα μπορούσε να γίνει εφικτή χωρίς την συνεχή μεταλαμπάδευση γνώσεων της επιβλέπουσας Καθηγήτριας, κα. Βλάχου Ακριβής, την οποία και ευχαριστώ θερμά για την καθοδήγησή της.

Θα ήθελα επίσης να εκφράσω την απέραντη ευγνωμοσύνη στους γονείς μου, οι οποίοι αποτέλεσαν τον ακρογωνιαίο λίθο της ακαδημαϊκής μου πορείας. Η αδιάλειπτη υποστήριξη, η καθοδήγηση και η αστείρευτη αγάπη τους υπήρξαν το σταθερό θεμέλιο πάνω στο οποίο οικοδόμησα της σπουδές μου όλα αυτά τα χρόνια. Χωρίς αυτούς, το ταξίδι αυτό δεν θα ολοκληρωνόταν ποτέ.

© 2024

του/της

Σαλαμούρα Νικόλαου

Τμήμα Μηχανικών Πληροφοριακών και Επικοινωνιακών Συστημάτων

ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΙΓΑΙΟΥ

Η σελίδα αυτή είναι σκόπιμα λευκή.

Πίνακας περιεχομένων

1	Εισαγωγή	1
1.1	Ορισμός προβλήματος.....	1
1.2	Αντικείμενο διπλωματικής.....	1
1.3	Δομή της διπλωματικής	2
2	Θεωρητικό υπόβαθρο	3
2.1	Χωρικά δεδομένα και Χάρτες Google.....	3
2.1.1	Σύντομη ιστορική αναδρομή	3
2.2	Χρονολόγιο της Google	4
2.2.1	JSON Αρχεία	4
2.2.2	Πληροφορίες που περιέχονται στο Google Location History	4
2.3	Επίλυση προβλημάτων τοποθεσίας και διαδρομής	5
2.3.1	Αναγνώριση ακριβούς τοποθεσίας χρήστη	5
2.3.2	Εύρεση συντομότερης διαδρομής προς σημείο	7
2.3.3	Υπολογισμός εκτιμώμενης ώρας άφιξης	9
2.3.4	Μέθοδοι αναγνώρισης μεταφορικού μέσου του χρήστη	10
3	Πρόβλημα που θα μελετηθεί στη διπλωματική.....	12
3.1	Περιγραφή προβλήματος	12
3.2	Σύγκριση Αρχείων Ιστορικού Τοποθεσίας JSON	13
4	Σχετικές Εργασίες	15
5	Πειραματική αποτίμηση της λύσης.....	19
5.1	Δεδομένα εργασίας	19
5.2	Λεπτομέρειες υλοποίησης	19
5.3	Αποτελέσματα	3
6	Συμπεράσματα.....	8
	Βιβλιογραφία.....	10

Λίστα Σχημάτων

Εικόνα 1α: 2-D Trilateration.....	7
Εικόνα 2: Συνάρτηση τοποθέτησης σημείων πλάτους και μήκους σε Πλέγμα.....	20
Εικόνα 3: Συνάρτηση υπολογισμού απόκλισης.....	2
Εικόνα 4: Αποτελέσματα πλεγμάτων εφαρμογής (αριστερά πλέγμα Αυγούστου-δεξιά πλέγμα Δεκέμβρη)..	3
Εικόνα 5: Αποτελέσματα πλεγμάτων εφαρμογής (αριστερά πλέγμα Οκτώβρης-δεξιά πλέγμα Νοέμβρης)...	4
Εικόνα 6: Πλέγμα μεγέθους 40x40.....	5
Εικόνα 7: Εμφάνιση χάρτη εφαρμογής.....	5
Εικόνα 8: Εμφάνιση οπτικοποίησης από το αρχείο (Ρόδος)	6
Εικόνα 9: Οπτικοποίηση αρχείου με διαφορετική τοποθεσία (Άρτα).....	6
Εικόνα 10: Εμφάνιση δισδιάστατου ιστογράμματος.....	7

Ακρωνύμια

ETA	Estimated Time of Arrival
GPS	Global Positioning System
GNSS	Global Navigation Satellite System Network
JSON	JavaScript Object Notation
RMSD	Root Mean Square Deviation

Περίληψη

Η ραγδαία αύξηση στην χρήση των έξυπνων συσκευών, όπως τα έξυπνα τηλέφωνα, οι διασυνδεδεμένες συσκευές κλπ, δημιούργησαν έναν τεράστιο όγκο δεδομένων για επεξεργασία. Εκτός των δεδομένων του χρήστη που σκοπίμως αποθηκεύει ο ίδιος στο νέφος, υπάρχει και μια πληθώρα άλλων δεδομένων που συλλέγει η ίδια η συσκευή μέσω των αισθητήρων της, όπως το επιταχυνσιόμετρο, το γυροσκόπιο και το βαρόμετρο. Πιο συγκεκριμένα, το επιταχυνσιόμετρο σε συνδυασμό με τον αισθητήρα γεωμαγνητικού πεδίου είναι σε θέση να συλλέγουν συνεχώς δεδομένα για την τοποθεσία του χρήστη, τα οποία, κυρίως στις συσκευές Android, αποθηκεύονται και επεξεργάζονται από την Google. Η Google από μεριά της παρέχει έναν αξιόπιστο τρόπο για την εμφάνιση δεδομένων του χρήστη μέσω της υπηρεσίας χρονολογίου, η οποία κάνει μια ανασκόπηση στον χρήστη και στα σημεία τα οποία έχει επισκεφτεί. Ωστόσο η υπηρεσία αυτή παρέχει για την ώρα μειωμένες δυνατότητες, όπως για παράδειγμα την έλλειψη της άμεσης εμφάνισης συχνότητας επίσκεψης ενός σημείου. Στην παρούσα εργασία, παρουσιάζεται μια μεθοδολογία για την επεξεργασία των δεδομένων τοποθεσίας που προκύπτουν από τα αρχεία του χρονολογίου, εμφανίζοντας πιο ολοκληρωμένες πληροφορίες για τις γεωγραφικές προτιμήσεις του χρήστη στο παρελθόν. Επιπλέον, γίνεται σύγκριση μεταξύ δύο αρχείων χρονολογίου, ώστε να μπορούμε να εξάγουμε περισσότερα συμπεράσματα για τις διαφορές μεταξύ της μετακίνησης του χρήστη σε δύο διαφορετικούς μήνες.

Λέξεις Κλειδιά: *γεωχωρικά δεδομένα, χρονολόγιο,*

Abstract

The rapid increase in the usage of smart devices, such as smartphones, connected devices, etc., has created a huge volume of data for processing. Apart from the user data deliberately stored by the user in the cloud, there is also a plethora of other data collected by the device itself through its sensors, such as the accelerometer, gyroscope, and barometer. More specifically, the accelerometer, combined with the geomagnetic field sensor, is capable of continuously collecting data about the user's location, which, mainly on Android devices, is stored and processed by Google. Google, on its part, provides a reliable way to display user data through the Timeline service, which gives an overview of the user and the places they have visited. However, this service currently offers limited capabilities, such as the lack of direct display of visit frequency to a point. This thesis presents a methodology for processing location data derived from Timeline files, providing more comprehensive information about the user's geographical preferences in the past. Additionally, a comparison is made between two Timeline files to extract further insights into the differences in user movement between two different months.

Keywords: *geological data, spatial data, Timeline, gps*

1

Εισαγωγή

1.1 Ορισμός προβλήματος

Τα χωρο-κειμενικά δεδομένα περιγράφονται από γεωγραφικές συντεταγμένες (X,Y). Το χρονολόγιο (Timeline) των χαρτών της Google συλλέγουν από κινητές θέσεις την τοποθεσία τους κατά την διάρκεια του χρόνου, συσχετίζοντας την διαδρομή και την τοποθεσία με ενδιαφέρουσες πληροφορίες. Σκοπός είναι η ανάπτυξη εφαρμογής που να διαχειρίζεται το ιστορικό ενός χρονολογίου. Συγκεκριμένα, η εφαρμογή θα επεξεργάζεται τα δεδομένα ενός χρήστη και να ψάχνει μοτίβα στις διαδρομές του χρήστη, πχ επαναλαμβανόμενες διαδρομές.

1.2 Αντικείμενο διπλωματικής

Ο σκοπός της διπλωματικής βασίζεται σε πολλές πτυχές. Αφενός, το χρονολόγιο μπορεί να βοηθήσει τους χρήστες να δούνε πληροφορίες σχετικά με το ιστορικό των διαδρομών τους και των σημείων που έχουν επισκεφτεί. Αφετέρου, μπορούν εξαχθούν πιο λεπτομερή συμπεράσματα, όπως η κατανόηση των συνηθειών τους ή ακόμα και των διαδρομών που ακολουθούν πιο συχνά. Επιπλέον, η ανάλυση των χρονολογίων μπορεί να βοηθήσει σε τομείς πέραν των συνηθειών των χρηστών. Για παράδειγμα, μπορεί να συμβάλλει σε σημαντικά ευρήματα, όπως είναι η χρήση τους σε περιπτώσεις εγκληματικότητας βοηθώντας στην εξιχνίαση εγκλημάτων. Επίσης, τα συγκεκριμένα δεδομένα μπορούν να εξαγουν πληροφορίες σχετικά με τις προτιμήσεις των χρηστών όσον αφορά τα σημεία που επισκέπτονται, βοηθώντας στην δημιουργία νέων ευκαιριών εργασίας, διαφημιστικών πακέτων κ.λπ.

1.3 Δομή της διπλωματικής

Στο επόμενο κεφάλαιο, γίνεται μια εκτενής περιγραφή του θεωρητικού υποβάθρου της εργασίας. Παρέχεται μια σύντομη ιστορική αναδρομή στους χάρτες της Google και παρουσιάζεται και ο τρόπος λειτουργίας του χρονολογίου. Στην συνέχεια, αναλύεται ο τρόπος υπολογισμού της ακριβούς τοποθεσίας του χρήστη αλλά και ο υπολογισμός της συντομότερης διαδρομής προς κάποιο σημείο. Στο Κεφάλαιο 3, συζητείται το πρόβλημα που προσπαθεί να λύσει η συγκεκριμένη διπλωματική. Παρουσιάζεται επίσης ο τρόπος υπολογισμού της διαφοράς δύο αρχείων χρονολογίων. Το Κεφάλαιο 4 εστιάζει στις σχετικές εργασίες που έχουν αναρτηθεί στην βιβλιογραφία και συζητούν τον τρόπο εκμετάλλευσης της επεξεργασία και ανάλυσης των δεδομένων που προέρχονται από τα αρχεία τοποθεσίας των χρηστών. Στο Κεφάλαιο 5, παρουσιάζονται τα αποτελέσματα της εργασίας, καθώς και η εφαρμογή που υλοποιήθηκε για την ανάλυση και σύγκριση των αρχείων χρονολογίου. Τέλος, στο τελευταίο κεφάλαιο, εξάγονται τα συμπεράσματα της εργασίας και παρουσιάζονται κάποιες μελλοντικές κατευθύνσεις για την βελτίωση και επέκτασή της.

2

Θεωρητικό υπόβαθρο

2.1 Χωρικά δεδομένα και Χάρτες Google

2.1.1 Σύντομη ιστορική αναδρομή

Το χρονολόγιο [1] παρουσιάστηκε από την Google το 2015, καταγράφοντας τις κινήσεις εκατομμυρίων χρηστών κάθεμερινά. Συγκεκριμένα, αυτή η υπηρεσία είναι υπεύθυνη για το ιστορικό των τοποθεσιών που επισκέφτηκε κάθε χρήστης μέσα σε μια μέρα παρουσιάζοντάς ένα μοτίβο διαδρομών. Ο χρήστης της υπηρεσίας μπορεί να δει πολλές πληροφορίες για την προσωπική του ζωή, εξάγοντας χρήσιμα συμπεράσματα για απορίες σχετικά με το που κινήθηκε, τον τρόπο που κινήθηκε αλλά και πόση ώρα διήρκεσε η μετακίνησή του. Φυσικά, για να λειτουργήσει η συγκεκριμένη εφαρμογή ο χρήστης θα πρέπει να έχει ενεργοποιημένη την τοποθεσία του κινητού ή του tablet του. Με άλλα λόγια μέσω των χωρικών δεδομένων που λαμβάνει η συγκεκριμένη εφαρμογή από το στίγμα του χρήστη της μπορεί και ενημερώνεται συνεχώς εμφανίζοντας του πληροφορίες σχετικές με το που πήγε μέσα σε μια μέρα, αν κινήθηκε με αυτοκίνητο, ποδήλατο, τρένο, λεωφορείο ή κάποιο άλλο μέσο. Επίσης, εμφανίζει την συνολική ώρα που διήρκεσε κάθε μετακίνησή του.

Επισημαίνεται ότι το χρονολόγιο συνεχώς εξελίσσεται και μέχρι στιγμής έχουν προστεθεί αρκετές νέες και πρωτότυπες λειτουργίες. Μια από αυτές είναι το Insights [2], το οποίο συγκεντρώνει τις μετακινήσεις των ανθρώπων ανάλογα με τους τρόπους μετακίνησης που χρησιμοποιούν. Το Insights βοηθάει τους χρήστες να παρακολουθούν τα μέρη που έχουν επισκεφτεί παρέχοντας στατιστικά στοιχεία σχετικά με το χρόνο που έχουν αφιερώσει σε καταστήματα, εστιατόρια, μπαρ και άλλα καταστήματα. Επιπλέον, παρέχει ειδοποιήσεις σχετικά με τα πιο σημαντικά στατιστικά, όπως την ημέρα του μήνα του κάθε χρήστη με τις περισσότερες μετακινήσεις. Το Insights μπορεί να παρουσιάσει τα μέρη στα οποία οι χρήστες έχουν ταξιδέψει και το μεταφορικό μέσο το οποίο χρησιμοποίησαν, δηλαδή αν οι χρήστες ήταν με τα πόδια, με ποδήλατο ή με αυτοκίνητο. Παρέχει επίσης ένα ολοκληρωμένο ιστορικό του κάθε ταξιδιού με ενδιαφέρουσες πληροφορίες που αποκαλύπτουν την ημέρα του μήνα με τις περισσότερες μετακινήσεις, το όνομα των αξιοθέατων που έχουν επισκεφτεί, τα ξενοδοχεία στα οποία έμειναν και τα καταστήματα όπου ξοδεύτηκαν τα περισσότερα χρήματα.

