



ΧΑΡΟΚΟΠΕΙΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

Π.Μ.Σ. ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΤΗΛΕΜΑΤΙΚΗΣ

ΥΠΟΛΟΓΙΣΤΙΚΕΣ ΚΑΙ ΔΙΑΔΙΚΤΥΑΚΕΣ ΤΕΧΝΟΛΟΓΙΕΣ ΚΑΙ ΕΦΑΡΜΟΓΕΣ

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Ανίχνευση συνηθειών χρήστη με χρήση κινητών τηλεφώνων

Τραγοπούλου Σπυριδούλα

Επιβλέπων: **Βαρλάμης Ηρακλής**, Επίκουρος Καθηγητής

Μέλη Τριμελούς Επιτροπής: **Δημητρακόπουλος Γεώργιος**, Επίκουρος Καθηγητής
Τσερπές Κωνσταντίνος, Λέκτορας

ΑΘΗΝΑ

ΦΕΒΡΟΥΑΡΙΟΣ 2016

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Ανίχνευση συνηθειών χρήστη με χρήση κινητών τηλεφώνων

Τραγοπούλου Σπυριδούλα

ΑΜ: 13115

Επιβλέπων:

Βαρλάμης Ηρακλής, Επίκουρος Καθηγητής

Μέλη Τριμελούς Επιτροπής:

Δημητρακόπουλος Γεώργιος, Επίκουρος Καθηγητής

Τσερπές Κωνσταντίνος, Λέκτορας

Περίληψη

Οι χρήστες έξυπνων κινητών τηλεφώνων (smartphones) έχουν πρόσβαση σε πληθώρα πληροφοριών μέσω των κινητών τους συσκευών, η διαχείριση των οποίων απαιτεί χρόνο και κόπο από μέρους των χρηστών. Κατά συνέπεια, είναι αναγκαία η παροχή προσωποποιημένων υπηρεσιών στους χρήστες, οι οποίες έχουν ως στόχο να διευκολύνουν την πρόσβαση σε πληροφορίες που ο χρήστης έχει ανάγκη σε συγκεκριμένες χρονικές στιγμές. Η ανάπτυξη προσωποποιημένων εφαρμογών προϋποθέτει την κατανόηση της συμπεριφοράς του χρήστη και τον γενικότερων ενδιαφερόντων του. Η παρούσα εργασία στοχεύει στην ανάπτυξη μια εφαρμογής σε γλώσσα Java, η οποία θα εντοπίζει τις καθημερινές συνήθειες του χρήστη εφαρμόζοντας τεχνικές εξόρυξης δεδομένων και συνδυάζοντας γεωχωρικά δεδομένα σε δεδομένα κίνησης του χρήστη. Η συλλογή των δεδομένων κίνησης πραγματοποιείται μέσω της υπάρχουσας εφαρμογής GPSTracker χωρίς να προϋποθέτει την αλληλεπίδραση του χρήστη. Αναλυτικότερα, η εφαρμογή εντοπίζει τα σημεία ενδιαφέροντος με την τεχνική της συσταδοποίησης από τα δεδομένα κίνησης του χρήστη, ενώ στη συνέχεια χαρακτηρίζει σημασιολογικά τα σημεία ενδιαφέροντος με χρήση της εφαρμογής OpenStreetMaps. Επιπλέον, χρησιμοποιεί τα αναγνωρισμένα σημεία ενδιαφέροντος για να εξαγάγει τις συνήθειες του χρήστη με χρήση της τεχνικής των κανόνων συσχέτισης. Στην παρούσα εργασία αναγνωρίστηκαν επιτυχώς μερικές από τις καθημερινές συνήθειες του χρήστη, ενώ ταυτόχρονα μελετήθηκαν οι παράμετροι και οι προκλήσεις της συνολικής διαδικασίας.

ΘΕΜΑΤΙΚΗ ΠΕΡΙΟΧΗ: *Εξόρυξη Δεδομένων*

ΛΕΞΕΙΣ ΚΛΕΙΔΙΑ: *Ανίχνευση Συνηθειών, Συσταδοποίηση, Κανόνες Συσχέτισης, OpenStreetMaps*

Abstract

Smartphone users have access to a plethora of information via their mobile devices, which requires time and effort in order to be managed appropriately. Consequently, the personalization services are necessary, so as to facilitate access to information that the user needs at particular moments. The development of personalized applications requires the knowledge of user behavior and interests. This study aims to develop a Java application, which identifies the user's daily habits by using data mining techniques and by combining geospatial data with user's movement data. The collection of user movement data is performed via the existing application GPSTracker without requiring the user's interaction. Specifically, the application identifies the points of interest using a clustering technique on user movement data, followed by characterization of points of interest with semantic labels using the OpenStreetMaps API. Furthermore, using association rules on the recognized points of interest, the application extracts the user habit's and interests. In this study, part of user daily habits were successfully discovered. Moreover, the parameters and challenges of the complete process were identified and studied.

SUBJECT AREA: *Data mining*

KEYWORDS: *Habits recognition, Clustering, Association Rules, OpenStreetMaps*

Ευχαριστίες

Η παρούσα διπλωματική εργασία εκπονήθηκε στα πλαίσια του Π.Μ.Σ Πληροφορικής και Τηλεματικής στην κατεύθυνση Ύπολογιστικές και Διαδικτυακές Τεχνολογίες και Εφαρμογές", του Χαροκοπείου Πανεπιστημίου. Σε αυτό το σημείο, θα ήθελα να εκφράσω τις ευχαριστίες μου σε όλους όσους συνέβαλαν άμεσα ή έμμεσα στην ολοκλήρωση της διπλωματικής εργασίας μου και γενικότερα των μεταπτυχιακών σπουδών μου.

Αρχικά θα ήθελα να εκφράσω τις ιδιαίτερες ευχαριστίες μου στον επιβλέποντα καθηγητή της διπλωματικής μου εργασίας, Επίκουρο Καθηγητή Ηρακλή Βαρλάμη, για την συνεχή καθοδήγηση, την άμεση υποστήριξη και ενθάρρυνση, τις πολύτιμες συμβουλές και την υπομονή που έδειξε στο πρόσωπο μου, καθόλη τη διάρκεια εκπόνησης της διπλωματικής μου εργασίας αλλά και για την συμβολή των γνώσεων και των συμβουλών του κατά τη διάρκεια των μεταπτυχιακών σπουδών μου.

Στη συνέχεια, θα ήθελα να ευχαριστήσω τα μέλη της τριμελούς εξεταστικής επιτροπής, τον Επίκουρο Καθηγητή, Δημητράκοπουλο Γεώργιο και τον Λέκτορα, Κωνσταντίνο Τσερπέ, για την αξιολόγηση της προσπάθειας μου αλλά και για την συμβολή των πολύτιμων γνώσεων τους κατά τη διάρκεια των σπουδών μου.

Τέλος, δεν θα μπορούσα να παραλείψω να ευχαριστήσω το κοντινό μου περιβάλλον και ιδιαίτερα την οικογένεια μου, για την κατανόηση, την υπομονή και την ψυχολογική υποστήριξη στις δύσκολες στιγμές που υπήρξαν κατά τη διάρκεια εκπόνησης της διπλωματικής μου εργασίας αλλά και την συνολική συμπαράσταση και υποστήριξη καθόλη τη διάρκεια των σπουδών μου.

Περιεχόμενα

1	Εισαγωγή	15
1.1	Σκοπός της διπλωματικής εργασίας	16
1.2	Προτεινόμενη Προσέγγιση	16
1.3	Συνήθειες Χρήστη και Mobile Marketing	17
2	Θεωρητικό Υπόβαθρο	19
2.1	Εφαρμογές Ανίχνευσης Συνηθειών	20
2.1.1	Moves	20
2.1.2	Frequent Locations iOS	21
2.2	Σχετικές Έρευνες	22
2.2.1	Αισθητήρες Ανίχνευσης Συνηθειών	22
2.2.2	Τεχνικές Ανίχνευσης Συνηθειών	24
2.3	Προτεινόμενη Τεχνική	26
3	Σχεδιασμός	29
3.1	Αρχιτεκτονική Εφαρμογής	29
3.2	Μεθοδολογία Υλοποίησης Εφαρμογής	30
3.3	Εργαλεία Υλοποίησης Εφαρμογής	33
3.3.1	Okeanos IAAS Cloud	33
3.3.2	Apache HTTP Server	33

3.3.3	PostgreSQL	35
3.3.4	Apache Commons Mathematics	35
3.3.5	OpenStreetMap	36
3.3.6	Το περιβάλλον Weka	36
3.4	Εργαλεία Ανάπτυξης Εφαρμογής	36
3.4.1	Το περιβάλλον Android Studio	36
3.4.2	Το περιβάλλον Eclipse	37
3.5	Αρχιτεκτονική Εργαλείων	38
4	Υλοποίηση	39
4.1	Συλλογή Δεδομένων σε Κεντρικό Αποθετήριο	39
4.1.1	Δεδομένα Κίνησης	39
4.1.2	Κεντρικό Αποθετήριο Δεδομένων	41
4.1.3	Μεταφόρτωση Δεδομένων στο Κεντρικό Αποθετήριο	42
4.1.3.1	Χειροκίνητη	42
4.1.3.2	Αυτόματη	43
4.2	Ανάκτηση Δεδομένων Χρήστη από Κεντρικό Αποθετήριο	44
4.3	Ανίχνευση Σημείων Ενδιαφέροντος Χρήστη	45
4.4	Χαρακτηρισμός Σημείων Ενδιαφέροντος Χρήστη	48
4.5	Εξαγωγή Συνηθειών	51
5	Αποτελέσματα	53
5.1	Αποτελέσματα Ανίχνευσης Σημείων Ενδιαφέροντος	53
5.2	Αποτελέσματα Χαρακτηρισμού Σημείων Ενδιαφέροντος	57
5.3	Αποτελέσματα Εξαγωγής Συνηθειών	59
6	Συμπεράσματα	63

6.1	Συμπεράσματα	63
6.2	Μελλοντικές Επεκτάσεις	65
	Βιβλιογραφία	70

Κατάλογος Σχημάτων

2.1	Στιγμιότυπα χρήσης της εφαρμογής Moves (Πηγή: play.google.com ¹).	20
2.2	Στιγμιότυπα χρήσης της εφαρμογής Frequent Locations (Πηγή: https://aaronparecki.com ²).	21
3.1	Η αρχιτεκτονική της προτεινόμενης προσέγγισης.	30
3.2	Διάγραμμα Ροής Διαδικασίας Ανίχνευσης Συνηθειών.	31
3.3	Το περιβάλλον του Apache Web Server.	34
3.4	Η αρχιτεκτονική της πλατφόρμας Eclipse. (Πηγή: www.eclipse.org ³)	37
4.1	Στιγμιότυπα χρήσης της εφαρμογής GPSTracker.	40
4.2	Δείγμα καταγραφής δεδομένων κατά τη διάρκεια κίνησης του χρήστη.	41
4.3	Το σχήμα της Βάσης Δεδομένων του Κεντρικού Αποθετηρίου Δεδομένων.	42
4.4	Στιγμιότυπα χρήσης της μεταφόρτωσης αρχείου στην εφαρμογή GPSTracker.	43
4.5	Bounding box ενός σημείου ενδιαφέροντος.	49
4.6	Παράδειγμα αρχείου xml όπως εξάγεται από το OpenStreetMaps API.	50
4.7	Δείγμα αρχείου δεδομένων arff όπως εξάγεται από την εφαρμογή.	51
5.1	Τα αποτελέσματα του εντοπισμού σημείων ενδιαφέροντος όπως αποτυπώνονται σε χάρτη.	54
5.2	Οι συστάδες με σχετικά μεγάλη τιμή διαμέτρου που εντοπίστηκαν όπως αποτυ- πώνονται σε χάρτη.	56
5.3	Σημείο ενδιαφέροντος το οποίο χαρακτηρίστηκε ως public_transport.	58

5.4	Σημείο ενδιαφέροντος το οποίο χαρακτηρίστηκε ως amenity.	58
5.5	Διάγραμμα σχετικής συχνότητας εμφάνισης tag στα σημεία ενδιαφέροντος. . . .	59

Κατάλογος Πινάκων

4.1	Κατηγορίες χαρακτηριστικών που καταγράφονται από την εφαρμογή GPSTracker.	40
5.1	Αριθμός συστάδων ανά κατηγορία μήκους διαμέτρου.	55
5.2	Αριθμός συστάδων ανά κατηγορία μήκους μέσης ακτίνας από το centroid.	55
5.3	Κατανομή του αριθμού των συστάδων ανά tag του OpenStreetMaps.	59

Κεφάλαιο 1

Εισαγωγή

Τα έξυπνα κινητά τηλέφωνα επιτρέπουν την πρόσβαση του χρήστη σε μια πληθώρα πληροφοριών. Πληροφορίες για τα μέσα μαζικής μεταφοράς, καταστήματα με εκπτώσεις, νέα εστιατόρια, εκδηλώσεις είναι διαθέσιμες μέσω των smartphones. Η ποσότητα των πληροφοριών που είναι διαθέσιμες στον χρήστη μέσω ενός smartphone είναι συντριπτική και αυτός είναι ένας από τους κυριότερους παράγοντες της ευρείας υιοθέτησης των κινητών συσκευών στη σύγχρονη κοινωνία[10]. Όμως, οι συσκευές οι οποίες έχουν ως στόχο να διευκολύνουν την καθημερινή ζωή του χρήστη, για να είναι πραγματικά χρήσιμες πρέπει να διαθέτουν γνώσεις για τον συγκεκριμένο χρήστη, δηλαδή να μπορούν να κατανοήσουν τις ενέργειες που πρόκειται να κάνει ο χρήστης και για ποιο λόγο[6]. Δημιουργείται λοιπόν η ανάγκη παροχής προσωποποιημένων υπηρεσιών στους χρήστες έξυπνων κινητών τηλεφώνων, διαθέτοντας πιο εύκολα και γρήγορα προσβάσιμες τις πληροφορίες που ο χρήστης έχει ανάγκη την συγκεκριμένη χρονική στιγμή.

Η ανάπτυξη προσωποποιημένων εφαρμογών προϋποθέτει την επίβλεψη της συμπεριφοράς του χρήστη και την γνώση των συνηθειών του. Τα έξυπνα κινητά τηλέφωνα προσφέρουν πανταχού παρούσα πρόσβαση και επεξεργασία πληροφοριών, ενώ τα περισσότερα από αυτά διαθέτουν πλήθος ενσωματωμένων αισθητήρων που τα καθιστά ικανά να παράγουν πληροφορίες με υψηλή ακρίβεια[9]. Κατά συνέπεια, τα smartphones αποτελούν την ιδανική πλατφόρμα για την παρακολούθηση της συμπεριφοράς και την εξαγωγή των συνηθειών του χρήστη, η γνώση των οποίων μπορεί να είναι χρήσιμη για την διευκόλυνση της καθημερινής του ζωής.

1.1 Σκοπός της διπλωματικής εργασίας

Η παρούσα διπλωματική εργασία έχει στόχο να επεκτείνει την Android εφαρμογή GPSTracker που αναπτύχθηκε στα πλαίσια πτυχιακής εργασίας η οποία καταγράφει μέσω του αισθητήρα GPS δεδομένα θέσης του χρήστη και κατηγοριοποιεί σε πραγματικό χρόνο την κίνησή του. Σκοπός της εργασίας είναι αρχικά να συλλέγει τα δεδομένα σε κεντρικό αποθετήριο και να αναλύει σε δεύτερο χρόνο τα κατηγοριοποιημένα δεδομένα κίνησης έτσι ώστε να εντοπίζει σημεία ενδιαφέροντος στα οποία ο χρήστης περνά αρκετό χρόνο. Στη συνέχεια, συνδυάζει τα σημεία ενδιαφέροντος με γεωχωρικά δεδομένα ώστε να εξάγει γνώση για τα μέρη όπου περνάει χρόνο ο χρήστης και αναλύει τα δεδομένα ώστε να εξάγει τα γενικότερα ενδιαφέροντα και τις καθημερινές συνήθειες του χρήστη.

Η γνώση που εξάχθηκε από την συγκεκριμένη εργασία, είναι χρήσιμη για την εξατομίκευση των χρηστών και την παροχή υπηρεσιών στον χρήστη που βασίζονται στην συμπεριφορά, τις προτιμήσεις και τις συνήθειες του. Για παράδειγμα, αν ένας χρήστης επισκέπτεται συχνά εστιατόρια, τότε είναι πιθανό να τον ενδιαφέρουν προσφορές σε άλλα εστιατόρια, είτε η πληροφορία για ένα εστιατόριο που άνοιξε πρόσφατα. Αντίστοιχα, σε περίπτωση που ο χρήστης εντοπίζεται συχνά κοντά σε στάσεις λεωφορείου ή τρένου, θα ήταν χρήσιμο να έχει άμεση πρόσβαση στις διαδρομές και το πρόγραμμα δρομολογίων των Μέσων Μαζικής Μεταφοράς.

1.2 Προτεινόμενη Προσέγγιση

Η υπάρχουσα εφαρμογή GPSTracker λειτουργεί σε κινητές συσκευές με λειτουργικό σύστημα Android καταγράφοντας κατά τη διάρκεια της ημέρας δεδομένα θέσης του χρήστη με χρήση του αισθητήρα GPS. Έπειτα, επεξεργάζεται τοπικά τα δεδομένα και αναγνωρίζει τον τύπο κίνησης του χρήστη με τη μέθοδο της κατηγοριοποίησης (classification). Η προτεινόμενη προσέγγιση της παρούσας εργασίας έχει ως αρχικό στόχο να συλλέξει τα κατηγοριοποιημένα δεδομένα του χρήστη σε ένα κεντρικό αποθετήριο ώστε να είναι δυνατή η επεξεργασία τους σε δεύτερο χρόνο. Αναλυτικότερα, η συλλογή των δεδομένων πραγματοποιείται μεταφορτώνοντας αρχεία σε έναν Apache Web Server με τη χρήση ενός Web Service το οποίο στη συνέχεια αποθηκεύει τα δεδομένα σε μια Βάση Δεδομένων PostgreSQL με στόχο την καλύτερη διαχείρισή τους. Η διαδικασία της συλλογής των δεδομένων αποτελεί μια διάφανη (transparent) διαδικασία, η οποία δεν προϋποθέτει κάποια ενέργεια από μέρος του χρήστη και πραγματοποιείται μόνο όταν η κινητή συσκευή είναι συνδεδεμένη σε κάποιο δίκτυο Wi-fi χωρίς να επιβαρύνει τον χρήστη με χρεώσεις χρήσης δεδομένων κινητής τηλεφωνίας.

Στη συνέχεια, υλοποιήθηκε μια εφαρμογή σε γλώσσα Java η οποία στοχεύει στην ανάλυση των δεδομένων κίνησης χρήστη και στην εξαγωγή των συνηθειών του. Η εφαρμογή κάνει ανάκτηση των δεδομένων ανά ID χρήστη και εφόσον ολοκληρωθεί η απαραίτητη προεπεξεργασία των δεδομένων πραγματοποιεί ανίχνευση των σημείων ενδιαφέροντος του χρήστη, δηλαδή των σημείων όπου παρατηρείται ότι ο χρήστης παραμένει ακίνητος για ένα μεγάλο χρονικό διάστημα. Ο εντοπισμός των σημείων ενδιαφέροντος πραγματοποιήθηκε με χρήση τεχνικών εξόρυξης δεδομένων και πιο συγκεκριμένα με την τεχνική της συσταδοποίησης (clustering) χρησιμοποιώντας τους αλγορίθμους συσταδοποίησης που παρέχει η βιβλιοθήκη Apache Commons Math για Java.

Ο χαρακτηρισμός των σημείων ενδιαφέροντος υλοποιήθηκε συνδυάζοντας τα σημεία που ανιχνεύθηκαν, με γεωχωρικά δεδομένα μέσω της εφαρμογής OpenStreetMap[34]. Αναλυτικότερα, η Java εφαρμογή κάνει κλήση του OpenStreetMaps API ορίζοντας το σημείο ενδιαφέροντος ως ένα bounding box, το οποίο επιστρέφει ως απάντηση πληροφορίες σε μορφή XML για το σημείο ενδιαφέροντος και μετά από επεξεργασία αποδίδεται στο σημείο ένας χαρακτηρισμός, για παράδειγμα σταθμός τρένου, supermarket, εστιατόριο.

Τέλος, η αναγνώριση των συνηθειών του χρήστη πραγματοποιήθηκε με χρήση αλγορίθμων κανόνων συσχέτισης μέσω της πλατφόρμας Weka[15], εφόσον πρώτα τα δεδομένα κίνησης συνδυαστούν με τα σημεία ενδιαφέροντος που εντοπίστηκαν και εξαχθούν σε κατάλληλη μορφή για περαιτέρω ανάλυση.

1.3 Συνήθειες Χρήστη και Mobile Marketing

Η εμφάνιση των κινητών συσκευών έχει διαμορφώσει έναν εξ' ολοκλήρου νέο τρόπο ζωής στη σύγχρονη κοινωνία και όπως είναι φυσικό δεν θα μπορούσε να αφήσει ανεπηρέαστη την καταναλωτική συμπεριφορά. Παρατηρείται λοιπόν, ότι οι καταναλωτές τείνουν να παίρνουν στιγμιαίες καταναλωτικές αποφάσεις (micro-moments) καθώς η χρήση των κινητών συσκευών τους επιτρέπει να ενεργούν άμεσα, αναζητώντας είτε αγοράζοντας προϊόντα οποιαδήποτε στιγμή[12]. Όπως αναφέρει η Serena Poter, αντιπρόεδρος της ομάδας στρατηγικής marketing στην Macys.com, οι καταναλωτές σήμερα, ειδικά με την τεχνολογία που έχουν στη διάθεσή τους, αναμένουν την αμεσότητα, ενώ η αναζήτηση είναι το εργαλείο που τους επιτρέπει να βρουν γρήγορα ακριβώς αυτό που ψάχνουν[11]. Αυτή η συμπεριφορά, όπως περιγράφεται από την Google, αποτελεί μια σημαντική ευκαιρία για τις εταιρείες καθώς μπορούν μέσω των κινητών συσκευών να διαμορφώσουν τις αποφάσεις και τις προτιμήσεις των καταναλωτών[12].

Οι καταναλωτές οι οποίοι κάνουν χρήση των κινητών συσκευών τους, διακρίνονται από τις

υψηλές προσδοκίες στην επιλογή του προϊόντος που θα αγοράσουν ενώ ταυτόχρονα χαρακτηρίζονται από περιορισμένη υπομονή. Αυτό καθιστά σημαντικούς παράγοντες του mobile marketing, την ποιότητα και την συνάφεια με τις προτιμήσεις των καταναλωτών. Συνεπώς, οι εταιρείες που μπορούν να αντιληφθούν και να αντιμετωπίσουν καλύτερα τις προτιμήσεις ενός χρήστη καταλήγουν να είναι πιο ελκυστικές από άλλες που απαιτούν περισσότερο χρόνο αναζήτησης από τον χρήστη για την εύρεση του κατάλληλου προϊόντος[12].

