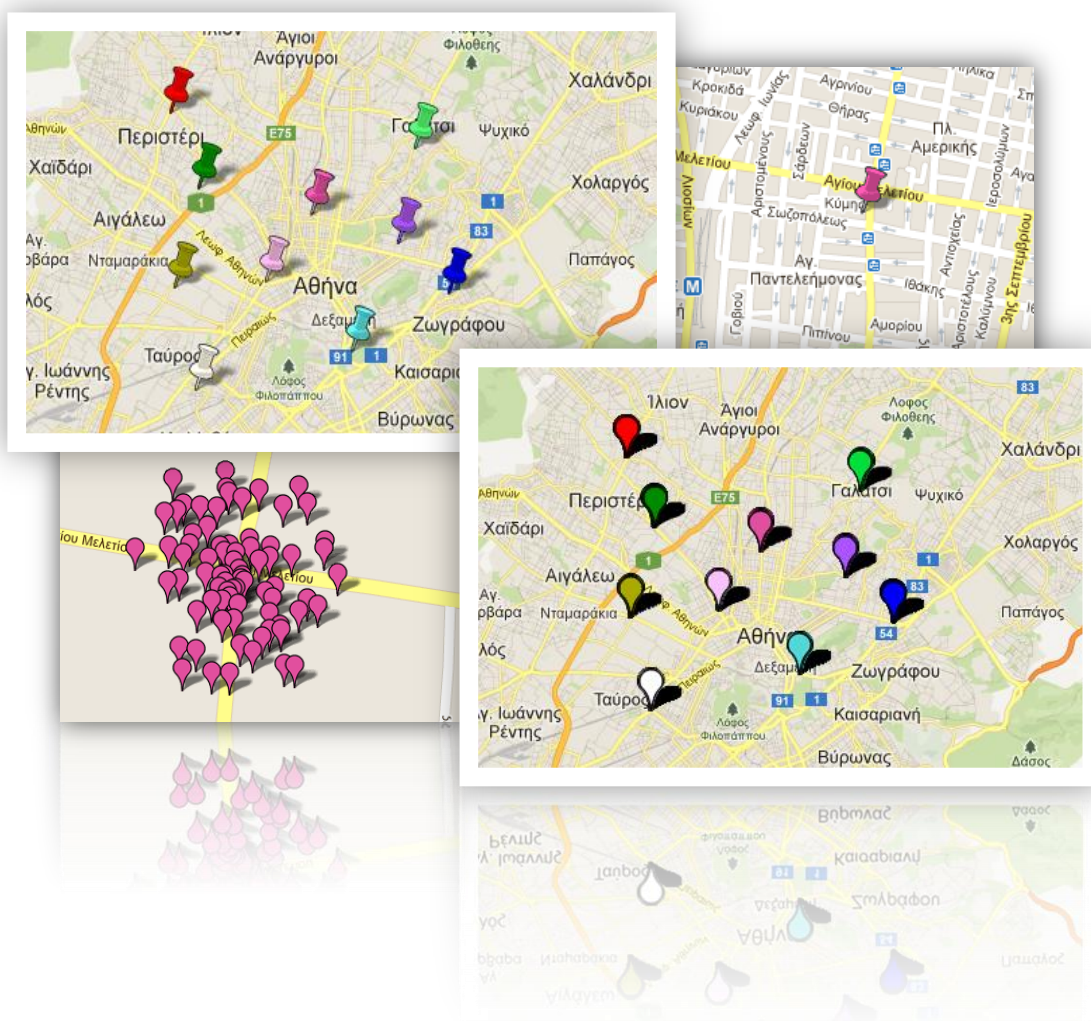


«Αξιοποίηση δεδομένων Crowdsourcing με χωρική αναφορά: μια εφαρμογή για καταγραφή προβλημάτων σε δημόσιες υποδομές»



ΜΕΤΑΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ
ΜΑΡΙΑ Γ.ΣΤΡΙΑΛΙΓΚΑ
ΜΑΡΤΙΟΣ 2013

Επιβλέπων Καθηγητής: Βεσκούκης Βασίλειος

ΠΕΡΙΛΗΨΗ

Ο σκοπός της μεταπτυχιακής εργασίας είναι η δημιουργία μιας εφαρμογής συλλογής και αξιοποίησης δεδομένων πλήθους (crowdsourcing), με συνδυασμό εργαλείων και υπηρεσιών που παρέχονται στο διαδίκτυο. Η εφαρμογή μας θα περιέχει την προδιαγραφή και υλοποίηση μεθόδων καταγραφής και άντλησης δεδομένων με χωρική αναφορά, από τα οποία θα μπορούν να παραχθούν νέα δεδομένα αλλά και χάρτες για την απεικόνιση των αποτελεσμάτων.

Για τη συλλογή δεδομένων από το πλήθος βασικό εργαλείο της εργασίας αποτέλεσε η crowdsourcing εφαρμογή Fulcrum App για Android και iOS κινητά τηλέφωνα, η οποία προσφέρεται δωρεάν στο διαδίκτυο. Η εφαρμογή διαμορφώθηκε και χρησιμοποιήθηκε για τη δημιουργία μίας φόρμας καταγραφής προβλημάτων σε δημόσιες υποδομές και τη συλλογή των αρχικών δεδομένων με τη βοήθεια «έξυπνου» κινητού τηλεφώνου. Τα δεδομένα που συλλέγονται από την εφαρμογή αποθηκεύονται σε πίνακες σε μία βάση δεδομένων για τη μετέπειτα χρήση και αξιοποίησή τους. Υπολογίζονται νέα δεδομένα με βάση τον καταγεγραμμένο τύπο προβλημάτων σε δημόσιες υποδομές τα οποία έχουν αναφερθεί από το πλήθος, καθώς επίσης και στατιστικά στοιχεία των καταγραφών, προκειμένου να παραχθούν χάρτες στο διαδίκτυο.

Για την οπτικοποίηση των αποτελεσμάτων δημιουργούνται δύο χάρτες για κάθε τύπο προβλήματος, εκ των οποίων ο πρώτος περιλαμβάνει τα δεδομένα των καταγραφών πλήθους ως σύνολο και ο δεύτερος τα δεδομένα των προβλημάτων μεμονωμένα που προέκυψαν από επεξεργασία των αρχικών δεδομένων με κατάλληλα επιλεγμένο και εφαρμοσμένο αλγόριθμο.

Για την επίτευξη του παραπάνω σκοπού της παρούσας εργασίας αναπτύχθηκε μια ροή εργασιών για την εξόρυξη και εκμετάλλευση των δεδομένων σε δεδομένη περιοχή με τη χρήση της γλώσσας προγραμματισμού PHP, της ΒΔ Postgres με μια επέκταση που υλοποιεί τον αλγόριθμο k-means, καθώς και τεχνολογίες για την απεικόνιση των αποτελεσμάτων (html, javascript, kml, google maps). Μελετήθηκαν τρεις διαφορετικές περιπτώσεις ως προς τα αρχικά δεδομένα. Στην πρώτη περίπτωση υλοποιήθηκε η εφαρμογή για μικρό αριθμό δεδομένων, στη δεύτερη περίπτωση για μεγάλο όγκο δεδομένων και στην τρίτη περίπτωση μελετήθηκε το ενδεχόμενο της γειννίασης των σημείων διαφορετικών τύπων προβλημάτων.

Τα αποτελέσματα της εφαρμογής αυτής μπορούν να αξιοποιηθούν από δημόσιους φορείς με στόχο τον άμεσο εντοπισμό προβλημάτων με τη βοήθεια των πολιτών, ο οποίος αποτελεί προϋπόθεση για την έγκαιρη και συστηματική εξάλειψη των καταγεγραμμένων προβλημάτων σε δημόσιες υποδομές, παρέχοντας τον πλήρη έλεγχο του συνόλου των δεδομένων στους ενδιαφερόμενους φορείς. Η εφαρμογή μπορεί να αποτελέσει κίνητρο και εργαλείο για τους πολίτες για μεγαλύτερη συμμετοχικότητα στα προβλήματα της πόλης τους. Ηδη σε άλλα πεδία εφαρμογών η ιδέα του crowdsourcing έχει βρει εκατομμύρια χρήστες.

Λέξεις κλειδιά:

Crowdsourcing, καταγραφή προβλημάτων, δημόσιες υποδομές, k-means, βάση δεδομένων, Fulcrum App, KML, PostgreSQL, Quantum GIS, ομαδοποίηση, συλλογή δεδομένων, Javascript, PHP, SQL, KML, Postgis.

ABSTRACT

The purpose of this thesis is to create an application for collecting and utilizing crowd data (crowdsourcing), combining tools and services offered in the web. Our application will contain specification and implementation of reporting and mining methods with spatial reference, from which new data and maps will be generated for the visualization of the results.

A fundamental tool for the crowd data collection of this project was a crowdsourcing application for iOS and Android mobile phones named Fulcrum, which is offered free on the internet. The application was developed and used for creating a damage report form regarding public infrastructure and collecting original data via a smartphone. The data collected by the application is stored in the tables of a database for subsequent use and exploitation. New data is then being calculated based on the reported type of damage in public infrastructure reported by the crowd, as well as statistics of the reports, in order to produce web maps.

Two maps are created for each type of problem for the visualization of the results, the first of which includes the crowd data of the reports as a whole and the second one includes the data of individual problems arising from initial data processing with appropriately selected and implemented algorithm.

In order to achieve the above purpose of this project, a workflow was developed for extracting and exploiting the data in a given region using the PHP programming language, the Postgres Database including an extension which implements the k-means algorithm, as well as particular technologies for the optimization of the results (html, javascript, kml, google maps). Three different cases were studied with respect to the original data. In the first case, the application was implemented for a small amount of data, in the second case for a large amount of data and in the third case we studied the possibility of neighboring points of different types of problems.

The results of this application can be used by public entities with a view to directly identifying problems with the help of the citizens, which is a prerequisite for timely and systematic elimination of reported problems in public infrastructure providing full control of all data to stakeholders. The application can be an incentive and a tool for citizens for greater inclusivity to the problems of their city. In other application fields the idea of crowdsourcing has already found millions of users.

Keywords:

Crowdsourcing, damage report, public infrastructure, k-means, database, Fulcrum App, KML, PostgreSQL, Quantum GIS, clustering, data collection, Javascript, PHP, SQL, KML, Postgis.

ΕΥΧΑΡΙΣΤΙΕΣ

Πριν την παρουσίαση της παρούσας μεταπτυχιακής εργασίας, αισθάνομαι την ανάγκη να ευχαριστήσω τους ανθρώπους που κατείχαν πολύ σημαντικό ρόλο κατά υλοποίησή της.

Πρώτο από όλους θα ήθελα να ευχαριστήσω τον επιβλέποντα της μεταπτυχιακής μου εργασίας, Επίκουρο Καθηγητή Βασίλειο Βεσκούκη για την πολύτιμη συμβολή και καθοδήγησή του, την εμπιστοσύνη και εκτίμηση που μου έδειξε, αλλά και τη συνέπεια και την άψογη συνεργασία που επίσης έδειξε καθ' όλη τη διάρκεια της εκπόνησης της μεταπτυχιακής εργασίας.

Τέλος, θέλω να ευχαριστήσω την οικογένειά μου για την ηθική συμπαράσταση που μου παρείχε όλο αυτόν τον καιρό για την ολοκλήρωση της μεταπτυχιακής μου εργασίας.

ΚΑΤΑΛΟΓΟΣ ΠΕΡΙΕΧΟΜΕΝΩΝ

| | |
|--|-----------|
| ΕΙΣΑΓΩΓΗ | 17 |
| 1 CROWDSOURCING | 19 |
| 1.1 Η Ιδέα..... | 19 |
| 1.1.1 Ιστορία..... | 22 |
| 1.1.2 Πλεονεκτήματα | 22 |
| 1.1.3 Ταξινόμηση των crowdsourcing συστημάτων | 23 |
| 1.1.4 Crowdsourcing συστήματα στο Διαδίκτυο | 25 |
| 1.1.5 Προκλήσεις και Λύσεις..... | 27 |
| 1.1.6 Γεωγραφική Πληροφορία..... | 28 |
| 1.1.7 Κυβερνήσεις και Μη Κερδοσκοπική Χρήση | 29 |
| 1.2 Βασικές Αρχές..... | 30 |
| 1.3 Συσκευές..... | 31 |
| 1.4 Παραδείγματα Εφαρμογών..... | 32 |
| 2 ΑΠΑΙΤΗΣΕΙΣ ΑΠΟ ΕΝΑ ΠΕΔΙΟ ΕΦΑΡΜΟΓΗΣ | 35 |
| 2.1 Καταγραφή Προβλημάτων σε δημόσιες Υποδομές..... | 35 |
| 2.2 Τρόπος Εφαρμογής του Crowdsourcing..... | 36 |
| 2.3 Δεδομένα που θα συλλέγονται και Τρόπος Συλλογής..... | 38 |
| 2.4 Απαιτήσεις Επεξεργασίας..... | 39 |
| 2.5 Απαιτήσεις Εμφάνισης Αποτελεσμάτων | 40 |
| 2.6 Περιορισμοί – Παραδοχές..... | 42 |
| 3 Ο ΑΛΓΟΡΙΘΜΟΣ K-MEANS | 45 |
| 3.1 Γενικά..... | 45 |
| 3.1.1 Ομαδοποίηση..... | 45 |
| 3.1.2 Αλγόριθμοι ομαδοποίησης..... | 46 |
| 3.1.3 Ιστορία..... | 46 |
| 3.1.4 K-means..... | 46 |
| 3.1.5 Ο βασικός αλγόριθμος k-means | 48 |
| 3.1.6 Μέθοδοι αρχικοποίησης | 50 |
| 3.1.7 Εφαρμογές του αλγορίθμου..... | 51 |
| 3.2 Παραλλαγές | 51 |
| 3.3 Απαιτήσεις για το δικό μας Πρόβλημα | 53 |
| 3.4 Περιορισμοί | 55 |
| 4 ΤΕΧΝΟΛΟΓΙΕΣ ΠΟΥ ΧΡΗΣΙΜΟΠΟΙΗΘΗΚΑΝ | 57 |
| 4.1 PHP | 57 |
| 4.1.1 Τι είναι η PHP..... | 57 |
| 4.1.2 Πριν από την Εμφάνιση της PHP..... | 57 |
| 4.1.3 Οι Πρώτες Εκδόσεις της PHP | 58 |
| 4.1.4 Η Τρέχουσα Έκδοση της PHP..... | 59 |
| 4.1.5 Η Σύνδεση με την HTML..... | 59 |
| 4.1.6 Η Διεργασία και η Μεταγλώττιση..... | 60 |
| 4.1.7 Πώς γράφεται η PHP..... | 60 |
| 4.1.8 Πώς δουλεύει η PHP | 61 |
| 4.1.9 Τι μπορεί να κάνει η PHP | 61 |
| 4.2 POSTGRES – POSTGIS – QGIS..... | 62 |

| | | |
|----------|--|------------|
| 4.2.1 | PostgreSQL..... | 63 |
| 4.2.2 | PostGIS..... | 65 |
| 4.2.3 | Quantum GIS | 66 |
| 4.3 | JavaScript | 67 |
| 4.3.1 | Javascript και Java | 69 |
| 4.3.2 | Ιστορία..... | 69 |
| 4.3.3 | Τρέχοντας την JavaScript..... | 70 |
| 4.3.4 | Εισάγοντας JavaScript σε μια HTML σελίδα | 70 |
| 4.3.5 | Non-JavaScript browsers | 71 |
| 4.3.6 | Events | 71 |
| 4.4 | KML | 72 |
| 4.4.1 | Βασικά αρχεία KML | 72 |
| 4.4.2 | Συστατικά μέρη ενός KML αρχείου | 73 |
| 4.4.3 | Διαδρομές - Paths | 81 |
| 4.4.4 | Πολύγωνα - Polygons..... | 82 |
| 4.4.5 | Στυλ για Γεωμετρία | 83 |
| 5 | ΥΛΟΠΟΙΗΣΗ ΤΟΥ ΣΥΣΤΗΜΑΤΟΣ | 85 |
| 5.1 | Βάση Δεδομένων | 86 |
| 5.1.1 | Σύνδεση με την PostgreSQL - Δημιουργία νέας βάσης δεδομένων | 86 |
| 5.1.2 | Επιλογή των κατηγοριών δεδομένων | 88 |
| 5.1.3 | Δημιουργία πινάκων | 88 |
| 5.2 | Ροή Συλλογής και Επεξεργασίας Δεδομένων | 91 |
| 5.2.1 | Σύνδεση με το Fulcrum..... | 92 |
| 5.2.2 | Δημιουργία της εφαρμογής καταγραφής προβλημάτων στο Fulcrum..... | 94 |
| 5.2.3 | Τρόπος συλλογής των δεδομένων | 97 |
| 5.2.4 | Εισαγωγή των δεδομένων στη βάση δεδομένων | 100 |
| 5.3 | Εφαρμογή του αλγορίθμου k-means | 104 |
| 5.4 | Δομή και Εμφάνιση των Αποτελεσμάτων | 110 |
| 5.4.1 | Απεικόνιση αποτελεσμάτων στο Quantum GIS | 110 |
| 5.4.2 | Απεικόνιση αποτελεσμάτων στο Google Maps και σε προσωπική ιστοσελίδα | 113 |
| 6 | ΠΑΡΑΔΕΙΓΜΑ ΜΕ ΕΝΔΕΙΚΤΙΚΑ ΔΕΔΟΜΕΝΑ..... | 117 |
| 6.1 | Περίπτωση 1 | 117 |
| 6.1.1 | Παράμετροι της εφαρμογής | 117 |
| 6.1.2 | Δεδομένα εισόδου | 118 |
| 6.1.3 | Υλοποίηση της εφαρμογής | 119 |
| 6.1.4 | Απεικόνιση ακατέργαστων αποτελεσμάτων στο Quantum GIS..... | 120 |
| 6.1.5 | Τελικά αποτελέσματα | 124 |
| 6.2 | Περίπτωση 2 | 127 |
| 6.2.1 | Παράμετροι της εφαρμογής | 127 |
| 6.2.2 | Δεδομένα εισόδου | 128 |
| 6.2.3 | Υλοποίηση της εφαρμογής | 129 |
| 6.2.4 | Απεικόνιση ακατέργαστων αποτελεσμάτων στο Quantum GIS..... | 130 |
| 6.2.5 | Τελικά αποτελέσματα | 134 |
| 6.3 | Περίπτωση 3 | 139 |
| 6.3.1 | Παράμετροι της εφαρμογής | 139 |
| 6.3.2 | Δεδομένα εισόδου | 139 |
| 6.3.3 | Υλοποίηση της εφαρμογής | 141 |

| | | |
|----------|--|------------|
| 6.3.4 | Απεικόνιση ακατέργαστων αποτελεσμάτων στο Quantum GIS..... | 142 |
| 6.3.5 | Τελικά αποτελέσματα | 146 |
| 7 | ΣΥΜΠΕΡΑΣΜΑΤΑ ΚΑΙ ΠΡΟΤΑΣΕΙΣ..... | 157 |
| 7.1 | Συμπεράσματα..... | 157 |
| 7.2 | Προτάσεις..... | 158 |
| | ΒΙΒΛΙΟΓΡΑΦΙΑ..... | 161 |

ΚΑΤΑΛΟΓΟΣ ΠΙΝΑΚΩΝ

| | | |
|--------------|--|-----|
| Πίνακας 3.1: | Παρουσίαση της λειτουργίας του αλγορίθμου | 49 |
| Πίνακας 4.1: | Συστατικά μέρη ενός KML αρχείου | 74 |
| Πίνακας 5.1: | Οι αρχικοί πίνακες (crowd_data και damage_types) της βάσης δεδομένων, οι στήλες, οι τύποι δεδομένων και τα κλειδιά τους..... | 91 |
| Πίνακας 5.2: | Τύποι δεδομένων που επιλέχθηκαν για κάθε πεδίο..... | 96 |
| Πίνακας 5.3: | Παράδειγμα της δομής των δεδομένων όπως εξάγονται από το csv αρχείο | 100 |
| Πίνακας 5.4: | Αντιστοιχία τύπου προβλήματος και χρώματος απεικόνισης | 102 |
| Πίνακας 5.5: | Οι πίνακες kmeans_polygons, kmeans_points και unique_cases της βάσης δεδομένων, οι στήλες, οι τύποι δεδομένων και τα κλειδιά τους, καθώς και η προσθήκη ξένου κλειδιού στον πίνακα crowd_data..... | 110 |
| Πίνακας 6.1: | Γεωγραφικές συντεταγμένες των δέκα αρχικών σημείων..... | 141 |

ΚΑΤΑΛΟΓΟΣ ΕΙΚΟΝΩΝ

| | | |
|-------------|---|----|
| Εικόνα 2.1: | Στάδια υλοποίησης του συστήματος της καταγραφής προβλημάτων σε δημόσιες υποδομές..... | 37 |
| Εικόνα 2.2: | Ροή σταδίων υλοποίησης του συστήματος της καταγραφής προβλημάτων σε δημόσιες υποδομές..... | 38 |
| Εικόνα 2.3: | Στάδια ομαδοποίησης των δεδομένων | 40 |
| Εικόνα 2.4: | Απεικόνιση του συνόλου των καταγραφών για δεδομένο τύπο προβλήματος..... | 40 |
| Εικόνα 2.5: | Απεικόνιση του συνόλου των καταγραφών για δεδομένη περίπτωση προβλήματος (μεγέθυνση προηγούμενου χάρτη) | 41 |
| Εικόνα 2.6: | Απεικόνιση των μοναδικών περιπτώσεων για δεδομένο τύπο προβλήματος..... | 41 |
| Εικόνα 3.1: | Σύνολο των σημείων πριν την υλοποίηση της ομαδοποίησης | 54 |
| Εικόνα 3.2: | Ιδανικό αποτέλεσμα της ομαδοποίησης k-means των σημείων για k=3..... | 55 |
| Εικόνα 4.1: | Τρόπος λειτουργίας της PHP..... | 60 |
| Εικόνα 4.2: | Τύποι δεδομένων που υποστηρίζει η PostgreSQL | 65 |
| Εικόνα 4.3: | Απεικόνιση αντικειμένου στο Google Earth και εμφάνιση ονόματος αντικειμένου..... | 75 |
| Εικόνα 4.4: | Απεικόνιση εικονιδίων στο Google Earth με διαφορετικά μεγέθη..... | 75 |
| Εικόνα 4.5: | Απεικόνιση της Αθήνας στο Google Earth με tilt=0 (αριστερά) και tilt=67 (δεξιά)..... | 78 |
| Εικόνα 4.6: | Απεικόνιση της Αθήνας στο Google Earth με range=9530.85758459179 | 78 |
| Εικόνα 4.7: | Απεικόνιση της Αθήνας στο Google Earth με range=13530.85758459179 | 78 |
| Εικόνα 4.8: | Απεικόνιση των στοιχείων <range>, <tilt>, <altitude> | 79 |

| | |
|---|-----|
| Εικόνα 4.9: Απεικόνιση στοιχείου <heading> | 80 |
| Εικόνα 4.10: Απεικόνιση διαδρομής στην περιοχή της Αθήνας στο Google Earth | 82 |
| Εικόνα 4.11: Απεικόνιση Πενταγώνου στο Google Earth πριν (αριστερά) και μετά (δεξιά) την δημιουργία πολυγώνου..... | 83 |
| Εικόνα 5.1: Ροή υλοποίησης των rhr αρχείων και τα αποτελέσματα που προέκυψαν από το καθένα..... | 86 |
| Εικόνα 5.2: Δημιουργία νέας βάσης δεδομένων | 87 |
| Εικόνα 5.3: Παράθυρο δημιουργίας των SQL ερωτημάτων | 87 |
| Εικόνα 5.4: Δημιουργία προσωπικού λογαριασμού στην ιστοσελίδα του Fulcrum..... | 92 |
| Εικόνα 5.5: Σύνδεση στο λογαριασμό της εφαρμογής FulcrumApp από το κινητό τηλέφωνο..... | 93 |
| Εικόνα 5.6: Σύνδεση στον προσωπικό λογαριασμό στην ιστοσελίδα του Fulcrum | 93 |
| Εικόνα 5.7: Δημιουργία πεδίων και επιλογή τύπων δεδομένων στην εφαρμογή του Fulcrum..... | 94 |
| Εικόνα 5.8: Τύποι προβλήματος που καταχωρήθηκαν στο πεδίο απλής επιλογής «Τύπος προβλήματος» | 95 |
| Εικόνα 5.9: Βαθμολόγηση της σοβαρότητας του προβλήματος από 1 έως 5 στο πεδίο κειμένου «Αξιολόγηση προβλήματος» | 96 |
| Εικόνα 5.10: Prinscreen από την εφαρμογή Damage Report στο κινητό τηλέφωνο – Αριστερά οι διαθέσιμες εφαρμογές, δεξιά χάρτης επιλογής θέσης..... | 97 |
| Εικόνα 5.11: Prinscreen από την εφαρμογή Damage Report στο κινητό τηλέφωνο – Συμπλήρωση προσωπικών δεδομένων | 97 |
| Εικόνα 5.12: Prinscreen από την εφαρμογή Damage Report στο κινητό τηλέφωνο – Συμπλήρωση πληροφοριών προβλήματος | 98 |
| Εικόνα 5.13: Στάδιο 1 – Επιλογή τύπου των εξαγόμενων δεδομένων | 99 |
| Εικόνα 5.14: Στάδιο 2 - Επιλογή των πεδίων και της ονομασίας τους στο εξαγόμενο αρχείο | 99 |
| Εικόνα 5.15: Στάδιο 3 – Επιβεβαίωση εξαγωγής..... | 99 |
| Εικόνα 5.16: Εισαγωγή των δεδομένων των τύπων προβλημάτων στον πίνακα «damage_types» | 101 |
| Εικόνα 5.17: Εισαγωγή των δεδομένων των καταγραφών στον πίνακα «crowd_data» .. | 104 |
| Εικόνα 5.18: Απεικόνιση της ομαδοποίησης των σημείων με τον αλγόριθμο k-means χωρίς τη χρήση διανυσμάτων αρχικοποίησης | 106 |
| Εικόνα 5.19: Εσφαλμένη (αριστερά και κέντρο) και φαινομενικά ορθή (δεξιά) ομαδοποίηση των σημείων | 106 |
| Εικόνα 5.20: Τελικώς εσφαλμένη ομαδοποίηση των σημείων | 106 |
| Εικόνα 5.21: Δημιουργία του πίνακα «kmeans_polygons»..... | 108 |
| Εικόνα 5.22: Δημιουργία του πίνακα «kmeans_points»..... | 109 |
| Εικόνα 5.23: Δημιουργία του πίνακα «unique_cases» | 109 |
| Εικόνα 5.24: Δημιουργία PostGIS σύνδεσης στο περιβάλλον του QuantumGIS | 111 |
| Εικόνα 5.25: Εισαγωγή ονόματος και κωδικού χρήστη..... | 112 |
| Εικόνα 5.26: Επιλογή των πινάκων και προσθήκη αυτών στη βάση του PostGIS..... | 112 |
| Εικόνα 6.1: Απεικόνιση των δέκα αρχικών σημείων στο περιβάλλον του Google Maps . | 118 |
| Εικόνα 6.2: Γεωγραφικές συντεταγμένες των δέκα αρχικών σημείων..... | 119 |
| Εικόνα 6.3: Οι πίνακες που δημιουργήθηκαν στη βάση δεδομένων για την «Περίπτωση 1» | 120 |
| Εικόνα 6.4: Απεικόνιση του συνόλου των αρχικών δεδομένων των καταγραφών..... | 121 |
| Εικόνα 6.5: Μεγέθυνση της περιοχής των καταγραφών συγκεκριμένου τύπου προβλήματος..... | 122 |
| Εικόνα 6.6: Σύνολο των καταγραφών για τον τύπο προβλήματος «χαλασμένο φανάρι» (πράσινο χρώμα) | 122 |

| | |
|--|-----|
| Εικόνα 6.7: Απεικόνιση πολυγώνων ομαδοποίησης σε δεδομένη έκταση | 122 |
| Εικόνα 6.8: Εστίαση σε ένα πολύγωνο ομαδοποίησης..... | 123 |
| Εικόνα 6.9: Χρωματική διαφοροποίηση των δεδομένων με βάση την ομαδοποίηση k-means..... | 123 |
| Εικόνα 6.10: Απεικόνιση των μοναδικών σημείων που αντιπροσωπεύουν το σύνολο των καταγραφών κάθε περίπτωσης | 123 |
| Εικόνα 6.11: Απεικόνιση του συνόλου των καταγραφών ανά τύπο προβλήματος στο περιβάλλον του Google Maps για την «Περίπτωση 1» | 125 |
| Εικόνα 6.12: Μεγέθυνση της απεικόνισης των καταγραφών του προβλήματος «χαλασμένο φανάρι» στο περιβάλλον του Google Maps για την «Περίπτωση 1» | 125 |
| Εικόνα 6.13: Απεικόνιση του συνόλου των μοναδικών περιπτώσεων ανά τύπο προβλήματος στο περιβάλλον του Google Maps για την «Περίπτωση 1» | 126 |
| Εικόνα 6.14: Μεγέθυνση της απεικόνισης της μοναδικής περίπτωσης του προβλήματος «χαλασμένο φανάρι» στο περιβάλλον του Google Maps για την «Περίπτωση 1» | 126 |
| Εικόνα 6.15: Απεικόνιση του συνόλου των καταγραφών ανά τύπο προβλήματος σε προσωπική ιστοσελίδα για την «Περίπτωση 1» | 127 |
| Εικόνα 6.16: Απεικόνιση του συνόλου των μοναδικών περιπτώσεων ανά τύπο προβλήματος σε προσωπική ιστοσελίδα για την «Περίπτωση 1» | 127 |
| Εικόνα 6.17: Απεικόνιση των 161 αρχικών σημείων στο περιβάλλον του Google Maps. | 129 |
| Εικόνα 6.18: Οι πίνακες που δημιουργήθηκαν στη βάση δεδομένων για την «Περίπτωση 2» | 130 |
| Εικόνα 6.19: Απεικόνιση του συνόλου των αρχικών δεδομένων των καταγραφών..... | 131 |
| Εικόνα 6.20: Σύνολο των καταγραφών για τον τύπο προβλήματος «χαλασμένο φανάρι» (πράσινο χρώμα) | 132 |
| Εικόνα 6.21: Απεικόνιση πολυγώνων ομαδοποίησης σε δεδομένη έκταση | 132 |
| Εικόνα 6.22: Εστίαση σε ένα πολύγωνο ομαδοποίησης..... | 133 |
| Εικόνα 6.23: Χρωματική διαφοροποίηση των δεδομένων με βάση την ομαδοποίηση k-means..... | 133 |
| Εικόνα 6.24: Απεικόνιση των μοναδικών σημείων που αντιπροσωπεύουν το σύνολο των καταγραφών κάθε περίπτωσης | 134 |
| Εικόνα 6.25: Απεικόνιση του συνόλου των καταγραφών ανά τύπο προβλήματος στο περιβάλλον του Google Maps για την «Περίπτωση 2» | 136 |
| Εικόνα 6.26: Μεγέθυνση της απεικόνισης των καταγραφών του προβλήματος «χαλασμένο φανάρι» στο περιβάλλον του Google Maps για την «Περίπτωση 2» | 136 |
| Εικόνα 6.27: Απεικόνιση του συνόλου των μοναδικών περιπτώσεων ανά τύπο προβλήματος στο περιβάλλον του Google Maps για την «Περίπτωση 2» | 137 |
| Εικόνα 6.28: Μεγέθυνση της απεικόνισης των μοναδικών περιπτώσεων του προβλήματος «χαλασμένο φανάρι» στο περιβάλλον του Google Maps για την «Περίπτωση 2» | 137 |
| Εικόνα 6.29: Απεικόνιση του συνόλου των καταγραφών ανά τύπο προβλήματος σε προσωπική ιστοσελίδα για την «Περίπτωση 2» | 138 |
| Εικόνα 6.30: Μεγέθυνση της απεικόνισης των καταγραφών του προβλήματος «λακκούβα» σε προσωπική ιστοσελίδα για την «Περίπτωση 2»..... | 138 |
| Εικόνα 6.31: Απεικόνιση του συνόλου των μοναδικών περιπτώσεων ανά τύπο προβλήματος σε προσωπική ιστοσελίδα για την «Περίπτωση 2» | 138 |
| Εικόνα 6.32: Απεικόνιση των 10 αρχικών σημείων σε πολύ μικρή έκταση σε σχέση με τις προηγούμενες περιπτώσεις το περιβάλλον του Google Maps | 140 |
| Εικόνα 6.33: Μεγέθυνση των 10 επιλεγμένων αρχικών σημείων στο περιβάλλον του Google Maps | 140 |
| Εικόνα 6.34: Οι πίνακες που δημιουργήθηκαν στη βάση δεδομένων για την «Περίπτωση 3» | 142 |
| Εικόνα 6.35: Απεικόνιση του συνόλου των αρχικών δεδομένων των καταγραφών..... | 143 |

| | |
|---|-----|
| Εικόνα 6.36: Απεικόνιση του συνόλου των αρχικών δεδομένων των καταγραφών..... | 144 |
| Εικόνα 6.37: Σύνολο των καταγραφών για τον τύπο προβλήματος «σπασμένο πεζοδρόμιο» (πράσινο χρώμα)..... | 144 |
| Εικόνα 6.38: Απεικόνιση πολυγώνων ομαδοποίησης και των σημείων των καταγραφών | 144 |
| Εικόνα 6.39: Χρωματική διαφοροποίηση των δεδομένων με βάση την ομαδοποίηση k-means..... | 145 |
| Εικόνα 6.40: Απεικόνιση των μοναδικών σημείων που αντιπροσωπεύουν το σύνολο των καταγραφών κάθε περίπτωσης..... | 145 |
| Εικόνα 6.41: Απεικόνιση των μοναδικών σημείων σε σχέση με το σύνολο των καταγραφών κάθε περίπτωσης..... | 145 |
| Εικόνα 6.42: Απεικόνιση του συνόλου των μοναδικών περιπτώσεων ανά τύπο προβλήματος για απόσταση ομαδοποίησης 10 m στο περιβάλλον του Google Maps για την «Περίπτωση 3»..... | 147 |
| Εικόνα 6.43: Απεικόνιση του συνόλου των μοναδικών περιπτώσεων ανά τύπο προβλήματος για απόσταση ομαδοποίησης 10 m στο περιβάλλον του Google Maps για την «Περίπτωση 3»..... | 148 |
| Εικόνα 6.44: Απεικόνιση του συνόλου των μοναδικών περιπτώσεων ανά τύπο προβλήματος για απόσταση ομαδοποίησης 15 m στο περιβάλλον του Google Maps για την «Περίπτωση 3»..... | 148 |
| Εικόνα 6.45: Απεικόνιση του συνόλου των μοναδικών περιπτώσεων ανά τύπο προβλήματος για απόσταση ομαδοποίησης 15 m στο περιβάλλον του Google Maps για την «Περίπτωση 3»..... | 149 |
| Εικόνα 6.46: Απεικόνιση του συνόλου των μοναδικών περιπτώσεων ανά τύπο προβλήματος για απόσταση ομαδοποίησης 20 m στο περιβάλλον του Google Maps για την «Περίπτωση 3»..... | 149 |
| Εικόνα 6.47: Απεικόνιση του συνόλου των μοναδικών περιπτώσεων ανά τύπο προβλήματος για απόσταση ομαδοποίησης 20 m στο περιβάλλον του Google Maps για την «Περίπτωση 3»..... | 150 |
| Εικόνα 6.48: Απεικόνιση του συνόλου των μοναδικών περιπτώσεων ανά τύπο προβλήματος για απόσταση ομαδοποίησης 25 m στο περιβάλλον του Google Maps για την «Περίπτωση 3»..... | 150 |
| Εικόνα 6.49: Απεικόνιση του συνόλου των μοναδικών περιπτώσεων ανά τύπο προβλήματος για απόσταση ομαδοποίησης 25 m στο περιβάλλον του Google Maps για την «Περίπτωση 3»..... | 151 |
| Εικόνα 6.50: Απεικόνιση του συνόλου των μοναδικών περιπτώσεων ανά τύπο προβλήματος για απόσταση ομαδοποίησης 30 m στο περιβάλλον του Google Maps για την «Περίπτωση 3»..... | 151 |
| Εικόνα 6.51: Απεικόνιση του συνόλου των μοναδικών περιπτώσεων ανά τύπο προβλήματος για απόσταση ομαδοποίησης 30 m στο περιβάλλον του Google Maps για την «Περίπτωση 3»..... | 152 |
| Εικόνα 6.52: Απεικόνιση του συνόλου των μοναδικών περιπτώσεων ανά τύπο προβλήματος για απόσταση ομαδοποίησης 35 m στο περιβάλλον του Google Maps για την «Περίπτωση 3»..... | 152 |
| Εικόνα 6.53: Απεικόνιση του συνόλου των μοναδικών περιπτώσεων ανά τύπο προβλήματος για απόσταση ομαδοποίησης 35 m στο περιβάλλον του Google Maps για την «Περίπτωση 3»..... | 153 |
| Εικόνα 6.54: Απεικόνιση του συνόλου των μοναδικών περιπτώσεων ανά τύπο προβλήματος για απόσταση ομαδοποίησης 40 m στο περιβάλλον του Google Maps για την «Περίπτωση 3»..... | 153 |

| | |
|---|-----|
| Εικόνα 6.55: Απεικόνιση του συνόλου των μοναδικών περιπτώσεων ανά τύπο προβλήματος για απόσταση ομαδοποίησης 40 m στο περιβάλλον του Google Maps για την «Περίπτωση 3»..... | 154 |
| Εικόνα 6.56: Περιπτώσεις προβλημάτων με όμοιο τύπο προβλήματος ανά αριθμό | 155 |
| Εικόνα 6.57: Απεικόνιση του συνόλου των μοναδικών περιπτώσεων ανά τύπο προβλήματος για απόσταση ομαδοποίησης 20 m σε προσωπική ιστοσελίδα για την «Περίπτωση 3»..... | 155 |
| Εικόνα 6.58: Απεικόνιση του συνόλου των μοναδικών περιπτώσεων ανά τύπο προβλήματος για απόσταση ομαδοποίησης 20 m σε προσωπική ιστοσελίδα για την «Περίπτωση 3»..... | 156 |

ΕΙΣΑΓΩΓΗ

Προσέγγιση του Προβλήματος

Η ολοένα αυξανόμενη ανάπτυξη των crowdsourcing τεχνολογιών ανοίγει ένα νέο ορίζοντα στη συλλογή μαζικής πληροφορίας και την επίλυση προβλημάτων σε πολλούς και διαφορετικούς τομείς. Οι τεχνολογίες crowdsourcing σε συνδυασμό με την τεχνολογική καινοτομία προσφέρουν τη δυνατότητα αποτελεσματικής αξιοποίησης δεδομένων που παρέχονται σε σχεδόν πραγματικό χρόνο από το πλήθος με χαμηλό κόστος, και αποτέλεσαν το εφαλτήριο για την υλοποίηση της εργασίας. Ο στόχος της μεταπτυχιακής εργασίας ήταν η δημιουργία μιας εφαρμογής η οποία θα στηρίζεται στη μαζική πληροφορία που θα προέρχεται από την αξιοποίηση crowdsourcing δεδομένων για την καταγραφή προβλημάτων σε δημόσιες υποδομές. Το αντικείμενο αυτό είναι ενδεικτικό και κατάλληλο για να αναδειχθούν οι δυνατότητες των συναφών τεχνολογιών σε δεδομένα με χωρική αναφορά.

Αντικείμενο της Μεταπτυχιακής Εργασίας

Το αντικείμενο της μεταπτυχιακής εργασίας είναι η δημιουργία μίας ολοκληρωμένης εφαρμογής για τη συλλογή χρήσιμης πληροφορίας με crowdsourcing τεχνολογίες και την ταυτόχρονη αξιοποίηση της πληροφορίας αυτής για την ανάδειξη των ήδη υπαρχόντων αλλά και των αναδιδόμενων προβλημάτων σε δημόσιες υποδομές με σκοπό την έγκαιρη εξάλειψή τους.

Η εργασία αυτή προσπαθεί να λύσει το πρόβλημα της συστηματοποίησης του εντοπισμού, της συγκέντρωσης, της ταξινόμησης και της εκτίμησης των διαφόρων προβλημάτων δημόσιων υποδομών εύκολα και οικονομικά, αλλά και της απεικόνισης των προβλημάτων αυτών σε διαδραστικούς χάρτες στο διαδίκτυο σύμφωνα με συγκεκριμένα κριτήρια.

Αναλυτικότερα, στην παρούσα εργασία υλοποιήθηκε ένα σύστημα διαχείρισης χωρικών δεδομένων, το οποίο ενσωματώνει, αποθηκεύει, προσαρμόζει, αναλύει και παρουσιάζει γεωγραφικά συσχετισμένες πληροφορίες. Οι πληροφορίες με τις οποίες εμπλουτίζεται το σύστημα αυτό προέρχονται από την αξιοποίηση της crowdsourcing εφαρμογής Fulcrum App. Το πλήθος μπορεί να καταγράψει προβλήματα που συναντά σε δημόσιες υποδομές με χρήση της εφαρμογής Fulcrum App από την Android/iOS κινητή συσκευή του, επιλέγοντας από μία λίστα πιθανών προβλημάτων και καταχωρώντας προαιρετικά τα προσωπικά του στοιχεία. Η πληροφορία θα εισάγεται σε μία βάση δεδομένων και στη συνέχεια θα αντλούνται και θα ομαδοποιούνται δεδομένα με βάση τον τύπο του προβλήματος για την απεικόνιση των διαφόρων καταγραφών σε διαδικτυακούς χάρτες και συγκεκριμένα στο περιβάλλον Google maps.

Η εφαρμογή αυτή μπορεί να αξιοποιηθεί για την αντιμετώπιση των καταγεγραμμένων προβλημάτων με εκτίμηση των εξαγόμενων στατιστικών στοιχείων για το εκάστοτε πρόβλημα. Η εφαρμογή υλοποιήθηκε στα πλαίσια της μεταπτυχιακής εργασίας με

προοπτικές περαιτέρω βελτίωσης και εμπλουτισμού για την αξιοποίηση από τις δημόσιες αρχές.

Συνεισφορά

Η συνεισφορά της μεταπτυχιακής εργασίας συνοψίζεται ως εξής:

1. Σχεδιάστηκε μια ολοκληρωμένη ροή συλλογής δεδομένων με χωρική αναφορά από το πλήθος, με χρήση διαφορετικών εργαλείων σε μη ομοιογενή υπολογιστικά περιβάλλοντα.
2. Μελετήθηκε η εφαρμογή αλγορίθμου k-means για τη φύση του συγκεκριμένου προβλήματος.
3. Υλοποιήθηκε η ΒΔ και η εφαρμογή για τη συγκέντρωση των δεδομένων.
4. Υλοποιήθηκε ο αλγόριθμος, ο οποίος στηρίζεται στην καταχώρηση της πληροφορίας στη χωρική βάση δεδομένων και την αξιοποίηση των δεδομένων της για την παραγωγή χαρτών.
5. Αξιολογήθηκαν τα αποτελέσματα και η συμπεριφορά των συγκεκριμένων αλγορίθμων.
6. Εξήχθησαν ανάλογα συμπεράσματα.

Οργάνωση Κειμένου

Συνοπτικά, η μεταπτυχιακή εργασία περιλαμβάνει την εισαγωγή, επτά κεφάλαια, τη βιβλιογραφία και το παράρτημα. Τα κεφάλαια της εργασίας αναπτύσσονται ως εξής:

Στο πρώτο κεφάλαιο, γίνεται μια γενική περιγραφή των crowdsourcing τεχνολογιών.

Στο δεύτερο κεφάλαιο, γίνεται μια γενική περιγραφή των απαιτήσεων του πεδίου εφαρμογής.

Στο τρίτο κεφάλαιο, περιγράφεται πλήρως ο αλγόριθμος k-means γενικά αλλά και ειδικά σε σχέση με το δικό μας πρόβλημα.

Στο τέταρτο κεφάλαιο, γίνεται μια λεπτομερής περιγραφή των τεχνολογιών που χρησιμοποιήθηκαν κατά την εκπόνηση της εργασίας.

Στο πέμπτο κεφάλαιο, αναλύεται διεξοδικά το σύστημα που υλοποιήθηκε και η ροή των εργασιών.

Στο έκτο κεφάλαιο, παρατίθενται τρία παραδείγματα που υλοποιήθηκαν με ενδεικτικά δεδομένα.

Στο έβδομο κεφάλαιο παρουσιάζονται τα συμπεράσματα που εξήχθησαν από την διεκπεραίωση της εργασίας και προτείνονται ιδέες για περεταίρω βελτίωση της εφαρμογής.

1

CROWDSOURCING

1.1 Η Ιδέα

Ως **crowdsourcing** αναφέρεται η πράξη της εξωτερικής ανάθεσης καθηκόντων, που παραδοσιακά εκτελούνταν από υπάλληλο ή εργολάβο, σε ένα μεγάλο πλήθος εθελοντών ή μία κοινότητα μέσω ανοικτής πρόσκλησης. Πρόκειται δηλαδή για την **εποχή του πλήθους**. Τα συστήματα crowdsourcing στρατολογούν μια πληθώρα ανθρώπων να βοηθήσουν στην επίλυση ενός μεγάλου φάσματος προβλημάτων. Όπως είναι σύνηθες για έναν ανερχόμενο τομέα, η προσπάθεια αυτή εμφανίστηκε με πολλές ονομασίες, συμπεριλαμβανομένου των peer production, user-powered systems, user-generated content, collaborative systems, community systems, social systems, social search, social media, collective intelligence, wikinomics, crowd wisdom, smart mobs, mass collaboration και human computation, όμως επικράτησε ο όρος crowdsourcing.

Οι Estellés-Arolas και González Ladrón-de- Guevara έδωσαν τον εξής ορισμό για το crowdsourcing:

«Το crowdsourcing είναι μία μορφή συλλογικής διαδικτυακής δραστηριότητας στην οποία ένα άτομο, ένα ίδρυμα, ένας μη κερδοσκοπικός οργανισμός ή μία εταιρεία προτείνει σε μία ομάδα ατόμων με ποικίλλες γνώσεις, ετερογένεια και αριθμό, μέσω μίας ανοικτής πρόσκλησης, να αναλάβουν εθελοντικά μια εργασία».

Η ανάληψη της εργασίας, η οποία ποικίλλει σε πολυπλοκότητα και στο βαθμό στον οποίο είναι χωρισμένη και στην οποία το πλήθος πρέπει να συμμετάσχει με προσωπική εργασία, χρήματα, γνώση, εμπειρία, περιλαμβάνει πάντοτε αμοιβαίο όφελος και για τις δύο πλευρές. Οι χρήστες λαμβάνουν την ικανοποίηση κάποιας ανάγκης τους, είτε αυτή είναι οικονομική, είτε κοινωνική αναγνώριση, προσωπική ικανοποίηση, ανάπτυξη ατομικών ικανοτήτων σε κάποιο τομέα, ενώ ο εκκινήτης της πρωτοβουλίας αποκτά και χρησιμοποιεί προς όφελός του, αυτά που έχει συνεισφέρει ο χρήστης στο εγχείρημα, τα οποία εξαρτώνται από τη δραστηριότητα που έχει αναλάβει ο χρήστης.

Ο Jeff Howe, ήταν ένας από τους πρώτους που χρησιμοποίησαν τον όρο, και καθόρισε πώς η έννοια του crowdsourcing εξαρτάται ουσιαστικά από το γεγονός πως εφόσον πρόκειται για ανοικτή πρόσκληση σε ένα άγνωστο πλήθος ατόμων, συγκεντρώνει αυτούς που είναι οι πιο ικανοί για να αναλάβουν καθήκοντα, να λύσουν πολύπλοκα προβλήματα και να συνεισφέρουν με τις πλέον σχετικές και φρέσκιες ιδέες.

Για παράδειγμα, το κοινό μπορεί να προσκληθεί να αναπτύξει μια νέα τεχνολογία, να αναλάβει μια εργασία σχεδίασης (γνωστό και ως συνεργατικός σχεδιασμός ή δημοκρατικός σχεδιασμός και κατανεμημένος συνεργατικός σχεδιασμός) να εξειδικεύσει ή να υλοποιήσει τα βήματα ενός αλγορίθμου (ανθρωπογενής υπολογιστική), ή να βοηθήσει στην αποτύπωση, συστηματοποίηση ή ανάλυση δεδομένων μεγάλου μεγέθους.

Ο όρος *crowdsourcing* χρησιμοποιείται συχνά για εκείνες τις μεθόδους δημιουργίας δεδομένων κατά τις οποίες η παρουσία ενός μεγάλου αριθμού χρηστών, οι οποίοι δεν είναι κεντρικά οργανωμένοι, αναπαράγει περιεχόμενο το οποίο κοινοποιείται. Σε ένα **γεωγραφικό πλαίσιο**, η εθελοντική γεωγραφική πληροφορία είναι ο όρος που χρησιμοποιείται όλο και περισσότερο για συστήματα όπως χάρτες που κατασκευάζονται από χρήστες (Goodchild, 2007).

Η ιδέα της χρήσης ατομικής γνώσης των χρηστών και η καταγραφή αυτής έτσι ώστε να προκύπτουν νέα πρότυπα δεδομένων, επαφίεται στην ιδέα ότι τα πλήθη των χρηστών έχουν τη δική τους λογική η οποία δεν μπορεί να καταγραφεί αλλιώς, παρά μόνο το πλήθος που εξασκεί ατομικά τη λογική αυτή. Αυτή η ιδέα βασίζεται στο γεγονός ότι παρόλο που ένας μεγάλος αριθμός ατομικών εκτιμήσεων μπορεί να είναι εσφαλμένος, ο μέσος όρος τους μπορεί να προσεγγίζει καλύτερα το σημείο από οποιαδήποτε ατομική εκτίμηση. Με συνετό χειρισμό, η τυχαία λήψη δειγμάτων από γνώμες ή υπολογισμούς ενός μεγάλου αριθμού χρηστών μπορεί να οδηγήσει σε δεδομένα και πληροφορία εκπληκτικά ακριβή, τα οποία σε ορισμένες περιπτώσεις δε μπορούν να καταγραφούν με κάποιον άλλο τρόπο (Surowiecki, 2004).

Όταν αυτό το είδος της ανάπτυξης εφαρμόζεται στη δημιουργία δεδομένων τα οποία στη συνέχεια είναι προσβάσιμα και κοινοποιήσιμα όπως μία διαδικτυακή υπηρεσία, πρόκειται για **crowdsourcing**. Αυτό βασίζεται στην ιδέα ότι οι Web 2.0¹ τεχνολογίες είναι ικανές να μοχλεύουν τεράστιες ποσότητες δεδομένων, ιδιαίτερα για αγορές από εκείνους που αλληλεπιδρούν στην αγορά. Ο όρος αγορά αναφέρεται σε ένα forum όπου οι χρήστες μπορούν να αλληλεπιδρούν και να κοινοποιούν αντί να αγοράζουν και να πωλούν, το οποίο θα αποτελέσει ένα υποσύνολο αυτών των αγορών. Κατά μία έννοια, επιτρέποντας στους χρήστες να αναπαράγουν το δικό τους περιεχόμενο, κάτι το οποίο δεν είναι καινούριο, δημιουργείται η βάση για ομαδική ψυχανάλυση και επίλυση προβλημάτων. Αυτό που είναι καινούριο είναι η αντίληψη ότι αυτού του είδους η ομαδική δυναμική είναι ασυντόνιστη και εθελοντική, με τη δύναμη του μέσου να εξαρτάται από το πόσο χρήσιμη θεωρεί ο χρήστης ότι είναι η εργασία της διάθεσης πληροφορίας. Σε πολλές περιπτώσεις, τα διατιθέμενα δεδομένα απλά προστίθενται στο σωρό και μέχρι να πραγματοποιηθεί αλληλεπίδραση μεταξύ των χρηστών, η μόνη προστιθέμενη αξία οφείλεται στην έκταση στην οποία ο προστιθέμενος σωρός δεδομένων αντιπροσωπεύει κάποια φαινόμενα ενδιαφέροντος. Όταν η αλληλεπίδραση ξεκινάει μεταξύ των χρηστών και καινούρια δεδομένα παράγονται από αυτήν την αλληλεπίδραση και προκύπτουν κοινωνικά

¹ Ο όρος Web 2.0 (Ιστός 2.0), χρησιμοποιείται για να περιγράψει τη νέα γενιά του Παγκόσμιου Ιστού η οποία βασίζεται στην όλο και μεγαλύτερη δυνατότητα των χρηστών του Διαδικτύου να μοιράζονται πληροφορίες και να συνεργάζονται online. Αυτή η νέα γενιά είναι μια δυναμική διαδικτυακή πλατφόρμα στην οποία μπορούν να αλληλεπιδρούν χρήστες χωρίς εξειδικευμένες γνώσεις σε θέματα υπολογιστών και δικτύων.

δίκτυα , τότε η προστιθέμενη αξία εξαρτάται από το ίδιο το πλήθος, παρόλο που σχεδιάζεται με αμέτρητους τρόπους από σχεδιαστές και διαχειριστές των διαδικτυακών υπηρεσιών που εμπλέκονται.

Στις αναπτυσσόμενες τεχνολογίες οι οποίες εξαρτώνται από τέτοιου είδους άντληση πληροφοριών, η κοινωνία των χρηστών παραμένει ακαθόριστη. Συνήθεις Web 2.0 ιστοσελίδες κοινωνικών δικτύων όπως το Facebook και κοινόχρηστες βάσεις δεδομένων όπως το Flickr καθιστούν τη διάδοση των χρηστών από στόμα σε στόμα, χρησιμοποιώντας ένα ολόκληρο φάσμα μέσων στα οποία οι πιθανοί χρήστες ανταποκρίνονται αλλά υπάρχει και ο παράγοντας της άμεσης ώθησης από τους κατόχους ή τους διαχειριστές ή τους σχεδιαστές αυτών των ιστοσελίδων. Όταν υπάρχει πιο άμεση πίεση ψηφιακά με τη χρήση online συστημάτων, αυτό αναφέρεται ως **crowdcasting** το οποίο είναι η ένωση του broadcasting και του crowdsourcing, με το εν δυνάμει πλήθος να «απομακρύνεται» ως άμεσο αποτέλεσμα της «πίεσης». Τα κίνητρα για την ανάπτυξη συστημάτων κατ' αυτόν τον τρόπο εξαρτώνται καθαρά από την αξία που διαβλέπει το πλήθος από την αλληλεπίδραση με αυτόν τον τρόπο.

Ο όρος crowdsourcing έχει γίνει δημοφιλής στις επιχειρήσεις και όχι μόνο, ως μια συντόμευση της τάσης για την ανάδειξη της μαζικής συνεργασίας που προσφέρουν οι τεχνολογίες του Web 2.0, προκειμένου να επιτευχθούν συγκεκριμένοι επιχειρηματικοί στόχοι. Επιπλέον, το crowdsourcing αξιοποιείται από κυβερνήσεις και μη κερδοσκοπικούς οργανισμούς για κοινωφελείς λόγους, διευρύνοντας τη συμμετοχικότητα και ενισχύοντας την κοινωνική συνοχή.

Πρέπει να σημειωθεί η σχέση των crowdsourcing συστημάτων με το Διαδίκτυο. Τα crowdsourcing συστήματα αλλά και γενικότερα τα κατανεμημένα δίκτυα εργασίας χρησιμοποιούν το Διαδίκτυο για να εκμεταλλευτούν τη διαθέσιμη επεξεργαστική ισχύ εκατομμυρίων ανθρώπινων μυαλών ακριβώς όπως τα κατανεμημένα υπολογιστικά έργα έχουν αξιοποιήσει την αχρησιμοποίητη επεξεργαστική ισχύ από εκατομμύρια μεμονωμένους υπολογιστές.

Η Wikipedia έδειξε ότι το crowdsourcing μοντέλο μπορεί να χρησιμοποιηθεί για τη δημιουργία μιας εκπληκτικά ολοκληρωμένης ηλεκτρονικής εγκυκλοπαίδειας. Αλλά και εταιρείες όπως η eBay και MySpace έχουν δημιουργήσει επικερδείς επιχειρήσεις που δεν θα μπορούσαν να υπάρξουν χωρίς τη συνεισφορά των χρηστών. Όλες αυτές οι εταιρείες αναπτύχθηκαν στην εποχή του Διαδικτύου και έχουν σχεδιαστεί για να επωφεληθούν από το δικτυωμένο κόσμο. Ωστόσο τώρα το παραγωγικό δυναμικό εκατομμυρίων ατόμων προσελκύει την προσοχή των old-line επιχειρήσεων. Την τελευταία δεκαετία, οι εταιρείες έχουν ψάξει στο εξωτερικό, στην Ινδία ή την Κίνα, για φθινό εργατικό δυναμικό. Όμως πλέον δεν έχει σημασία που βρίσκονται οι εργαζόμενοι αρκεί να είναι συνδεδεμένοι στο δίκτυο.

Η τεχνολογική πρόοδος σε όλα, από το λογισμικό σχεδιασμού προϊόντων μέχρι και τις ψηφιακές φωτογραφικές μηχανές, σπάει τα εμπόδια του κόστους που κάποτε χώριζαν τους ερασιτέχνες από τους επαγγελματίες.

Οι ερασιτέχνες αλλά και οι εργαζόμενοι μερικής απασχόλησης απέκτησαν μια αγορά για τις προσπάθειές τους. Η αγορά εργασίας δεν είναι πάντα δωρεάν, αλλά κοστίζει πολύ λιγότερο από ό, τι κοστίζουν οι παραδοσιακοί υπάλληλοι.

1.1.1 Ιστορία

Ο όρος «crowdsourcing» προέρχεται από την ένωση των λέξεων crowd (πλήθος) και outsourcing (εξωτερική ανάθεση εργασιών) και χρησιμοποιήθηκε για πρώτη φορά από τον Jeff Howe τον Ιούνιο του 2006 στην έκδοση του Wired με τίτλο «The Rise of Crowdsourcing». Ο Howe εξηγεί πως επειδή η εξέλιξη της τεχνολογίας έχει οδηγήσει σε φθηνότερες καταναλωτικές ηλεκτρονικές συσκευές, η απόσταση μεταξύ επαγγελματιών και ερασιτεχνών έχει ελαχιστοποιηθεί. Οι εταιρείες μπορούν πλέον να αξιοποιήσουν τα ταλέντα του κοινού και ο Howe σημειώνει πως «Δεν πρόκειται για outsourcing; είναι crowdsourcing». Μια λιγότερο εμπορική προσέγγιση δόθηκε από τον Henk van Ess το Σεπτέμβριο του 2010: «Το Crowdsourcing είναι η δρομολόγηση της επιθυμίας των ειδικών να λύσουν ένα πρόβλημα και να μοιραστούν ελεύθερα την απάντησή του με όλους».

Τα έργα που χρησιμοποιούν τη συλλογική ευφυΐα, όπως το LazyWeb ή το ESP Game του Luis von Ahn, προηγούνται του όρου crowdsourcing για πάρα πολλά χρόνια. Πρόσφατα, το Διαδίκτυο χρησιμοποιείται για τη δημοσίευση και διαχείριση έργων crowdsourcing.

1.1.2 Πλεονεκτήματα

Το crowdsourcing είναι ένα μοντέλο κατανεμημένης επίλυσης προβλημάτων και λειτουργίας. Με την τυπική έννοια του όρου, τα προβλήματα δημοσιεύονται σε ένα άγνωστο πλήθος χρηστών, μέσω μιας ανοικτής πρόσκλησης για την παροχή λύσεων. Οι χρήστες - οι οποίοι αναφέρονται και ως το πλήθος, δηλαδή crowd - σχηματίζουν κοινότητες στο Διαδίκτυο και υποβάλλουν λύσεις. Το πλήθος αξιολογεί επίσης τις λύσεις, επιλέγοντας τις καλύτερες. Αυτές οι λύσεις επιλέγονται από την αρχική οντότητα που έθεσε το πρόβλημα, ενώ τα φυσικά πρόσωπα που της πρότειναν, ανταμείβονται κάποιες φορές. Σε ορισμένες περιπτώσεις, αυτή η εργασία ανταμείβεται είτε χρηματικά, είτε με βραβεία, είτε με αναγνώριση. Σε κάποιες άλλες περιπτώσεις, η ανταμοιβή είναι απλά η φήμη, είτε η διανοητική ικανοποίηση. Το crowdsourcing μπορεί να παράγει λύσεις από ερασιτέχνες ή εθελοντές που δουλεύουν στον ελεύθερο χρόνο τους, ή από ειδικούς και μικρές επιχειρήσεις που είναι άγνωστες στο διοργανωτή. Ο Jeff Howe έχει διακρίνει τέσσερις τύπους στρατηγικών για crowdsourcing:

- Crowdfunding (συγκέντρωση χρηματικών κεφαλαίων)
- Crowdcreation (συλλογική δημιουργία)
- Crowdvoting (συλλογική ψηφοφορία)
- Crowd wisdom (συλλογική ευφυΐα)

Τα **πλεονεκτήματα** του crowdsourcing είναι τα παρακάτω:

- Τα προβλήματα μπορούν να επιλυθούν με συγκριτικά μικρό κόστος και συχνά πολύ γρήγορα.

- Η ανταμοιβή γίνεται με βάση τα αποτελέσματα ή δεν υπάρχει καν
- Ο οργανισμός μπορεί να αξιοποιήσει ένα μεγαλύτερο εύρος δεξιοτήτων από αυτό που διαθέτει.
- Ακούγοντας το πλήθος, οι οργανισμοί αποκτούν άμεση επίγνωση των επιθυμιών των πελατών τους.

Το crowdsourcing έχει δεχθεί κριτική για διάφορους λόγους, όπως για το ότι δεν παράγει πάντα ποιοτικά αποτελέσματα, ή επειδή χρησιμοποιείται για την ανάθεση φθηνής - ή ακόμη δωρεάν - εργασίας. Όταν το Facebook ξεκίνησε το πρόγραμμα προσαρμογής του σε τοπικές γλώσσες το 2008, αντιμετώπισε και τα δύο είδη κριτικής.

Η διαφορά μεταξύ του **crowdsourcing** και του τυπικού **outsourcing** είναι πως το πρόβλημα ανατίθεται σε ένα αδιευκρίνιστο κοινό, αντί σε έναν συγκεκριμένο οργανισμό. Η διαφορά μεταξύ του crowdsourcing και του ανοικτού λογισμικού είναι πως η παραγωγή ανοικτού λογισμικού είναι μια συνεργατική διαδικασία που εκκινείται και αναλαμβάνεται εθελοντικά από μέλη του κοινού. Στο crowdsourcing η δραστηριότητα ξεκινά από έναν οργανισμό και η δουλειά μπορεί να αναληφθεί από ένα άτομο ή μια ομάδα.

1.1.3 Ταξινόμηση των crowdsourcing συστημάτων

Τα συστήματα crowdsourcing αντιμετωπίζουν τέσσερις μεγάλες προκλήσεις: πώς να προσλαμβάνουν συνεργάτες, τι μπορούν να κάνουν οι συνεργάτες αυτοί, πώς να συνδυάσουν τις συνεισφορές τους, και πώς να διαχειριστούν τις καταχρήσεις. Πολλά συστήματα πρέπει να εξισορροπήσουν προσεκτικά την διαφάνεια με την ποιότητα. Ο αγώνας είναι στη δόμηση γενικότερα crowdsourcing πλατφόρμων που μπορούν να χρησιμοποιηθούν για να δημιουργήσουν γρήγορα crowdsourcing εφαρμογές σε πολλούς τομείς. Χρησιμοποιώντας τις εφαρμογές αυτές, μπορούμε να δημιουργήσουμε βάσεις δεδομένων, η δημιουργία των οποίων στο παρελθόν ήταν αδιανόητη σε αστραπιαία ταχύτητα.