Μια άλλη λειτουργία του χρονολογίου της Google είναι το «Google Photos» [3], μια σχετικά νέα λειτουργία της Google που επιτρέπει στους χρήστες να βλέπουν τα μέρη που ταξίδεψαν και τις φωτογραφίες που τραβήχτηκαν κατά τη διάρκεια της ημέρας.

2.2 Χρονολόγιο της Google

2.2.1 JSON Αρχεία

Το JSON (JavaScript Object Notation) είναι μια μορφή αρχείου ανοικτού τύπου για κοινή χρήση δεδομένων που χρησιμοποιεί κείμενο, το οποίο μπορεί να αναγνωστεί από τον χρήστη και χρησιμεύει στην αποθήκευση και στην μετάδοση δεδομένων [4]. Δεδομένου ότι το JSON πλεονεκτεί σε ότι αφορά τη μορφοποίηση, καθώς αυτή είναι συγκριτικά λιγότερη σε σχέση με άλλες ανάλογες μορφές αρχείων, αποτελεί μια καλή εναλλακτική λύση έναντι του XML. Τα αρχεία JSON αναγνωρίζονται μέσω της επέκτασης .json. Παρά το γεγονός ότι προέρχεται από την JavaScript, το JSON είναι μια μορφή δεδομένων που είναι ανεξάρτητη από τη γλώσσα. Το JSON έχει σημαντικά πλεονεκτήματα στη δημιουργία και ανάλυσή του, καθώς υποστηρίζεται από πολλές σύγχρονες γλώσσες προγραμματισμού. Ο τύπος μέσου που χρησιμοποιείται για τα JSON αρχεία είναι το application/json.

Το JSON αποτελεί μια απλή μορφή αρχείου που κατασκευάστηκε από ανθρώπους που είχαν την φιλοσοφία να κρατούν τα πράγματα απλά. Συγκεκριμένα, το JSON είναι μια πολύ καλή λύση για μετατροπή ή μετεγκατάσταση βάσεων δεδομένων. Ένα παράδειγμα χρήσης αυτής της μορφής αρχείου είναι η εξαγωγή δεδομένων από μια εφαρμογή στο διαδίκτυο, η οποία είναι πολύ πιο συνηθισμένη από άλλες μορφές, όπως η CSV και Excel. Αξιοσημείωτο είναι ότι το JSON μπορεί να αναλύεται από κάθε γλώσσα προγραμματισμού και επίσης είναι ένα ανοικτής μορφής αρχείο που μπορεί να αναλυθεί από οποιαδήποτε τεχνολογία είναι σε χρήση [5].

2.2.2 Πληροφορίες που περιέχονται στο Google Location History

Η Google, για να επιτρέψει στους χρήστες να διατηρούν ένα πλήρες ιστορικό των μετακινήσεών τους, αποθηκεύει αναλυτικές πληροφορίες σχετικά με τις τοποθεσίες που επισκέπτεται κάθε χρήστης σε αρχεία JSON. Στην συνέχεια, χρησιμοποιεί αυτά τα αρχεία στην εφαρμογή του χρονολογίου για να εμφανίσει στον χρήστη μια πιο κατανοητή εικόνα των μετακινήσεών του μέσω του χάρτη της Google. Τα δεδομένα τοποθεσίας της Google που είναι αποθηκευμένα σε αρχεία JSON παρέχουν μια ολοκληρωμένη καταγραφή του ιστορικού τοποθεσίας σας, προσφέροντας πληροφορίες για τις προηγούμενες κινήσεις και τις δραστηριότητές των χρηστών.

Αυτά τα αρχεία περιέχουν χρονολογικές πληροφορίες σχετικά με τα μέρη που επισκέφτηκαν οι χρήστες, συμπεριλαμβανομένων χρονικών σημάνσεων, συντεταγμένων γεωγραφικού πλάτους και μήκους και εκτιμήσεων ακρίβειας. Επιπλέον, οι χρήστες μπορούν να βρουν λεπτομέρειες σχετικά με τις δραστηριότητές τους σε συγκεκριμένες τοποθεσίες, δεδομένα υψομέτρου, ταχύτητα συσκευής και την πηγή δεδομένων τοποθεσίας. Η Google επιχειρεί να κατηγοριοποιήσει και να ονομάσει συγκεκριμένα μέρη που επισκέπτονται οι χρήστες και τα γεωγραφικά φράγματα μπορούν να εμφανίζονται όταν οι χρήστες εισέρχονται ή φεύγουν από συγκεκριμένες περιοχές. Μπορεί επίσης να περιλαμβάνονται πληροφορίες

σχετικά με τη συσκευή, όπως το αναγνωριστικό και το μοντέλο της συσκευής. Αυτά τα δεδομένα μπορούν να χρησιμοποιηθούν για τη δημιουργία χρονοδιαγραμμάτων των κινήσεων των χρηστών και είναι πολύτιμα για την τήρηση προσωπικών αρχείων, την ανάλυση ταξιδιών και την έρευνα βάσει τοποθεσίας, αλλά θα πρέπει να λαμβάνονται πάντα υπόψη τα ζητήματα απορρήτου [6]. Στον Πίνακα 1 εμφανίζονται αναλυτικά οι πληροφορίες που εμπεριέχονται στα αρχεία .json τα οποία χρησιμοποιούνται από το χρονολόγιο της Google για την αποθήκευση και διαχείριση των δεδομένων τοποθεσίας του χρήστη.

<i>Timeline objects:</i>	Πίνακας με τα δεδομένα του χρονολογίου
<i>Place visit:</i>	Σύνολο από σημεία διαδρομής που ακολούθησε ο χρήστης
<i>Latitude:</i>	Το γεωγραφικό πλάτος
<i>Longitude:</i>	Το γεωγραφικό μήκος
<i>Place id:</i>	Είναι το id της τοποθεσίας
<i>Address:</i>	Η διεύθυνση του σημείου της τοποθεσίας
<i>Device Tag:</i>	Το id της συσκευής
<i>Location confidence:</i>	Επίπεδο αξιοπιστίας για την επίσκεψη του χρήστη στο σημείο
<i>startTimestamp:</i>	Η χρονοσφραγίδα άφιξης στο σημείο
<i>endTimestamp:</i>	Η χρονοσφραγίδα αναχώρησης από το σημείο

Πίνακας 1: Δεδομένα που εμφανίζονται στο αρχείο χρονολογίου του χρήστη

2.3 Επίλυση προβλημάτων τοποθεσίας και διαδρομής

2.3.1 Αναγνώριση ακριβής τοποθεσίας χρήστη

Για την ακριβή αναγνώριση της τοποθεσίας του χρήστη σε οποιοδήποτε σημείο στον πλανήτη, χρησιμοποιείται μια πληθώρα δορυφόρων του παγκόσμιου συστήματος εντοπισμού θέσης (Global Positioning System – GPS). Μπορεί η ιδέα πίσω από την χρήση του GPS να φαίνεται απλή, ωστόσο οι μαθηματικοί υπολογισμοί που εκτελούνται χρειάζονται τεράστια ακρίβεια [7].

Αρχικά, η αναγνώριση τοποθεσίας μέσω GPS βασίζεται σε δύο βασικά μαθηματικά σενάρια. Το πρώτο ονομάζεται «Μέθοδος ελαχίστων τετραγώνων» (Trilateration), το οποίο πρακτικά υπολογίζει ένα σημείο (π.χ. την τοποθεσία του χρήστη) μέσω της απόστασης του από τρία άλλα σημεία [8]. Το δεύτερο σενάριο αφορά τη σχέση της απόστασης που έχει διανυθεί, της ταχύτητας με την οποία διανύθηκε (ρυθμός) και του χρόνου που χρειάστηκε για να διανυθεί, δηλαδή βάσει του τύπου:

$$\text{Απόσταση} = \text{Ρυθμός} \times \text{Διάρκεια}$$

Το GPS είναι ένας τρόπος παρακολούθησης ο οποίος μπορεί να χρησιμοποιηθεί για στρατιωτικούς αλλά και για πολιτικούς σκοπούς. Το σύστημα παρακολούθησης που χρησιμοποιείται από το Παγκόσμιο Σύστημα Εντοπισμού Θέσης είναι το δίκτυο GNSS (Global Navigation Satellite System Network) [9]. Αυτό το σύστημα όχι μόνο αποθηκεύει την τοποθεσία, αλλά είναι επίσης υπεύθυνο για την διατήρηση της ταχύτητας, της κατεύθυνσης, του χρόνου και άλλων παραμέτρων. Το Παγκόσμιο Σύστημα Εντοπισμού

Θέσης αποτελείται από 27 δορυφόρους GPS σε τροχιά γύρω από τη Γη, εκ των οποίων οι 24 είναι λειτουργικοί και οι υπόλοιποι 3 είναι εφεδρικοί (για χρήση σε περίπτωση βλάβης). Οι δορυφόροι αυτοί περιστρέφονται γύρω από τη Γη κάθε 12 ώρες και εκπέμπουν σήματα πολιτικών και στρατιωτικών δεδομένων πλοήγησης, σε δύο συχνότητες ζώνης L. Αυτά τα σήματα εκπέμπονται στο διάστημα και συλλέγονται από τον δέκτη GPS.

Πέντε σταθμοί παρακολούθησης, μαζί με τέσσερις επίγειες κεραίες, αποτελούν τον έλεγχο του Παγκόσμιου Συστήματος Εντοπισμού. Αυτοί οι σταθμοί βρίσκονται σε όλη την υδρόγειο και συγκεντρώνουν τα δεδομένα σχετικά με τη θέση των δορυφόρων, τα οποία αναμεταδίδονται περαιτέρω στον κύριο σταθμό ελέγχου στην αεροπορική βάση Schriever στο Κολοράντο των ΗΠΑ [7].

Η παθητική παρακολούθηση μέσω GPS περιλαμβάνει την αποθήκευση των δεδομένων που λαμβάνονται κατά την παρακολούθηση, ενώ στην ενεργή παρακολούθηση ορισμένες πληροφορίες αναμεταδίδονται τακτικά μέσω ενός μόντεμ εντός της μονάδας συστήματος GPS (GPS δύο κατευθύνσεων) σε μια κεντρική βάση δεδομένων.

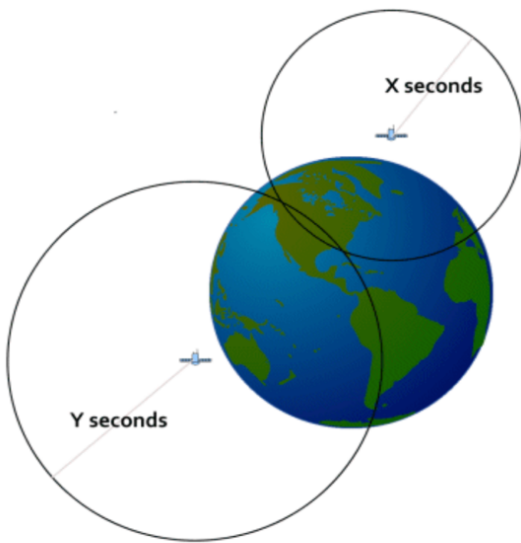
Η παρακολούθηση τοποθεσίας βασίζεται σε μια μαθηματική αρχή που ονομάζεται «Μέθοδος Ελαχίστων Τετραγώνων». Υπάρχουν δύο τύποι: η μέθοδος 2-D, η οποία χρησιμοποιεί δύο δορυφόρους και υπολογίζει προσεγγιστικά την τοποθεσία σε ένα δυσδιάστατο πλαίσιο, και η μέθοδος 3-D, η οποία κάνει χρήση τριών δορυφόρων και υπολογίζει την τοποθεσία του χρήστη με τεράστια ακρίβεια. Αυτό απαιτεί δύο παραμέτρους:

- A) Πρώτον, απαιτείται να είναι γνωστή η τοποθεσία της θέσης που βρίσκεται ο χρήστης, η οποία στη συνέχεια θα εντοπιστεί από τουλάχιστον τρεις δορυφόρους. Η εμβέλεια αυτών των δορυφόρων θα πρέπει να μπορεί να καλύπτει τη συγκεκριμένη θέση.
- B) Στην συνέχεια, η απόσταση μεταξύ ενός αντικειμένου (π.χ. ενός οχήματος) και ενός από τους δορυφόρους θα πρέπει να είναι γνωστή [8].

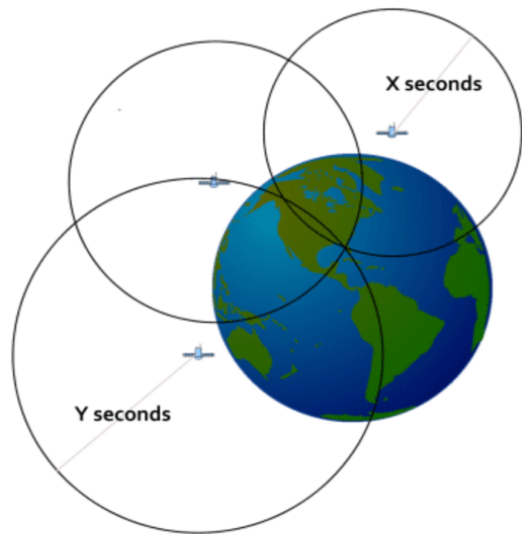
Για το λόγο αυτό, ένας μικροσκοπικός δέκτης/συσκευή εγκαθίσταται στα οχήματα ή στα τηλέφωνα για τη λήψη του ραδιοφώνου (ηλεκτρομαγνητικά σήματα) που εκπέμπονται από τον δορυφόρο. Αυτό βοηθά στον υπολογισμό της απόστασης μεταξύ του οχήματος και του δορυφόρου, καθώς και της θέσης του δορυφόρου.

Γνωρίζοντας λοιπόν την απόσταση που έχουνε αποδώσει οι τρεις δορυφόροι, η θέση του χρήστη μπορεί να μετρηθεί μέσω του «τριγωνισμού» (Triangulation), η οποία επεκτείνει την μαθηματική αρχή της Μεθόδου Ελαχίστων Τετραγώνων. Στην περίπτωση της 3-D μεθόδου ελαχίστων τετραγώνων, χρησιμοποιούνται τρεις σφαίρες, όπως απεικονίζεται στην Εικόνα 1β, ενώ στην μέθοδο 2-D, ο αριθμός των σφαιρών είναι μόλις δύο, όπως απεικονίζεται και στην Εικόνα 1α [7].