Ταυτόχρονα, η γνώση των συνηθειών των καταναλωτών φαίνεται να αποτελεί γνώμονα για την ανάπτυξη του mobile marketing καθώς παρατηρείται ότι οι χρήστες έχουν διάθεση να δοκιμάσουν νέα πράγματα σε στιγμές ρουτίνας. Πιο συγκεκριμένα, το 91% των χρηστών κινητών συσκευών κάνουν αναζήτηση νέων ιδεών από το τηλέφωνο τους ενώ κάνουν μια συγκεκριμένη δραστηριότητα[12]. Για παράδειγμα, οι χρήστες αναζητούν μέρη, καταστήματα ή επιχειρήσεις που βρίσκονται κοντά τους συνήθως όταν ταξιδεύουν είτε κατά τη διάρκεια του Σαββατοκύριακου που βρίσκονται έξω για τις συνηθισμένες τους δραστηριότητες. Επιπλέον, παρατηρείται ότι κατά τη διάρκεια ενός συνηθισμένου Σαββατοκύριακου, υπάρχουν πολλές στιγμές οι οποίες είναι κατάλληλες για την πυροδότηση μιας νέας ανάγκης για τον χρήστη. Αναλυτικότερα, ρίχνοντας μια πιο προσεκτική ματιά σε αναζητήσεις με βάση την τοποθεσία κατά τη διάρκεια του Σαββατοκύριακου, διαπιστώνεται ότι το Σάββατο είναι συχνότερες οι αναζητήσεις για κινηματογραφικές αίθουσες και ινστιτούτα αισθητικής, ενώ πιο αργά το βράδυ της ίδιας μέρας πληθαίνουν οι αναζητήσεις εξόδου για ποτό και για πίτσα[11].

Κατά συνέπεια, η συγκεκριμένη διπλωματική εργασία θα μπορούσε να αποτελέσει εργαλείο για την καλύτερη ανάπτυξη του mobile marketing όσον αφορά τον μηχανισμό προτάσεων προς τον χρήστη. Η γνώση των συνηθειών του χρήστη θα μπορούσε να παράγει συναφείς και άμεσες προτάσεις με βάση τις δραστηριότητες του και την χρονική στιγμή που επιλέγει να κάνει κάτι. Με αυτόν τον τρόπο, περιορίζονται οι ενέργειες από μέρος του χρήστη καθώς οι προτάσεις είναι αυτόματες, άμεσες και καλύπτουν τις ανάγκες του, προτείνοντας του προϊόντα που ταιριάζουν στη διάθεση του και τις προτιμήσεις του.

Κεφάλαιο 2

Θεωρητικό Υπόβαθρο

Η γνώση των δραστηριοτήτων και των συνηθειών του χρήστη είναι πολύ σημαντική για την ανάπτυξη προσωποποιημένων εφαρμογών, οι οποίες είναι ωφέλιμες σε πολλούς τομείς της καθημερινής του ζωής [24]. Όπως αναφέρουν οι Maggiore και συν. 2014, τα σημεία ενδιαφέροντος που επισκέπτεται ένας χρήστης, τα οποία παρατηρούνται από τα GPS ίχνη κατά την κίνηση του, μπορούν να χρησιμεύσουν στην παροχή κατάλληλων πληροφοριών σε συγκεκριμένες χρονικές στιγμές ή σε συγκεκριμένες τοποθεσίες. Πιο συγκεκριμένα, η εξαγωγή σημείων ενδιαφέροντος, επιτρέπει σε μια έξυπνη συσκευή να προσφέρει εξατομικευμένες και συναφείς πληροφορίες στον χρήστη χωρίς την ανάγκη εκτέλεσης ερωτήματος αναζήτησης από τον χρήστη.

Παράλληλα, σύμφωνα με τους Parliatseyeu και συν. 2009, η αναγνώριση δραστηριότητας είναι σημαντική για την υποβοηθούμενη διαβίωση και την υγειονομική περίθαλψη. Αυτές οι μέθοδοι μπορούν να χρησιμοποιηθούν για να υποστηρίξουν τη μνήμη των ηλικιωμένων, ενώ επιτρέπουν την έγκαιρη διάγνωση πιθανών προβλημάτων υγείας. Τέλος, η πρόβλεψη της δραστηριότητας του χρήστη αλλά και η θέση του μπορεί να χρησιμοποιηθεί για τη βελτίωση της δρομολόγησης και τη συνολική απόδοση των ασύρματων δικτύων.

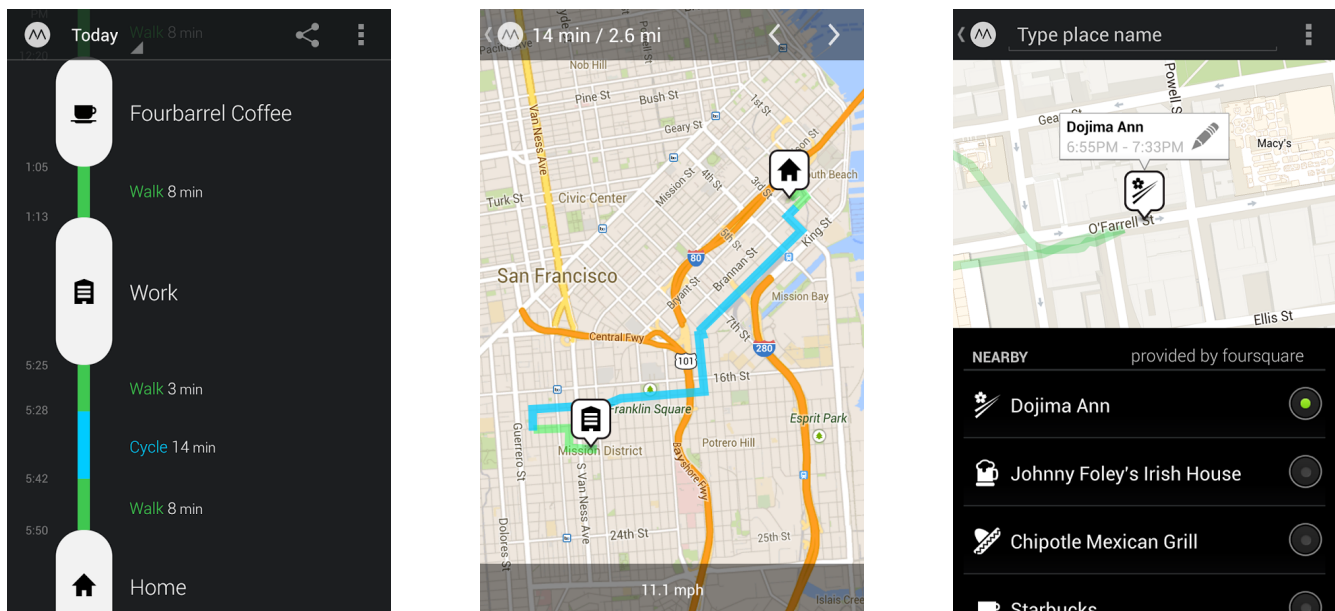
Επιπλέον, οι Phithakkitnukoon και συν. 2014, πραγματοποιούν έρευνα ώστε να εξετάσουν την κινητικότητα των τουριστών, καθώς όπως αναφέρουν οι καθημερινές μετακινήσεις των χρηστών είναι αρκετά προβλέψιμες σε αντίθεση με τις μετακινήσεις των τουριστών οι οποίες παρουσιάζουν ενδιαφέρον. Συνεπώς, η κατανόηση της συμπεριφοράς των τουριστών είναι σημαντική για τον αστικό σχεδιασμό, τη διαχείριση των μεταφορών και των τουριστικών αρχών.

Τα έξυπνα κινητά τηλέφωνα αποτελούν την ιδανική πλατφόρμα για την αναγνώριση των συνηθειών του χρήστη, καθώς διαθέτουν επαρκείς υπολογιστικούς πόρους, ασύρματη επικοινωνία και δυνατότητες πολυμέσων. Ακόμα, οι άνθρωποι συνηθίζουν να φέρουν μαζί τους τα κινητά

τους τηλέφωνα, συνεπώς ένα smartphone μπορεί να χρησιμοποιηθεί ως ο κύριος αισθητήρας και η μονάδα επεξεργασίας για την αναγνώριση της συμπεριφοράς του χρήστη[24].

2.1 Εφαρμογές Ανίχνευσης Συνηθειών

2.1.1 Moves



Σχήμα 2.1 Στιγμιότυπα χρήσης της εφαρμογής Moves (Πηγή: play.google.com¹).

Η εφαρμογή Moves καταγράφει την καθημερινή δραστηριότητα του χρήστη, ενώ υπολογίζει την κατανάλωση των θερμίδων κατά τη διάρκεια της ημέρας. Αναπτύχθηκε από την ProtoGeo, με στόχο να δώσει κίνητρο στον χρήστη να υιοθετήσει υγιεινές συνήθειες και την φυσική απώλεια βάρους, δίνοντας του την δυνατότητα να παρακολουθεί τις καθημερινές του δραστηριότητες[25].

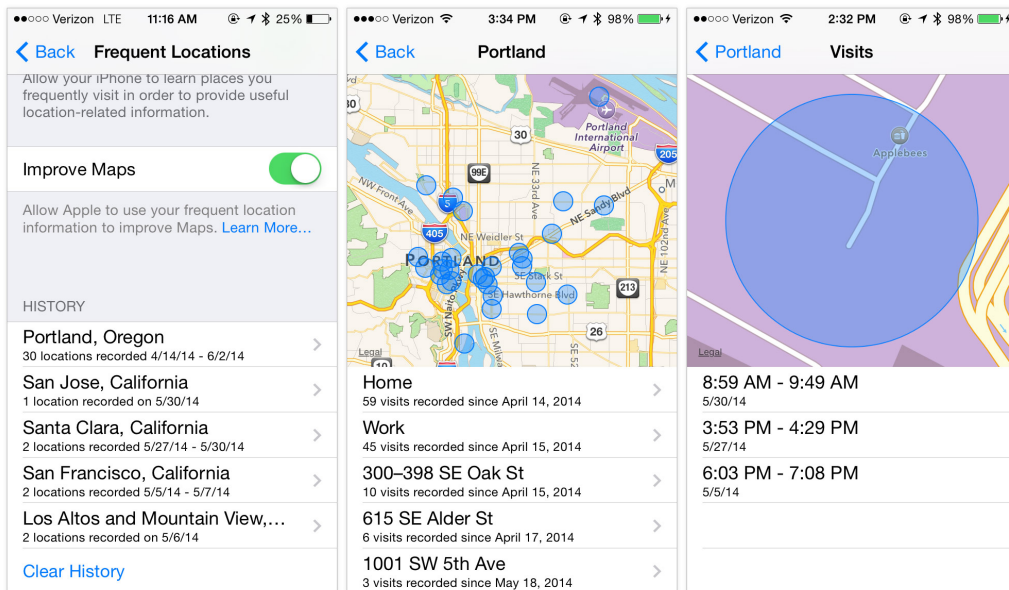
Κάνει χρήση των αισθητήρων της κινητής συσκευής καταγράφοντας τη θέση του χρήστη και αναγνωρίζει την δραστηριότητα, τις διαδρομές και τις τοποθεσίες από όπου διέρχεται ο χρήστης. Πιο συγκεκριμένα, η εφαρμογή αναγνωρίζει επιτυχώς το περπάτημα, την ποδηλασία και το τρέξιμο, ενώ κατηγοριοποιεί όλες τις μετακινήσεις με όχημα ως “μεταφορά”[13].

Η εφαρμογή καταγράφει αυτόματα τη θέση του χρήστη λειτουργώντας στο background χωρίς να προϋποθέτει ενέργεια από τον χρήστη. Τα δεδομένα που συλλέγει, μεταφορτώνονται σε

¹Άρθρο στο διαδίκτυο <https://play.google.com/store/apps/details?id=com.protogeo.moves>. Ελέγχθηκε 11/01/2016

servers όπου υπόκεινται σε επιπλέον επεξεργασία, για να υπολογιστούν τα καθημερινά στατιστικά του χρήστη και η απεικόνιση της κίνησης του σε χρονοδιάγραμμα. Η Moves βασίζεται σε αλγορίθμους αναγνώρισης δραστηριότητας και σημείων ενδιαφέροντος, οι οποίοι βέβαια παρουσιάζουν συχνά μικρές αποκλείσεις. Για αυτό τον λόγο, επιτρέπει στον χρήστη να διορθώσει τα σφάλματα που εμφανίζονται στην καθημερινή δραστηριότητα του, χρησιμοποιώντας τα σωστά δεδομένα ως δεδομένα εκπαίδευσης ώστε να αυξήσει το ποσοστό επιτυχίας της εφαρμογής[25].

2.1.2 Frequent Locations iOS



Σχήμα 2.2 Στιγμιότυπα χρήσης της εφαρμογής Frequent Locations (Πηγή: <https://aaronparecki.com>²).

Η Apple από την έκδοση 7 του λειτουργικού συστήματος iOS και μετά, έχει προεγκατεστημένη την εφαρμογή Frequent Locations, η οποία κάνει χρήση της υπηρεσίας εντοπισμού θέσης με στόχο να καταγράψει τα σημεία ενδιαφέροντος του χρήστη[3]. Η εφαρμογή εκτελείται όταν η υπηρεσία τοποθεσίας είναι ενεργοποιημένη και αναγνωρίζει τα μέρη που επισκέπτεται ο χρήστης, τη διάρκεια παραμονής του χρήστη σε κάθε σημείο αλλά και πόσο συχνά επισκέπτεται

²Άρθρο στο διαδίκτυο <https://aaronparecki.com/articles/2014/06/03/1/frequent-locations-api-in-ios-8>. Ελέγχθηκε 16/01/2016

κάθε σημείο[3][5]. Επιπλέον, ονοματίζει με ετικέτες κάποια από τα σημεία όπως "σπίτι" ή "γραφείο" χωρίς την εισαγωγή των διευθύνσεων από τον χρήστη. Η εφαρμογή, δημιουργεί επίσης μια λίστα από τις πόλεις που έχει επισκεφθεί ο χρήστης, και επιλέγοντας μια από αυτές εμφανίζει όλα τα σημεία που έχει επισκεφθεί σε αυτή την πόλη.

Όσον αφορά τη χρήση των δεδομένων που καταγράφονται μέσω αυτής της υπηρεσίας, η Apple αναφέρει ότι τα δεδομένα αυτά διατηρούνται αποκλειστικά και μόνο στη συσκευή του χρήστη και δεν αποστέλλονται στην Apple χωρίς τη συγκατάθεση του. Τα δεδομένα χρησιμοποιούνται για να παρέχουν στον χρήστη εξατομικευμένες υπηρεσίες, όπως για παράδειγμα την ενημέρωση του για τον χρόνο που απαιτείται για να επιστρέψει από το γραφείο στο σπίτι με βάση την κίνηση στους δρόμους[5].

2.2 Σχετικές Έρευνες

2.2.1 Αισθητήρες Ανίχνευσης Συνηθειών

Όσον αφορά στους αισθητήρες που χρησιμοποιούνται για την αναγνώριση της δραστηριότητας του χρήστη, έχει πραγματοποιηθεί πλήθος ερευνών στις οποίες παρουσιάζονται διαφορετικές προσεγγίσεις χρήσης των αισθητήρων ενός smartphone, ανάλογα με τον σκοπό που έχει ο κάθε ερευνητής. Σύμφωνα με τους Papliatseyeu και συν. 2009, δεν υπάρχει ιδανική τεχνολογία προσδιορισμού θέσης. Ο αισθητήρας GPS παρέχει ακρίβεια μεγαλύτερη των 10 μέτρων, αλλά είναι διαθέσιμος μόνο σε εξωτερικούς χώρους χωρίς να παρεμποδίζεται η θέα του ουρανού. Το δίκτυο Wi-Fi ακόμα και αν παρουσιάζει μειωμένη μέση τιμή σφάλματος (2 μέτρα), διαθέτει πολύ περιορισμένη κάλυψη σε αραιοκατοικημένες περιοχές και σε αναπτυσσόμενες χώρες, ενώ τέλος η τεχνολογία GSM διαθέτει υψηλή κάλυψη αλλά πολύ χαμηλή ακρίβεια εκατοντάδων μέτρων.

Σε υπάρχουσες έρευνες που στοχεύουν στην αναγνώριση δραστηριότητας εντοπίζοντας τα σημεία ενδιαφέροντος του χρήστη, όπως οι Liu και συν. 2006 και οι Maggiore και συν. 2014, γίνεται χρήση μόνο του αισθητήρα GPS της κινητής συσκευής καθώς στηρίζονται αποκλειστικά σε προσδιορισμό συντεταγμένων. Τα αποτελέσματα των δύο ερευνών αποδεικνύουν ότι η υπηρεσία GPS είναι αποτελεσματική σε τέτοιου είδους υλοποιήσεις, καθώς εστιάζουν στην κίνηση του χρήστη σε εξωτερικούς χώρους, όπου τα δεδομένα GPS έχουν υψηλή ακρίβεια.

Από την άλλη πλευρά, οι Feng και συν. 2013 ισχυρίζονται ότι τα δεδομένα GPS περιέχουν συχνά σφάλμα καθώς είτε το σήμα χάνεται είτε διαθέτει μειωμένη ακρίβεια. Για αυτό τον λόγο, οι προσεγγίσεις που στοχεύουν να αναγνωρίσουν τον τύπο κίνησης του χρήστη με βάση την ταχύτητα κίνησης του, δεν αρκεί να στηρίζονται στον αισθητήρα GPS καθώς συχνά τα δεδομένα

που καταγράφουν είναι ανακριβή. Ακόμα, οι Hemminki και συν. 2013 αναφέρουν ότι ένας επιπλέον περιορισμός της υπηρεσίας GPS, είναι η υψηλή κατανάλωση της μπαταρίας της κινητής συσκευής που προκαλεί η χρήση της. Οι Phithakkitnukoon και συν. 2014, οι οποίοι στην έρευνα τους μελετούν την κινητικότητα των τουριστών με GPS, κάνουν λόγο για χρήση του αισθητήρα του επιταχυνσιόμετρου όταν ο χρήστης βρίσκεται σε περίοδο σχετικής στάσης κατά την διάρκεια της οποίας μπορεί να ανασταλεί η χρήση της υπηρεσίας GPS με στόχο να μειωθεί η κατανάλωση της μπαταρίας. Ταυτόχρονα, παρατηρείται ότι η ανάγκη της απρόσκοπτης επικοινωνίας με τους δορυφόρους του αισθητήρα GPS, παρουσιάζει προβλήματα όταν ο χρήστης μετακινείται με μέσα μαζικής μεταφοράς, ενώ καθιστά δύσκολη την διάκριση κίνησης ανάμεσα στα μηχανοκίνητα μέσα μεταφοράς [16].

Οι Feng και συν. 2013 θεωρούν ότι το επιταχυνσιόμετρο (accelerometer) ενός smartphone, μπορεί να εμπλουτίσει και να συμπληρώσει τα δεδομένα κίνησης που καταγράφονται από τον αισθητήρα GPS, εφόσον η συγκεκριμένη τεχνολογία δεν είναι το ίδιο ευαίσθητη σε σφάλματα όπως το GPS ενώ μπορεί να εκτιμήσει την πραγματική ταχύτητα και την μετατόπιση του χρήστη κατά μήκος των αξόνων x, y και z που διαθέτει. Ο σκοπός της έρευνας τους είναι να εξετάσει τα πλεονεκτήματα της χρήσης δεδομένων επιταχυνσιόμετρου σε συνδυασμό με δεδομένα GPS για την αναγνώριση του τρόπου μεταφοράς του χρήστη. Πιο συγκεκριμένα, εξετάζουν τρεις διαφορετικές προσεγγίσεις : α) μόνο δεδομένα GPS β) μόνο δεδομένα επιταχυνσιόμετρου και γ) συνδυασμός δεδομένων GPS και επιταχυνσιόμετρου. Τα αποτελέσματα τους δείχνουν ότι ο συνδυασμός των δεδομένων των δυο αισθητήρων ξεπερνούν σε ακρίβεια τις άλλες δυο προσεγγίσεις στην αναγνώριση του μέσου μεταφοράς που βρίσκεται ο χρήστης.

Επιπλέον, Hemminki και συν. 2013 στην έρευνα τους κάνουν χρήση του επιταχυνσιόμετρου για να αναγνωρίσουν το μέσο μεταφοράς με το οποίο κινείται ο χρήστης. Όπως περιγράφουν, το επιταχυνσιόμετρο έχει πολύ χαμηλή κατανάλωση ενέργειας που επιτρέπει τη συνεχή παρακολούθηση του χρήστη στα μέσα μαζικής μεταφοράς. Ακόμα, το επιταχυνσιόμετρο είναι ένας αισθητήρας που υπολογίζει άμεσα τις κινήσεις του χρήστη χωρίς να εξαρτάται από εξωτερικές πηγές σήματος, ενώ καταγράφει εξαιρετικά λεπτομερείς πληροφορίες για την κίνηση της συσκευής επιτρέποντας την διάκριση μεταξύ των μηχανοκίνητων μέσων μεταφοράς παρά τις λεπτές διαφορές τους.

Οι Parliatseyeu και συν. 2009, προτείνουν μια διαφορετική τεχνική αναγνώρισης δραστηριότητας, συνδυάζοντας πολλαπλές τεχνολογίες προσδιορισμού τοποθεσίας όπως GPS, GSM, Wi-Fi και Bluetooth, στοχεύοντας σε πλήρη κάλυψη και ακρίβεια στον εντοπισμό της θέσης του χρήστη. Πιο συγκεκριμένα, η προσέγγιση που ακολουθούν είναι γνωστή ως «sensor fusion» (συγχώνευση αισθητήρων) και συνδυάζει δεδομένα από πολλαπλούς αισθητήρες θέσης με σκο-

πό να αυξήσει την ακρίβεια του συστήματος. Όπως αναφέρουν, το συγκεκριμένο σύστημα είναι ικανό να εντοπίζει νέα είδη δραστηριότητας καθώς αυξάνει τη συνολική ακρίβεια αναγνώρισης κίνησης του χρήστη.

2.2.2 Τεχνικές Ανίχνευσης Συνηθειών

Όσον αφορά την ανίχνευση των συνηθειών ενός χρήστη, έχει διεξαχθεί πλήθος σχετικών ερευνών, οι οποίες προσπαθούν να ανιχνεύσουν την κινητικότητα του χρήστη βασιζόμενες κυρίως στις τοποθεσίες από όπου διέρχεται ο χρήστης. Η μεθοδολογία που ακολουθείται από τις σχετικές έρευνες, εστιάζει αρχικά στον εντοπισμό σημείων ενδιαφέροντος του χρήστη και στη συνέχεια στις μεταβάσεις του χρήστη μεταξύ αυτών των σημείων για να εξαχθούν οι συχνές μετακινήσεις του χρήστη και κατά συνέπεια τις συνήθειες του. Τα σημεία ενδιαφέροντος ορίζονται ως οι τοποθεσίες που είναι σημαντικές για έναν μεμονωμένο χρήστη και μπορεί να τους εκχωρηθεί μια σημασιολογική έννοια για παράδειγμα σπίτι, δουλειά[24].