Τα crowdsourcing συστήματα μπορούν να ταξινομηθούν με πολλούς τρόπους. Εξετάζονται 9 τρόποι που θεωρούνται ως πιο σημαντικοί. Οι δύο τρόποι που έρχονται άμεσα στο μυαλό είναι *η φύση της συνεργασίας* και *ο τύπος του προβλήματος*. Η συνεργασία μπορεί να είναι άμεση ή έμμεση, και το πρόβλημα μπορεί να είναι οποιοδήποτε πρόβλημα οριστεί από τους ιδιοκτήτες του συστήματος. Οι επόμενοι τέσσερις τρόποι ταξινόμησης αναφέρονται στο πως ένα crowdsourcing σύστημα επιλύει τις τέσσερις βασικές προκλήσεις που περιγράφηκαν νωρίτερα: πώς θα στρατολογηθούν και θα διατηρηθούν οι χρήστες, τι θα μπορούν να κάνουν, πώς θα συνδεθούν τα δεδομένα τους και πώς θα αξιολογηθούν. Οι υπόλοιποι τρεις τρόποι ταξινόμησης οι οποίοι και αναλύονται στη συνέχεια ως πιο σημαντικοί είναι: ο βαθμός της ανθρώπινης προσπάθειας, ο ρόλος των χρηστών, και αυτόνομες έναντι εξαρτημένων αρχιτεκτονικών.

Βαθμός χειρωνακτικής προσπάθειας. Όταν φτιάχνουμε ένα crowdsourcing σύστημα, πρέπει να αποφασίσουμε πόση χειρωνακτική προσπάθεια χρειάζεται για την επίλυση καθεμιάς από τις τέσσερις crowdsourcing προκλήσεις. Μπορεί να κυμαίνεται από

σχετικά μικρή έως ουσιώδης και εξαρτάται καθαρά από το βαθμό αυτοματοποίησης του συστήματος. Πρέπει να αποφασίσουμε πως θα διαχωρίσουμε τη χειρωνακτική προσπάθεια μεταξύ των χρηστών και τους ιδιοκτήτες τους συστήματος. Κάποια συστήματα ζητούν από τους χρήστες να συνεισφέρουν σχετικά λίγο και οι ιδιοκτήτες να κάνουν την περισσότερη δουλειά. Για παράδειγμα, για την ανίχνευση κακόβουλων χρηστών, οι χρήστες μπορούν απλά να πατήσουν ένα κουμπί για να αναφέρουν την ύποπτη συμπεριφορά, ωστόσο οι ιδιοκτήτες πρέπει να εξετάσουν προσεκτικά όλα τα σχετικά στοιχεία για να προσδιορίσουν εάν ένας χρήστης είναι πραγματικά κακόβουλος. Κάποια συστήματα κάνουν το ανάποδο. Για παράδειγμα, το μεγαλύτερο χειρωνακτικό βάρος της συγχώνευσης των επεξεργασιών της Wikipedia ανατίθεται στους χρήστες (οι οποίοι επεξεργάζονται) και όχι στους διαχειριστές.

Ο ρόλος των χρηστών. Θεωρούμε τέσσερις βασικούς ρόλους των ανθρώπων σε ένα crowdsourcing σύστημα:

Slaves: οι άνθρωποι βοηθούν στην επίλυση του προβλήματος με έναν «διαίρει και βασίλευε» τρόπο, για την ελαχιστοποίηση των πόρων (όπως χρόνος, προσπάθεια) των χρηστών. Κάποια παραδείγματα είναι το παιχνίδι ESP και η αναζήτηση ενός αγνοούμενου πλοίου σε δορυφορικές εικόνες με τη χρήση του Mechanical Turk.

Perspective providers: οι άνθρωποι συνεισφέρουν διάφορες προοπτικές, οι οποίες όταν συνδυαστούν συχνά παράγουν μία καλύτερη λύση απ' ό,τι με έναν μόνο άνθρωπο. Παραδείγματα αποτελούν η κριτική βιβλίων και η συγκέντρωση στοιχημάτων ανθρώπων για τη διεξαγωγή προβλέψεων.

Content providers: οι άνθρωποι συνεισφέρουν με αυτοδημιούργητο περιεχόμενο όπως συμβαίνει με τα βίντεο στο YouTube και τις εικόνες στο Flickr.

Component providers: οι άνθρωποι λειτουργούν ως συστατικά του έργου-στόχου, όπως ένα κοινωνικό δίκτυο, ή απλά μια κοινωνία χρηστών.

Οι άνθρωποι συχνά διαδραματίζουν πολλαπλούς ρόλους μέσα σε ένα crowdsourcing σύστημα. Είναι σημαντικό να γνωρίζεις κανείς αυτούς τους ρόλους διότι αυτό θα καθορίσει τη στρατολόγηση των χρηστών.

Αυτόνομο έναντι εξαρτημένου. Κατά τη δημιουργία ενός crowdsourcing συστήματος, μπορούμε να αποφασίσουμε να στηριχθούμε σε ένα καλά οργανωμένο σύστημα, εκμεταλλευόμενοι στοιχεία τα οποία οι χρήστες αφήνουν στο σύστημα αυτό για να λύσουν το πρόβλημά τους. Για παράδειγμα, το Search Assist του Yahoo χρησιμοποιεί το αρχείο καταγραφής αναζήτησης και τα κλικ του χρήστη σε μια μηχανής αναζήτησης για να διορθώσει τα ορθογραφικά λάθη. Ένα άλλο σύστημα μπορεί να εκμεταλλευτεί τις αγορές του χρήστη σε ένα online κατάστημα βιβλίων (Amazon) για να προτείνει βιβλία. Σε αντίθεση με τα αυτόνομα συστήματα, τέτοια εξαρτημένα συστήματα δε χρειάζεται να επιλύσουν τις προκλήσεις της στρατολόγησης των χρηστών και της απόφασης του τι μπορούν να κάνουν. Παρόλα αυτά πρέπει να αποφασίσουν πώς να αξιολογήσουν τους χρήστες και τα δεδομένα εισόδου τους, και πώς να συνδυάσουν τέτοια δεδομένα για να λύσουν το πρόβλημα.

1.1.4 Crowdsourcing συστήματα στο Διαδίκτυο

Βασισμένο στις στατικές σελίδες HTML, το Διαδίκτυο προσέφερε πολλές διαδραστικές υπηρεσίες. Ορισμένες υπηρεσίες εξυπηρετούν μηχανές (όπως το Google Map API), αλλά οι περισσότερες εξυπηρετούν τον άνθρωπο. Πολλές από τις υπηρεσίες αυτές δεν χρειάζεται να «προσλάβουν» χρήστες, με την έννοια όσο περισσότεροι τόσο καλύτερα. Κάποια παραδείγματα περιλαμβάνουν υπηρεσίες πληρωμής στάθμευσης για τους κατοίκους της πόλης και υπηρεσίες κρατήσεων δωματίου. Πολλές υπηρεσίες, ωστόσο, αντιμετωπίζουν crowdsourcing προκλήσεις, συμπεριλαμβανομένης της ανάγκης να αναπτύξουν μεγάλο αριθμό βάσεων χρηστών. Αυτό το Δίκτυο αυξάνεται ταχύτατα, με πολλά νέα crowdsourcing συστήματα να αναπτύσσονται και τα μη-crowdsourcing συστήματα να προσθέτουν crowdsourcing χαρακτηριστικά.

Στην συνέχεια εστιάζουμε σε **συστήματα crowdsourcing** τα οποία διαχειρίζονται το διαδίκτυο για να επιλύσουν τις τέσσερις προκλήσεις που αναφέρθηκαν ή ένα σημαντικό υποσύνολό τους. Το διαδίκτυο είναι μοναδικό στο ότι μπορεί να βοηθήσει στη συγκέντρωση ενός μεγάλου αριθμού χρηστών, να συμβάλλει σε μεγάλο βαθμό στην αυτοματοποίηση της διαδικασίας και να προσφέρει μεγάλη ποσότητα κοινωνικού λογισμικού (όπως email, wiki, blogging και tagging) τα οποία μπορούν να χρησιμοποιήσουν τα crowdsourcing συστήματα για να χειριστούν τους χρήστες τους. Έτσι, σε σύγκριση με το φυσικό κόσμο, το διαδίκτυο μπορεί να βελτιώσει δραματικά τα υπάρχοντα crowdsourcing συστήματα και να δώσει το έναυσμα για καινοτόμους τύπους συστημάτων.

Αναλύοντας τις διαστάσεις του crowdsourcing, εστιάζουμε στα crowdsourcing συστήματα στο διαδίκτυο, περιγράφοντας ένα σύνολο βασικών τύπων συστημάτων. Στη συνέχεια περιγράφονται τα σύνολα ξεκινώντας από τα άμεσα συστήματα.

Άμεσα συστήματα. Αυτά τα αυτόνομα συστήματα επιτρέπουν στους χρήστες να συνεργάζονται άμεσα. Συγκεκριμένα, οι χρήστες μπορούν να αξιολογήσουν, να κοινοποιήσουν, να δικτυωθούν, να δημιουργήσουν και να διεκπεραιώσουν εργασίες. Τα συστήματα αυτά είναι:

Αξιολόγησης. Τα συστήματα αυτά επιτρέπουν στους χρήστες να αξιολογήσουν τα «αντικείμενα», για παράδειγμα βιβλία, ταινίες, ιστοσελίδες, άλλους χρήστες, χρησιμοποιώντας λεκτικά σχόλια, αριθμητικά σκορ ή επισημάνσεις.

Κοινοποίησης. Τα συστήματα αυτά επιτρέπουν στους χρήστες να κοινοποιούν «αντικείμενα» όπως προϊόντα, υπηρεσίες, λεκτική γνώση και δομημένη γνώση. Στα συστήματα που μοιράζονται προϊόντα και υπηρεσίες ανήκει και το YouTube, ενώ στα συστήματα που κοινοποιούν λεκτική γνώση ανήκει το Twitter. Στα συστήματα που κοινοποιούν δομημένη γνώση, για παράδειγμα XML ή RDF δεδομένα, ανήκουν το Swivel και το Google Fusion Tables. Σε γενικές γραμμές, η κοινοποίηση των συστημάτων μπορεί να είναι κεντρική (όπως το YouTube και το Google Fusion Tables) ή κατακεντρωμένη (όπως το Napster ή το Orchestra).

Δικτύωσης. Τα συστήματα αυτά επιτρέπουν στους χρήστες συνεργαζόμενοι να δημιουργήσουν ένα μεγάλο γράφο κοινωνικού δικτύου, προσθέτοντας κόμβους και

ακμές. Έπειτα εκμεταλλεύονται το γράφο για να προσφέρουν υπηρεσίες. Σε μικρότερο βαθμό, τα συστήματα blogging είναι επίσης δικτυακά συστήματα καθώς οι bloggers μπορούν να συνδέονται και με άλλους bloggers. Ένα κλειδί για τη διάκριση των συστημάτων που αξιολογούν, κοινοποιούν ή δικτυώνονται είναι ότι δεν συγχωνεύουν τα δεδομένα εισόδου των χρηστών ή το κάνουν αυτόματα με σχετικά απλούς τρόπους.

Δημιουργίας έργων. Τα συστήματα που επιτρέπουν στους χρήστες να δημιουργούν έργα όπως το Wikipedia συχνά συγχωνεύουν τα δεδομένα εισόδου των χρηστών, και ζητούν από τους χρήστες να επεξεργαστούν και να συγχωνεύσουν μεταξύ τους τα δεδομένα, σε αντίθεση με τα προηγούμενα συστήματα. Ένα τέτοιο δημοφιλές έργο είναι οι λεκτικές Γνωσιακές Βάσεις (Knowledge Bases – KBs). Για να δημιουργηθούν τέτοιες KBs όπως το Wikipedia, οι χρήστες συνεισφέρουν με δεδομένα όπως προτάσεις, παραγράφους, ιστοσελίδες, κι έπειτα επεξεργάζονται και συγχωνεύουν ο ένας τις συνεισφορές του άλλου. Ένα ακόμα δημοφιλές έργο είναι οι δομημένες Γνωσιακές Βάσεις. Για παράδειγμα, το σύνολο όλων των πινάκων πληροφοριών του Wikipedia μπορούν να θεωρηθούν ως μία δομημένη Γνωσιακή Βάση που δημιουργήθηκε υπό τη συνεργασία των χρηστών του Wikipedia. Ένα άλλο ενδιαφέρον πρόβλημα είναι η δημιουργία και η βελτίωση συστημάτων που τρέχουν στο ίντερνετ. Γενικά, οι χρήστες μπορούν να βοηθήσουν στην κατασκευή και βελτίωση ενός συστήματος που εκτελείται στο Web.

Εκτέλεσης εργασιών. Ο τελευταίος τύπος άμεσων συστημάτων που θεωρούμε είναι ο τύπος που εκτελεί εργασίες. Τα παραδείγματα περιλαμβάνουν την εξόρυξη του χρυσού, την αναζήτηση εξαφανισθέντων ατόμων και τον συνεργατικό εντοπισμό σφαλμάτων. Οι εκλογές του 2008 είναι ένα πολύ γνωστό παράδειγμα, όπου η ομάδα του Obama «έτρεξε» μία μεγάλη ηλεκτρονική crowdsourcing λειτουργία ζητώντας από πολλούς εθελοντές να βοηθήσουν στην κινητοποίηση των ψηφοφόρων. Για να εφαρμόσουμε crowdsourcing σε μια εργασία, πρέπει να βρούμε τα μέρη της εργασίας στα οποία μπορεί να εφαρμοστεί «crowdsourcing», έτσι ώστε κάθε χρήστης να μπορεί να κάνει μια συνεισφορά και οι συνεισφορές με τη σειρά τους να μπορούν να συνδυαστούν για την επίλυση των μερών.

Έμμεσα συστήματα. Τα συστήματα αυτά επιτρέπουν στους χρήστες να συνεργάζονται έμμεσα για να λύσουν ένα πρόβλημα των ιδιοκτητών του συστήματος. Εμπίπτουν σε δύο ομάδες: τα αυτόνομα και τα εξαρτημένα. Ένα αυτόνομο σύστημα παρέχει μια υπηρεσία έτσι ώστε τα συστήματα να επιτρέπουν στους χρήστες να συνεργάζονται έμμεσα (δευτερευόντως) για την επίλυση ενός προβλήματος. Το ESP game επιτρέπει στους χρήστες να παίξουν ένα παιχνίδι με το οποίο μαντεύουν κοινές λέξεις που περιγράφουν εικόνες, οι οποίες εμφανίζονται ανεξάρτητα για κάθε χρήστη, και στη συνέχεια χρησιμοποιεί αυτές τις λέξεις για να δώσει ετικέτες στις εικόνες. Το νόημα είναι ότι η «συλλογική σοφία» συχνά είναι ακριβής υπό ορισμένες προϋποθέσεις και ότι αυτό βοηθά να ενσωματωθούν εμπιστευτικές πληροφορίες που είναι διαθέσιμες από τους χρήστες. Τελικά, μπορεί να υποστηριχθεί ότι το πρόβλημα πολλών συστημάτων που παρέχουν υπηρεσίες χρήστη είναι απλά να αναπτυχθεί μια μεγάλη κοινότητα χρηστών για διάφορους λόγους, όπως προσωπική ικανοποίηση, χρέωση τελών εγγραφής, πώληση διαφημίσεων, πώληση των συστημάτων σε άλλες εταιρείες. Το δεύτερο είδος έμμεσου συστήματος θεωρούμε ότι είναι ένα εξαρτημένο

σύστημα που εκμεταλλεύεται τα ίχνη του χρήστη ενός άλλου συστήματος (υποχρεώνοντας έτσι τους χρήστες αυτού του τελευταίου συστήματος να συνεργάζονται έμμεσα) για την επίλυση ενός προβλήματος. Τα συστήματα αυτά εκμεταλλεύονται τα ίχνη της μηχανής αναζήτησης των χρηστών (όπως η αναζήτηση αρχείων καταγραφής και τα κλικ του χρήστη) για ένα ευρύ φάσμα εργασιών (όπως η διόρθωση ορθογραφίας, η εύρεση συνωνύμων, η πρόβλεψη επιδημίας γρίπης και η αναπαραγωγή λέξεων-κλειδιών για διαφημίσεις).

1.1.5 Προκλήσεις και Λύσεις

Οι βασικές προκλήσεις των crowdsourcing συστημάτων οι οποίες αναφέρθηκαν και παραπάνω, αναλύονται στη συνέχεια και είναι οι εξής:

Πώς να στρατολογήσουμε και να διατηρήσουμε τους χρήστες; Η στρατολόγηση των χρηστών είναι μία από τις πιο σημαντικές crowdsourcing προκλήσεις, για τις οποίες υπάρχουν πέντε μεγάλες λύσεις. Πρώτον, μπορούμε να απαιτήσουμε από τους χρήστες να συνεισφέρουν, αν έχουμε τη δικαιοδοσία. Δεύτερον, μπορούμε να πληρώνουμε τους χρήστες. Τρίτον, μπορούμε να ζητήσουμε εθελοντές. Αυτή η λύση είναι δωρεάν και εύκολο να εκτελεστεί, και ως εκ τούτου είναι η πιο δημοφιλής. Τα πιο πρόσφατα crowdsourcing συστήματα στο Διαδίκτυο (όπως το Wikipedia και το YouTube), χρησιμοποιούν αυτή τη λύση. Το μειονέκτημα του εθελοντισμού είναι ότι είναι δύσκολο να προβλέψει τον αριθμό των χρηστών που συμμετέχουν σε μια συγκεκριμένη εφαρμογή. Η τέταρτη λύση είναι να ζητήσει από τους χρήστες να πληρώσουν για την υπηρεσία. Η βασική ιδέα είναι να απαιτήσουμε από τους χρήστες του συστήματος A να «πληρώσουν» για τη χρήση του, συμβάλλοντας σε ένα crowdsourcing σύστημα B. Η πέμπτη λύση είναι να στηριχτούμε στα ίχνη των χρηστών ενός καλά οργανωμένου συστήματος. Αυτό μας δίνει μια σταθερή ροή των χρηστών. Αλλά ακόμα πρέπει να λύσουμε τη δύσκολη πρόκληση του καθορισμού του τρόπου εκμετάλλευσης των ιχνών για το δικό μας σκοπό.

Αφού έχουμε επιλέξει μια στρατηγική στρατολόγησης, θα πρέπει να εξετάσουμε πώς να ενθαρρύνουμε περαιτέρω και να διατηρήσουμε τους χρήστες. Υπάρχουν πολλά προγράμματα ενθάρρυνσης και τη διατήρησης. Αναφέρονται εν συντομία τα πιο δημοφιλή. Πρώτον, μπορούμε να παρέχουμε άμεση ικανοποίηση, δείχνοντας άμεσα πώς η συμβολή ενός χρήστη κάνει τη διαφορά. Δεύτερον, μπορούμε να παρέχουμε μια ευχάριστη εμπειρία ή μια αναγκαία υπηρεσία, όπως η συμμετοχή σε ένα παιχνίδι (κάνοντας μια συνεισφορά). Τρίτον, μπορούμε να παρέχουμε τρόπους για τη δημιουργία, τη μέτρηση, και την επίδειξη εμπιστοσύνης. Τέταρτον, μπορούμε να στήσουμε διαγωνισμούς, όπως την ανάδειξη των κορυφαίων σε βαθμολογία χρηστών. Τέλος, μπορούμε να παρέχουμε περιπτώσεις ιδιοκτησίας, όπου ο χρήστης μπορεί να αισθάνεται μέρος του συστήματος, και ως εκ τούτου είναι υποχρεωμένος να «καλλιεργήσει» αυτό το μέρος. Αυτά τα προγράμματα ενσωματώνονται με φυσικό τρόπο στον εθελοντισμό, αλλά μπορούν επίσης να λειτουργήσουν καλά και για άλλες λύσεις στρατολόγησης. Για παράδειγμα, αφού ζητήσουμε από ένα σύνολο χρηστών να συνεισφέρουν, μπορούμε ακόμα να παρέχουμε άμεση ικανοποίηση, απολαυστική εμπειρία, φήμη, και πολλά άλλα, για να μεγιστοποιήσουμε τη συμμετοχή των χρηστών. Τέλος, πρέπει να σημειωθεί ότι τα ανεπτυγμένα crowdsourcing συστήματα συχνά χρησιμοποιούν ένα μίγμα των μεθόδων στρατολόγησης.

Τι συνεισφορές μπορούν να κάνουν οι χρήστες; Σε πολλά crowdsourcing συστήματα τα είδη των συνεισφορών που μπορούν να προσφέρουν οι χρήστες είναι κάπως περιορισμένα. Για παράδειγμα, για να γίνει αξιολόγηση, οι χρήστες εκτιμούν, βαθμολογούν ή επισημαίνουν. Για να γίνει κοινοποίηση, οι χρήστες προσθέτουν στοιχεία σε μία κεντρική Ιστοσελίδα. Για να γίνει δικτύωση, οι χρήστες συνδέονται με άλλους χρήστες. Σε πιο πολύπλοκα crowdsourcing συστήματα ωστόσο, οι χρήστες μπορούν να υλοποιήσουν ένα πολύ ευρύτερο φάσμα εισφορών. Η πρόκληση είναι να καθοριστεί το εύρος των πιθανών συνεισφορών και να σχεδιαστεί το σύστημα, έτσι ώστε να μπορεί να συγκεντρώσει ένα πλήθος αυτών των συνεισφορών. Για την επίτευξη του στόχου αυτού, θα πρέπει να εξεταστούν τέσσερις σημαντικοί παράγοντες. Πρώτον, γνωστικά πόσο απαιτητικές είναι οι συνεισφορές; Δεύτερον, ποιος θα πρέπει να είναι ο αντίκτυπος της συνεισφοράς; Τρίτον, τι θα συμβαίνει με τις μηχανικές συνεισφορές; Εάν ένα crowdsourcing σύστημα χρησιμοποιεί έναν αλγόριθμο για μια εργασία, τότε θέλουμε ανθρώπους να κάνουν συνεισφορές, οι οποίες γίνονται εύκολα από τους ανθρώπους, αλλά δύσκολα από τις μηχανές.

Πώς θα συνδυαστούν οι συνεισφορές των χρηστών; Πολλά crowdsourcing συστήματα δεν συνδυάζουν τις συνεισφορές, ή το κάνουν με ένα χαλαρό τρόπο. Για παράδειγμα, τα σημερινά συστήματα αξιολόγησης δε συνδυάζουν σχόλια, αλλά συνδυάζουν αριθμητικές αξιολογήσεις χρησιμοποιώντας σχετικά απλούς τύπους. Τα συστήματα δικτύωσης απλά συνδέουν τις συνεισφορές για να σχηματίσουν ένα γράφο κοινωνικού δικτύου. Τα πιο πολύπλοκα crowdsourcing συστήματα, ωστόσο, όπως εκείνα που κατασκευάζουν λογισμικό, KBS, συστήματα και παιχνίδια, συνδυάζουν τις συνεισφορές πιο σφιχτά. Ο ακριβής τρόπος με τον οποίο συμβαίνει αυτό εξαρτάται από την εφαρμογή.

Πώς θα αξιολογηθούν οι χρήστες και οι συνεισφορές; Τα crowdsourcing συστήματα συχνά πρέπει να διαχειριστούν κακόβουλους χρήστες. Για να γίνει αυτό, χρησιμοποιείται ένας συνδυασμός τεχνικών που μπλοκάρουν, εντοπίζουν και αποτρέπουν. Ανεξάρτητα από το πόσο καλά διαχειριζόμαστε τους κακόβουλους χρήστες, οι κακόβουλες συνεισφορές συχνά εξακολουθούν να διαρρέουν στο σύστημα. Στην περίπτωση αυτή, το σύστημα πρέπει να βρει έναν τρόπο να τις αναιρέσει. Εάν οι συνεισφορές είναι καλά ριζωμένες στο σύστημα, η αναιρέση αυτών μπορεί να είναι πολύ δύσκολη. Στο άλλο άκρο του φάσματος των χρηστών, πολλά συστήματα crowdsourcing προσδιορίζουν και μοχλεύουν τους επιρρεπείς χρήστες, χρησιμοποιώντας τόσο αυτόματες όσο και χειροκίνητες τεχνικές. Κατά συνέπεια, ένα μέρος της εργασίας εξετάζει πώς θα εντοπίζονται αυτόματα αυτοί οι χρήστες.

1.1.6 Γεωγραφική Πληροφορία

Σημαντικός είναι ο τρόπος που προέκυψε η χαρτογράφηση για το κοινό ως μέρος των τεχνολογιών Web 2.0. Η ανάπτυξη και χρήση τεχνολογιών σε σχέση με τη γεωγραφική πληροφορία είναι καθοριστική για την ανάπτυξη αυτού του τομέα.

Η πρόσβαση στο ίντερνετ είναι τεράστια μέσω του διαδικτύου ή του ηλεκτρονικού ταχυδρομείου. Από την άποψη του διαδικτύου, η πρόσβαση επιτυγχάνεται κυρίως μέσω γραφικών διεπαφών χρηστών (graphical user interfaces - GUIs) με τη μορφή

της «περιήγησης» η οποία επιτρέπει στους χρήστες να ελέγχουν γραφικά τις λειτουργίες στο περιβάλλον του διαδικτύου με τον ίδιο τρόπο που ελέγχει τη σύγχρονη πρόσβαση στην επιφάνεια εργασίας. Ήταν αναμενόμενο ότι οι χωρικές σχέσεις θα γίνονταν πιο σημαντικές με την άνοδο των GUIs και παράλληλα με τα διαγράμματα και τις εικόνες καθώς οι χάρτες αντιπροσωπεύουν μία από τις πιο δημοφιλείς εικονικές μορφές στο διαδίκτυο.

Η θεώρηση των σχέσεων ως δίκτυα είναι μία παλιά ιδέα η οποία αποτέλεσε επείγουσα ανάγκη σε μια εποχή που οι αλληλεπιδράσεις ολοένα και αυξάνονται καθώς οι νέες τεχνολογίες μας επιτρέπουν να επικοινωνούμε με τρόπους που ήταν αδύνατο παλιότερα. Η επιστήμη των δικτύων ίσως μας έχει βοηθήσει να διατυπώσουμε τις άμεσες σχέσεις με έναν πολύ πιο συνεκτικό τρόπο αλλά ακόμα περιορίζεται στο ότι το βασικό εμπόδιο για την κατανόηση του καλειδοσκοπίου των κοινωνικών σχέσεων είναι ότι η ικανότητά μας να αντιμετωπίζουμε δίκτυα περιορίζεται σε αμφίδρομες σχέσεις: πολλές από αυτές βρίσκονται σε μεγαλύτερες ομάδες αλλά παρόλα αυτά γίνονται πάντα αντιληπτές ως διπλή κυκλοφορία.

Οι χάρτες που κατασκευάζονται από κάτω προς τα πάνω είναι πιθανό να είναι τόσο καλοί, αν όχι καλύτεροι από οτιδήποτε μπορεί να παραχθεί από πάνω προς τα κάτω, όσο είναι πιθανό να είναι περισσότερο κατατοπιστικοί και να υπόκεινται σε συνεχείς βελτιώσεις ακολουθώντας το παράδειγμα και τη λογική των διαδικτυακών πηγών όπως το Wikipedia. Η υπόσχεση αυτών των τεχνολογιών θα γίνει αντιληπτή μόνο όταν προκύψουν καλύτεροι τρόποι για την καταγραφή πληροφοριών εντοπισμού καθώς η τρέχουσα γενιά των εξατομικευμένων GPSs είναι πρωτόγονη. Παρόλα αυτά οι crowdsourcing χάρτες βρίσκονται στα πρώτα τους βήματα και όταν θα αρχίσουν να απογειώνονται (και να περιέχουν πολλά περισσότερα για τα τοπικά περιβάλλοντα), τότε η οπτική της αναζήτησης στα πρότυπα της συσχέτισης που διέπουν τα κοινωνικά δίκτυα που δημιουργήθηκαν σε αυτά τα περιβάλλοντα μπορούν να ξεκινήσουν στα σοβαρά. Πρόκειται για έναν τομέα που υπόσχεται να ενημερώσει την κοινωνική επιστήμη και την κοινωνική δράση στο κοντινό μέλλον με τρόπους που προς το παρόν προβλέπουμε μετά βίας.

1.1.7 Κυβερνήσεις και Μη Κερδοσκοπική Χρήση

Το crowdsourcing προσφέρει επιπλέον τη δυνατότητα να αποτελέσει ένα μηχανισμό επίλυσης προβλημάτων για κυβερνήσεις και μη κερδοσκοπική χρήση. Η διαφορά σε σχέση με την αξιοποίηση του crowdsourcing για επιχειρηματικούς σκοπούς είναι πως ο αρχικός δημιουργός της διαδικασίας επίλυσης του προβλήματος (κυβέρνηση, μη κερδοσκοπικοί οργανισμοί) και οι συμμετέχοντες στη διαδικασία (πολίτες), έχουν αμοιβαία και κοινά οφέλη. Για παράδειγμα, ένα έργο crowdsourcing για την καταγραφή των παράνομων διαφημιστικών πινακίδων, βοηθά την κυβέρνηση να εντοπίσει το πρόβλημα άμεσα, με μικρό κόστος και να το λύσει. Κατά συνέπεια, οι πολίτες έχουν συνεισφέρει τη συλλογική εθελοντική εργασία τους και ωφελούνται οι ίδιοι βελτιώνοντας την ποιότητα ζωής τους, καθώς βοηθούν στην επίλυση του προβλήματος και στην αποφυγή των τροχαίων ατυχημάτων.

Αντίστοιχα, η εθελοντική συνεισφορά και τα εργαλεία του crowdsourcing μπορούν να προσφέρουν άμεσες λύσεις σε περιπτώσεις κρίσεων και φυσικών καταστροφών, σε

παγκόσμιο επίπεδο. Στην περίπτωση του σεισμού της Αϊτής στις 12 Ιανουαρίου 2010, τα σωστικά συνεργεία άλλων χωρών, αντιμετώπισαν δύο πρακτικά και κρίσιμα προβλήματα. Αφενός δεν υπήρχε διαθέσιμο σε ψηφιακή μορφή το οδικό δίκτυο της Αϊτής, αφετέρου δεν υπήρχε τρόπος ενημέρωσης των σωστικών συνεργείων και του γενικού πληθυσμού για τις περιοχές με πόσιμο νερό, στέγη, φαγητό. Τη λύση έδωσαν εθελοντές από όλο τον κόσμο, μέσω της κοινότητας του OpenStreetMap.

Γενικότερα, η ανάπτυξη των συνεργατικών υπηρεσιών στο Διαδίκτυο, η κουλτούρα της εθελοντικής συμμετοχής, τα κοινωνικά δίκτυα και τα έργα ανοικτού περιεχομένου και δεδομένων, οδηγούν σε μια νέα συλλογική πραγματικότητα στο Διαδίκτυο. Εθελοντές, αναλαμβάνουν συλλογικά να εκτελέσουν ένα έργο ή να προσφέρουν μια υπηρεσία, μοιράζοντας μεταξύ τους την προσπάθεια και προσφέροντας ελεύθερα τα αποτελέσματά της. Δηλαδή το πλήθος μπορεί να προσφέρει τη συλλογική εμπειρία και ευφυΐα για το κοινό καλό. Αυτή η παγκόσμια δυναμική συμμετοχής και αλληλοβοήθειας, είναι ευκαιρία για να παρακαμφθούν αργοκίνητες και πολύπλοκες γραφειοκρατικές διαδικασίες ή/και να αντιμετωπιστεί η έλλειψη πόρων, ώστε να δοθούν άμεσες λύσεις σε σημαντικά προβλήματα, ενισχύοντας παράλληλα την αλληλεγγύη και την κοινωνική συνοχή.

1.2 Βασικές Αρχές

Η αρχή του crowdsourcing είναι ότι τα περισσότερα κεφάλια είναι καλύτερα από ένα. Η αναζήτηση ενός μεγάλου πλήθους ανθρώπων για ιδέες, δεξιότητες, ή συμμετοχή, καθιστά την ποιότητα του περιεχομένου και την αναπαραγωγή ιδεών ανώτερο.

Η πραγματοποίηση μιας επιτυχούς μετάβασης στην καινοτομία απαιτεί ορισμένες αλλαγές και υιοθέτηση νέας τεχνολογίας, αλλά όταν γίνει σωστά οι ανταμοιβή μπορεί να είναι τεράστια.

Υπάρχουν ορισμένες αναδυόμενες βασικές αρχές που οδηγούν σε ένα επιτυχημένο crowdsourcing. Επιφανειακά, το crowdsourcing φαίνεται εύκολο – ανακοινώνει κάποιος μία ερώτηση στο κοινό, παίρνει τις ιδέες του και του επιτρέπει να ψηφίσουν την καλύτερη. Ωστόσο για να θεωρηθεί ένα crowdsourcing σύστημα επιτυχημένο θα πρέπει να δοθεί προσοχή στις παρακάτω βασικές αρχές:

- Σωστός σκοπός. Το πλήθος θα πρέπει να κληθεί να δώσει ιδέες για το αντικείμενο του crowdsourcing που πρόκειται να υλοποιηθεί.
- Σωστή ενημέρωση του πλήθους. Το πλήθος θα πρέπει να ενημερωθεί για το αντικείμενο, τις απαιτήσεις και τις φιλοδοξίες του crowdsourcing. Πρέπει να δοθούν σαφείς οδηγίες προς το πλήθος ως προς το τι αποζητά ο δημιουργός του crowdsourcing από αυτό. Το πλήθος μπορεί να βοηθήσει απαντώντας σε ερωτήσεις οι οποίες θα είναι ενθαρρυντικές και συγκεκριμένες. Εάν τα ερωτήματα είναι ασαφή, τότε και τα αποτελέσματα θα είναι ασαφή. Το πλήθος θα πρέπει να έχει το χώρο για να γίνει δημιουργικό, ενώ παράλληλα τα ερωτήματα θα πρέπει να οδηγούν προς την επιθυμητή κατεύθυνση.
- Επιλογή του σωστού πλήθους. Το πλήθος θα πρέπει να ποικίλει και να έχει τα κατάλληλα προσόντα. Το crowdsourcing βασίζεται στην παραδοχή ότι οι πιο ακριβείς απαντήσεις προέρχονται από ένα πλήθος αρκετά διαφορετικό ώστε να

δοθούν απαντήσεις από διαφορετικές οπτικές γωνίες χωρίς την ύπαρξη εγγενούς προκατάληψης καλοθελητών.

- Κατάλληλα κίνητρα. Η δόξα, ο αλτρουισμός και η φιλοδοξία τροφοδοτούν τις ανάγκες του πλήθους. Αποδεικνύεται ότι τα κατάλληλα κίνητρα προσελκύουν και διατηρούν υψηλή τη συμμετοχή. Η χρηματική βράβευση του πλήθους αποτελεί ένα από τα κίνητρα. Ωστόσο υπάρχουν και άλλες μορφές κινήτρων όπως η ανάπτυξη συνεταιρικής σχέσης, η επιδότηση μιας ιδέας, η παρουσίαση σε μία επιχειρηματική εκδήλωση ή ακόμα και στα μέσα ενημέρωσης.
- Κατάλληλο μοντέλο. Τα μοντέλα που χρησιμοποιούνται στο crowdsourcing περιλαμβάνουν την αξιοποίηση της σοφίας του πλήθους, την παραγωγή του πλήθους και τη χρηματοδότηση του πλήθους. Ανάλογα με τις απαιτήσεις της εφαρμογής – ιδέες, λύση σε ένα πρόβλημα ή χρηματοδότηση – επιλέγεται και το κατάλληλο μοντέλο.
- Κατάλληλη προώθηση. Είναι απαραίτητη η σωστή διάδοση της εφαρμογής στον κόσμο ώστε να έχει την αντίστοιχη απήχηση και καλύτερα αποτελέσματα. Ορισμένοι τρόποι διαφήμισης είναι οι ιστοσελίδες, προωθητικά email, blogs, social media κλπ.
- Κατάλληλη διαχείριση. Απαραίτητο στοιχείο της κατάλληλης διαχείρισης αποτελεί η φροντίδα για συμμετοχή του πλήθους. Η επιτυχημένη διαχείριση κρίνεται στις συναλλαγές μεταξύ του διαχειριστή και του πλήθους.
- Κατάλληλη τεχνολογία. Ακόμη και αν όλα τα παραπάνω γίνουν σωστά, χωρίς τη σωστή crowdsourcing τεχνολογία η καινοτομία που σχεδιάζεται μπορεί να ναυαγήσει. Ανάλογα με τη το είδος της εφαρμογής επιλέγεται και η κατάλληλη τεχνολογία που ταιριάζει καλύτερα στα δεδομένα της εφαρμογής.

1.3 Συσσκευές

Ο όρος crowdsourcing εμφανίστηκε σχετικά πρόσφατα και αυτό οφείλεται κυρίως στο γεγονός πως εκείνη την περίοδο εισήχθησαν δημοφιλείς Web 2.0 υπηρεσίες, αναπτύχθηκαν τα κοινωνικά δίκτυα και η τεχνολογική πρόοδος στον τομέα των κινητών συσκευών επέτρεψε στους χρήστες τους να έχουν μαζική πρόσβαση στο Διαδίκτυο από τα smartphones τους.

Η υψηλή διείσδυση των κινητών συσκευών είναι ένας δείκτης που δείχνει τις επιπτώσεις τους στην καθημερινή ζωή μας, η οποία μαρτυρείται μέσω των διάφορων υπηρεσιών που παρέχονται μέσω αυτών. Οι περισσότερες δοκιμές σήμερα απαιτούν πρόσθετες εγκαταστάσεις υλικού από τη μεριά του χρήστη, προκειμένου να συλλέξει στοιχεία και να τα χρησιμοποιήσει για βελτιστοποίηση των αποτελεσμάτων. Ωστόσο, παρόμοιες πληροφορίες μπορούν να αποκτηθούν από τους χρήστες και σε πολλές περιπτώσεις η ποιότητα των πληροφοριών μπορεί να είναι συγκρίσιμη ή και καλύτερη από τις εφαρμοζόμενες παθητικές μεθόδους σήμερα. Πιστεύεται ότι κινητές συσκευές μπορούν να παίξουν καθοριστικό ρόλο στην υποβοηθούμενη από το χρήστη απόκτηση και πρόβλεψη δεδομένων, καθώς και την ενδυνάμωση των προσεγγίσεων που ασχολούνται με what-if προσομοιώσεις κλπ. Με αυτή την ιδιότητα οι crowdsourcing πληροφορίες μέσω της χρήσης των κινητών συσκευών μπορεί να χρησιμεύσει ως μία τεχνολογία πολλών δυνατοτήτων.

1.4 Παραδείγματα Εφαρμογών

Κατά την τελευταία δεκαετία, εμφανίστηκαν πολλά συστήματα crowdsourcing στο παγκόσμιο διαδίκτυο. Μερικά τέτοια παραδείγματα είναι τα ακόλουθα:

Το Amazon Mechanical Turk είναι μία διαδικτυακή αγορά η οποία δίνει την ευκαιρία σε προγραμματιστές να συντονίσουν την ανθρώπινη ευφυΐα ώστε να λύσουν προβλήματα τα οποία δεν μπορούν να λυθούν από υπολογιστή.

Το «2009 DARPA Network challenge» ήταν ένας διαγωνισμός που έγινε στις ΗΠΑ. Ο διαγωνισμός χορηγήθηκε από το Defense Advanced Research Projects Agency (DARPA). Οι ομάδες που συμμετείχαν στο διαγωνισμό έπρεπε να εντοπίσουν 10 κόκκινα μετεωρολογικά μπαλόνια τα οποία τοποθετήθηκαν σε 10 σημεία στις ΗΠΑ. Οι ομάδες χρησιμοποίησαν τα κοινωνικά δίκτυα για να εντοπίσουν τα μπαλόνια.

Το Johnny Cash Project είναι ένα συλλογικό έργο τέχνης. Οποιοσδήποτε μπορεί να λάβει μέρος σχεδιάζοντας ένα πλαίσιο από ένα video clip του Johnny Cash.

Το Wikipedia δημιουργεί μια λεκτική KB. Αλλά επίσης δημιουργεί μια δομημένη γνωσιακή βάση και φιλοξενεί πολλά forums κοινοποίησης γνώσης όπως ομάδες συζήτησης.

Το YouTube επιτρέπει στους χρήστες να κοινοποιήσουν αλλά και να αξιολογήσουν βίντεο.

Το Awardesigns.com είναι μία ισπανική crowdsourcing πλατφόρμα η οποία ξεκίνησε το 2011. Περισσότεροι από 2000 σχεδιαστές και συγγραφείς από τουλάχιστον 15 χώρες βοηθούν τους επιχειρηματίες και τις μικρές επιχειρήσεις να επιλέξουν λογότυπο, web design, εικονογράφηση και άλλα είδη γραφικού σχεδιασμού.

Το California Digital Newspaper Collection εφάρμοσε crowdsourcing για τη διόρθωση κειμένου από τα ψηφιοποιημένα ιστορικά αρχεία εφημερίδων από το 1846.

Το CloudCrowd, το οποίο ιδρύθηκε το 2009, είναι μια διαδικτυακή πλατφόρμα εργασίας που σπάει το έργο σε ροές εργασιών, και καθιστά τις μεμονωμένες εργασίες διαθέσιμες στους crowdsourced εργαζόμενους.

Το Crowdsourcing.org ιδρύθηκε το 2010 και είναι μια ουδέτερη οργάνωση που αφιερώνεται αποκλειστικά στο crowdsourcing και το crowdfunding. Ως μία από τις πλέον αξιόπιστες αρχές του crowdsourcing χώρου, το Crowdsourcing.org αναγνωρίζεται παγκοσμίως για την crowdsourcing και crowdfunding τεχνογνωσία και την αμερόληπτη ηγετική σκέψη.

Το LEGO Design byME ήταν μια υπηρεσία που συνδέεται με την κατασκευή παιχνιδιών Lego. Ξεκίνησε το 2005 με την ονομασία Lego Factory. Η υπηρεσία επέτρεψε στους ανθρώπους να σχεδιάσουν τα δικά τους μοντέλα Lego με τη χρήση ενός προγράμματος ηλεκτρονικού υπολογιστή, στη συνέχεια να τα φορτώσουν στην

ιστοσελίδα της Lego, να σχεδιάσουν το δικό τους κουτί, και να τα παραγγείλουν για πραγματική αγορά.

Με το duolingo μπορεί να μάθει κάποιος μία ξένη γλώσσα δωρεάν βοηθώντας στη μετάφραση κειμένου.

Το παιχνίδι με ESP ξεκίνησε το 2003 και βάζει το πλήθος να δώσει ονόματα εικόνες ως αποτέλεσμα ενός παιχνιδιού. Οι ετικέτες των εικόνων μπορούν να χρησιμοποιηθούν για να βελτιωθεί η αναζήτηση εικόνων στον Παγκόσμιο Ιστό. Αυτό το παιχνίδι οδήγησε στην έννοια της ψυχαγωγίας με ένα σκοπό.

Η канаδική ομάδα εξόρυξης χρυσού Goldcorp δημοσίευσε 400 megabyte δεδομένων των γεωλογικών ερευνών στο Red Lake του Οντάριο μέσω του Διαδικτύου. Πρόσφεραν ένα βραβείο 575.000 δολαρίων σε όποιον θα μπορούσε να αναλύσει τα δεδομένα και να προτείνει τοποθεσίες όπου θα μπορούσε να βρεθεί χρυσός. Η εταιρεία ισχυρίζεται ότι ο διαγωνισμός παράγαγε 110 στόχους, πάνω από το 80% των οποίων αποδείχθηκαν παραγωγικοί, παράγοντας 8 εκατομμύρια ουγκιές χρυσού, αξίας άνω των 3 δισεκατομμυρίων δολαρίων. Το βραβείο κέρδισε μία μικρή ομάδα συμβούλων στο Περθ, στη Δυτική Αυστραλία.

Το OpenStreetMap είναι άλλο ένα crowdsourcing παράδειγμα. Πρόκειται για έναν ελεύθερο επεξεργάσιμο χάρτη του κόσμου, ο οποίος είχε πάνω από 100.000 συμμετέχοντες έως τα μέσα του 2009. Η δημιουργία και συντήρηση των γεωχωρικών δεδομένων είναι μία κοπιαστική εργασία η οποία είναι ακριβή χρησιμοποιώντας τις παραδοσιακές προσεγγίσεις.

Η Pepsi ξεκίνησε μια εκστρατεία μάρκετινγκ στις αρχές του 2007 που επέτρεψε στους καταναλωτές να σχεδιάσουν την εμφάνιση του κουτιού της Pepsi. Οι νικητές θα έπαιρναν ένα βραβείο των 10.000 δολαρίων, και το έργο τους θα εμφανιζόταν σε 500 εκατομμύρια κουτιά Pepsi στις Ηνωμένες Πολιτείες.

Το reCAPTCHA έχει στόχο να βοηθήσει στην ψηφιοποίηση κειμένου βιβλίων, ενώ προστατεύει τις ιστοσελίδες από όσους προσπαθούν να αποκτήσουν πρόσβαση σε απαγορευμένες περιοχές. Παρουσιάζονται εικόνες του βιβλίου στο πλήθος, και του ζητείται να συμπληρώσει το αντίστοιχο κείμενο. Είκοσι χρόνια των New York Times έχουν ήδη ψηφιοποιηθεί.

Το Student of Fortune είναι μια διαδικτυακή υπηρεσία που επιτρέπει στους μαθητές να υποβάλουν ερωτήματα που αφορούν στις εργασίες στους δασκάλους για να απαντήσουν μέσω μιας υπηρεσίας διδασκαλίας για συγκεκριμένο κόστος.

Αυτά είναι μόνο μερικά από τα crowdsourcing παραδείγματα που υπάρχουν. Το crowdsourcing είναι πλέον πολύ διαδεδομένο στο ευρύ κοινό και υπάρχει τεράστιος αριθμός crowdsourcing εφαρμογών, άλλες μεγαλύτερης και άλλες μικρότερης εμβέλειας, οι οποίες μπορεί να αφορούν στο οτιδήποτε.

2

ΑΠΑΙΤΗΣΕΙΣ ΑΠΟ ΕΝΑ ΠΕΔΙΟ ΕΦΑΡΜΟΓΗΣ

2.1 Καταγραφή Προβλημάτων σε δημόσιες Υποδομές

Η παρούσα μεταπτυχιακή εργασία καταπιάνεται με το θέμα της συλλογής αλλά και της αξιοποίησης καταγεγραμμένων προβλημάτων που αφορούν σε δημόσιες υποδομές. Το ζητούμενο της εργασίας ήταν να υλοποιηθεί ένα σύστημα καταγραφής προβλημάτων στο οποίο θα μπορούν να λάβουν μέρος τα μέλη μίας κοινότητας, με σκοπό την ταχύτερη και συστηματική διευθέτησή τους.

Οι δημόσιες υποδομές αποτελούνται από όλα εκείνα τα έργα που υλοποιούν οι αρμόδιοι δημόσιοι φορείς προς όφελος της κοινωνίας. Μερικά παραδείγματα δημόσιων υποδομών είναι το οδόστρωμα, οι στάσεις λεωφορείων, οι ταμπέλες, τα παγκάκια, οι τηλεφωνικοί θάλαμοι και πολλά άλλα.

Η γενική ιδέα είναι να έχει τη δυνατότητα ο καθένας να καταγράψει επιτόπου σε πραγματικό χρόνο προβλήματα που συναντά στην πόλη του και σχετίζονται με την αλλοίωση δημοσίων υποδομών, χρησιμοποιώντας ένα «έξυπνο» τηλέφωνο. Πιο συγκεκριμένα, θα καταγράφονται από το πλήθος όλες οι απαραίτητες πληροφορίες για την περιγραφή και τον προσδιορισμό προβλημάτων σε υποδομές δημόσιας περιουσίας, με στόχο τη μετέπειτα εκμετάλλευση των πληροφοριών αυτών για την εξαγωγή νέων δεδομένων και τη δημιουργία χαρτών, που θα προσφέρουν μία γενική εικόνα για το σύνολο των προβλημάτων αλλά και για τα προβλήματα μεμονωμένα.

Η καταγραφή προβλημάτων σε δημόσιες υποδομές μπορεί να καλύπτει ένα ευρύ φάσμα υπαρκτών προβλημάτων, ωστόσο ο στόχος της εργασίας δεν ήταν η διερεύνηση όλων των πιθανών υπαρχόντων προβλημάτων, αλλά η προσέγγιση και επίδειξη της συγκεκριμένης οικογένειας εφαρμογών και τεχνολογιών που γενικά χαρακτηρίζεται ως «crowdsourcing». Για το λόγο αυτό, το η καταγραφή των προβλημάτων περιορίζεται σε ενδεικτικούς τύπους προβλημάτων που θα μπορεί να καταγράψει το πλήθος. Στα παραδείγματα δημιουργήθηκαν εικονικά δεδομένα πλήθους που σχετίζονται με τις παρακάτω υποδομές:

- ✓ το οδόστρωμα
- ✓ το πεζοδρόμιο
- ✓ τα φανάρια
- ✓ τα δέντρα
- ✓ τις ταμπέλες

- ✓ τα φρεάτια
- ✓ τους αγωγούς
- ✓ τις λάμπες των στύλων
- ✓ τα παγκάκια
- ✓ τους τηλεφωνικούς θαλάμους
- ✓ τις στάσεις των λεωφορείων
- ✓ τα γκράφιτι για την καταστροφή δημόσιας περιουσίας
- ✓ η παράνομη αφισοκόλληση για την αλλοίωση της εικόνας δημόσιων κτιρίων και εγκαταστάσεων
- ✓ τα εγκαταλελειμμένα οχήματα για παρεμπόδιση σε δημόσιους χώρους

Η αρχιτεκτονική της εφαρμογής επιτρέπει την πλήρη παραμετροποίηση των προβλημάτων που θα καταγράφονται.

Για να συλλεχθεί η πληροφορία των καταγραφών με εύκολο και αποτελεσματικό τρόπο, θα πρέπει να χρησιμοποιηθούν crowdsourcing μέθοδοι, έτσι ώστε να έχει ο καθένας τη δυνατότητα να καταγράφει εύκολα και γρήγορα προβλήματα σε υποδομές δημόσιας περιουσίας. Η καταγραφή των προβλημάτων θα μπορεί να υλοποιηθεί από το πλήθος, υπό την προϋπόθεση να απαντάει σε συγκεκριμένα ερωτήματα που θα αφορούν στο εκάστοτε πρόβλημα και τα οποία θα διευκρινίζουν:

- ✓ το είδος του προβλήματος
- ✓ την τοποθεσία του προβλήματος
- ✓ την επικινδυνότητα του προβλήματος
- ✓ την ημερομηνία που συναντήθηκε το πρόβλημα

Η ακριβής και έγκαιρη αναφορά ενός προβλήματος θα είναι ουσιώδης στην περίπτωση διαχείρισης των προβλημάτων από δημόσιους φορείς. Ο δημόσιος φορέας θα πρέπει να είναι σε θέση να βλέπει και να διαχειρίζεται τα δεδομένα των καταγραφών, να γνωρίζει πόσες καταγραφές πραγματοποιήθηκαν για ένα συγκεκριμένο πρόβλημα αλλά και τον βαθμό επικινδυνότητας του προβλήματος, ώστε να είναι σε θέση να εκτιμήσει και να ιεραρχήσει τα προβλήματα με βάση την αμεσότητα της ανάγκης για επίλυση.

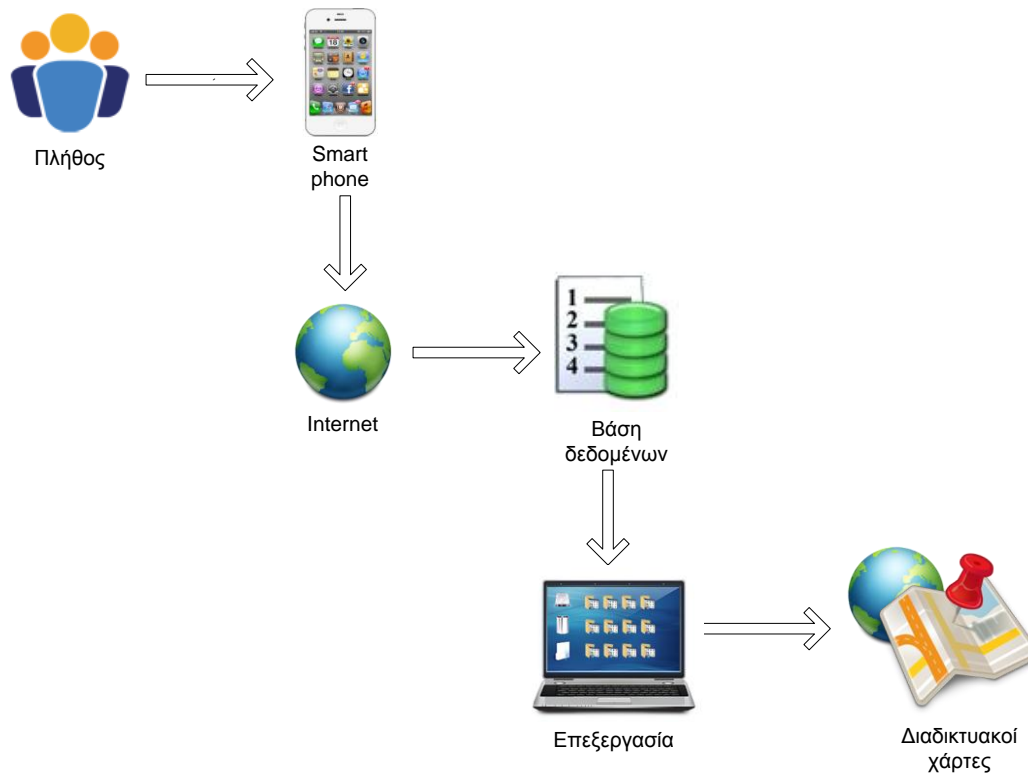
2.2 Τρόπος Εφαρμογής του Crowdsourcing

Η κεντρική ιδέα της υλοποίησης της εφαρμογής είναι να μπορεί κανείς να καταγράψει ανά πάσα ώρα και μέρος κάποιο πρόβλημα που συνάντησε και η πληροφορία να καταχωρείται σε μία βάση δεδομένων. Με τον τρόπο αυτό, οι πολίτες αντικαθιστούν το ρόλο των δημόσιων φορέων όσον αφορά στην αναζήτηση των προβλημάτων σε δημόσιες υποδομές και κινητοποιούνται προσφέροντας δωρεάν πληροφορία συμβάλλοντας έτσι στην επίλυση του προβλήματος.

Η καταγραφή των προβλημάτων σε δημόσιες περιουσίες μπορεί να πραγματοποιηθεί με διάφορους τρόπους. Στην εργασία αυτή, η καταγραφή ενός προβλήματος θα μπορεί να πραγματοποιηθεί μέσω της τυποποιημένης περιγραφής του προβλήματος με τη βοήθεια μίας crowdsourcing εφαρμογής στο κινητό τηλέφωνο.

Θα υλοποιηθεί ένα σύστημα crowdsourcing με τη βοήθεια του οποίου θα συγκεντρώνεται όλη η απαραίτητη πληροφορία για κάθε πρόβλημα. Το crowdsourcing σύστημα που θα χρησιμοποιηθεί για τη συλλογή των δεδομένων είναι η εφαρμογή Fulcrum App, η οποία περιγράφεται σε επόμενο κεφάλαιο.

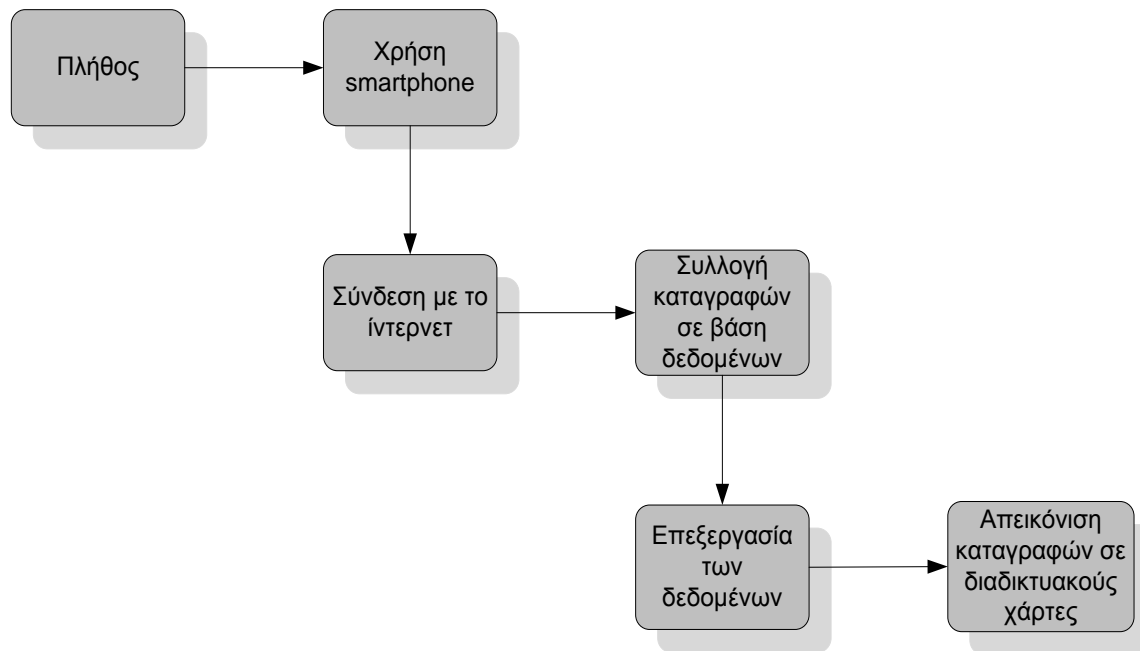
Η ροή υλοποίησης του συστήματος, η οποία αποτελεί και το σκελετό της εφαρμογής, παρουσιάζεται στο σχήμα που ακολουθεί.



Εικόνα 2.1: Στάδια υλοποίησης του συστήματος της καταγραφής προβλημάτων σε δημόσιες υποδομές

Το σχήμα των σταδίων υλοποίησης του συστήματος ερμηνεύεται συνοπτικά ως εξής:

- ✓ Αρχικά είναι απαραίτητη η διαθεσιμότητα και η συμμετοχή του πλήθους προκειμένου να καταγράψει τα προβλήματα που συναντάει.
- ✓ Εκείνος που θα καταγράψει ένα πρόβλημα θα πρέπει να διαθέτει και να χρησιμοποιεί smartphone κινητό τηλέφωνο.
- ✓ Επιπλέον θα πρέπει να συνδεθεί στο ίντερνετ μέσω του κινητού τηλεφώνου ώστε να συγχρονιστούν τα δεδομένα με την crowdsourcing εφαρμογή όταν πραγματοποιηθεί η καταγραφή.
- ✓ Πραγματοποιείται η καταγραφή του προβλήματος μέσω μίας crowdsourcing εφαρμογής στο κινητό τηλέφωνο.
- ✓ Τα δεδομένα αποθηκεύονται σε μία βάση δεδομένων και υφίστανται τις κατάλληλες επεξεργασίες.
- ✓ Τα δεδομένα των καταγραφών των προβλημάτων απεικονίζονται σε χάρτες στο διαδίκτυο.



Εικόνα 2.2: Ροή σταδίων υλοποίησης του συστήματος της καταγραφής προβλημάτων σε δημόσιες υποδομές

2.3 Δεδομένα που θα συλλέγονται και Τρόπος Συλλογής

Για τη συλλογή όλων των απαραίτητων δεδομένων θα χρησιμοποιηθεί crowdsourcing εφαρμογή με τη βοήθεια ενός κινητού τηλεφώνου. Τα δεδομένα που θα συλλέγονται σχετίζονται άμεσα με τη θεματολογία και τις απαιτήσεις της εφαρμογής.

Στην παρούσα εργασία, το θέμα της εφαρμογής όπως ήδη αναλύθηκε σε προηγούμενη ενότητα είναι η καταγραφή προβλημάτων που συναντάει ένας πολίτης σε δημόσιες υποδομές. Επομένως, τα είδη των δεδομένων που θα συλλέγονται διαμορφώνονται με βάση το θέμα της εφαρμογής και θα χωρίζονται σε δύο κατηγορίες:

- ✓ Προσωπικά δεδομένα και
- ✓ Πληροφορίες προβλήματος

Στην πρώτη κατηγορία θα καταγράφονται τα στοιχεία επικοινωνίας του ατόμου, ενώ στη δεύτερη κατηγορία όλες οι απαραίτητες πληροφορίες για τον προσδιορισμό του προβλήματος και της επικινδυνότητας αυτού, έτσι ώστε ο διαχειριστής της εφαρμογής να μπορεί να εκτιμήσει και να ιεραρχήσει τα προβλήματα ως προς τη σοβαρότητα αυτών.

Οι δύο παραπάνω βασικές κατηγορίες δεδομένων θα χωρίζονται σε πολλές υποκατηγορίες και πιο συγκεκριμένα:

- ✓ Προσωπικά δεδομένα
 - Όνομα

- Επίθετο
- Τηλέφωνο
- e-mail
- Πόλη
- ✓ Πληροφορίες προβλήματος
 - Τύπος προβλήματος
 - Συνοπτική περιγραφή προβλήματος
 - Ημερομηνία προβλήματος
 - Αξιολόγηση προβλήματος
 - Φωτογραφία

Ο χρήστης του κινητού τηλεφώνου θα μπορεί να ξεκινήσει να καταγράφει τα προβλήματα που συναντάει, διευκρινίζοντας τις παραπάνω πληροφορίες για κάθε καταγραφόμενο πρόβλημα. Με τον τρόπο αυτό θα προστίθεται στο σύστημα ένα νέο σημείο και πληροφορία για κάθε καινούρια καταγραφή.

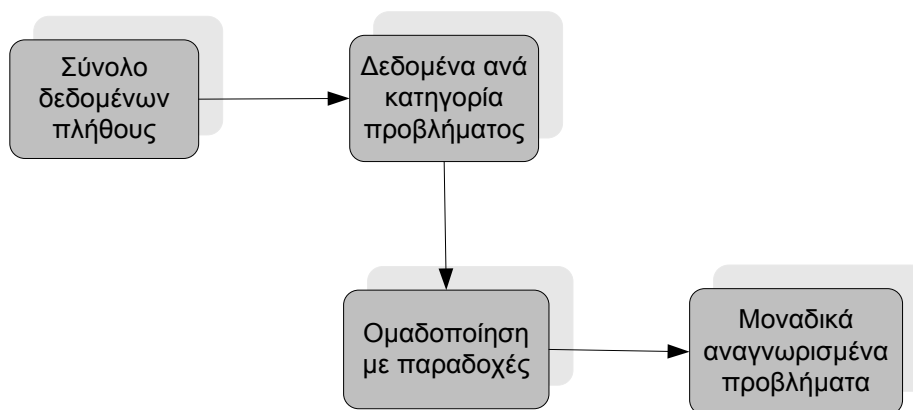
Για κάθε περίπτωση προβλήματος θα μπορούν να πραγματοποιηθούν πολλές καταγραφές. Επιπλέον, θα μπορούν να καταγραφούν πολλές και διαφορετικές περιπτώσεις προβλημάτων σε μία δεδομένη περιοχή. Εφόσον θα ολοκληρώνεται η διαδικασία των καταγραφών, τα δεδομένα θα είναι διαθέσιμα προς αποθήκευση.

2.4 Απαιτήσεις Επεξεργασίας

Η συλλογή των δεδομένων των καταγραφών των προβλημάτων σε δημόσιες υποδομές δεν τα καθιστά από μόνη της αξιοποιήσιμα. Αυτό συμβαίνει διότι με την αδιάλειπτη, άτακτη και ακανόνιστη καταγραφή προβλημάτων, συγκεντρώνεται ένας μεγάλος όγκος πληροφορίας και πολλαπλών σημείων που είναι δύσκολο να τα διαχειριστεί κάποιος και να εξάγει συμπεράσματα για τα δεδομένα.

Για κάθε πρόβλημα που υφίσταται θα μπορούν να πραγματοποιηθούν παραπάνω από μία καταγραφές, από διαφορετικά πρόσωπα η κάθε μία, με αποτέλεσμα να επικρατήσει σύγχυση κατά τη λήψη όλων των δεδομένων που θα συλλέγονται κατά τη διαδικασία των καταγραφών, λόγω της διαφορετικής θέσης της κάθε καταγραφής για το ίδιο πρόβλημα αλλά και ενδεχόμενου σφάλματος του GPS κατά τον υπολογισμό των συντεταγμένων της θέσης της καταγραφής. Επομένως κρίνεται απαραίτητη η παρέμβαση στα δεδομένα, προκειμένου να υλοποιηθεί ομαδοποίηση της μαζικής πληροφορίας σύμφωνα με το καταγραφόμενο πρόβλημα, ώστε να είναι διαθέσιμα για περαιτέρω αξιοποίηση.