Σε αντίθεση με την 2-D μέθοδο, οι ακτίνες της σφαίρας εξαπλώνονται προς όλες τις κατευθύνσεις, σχηματίζοντας έτσι φανταστικές τρισδιάστατες σφαίρες γύρω από ένα δεδομένο σημείο. Η τομή αυτών των τριών σφαιρών, δίνει τη θέση του αντικειμένου. Στη περίπτωση της 3-D μεθόδου, λαμβάνονται δύο σημεία τομής λόγω των τριών σφαιρών [8]. Η Γη βοηθάει εξαλείφοντας ένα από τα σημεία, ενεργώντας ως η τέταρτη σφαίρα. Αυτό ενισχύει την αναζήτηση και καθιστά δυνατό τον εντοπισμό της ακριβούς θέσης του συγκεκριμένου αντικειμένου, τόπου ή προσώπου.



Εικόνα 1α: 2-D Trilateration



Εικόνα 1β: 3-D Trilateration

Πηγή: <http://www.my-car-computer.com/trilateration.html>

Όπως εξηγήθηκε παραπάνω, η μέθοδος ελαχίστων τετραγώνων υπολογίζει την τοποθεσία του χρήστη, μέσω της απόστασής του από το στίγμα τριών δορυφόρων. Ωστόσο, ο υπολογισμός της πραγματικής απόστασης από τον δορυφόρο δεν είναι κάτι το οποίο μπορεί να μετρηθεί άμεσα, αλλά πρέπει να εκτιμηθεί μέσω υπολογισμών. Σε αυτό το σημείο απαραίτητη είναι η εφαρμογή του δεύτερου σεναρίου, όπως αναφέρθηκε παραπάνω, που συνδυάζει την ταχύτητα και την χρόνο που χρειάστηκε να διανυθεί μια απόσταση [7]. Το σημαντικό κομμάτι της εφαρμογής του παραπάνω τύπου είναι το ότι οι δορυφόροι GPS αποστέλλουν συνεχώς ραδιοφωνικά κύματα. Στο σύστημα GPS, ο ρυθμός είναι η ταχύτητα με την οποία ταξιδεύει το ραδιοφωνικό κύμα, που ισούται με την ταχύτητα του φωτός (299.792.458 μέτρα ανά δευτερόλεπτο).

2.3.2 Εύρεση συντομότερης διαδρομής προς σημείο

Οι χάρτες Google κάνουν χρήση γράφων ως δομών δεδομένων για τον υπολογισμό του συντομότερου μονοπατιού από ένα σημείο Α σε ένα σημείο Β. Ένας γράφος αποτελείται από μια σειρά κόμβων και ακμών που συνδέουν αυτούς τους κόμβους μεταξύ τους. Ένας αλγόριθμος που επιτυγχάνει την εύρεση του συντομότερου μονοπατιού προς κάποιον κόμβο, είναι ο αλγόριθμος του Dijkstra. Ο Dijkstra εφαρμόζεται σε έναν βεβαρυμμένο γράφο, στον οποίο κάθε ακμή από έναν κόμβο σε κάποιον άλλον χαρακτηρίζεται από ένα βάρος, δηλαδή την απόσταση που χωρίζει τους δύο κόμβους [7].

- Βήμα 1:** Επιλέγουμε έναν αρχικό κόμβο στον βεβαρυσμένο γράφο ο οποίος παίζει τον ρόλο του «πηγαίου κόμβου».
- Βήμα 2:** Η απόσταση των άλλων κόμβων αρχικοποιείται με άπειρο σχετικά με τον πηγαίο κόμβο.
- Βήμα 3:** Μία άδεια συλλογή N περιέχει τους κόμβους στους οποίους το συντομότερο μονοπάτι έχει βρεθεί.
- Βήμα 4:** Η απόσταση του πηγαίου κόμβου από τον εαυτό του είναι μηδενική. Η απόσταση προστίθεται στην συλλογή N και στην συνέχεια ορίζεται ως «τρέχων κόμβος».
- Βήμα 5:** Εξετάζουμε τους γειτονικούς κόμβους του τρέχων κόμβου που συνδέονται με αυτόν με μια βεβαρυσμένη άκρη. Αθροίζουμε την απόσταση με το βάρος της άκρης.
- Βήμα 6:** Εάν στους γειτονικούς κόμβους έχει ανατεθεί ήδη κάποια απόσταση-βάρος τότε ανανεώνουμε την απόσταση σε αυτή που υπολογίστηκε στο Βήμα 5, σε περίπτωση που αυτή είναι μικρότερη από την τρέχουσα τιμή.
- Βήμα 7:** Επιλέγουμε τον κόμβο με την μικρότερη απόσταση και τον προσθέτουμε στην συλλογή N . Ο νέος κόμβος που προστέθηκε είναι πλέον ο καινούριος τρέχων κόμβος.
- Βήμα 8:** Επαναλαμβάνουμε τα βήματα 5 έως 7 έως ότου ο κόμβος-προορισμός προστεθεί στην συλλογή N ή μέχρι να μην υπάρχουν άλλοι μη-εξερευνημένοι κόμβοι στην N .

Οι χάρτες Google χειρίζονται έναν τεράστιο όγκο δεδομένων και πληροφοριών ώστε να είναι ικανοί να ανακαλύψουν κάθε πιθανή διαδρομή από ένα σημείο της γης σε κάποιο άλλο. Από τα υπερμεγέθη αυτά δεδομένα δημιουργούνται γράφοι οι οποίοι περιέχουν ένα πολύ μεγάλο πλήθος κόμβων, καθιστώντας το πρόβλημα εύρεσης συντομότερου μονοπατιού χρονοβόρο και με τεράστια πολυπλοκότητα, χρονική και χωρική. Σε μια τέτοια περίπτωση, οι δυνατότητες του αλγορίθμου Dijkstra μειώνονται, με τον αλγόριθμο να έχει πολλές πιθανότητες να αποτύχει.

Το παραπάνω πρόβλημα διαχειρίζονται πολύ καλύτερα ευρετικοί αλγόριθμοι, όπως ο A^* . Ο A^* είναι ένας αλγόριθμος παρόμοιος με τον Dijkstra, ο οποίος χρησιμοποιεί μια ευρετική συνάρτηση για να ανακαλύψει το καλύτερο και συντομότερο μονοπάτι. Πιο συγκεκριμένα, δίνει προτεραιότητα σε κόμβους οι οποίοι θεωρούνται καλύτεροι, δηλαδή είναι πιο κοντά θεωρητικά προς τον στόχο ή χρειάζεται λιγότερος χρόνος για να τους επισκεφτούμε κλπ. Αντιθέτως, ο Dijkstra είναι πιο αργός, εφόσον είναι σχεδιασμένος ώστε να εξερευνά όλους τους κόμβους του γράφου.

Ο αλγόριθμος A^* είναι ένας ευέλικτος και πιο αποτελεσματικός αλγόριθμος που χρησιμοποιούν οι χάρτες Google, με σκοπό την εύρεση της συντομότερης διαδρομής μεταξύ δύο κόμβων. Έχει μεγαλύτερο εύρος πλαισίων και μπορεί να διαχειριστεί με μεγαλύτερη ευκολία γράφους που είναι αρκετά μεγάλοι σε όγκο, και για αυτόν τον λόγο προτιμάται περισσότερο σε σχέση με τον Dijkstra. Ο A^* συνδυάζει τα πιο δυνατά μέρη δύο πολύ διαδεδομένων αλγορίθμων, του Dijkstra και του Greedy Best-First. Ο πρώτος ευνοεί ακμές που είναι κοντά στον κόμβο-πηγή, ενώ ο δεύτερος επιλέγει κόμβους που είναι πιο κοντά στον κόμβο προορισμό.

Η βασική διαφορά του A^* με τους υπόλοιπους αλγορίθμους, είναι ότι ο A^* , ως ευρετικός αλγόριθμος αναζήτησης, χρησιμοποιεί δύο συναρτήσεις κόστους:

- Η πρώτη είναι η $g(x)$, η οποία υπολογίζει το πραγματικό κόστος προς κάποιον κόμβο.
- Η δεύτερη είναι η $h(x)$, η οποία βάσει κάποιας ευρετικής συνάρτησης, υπολογίζει το εκτιμώμενο κόστος προς κάποιον κόμβο.

- Η $f(x)$, είναι το άθροισμα των δύο παραπάνω συναρτήσεων ώστε ο αλγόριθμος να αποφασίσει τον επόμενο καλύτερο κόμβο για να επισκεφτεί.

Η ευρετική συνάρτηση αποτελεί πολύ σημαντικό κομμάτι στον αλγόριθμο A^* . Εφόσον ο αλγόριθμος είναι ευρετικός, τότε το κόστος δεν πρέπει να υπερεκτιμάται. Συνεπώς, το κόστος $h(x)$, που αποτελεί μια βασική εκτίμηση του κόμβου x προς τον κόμβο στόχο, δεν θα πρέπει να υπερβαίνει το πραγματικό κόστος της απόστασης. Εάν αυτό συμβεί, ο αλγόριθμος A^* είναι πολύ πιθανό να αποτύχει να δώσει τα επιθυμητά αποτελέσματα.

Στην αναζήτηση με χρήση A^* , ένας κόμβος επεκτείνεται μόνο εάν θεωρείται πολλά υποσχόμενος, δηλαδή έχει μικρό πραγματικό κόστος προς αυτόν και μικρό εκτιμώμενο κόστος προς τον κόμβο στόχο. Αυτός ο αλγόριθμος εστιάζει αποκλειστικά στην επίτευξη του κόμβου στόχου από τον κόμβο που εξετάζεται εκείνη τη στιγμή, χωρίς να προσπαθεί να φτάσει σε άλλους κόμβους. Συνεπώς, εάν η ευρετική συνάρτηση προσεγγίζει το μελλοντικό κόστος με ακρίβεια, τότε χρειάζεται να εξερευνήσουμε πολύ λιγότερους κόμβους σε σύγκριση με τον αλγόριθμο του Dijkstra.

Ο αλγόριθμος A^* τερματίζεται όταν έχει επιτευχθεί ο κόμβος προορισμού, διασφαλίζοντας ότι το συνολικό κόστος για την επίτευξη του κόμβου προορισμού είναι το ελάχιστο σε σύγκριση με όλες τις υπόλοιπες διαδρομές. Μπορεί επίσης να τερματιστεί εάν δεν έχει επιτευχθεί ο κόμβος προορισμού παρά την εξερεύνηση όλων των κόμβων. Σε αντίθεση με τον αλγόριθμο του Dijkstra, ο A^* δεν εξερευνά όλους τους δυνατούς κόμβους, επομένως αποδεικνύεται πιο ικανός και αποδοτικός [7].

Λόγω της υψηλής ακρίβειας, της ευελιξίας, της ικανότητας αντιμετώπισης γραφημάτων με τεράστιο πλήθος κόμβων και ακμών, καθώς και των ιδιαίτερων πλεονεκτημάτων του σε σχέση με τον αλγόριθμο του Dijkstra, οι χάρτες Google έχουν στραφεί στην ανάπτυξη αλγορίθμου A^* για την εύρεση της συντομότερης διαδρομής μεταξύ της πηγής και του προορισμού.

2.3.3 Υπολογισμός εκτιμώμενης ώρας άφιξης

Όπως και σε άλλα παρόμοια προβλήματα, η εκτιμώμενη ώρα άφιξης (Estimated Time of Arrival-ETA), βασίζεται σε πολλές και διαφορετικές παραμέτρους και δεδομένα που είναι διαθέσιμα σε κάποια περιοχή. Τα δεδομένα αυτά μπορούν να ποικίλουν, από τα επίσημα όρια ταχύτητας που εμπλέκονται στην διαδρομή ή την προτεινόμενη ταχύτητα ανάλογα με το είδος του δρόμου (ταχείας κυκλοφορίας, αυτοκινητόδρομος, κλπ.) έως και τους χρόνους ταξιδιού προηγούμενων χρηστών την ίδια ημέρα ή σε άλλες ημέρες σε παρόμοιες ώρες, καθώς και δεδομένα που αφορούν πληροφορίες σχετικά με την τρέχουσα κίνηση στους αντίστοιχους δρόμους της διαδρομής.

Προφανώς, ένας εκτιμώμενος χρόνος άφιξης μπορεί να έχει πάρα πολύ μικρή απόκλιση από την πραγματική ώρα άφιξης, εάν τα δεδομένα που χρησιμοποιήθηκαν είναι επαρκή, είτε όσον αφορά πληροφορίες για τους δρόμους της διαδρομής είτε τα αντίστοιχα δεδομένα άλλων χρηστών που την ακολούθησαν στο παρελθόν [10].

Ωστόσο, παρά το γεγονός ότι οι αλγόριθμοι που χρησιμοποιούνται μπορούν να αποδειχτούν αρκετά ακριβείς, ο υπολογισμός του ETA είναι ένα πρόβλημα που αφορά την πρόβλεψη του μέλλοντος, ενώ η κίνηση στους δρόμους είναι κάτι που είναι σχεδόν αδύνατο να προβλεφθεί. Για παράδειγμα, ακόμα και αν υπάρχει πλήρη ενημέρωση σχετικά με την κατάσταση της κίνησης στους δρόμους την τρέχουσα στιγμή αλλά και τις επόμενες ώρες, για παράδειγμα κάποια έργα στον δρόμο ή την λήξη ενός ποδοσφαιρικού αγώνα, δεν υπάρχει άμεσος τρόπος να προβλεφθεί ένα ατύχημα ή η εμπλοκή ενός αργού φορτηγού στην διαδρομή.

2.3.4 Μέθοδοι αναγνώρισης μεταφορικού μέσου του χρήστη

GPS

Η εφαρμογή χαρτών της Google μπορεί να χρησιμοποιήσει διαφορετικές μεθόδους για να διακρίνει πότε ο χρήστης βρίσκεται μέσα σε ένα αυτοκίνητο ή πότε κινείται με τα πόδια. Μία από τις βασικές μεθόδους που χρησιμοποιεί η Google είναι η χρήση των GPS δεδομένων. Το GPS είναι ένα σύστημα πλοήγησης βασισμένο σε δορυφόρους χάρη στο οποίο η τοποθεσία μιας συσκευής μπορεί να προσδιοριστεί με μεγάλη ακρίβεια. Ωστόσο το GPS δεν έχει την ικανότητα να αναγνωρίσει την ταχύτητα του χρήστη, καθώς αυτό μπορεί να γίνει μόνο μέσω κάποιου αισθητήρα της ίδιας της συσκευής και αυτό μπορεί να γίνει έμμεσα βάσει υπολογισμών.