Οι Maggiore και συν. 2014 στην έρευνα τους, υλοποιούν μια εφαρμογή για κινητά τηλέφωνα, η οποία καταγράφει την κίνηση του χρήστη με χρήση του αισθητήρα GPS και στη συνέχεια εντοπίζει σημεία ενδιαφέροντος του χρήστη. Το σύστημα που ανέπτυξαν βασίζεται σε density - based συσταδοποίηση για να εντοπίσουν χρονικά διαστήματα κατά τα οποία ο χρήστης κινούνταν ή ήταν ακίνητος. Παρατηρείται ότι οι συστάδες στις οποίες ο χρήστης είχε σταθερή ταχύτητα, δηλαδή ήταν ακίνητος, είχαν κυκλικό σχήμα, ενώ όταν ο χρήστης βρισκόταν σε κίνηση, οι αντίστοιχες συστάδες είχαν ελλειψοειδές σχήμα. Οι συστάδες με κυκλικό σχήμα, θεωρούνται τα σημεία ενδιαφέροντος του χρήστη, ενώ στη συνέχεια χρησιμοποιείται η διαδικασία Markov σε μια σειρά πιθανοτήτων για κάθε πιθανή σημασιολογική ετικέτα ώστε να εκχωρηθεί μια έννοια σε αυτές τις τοποθεσίες.

Μια διαφορετική προσέγγιση παρουσιάζεται από τους Ashbrook και συν. 2003, οι οποίοι θεωρούν ότι ένα σημείο ενδιαφέροντος χαρακτηρίζεται από μεγάλη διάρκεια παραμονής του χρήστη στην ίδια τοποθεσία. Στην έρευνα τους χρησιμοποιούν την υπηρεσία GPS για τον προσδιορισμό της θέσης του χρήστη, συνεπώς ορίζουν μια σημαντική τοποθεσία όταν παρατηρήσουν ότι η ροή των καταγεγραμμένων δεδομένων διακόπτεται καθώς χάνεται το σήμα GPS και συμπεραίνουν ότι ο χρήστης εισήλθε σε κτίριο όπου το GPS δεν είναι διαθέσιμο. Θεωρούν μια τέτοια τοποθεσία ως σημείο ενδιαφέροντος όταν η στάση του χρήστη στο συγκεκριμένο σημείο διαρκεί περισσότερο από δέκα λεπτά της ώρας. Παράλληλα, όπως αναφέρουν πολλές μετρήσεις GPS που λαμβάνονται στην ίδια φυσική θέση, συχνά δεν καταγράφονται ως το ίδιο σημείο συντεταγμένων, το οποίο σημαίνει ότι ο χρήστης μπορεί βρίσκεται σε σταθερό σημείο για μεγάλο χρονικό

διάστημα αλλά δεν μπορεί να ανιχνευτεί με την ίδια τεχνική ως σημαντικό σημείο. Για αυτό τον λόγο, εφαρμόζουν τον αλγόριθμο συσταδοποίησης k-means στις θέσεις που καταγράφονται και θεωρούν ως σημεία ενδιαφέροντος τα κέντρα κάθε συστάδας.

Οι Phithakkitnukoon και συν. 2014, διερευνούν τα σημεία ενδιαφέροντος με μια προσέγγιση τριών βημάτων, προσπαθώντας να ανιχνεύσουν την κατοικία και τον χώρο εργασίας του χρήστη. Αρχικά, προσδιορίζουν ως στάση ένα σύνολο καταγεγραμμένων θέσεων GPS σε κοντινή απόσταση, δηλαδή μια θέση στην οποία ο χρήστης ξοδεύει ένα σημαντικό ποσό του χρόνου. Στη συνέχεια, εφαρμόζουν spatial clustering των στάσεων του χρήστη και θεωρούν ότι το κέντρο (centroid) κάθε cluster είναι ένα σημείο ενδιαφέροντος. Έπειτα από πειράματα σε τεχνικές συσταδοποίησης συμπέραναν ότι ο αλγόριθμος DBSCAN ήταν ο πιο αποτελεσματικός στα δεδομένα τους. Το τελευταίο στάδιο έχει ως στόχο να χαρακτηρίσει τα σημεία ενδιαφέροντος που ανιχνεύθηκαν είτε ως «σπίτι» είτε ως «δουλειά». Για αυτό τον λόγο, χρησιμοποιείται ο αλγόριθμος κατηγοριοποίησης Random Forest για να χαρακτηρίσει τις τοποθεσίες με βάση μερικά χαρακτηριστικά όπως τον αριθμό των ωρών ή τον αριθμό των ημερών που παρέμεινε σε εκείνο το σημείο ο χρήστης.

Αφού ολοκληρωθεί ο εντοπισμός των σημείων ενδιαφέροντος του χρήστη, ακολουθεί η διερεύνηση των συχνών μετακινήσεων που πραγματοποιεί ο χρήστης μεταξύ αυτών των σημείων, με στόχο την ανίχνευση των καθημερινών συνηθειών του. Οι Maggiore και συν. 2014 χρησιμοποιούν το ιστορικό των τοποθεσιών που επισκέφθηκε ο χρήστης για να δημιουργήσουν μια διαδικασία Markov. Οι κόμβοι της αλυσίδας Markov, αποτελούνται από τα clusters που εντοπίστηκαν στο προηγούμενο βήμα, ενώ οι μεταβάσεις μεταξύ των κόμβων είναι οι μετακινήσεις του χρήστη που έχουν καταγραφεί μέχρι εκείνη τη στιγμή, για τις οποίες ταυτόχρονα διατηρείται και σε ξεχωριστό πίνακα η διάρκεια τους. Το σύστημα τους δοκιμάστηκε σε δεδομένα κίνησης μαθητών και τα αποτελέσματα τους έδειξαν ότι με τη συγκεκριμένη τεχνική αναγνωρίστηκαν με επιτυχία μετακινήσεις των χρηστών από την πρωτεύουσα στη δευτερεύουσα κατοικία τους, από το σπίτι στο σχολείο και από το σπίτι στη βραδινή έξοδο.

Οι Ashbrook και συν. 2003, ακολουθούν μια παρόμοια μεθοδολογία ανίχνευσης συνηθειών, δημιουργώντας ένα μοντέλο που εντοπίζει συχνά εμφανιζόμενα μοτίβα κίνησης και κατά συνέπεια είναι σημαντικά για τον χρήστη. Μετά την εξαγωγή των σημείων ενδιαφέροντος από τα δεδομένα του κάθε χρήστη, εκχωρούν σε κάθε τοποθεσία ένα μοναδικό αναγνωριστικό ενώ διατηρούν την χρονολογική σειρά επίσκεψής τους. Στη συνέχεια, δημιουργούν ένα μοντέλο Markov για κάθε τοποθεσία με μεταβάσεις για κάθε άλλο καταγεγραμμένο σημείο. Κάθε κόμβος στο μοντέλο Markov αποτελείται από μια τοποθεσία και κάθε μετάβαση αντιπροσωπεύει την πιθανότητα ο χρήστης να μετακινηθεί μεταξύ αυτών των δυο σημείων. Η πιθανότητα μετάβασης

μεταξύ δύο σημείων αυξάνεται όταν πρόκειται για μια συνηθισμένη μετακίνηση του χρήστη, ενώ αν ο χρήστης δεν έχει ταξιδέψει ποτέ μεταξύ δυο σημείων η πιθανότητα είναι μηδενική.

Τέλος, οι Papliatseyeu και συν. 2009, υποστηρίζουν ότι οι υπάρχουσες τεχνικές αναγνώρισης συνηθειών που αναπαριστούν τα σημεία ενδιαφέροντος με κόμβους και τις διαδρομές μεταξύ αυτών με ακμές, αγνοούν μερικούς σημαντικούς παράγοντες των μετακινήσεων, όπως η χρονική στιγμή και η συχνότητα των διαδρομών. Για παράδειγμα, αναφέρουν ότι η πιθανότητα μετάβασης από το σπίτι στον χώρο εργασίας κατά τη διάρκεια του Σαββατοκύριακου είναι αδιαμφισβήτητα μικρότερη από τις καθημερινές μέρες, παράγοντας που δεν λαμβάνεται υπόψιν στις υπάρχουσες έρευνες. Για αυτό τον λόγο, προτείνουν τη διερεύνηση χρήσης του μετασχηματισμού Fourier για την αναγνώριση των συνηθειών των χρηστών.

2.3 Προτεινόμενη Τεχνική

Η ανίχνευση των συνηθειών του χρήστη στην παρούσα εργασία υλοποιείται μέσω της αναγνώρισης των σημείων ενδιαφέροντος του χρήστη τα οποία σε δεύτερο επίπεδο χαρακτηρίζονται ώστε να εξαχθούν οι τοποθεσίες που συνηθίζει να επισκέπτεται ο χρήστης. Ταυτόχρονα, κάθε σημείο ενδιαφέροντος που αναγνωρίζεται θα συνδυάζεται με χαρακτηριστικά σχετικά με την χρονική στιγμή που βρίσκεται ο χρήστης στην συγκεκριμένη τοποθεσία, ώστε να μπορούμε να καταλήξουμε σε συνήθειες που έχει στην καθημερινή του ζωή.

Η συγκεκριμένη προσέγγιση διαφέρει από τις ήδη υπάρχουσες έρευνες που περιγράφηκαν παραπάνω, καθώς δεν εξάγει τις συνήθειες του χρήστη εξετάζοντας τις μεταβάσεις μεταξύ των σημείων ενδιαφέροντος αλλά εστιάζει στον χαρακτηρισμό των σημείων ενδιαφέροντος και στη χρονική στιγμή που ο χρήστης συνηθίζει να βρίσκεται σε κάθε σημείο.

Η συλλογή των δεδομένων πραγματοποιείται μέσω της εφαρμογής GPSTracker η οποία καταγράφει χαρακτηριστικά της κίνησης και της θέσης του χρήστη όπως τις συντεταγμένες της θέσης του, την στιγμιαία και τη σχετική ταχύτητα του και το υψόμετρο που βρίσκεται χρησιμοποιώντας τους αισθητήρες GPS και του επιταχυνσιόμετρου (accelerometer). Επιπλέον, η εφαρμογή καταγράφει χαρακτηριστικά της χρονικής στιγμής που ο χρήστης διέρχεται από ένα σημείο όπως η μέρα της βδομάδας, αν πρόκειται για εργάσιμη μέρα ή Σαββατοκύριακο αλλά και την συγκεκριμένη χρονική στιγμή που ο χρήστης βρέθηκε σε κάθε σημείο.

Για την υλοποίηση της συγκεκριμένης εφαρμογής, θα χρησιμοποιηθούν χαρακτηριστικά της θέσης του χρήστη που προκύπτουν από τον αισθητήρα GPS, δηλαδή οι συντεταγμένες που βρίσκεται ο χρήστης σε κάθε στιγμιότυπο, καθώς σύμφωνα με παρόμοιες έρευνες των Liu και συν. 2006 και Maggiore και συν. 2014, ο αισθητήρας GPS είναι αποτελεσματικός στην ανα-

γνώριση σημείων ενδιαφέροντος διότι παρέχει υψηλή ακρίβεια στην θέση του χρήστη σε εξωτερικούς χώρους. Επίσης, η παρούσα έρευνα δεν στοχεύει στην αναγνώριση του μέσου μεταφοράς του χρήστη καθώς αγνοεί τις μεταβάσεις του μεταξύ των σημείων ενδιαφέροντος, συνεπώς η χρήση του αισθητήρα του επιταχυνσιόμετρου δεν προσδίδει κάποια επιπλέον χρήσιμη πληροφορία για την ανάλυση που απαιτεί η συγκεκριμένη εργασία.

Η ανίχνευση των σημείων ενδιαφέροντος του χρήστη θα υλοποιηθεί εφαρμόζοντας density-based συσταδοποίηση στα σημεία που ορίζονται από τα ζεύγη συντεταγμένων που καταγράφονται. Για τις συστάδες που θα προκύψουν, θα υπολογιστεί το centroid (κέντρο βάρους) της κάθε συστάδας και στη συνέχεια θα δημιουργηθεί ένα bounding box (παραλληλόγραμμο) γύρω από το centroid. Το bounding box που θα οριστεί θα χρησιμοποιηθεί για τον χαρακτηρισμό του σημείου ενδιαφέροντος μέσω του OpenStreetMaps API. Αξιοποιώντας δεδομένα που καταγράφονται από την εφαρμογή GPSTracker, τα σημεία ενδιαφέροντος που θα ανιχνευθούν θα συνδυαστούν με χαρακτηριστικά της εκάστοτε τρέχουσας χρονικής στιγμής, όπως η μέρα της εβδομάδας και αν πρόκειται για εργάσιμη μέρα ή όχι. Τέλος, θα εφαρμοστούν τεχνικές κανόνων συσχέτισης στα συνδυασμένα δεδομένα, με στόχο την εύρεση συχνών στοιχείων του συνόλου δεδομένων, τα οποία κατά συνέπεια αποτελούν τις συνήθειες του χρήστη.

Αναλύοντας με αυτόν τον τρόπο τα δεδομένα κίνησης του χρήστη σε καθημερινή βάση, μπορούν να εξαχθούν συμπεράσματα για τις δραστηριότητες και τις προτιμήσεις του χρήστη. Η γνώση αυτών των πληροφοριών μπορεί να χρησιμοποιηθεί για την ανάπτυξη ενός συστήματος συστάσεων, το οποίο επεκτείνοντας την υπάρχουσα εργασία θα προτείνει στον χρήστη νέες δραστηριότητες και πιθανά σημεία ενδιαφέροντος με βάση τις προσωπικές προτιμήσεις του χρήστη.

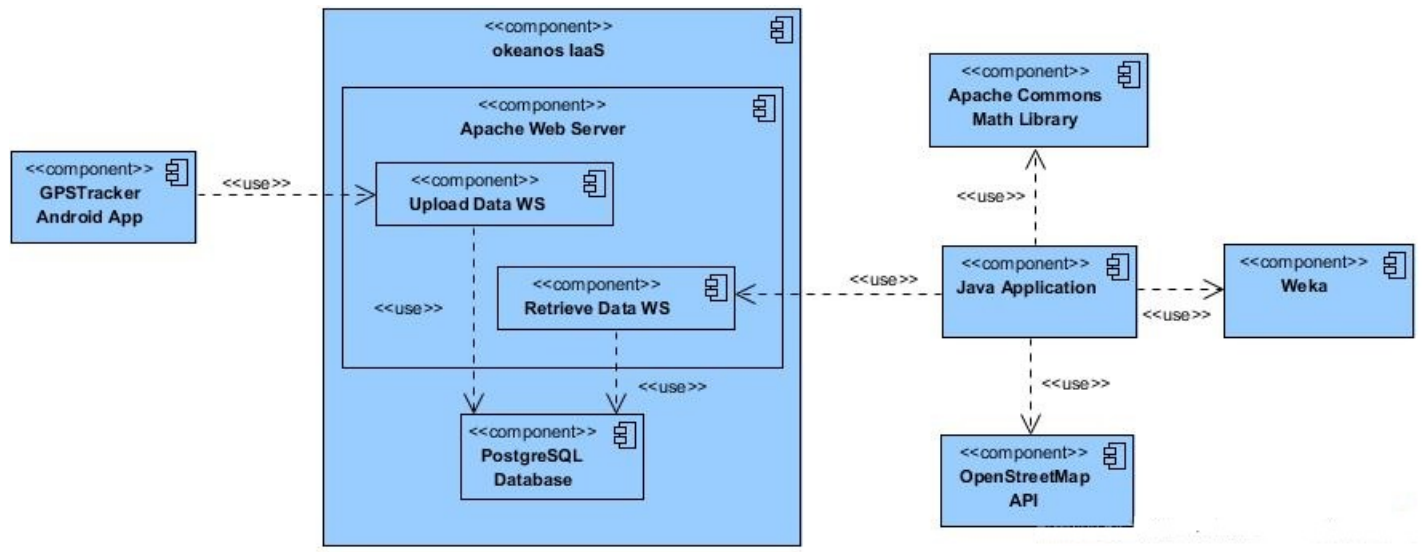
Κεφάλαιο 3

Σχεδιασμός

3.1 Αρχιτεκτονική Εφαρμογής

Η αρχιτεκτονική της εφαρμογής της παρούσας διπλωματικής εργασίας, βασίζεται σε τέσσερα συστατικά που υλοποιούν τις λειτουργίες της μεταφόρτωσης των δεδομένων κίνησης σε ένα κεντρικό αποθετήριο, της ανάκτησης των δεδομένων από το αποθετήριο με βάση τον χρήστη, την ανάλυση των δεδομένων και τον εντοπισμό των σημείων ενδιαφέροντος του χρήστη και τέλος τον χαρακτηρισμό των σημείων ενδιαφέροντος με στόχο την ανίχνευση των συνηθειών του χρήστη. Πιο συγκεκριμένα, η αρχιτεκτονική της παρούσας εφαρμογής αποτελείται από μια εφαρμογή κινητής συσκευής σε Android (GPSTracker), από έναν Apache Web Server και μια Βάση Δεδομένων PostgreSQL και τέλος από μια εφαρμογή Java όπου πραγματοποιείται η ανάλυση των δεδομένων και η εξαγωγή γνώσης.

Στο σχήμα 3.1 παρουσιάζεται η αρχιτεκτονική της εφαρμογής με τα στοιχεία που την απαρτίζουν αλλά και οι αλληλεπιδράσεις μεταξύ τους. Αρχικά, τα δεδομένα κίνησης που καταγράφονται από την εφαρμογή GPSTracker μεταφορτώνονται σε μια Βάση Δεδομένων PostgreSQL με την κλήση του Upload Data Web Service. Για την ανάλυση των δεδομένων, γίνεται κλήση του Retrieve Data Web Service και ανακτώνται τα δεδομένα κίνησης ανά χρήστη. Στη συνέχεια, πραγματοποιείται ο εντοπισμός των σημείων ενδιαφέροντος κάνοντας χρήση της βιβλιοθήκης Apache Commons Math. Εφόσον εντοπιστούν τα σημεία ενδιαφέροντος χρησιμοποιείται το OpenStreetMap API, το οποίο προσδίδει μια περιγραφή χαρακτηρίζοντας τα σημεία ενδιαφέροντος. Τέλος, συνδυάζοντας τα σημεία ενδιαφέροντος με χαρακτηριστικά της κίνησης του χρήστη εξάγονται οι συνηθείες του κάνοντας χρήση τεχνικών εξόρυξης δεδομένων μέσω του περιβάλλοντος Weka.



Σχήμα 3.1 Η αρχιτεκτονική της προτεινόμενης προσέγγισης.

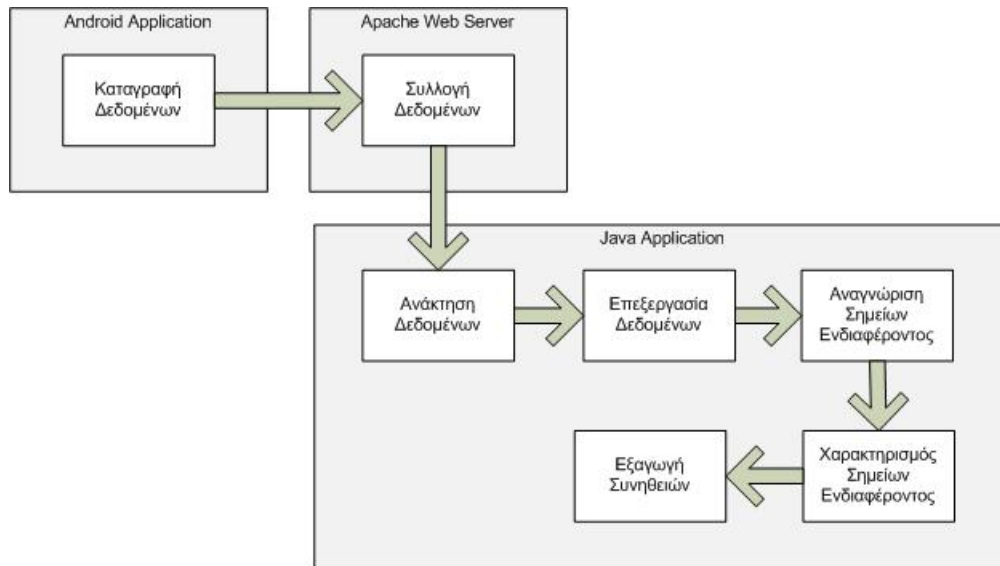
3.2 Μεθοδολογία Υλοποίησης Εφαρμογής

Στην παρούσα ενότητα, θα γίνει περιγραφή της μεθοδολογίας που ακολουθείται για την αναγνώριση συνηθειών ενός χρήστη. Η προσέγγιση που χρησιμοποιείται περιλαμβάνει διακριτές διαδικασίες οι οποίες παρουσιάζονται στο Σχήμα 3.2 και θα αναλυθούν στη συνέχεια.

Το πρώτο στάδιο αποτελεί η *Καταγραφή των Δεδομένων* η οποία πραγματοποιείται στην κινητή συσκευή μέσω της εφαρμογής GPSTracker, προϋποθέτοντας την αλληλεπίδραση με τον χρήστη. Ο επόμενος στόχος είναι η κεντριοποιημένη *Συλλογή των Δεδομένων* από την κινητή συσκευή σε απομακρυσμένο αποθετήριο, υλοποιημένη με διάφανο τρόπο διευκολύνοντας την εμπειρία του χρήστη κατά τη χρήση της εφαρμογής.

Όσον αφορά τη διαδικασία ανάλυσης των δεδομένων που καταγράφονται με σκοπό την εξαγωγή συνηθειών, είναι σημαντικό να αναλυθεί λεπτομερώς, ώστε να γίνει κατανοητή η προσέγγιση της παρούσας εργασίας. Επίσης, η διαδικασία περιγράφεται σε μορφή αλγορίθμου (Αλγόριθμος 3.1). Αρχικά, πραγματοποιείται η *Ανάκτηση των Δεδομένων* από το κεντρικό αποθετήριο σε επίπεδο χρήστη, έτσι ώστε να εξαχθούν συμπεράσματα για κάθε χρήστη. Στη συνέχεια, τα δεδομένα επεξεργάζονται (*Επεξεργασία Δεδομένων*) και απομονώνονται τα σημαντικά χαρακτηριστικά κάθε στίγματος, δηλαδή πληροφορίες που αποδίδουν γνώση χρήσιμη για τα επόμενα στάδια της ανάλυσης τους. Πιο συγκεκριμένα, για κάθε καταγεγραμμένο στίγμα του χρήστη δημιουργείται ένα σημείο με βάση τις υπάρχουσες συντεταγμένες, ενώ ταυτόχρονα χρησιμο-

ποιούνται χαρακτηριστικά όπως η ώρα της ημέρας, το είδος της κίνησης του χρήστη και το αν το στίγμα καταγράφηκε σε εργάσιμη ημέρα ή κατά τη διάρκεια Σαββατοκύριακου. Τα συγκεκριμένα χαρακτηριστικά επιλέχθηκαν καθώς αποδίδουν πληροφορίες κατάλληλες για την εξαγωγή συνηθειών.