Η ομαδοποίηση των δεδομένων με βάση το πρόβλημα που θα καταγράφεται κάθε φορά θα πραγματοποιείται με χρήση του αλγορίθμου ομαδοποίησης k-means. Ο αλγόριθμος k-means θα ομαδοποιεί τα σημειακά δεδομένα με βάση την κατανομή τους στο χώρο, τον αριθμό των ομάδων k στον οποίο επιθυμούμε να διαχωριστούν τα δεδομένα, και προαιρετικά τα σημεία αρχικοποίησης τα οποία είναι τόσα όσες και οι ομάδες k που επιθυμούμε να δημιουργηθούν.



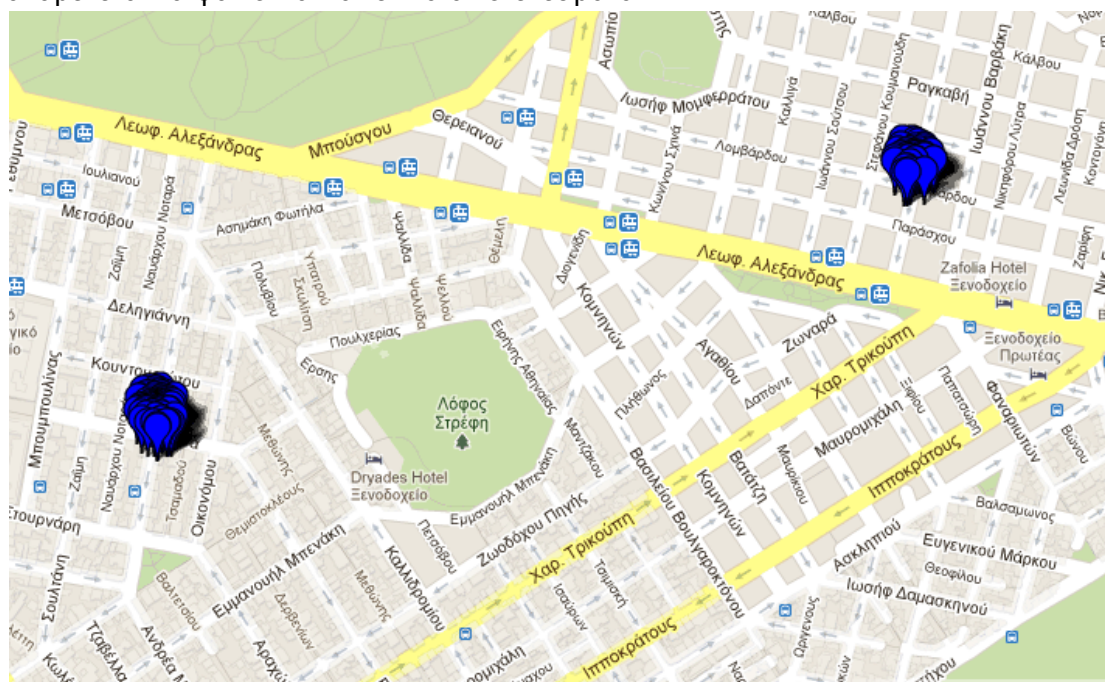
Εικόνα 2.3: Στάδια ομαδοποίησης των δεδομένων

2.5 Απαιτήσεις Εμφάνισης Αποτελεσμάτων

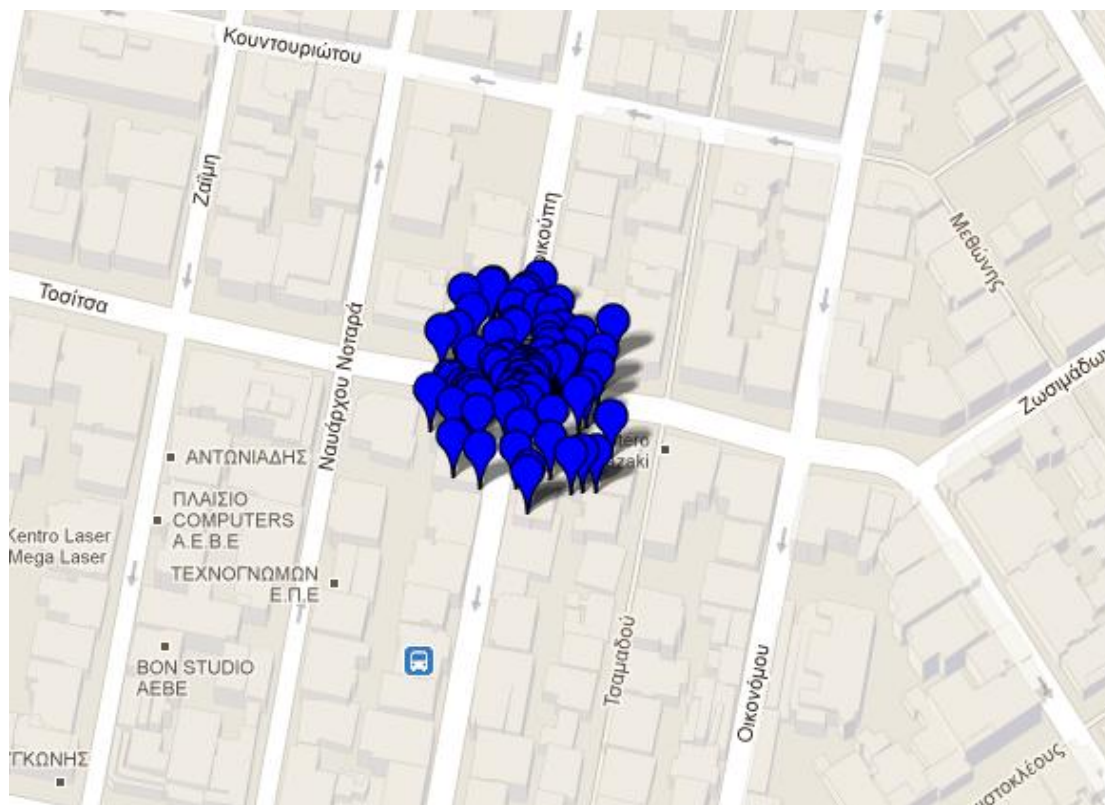
Μετά την ολοκλήρωση της συλλογής των δεδομένων των καταγραφών, τα αποτελέσματα θα απεικονίζονται σε χάρτες στο διαδίκτυο. Μέσα από τους χάρτες αυτούς θα μπορεί ο χρήστης να παρατηρήσει και να εκτιμήσει τη σοβαρότητα κάθε προβλήματος και να πάρει αποφάσεις ως προς την αντιμετώπισή του. Συγκεκριμένα θα σχηματιστούν δύο κατηγορίες χάρτων, οι οποίες είναι:

- ✓ ο χάρτης του συνόλου των δεδομένων: ο χάρτης αυτός θα παρουσιάζει το σύνολο των καταγραφών για κάθε τύπο προβλήματος.
- ✓ ο χάρτης των μοναδικών περιπτώσεων: ο χάρτης αυτός θα παρουσιάζει όλα τα μεμονωμένα σημεία που θα αντιπροσωπεύουν κάθε περίπτωση προβλήματος ανά τύπο προβλήματος.

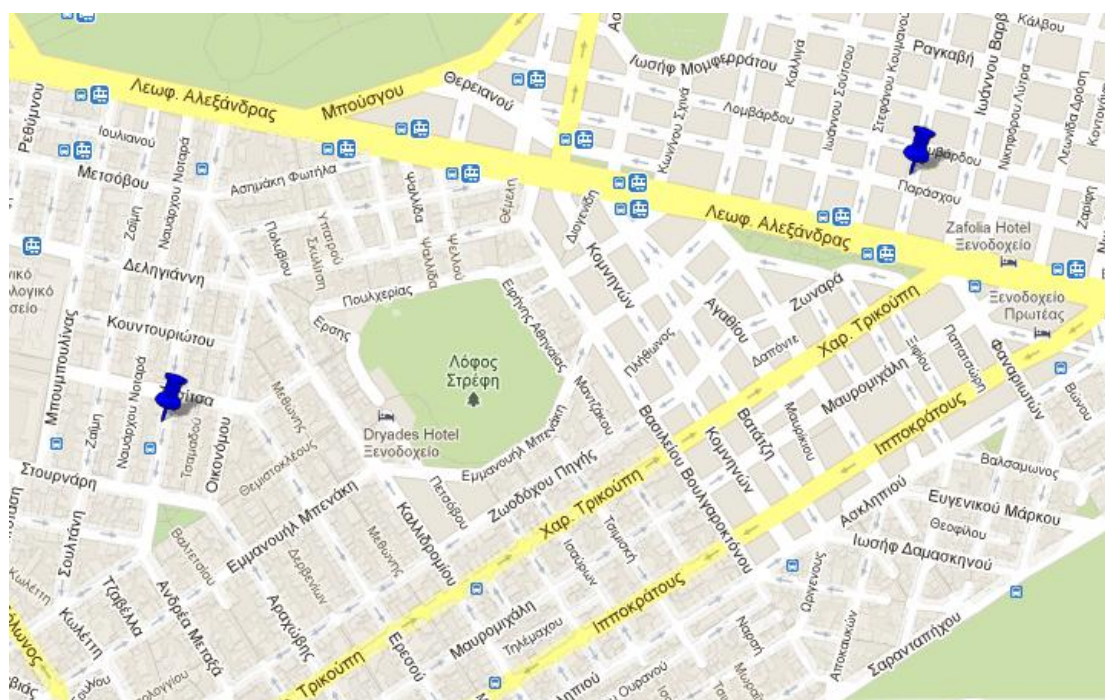
Στη συνέχεια παρουσιάζονται τρεις χάρτες οι οποίοι δείχνουν ενδεικτικά πως αναμένεται να φαίνονται τα τελικά αποτελέσματα.



Εικόνα 2.4: Απεικόνιση του συνόλου των καταγραφών για δεδομένο τύπο προβλήματος



Εικόνα 2.5: Απεικόνιση του συνόλου των καταγραφών για δεδομένη περίπτωση προβλήματος (μεγέθυνση προηγούμενου χάρτη)



Εικόνα 2.6: Απεικόνιση των μοναδικών περιπτώσεων για δεδομένο τύπο προβλήματος

Στο διαδικτυακό περιβάλλον όπου θα παρουσιάζονται οι χάρτες, θα υπάρχει η δυνατότητα επιλεκτικής απεικόνισης των καταγεγραμμένων προβλημάτων βάσει συγκεκριμένων κριτηρίων με τη χρήση κουμπιών επιλογής. Ο χρήστης θα έχει τη

δυνατότητα να επιλέγει τα δεδομένα που θα εμφανίζει ο χάρτης και αξιολογώντας τα στατιστικά στοιχεία των καταγραφών θα μπορεί να εξάγει τα αντίστοιχα συμπεράσματα.

2.6 Περιορισμοί – Παραδοχές

Για την υλοποίηση του συστήματος που περιγράφηκε στις προηγούμενες ενότητες, αποτελεί απαραίτητη προϋπόθεση η αποδοχή ορισμένων περιορισμών, η απουσία των οποίων δυσχεραίνει την έκβαση της υλοποίησης της εφαρμογής στα πλαίσια της μεταπτυχιακής εργασίας. Για το λόγο αυτό, κρίθηκε αναγκαίο να πραγματοποιηθούν οι ακόλουθες παραδοχές:

- Η συλλογή των δεδομένων των καταγραφών των προβλημάτων θα πραγματοποιηθεί με τη βοήθεια της crowdsourcing εφαρμογής Fulcrum. Η εφαρμογή αυτή διαθέτει έτοιμα εργαλεία για τη δημιουργία φόρμας συμπλήρωσης δεδομένων μέσω της οποίας θα πραγματοποιούνται οι καταγραφές από το πλήθος. Η δημιουργία εκ νέου εφαρμογής συλλογής δεδομένων δεν εντάσσεται μέσα στα πλαίσια υλοποίησης της παρούσας εργασίας.
- Η καταγραφή των προβλημάτων απαιτεί τη χρήση κινητού τηλεφώνου iOS ή Android λογισμικού, καθώς επίσης και την εγκατάσταση της εφαρμογής Fulcrum App μέσω της οποίας θα καταγράφονται τα προβλήματα. Αυτό σε μια πραγματική εφαρμογή ενός τέτοιου συστήματος θα πρέπει να αντικατασταθεί από εφαρμογή smartphone ειδικά κατασκευασμένη για το σκοπό αυτό.
- Η καταγραφή των προβλημάτων περιορίζεται στους ενδεικτικούς προκαθορισμένους τύπους προβλημάτων εκ των οποίων θα μπορεί να επιλέγει το πλήθος, ώστε να υπάρχει μία τυποποίηση στα δεδομένα και να είναι πιο άμεση η επεξεργασία τους. Ο χρήστης δεν έχει τη δυνατότητα εισαγωγής του δικού του τύπου, διότι αυτό θα απαιτούσε ένα επιπλέον στάδιο ταξινόμησης των διαφορετικών διατυπώσεων που χρησιμοποιούν διαφορετικοί χρήστες για το ίδιο πρόβλημα.
- Το ίδιο πρόβλημα δεν θα παρουσιάζεται περισσότερες από μία φορά σε απόσταση μικρότερη των 50 μέτρων, για αποφυγή της σύγχυσης που θα προκύψει από την ομαδοποίηση των δεδομένων με τον αλγόριθμο k-means στην περίπτωση αυτή. Η απόσταση των 50 μέτρων προκύπτει στην περίπτωση που δύο ομάδες σημείων δύο διαφορετικών περιπτώσεων προβλημάτων εφάπτονται. Εφόσον η μακρινότερη καταγραφή για ένα πρόβλημα πραγματοποιείται σε απόσταση 25 μέτρων, στην περίπτωση των εφάπτομενων ομάδων η ελάχιστη δυνατή απόσταση μεταξύ δύο προβλημάτων θα είναι 50 μέτρα.
- Η συλλογή των δεδομένων των καταγραφών αυτή καθεαυτή ξεφεύγει από το σκοπό της εργασίας. Για το λόγο αυτό, χρησιμοποιήθηκαν πλασματικά τυχαία δεδομένα προσομοιώνοντας ωστόσο τη δομή των δεδομένων που προέρχονται από τη συλλογή των πραγματικών δεδομένων της εφαρμογής Fulcrum App.

- Για τη δημιουργία των πλασματικών δεδομένων και πιο συγκεκριμένα των πλασματικών σημείων που θα αντιπροσωπεύουν τις καταγραφές, έγινε η παραδοχή ότι όλες οι καταγραφές που αφορούν σε μία συγκεκριμένη περίπτωση προβλήματος θα περιέχονται μέσα σε κύκλο ακτίνας 25 μέτρων.

3

Ο ΑΛΓΟΡΙΘΜΟΣ K-MEANS

3.1 Γενικά

3.1.1 Ομαδοποίηση

Συσταδοποίηση ή ομαδοποίηση (clustering) είναι η διαδικασία εκείνη κατά την οποία ένα σύνολο από «αντικείμενα», διαχωρίζονται σε ένα σύνολο από λογικές ομάδες (clusters). Η καταχώρηση αντικειμένων σε ίδια ομάδα μεταφράζεται ως ομοιότητα των αντικειμένων αυτών και αντίστροφα (αντικείμενα που ανήκουν σε διαφορετικές ομάδες είναι ανόμοια). Η ομοιότητα ή μη, μεταξύ των αντικειμένων, ουσιαστικά εξαρτάται από το συγκεκριμένο πρόβλημα και τη μορφή των «αντικειμένων». Στη βιβλιογραφία συναντάται ως ομαδοποίηση και μη επιβλεπόμενη μάθηση. Τα αντικείμενα μπορούν να αναφερθούν και αυτά με διαφορετικούς όρους: πρότυπα, διανύσματα.

Η διαδικασία της ομαδοποίησης μπορεί να αναλυθεί σε τρία βασικά χαρακτηριστικά. Προηγείται η διαδικασία σχηματισμού διανυσμάτων. Κατά την διαδικασία αυτή επιλέγονται τα k –χαρακτηριστικά ενός διανύσματος με τέτοιο τρόπο ώστε να καθιστά τη διαδικασία της ομαδοποίησης εύκολη και δυνατή. Δηλαδή τα διανύσματα σχηματίζονται με σκοπό να μεταφέρουν την μέγιστη δυνατή πληροφορία, να χρησιμοποιούν την ελάχιστη δυνατή πληροφορία αναπαράστασης $-k$ ελάχιστο δυνατό ακέραιο και η αναπαράσταση τους να τα καθιστά εύκολα στην επεξεργασία. Τα τρία βασικά χαρακτηριστικά είναι:

- Εύρεση μίας συνάρτησης εγγύτητας η οποία ποσοτικοποιεί τους όρους όμοια και ανόμοια μεταξύ των διανυσμάτων. Συνήθως η συνάρτηση αυτή αποτελεί μετρική.
- Εύρεση ενός κριτηρίου με βάση το οποίο τα διάφορα αντικείμενα θα καταχωρούνται σε διάφορες ομάδες. Το κριτήριο αυτό αποτελεί οδηγό για τον αλγόριθμο ομαδοποίησης. Ένα απλό κριτήριο, για δεδομένο μέτρο ανομοιότητας $d()$ μεταξύ διανύσματος και συνόλου διανυσμάτων, ένα διάνυσμα x_i θα καταχωρείται στην ομάδα c_j αν ισχύει: $d(x_i, c_j) \leq d(x_i, c_q)$ για κάθε άλλη ομάδα q .
- Τέλος, επιλογή ενός κατάλληλου αλγόριθμου ομαδοποίησης ο οποίος θα εμπεριέχει και θα κάνει χρήση των δύο παραπάνω εννοιών και θα έχει ως έξοδο το ζητούμενο: ένα σύνολο m ομάδων (έπεται η διαδικασία ελέγχου).

3.1.2 Αλγόριθμοι ομαδοποίησης

Υπάρχει ένα μεγάλο πλήθος από αλγόριθμους ομαδοποίησης που έχουν προταθεί και ο καθένας τους βασίζεται σε διαφορετική φιλοσοφία. Σχεδόν όλοι τους δέχονται ένα σύνολο παραμέτρων που μπορεί να είναι το πλήθος των ομάδων, διανύσματα αρχικοποίησης που απαιτούνται από τον αλγόριθμο κάποιες υποθέσεις για την πυκνότητα των διανυσμάτων στο χώρο και άλλες διάφορες παραμέτρους. Διαφοροποιώντας αυτές τις παραμέτρους προκύπτει ένα σύνολο από αλγόριθμους σε κάθε βασική κατηγορία. Υπάρχουν τρεις κατηγορίες αλγορίθμων ομαδοποίησης:

- K-means
- Ιεραρχικοί
- Ανταγωνιστικής Μάθησης

Στην εργασία αυτή χρησιμοποιείται ο αλγόριθμος **kmeans** και αναλύεται στη συνέχεια.

3.1.3 Ιστορία

Ο όρος «k-means» χρησιμοποιήθηκε πρώτα από τον James MacQueen το 1967 αν και η ιδέα χρονολογείται στον Hugo Steinhaus το 1957. Ο βασικός αλγόριθμος αρχικά προτάθηκε από τον Stuart Lloyd το 1957 ως μια τεχνική για παλμοκωδική διαμόρφωση (pulse-code modulation - PCM)², παρόλο που δεν δημοσιεύθηκε έξω από τα εργαστήρια Μπελ μέχρι το 1982. Το 1965 ο E.W.Forgy δημοσίευσε ουσιαστικά την ίδια μέθοδο, γι ' αυτό μερικές φορές αυτή αναφέρεται ως Lloyd-Forgy επίσης. Μια πιο αποδοτική έκδοση προτάθηκε και δημοσιεύθηκε σε Fortran από τους Hartigan και Wong το 1975-1979.

3.1.4 K-means

Η **ομαδοποίηση k-means** είναι ένας παλιός αλλά δημοφιλής αλγόριθμος ομαδοποίησης γνωστός για την ταχύτητα και την απλότητά του. Είναι ένας πολύ γνωστός αλγόριθμος γεωμετρικής ομαδοποίησης που βασίζεται στο έργο του Lloyd το 1982. Λαμβάνοντας υπόψη ένα σύνολο n σημειακών δεδομένων, ο αλγόριθμος χρησιμοποιεί μια τοπική προσέγγιση αναζήτησης για να ταξινομήσει τα σημεία σε k ομάδες, επιλέγοντας αυθαίρετα ένα σύνολο από k αρχικά κέντρα ομαδοποίησης. Κάθε σημείο τότε αντιστοιχίζεται στο πλησιέστερο σε αυτό κέντρο, και τα κέντρα επαναυπολογίζονται ως κέντρα μάζας των αντιστοιχισμένων σημείων τους. Αυτό επαναλαμβάνεται έως ότου η διαδικασία σταθεροποιηθεί. Μπορεί να δείχθει ότι δεν μπορεί να συμβεί δύο φορές η ίδια ταξινόμηση κατά τη διάρκεια του αλγορίθμου, και έτσι ο αλγόριθμος είναι σίγουρο ότι θα τερματίσει. Αυτό έχει ως αποτέλεσμα μια κατάτμηση του χώρου δεδομένων σε κελιά Voronoi.

² Η PCM είναι ένας στοιχειώδης τρόπος διαμόρφωσης που δεν χρησιμοποιεί φέρον. Το μεταδιδόμενο (διαμορφωμένο) σήμα PCM είναι μια ψηφιακή αναπαράσταση του αναλογικού σήματος όπου το πλάτος του σήματος δειγματοληπτείται, κβαντίζεται και μεταδίδεται ως σειρά συμβόλων, συνήθως, δυαδικών (bit). Ο δέκτης από τους λαμβανόμενους παλμούς ανακτά την ψηφιακή ακολουθία συμβόλων και ανακατασκευάζει το αναλογικό σήμα μέσω μετατροπής D/A.

Το πρόβλημα είναι υπολογιστικά δύσκολο (NP-hard), ωστόσο υπάρχουν αποτελεσματικοί ευρετικοί αλγόριθμοι που χρησιμοποιούνται συχνά και συγκλίνουν γρήγορα σε ένα τοπικά βέλτιστο. Για μίξεις Γκαουσιανών κατανομών μέσω μιας επαναληπτικής προσέγγισης τελειοποίησης, οι αλγόριθμοι αυτοί είναι συνήθως παρόμοιοι με τον αλγόριθμο EM (Expectation Maximization -μεγιστοποίηση αναμονής). Επιπλέον, και οι δύο χρησιμοποιούν κέντρα των ομάδων για να μοντελοποιήσουν τα δεδομένα, ωστόσο η ομαδοποίηση k-means τείνει να βρει ομάδες συγκρίσιμης χωρικής έκτασης, ενώ ο μηχανισμός EM επιτρέπει στις ομάδες να έχουν διάφορα σχήματα.

Η k-means μέθοδος εξακολουθεί να είναι πολύ δημοφιλής σήμερα, και έχει εφαρμοστεί σε διάφορους τομείς που κυμαίνονται από την υπολογιστική βιολογία έως τα γραφικά υπολογιστών. Το βασικό **πλεονέκτημα** του αλγορίθμου έγκειται στην απλότητα και την ταχύτητα. Ο χρόνος εκτέλεσης του k-means έχει μελετηθεί εκτενώς πειραματικά.

Τα βασικά χαρακτηριστικά του k-means που τον καθιστούν αποτελεσματικό συχνά θεωρούνται ως τα μεγαλύτερα μειονεκτήματα:

- Η Ευκλείδεια απόσταση χρησιμοποιείται ως ένα μέτρο σύγκρισης και η διακύμανση χρησιμοποιείται ως μέτρο της διασποράς της ομάδας.
- Το πλήθος των ομάδων k είναι μια παράμετρος εισόδου: μια ακατάλληλη επιλογή του k μπορεί να δώσει φτωχά αποτελέσματα. Αυτό συμβαίνει γιατί, όταν εκτελείται ο k-means, είναι σημαντικό να εκτελούνται διαγνωστικοί έλεγχοι για τον προσδιορισμό του αριθμού των ομάδων του συνόλου των δεδομένων.
- Η σύγκλιση στο τοπικό ελάχιστο μπορεί να παράγει προφανή («λάθος») αποτελέσματα.

Ένας βασικός περιορισμός του k-means είναι το μοντέλο της ομαδοποίησης. Η ιδέα βασίζεται σε σφαιρικά συμπλέγματα σημείων (ομάδες) τα οποία διαχωρίζονται κατά τέτοιο τρόπο ώστε η μέση τιμή να συγκλίνει προς το κέντρο της ομάδας. Οι ομάδες αναμένεται να είναι ανάλογου μεγέθους, έτσι ώστε η αντιστοίχιση στο πλησιέστερο κέντρο να είναι σωστή.

Το αποτέλεσμα του k-means όπως αναφέρθηκε και παραπάνω μπορεί να θεωρηθεί ως κελιά Voronoi των μέσων των ομάδων. Δεδομένου ότι τα δεδομένα είναι χωρισμένα στη μέση μεταξύ των μέσων των ομάδων, αυτό μπορεί να οδηγήσει σε μη ευνοϊκές διαιρέσεις. Τα Γκαουσιανά μοντέλα που χρησιμοποιήθηκαν από τον αλγόριθμο EM (ο οποίος μπορεί να θεωρηθεί ως μια γενίκευση του k-means) είναι πιο ευέλικτα εδώ έχοντας τόσο διακυμάνσεις όσο και συνδιακυμάνσεις. Το αποτέλεσμα EM είναι έτσι σε θέση να χωρέσει ομάδες μεταβλητού μεγέθους πολύ καλύτερα από τον k-means, καθώς επίσης και συσχετιζόμενες ομάδες.

Στην ομαδοποίηση k-means σε **πολυδιάστατο χώρο**, μας δίνεται ένα σύνολο n σημειακών δεδομένων στον d -διάστατο χώρο R^d και ένας ακέραιος k και το πρόβλημα είναι να προσδιοριστεί το σύνολο σημείων k στον R^d , που ονομάζονται

κέντρα, έτσι ώστε να ελαχιστοποιείται η μέση τετραγωνική απόσταση από κάθε σημείο στο πλησιέστερο κέντρο της. Το πρόβλημα αυτό συχνά αποκαλείται παραμόρφωση τετραγωνικού σφάλματος και αυτό το είδος ομαδοποίησης εμπίπτει στη γενική κατηγορία της βασισμένης στη διακύμανση ομαδοποίησης.

Η ομαδοποίηση με βάση τον k-means είναι στενά συνδεδεμένη με μια σειρά άλλων προβλημάτων ομαδοποίησης και θέσης. Αυτά περιλαμβάνουν τα Ευκλείδεια k-medians, στα οποία στόχος είναι να ελαχιστοποιηθεί το άθροισμα των αποστάσεων από το πλησιέστερο κέντρο, και τη γεωμετρικό k-center πρόβλημα, στο οποίο στόχος είναι η ελαχιστοποίηση της μέγιστης απόστασης από κάθε σημείο από το πλησιέστερο κέντρο της.

3.1.5 Ο βασικός αλγόριθμος k-means

Όπως ήδη αναφέρθηκε, ο βασικός αλγόριθμος k-means εισήχθη από τον Lloyd το 1982 και πάνω σε αυτόν στηρίχθηκαν οι παραλλαγές που εισήχθησαν στη συνέχεια.

Ο **αλγόριθμος Lloyd** βασίζεται στην απλή παρατήρηση ότι η βέλτιστη θέση για ένα κέντρο είναι στο κεντροειδές του σχετικού cluster. Λόγω της απλότητας και της ευελιξίας του, ο αλγόριθμος Lloyd είναι πολύ δημοφιλής στη στατιστική ανάλυση. Συγκεκριμένα, δεδομένου κάθε άλλου αλγορίθμου, ο αλγόριθμος Lloyd μπορεί να εφαρμοστεί ως ένα στάδιο τελικής επεξεργασίας για τη βελτίωση της τελικής παραμόρφωση. Όπως μπορεί να διαπιστωθεί από πειράματα, αυτό μπορεί να οδηγήσει σε σημαντικές βελτιώσεις. Ωστόσο, μια απλή εφαρμογή του αλγορίθμου Lloyd μπορεί να είναι αρκετά αργή. Αυτό οφείλεται κυρίως στο κόστος του υπολογισμού των πλησιέστερων γειτόνων. Μια απλή και ταυτόχρονα αποτελεσματική υλοποίηση του Lloyd, ονομάζεται **αλγόριθμος φιλτραρίσματος**. Αυτός ο αλγόριθμος ξεκινά αποθηκεύοντας τα δεδομένα σε ένα kd-tree. Σε κάθε στάδιο του αλγορίθμου Lloyd, υπολογίζεται το κοντινότερο κέντρο κάθε σημείου και κάθε κέντρο μετακινείται στο κεντροειδές του σχετικού γείτονα. Η ιδέα είναι να διατηρείται, για κάθε κόμβο του δέντρου ένα υποσύνολο υποψήφιων κέντρων. Αυτά τα υποψήφια κέντρα, για κάθε κόμβο κλαδεύονται, ή φιλτράρονται, καθώς μεταβιβάζονται στους κόμβους παιδιά. Επειδή το kd-tree υπολογίζεται για τα στοιχεία αντί για τα κέντρα, δεν υπάρχει η ανάγκη ενημέρωσης αυτής της δομής σε κάθε στάδιο του αλγορίθμου Lloyd.

Δεδομένου ενός αρχικού συνόλου k με μέσα m_1, \dots, m_k , ο αλγόριθμος εξελίσσεται εναλλασσόμενος μεταξύ δύο βημάτων:

Βήμα ανάθεσης: στο στάδιο αυτό πραγματοποιείται αντιστοίχιση κάθε παρατήρησης στην ομάδα της οποίας η μέση τιμή βρίσκεται πιο κοντά σε αυτήν, δηλαδή γίνεται διαχωρισμός των παρατηρήσεων σύμφωνα με το διάγραμμα Voronoi που δημιουργούνται από τις μέσες τιμές.

$$S_i^{(t)} = \{x_p : \|x_p - m_i^{(t)}\| \leq \|x_p - m_j^{(t)}\| \forall 1 \leq j \leq k\}$$

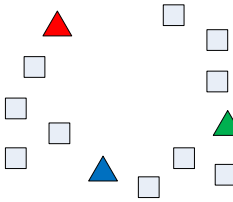
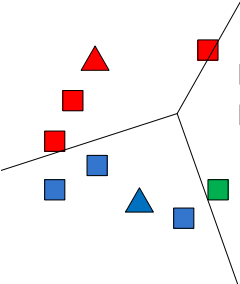
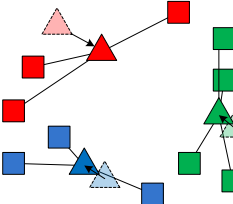
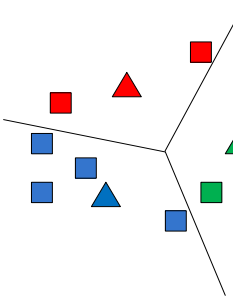
όπου κάθε x_p έχει αντιστοιχιστεί σε ακριβώς ένα $S_i^{(t)}$, ακόμη και αν θα μπορούσε να έχει αντιστοιχιστεί σε δύο ή περισσότερα από αυτά.

Βήμα ενημέρωσης: υπολογίζει τις νέες μέσες τιμές που θα είναι τα κεντροειδή των παρατηρήσεων στις νέες ομάδες. Ο αλγόριθμος Lloyd βασίζεται στην απλή παρατήρηση ότι η βέλτιστη τοποθέτηση του κέντρου είναι στο κεντροειδές της συναφούς ομάδας. Για τα σημεία σε γενική θέση (ειδικότερα, εάν δεν υπάρχει σημείο το οποίο να ισαπέχει από δύο κέντρα), ο αλγόριθμος τελικά θα συγκλίνει σε ένα σημείο το οποίο είναι τοπικό ελάχιστο. Ωστόσο, το αποτέλεσμα δεν είναι κατ' ανάγκη ένα καθολικό ελάχιστο.

$$m_i^{(t+1)} = \frac{1}{|S_i^{(t)}|} \sum_{x_j \in S_i^{(t)}} x_j$$

Ο αλγόριθμος συγκλίνει, όταν οι αντιστοιχίες δεν αλλάζουν.

Πρέπει να σημειωθεί ότι ο αλγόριθμος Lloyd δεν προσδιορίζει την αρχική τοποθέτηση των κέντρων, αλλά χρησιμοποιούνται αλγόριθμοι αρχικοποίησης οι οποίοι το κάνουν και περιγράφονται σε επόμενη ενότητα. Το διάγραμμα που ακολουθεί επιδεικνύει συνοπτικά τη λειτουργία του αλγορίθμου Lloyd.

| | |
|---|---|
|  | <p>1) Τα αρχικά μέσα (means) κ αναπαράγονται τυχαία εντός της περιοχής των δεδομένων. Στη συγκεκριμένη περίπτωση είναι $k=3$ και παρουσιάζονται χρωματισμένα.</p> |
|  | <p>2) Δημιουργούνται k ομάδες (clusters) συσχετίζοντας κάθε παρατήρηση με το κοντινότερο μέσο. Οι ταξινομήσεις εδώ αντιπροσωπεύουν το διάγραμμα Voronoi που δημιουργείται με βάση τα μέσα.</p> |
|  | <p>3) Το κεντροειδές κάθε μίας από τις k ομάδες γίνεται το νέο μέσο.</p> |
|  | <p>4) Τα βήματα 2 και 3 επαναλαμβάνονται έως ότου επέλθει σύγκλιση.</p> |

Πίνακας 3.1: Παρουσίαση της λειτουργίας του αλγορίθμου

Ο πιο κοινός τρόπος να βελτιωθεί η αποτελεσματικότητα του αλγορίθμου του Lloyd είναι να υπολογιστούν οι πλησιέστεροι γείτονες πιο αποτελεσματικά, ας πούμε από την προεπεξεργασία των κεντρικών σημείων κατά την έναρξη κάθε σταδίου σε μια δομή δεδομένων για την απάντηση στα ερωτήματα του πλησιέστερου γείτονα. Αυτός ο αλγόριθμος αρχίζει με την αποθήκευση των σημείων σε ένα kd-δέντρο. Υπενθυμίζεται ότι, σε κάθε στάδιο του αλγορίθμου Lloyd, υπολογίζεται το κοντινότερο σε κάθε σημείο κέντρο και κάθε κέντρο μετακινείται προς το κεντροειδές των αντίστοιχων γειτόνων. Η ιδέα είναι να διατηρηθεί, για κάθε κόμβο του δέντρου, ένα υποσύνολο υποψήφιων κέντρων. Οι υποψήφιοι για κάθε κόμβο «κλαδεύονται», ή φιλτράρονται, καθώς μεταδίδονται στα παιδιά του κόμβου. Εφόσον το kd-δέντρο υπολογίζεται για τα σημειακά δεδομένα αντί για τα κέντρα, δεν υπάρχει καμία ανάγκη για ενημέρωση αυτής της δομής με κάθε στάδιο του αλγορίθμου Lloyd. Επίσης, δεδομένου ότι συνήθως υπάρχουν πολλά περισσότερα σημεία από ό,τι κέντρα, υπάρχουν μεγαλύτερες οικονομίες κλίμακας για να πραγματοποιηθούν.

3.1.6 Μέθοδοι αρχικοποίησης

Η διαδικασία της αρχικοποίησης είναι ο ορισμός των k αρχικών κέντρων των ομάδων. Ο αρχικός κλασικός k -means ορίζει τα k πρώτα αντικείμενα του συνόλου των προς ομαδοποίηση δεδομένων ως τα k πρώτα κέντρα των ομάδων. Αυτή η επιλογή είναι μια λύση του προβλήματος της αρχικοποίησης, αλλά αν λάβουμε υπ' όψιν μας πρώτον, την ευαισθησία του k -means από τα αρχικά μέσα και δεύτερον, ότι ο αλγόριθμος αυτός συνήθως συγκλίνει σε ένα τοπικό βέλτιστο και όχι στο ολικό βέλτιστο μας κάνει να αναζητήσουμε μια πιο αποδοτική τεχνική επιλογής των αρχικών μέσων.

Δύο πιο αποτελεσματικές μέθοδοι αρχικοποίησης που χρησιμοποιούνται συνήθως είναι οι **Forgy** και **Random Partition**. Η μέθοδος Forgy επιλέγει τυχαία k παρατηρήσεις από το σύνολο δεδομένων και τις χρησιμοποιεί ως το αρχική μέση τιμή. Η μέθοδος Random Partition αρχικά εκχωρεί τυχαία μια ομάδα σε κάθε παρατήρηση και προχωρά στη συνέχεια στο βήμα ενημέρωσης, υπολογίζοντας έτσι τα αρχικά μέσα ως κεντροειδή των τυχαία αντιστοιχισμένων σημείων της ομάδας. Η μέθοδος Forgy τείνει να εξαπλώνει τα αρχικά μέσα, ενώ η Random Partition τα τοποθετεί όλα κοντά στο κέντρο του συνόλου δεδομένων. Σύμφωνα με τον Hamerly, η μέθοδος Random Partition είναι γενικά προτιμότερη για αλγόριθμους όπως τον k -harmonic means και τον fuzzy k -means. Για τους αλγόριθμους EM και τον πρότυπο k -means, η Forgy μέθοδος αρχικοποίησης είναι προτιμότερη.

Καθώς πρόκειται για έναν ευρετικό αλγόριθμο, δεν υπάρχει εγγύηση ότι θα συγκλίνει σε ένα ολικό βέλτιστο, και το αποτέλεσμα μπορεί να εξαρτάται από τις αρχικές ομάδες. Είναι συνηθισμένο να εκτελέσει κανείς τον αλγόριθμο πολλές φορές με διαφορετικές συνθήκες εκκίνησης, καθώς συνήθως είναι πολύ γρήγορος. Ωστόσο, στη χειρότερη περίπτωση, ο k -means μπορεί να είναι πολύ αργός έως ότου να συγκλίνει: συγκεκριμένα έχει αποδειχθεί ότι υπάρχουν ορισμένα σύνολα σημείων, ακόμη και στις 2 διαστάσεις, στα οποία ο k -means καθυστερεί εκθετικά να συγκλίνει. Αυτά τα σύνολα σημείων δεν φαίνεται να προκύπτουν στην πράξη: αυτό επιβεβαιώνεται από το γεγονός ότι ο επίπεδος χρόνος λειτουργίας του k -means είναι πολυωνυμικός. Το βήμα «ανάθεσης» αναφέρεται επίσης ως **βήμα αναμονής** και το

βήμα «ενημέρωσης» ως **βήμα μεγιστοποίησης**, καθιστώντας έτσι τον αλγόριθμο μια παραλλαγή του γενικευμένου αλγορίθμου EM.

3.1.7 Εφαρμογές του αλγορίθμου

Η ομαδοποίηση k-means ιδίως όταν χρησιμοποιεί ευρετικές μεθόδους όπως τον αλγόριθμο Lloyd είναι μάλλον εύκολο να υλοποιηθεί και να εφαρμοστεί ακόμα και σε μεγάλα σύνολα δεδομένων. Ως εκ τούτου, έχει χρησιμοποιηθεί με επιτυχία σε διάφορα θέματα, που κυμαίνονται από την κατάτμηση της αγοράς, τη μηχανική όραση, τη γεωστατιστική, και την Αστρονομία μέχρι και τη γεωργία. Συχνά χρησιμοποιείται ως ένα βήμα προεπεξεργασίας για άλλους αλγόριθμους, για παράδειγμα, για την εύρεση παραμέτρων εκκίνησης.

Τα προβλήματα ομαδοποίησης προκύπτουν σε άλλες εφαρμογές, όπως η εξόρυξη δεδομένων και η ανακάλυψη γνώσης, η συμπίεση δεδομένων και ο κβαντισμός διανυσμάτων, η αναγνώριση προτύπων και η ταξινόμηση προτύπων. Η αντίληψη του τι συνιστά μια καλή ομάδα εξαρτάται από την εφαρμογή. Υπάρχουν πολλές μέθοδοι για την εύρεση ομάδων που υπόκεινται σε διάφορα κριτήρια. Αυτά περιλαμβάνουν προσεγγίσεις που βασίζονται στον διαχωρισμό και τη συγχώνευση, τις τυχαιοποιημένες προσεγγίσεις, μεθόδους που βασίζονται σε νευρωνικά δίκτυα, καθώς και μεθόδους που έχουν σχεδιαστεί για να προσαρμόζονται σε μεγάλες βάσεις δεδομένων.

3.2 Παραλλαγές

Υπάρχει ένας μεγάλος αριθμός παραλλαγών που έχουν προταθεί για αυτόν τον αλγόριθμο, έτσι ώστε να μπορεί να επιλέξει κανείς την παραλλαγή εκείνη που ταιριάζει καλύτερα στους σκοπούς της εφαρμογής του.

Ορισμένες γνωστές παραλλαγές του αλγορίθμου k-means είναι:

- ✓ Ο αλγόριθμος **ISODATA** ο οποίος περιλαμβάνει μία διαδικασία αναζήτησης του καλύτερου αριθμού ομάδων με βάση κάποιο κόστος εκτέλεσης. Ο αλγόριθμος ISODATA έχει κάποιες περαιτέρω βελτιώσεις που αφορούν στη διάσπαση και τη συγχώνευση των ομάδων (Jensen, 1996). Οι ομάδες συγχωνεύονται είτε εάν ο αριθμός των στοιχείων σε μια ομάδα είναι μικρότερος από ένα ορισμένο όριο είτε εάν τα κέντρα δύο ομάδων είναι πιο κοντά από ένα ορισμένο όριο. Οι ομάδες χωρίζονται σε δύο διαφορετικές ομάδες, αν η τυπική απόκλιση της ομάδας υπερβαίνει μια προκαθορισμένη τιμή και ο αριθμός των μελών είναι διπλάσιος από το όριο για τον ελάχιστο αριθμό των μελών. Ο αλγόριθμος ISODATA είναι παρόμοιος με τον αλγόριθμο k-means με τη διαφορά ότι ο αλγόριθμος ISODATA επιτρέπει διαφορετικό αριθμό ομάδων, ενώ ο k-means υποθέτει ότι ο αριθμός των ομάδων είναι γνωστός εκ των προτέρων.
- ✓ Ο **Fuzzy C-Means** ο οποίος επεκτείνει τον κλασικό αλγόριθμο k-means χρησιμοποιώντας την θεωρία της ασαφούς λογικής. Αυτός ο αλγόριθμος λειτουργεί αναθέτοντας συμμετοχή σε κάθε σημείο δεδομένων που αντιστοιχεί στο κέντρο κάθε ομάδας με βάση την απόσταση μεταξύ του κέντρου της ομάδας και του σημείου. Όσο πιο κοντά είναι τα δεδομένα στο κέντρο της ομάδας τόσο

πιο μεγάλη είναι η συμμετοχή τους στο συγκεκριμένο κέντρο ομάδας. Σαφώς, η άθροιση της συμμετοχής κάθε σημείου θα πρέπει να είναι ίση με ένα. Μετά από κάθε επανάληψη η συμμετοχή και τα κέντρα των ομάδων ενημερώνονται. Ο αλγόριθμος σταματάει όταν τα αποτελέσματα συγκλίνουν. Στην ασαφή ομαδοποίηση τα στοιχεία μπορεί να ανήκουν σε περισσότερες από μία ομάδες και η συσχέτιση με κάθε στοιχείο είναι ένα σύνολο επιπέδων συμμετοχής.

- ✓ Ο **SAS PROC FASTCLUS**. Η διαδικασία αυτή συνδυάζει μία αποτελεσματική μέθοδο για την εύρεση των αρχικών ομάδων με ένα πρότυπο επαναληπτικό αλγόριθμο για την ελαχιστοποίηση του αθροίσματος των τετραγώνων των αποστάσεων από τα μέσα των ομάδων. Με τη μέθοδο αυτή ένα σύνολο σημείων που ονομάζεται *cluster seeds* επιλέγεται ως μία πρώτη εικασία των μέσων των ομάδων. Κάθε παρατήρηση εκχωρείται στο πλησιέστερο *cluster seed* για να σχηματισμό προσωρινών ομάδων. Τα *cluster seeds* αντικαθίστανται τότε από τα μέσα των προσωρινών ομάδων, και η διαδικασία επαναλαμβάνεται έως ότου δεν υπάρχουν περαιτέρω αλλαγές στις ομάδες. Η FASTCLUS διαδικασία διαφέρει από άλλες αντίστοιχες μεθόδους ταξινόμησης στον τρόπο που επιλέγονται τα αρχικά σημεία.
- ✓ Το μοντέλο της μείξης Γκαουσιανών κατανομών ή αλλιώς **Gaussian mixture models**, εκπαιδευμένο με το αλγόριθμο Expectation – Maximization (EM), διατηρεί πιθανολογικές αντιστοιχίσεις σε ομάδες, αντί των ντετερμινιστικών αντιστοιχίσεων, καθώς επίσης και πολυπαραγοντικές Γκαουσιανές κατανομές, αντί των μέσων. Η model-based προσέγγιση αποτελεί έναν άλλο τρόπο αντιμετώπισης των προβλημάτων ομαδοποίησης. Στηρίζεται στη χρήση έτοιμων μοντέλων για τις ομάδες και την προσπάθεια για βελτιστοποίηση του ταιριάσματος μεταξύ των δεδομένων και του μοντέλου. Μαθηματικά κάθε ομάδα μπορεί να αντιπροσωπευτεί από μία παραμετρική κατανομή όπως η Γκαουσιανή ή η Poisson. Το σύνολο των δεδομένων μοντελοποιείται επομένως με βάση μίας μείξης κατανομών. Με το μοντέλο της μείξης Γκαουσιανών κατανομών θεωρούμε τις ομάδες ως Γκαουσιανές κατανομές με κέντρο τα κεντροειδή τους. Ο αλγόριθμος που χρησιμοποιείται για την εύρεση της μείξης των Γκαουσιανών κατανομών είναι ο Expectation – Maximization (EM).
- ✓ Πολλές μέθοδοι έχουν προταθεί για την επιλογή καλύτερων αρχικών ομάδων. Μία πρόσφατη πρόταση αποτελεί ο **k-means++**. Ο k-means++ είναι ένας αλγόριθμος για την επιλογή των αρχικών τιμών (*seeds*) για τον αλγόριθμο k-means. Προτάθηκε το 2007 από τους David Arthur και Sergei Vassilvitskii, ως ένας προσεγγιστικός αλγόριθμος για το NP-hard πρόβλημα του k-means. Αποτελεί έναν τρόπο για την αποφυγή των φτωχών ομαδοποιήσεων που δίνει μερικές φορές ο πρότυπος αλγόριθμος k-means. Η προσέγγιση του k-means μπορεί να είναι κακή σε σχέση με την αντικειμενική συνάρτηση σε σύγκριση με τη βέλτιστη ομαδοποίηση. Ο αλγόριθμος k-means++ αντιμετωπίζει το πρόβλημα αυτό, καθορίζοντας μια διαδικασία για την αρχικοποίηση των κέντρων των ομάδων πριν την υλοποίηση των επαναλήψεων βελτιστοποίησης με τον πρότυπο αλγόριθμο k-means. Με την αρχικοποίηση k-means++, ο αλγόριθμος βρίσκει αποδεδειγμένα μια λύση που είναι $O(\log k)$ ανταγωνιστική ως προς τη βέλτιστη λύση k-means.
- ✓ Ο αλγόριθμος ομαδοποίησης **Spherical k-means** είναι κατάλληλος για διανυσματικά δεδομένα. Όταν τα δεδομένα κανονικοποιούνται είναι πιο κατάλληλη μια παραλλαγή του αλγόριθμου k-means η οποία λειτουργεί με τα

κανονικοποιημένα δεδομένα. Δεδομένου ότι τα σημεία δεδομένων βρίσκονται σε μία μονάδα υπερσφαίρας, ο αλγόριθμος ονομάζεται Spherical k-means.

- ✓ Ο αλγόριθμος **Minkowski metric weighted k-means** αντιμετωπίζει το πρόβλημα των στοιχείων θορύβου αναθέτοντας βάρη σε κάθε στοιχείο ανά ομάδα. Ο αλγόριθμος αυτός ελαχιστοποιεί το άθροισμα της απόστασης του κάθε στοιχείου από το αντίστοιχο κέντρο βάρους του. Υπολογίζει αυτόματα τα βάρη των στοιχείων για κάθε ομάδα. Το πρόβλημα που αντιμετωπίζεται με τον αλγόριθμο αυτό σχετίζεται με το γεγονός ότι ο k-means αντιμετωπίζει όλα τα στοιχεία σε ένα σύνολο δεδομένων σα να έχουν τον ίδιο βαθμό συσχέτισης. Ωστόσο, στα περισσότερα σύνολα δεδομένων, διαφορετικά στοιχεία θα έχουν διαφορετικούς βαθμούς συσχέτισης. Δεν είναι μόνο θέμα επιλογής στοιχείων, αλλά απόδοσης βαρών στα στοιχεία. Ακόμα κι αν δύο στοιχεία σχετίζονται, δε θα έχουν πάντα τον ίδιο βαθμό συσχέτισης. Γι' αυτό και έχει ενσωματωθεί η απόδοση βαρών στα στοιχεία στον k-means.

Ένα κοινό πρόβλημα των διαφόρων k-means αλγορίθμων είναι ότι χρειάζονται πολύ χρόνο για να τρέξουν. Οι λόγοι είναι δύο. Πρώτον, συχνά εφαρμόζονται σε μετρίων έως υψηλών διαστάσεων χώρους, και ο υπολογισμός των πλησιέστερων γειτόνων σε τέτοιους χώρους δεν είναι ένα ασήμαντο πρόβλημα. Ο δεύτερος λόγος είναι ότι συχνά απαιτούνται πολλές επαναλήψεις μέχρι οι προϋποθέσεις τερματισμού του αλγορίθμου να πληρούνται.

3.3 Απαιτήσεις για το δικό μας Πρόβλημα

Η υλοποίηση της εφαρμογής που μελετάται στην παρούσα εργασία απαιτεί την επεξεργασία των δεδομένων των καταγραφών των προβλημάτων που συλλέγονται από το πλήθος. Τα δεδομένα των καταγραφών όπως λαμβάνονται από το πλήθος δεν είναι αξιοποιήσιμα. Αυτό συμβαίνει για δύο βασικούς λόγους:

- Καταρχάς συγκεντρώνεται ένας μεγάλος όγκος πληροφορίας και πολλαπλών σημείων με αποτέλεσμα να επικρατήσει σύγχυση κατά τη λήψη όλων των δεδομένων που θα συλλέγονται κατά τη διαδικασία των καταγραφών, λόγω της διαφορετικής θέσης της κάθε καταγραφής για το ίδιο πρόβλημα.
- Επιπλέον το πιθανό σφάλμα του GPS κατά τον υπολογισμό των συντεταγμένων της θέσης της καταγραφής προσδίδει μία μετατόπιση στην πραγματική θέση του προβλήματος.

Επομένως κάθε καταγραφή θα έχει διαφορετικές συντεταγμένες με αποτέλεσμα η παρέμβαση στα δεδομένα να είναι απαραίτητη, προκειμένου να υλοποιηθεί ομαδοποίηση της μαζικής πληροφορίας σύμφωνα με το καταγραφόμενο πρόβλημα, ώστε να είναι διαθέσιμα για περαιτέρω αξιοποίηση.

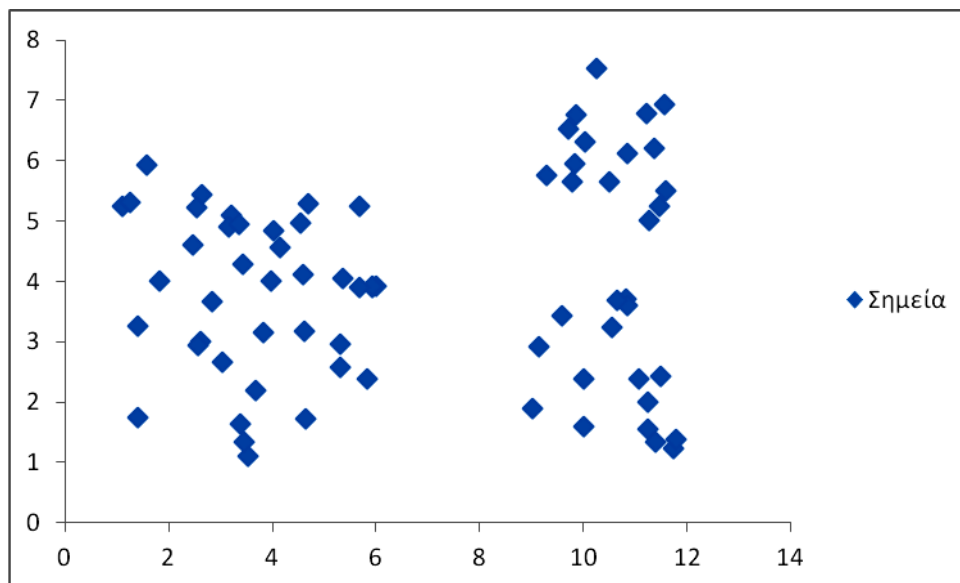
Η ομαδοποίηση των δεδομένων με βάση το πρόβλημα που θα καταγράφεται κάθε φορά θα πραγματοποιείται με χρήση του αλγορίθμου ομαδοποίησης k-means. Ο αλγόριθμος k-means θα ομαδοποιεί τα σημειακά δεδομένα με βάση την κατανομή τους στο χώρο, τον αριθμό των ομάδων k στον οποίο επιθυμούμε να διαχωριστούν τα δεδομένα, και προαιρετικά τόσα σημεία αρχικοποίησης όσες και οι ομάδες k που

επιθυμούμε να δημιουργηθούν. Επομένως, ο αλγόριθμος k-means για να υλοποιηθεί θα δέχεται τρεις παραμέτρους εισόδου:

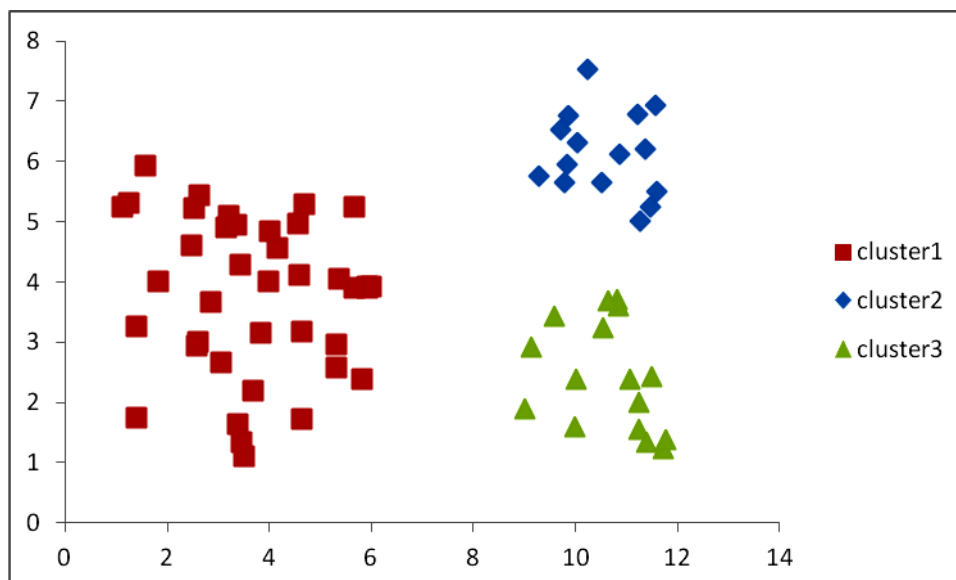
- Τα αρχικά σημειακά δεδομένα που θα συλλέγονται από τις καταγραφές των προβλημάτων.
- Τον αριθμό k των ομάδων που πρόκειται να διαμορφωθούν για την ταξινόμηση των σημείων.
- Τα διανύσματα αρχικοποίησης, τα οποία θα αποτελούν οδηγό για τον αλγόριθμο ώστε το αποτέλεσμα της ομαδοποίησης να είναι πιο ακριβές και πιο κοντά στην πραγματικότητα.

Το πρόβλημα που προκύπτει στο σημείο αυτό είναι η δημιουργία των διανυσμάτων αρχικοποίησης. Για την αντιμετώπιση του προβλήματος αυτού, θα εφαρμοστεί μία συνάρτηση για τη δημιουργία πολυγώνων μέσα στα οποία θα περιλαμβάνονται τα σημεία κάθε ομάδας. Οι συντεταγμένες των κεντροειδών των πολυγώνων αυτών θα αποτελούν στη συνέχεια τα διανύσματα αρχικοποίησης με βάση τα οποία θα υλοποιηθεί η ομαδοποίηση του αλγορίθμου k-means.

Τα αποτελέσματα που αναμένεται να προκύψουν από την ομαδοποίηση των σημείων με τον αλγόριθμο k-means πρέπει να έχουν παραπλήσια μορφή με τα αποτελέσματα της ιδανικής ομαδοποίησης που παρουσιάζεται ενδεικτικά στις παρακάτω εικόνες.



Εικόνα 3.1: Σύνολο των σημείων πριν την υλοποίηση της ομαδοποίησης



Εικόνα 3.2: Ιδανικό αποτέλεσμα της ομαδοποίησης k-means των σημείων για k=3

3.4 Περιορισμοί

Η συνάρτηση που περιγράφηκε στην προηγούμενη ενότητα για την υλοποίηση της πρώτης ομαδοποίησης των δεδομένων με τη δημιουργία πολυγώνων ομαδοποίησης, δέχεται ως μοναδική παράμετρο την απόσταση επιρροής των ομάδων, η τιμή της οποίας καθορίζει ποια από τα σημεία θα ταξινομηθούν σε μία ομάδα. Πρόκειται για τη μέγιστη απόσταση μεταξύ των σημείων.

Στην εφαρμογή που μελετάμε η μέγιστη απόσταση των σημείων μεταξύ τους επιλέχθηκε να είναι ίση με 30 μέτρα. Αυτό σημαίνει ότι θα ταξινομούνται στην ίδια ομάδα όλα τα σημεία που απέχουν το ένα από το άλλο έως 30 μέτρα. Η τιμή αυτή αρχικά προσεγγιστικά μετρώντας τη μέγιστη απόσταση μεταξύ δύο σημείων στην ίδια ομάδα και επιβεβαιώθηκε έπειτα από δοκιμές. Επίσης είναι μεγαλύτερη κατά 5 μέτρα από την τιμή της μέγιστης απόστασης από τα αρχικά σημεία που ορίστηκε κατά τη δημιουργία των τυχαίων σημείων, έτσι ώστε στην ακραία περίπτωση της παρουσίας μίας καταγραφής σε απόσταση 25 μέτρα από το πρόβλημα να συμπεριληφθεί στη συγκεκριμένη ομάδα περίπτωσης κατά την ταξινόμηση των σημείων.

Επιπλέον, κάθε περίπτωση προβλήματος βρίσκεται σε εύλογη απόσταση από κάθε άλλο πρόβλημα κι επομένως δεν υπάρχει περίπτωση σύγχυσης του αλγορίθμου. Η απόσταση μεταξύ των κοντινότερων σημείων καταγραφών δύο περιπτώσεων προβλημάτων είναι μεγαλύτερη από 60 μέτρα, δηλαδή το άθροισμα των αποστάσεων επιρροής δύο γειτονικών ομάδων. Εξαιρέση αποτελεί η περίπτωση μελέτης της αντίδρασης του αλγορίθμου k-means για γειτονικά προβλήματα όμοιου τύπου, όπου εφαρμόζονται διάφορες τιμές απόστασης για τη δημιουργία των πολυγώνων ομαδοποίησης.

4

ΤΕΧΝΟΛΟΓΙΕΣ ΠΟΥ ΧΡΗΣΙΜΟΠΟΙΗΘΗΚΑΝ

4.1 PHP

4.1.1 Τι είναι η PHP

Ο όρος PHP αποτελεί ακρωνύμιο των λέξεων *Hypertext Preprocessor* και είναι μία ευρέως διαδεδομένη, ανοικτού κώδικα, γενικής χρήσης χειρόγραφη γλώσσα η οποία είναι ειδικά κατάλληλη για διαδικτυακό προγραμματισμό και μπορεί να ενσωματωθεί στην HTML. Η μεγάλη διαφορά της με πανομοιότυπες γλώσσες προγραμματισμού (π.χ. C, Perl, Python) είναι ότι ο κώδικας εκτελείται στον διακομιστή, παράγοντας και κατά συνέπεια συνδυάζοντας HTML το οποίο αποστέλλεται στη συνέχεια στον πελάτη. Ο πελάτης θα λάβει τα αποτελέσματα από την εφαρμογή αυτών των ενεργειών, αλλά δε θα ξέρει τον κώδικα που κρύβεται από πίσω. Επίσης, πρέπει να σημειωθεί ότι η χρήση της είναι ιδιαίτερα εύκολη για τη δημιουργία εφαρμογών.

Οι σελίδες PHP περιέχουν HTML με ενσωματωμένο κώδικα. Ο κώδικας PHP περικλείεται σε ειδικές οδηγίες επεξεργασίας αρχής και τέλους `<?php και ?>` αντίστοιχα που επιτρέπουν στον ενδιαφερόμενο να μεταπηδήσει προς και από την PHP λειτουργία.

Το μεγαλύτερο πλεονέκτημα στη χρήση της PHP είναι ότι είναι πάρα πολύ απλή για έναν αρχάριο, αλλά ταυτόχρονα προσφέρει πολλά προηγμένα χαρακτηριστικά για έναν επαγγελματία προγραμματιστή.

4.1.2 Πριν από την Εμφάνιση της PHP

Πριν ακόμα την εμφάνιση της PHP, υπήρχε μεγάλη δραστηριότητα στο Web programming. Παλαιότερα, ο κώδικας για την ευκολότερη επεξεργασία των δεδομένων μιας φόρμας γραφόταν στη γλώσσα C. Αλλά, ενώ ο κώδικας αυτός ήταν πολύ γρήγορος στην εκτέλεσή του, ήταν πολύ δύσκολος και πολύπλοκος στο γράψιμό του. Ο κύριος λόγος γι' αυτό ήταν ότι η C δεν είχε σχεδιασθεί ειδικά για το Web και έτσι δεν υπήρχε έτοιμος κώδικας για την εκτέλεση κάποιων κοινών εργασιών και ο προγραμματιστής (Web developer) έπρεπε να τα κάνει όλα μόνος του.

Ο προγραμματισμός στην C βέβαια ήταν πολύ καλύτερος από την κατάσταση που υπήρχε μέχρι τότε, καθώς το πρωτόκολλο HTTP είναι ένα stateless σύστημα, που σημαίνει ότι δεν αποθηκεύει καθόλου δεδομένα ανάμεσα στις σελίδες και έτσι και

αυτό ακόμα το γράψιμο κώδικα σε C για την αποστολή δεδομένων ανάμεσα στις σελίδες ήταν ένα σημαντικό βήμα μπροστά.

Αυτό το πρόβλημα αντιμετωπίστηκε κάπως με μια ευκολότερη γλώσσα υψηλού επιπέδου, την Perl, όπου τα αρχικά προέρχονται από τις λέξεις "Practical Extraction and Report Language". Η Perl, αν και αρχικά δημιουργήθηκε ως μια γλώσσα επεξεργασίας κειμένου, είχε δυνατότητες για επεξεργασία των καταχωρήσεων μιας HTML φόρμας και όχι μόνο. Η σχεδίαση της Perl ήταν απλή : ένα script της Perl μπορούσε να κάνει όλες τις απαραίτητες λειτουργίες μιας ιστοσελίδας και μπορούσαμε να ενσωματώσουμε σ' αυτό όποιον κώδικα της HTML θέλαμε. Η Perl διέθετε ακόμη πάρα πολλές συναρτήσεις (functions) για να γίνονται εύκολα πολλές εργασίες και έτσι έγινε σύντομα πολύ δημοφιλής ανάμεσα στους Web developers.

Αν και αποτέλεσε ένα σημαντικό βήμα μπροστά για το Web development, η Perl απείχε ακόμα πολύ από το ιδανικό. Ο τρόπος εργασίας της, δηλαδή το ότι «μια γλώσσα μπορεί να τα κάνει όλα» σήμαινε ότι δεν ήταν σχεδιασμένη για το Web και πολλοί προγραμματιστές της Perl προτιμούσαν τον δομημένο, εύκολο στην ανάγνωση προγραμματισμό από τον προγραμματισμό της «μίας γραμμής», όπου υπήρχε συμπυκνωμένος και δύσκολος στην κατανόηση κώδικας. Ίσως το μεγαλύτερο ελάττωμά της ήταν ότι η Perl ήταν Perl-centric, που σημαίνει ότι για να δημιουργηθεί HTML κώδικας, έπρεπε να ενσωματώσουμε τον HTML κώδικα μέσα στην Perl.