Η συσκευή χρησιμοποιεί το GPS για να ελέγξει την τοποθεσία της συσκευής ανάμεσα σε δύο χρονικά πλαίσια και στην συνέχεια να αναγνωρίσει την σχετική ταχύτητα, υπολογίζοντας τον χρόνο που χρειάστηκε η συσκευή να μεταβεί από ένα σημείο σε ένα άλλο. Για παράδειγμα, η συσκευή του χρήστη ανακαλύπτει την ακριβή τοποθεσία x_1 την χρονική στιγμή t_1 και στην συνέχεια σε μια δεύτερη χρονική στιγμή t_2 την νέα τοποθεσία του χρήστη x_2 . Το διάστημα στο οποίο αναγνωρίστηκαν οι δύο αυτές τοποθεσίες είναι $t=t_2-t_1$, ενώ η απόσταση x των δυο σημείων υπολογίζεται μέσω του Πυθαγόρειου θεωρήματος. Συνεπώς η μέση ταχύτητα της συσκευής μπορεί εύκολα να εκφραστεί ως το πηλίκο της απόστασης x που διανύθηκε προς το χρονικό διάστημα t .

Επιταχυνσιόμετρο- Γυροσκόπιο

Ωστόσο, τα GPS δεδομένα δεν είναι η μόνη πηγή υπολογισμού της ταχύτητας. Οι χάρτες Google χρησιμοποιούν και άλλους αισθητήρες του τηλεφώνου, όπως το επιταχυνσιόμετρο, το γυροσκόπιο και την πυξίδα του κινητού, χάρη στα οποία μπορεί να εξακριβώσει το μοτίβο κίνησης και την κατεύθυνση του χρήστη.

Το γυροσκόπιο σε ένα κινητό τηλέφωνο είναι μια μικρή συσκευή που χρησιμοποιεί τις αρχές της γωνιακής ορμής για να μετρήσει και να διατηρήσει τον προσανατολισμό της συσκευής. Αποτελείται από έναν περιστρεφόμενο τροχό ή ρότορα που είναι τοποθετημένος σε μια περιστρεφόμενη βάση, τα οποία επιτρέπουν στον τροχό να περιστρέφεται ελεύθερα προς οποιαδήποτε κατεύθυνση. Όταν το κινητό τηλέφωνο περιστρέφεται ή μετακινείται, ο περιστρεφόμενος τροχός υφίσταται μια δύναμη που ονομάζεται γυροσκοπική μετάπτωση. Αυτή η δύναμη προκαλεί τη μετατόπιση του άξονα του τροχού προς την κατεύθυνση της ασκούμενης δύναμης, η οποία με τη σειρά της προκαλεί αλλαγή στον προσανατολισμό της συσκευής. Το γυροσκόπιο είναι χρήσιμο στην αναγνώριση της κατεύθυνσης που είναι τοποθετημένο το κινητό. Το κινητό στους περισσότερους χρήστες θα είναι σε κάθετη κατεύθυνση όταν περπατάει, εάν για παράδειγμα είναι τοποθετημένο στην τσέπη του, ενώ θα έχει μια πιο τυχαία κατεύθυνση όταν οδηγεί.

Από την άλλη, ένα επιταχυνσιόμετρο είναι ένα εργαλείο που μετρά τη σωστή επιτάχυνση. Η σωστή επιτάχυνση ορίζεται ως η επιτάχυνση (ρυθμός μεταβολής της ταχύτητας) ενός σώματος στο δικό του πλαίσιο στιγμιαίας ανάπαυσης. Με πιο απλά λόγια, η μέτρηση σε πραγματικό χρόνο του ρυθμού ενός αντικειμένου πηγαίνει πιο γρήγορα ή πιο αργά. Αυτό έρχεται σε αντίθεση με την επιτάχυνση συντεταγμένων η οποία είναι το μέτρο της επιτάχυνσης σε ένα σταθερό σύστημα συντεταγμένων. Το επιταχυνσιόμετρο ουσιαστικά μετράει αλλαγές στην κίνηση του χρήστη αλλά και την τοποθέτηση του κινητού με την βοήθεια του γυροσκοπίου. Για παράδειγμα, εάν το κινητό είναι τοποθετημένο κάθετα και η ταχύτητα του χρήστη είναι

μικρή, τότε πιθανότατα η συσκευή είναι στην τσέπη του χρήστη ο οποίος μάλλον περπατάει. Επιπλέον αναγνωρίζονται μέσω του επιταχυνσιόμετρου οι δονήσεις που δέχεται η συσκευή. Εάν οι δονήσεις είναι απότομες και μη σταθερές ο χρήστης πιθανότητα βρίσκεται στο αυτοκίνητο, ενώ εάν οι δονήσεις είναι πιο ομαλές και ακολουθούν κάποιο επαναλαμβανόμενο μοτίβο, τότε ο χρήστης πιθανότητα είναι με τα πόδια.

Δεδομένα Δρόμων

Επιπλέον, οι χάρτες Google χρησιμοποιούν και άλλες πληροφορίες για να καταλήξουν με περισσότερη βεβαιότητα στο αν ο χρήστης περπατάει ή βρίσκεται σε κάποιο όχημα, όπως οι πληροφορίες από το Google Street. Οι χάρτες διαθέτουν εκτενείς πληροφορίες σχετικά με τους δρόμους, τα μονοπάτια και τα πεζοδρόμια σε μια περιοχή. Αναγνωρίζοντας το μοτίβο κίνησης του χρήστη, δηλαδή εάν προτιμάει δρόμους αυτοκινήτων είτε πεζοδρόμια, μπορούν να προσδιορίσουν αν ο χρήστης χρησιμοποιεί κάποιο μέσο μεταφοράς ή όχι.

Activity Recognition API

Το Activity Recognition API [11] είναι ένα λογισμικό που χρησιμοποιεί η Google και είναι υλοποιημένο πάνω από τους αισθητήρες που είναι διαθέσιμοι σε μια συσκευή για να παρέχει πληροφορίες στους χάρτες Google σχετικά με το τι κάνουν οι χρήστες σε πραγματικό χρόνο. Χρησιμοποιώντας μοντέλα μηχανικής εκμάθησης, το API διαβάζει περιοδικά δεδομένα από αυτούς τους αισθητήρες σε σύντομες εκρήξεις και τα επεξεργάζεται. Το API αυτό είναι ιδιαίτερα ισχυρό όταν χρησιμοποιείται ως συμπληρωματικό βοήθημα σε επιταχυνσιόμετρα και γυροσκόπια, καθώς μπορεί να ανιχνεύσει μικρές διακυμάνσεις στα δεκάδες σήματα από τους εν λόγω αισθητήρες.

Για παράδειγμα, για τη βελτιστοποίηση πόρων, το API σταματάει κάθε εγγραφή εάν το τηλέφωνό παραμένει ακίνητο για λίγο και χρησιμοποιεί σταδιακά αισθητήρες χαμηλότερης ισχύος για να συνεχίσει την εγγραφή όταν επανέλθει η κίνηση.

3

Πρόβλημα που θα μελετηθεί στη διπλωματική

3.1 Περιγραφή προβλήματος

Το πρόβλημα που θα μελετηθεί στην παρούσα διπλωματική αφορά τη διαχείριση και ανάλυση των δεδομένων που μπορούν να εξαχθούν από το χρονολόγιο της Google για έναν ή και για πολλούς χρήστες. Το χρονολόγιο μπορεί να βοηθήσει τους χρήστες να κάνουν ουσιαστικά μια μικρή «ιστορική αναδρομή» για τα σημεία τα οποία έχουν επισκεφτεί, τον χρόνο παραμονής τους σε αυτά, τις διαδρομές που εκτέλεσαν κλπ.

Η παραπάνω πληροφορία μπορεί να βοηθήσει στην εξαγωγή διάφορων σημαντικών αποτελεσμάτων για τους χρήστες αλλά και για επιχειρήσεις που ενδεχομένως να χρησιμοποιούσαν τα δεδομένα αυτά προς όφελός τους. Πιο συγκεκριμένα, η διαχείριση τέτοιων αρχείων μπορεί να εξάγει σημαντικά συμπεράσματα, όπως τις διαδρομές που ακολουθεί ο χρήστης σε μια καθημερινή βάση, όπως για παράδειγμα την διαδρομή προς την δουλειά του, ή να εντοπίσει μέρη τα οποία επισκέπτεται συχνά. Τα παραπάνω συμπεράσματα θα μπορούσαν να τον βοηθήσουν να βελτιστοποιήσει τον χρόνο διαδρομής που εκτελεί, προτού χρειαστεί να την εκτελέσει. Έτσι για παράδειγμα κάποια εφαρμογή που έχει υλοποιηθεί χρησιμοποιώντας τα δεδομένα του χρονολογίου θα μπορούσε να ειδοποιεί τον χρήστη για κάποια πιθανή κίνηση, λόγω ενός ατυχήματος λόγω χάρη, στην συνηθισμένη διαδρομή που ακολουθεί, ώστε εξαρχής να ακολουθήσει κάποια διαφορετική.

Η χρήση όμως των παραπάνω δεδομένων συνδυαστικά, δηλαδή η συλλογή αρχείων χρονολογίου από περισσότερους χρήστες μια περιοχής ή και πόλης, μπορεί να οδηγήσει σε βελτίωση διάφορων τομέων για το κοινωνικό συμφέρον. Έτσι για παράδειγμα, η συλλογή των δεδομένων και η ανάλυσή τους μπορεί να βελτιστοποιήσει σε τεράστιο βαθμό την κίνηση στους δρόμους. Τα συμπεράσματα από μια τέτοια ανάλυση θα μπορούσαν να εμφανίσουν κοινές διαδρομές που οι χρήστες χρησιμοποιούν και σε ποιες ώρες τις ημέρας, βάσει των οποίων θα ήταν εφικτό να αποφευχθεί η κίνηση στους δρόμους μέσω χρήσης της τροχαίας ή προτείνοντας στους χρήστες εναλλακτικές διαδρομές.

Τα αποτελέσματα της ανάλυσης θα μπορούσαν επίσης να χρησιμοποιηθούν και για τον πολεοδομικό σχεδιασμό, όπου μέρη και διαδρομές που είναι πολυσύχναστες, μπορούν να βοηθήσουν στην κατασκευή καλύτερων δρόμων, πάρκων αναψυχής, πολυκαταστημάτων κλπ.

Από την άλλη πλευρά, οι τοπικές επιχειρήσεις θα μπορούσαν να ενημερωθούν για τις προτιμήσεις των χρηστών στο παρελθόν, και να προωθήσουν στοχευμένες διαφημίσεις στους χρήστες που έχουν δείξει ενδιαφέρον σε παρόμοιες επιλογές. Για παράδειγμα, ένας χρήστης που αγοράζει συνεχώς μπουγάτσες

Θεσσαλονίκης, μπορεί να δεχτεί διαφημίσεις από κάποιον καινούριο τοπικό φούρνο για τα αντίστοιχα προϊόντα που παράγει.

Στο τρέχων κεφάλαιο θα αναλύσουμε τις μεθόδους που μπορούν να χρησιμοποιηθούν για την σωστή και αποδοτική εκμετάλλευση των αρχείων χρονολογίου, συνεισφέροντας στην εξαγωγή χρήσιμων συμπερασμάτων.

3.2 Σύγκριση Αρχείων Ιστορικού Τοποθεσίας JSON

Η μέση τετραγωνική απόκλιση ρίζας (Root Mean Square Deviation-RMSD) είναι ένα στατιστικό μέτρο που χρησιμοποιείται συνήθως για την αξιολόγηση της ομοιότητας μεταξύ δύο συνόλων δεδομένων. Είναι ιδιαίτερα χρήσιμο για τη σύγκριση της καταλληλότητας ενός μοντέλου με πειραματικά δεδομένα.

Το RMSD υπολογίζεται λαμβάνοντας πρώτα τις διαφορές μεταξύ των τιμών των δύο συνόλων δεδομένων, τετραγωνίζοντας αυτές τις διαφορές, λαμβάνοντας τον μέσο όρο των τετραγωνικών διαφορών και, τέλος, λαμβάνοντας την τετραγωνική ρίζα αυτού του μέσου όρου. Μαθηματικά, ο τύπος για το RMSD είναι:

$$\sqrt{\frac{1}{N} \sum (y_i - x_i)^2}$$

όπου N είναι ο αριθμός των σημείων δεδομένων, y_i είναι η πειραματική τιμή, και x_i είναι η τιμή που προβλέπεται από το μοντέλο.

Το RMSD εκφράζεται στις ίδιες μονάδες με τα αρχικά δεδομένα, διευκολύνοντας την ερμηνεία του. Μια μικρότερη τιμή RMSD δείχνει ότι το μοντέλο ταιριάζει καλύτερα στα πειραματικά δεδομένα. Ωστόσο, είναι σημαντικό να σημειωθεί ότι το RMSD επηρεάζεται από το μέγεθος των δεδομένων και, ως εκ τούτου, μπορεί να μην είναι πάντα το καταλληλότερο μέτρο ομοιότητας για όλους τους τύπους δεδομένων.

Το RMSD χρησιμοποιείται ευρέως σε διάφορους τομείς, συμπεριλαμβανομένης της χημείας, της φυσικής, της βιολογίας, της μηχανικής και της οικονομίας. Χρησιμοποιείται συνήθως για την αξιολόγηση της ακρίβειας των υπολογιστικών μοντέλων στην πρόβλεψη της συμπεριφοράς μορίων, υλικών ή συστημάτων. Επιπλέον, το RMSD μπορεί να χρησιμοποιηθεί για τον ποσοτικό προσδιορισμό της μεταβλητότητας των μετρήσεων ή των συνόλων δεδομένων, παρέχοντας πληροφορίες σχετικά με το επίπεδο αβεβαιότητας ή θορύβου στα δεδομένα.