Σχήμα 3.2 Διάγραμμα Ροής Διαδικασίας Ανίχνευσης Συνηθειών.

Κατόπιν, πραγματοποιείται η *Αναγνώριση των Σημείων Ενδιαφέροντος*, με χρήση της τεχνικής της συσταδοποίησης στα σημεία που δημιουργήθηκαν στο προηγούμενο βήμα. Ο αλγόριθμος DBSCAN επιλέχθηκε για την διαδικασία συσταδοποίησης των δεδομένων καθώς θεωρήθηκε αποτελεσματικότερος για τα συγκεκριμένα δεδομένα για πολλαπλούς λόγους. Αρχικά, ο DBSCAN αποτελεί έναν αλγόριθμο συσταδοποίησης με βάση την πυκνότητα (density-based), το οποίο σημαίνει ότι είναι καταλληλός για την αναγνώριση σημείων ενδιαφέροντος, δηλαδή σημείων στα οποία ο χρήστης ξοδεύει αρκετό χρόνο και κατά συνέπεια σημείων που περιέχουν πυκνά ίχνη του χρήστη. Επιπλέον, ο συγκεκριμένος αλγόριθμος αντιμετωπίζει τις αραιές περιοχές ως θόρυβο, προσέγγιση που συμβαδίζει με τον στόχο της εργασίας. Ακόμη, η εκτέλεση του αλγορίθμου δεν προϋποθέτει την εισαγωγή συγκεκριμένου αριθμού συστάδων προς εντοπισμό όπως συμβαίνει με τον αλγόριθμο k-means. Τέλος, ο DBSCAN έχει την ιδιότητα να αναγνωρίζει συστάδες με αυθαίρετο σχήμα, γεγονός που αυξάνει την ακρίβεια εντοπισμού σημείων ενδιαφέροντος, καθώς ο χρήστης μπορεί να κινείται γύρω από ένα σημείο με διάφορους τρόπους και προς πολλές κατευθύνσεις[18].

Έπειτα, εκτελείται η διαδικασία του χαρακτηρισμού των σημείων ενδιαφέροντος που ήδη

αναγνωρίστηκαν, διεξάγοντας κλήση στην πλατφόρμα OpenStreetMaps για ένα bounding box που ορίζεται από το σημείο ενδιαφέροντος με βάση το κέντρο κάθε συστάδας. Σε περίπτωση που ανακτηθούν περισσότερες από μια πληροφορίες για ένα σημείο ενδιαφέροντος, θεωρείται ως πιο κατάλληλη η πληροφορία που βρίσκεται πιο κοντά στο κέντρο του σημείου ενδιαφέροντος.

Algorithm 3.1 Detection of user habits using movement data

Input: users[i : 1...N], files[i,f : 1...M], lines [i,f,l : 1...L]**Output:** user_habits[i,j : 1...K]

```
for  $i \leftarrow 1$  to  $N$  do                                     /* for every user */
  for  $f \leftarrow 1$  to  $M$  do                                     /* files of users */
    for  $l \leftarrow 1$  to  $L$  do                                     /* lines of file */
       $attributes[] \leftarrow tokenizeByAttributes(line[i, f, l])$ 
       $point \leftarrow attributes[]$ 
       $pointList[] \leftarrow add(point)$ 
    end
  end
   $clusters \leftarrow performDBScan(pointList[], with clusters[j:1...C])$ 
  for  $j \leftarrow 1$  to  $C$  do                                     /* for every cluster */
     $centroid \leftarrow findCentroid(clusters[j])$ 
     $boundingBox \leftarrow calculateBoundingBox(centroid)$ 
     $xmlFile \leftarrow callOpenStreetMaps(boundingBox)$ 
     $tags[] \leftarrow getTags(xmlFile)$ 
     $closestTag \leftarrow getClosestTag(centroid, tags[])$ 
     $clusters[j] \leftarrow addTagToCluster(closestTag, clusters[j])$ 
  end
   $arffFile \leftarrow exportClustersToArff(clusters[])$ 
   $user\_habits[i] \leftarrow performApriori(arffFile)$ 
end
```

Τέλος, αφότου ολοκληρωθεί και η αναγνώριση των σημείων ενδιαφέροντος υλοποιείται η *Εξαγωγή Συνηθειών*, κάνοντας χρήση των στιγμάτων που απαρτίζουν τα σημεία ενδιαφέροντος. Πιο συγκεκριμένα, ο στόχος της συγκεκριμένης ενέργειας είναι να εξαχθούν οι πληροφορίες των σημείων ενδιαφέροντος που αναγνωρίστηκαν συνδυασμένες με χαρακτηριστικά που περιγράφουν την χρονική στιγμή που ο χρήστης βρέθηκε σε κάθε σημείο. Με αυτόν τον τρόπο, μπορούν να εξαχθούν οι συνήθειες του χρήστη εκτελώντας τεχνικές εύρεσης κανόνων συσχέτισης στο σύνο-

λο δεδομένων που περιγράφηκε. Για την εύρεση συνηθειών, επιλέχθηκε ο αλγόριθμος Apriori καθώς αρχικά είναι ο καταλληλότερος για τον εντοπισμό των στοιχείων που εμφανίζονται συχνά σε μια βάση δεδομένων, ενώ επιπλέον έχει σχεδιαστεί για σύνολα δεδομένων που εκφράζουν συναλλαγές με την γενικότερη έννοια [4]. Συνεπώς, ο αλγόριθμος Apriori μπορεί να εντοπίσει με επιτυχία τα συχνά στοιχεία σε ένα σύνολο από δεδομένα κίνησης και με αυτό τον τρόπο να εξαχθούν οι συνήθειες του χρήστη.

3.3 Εργαλεία Υλοποίησης Εφαρμογής

3.3.1 Okeanos IAAS Cloud

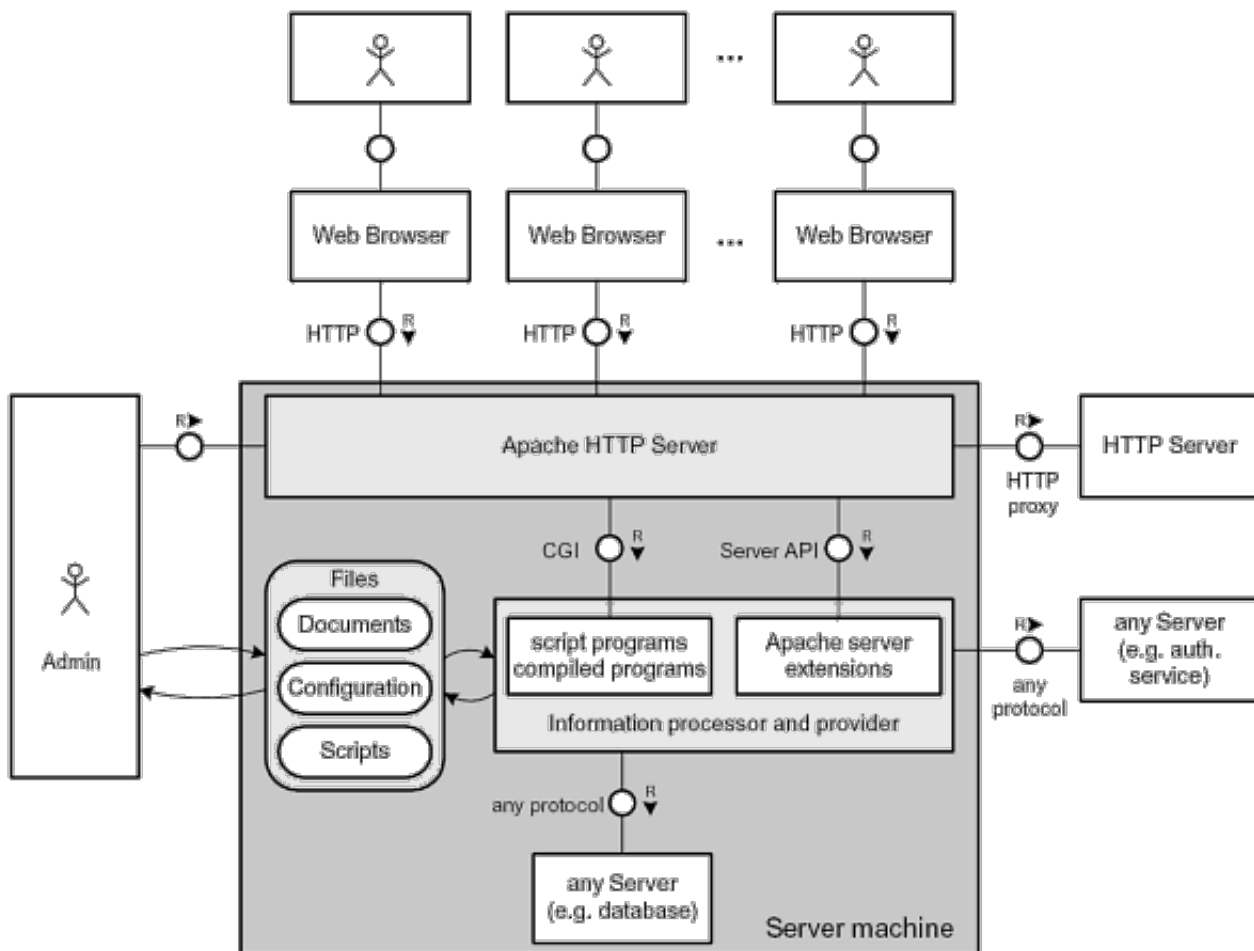
Το okeanos αποτελεί μια IAAS Cloud υπηρεσία, η οποία σχεδιάστηκε και αναπτύχθηκε από το Εθνικό Δίκτυο Έρευνας και Τεχνολογίας ΑΕ (ΕΔΕΤ) και παρέχεται δωρεάν στα Ελληνικά Πανεπιστήμια και στα δημόσια Ερευνητικά Κέντρα. Το λογισμικό okeanos αναπτύχθηκε βασισμένο σε υπάρχοντα λογισμικά ανοιχτού κώδικα, όπως το Google Ganeti, το οποίο επεκτάθηκε ώστε να παρέχει μια ολοκληρωμένη πλατφόρμα IAAS. Η πλατφόρμα okeanos δίνει τη δυνατότητα δημιουργίας εικονικών μηχανών αξιοποιώντας τα κεντρικά data center της ΕΔΕΤ ΑΕ, εξοικονομώντας τις προμήθειες διακομιστών (servers) στα εκπαιδευτικά ιδρύματα της χώρας [1].

3.3.2 Apache HTTP Server

Ο Apache Web Server, αποτελεί παγκοσμίως το πλέον χρησιμοποιούμενο λογισμικό εξυπηρετητή παγκόσμιου ιστού. Κυκλοφόρησε υπό την άδεια λογισμικού Apache και είναι ένα λογισμικό ανοιχτού κώδικα. Η πρώτη έκδοση του αναπτύχθηκε το 1995 από μια ομάδα προγραμματιστών βασισμένη στον NCSA httpd Web Server. Ο Apache έπαιξε καθοριστικό ρόλο στην αρχική ανάπτυξη του Παγκόσμιου Ιστού προσπερνώντας σύντομα τον NCSA HTTPd ως το κυρίαρχο διακομιστή HTTP και έχει παραμείνει ο δημοφιλέστερος από τον Απρίλιο του 1996 [;]. Η ανάπτυξη του λογισμικού του ακολούθησε την ίδια διαδρομή με το λειτουργικό σύστημα Linux καθώς δημιουργήθηκε αρχικά για το λειτουργικό σύστημα Unix ενώ οι νεότερες εκδόσεις υποστηρίζουν και διαφορετικά λειτουργικά συστήματα [21].

Ο Apache αναπτύχθηκε και υποστηρίζεται από μια ανοιχτή κοινότητα προγραμματιστών υπό την αιγίδα του Apache Software Foundation. Ο πηγαίος κώδικας του λογισμικού είναι διαθέσιμος δωρεάν και έτσι οι χρήστες μπορούν να προσαρμόσουν το λογισμικό στις ιδιαίτερες ανάγκες τους [21]. Χρησιμοποιείται κυρίως σε συστήματα που βασίζονται στο Unix (Linux πιο

συχνά), αν και το λογισμικό είναι διαθέσιμο για ένα ευρύ φάσμα λειτουργικών συστημάτων εκτός του Unix, συμπεριλαμβανομένων των eComStation, Microsoft Windows, NetWare, OpenVMS, OS / 2, και TPF [32].



Σχήμα 3.3 Το περιβάλλον του Apache Web Server.

Όλες οι διαθέσιμες εκδόσεις του Apache Web Server χρησιμοποιούνται από το μεγαλύτερο μερίδιο της αγοράς. Ακόμα και αν το γεγονός ότι είναι ελεύθερο λογισμικό αποτελεί ένα σημαντικό παράγοντα, ο κύριος λόγος για την επιτυχία του Apache είναι οι πολλαπλές λειτουργίες που παρέχει. Αρχικά, ο συγκεκριμένος server υποστηρίζει ταυτοχρονισμό και συνεπώς μπορεί να εξυπηρετήσει μεγάλο αριθμό client. Ο αριθμός των client που μπορούν να εξυπηρετηθούν ταυτόχρονα περιορίζεται μόνο από το hardware και το λειτουργικό του συστήματος όπου εκτελείται. Ο διακομιστής μπορεί εύκολα να ρυθμιστεί με επεξεργασία συγκεκριμένων αρχείων κειμένου ή χρησιμοποιώντας ένα από τα πολλά γραφικά περιβάλλοντα που είναι διαθέσιμα για τη δια-

χείρισή του. Επιπλέον, ο Apache είναι εύκολα επεκτάσιμος καθώς επιπρόσθετες λειτουργίες (extensions) μπορούν να αναπτυχθούν και να προστεθούν στον server δημιουργώντας έτσι έναν ισχυρό web server με πολλές δυνατότητες [14]. Ο Apache υποστηρίζει server scripting γλώσσες όπως η Perl, η Python και η PHP με τις οποίες μπορούν εύκολα να αναπτυχθούν web-based εφαρμογές [21]. Το σχήμα 3.3 δείχνει το περιβάλλον του Apache Web Server.

3.3.3 PostgreSQL

Η PostgreSQL είναι μια σχεσιακή βάση δεδομένων (ORDBMS) ανοιχτού κώδικα με κύριο χαρακτηριστικό την επεκτασιμότητα και την συμμόρφωση στα πρότυπα. Οι κύριες λειτουργίες της είναι η ασφαλής αποθήκευση των δεδομένων με τον βέλτιστο δυνατό τρόπο και η ανάκτηση των δεδομένων μέσω άλλων εφαρμογών λογισμικού. Έχει δυνατότητα υποστήριξης φόρτου που κυμαίνεται από μικρές εφαρμογές που εκτελούνται σε ένα μηχάνημα μέχρι μεγάλες web-based εφαρμογές με πολλαπλούς χρήστες.

Υλοποιεί το μεγαλύτερο μέρος του προτύπου SQL:2011, δηλαδή ακολουθεί το πρότυπο ACID (Atomicity, Consistency, Isolation, Durability), είναι transactional (συναλλακτική), αποφεύγει τα κλειδώματα με τη μέθοδο MVCC (Multiversion Concurrency Control), χειρίζεται με επιτυχία σύνθετα SQL ερωτήματα χρησιμοποιώντας τεχνικές ευρετήριασης που δεν είναι διαθέσιμες σε άλλες Βάσεις Δεδομένων. Επιπλέον, υποστηρίζει όψεις (views), σκανδάλες (triggers), ξένα κλειδιά (foreign Keys) και διαδικασίες (procedures) βάσεων δεδομένων.

Η PostgreSQL είναι λειτουργεί σε πολλά λειτουργικά συστήματα, συμπεριλαμβανομένων των Linux, FreeBSD, OS X, Solaris και Microsoft Windows. Αναπτύχθηκε από το PostgreSQL Global Development Group, μια ομάδα από πολλές εταιρείες και μεμονωμένους φορείς. Είναι ελεύθερο λογισμικό ανοιχτού κώδικα, και κυκλοφορεί υπό τους όρους της άδειας PostgreSQL, μια ειδική άδεια ελεύθερου λογισμικού [27][30].

3.3.4 Apache Commons Mathematics

Το Apache Commons είναι ένα έργο του Apache Software Foundation, που έχει ως σκοπό να παρέχει επαναχρησιμοποιούμενο λογισμικό ανοιχτού κώδικα Java. Η βιβλιοθήκη Commons Math πιο συγκεκριμένα, περιέχει μαθηματικές και στατιστικές μεθόδους που αφορούν τα πιο κοινά προβλήματα και δεν είναι διαθέσιμα στη γλώσσα προγραμματισμού Java[29][26].

3.3.5 OpenStreetMap

Το OpenStreetMap (OSM) είναι ένα συνεργατικό έργο για τη δημιουργία ενός επεξεργάσιμου παγκόσμιου χάρτη. Δημιουργήθηκε από τον Steve Coast το 2004 και συνεχίζει να αναπτύσσεται από μια κοινότητα χαρτογράφων που συνεισφέρουν και διατηρούν δεδομένα σχετικά με δρόμους, μονοπάτια, καφετέριες, σιδηροδρομικούς σταθμούς κτλ σε όλον τον κόσμο. Οι χρήστες συλλέγουν δεδομένα χρησιμοποιώντας αεροφωτογραφίες, συσκευές GPS, και τοπικούς χάρτες χαμηλής τεχνολογίας ώστε να εξασφαλιστεί η ακρίβεια των δεδομένων. Το OpenStreetMap είναι βάση ανοικτών δεδομένων που διατίθενται με άδεια Open Data Commons Open Database License (ODbL)[34][23].

3.3.6 Το περιβάλλον Weka

Το περιβάλλον WEKA (Waikato Environment for Knowledge Analysis) είναι ένα λογισμικό ανοιχτού κώδικα και αποτελεί ένα από τα πιο σημαντικά συστήματα για μεθόδους εξόρυξης δεδομένων και μηχανικής μάθησης. Αναπτύχθηκε το 1992, εξαιτίας της ανάγκης για ένα λογισμικό το οποίο θα επιτρέψει στους ερευνητές να έχουν πρόσβαση σε τεχνολογικά εξελιγμένες τεχνικές μηχανικής μάθησης. Χρησιμοποιείται ευρέως σε ακαδημαϊκό και επιχειρηματικό επίπεδο ως εργαλείο εξόρυξης δεδομένων για έρευνα και ένα από τα κύρια πλεονεκτήματα του είναι ότι επιτρέπει στους ερευνητές να υλοποιήσουν νέες τεχνικές και αλγορίθμους χωρίς να χρειάζεται να δημιουργήσουν υποδομές για την επεξεργασία των δεδομένων και την αξιολόγηση των αποτελεσμάτων τους [15].

Το WEKA παρέχει μια συλλογή αλγορίθμων μηχανικής μάθησης και προεπεξεργασίας δεδομένων. Υποστηρίζει αρχεία σε format (arff), αλλά και μορφής CSV. Ακόμα, περιλαμβάνει πολλούς αλγορίθμους για τεχνικές κατηγοριοποίησης (classification), παλινδρόμησης (regression), συσταδοποίησης (clustering) και κανόνων συσχέτισης (associate). Τέλος, αξιολογεί τα αποτελέσματα κάθε αλγόριθμου που εφαρμόζεται στα δεδομένα, ώστε να μπορεί ο χρήστης να συγκρίνει τους διαφορετικούς αλγορίθμους που χρησιμοποιεί.

3.4 Εργαλεία Ανάπτυξης Εφαρμογής

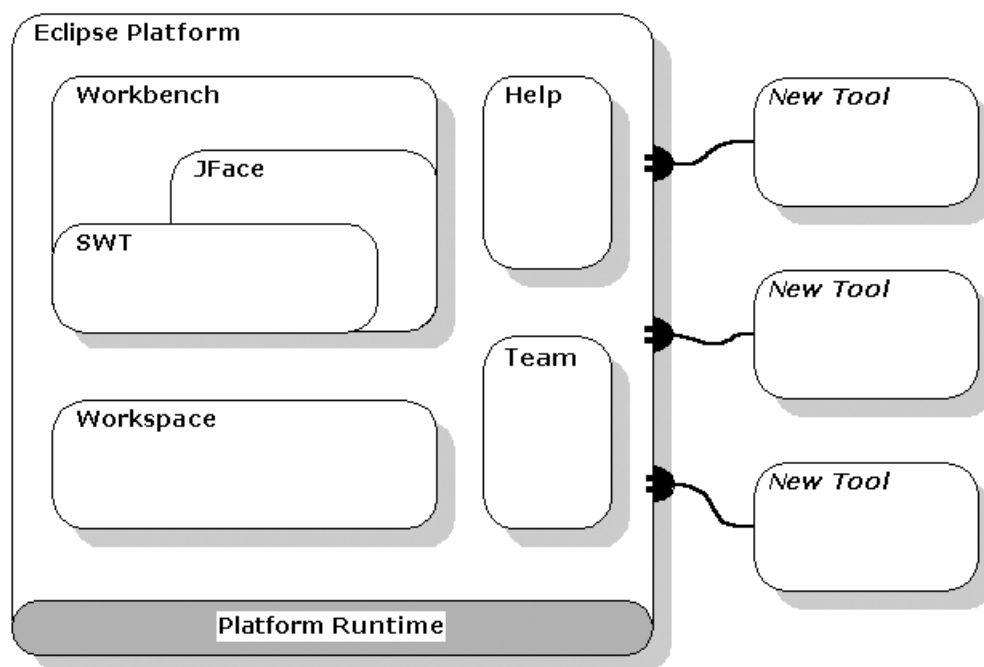
3.4.1 Το περιβάλλον Android Studio

Το Android Studio είναι το επίσημο ολοκληρωμένο περιβάλλον ανάπτυξης λογισμικού για την πλατφόρμα Android. Αποτελεί ελεύθερο λογισμικό και κυκλοφορεί υπό την άδεια λογισμικού

Apache2.0.

Βασίζεται στο λογισμικό IntelliJ IDEA της JetBrains και είναι σχεδιασμένο ειδικά για ανάπτυξη Android εφαρμογών. Είναι διαθέσιμο για λειτουργικά συστήματα Windows, Mac OS X και Linux, ενώ έχει αντικαταστήσει το περιβάλλον Eclipse με τα πρόσθετα εργαλεία για ανάπτυξη εφαρμογών Android (ADT), ως το κύριο περιβάλλον της Google για την ανάπτυξη λογισμικού Android[28].