4.1.3 Οι Πρώτες Εκδόσεις της PHP

Η αρχική έκδοση της PHP σχεδιάστηκε και δημιουργήθηκε από τον Rasmus Lerdorf στα μέσα της δεκαετίας του 1990 ως ένας τρόπος για να μπορούν να γίνονται διάφορες κοινές εργασίες στο Web ευκολότερα και με λιγότερες επαναλήψεις. Τότε, ο κύριος σκοπός ήταν να υπάρχει η λιγότερη δυνατή ποσότητα λογικής στην επίτευξη του αποτελέσματος και αυτό οδήγησε την PHP στο να γίνει HTML-centric, δηλαδή ο κώδικας της PHP ήταν ενσωματωμένος μέσα στον κώδικα της HTML.

Η πρώτη δημοφιλής έκδοση της PHP ονομάστηκε PHP/FI 2.0, από τα αρχικά Personal Home Page/Form Interpreter. Το κύριο χαρακτηριστικό αυτής της έκδοσης ήταν ότι ο PHP/FI parser ήταν γραμμένος κυρίως με το χέρι και έτσι δημιουργούνταν συχνά λάθη. Ο όρος parser (αναλυτής/επεξεργαστής) αναφέρεται στον μηχανισμό ο οποίος δέχεται ένα script (κώδικα) και το μετατρέπει σε κάτι που μπορεί να κατανοήσει ο υπολογιστής.

Μερικά από αυτά τα προβλήματα επιλύθηκαν στην έκδοση 3, όταν ο Zeev Suraski και ο Andi Gutmans ξαναέγραψαν την PHP από την αρχή χρησιμοποιώντας καινούργια εργαλεία. Η PHP απέκτησε πολλούς οπαδούς και όταν εμφανίσθηκε η καινούργια έκδοση στα μέσα του 2000, είχε ήδη εγκατασταθεί σε περισσότερα από 2,5 εκατομμύρια Web-site domains, σε σύγκριση με τα 250.000 μόλις 18 μήνες νωρίτερα.

Στα μέσα του έτους 2000, εμφανίσθηκε η έκδοση PHP 4, που είχε μεγάλες διαφορές από την PHP 3. Πολύ δουλειά έγινε στο να διασφαλισθεί η προς τα πίσω

συμβατότητα του κώδικα με τα παλιά scripts της PHP και έτσι η αναβάθμιση από την PHP 3 στην PHP 4 ήταν πολύ πιο ομαλή απ' ό,τι ήταν η αναβάθμιση από την PHP/FI στην PHP 3.

Η σημαντικότερη ίσως αλλαγή που έγινε στην PHP 4 ήταν η καθιέρωση της Μηχανής Zend (Zend Engine), η οποία δημιουργήθηκε από την εταιρεία Zend, των Zeev Suraski και Andi Gutmans. Το όνομα Zend προέρχεται από τις λέξεις ZEEν και aNDi και ο σκοπός της μηχανής ήταν να προωθήσει την PHP στο εταιρικό περιβάλλον, ώστε να υπάρχει πολύ περισσότερη ευελιξία στη γλώσσα απ' ό,τι παλαιότερα.

Μια άλλη σημαντική καινοτομία ήταν ότι η PHP μπορούσε τώρα να εκτελεστεί σε πολλούς Web servers, όπως Apache 1.3.x, Apache 2, Microsoft's IIS, Zeus, AOLServer κ.ά. Επίσης, η απόδοση της γλώσσας έκανε ένα πολύ μεγάλο άλμα μπροστά εξαιτίας δύο παραγόντων. Πρώτα, ενώ η PHP 3 χρησιμοποιούσε τη λογική "εκτέλεση ενώ γίνεται διερμήνευση", που σήμαινε ότι η PHP διάβαζε μια γραμμή πηγαίου κώδικα, τον διερμήνευε, τον εκτελούσε, διάβαζε μια άλλη γραμμή κώδικα, τον διερμήνευε, τον εκτελούσε, διάβαζε την επόμενη γραμμή κωκ. Αυτό σήμαινε ότι ο κώδικας διαβαζόταν και διερμηνευόταν πολλές φορές, χωρίς να υπάρχει κανένας απολύτως λόγος.

Η PHP 4 υιοθέτησε τη λογική "μεταγλώττιση πρώτα, εκτέλεση αργότερα", όπου πρώτα διάβαζε ολόκληρο το script και το μεταγλώττιζε σε ενδιάμεσο κώδικα (byte code) πριν το εκτελέσει. Αυτό είχε ως αποτέλεσμα μεγάλη αύξηση στην ταχύτητα. Ο κώδικας "byte code" αποτελεί μια εσωτερική αναπαράσταση ενός script που η PHP μπορεί να κατανοήσει εύκολα και είναι συνήθως πολύ μεγαλύτερος σε μήκος από το ίδιο το script καθώς η κάθε εντολή της PHP διασπάται (αναλύεται) σε πολλές άλλες πιο απλές εντολές.

Επίσης, η PHP 4 εισήγαγε την πολυεπεξεργασία (multi-threading), όπου μπορούν κάποιες συναρτήσεις να εκτελούνται ανεξάρτητα από το κυρίως script. Η PHP συνεχίζει να προχωράει ατάραχη και η τρέχουσα έκδοσή της είναι η 4.3.9, είναι δε εγκατεστημένη σε 9,5 εκατομμύρια περίπου Web servers σ' όλον τον κόσμο.

4.1.4 Η Τρέχουσα Έκδοση της PHP

Η PHP 5 ήταν ένα μεγάλο βήμα μπροστά για τη γλώσσα, αν και όχι τόσο μεγάλο όσο η μετάβαση από την PHP 3 στην PHP 4. Η PHP 5 προσφέρει scripts για αντικειμενοστραφή προγραμματισμό (object-oriented). Επίσης, υπάρχει μια μεγάλη ποικιλία από συναρτήσεις για αντικείμενα (objects) που τα κάνει πολύ πιο ευέλικτα και εύκολα στη χρήση τους. Ακόμη, τα αντικείμενα αντιμετωπίζονται πάντα ως αναφορές (references) ώστε να βοηθηθούν οι προγραμματιστές που δυσκολεύονται να εργαστούν με τα αντικείμενα.

4.1.5 Η Σύνδεση με την HTML

Ο κώδικας της PHP ενσωματώνεται μέσα στον κώδικα της HTML σε ειδικά μπλοκ κώδικα, που περικλείονται από τα σύμβολα `<?php` και `?>`, όπως φαίνεται παρακάτω:

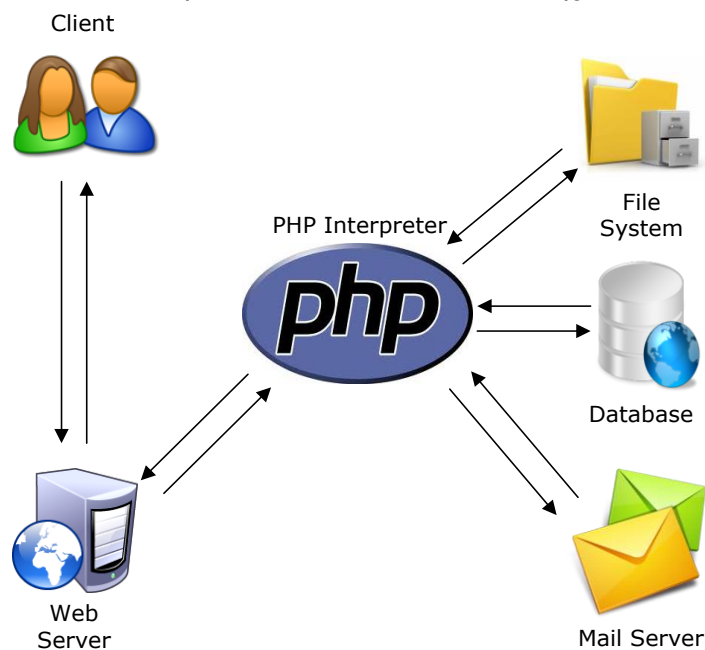
```
<html>
```

```

<body>
<p>Καλώς ήρθες,<?php print $Name; ?></p>
</body>
</html>

```

Δεν θα πρέπει να ξεχνάμε ότι ο κώδικας της PHP εκτελείται εξ ολοκλήρου στον διακομιστή (server) και έτσι ο πελάτης (client) λαμβάνει μόνο το τελικό αποτέλεσμα από την εκτέλεση του script. Αυτό σημαίνει με απλά λόγια ότι οι τελικοί χρήστες δεν μπορούν ποτέ να δουν τον πηγαίο κώδικα (source code) της PHP.



Εικόνα 4.1: Τρόπος λειτουργίας της PHP

4.1.6 Η Διερμηνευση και η Μεταγλώττιση

Η PHP χρησιμοποιεί μια μίξη από διερμηνευση (interpretation) και μεταγλώττιση (compilation) έτσι ώστε να μπορέσει να δώσει στους προγραμματιστές τον καλύτερο δυνατό συνδυασμό απόδοσης και ευελιξίας. Στο παρασκήνιο, η PHP μεταγλωττίζει το script σε μια σειρά από εντολές (instructions), που είναι γνωστές με τον όρο opcodes, οι οποίες εντολές εκτελούνται μία-μία μέχρι να τελειώσει το script.

Αυτό είναι κάτι διαφορετικό από τις παραδοσιακές γλώσσες που μεταγλωττίζονται, όπως είναι η C++, όπου ο κώδικας μεταγλωττίζεται σε εκτελέσιμο κώδικα μηχανής, ενώ η PHP μεταγλωττίζει εκ νέου το script κάθε φορά που αυτό απαιτείται. Αυτή η συνεχής μεταγλώττιση μπορεί να φαίνεται ως απώλεια χρόνου, αλλά δεν είναι καθόλου κακή καθώς δεν χρειάζεται να κάνουμε συνέχεια εμείς τη μεταγλώττιση των scripts όταν γίνονται κάποιες αλλαγές σ' αυτά.

4.1.7 Πώς γράφεται η PHP

Όπως είδαμε και νωρίτερα, ο κώδικας της PHP ενσωματώνεται μέσα στην HTML και τα στοιχεία της PHP που υπάρχουν σ' ένα script είναι σαν νησίδες κώδικα (code islands), δηλαδή αυτόνομα κομμάτια κώδικα που μπορούν να εκτελεστούν ανεξάρτητα από την περιβάλλουσα HTML. Αυτό βέβαια δεν σημαίνει ότι ο κώδικας της PHP δεν μπορεί να επηρεάσει την HTML, κάθε άλλο μάλιστα.

Τα scripts της PHP αποθηκεύονται συνήθως με την επέκταση .php και κάθε φορά που ο Web server πρέπει να στείλει ένα αρχείο που τελειώνει σε .php, πρώτα το στέλνει στον διερμηνευτή (interpreter) της PHP, ο οποίος εκτελεί τον κώδικα της PHP που υπάρχει στο script πριν επιστρέψει το παραγόμενο αρχείο στον τελικό χρήστη. Η κάθε γραμμή του PHP κώδικα είναι γνωστή ως εντολή (statement) και τελειώνει με το σύμβολο ;.

Οι μεταβλητές (variables) της PHP ξεκινούν με το σύμβολο \$ ακολουθούμενο από ένα γράμμα ή τον χαρακτήρα _ (underscore) και μετά από έναν συνδυασμό από γράμματα, ψηφία και τον χαρακτήρα _. Αυτό σημαίνει ότι δεν μπορούμε να ξεκινήσουμε το όνομα μιας μεταβλητής με ψηφίο.

4.1.8 Πώς δουλεύει η PHP

Η PHP είναι μια γλώσσα "server-side". Αυτό σημαίνει ότι ο κώδικας PHP που περιέχει μια σελίδα εκτελείται στον server (όπου είναι αποθηκευμένη η σελίδα), ενώ τα αποτελέσματα εμφανίζονται με μορφή HTML στον τελικό χρήστη.

Ο τρόπος με τον οποίο δουλεύει ένας web server (απαραίτητο λογισμικό για την επεξεργασία και τη λειτουργία μιας ιστοσελίδας) στον οποίο υπάρχει εγκατεστημένη η PHP είναι ο εξής: ο χρήστης "καλεί" μια σελίδα και ο server κάνει τις αντίστοιχες διεργασίες, για να παρουσιάσει το επιθυμητό αποτέλεσμα πίσω στο χρήστη. Μια απλή σελίδα HTML παρακάμπτει το εγκατεστημένο λογισμικό της PHP στον web server και εμφανίζεται όπως ακριβώς είναι στο χρήστη.

4.1.9 Τι μπορεί να κάνει η PHP

Η PHP κυρίως επικεντρώνεται στον προγραμματισμό και τις ενέργειες του διακομιστή (server-side scripting), επομένως μπορεί ο χρήστης να κάνει οτιδήποτε, μπορεί να κάνει οποιοδήποτε άλλο CGI πρόγραμμα. Αλλά η PHP μπορεί να κάνει πολλά περισσότερα.

Υπάρχουν τρεις βασικές περιπτώσεις στις οποίες χρησιμοποιούνται PHP σενάρια:

- Server-side scripting. Είναι το πιο συνηθισμένο. Χρειάζονται τρία πράγματα για να δουλέψει αυτό, ο PHP parser, ένας web server κι ένας web browser. Ο χρήστης πρέπει να "τρέξει" το διακομιστή ιστοσελίδων. Ο χρήστης μπορεί να έχει πρόσβαση στο αποτέλεσμα του προγράμματος PHP μέσω ενός προγράμματος περιήγησης. Όλα αυτά μπορούν να εκτελούνται στο σταθερό υπολογιστή καθώς ο ενδιαφερόμενος θα πειραματίζεται με τον προγραμματισμό σε PHP.
- Command line scripting. Μπορεί να τρέξει ένα PHP πρόγραμμα χωρίς κάποιον διακομιστή ή πρόγραμμα περιήγησης. Χρειάζεται μόνο ένα πρόγραμμα ανάλυσης PHP (PHP parser).
- Writing desktop applications. Η PHP δεν είναι η καλύτερη γλώσσα για τη δημιουργία εφαρμογών επιφάνειας εργασίας με ένα γραφικό περιβάλλον χρήστη, αλλά ο χρήστης που γνωρίζει καλά την PHP μπορεί να χρησιμοποιήσει κάποια προηγμένα PHP χαρακτηριστικά στις εφαρμογές του καθώς επίσης και PHP-GTK

για να γράψει κάποια προγράμματα. Η PHP-GTK είναι μία επέκταση της PHP, μη διαθέσιμη όμως στην κεντρική διανομή.

Η PHP μπορεί να χρησιμοποιηθεί σε όλα τα βασικά λειτουργικά συστήματα, όπως Linux, Microsoft Windows, Mac OS X, RISC OS. Η PHP υποστηρίζει επίσης τους περισσότερους διακομιστές δικτύου (web servers), μεταξύ των οποίων Apache, IIS, και πολλοί άλλοι. Η PHP λειτουργεί είτε ως μία λειτουργική μονάδα είτε ως CGI επεξεργαστής.

Επομένως με την PHP, ο προγραμματιστής έχει την ελευθερία να διαλέξει ένα λειτουργικό σύστημα κι έναν διακομιστή δικτύου. Επιπλέον, έχει την ευκαιρία να χρησιμοποιήσει δομημένο προγραμματισμό ή αντικειμενοστραφή προγραμματισμό, ή ένα μίγμα και των δύο.

Με την PHP ο χρήστης δεν περιορίζεται στην εξαγωγή HTML. Οι δυνατότητες της PHP περιλαμβάνουν την εξαγωγή εικόνων, PDF αρχείων, ακόμη και ταινίες Flash. Μπορεί επίσης να εξάγει εύκολα ένα κείμενο, όπως για παράδειγμα XHTML και οποιοδήποτε άλλο XML αρχείο. Η PHP μπορεί να παράγει μόνη της αυτά τα αρχεία, και να τα αποθηκεύσει στο σύστημα αρχείων (file system), αντί να τα εκτυπώσει, δημιουργώντας μία κρυφή μνήμη για το δυναμικό περιεχόμενο.

Ένα από τα πιο δυνατά και σημαντικά χαρακτηριστικά της PHP είναι ότι υποστηρίζει ένα ευρύ φάσμα βάσεων δεδομένων. Η δημιουργία μίας ιστοσελίδας στηριζόμενη σε βάσεις δεδομένων είναι πολύ απλή με τη χρήση συγκεκριμένων επεκτάσεων των βάσεων δεδομένων (π.χ. mysql), ή με τη σύνδεση σε οποιαδήποτε βάση δεδομένων υποστηρίζοντας την Open Database Connection μέσω της ODBC επέκτασης.

Η PHP υποστηρίζει επίσης την επικοινωνία με άλλες υπηρεσίες χρησιμοποιώντας πρωτόκολλα όπως LDAP, IMAP, SNMP, NNTP, POP3, HTTP, COM (στα Windows) και πολλά άλλα. Όσον αφορά τη διασύνδεση, η PHP υποστηρίζει επίσης τη δημιουργία αντικειμένων και τη χρήση αυτών ως PHP αντικείμενα.

Η PHP έχει χρήσιμες δυνατότητες επεξεργασίας κειμένου, οι οποίες περιλαμβάνουν επεκτάσεις και εργαλεία για την ανάλυση και την πρόσβαση του χρήστη σε XML έγγραφα. Η PHP τυποποιεί όλες τις XML επεκτάσεις με σταθερή βάση την libxml2, και επεκτείνει τη δυνατότητα του χρήστη να ορίσει την προσθήκη SimpleXML, XMLReader και XMLWriter υποστήριξης.

Τα χαρακτηριστικά και τα προτερήματα της PHP είναι πολλά περισσότερα από αυτά που αναφέρθηκαν παραπάνω. Αυτός είναι και ο λόγος που είναι τόσο διαδεδομένη και τόσο συνηθισμένη γλώσσα προγραμματισμού.

4.2 POSTGRES – POSTGIS – QGIS

Στην εποχή μας η χρήση χωρικών πληροφοριών γίνεται ολοένα και περισσότερο μέρος της δραστηριότητας μιας μεγάλης γκάμας επιχειρήσεων, επαγγελματιών και ιδρυμάτων. Ο αυξανόμενος όγκος αυτών των χωρικών πληροφοριών οδήγησε στην

ανάπτυξη συστημάτων που να είναι σε θέση να τις αποθηκεύσουν, διαχειριστούν και ανακτήσουν (να υποβάλουν ερωτήματα) με αποτελεσματικό τρόπο. Ανάμεσα σε αυτού του είδους τα συστήματα συγκαταλέγονται και οι Χωρικές Βάσεις Δεδομένων (ΧΒΔ). Ένα τέτοιο σύστημα είναι η χωρική βάση δεδομένων που αποτελείται από την βάση δεδομένων PostgreSQL και την επέκταση χωρικών λειτουργιών PostGIS. Τα λογισμικά αυτά τείνουν να χρησιμοποιούνται μαζί με ένα Σύστημα Γεωγραφικών Πληροφοριών Γραφείου, συνήθως το QuantumGIS, προκειμένου να επιτρέπεται η οπτικοποίηση των δεδομένων και η επέμβαση σε αυτά σε γραφικό περιβάλλον. Συνολικά το σύστημα που αποτελείται από τα παραπάνω λογισμικά επιτρέπει μία πλήρη γκάμα λειτουργιών αποθήκευσης, διαχείρισης και ανάκτησης χωρικών δεδομένων.

4.2.1 PostgreSQL

Η PostgreSQL αποτελεί ένα ανοιχτού κώδικα λογισμικό για τη δημιουργία σχεσιακών βάσεων δεδομένων με πολλές δυνατότητες και είναι διαθέσιμη χωρίς κόστος. Η ανάπτυξή της ήδη μετράει πάνω από 20 χρόνια και βασίζεται σε μια αποδεδειγμένα καλή αρχιτεκτονική η οποία έχει δημιουργήσει μια ισχυρή αντίληψη των χρηστών της γύρω από την αξιοπιστία, την ακεραιότητα δεδομένων και την ορθή λειτουργία. Αναπτύχθηκε αρχικά στο UC Berkeley στο Τμήμα Επιστήμης Υπολογιστών και εισήγαγε πολλές από τις αντικείμενο-σχεσιακές έννοιες οι οποίες τώρα είναι διαθέσιμες σε ορισμένες εμπορικές βάσεις δεδομένων. Παρέχει συναλλαγές, ακεραιότητα αναφορών, αποθηκευμένες διαδικασίες και επεκτασιμότητα τύπου. Η PostgreSQL είναι ένας open source απόγονος του αρχικού κώδικα Berkeley.

Η PostgreSQL τρέχει σε όλα τα βασικά λειτουργικά συστήματα, περιλαμβάνοντας τα Linux, UNIX, και Windows. Είναι ACID συμβατή, έχει ολοκληρωμένη υποστήριξη για foreign keys, joins, views, triggers, και stored procedures σε διάφορες γλώσσες προγραμματισμού. Συμπεριλαμβάνει τους περισσότερους τύπους δεδομένων SQL92 και SQL99, συμπεριλαμβανομένων INTEGER, NUMERIC, BOOLEAN, CHAR, VARCHAR, DATE, INTERVAL, και TIMESTAMP. Επίσης υποστηρίζει αποθήκευση binary large objects, όπως εικόνες, ήχοι ή video. Διαθέτει native programming interfaces για C/C++, Java, .Net, Perl, Python, Ruby, Tcl, ODBC, κ.ά.

Γενικά χαρακτηριστικά και συμβατότητα με προδιαγραφές

Η PostgreSQL είναι συνεπής με τις προδιαγραφές. Η υλοποίησή της είναι απολύτως σύμφωνη με τις προδιαγραφές ANSI-SQL 92/99. Έχει ολοκληρωμένη υποστήριξη για subqueries, read-committed και serializable transaction isolation levels. Η PostgreSQL αποτελεί ένα πλήρες σχεσιακό σύστημα που υποστηρίζει πολλαπλά σχήματα ανά βάση δεδομένων.

Στα Data integrity χαρακτηριστικά συμπεριλαμβάνονται: primary keys, foreign keys με υποστήριξη restricting και cascading updates/deletes, check constraints, unique constraints, και not null constraints.

Η PostgreSQL έχει αρκετά προηγμένα χαρακτηριστικά όπως: auto-increment columns μέσω sequences, LIMIT/OFFSET που επιτρέπουν την επιστροφή partial result sets. Όσον αφορά τα indexes υποστηρίζει compound, unique, partial, και

functional indexes τα οποία μπορούν να χρησιμοποιήσουν οποιονδήποτε από τους B-tree, R-tree, hash, ή GiST αλγόριθμους.

Το GiST (*Generalized Search Tree*) indexing αποτελεί ένα προηγμένο σύστημα το οποίο συνδυάζει ένα μεγάλο εύρος από διαφορετικούς αλγόριθμους ταξινόμησης και αναζήτησης όπως B-tree, B+-tree, R-tree, partial sum trees, ranked B+-trees και αρκετούς ακόμα. Επίσης διαθέτει interface το οποίο επιτρέπει τόσο την δημιουργία τύπων δεδομένων όσο και επεκτάσιμους τρόπους ερωτημάτων (query) για την αναζήτησή τους.

Βασικά χαρακτηριστικά

- ANSI SQL 89, 92, 93
- 100% συμβατή με ACID και πλήρη υποστήριξη commit και rollback.
- Online αντίγραφα ασφαλείας: υψηλότερη ασφάλεια και διαθεσιμότητα των δεδομένων.
- Τύποι δεδομένων: numeric, decimal, smallint, integer, bigint, real, double, serial, char, varchar, bit, text, date, time, timestamp, interval, boolean, network address, geometric types και πολλά άλλα.
- Δυνατότητα δημιουργίας νέων τύπων δεδομένων από τους χρήστες.
- Αποθήκευση BLOBS (binary large objects), συμπεριλαμβανομένων αρχείων κειμένου, ήχου, εικόνων ή βίντεο.
- Πλήρης υποστήριξη συναρτήσεων συγκεντρωτικών αποτελεσμάτων (GROUP_BY) όπως COUNT, SUM, AVG, MIN, MAX, STDDEV and VARIANCE. Δυνατότητα δημιουργίας νέων συγκεντρωτικών συναρτήσεων εφόσον χρειαστεί.
- Υποστήριξη όλων των τύπων ενώσεων (cross, inner, outer, left, right, full, natural).
- Συναρτήσεις ορισμένες από τον χρήστη, οι οποίες μπορούν να γραφούν σε πολλές γλώσσες προγραμματισμού όπως C, SQL, PL/pgSQL, TCL, Perl, Python and Ruby.
- Περιβάλλον ανάπτυξης γλωσσών προγραμματισμού όπως Perl, Python, Zope, PHP, TCL/TK, ODBC, JDBC, C/C++, Embedded SQL, Delphi/Kylix/Pascal, VB, ASP, Java.
- Βιβλιοθήκη συναρτήσεων και τελεστών με ορισμένες προεγκατεστημένες συναρτήσεις όπως math, date/time, string, geometric, formatting κ.α.
- Συναρτήσεις trigger μπορούν να οριστούν από οποιαδήποτε γλώσσα προγραμματισμού που υποστηρίζει server όπως C ή PL/pgQL.
- Προσωρινοί πίνακες οι οποίοι σβήνονται αυτόματα μετά το τέλος της συνόδου.
- Μοντέλο ασφαλείας ομάδας/χρήστη. Η πρόσβαση στο διακομιστή της βάσης δεδομένων μπορεί να περιορίζεται από το χρήστη, κεντρικό υπολογιστή ή μια βάση δεδομένων.
- Το μέγεθος του πίνακα και της βάσης δεδομένων είναι σχεδόν απεριόριστο. Απεριόριστες καταχωρήσεις και ευρετήρια ανά πίνακα.

Υποστηριζόμενοι τύποι δεδομένων

Η PostgreSQL διαθέτει μια ευρεία ποικιλία τύπων δεδομένων. Ο χρήστης έχει την δυνατότητα να προσθέσει ένα νέο τύπο δεδομένων χρησιμοποιώντας την εντολή CREATE TYPE.

| | | | | | | | | |
|--|------------------------|----------------------------------|---------------------------|--------------------------------|---------------------------------|----------------------------|---------|--------|
| Λογικοί και δυαδικοί (Boolean and binary types) | boolean, bool | bit(n) | bit varying(n), varbit(n) | | | | | |
| Χαρακτήρες (Character types) | character (n), char(n) | character varying(n), varchar(n) | text | | | | | |
| Αριθμητικοί (Numeric types) | smallint, int2 | integer, int, int4 | bigint, int8 | real, float4 | double precision, float8, float | numeric(p,s), decimal(p,s) | money | serial |
| Ημερομηνία και ώρα (Date and time types) | date | time | time with time zone | timestamp (includes time zone) | interval | | | |
| Γεωμετρικοί (Geometric types) | box | line | lseg | circle | path | point | polygon | |
| Δικτυακοί (Network types) | cidr | inet | macaddr | | | | | |
| Συστήματος (System types) | oid | xid | | | | | | |

Εικόνα 4.2: Τύποι δεδομένων που υποστηρίζει η PostgreSQL

4.2.2 PostGIS

Είναι μία επέκταση της PostgreSQL για να υποστηρίζει χωρικά δεδομένα, σύμφωνα με το πρότυπο του OGC. Παρέχει ειδικούς τελεστές για τη σύνταξη ερωτημάτων, λειτουργίες συνάθροισης επάνω σε χωρικά δεδομένα καθώς και χωρικές συναρτήσεις. Επιτρέπει επίσης την ανάθεση προβολικών συστημάτων στα χωρικά δεδομένα. Τέλος, μπορεί να χρησιμοποιηθεί για να οπτικοποιηθούν τα δεδομένα μέσω ειδικών εφαρμογών όπως το Quantum GIS και ο Geo Server.

Η PostGIS ενεργοποιεί χωρικά την δημοφιλή αντικειμενοστρεφή-σχεσιακή βάση δεδομένων PostgreSQL, επιτρέποντάς της να χρησιμοποιηθεί σαν βάση δεδομένων σε Γεωγραφικά Συστήματα Πληροφοριών (GIS) και εφαρμογές διαδικτυακής χαρτογράφησης, με τον ίδιο τρόπο που η Oracle Spatial ενεργοποιεί χωρικά την βάση δεδομένων Oracle.

Η PostGIS είναι σταθερή, γρήγορη, συμβατή με τα διεθνή πρότυπα, παρέχει εκατοντάδες χωρικές συναρτήσεις και είναι η πιο δημοφιλής αυτή τη στιγμή χωρική βάση δεδομένων ανοιχτού κώδικα. Η PostGIS χρησιμοποιείται από ποικίλους οργανισμούς ανά τον κόσμο, περιλαμβανομένων υψηλού ρίσκου κυβερνητικών υπηρεσιών και οργανισμών, αποθηκεύοντας terrabytes δεδομένων και εξυπηρετώντας εκατομμύρια διαδικτυακές κλήσεις ανά ημέρα.

Η διαχείριση της βάσης δεδομένων γίνεται μεταξύ άλλων μέσω των pgAdmin και rhpPgAdmin. Η είσοδος και έξοδος δεδομένων παρέχεται από πληθώρα εργαλείων μετατροπής (shp2pgsql, pgsq12shp, ogr2ogr, dxf2postgis). Επιπλέον υπάρχουν πολλά λογισμικά GIS (desktop και διαδικτυακά) για επισκόπηση δεδομένων σε PostGIS.

Βασικά Χαρακτηριστικά

- Εκατοντάδες χωρικές λειτουργίες
 - Εργαλεία χωρικής ανάλυσης όπως: Ζώνες επιρροής, ένωση, τομή, επίθεση, απόσταση και πολλά περισσότερα
- Ολοκλήρωση συναλλαγών ACID
- Χωρικοί κατάλογοι R-Tree
- Υποστήριξη πολλών ταυτόχρονων χρηστών
- Κλείδωμα σε επίπεδο γραμμής

- Δυνατότητα αντιγραφής
- Στεγανοποίηση
- Ασφάλεια με βάση ρόλους
- Χώροι πινάκων, σχήματα βάσης

4.2.3 Quantum GIS

Το Quantum GIS (QGIS) είναι ένα φιλικό προς το χρήστη Ανοιχτού Κώδικα Σύστημα Γεωγραφικών Πληροφοριών που χρησιμοποιείται για τη δημιουργία, τροποποίηση, ανάλυση και παρουσίαση των γεωγραφικών δεδομένων. Στο περιβάλλον λειτουργίας του QGIS υπάρχει μια περιοχή εμφάνισης των χωρικών δεδομένων (map view), και μια περιοχή διαχείρισης του υπομνήματος (legend) στην οποία παρατίθενται τα ονόματα και τα σύμβολα των δεδομένων, περίπου με τον ίδιο τρόπο που αυτά παρουσιάζονται στο ArcGIS. Επίσης περιέχονται διάφορα εργαλεία και μενού επιλογών. Στο QGIS κάθε εργασία αποθηκεύεται σε ένα αρχείο το οποίο καλείται qgis project (.qgs). Το βασικό προτέρημα του QGIS είναι ότι μπορεί να εισάγει, να οπτικοποιήσει, διαχειριστεί, τροποποιήσει – δημιουργήσει δεδομένα που περιέχονται σε μία βάση δεδομένων PostgreSQL, σύμφωνα με το πρότυπο της επέκτασης PostGIS. Το QGIS δεν υποστηρίζει τα χωρικά δεδομένα που μπορεί να υπάρχουν στη PostgreSQL χωρίς το PostGIS.

Το QGIS αποτελεί επίσημο πρόγραμμα του Open Source Geospatial Foundation (OSGeo). Τρέχει σε Linux, Unix, Mac OSX και Windows και υποστηρίζει πολλές φορέα, ράστερ, και μορφές βάσεων δεδομένων και λειτουργικών δυνατοτήτων.

Το Quantum GIS παρέχει έναν συνεχώς αυξανόμενο αριθμό των δυνατοτήτων που παρέχονται από τις βασικές λειτουργίες και plugins. Μπορεί να απεικονίσει, διαχειριστεί, επεξεργαστεί, αναλύσει δεδομένα, και συνθέτει εκτυπώσιμους χάρτες.

Βασικά Χαρακτηριστικά

- Γραφικό περιβάλλον διεπαφής.
 - Αναγνώριση και επιλογή χαρακτηριστικών,
 - επεξεργασία/οπτικοποίηση/αναζήτηση περιγραφικών χαρακτηριστικών,
 - Άμεση αλλαγή προβολικού συστήματος,
 - συνθέτη εκτυπώσεων,
 - σύμβολα χαρακτηριστικών,
 - αλλαγές συμβόλων για διανυσματικά και εικονιστικά δεδομένα,
 - προσθήκη νέων επιπέδων,
 - και πολλά περισσότερα...
- Εύκολη προεπισκόπηση πολλών διανυσματικών και εικονιστικών προτύπων ψηφιακών αρχείων
 - Πίνακες της βάσης δεδομένων PostgreSQL
 - Υποστήριξη για τα περισσότερα διανυσματικά πρότυπα: περιλαμβανομένου των ESRI shapefiles, MapInfo, SDTS και GML.
 - Υποστήριξη για εικονιστικά δεδομένα όπως Ψηφιακά Μοντέλα Εδάφους, αεροφωτογραφίες και δορυφορικές εικόνες,
 - Υποστήριξη προτύπων του GRASS,
 - Υποστήριξη ανάγνωσης διαδικτυακών υπηρεσιών θέασης και μεταφόρτωσης του OGC (WMS ή WFS);

- Δημιουργία, επεξεργασία και εξαγωγή χωρικών δεδομένων με τη χρήση:
 - Εργαλείων ψηφιοποίησης του GRASS και το πρότυπο shapefile,
 - Του πρόσθετου γεωαναφοράς (plugin),
 - Εργαλεία GPS για την εισαγωγή και εξαγωγή αρχείων GPX, τον μετασχηματισμό από άλλα πρότυπα GPS σε GPX , ή η μεταφόρτωση απευθείας αρχείων σε δέκτη GPS
- Η εκτέλεση χωρικών αναλύσεων με τα πρόσθετα fTools και GRASS
 - Άλγεβρα χαρτών,
 - Ανάλυση τοπίου,
 - Υδρολογική μοντελοποίηση,
 - Ανάλυση δικτύων,
 - και πολλά περισσότερα
- Δημοσιοποίηση στο διαδίκτυο
- Αρχιτεκτονική με πρόσθετα (plugins).

4.3 JavaScript

Η JavaScript ήρθε περίπου ως μια κοινή προσπάθεια μεταξύ της Netscape Communications Corporation και Sun Microsystems. Το δελτίο τύπου της νέας γλώσσας ήρθε στις 4 Δεκεμβρίου 1995, όταν ο Netscape Navigator 2.0 ήταν ακόμα σε δοκιμαστική έκδοση. Η έκδοση 1.0 της JavaScript έγινε διαθέσιμη με το νέο πρόγραμμα περιήγησης.

Η **JavaScript** είναι μια γλώσσα συγγραφής σεναρίων (scripting language) που χρησιμοποιείται για να προσθέσει εφέ και διαλογικότητα στις ιστοσελίδες και είναι ανταγωνιστική της γλώσσας προγραμματισμού VBScript. Το αρχικό της όνομα ήταν LiveScript. Με την JavaScript μπορεί εύκολα να αναπτυχθεί μια interactive σελίδα (μια σελίδα δηλαδή που αλληλεπιδρά στον εαυτό της).

Ο κώδικας της JavaScript γράφεται σε καθαρό κείμενο (ASCII μορφή) και ενσωματώνεται μέσα στον κώδικα της HTML, μπορεί δε να εκτελεστεί αμέσως αρκεί μόνο ο browser να υποστηρίζει την JavaScript.

Η JavaScript είναι αντικειμενοστραφής, client-side scripting γλώσσα που χρησιμοποιείται για να κάνει τις ιστοσελίδες πιο δυναμικές. Για να έχει νόημα ένας τέτοιος ορισμός, ας ρίξουμε μια ματιά στα σημαντικά τμήματά του ένα προς ένα.

Object-based (αντικειμενοστραφής): αντικειμενοστραφής σημαίνει ότι η JavaScript μπορεί να χρησιμοποιήσει αντικείμενα. Ωστόσο, τα αντικείμενα δεν βασίζονται σε κλάσεις, αντιθέτως είναι απλά γενικά αντικείμενα.

Client Side: client side σημαίνει ότι το JavaScript τρέχει στον client (λογισμικό) που χρησιμοποιεί ο χρήστης, και όχι στο διακομιστή (Web server) της ιστοσελίδας. Σε αυτή την περίπτωση, ο client θα ήταν ένα πρόγραμμα περιήγησης στο διαδίκτυο. Για γίνει περισσότερο κατανοητό, περιγράφεται πώς λειτουργεί μία server-side γλώσσα και πώς μία client-side γλώσσα.

Server-side γλώσσες. Μία server-side γλώσσα παίρνει πληροφορίες από την ιστοσελίδα ή το πρόγραμμα περιήγησης στο διαδίκτυο, τις στέλνει σε ένα πρόγραμμα που εκτελείται στον διακομιστή (host server), και στη συνέχεια στέλνει τις πληροφορίες πίσω στο πρόγραμμα περιήγησης. Ως εκ τούτου, ένα ενδιάμεσο βήμα πρέπει να στείλει και να ανακτήσει πληροφορίες από το διακομιστή πριν εμφανιστούν τα αποτελέσματα στο πρόγραμμα περιήγησης.

Client-Side γλώσσες. Μια client-side γλώσσα τρέχει απευθείας μέσω του client που χρησιμοποιείται από το χρήστη. Στην περίπτωση της JavaScript, ο client είναι ένα πρόγραμμα περιήγησης στο διαδίκτυο. Ως εκ τούτου, η JavaScript εκτελείται απευθείας στο πρόγραμμα περιήγησης και δεν χρειάζεται να περάσει μέσα από το επιπλέον βήμα της αποστολής και ανάκτησης πληροφοριών από τον server. Με μια client side γλώσσα, το πρόγραμμα περιήγησης διαβάζει και ερμηνεύει τον κώδικα, και τα αποτελέσματα μπορούν να δοθούν στο χρήστη χωρίς της λήψη πληροφοριών από τον server πρώτα. Αυτή η διαδικασία βοηθάει ορισμένες εργασίες να τρέχουν πιο γρήγορα. Μια client-side γλώσσα μπορεί επίσης να έχει πρόσβαση σε ειδικά χαρακτηριστικά σε ένα παράθυρο περιήγησης που μπορεί να μην είναι προσβάσιμα με μία server side γλώσσα.

Scripting Language: Μια scripting γλώσσα δεν απαιτεί ένα πρόγραμμα που χρειάζεται compile (μεταγλώττιση) πριν από την εκτέλεσή του. Όλη η ερμηνεία γίνεται σε σύντομο χρονικό διάστημα από τον client. Με μία συνήθη γλώσσα προγραμματισμού, για την εκτέλεση ενός κώδικα, θα πρέπει να γίνει compile χρησιμοποιώντας έναν ειδικό compiler για να βεβαιωθούμε ότι δεν υπάρχουν συντακτικά λάθη. Με μία scripting γλώσσα, ο κώδικας ερμηνεύεται καθώς φορτώνεται στον client. Ωστόσο, τα λάθη δεν εντοπίζονται πριν τρέξει το script και αυτό θα μπορούσε να προκαλέσει προβλήματα με τον client, αν δεν μπορεί να χειριστεί καλά τα σφάλματα. Στην περίπτωση της JavaScript, η αντιμετώπιση των λαθών εξαρτάται από το πρόγραμμα περιήγησης που χρησιμοποιείται από τον χρήστη.

Συνθέτοντας όλα τα παραπάνω, καταλήγουμε στον τρόπο λειτουργίας της JavaScript. Η JavaScript τρέχει στο πρόγραμμα περιήγησης με την προσθήκη της σε ένα υπάρχον έγγραφο HTML (είτε άμεσα είτε με αναφορά σε ένα εξωτερικό script αρχείο). Μπορούμε να προσθέσουμε ειδικά tags και εντολές στον κώδικα HTML που θα οδηγήσει το πρόγραμμα περιήγησης στην εκτέλεση του script. Μόλις το πρόγραμμα περιήγησης δει αυτά τα ειδικά tags, ερμηνεύει τις εντολές JavaScript και κάνει ό, τι του λέει ο κώδικας. Έτσι, με την απλή επεξεργασία ενός εγγράφου HTML, μπορεί κανείς να αρχίσει να χρησιμοποιεί JavaScript στις ιστοσελίδες και να δει τα αποτελέσματα.

Η JavaScript διαθέτει τα παρακάτω πλεονεκτήματα:

- Δεν απαιτεί την αγορά ειδικού λογισμικού (εργαλείου ανάπτυξης).
- Δεν χρειάζεται μεταγλώττιση (compilation) και μεταγλωττιστές (compilers).
- Η συγγραφή του κώδικα είναι πολύ απλή υπόθεση. Δεν υπάρχουν ειδικές ρουτίνες εκκίνησης ή άλλες προαπαιτούμενες δομές.
- Δεν είναι μονοδιάστατη γλώσσα.

- Τα προγράμματα θα τρέχουν σε όλα τα λειτουργικά συστήματα.
- Αποτελεί διαχρονική αξία. Δεν τίθεται κανένα παράπονο για τις δυνατότητές της ως γλώσσας σε σχέση με το σκοπό που επιτελεί. Άλλες νεώτερες τεχνολογίες γεννιούνται βασιζόμενες στη ύπαρξή της (π.χ. DOM3).
- Η JavaScript έχει πολλά κοινά δομικά στοιχεία με άλλες δημοφιλείς γλώσσες όπως C, C#, Java, PHP, Perl κ.ά.
- Είναι η μοναδική γλώσσα που υποστηρίζουν όλοι οι browsers, και επομένως στο γήπεδο του client-side web scripting παίζει κυριολεκτικά χωρίς αντίπαλο.

4.3.1 Javascript και Java

Η Javascript δεν θα πρέπει να συγχέεται με τη Java. Η Java είναι διαφορετική γλώσσα προγραμματισμού και με διαφορετικές εφαρμογές. Η χρήση της λέξης «Java» στο όνομα της γλώσσας έχει περισσότερη σχέση με το προφίλ του προϊόντος που έπρεπε να έχει και λιγότερο με κάποια πιθανή συμβατότητα ή άλλη στενή σχέση με τη Java. Ρόλο σε αυτήν τη σύγχυση έπαιξε και ότι η Java και η Javascript έχουν δεχτεί σημαντικές επιρροές από τη γλώσσα C, ειδικά στο συντακτικό μέρος, ενώ είναι και οι δύο αντικειμενοστρεφείς γλώσσες. Τονίζεται ότι ο σωστός τρόπος γραφής της είναι «Javascript» και όχι «Java script» σαν δύο λέξεις, όπως λανθασμένα γράφεται ορισμένες φορές.

4.3.2 Ιστορία

Η γλώσσα προγραμματισμού JavaScript δημιουργήθηκε αρχικά από τον Brendan Eich της εταιρείας Netscape με την επωνυμία Mocha με σκοπό να ζωντανέψει το - τότε - στατικό Web. Αργότερα, η Mocha μετονομάστηκε σε LiveScript, και τελικά σε JavaScript, κυρίως επειδή η ανάπτυξή της επηρεάστηκε περισσότερο από τη γλώσσα προγραμματισμού Java. Πέρα από το όνομα δεν έχει ιδιαίτερες ομοιότητες με τη γλώσσα Java που παρουσιάστηκε από τη Sun την ίδια εποχή. LiveScript ήταν το επίσημο όνομα της γλώσσας όταν για πρώτη φορά κυκλοφόρησε στην αγορά τον Σεπτέμβριο του 1995. Η LiveScript μετονομάστηκε σε JavaScript στις 4 Δεκεμβρίου 1995.

Η JavaScript απέκτησε μεγάλη επιτυχία ως client-side γλώσσα για υλοποίηση κώδικα σε ιστοσελίδες, και χρησιμοποιήθηκε σε διάφορα προγράμματα περιήγησης στο διαδίκτυο. Πολύ σύντομα η Microsoft αναγνώρισε τις δυνατότητες που έδινε η JavaScript στον αντίπαλο browser και αποφάσισε να εφοδιάσει το δικό της με μία παρόμοια γλώσσα. Κατά συνέπεια, η εταιρεία Microsoft ονόμασε την εφαρμογή της σε JScript για να αποφύγει δύσκολα θέματα εμπορικών σημάτων. Η JScript περιλήφθηκε στο πρόγραμμα Internet Explorer εκδοχή 3.0, το οποίο κυκλοφόρησε τον Αύγουστο του 1996.

Τον Νοέμβριο του 1996, η Netscape ανακοίνωσε ότι είχε υποβάλει τη γλώσσα JavaScript στο Ecma International (μια οργάνωση της τυποποίησης των γλωσσών προγραμματισμού) για εξέταση ως βιομηχανικό πρότυπο, και στη συνέχεια το έργο είχε ως αποτέλεσμα την τυποποιημένη μορφή που ονομάζεται ECMAScript.

Η JavaScript έχει γίνει μία από τις πιο δημοφιλείς γλώσσες προγραμματισμού ηλεκτρονικών υπολογιστών στον Παγκόσμιο Ιστό. Αρχικά, όμως, πολλοί επαγγελματίες προγραμματιστές υποτίμησαν τη γλώσσα διότι, μεταξύ άλλων λόγων, το κοινό της ήταν ερασιτέχνες συγγραφείς ιστοσελίδων και όχι επαγγελματίες προγραμματιστές. Με τη χρήση της τεχνολογίας Ajax, η γλώσσα JavaScript επέστρεψε στο προσκήνιο και έφερε πιο επαγγελματική προσοχή προγραμματισμού. Το αποτέλεσμα ήταν έναν καινοτόμο αντίκτυπο στην εξάπλωση των βιβλιοθηκών, τη βελτίωση του προγραμματισμού με JavaScript, καθώς και την αυξημένη της χρήσης της JavaScript έξω από τα προγράμματα περιήγησης.

Τον Ιανουάριο του 2009, το CommonJS ιδρύθηκε με στόχο τον καθορισμό ενός κοινού προτύπου βιβλιοθήκης κυρίως για την ανάπτυξη της JavaScript έξω από το πρόγραμμα περιήγησης.

4.3.3 Τρέχοντας την JavaScript

Τι χρειάζεται ώστε να τρέξουν scripts γραμμένα σε JavaScript; Χρειάζεται ένας browser που υποστηρίζει JavaScript - παραδείγματος χάριν το Netscape Navigator (από την έκδοση 2.0) ή το Microsoft Internet Explorer (MSIE εν συντομία - από την έκδοση 3.0). Εφόσον αυτοί οι δύο browsers είναι πολύ διαδεδομένοι, πολλοί χρήστες έχουν την δυνατότητα να τρέξουν scripts γραμμένα σε JavaScript.

4.3.4 Εισάγοντας JavaScript σε μια HTML σελίδα

Χρησιμοποιώντας τις HTML ετικέτες (tags) Script. Τα script tags χρησιμοποιούνται για να πουν στο πρόγραμμα περιήγησης σε ποιο σημείο θα ξεκινήσει κάποιος τύπος scripting γλώσσας και σε ποιο σημείο θα τελειώσει σε ένα έγγραφο HTML. Στην πιο βασική τους μορφή, τα script tags φαίνονται ακριβώς όπως και κάθε άλλο σύνολο HTML tags:

```
<script>  
JavaScript κώδικας εδώ  
</ script>
```

Όπως μπορεί να δει κανείς, υπάρχει το <script> που ανοίγει το tag, ο κώδικας JavaScript, και στη συνέχεια, το κλείσιμο </ script> του tag. Όταν χρησιμοποιούνται μόνο τα βασικά tags ανοίγματος και κλεισίματος, όπως παραπάνω, πολλά προγράμματα περιήγησης υποθέτουν ότι η scripting γλώσσα που θα ακολουθήσει είναι η JavaScript. Ωστόσο, σε ορισμένα προγράμματα περιήγησης μπορεί να χρειαστεί να δηλωθεί ποια scripting γλώσσα χρησιμοποιείται. Εκτός από τη διάκριση τότε αρχίζει και τότε τελειώνει ένα script για το πρόγραμμα περιήγησης, τα script tags μπορούν επίσης να πουν στο πρόγραμμα περιήγησης ποια scripting γλώσσα θα χρησιμοποιηθεί και να καθορίσει τη διεύθυνση ενός εξωτερικού JavaScript αρχείου.

Όπως ήδη περιγράφηκε, ο κώδικας JavaScript εισάγεται απευθείας στην HTML σελίδα. Ακολουθεί ένα απλό παράδειγμα:

```
<html> <body>  
<script language="JavaScript">
```

```
document.write("Hello World with JavaScript");  
</script> </body> </html>
```

Για να δουλέψει αυτό το script, πρέπει να αποθηκευθεί ο παραπάνω κώδικας ένα απλό HTML έγγραφο και να ανοιχθεί στη συνέχεια σε ένα JavaScript-enabled browser. Το αποτέλεσμα που θα εμφανιστεί θα είναι:

Hello World with JavaScript

Οτιδήποτε μεταξύ <script> και </script> μεταφράζεται σαν JavaScript κώδικας.

4.3.5 Non-JavaScript browsers

Ένας non-JavaScript browser δεν ξέρει τι σημαίνει <script>. Αγνοεί το <script> και εμφανίζει όλο τον ακόλουθο κώδικα σαν κανονικό κείμενο. Αυτό σημαίνει ότι ο χρήστης θα δει τον Javascript κώδικα μέσα στο HTML έγγραφο.

4.3.6 Events

Κατά τη δημιουργία των scripts, συχνά διαπιστώνουμε ότι υπάρχουν «συμβάντα» (events) από τον χρήστη (όπως η μετακίνηση του ποντικιού πάνω σε ένα συγκεκριμένο στοιχείο ή ένα κλικ σε ένα συγκεκριμένο στοιχείο) στα οποία θέλουμε να αντιδράσει το script. Αυτό πραγματοποιείται μέσω της χρήσης του event handler (χειριστές συμβάντων).

Οι event handlers είναι χρήσιμοι, επειδή δίνουν τη δυνατότητα στο χρήστη να αποκτήσει πρόσβαση στα events που μπορεί να συμβούν στη σελίδα. Για παράδειγμα, αν θέλουμε να στείλουμε μια ειδοποίηση στο χρήστη όταν μετακινεί το ποντίκι πάνω σε ένα link, μπορούμε να χρησιμοποιήσουμε τον event-handler για να επικαλεσθεί την ειδοποίηση JavaScript που έχουμε κωδικοποιήσει για να αντιδράσει στο event. Με αυτόν τον τρόπο ό,τι συμβαίνει βασίζεται στις ενέργειες του χρήστη, το οποίο μας δίνει τη δυνατότητα να φτιάξουμε περισσότερο διαδραστικές ιστοσελίδες.

Για να δούμε πώς λειτουργούν οι event handlers, θα πρέπει να ξέρουμε πού μπορούμε να τους τοποθετήσουμε σε ένα έγγραφο και πώς να τους χρησιμοποιήσουμε για να προσθέσουμε τον JavaScript κώδικα για ένα event. Οι event handlers μπορούν να χρησιμοποιηθούν σε έναν αριθμό θέσεων. Μπορούν να χρησιμοποιηθούν απ' ευθείας εντός των HTML στοιχείων με την προσθήκη ειδικών χαρακτηριστικών στα στοιχεία αυτά. Μπορούν επίσης να χρησιμοποιηθούν εντός των <script> και </script> ετικετών ή σε ένα εξωτερικό αρχείο JavaScript.

Για να χρησιμοποιηθεί ένας event handler απευθείας σε ένα στοιχείο HTML, πρέπει να γνωρίζουμε τη λέξη-κλειδί για τον event handler και πού θα τοποθετηθεί μέσα στον κώδικα HTML. Ακολουθεί ένα παράδειγμα, στο οποίο συμβαίνει κάτι όταν ο χρήστης κάνει κλικ σε μία συγκεκριμένη περιοχή του εγγράφου. Ένα στοιχείο που μπορεί να πατηθεί κουμπί είναι μια φόρμα. Έτσι, ας υποθέσουμε ότι θέλουμε να ειδοποιηθεί ο χρήστης για κάτι, όταν κάνει κλικ σε ένα κουμπί της φόρμας. Ο κώδικας που αντιστοιχεί σε αυτή τη λειτουργία είναι:

```
<input type="button" value="Click Me!" onclick="JavaScript code here" />
```

Για να χρησιμοποιήσουμε έναν event handler, τον προσθέτουμε ως ένα επιπλέον χαρακτηριστικό σε μια ετικέτα HTML. Η μόνη διαφορά μεταξύ ενός event handler “attribute” και ένα HTML attribute είναι ότι μπορούμε να προσθέσουμε κώδικα JavaScript μέσα σε ένα event handler attribute.

4.4 KML

Το KML (Keyhole Markup Language) είναι μία μορφή αρχείου για την απεικόνιση γεωγραφικών πληροφοριών σε έναν Earth browser, όπως τα Google Earth, Google Maps και Google Maps για κινητά. Η KML χρησιμοποιεί μία δομή βασισμένη σε ετικέτες (tags) με ένθετα στοιχεία (elements) και χαρακτηριστικά (attributes) και είναι βασισμένη στο πρότυπο XML. Όλες οι ετικέτες είναι ευαίσθητες στη διάκριση πεζών-κεφαλαίων και πρέπει να εμφανίζονται ακριβώς όπως αναφέρονται στον ορισμό KML. Ο ορισμός δηλώνει ποιες ετικέτες είναι προαιρετικές. Μέσα σε ένα συγκεκριμένο στοιχείο, οι ετικέτες πρέπει να εμφανίζονται με τη σειρά που εμφανίζονται στον ορισμό.

4.4.1 Βασικά αρχεία KML

Το απλούστερο είδος KML αρχείων είναι αυτά που μπορούν να συνταχθούν άμεσα στο Google Earth, δηλαδή ο ενδιαφερόμενος δεν χρειάζεται να επεξεργαστεί ή να δημιουργήσει κάθε KML σε ένα πρόγραμμα επεξεργασίας κειμένου. Σύνταξη σημάνσεων μερών (Placemarks), επικαλύψεις εδάφους (ground overlays), διαδρομές (paths) και πολύγωνα (polygons) μπορούν όλα να συνταχθούν άμεσα στο Google Earth. Ένα παράδειγμα ενός KML αρχείου είναι το ακόλουθο:

```
<?xml version="1.0" encoding="UTF-8"?>
<kml xmlns="http://www.opengis.net/kml/2.2"
xmlns:gx="http://www.google.com/kml/ext/2.2"
xmlns:kml="http://www.opengis.net/kml/2.2"
xmlns:atom="http://www.w3.org/2005/Atom">
<Document>
  <name>test_simeio.kml</name>
  <Style id="sn_ylw-pushpin">
    <IconStyle>
      <scale>1.1</scale>
      <Icon>
        <href>http://maps.google.com/mapfiles/kml/
pushpin/ylw-pushpin.png</href>
      </Icon>
      <hotSpot x="20" y="2" xunits="pixels" yunits="pixels"/>
    </IconStyle>
  </Style>
  <StyleMap id="msn_ylw-pushpin">
    <Pair>
      <key>normal</key>
      <styleUrl>#sn_ylw-pushpin</styleUrl>
    </Pair>
    <Pair>
      <key>highlight</key>
```



```

        <styleUrl>#sh_ylw-pushpin</styleUrl>
    </Pair>
</StyleMap>
<Style id="sh_ylw-pushpin">
    <IconStyle>
        <scale>1.3</scale>
        <Icon>
            <href>http://maps.google.com/mapfiles/kml/
pushpin/ylw-pushpin.png</href>
        </Icon>
        <hotSpot x="20" y="2" xunits="pixels" yunits="pixels"/>
    </IconStyle>
</Style>
<Placemark>
    <name>Point 1</name>
    <LookAt>
        <longitude>-122.444743</longitude>
        <latitude>37.7526445</latitude>
        <altitude>0</altitude>
        <heading>0.2554718671115674</heading>
        <tilt>0</tilt>
        <range>193836.0400797448</range>
        <altitudeMode>relativeToGround</altitudeMode>
        <gx:altitudeMode>relativeToSeaFloor</gx:altitudeMode>
    </LookAt>
    <styleUrl>#msn_ylw-pushpin</styleUrl>
    <Point>
        <altitudeMode>clampToGround</altitudeMode>
        <gx:altitudeMode>clampToSeaFloor</gx:altitudeMode>
        <coordinates>-122.50678013203,37.769149792429,0</coordinates>
    </Point>
</Placemark>
<Placemark>
    <name>OUT</name>
    <styleUrl>#msn_ylw-pushpin</styleUrl>
    <Polygon>
        <tessellate>1</tessellate>
        <outerBoundaryIs>
            <LinearRing>
                <coordinates>
                    -122.50854,37.777099,0 -122.380946,37.777099,0 -
122.380946,37.72819,0 -122.50854,37.72819,0
                </coordinates>
            </LinearRing>
        </outerBoundaryIs>
    </Polygon>
</Placemark>
</Document>
</kml>

```

4.4.2 Συστατικά μέρη ενός KML αρχείου

Κάθε KML αρχείο ξεκινάει με την εξής δομή:

- Μία κεφαλίδα XML. Αυτή βρίσκεται στη γραμμή 1 σε κάθε KML αρχείο. Κανένα κενό ούτε άλλοι χαρακτήρες μπορούν να εμφανιστούν πριν από αυτή τη γραμμή.
- Μια δήλωση χώρου ονομάτων KML. Αυτό πραγματοποιείται στη γραμμή 2 σε κάθε KML αρχείο.

Σε ένα KML αρχείο μπορεί να εμφανίζονται τα εξής στοιχεία:

| | | | | | | | |
|----------|-----------------|-----------|------------|-----------------|-----------------|--|--|
| kml | Document | name | | | | | |
| | | style id | IconStyle | scale | | | |
| | | | | Icon | href | | |
| | | | | hotspot | | | |
| | | StyleMap | Pair | key | | | |
| | | | | StyleUrl | | | |
| | | Placemark | name | | | | |
| | | | | LookAt | longitude | | |
| | | | | | latitude | | |
| | | | | | altitude | | |
| | | | | | heading | | |
| | | | | | tilt | | |
| | | | | | range | | |
| | | | | | altitudeMode | | |
| | | | | | gx:altitudeMode | | |
| | | | | styleUrl | | | |
| | | Point | | altitudeMode | | | |
| | | | | gx:altitudeMode | | | |
| | | | | coordinates | | | |
| | | Placemark | name | | | | |
| styleUrl | | | | | | | |
| Polygon | tessellate | | | | | | |
| | outerBoundaryIs | | LinearRing | coordinates | | | |

Πίνακας 4.1: Συστατικά μέρη ενός KML αρχείου

Αναλυτικότερα τα βασικά στοιχεία που αποτελούν ένα KML αρχείο είναι τα εξής:

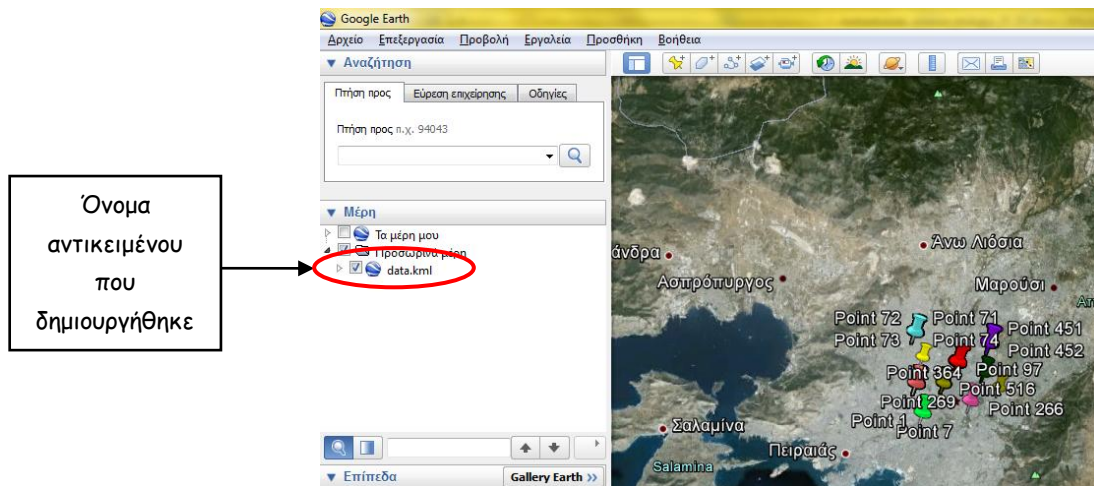
✓ **Έγγραφο – <Document>**

Ένα στοιχείο <Document> περιέχει ένα σύνολο από χαρακτηριστικά και στυλ. Το στοιχείο αυτό απαιτείται εάν το KML αρχείο χρησιμοποιεί κοινόχρηστα στυλ (*shared styles*). Ενδείκνυται η χρήση κοινόχρηστων στυλ, τα οποία απαιτούν τα παρακάτω βήματα:

- ❖ Προσδιορισμό όλων των στυλ μέσα σε ένα <Document>
- ❖ Απόδοση μοναδικής ταυτότητας ID για κάθε στυλ
- ❖ Εντός ενός δοσμένου στοιχείου Feature ή StyleMap, αναφορά στην ταυτότητα ID του στυλ χρησιμοποιώντας ένα <styleUrl> στοιχείο.

✓ **Όνομα - <name>**

Το πεδίο αυτό καθορίζεται από τον χρήστη και απεικονίζεται στην τρισδιάστατη απεικόνιση ως ετικέτα για το αντικείμενο (για παράδειγμα για ένα σημείο σήμανσης μέρους, ένα φάκελο ή ένα διαδικτυακό σύνδεσμο).



Εικόνα 4.3: Απεικόνιση αντικειμένου στο Google Earth και εμφάνιση ονόματος αντικειμένου

✓ Στυλ εικονιδίων - <IconStyle>

Καθορίζει πως απεικονίζονται τα εικονίδια για τα Placemarks. Το στοιχείο <Icon> καθορίζει την εμφάνιση του εικονιδίου. Το στοιχείο <scale> καθορίζει την κατά x και y κλίμακα του εικονιδίου. Το χρώμα που έχει ορίζεται στο πεδίο <color> του <IconStyle> αναμιγνύεται με το χρώμα του πεδίου <Icon>.

Στοιχεία σχετικά με το στυλ των εικονιδίων - <IconStyle>

<scale>

Επαναπροσδιορίζει το μέγεθος του εικονιδίου.



Εικόνα 4.4: Απεικόνιση εικονιδίων στο Google Earth με διαφορετικά μεγέθη

<heading>

Κατεύθυνση (Βοράς, Νότος, Ανατολή, Δύση), σε μοίρες. Από προεπιλογή είναι μηδέν (Βοράς). Η διακύμανση των τιμών είναι από 0 έως 360 μοίρες.

<Icon>

Στο <IconStyle>, το μόνο θυγατρικό στοιχείο του <Icon> είναι το <href>:

<href>

Εδώ βρίσκεται μία HTTP διεύθυνση ή ο εντοπισμός ενός τοπικού αρχείου που χρησιμοποιείται για να φορτωθεί ένα εικονίδιο (icon).

<hotSpot x="0.5" y="0.5" xunits="fraction" yunits="fraction">

Η γραμμή αυτή του κώδικα προσδιορίζει τη θέση στο εικονίδιο που είναι «συνδεδεμένη» με το στοιχείο <Point> που προσδιορίζεται στο <Placemark>. Οι τιμές x και y μπορούν να προσδιοριστούν με τρεις διαφορετικούς τρόπους: ως pixels, ως fractions (κλάσματα) του εικονιδίου, ή ως inset pixels, το οποίο είναι μία μετατόπιση σε pixels από την ανώτερη δεξιά γωνία του εικονιδίου. Οι x και y θέσεις μπορούν να προσδιοριστούν με διάφορους τρόπους – για παράδειγμα, το x μπορεί να είναι σε pixels και το y να είναι κλάσμα. Η αρχή του συστήματος συντεταγμένων βρίσκεται στην χαμηλότερη αριστερή γωνία του εικονιδίου.

<StyleMap>

Το στοιχείο <StyleMap> σχεδιάζει μεταξύ δύο διαφορετικών στυλ. Τυπικά ένα <StyleMap> στοιχείο χρησιμοποιείται για να δώσει ξεχωριστά φυσιολογικά και τονισμένα στυλ για μία σήμανση Placemark, έτσι ώστε η τονισμένη εκδοχή να εμφανίζεται όταν ο χρήστης τοποθετήσει τον κέρσορα του ποντικού επάνω στο εικονίδιο στο Google Earth.