Ακολουθεί η διαδικασία βήμα προς βήμα για τον υπολογισμό της ομοιότητας μεταξύ δύο πινάκων χρησιμοποιώντας την αναλογική συχνότητα κάθε κελιού που χρησιμοποιήσαμε:

1. Για αρχή προσδιορίζουμε τις διαστάσεις των δύο πινάκων: Προτού μπορέσουμε να υπολογίσουμε την ομοιότητα μεταξύ δύο πινάκων, πρέπει να βεβαιωθούμε ότι έχουν τις ίδιες διαστάσεις. Αν δεν ισχύει η ισότητα αυτή, η σύγκριση δεν είναι εφικτή.
2. Υπολογίζουμε τον συνολικό αριθμό στοιχείων σε κάθε πίνακα: Για να προσδιορίσουμε την αναλογική συχνότητα κάθε κελιού, πρέπει να γνωρίζουμε πόσα στοιχεία υπάρχουν σε κάθε πίνακα. Αυτό το

επιτυγχάνουμε αθροίζοντας τις εμφανίσεις που εμπεριέχονται σε κάθε κελί, δημιουργώντας έτσι το S_1 άθροισμα για τον πρώτο πίνακα και το άθροισμα S_2 για τον δεύτερο πίνακα.

3. Υπολογίζουμε τη συχνότητα κάθε κελιού στον πρώτο πίνακα: Για τον υπολογισμό συχνότητας στον πίνακα X , διαιρούμε το πλήθος X_{ij} των σημείων που υπάρχουν ως σύνολο στο εκάστοτε κελί i,j με το συνολικό πλήθος των στοιχείων S_1 που τοποθετήθηκαν στον πίνακα μετά το διάβασμα του αρχείου JSON. Έτσι δημιουργούμε την αναλογική συχνότητα του κάθε κελιού X_{ij}/S_1 . Στο τέλος επαναλαμβάνουμε την ίδια διαδικασία για τον δεύτερο πίνακα Y .

4. Διαφορά αναλογικών συχνοτήτων : Για να προσδιορίσουμε τη διαφορά μεταξύ των δύο πινάκων, αφαιρούμε τη συχνότητα κάθε κελιού στον δεύτερο πίνακα από τη συχνότητα του αντίστοιχου κελιού στον πρώτο πίνακα, για όλα τα κελιά στις γραμμές και τις στήλες του πίνακα. Έτσι για κάθε κελί i,j δημιουργούμε την διαφορά $(X_{ij} - Y_{ij})$.

5. Δημιουργία Τελικού Αθροίσματος: Για να μπορέσουμε να βρούμε την ομοιότητα των πινάκων, θα πρέπει να δημιουργήσουμε το συνολικό άθροισμα του τετραγώνου όλων των διαφορών αναλογικών συχνοτήτων. Ο τύπος που θα χρησιμοποιήσουμε είναι ο:

$$\sqrt{\sum_i^N \sum_j^M (X_{i,j}/S_1 - Y_{i,j}/S_2)^2}$$

4

Σχετικές Εργασίας

Στο άρθρο [12] οι συγγραφείς χρησιμοποίησαν διάφορες μεθόδους στην προσπάθεια να κατανοήσουν την κινητικότητα των χρηστών, χωρίς την χρήση ζωντανών δεδομένων από το GPS, αλλά μόνο με την χρήση των αρχείων καταγραφής με την βοήθεια της εποπτευόμενης μάθησης (supervised learning). Σκοπός του άρθρου ήταν οι διαφοροποίηση των τεσσάρων μεθόδων μεταφοράς που χρησιμοποιούν οι χρήστες, μέσω των παραπάνω δεδομένων, δηλαδή το περπάτημα, την οδήγηση αυτοκινήτου, την οδήγηση ποδηλάτου και την χρήση λεωφορείου.

Οι τέσσερις παραπάνω τρόποι μεταφοράς χαρακτηρίζουν την κινητικότητα ενός χρήστη και μπορούν να υποστηρίξουν μια ποικιλία από διάχτυα υπολογιστική συστήματα, όπως την αναγνώριση ανθρώπινης συμπεριφοράς, την κοινή χρήση πορείας και την έξυπνη σύσταση διαδρομής. Οι συγγραφείς χρησιμοποίησαν δεδομένα από 65 άτομα για μια περίοδο 10 μηνών. Στην συνέχεια, προτείνανε μια μέθοδο τμηματοποίησης η οποία βασίζεται σε σημεία αλλαγής. Η μέθοδος αυτή χωρίζει κάθε τροχιά GPS σε ξεχωριστά τμήματα διαφορετικών τρόπων μεταφοράς. Επιπλέον, για κάθε μέθοδο, προσδιόρισαν ένα σύνολο εξελιγμένων χαρακτηριστικών, όπως τον ρυθμό αλλαγής κατεύθυνσης, τον ρυθμό αλλαγής ταχύτητας και τον ρυθμό διακοπής κίνησης του χρήστη.

Αυτά τα χαρακτηριστικά δεν σχετίζονται αρκετά με την ταχύτητα, επομένως δεν επηρεάζονται από τις διαφορετικές συνθήκες κυκλοφορίας, δηλαδή εάν υπάρχει πολύ κίνηση στους δρόμους ή όχι. Με βάση τη μέθοδο τμηματοποίησης που βασίζεται στα σημεία αλλαγής και το μοντέλο συμπερασμάτων που βασίζεται σε δέντρο αποφάσεων (Decision Trees), οι συγγραφείς παρουσιάζουν μια ακρίβεια συμπερασμάτων της τάξης του 71%. Οι συγγραφείς υποστηρίζουν επίσης πως το σύνολο αυτών των χαρακτηριστικών μπορεί να επεκταθεί σε άλλα διάχτυα υπολογιστικά συστήματα με σκοπό την αναγνώριση της ανθρώπινης συμπεριφοράς και την κατανόηση της κινητικότητας των χρηστών. Τέλος, οι συγγραφείς χρησιμοποίησαν έναν αλγόριθμο μετεπεξεργασίας βασισμένο σε γράφημα για την βελτίωση απόδοσης των συμπερασμάτων τους.

Οι συγγραφείς στο [13] αναφέρουν την σημασία της συλλογής και ανάλυσης δεδομένων των ηλικιωμένων ανθρώπων, οι οποίοι μέχρι το 2030 αναμένεται να αποτελούν περισσότερο από το 16% του πληθυσμού της Μαλαισίας. Καθώς ο πληθυσμός συνεχίζει να αυξάνεται, οι ρυθμοί ζωής αλλάζουν και γίνονται πιο απαιτητικοί, η μελέτη και κατανόηση των καθημερινών μοτίβων μετακίνησης των ηλικιωμένων ανθρώπων θα μπορέσει να παρέχει μια πιο ξεκάθαρη οπτική για τους τρόπους βελτίωσης των υπηρεσιών μεταφοράς και των πολιτικών που παρέχουν βιώσιμες και προσβάσιμες μετακινήσεις για τους ηλικιωμένους,

διατηρώντας και αυξάνοντας έτσι την ευημερία τους. Για το σκοπό της εργασίας αυτής, στρατολογήθηκαν 32 εθελοντές και καταγράφηκαν οι μετακινήσεις κατά των καθημερινών τους δραστηριοτήτων για 14 συνεχόμενες ημέρες. Οι 24 από αυτούς ολοκλήρωσαν το στάδιο συλλογής δεδομένων όπου οι ταξιδιωτικές τους δραστηριότητες καταγράφηκαν με επιτυχία στο χρονοδιάγραμμα των Χαρτών Google και συμπλήρωσαν την ηλεκτρονική φόρμα για 14 συνεχόμενες ημέρες. Ο Πίνακας 2 δείχνει τα δημογραφικά στοιχεία των 24 συμμετεχόντων. Το 71% των ερωτηθέντων ήταν γυναίκες και το 29% ήταν άνδρες.

Age range	Male	Female	Total	Malay	Chinese	Indian
< 60	1	2	3	2	1	0
60 to 64	0	8	8	3	4	1
65 to 69	1	2	3	1	2	0
70 to 75	2	2	4	1	2	1
> 75	3	3	6	2	3	1
Total	7	17	24	9	12	3

Πίνακας 2: Δημογραφικά στοιχεία εθελοντών (πηγή [12])

Ο Πίνακας 2 δείχνει επίσης ότι περισσότερο από το 90% των συμμετεχόντων βγήκαν για να λάβουν υπηρεσίες, να αγοράσουν είδη πρώτης ανάγκης, να κάνουν άλλες (κοινωνικές) δραστηριότητες και να κάνουν ψυχαγωγικές δραστηριότητες. Εν τω μεταξύ, λιγότεροι συμμετέχοντες έβγαιναν για δουλειά (37,6%) και έκαναν θρησκευτικές δραστηριότητες (45,9%). Περισσότεροι συμμετέχοντες επιλέγουν να περπατήσουν (16,7%) και να κάνουν ποδήλατο (4,2%) για ψυχαγωγικές δραστηριότητες σε σύγκριση με άλλους σκοπούς ταξιδιού.

activities	Purpose of Travel					
	Work	Goods	Services	Recreation	Religious	Others
Private Car	29.20%	70.80%	83.30%	41.70%	29.20%	58.30%
Pri- vate Car & Public	4.20%	8.40%	8.40%	0%	0%	21%
Private Car, Cycling & Walking	0%	0%	0%	4.20%	0%	0%
Private Car & Walking	0%	16.70%	8.30%	29.20%	4.20%	16.70%
Walking	4.20%	0%	0%	16.70%	12.50%	0%
Total	37.6%	95.9%	100%	91.8%	45.9%	96%

Πίνακας 3: Τρόπος μεταφοράς για διαφορετικούς σκοπούς ταξιδιού μεταξύ των συμμετεχόντων (πηγή [12])

Στο άρθρο [14] οι συγγραφείς παρουσιάζουν μια μεθοδολογία μέσω της οποίας αναλύουν τα δεδομένα τοποθεσίας των χρηστών και ως αποτέλεσμα εξάγουν τις συνήθειες που ακολουθούν οι χρήστες. Χρησιμοποιούν τα δεδομένα που παρέχονται από το ιστορικό τοποθεσίας της Google και συμπεραίνουν μέσω αυτών τους χρόνους που ο κάθε χρήστης έχει δαπανήσει σε διάφορες τοποθεσίες, καθώς και τα μέσα μεταφοράς που χρησιμοποιεί για να τις επισκεφτεί, ανακαλύπτοντας τις τοποθεσίες όπου ο χρήστης ξοδεύει τον περισσότερο χρόνο του.

Πιο συγκεκριμένα, στο άρθρο εξετάζονται δύο διαφορετικοί τύποι πληροφοριών που μπορούν να εξαχθούν από δεδομένα GPS ενός χρήστη. Αρχικά εξάγονται οι πληροφορίες τοποθεσίας για μέρη που έχει επισκεφτεί ο χρήστης, μαζί με τη χρονοσφραγίδα τους, τη διάρκεια της επίσκεψης και πληροφορίες της πιθανής διαδρομής που πιθανώς ακολούθησαν για να μετακινηθούν από μια τοποθεσία σε άλλη. Από τα αρχεία καταγραφής GPS των χρηστών, ύστερα από μια επεξεργασία εμπλουτισμού και απομόνωσης δεδομένων, εξάγουν ενδιαφέροντα μοτίβα συμπεριφοράς χρηστών, που περιλαμβάνουν σημεία που αυτοί επισκέπτονται συχνά, συχνές διαδρομές που ακολουθούν και προτιμώμενο μέσο μεταφοράς καθώς και χρονικά μοτίβα που ακολουθούν για την εκτέλεση συγκεκριμένων διαδρομών.

Η μέθοδος των συγγραφέων αποτελείται αρχικά από την συλλογή πληροφοριών από το GPS του χρήστη χρησιμοποιώντας δεδομένα από το ιστορικό των Χαρτών Google, τα οποία εισάγονται ως αρχεία KML (Keyhole Markup Language) στο σύστημα. Το KML είναι μια μορφή αρχείου που βασίζεται στο πρότυπο XML και χρησιμοποιεί μια δομή βασισμένη σε ετικέτες με ένθετα στοιχεία και χαρακτηριστικά για την έκφραση γεωγραφικών δεδομένων. Στην συνέχεια, εξάγονται πληροφορίες από τα δεδομένα τοποθεσίας του χρήστη, χρησιμοποιώντας έναν αλγόριθμο ομαδοποίησης με βάση την πυκνότητα, και εντοπίζονται περιοχές με πολύ πυκνά καταγεγραμμένα σημεία αλλά και περιοχές όπου ο χρήστης πέρασε με γρήγορο ρυθμό. Οι συγγραφείς ισχυρίζονται ότι οι πληροφορίες που εξάγονται από την μεθοδολογία αυτή μπορούν να χρησιμοποιηθούν σε πολλούς τύπου εφαρμογών και σε πεδία έρευνας, όπως συστήματα συστάσεων για παροχή προσποιημένων συστάσεων.

Στο άρθρο [15] παρουσιάζεται το πρωτότυπο του GeoLife, το οποίο είναι μια υπηρεσία κοινωνικής δικτύωσης που σκοπεύει να κατανοήσει τις διαδρομές, τις τοποθεσίες και τους χρήστες και να εξορύξει τη συσχέτιση μεταξύ χρηστών και τοποθεσιών όσον αφορά τις τροχιές GPS που δημιουργούνται από τους χρήστες. Οι χρήστες έχουν πρόσβαση σε μια ακολουθία τοποθεσιών στον πραγματικό κόσμο και δημιουργούν πολλές τροχιές με τη μορφή αρχείων καταγραφής GPS. Οι συγγραφείς στοχεύουν να κατανοήσουν μέσω του GeoLife τις συχνότερες διαδρομές που ακολουθούν και τις τοποθεσίες που επισκέπτονται οι χρήστες και να εκτελέσουν τρία βασικά σενάρια εφαρμογών: Το πρώτο σενάριο, η ανταλλαγή εμπειριών ζωής με βάση την τροχιά GPS, εστιάζει στην κατανόηση των τροχιών GPS. Στο δεύτερο σενάριο, τη γενική ταξιδιωτική σύσταση, οι συγγραφείς συμπεραίνουν την ταξιδιωτική εμπειρία ενός χρήστη (κατανόηση των χρηστών) και το επίπεδο ενδιαφέροντος μιας τοποθεσίας (κατανόηση τοποθεσιών) με επαναληπτικό τρόπο. Στην συνέχεια, οι ακολουθίες διαδρομών (συσχέτιση τοποθεσίας) ανιχνεύονται με βάση την συναγόμενη εμπειρία χρήστη και το ενδιαφέρον τοποθεσίας. Οι δραστηριότητες των χρηστών σε μια τοποθεσία καταγράφονται (συσχέτιση τοποθεσίας χρήστη) και ενεργοποιείται μια σύσταση δραστηριότητας-τοποθεσίας. Στο τρίτο σενάριο, στην εξατομικευμένη πρόταση φίλων και τοποθεσίας, γίνεται εκτίμηση της ομοιότητας μεταξύ των χρηστών (συσχέτιση χρήστη) με βάση την ομοιότητα μεταξύ των ιστορικών τοποθεσίας (συσχέτιση τοποθεσίας). Στην συνέχεια, διεξάγεται μια εξατομικευμένη σύσταση η οποία προβλέπει τα ενδιαφέροντα ενός χρήστη σε μια τοποθεσία που δεν έχει

επισκεφτεί (συσχέτιση τοποθεσίας χρήστη) ενσωματώνοντας την ομοιότητα σε ένα μοντέλο συνεργατικού φιλτραρίσματος. Συνολικά, ο χρήστης, η τοποθεσία και η τροχιά έχουν μια σχέση συνεργασίας και αμοιβαίας ενίσχυσης μεταξύ τους.