3.4.2 Το περιβάλλον Eclipse



Σχήμα 3.4 Η αρχιτεκτονική της πλατφόρμας Eclipse. (Πηγή: www.eclipse.org¹)

Το Eclipse IDE είναι ένα περιβάλλον ανάπτυξης λογισμικού, το οποίο αρχικά αναπτύχθηκε από την IBM ως εργαλείο ανάπτυξης Java με σκοπό να αντικαταστήσει το ήδη υπάρχον περιβάλλον Visual Age, αλλά κυκλοφόρησε ως λογισμικό ανοιχτού κώδικα τον Νοέμβριο του 2001 [7]. Το 2004, ο μη κερδοσκοπικός οργανισμός Eclipse Foundation και το επιστημονικό του προσωπικό ανέλαβε ολοκληρωτικά τον έλεγχο της πλατφόρμας Eclipse. Το Eclipse σήμερα

¹Άρθρο στο διαδίκτυο <http://www.eclipse.org/articles/Whitepaper-Platform-3.1/eclipse-platform-whitepaper.html>. Ελέγχθηκε 3/1/2016

αποτελεί το πιο διαδεδομένο περιβάλλον ανάπτυξης για Java, ενώ ταυτόχρονα εξαιτίας της επεκτάσιμης αρχιτεκτονικής του χρησιμοποιείται ως εργαλείο ανάπτυξης για πολλές άλλες γλώσσες προγραμματισμού [8].

Η πλατφόρμα Eclipse σχεδιάστηκε με στόχο την ανάπτυξη ολοκληρωμένων περιβαλλόντων ανάπτυξης. Αποτελεί ένα εύκολα επεκτάσιμο περιβάλλον, παρέχοντας μηχανισμούς που επιτρέπουν την εγκατάσταση επιπρόσθετων εργαλείων τα οποία μπορούν να κάνουν χρήση άλλων διεπαφών ανάπτυξης εφαρμογών (APIs). Η αρχιτεκτονική της πλατφόρμας Eclipse είναι δομημένη με τέτοιο τρόπο ώστε να επιτρέπει την χρήση πρόσθετων εργαλείων, όπως φαίνεται στο Σχήμα 3.4 [8]. Τα πρόσθετα εργαλεία αποτελούν την μικρότερη μονάδα της πλατφόρμας Eclipse, τα οποία είναι δομημένα κομμάτια κώδικα που προσφέρουν επιπλέον λειτουργικότητα στην πλατφόρμα.

3.5 Αρχιτεκτονική Εργαλείων

Για την υλοποίηση της εφαρμογής της παρούσας εργασίας χρησιμοποιήθηκε η πλατφόρμα Okeanos, όπου δημιουργήθηκε virtual machine (εικονικό μηχάνημα) με λειτουργικό σύστημα Ubuntu 12.04 LTS. Στο virtual machine εγκαταστάθηκε λογισμικό Apache Web Server Version 2.2.22 για Ubuntu, το οποίο χρησιμοποιήθηκε ως server για την συλλογή των δεδομένων κίνησης. Επιπλέον, στο ίδιο μηχάνημα εγκαταστάθηκε για την αποθήκευση και την διαχείριση των δεδομένων, βάση δεδομένων Postgresql Version 9.1.

Για την ανάπτυξη της εφαρμογής, χρησιμοποιήθηκε το περιβάλλον προγραμματισμού Eclipse IDE for Java Developers Version: 4.2.1 με τα πρόσθετα εργαλεία Android (ADT) Version 23.0.2. Τέλος, έγινε χρήση της βιβλιοθήκης Apache Commons Math Version 2.2 και του προγράμματος Weka Version 3.6.13 ώστε να πραγματοποιηθούν οι διαδικασίες εξόρυξης δεδομένων της εφαρμογής.

Κεφάλαιο 4

Υλοποίηση

Στο παρόν κεφάλαιο, περιγράφεται η μέθοδος υλοποίησης της εφαρμογής της παρούσας διπλωματικής εργασίας, η οποία έχει στόχο την συλλογή των δεδομένων κίνησης του χρήστη και την επεξεργασία των δεδομένων σε δεύτερο χρόνο ώστε να εξαχθούν οι συνήθειες του χρήστη. Πιο συγκεκριμένα, θα παρουσιαστούν αναλυτικότερα οι διαδικασίες συλλογής των δεδομένων από την κινητή συσκευή στο κεντρικό αποθετήριο, η ανάκτηση των δεδομένων σε επίπεδο χρήστη και στη συνέχεια η ανίχνευση των σημείων ενδιαφέροντος και ο χαρακτηρισμός των σημείων και η εξαγωγή των συνηθειών του χρήστη.

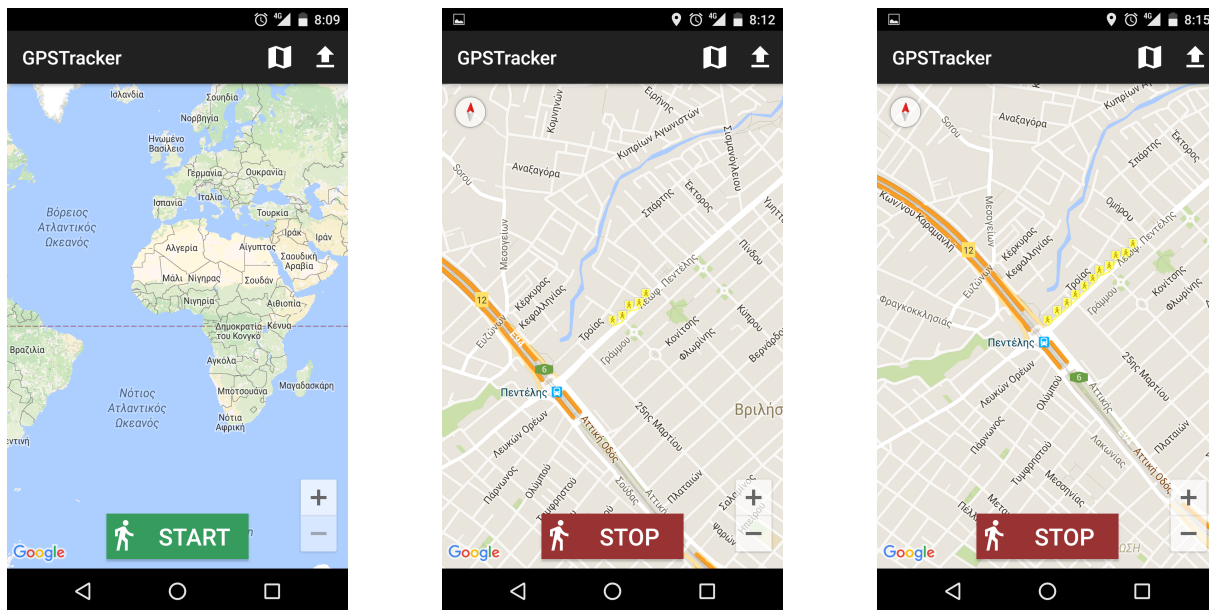
4.1 Συλλογή Δεδομένων σε Κεντρικό Αποθετήριο

4.1.1 Δεδομένα Κίνησης

Τα δεδομένα κίνησης του χρήστη καταγράφονται από την εφαρμογή GPSTracker που αναπτύχθηκε στα πλαίσια πτυχιακής εργασίας. Η καταγραφή των δεδομένων προϋποθέτει την αλληλεπίδραση του χρήστη με την εφαρμογή, καθώς ξεκινά με το πάτημα του κουμπιού *START* και διακόπτεται πατώντας το κουμπί *STOP*. Παράδειγμα χρήσης της εφαρμογής παρουσιάζεται στο Σχήμα 4.1.

Τα δεδομένα κίνησης αποθηκεύονται στη μνήμη της κινητής συσκευής σε αρχεία μορφής μορφής *arff*. Κάθε αρχείο δεδομένων αποτελείται από στιγμιότυπα κίνησης τα οποία καταγράφονται ανά 30 δευτερόλεπτα. Η θέση του χρήστη περιγράφεται σε κάθε στιγμιότυπο με γεωγραφικές συντεταγμένες, όπως ορίζονται από την υπηρεσία GPS της κινητής συσκευής. Επιπλέον, σε κάθε στιγμιότυπο καταγράφονται χαρακτηριστικά της κίνησης του χρήστη, με χρήση των αισθητήρων GPS και *accelerometer* (επιταχυνσιόμετρο) αλλά και με τη χρήση γεωχωρικών δεδομένων

που είναι διαθέσιμα μέσω μιας SQLite βάσης δεδομένων της εφαρμογής. Τα γνωρίσματα της κίνησης που καταγράφονται μπορούν εύκολα να κατηγοριοποιηθούν σε χαρακτηριστικά θέσης, χρονικής στιγμής και κίνησης. Στον Πίνακα 4.1 παρουσιάζονται τα χαρακτηριστικά της κίνησης του χρήστη που καταγράφονται από την εφαρμογή GPSTracker στις κατηγορίες που αναφέρθηκαν προηγουμένως.



Σχήμα 4.1 Στιγμιότυπα χρήσης της εφαρμογής GPSTracker.

Πίνακας 4.1 Κατηγορίες χαρακτηριστικών που καταγράφονται από την εφαρμογή GPSTracker.

Χαρακτηριστικά θέσης	Χαρακτηριστικά χρονικής στιγμής	Χαρακτηριστικά κίνησης
Longitude	Timestamp	CurrentSpeed
Latitude	Hour	RelativeSpeed
Altitude	TimeZone	MoveType
InsideThePolygon	DayOfWeek	
onBusLine	isWorkingDay	
onMetroLine		
isGpsFixed		

Στο Σχήμα 4.2 φαίνεται ένα δείγμα δεδομένων κίνησης σε μορφή αρχείου *arff*, όπως καταγράφεται από την εφαρμογή GPSTracker κατά τη διάρκεια κίνησης του χρήστη. Ωστόσο, η

υλοποίηση της παρούσας εφαρμογής εστιάζει στη χρήση χαρακτηριστικών θέσης και πιο συγκεκριμένα των γεωγραφικών συντεταγμένων αλλά και στα χαρακτηριστικά της χρονικής στιγμής που ο χρήστης βρίσκεται σε μια συγκεκριμένη θέση.

```
@relation gps_tracking

@attribute Longitude numeric
@attribute Latitude numeric
@attribute CurrentSpeed numeric
@attribute RelativeSpeed numeric
@attribute Altitude numeric
@attribute Timestamp date "yyyy-MM-dd HH:mm:ss"
@attribute Timestamp_H date "HH"
@attribute TimeZoneOffset numeric
@attribute DayOfWeek {Mon,Tue,Wed,Thu,Fri,Sat,Sun}
@attribute IsWorkingDay {yes,no}
@attribute InsideThePolygon {yes,no}
@attribute OnBusLine {yes,no}
@attribute OnMetroLine {yes,no}
@attribute MoveType {Walking,Running,Biking,Driving,Metro,Bus,Motionless}
@attribute IsGpsFixed {yes,no}

@data
38.75913401,22.85234022,0.36686681111653646,0.0,25.0,"2015-7-26 1:10:6","1",2,Sun,no,no,no,no,Biking,yes
38.75907116,22.85228001,0.30074747677507074,0.0,-24.0,"2015-7-26 1:10:37","1",2,Sun,no,no,no,no,Biking,yes
38.7590612,22.85230779,0.08852431774139405,0.0,0.0,"2015-7-26 1:11:7","1",2,Sun,no,no,no,no,Bus,no
38.75906071,22.85231046,0.007689174625181382,0.0,0.0,"2015-7-26 1:11:37","1",2,Sun,no,no,no,no,Motionless,no
```

Σχήμα 4.2 Δείγμα καταγραφής δεδομένων κατά τη διάρκεια κίνησης του χρήστη.

4.1.2 Κεντρικό Αποθετήριο Δεδομένων

Για την ανάπτυξη του κεντρικού αποθετηρίου δεδομένων, όπου θα συλλέγονται τα δεδομένα κίνησης των χρηστών που καταγράφονται από την εφαρμογή GPSTracker, αρχικά δημιουργήθηκε ένα εικονικό μηχάνημα (virtual machine) στην πλατφόρμα okeanos. Για την φιλοξενία του αποθετηρίου, χρησιμοποιήθηκε μηχάνημα με λειτουργικό σύστημα Ubuntu 12.04 LTS.

Στη συνέχεια, στο εικονικό μηχάνημα έγινε εγκατάσταση λογισμικού εξυπηρετητή (server) και πιο συγκεκριμένα Apache Web Server Version 2.2.22 και βάσης δεδομένων PostgreSQL Version 9.1.

Για την αποθήκευση των δεδομένων κίνησης στο κεντρικό αποθετήριο δεδομένων, δημιουργήθηκε στην Βάση Δεδομένων πίνακας στον οποίο εισάγονται τα δεδομένα σε μορφή αρχείων arff, όπως ακριβώς καταγράφονται από την εφαρμογή GPSTracker. Αναλυτικότερα, δημιουργήθηκε ο πίνακας routes, με τα εξής πεδία: *UUID*, το οποίο είναι το μοναδικό αναγνωριστικό κάθε χρήστη και περιγράφεται από το ID της κινητής συσκευής του χρήστη, *START_TIME*, το timestamp (χρονοσφραγίδα) της πρώτης εγγραφής που περιέχει το αρχείο arff, *END_TIME*, το

timestamp (χρονοσφραγίδα) της τελευταίας εγγραφής που περιέχει το αρχείο arff και *ROUTE_FILE*, το πεδίο που αποθηκεύεται το αρχείο arff σε μορφή binary. Το σχήμα της βάσης δεδομένων του αποθετηρίου δεδομένων φαίνεται στο Σχήμα 4.3.

Routes	
PK	ID
	START_TIME END_TIME UUID ROUTE_FILE

Σχήμα 4.3 Το σχήμα της Βάσης Δεδομένων του Κεντρικού Αποθετηρίου Δεδομένων.

4.1.3 Μεταφόρτωση Δεδομένων στο Κεντρικό Αποθετήριο

Για την διαδικασία της μεταφόρτωσης των αρχείων δεδομένων στο κεντρικό αποθετήριο, αναπτύχθηκε Web Service (UploadRoute) σε γλώσσα PHP, το οποίο δέχεται σαν input τις παραμέτρους: start_time, end_time, uuid και το αρχείο δεδομένων arff (routeFile) και αναλαμβάνει να εισάγει τα παραπάνω δεδομένα στη βάση δεδομένων του server. Όταν ολοκληρωθεί η διαδικασία επιστρέφει ένα αντικείμενο JSON το οποίο περιέχει το element *status* με τιμή success αν η εισαγωγή ήταν επιτυχής, είτε με τιμή failure αν υπήρξε κάποιο σφάλμα.

4.1.3.1 Χειροκίνητη

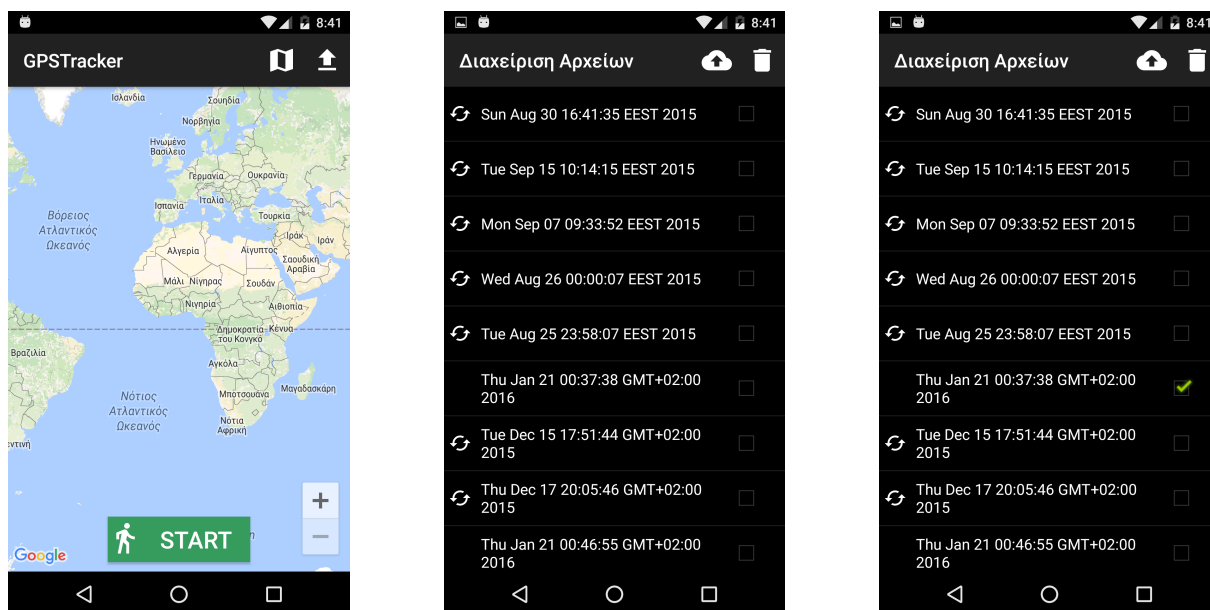
Όσον αφορά τη διαδικασία μεταφόρτωσης των αρχείων από την κινητή συσκευή στο κεντρικό αποθετήριο δεδομένων, στα πλαίσια της παρούσας εργασίας τροποποιήθηκε η υπάρχουσα λειτουργικότητα της εφαρμογής GPSTracker.

Χρησιμοποιήθηκε η Android κλάση Async Task, η οποία επιτρέπει την εκτέλεση HTTP κλήσεων στο παρασκήνιο (background) και την εμφάνιση των αποτελεσμάτων στη βασική διεργασία της εφαρμογής. Δημιουργήθηκε η κλάση UploadFileToServerTask, η οποία χρησιμοποιεί την μέθοδο doInBackground της AsyncTask για την διαδικασία μεταφόρτωσης των αρχείων και την μέθοδο onProgressUpdate για να ενημερώνει τον χρήστη για την πρόοδο της διαδικασίας μέσω ενός ProgressDialog. Όταν ολοκληρωθεί η διαδικασία, ενημερώνει τον χρήστη με μήνυμα επιτυχίας ή σφάλματος με τη μέθοδο onPostExecute.

Πιο συγκεκριμένα, κατά την εκτέλεση της μεθόδου doInBackground καλείται το Web Service UploadRoute χρησιμοποιώντας τη μέθοδο HTTP POST, με παραμέτρους: το timestamp του

πρώτου instance του αρχείου, το timestamp του τελευταίου instance του αρχείου, το uuid της συσκευής του χρήστη και το αρχείο arff. Κατά τη διάρκεια της HTTP κλήσης εμφανίζεται στον χρήστη παράθυρο που τον ενημερώνει για την πρόοδο της διαδικασίας μεταφόρτωσης των αρχείων στον server. Τέλος, όταν η διαδικασία ολοκληρωθεί ο χρήστης ενημερώνεται για την επιτυχία ή αποτυχία της μεταφόρτωσης, με βάση το αντικείμενο JSON που επιστρέφει το Web Service.

Αναλυτικότερα, όπως φαίνεται στο Σχήμα 4.4 ο χρήστης μεταβαίνει στην οθόνη της διαχείρισης αρχείων πατώντας το κουμπί μεταφόρτωσης στην αρχική οθόνη της εφαρμογής. Στη συνέχεια, εμφανίζεται η λίστα με τα αποθηκευμένα αρχεία στη συσκευή, ενώ παράλληλα τα αρχεία που έχουν ήδη μεταφορτωθεί σημειώνονται με το εικονίδιο του συγχρονισμού. Ο χρήστης μπορεί να επιλέξει πολλαπλά αρχεία προς μεταφόρτωση και πατώντας το κουμπί της μεταφόρτωσης η διαδικασία ξεκινά. Αφού ολοκληρωθεί η διαδικασία της μεταφόρτωσης του αρχείου, το αρχείο παραμένει και στη μνήμη της κινητής συσκευής και ο χρήστης έχει την επιλογή της διαγραφής του από την ίδια δραστηριότητα.



Σχήμα 4.4 Στιγμιότυπα χρήσης της μεταφόρτωσης αρχείου στην εφαρμογή GPSTracker.

4.1.3.2 Αυτόματη

Ωστόσο, ο στόχος της παρούσας εργασίας ορίζει την διαδικασία συλλογής των δεδομένων ως μια διάφανη διαδικασία, η οποία δεν προϋποθέτει την αλληλεπίδραση του χρήστη της κινητής συσκευής. Για αυτό τον λόγο, η μεταφόρτωση των αρχείων δεδομένων υλοποιήθηκε με έναν επι-

πλέον τρόπο ώστε τα νέα αρχεία να μεταφορτώνονται αυτόματα όταν η κινητή συσκευή συνδέεται σε ασύρματο δίκτυο Wi-Fi.

Υλοποιήθηκε ακόμα μια κλάση UploadFilesBackgroundTask που επεκτείνει την AsyncTask. Η κλάση UploadFilesBackgroundTask, παρόμοια με την κλάση UploadFileToServerTask, χρησιμοποιεί την μέθοδο doInBackground για την μεταφόρτωση του αρχείου στο κεντρικό αποθετήριο μέσω του Web Service UploadRoute, αλλά δεν ενημερώνει τον χρήστη για το αποτέλεσμα της διαδικασίας καθώς η συγκεκριμένη λειτουργία απαιτείται να μην γίνεται αντιληπτή από τον χρήστη.

Επιπλέον, δημιουργήθηκε η κλάση WiFiReceiver η οποία επεκτείνει την κλάση BroadcastReceiver του Android, και λαμβάνει ειδοποίηση κάθε φορά που η κινητή συσκευή συνδέεται ή αποσυνδέεται από κάποιο ασύρματο δίκτυο. Για να επιτευχθεί αυτό, είναι απαραίτητο να καταχωρηθεί στο αρχείο AndroidManifest.xml η συγκεκριμένη κλάση και να εγγραφεί ώστε να λαμβάνει ειδοποιήσεις σχετικές με την κατάσταση του ασύρματου δικτύου, με τον τρόπο που παρατίθεται παρακάτω.

Listing 4.1 Καταχώρηση της κλάσης WiFiReceiver στο AndroidManifest.xml

```
1 <receiver android:name="com.example.gpstracker.WiFiReceiver" >
2     <intent-filter android:priority="100" >
3         <action android:name="android.net.wifi.STATE_CHANGE" />
4     </intent-filter>
5 </receiver>
```

Αναλυτικότερα, όταν ληφθεί ειδοποίηση ότι η κινητή συσκευή συνδέθηκε σε ασύρματο δίκτυο, αναζητούνται τα αρχεία που υπάρχουν στην μνήμη της συσκευής αλλά δεν έχουν αποσταλλεί στο κεντρικό αποθετήριο και κάνοντας χρήση της UploadFilesBackgroundTask μεταφορτώνονται στον server.