<Pair>

Καθορίζει ένα ζευγάρι κλειδιού/τιμής το οποίο χαρτογραφεί μία λειτουργία (κανονική ή τονισμένη) στο προκαθορισμένο <styleUrl>. Το στοιχείο <Pair> περιέχει δύο στοιχεία, που είναι απαραίτητα και τα δύο:

- **<key>**. Αναγνωρίζει το κλειδί.
- **<styleUrl>** ή **<Style>**. Σχετίζεται με το στυλ. Στο <styleUrl>, για σχετικά στοιχεία στυλ που βρίσκονται τοπικά στο έγγραφο KML, χρησιμοποιείται ένα απλό σύμβολο αναφοράς #. Για στυλ που περιέχονται σε εξωτερικά αρχεία, χρησιμοποιείται ένα πλήρες URL μαζί με το σύμβολο #.

✓ Σήμανση μέρους - Placemarks

Τα Placemarks είναι μία από τις πιο συχνά χρησιμοποιούμενες δυνατότητες του Google Earth. Σηματοδοτούν μια θέση στην επιφάνεια της γης, χρησιμοποιώντας μία κίτρινη πινέζα ως εικονίδιο. Η απλούστερη σήμανση μέρους περιλαμβάνει μόνο ένα στοιχείο <point>, το οποίο καθορίζει τη θέση της σημάνσεως. Το Google Earth δίνει τη δυνατότητα στον χρήστη να καθορίσει ένα όνομα και ένα προσαρμοσμένο εικονίδιο για την σήμανση μέρους, και επίσης να προσθέσει και άλλα στοιχεία γεωμετρίας σε αυτό. Ανοίγοντας ένα αρχείο KML στο Google Earth, διακρίνουμε διαφορετικούς τύπους σημάνσεως μέρους.

Η δομή του placemark φαίνεται στο παράδειγμα kml που παρουσιάστηκε στην αρχή της ενότητας. Η δομή του αρχείου αυτού αναλύεται ως εξής:

- Ένα όνομα (<name>) που χρησιμοποιείται ως ετικέτα για την σήμανση μέρους.

- Μια περιγραφή (<description>) που εμφανίζεται στο παράθυρο "Βοήθειας" που επισυνάπτεται στο σημείο σήμανσης.
- Ένα σημείο (<Point>) που καθορίζει τη θέση του σημείου σήμανσεως στην επιφάνεια της γης. (γεωγραφικό μήκος- *longitude*, γεωγραφικό πλάτος- *latitude* και προαιρετικά υψόμετρο- *altitude*).

Ένα σημείο σήμανσης μέρους είναι ο μόνος τρόπος για να σχεδιαστεί ένα εικονίδιο και ετικέτα σε 3D Viewer του Google Earth. Από προεπιλογή, το εικονίδιο είναι η κίτρινη πινέζα. Στην KML, ένα <placemark> μπορεί να περιέχει ένα ή περισσότερα στοιχεία γεωμετρίας, όπως ένα LineString, ένα Πολύγωνο (Polygon), ή ένα μοντέλο (model). Αλλά μόνο ένα <Placemark> με ένα σημείο (Point) μπορεί να έχει εικονίδιο και ετικέτα. Το <Point> χρησιμοποιείται για να τοποθετηθεί το εικονίδιο, αλλά δεν υπάρχει καμία γραφική αναπαράσταση του σημείου του ίδιου.

<LookAt>

Καθορίζει μία εικονική κάμερα που σχετίζεται με οποιοδήποτε στοιχείο που προέρχεται από το πεδίο <Feature>. Το στοιχείο LookAt τοποθετεί την κάμερα σε σχέση με το αντικείμενο που παρατηρείται. Στο Google Earth η εικόνα "πετάει" σε αυτό το σημείο this LookAt όταν ο χρήστης πατήσει διπλό κλικ σε ένα αντικείμενο στο Places panel (παράθυρο μερών) ή σε ένα εικονίδιο στο 3D viewer (3D προβολή).

Στοιχεία σχετικά με το <LookAt>

<longitude>

Γεωγραφικό μήκος του σημείου από το οποίο κοιτάζει η κάμερα. Γωνιακή απόσταση σε μοίρες, ως προς τον Πρώτο Μεσημβρινό. Οι τιμές δυτικά του κεντρικού Μεσημβρινού κυμαίνονται από -180 έως 0 μοίρες. Οι τιμές ανατολικά του κεντρικού Μεσημβρινού κυμαίνονται από 0 έως 180 μοίρες.

<latitude>

Γεωγραφικό πλάτος του σημείου από το οποίο κοιτάζει η κάμερα. Σε μοίρες βόρεια και νότια του Ισημερινού (0 μοίρες). Οι τιμές κυμαίνονται από -90 έως 90 μοίρες.

<altitude>

Απόσταση από την επιφάνεια της γης σε μέτρα. Distance from the earth's surface, in meters. Ερμηνεύεται σύμφωνα με το <altitudeMode> που περιγράφεται παρακάτω.

<heading>

Κατεύθυνση (Βόρεια, Νότια, Ανατολικά, Δυτικά), σε μοίρες. Από προεπιλογή είναι 0 (Βόρεια). Οι τιμές κυμαίνονται από 0 έως 360 μοίρες.

<tilt>

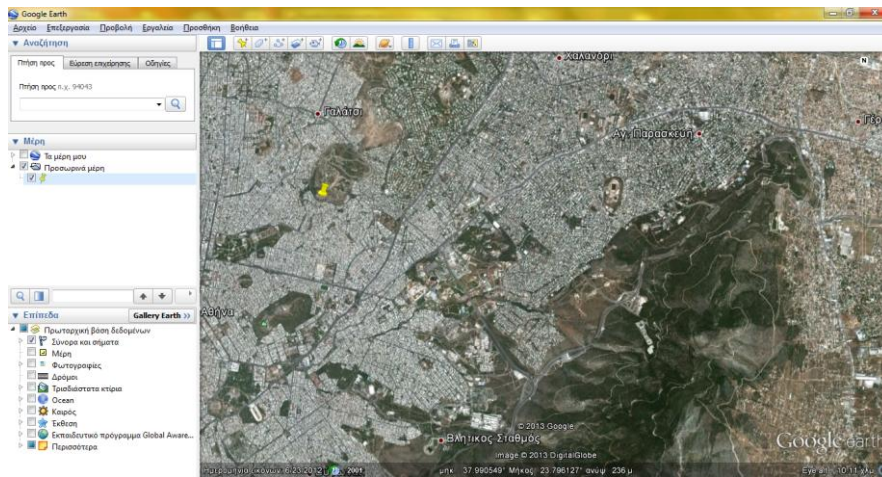
Η γωνία μεταξύ της κατεύθυνσης της LookAt θέσης και της κάθετης στην επιφάνεια της γης. Οι τιμές κυμαίνονται από 0 έως 90 μοίρες. Οι τιμές για το στοιχείο <tilt> δε μπορούν να είναι αρνητικές. Μία τιμή <tilt> των 0 μοιρών δείχνει την προβολή ακριβώς από πάνω. Μία τιμή <tilt> των 90 μοιρών δείχνει την προβολή κατά μήκος του ορίζοντα.



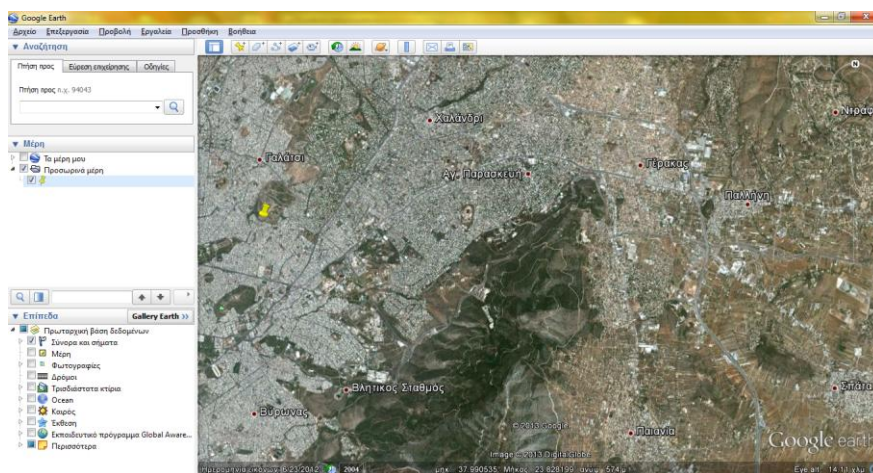
Εικόνα 4.5: Απεικόνιση της Αθήνας στο Google Earth με tilt=0 (αριστερά) και tilt=67 (δεξιά)

<range>

Απόσταση σε μέτρα από το σημείο που προσδιορίζεται από το γεωγραφικό μήκος, το γεωγραφικό πλάτος και το υψόμετρο στην LookAt θέση.



Εικόνα 4.6: Απεικόνιση της Αθήνας στο Google Earth με range=9530.85758459179



Εικόνα 4.7: Απεικόνιση της Αθήνας στο Google Earth με range=13530.85758459179

<altitudeMode>

Καθορίζει πως ερμηνεύεται το στοιχείο <altitude> που προσδιορίζεται για ο σημείο LookAt. Πιθανές τιμές είναι οι ακόλουθες:

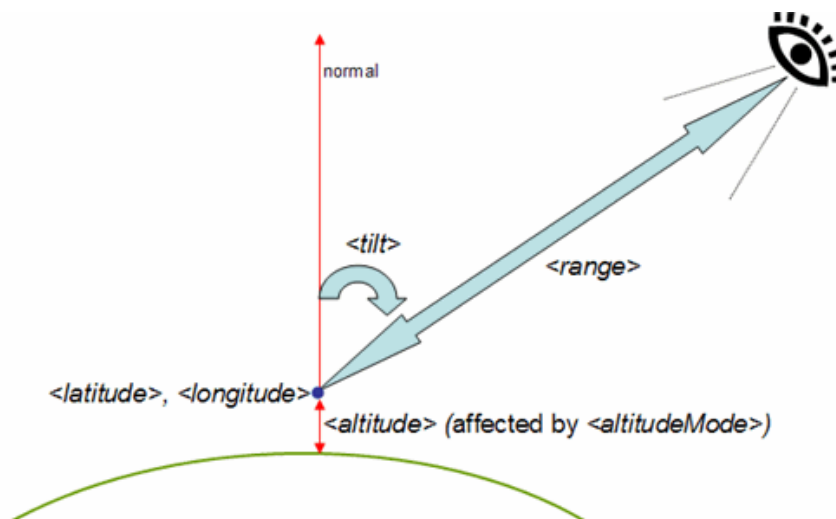
- **clampToGround** - (προεπιλογή) αγνοεί τον προσδιορισμό υψομέτρου και τοποθετεί τη θέση LookAt στο έδαφος.
- **relativeToGround** – ερμηνεύει το υψόμετρο ως μία τιμή σε μέτρα πάνω από το έδαφος.
- **absolute** – ερμηνεύει το υψόμετρο ως μία τιμή σε μέτρα πάνω από την επιφάνεια της θάλασσας.

<gx:altitudeMode>

Αποτελεί μία επέκταση του KML, που επιτρέπει υψόμετρα ως προς τον πυθμένα της θάλασσας. Οι πιθανές τιμές είναι:

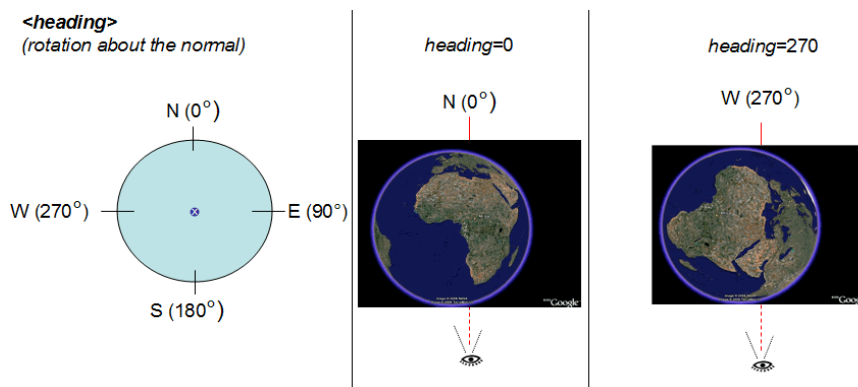
- **relativeToSeaFloor (ως προς τον πυθμένα της θάλασσας)** – ερμηνεύει το υψόμετρο ως μία τιμή σε μέτρα πάνω από τον πυθμένα της θάλασσας. Αν το σημείο είναι πάνω από ξηρά αντί για θάλασσα, το υψόμετρο θα ερμηνευθεί να είναι πάνω από το έδαφος.
- **clampToSeaFloor (κολλημένο στο βυθό της θάλασσας)** – ο προσδιορισμός υψομέτρου αγνοείται, και το σημείο θα τοποθετηθεί στον πυθμένα της θάλασσας. Αν το σημείο βρίσκεται σε ξηρά και όχι σε θάλασσα, το σημείο θα τοποθετηθεί στο έδαφος.

Το παρακάτω διάγραμμα απεικονίζει τα στοιχεία <range>, <tilt>, και <altitude>:



Εικόνα 4.8: Απεικόνιση των στοιχείων <range>, <tilt>, <altitude>

Το ακόλουθο διάγραμμα απεικονίζει το στοιχείο <heading>:



Εικόνα 4.9: Απεικόνιση στοιχείου <heading>

<Point>

Αποτελεί την γεωγραφική θέση, η οποία καθορίζεται από το γεωγραφικό μήκος και πλάτος και προαιρετικά το ύψος. Όταν ένα σημείο περιέχεται στο στοιχείο Placemark, το σημείο το ίδιο καθορίζει τη θέση του ονόματος και του εικονιδίου της σημάσεως Placemark. Όταν εξαγεται ένα σημείο, συνδέεται με το έδαφος με μία γραμμή. Αυτή η «πρόσδεση» χρησιμοποιεί το τρέχον στυλ γραμμής LineStyle.

Στοιχεία σχετικά με το στοιχείο <Point>

<altitudeMode>

Καθορίζει πως ερμηνεύονται τα συστατικά του υψομέτρου στο στοιχείο <coordinates>, όπως περιγράφηκε παραπάνω.

<coordinates>

Πρόκειται για ένα σύνολο που αποτελείται από τιμές κινητής υποδιαστολής του γεωγραφικού μήκους (*longitude*), του γεωγραφικού πλάτους (*latitude*) και του υψομέτρου (*altitude*) κατά σειρά. Το γεωγραφικό μήκος και πλάτος είναι σε μοίρες, όπου:

- $-180 \leq longitude \leq 180$
- $-90 \leq latitude \leq 90$
- *altitude* (προαιρετικά) σε μέτρα ως προς την επιφάνεια της θάλασσας.

<Polygon>

Ένα πολύγωνο καθορίζεται από ένα εξωτερικό όριο και κανένα ή περισσότερα εσωτερικά όρια. Τα όρια με τη σειρά τους καθορίζονται με το στοιχείο <LinearRings>. Όταν εξαγεται ένα πολύγωνο, τα όριά του συνδέονται στο έδαφος για να σχηματιστούν πρόσθετα πολύγωνα, γεγονός που του δίνει την εμφάνιση κτιρίου ή κουτιού. Τα εξαγόμενα πολύγωνα χρησιμοποιούν το στοιχείο <PolyStyle> για το χρώμα, τις χρωματικές λειτουργίες και το γέμισμά τους.

Το στοιχείο <coordinates> για τα πολύγωνα πρέπει να καθορισθεί με φορά αντίθετη του ρολογιού. Τα πολύγωνα ακολουθούν τον κανόνα του δεξιού χεριού, ο οποίος

ορίζει ότι αν τοποθετηθούν τα δάχτυλα του δεξιού χεριού προς την κατεύθυνση στην οποία καθορίζονται οι συντεταγμένες, ο αντίχειρας δείχνει στη γενική κατεύθυνση της γεωμετρικής καθέτου του πολυγώνου. Η Google Earth γεμίζει μόνο την εμπρόσθια επιφάνεια των πολυγώνων κι έτσι το επιθυμητό αποτέλεσμα επιτυγχάνεται μόνο όταν οι συντεταγμένες προσδιορίζονται με την σωστή σειρά. Διαφορετικά το πολύγωνο θα εμφανιστεί γκρι.

Στοιχεία σχετικά με το <Polygon>

<extrude>

Τελεστές boolean. Προσδιορίζει αν το σημείο θα συνδεθεί με το έδαφος με γραμμή. Για να εξαχθεί ένα σημείο, η τιμή του στοιχείου <altitudeMode> πρέπει να είναι είτε σε σχέση με το έδαφος (relativeToGround), είτε σε σχέση με τη στάθμη της θάλασσας (relativeToSeaFloor), ή απόλυτη (absolute). Το σημείο τοποθετείται με φορά προς το κέντρο της γήινης σφαίρας.

<tessellate>

Αυτό το πεδίο δε χρησιμοποιείται από το πολύγωνο. Για να ακολουθήσει ένα πολύγωνο το έδαφος ώστε να ενεργοποιηθεί η δημιουργία τρισδιάστατου μοντέλου με πολλά πολύγωνα τα οποία συνδέονται μεταξύ τους (tessellation), πρέπει να προσδιοριστεί μία υψομετρική λειτουργία clampToGround ή clampToSeaFloor.

<innerBoundaryIs>

Περιέχει ένα στοιχείο <LinearRing>. Ένα πολύγωνο μπορεί να περιέχει πολλαπλά <innerBoundaryIs> στοιχεία, τα οποία δημιουργούν πολλαπλές εγκοπές στο πολύγωνο.

<LinearRing>

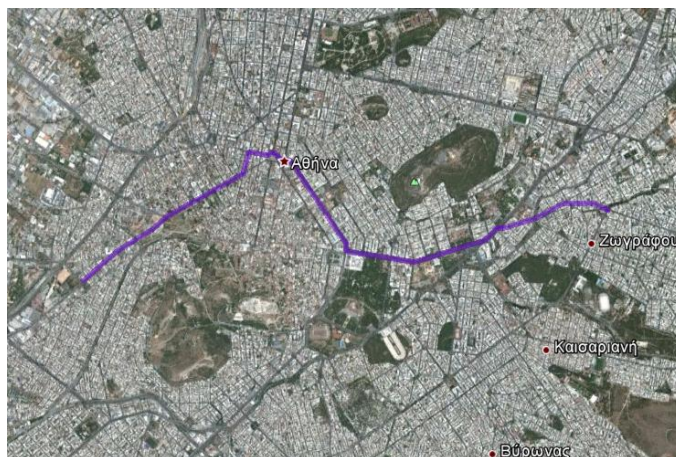
Καθορίζει μία κλειστή γραμμική χορδή, συνήθως το εξωτερικό όριο του πολυγώνου. Προαιρετικά, το στοιχείο αυτό μπορεί να χρησιμοποιηθεί ως το εσωτερικό όριο ενός πολυγώνου για να δημιουργηθούν τρύπες στο πολύγωνο. Ένα πολύγωνο μπορεί να περιέχει πολλαπλά <LinearRing> στοιχεία που χρησιμοποιούνται ως εσωτερικά όρια.

4.4.3 Διαδρομές - Paths

Πολλοί διαφορετικοί τύποι διαδρομών μπορούν να δημιουργηθούν στο Google Earth. Στο KML, δημιουργείται μια διαδρομή με ένα στοιχείο <LineString> .

Η <tessellate> ετικέτα διασπά τη γραμμή σε μικρότερα κομμάτια, και η <extrude> ετικέτα επεκτείνει τη γραμμή προς το έδαφος.

Ακολουθεί ένα παράδειγμα απεικόνισης διαδρομών στην ευρύτερη περιοχή της Αθήνας.



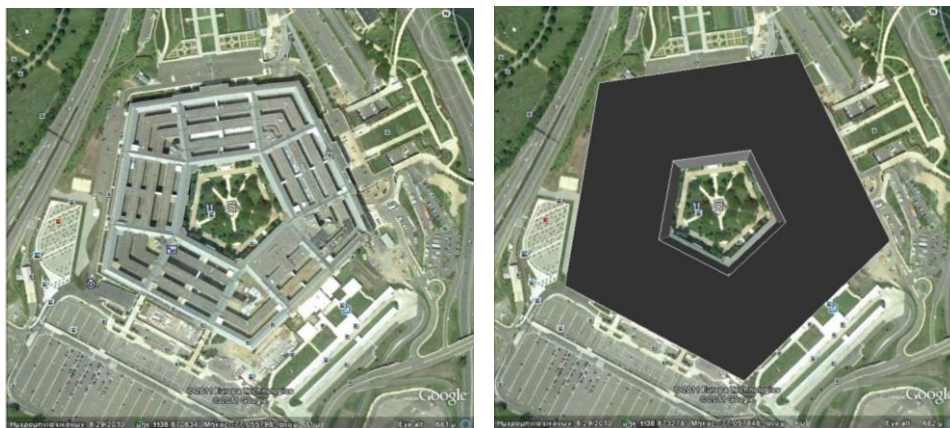
Εικόνα 4.10: Απεικόνιση διαδρομής στην περιοχή της Αθήνας στο Google Earth

4.4.4 Πολύγωνα - Polygons

Τα πολύγωνα είναι χρήσιμα για τη δημιουργία απλών κτιρίων και άλλων σχημάτων. Ακολουθεί ένα παράδειγμα.

```
<?xml version="1.0" encoding="UTF-8"?>
<kml xmlns="http://www.opengis.net/kml/2.2">
  <Placemark>
    <name>The Pentagon</name>
    <Polygon>
      <extrude>1</extrude>
      <altitudeMode>relativeToGround</altitudeMode>
      <outerBoundaryIs>
        <LinearRing>
          <coordinates>
            -77.05788457660967,38.87253259892824,100
            -77.05465973756702,38.87291016281703,100
            -77.05315536854791,38.87053267794386,100
            -77.05552622493516,38.868757801256,100
            -77.05844056290393,38.86996206506943,100
            -77.05788457660967,38.87253259892824,100
          </coordinates>
        </LinearRing>
      </outerBoundaryIs>
      <innerBoundaryIs>
        <LinearRing>
          <coordinates>
            -77.05668055019126,38.87154239798456,100
            -77.05542625960818,38.87167890344077,100
            -77.05485125901024,38.87076535397792,100
            -77.05577677433152,38.87008686581446,100
            -77.05691162017543,38.87054446963351,100
            -77.05668055019126,38.87154239798456,100
          </coordinates>
        </LinearRing>
      </innerBoundaryIs>
    </Polygon>
  </Placemark>
</kml>
```

Ακολουθεί η απεικόνιση του παραπάνω παραδείγματος στο Google Earth.



Εικόνα 4.11: Απεικόνιση Πενταγώνου στο Google Earth πριν (αριστερά) και μετά (δεξιά) την δημιουργία πολυγώνου

4.4.5 Στυλ για Γεωμετρία

Το στυλ είναι ένα σημαντικό μέρος στον τρόπο εμφάνισης των δεδομένων. Ορίζοντας ένα στυλ στην αρχή ενός εγγράφου KML και επίσης καθορίζοντας μία ταυτότητα ID για αυτό, μπορεί ο χρήστης να χρησιμοποιήσει αυτό το στυλ στη γεωμετρία, τις σημάνσεις μερών και τις επικαλύψεις που έχουν οριστεί αλλού στο έγγραφο. Επειδή περισσότερα από ένα στοιχεία μπορούν να χρησιμοποιήσουν το ίδιο στυλ, στυλ που ορίζονται και χρησιμοποιούνται με αυτόν τον τρόπο αναφέρονται ως κοινόχρηστα στυλ. Ο χρήστης ορίζει ένα δεδομένο στυλ μία φορά, και στη συνέχεια κάνει αναφορά σε αυτό πολλές φορές, χρησιμοποιώντας το στοιχείο `<styleUrl>`. Εάν ο ορισμός στυλ είναι μέσα στο ίδιο αρχείο, προηγείται το Αναγνωριστικό στυλ ID με ένα σύμβολο #. Εάν ο ορισμός στυλ είναι σε ένα εξωτερικό αρχείο, περιλαμβάνει την πλήρη διεύθυνση στο στοιχείο `<styleUrl>`.

Εδώ πρέπει να σημειωθεί ότι το στοιχείο `<Style>` είναι θυγατρικό του `<Placemark>`.

5

ΥΛΟΠΟΙΗΣΗ ΤΟΥ ΣΥΣΤΗΜΑΤΟΣ

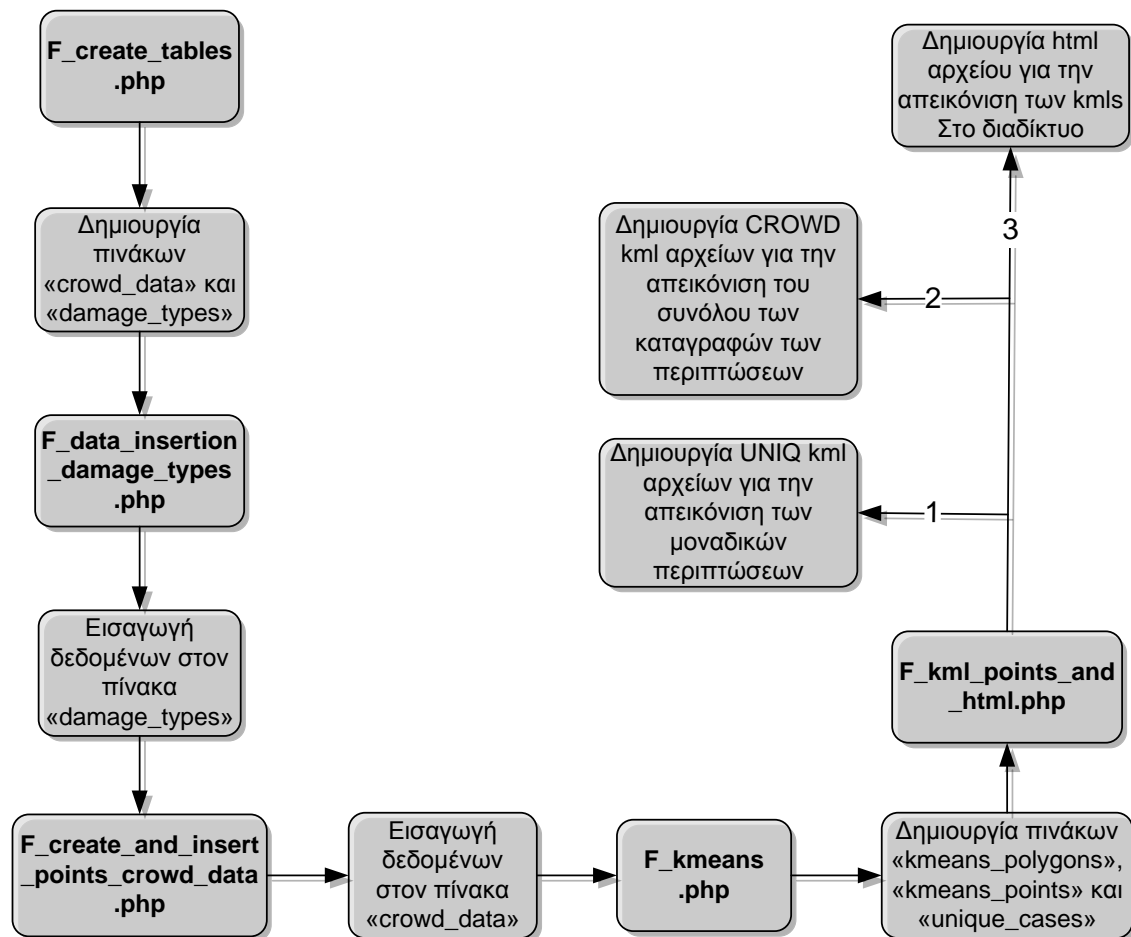
Στο κεφάλαιο αυτό αναλύεται το στάδιο της υλοποίησης του συστήματος που περιλαμβάνει τη συλλογή, αποθήκευση και επεξεργασία των καταγεγραμμένων δεδομένων των προβλημάτων. Αφότου ολοκληρωθεί η διαδικασία της συλλογής των καταγραφών με τη βοήθεια του crowdsourcing, σειρά έχει η υλοποίηση ενός συστήματος για την συγκέντρωση και αξιοποίηση της διαθέσιμης πληροφορίας.

Η υλοποίηση του συστήματος πραγματοποιήθηκε με χρήση της γλώσσας προγραμματισμού PHP, περιλαμβάνοντας ωστόσο και τμήματα κώδικα σε γλώσσα SQL και javascript. Συγκροτήθηκαν πέντε php αρχεία καθένα εκ των οποίων διατελεί και μία διαφορετική λειτουργία. Ο κώδικας βρίσκεται αυτούσιος στο παράρτημα της εργασίας. Ο αλγόριθμος που υλοποιήθηκε εκτελεί τις ακόλουθες εργασίες:

- Δημιουργία των πινάκων «`damage_types`» και «`crowd_data`» στη βάση δεδομένων με υλοποίηση του αρχείου **F_create_tables.php**.
- Εισαγωγή των δεδομένων στους πίνακες «`damage_types`» και «`crowd_data`» με υλοποίηση των αρχείων **F_data_insertion_damage_types.php** και **F_create_and_insert_points_crowd_data.php** αντίστοιχα.
- Εφαρμογή του αλγορίθμου k-means για την ομαδοποίηση των δεδομένων και επεξεργασία των δεδομένων και δημιουργία νέων πινάκων με υλοποίηση του αρχείου **F_kmeans.php**.
- Σύνδεση της βάσης δεδομένων με το λογισμικό Quantum GIS για την απεικόνιση των αποτελεσμάτων.
- Δημιουργία kml αρχείων για την απεικόνιση των δεδομένων στο Google Maps και το Διαδίκτυο με υλοποίηση του αρχείου **F_kml_points_and_html.php**.

Τα στάδια υλοποίησης του αλγορίθμου αναλύονται διεξοδικά στις επόμενες ενότητες. Όλες οι διεργασίες που διατελούνται σε σχέση με τη βάση δεδομένων, υλοποιούνται με κώδικα σε γλώσσα SQL η οποία ωστόσο έχει ενσωματωθεί στον τελικό κώδικα σε PHP.

Στο διάγραμμα που ακολουθεί παρουσιάζεται συνοπτικά η ροή εκτέλεσης των πέντε php αρχείων και τα αποτελέσματα που προέκυψαν από την υλοποίηση κάθε αρχείου.



Εικόνα 5.1: Ροή υλοποίησης των php αρχείων και τα αποτελέσματα που προέκυψαν από το καθένα

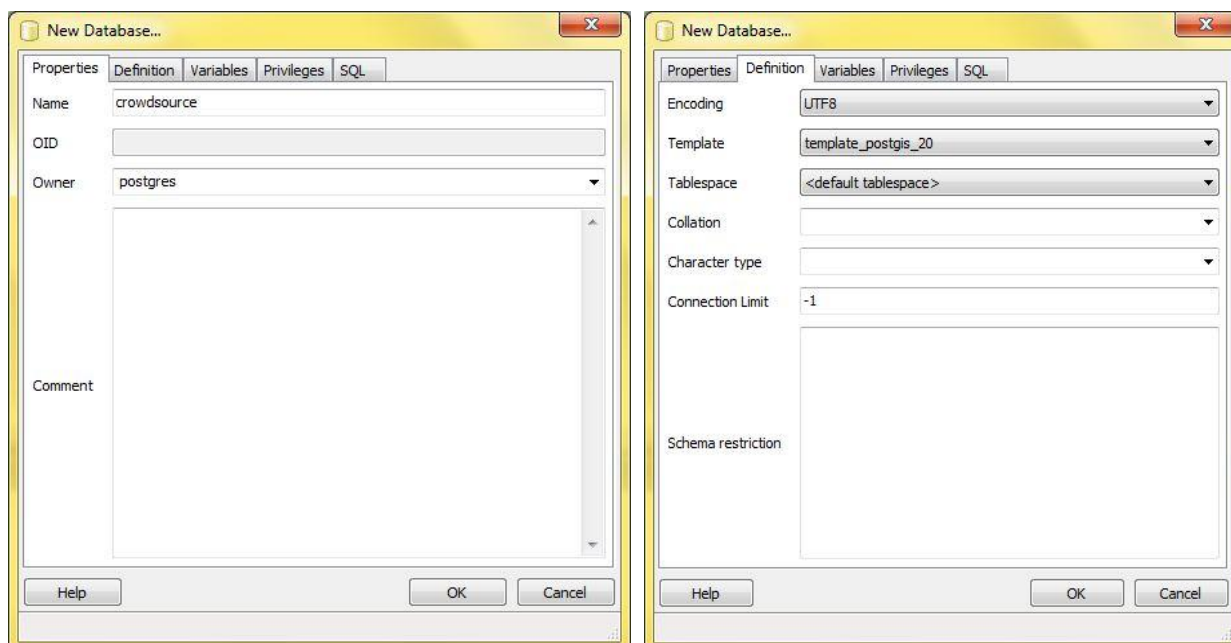
5.1 Βάση Δεδομένων

Το πρώτο στάδιο της υλοποίησης της εφαρμογής αποτέλεσε η δημιουργία μιας βάσης δεδομένων στο λογισμικό ανοικτού κώδικα PostgreSQL με σκοπό την καταχώρηση και αποθήκευση των καταγεγραμμένων προβλημάτων και των συναφών πληροφοριών που περιγράφουν το εκάστοτε πρόβλημα.

5.1.1 Σύνδεση με την PostgreSQL - Δημιουργία νέας βάσης δεδομένων

Για τη δημιουργία της βάσης δεδομένων είναι απαραίτητη η πραγματοποίηση σύνδεσης με το λογισμικό της PostgreSQL. Επιλέγεται διαδοχικά Start → Programs → PostgreSQL 9.1 → pgAdmin III. Με δεξί κλικ στο PostgreSQL Database Server 9.1 επιλέγουμε Connect και δίνουμε ως username: postgres και ως password αυτό που ορίστηκε κατά την εγκατάσταση.

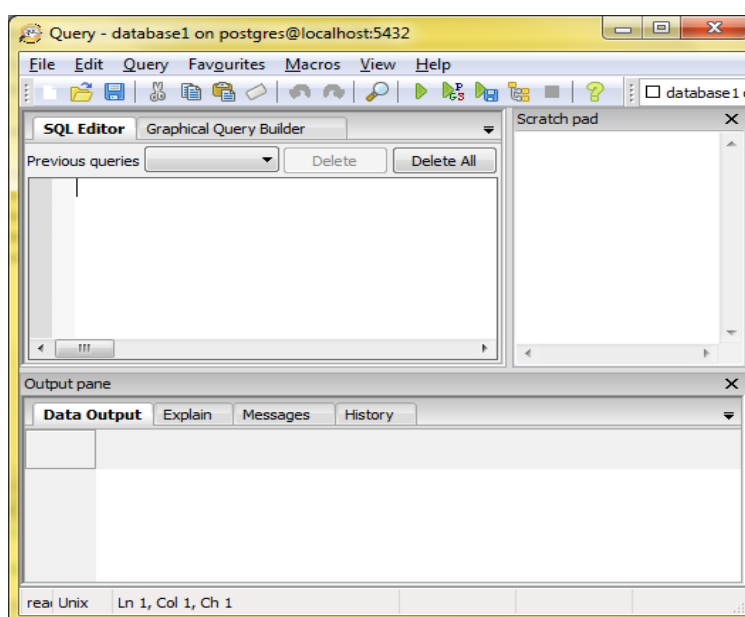
Η δημιουργία μιας νέας βάσης δεδομένων με την ονομασία crowdsourcing υλοποιείται σε παραθυρικό περιβάλλον δίνοντας τις παραμέτρους που φαίνονται παρακάτω:



Εικόνα 5.2: Δημιουργία νέας βάσης δεδομένων

Η βάση δεδομένων θα πρέπει να βασίζεται στο πρότυπο `template_postgis_20` και να ακολουθεί την κωδικοποίηση UTF-8, ώστε να λειτουργούν οι συναρτήσεις του PostGIS. Εάν δεν επιλεγεί ως `template` το `template_postgis_20`, παρατηρείται ότι δεν δημιουργούνται οι πίνακες `geometry_columns` και `spatial reference system` στο σχήμα και κατ' επέκταση δεν μπορεί να εισαχθεί το πεδίο της γεωμετρίας των δεδομένων, σε επόμενο στάδιο.

Στη συνέχεια από το μενού `Tools` → `Query tool` ανοίγει το παράθυρο δημιουργίας των ερωτημάτων. Στο υποπαράθυρο `SQL Editor` υλοποιήθηκε ο κώδικας για την υποβολή των ερωτημάτων, ενώ από το μενού `Query` → `Execute` (F5) εκτελούνται τα ερωτήματα. Όλα τα ερωτήματα που αφορούν στη βάση δεδομένων υλοποιήθηκαν αρχικά στο παράθυρο `Query Tool` και στη συνέχεια ενσωματώθηκαν στον τελικό κώδικα PHP.



Εικόνα 5.3: Παράθυρο δημιουργίας των SQL ερωτημάτων

5.1.2 Επιλογή των κατηγοριών δεδομένων

Τα είδη των δεδομένων που συλλέγονται, όπως περιγράφηκε και σε προηγούμενη ενότητα, διαμορφώθηκαν με βάση το θέμα της εφαρμογής και χωρίζονται σε δύο κατηγορίες:

- ✓ Προσωπικά δεδομένα και
- ✓ Πληροφορίες προβλήματος

Στην πρώτη κατηγορία καταγράφονται τα στοιχεία επικοινωνίας του ατόμου, ενώ στη δεύτερη κατηγορία όλες οι απαραίτητες πληροφορίες για τον προσδιορισμό του προβλήματος και της επικινδυνότητας αυτού, έτσι ώστε ο διαχειριστής της εφαρμογής να μπορεί να εκτιμήσει και να ιεραρχήσει τα προβλήματα ως προς τη σοβαρότητα αυτών.

Οι δύο παραπάνω βασικές κατηγορίες δεδομένων χωρίζονται σε πολλές υποκατηγορίες και πιο συγκεκριμένα:

- ✓ Προσωπικά δεδομένα και
 - Όνομα
 - Επίθετο
 - Τηλέφωνο
 - e-mail
 - Πόλη
- ✓ Πληροφορίες προβλήματος
 - Τύπος προβλήματος
 - Συνοπτική περιγραφή προβλήματος
 - Ημερομηνία προβλήματος
 - Αξιολόγηση προβλήματος
 - Φωτογραφία (προαιρετικά)

Με βάση τις προαναφερθείσες κατηγορίες δεδομένων θα δημιουργηθούν στη συνέχεια τα πεδία συμπλήρωσης δεδομένων στην εφαρμογή Fulcrum καθώς επίσης και οι πίνακες στη βάση δεδομένων στους οποίους θα καταχωρηθούν τα δεδομένα που θα συλλεχθούν.

5.1.3 Δημιουργία πινάκων

Οι πίνακες της βάσης δεδομένων που επρόκειτο να κατασκευαστούν σχεδιάστηκαν με βάση τις κατηγορίες δεδομένων που ορίστηκαν νωρίτερα. Από την εξαγωγή των συλλεγμένων δεδομένων στην εφαρμογή του Fulcrum προκύπτει ένα csv αρχείο με πεδία τις αντίστοιχες κατηγορίες δεδομένων που ορίστηκαν. Το csv αρχείο διαθέτει τόσα πεδία όσα και τα πεδία που συμπληρώνει ο χρήστης στην εφαρμογή Fulcrum από το κινητό του τηλέφωνο για να καταγράψει ένα πρόβλημα, και επιπλέον έξι πεδία τα οποία αναφέρονται στις γεωγραφικές συντεταγμένες της θέσης του προβλήματος, την ημερομηνία και ώρα της δημιουργίας της αναφοράς, την ημερομηνία και ώρα της ενημέρωσης της αναφοράς, και το χρήστη από τον οποίο δημιουργήθηκε και ενημερώθηκε η αναφορά. Το csv αρχείο περιλαμβάνει τα ακόλουθα πεδία:

- | | |
|-----------------------------------|----------------------------|
| ➤ Latitude | ➤ Ημ/νία Προβλήματος |
| ➤ Longitude | ➤ Αξιολόγηση προβλήματος |
| ➤ Όνομα | ➤ Φωτογραφία (προαιρετικά) |
| ➤ Επίθετο | ➤ Caption |
| ➤ Τηλέφωνο | ➤ Updated |
| ➤ e_mail | ➤ Created |
| ➤ Πόλη | ➤ Project |
| ➤ Τύπος προβλήματος | ➤ created_by |
| ➤ Συνοπτική περιγραφή προβλήματος | ➤ updated_by |

Με βάση το csv αρχείο που θα εξαχεται για την παραλαβή των δεδομένων, σχεδιάστηκε η δομή της βάσης δεδομένων. Αρχικά δημιουργήθηκε μία βάση δεδομένων με το όνομα «crowdsourcing» εντός της οποίας υλοποιήθηκαν δύο πίνακες οι οποίοι είναι:

- **crowd_data.** Στον πίνακα «crowd_data» θα καταχωρούνται όλα τα δεδομένα αυτούσια όπως παραλαμβάνονται από το csv αρχείο με τα αντίστοιχα πεδία, με την προσθήκη τριών ακόμη στηλών: ενός πρωτεύοντος κλειδιού id, της γεωμετρίας geom που θα προκύπτει από τις γεωγραφικές συντεταγμένες του σημείου και θα του προσδίδει γεωμετρικά χαρακτηριστικά, και ενός ξένου κλειδιού (unique_cases_id) το οποίο θα συμπληρώνεται στη συνέχεια και θα καθορίζει την ταξινόμηση των σημείων με βάση τον αλγόριθμο k-means. Οι ονομασίες των στηλών του πίνακα «crowd_data» είναι οι παρακάτω:

- | | |
|---------------------|-------------------|
| ➤ id | ➤ date |
| ➤ latitude | ➤ severity |
| ➤ longitude | ➤ capture |
| ➤ name | ➤ updated |
| ➤ surname | ➤ created |
| ➤ phone_number | ➤ project |
| ➤ e_mail | ➤ created_by |
| ➤ city | ➤ updated_by |
| ➤ type_of_damage | ➤ geom. |
| ➤ short_description | ➤ Unique_cases_id |

- **damage_types.** Στον πίνακα «damage_types» θα καταγράφονται όλοι οι πιθανοί τύποι προβλημάτων έτσι όπως ορίζονται κατά τη δημιουργία της φόρμας καταγραφής προβλημάτων στην εφαρμογή Fulcrum. Ο πίνακας αυτός περιλαμβάνει οκτώ στήλες:

- | | |
|---------------------|--------------------|
| ➤ t_code | ➤ kml_crowd_colour |
| ➤ description | ➤ styleurl_unique |
| ➤ colour | ➤ styleurl_crowd |
| ➤ kml_unique_colour | ➤ icon |

Το πεδίο «t_code» το οποίο είναι και το πρωτεύον κλειδί του πίνακα αντιστοιχεί στον κωδικό αριθμό κάθε προβλήματος, ενώ το πεδίο «description» αντιστοιχεί στην ονομασία του προβλήματος. Το πεδίο «colour» αναφέρεται στο χρώμα που θα αντιστοιχεί κάθε τύπος προβλήματος κατά την απεικόνιση των αποτελεσμάτων σε χάρτες. Τα πεδία «kml_unique_colour» και

«kml_crowd_colour» καθορίζουν το χρώμα και το στυλ των εικονιδίων που θα αντιπροσωπεύουν κάθε πρόβλημα στους χάρτες, ενώ τα πεδία «styleurl_unique» και «styleurl_crowd» προσδίδουν την ιδιότητα του χρώματος σε κάθε ένα από τα σημεία των δεδομένων για τον τύπο προβλήματος που αντιστοιχούν. Τέλος, το πεδίο «icon» είναι προαιρετικό και αντιστοιχεί κάθε τύπο προβλήματος σε μία εικόνα σχετική με το πρόβλημα.

Στη συνέχεια δημιουργούνται άλλοι τρεις πίνακες με βάση τα αρχικά δεδομένα και τον υπολογισμό του k-means. Οι πίνακες αυτοί σχηματίζονται με τη σειρά που περιγράφονται στη συνέχεια:

- **kmeans_polygons.** Στον πίνακα αυτό υπολογίζεται η γεωμετρία k πολυγώνων με βάση την ομαδοποίηση k-means, καθώς επίσης και τα κεντροειδή των πολυγώνων αυτών.
- **kmeans_points.** Στον πίνακα «kmeans_points» υπολογίζονται οι τιμές k-means που αντιστοιχούν σε κάθε μία καταγραφή. Ο υπολογισμός των τιμών k-means πραγματοποιείται με την αξιοποίηση των συντεταγμένων των κεντροειδών των πολυγώνων ως μίας εκ των παραμέτρων εισόδου του αλγορίθμου k-means.
- **unique_cases.** Ο πίνακας «unique_cases» προκύπτει από τον πίνακα «kmeans_points» ως αποτέλεσμα της ομαδοποίησης των σημείων με ίδια τιμή k-means. Εδώ εισάγονται οι μοναδικές τιμές k-means που προέκυψαν σε αντιστοιχία με τη μέση τιμή των σημείων με ίδια τιμή k-means, δηλαδή των σημείων που ανήκουν στην ίδια ταξινόμηση.

Οι τρεις τελευταίοι πίνακες περιγράφονται αναλυτικότερα σε επόμενες ενότητες. Οι τύποι δεδομένων που χρησιμοποιήθηκαν κατά τη δημιουργία των πινάκων είναι:

Character varying. Συμβολοσειρά μεταβλητού αριθμού χαρακτήρων.

Integer. Ακέραιος αριθμός 4-byte.

Bigint. Ακέραιος αριθμός 8-byte.

Double precision. Δεκαδικός αριθμός διπλής ακρίβειας (8-bytes).

Text. Συμβολοσειρά μεταβλητού αριθμού χαρακτήρων.

Serial. Ακέραιος αριθμός με αυτόματη αρίθμηση.

Numeric. Ακριβής αριθμός με επιλεγμένη ακρίβεια.

Date. Ημερομηνία

Στον πίνακα που ακολουθεί παρουσιάζονται οι πίνακες «crowd_data» και «damage_types» οι οποίοι δημιουργούνται σε αυτό το στάδιο στη βάση δεδομένων, το όνομα των στηλών, οι τύποι δεδομένων κάθε στήλης, καθώς επίσης και το πρωτεύον και τα δευτερεύοντα κλειδιά. Οι υπόλοιποι τρεις πίνακες παρουσιάζονται στις επόμενες ενότητες, στο στάδιο που δημιουργείται ο καθένας.

| Πίνακας | Στήλη | Τύπος Δεδομένων | Πρωτεύον κλειδί | Ξένο κλειδί |
|---------------------|-------------------|--------------------|--------------------|--|
| crowd_data | id | serial | ✓ | |
| | latitude | double precision | | |
| | longitude | double precision | | |
| | name | varchar | | |
| | surname | varchar | | |
| | phone_number | varchar | | |
| | e_mail | varchar | | |
| | city | varchar | | |
| | type_of_damage | integer not Null | | REFERENCES damage_types (t_code) |
| | short_description | Varchar | | |
| | date | date | | |
| | severity | integer | | |
| | capture | varchar | | |
| | updated | varchar | | |
| | created | varchar | | |
| | project | varchar | | |
| | created_by | varchar | | |
| updated_by | varchar | | | |
| geom. | geometry | | | |
| unique_cases_id | integer | | | |
| damage_types | t_code | serial | ✓ | |
| | description | double precision | | |
| | colour | double precision | | |
| | kml_unique_colour | varchar | | |
| | kml_crowd_colour | varchar | | |
| | styleurl_unique | varchar | | |
| | styleurl_crowd | varchar | | |
| | icon | varchar | | |

Πίνακας 5.1: Οι αρχικοί πίνακες (crowd_data και damage_types) της βάσης δεδομένων, οι στήλες, οι τύποι δεδομένων και τα κλειδιά τους

Εφόσον δημιουργήθηκαν οι δύο πρώτοι πίνακες «crowd_data» και «damage_types», σειρά έχει η καταχώρηση των δεδομένων στους δύο πίνακες και η επεξεργασία αυτών για τη δημιουργία νέων πινάκων δεδομένων.

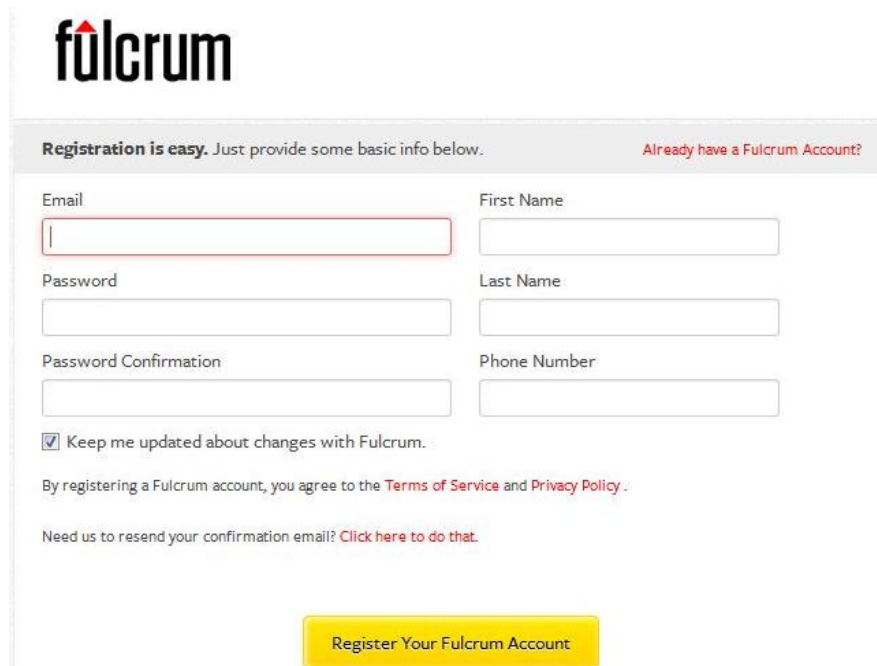
5.2 Ροή Συλλογής και Επεξεργασίας Δεδομένων

Το στάδιο αυτό αποτελεί τη φάση υλοποίησης των καταγραφών από το πλήθος, καθώς επίσης και τη συλλογή και επεξεργασία των δεδομένων αυτών. Στο σημείο αυτό υπενθυμίζεται ότι η καταγραφή των προβλημάτων πραγματοποιείται ενδεικτικά για ένα μικρό αριθμό προβλημάτων με χρήση της εφαρμογής Fulcrum App, ενώ το σύνολο των δεδομένων κατασκευάζεται με πρότυπο τη δομή αυτών των δεδομένων που εξήχθησαν από το Fulcrum.

5.2.1 Σύνδεση με το Fulcrum

Για να εφαρμοστεί crowdsourcing με τη βοήθεια της εφαρμογής Fulcrum App, αρχικά ήταν απαραίτητο να πραγματοποιηθούν οι ακόλουθες διεργασίες:

- Εγγραφή στο Fulcrum App με τη δημιουργία προσωπικού λογαριασμού στην επίσημη ιστοσελίδα της εφαρμογής (<http://fulcrumapp.com/>). Κατά την εγγραφή και τη σύνδεση απαιτείται η εισαγωγή email και κωδικού που ορίζει ο χρήστης.
- Χρήση φορητής συσκευής iPhone, iPad ή Android.
- Εγκατάσταση της εφαρμογής Fulcrum App στην κινητή συσκευή.

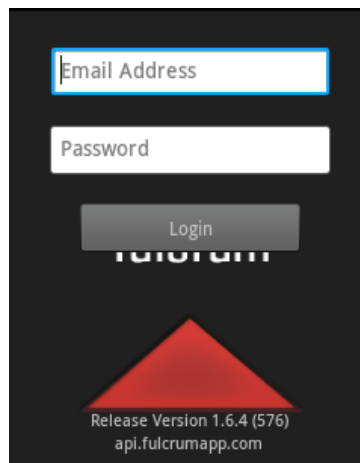


The image shows a registration form for Fulcrum. At the top left is the Fulcrum logo. Below it, the text reads "Registration is easy. Just provide some basic info below." and "Already have a Fulcrum Account?". The form contains six input fields: Email, First Name, Password, Last Name, Password Confirmation, and Phone Number. There is a checkbox labeled "Keep me updated about changes with Fulcrum." and a link "Click here to do that." below it. At the bottom center is a yellow button that says "Register Your Fulcrum Account".

Εικόνα 5.4: Δημιουργία προσωπικού λογαριασμού στην ιστοσελίδα του Fulcrum

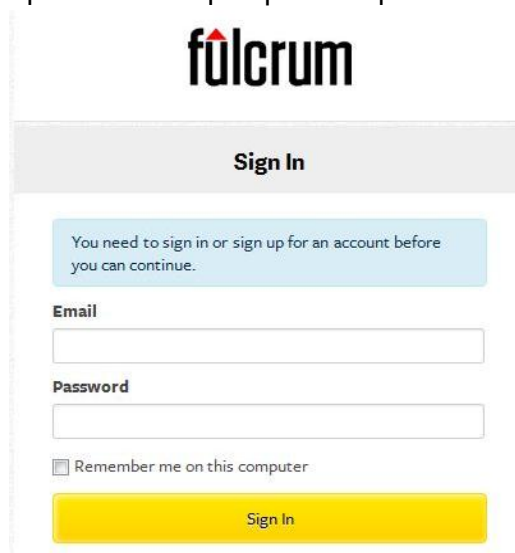
Μετά την ολοκλήρωση των διεργασιών που περιγράφηκαν, δημιουργήθηκε στην ιστοσελίδα του Fulcrum μία φόρμα με διάφορα πεδία δεδομένων με βάση την οποία θα συλλέγονται συγκεκριμένες πληροφορίες για τα προβλήματα που θα καταγράφονται. Ο χρήστης της εφαρμογής θα μπορεί να καταγράφει το πρόβλημα που επιθυμεί απλά συμπληρώνοντας τη φόρμα αυτή. Στο σημείο αυτό η πρώτη καταγραφή προβλημάτων είναι έτοιμη να υλοποιηθεί. Η καταγραφή ολοκληρώνεται σε δύο βήματα:

- Αρχικά πρέπει να πραγματοποιηθεί εκκίνηση της εγκατεστημένης εφαρμογής στην κινητή συσκευή και εισαγωγή του email και του κωδικού για πρόσβαση στο λογαριασμό που δημιουργήθηκε.
- Στη συνέχεια πρέπει να συμπληρωθούν τα πεδία για την περιγραφή του προβλήματος, αλλά και τα προσωπικά στοιχεία του χρήστη της εφαρμογής.



Εικόνα 5.5: Σύνδεση στο λογαριασμό της εφαρμογής FulcrumApp από το κινητό τηλέφωνο

Ο διαχειριστής της εφαρμογής μπορεί να συνδεθεί στο λογαριασμό του στο Fulcrum μέσω της αντίστοιχης ιστοσελίδας προκειμένου να επεξεργαστεί, να απεικονίσει ή να αποθηκεύσει στον ηλεκτρονικό υπολογιστή τα δεδομένα των καταγραφών.



Εικόνα 5.6: Σύνδεση στον προσωπικό λογαριασμό στην ιστοσελίδα του Fulcrum

Το όνομα που δόθηκε στην εφαρμογή καταγραφής των προβλημάτων είναι «**Damage Report**» και της αποδόθηκε η εξής περιγραφή:

«Εφαρμογή για την άμεση και σύντομη υποβολή καταγγελιών που αφορούν σε προβλήματα δημόσιων υποδομών».

Θα πρέπει να σημειωθεί ότι η φόρμα που δημιουργήθηκε για την καταγραφή των προβλημάτων είναι διαθέσιμη μόνο στο δημιουργό του συγκεκριμένου λογαριασμού στο Fulcrum. Επομένως, προκειμένου η εφαρμογή να είναι διαθέσιμη στο πλήθος και όχι μόνο στο δημιουργό ή το διαχειριστή της, ώστε να μπορεί να καταγράψει όποιος θέλει ένα υπάρχον πρόβλημα, υπάρχουν δύο πιθανές λύσεις:

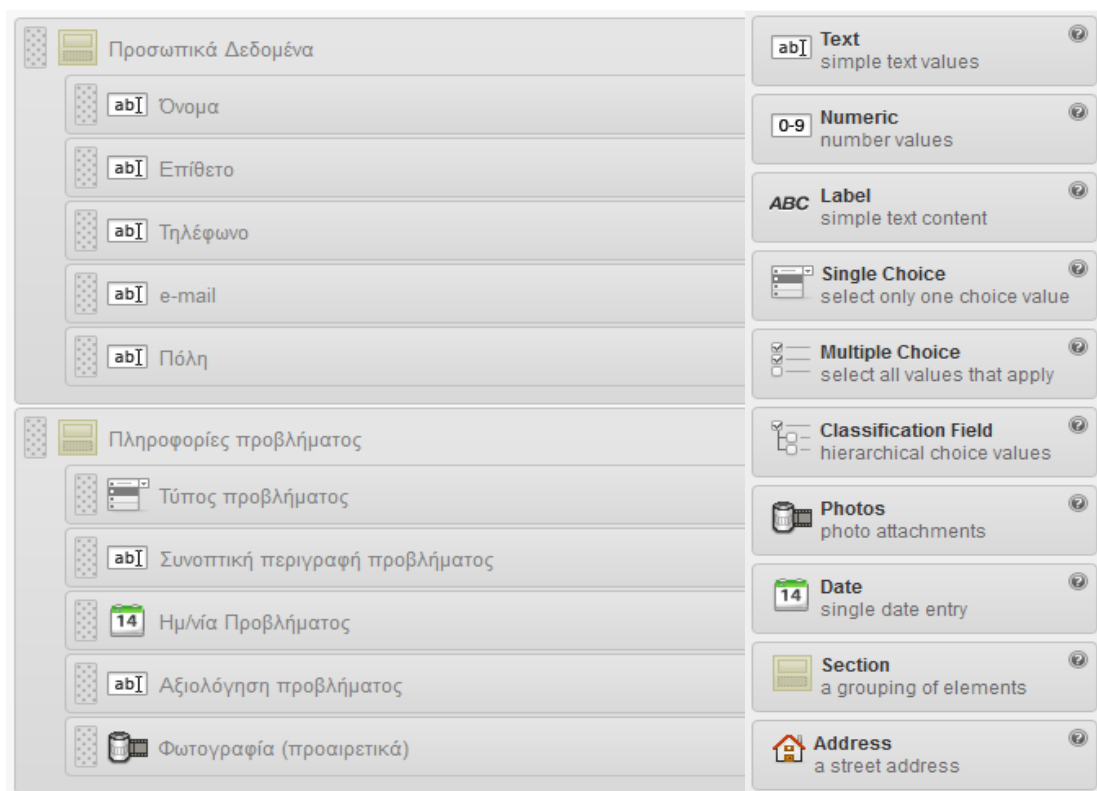
- ✓ Η παραχώρηση στο πλήθος του κωδικού πρόσβασης του λογαριασμού της εφαρμογής.

- ✓ Η δημιουργία ομάδων ατόμων μέσω της εφαρμογής του Fulcrum, οι οποίες θα έχουν δικαιοδοτηθεί από το διαχειριστή της εφαρμογής να καταγράψουν προβλήματα σε δημόσιες υποδομές συμπληρώνοντας τη φόρμα της εφαρμογής.

5.2.2 Δημιουργία της εφαρμογής καταγραφής προβλημάτων στο Fulcrum

Η crowdsourcing εφαρμογή Fulcrum App, η οποία χρησιμοποιήθηκε για τη συλλογή των αρχικών δεδομένων, διαθέτει ένα μεγάλο πλήθος έτοιμων εργαλείων αλλά και πρότυπων φορμών, με βάση τα οποία μπορεί κάποιος να δημιουργήσει τη δική του εφαρμογή.

Τα εργαλεία αυτά χρησιμοποιήθηκαν για τη δημιουργία μίας εφαρμογής καταγραφής προβλημάτων. Αφότου επιλέχθηκαν τα είδη των δεδομένων που θα συλλέγονται από την εφαρμογή, σειρά είχε η υλοποίηση των κατηγοριών δεδομένων που ήδη περιγράφηκαν, στην εφαρμογή του Fulcrum. Δημιουργήθηκαν δύο αρχικά πεδία σύμφωνα με τον αρχικό διαχωρισμό σε δύο υπερκατηγορίες και στη συνέχεια δημιουργήθηκαν πέντε υποπεδία για κάθε ένα από τα δύο αρχικά πεδία. Η δόμηση των πεδίων φαίνεται στην εικόνα που ακολουθεί.



Εικόνα 5.7: Δημιουργία πεδίων και επιλογή τύπων δεδομένων στην εφαρμογή του Fulcrum

Ο τύπος των δεδομένων του κάθε πεδίου επιλέγεται από ένα φάσμα προκαθορισμένων τύπων δεδομένων που διαθέτει το Fulcrum και είναι οι εξής:

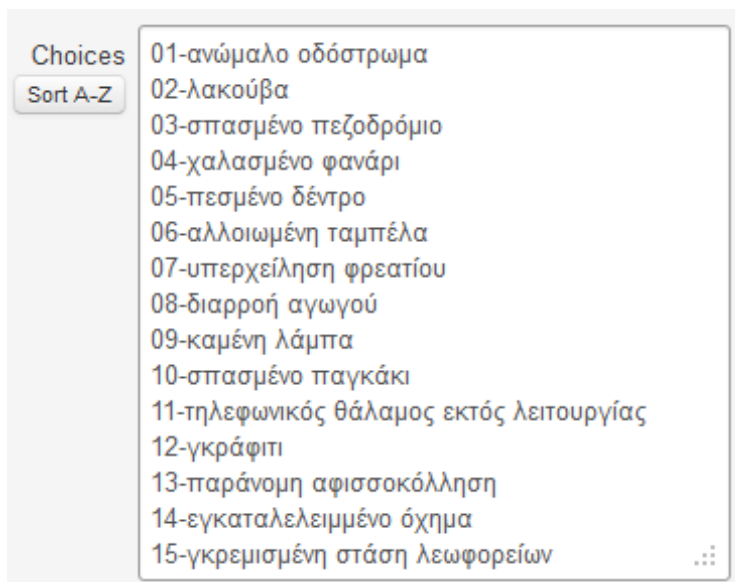
- Κείμενο
- Αριθμός

- Ετικέτα
- Απλής επιλογής
- Πολλαπλής επιλογής
- Ταξινόμησης
- Φωτογραφία
- Ημερομηνία
- Τομέας
- Διεύθυνση

Από τους προκαθορισμένους τύπους δεδομένων χρησιμοποιήθηκαν:

- ✓ Κείμενο για την πλειοψηφία των πεδίων και συγκεκριμένα για τα πεδία «Όνομα», «Επίθετο», «Τηλέφωνο», «e-mail», «Πόλη», «Συνοπτική περιγραφή» και «Αξιολόγηση προβλήματος».
- ✓ Απλής επιλογής για το πεδίο «Τύπος προβλήματος».
- ✓ Ημερομηνία για το πεδίο «Ημερομηνία προβλήματος».
- ✓ Και φωτογραφία για το πεδίο «Φωτογραφία».

Ο τύπος «ημερομηνία» δίνει τη δυνατότητα για αυτόματη καταχώρηση της τρέχουσας ημερομηνίας στο αντίστοιχο πεδίο. Ο τύπος «φωτογραφία» δίνει τη δυνατότητα για λήψη φωτογραφίας του προβλήματος τη στιγμή της καταγραφής, αλλά και για καταχώρηση φωτογραφίας από τις ήδη υπάρχουσες στη συσκευή του τηλεφώνου. Τέλος, ο τύπος «απλής επιλογής» που επιλέχθηκε για το πεδίο «Τύπος προβλήματος» δίνει τη δυνατότητα επιλογής του προβλήματος από μία προκαθορισμένη λίστα τύπων προβλημάτων. Εισήχθησαν 15 τύποι προβλημάτων δημοσίων υποδομών εκ των οποίων ο χρήστης θα μπορεί να επιλέγει και να καταγράφει έναν κάθε φορά.



Εικόνα 5.8: Τύποι προβλήματος που καταχωρήθηκαν στο πεδίο απλής επιλογής «Τύπος προβλήματος»

Το πεδίο «Αξιολόγηση προβλήματος» δίνει ως οδηγία τη βαθμολόγηση της σοβαρότητας του προβλήματος σε κλίμακα 1 ως 5, όπου με 1 θα βαθμολογείται το

λιγότερο σοβαρό/επικίνδυνο πρόβλημα ενώ με 5 θα βαθμολογείται το πολύ σοβαρό πρόβλημα σύμφωνα με την κρίση του ατόμου.