Το κοινωνικό δίκτυο GeoLife προτείνεται από τους συγγραφείς στο άρθρο [16], με σκοπό την οπτικοποίηση δεδομένων GPS σε ψηφιακούς χάρτες και προβολή των προηγούμενων γεγονότων ζωντανά, την ευρετηρίαση των τροχιών GPS που υποβάλλουν οι χρήστες βάσεις της συμπεριφορά μεταφόρτωσης τους, την αναζήτηση τροχιών GPS σε ψηφιακούς χάρτες χρησιμοποιώντας μια χωρική περιοχή ή/και ένα χρονικό διάστημα ως ερώτημα και την κατανόηση του μοτίβου ζωής των ανθρώπων από ακατέργαστα δεδομένα GPS. Επιπλέον στο άρθρο, προτείνεται μια προσέγγιση που βασίζεται στην εποπτευόμενη μάθηση για αυτόματη εκμάθηση των τρόπων μεταφοράς των χρηστών, όπως περπάτημα, μεταφορά με λεωφορείο, ποδηλασία και οδήγηση από ακατέργαστα δεδομένα GPS.

5

Πειραματική αποτίμηση της λύσης

5.1 Δεδομένα εργασίας

Τα δεδομένα που χρησιμοποιήσαμε για τον έλεγχο της εφαρμογής αφορούσαν 5 διαφορετικούς χρήστες σε 2 διαφορετικές περιόδους του χρόνου. Συγκεκριμένα, η μια περίοδος ήταν η καλοκαιρινή (τον μήνα Αύγουστο), όπου οι χρήστες τείνουν να ταξιδεύουν και να μετακινούνται περισσότερο, επισκεπτόμενοι διάφορους προορισμούς και τοποθεσίες. Αυτό μας επέτρεψε να δοκιμάσουμε την απόδοση του αλγορίθμου μας σε σύνολα δεδομένων με υψηλή κινητικότητα και ποικιλία επισκέψεων.

Η δεύτερη περίοδος που επιλέξαμε αντιπροσώπευαν καταστάσεις όπου οι χρήστες μετακινούνται λιγότερο, όπως για παράδειγμα κατά τη διάρκεια του φθινοπώρου ή του χειμώνα, όταν οι καιρικές συνθήκες και οι καθημερινές δραστηριότητες τείνουν να περιορίζουν τις μετακινήσεις. Αυτά τα σύνολα δεδομένων μας επέτρεψαν να αξιολογήσουμε την απόδοση του αλγορίθμου σε περιπτώσεις με λιγότερη ποικιλία και συχνότητα επισκέψεων.

Χρησιμοποιώντας δεδομένα από περιόδους με διαφορετικά επίπεδα μετακινήσεων, θα μπορούσαμε να αξιολογήσουμε πιο ολοκληρωμένα την απόδοση του αλγορίθμου. Αυτό θα γίνει μέσω των αναπαραστάσεων που παρέχει η εφαρμογή, όπως τα χάρτες θερμότητας (heatmaps) και τα πλέγματα που απεικονίζουν τη συχνότητα εμφάνισης των σημείων. Με περιόδους υψηλής κινητικότητας, όπως το καλοκαίρι, αναμένουμε πιο πυκνά και ποικίλα πρότυπα στις αναπαραστάσεις, ενώ σε περιόδους χαμηλότερης κινητικότητας, τα πρότυπα αναμένεται να είναι πιο αραιά και περιορισμένα. Αυτή η ποικιλία στα δεδομένα θα μας επιτρέψει να αξιολογήσουμε την ικανότητα του αλγορίθμου να ανιχνεύει ομοιότητες και να κανονικοποιεί σωστά τα δεδομένα υπό διαφορετικές συνθήκες κινητικότητας.

5.2 Λεπτομέρειες υλοποίησης

Για τις ανάγκες της εργασίας δημιουργήσαμε έναν αλγόριθμο σε γλώσσα Java, με την χρήση του JDK 21, ο οποίος από ένα αρχείο χρονολογίου, εμφανίζει την συχνότητα την οποία επισκέπτεται ο χρήστης τα σημεία του αρχείου αυτού. Με την είσοδο δύο αντίστοιχων αρχείων χρονολογίου, το πρόγραμμα μπορεί να συγκρίνει την ομοιότητα των αρχείων αναζητώντας πόσα παρόμοια σημεία εμφανίζονται πολλές φορές και στα δύο αρχεία.

Εν ολίγοις, το πρόγραμμα με την είσοδο ενός αρχείου αναγνωρίζει τα σημεία τα οποία έχει επισκεφτεί ο χρήστης και τα τοποθετεί μέσα σε ένα δισδυάστατο πίνακα-πλαίσιο, το μέγεθος του οποίου δηλώνεται από τον χρήστη, μέσω της μεθόδου κανονικοποίησης. Η εφαρμογή ζητάει από τον χρήστη την είσοδο δύο αρχείων τύπου .json και δημιουργεί από αυτά δύο αντίστοιχα πλέγματα. Ως αποτέλεσμα, το πρόγραμμα εμφανίζει τα πλέγματα με την συχνότητα επίσκεψης κάθε κελιού, ενώ παράλληλα εμφανίζει και την ομοιότητα των δυο αυτών πλεγμάτων μέσω της μεθόδου απόκλισης.

Παρακάτω, θα εξηγήσουμε τον τρόπο με τον οποίο λειτουργεί το πρόγραμμά μας:

```

/*συνάρτηση που μετατρέπει λίστα με latitude και longitude, σε 2d grid βάσει
/ συχνότητας επίσκεψης με την μέθοδο της κανονικοποίησης.
*/
public static void addPointsToGrid(int grid[][],
    ArrayList<Long> lati_list,
    ArrayList<Long> long_list) {

    //δημιουργία grid με αρχικοποίηση των τιμών σε 0
    for (int i = 0; i < N; i++) {

        for (int j = 0; j < N; j++) {
            grid[i][j] = 0;
        }
    }

    //εύρεση ελάχιστου και μέγιστου σημείου γεωγραφικού πλάτους και μήκους
    Long maxLat = lati_list.get(0);
    Long minLat = lati_list.get(0);
    Long maxLon = long_list.get(0);
    Long minLon = long_list.get(0);
    for (int i = 0; i < lati_list.size(); i++) {
        if (lati_list.get(i) > maxLat) {
            maxLat = lati_list.get(i);
        }
        if (lati_list.get(i) < minLat) {
            minLat = lati_list.get(i);
        }
        if (long_list.get(i) > maxLon) {
            maxLon = long_list.get(i);
        }
        if (long_list.get(i) < minLon) {
            minLon = long_list.get(i);
        }
    }

    //εύρεση διαφοράς max-min του πλάτους
    Long maxminLat = maxLat - minLat;
    //εύρεση διαφοράς max-min του μήκους
    Long maxminLong = maxLon - minLon;
    int x, y;
    for (int i = 0; i < lati_list.size(); i++) {
        x = (int) ((lati_list.get(i) - minLat) * (N - 1) / maxminLat);
        y = (int) ((long_list.get(i) - minLon) * (N - 1) / maxminLong);
        //αύξηση συχνότητας του σημείου στο πλέγμα
        grid[x][y] += 1;
    }
}

```

Εικόνα 2: Συνάρτηση τοποθέτησης σημείων πλάτους και μήκους σε Πλέγμα

Η συνάρτηση `addPointsToGrid`, όπως εμφανίζεται στην Εικόνα 2, εκτελεί πολλές εργασίες που σχετίζονται με την επεξεργασία ενός πλέγματος με βάση δεδομένα γεωγραφικού πλάτους και μήκους. Ακολουθεί μια ανάλυση του τρόπου λειτουργίας:

1. **Αρχικοποίηση πίνακα grid:**

Λαμβάνει ως παράμετρο ένα πλέγμα πίνακα 2D και μηδενίζει όλα τα στοιχεία του. Το μέγεθος του πλέγματος ορίζεται από μια σταθερά `N`.

2. **Εύρεση μέγιστου και ελάχιστου γεωγραφικού πλάτους και μήκους:**

Η συνάρτηση υπολογίζει τις μέγιστες και ελάχιστες τιμές για το γεωγραφικό πλάτος και το μήκος με βάση τις λίστες `lati_list` και `long_list`. Αρχικοποιεί τα `maxLat`, `minLat`, `maxLon` και `minLon` με τις πρώτες τιμές από αυτές τις λίστες και στη συνέχεια επαναλαμβάνει τις λίστες για να βρει τις πραγματικές μέγιστες και ελάχιστες τιμές.

3. **Υπολογισμός διαφορών γεωγραφικού πλάτους και μήκους:**

Υπολογίζει τις διαφορές μεταξύ του μέγιστου και του ελάχιστου γεωγραφικού πλάτους (`maxminLat`) και του γεωγραφικού μήκους (`maxminLong`). Αυτές οι διαφορές θα χρησιμοποιηθούν για τη χαρτογράφηση των τιμών γεωγραφικού πλάτους και μήκους στο πλέγμα.

4. **Τοποθέτηση σημείων στο grid με χρήση κανονικοποίησης:**

Επαναλαμβάνεται ξανά μέσω του `lati_list` και του `long_list` και αντιστοιχίζει κάθε τιμή γεωγραφικού πλάτους και μήκους σε ένα αντίστοιχο κελί στο πλέγμα. Αυτή η αντιστοίχιση περιλαμβάνει τη διαίρεση της διαφοράς μεταξύ της τιμής και της ελάχιστης τιμής με τη διαφορά μεταξύ της μέγιστης και της ελάχιστης τιμής, κλιμακώνοντας αποτελεσματικά τις τιμές ώστε να ταιριάζουν στις διαστάσεις του πλέγματος.

5. **Αύξηση συχνότητας πλέγματος:**

Για κάθε σημείο, αυξάνει την τιμή στο αντίστοιχο κελί του πλέγματος κατά 1. Με αυτόν τον τρόπο καταγράφεται η συχνότητα εμφάνισης κάθε σημείου.

Συνοπτικά, αυτή η συνάρτηση λαμβάνει δεδομένα γεωγραφικού πλάτους και μήκους, υπολογίζει τις μέγιστες και ελάχιστες τιμές, αντιστοιχίζει αυτές τις τιμές σε ένα πλέγμα και παρακολουθεί τη συχνότητα των σημείων σε κάθε κελί του πλέγματος. Το πλέγμα είναι ουσιαστικά ένα διδιάστατο ιστόγραμμα που αντιπροσωπεύει την κατανομή των σημείων με βάση το γεωγραφικό πλάτος και το μήκος τους.

Η συνάρτηση `findDeviation`, όπως φαίνεται και στην Εικόνα 3, υπολογίζει την απόκλιση μεταξύ δύο πλεγμάτων που αντιπροσωπεύονται από δύο πλέγματα (`πλέγμα1` και `πλέγμα2`). Η απόκλιση υπολογίζεται ως μέτρο της διαφοράς μεταξύ των δύο πλεγμάτων. Ακολουθεί μια εξήγηση βήμα προς βήμα για την παραπάνω λειτουργία:

```

public static void findDeviation(int grid1[], int grid2[][]) {
    //εύρεση συνολικής συχνότητας σημείων του κάθε πλέγματος
    int sunolikaSimeia1 = 0;
    int sunolikaSimeia2 = 0;
    for (int i = 0; i < N; i++) {
        for (int j = 0; j < N; j++) {
            sunolikaSimeia1 += grid1[i][j];
            sunolikaSimeia2 += grid2[i][j];
        }
    }

    double sum = 0;
    for (int i = 0; i < N; i++) {
        for (int j = 0; j < N; j++) {
            //δημιουργία συνολικού αθροίσματος του τετραγώνου της κάθε συχνότητας
            //του πρώτου πλέγματος προς το πλήθος συνολικών σημείων του μείων
            //της συχνότητας του ίδιου σημείου του δεύτερου πλέγματος προς το
            //συνολικό πλήθος
            sum += Math.pow(
                grid1[i][j]/(double)sunolikaSimeia1
                - grid2[i][j]/(double)sunolikaSimeia2
                , 2);
        }
    }
    double deviation = Math.sqrt(sum);
    System.out.println("Deviation = " + deviation);
}

```

Εικόνα 3: Συνάρτηση υπολογισμού απόκλισης

1. Υπολογισμός του συνολικού αριθμού σημείων σε κάθε πλέγμα:

Αρχικοποιεί δύο μεταβλητές, numOfPoints1 και numOfPoints2, στο μηδέν. Στη συνέχεια, ο υπολογισμός επαναλαμβάνεται τόσο για το πλέγμα1 όσο και για το πλέγμα2, αθροίζοντας τις τιμές σε όλα τα κελιά. Αυτά τα αθροίσματα αντιπροσωπεύουν τον συνολικό αριθμό πόντων σε κάθε πλέγμα.

2. Υπολογισμός Απόκλισης:

Αρχικοποιούμε ένα μεταβλητό άθροισμα στο μηδέν. Αυτή η μεταβλητή θα αποθηκεύει τις τετραγωνικές διαφορές μεταξύ των αντίστοιχων κελιών στα δύο πλέγματα. Για κάθε κελί, υπολογίζουμε την απόκλιση ως εξής:

- Υπολογίζουμε τον λόγο της τιμής στο κελί του πλέγματος διαιρεμένος με τον συνολικό αριθμό σημείων στο grid1 και τον λόγο της τιμής στο αντίστοιχο κελί του grid2 διαιρεμένος με τον συνολικό αριθμό σημείων στο grid2. Αυτό κανονικοποιεί τις τιμές σε κάθε πλέγμα.
- Στην συνέχεια υπολογίζουμε την τετραγωνική διαφορά μεταξύ των δύο κανονικοποιημένων τιμών, προσθέτοντας αυτή την τετραγωνική διαφορά στο άθροισμα.
- Αφού συγκεντρώσουμε όλες τις τετραγωνικές διαφορές στη μεταβλητή αθροίσματος, υπολογίζουμε την τετραγωνική ρίζα αυτού του αθροίσματος χρησιμοποιώντας το Math.sqrt(sum). Αυτό δίνει την τελική απόκλιση μεταξύ των δύο πλεγμάτων.