Η προσέγγιση που περιγράφηκε, επιτυγχάνει την αυτόματη συλλογή των δεδομένων κίνησης από την κινητή συσκευή, χωρίς να απαιτείται αλληλεπίδραση από τον χρήστη, ενώ ταυτόχρονα διασφαλίζει ότι δεν θα επιφέρει επιπλέον χρεώσεις στον χρήστη καθώς δεν χρησιμοποιεί το δίκτυο κινητής τηλεφωνίας.

4.2 Ανάκτηση Δεδομένων Χρήστη από Κεντρικό Αποθετήριο

Αφού ολοκληρώθηκε η διαδικασία συλλογής των δεδομένων κίνησης από την κινητή συσκευή, στόχος της παρούσας εργασίας είναι η ανάλυση των δεδομένων κίνησης του χρήστη και

η εξαγωγή συμπερασμάτων που αφορούν τις συνήθειες του χρήστη. Η ανάλυση των δεδομένων κίνησης πραγματοποιήθηκε σε γλώσσα Java με επεξεργασία σε πολλαπλά επίπεδα και κάνοντας χρήση τεχνικών εξόρυξης δεδομένων.

Το πρώτο βήμα για την ανάλυση των δεδομένων, αποτελεί η διαδικασία ανάκτησης των αρχείων από το κεντρικό αποθετήριο δεδομένων κάνοντας χρήση του JDBC API της Oracle, με το οποίο γίνεται δυνατή η σύνδεση στη Βάση Δεδομένων του server. Αρχικά, πραγματοποιείται επερώτημα στη βάση δεδομένων για την εύρεση των διαφορετικών χρηστών (με βάση το UUID) για τους οποίους υπάρχουν δεδομένα στο αποθετήριο, ενώ στη συνέχεια γίνεται ανάκτηση του συνόλου των δεδομένων κίνησης ανά χρήστη. Τα δεδομένα επιστρέφονται σε μορφή δυαδικών αρχείων στα οποία διεξάγεται επεξεργασία για να καταστούν σε κατάλληλη μορφή για περαιτέρω ανάλυση.

4.3 Ανίχνευση Σημείων Ενδιαφέροντος Χρήστη

Η ανίχνευση των σημείων ενδιαφέροντος του χρήστη υλοποιήθηκε με χρήση της τεχνικής συσταδοποίησης (clustering) στα δεδομένα κίνησης του χρήστη και πιο συγκεκριμένα με τον αλγόριθμο DBSCAN χρησιμοποιώντας την βιβλιοθήκη Apache Commons Math.

Ωστόσο, παρόλα τα πλεονεκτήματα που παρουσιάζει ο αλγόριθμος DBSCAN για τον εντοπισμό σημείων ενδιαφέροντος σε δεδομένα GPS, εντοπίζει συστάδες με βάση τα σημεία στον χώρο, αγνοώντας την παράμετρο του χρόνου. Παράλληλα, χρησιμοποιεί εξ ορισμού ως μέτρο απόστασης μεταξύ των σημείων την Ευκλείδεια απόσταση, η οποία δεν λαμβάνει υπόψιν ότι η γη δεν είναι επίπεδη, συνεπώς η συγκεκριμένη μέθοδος έχει μειωμένη ακρίβεια σε γεωγραφικά δεδομένα.

Για τις ανάγκες της συγκεκριμένης εφαρμογής, υλοποιήθηκε ένα custom DistanceMeasure για τον αλγόριθμο DBSCAN, το οποίο υπολογίζει την γεωγραφική απόσταση δύο σημείων και επιπλέον απομακρύνει τα σημεία τα οποία έχουν κοντινή απόσταση αλλά μεγάλη χρονική διαφορά.

Αρχικά, υλοποιήθηκε η κλάση GeoDistanceCalculator, στην οποία δημιουργήθηκε η μέθοδος distance που παρατίθεται παρακάτω 4.2. Η μέθοδος λαμβάνει σαν παραμέτρους τις συντεταγμένες δυο σημείων (lat1, lon1, lat2, lon2) και υπολογίζει την απόσταση των δύο τοποθεσιών σε χιλιόμετρα με βάση την Haversine formula[33]. Η haversine formula, είναι μια μαθηματική εξίσωση η οποία υπολογίζει την απόσταση μεταξύ δύο σημείων πάνω σε μια σφαίρα με βάση τα γεωγραφικά μήκη και πλάτη τους [33], με τον τρόπο που παρουσιάζεται στην εξίσωση 4.1.

$$d = 2r \arcsin \sqrt{\sin^2\left(\frac{\phi_2 - \phi_1}{2}\right) + \cos(\phi_1) \cos(\phi_2) \sin^2\left(\frac{\lambda_2 - \lambda_1}{2}\right)} \quad (4.1)$$

όπου :

- d = η απόσταση μεταξύ των δυο σημείων,
- r = η ακτίνα της σφαίρας,
- λ_1, λ_2 = το γεωγραφικό μήκος του σημείου 1 και του σημείου 2,
- ϕ_1, ϕ_2 = το γεωγραφικό πλάτος του σημείου 1 και του σημείου 2.

Η υλοποίηση της μεθόδου distance πραγματοποιήθηκε χρησιμοποιώντας ως ακτίνα την ακτίνα της γης σε χιλιόμετρα (κατά προσέγγιση), καθώς ο στόχος είναι να υπολογιστεί η γεωγραφική απόσταση μεταξύ δυο τοποθεσιών.

Listing 4.2 Υπολογισμός γεωγραφικής απόστασης μεταξύ δύο σημείων

```

1  public static double distance(double lat1, double lon1, double lat2,
2      double lon2) {
3      double earthRadius = 6371.01; //kilometers
4      double dLat = Math.toRadians(lat2 - lat1);
5      double dLng = Math.toRadians(lon2 - lon1);
6      double a = Math.sin(dLat/2) * Math.sin(dLat/2) +
7          Math.cos(Math.toRadians(lat1)) * Math.cos(Math.toRadians(lat2)) *
8          Math.sin(dLng/2) * Math.sin(dLng/2);
9      double c = 2 * Math.atan2(Math.sqrt(a), Math.sqrt(1-a));
10     float dist = (float) (earthRadius * c);
11     return dist;
12 }
```

Όπως αναφέρθηκε παραπάνω, υλοποιήθηκε η κλάση POIDistanceMeasure, ένα μέτρο απόστασης το οποίο υπολογίζει την γεωγραφική απόσταση μεταξύ δυο σημείων κάνοντας χρήση της μεθόδου distance που περιγράφηκε προηγουμένως.

Επιπλέον, λαμβάνει υπόψιν τη χρονική διαφορά μεταξύ δυο στιγμάτων, αυξάνοντας την τιμή της απόστασης με βάση την απόσταση που θα μπορούσε να είχε διανύσει ο χρήστης στον ενδιάμεσο χρόνο. Αναλυτικότερα, δεδομένου ότι ο άνθρωπος διανύει κατά μέσο όρο πέντε χιλιόμετρα την ώρα περπατώντας [31], υπολογίζεται η απόσταση που θα είχε διανύσει ο χρήστης περπατώντας στην χρονική διάρκεια που μεσολαβεί μεταξύ των στιγμάτων. Τελικά, η γεωγραφική απόσταση δυο σημείων ορίζεται ως την μεταξύ τους απόσταση, προσαυξημένη με την απόσταση που υποθετικά θα είχε διανύσει ο χρήστης αν βρισκόταν σε κίνηση σε χρόνο που αντιστοιχεί

στη χρονική τους διαφορά. Η ακριβής μεθοδολογία της υλοποίησης της POIDistanceMeasure παρουσιάζεται παρακάτω (4.3).

Listing 4.3 Μέτρο Απόστασης για την αναγνώριση των σημείων ενδιαφέροντος

```

1 public class POIDistanceMeasure implements DistanceMeasure{
2     double walkingDistPerHour = 5.0;
3
4     public double compute(double[] pointA, double[] pointB) {
5         double latA = pointA[0];
6         double lonA = pointA[1];
7         double latB = pointB[0];
8         double lonB = pointB[1];
9         double timestampA = pointA[2];
10        double timestampB = pointB[2];
11        double additionFactor = 1.0;
12        Date dateA = new Date((long)timestampA * 1000);
13        Date dateB = new Date((long)timestampB * 1000);
14        long timeDiff = Math.abs(((dateB.getTime()/(60 * 60 * 1000)) - (dateA.getTime()/(60 * 60 *
15            1000))));
16        additionFactor = timeDiff * walkingDistPerHour;
17        return GeoDistanceCalculator.distance(latA, lonA, latB, lonB) + additionFactor;
18    }

```

Τέλος, για την εκτέλεση του DBSCAN πρέπει να καθοριστούν δυο παράμετροι: ϵ (*eps*), το οποίο περιγράφει την απόσταση που πρέπει να υπάρχει μεταξύ των σημείων για να θεωρηθεί ένα cluster και *minPts*, που ορίζει τον ελάχιστο αριθμό σημείων που απαιτούνται για να δημιουργηθεί ένα cluster. Για να πραγματοποιηθεί σωστά η παραμετροποίηση καθορίστηκαν σαν παραδοχές, ότι αρχικά θεωρείται σημείο ενδιαφέροντος μια τοποθεσία όπου ο χρήστης παρέμεινε τουλάχιστον πέντε λεπτά και επιπλέον ότι ο χρήστης μπορεί να κινείται σε ακτίνα δέκα μέτρων γύρω από μια τοποθεσία η οποία αποτελεί σημείο ενδιαφέροντος. Συνεπώς, εφόσον το στίγμα του χρήστη καταγράφεται ανά τριάντα δευτερόλεπτα αυτό μετατρέπεται εύκολα σε δέκα στίγματα στην διάρκεια των πέντε λεπτών, άρα η παράμετρος *minPts* πρέπει να έχει την τιμή 10 και η παράμετρος *eps* έχει την τιμή 0.01 καθώς η μέτρηση της απόστασης γίνεται σε χιλιόμετρα.

Listing 4.4 Εκτέλεση του αλγορίθμου DBSCAN.

```

1 DBSCANClusterer dbscan = new DBSCANClusterer(.01, 10, new POIDistanceMeasure());
2 List clusters = dbscan.cluster(points);

```

Συνοψίζοντας, η μεθοδολογία ανίχνευσης σημείων ενδιαφέροντος με τον αλγόριθμο DB-SCAN, στοχεύει στη δημιουργία ξεχωριστών συστάδων για τις διαφορετικές χρονικές στιγμές που ο χρήστης επισκέπτεται ένα συγκεκριμένο σημείο ενδιαφέροντος. Με αυτόν τον τρόπο, διευκολύνεται η διαδικασία εύρεσης συνηθειών, καθώς καταγράφονται οι πολλαπλές επισκέψεις του χρήστη στα σημεία ενδιαφέροντος.

4.4 Χαρακτηρισμός Σημείων Ενδιαφέροντος Χρήστη

Για την αναγνώριση των σημείων ενδιαφέροντος με χρήση του OpenStreetMaps API, ήταν απαραίτητο αρχικά να καθοριστεί η γεωγραφική περιοχή (bounding box). Για κάθε περιοχή, ανακτήθηκαν πληροφορίες, οι οποίες στην συνέχεια επεξεργάστηκαν και τελικά χρησιμοποιήθηκαν για να χαρακτηρίσουν τα σημεία ενδιαφέροντος.

Η δημιουργία του bounding box για ένα σημείο ενδιαφέροντος προϋποθέτει τον καθορισμό της ακτίνας που πρέπει να απέχει η κάθε πλευρά του bounding box από το κέντρο της συστάδας. Ο προσδιορισμός της κατάλληλης τιμής της απόστασης, αποτελεί μια σύνθετη διαδικασία καθώς δεν υπάρχουν καθορισμένα κριτήρια επιλογής της συγκεκριμένης παραμέτρου.

Για την επιλογή της κατάλληλης τιμής της παραμέτρου distance, πραγματοποιήθηκε διερεύνηση με βάση την υπάρχουσα βιβλιογραφία. Αρχικά, σύμφωνα με έρευνα των Helbich και συν. 2012 ο χάρτης OpenStreetMaps εμφανίζει σφάλμα σε σχέση με την πραγματική θέση των σημείων το οποίο κυμαίνεται από 0,2 έως 18,7 μέτρα[19]. Επιπλέον, τα αποτελέσματα της έρευνας των Kim και συν. 2013, που πραγματοποιούν για την αξιολόγηση της υπηρεσίας τοποθεσίας των έξυπνων κινητών τηλεφώνων, δείχνουν ότι το A-GPS (assisted GPS) δηλαδή η ενσωματωμένη συσκευή GPS σε ένα smartphone έχει ακρίβεια εντοπισμού περίπου 10 μέτρα[17]. Λαμβάνοντας υπόψιν τις παραπάνω πληροφορίες ώστε να ανακτηθούν έγκυρες πληροφορίες για κάθε σημείο ενδιαφέροντος, η παράμετρος distance θα πρέπει να διαμορφωθεί με τέτοιο τρόπο ώστε να υπερκαλύπτει αρχικά το σφάλμα του A-GPS κατά την καταγραφή των δεδομένων κίνησης αλλά και το σφάλμα της αποτύπωσης των σημείων στον χάρτη του OpenStreetMaps. Συνεπώς, μπορούμε να ορίσουμε την ακτίνα του bounding box στα 30 μέτρα απόστασης από το κέντρο κάθε συστάδας αν υπολογίσουμε ότι το συνολικό σφάλμα είναι περίπου 29 μέτρα, ενώ σε περίπτωση που ανακτηθούν περισσότερες από μια πληροφορίες για ένα σημείο ενδιαφέροντος θα θεωρήσουμε ως έγκυρη την πληροφορία που βρέθηκε σε πιο κοντινή απόσταση από το κέντρο κάθε συστάδας.

Αφού ολοκληρώθηκε η διαδικασία καθορισμού της παραμέτρου distance, αρχικά βρέθηκε το κέντρο κάθε συστάδας (centroid), υπολογίζοντας τον μέσο όρο των συντεταγμένων (γεωγραφικό

πλάτος και μήκος) που αποτελούν την συστάδα. Στη συνέχεια, για τον καθορισμό του bounding box γύρω από το κέντρο κάθε συστάδας, έγινε χρήση της μεθόδου `boundingCoordinates`¹, η οποία υπολογίζει για μια γεωγραφική τοποθεσία, τις ελάχιστες και μέγιστες συντεταγμένες που ορίζουν ένα ορθογώνιο με ακτίνα που καθορίζεται από την παράμετρο `distance`. Όλα τα σημεία του bounding box έχουν γεωγραφική απόσταση μικρότερη ή ίση με την παράμετρο `distance` από το κέντρο της συστάδας. Στο σχήμα 4.5, φαίνεται το bounding box ενός σημείου ενδιαφέροντος όπως υπολογίστηκε από την μέθοδο `boundingCoordinates`.



Σχήμα 4.5 Bounding box ενός σημείου ενδιαφέροντος.

Η συλλογή πληροφοριών για την γεωγραφική περιοχή που ορίζεται από το bounding box, πραγματοποιείται με κλήση στο OpenStreetMaps API, με παραμέτρους τις ελάχιστες και μέγιστες συντεταγμένες από τις οποίες ορίζεται το εκάστοτε bounding box. Το OpenStreetMaps API επιστρέφει τα δεδομένα σε μορφή αρχείου XML (Σχήμα 4.6) το οποίο υφίσταται περαιτέρω επεξεργασία σε Java, ώστε να εξαχθούν πληροφορίες για το σημείο ενδιαφέροντος.

Αναλυτικότερα, αφού εξετάστηκαν τα πιθανά tags που είναι πιθανό να βρεθούν στο XML αρχείο κάθε σημείου ενδιαφέροντος [22], επιλέχθηκαν μερικά από αυτά με κριτήριο το αν

¹Υλοποίηση στο διαδίκτυο <http://janmatuschek.de/LatitudeLongitudeBoundingCoordinates>. Ελέγχθηκε 06/02/2016

μπορούν να περιγράψουν καθημερινές συνήθειες του χρήστη. Πιο συγκεκριμένα, τα tags του OpenStreetMaps που χρησιμοποιήθηκαν για τον χαρακτηρισμό των σημείων ενδιαφέροντος είναι τα εξής:

- *amenity*, περιγράφει χαρακτηριστικά ενός κτιρίου ή τοποθεσίας,
- *public_transport*, περιέχει χαρακτηριστικά της δημόσιας συγκοινωνίας,
- *shop*, αναφέρεται σε επιχείρηση που διαθέτει προϊόντα προς πώληση,
- *sport*, περιγράφει τοποθεσίες κατάλληλες για άθληση,
- *leisure*, χαρακτηρίζει μέρη που είναι κατάλληλα για ξεκούραση ή άθληση.

```
<?xml version="1.0" encoding="UTF-8"?>
<osm version="0.6" generator="CGImap 0.4.0 (22136 thorn-03.openstreetmap.org)" copyright="OpenStreetMap contributors" attribution="http://www.openstreetmap.org/copyright" license="http://opendatacommons.org/licenses/odbl/1-0/">
  <bounds minlat="38.0340500" minlon="23.7369800" maxlat="38.0346400" maxlon="23.7375200"/>
  <node id="519954080" visible="true" version="3" changeset="10846515" timestamp="2012-03-02T10:12:51Z" user="armitatz" uid="414661" lat="38.0357216" lon="23.7355228">
    <tag k="highway" v="traffic_signals"/>
  </node>
  <node id="2750852492" visible="true" version="1" changeset="21374126" timestamp="2014-03-28T22:34:57Z" user="Chris Makridis" uid="1227858" lat="38.0341596" lon="23.7374317">
    <tag k="addr:housenumber" v="4"/>
    <tag k="addr:postcode" v="14341"/>
    <tag k="addr:street" v="Βρυούλων"/>
    <tag k="amenity" v="restaurant"/>
    <tag k="name" v="Εν Αιθρία"/>
    <tag k="website" v="www.enaithria.com"/>
  </node>
  <node id="3149097754" visible="true" version="1" changeset="26321623" timestamp="2014-10-25T13:44:12Z" user="athinaios" uid="2063365" lat="38.0347367" lon="23.7378470"/>
  <node id="3149097755" visible="true" version="1" changeset="26321623" timestamp="2014-10-25T13:44:12Z" user="athinaios" uid="2063365" lat="38.0347239" lon="23.7375604"/>
```

Σχήμα 4.6 Παράδειγμα αρχείου xml όπως εξάγεται από το OpenStreetMaps API.

Τέλος, για κάθε συστάδα ανιχνεύθηκαν όλα τα xml elements που περιέχουν κάποια από τα tags που προαναφέρθηκαν και σε περίπτωση που βρέθηκαν περισσότερα από ένα επιλέχθηκε εκείνο που είχε τη μικρότερη γεωγραφική απόσταση από το κέντρο κάθε συστάδας. Με αυτόν τον τρόπο, ολοκληρώθηκε ο χαρακτηρισμός των σημείων ενδιαφέροντος του χρήστη, για όποια από τα σημεία υπήρχε διαθέσιμη πληροφορία στην εφαρμογή OpenStreetMaps.

4.5 Εξαγωγή Συνηθειών

Όσον αφορά την διαδικασία αναγνώρισης συνηθειών, πραγματοποιήθηκε με χρήση τεχνικών εξόρυξης δεδομένων και πιο συγκεκριμένα με την εξαγωγή κανόνων συσχέτισης. Γενικότερα, οι κανόνες συσχέτισης στοχεύουν στην εύρεση κρυμμένων συσχετίσεων μεταξύ των γνωρισμάτων ενός συνόλου δεδομένων [2]. Για την εξαγωγή συνηθειών χρησιμοποιήθηκε ο χαρακτηρισμός των σημείων ενδιαφέροντος που πραγματοποιήθηκε αλλά και η πληροφορία της δραστηριότητας που καταγράφεται από την εφαρμογή GPSTracker.

```
@relation habits

@attribute DayZone {Morning,Afternoon,Evening,Night}
@attribute IsWorkingDay {true,false}
@attribute MoveType {Walking,Running,Biking,Driving,Metro,Bus,Motionless}
@attribute Tag {amenity,public_transport,shop,sport,leisure}

@data
Morning,true,Metro,public_transport
Morning,true,Metro,public_transport
Morning,true,Metro,public_transport
Morning,true,Metro,public_transport
Morning,true,Metro,public_transport
Morning,true,Metro,public_transport
```

Σχήμα 4.7 Δείγμα αρχείου δεδομένων arff όπως εξάγεται από την εφαρμογή.

Πιο συγκεκριμένα, έγινε εξαγωγή των στιγμάτων τα οποία αποτελούν κάποιο από τα σημεία ενδιαφέροντος που αναγνωρίστηκαν, αφού πρώτα συνδυάστηκαν με περισσότερες πληροφορίες όπως η ώρα της ημέρας που καταγράφηκαν η οποία μετατράπηκε σε ζώνη ώρας (πρωί, μεσημέρι, απόγευμα, βράδυ), ο τύπος κίνησης του χρήστη τη συγκεκριμένη στιγμή, αν καταγράφηκαν κατά τη διάρκεια εργάσιμης ημέρας ή Σαββατοκύριακου αλλά και την ιδιότητα του συγκεκριμένου σημείου ενδιαφέροντος. Πραγματοποιήθηκε μετατροπή των γνωρισμάτων σε κατηγορικά ώστε να είναι δυνατή η εύρεση κανόνων συσχέτισης με τον αλγόριθμο Apriori. Η εξαγωγή των δεδομένων έγινε σε αρχείο μορφής arff, με τον τρόπο που παρουσιάζεται στο Σχήμα 4.7. Τέλος, μέσω του προγράμματος Weka εκτελέστηκε ο αλγόριθμος Apriori στο αρχείο arff που εξάχθηκε, ορίζοντας την παράμετρο confidence ίση με 1 και περιορίζοντας τον αριθμό των αποτελεσμάτων σε 20, ώστε να εξαχθούν όσο το δυνατό πιο έγκυροι κανόνες και κατά επέκταση συνήθειες.