Εικόνα 5.9: Βαθμολόγηση της σοβαρότητας του προβλήματος από 1 έως 5 στο πεδίο κειμένου «Αξιολόγηση προβλήματος»

Ο τύπος δεδομένων των πεδίων σε αντιστοιχία με το εκάστοτε πεδίο φαίνεται στον πίνακα που ακολουθεί.

| Πεδίο | Υποπεδίο | Τύπος Δεδομένων |
|-------------------------|--------------------------|-----------------|
| Προσωπικά δεδομένα | Όνομα | text |
| | Επίθετο | text |
| | Τηλέφωνο | text |
| | e-mail | text |
| | Πόλη | text |
| | Τύπος προβλήματος | single choice |
| Πληροφορίες Προβλήματος | Συνοπτική περιγραφή | text |
| | Ημερομηνία προβλήματος | date |
| | Αξιολόγηση προβλήματος | text |
| | Φωτογραφία (προαιρετικά) | photos |

Πίνακας 5.2: Τύποι δεδομένων που επιλέχθηκαν για κάθε πεδίο

Η συμπλήρωση όλων των πεδίων δεδομένων είναι προαιρετική εκτός από τα πεδία «Τηλέφωνο» ώστε να είναι δυνατή η επικοινωνία με τον χρήστη για τυχόν διευκρινίσεις, «Τύπος προβλήματος» και «Ημερομηνία», των οποίων η συμπλήρωση είναι **υποχρεωτική**.

Πέραν από τα πεδία που περιγράφηκαν και υλοποιήθηκαν, απαραίτητη για τον εντοπισμό των προβλημάτων είναι η απόδοση συντεταγμένων σε κάθε καταγραφή. Η εφαρμογή Fulcrum δίνει τη δυνατότητα αυτόματης καταχώρησης των γεωγραφικών συντεταγμένων της θέσης της καταγραφής με την επιλογή «set current location», αλλά και με επιλογή της θέσης από χάρτη.

Επιπλέον, η εφαρμογή προσθέτει στα τελικά εξαγόμενα δεδομένα τέσσερα επιπλέον πεδία, τα οποία δε συμπληρώνει ο χρήστης αλλά αποθηκεύονται αυτόματα, τα οποία αφορούν:

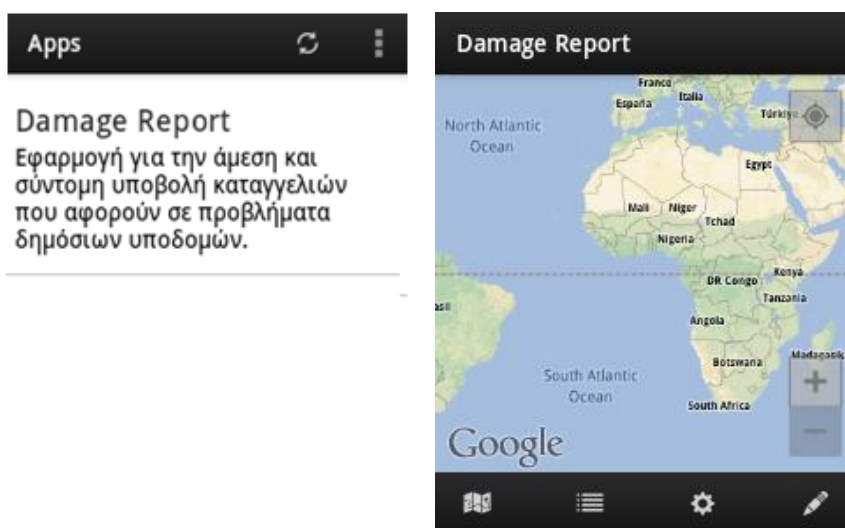
- ✓ στην ημερομηνία και ώρα που πραγματοποιήθηκε η καταγραφή
- ✓ στην ημερομηνία και ώρα που ανανεώθηκε η καταγραφή
- ✓ από ποιον πραγματοποιήθηκε η καταγραφή, το οποίο συμπληρώνεται με βάση το email εγγραφής του χρήστη στο Fulcrum

- ✓ από ποιον ανανεώθηκε η καταγραφή, το οποίο επίσης συμπληρώνεται από το email του χρήστη

5.2.3 Τρόπος συλλογής των δεδομένων

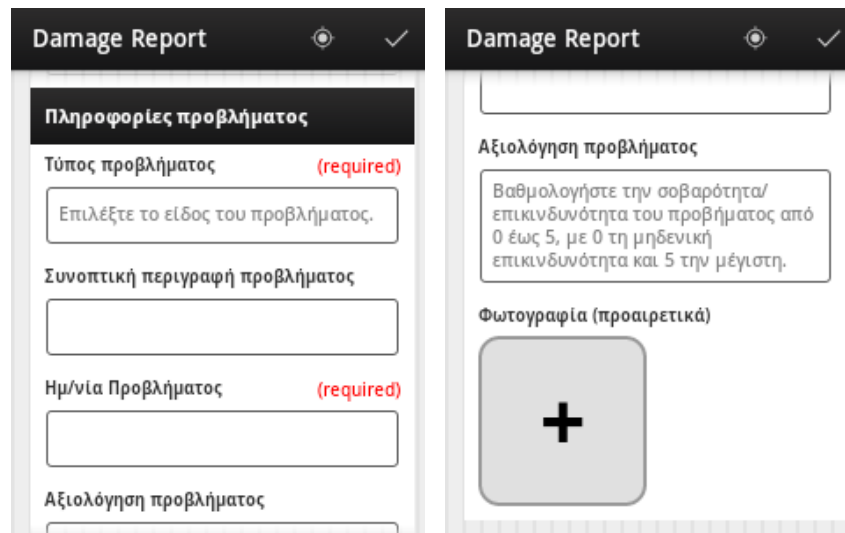
Μετά την ολοκλήρωση του ορισμού των πεδίων της εφαρμογής, η εφαρμογή είναι διαθέσιμη προς χρήση. Το μόνο που απαιτείται είναι ένα κινητό τηλέφωνο με λογισμικό Android ή iOS και η εγκατάσταση της εφαρμογής στην κινητή συσκευή. Ο χρήστης μπορεί στη συνέχεια να ξεκινήσει να καταγράφει τα προβλήματα που συναντάει είτε online, είτε offline και συγχρονίζοντας τα δεδομένα όταν βρεθεί online. Με τον τρόπο αυτό προστίθεται στην εφαρμογή του Fulcrum ένα νέο σημείο και πληροφορία για κάθε καινούρια καταγραφή.

Οι εικόνες που ακολουθούν αποτελούν printscreen της εφαρμογής της καταγραφής προβλημάτων στο κινητό τηλέφωνο η οποία δημιουργήθηκε στο προηγούμενο στάδιο για το σκοπό της εργασίας.



Εικόνα 5.10: Prinscreen από την εφαρμογή Damage Report στο κινητό τηλέφωνο – Αριστερά οι διαθέσιμες εφαρμογές, δεξιά χάρτης επιλογής θέσης

Εικόνα 5.11: Prinscreen από την εφαρμογή Damage Report στο κινητό τηλέφωνο – Συμπλήρωση προσωπικών δεδομένων



Εικόνα 5.12: Prinscreen από την εφαρμογή Damage Report στο κινητό τηλέφωνο – Συμπλήρωση πληροφοριών προβλήματος

Η εφαρμογή είναι πλέον διαθέσιμη για την καταγραφή προβλημάτων σε δημόσιες υποδομές. Υλοποιήθηκε ενδεικτικά ένας μικρός αριθμός καταγραφών. Εδώ πρέπει να σημειωθεί ότι για μία συγκεκριμένη περίπτωση προβλήματος μπορούν να πραγματοποιηθούν πολλές καταγραφές. Επιπλέον, μπορούν να καταγραφούν πολλές και διαφορετικές περιπτώσεις προβλημάτων σε μία δεδομένη περιοχή. Εφόσον ολοκληρωθεί η διαδικασία των καταγραφών, τα δεδομένα είναι διαθέσιμα προς αποθήκευση.

Η πληροφορία που καταγράφεται στο Fulcrum είναι διαθέσιμη με τρεις διαφορετικούς τρόπους. Το Fulcrum επιτρέπει την παρουσίαση των αποτελεσμάτων σε δυναμικούς χάρτες στο διαδίκτυο, στον προσωπικό λογαριασμό του διαχειριστή της εφαρμογής. Τα δεδομένα μπορούν να συγχρονιστούν ανά πάσα στιγμή και να εμφανιστούν όλες οι καταγραφές συνοδευόμενες από την καταγεγραμμένη πληροφορία στον χάρτη του Fulcrum. Επιπλέον, υπάρχει η δυνατότητα για Google Earth Streaming, το οποίο επιτρέπει την απεικόνιση των δεδομένων live στο Google Earth. Αυτό σημαίνει ότι με κάθε νέα προσθήκη καταγραφής, ο χάρτης του Google Earth ανανεώνεται με νέο σημείο. Τέλος, τα δεδομένα μπορούν να εξαχθούν και να αποθηκευτούν στον ηλεκτρονικό υπολογιστή σε διάφορες μορφές αρχείων. Οι τύποι αρχείων που διαθέτει η εφαρμογή για την εξαγωγή των δεδομένων είναι:

- CSV
- KML
- ESRI Shapefile
- GeoJSON

Στην εργασία αυτή, επιλέχθηκε τα δεδομένα να εξαγονται σε μορφή .csv έτσι ώστε να είναι πιο απλή η επεξεργασία τους. Η εξαγωγή των δεδομένων πραγματοποιείται σε τρία στάδια:

- Το πρώτο στάδιο περιλαμβάνει την επιλογή της μορφής του εξαγόμενου αρχείου δεδομένων.

- Το δεύτερο στάδιο επιτρέπει την επιλογή των πεδίων που επιθυμεί ο χρήστης να συμπεριληφθούν στο εξαγόμενο αρχείο, καθώς επίσης και να καθορίσει το όνομα που θα έχει κάθε πεδίο στο εξαγόμενο αρχείο.
- Το τρίτο στάδιο είναι το στάδιο της επιβεβαίωσης της εξαγωγής των δεδομένων.

With your current filters, the export will contain 4 record(s).

File Format:

Date Range: to

Date Time Zone:

Area Filter: [Select Area](#)

Include Photos:

Include Record ID:

Apps: [Toggle all](#)
 Damage Report

Select the desired format of your data export. Fulcrum can export in the following formats:

- CSV
- KML
- ESRI Shapefile
- GeoJSON

Optionally you may export your photos. It will be included in a folder, and the photo names will be referenced in the export.

Εικόνα 5.13: Στάδιο 1 – Επιλογή τύπου των εξαγόμενων δεδομένων

Damage Report

| Field Label | Exported Column Name |
|--|--------------------------------------|
| <input checked="" type="checkbox"/> Όνομα | <input type="text"/> |
| <input checked="" type="checkbox"/> Επώνυμο | <input type="text"/> |
| <input checked="" type="checkbox"/> Τηλέφωνο | <input type="text"/> |
| <input checked="" type="checkbox"/> e-mail | <input type="text" value="e_mail"/> |
| <input checked="" type="checkbox"/> Πόλη | <input type="text"/> |
| <input checked="" type="checkbox"/> Τύπος προβλήματος | <input type="text"/> |
| <input checked="" type="checkbox"/> Συνοπτική περιγραφή προβλήματος | <input type="text"/> |
| <input checked="" type="checkbox"/> Ημ/νία Προβλήματος | <input type="text"/> |
| <input checked="" type="checkbox"/> Αξιολόγηση προβλήματος | <input type="text"/> |
| <input checked="" type="checkbox"/> Φωτογραφία (προαιρετικά) | <input type="text"/> |
| <input checked="" type="checkbox"/> Φωτογραφία (προαιρετικά) Caption | <input type="text" value="caption"/> |

Εικόνα 5.14: Στάδιο 2 - Επιλογή των πεδίων και της ονομασίας τους στο εξαγόμενο αρχείο

Apps: Damage Report

Export format: CSV

Timezone for exported dates: (GMT+00:00) UTC

Exporting photos: No

Number of records to be exported: 4

Εικόνα 5.15: Στάδιο 3 – Επιβεβαίωση εξαγωγής

Στους παρακάτω πίνακες παρουσιάζονται τα περιεχόμενα του εξαγόμενου csv αρχείου το οποίο περιλαμβάνει το όνομα των πεδίων δεδομένων και δύο ενδεικτικές καταγραφές. Το στοιχείο που διαφοροποιείται στους πίνακες σε σχέση με το εξαγόμενο προϊόν είναι η δομή των δεδομένων. Συγκεκριμένα, το όνομα των πεδίων αλλά και τα δεδομένα που έχουν καταγραφεί στο csv αρχείο, έχουν τη δομή που παρουσιάζεται στους πίνακες με τη διαφορά ότι διαχωρίζονται μεταξύ τους με κόμμα. Αυτό καθιστά την ανάγνωσή τους ιδιαίτερα εύκολη.

| latitude | longitude | Όνομα | Επίθετο | Τηλέφωνο | e_mail | Πόλη |
|-----------|-----------|---------|--------------|------------|------------|-------|
| 37.979691 | 23.763385 | Γιώργος | Παπαδάκης | 6977777777 | g@yahoo.gr | αθήνα |
| 37.96944 | 23.753214 | Ελένη | Παπαδοπούλου | 6922222222 | p@yahoo.gr | αθήνα |

| Τύπος προβλήματος | Συνοπτική περιγραφή προβλήματος | Ημ/νία Προβλήματος | Αξιολόγηση προβλήματος | Φωτογραφία (προαιρετικά) | caption |
|---------------------|--|--------------------|------------------------|--------------------------|---------|
| 04-χαλασμένο φανάρι | το φανάρι δε λειτουργεί | 13/11/2012 | 3 | - | - |
| 02-λακούβα | επικίνδυνη λακούβα στη μέση του δρόμου | 10/11/2012 | 5 | - | - |

| updated | created | project | created_by | updated_by |
|---------------------|---------------------|---------|------------|------------|
| 13/11/2012 15:24 | 13/11/2012 15:24 | - | mairi | - |
| 13/11/2012 16:25 | 13/11/2012 15:27 | - | mairi | mairi |

Πίνακας 5.3: Παράδειγμα της δομής των δεδομένων όπως εξάγονται από το csv αρχείο

5.2.4 Εισαγωγή των δεδομένων στη βάση δεδομένων

Στο παρόν στάδιο, αρχικά στόχος είναι η εισαγωγή των δεδομένων που εξήχθησαν από την εφαρμογή Fulcrum σε μορφή csv, στη βάση δεδομένων που δημιουργήθηκε προηγουμένως.

Επειδή ήταν χρονικά και πρακτικά δύσκολο να συλλεχθεί ο απαραίτητος όγκος πραγματικών καταγραφών προβλημάτων στα πλαίσια της μεταπτυχιακής εργασίας με τη βοήθεια του crowdsourcing, αποφασίσθηκε να χρησιμοποιηθούν πλαστά δεδομένα. Τα δεδομένα αυτά δημιουργήθηκαν με τυχαίο τρόπο, με γνώμονα όμως την προσέγγιση σε πραγματικά δεδομένα, και στη συνέχεια καταχωρήθηκαν σε ένα csv αρχείο έχοντας ακριβώς την ίδια δομή με το csv αρχείο των ενδεικτικών καταγραφών που εξάγεται από την εφαρμογή του Fulcrum.

Τα βήματα υλοποίησης του σταδίου αυτού περιγράφονται λεπτομερώς στη συνέχεια:

- ✓ **Εισαγωγή δεδομένων στον πίνακα «damage_types».** Η εισαγωγή των δεδομένων στον πίνακα «damage_types» αποτελεί το πρώτο βήμα μετά τη δημιουργία της βάσης δεδομένων και τη δημιουργία των αρχικών πινάκων. Ο πίνακας «damage_types» περιλαμβάνει όλους τους πιθανούς τύπους προβλημάτων και τα στοιχεία που καθορίζουν το χρώμα που αντιστοιχεί σε κάθε

τύπο προβλήματος για τη σωστή απεικόνισή τους σε χάρτες. Στο στάδιο αυτό εκτελούνται οι ακόλουθες διεργασίες:

- Δημιουργία txt αρχείου με τους τύπους των υπαρχόντων προβλημάτων. Οι τύποι προβλημάτων που ορίστηκαν αρχικά κατά τη δημιουργία της φόρμας καταγραφής προβλημάτων στην crowdsourcing εφαρμογή, καταγράφονται σε ένα αρχείο txt κειμένου. Το αρχείο αυτό περιλαμβάνει τον κωδικό και την ονομασία κάθε τύπου προβλήματος.
- Δημιουργία txt αρχείου με τα χρώματα που πρόκειται να αντιστοιχηθούν σε κάθε τύπο προβλήματος. Για κάθε έναν από τους 15 τύπους προβλημάτων που χρησιμοποιούνται, καταγράφεται και ένα διαφορετικό χρώμα, το οποίο θα αποτελέσει οδηγό για τη χρωματική απεικόνιση των σημείων.
- Δημιουργία txt αρχείων με την κωδικοποιημένη σε γλώσσα KML περιγραφή των εικονιδίων που θα αντιπροσωπεύουν τις καταγραφές των προβλημάτων στους διαδικτυακούς χάρτες που πρόκειται να κατασκευαστούν στη συνέχεια. Τα αρχεία αυτά θα συμβάλλουν στη σύνθεση των χαρτών για τη σωστή απεικόνιση των σημείων με βάση την κατηγορία προβλήματος στην οποία ανήκει κάθε σημείο.
- Δημιουργία txt αρχείων που περιέχουν την κωδικοποίηση κάθε χρώματος για τη σωστή αντιστοιχία του χρώματος με κάθε σημείο δεδομένων.
- Ανάγνωση των περιεχομένων όλων των txt αρχείων γραμμή προς γραμμή με χρήση κώδικα σε γλώσσα PHP. Τα δεδομένα καταχωρούνται σε πίνακες της γλώσσας PHP.
- Πραγματοποιείται σύνδεση με το λογισμικό PostgreSQL και τη βάση δεδομένων. Στο σημείο αυτό εισάγονται τα περιεχόμενα των αρχείων στον πίνακα «damage_types» της βάσης δεδομένων με την εντολή INSERT INTO σε γλώσσα SQL, η οποία έχει ενσωματωθεί στον τελικό αλγόριθμο σε γλώσσα PHP.

Ο πίνακας έχει τη δομή που φαίνεται στην εικόνα που ακολουθεί και περιέχει τα αντίστοιχα δεδομένα.

| | t_code [PK] int | description character varying(200) | colour character varying(1 | kml_unique_colour text | kml_crowd_colour text | styleurl_unique character varyir | styleurl_crowd character varyir | icon bytea |
|----|--------------------|---------------------------------------|-------------------------------|---------------------------|--------------------------|-------------------------------------|------------------------------------|---------------|
| 1 | 1 | ανάμυλο οδόστρωμα | μπλε | <StyleMap id="ms | <Style id="sh_w | msn_wht-pushp | msn_wht-blanc | <binar |
| 2 | 2 | λακούβα | πράσινο ανοιχτό | <Style id="sn_wh | <Style id="sn_g | msn_wht-pushp | msn_grn-blanc | <binar |
| 3 | 3 | σπασμένο πεζοδρόμιο | γαλάζιο1 | <Style id="sn_lt | <Style id="sn_l | msn_ltblu-pus | msn_ltblu-bl | <binar |
| 4 | 4 | χαλασμένο φανάρι | φούξια | <StyleMap id="ms | <Style id="sn_p | msn_pink-push | msn_pink-blanc | <binar |
| 5 | 5 | πεσμένο δέντρο | άσπρο | <StyleMap id="ms | <StyleMap id="m | msn_wht-pushp | msn_wht-blanc | <binar |
| 6 | 6 | αλλοιωμένη τομπέλα | μωβ ανοιχτό | <Style id="sh_wh | <StyleMap id="m | msn_wht-pushp | msn_wht-blanc | <binar |
| 7 | 7 | υπερχείληση φρεατίου | ροζ | <Style id="sn_wh | <Style id="sn_w | msn_wht-pushp | msn_wht-blanc | <binar |
| 8 | 8 | διαρροή αγωγού | κόκκινο | <StyleMap id="ms | <Style id="sh_w | msn_wht-pushp | msn_wht-blanc | <binar |
| 9 | 9 | καμένη λάμπα | πράσινο σκούρο | <StyleMap id="ms | <Style id="sn_w | msn_wht-pushp | msn_wht-blanc | <binar |
| 10 | 10 | σπασμένο παγκάκι | χακί | <Style id="sn_wh | <Style id="sn_w | msn_wht-pushp | msn_wht-blanc | <binar |
| 11 | 11 | τηλεφωνικός θάλαμος εκτός λειτουργίας | γαλάζιο | <Style id="sh_wh | <Style id="sn_w | msn_wht-pushp | msn_wht-blanc | <binar |
| 12 | 12 | γκράφιτι | πορτοκαλί | <Style id="sn_wh | <Style id="s_yl | msn_wht-pushp | m_ylw-pushp | <binar |
| 13 | 13 | παράνομη αφισσοκόλληση | κίτρινο | <StyleMap id="ms | <StyleMap id="m | msn_ylw-pushp | msn_wht-blanc | <binar |
| 14 | 14 | εγκαταλελειμμένο όχημα | μύρο | <Style id="sn_wh | <Style id="sh_w | msn_wht-pushp | msn_wht-blanc | <binar |
| 15 | 15 | γκρεμισμένη στάση λεωφορείων | μωβ σκούρο | <StyleMap id="ms | <Style id="sn_w | msn_wht-pushp | msn_wht-blanc | <binar |

Εικόνα 5.16: Εισαγωγή των δεδομένων των τύπων προβλημάτων στον πίνακα «damage_types»

Όπως φαίνεται και στην εικόνα η αντιστοιχία των τύπων προβλημάτων και των χρωμάτων απεικόνισης του κάθε προβλήματος έχει ως εξής:

| Τύπος προβλήματος | Χρώμα |
|---------------------------------------|-----------------|
| ανώμαλο οδόστρωμα | μπλε |
| λακκούβα | πράσινο ανοιχτό |
| σπασμένο πεζοδρόμιο | γαλάζιο1 |
| χαλασμένο φανάρι | φούξια |
| πεσμένο δέντρο | άσπρο |
| αλλοιωμένη ταμπέλα | μωβ ανοιχτό |
| υπερχείλιση φρεατίου | ροζ |
| διαρροή αγωγού | κόκκινο |
| καμένη λάμπα | πράσινο σκούρο |
| σπασμένο παγκάκι | χακί |
| τηλεφωνικός θάλαμος εκτός λειτουργίας | γαλάζιο |
| γκράφιτι | πορτοκαλί |
| παράνομη αφισκοκόλληση | κίτρινο |
| εγκαταλελειμμένο όχημα | μαύρο |
| γκρεμισμένη στάση λεωφορείων | μωβ σκούρο |

Πίνακας 5.4: Αντιστοιχία τύπου προβλήματος και χρώματος απεικόνισης

- ✓ **Δημιουργία και εισαγωγή των δεδομένων των καταγραφών στον πίνακα «crowd_data».** Στο στάδιο αυτό, επειδή όπως ήδη ειπώθηκε η συλλογή πραγματικών δεδομένων δεν ήταν χρονικά εφικτή, τα δεδομένα κατασκευάστηκαν έχοντας τη δομή του εξαγόμενου csv αρχείου και στη συνέχεια εισήχθησαν στον πίνακα «crowd_data». Αναλυτικότερα:
 - Δημιουργήθηκε ένα αρχείο excel με όλα τα πεδία της εφαρμογής όπως αυτά εξάγονται από το Fulcrum, με εξαίρεση τα πεδία του γεωγραφικού πλάτους και μήκους καθώς επίσης και του τύπου προβλήματος:

| | |
|-------------------|------------|
| name | severity |
| surname | capture |
| phone_number | updated |
| e_mail | created |
| city | project |
| short_description | created_by |
| date | updated_by |

 Τα δεδομένα σχηματίστηκαν με μία σχετική τυχαιότητα. Τα ονοματεπώνυμα αναπαρήχθησαν με τυχαία ονόματα και επώνυμα. Ενώ τα αριθμητικά δεδομένα δημιουργήθηκαν με τη βοήθεια της συνάρτησης του excel randbetween(), ώστε αν και τυχαία να είναι ρεαλιστικά. Στη συνέχεια, το φύλλο excel με τα δεδομένα που αντιστοιχούν στα δεδομένα πραγματικών καταγραφών της εφαρμογής Fulcrum, αποθηκεύτηκε ως csv αρχείο προσεγγίζοντας την επιθυμητή δομή που περιγράφηκε νωρίτερα.
 - Για την υλοποίηση της εφαρμογής, προσομοιώνοντας τις πραγματικές συνθήκες καταγραφής προβλημάτων, κρίθηκε απαραίτητο να δημιουργηθεί ένας αριθμός τυχαίων αρχικών σημείων στην ευρύτερη περιοχή της Αθήνας, καθένα εκ των οποίων θα αντιπροσωπεύει ένα συγκεκριμένο πρόβλημα και με βάση τα οποία θα δημιουργηθούν στη συνέχεια τα σημειακά δεδομένα των καταγραφών για κάθε πρόβλημα. Για κάθε περίπτωση προβλήματος δημιουργήθηκαν από 1 έως 100 ζεύγη

συντεταγμένων καταγραφών. Τα αρχικά σημεία εντοπίστηκαν και σημειώθηκαν στο περιβάλλον του Google Earth κι έπειτα αποθηκεύτηκαν στα αντίστοιχα kml αρχεία. Με τη βοήθεια λογισμικού μετατροπής αρχείων από KML σε csv και αντίστροφα, εξήχθησαν οι γεωγραφικές συντεταγμένες σε μορφή csv. Τα σημεία αυτά καταχωρήθηκαν σε πίνακες στον αλγόριθμο της PHP ως γεωγραφικές συντεταγμένες (φ,λ).

- Τα αρχικά σημεία θα αποτελέσουν τα προβλήματα γύρω από τα οποία θα πραγματοποιηθούν οι καταγραφές. Οι γεωγραφικές συντεταγμένες των αρχικών σημείων μετατράπηκαν από (φ₀,λ₀) WGS84 σε προβολικές συντεταγμένες (x₀,y₀) ΕΓΣΑ87.
- Στο στάδιο αυτό δημιουργήθηκε ο αλγόριθμος για τη δημιουργία τυχαίων σημείων γύρω από τα αρχικά, τα οποία θα αντιπροσωπεύουν όλες τις καταγραφές που έχουν γίνει για κάθε πρόβλημα. Τα τυχαία σημεία δημιουργούνται από τον υπολογισμό των προβολικών συντεταγμένων (x,y) για κάθε καταγραφή. Οι προβολικές συντεταγμένες των σημείων υπολογίζονται με βάση τις πολικές συντεταγμένες (r,θ) θεωρώντας ως αρχή των αξόνων για κάθε καταγραφή το εκάστοτε από τα αρχικά σημεία, με πολικές συντεταγμένες (0,0). Έτσι υπολογίζεται η μετατόπιση κάθε καταγραφής από την αρχή των αξόνων ως:

$$\delta x = r \cdot \cos \theta$$

$$\delta y = r \cdot \sin \theta$$

όπου η ακτίνα κυμαίνεται μεταξύ 0<r<25 (m) και η πολική γωνία μεταξύ 0<θ<360°. Αυτό σημαίνει ότι δημιουργούνται τυχαία σημεία γύρω από τα αρχικά με ακτίνα έως 25 m και γωνία έως 360 μοίρες. Η ακτίνα επιλέχθηκε με κριτήριο τη μέγιστη απόσταση από τη θέση ενός υπάρχοντος προβλήματος από την οποία μπορεί να πραγματοποιηθεί μία καταγραφή για το συγκεκριμένο πρόβλημα.

- Οι προβολικές συντεταγμένες των τυχαίων σημείων που δημιουργήθηκαν υπολογίζονται στο σύστημα ΕΓΣΑ87 ως το άθροισμα των προβολικών συντεταγμένων των αρχικών σημείων και της εκάστοτε μετατόπισης δx και δy κάθε καταγραφής από το ίδιο το πρόβλημα:

$$x = x_0 + \delta x$$

$$y = y_0 + \delta y$$

- Τέλος, οι προβολικές συντεταγμένες των τυχαίων σημείων μετατρέπονται σε γεωγραφικές συντεταγμένες (φ,λ) του συστήματος WGS84.
- Για το σύνολο των τυχαίων σημείων που δημιουργούνται για κάθε αρχικό σημείο, αντιστοιχίζεται και ένας τύπος προβλήματος.
- Έτσι το τελικό csv αρχείο που παράγεται θα περιέχει όλα τα απαραίτητα δεδομένα, συμπεριλαμβανομένων των αρχικών δεδομένων που προέκυψαν από το λογισμικό του excel, του συνόλου των συντεταγμένων (φ,λ) των τυχαίων σημείων σε WGS84, καθώς επίσης και των τύπων προβλημάτων που τους αντιστοιχούν. Το csv αρχείο είναι έτοιμο για ανάγνωση.
- Αφού διαμορφώθηκαν τα δεδομένα των καταγραφών, σειρά έχει η προσπέλαση του csv αρχείου και η καταχώριση των δεδομένων του σε πίνακες με κώδικα PHP. Έπειτα πραγματοποιείται σύνδεση με τη βάση δεδομένων του λογισμικού PostgreSQL και εισάγονται τα δεδομένα του

csv αρχείου στα αντίστοιχα πεδία του πίνακα «crowd_data» με την εντολή INSERT INTO της γλώσσας SQL ενσωματωμένη στον κώδικα PHP.

- Τέλος, ενημερώνεται η γεωμετρία των σημείων στη στήλη geom του πίνακα «crowd_data», με βάση τις ήδη καταχωρημένες γεωγραφικές συντεταγμένες, η οποία προσδίδει τη γεωμετρία της θέσης στα σημεία και καθιστά δυνατή την απεικόνισή τους στο λογισμικό Quantum GIS.

Στους πίνακες που ακολουθούν παρουσιάζονται ενδεικτικά έξι καταγραφές προβλημάτων από τον πίνακα «crowd_data».

| | id [PK] serial | latitude double precis | longitude double precis | name character vai | surname character vai | phone_num character vai | email character vai | city character vai | type_of_dam integer | short_descri character vai |
|---|----------------|------------------------|-------------------------|--------------------|-----------------------|-------------------------|---------------------|--------------------|---------------------|----------------------------|
| 1 | 1 | 37.98652849 | 23.76156168 | ΓΙΩΡΓΟΣ | ΠΑΠΑΔΑΚΗΣ | 6980799076 | txf@yahoo.g | Αθήνα | 1 | '' |
| 2 | 2 | 37.98639239 | 23.76138437 | ΜΑΡΙΝΑ | ΜΑΡΙΝΗ | 6947165896 | sag@yahoo.g | Αθήνα | 1 | '' |
| 3 | 3 | 37.98647309 | 23.76122645 | ΕΛΕΝΗ | ΠΑΝΤΟΥ | 6969325248 | hbd@yahoo.g | Αθήνα | 1 | '' |
| 4 | 4 | 37.98655674 | 23.76113643 | ΑΝΔΡΕΑΣ | ΑΝΔΡΕΟΥ | 6931857379 | uxj@yahoo.g | Αθήνα | 1 | '' |
| 5 | 5 | 37.98656770 | 23.76136754 | ΣΤΕΛΙΟΣ | ΡΟΚΚΟΣ | 6994390328 | cjh@yahoo.g | Αθήνα | 1 | '' |
| 6 | 6 | 37.98644048 | 23.76141076 | ΚΑΤΕΡΙΝΑ | ΛΕΧΟΥ | 6987523485 | wcg@yahoo.g | Αθήνα | 1 | '' |

| date date | severity integer | capture character vai | updated character vai | created character vai | project character vai | created_by character vai | updated_by character vai | geom geometry(Po | unique_cases integer |
|------------|------------------|-----------------------|-----------------------|-----------------------|-----------------------|--------------------------|--------------------------|------------------|----------------------|
| 2013-03-04 | 2 | '' | 2013-01-25 | 2013-05-03 | '' | ΓΙΩΡΓΟΣ | '' | 0101000020E | 2 |
| 2013-08-17 | 3 | '' | 2013-03-16 | 2013-04-05 | '' | ΜΑΡΙΝΑ | '' | 0101000020E | 2 |
| 2013-04-18 | 2 | '' | 2013-07-30 | 2013-04-15 | '' | ΕΛΕΝΗ | '' | 0101000020E | 2 |
| 2013-12-16 | 4 | '' | 2013-07-01 | 2013-04-13 | '' | ΑΝΔΡΕΑΣ | '' | 0101000020E | 2 |
| 2013-03-02 | 2 | '' | 2013-04-22 | 2013-05-20 | '' | ΣΤΕΛΙΟΣ | '' | 0101000020E | 2 |
| 2013-06-28 | 3 | '' | 2013-11-03 | 2013-05-18 | '' | ΚΑΤΕΡΙΝΑ | '' | 0101000020E | 2 |

Εικόνα 5.17: Εισαγωγή των δεδομένων των καταγραφών στον πίνακα «crowd_data»

Στο σημείο αυτό ολοκληρώνεται η εισαγωγή των δεδομένων στους δύο πίνακες και η επεξεργασία αυτών, και είναι δυνατή η μετάβαση στο επόμενο στάδιο που αφορά στην ομαδοποίηση των δεδομένων με εφαρμογή του αλγορίθμου k-means.

5.3 Εφαρμογή του αλγορίθμου k-means

Η συλλογή των δεδομένων των καταγραφών από μόνη της δεν αρκεί για να εξαχθούν συμπεράσματα και να αναπαρασταθούν σωστά τα δεδομένα σε χάρτες. Θα πρέπει να πραγματοποιηθεί ομαδοποίηση των δεδομένων του πίνακα «crowd_data» της βάσης δεδομένων, με βάση την κάθε περίπτωση προβλήματος, έτσι ώστε όλες εκείνες οι καταγραφές που αφορούν στο ίδιο πρόβλημα να ανήκουν στην ίδια ομάδα σημείων.

Η ταξινόμηση των σημείων σε ομάδες υλοποιείται με τη χρήση του αλγορίθμου ομαδοποίησης k-means ως συνάρτηση η οποία εγκαταστάθηκε ως extension στο λογισμικό PostgreSQL. Το extension του αλγορίθμου k-means ήταν συμβατό μόνο με την έκδοση 9.1 x86 της PostgreSQL.

Ο αλγόριθμος k-means δεν εφαρμόζεται άμεσα στα δεδομένα καθώς απαιτείται μία προεργασία, ώστε να εισαχθούν οι κατάλληλες παράμετροι στον αλγόριθμο προκειμένου να βελτιστοποιηθεί το αποτέλεσμα της ομαδοποίησης και να

προσεγγίζει σε μεγαλύτερο βαθμό την πραγματικότητα. Οι παράμετροι εισόδου του αλγορίθμου k-means είναι:

- ο πίνακας των διανυσμάτων των συντεταγμένων των σημείων που επιθυμεί ο χρήστης να ομαδοποιήσει.
- ο αριθμός των ομάδων k στις οποίες επιθυμεί ο χρήστης να ταξινομηθούν τα σημειακά δεδομένα.
- ο πίνακας ενός συνόλου διανυσμάτων αρχικοποίησης, τα οποία είναι όσες και οι ομάδες που αποφασίστηκε να δημιουργηθούν. Τα διανύσματα αρχικοποίησης είναι σημεία που ορίζονται από τον χρήστη με σκοπό την καθοδήγηση του αλγορίθμου για τη δημιουργία ομάδων λαμβάνοντας υπόψη τη θέση των διανυσμάτων αρχικοποίησης. Η παράμετρος αυτή είναι προαιρετική.

Ο αλγόριθμος k-means εξαρτάται από τον τρόπο που ορίζονται τα διανύσματα αρχικοποίησης. Έτσι η συνάρτηση k-means που χρησιμοποιήθηκε προσφέρει δύο τρόπους εφαρμογής σε σχέση με τον ορισμό των διανυσμάτων αρχικοποίησης:

- την αυτόματη αρχικοποίηση των διανυσμάτων
- και τον καθορισμό των διανυσμάτων αρχικοποίησης από τον χρήστη.

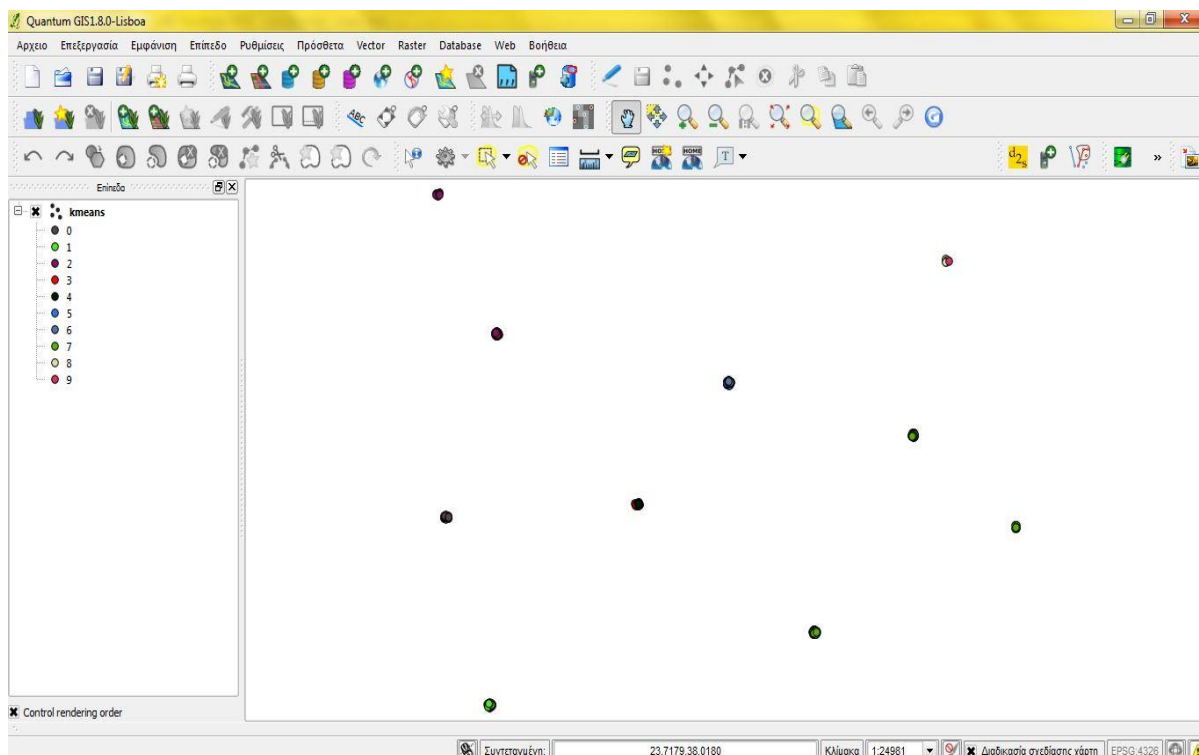
Στην περίπτωση της αυτόματης αρχικοποίησης των διανυσμάτων, η τρίτη παράμετρος του αλγορίθμου παραλείπεται, και ο αλγόριθμος εξετάζει όλα τα διανύσματα εισόδου και αποθηκεύει το μέγιστο (max) και το ελάχιστο (min) για κάθε στοιχείο του διανύσματος. Έπειτα αποφασίζει ποια θα είναι τα αρχικά υποθετικά διανύσματα με τη βοήθεια του τύπου:

$$\text{init} = (\text{max} - \text{min}) * (i + 1) / (k + 1) + \text{min}$$

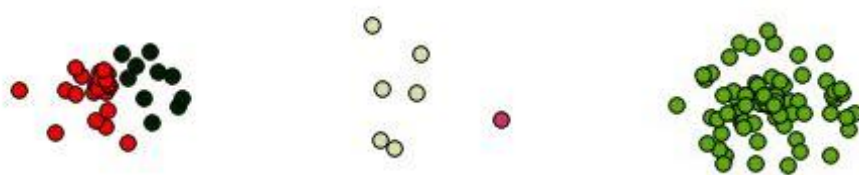
όπου το i κυμαίνεται από 0 έως k-1. Στη συνέχεια, επιλέγεται ένα από τα διανύσματα εισόδου ως το πλησιέστερο στο υποθετικό διάνυσμα. Τα διανύσματα που επιλέγονται αποτελούν τα διανύσματα αρχικοποίησης που επέλεξε ο αλγόριθμος.

Στην περίπτωση της συγκεκριμένης εφαρμογής, **αρχικά** εφαρμόστηκε ο αλγόριθμος k-means χωρίς την εισαγωγή της παραμέτρου των διανυσμάτων αρχικοποίησης. Έτσι ο αλγόριθμος δημιούργησε αυτόματα τα διανύσματα αρχικοποίησης, και στη συνέχεια ομαδοποίησε τα δεδομένα δίνοντας μη ικανοποιητική ταξινόμηση των σημείων.

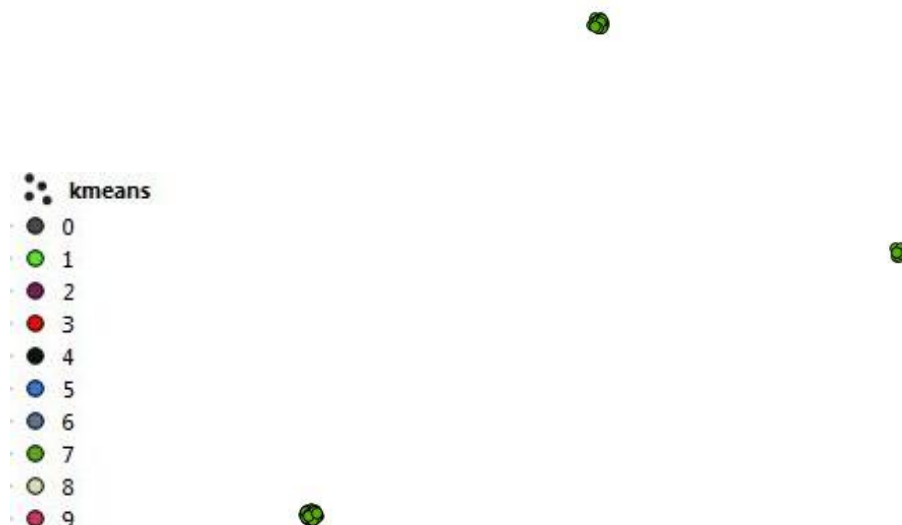
Τα αποτελέσματα της εφαρμογής του αλγορίθμου k-means χωρίς τη χρήση της παραμέτρου των διανυσμάτων αρχικοποίησης απεικονίστηκαν στο λογισμικό Quantum GIS. Ο αλγόριθμος ομαδοποίησε σωστά κάποια από τα σημεία, ενώ κάποια τα ομαδοποίησε λάθος.



Εικόνα 5.18: Απεικόνιση της ομαδοποίησης των σημείων με τον αλγόριθμο k-means χωρίς τη χρήση διανυσμάτων αρχικοποίησης



Εικόνα 5.19: Εσφαλμένη (αριστερά και κέντρο) και φαινομενικά ορθή (δεξιά) ομαδοποίηση των σημείων



Εικόνα 5.20: Τελικώς εσφαλμένη ομαδοποίηση των σημείων

Όπως φαίνεται στις μεγεθυμένες εικόνες, η ταξινόμηση των σημείων δεν είναι η αναμενόμενη. Για το λόγο αυτό, αποφασίσθηκε η χρήση και της τρίτης παραμέτρου των διανυσμάτων αρχικοποίησης στον αλγόριθμο k-means για τη διεξαγωγή ορθότερων αποτελεσμάτων. Η τρίτη παράμετρος θα υπολογιστεί με τη βοήθεια της συνάρτησης «get_domains_n» που εγκαταστάθηκε στο περιβάλλον της PostgreSQL.

Τα βήματα υλοποίησης του σταδίου εφαρμογής του αλγορίθμου k-means περιγράφονται λεπτομερώς στη συνέχεια:

✓ **Δημιουργία των διανυσμάτων αρχικοποίησης και του «πίνακα kmeans_polygons».** Εγκαταστάθηκε στο λογισμικό PostgreSQL η συνάρτηση «get_domains_n» η οποία υπολογίζει ομάδες σημείων με βάση την απόσταση των σημείων μεταξύ τους. Με τη συνάρτηση αυτή πραγματοποιείται μία πρώτη ομαδοποίηση χωρίς την απαίτηση του ορισμού του αριθμού των ομάδων που θα δημιουργηθούν. Αναλυτικότερα εκτελούνται οι εξής διεργασίες:

- Η συνάρτηση υπολογίζει τις ομάδες με μοναδική παράμετρο τη μέγιστη απόσταση των σημείων μεταξύ τους σε m . Από την εφαρμογή της συνάρτησης προκύπτει k αριθμός πολυγώνων ομαδοποίησης.
- Υπολογίζεται η γεωμετρία των κεντροειδών των πολυγώνων ομαδοποίησης καθώς επίσης και οι γεωγραφικές τους συντεταγμένες.
- Δημιουργείται ο πίνακας «kmeans_points» ο οποίος περιλαμβάνει τη γεωμετρία των πολυγώνων ομαδοποίησης, τη γεωμετρία των κεντροειδών των πολυγώνων και τις γεωγραφικές τους συντεταγμένες.

Οι γεωγραφικές συντεταγμένες των k πολυγώνων ομαδοποίησης θα αποτελέσουν τα διανύσματα αρχικοποίησης με βάση τα οποία θα πραγματοποιηθεί η τελική ταξινόμηση με τον αλγόριθμο k-means. Επίσης ο αριθμός k των πολυγώνων που δημιουργήθηκαν θα αποτελέσει τη δεύτερη παράμετρο που ορίζει ο χρήστης στον αλγόριθμο k-means δηλαδή τον αριθμό των ομάδων που επιθυμεί να δημιουργηθούν.

✓ **Εφαρμογή του αλγορίθμου k-means και δημιουργία του πίνακα «kmeans_points».** Εφόσον υπολογίστηκαν τα διανύσματα αρχικοποίησης και ο αριθμός των ομάδων που πρόκειται να δημιουργηθούν, μπορεί πλέον να εφαρμοστεί ο αλγόριθμος k-means για το σύνολο των σημειακών δεδομένων των καταγραφών. Το κάθε σημείο κατατάσσεται σε μία ομάδα με συγκεκριμένο κωδικό. Επομένως, κάθε υποσύνολο σημείων που ανήκουν στην ίδια ομάδα έχουν τον ίδιο κωδικό k-means. Ο πίνακας «kmeans_points» δημιουργείται ως αποτέλεσμα της εφαρμογής του αλγορίθμου k-means στο σύνολο των σημείων των καταγραφών. Πέραν από τον κωδικό k-means της ομάδας που αντιστοιχεί σε κάθε σημείο, ο πίνακας περιλαμβάνει και τα πεδία που καθορίζουν τις γεωγραφικές συντεταγμένες, την αξιολόγηση και τον τύπο προβλήματος κάθε σημείου.

✓ **Επιλογή των μεμονωμένων περιπτώσεων προβλημάτων και δημιουργία του πίνακα «unique_cases».** Ο πίνακας «unique_cases» προκύπτει ως αποτέλεσμα της επιλογής των μεμονωμένων περιπτώσεων προβλημάτων με βάση την τιμή k-means κάθε σημείου από τον πίνακα «kmeans_points». Για κάθε μία τιμή k-means, δηλαδή για κάθε διαφορετική περίπτωση προβλήματος,

υπολογίζεται ένα νέο σημείο ως ο μέσος όρος των σημείων των καταγραφών που πραγματοποιήθηκαν για το συγκεκριμένο πρόβλημα. Επιπλέον, υπολογίζονται η μέση επικινδυνότητα κάθε περίπτωσης που προκύπτει από το μέσο όρο των αξιολογήσεων των καταγραφών για το εκάστοτε πρόβλημα, καθώς επίσης και ο αριθμός των καταγραφών που καταχωρήθηκαν για κάθε περίπτωση προβλήματος. Τέλος, ενημερώνεται η στήλη «unique_cases_id» με την τιμή k-means που αντιστοιχεί σε κάθε σημείο.

Συνοψίζοντας υλοποιήθηκαν τρεις πίνακες στο στάδιο αυτό:

- **kmeans_polygons.** Στον πίνακα «kmeans_polygons» καταχωρούνται τα δεδομένα της πρώτης ταξινόμησης, δηλαδή η γεωμετρία των πολυγώνων ομαδοποίησης, καθώς επίσης η γεωμετρία και οι γεωγραφικές συντεταγμένες των κεντροειδών τους. Οι ονομασίες των πεδίων του πίνακα «kmeans_polygons» είναι:

- id
- polygons_gm
- centroids_geom
- centroids_latitude
- centroids_longitude

Η ακόλουθη εικόνα παρουσιάζει ενδεικτικά αποτελέσματα της δημιουργίας του πίνακα «kmeans_polygons».

| | polygons_gm geometry | centroids_geom geometry(Point,4326) | centroids_latitude double precision | centroids_longitude double precision |
|-----------|---------------------------------|--|--|---|
| 1 | 0103000020E6 | 0101000020E6100000 | 38.019545555597 | 23.6933292264938 |
| 2 | 0103000020E6 | 0101000020E6100000 | 38.000781235241 | 23.7275508434095 |
| 3 | 0103000020E6 | 0101000020E6100000 | 37.986440625682 | 23.7612895340106 |
| 4 | 0103000020E6 | 0101000020E6100000 | 37.968731034436 | 23.6994792121014 |
| 5 | 0103000020E6 | 0101000020E6100000 | 37.975979057574 | 23.7375979593013 |
| 6 | 0103000020E6 | 0101000020E6100000 | 37.987454112026 | 23.6944327064328 |
| 7 | 0103000020E6 | 0101000020E6100000 | 38.005646084028 | 23.7003391202935 |
| 8 | 0103000020E6 | 0101000020E6100000 | 37.995611144312 | 23.7491798078831 |
| 9 | 0103000020E6 | 0101000020E6100000 | 38.012928643365 | 23.7531063451771 |
| 10 | 0103000020E6 | 0101000020E6100000 | 37.988720410056 | 23.7168131710636 |

Εικόνα 5.21: Δημιουργία του πίνακα «kmeans_polygons»

- **kmeans_points.** Στον πίνακα «kmeans_points» υπολογίζεται η τιμή k-means για κάθε σημείο που έχει καταγραφεί και περιλαμβάνει κάποια από τα πεδία δεδομένα που περιγράφουν την κάθε καταγραφή. Ο πίνακας αυτός περιλαμβάνει οκτώ στήλες:

- kmeans
- latitude
- longitude
- severity
- type_of_damage
- cd_id
- id
- geom

Η ακόλουθη εικόνα παρουσιάζει ενδεικτικά αποτελέσματα της δημιουργίας του πίνακα «kmeans_points».

| | kmeans integer | latitude double precis | longitude double precis | severity integer | type_of_damage integer | cd_id integer | id [PK] serial | geom geometry |
|---|-------------------|---------------------------|----------------------------|---------------------|---------------------------|------------------|-------------------|------------------|
| 1 | 0 | 38.01933861 | 23.69322673 | 2 | 8 | 557 | 1 | 010100000: |
| 2 | 0 | 38.01937843 | 23.69329326 | 5 | 8 | 519 | 2 | 010100000: |
| 3 | 0 | 38.01938282 | 23.69351607 | 1 | 8 | 531 | 3 | 010100000: |
| 4 | 0 | 38.01941590 | 23.69332097 | 1 | 8 | 503 | 4 | 010100000: |
| 5 | 0 | 38.01942529 | 23.69346432 | 1 | 8 | 504 | 5 | 010100000: |
| 6 | 0 | 38.01944348 | 23.69313221 | 2 | 8 | 559 | 6 | 010100000: |

Εικόνα 5.22: Δημιουργία του πίνακα «kmeans_points»

➤ **unique_cases.** Στον πίνακα «unique_cases» καταχωρούνται όλες οι μοναδικές περιπτώσεις προβλημάτων έτσι όπως προέκυψαν από τον πίνακα «unique_cases». Ο πίνακας αυτός περιλαμβάνει οκτώ στήλες:

- kmeans
- uc_damage_types
- avg_latitude
- avg_longitude
- avg_severity
- num_of_records
- uc_id
- geom

Η ακόλουθη εικόνα παρουσιάζει ενδεικτικά αποτελέσματα της δημιουργίας του πίνακα «unique_cases».

| | kmeans integer | uc_damage_ integer | avg_latitude double precis | avg_longitud double precis | avg_severity numeric | num_of_reco bigint | uc_id [PK] serial | geom geometry(Po |
|----|-------------------|-----------------------|-------------------------------|-------------------------------|-------------------------|-----------------------|----------------------|---------------------|
| 1 | 0 | 8 | 38.01955713 | 23.69335511 | 2.970588235 | 68 | 1 | 0101000020E |
| 2 | 1 | 4 | 38.00078886 | 23.72756240 | 3.065573770 | 61 | 2 | 0101000020E |
| 3 | 2 | 1 | 37.98644328 | 23.76128932 | 3.087912087 | 91 | 3 | 0101000020E |
| 4 | 3 | 5 | 37.96872025 | 23.69950834 | 3.125000000 | 88 | 4 | 0101000020E |
| 5 | 4 | 3 | 37.97598742 | 23.73760124 | 3.072916666 | 96 | 5 | 0101000020E |
| 6 | 5 | 10 | 37.98746835 | 23.69441041 | 3.392857142 | 28 | 6 | 0101000020E |
| 7 | 6 | 9 | 38.00564595 | 23.70033455 | 3.039473684 | 76 | 7 | 0101000020E |
| 8 | 7 | 6 | 37.99561593 | 23.74916278 | 2.968750000 | 64 | 8 | 0101000020E |
| 9 | 8 | 2 | 38.01292195 | 23.75310632 | 3.000000000 | 34 | 9 | 0101000020E |
| 10 | 9 | 7 | 37.98873417 | 23.71681112 | 2.737704918 | 61 | 10 | 0101000020E |

Εικόνα 5.23: Δημιουργία του πίνακα «unique_cases»

Στον πίνακα που ακολουθεί παρουσιάζεται η δομή των πινάκων «kmeans_polygons», «kmeans_points» και «unique_cases» οι οποίοι δημιουργήθηκαν στο στάδιο αυτό, καθώς επίσης και το όνομα των στηλών, οι τύποι δεδομένων κάθε στήλης, καθώς επίσης και τα πρωτεύοντα και τα δευτερεύοντα κλειδιά. Μετά την ολοκλήρωση της δημιουργίας των τριών νέων πινάκων στη βάση δεδομένων, προστίθεται στο πεδίο «unique_cases_id» του πίνακα «crowd_data» το πεδίο «kmeans» του πίνακα «unique_cases» ως ξένο κλειδί και ενημερώνεται το πεδίο «unique_cases_id» με βάση την τιμή «k-means» του πίνακα «kmeans_points».

| Πίνακας | Στήλη | Τύπος Δεδομένων | Πρωτεύον κλειδί | Ξένο κλειδί |
|------------------------|---------------------|------------------|-----------------|----------------------------------|
| kmeans_polygons | id | serial | ✓ | |
| | polygons_geom | geometry | | |
| | centroids_geom | geometry | | |
| | centroids_latitude | double precision | | |
| | centroids_longitude | double precision | | |
| kmeans_points | kmeans | integer | | |
| | latitude | double precision | | |
| | longitude | double precision | | |
| | severity | integer | | |
| | type_of_damage | integer | | |
| | cd_id | integer | | REFERENCES crowd_data (id) |
| | id | serial | ✓ | |
| unique_cases | geom | geometry | | |
| | kmeans | integer | | |
| | uc_damage_types | integer | | REFERENCES damage_types (t_code) |
| | avg_latitude | double precision | | |
| | avg_longitude | double precision | | |
| | avg_severity | numeric | | |
| | num_of_records | bigint | | |
| | uc_id | serial | ✓ | |
| crowd_data | geom | geometry | | |
| | unique_cases_id | integer | | REFERENCES unique_cases (kmeans) |

Πίνακας 5.5: Οι πίνακες **kmeans_polygons**, **kmeans_points** και **unique_cases** της βάσης δεδομένων, οι στήλες, οι τύποι δεδομένων και τα κλειδιά τους, καθώς και η προσθήκη ξένου κλειδιού στον πίνακα **crowd_data**

5.4 Δομή και Εμφάνιση των Αποτελεσμάτων

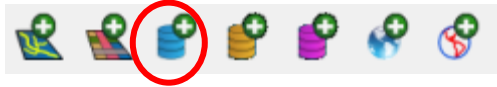
Τα νέα δεδομένα που προέκυψαν από την εφαρμογή του αλγορίθμου k-means στα αρχικά δεδομένα θα απεικονιστούν με δύο τρόπους: στο περιβάλλον του Quantum GIS και στο διαδίκτυο μετά τη δημιουργία KML αρχείων.

5.4.1 Απεικόνιση αποτελεσμάτων στο Quantum GIS

Η πρώτη οπτικοποίηση των αποτελεσμάτων πραγματοποιείται στο λογισμικό ανοικτού κώδικα Quantum GIS. Το μόνο που απαιτείται για την απεικόνιση των αποτελεσμάτων στο QGIS είναι η δημιουργία μίας σύνδεσης με τη βάση δεδομένων του λογισμικού PostgreSQL κατά την προσθήκη των postgis επιπέδων.

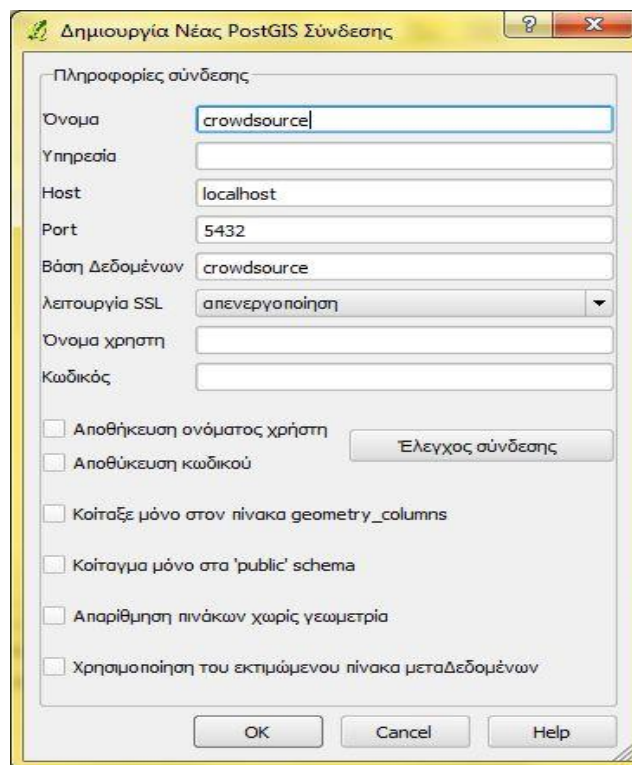
Στην συνέχεια περιγράφονται οι κατάλληλες ρυθμίσεις (configuration) που πρέπει να γίνουν στο γεωγραφικό σύστημα ώστε να μπορεί να αποκατασταθεί η σύνδεση με την βάση δεδομένων για την ανάκτηση χωρικής πληροφορίας.

Οι πίνακες που προέκυψαν στη βάση δεδομένων εισάγονται στο περιβάλλον του Quantum GIS από το μενού επίπεδο → add PostGIS layer είτε από το αντίστοιχο εικονίδιο:



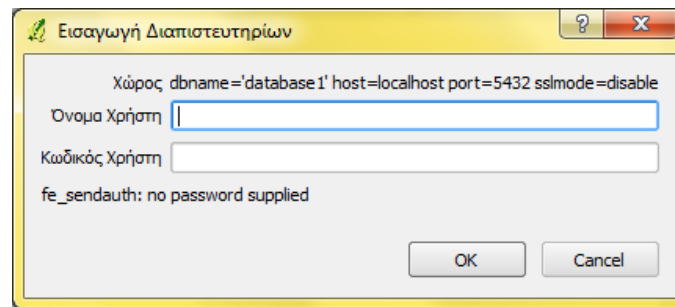
Από το κουμπί «νέο» εμφανίζεται το παράθυρο «*Create a new PostGIS connection*». Στο παράθυρο αυτό συμπληρώνονται οι παράμετροι:

- Όνομα: Επιλέγεται το όνομα που επιθυμούμε να δοθεί στη σύνδεση.
- Host: Πληκτρολογείται localhost.
- Database: Πληκτρολογείται το όνομα της βάσης δεδομένων database1, από την οποία θα αντληθούν τα δεδομένα.
- Port: Επιλέγεται η προεπιλεγμένη θύρα 5432.
- Username: Πληκτρολογείται το όνομα χρήστη το οποίο είναι postgres.
- Password: Πληκτρολογείται ο κωδικός που έχει οριστεί από τον χρήστη.



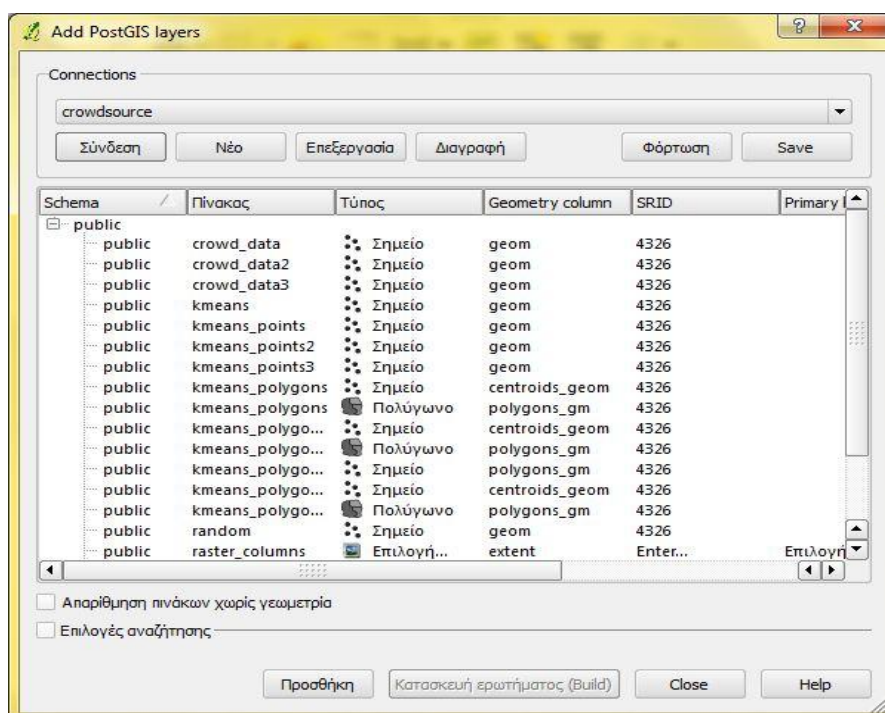
Εικόνα 5.24: Δημιουργία PostGIS σύνδεσης στο περιβάλλον του QuantumGIS

Εφόσον υλοποιηθεί η σύνδεση την πρώτη φορά, επιλέγοντας το όνομα που δόθηκε στην παραπάνω σύνδεση μπορεί να πραγματοποιείται η σύνδεση μεταξύ Quantum GIS και PostGIS στις μετέπειτα χρήσεις του λογισμικού για την εισαγωγή των δεδομένων από το κουμπί *σύνδεση*.



Εικόνα 5.25: Εισαγωγή ονόματος και κωδικού χρήστη

Αφότου γίνει η σύνδεση εμφανίζεται η λίστα με τα επίπεδα της βάσης δεδομένων που δημιουργήθηκε στο λογισμικό του PostGIS. Ποιο συγκεκριμένα εμφανίζονται οι οντότητες οι οποίες έχουν χωρική διάσταση, δηλαδή συγκεκριμένη γεωμετρία. Οι πίνακες που δεν έχουν γεωμετρία δεν μπορούν να απεικονιστούν.