3. Εκτύπωση του αποτελέσματος:

Η συνάρτηση εκτυπώνει την υπολογισμένη απόκλιση στην κονσόλα.

Συνοπτικά, αυτή η συνάρτηση μετρά την απόκλιση μεταξύ δύο πλεγμάτων συγκρίνοντας τις κανονικοποιημένες τιμές στα αντίστοιχα κελιά και αθροίζοντας τις τετραγωνικές διαφορές. Παρέχει μια ενιαία αριθμητική τιμή που αντιπροσωπεύει πόσο διαφορετικά είναι τα δύο πλέγματα μεταξύ τους.

5.3 Αποτελέσματα

Η εφαρμογή που υλοποιήσαμε εμφανίζει τα δύο πλέγματα από τα αρχεία τα οποία εισάγει ο χρήστης. Το κάθε πλέγμα περιέχει την συχνότητα εμφάνισης του κάθε σημείου μετά την κανονικοποίηση. Στο συγκεκριμένο παράδειγμα δώσαμε δύο αρχεία χρονολογίου από δύο ημερολογιακούς μήνες, τον Αύγουστο και τον Δεκέμβρη, εφόσον τους συγκεκριμένους μήνες είναι λογικό λόγω των καλοκαιρινών διακοπών και των χειμερινών εκδρομών των Χριστουγέννων να εμφανίζονται μεγάλες διαφορές στα σημεία επίσκεψης. Η κανονικοποίηση έγινε σε πλέγματα μεγέθους 10x10. Παρατηρούμε στην Εικόνα 4 ότι τα συχνότερα σημεία εμφάνισης στους δύο πίνακες είναι σε διαφορετικές συντεταγμένες του πίνακα.

PRINTING										PRINTING									
0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	174	1	
0	0	0	0	0	63	2	0	0	0	0	0	0	0	0	0	0	0	0	
0	0	0	0	0	4	3	0	0	0	0	0	0	0	0	0	0	0	0	
0	0	0	0	0	0	28	0	0	0	0	0	0	0	0	0	0	0	0	
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
2	0	0	0	1	0	1	0	1	0	0	0	0	0	0	0	0	0	0	
0	0	0	0	0	0	1	2	1	1	0	0	0	0	0	0	0	0	0	
0	0	0	0	0	0	0	12	2	0	15	0	0	0	0	0	0	0	0	
0	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	

Εικόνα 4: Αποτελέσματα πλεγμάτων εφαρμογής (αριστερά: πλέγμα Αυγούστου - δεξιά: πλέγμα Δεκέμβρη)

Στο συγκεκριμένο παράδειγμα η εφαρμογή εμφανίζει ένα αποτέλεσμα απόκλισης της τάξης του 0.9845. Αυτό σημαίνει ότι τα δύο αρχεία έχουν μια σημαντική απόκλιση. Αντίστοιχα, μπορούμε να δώσουμε δύο σημεία που αφορούν μήνες όπου ο χρήστης είναι πιθανότερο να βρίσκεται σε παρόμοιες τοποθεσίες. Για παράδειγμα, κατά την είσοδο των μηνών Οκτώβρη και Νοέμβρη τα αποτελέσματα παρουσιάζονται παρακάτω στην Εικόνα 5.

PRINTING										PRINTING									
0	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	13	1	0	0	0	0	0	0	0	0	8	1
0	0	0	0	0	0	0	0	38	0	0	0	0	0	0	0	0	0	27	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
25	0	0	0	0	0	0	0	0	0	25	0	0	0	0	0	0	0	0	0
22	0	0	0	0	0	0	0	0	0	22	0	0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0

Εικόνα 5: Αποτελέσματα πλεγμάτων εφαρμογής (αριστερά: πλέγμα Οκτώβρης - δεξιά: πλέγμα Νοέμβρης)

Στην περίπτωση αυτή η εφαρμογή υπολόγισε απόκλιση μικρότερη του 0.1 και πιο συγκεκριμένα της τάξης του 0.0921. Όπως άλλωστε παρατηρούμε και από τα δύο πλέγματα η εμφάνιση των πιο συχνά επισκεπτόμενων σημείων είναι παρόμοια, δηλαδή αφορά τις ίδιες συντεταγμένες.

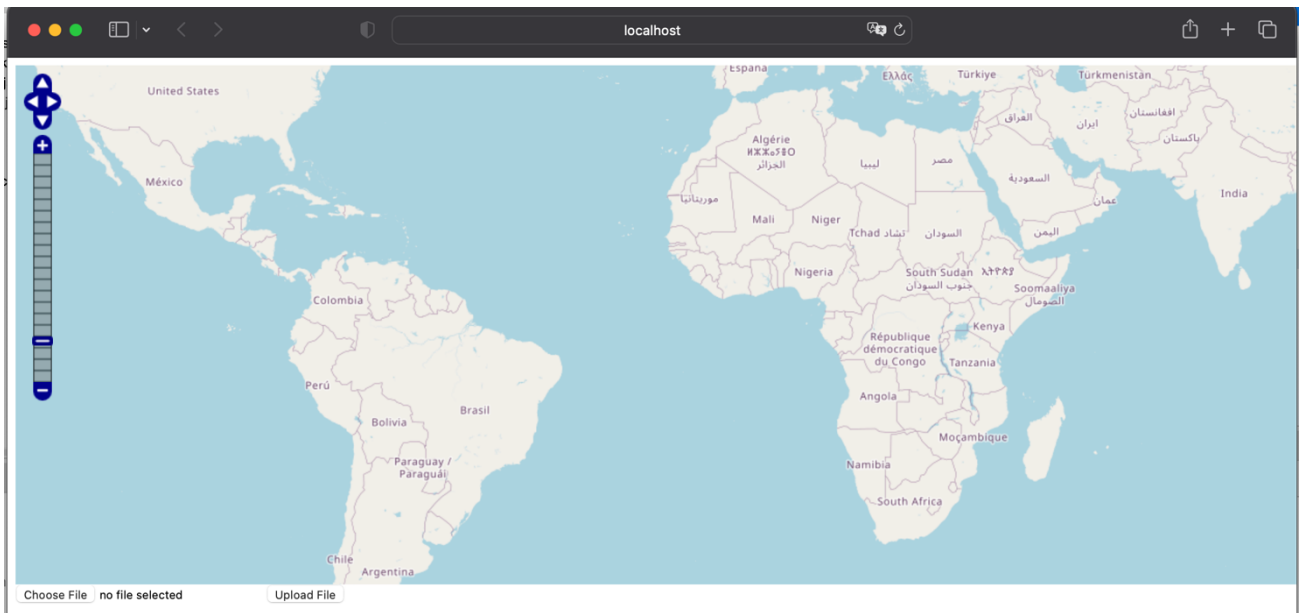
Ενώ τα παραπάνω αποτελέσματα εμφανίζονται αρκετά ορθολογικά, υπάρχουν περιπτώσεις που η απόκλιση μπορεί να αποδώσει πολύ μικρές τιμές σε αρχεία που στην πραγματικότητα αφορούν εντελώς διαφορετικά αρχεία χρονολογίου. Το παραπάνω συμπέρασμα προκύπτει βάσει του γεγονότος ότι τα συγκεκριμένα πλέγματα είναι αρκετά μικρά σε σχέση με το πλήθος των σημείων που μπορούν να υπάρξουν μέσα σε ένα αρχείο χρονολογίου. Άρα λοιπόν για την ακριβέστερη εξαγωγή συμπερασμάτων μέσω της τιμής της απόκλισης, απαραίτητη προϋπόθεση αποτελεί το πλέγμα να είναι όσο το δυνατόν μεγαλύτερο γίνεται, δηλαδή 1.000x1.000 ή 10.000x10.000 κλπ. Σε αυτήν την περίπτωση όμως, παρά το γεγονός πως η απόκλιση θα εμφανίζει με αρκετά καλή προσέγγιση την διαφορά των αρχείων, η εμφάνιση των πλεγμάτων θα καθιστά πολύ δύσκολη την ανάγνωσή του από τους χρήστες.

Παρακάτω, στην Εικόνα 6, εμφανίζουμε ένα παράδειγμα πλέγματος 40x40, το οποίο είναι πολύ μικρότερο από τις επιθυμητές τιμές που αναφέραμε παραπάνω. Όπως μπορούμε να παρατηρήσουμε, υπάρχει μεγαλύτερη διασπορά σε τιμές σε περισσότερα κελιά, το οποίο είναι φυσιολογικό, αλλά όσο το μέγεθος του πίνακα μικραίνει, η οπτικοποίηση γίνεται πολύ πιο δύσκολη.

[illegible]

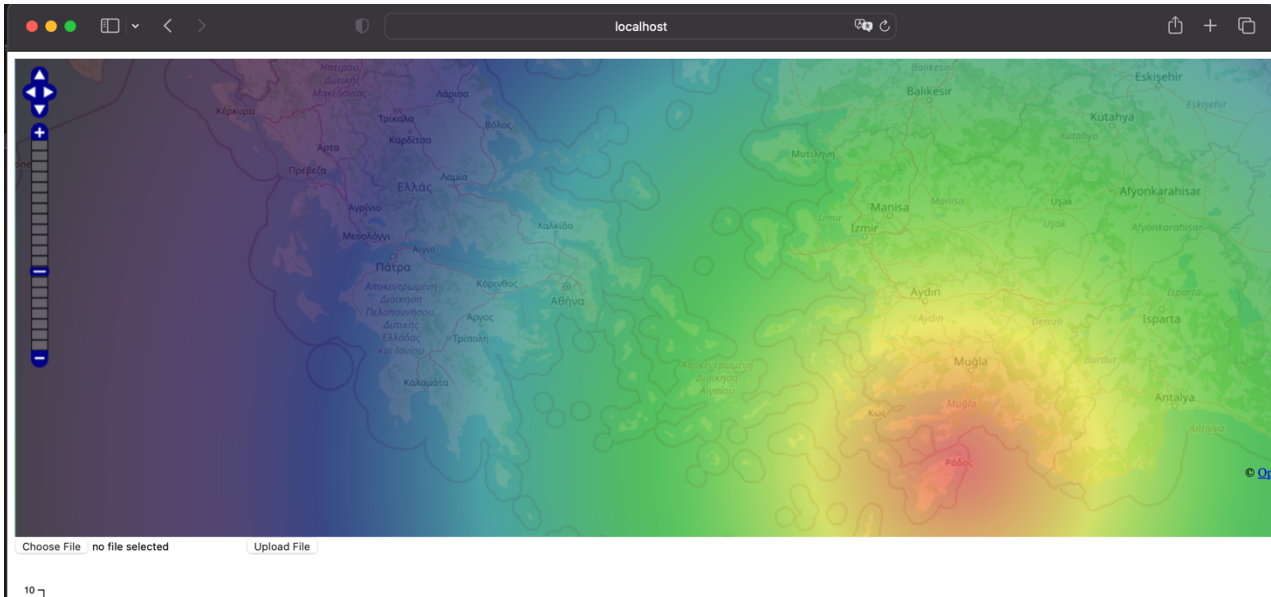
Εικόνα 6: Πλέγμα μεγέθους 40x40

Για τον λόγο αυτό, για την εμφάνιση των αποτελεσμάτων του αλγορίθμου με πολύ μεγαλύτερη ευκολία και εξαγωγή καλύτερων συμπερασμάτων, δημιουργήσαμε 2 διαφορετικούς τρόπους οπτικοποίησης. Ο πρώτος αφορά την εμφάνιση των σημείων με την χρήση ενός heatmap, το οποίο όμως εμφανίζεται πάνω στον παγκόσμιο χάρτη.



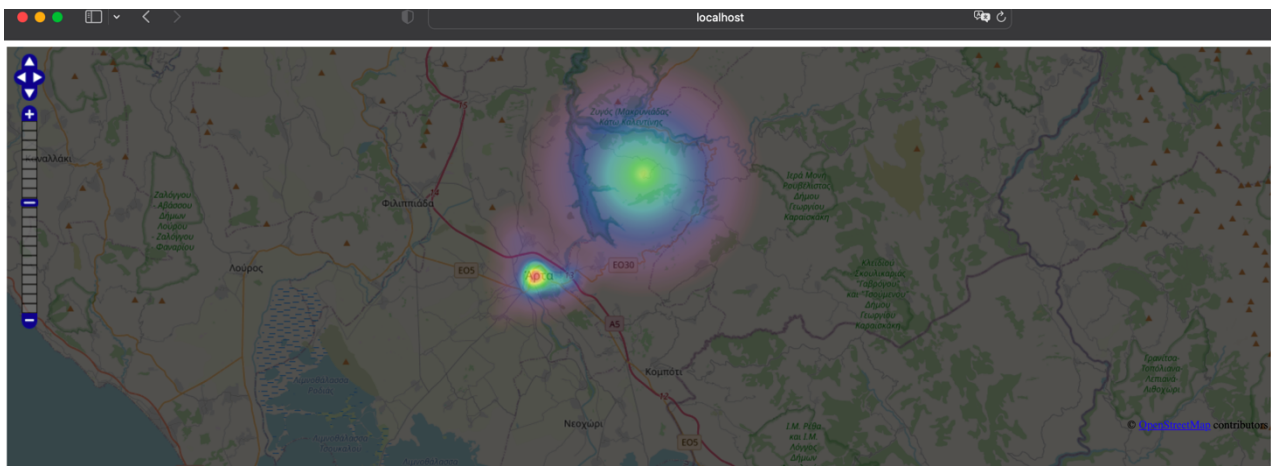
Εικόνα 7: Εμφάνιση χάρτη εφαρμογής

Στην παραπάνω web εφαρμογή υπάρχει η επιλογή ενός κουμπιού όπου με το πάτημά του ο χρήστης μπορεί να επιλέξει ένα αρχείο .json το οποίο έχει εξάγει από το χρονολόγιό του. Στην προκειμένη περίπτωση, αν για παράδειγμα εισάγουμε ένα αρχείο το οποίο περιέχει δεδομένα από κάποιον μήνα που βρεθήκαμε στην Ρόδο, όπως εμφανίζεται στην Εικόνα 8 ο χάρτης θα εμφανίσει με έντονο κόκκινο χρώμα το σημείο που εμφανίζεται τις περισσότερες φορές στο αρχείο.