Κεφάλαιο 5

Αποτελέσματα

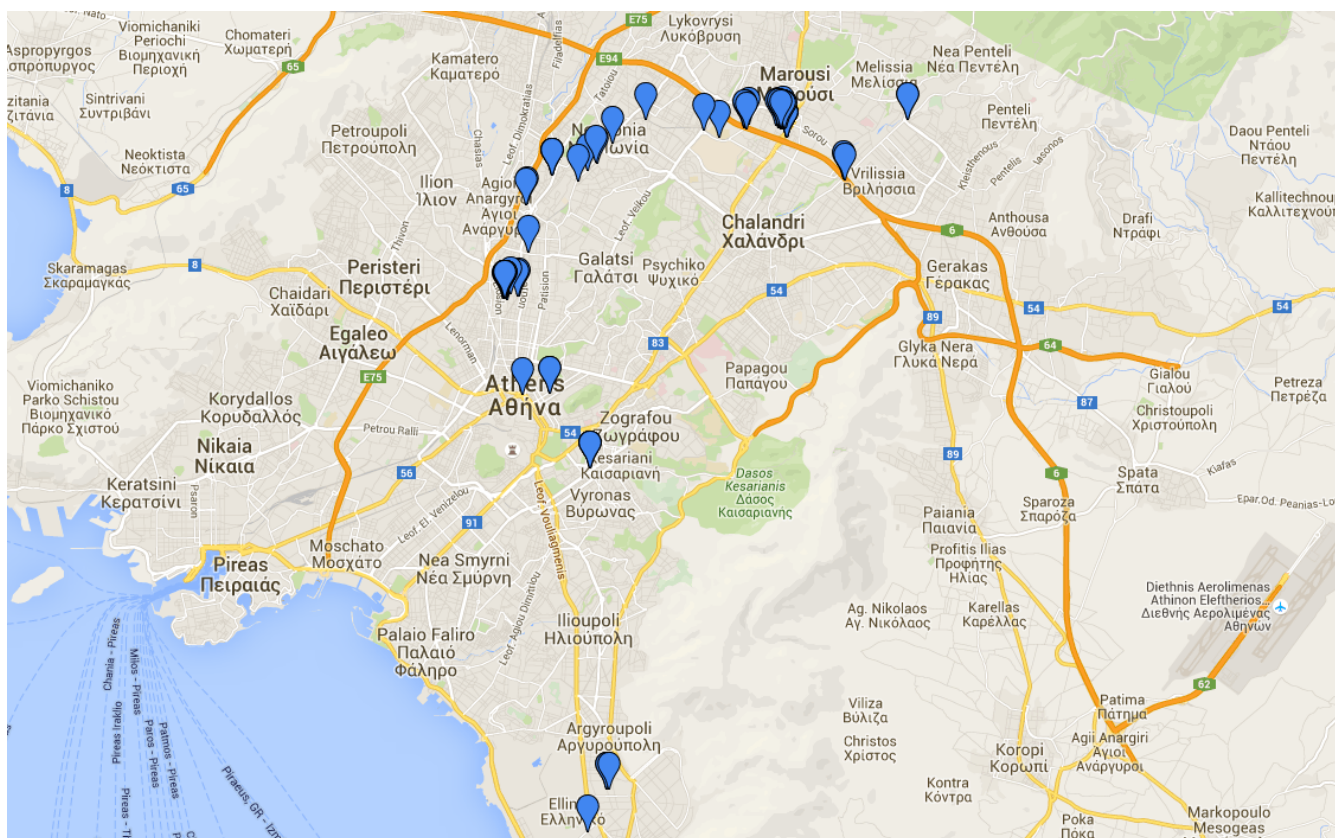
Στο παρόν κεφάλαιο, παρουσιάζονται και στη συνέχεια αξιολογούνται τα αποτελέσματα της εφαρμογής όσον αφορά τις διαδικασίες της ανίχνευσης των σημείων ενδιαφέροντος, του χαρακτηρισμού των σημείων και της εξαγωγής των συνηθειών του χρήστη. Οι διαδικασίες εφαρμόστηκαν σε 12817 στιγμιότυπα θέσης ενός χρήστη, τα οποία αντιστοιχούν σε 10 ημερολογιακές ημέρες. Τα δεδομένα κίνησης καταγράφηκαν από την εφαρμογή GPSTracker και συλλέχθηκαν στο κεντρικό αποθετήριο δεδομένων, με τον τρόπο που περιγράφηκε παραπάνω.

5.1 Αποτελέσματα Ανίχνευσης Σημείων Ενδιαφέροντος

Ο εντοπισμός των σημείων ενδιαφέροντος στα δεδομένα κίνησης, όπως αναφέρθηκε, πραγματοποιήθηκε με τον αλγόριθμο DBSCAN. Για την εκτέλεση του DBSCAN, υλοποιήθηκε νέο μέτρο υπολογισμού απόστασης μεταξύ των στιγμάτων, το οποίο λαμβάνει υπόψιν όχι μόνο τη γεωγραφική απόσταση μεταξύ των σημείων αλλά και τη χρονική διαφορά μεταξύ δυο καταγεγραμμένων στιγμάτων του χρήστη. Η τεχνική αυτή έχει στόχο να επιτύχει τη δημιουργία διαφορετικής συστάδας σημείων, σε περίπτωση που ο χρήστης επισκεφθεί ένα σημείο ενδιαφέροντος σε διαφορετικές χρονικές στιγμές. Για παράδειγμα, αν το σημείο ενδιαφέροντος που εξετάζεται είναι ο χώρος εργασίας του χρήστη όπου βρίσκεται σε καθημερινή βάση, εκτελώντας τον αλγόριθμο DBSCAN θα δημιουργηθούν πολλαπλές συστάδες στο συγκεκριμένο σημείο ανάλογου πλήθους με των αριθμό των επισκέψεων του χρήστη.

Η εκτέλεση του αλγορίθμου DBSCAN πραγματοποιήθηκε με παραμέτρους: ϵ (eps) = 0.1 και $minPts$ = 10 και είχε ως αποτέλεσμα την δημιουργία 315 συστάδων. Τα αποτελέσματα της ανίχνευσης σημείων ενδιαφέροντος για τον συγκεκριμένο χρήστη αποτυπώνονται σε χάρτη, μέσω

της εφαρμογής Google Maps στο Σχήμα 5.1.



Σχήμα 5.1 Τα αποτελέσματα του εντοπισμού σημείων ενδιαφέροντος όπως αποτυπώνονται σε χάρτη.

Ο αλγόριθμος DBSCAN έχει την ιδιότητα να αναγνωρίζει συστάδες με αυθαίρετο σχήμα, γεγονός που αφενός αυξάνει την ακρίβεια εντοπισμού σημείων ενδιαφέροντος, αναγνωρίζοντας τις διαφορετικές κινήσεις του χρήστη γύρω από ένα σημείο ενδιαφέροντος [18], αφετέρου εμπεριέχει τον κίνδυνο εντοπισμού συστάδων με αρκετά μεγάλο μήκος οι οποίες στην πραγματικότητα δεν αποτελούν σημεία ενδιαφέροντος. Αναλυτικότερα, σε περίπτωση που ο χρήστης κινείται με πολύ μικρή ταχύτητα προς μια κατεύθυνση υπάρχει πιθανότητα ο DBSCAN να αναγνωρίσει τα συγκεκριμένα στίγματα ως συστάδα, χωρίς όμως ο χρήστης να κινείται γύρω από ένα σημείο. Κατά συνέπεια, κρίνεται χρήσιμο να αξιολογηθούν τα αποτελέσματα εντοπισμού σημείων ενδιαφέροντος, ώστε να ελεγχθεί η εγκυρότητα της διαδικασίας συσταδοποίησης.

Η αξιολόγηση της ποιότητας των συστάδων εκτιμήθηκε κάνοντας χρήση μέτρων απόστασης σε επίπεδο συστάδας. Πιο συγκεκριμένα, χρησιμοποιήθηκε η διάμετρος των συστάδων που εντοπίστηκαν, υπολογίζοντας την μέγιστη γεωγραφική απόσταση μεταξύ δυο οποιονδήποτε σημείων

που ανήκουν σε κάθε συστάδα και η μέση γεωγραφική απόσταση των σημείων κάθε συστάδας από το κέντρο της (centroid). Με αυτόν τον τρόπο, με σχετικά μεγάλες τιμές των δυο αυτών μέτρων, είναι δυνατό να εντοπιστούν συστάδες μεγάλου μήκους. Στον Πίνακα 5.1 παρουσιάζεται ο αριθμός των συστάδων όπως κατανέμονται στις κατηγορίες τιμών απόστασης, ενώ στον Πίνακα 5.2 αποτυπώνονται με τον ίδιο τρόπο οι συστάδες σε σχέση με την μέση απόσταση των σημείων τους από το centroid.

Πίνακας 5.1 Αριθμός συστάδων ανά κατηγορία μήκους διαμέτρου.

< 10 m	10 - 20 m	20 - 30 m	> 30 m
214	76	21	4

Πίνακας 5.2 Αριθμός συστάδων ανά κατηγορία μήκους μέσης ακτίνας από το centroid.

< 5 m	5 - 10 m	10 - 15 m	> 15 m
286	28	1	0

Εξετάζοντας τις τιμές διαμέτρου των συστάδων, παρατηρούμε ότι υπάρχουν τέσσερις συστάδες οι οποίες έχουν σχετικά μεγάλο μήκος (> 30 m) και οι οποίες χρήζουν περαιτέρω διερεύνησης, ώστε να διαπιστωθεί αν αποτελούν όντως σημεία ενδιαφέροντος του χρήστη είτε αν πρόκειται για διαδρομές κατά τις οποίες ο χρήστης κινούταν με μικρή ταχύτητα.





Σχήμα 5.2 Οι συστάδες με σχετικά μεγάλη τιμή διαμέτρου που εντοπίστηκαν όπως αποτυπώνονται σε χάρτη.

Για αυτό τον λόγο, οι συγκεκριμένες συστάδες αποτυπώθηκαν σε χάρτη ώστε να διευκολυνθεί η αξιολόγηση τους (Σχήμα 5.2). Αρχικά, παρατηρείται ότι όλες οι συστάδες βρίσκονται στην ίδια γεωγραφική περιοχή, ενώ το πλήθος των σημείων βρίσκονται στα όρια ενός συγκεκριμένου οικοδομικού τετραγώνου. Επομένως, τα συγκεκριμένα σημεία έχουν αυξημένες πιθανότητες να αποτελούν σημεία ενδιαφέροντος καθώς έχουν καταγραφεί παραπάνω από μια φορά. Επιπλέον, η πλειονότητα των στιγμάτων βρίσκονται σε σημεία όπου υπάρχουν διαφορετικά κτίρια, γεγονός που μας επιτρέπει να συμπεράνουμε ότι ο χρήστης δεν είναι δυνατό να κινείται συνεχόμενα σε αυτά τα σημεία. Συνεπώς, οι αναφερόμενες συστάδες αποτελούν σημεία ενδιαφέροντος για τον χρήστη, ενώ η σχετικά μεγάλη διάμετρος τους πιθανόν οφείλεται σε σφάλμα του αισθητήρα GPS το οποίο προσεγγίζει την απόσταση των 15 μέτρων όταν ο χρήστης βρίσκεται στο εσωτερικό ενός οικοδομικού τετραγώνου [20].

Όσον αφορά τις αποστάσεις των σημείων από το κέντρο κάθε συστάδας, παρατηρείται ότι το πλήθος των συστάδων έχουν μέση απόσταση μικρότερη των πέντε μέτρων, ενώ κάποιες κυμαίνονται από πέντε έως δέκα μέτρα και μόνο μια καταγράφηκε στην κατηγορία των δέκα με δεκαπέντε μέτρων. Λαμβάνοντας υπόψιν ότι η εκτέλεση του αλγορίθμου πραγματοποιήθηκε με ακτίνα (*eps*) δέκα μέτρων, η τιμή της μέσης απόστασης από το κέντρο θα πρέπει να προσεγγίζει την απόσταση των δέκα μέτρων. Συνεπώς, έχει νόημα να διερευνηθεί η μοναδική συστάδα που η μέση απόσταση των σημείων της από το κέντρο ξεπερνάει την τιμή των δέκα μέτρων καθώς

υπάρχει πιθανότητα να μην αποτελεί σημείο ενδιαφέροντος. Ωστόσο, η συγκεκριμένη συστάδα συμπίπτει με την τρίτη συστάδα που παρουσιάζεται στο Σχήμα 5.2 (μπλε χρώμα), η οποία έχει ήδη εξεταστεί και έχει χαρακτηριστεί ως ένα έγκυρο σημείο ενδιαφέροντος.

Συμπερασματικά, τα αποτελέσματα του εντοπισμού σημείων ενδιαφέροντος δείχνουν ότι η τεχνική που χρησιμοποιήθηκε αναγνώρισε συστάδες οι οποίες αποτελούν στην πραγματικότητα σημεία ενδιαφέροντος του χρήστη. Ωστόσο, υπάρχουν περιθώρια περαιτέρω αξιολόγησης της συγκεκριμένης μεθοδολογίας, εφαρμόζοντας την σε δεδομένα κίνησης μεγαλύτερου χρονικού παραθύρου και περισσότερων χρηστών, ώστε να εντοπιστούν πιθανοί περιορισμοί της συγκεκριμένης τεχνικής.

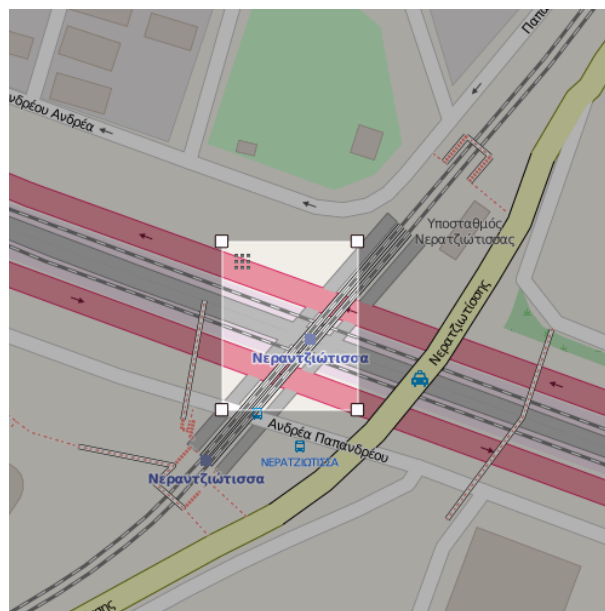
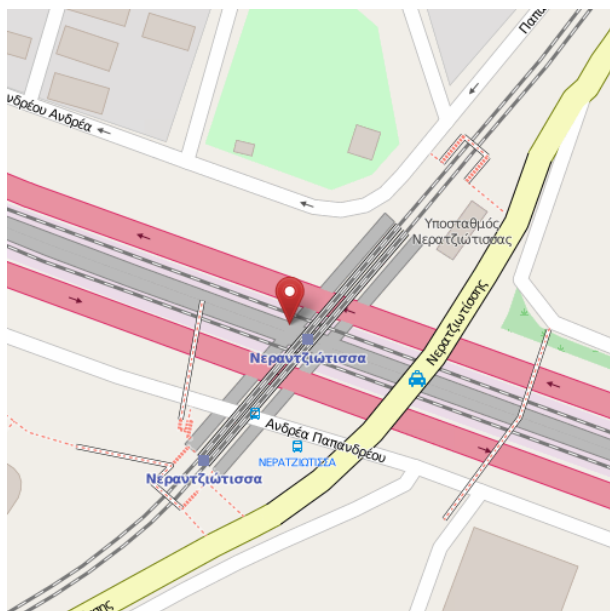
5.2 Αποτελέσματα Χαρακτηρισμού Σημείων Ενδιαφέροντος

Ο χαρακτηρισμός των σημείων ενδιαφέροντος που ανιχνεύθηκαν πραγματοποιήθηκε με χρήση του OpenStreetMaps API. Ορίσθηκε ένα ορθογώνιο οριοθετημένο (bounding box) σε ακτίνα 30 μέτρων γύρω από το κέντρο κάθε συστάδας, ανακτήθηκαν πληροφορίες για τη συγκεκριμένη γεωγραφική περιοχή και εκχωρήθηκε μια περιγραφή για το σημείο ενδιαφέροντος με βάση τις διαθέσιμες πληροφορίες που εξάχθηκαν.

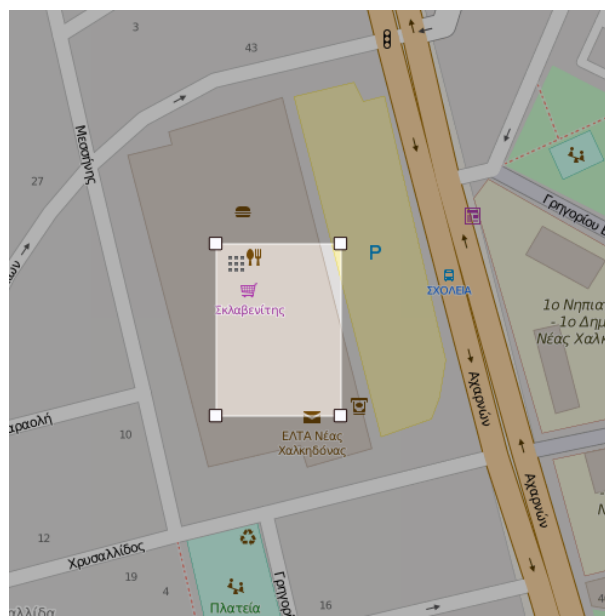
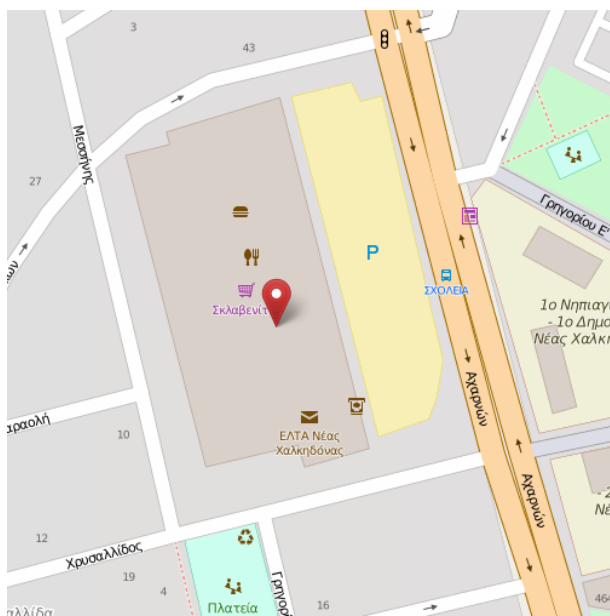
Η διαδικασία χαρακτηρισμού σημείων ενδιαφέροντος εφαρμόστηκε σε 315 συστάδες που εντοπίστηκαν από τον αλγόριθμο DBSCAN και είχε ως αποτέλεσμα την εξαγωγή πληροφορίας από το OpenStreetMaps για 33 συστάδες από αυτές. Για τα υπόλοιπα σημεία ενδιαφέροντος δεν ανακτήθηκαν διαθέσιμες πληροφορίες για την γεωγραφική περιοχή που ορίστηκε από το εκάστοτε bounding box. Συνεπώς, το ποσοστό επιτυχίας χαρακτηρισμού των σημείων ενδιαφέροντος ανέρχεται στο 10,48% του συνόλου των σημείων.

Στο Σχήμα 5.3 αποτυπώνεται στην εφαρμογή OpenStreetMaps ένα από τα σημεία ενδιαφέροντος που χαρακτηρίστηκε ως public transport. Στην πρώτη εικόνα φαίνεται το centroid του σημείου ενδιαφέροντος το οποίο βρίσκεται κοντά σε σταθμό τρένου, ενώ στην δεύτερη εικόνα το bounding box γύρω από το σημείο για το οποίο ανακτήθηκαν πληροφορίες. Παρόμοια, στο Σχήμα 5.4, το centroid του σημείου βρίσκεται εντός των ορίων ενός κτιρίου και το συγκεκριμένο σημείο ενδιαφέροντος αναγνωρίστηκε ως amenity.

Στον Πίνακα 5.3, παρουσιάζεται η κατανομή των συστάδων στα tags του OpenStreetMaps, τα οποία χρησιμοποιήθηκαν για τον χαρακτηρισμό των σημείων ενδιαφέροντος. Επιπλέον, στο Σχήμα 5.5 αποτυπώνεται το ιστόγραμμα που απεικονίζει την συχνότητα εμφάνισης κάθε tag του OpenStreetMaps στα σημεία ενδιαφέροντος που χαρακτηρίστηκαν.



Σχήμα 5.3 Σημείο ενδιαφέροντος το οποίο χαρακτηρίστηκε ως public_transport.

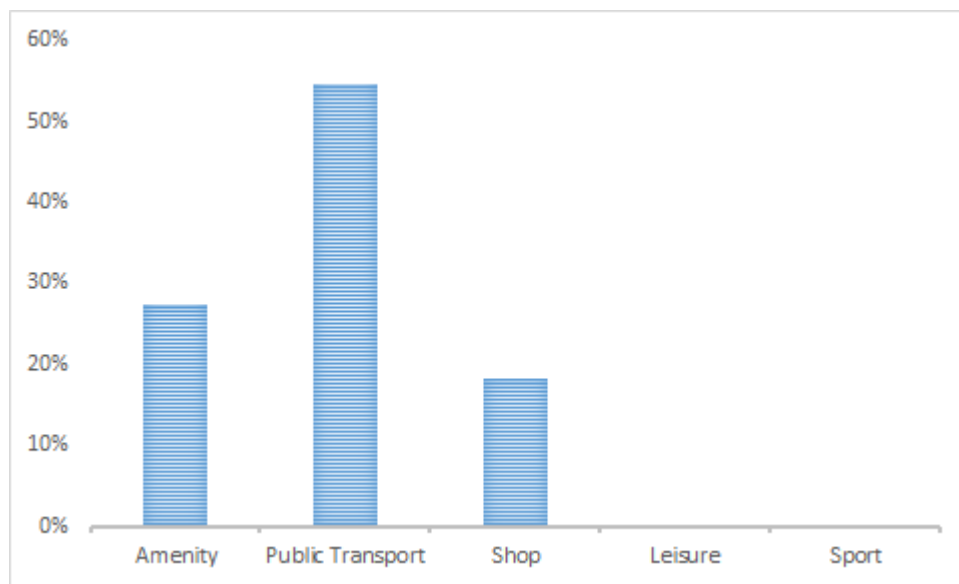


Σχήμα 5.4 Σημείο ενδιαφέροντος το οποίο χαρακτηρίστηκε ως amenity.

Πίνακας 5.3 Κατανομή του αριθμού των συστάδων ανά tag του OpenStreetMaps.

Amenity	Public Transport	Shop	Leisure	Sport
9	18	6	0	0

Επομένως, με βάση τα αποτελέσματα που παρουσιάστηκαν, παρατηρείται σχετικά μικρό ποσοστό επιτυχίας της συγκεκριμένης τεχνικής χαρακτηρισμού σημείων ενδιαφέροντος. Αυτό ενδεχομένως οφείλεται στον περιορισμένο αριθμό διαθέσιμων πληροφοριών του OpenStreetMaps για τα σημεία ενδιαφέροντος του συγκεκριμένου χρήστη, είτε στην ύπαρξη διαφορετικών tag τα οποία δεν διερευνήθηκαν με την παρούσα μεθοδολογία. Επιπλέον, η τεχνική που χρησιμοποιήθηκε δεν μπορεί να εξάγει πληροφορίες για σημεία ενδιαφέροντος όπως το σπίτι και η εργασία του χρήστη, στα οποία πιθανόν ο χρήστης να ξοδεύει αρκετό χρόνο, καθώς το OpenStreetMaps δεν διαθέτει τέτοιου είδους πληροφορίες.