Εικόνα 5.26: Επιλογή των πινάκων και προσθήκη αυτών στη βάση του PostGIS

Με κλικ στο κουμπί «προσθήκη» εισάγονται τα δεδομένα. Τα layers που εισάγονται στο Quantum GIS είναι οι πίνακες «crowd_data», «kmeans_polygons», και «unique_cases» για την ανάδειξη της πορείας των ακατέργαστων δεδομένων πριν τη δημιουργία των χαρτών με τη δημιουργία kml αρχείων. Αναλυτικότερα:

- **«crowd_data».** Το πλέον ενημερωμένο με τον k-means επίπεδο crowd_data αναπαριστά όλα τα αρχικά σημεία των καταγραφών των προβλημάτων που πραγματοποιήθηκαν. Με δεξί κλικ επάνω στο layer και επιλογή properties, επιλέγεται η απεικόνιση των σημείων κατηγοριοποιημένων με βάση την τιμή unique_cases_id. Έτσι, τα σημεία που ανήκουν στην ίδια ομάδα απεικονίζονται με όμοιο χρώμα και αφορούν στην ίδια περίπτωση προβλήματος.

- **«kmeans_polygons»**. Το επίπεδο kmeans_polygons απεικονίζει τα πολύγωνα που προκύπτουν από την πρώτη ομαδοποίηση των δεδομένων, εντός των οποίων περιέχονται όλα εκείνα τα σημεία που ανήκουν στην ίδια ομάδα ταξινόμησης.
- **«unique_cases»**. Το επίπεδο unique_cases απεικονίζει όλες τις μεμονωμένες περιπτώσεις προβλημάτων με βάση τον μέσο όρο των σημείων κάθε ομάδας προβλημάτων. Η απεικόνιση των σημείων μπορεί να κατηγοριοποιηθεί με βάση τον k-means αντίστοιχα με το επίπεδο «crowd_data».

5.4.2 Απεικόνιση αποτελεσμάτων στο Google Maps και σε προσωπική ιστοσελίδα

Πέραν από την απεικόνιση των αποτελεσμάτων στο λογισμικό Quantum GIS, τα τελικά δεδομένα της εφαρμογής θα απεικονιστούν επίσης σε χάρτες στο περιβάλλον του Google Maps αλλά και σε προσωπική ιστοσελίδα που δημιουργείται στη συνέχεια. Για να υλοποιηθούν οι χάρτες πρέπει αρχικά να δημιουργηθούν KML αρχεία που θα περιέχουν όλα τα απαραίτητα δεδομένα για την αναπαράσταση των αποτελεσμάτων. Για κάθε μία περίπτωση προβλήματος δημιουργούνται δύο KML αρχεία, ένα για την απεικόνιση όλων των καταγραφών για το συγκεκριμένο πρόβλημα και ένα για την απεικόνιση του σημείου που αντιστοιχεί στο μέσο όρο των σημείων της συγκεκριμένης περίπτωσης προβλήματος.

Για τη δημιουργία των δύο KML αρχείων για κάθε περίπτωση προβλήματος, υλοποιήθηκαν ένα αρχείο με τον αντίστοιχο αλγόριθμο σε γλώσσα PHP. Η λειτουργία του αρχείου αναλύεται στη συνέχεια:

- ✓ **Δημιουργία kml αρχείων για την απεικόνιση του συνόλου των καταγραφών για κάθε περίπτωση προβλήματος.** Στόχος είναι η δημιουργία ενός KML αρχείου για κάθε περίπτωση προβλήματος, με σκοπό την απεικόνιση του συνόλου των σημείων σε χάρτη με βάση την τιμή k-means κάθε σημείου. Τα στάδια παραγωγής του KML αρχείου περιγράφονται στη συνέχεια:
 - Αρχικά ανακτώνται όλες οι απαραίτητες πληροφορίες για το σύνολο των σημείων των καταγεγραμμένων προβλημάτων από τον πίνακα «crowd_data». Τα πεδία που είναι απαραίτητα για τη δόμηση του KML αρχείου είναι τα «latitude», «longitude», «name», «surname», «type_of_damage», «phone_number» και «e_mail», προκειμένου στη συνέχεια να δημιουργηθεί ένα KML αρχείο για κάθε περίπτωση προβλήματος το οποίο να διαθέτει τις πληροφορίες αυτές για την κάθε καταγραφή.
 - Έπειτα ανακτώνται από τον πίνακα «damage_types» οι πληροφορίες που περιγράφουν το χρώμα και το στυλ κάθε εικονιδίου που θα αντιπροσωπεύει τον εκάστοτε τύπο προβλήματος στο χάρτη. Σε κάθε τύπο προβλήματος αντιστοιχίζεται το χρώμα που έχει οριστεί στον πίνακα «damage_types».
 - Στη συνέχεια δημιουργείται ένα Placemark για κάθε σημείο, το οποίο περιέχει όλα τα δεδομένα που ανακτήθηκαν από τους πίνακες για το συγκεκριμένο σημείο.
 - Δημιουργείται ένα KML αρχείο για κάθε τύπο προβλήματος όπου εισάγονται τα Placemarks. Αφού το KML αρχείο αποκτήσει την κατάλληλη δομή, μπορεί

να απεικονιστεί στο Google Maps και στην προσωπική ιστοσελίδα που περιγράφεται στη συνέχεια.

- ✓ **Δημιουργία kml αρχείων για την απεικόνιση του σημείου του μέσου όρου των καταγραφών για κάθε περίπτωση προβλήματος.** Στόχος είναι η δημιουργία ενός KML αρχείου για κάθε περίπτωση προβλήματος, με σκοπό την απεικόνιση σε χάρτη του σημείου που αντιστοιχεί στη μέση τιμή των καταγραφών για το συγκεκριμένο πρόβλημα με βάση την τιμή k-means κάθε σημείου. Τα στάδια παραγωγής του KML αρχείου είναι όμοια με εκείνα της παραγωγής του πρώτου KML αρχείου.

Αφότου δημιουργηθούν όλα τα kml αρχεία που περιγράφηκαν παραπάνω, ακολουθεί η απεικόνιση των αποτελεσμάτων στο Google Maps και σε προσωπική ιστοσελίδα μέσω της δημιουργίας ενός html αρχείου. Για να απεικονιστούν τα αποτελέσματα στο διαδίκτυο πρέπει να πραγματοποιηθούν με τη σειρά οι παρακάτω ενέργειες:

- **Δημιουργία προσωπικής ιστοσελίδας.** Κάθε φοιτητής του πολυτεχνείου έχει δικαίωμα να ενεργοποιήσει την προσωπική του ιστοσελίδα από την υπηρεσία my.ntua.gr αφού κάνει login στην υπηρεσία με το username και το password που του έχει αποδοθεί από το κέντρο Η/Υ, επιλέγοντας ως mail server τον mail.ntua.gr. Στη σελίδα αυτή θα φιλοξενηθούν οι χάρτες με τα αποτελέσματα που προέκυψαν από την υλοποίηση της εφαρμογής.
- **Δημιουργία του αρχείου index.html.** Στο στάδιο αυτό δημιουργείται το αρχείο index.html το οποίο περιέχει όλα τα απαραίτητα δεδομένα για την απεικόνιση των αποτελεσμάτων σε χάρτες στο διαδίκτυο. Η δημιουργία του αρχείου πραγματοποιείται μέσα από την υλοποίηση του κώδικα του rhr αρχείου που παράγει και τα kml αρχεία. Ο κώδικας **ενημερώθηκε κατάλληλα** ώστε πέραν από τα kml αρχεία να δημιουργείται και το αντίστοιχο html αρχείο. Το αρχείο αυτό περιλαμβάνει στο μεγαλύτερο μέρος του ένα τμήμα κώδικα σε γλώσσα JavaScript, το οποίο καθορίζει αναλυτικά την εμφάνιση του χάρτη και των επιπέδων (layers) που θα απεικονιστούν σε αυτόν στο διαδίκτυο. Τα δεδομένα των layers προέρχονται από τα kml αρχεία που δημιουργήθηκαν στο προηγούμενο στάδιο.
- **Πρόσβαση στον αποθηκευτικό χώρο.** Η πρόσβαση στον αποθηκευτικό χώρο γίνεται μέσω οποιασδήποτε FTP εφαρμογής στη διεύθυνση users.ntua.gr με χρήση του κωδικού ασφαλείας που έχει παραχωρηθεί στο χρήστη από το κέντρο ηλεκτρονικού υπολογιστή. Το λειτουργικό σύστημα Windows στις εκδόσεις XP, Vista, 2007 παρέχει ενσωματωμένη στο σύστημα εφαρμογή που επιτρέπει την εύκολη πρόσβαση στις προσωπικές ιστοσελίδες μέσω της δημιουργίας μίας θέσης δικτύου (My Network Places). Αφότου δημιουργηθεί η θέση δικτύου μπορούμε μέσα από το My Network places να συνδεόμαστε στην υπηρεσία πληκτρολογώντας το username και το password.
- **Προσθήκη αρχείων στο My Network Places.** Στο στάδιο αυτό προστίθενται όλα τα εξαγόμενα kml αρχεία και το αρχείο index.html στη θέση του My Network

Places που δημιουργήσαμε, ώστε να τα απεικονίσουμε στη συνέχεια στο Google Maps και στην προσωπική ιστοσελίδα που δημιουργήθηκε νωρίτερα.

- **Απεικόνιση στην προσωπική ιστοσελίδα.** Ανατρέχοντας σε οποιονδήποτε διαδικτυακό browser και πληκτρολογώντας τη διεύθυνση της ιστοσελίδας που δημιουργήθηκε νωρίτερα <http://users.ntua.gr/ma1r1str/>, τα αποτελέσματα των καταγραφών απεικονίζονται σε διαδικτυακό χάρτη.

- **Απεικόνιση στο Google Maps.** Για την απεικόνιση των αποτελεσμάτων σε χάρτες του Google Maps το μόνο που χρειάζεται είναι η ύπαρξη των kml αρχείων με τα αποτελέσματα στη θέση του My Network Places. Αφού έχουμε προσθέσει τα kml στη θέση δικτύου, ανοίγουμε μία σελίδα περιήγησης και πληκτρολογούμε στον browser τη διεύθυνση τη <http://maps.google.com/maps?q=> και στη συνέχεια επικολλάμε το public link του kml αρχείου που επιθυμούμε να απεικονίσουμε κάθε φορά.

6

ΠΑΡΑΔΕΙΓΜΑ ΜΕ ΕΝΔΕΙΚΤΙΚΑ ΔΕΔΟΜΕΝΑ

Στο κεφάλαιο αυτό παρουσιάζονται τρεις περιπτώσεις υλοποίησης της εφαρμογής που περιγράφηκε στις προηγούμενες ενότητες ως προς τα δεδομένα που χρησιμοποιήθηκαν. Στην πρώτη περίπτωση, το σύστημα υλοποιήθηκε για ένα σχετικά μικρό αριθμό δεδομένων καταγραφών, στη δεύτερη περίπτωση εφαρμόστηκε για έναν αισθητά μεγαλύτερο όγκο δεδομένων, ενώ στην τρίτη περίπτωση εξετάστηκε το ενδεχόμενο της γεινίασης των καταγραφών δύο ή και παραπάνω όμοιων τύπων προβλημάτων και η αντίδραση του αλγορίθμου k-means.

6.1 Περίπτωση 1

Στο πρώτο παράδειγμα της εφαρμογής μελετήθηκε η περίπτωση αναφοράς 10 διαφορετικών περιπτώσεων προβλημάτων, για κάθε μία εκ των οποίων έχουν πραγματοποιηθεί 1 έως 100 καταγραφές.

6.1.1 Παράμετροι της εφαρμογής

Οι βασικές παράμετροι που είναι απαραίτητες για την υλοποίηση της εφαρμογής, ορισμένες εκ των οποίων διαφοροποιούνται ανά περίπτωση ενώ κάποιες παραμένουν σταθερές και στις τρεις περιπτώσεις, είναι οι εξής:

- ☞ Οι γεωγραφικές συντεταγμένες (φ, λ) **10 αρχικών σημείων**, τα οποία θα αποτελέσουν τις δέκα υπό μελέτη περιπτώσεις προβλημάτων που θα καταγράψει το πλήθος.
- ☞ Η μέγιστη και η ελάχιστη τιμή της γωνίας και της ακτίνας των πολικών συντεταγμένων των σημείων που πρόκειται να δημιουργηθούν και θα αποτελέσουν τις καταγραφές των προβλημάτων για κάθε ένα από τα δέκα προβλήματα. Η τιμή της γωνίας κυμαίνεται από 0 έως 360 μοίρες ενώ η τιμή της ακτίνας κυμαίνεται από 0 έως 25 μέτρα. Οι τιμές αυτές υποδηλώνουν το εύρος της περιοχής γύρω από τις 10 περιπτώσεις προβλημάτων εντός της οποίας θα δημιουργηθούν τα τυχαία σημεία που θα αποτελέσουν τις καταγραφές του πλήθους.
- ☞ Ο μέγιστος αριθμός τυχαίων σημείων – καταγραφών που θα δημιουργηθούν για κάθε περίπτωση προβλήματος. Για κάθε περίπτωση προβλήματος επιλέχθηκε να πραγματοποιούνται έως 100 καταγραφές. Ο αριθμός των καταγραφών επιλέγεται τυχαία με τη βοήθεια της συνάρτησης $\text{rand}(1,100)$.
- ☞ Οι παράμετροι μετατροπής των επίπεδων συντεταγμένων x, y του συστήματος ΕΓΣΑ87 σε γεωγραφικές συντεταγμένες φ, λ του συστήματος WGS84 και αντίστροφα,

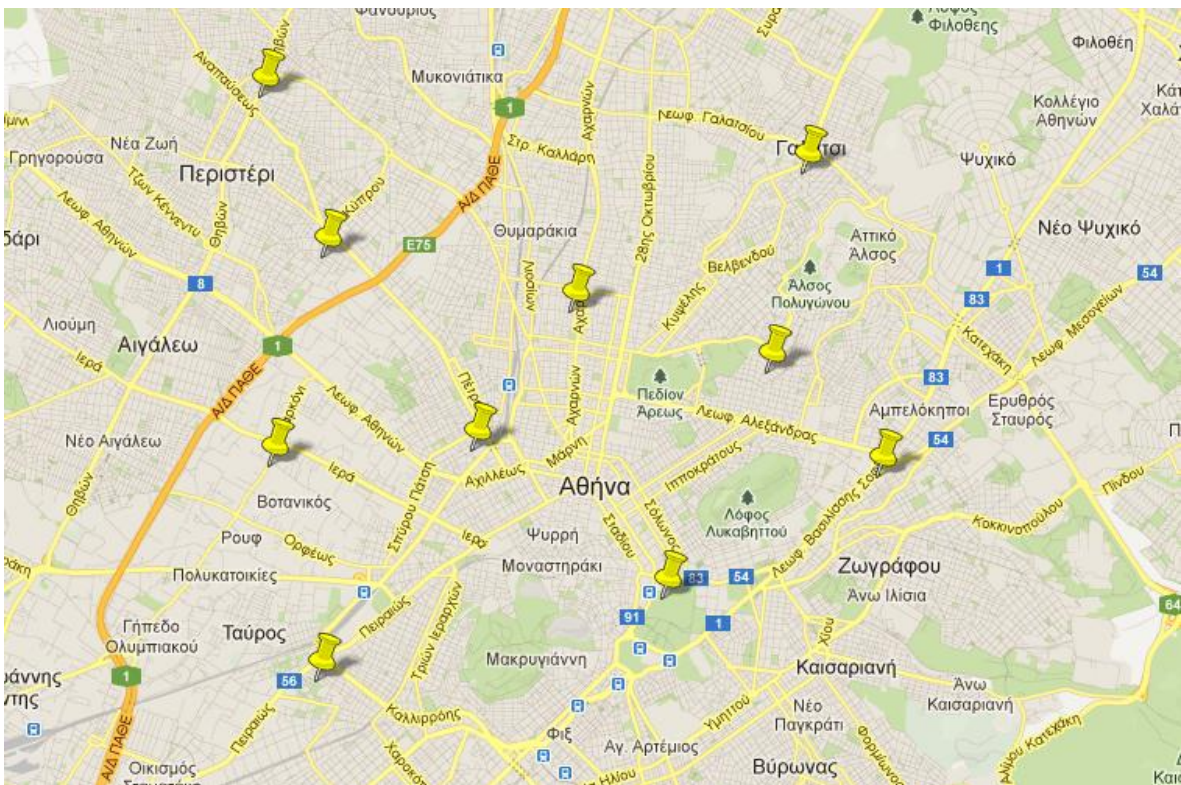
οι οποίες είναι σταθερές και δεν μεταβάλλονται από περίπτωση σε περίπτωση για τα δεδομένα της Ελλάδας.

- ☞ Η απόσταση που καθορίζει το εύρος της ομαδοποίησης των σημείων για τη δημιουργία των πολυγώνων ομαδοποίησης. Στην περίπτωση που μελετάμε η απόσταση πήρε τιμή ίση με 30 m.

6.1.2 Δεδομένα εισόδου

Στην πρώτη περίπτωση που υλοποιήθηκε, μελετήθηκε η εφαρμογή του συστήματος με τη χρήση ιδανικών αρχικών δεδομένων. Έτσι, επιλέχθηκαν χειροκίνητα στο περιβάλλον του Google Earth δέκα σημεία σε ικανοποιητική απόσταση μεταξύ τους, καθένα εκ των οποίων αντιπροσωπεύει ένα διαφορετικό τύπο προβλήματος. Γύρω από κάθε μία από τις δέκα περιπτώσεις προβλήματος θα δημιουργηθούν στη συνέχεια τυχαία σημεία, τα οποία θα αντιπροσωπεύουν τις καταγραφές που πραγματοποιήθηκαν για το συγκεκριμένο πρόβλημα.

Αποφασίστηκε τα δέκα αρχικά σημεία να μη δημιουργηθούν με την αναπαραγωγή τυχαίων συντεταγμένων έτσι ώστε να μη συμπίπτουν με κτίρια ή άλλα αντικείμενα. Για το λόγο αυτό επιλέχθηκε να βρίσκονται κοντά σε διασταυρώσεις δρόμων, ώστε να προσομοιάζουν πραγματικές περιπτώσεις καταγραφών. Τα σημεία αυτά παρουσιάζονται στο απόσπασμα του χάρτη του Google Maps που φαίνεται στη συνέχεια.



Εικόνα 6.1: Απεικόνιση των δέκα αρχικών σημείων στο περιβάλλον του Google Maps

Οι γεωγραφικές συντεταγμένες των δέκα αρχικών σημείων που χρησιμοποιήθηκαν ως κέντρα για την αναπαραγωγή νέων τυχαίων σημείων σε μέγιστη απόσταση 25 μέτρα παρουσιάζονται στον πίνακα που ακολουθεί.

| α/α | φ | λ |
|-----|-----------|-----------|
| 1 | 37.986499 | 23.761300 |
| 2 | 38.012974 | 23.753113 |
| 3 | 37.976029 | 23.737596 |
| 4 | 38.000833 | 23.727542 |
| 5 | 37.968781 | 23.699485 |
| 6 | 37.995662 | 23.749169 |
| 7 | 37.988790 | 23.716786 |
| 8 | 38.019600 | 23.693327 |
| 9 | 38.005699 | 23.700340 |
| 10 | 37.987502 | 23.694391 |

Εικόνα 6.2: Γεωγραφικές συντεταγμένες των δέκα αρχικών σημείων

6.1.3 Υλοποίηση της εφαρμογής

Αφού δημιουργήθηκε η βάση δεδομένων και διαμορφώθηκαν τα αναγκαία ερωτήματα για τη δημιουργία των πινάκων, την εισαγωγή των δεδομένων, και την επιλογή δεδομένων από τους πίνακες σε γλώσσα SQL, ενσωματώθηκε ο κώδικας αυτός στον τελικό κώδικα στη γλώσσα PHP, και τέθηκε το σύστημα σε λειτουργία αφήνοντάς το να «τρέξει» μέσω του παραθύρου της γραμμής εντολών.

Τα αρχεία του αλγορίθμου που τέθηκαν σε λειτουργία είναι πέντε:

- **F1_create_tables.php**, για τη δημιουργία των πινάκων «damage_types» και «crowd_data» στη βάση δεδομένων.
- **F1_data_insertion_damage_types.php**, για την εισαγωγή των δεδομένων στον πίνακα «damage_types».
- **F1_create_and_insert_points_crowd_data.php**, για τη δημιουργία των τυχαίων σημείων των καταγραφών γύρω από τα δέκα αρχικά σημεία, τη διαμόρφωση του τελικού csv αρχείου με όλες τις πληροφορίες των καταγραφών, και την εισαγωγή όλων των πληροφοριών στον πίνακα «crowd_data».
- **F1_kmeans.php**, για την υλοποίηση της ομαδοποίησης των καταγραφών των προβλημάτων με χρήση του αλγορίθμου `-kmeans` και την προσθήκη των πινάκων «kmeans_polygons», «kmeans_points» και «unique_cases» στη βάση δεδομένων.
- **F1_kml_points_and_html.php**, για τον υπολογισμό των μοναδικών περιπτώσεων προβλημάτων και τη δημιουργία ενός kml αρχείου για κάθε τύπο προβλήματος, τη δημιουργία ενός kml αρχείου για κάθε τύπο προβλήματος για το σύνολο των δεδομένων, και τη δημιουργία ενός html αρχείου για την απεικόνιση των δεδομένων των kmls σε διαδικτυακούς χάρτες.

Η εισαγωγή των δεδομένων στον πίνακα «damage_types» πραγματοποιήθηκε με άντληση της πληροφορίας από txt αρχεία, ενώ η εισαγωγή των δεδομένων στον πίνακα «crowd_data» πραγματοποιήθηκε με άντληση της πληροφορίας από κατάλληλα διαμορφωμένο csv αρχείο.

Μετά την ολοκλήρωση του «τρεξίματος» των πέντε php αρχείων, προέκυψαν πέντε διαφορετικοί πίνακες με τα αντίστοιχα δεδομένα μέσα στη βάση δεδομένων του λογισμικού PostgreSQL. Επιπλέον, εξήχθησαν δύο kml αρχεία για κάθε τύπο προβλήματος με τα αποτελέσματα που προέκυψαν από την υλοποίηση του κώδικα.

Στην εικόνα που ακολουθεί παρουσιάζεται η δομή των πινάκων και τα πεδία κάθε πίνακα που δημιουργήθηκαν στη βάση δεδομένων έπειτα από την υλοποίηση του αλγορίθμου.

| crowd data | damage types | kmeans polygons | kmeans points | unique cases |
|--|--|--|---|--|
| <input type="checkbox"/> id | <input type="checkbox"/> t_code | <input type="checkbox"/> polygons_gm | <input type="checkbox"/> kmeans | <input type="checkbox"/> kmeans |
| <input type="checkbox"/> latitude | <input type="checkbox"/> description | <input type="checkbox"/> centroids_geom | <input type="checkbox"/> latitude | <input type="checkbox"/> uc_damage_types |
| <input type="checkbox"/> longitude | <input type="checkbox"/> colour | <input type="checkbox"/> centroids_latitude | <input type="checkbox"/> longitude | <input type="checkbox"/> avg_latitude |
| <input type="checkbox"/> name | <input type="checkbox"/> kml_unique_colour | <input type="checkbox"/> centroids_longitude | <input type="checkbox"/> severity | <input type="checkbox"/> avg_longitude |
| <input type="checkbox"/> surname | <input type="checkbox"/> kml_crowd_colour | | <input type="checkbox"/> type_of_damage | <input type="checkbox"/> avg_severity |
| <input type="checkbox"/> phone_number | <input type="checkbox"/> styleurl_unique | | <input type="checkbox"/> cd_id | <input type="checkbox"/> num_of_records |
| <input type="checkbox"/> email | <input type="checkbox"/> styleurl_crowd | | <input type="checkbox"/> id | <input type="checkbox"/> uc_id |
| <input type="checkbox"/> city | <input type="checkbox"/> icon | | <input type="checkbox"/> geom | <input type="checkbox"/> geom |
| <input type="checkbox"/> type_of_damage | | | | |
| <input type="checkbox"/> short_description | | | | |
| <input type="checkbox"/> date | | | | |
| <input type="checkbox"/> severity | | | | |
| <input type="checkbox"/> capture | | | | |
| <input type="checkbox"/> updated | | | | |
| <input type="checkbox"/> created | | | | |
| <input type="checkbox"/> project | | | | |
| <input type="checkbox"/> created_by | | | | |
| <input type="checkbox"/> updated_by | | | | |
| <input type="checkbox"/> geom | | | | |
| <input type="checkbox"/> unique_cases_id | | | | |

Εικόνα 6.3: Οι πίνακες που δημιουργήθηκαν στη βάση δεδομένων για την «Περίπτωση 1»

6.1.4 Απεικόνιση ακατέργαστων αποτελεσμάτων στο Quantum GIS

Στην παρούσα ενότητα παρουσιάζεται η πορεία των ακατέργαστων δεδομένων στο περιβάλλον του λογισμικού Quantum GIS, πριν την απεικόνισή τους στους τελικούς διαδικτυακούς χάρτες μέσω των kml αρχείων. Για την απεικόνιση των δεδομένων της βάσης δεδομένων στο Quantum GIS απαιτείται η υλοποίηση σύνδεσης με την συγκεκριμένη βάση δεδομένων όπως περιγράφηκε σε προηγούμενη ενότητα.

Όλοι οι χάρτες που παρουσιάζονται έχουν ως υπόβαθρο χάρτη των καλλικρατικών δήμων της Ελλάδος. Ο χάρτης των δήμων προέρχεται από τη σελίδα geodata.gov.gr η οποία έχει σχεδιαστεί, αναπτυχθεί και συντηρείται από το Ινστιτούτο Πληροφοριακών Συστημάτων του Ερευνητικού Κέντρου «Αθηνά» με σκοπό να αποτελέσει ένα κεντρικό σημείο συλλογής, αναζήτησης, διάθεσης και απεικόνισης της ανοικτής δημόσιας γεωχωρικής πληροφορίας. Ο χάρτης αποθηκεύτηκε ως shapefile αρχείο και το σύστημα αναφοράς των δεδομένων του αρχείου είναι το ΕΓΣΑ87. Πραγματοποιήθηκε επαναπροβολή του χάρτη στο σύστημα WGS84 και στη συνέχεια εισήχθη ως shapefile στη βάση δεδομένων. Έπειτα πραγματοποιήθηκε σύνδεση με το λογισμικό Quantum GIS όπου απεικονίστηκε μαζί με τα υπόλοιπα δεδομένα.

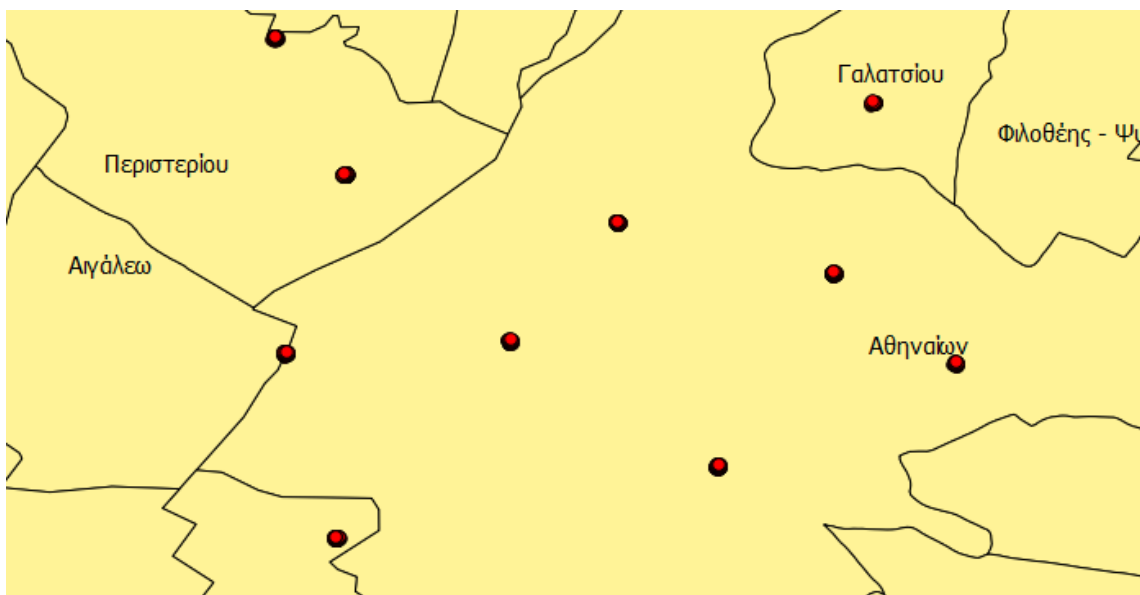
Στους χάρτες απεικονίζονται τα δεδομένα με τη σειρά που υλοποιήθηκαν στη βάση δεδομένων. Πιο συγκεκριμένα απεικονίζονται:

- ✓ τα αρχικά δεδομένα όπως προκύπτουν από τις καταγραφές των προβλημάτων από το πλήθος,
- ✓ τα δεδομένα των καταγραφών που πραγματοποιήθηκαν για ένα συγκεκριμένο τύπο προβλήματος,
- ✓ τα πολύγωνα ομαδοποίησης που προέκυψαν από την υλοποίηση της πρώτης ομαδοποίησης των σημείων για την ορισμό των διανυσμάτων αρχικοποίησης,
- ✓ οι ομάδες σημείων των καταγραφών που προέκυψαν από την υλοποίηση της ομαδοποίησης με τον αλγόριθμο k-means,

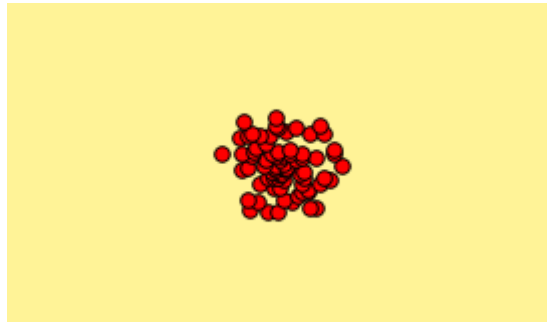
- ✓ και οι μοναδικές περιπτώσεις προβλημάτων που αντιπροσωπεύουν το σύνολο των καταγραφών για κάθε περίπτωση.

Στο σημείο αυτό πρέπει να σημειωθούν οι εξής παρατηρήσεις:

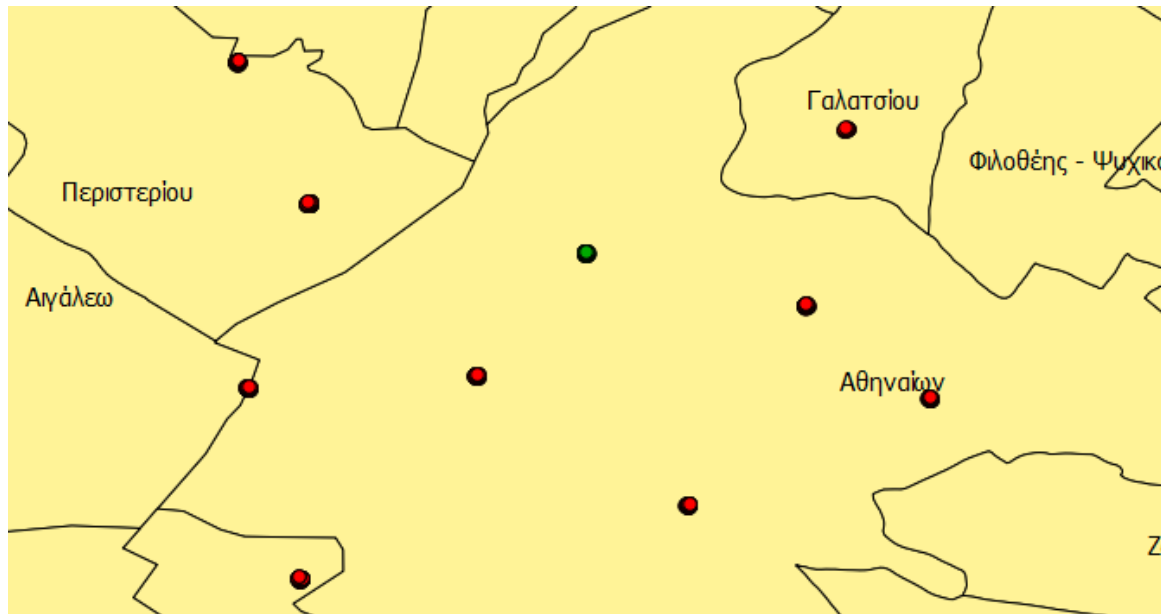
- Στις εικόνες 6.4, 6.6 και 6.9 παρουσιάζεται το σύνολο των σημείων των καταγραφών αλλά επειδή η έκταση που απεικονίζεται είναι μεγάλη με αποτέλεσμα να μη γίνεται μεγέθυνση του χάρτη ώστε να απεικονιστούν όλες οι καταγεγραμμένες περιπτώσεις, τα σημεία των καταγραφών κάθε περίπτωσης φαίνονται ως ένα σημείο. Για το λόγο αυτό αλλά και για να μην υπάρξει ταύτιση των χαρτών αυτών με το χάρτη της εικόνας 6.10, στην εικόνα 6.5 παρουσιάζεται η πραγματική εικόνα των καταγραφών σε μεγέθυνση.
- Στην εικόνα 6.6 παρουσιάζεται το σύνολο των καταγραφών για τον τύπο προβλήματος «χαλασμένο φανάρι» που απεικονίζεται με πράσινο χρώμα, σε σχέση με τις καταγραφές για τα υπόλοιπα προβλήματα που διαφοροποιούνται με κόκκινο χρώμα.
- Στην εικόνα 6.7 παρουσιάζεται μεγεθυμένο απόσπασμα του χάρτη των πολυγώνων ομαδοποίησης, ενώ στο χάρτη 6.8 παρουσιάζεται το πολύγωνο ομαδοποίησης μιας περίπτωσης προβλήματος σε ακόμα μεγαλύτερη μεγέθυνση ώστε να είναι περισσότερο ευδιάκριτο.
- Στην εικόνα 6.9 κάθε ομάδα σημείων που προέκυψε από την ομαδοποίηση k-means απεικονίζεται με το ίδιο χρώμα.
- Τέλος, στην εικόνα 6.10 απεικονίζονται με μπλε χρώμα οι μέσοι όροι των σημείων των καταγραφών ανά περίπτωση προβλήματος.



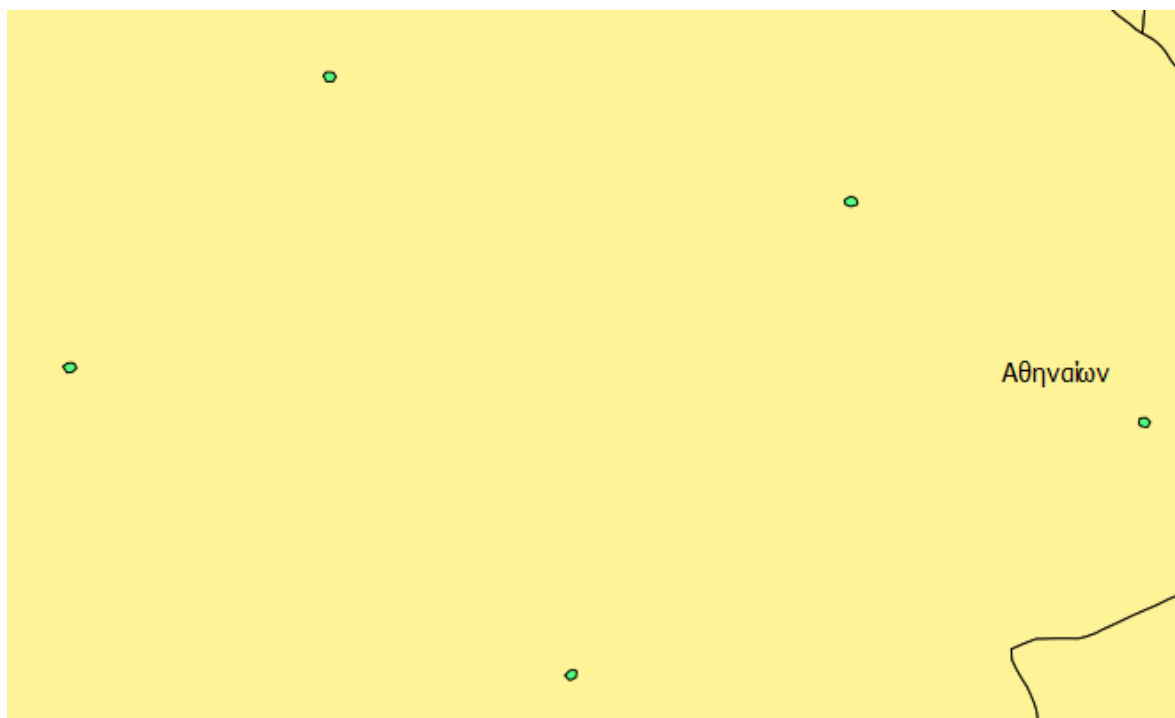
Εικόνα 6.4: Απεικόνιση του συνόλου των αρχικών δεδομένων των καταγραφών



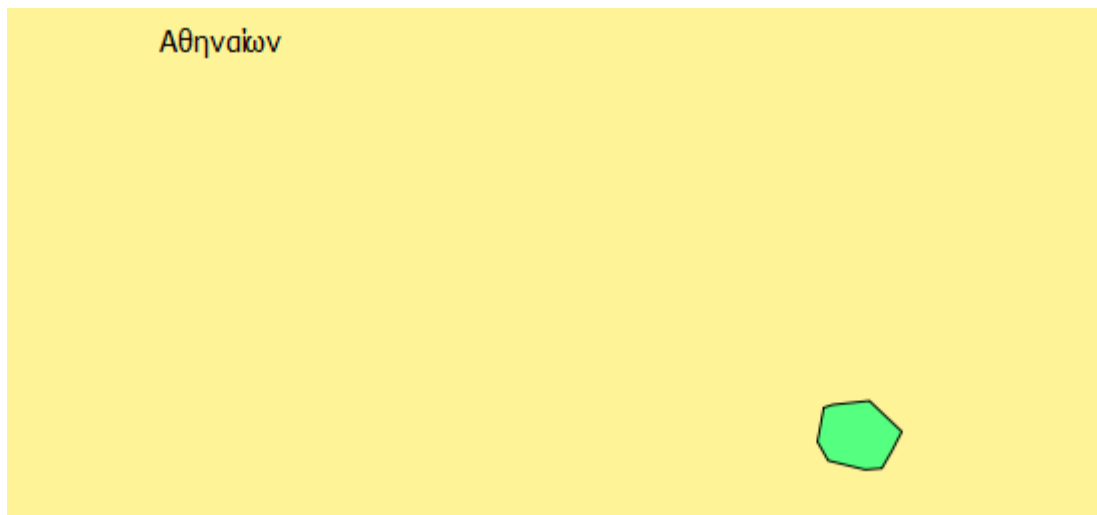
Εικόνα 6.5: Μεγέθυνση της περιοχής των καταγραφών συγκεκριμένου τύπου προβλήματος



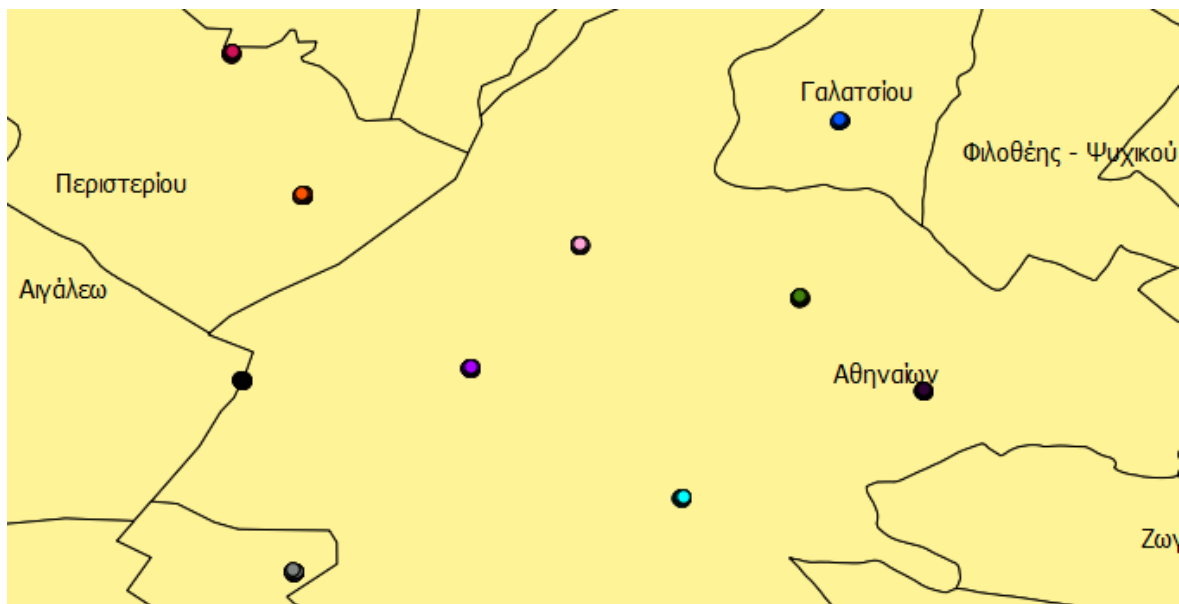
Εικόνα 6.6: Σύνολο των καταγραφών για τον τύπο προβλήματος «χαλασμένο φανάρι» (πράσινο χρώμα)



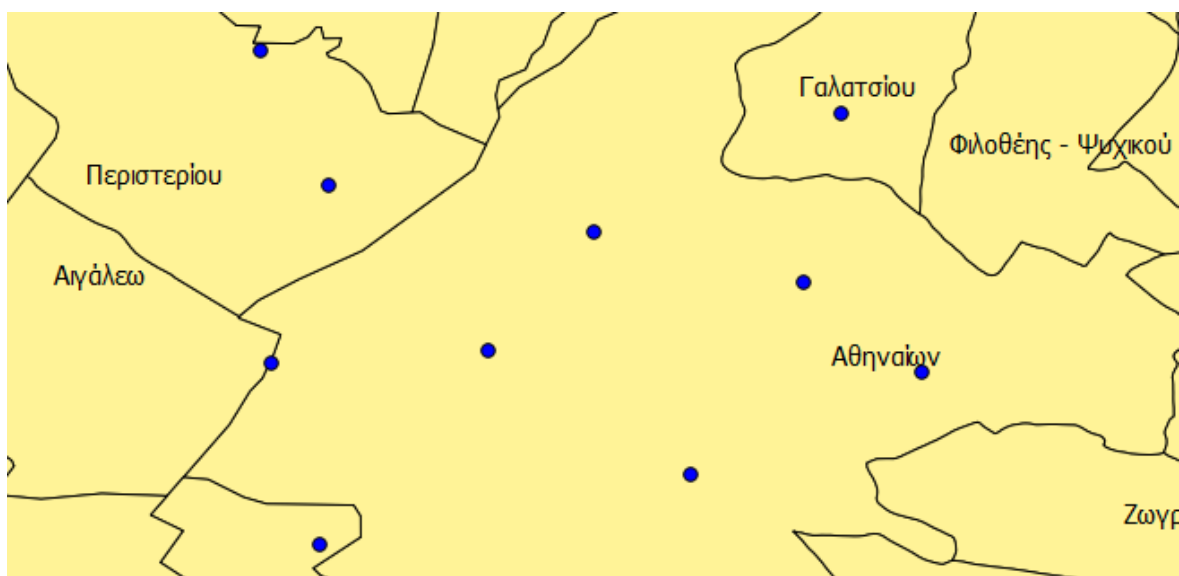
Εικόνα 6.7: Απεικόνιση πολυγώνων ομαδοποίησης σε δεδομένη έκταση



Εικόνα 6.8: Εστίαση σε ένα πολύγωνο ομαδοποίησης



Εικόνα 6.9: Χρωματική διαφοροποίηση των δεδομένων με βάση την ομαδοποίηση k-means



Εικόνα 6.10: Απεικόνιση των μοναδικών σημείων που αντιπροσωπεύουν το σύνολο των καταγραφών κάθε περίπτωσης

6.1.5 Τελικά αποτελέσματα

Στην ενότητα αυτή παρατίθενται τα τελικά αποτελέσματα του «τρεξίματος» της εφαρμογής για την πρώτη περίπτωση, όπως προέκυψαν από τη δημιουργία των kml αρχείων.

Από την εφαρμογή του αλγορίθμου δημιουργήθηκαν τα εξής KML αρχεία:

- | | |
|------------------------|-------------------------|
| ➤ 1-CROWD-20130226.kml | ➤ 6-CROWD-20130226.kml |
| ➤ 1-UNIQ-20130226.kml | ➤ 6-UNIQ-20130226.kml |
| ➤ 2-CROWD-20130226.kml | ➤ 7-CROWD-20130226.kml |
| ➤ 2-UNIQ-20130226.kml | ➤ 7-UNIQ-20130226.kml |
| ➤ 3-CROWD-20130226.kml | ➤ 8-CROWD-20130226.kml |
| ➤ 3-UNIQ-20130226.kml | ➤ 8-UNIQ-20130226.kml |
| ➤ 4-CROWD-20130226.kml | ➤ 9-CROWD-20130226.kml |
| ➤ 4-UNIQ-20130226.kml | ➤ 9-UNIQ-20130226.kml |
| ➤ 5-CROWD-20130226.kml | ➤ 10-CROWD-20130226.kml |
| ➤ 5-UNIQ-20130226.kml | ➤ 10-UNIQ-20130226.kml |

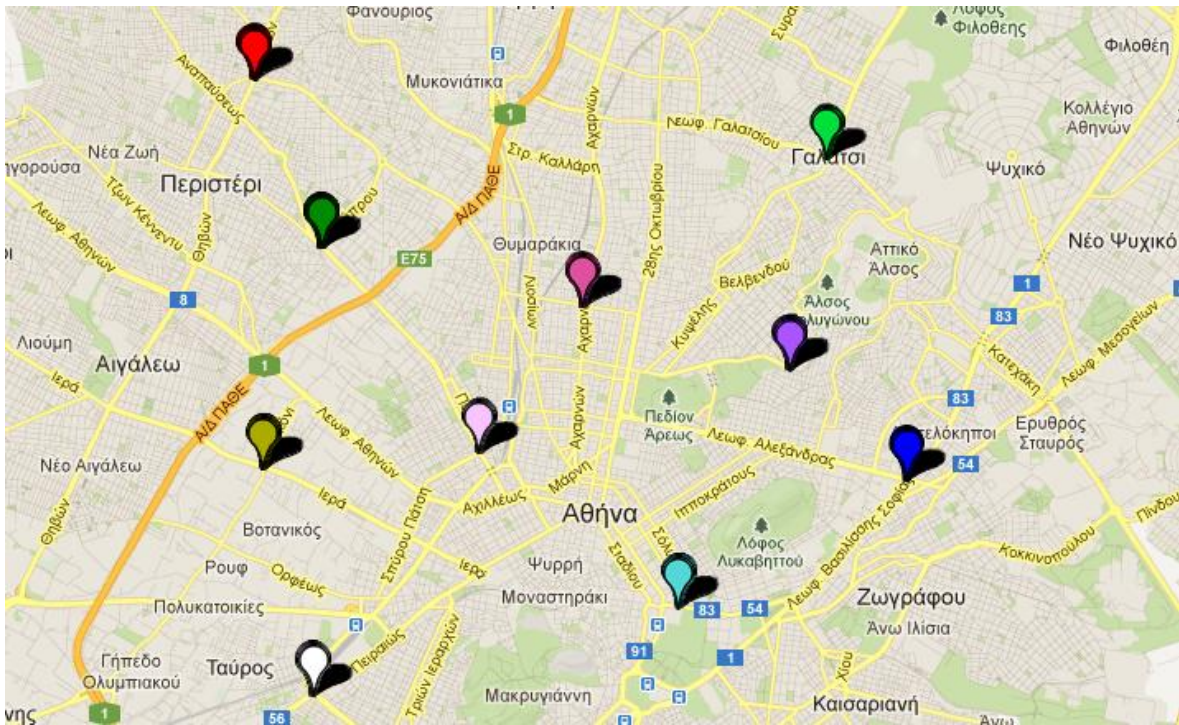
Κάθε αριθμός στο όνομα των αρχείων υποδηλώνει τον τύπο προβλήματος που απεικονίζεται. Οι αντιστοιχίες των αριθμών και των τύπων προβλήματος παρουσιάζονται στη συνέχεια.

- 01-ανώμαλο οδόστρωμα
- 02-λακκούβα
- 03-σπασμένο πεζοδρόμιο
- 04-χαλασμένο φανάρι
- 05-πεσμένο δέντρο
- 06-αλλοιωμένη ταμπέλα
- 07-υπερχείλιση φρεατίου
- 08-διαρροή αγωγού
- 09-καμένη λάμπα
- 10-σπασμένο παγκάκι

Εδώ πρέπει να σημειωθεί ότι επειδή ο αρχικός αριθμός σημείων που δημιουργήθηκαν ήταν δέκα, δηλαδή μικρότερος από τον αριθμό των 15 τύπων προβλημάτων που δηλώθηκαν αρχικά, στην πρώτη περίπτωση που μελετάται συμμετέχουν μόνο οι πρώτοι δέκα τύποι προβλήματος.

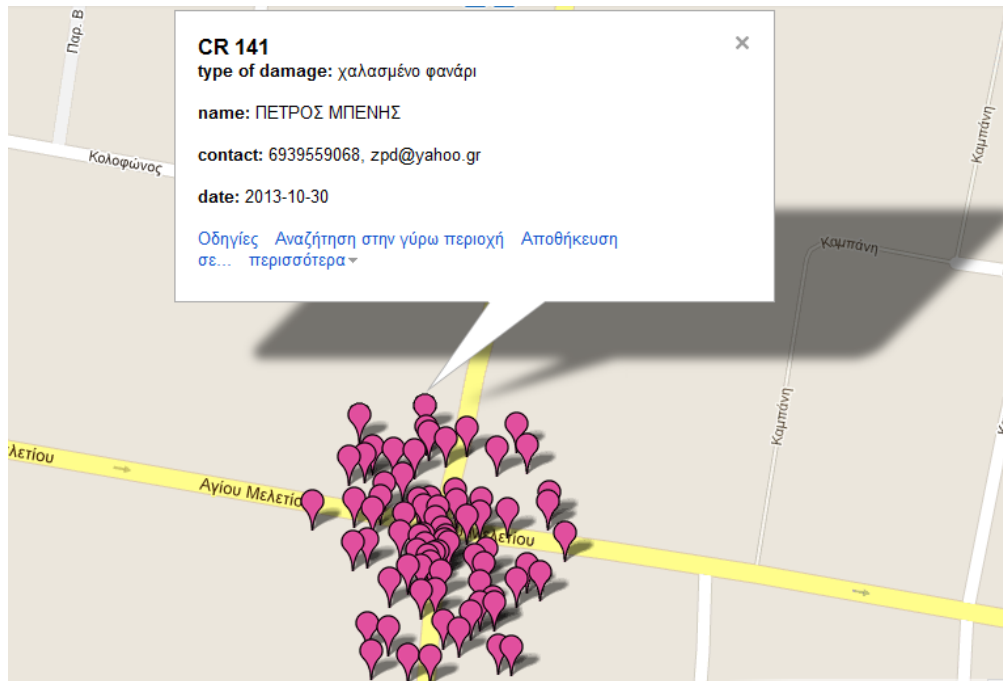
Για κάθε έναν από τους παραπάνω τύπους προβλημάτων δημιουργήθηκε ένα αρχείο CROWD για την απεικόνιση του συνόλου των καταγραφών ανά πρόβλημα, ενώ για την απεικόνιση των μέσων σημείων κάθε ομάδας καταγραφών ανά πρόβλημα δημιουργήθηκε από ένα αρχείο UNIQ.

Στη συνέχεια απεικονίστηκε το σύνολο των αποτελεσμάτων στο περιβάλλον του **Google Maps** αλλά και στην **προσωπική ιστοσελίδα**. Παρουσιάζεται ο χάρτης του συνόλου των καταγραφών, καθώς επίσης και των μοναδικών περιπτώσεων που αντιστοιχούν στη μέση τιμή των καταγραφών για κάθε περίπτωση προβλήματος, καθώς επίσης και οι τύποι προβλήματος που αντιστοιχούν σε κάθε χρώμα απεικόνισης.



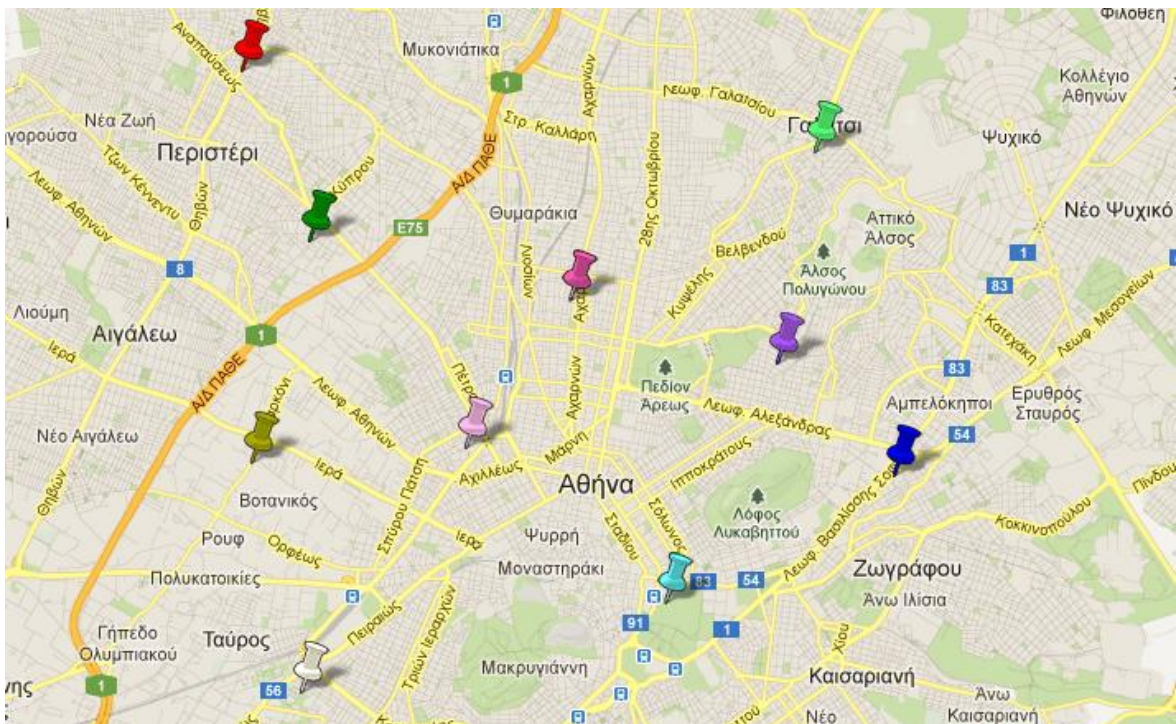
| | | | |
|--|---------------------|--|----------------------|
| | ανώμαλο οδόστρωμα | | αλλοιωμένη ταμπέλα |
| | λακούβα | | υπερχείλιση φρεατίου |
| | σπασμένο πεζοδρόμιο | | διαρροή αγωγού |
| | χαλασμένο φανάρι | | καμένη λάμπα |
| | πεσμένο δέντρο | | σπασμένο παγκάκι |











Εικόνα 6.11: Απεικόνιση του συνόλου των καταγραφών ανά τύπο προβλήματος στο περιβάλλον του Google Maps για την «Περίπτωση 1»



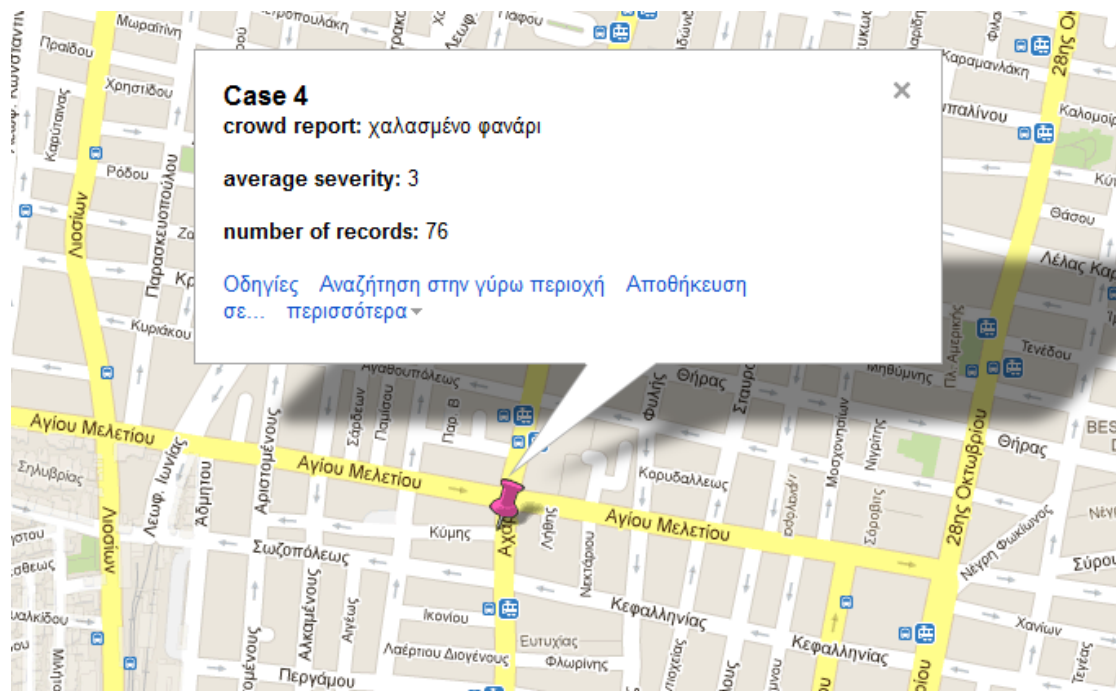
χαλασμένο φανάρι

Εικόνα 6.12: Μεγέθυνση της απεικόνισης των καταγραφών του προβλήματος «χαλασμένο φανάρι» στο περιβάλλον του Google Maps για την «Περίπτωση 1»



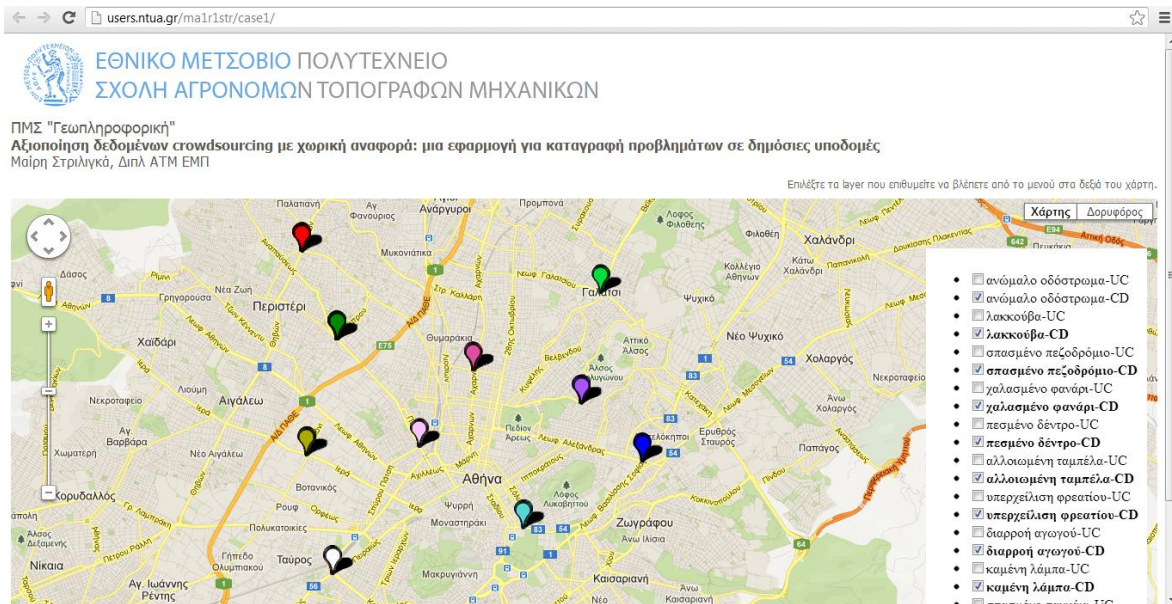
- | | | | |
|---|---------------------|---|----------------------|
|  | ανώμαλο οδόστρωμα |  | αλλοιωμένη ταμπέλα |
|  | λακούβα |  | υπερχείλιση φρεατίου |
|  | σπασμένο πεζοδρόμιο |  | διαρροή αγωγού |
|  | χαλασμένο φανάρι |  | καμένη λάμπα |
|  | πεσμένο δέντρο |  | σπασμένο παγκάκι |

Εικόνα 6.13: Απεικόνιση του συνόλου των μοναδικών περιπτώσεων ανά τύπο προβλήματος στο περιβάλλον του Google Maps για την «Περίπτωση 1»

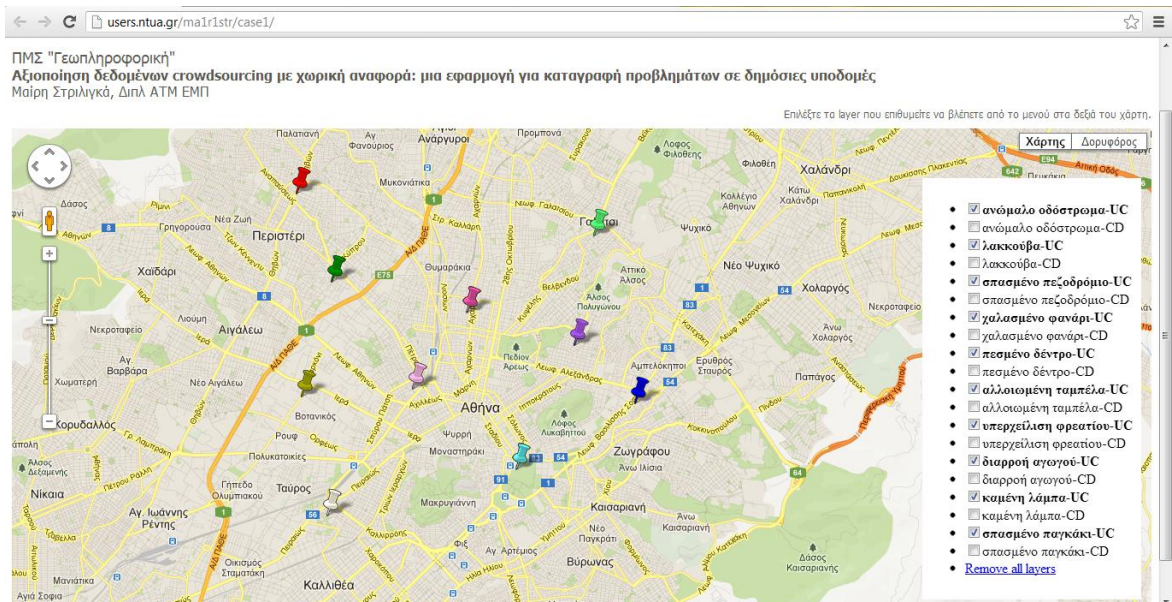


- | | |
|---|------------------|
|  | χαλασμένο φανάρι |
|---|------------------|

Εικόνα 6.14: Μεγέθυνση της απεικόνισης της μοναδικής περίπτωσης του προβλήματος «χαλασμένο φανάρι» στο περιβάλλον του Google Maps για την «Περίπτωση 1»



Εικόνα 6.15: Απεικόνιση του συνόλου των καταγραφών ανά τύπο προβλήματος σε προσωπική ιστοσελίδα για την «Περίπτωση 1»



Εικόνα 6.16: Απεικόνιση του συνόλου των μοναδικών περιπτώσεων ανά τύπο προβλήματος σε προσωπική ιστοσελίδα για την «Περίπτωση 1»

6.2 Περίπτωση 2

Στο δεύτερο παράδειγμα της εφαρμογής μελετήθηκε η περίπτωση αναφοράς 161 διαφορετικών περιπτώσεων προβλημάτων, για κάθε μία εκ των οποίων έχουν πραγματοποιηθεί 1 έως 100 καταγραφές.