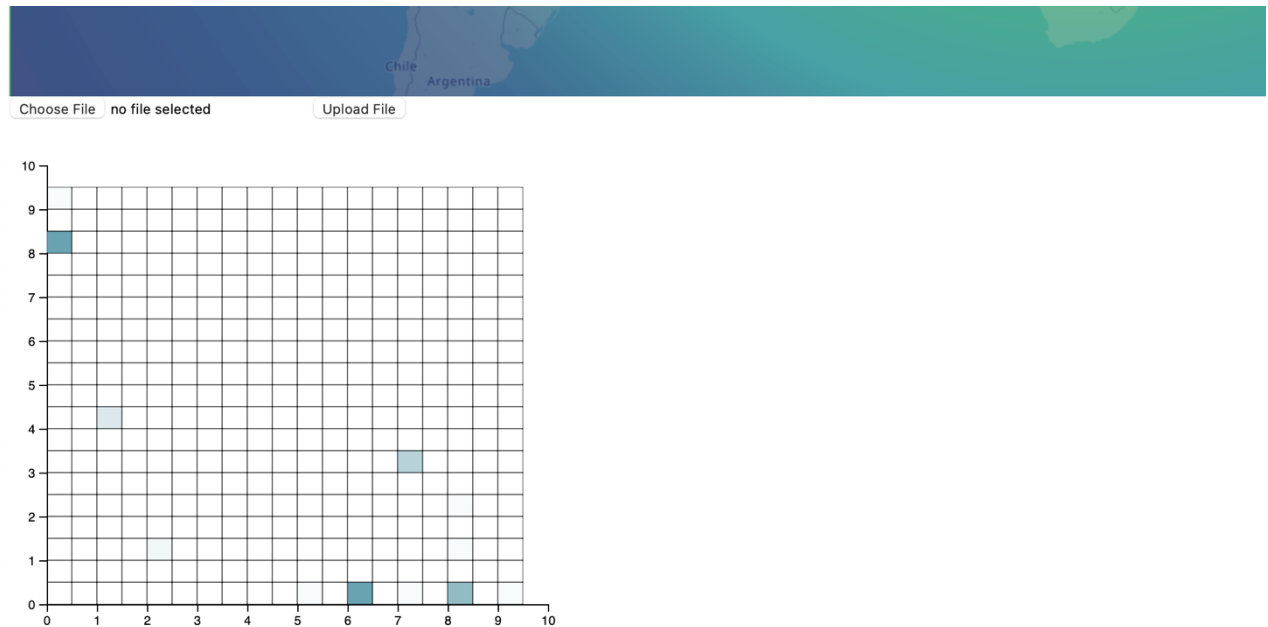
Εικόνα 8: Εμφάνιση οπτικοποίησης από αρχείο καλοκαιρινής περιόδου (Ρόδος)

Σε μία άλλη περίπτωση, όπου το αρχείο εισόδου αφορά διαφορετικές τοποθεσίες, μπορούμε να δούμε πως η εμφάνιση των χρωμάτων είναι λιγότερα έντονη. Για παράδειγμα, στην Εικόνα 9 εμφανίζεται στον χάρτη με πιο έντονο χρώμα η Άρτα, το οποίο σημαίνει πως υπήρχαν αρκετές επισκέψεις σε κάποια συγκεκριμένα σημεία. Από την άλλη, σε αντίστοιχες περιοχές πιο πάνω από την Άρτα εμφανίζεται πράσινος κύκλος με μεγαλύτερη ακτίνα, που σημαίνει πως υπήρχαν και εκεί επισκέψεις, σε περισσότερα σημεία, αλλά με λιγότερη συχνότητα.



Εικόνα 9: Οπτικοποίηση αρχείου με διαφορετική τοποθεσία από αρχείο χειμερινής περιόδου (Άρτα)

Τέλος, στην εφαρμογή εμφανίζεται και ένα δισδιάστατο ιστόγραμμα το οποίο ουσιαστικά είναι η οπτικοποίηση του πλέγματος, όπως εμφανίζεται στην Εικόνα 4 και Εικόνα 5. Ουσιαστικά αντί να εμφανίζονται απλά αριθμοί στον δισδιάστατο πίνακα, εμφανίζονται με την μορφή ιστογράμματος, όπου όσο μεγαλύτερος ο αριθμός σε κάποιο κελί, τόσο πιο έντονο το χρώμα που θα χρωματιστεί το κελί. Αυτό μας δίνει μια καλύτερη εικόνα της διασποράς των σημείων μέσα στο αρχείο .json.



Εικόνα 10: Εμφάνιση δισδιάστατου ιστογράμματος

6

Συμπεράσματα

Η σημασία και χρησιμότητα των δεδομένων χρονολογίου που συλλέγονται από έξυπνες συσκευές, έχει αναγνωριστεί από πολλούς ερευνητές τα τελευταία χρόνια οι οποίοι έχουν αξιοποιήσει τα δεδομένα αυτά για την εξαγωγή συμπερασμάτων και την ανάπτυξη νέων εφαρμογών. Στον τομέα της αστικής ανάλυσης και του σχεδιασμού πόλεων, οι ερευνητές χρησιμοποιούν τα δεδομένα του χρονολογίου για να μελετήσουν τα πρότυπα κινητικότητας των ανθρώπων και να κατανοήσουν καλύτερα τη ζήτηση για μεταφορικές υποδομές και υπηρεσίες. Αυτές οι πληροφορίες μπορούν να βοηθήσουν στον καλύτερο σχεδιασμό των δρόμων, των συστημάτων μαζικής μεταφοράς και άλλων αστικών υποδομών. Στον τομέα της δημόσιας υγείας, οι επιδημιολόγοι μπορούν να χρησιμοποιήσουν τα δεδομένα χρονολογίου για να παρακολουθήσουν και να μοντελοποιήσουν την εξάπλωση ασθενειών, προκειμένου να σχεδιάσουν αποτελεσματικότερες στρατηγικές πρόληψης και ελέγχου. Στη βιομηχανία του μάρκετινγκ, οι αναλυτές χρησιμοποιούν τα δεδομένα χρονολογίου και να κατανοήσουν τις συνήθειες και τις προτιμήσεις των καταναλωτών, προκειμένου να προσαρμόσουν καλύτερα τις στρατηγικές τους και να παρέχουν πιο εξατομικευμένες εμπειρίες. Συνολικά, τα δεδομένα χρονολογίου αποτελούν μια πολύτιμη πηγή πληροφοριών για πολλούς επιστημονικούς και ερευνητικούς τομείς, καθώς παρέχουν μια αναλυτική εικόνα των ανθρώπινων δραστηριοτήτων και μετακινήσεων σε διάφορα περιβάλλοντα και κλίμακες.

Με βάση την ανάλυση των αποτελεσμάτων από τα σύνολα δεδομένα με διαφορετικά επίπεδα μετακινήσεων, μπορούμε να εξάγουμε διάφορα συμπεράσματα σχετικά με την απόδοση και τις δυνατότητες του αλγορίθμου. Πιο συγκεκριμένα, τα αποτελέσματα έδειξαν πως η χρήση της απόκλισης για την σύγκριση δύο αρχείων, ύστερα από την μετατροπή τους σε αντίστοιχο πλέγμα μεγέθους μικρότερη του 100, ως κριτήριο σύγκρισης των δύο αρχείων παρουσιάζει υψηλά επίπεδα ακρίβειας. Για παράδειγμα, για τον ίδιο χρήστη ο οποίος εκτελούσε μετακινήσεις στην Ρόδο και στην Άρτα η ομοιότητα των αρχείων του ήταν αρκετά κοντά (με απόκλιση κοντά στο 0) σε σχέση με τα αρχεία άλλου χρήστη που κατοικεί στην Σάμο (όπου η απόκλιση πλησίαζε στο 1). Ωστόσο ο αλγόριθμος, όπως είναι φυσικό, επηρεάζεται αρκετά από το μέγεθος του N , το οποίο θα πρέπει να επιλεγεί σωστά, ώστε να κατανεμηθούν ομοιόμορφα οι διαφορετικές τοποθεσίες του χρήστη στο αντίστοιχο πλέγμα. Έτσι, ένα πλέγμα μεγέθους $N \gg 100$ μπορεί να οδηγήσει σε ένα πλέγμα με πολύ άδειες θέσεις, το οποίο ακόμα και σε αρχεία τα οποία δεν διαφέρουν αρκετά, η απόκλιση θα είναι πολύ μεγάλη, καθώς διαφορετικές, αν και γειτονικές θέσεις, θα είναι γεμάτες. Αντιθέτως, η επιλογή ενός N το οποίο είναι αρκετά μικρό, θα οδηγήσει σε τιμές απόκλισης αρκετά μικρές ακόμα και σε αρχεία τα

οποία διαφέρουν αρκετά, εφόσον ο αλγόριθμος δεν θα μπορεί να υπολογίσει την πραγματική διαφορά των σημείων από μικρό αριθμό κελιών.

Η παρούσα εργασία προσφέρει μια ολοκληρωμένη μεθοδολογία για την αποτελεσματική ανάλυση και επεξεργασία δεδομένων χρονολογίου. Μέσω της λεπτομερούς ανάλυσης των δεδομένων αυτών και της δυνατότητας σύγκρισης αρχείων χρονολογίου διαφορετικών περιόδων, η προτεινόμενη προσέγγιση παρέχει μια εμπειριστατωμένη κατανόηση των προτύπων κινητικότητας και των καθημερινών ρουτινών των χρηστών. Αυτή η γνώση μπορεί να αξιοποιηθεί σε τομείς όπως ο αστικός σχεδιασμός, οι στρατηγικές μάρκετινγκ, η μελέτη ανθρώπινης συμπεριφοράς και ο σχεδιασμός βιώσιμων περιβαλλοντικών λύσεων. Ως μελλοντική εργασία, θα μπορούσαν εκπονηθεί η υλοποίηση κάποιων εκ των παραπάνω προτάσεων. Για παράδειγμα, θα μπορούσαμε να συνδυάσουμε τα δεδομένα χρονολογίου με άλλα σύνολα δεδομένων, όπως δεδομένα καιρού, κυκλοφορίας ή δημόσιων εκδηλώσεων, για την καλύτερη κατανόηση παραγόντων που επηρεάζουν την κινητικότητα των χρηστών. Επιπλέον, θα μπορούσαν να προταθούν περισσότεροι τρόποι για την καλύτερη οπτικοποίηση των γεωχωρικών δεδομένων των χρηστών, ενώ θα μπορούσε να καταστεί η εφαρμογή πιο προσβάσιμη και εύχρηστη προς τους τελικούς χρήστες.

Βιβλιογραφία

- [1] ‘Google Maps Timeline - Android - Google Maps Help’. Accessed: Feb. 26, 2024. [Online]. Available: <https://support.google.com/maps/answer/6258979?hl=en&co=GENIE.Platform%3DAndroid>
- [2] A. Li, ‘Google Maps Timeline adds “Insights” & “Trips” tabs as transit crowdedness predictions expand’, 9to5Google. Accessed: Feb. 26, 2024. [Online]. Available: <https://9to5google.com/2021/07/21/google-maps-timeline-insights/>
- [3] ‘Google Photos’. Accessed: Feb. 26, 2024. [Online]. Available: <https://www.google.com/photos/about/>
- [4] ‘Working with JSON - Learn web development | MDN’. Accessed: Feb. 26, 2024. [Online]. Available: <https://developer.mozilla.org/en-US/docs/Learn/JavaScript/Objects/JSON>
- [5] ‘JSON File Format - What is a JSON file?’ Accessed: Feb. 26, 2024. [Online]. Available: <https://docs.fileformat.com/web/json/>
- [6] ‘Semantic Location History - Location History Format’. Accessed: Feb. 26, 2024. [Online]. Available: <https://locationhistoryformat.com/reference/semantic/>
- [7] H. Mehta, P. Kanani, and P. Lande, ‘Google Maps’, *Int. J. Comput. Appl.*, vol. 178, pp. 41–46, May 2019, doi: 10.5120/ijca2019918791.
- [8] ‘GPS.gov: Trilateration Exercise’. Accessed: Feb. 26, 2024. [Online]. Available: <https://www.gps.gov/multimedia/tutorials/trilateration/>
- [9] ‘What is GNSS? | EU Agency for the Space Programme’. Accessed: Feb. 26, 2024. [Online]. Available: <https://www.euspa.europa.eu/european-space/eu-space-programme/what-gnss>
- [10] A. Darrow-Pinion *et al.*, ‘ETA Prediction with Graph Neural Networks in Google Maps’, in *Proceedings of the 30th ACM International Conference on Information & Knowledge Management*, Oct. 2021, pp. 3767–3776. doi: 10.1145/3459637.3481916.
- [11] ‘Activity Recognition API’, Google for Developers. Accessed: Feb. 26, 2024. [Online]. Available: <https://developers.google.com/location-context/activity-recognition>
- [12] Y. Zheng, Y. Chen, Q. Li, X. Xie, and W.-Y. Ma, ‘Understanding transportation modes based on GPS data for web applications’, *ACM Trans. Web*, vol. 4, no. 1, pp. 1–36, Jan. 2010, doi: 10.1145/1658373.1658374.
- [13] A. R. M. Radzol *et al.*, ‘A Study of Older Persons’ Daily Travel Patterns in Klang Valley. | Malaysian Journal of Medicine & Health Sciences | EBSCOhost’. Accessed: Feb. 26, 2024. [Online]. Available: <https://openurl.ebsco.com/contentitem/gcd:160927220?sid=ebsco:plink:crawler&id=ebsco:gcd:160927220>
- [14] C. Sardanios, I. Varlamis, and G. Bouras, ‘Extracting User Habits from Google Maps History Logs’, in *2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, Barcelona: IEEE, Aug. 2018, pp. 690–697. doi: 10.1109/ASONAM.2018.8508442.
- [15] Y. Zheng, X. Xie, and W.-Y. Ma, ‘GeoLife: A Collaborative Social Networking Service among User, Location and Trajectory’.
- [16] Y. Zheng, L. Wang, R. Zhang, X. Xie, and W.-Y. Ma, ‘GeoLife: Managing and Understanding Your Past Life over Maps’, in *The Ninth International Conference on Mobile Data Management (mdm 2008)*, Apr. 2008, pp. 211–212. doi: 10.1109/MDM.2008.20.