Σχήμα 5.5 Διάγραμμα σχετικής συχνότητας εμφάνισης tag στα σημεία ενδιαφέροντος.

5.3 Αποτελέσματα Εξαγωγής Συνηθειών

Η ανίχνευση συνηθειών, όπως έχει ήδη περιγραφεί, πραγματοποιήθηκε με την εξαγωγή κανόνων συσχέτισης με χρήση του αλγορίθμου Apriori μέσω της πλατφόρμας Weka. Αφού

ολοκληρώθηκε η διαδικασία χαρακτηρισμού των σημείων ενδιαφέροντος, έγινε εξαγωγή των σημείων για κάθε αναγνωρισμένο σημείο ενδιαφέροντος σε μορφή: DayZone, IsWorkingDay, MoveType, Tag και εφαρμόστηκε ο αλγόριθμος Apriori στο αρχείο δεδομένων που παράχθηκε το οποίο περιείχε 790 εγγραφές. Τα αποτελέσματα των κανόνων που προέκυψαν είναι τα εξής:

1. MoveType = Metro ==> IsWorkingDay = true
2. DayZone = Evening ==> IsWorkingDay = true
3. Tag = public_transport ==> IsWorkingDay = true
4. DayZone = Evening MoveType = Metro ==> IsWorkingDay = true
5. MoveType = Metro Tag = public_transport 354 ==> IsWorkingDay = true
6. DayZone = Evening Tag = public_transport 331 ==> IsWorkingDay = true
7. DayZone = Evening MoveType = Metro Tag = public_transport ==> IsWorkingDay = true
8. DayZone = Morning ==> IsWorkingDay = true
9. Tag = amenity ==> IsWorkingDay = true
10. DayZone = Morning MoveType = Metro ==> IsWorkingDay = true
11. MoveType = Metro Tag = amenity ==> IsWorkingDay = true
12. DayZone = Evening Tag = amenity ==> IsWorkingDay = true
13. Tag = shop ==> IsWorkingDay = true
14. DayZone = Morning Tag = shop 127 ==> IsWorkingDay = true
15. MoveType = Metro Tag = shop ==> IsWorkingDay = true
16. DayZone = Morning MoveType = Metro Tag = shop ==> IsWorkingDay = true
17. DayZone = Evening MoveType = Metro Tag = amenity ==> IsWorkingDay = true
18. DayZone = Morning Tag = public_transport ==> IsWorkingDay = true

Συνολικά εξάχθηκαν 18 κανόνες με την παράμετρο confidence ίση με 1, όπως είχε καθοριστεί κατά την παραμετροποίηση του αλγορίθμου. Επιπλέον, παρατηρείται ότι δεν έχουν εξαχθεί συνηθειες του χρήστη από όλους τους κανόνες που εντοπίστηκαν από τον Apriori. Για παράδειγμα, οι κανόνες 2 και 8 δεν αντιπροσωπεύουν κάποια συνήθεια του χρήστη. Ωστόσο, οι 16 κανόνες που απομένουν φαίνεται να περιγράφουν συνήθειες του χρήστη. Ενδεικτικά, από τους κανόνες 10 και 18 μπορούμε να συμπεράνουμε ότι ο χρήστης χρησιμοποιεί μέσα μαζικής μεταφοράς τα πρωινά των εργάσιμων ημερών. Επιπλέον, από τους κανόνες 4,6 και 7 παρατηρείται ότι ο χρήστης συνηθίζει να μετακινείται με μετρό τα απογεύματα των εργάσιμων ημερών. Τέλος, με βάση τον κανόνα 13, ο χρήστης τις καθημερινές επισκέπτεται κάποιο κατάστημα, ενώ σύμφωνα με τους κανόνες 15 και 16 χρησιμοποιεί κάποιο μέσο μαζικής μεταφοράς για να φτάσει στο κατάστημα.

Ανακεφαλαιώνοντας, τα αποτελέσματα της εκτέλεσης του αλγορίθμου Apriori, περιγράφουν μερικές από τις συνήθειες του χρήστη, οι οποίες όμως δεν είναι αρκετές και δεν μπορούν να αξιολογηθούν ως προς την εγκυρότητα τους ώστε να καταλήξουμε σε συμπέρασμα των γενικότερων ενδιαφερόντων του χρήστη, καθώς τα δεδομένα που χρησιμοποιήθηκαν για την εξαγωγή τους είναι περιορισμένα. Αυτό οφείλεται, στον περιορισμένο αριθμό δεδομένων κίνησης αλλά και στο σχετικά μικρό ποσοστό επιτυχίας χαρακτηρισμού των σημείων ενδιαφέροντος, το οποίο έχει ως αποτέλεσμα την εξαγωγή κανόνων συσχέτισης από μικρό αριθμό στιγμάτων.

Κεφάλαιο 6

Συμπεράσματα

6.1 Συμπεράσματα

Στα πλαίσια της παρούσας διπλωματικής εργασίας, αναπτύχθηκε μια εφαρμογή σε γλώσσα Java, η οποία στοχεύει στην ανίχνευση συνηθειών του χρήστη μέσω δεδομένων κίνησης που καταγράφονται από την υπάρχουσα εφαρμογή GPSTracker για λειτουργικό σύστημα Android. Η ανίχνευση των συνηθειών, υλοποιήθηκε αρχικά με τον εντοπισμό των σημείων ενδιαφέροντος τα οποία στη συνέχεια χαρακτηρίστηκαν και με χρήση τεχνικών εξόρυξης δεδομένων έγινε εξαγωγή των συνηθειών σε μορφή κανόνων συσχέτισης.

Όσον αφορά την διαδικασία εντοπισμού σημείων ενδιαφέροντος, αξιολογήθηκε με βάση την ποιότητα των συστάδων που εντοπίστηκαν και σύμφωνα με τα αποτελέσματα οι συστάδες που παράχθηκαν αποτελούν σημεία ενδιαφέροντος, καθώς παρατηρείται ότι πρόκειται για σημεία όπου ο χρήστης κινείται γύρω τους. Ωστόσο, η διαδικασία συσταδοποίησης εκτελέστηκε σε μικρό αριθμό δεδομένων κίνησης ενός χρήστη και κατά συνέπεια, κρίνεται απαραίτητο να εφαρμοστεί σε μεγαλύτερο αριθμό δεδομένων διαφορετικών χρηστών και να πραγματοποιηθεί περαιτέρω αξιολόγηση ώστε να προκύψουν εγκυρότερα συμπεράσματα σε σχέση με την ακρίβεια των αποτελεσμάτων.

Ο χαρακτηρισμός των σημείων ενδιαφέροντος που εντοπίστηκαν, αξιολογήθηκε με βάση τον αριθμό των σημείων ενδιαφέροντος που αξιολογήθηκαν επιτυχώς. Το ποσοστό επιτυχίας που παρουσιάζει η συγκεκριμένη τεχνική σύμφωνα με την αξιολόγηση των αποτελεσμάτων προσεγγίζει το 10,48%, ένα σχετικά μικρό ποσοστό το οποίο διαθέτει πολλά περιθώρια βελτίωσης. Αρχικά, το ποσοστό αυτό μπορεί να δικαιολογηθεί αν λάβουμε υπόψιν το γεγονός ότι η εφαρμογή OpenStreetMaps διαθέτει περιορισμένα δεδομένα για τα σημεία ενδιαφέροντος ειδικά στα όρια της

Ελλάδας, όπου και πραγματοποιήθηκε η έρευνα. Παράλληλα, ένας σημαντικός περιορισμός της συγκεκριμένης τεχνικής είναι η εξαγωγή πληροφοριών με βάση ένα καθορισμένο bounding box, τα όρια του οποίου δεν απέχουν την ίδια απόσταση από το κέντρο κάθε συστάδας, ενώ ενδεχομένως το ποσοστό των χαρακτηρισμένων σημείων ενδιαφέροντος να παρουσίαζε αύξηση, αν η ανάκτηση πληροφοριών γινόταν με βάση μια κυκλική γεωγραφική περιοχή γύρω από το κέντρο κάθε συστάδας. Τέλος, θα πρέπει να σημειωθεί ότι η παρούσα μεθοδολογία δεν έχει τη δυνατότητα να αναγνωρίσει σημεία ενδιαφέροντος παρόμοια με το σπίτι και τον χώρο εργασίας του χρήστη, το οποίο έχει ως αποτέλεσμα να μην χαρακτηρίζονται καθόλου τέτοιου είδους τοποθεσίες.

Επιπλέον, η διαδικασία αναγνώρισης συνηθειών του χρήστη, είχε ως αποτέλεσμα την εξαγωγή κανόνων συσχέτισης που περιγράφουν μερικές από τις καθημερινές συνήθειες του χρήστη. Ωστόσο, ο περιορισμένος αριθμός στιγμάτων προς ανάλυση, καθιστά δύσκολη την αξιολόγηση των συνηθειών που προέκυψαν και την εξαγωγή συμπερασμάτων για την καθημερινότητα και τα ενδιαφέροντα του χρήστη. Όπως προκύπτει από τα αποτελέσματα, είναι απαραίτητο η μεθοδολογία εξαγωγής συνηθειών να εφαρμοστεί σε μεγαλύτερο αριθμό δεδομένων κίνησης πολλαπλών χρηστών και πιθανόν να αξιολογηθούν με την χρήση ερωτηματολογίων όπου οι χρήστες θα καταγράφουν τις πραγματικές τους συνήθειες. Με αυτό τον τρόπο, μπορούν επιπλέον να διερευνηθούν οι παράμετροι εκτέλεσης του αλγορίθμου *Apriori* και να αξιολογηθούν τα αποτελέσματα κανόνων συσχέτισης στα συγκεκριμένα δεδομένα.

Παράλληλα, θα ήταν καλό να αναφερθεί ότι σύμφωνα με τα αποτελέσματα εντοπισμού σημείων ενδιαφέροντος, η καταγραφή των δεδομένων κίνησης αποκλειστικά με τον αισθητήρα GPS αποτελεί έναν περιορισμό της συγκεκριμένης εφαρμογής, καθώς παρατηρήθηκε ότι σε συγκεκριμένες περιπτώσεις το σφάλμα του GPS είναι σημαντικό ώστε να προκύπτουν μη έγκυρα αποτελέσματα. Προτείνεται λοιπόν, η τροποποίηση της διαδικασίας καταγραφής της θέσης του χρήστη με την εφαρμογή GPSTracker, έτσι ώστε να χρησιμοποιούνται περισσότερες υπηρεσίες τοποθεσίας, με στόχο την αύξηση της ακρίβειας καθορισμού θέσης.

Συμπερασματικά, η εφαρμογή που αναπτύχθηκε ανταποκρίνεται στους αρχικούς στόχους της παρούσας διπλωματικής εργασίας, καθώς υλοποιεί την συλλογή και την ανάλυση των δεδομένων κίνησης του χρήστη εξαγοντας τα σημεία ενδιαφέροντος και τις συνήθειες του χρήστη, όπως αρχικά είχε περιγραφεί. Ωστόσο, αποτελεί την πρώτη έκδοση μιας ολοκληρωμένης εφαρμογής αναγνώρισης συνηθειών του χρήστη, καθώς σύμφωνα με την αξιολόγηση της εμφανίζει περιορισμένη ακρίβεια αποτελεσμάτων και έχει περιθώριο πολλαπλών βελτιώσεων ώστε να μπορεί να εξάγει με μεγάλη ακρίβεια τις συνήθειες και τα ενδιαφέροντα του χρήστη.

6.2 Μελλοντικές Επεκτάσεις

Η ανίχνευση συνηθειών μέσω κινητών τηλεφώνων είναι ένας τομέας που παρουσιάζει ενδιαφέρον στην ερευνητική κοινότητα και έχει διεξαχθεί πλήθος ερευνών όπου παρουσιάζονται διαφορετικές τεχνικές αναγνώρισης συνηθειών. Ωστόσο, οι υπάρχουσες τεχνικές ανίχνευσης δραστηριότητας του χρήστη εξετάζουν τις μεταβάσεις μεταξύ των σημείων ενδιαφέροντος, σε αντίθεση με την μεθοδολογία που περιγράφηκε στην παρούσα εργασία η οποία εστιάζει στην σημασιολογική αναγνώριση των σημείων ενδιαφέροντος και την χρονική στιγμή που ο χρήστης επισκέπτεται μια τοποθεσία. Παρόλα αυτά, η παρούσα έκδοση της εφαρμογής απαιτεί αρκετές βελτιώσεις της τρέχουσας τεχνικής και αρκετές προοπτικές επέκτασης της υπάρχουσας λειτουργικότητας ώστε να αυξηθεί η ακρίβεια των αποτελεσμάτων.

Αρχικά, μια σημαντική βελτίωση θα αποτελούσε η τροποποίηση της διαδικασίας καταγραφής των δεδομένων από την εφαρμογή GPSTracker, χρησιμοποιώντας έναν υβριδικό τρόπο καθορισμού θέσης. Αναλυτικότερα, η πληροφορία της θέσης του αισθητήρα GPS θα μπορούσε να συνδυαστεί με πληροφορίες από άλλες υπηρεσίες τοποθεσίας, περιορίζοντας έτσι το σφάλμα του GPS ειδικά όταν ο χρήστης βρίσκεται σε κλειστό χώρο και το σήμα του αισθητήρα GPS είναι περιορισμένο.

Όσον αφορά στην προσέγγιση της εφαρμογής στην διαδικασία χαρακτηρισμού σημείων ενδιαφέροντος, θα ήταν χρήσιμο να διερευνηθεί ο καθορισμός μιας κυκλικής γεωγραφικής περιοχής που να περικλύει το σημείο, αντί για την δημιουργία bounding box, το οποίο ενδέχεται να αυξήσει το ποσοστό επιτυχίας της διαδικασίας. Επιπλέον, ο χαρακτηρισμός των σημείων θα πρέπει με κάποιο τρόπο να επεκταθεί ώστε να είναι δυνατό να αναγνωρίζονται σημεία όπως το σπίτι και ο χώρος εργασίας του χρήστη, εφόσον είναι αδύνατο μέσω της εφαρμογής OpenStreetMaps.

Παράλληλα, προτείνεται το σύνολο της διαδικασίας εξαγωγής συνηθειών να δοκιμαστεί σε περισσότερους χρήστες, οι οποίοι θα καταγράφουν δεδομένα κίνησης για ένα μεγάλο χρονικό διάστημα και θα είναι σε θέση να περιγράψουν τις καθημερινές τους δραστηριότητες και με αυτόν τον τρόπο να αξιολογηθούν τα αποτελέσματα της εφαρμογής. Με αυτόν τον τρόπο, θα είναι δυνατό να εντοπιστούν οι περιορισμοί της συγκεκριμένης εφαρμογής και να βελτιωθούν ώστε να αυξηθεί η ακρίβεια της αναγνώρισης συνηθειών των χρηστών.

Τέλος, η μελλοντική κατεύθυνση της παρούσας προσέγγισης, αφορά την ανάπτυξη συστήματος συστάσεων, το οποίο κάνοντας χρήση τη γνώση για τις συνήθειες του χρήστη, θα μπορούσε να παρέχει προσωποποιημένες υπηρεσίες στους χρήστες κινητών τηλεφώνων. Με αυτό τον τρόπο, διευκολύνεται η πρόσβαση του χρήστη σε πληροφορίες που έχει ανάγκη κάθε χρονική στιγμή,

χωρίς να ξοδεύει χρόνο για την αναζήτηση της πληροφορίας που τον ενδιαφέρει.

Βιβλιογραφία

- [1] Εθνικό Δίκτυο Έρευνας και Τεχνολογίας. Okeanos IAAS Cloud. <https://www.grnet.gr/el/node/399>. [Online· τελευταία προσπέλαση 2-Ιανουαρίου-2016].
- [2] Χαλκίδη Μαρία Βαζιργιάννης Μιχάλης. *Εξόρυξη γνώσης από βάσεις δεδομένων και του παγκόσμιου ιστό*. Τυπωθήτω, 2005.
- [3] Aaron Parecki. Frequent Locations API in iOS 8. <https://aaronparecki.com/articles/2014/06/03/1/frequent-locations-api-in-ios-8>. [Online· τελευταία προσπέλαση 16-Ιανουαρίου-2016].
- [4] Rakesh Agrawal and Ramakrishnan Srikant. Fast algorithms for mining association rules in large databases. In *Proceedings of the 20th International Conference on Very Large Data Bases, VLDB '94*, pages 487–499, San Francisco, CA, USA, 1994. Morgan Kaufmann Publishers Inc.
- [5] Apple. About privacy and Location Services for iOS 8 and iOS 9. <https://support.apple.com/en-us/HT203033>. [Online· τελευταία προσπέλαση 16-Ιανουαρίου-2016].
- [6] Daniel Ashbrook and Thad Starner. Using gps to learn significant locations and predict movement across multiple users. *Personal Ubiquitous Comput.*, 7:275–286, 2003.
- [7] Ed Burnette. *Eclipse IDE Pocket Guide*. O'Reilly Media, Inc., 2005.
- [8] Onur Cinar. *Android Apps with Eclipse*. Apress, Berkely, CA, USA, 1st edition, 2012.
- [9] Francisco Duarte, André Lourenço, and Arnaldo Abrantes. Classification of physical activities using a smartphone: Evaluation study using multiple users. *Procedia Technology*, 17:239 – 247, 2014.

- [10] G. Maggiore, C. Santos, and A. Plaat. Smarter smartphones: Understanding and predicting user habits from gps sensor data. *Procedia Computer Science*, 34:297 – 304, 2014.
- [11] Google. I-Want-to-Go Moments: From Search to Store. <https://www.thinkwithgoogle.com/articles/i-want-to-go-micro-moments.html>. [Online· τελευταία προσπέλαση 9-Ιανουαρίου-2016].
- [12] Google. Micro-Moments. <https://www.thinkwithgoogle.com/micromoments/intro.html>. [Online· τελευταία προσπέλαση 9-Ιανουαρίου-2016].
- [13] Google Play. Moves. <https://play.google.com/store/apps/details?id=com.protogeo.moves>. [Online· τελευταία προσπέλαση 13-Ιανουαρίου-2016].
- [14] B. Grone. *The Apache Modeling Project*. Technische Berichte des Hasso-Plattner-Instituts für Softwaresystemtechnik an der Universität Potsdam. Hasso-Plattner-Inst. für Softwaresystemtechnik, 2004.
- [15] Mark Hall, Eibe Frank, Geoffrey Holmes, Bernhard Pfahringer, Peter Reutemann, and Ian H. Witten. The WEKA data mining software: an update. *SIGKDD Explor. Newsl.*, 11(1):10–18, 2009.
- [16] Samuli Hemminki, Petteri Nurmi, and Sasu Tarkoma. Accelerometer-based transportation mode detection on smartphones. In *Proceedings of the 11th ACM Conference on Embedded Networked Sensor Systems*, SenSys '13, 2013.
- [17] Jinsoo Kim, Seongkyu Lee, Hoyong Ahn, Dongju Seo, Doochun Seo, Jongchool Lee, and Chuluong Choi. accuracy evaluation of a smartphone-based technology for coastal monitoring. *Measurement*, 46(1):233 – 248, 2013.
- [18] Slava Kisilevich, Florian Mansmann, and Daniel Keim. p-dbscan: A density based clustering algorithm for exploration and analysis of attractive areas using collections of geo-tagged photos. In *Proceedings of the 11th ACM Conference on Embedded Networked Sensor Systems*, New York, NY, USA, 2010. ACM.
- [19] Marco Helbich, Christof Amelunxen, Pascal Neis, and Alexander Zipf. comparative spatial analysis of positional accuracy of openstreetmap and proprietary geodata. In *Proceedings of GI Forum 2012: Geovisualization, Society and Learning*, 2012.

- [20] Marko Modsching, Ronny Kramer, and Klaus ten Hagen. Field trial on gps accuracy in a medium size city: The influence of built-up. *3rd workshop on positioning, navigation and communication*, pages 209-218, 2006.
- [21] Srinivas Nanduri and Takwirira Tafireyi. a study on open source software development - an apache web server case study. In *International Conference on ICT for Africa 2013*, 2013.
- [22] OpenStreetMap Foundation (OSMF). Map Features. http://wiki.openstreetmap.org/wiki/Map_Features. [Online· τελευταία προσπέλαση 7-Φεβρουαρίου-2016].
- [23] OpenStreetMap Foundation (OSMF). OpenStreetMap. <https://www.openstreetmap.org/about>. [Online· τελευταία προσπέλαση 2-Ιανουαρίου-2016].
- [24] Andrei Papliatseyeu and Oscar Mayora. Mobile habits: Inferring and predicting user activities with a location-aware smartphone. *3rd Symposium of Ubiquitous Computing and Ambient Intelligence 2008*, 2009.
- [25] ProtoGeo. Activity Diary of Your Life. <http://moves-app.com/>. [Online· τελευταία προσπέλαση 13-Ιανουαρίου-2016].
- [26] The Apache Software Foundation. Commons Math: The Apache Commons Mathematics Library. <https://commons.apache.org/proper/commons-math/>. [Online· τελευταία προσπέλαση 3-Ιανουαρίου-2016].
- [27] The PostgreSQL Global Development Group. PostgreSQL. <http://www.postgresql.org/about/>. [Online· τελευταία προσπέλαση 2-Ιανουαρίου-2016].
- [28] Wikipedia, The Free Encyclopedia. Android Studio. https://en.wikipedia.org/wiki/Android_Studio. [Online· τελευταία προσπέλαση 9-Φεβρουαρίου-2016].
- [29] Wikipedia, The Free Encyclopedia. Apache Commons. https://en.wikipedia.org/wiki/Apache_Commons. [Online· τελευταία προσπέλαση 3-Ιανουαρίου-2016].
- [30] Wikipedia, The Free Encyclopedia. PostgreSQL. <https://en.wikipedia.org/wiki/PostgreSQL>. [Online· τελευταία προσπέλαση 2-Ιανουαρίου-2016].
- [31] Wikipedia, The Free Encyclopedia. Walking. <https://en.wikipedia.org/wiki/Walking>. [Online· τελευταία προσπέλαση 7-Φεβρουαρίου-2016].

- [32] Wikipedia, The Free Encyclopedia. apache http server. https://en.wikipedia.org/wiki/Apache_HTTP_Server. [Online· τελευταία προσπέλαση 21-Φεβρουαρίου-2016].
- [33] Wikipedia, The Free Encyclopedia. haversine formula. https://en.wikipedia.org/wiki/Haversine_formula. [Online· τελευταία προσπέλαση 4-Φεβρουαρίου-2016].
- [34] Wikipedia, The Free Encyclopedia. openstreetmapopenstreetmap. <https://en.wikipedia.org/wiki/OpenStreetMap>. [Online· τελευταία προσπέλαση 2-Ιανουαρίου-2016].