6.2.1 Παράμετροι της εφαρμογής

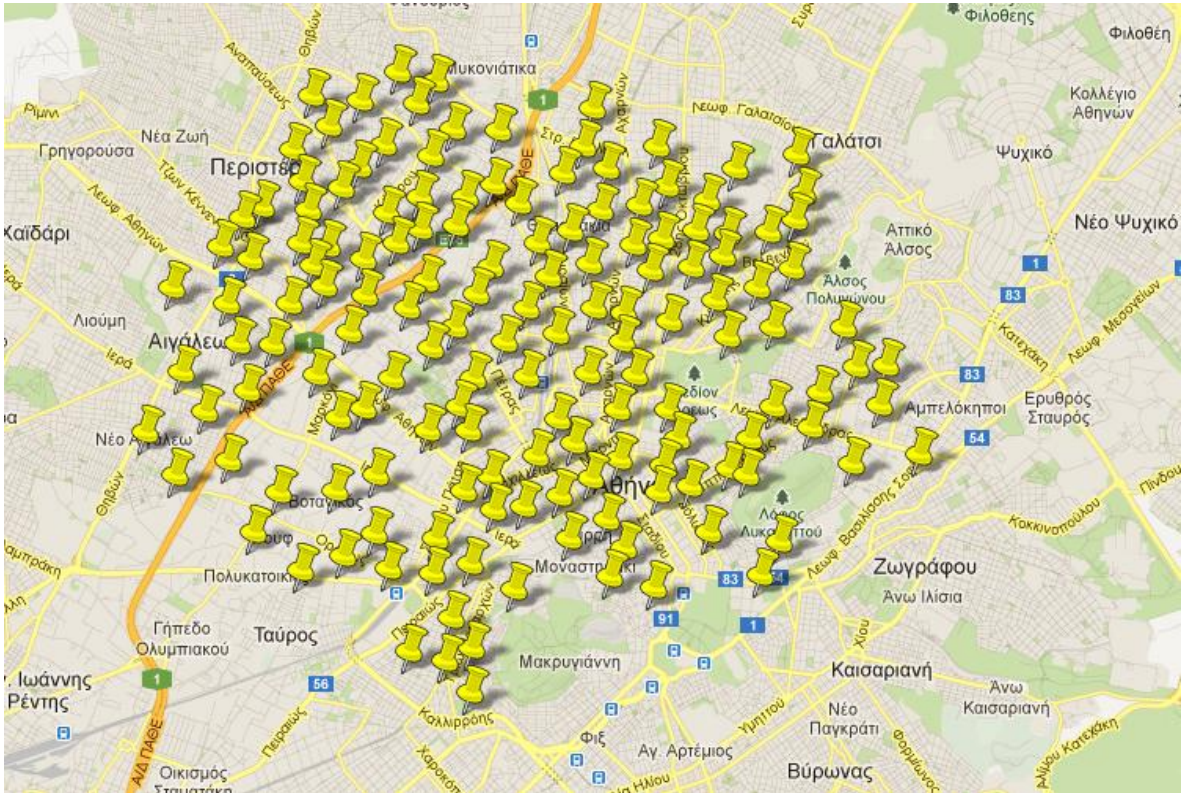
Οι βασικές παράμετροι αντίστοιχα για την υλοποίηση της δεύτερης περίπτωσης της εφαρμογής είναι οι εξής:

- ☞ Οι γεωγραφικές συντεταγμένες (φ,λ) **161 αρχικών σημείων**, τα οποία θα αποτελέσουν τις δέκα υπό μελέτη περιπτώσεις προβλημάτων που θα καταγράψει το πλήθος.
- ☞ Η μέγιστη και η ελάχιστη τιμή της γωνίας και της ακτίνας των πολικών συντεταγμένων των σημείων που πρόκειται να δημιουργηθούν και θα αποτελέσουν τις καταγραφές των προβλημάτων για κάθε ένα από τα 161 προβλήματα. Η τιμή της γωνίας κυμαίνεται από 0 έως 360 μοίρες ενώ η τιμή της ακτίνας κυμαίνεται από 0 έως 25 μέτρα. Οι τιμές αυτές υποδηλώνουν το εύρος της περιοχής γύρω από τις 161 περιπτώσεις προβλημάτων εντός της οποίας θα δημιουργηθούν τα τυχαία σημεία που θα αποτελέσουν τις καταγραφές του πλήθους.
- ☞ Ο μέγιστος αριθμός τυχαίων σημείων – καταγραφών που θα δημιουργηθούν για κάθε περίπτωση προβλήματος. Για κάθε περίπτωση προβλήματος επιλέχθηκε να πραγματοποιούνται έως 100 καταγραφές. Ο αριθμός των καταγραφών επιλέγεται τυχαία με τη βοήθεια της συνάρτησης $\text{rand}(1,100)$.
- ☞ Οι παράμετροι μετατροπής των επίπεδων συντεταγμένων x, y του συστήματος ΕΓΣΑ87 σε γεωγραφικές συντεταγμένες ϕ, λ του συστήματος WGS84 και αντίστροφα, οι οποίες είναι σταθερές και δεν μεταβάλλονται από περίπτωση σε περίπτωση για τα δεδομένα της Ελλάδας.
- ☞ Η απόσταση που καθορίζει το εύρος της ομαδοποίησης των σημείων για τη δημιουργία των πολυγώνων ομαδοποίησης. Στην περίπτωση που μελετάμε η απ πήρε τιμή ίση με 30 m.

6.2.2 Δεδομένα εισόδου

Στην δεύτερη περίπτωση που υλοποιήθηκε, μελετήθηκε η εφαρμογή του συστήματος με τη χρήση μεγάλου όγκου αρχικών δεδομένων. Έτσι, επιλέχθηκαν χειροκίνητα στο περιβάλλον του Google Earth 161 σημεία σε ικανοποιητική απόσταση μεταξύ τους, καθένα εκ των οποίων αντιπροσωπεύει ένα διαφορετικό τύπο προβλήματος. Γύρω από κάθε μία από τις 161 περιπτώσεις προβλήματος θα δημιουργηθούν στη συνέχεια τυχαία σημεία, τα οποία θα αντιπροσωπεύουν τις καταγραφές που πραγματοποιήθηκαν για το συγκεκριμένο πρόβλημα.

Όπως και στην πρώτη περίπτωση έτσι και στην περίπτωση που μελετάμε, αποφασίσθηκε τα 161 αρχικά σημεία να μη δημιουργηθούν με την αναπαραγωγή τυχαίων συντεταγμένων έτσι ώστε να μη συμπίπτουν με κτίρια ή άλλα αντικείμενα. Για το λόγο αυτό επιλέχθηκε να βρίσκονται κοντά σε διασταυρώσεις δρόμων, ώστε να προσομοιάζουν πραγματικές περιπτώσεις καταγραφών. Τα σημεία αυτά παρουσιάζονται στο απόσπασμα του χάρτη του Google Maps που φαίνεται στη συνέχεια.



Εικόνα 6.17: Απεικόνιση των 161 αρχικών σημείων στο περιβάλλον του Google Maps

6.2.3 Υλοποίηση της εφαρμογής

Αφού δημιουργήθηκε η βάση δεδομένων και διαμορφώθηκαν τα αναγκαία ερωτήματα για τη δημιουργία των πινάκων, την εισαγωγή των δεδομένων, και την επιλογή δεδομένων από τους πίνακες σε γλώσσα SQL, ενσωματώθηκε ο κώδικας αυτός στον τελικό κώδικα στη γλώσσα PHP, και τέθηκε το σύστημα σε λειτουργία αφήνοντάς το να «τρέξει» μέσω του παραθύρου της γραμμής εντολών.

Τα αρχεία του αλγορίθμου που τέθηκαν σε λειτουργία είναι πέντε:

- **F2_create_tables.php**, για τη δημιουργία των πινάκων «damage_types2» και «crowd_data2» στη βάση δεδομένων.
- **F2_data_insertion_damage_types.php**, για την εισαγωγή των δεδομένων στον πίνακα «damage_types2».
- **F2_create_and_insert_points_crowd_data.php**, για τη δημιουργία των τυχαίων σημείων των καταγραφών γύρω από τα δέκα αρχικά σημεία, τη διαμόρφωση του τελικού csv αρχείου με όλες τις πληροφορίες των καταγραφών, και την εισαγωγή όλων των πληροφοριών στον πίνακα «crowd_data2».
- **F2_kmeans.php**, για την υλοποίηση της ομαδοποίησης των καταγραφών των προβλημάτων με χρήση του αλγορίθμου `kmeans` και την προσθήκη των πινάκων «kmeans_polygons2», «kmeans_points2» και «unique_cases2» στη βάση δεδομένων.
- **F2_kml_points_and_html.php**, για τον υπολογισμό των μοναδικών περιπτώσεων προβλημάτων και τη δημιουργία ενός kml αρχείου για κάθε τύπο προβλήματος, τη δημιουργία ενός kml αρχείου για κάθε τύπο προβλήματος για το σύνολο των δεδομένων, και τη δημιουργία ενός html αρχείου για την απεικόνιση των δεδομένων των kmls σε διαδικτυακούς χάρτες.

Η εισαγωγή των δεδομένων στον πίνακα «damage_types2» πραγματοποιήθηκε με άντληση της πληροφορίας από txt αρχεία, ενώ η εισαγωγή των δεδομένων στον πίνακα «crowd_data2» πραγματοποιήθηκε με άντληση της πληροφορίας από κατάλληλα διαμορφωμένο csv αρχείο.

Μετά την ολοκλήρωση του «τρεξίματος» των πέντε rhr αρχείων, προέκυψαν πέντε διαφορετικοί πίνακες με τα αντίστοιχα δεδομένα μέσα στη βάση δεδομένων του λογισμικού PostgreSQL. Επιπλέον, εξήχθησαν δύο kml αρχεία για κάθε τύπο προβλήματος με τα αποτελέσματα που προέκυψαν από την υλοποίηση του κώδικα.

Στην εικόνα που ακολουθεί παρουσιάζονται οι πίνακες και τα πεδία κάθε πίνακα που δημιουργήθηκαν στη βάση δεδομένων έπειτα από την υλοποίηση του αλγορίθμου.

| crowd_data2 | damage_types2 | kmeans_polygons2 | kmeans_points2 | unique_cases2 |
|--|--|--|---|--|
| <input type="checkbox"/> id | <input type="checkbox"/> t_code | <input type="checkbox"/> polygons_gm | <input type="checkbox"/> kmeans | <input type="checkbox"/> kmeans |
| <input type="checkbox"/> latitude | <input type="checkbox"/> description | <input type="checkbox"/> centroids_geom | <input type="checkbox"/> latitude | <input type="checkbox"/> uc_damage_types |
| <input type="checkbox"/> longitude | <input type="checkbox"/> colour | <input type="checkbox"/> centroids_latitude | <input type="checkbox"/> longitude | <input type="checkbox"/> avg_latitude |
| <input type="checkbox"/> name | <input type="checkbox"/> kml_unique_colour | <input type="checkbox"/> centroids_longitude | <input type="checkbox"/> severity | <input type="checkbox"/> avg_longitude |
| <input type="checkbox"/> surname | <input type="checkbox"/> kml_crowd_colour | | <input type="checkbox"/> type_of_damage | <input type="checkbox"/> avg_severity |
| <input type="checkbox"/> phone_number | <input type="checkbox"/> styleurl_unique | | <input type="checkbox"/> cd_id | <input type="checkbox"/> num_of_records |
| <input type="checkbox"/> email | <input type="checkbox"/> styleurl_crowd | | <input type="checkbox"/> id | <input type="checkbox"/> uc_id |
| <input type="checkbox"/> city | <input type="checkbox"/> icon | | <input type="checkbox"/> geom | <input type="checkbox"/> geom |
| <input type="checkbox"/> type_of_damage | | | | |
| <input type="checkbox"/> short_description | | | | |
| <input type="checkbox"/> date | | | | |
| <input type="checkbox"/> severity | | | | |
| <input type="checkbox"/> capture | | | | |
| <input type="checkbox"/> updated | | | | |
| <input type="checkbox"/> created | | | | |
| <input type="checkbox"/> project | | | | |
| <input type="checkbox"/> created_by | | | | |
| <input type="checkbox"/> updated_by | | | | |
| <input type="checkbox"/> geom | | | | |
| <input type="checkbox"/> unique_cases_id | | | | |

Εικόνα 6.18: Οι πίνακες που δημιουργήθηκαν στη βάση δεδομένων για την «Περίπτωση 2»

6.2.4 Απεικόνιση ακατέργαστων αποτελεσμάτων στο Quantum GIS

Στην παρούσα ενότητα παρουσιάζεται η πορεία των ακατέργαστων δεδομένων στο περιβάλλον του λογισμικού Quantum GIS, πριν την απεικόνισή τους στους τελικούς διαδικτυακούς χάρτες μέσω των kml αρχείων. Για την απεικόνιση των δεδομένων της βάσης δεδομένων στο Quantum GIS απαιτείται η υλοποίηση σύνδεσης με την συγκεκριμένη βάση δεδομένων όπως περιγράφηκε σε προηγούμενη ενότητα.

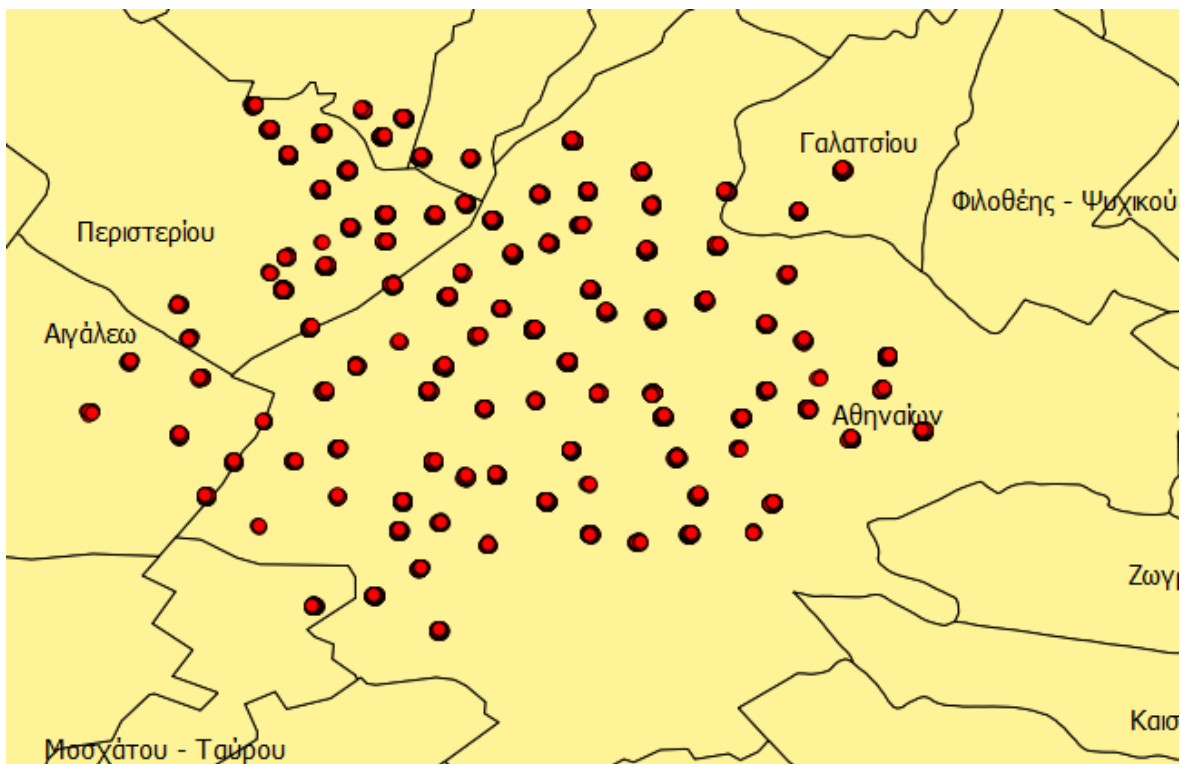
Όλοι οι χάρτες που παρουσιάζονται έχουν ως υπόβαθρο χάρτη των καλλικρατικών δήμων της Ελλάδος. Στους χάρτες απεικονίζονται τα δεδομένα με τη σειρά που υλοποιήθηκαν στη βάση δεδομένων. Πιο συγκεκριμένα απεικονίζονται:

- ✓ τα αρχικά δεδομένα όπως προκύπτουν από τις καταγραφές των προβλημάτων από το πλήθος,
- ✓ τα δεδομένα των καταγραφών που πραγματοποιήθηκαν για ένα συγκεκριμένο τύπο προβλήματος,
- ✓ τα πολύγωνα ομαδοποίησης που προέκυψαν από την υλοποίηση της πρώτης ομαδοποίησης των σημείων για την ορισμό των διανυσμάτων αρχικοποίησης,
- ✓ οι ομάδες σημείων των καταγραφών που προέκυψαν από την υλοποίηση της ομαδοποίησης με τον αλγόριθμο k-means,

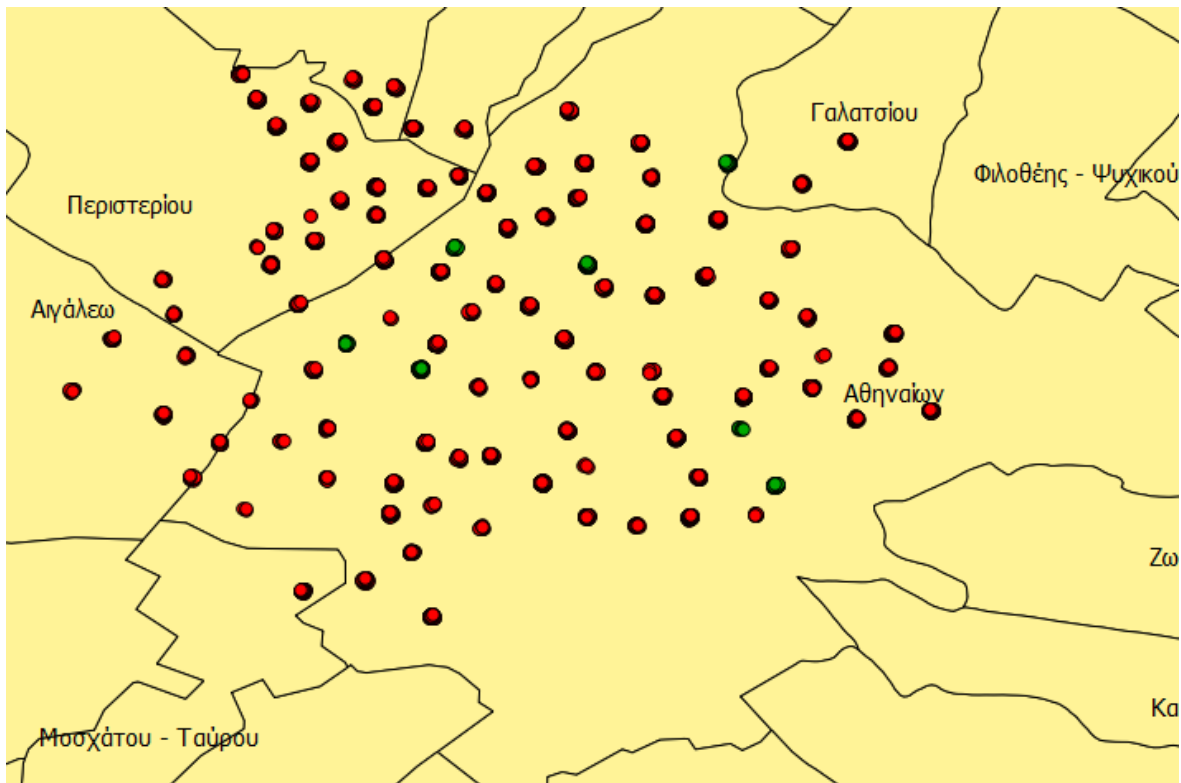
- ✓ και οι μοναδικές περιπτώσεις προβλημάτων που αντιπροσωπεύουν το σύνολο των καταγραφών για κάθε περίπτωση.

Στο σημείο αυτό πρέπει να σημειωθούν οι εξής παρατηρήσεις:

- Στις εικόνες 6.17, 6.18 και 6.21 παρουσιάζεται το σύνολο των σημείων των καταγραφών αλλά επειδή η έκταση που απεικονίζεται είναι μεγάλη με αποτέλεσμα να μη γίνεται μεγέθυνση του χάρτη ώστε να απεικονιστούν όλες οι καταγεγραμμένες περιπτώσεις, τα σημεία των καταγραφών κάθε περίπτωσης φαίνονται ως ένα σημείο. Για το λόγο αυτό δεν πρέπει να υπάρξει ταύτιση των χαρτών αυτών με το χάρτη της εικόνας 6.22 ο οποίος απεικονίζει μεμονωμένα σημεία. Η πραγματική εικόνα των καταγραφών σε μεγέθυνση φαίνεται στην εικόνα 6.20.
- Στην εικόνα 6.18 παρουσιάζεται το σύνολο των καταγραφών για τις περιπτώσεις προβλήματος «χαλασμένο φανάρι» που απεικονίζονται με πράσινο χρώμα, σε σχέση με τις καταγραφές για τα υπόλοιπα προβλήματα που διαφοροποιούνται με κόκκινο χρώμα.
- Στην εικόνα 6.19 παρουσιάζεται μεγεθυμένο απόσπασμα του χάρτη των πολυγώνων ομαδοποίησης, ενώ στο χάρτη 6.20 παρουσιάζεται το πολύγωνο ομαδοποίησης μιας περίπτωσης προβλήματος σε ακόμα μεγαλύτερη μεγέθυνση ώστε να είναι περισσότερο ευδιάκριτο, καθώς επίσης και τα σημεία των καταγραφών της περίπτωσης αυτής που ανήκουν στην ομάδα που ορίζει το πολύγωνο.
- Στην εικόνα 6.21 κάθε ομάδα σημείων που προέκυψε από την ομαδοποίηση k-means απεικονίζεται με το ίδιο χρώμα.
- Τέλος, στην εικόνα 6.22 απεικονίζονται με μπλε χρώμα οι μέσοι όροι των σημείων των καταγραφών ανά περίπτωση προβλήματος.



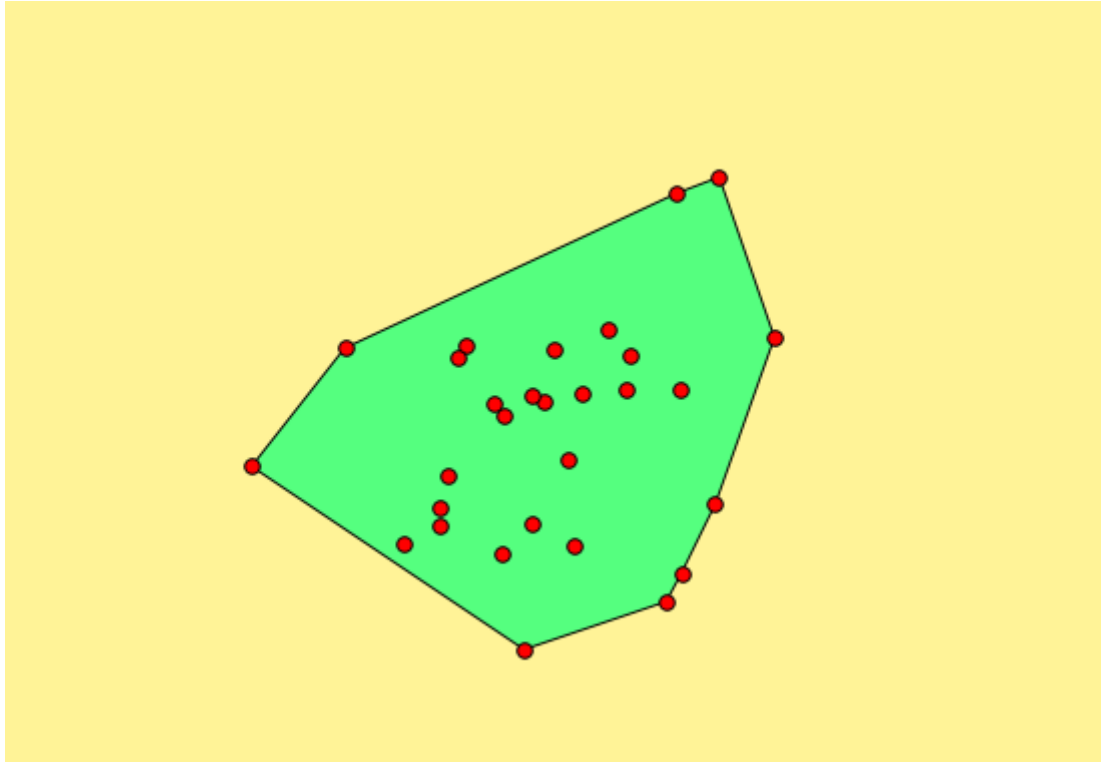
Εικόνα 6.19: Απεικόνιση του συνόλου των αρχικών δεδομένων των καταγραφών



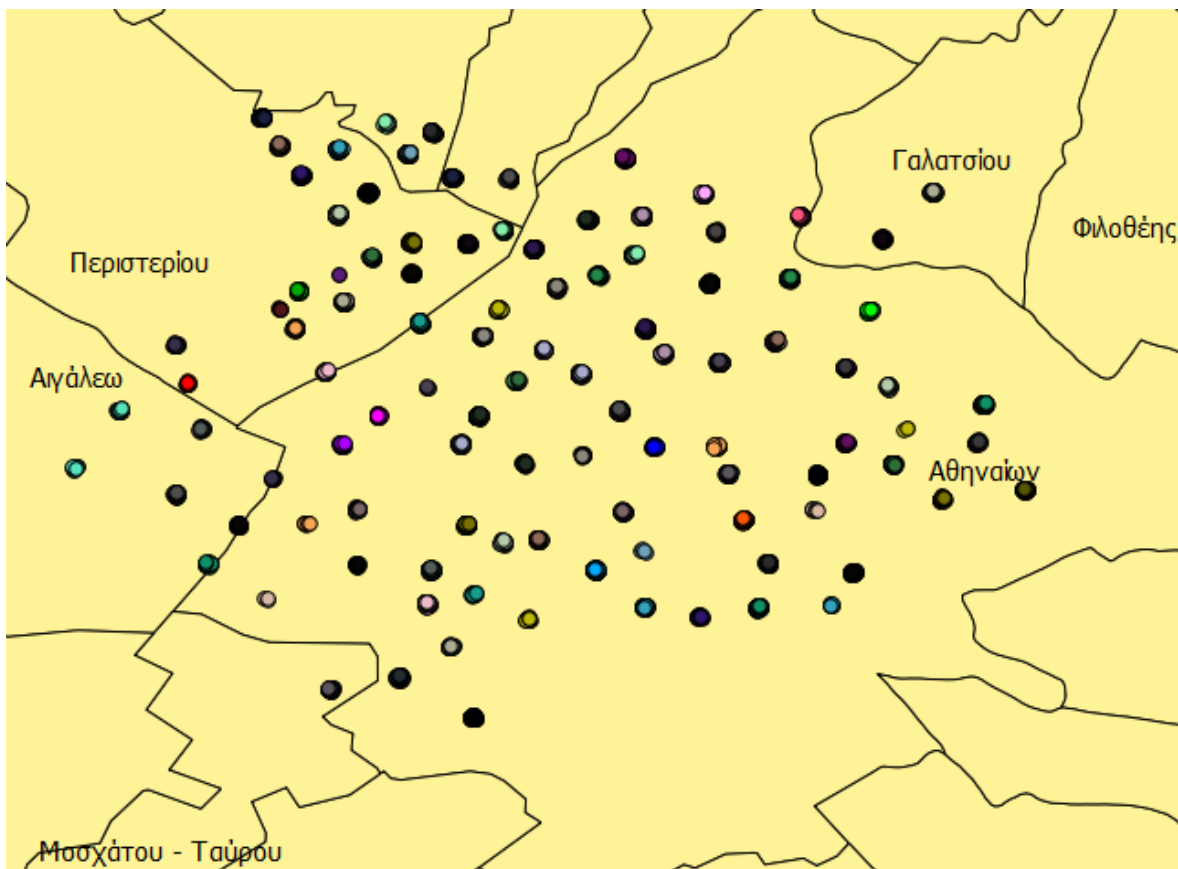
Εικόνα 6.20: Σύνολο των καταγραφών για τον τύπο προβλήματος «χαλασμένο φανάρι» (πράσινο χρώμα)



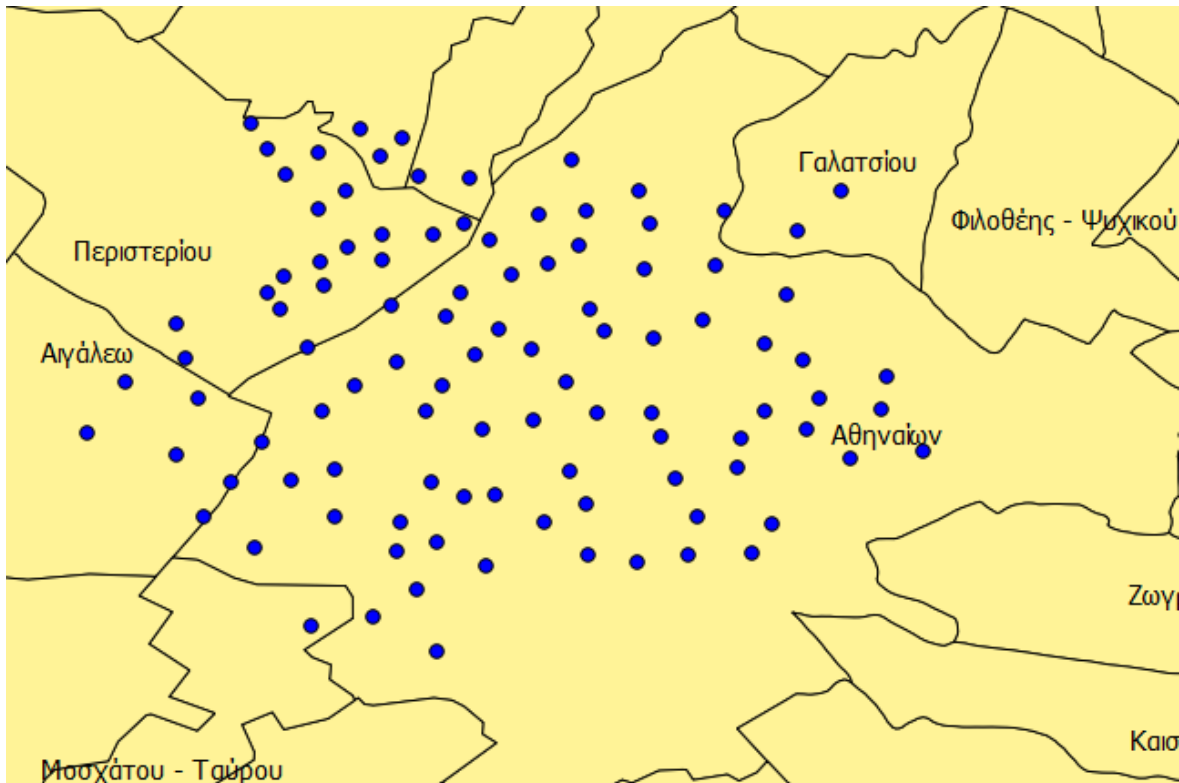
Εικόνα 6.21: Απεικόνιση πολυγώνων ομαδοποίησης σε δεδομένη έκταση



Εικόνα 6.22: Εστίαση σε ένα πολύγωνο ομαδοποίησης



Εικόνα 6.23: Χρωματική διαφοροποίηση των δεδομένων με βάση την ομαδοποίηση k-means



Εικόνα 6.24: Απεικόνιση των μοναδικών σημείων που αντιπροσωπεύουν το σύνολο των καταγραφών κάθε περίπτωσης

6.2.5 Τελικά αποτελέσματα

Στην ενότητα αυτή παρατίθενται τα αποτελέσματα του «τρεξίματος» της εφαρμογής για τη δεύτερη περίπτωση, όπως προέκυψαν από τη δημιουργία των kml αρχείων.

Από την εφαρμογή του αλγορίθμου δημιουργήθηκαν τα εξής KML αρχεία:

- | | |
|------------------------|-------------------------|
| ➤ 1-CROWD-20130226.kml | ➤ 6-CROWD-20130226.kml |
| ➤ 1-UNIQ-20130226.kml | ➤ 6-UNIQ-20130226.kml |
| ➤ 2-CROWD-20130226.kml | ➤ 7-CROWD-20130226.kml |
| ➤ 2-UNIQ-20130226.kml | ➤ 7-UNIQ-20130226.kml |
| ➤ 3-CROWD-20130226.kml | ➤ 8-CROWD-20130226.kml |
| ➤ 3-UNIQ-20130226.kml | ➤ 8-UNIQ-20130226.kml |
| ➤ 4-CROWD-20130226.kml | ➤ 9-CROWD-20130226.kml |
| ➤ 4-UNIQ-20130226.kml | ➤ 9-UNIQ-20130226.kml |
| ➤ 5-CROWD-20130226.kml | ➤ 10-CROWD-20130226.kml |
| ➤ 5-UNIQ-20130226.kml | ➤ 10-UNIQ-20130226.kml |

Κάθε αριθμός στο όνομα των αρχείων υποδηλώνει τον τύπο προβλήματος που απεικονίζεται. Οι αντιστοιχίες των αριθμών και των τύπων προβλήματος παρουσιάζονται στη συνέχεια.

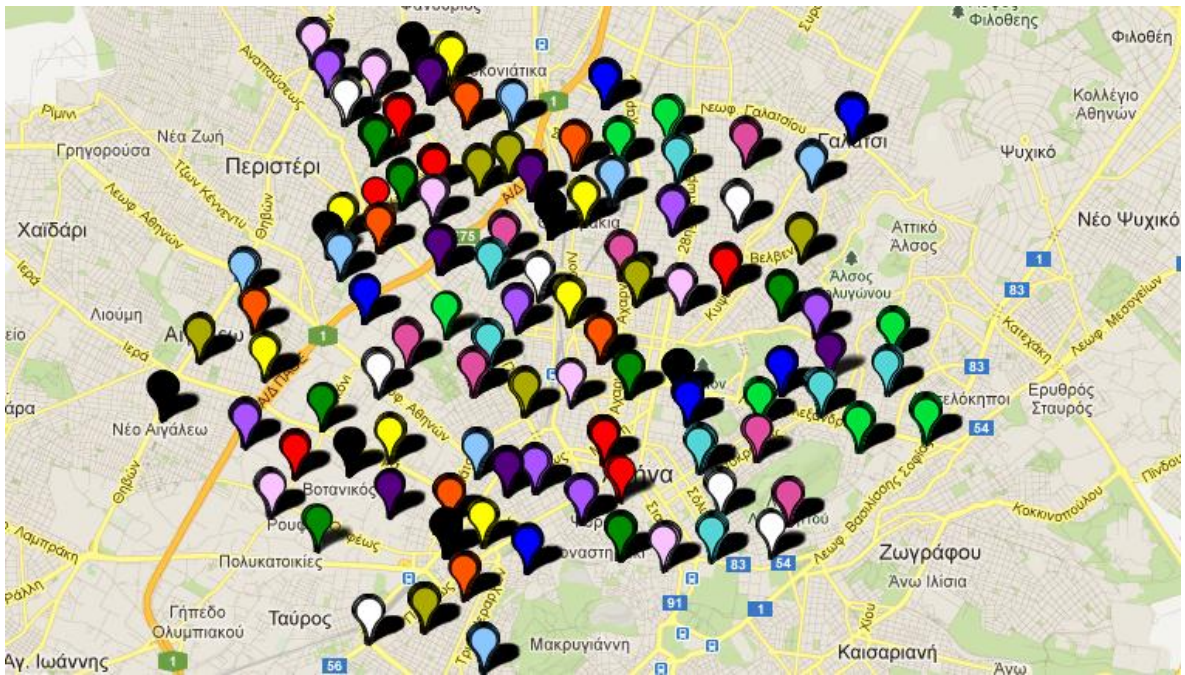
- 01-ανώμαλο οδόστρωμα
- 02-λακκούβα
- 03-σπασμένο πεζοδρόμιο
- 04-χαλασμένο φανάρι
- 05-πεςμένο δέντρο

- 06-αλλοιωμένη ταμπέλα
- 07-υπερχείλιση φρεατίου
- 08-διαρροή αγωγού
- 09-καμένη λάμπα
- 10-σπασμένο παγκάκι
- 11-τηλεφωνικός θάλαμος εκτός λειτουργίας
- 12-γκράφιτι
- 13-παράνομη αφισοκόλληση
- 14-εγκαταλελειμμένο όχημα
- 15-γκρεμισμένη στάση λεωφορείων

Εδώ πρέπει να σημειωθεί ότι επειδή ο αρχικός αριθμός σημείων που δημιουργήθηκαν ήταν 161, δηλαδή πολύ μεγαλύτερος από τον αριθμό των τύπων προβλημάτων που δηλώθηκαν αρχικά, στην δεύτερη περίπτωση που μελετάται συμμετέχουν και οι 15 τύποι προβλήματος και επαναλαμβάνονται σε πολλές περιπτώσεις προβλημάτων.

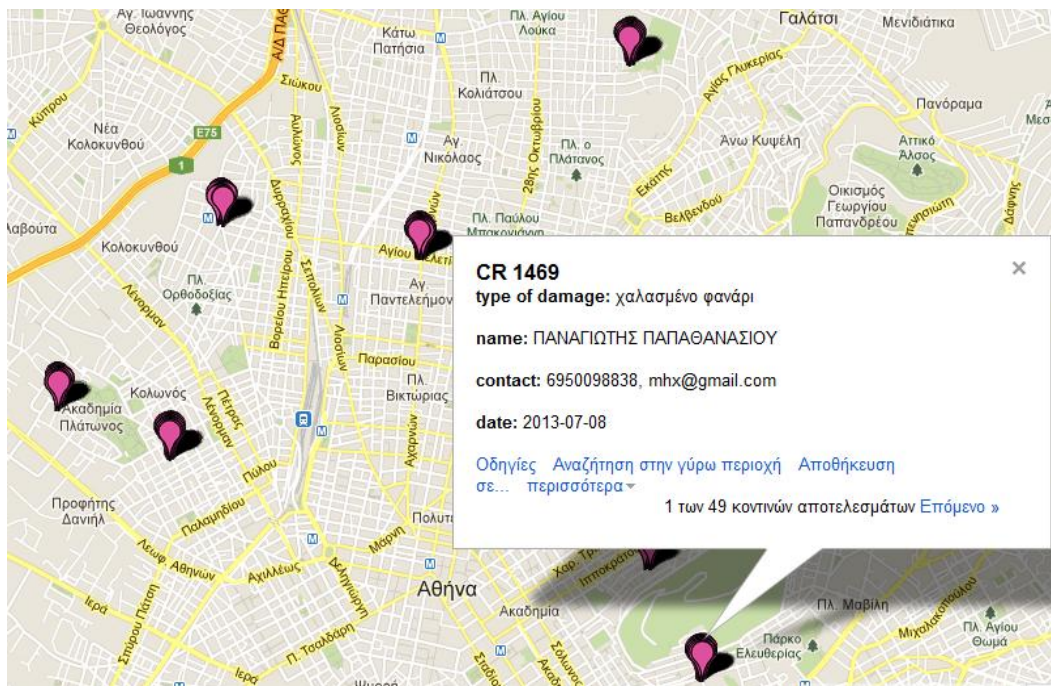
Για κάθε έναν από τους παραπάνω τύπους προβλημάτων δημιουργήθηκε ένα αρχείο CROWD για την απεικόνιση του συνόλου των καταγραφών ανά πρόβλημα, ενώ για την απεικόνιση των μέσων σημείων κάθε ομάδας καταγραφών ανά πρόβλημα δημιουργήθηκε από ένα αρχείο UNIQ.

Στη συνέχεια απεικονίστηκε το σύνολο των αποτελεσμάτων στο περιβάλλον του **Google Maps** αλλά και στην **προσωπική ιστοσελίδα**. Παρουσιάζεται ο χάρτης του συνόλου των καταγραφών, καθώς επίσης και των μοναδικών περιπτώσεων που αντιστοιχούν στη μέση τιμή των καταγραφών για κάθε περίπτωση προβλήματος, καθώς επίσης και οι τύποι προβλήματος που αντιστοιχούν σε κάθε χρώμα απεικόνισης.



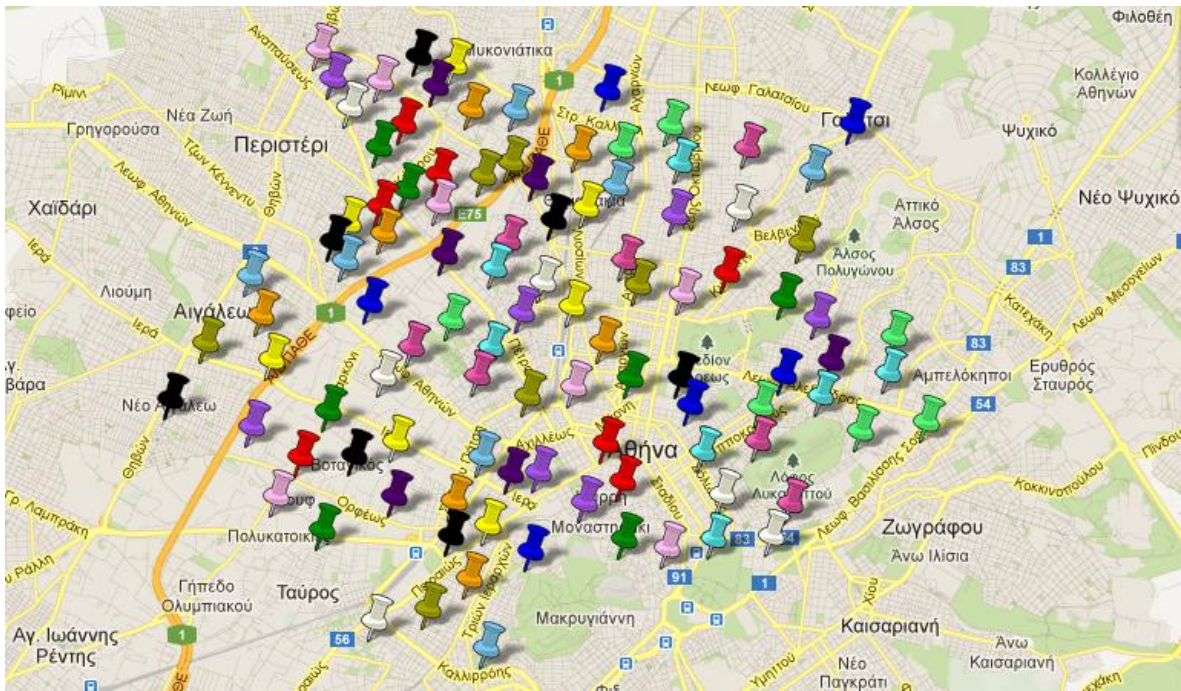
| | | | | | |
|--|---------------------|--|----------------------|--|---------------------------------------|
| | ανώμαλο οδόστρωμα | | αλλοιωμένη ταμπέλα | | τηλεφωνικός θάλαμος εκτός λειτουργίας |
| | λακκούβα | | υπερχείλιση φρεατίου | | γκράφιτι |
| | σπασμένο πεζοδρόμιο | | διαρροή αγωγού | | παράνομη αφισκοκόλληση |
| | χαλασμένο φανάρι | | καμένη λάμπα | | εγκαταλελειμμένο όχημα |
| | πεσμένο δέντρο | | σπασμένο παγκάκι | | γκρεμισμένη στάση λεωφορείων |

Εικόνα 6.25: Απεικόνιση του συνόλου των καταγραφών ανά τύπο προβλήματος στο περιβάλλον του Google Maps για την «Περίπτωση 2»



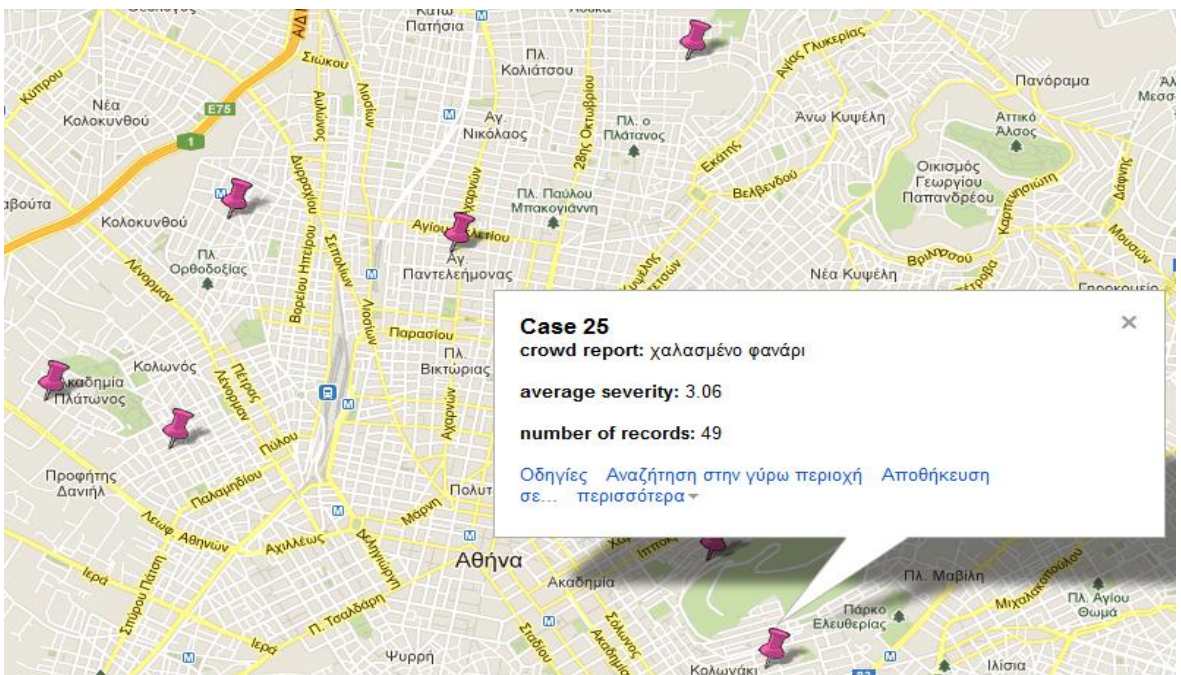
| | |
|--|------------------|
| | χαλασμένο φανάρι |
|--|------------------|

Εικόνα 6.26: Μεγέθυνση της απεικόνισης των καταγραφών του προβλήματος «χαλασμένο φανάρι» στο περιβάλλον του Google Maps για την «Περίπτωση 2»



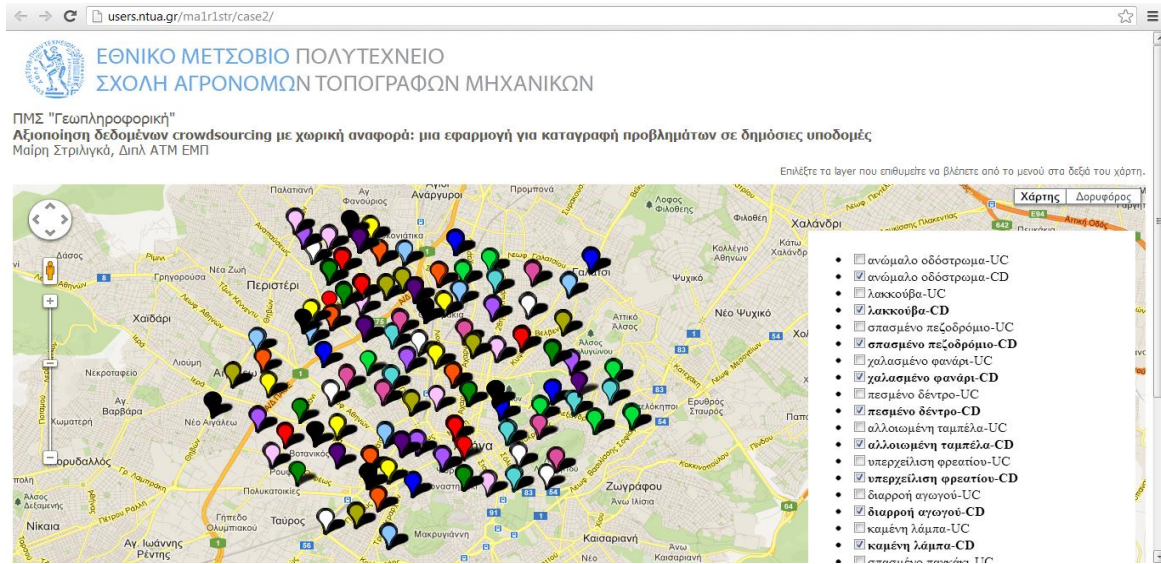
| | | | | | |
|---|---------------------|---|----------------------|--|---------------------------------------|
|  | ανώμαλο οδόστρωμα |  | αλλοιωμένη ταμπέλα |  | τηλεφωνικός θάλαμος εκτός λειτουργίας |
|  | λακούβια |  | υπερχείλιση φρεατίου |  | γκράφιτι |
|  | σπασμένο πεζοδρόμιο |  | διαρροή αγωγού |  | παράνομη αφισκοκόλληση |
|  | χαλασμένο φανάρι |  | καμένη λάμπα |  | εγκαταλελειμμένο όχημα |
|  | πεσμένο δέντρο |  | σπασμένο παγκάκι |  | γκρεμισμένη στάση λεωφορείων |

Εικόνα 6.27: Απεικόνιση του συνόλου των μοναδικών περιπτώσεων ανά τύπο προβλήματος στο περιβάλλον του Google Maps για την «Περίπτωση 2»

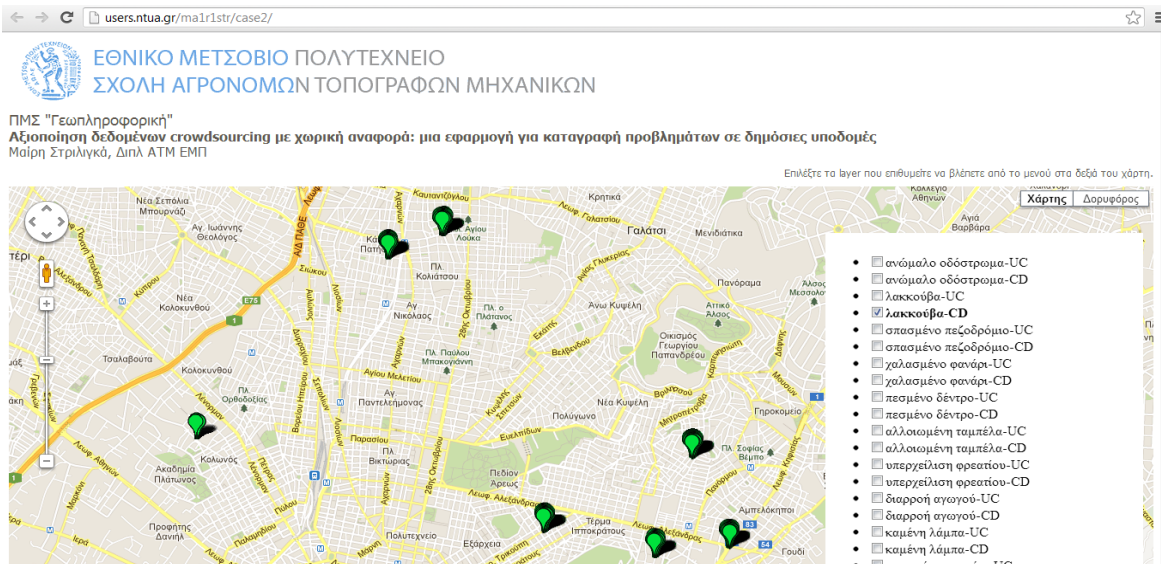


| | |
|---|------------------|
|  | χαλασμένο φανάρι |
|---|------------------|

Εικόνα 6.28: Μεγέθυνση της απεικόνισης των μοναδικών περιπτώσεων του προβλήματος «χαλασμένο φανάρι» στο περιβάλλον του Google Maps για την «Περίπτωση 2»



Εικόνα 6.29: Απεικόνιση του συνόλου των καταγραφών ανά τύπο προβλήματος σε προσωπική ιστοσελίδα για την «Περίπτωση 2»



Εικόνα 6.30: Μεγέθυνση της απεικόνισης των καταγραφών του προβλήματος «λακκούβα» σε προσωπική ιστοσελίδα για την «Περίπτωση 2»



Εικόνα 6.31: Απεικόνιση του συνόλου των μοναδικών περιπτώσεων ανά τύπο προβλήματος σε προσωπική ιστοσελίδα για την «Περίπτωση 2»

6.3 Περίπτωση 3

Στο τρίτο παράδειγμα της εφαρμογής μελετήθηκε η περίπτωση αναφοράς 10 διαφορετικών περιπτώσεων προβλημάτων, για κάθε μία εκ των οποίων έχουν πραγματοποιηθεί 1 έως 100 καταγραφές. Η τρίτη περίπτωση διαφοροποιείται από την πρώτη στο γεγονός ότι μελετάται η περίπτωση γεινίασης δύο ή τριών περιπτώσεων που αφορούν ίδιο τύπο προβλήματος με σκοπό τον έλεγχο της αξιοπιστίας του αλγορίθμου ομαδοποίησης k-means όταν τα σημεία βρίσκονται πολύ κοντά μεταξύ τους.

6.3.1 Παράμετροι της εφαρμογής

Οι βασικές παράμετροι αντίστοιχα για την υλοποίηση της τρίτης περίπτωσης της εφαρμογής είναι οι εξής:

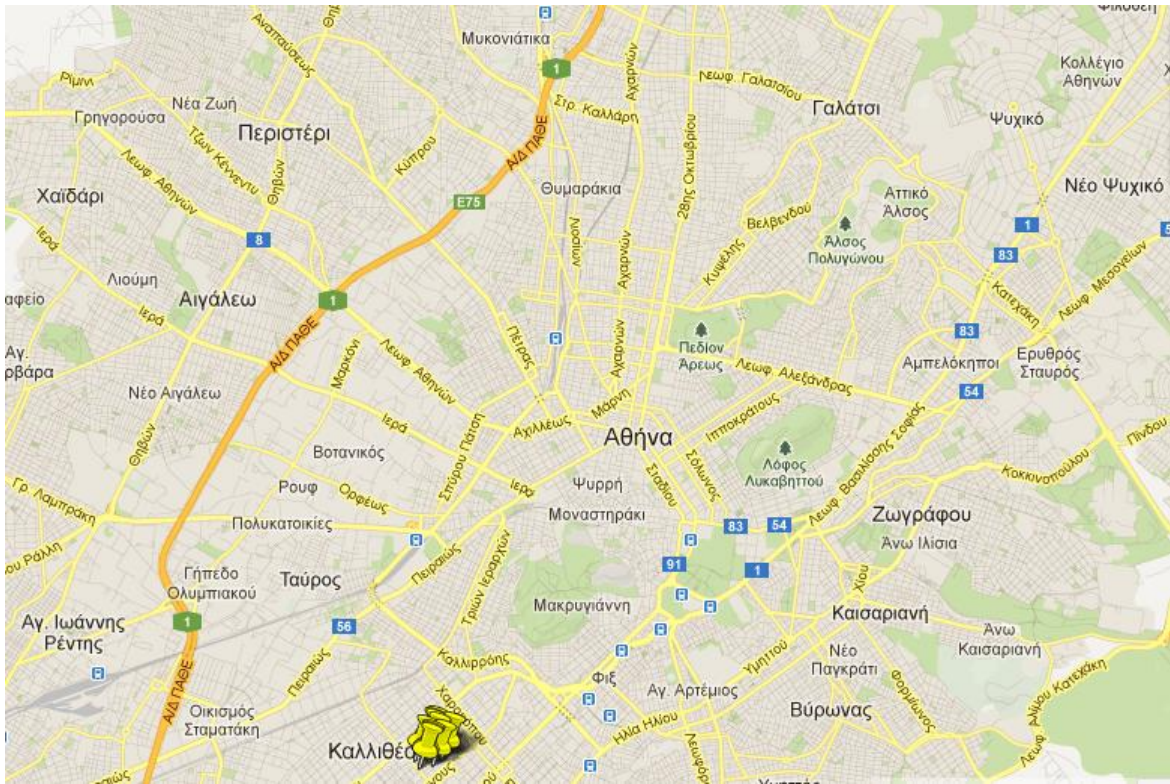
- ☞ Οι γεωγραφικές συντεταγμένες (φ,λ) **10 αρχικών σημείων**, τα οποία θα αποτελέσουν τις δέκα υπό μελέτη περιπτώσεις προβλημάτων που θα καταγράψει το πλήθος.
- ☞ Η μέγιστη και η ελάχιστη τιμή της γωνίας και της ακτίνας των πολικών συντεταγμένων των σημείων που πρόκειται να δημιουργηθούν και θα αποτελέσουν τις καταγραφές των προβλημάτων για κάθε ένα από τα δέκα προβλήματα. Η τιμή της γωνίας κυμαίνεται από 0 έως 360 μοίρες ενώ η τιμή της ακτίνας κυμαίνεται από 0 έως 25 μέτρα. Οι τιμές αυτές υποδηλώνουν το εύρος της περιοχής γύρω από τις 10 περιπτώσεις προβλημάτων εντός της οποίας θα δημιουργηθούν τα τυχαία σημεία που θα αποτελέσουν τις καταγραφές του πλήθους.
- ☞ Ο μέγιστος αριθμός τυχαίων σημείων – καταγραφών που θα δημιουργηθούν για κάθε περίπτωση προβλήματος. Για κάθε περίπτωση προβλήματος επιλέχθηκε να πραγματοποιούνται έως 100 καταγραφές. Ο αριθμός των καταγραφών επιλέγεται τυχαία με τη βοήθεια της συνάρτησης $\text{rand}(1,100)$.
- ☞ Οι παράμετροι μετατροπής των επίπεδων συντεταγμένων x, y του συστήματος ΕΓΣΑ87 σε γεωγραφικές συντεταγμένες φ, λ του συστήματος WGS84 και αντίστροφα, οι οποίες είναι σταθερές και δεν μεταβάλλονται από περίπτωση σε περίπτωση για τα δεδομένα της Ελλάδας.
- ☞ Η απόσταση που καθορίζει το εύρος της ομαδοποίησης των σημείων για τη δημιουργία των πολυγώνων ομαδοποίησης. Στην περίπτωση που μελετάμε η απόσταση πήρε διάφορες τιμές, ώστε να ελεγχθεί η αποτελεσματικότητα της ομαδοποίησης.

6.3.2 Δεδομένα εισόδου

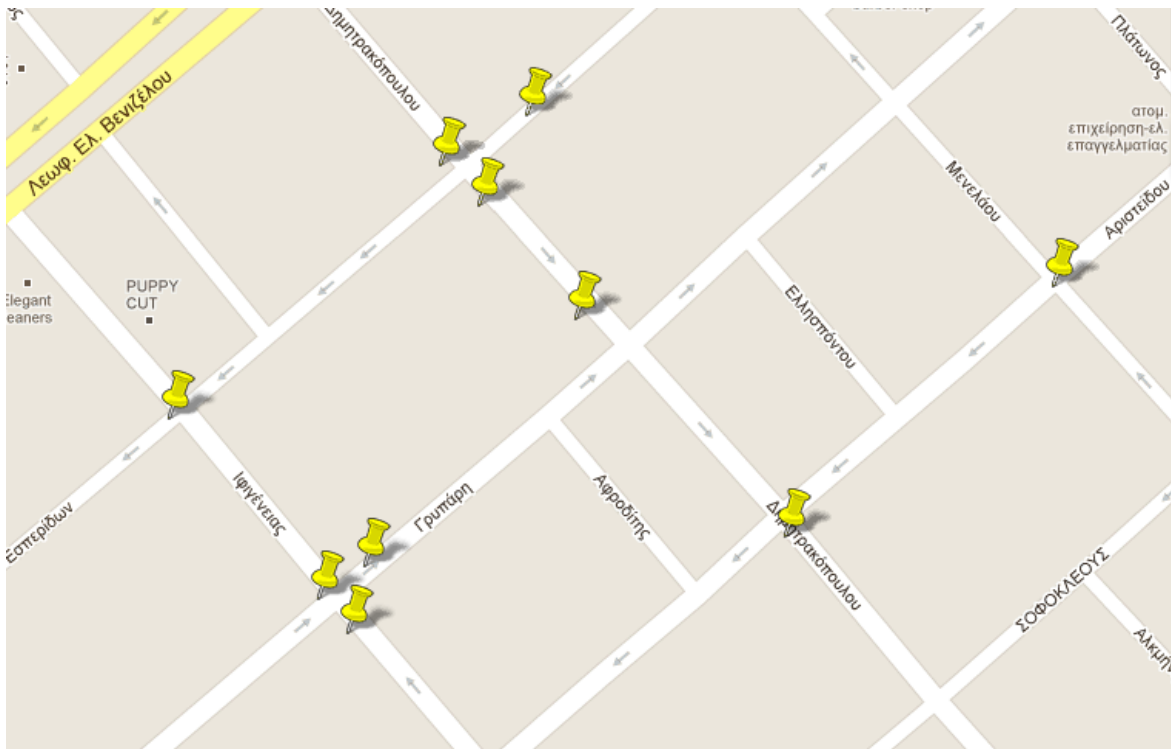
Στην τρίτη και τελευταία περίπτωση που υλοποιήθηκε, μελετήθηκε η εφαρμογή του συστήματος με τη χρήση κοντινών σε απόσταση αρχικών δεδομένων. Έτσι, επιλέχθηκαν χειροκίνητα στο περιβάλλον του Google Earth δέκα σημεία σε πολύ μικρή απόσταση μεταξύ τους, καθένα εκ των οποίων αντιπροσωπεύει ένα διαφορετικό τύπο προβλήματος. Γύρω από κάθε μία από τις δέκα περιπτώσεις προβλήματος θα δημιουργηθούν στη συνέχεια τυχαία σημεία, τα οποία θα αντιπροσωπεύουν τις καταγραφές που πραγματοποιήθηκαν για το συγκεκριμένο πρόβλημα.

Όπως και στις προηγούμενες περιπτώσεις, αποφασίσθηκε τα δέκα αρχικά σημεία να μη δημιουργηθούν με την αναπαραγωγή τυχαίων συντεταγμένων έτσι ώστε να μη

συμπίπτουν με κτίρια ή άλλα αντικείμενα. Για το λόγο αυτό επιλέχθηκε να βρίσκονται κοντά σε διασταυρώσεις δρόμων, ώστε να προσομοιάζουν πραγματικές περιπτώσεις καταγραφών. Τα σημεία αυτά παρουσιάζονται στο απόσπασμα του χάρτη του Google Maps που φαίνεται στη συνέχεια.



Εικόνα 6.32: Απεικόνιση των 10 αρχικών σημείων σε πολύ μικρή έκταση σε σχέση με τις προηγούμενες περιπτώσεις το περιβάλλον του Google Maps



Εικόνα 6.33: Μεγέθυνση των 10 επιλεγμένων αρχικών σημείων στο περιβάλλον του Google Maps

Οι γεωγραφικές συντεταγμένες των δέκα αρχικών σημείων που χρησιμοποιήθηκαν ως κέντρα για την αναπαραγωγή νέων τυχαίων σημείων σε μέγιστη απόσταση 25 μέτρα παρουσιάζονται στον πίνακα που ακολουθεί.

| α/α | φ | λ |
|-----|-----------|-----------|
| 1 | 37.986499 | 23.761300 |
| 2 | 38.012974 | 23.753113 |
| 3 | 37.976029 | 23.737596 |
| 4 | 38.000833 | 23.727542 |
| 5 | 37.968781 | 23.699485 |
| 6 | 37.995662 | 23.749169 |
| 7 | 37.988790 | 23.716786 |
| 8 | 38.019600 | 23.693327 |
| 9 | 38.005699 | 23.700340 |
| 10 | 37.987502 | 23.694391 |

Πίνακας 6.1: Γεωγραφικές συντεταγμένες των δέκα αρχικών σημείων

6.3.3 Υλοποίηση της εφαρμογής

Αφού διαμορφώθηκαν τα αναγκαία ερωτήματα για τη δημιουργία των πινάκων, την εισαγωγή των δεδομένων, και την επιλογή δεδομένων από τους πίνακες σε γλώσσα SQL, ενσωματώθηκε ο κώδικας αυτός στον τελικό κώδικα στη γλώσσα PHP, και τέθηκε το σύστημα σε λειτουργία αφήνοντάς το να «τρέξει» μέσω του παραθύρου της γραμμής εντολών.

Τα αρχεία του αλγορίθμου που τέθηκαν σε λειτουργία είναι πέντε:

- **F3_create_tables.php**, για τη δημιουργία των πινάκων «damage_types3» και «crowd_data3» στη βάση δεδομένων.
- **F3_data_insertion_damage_types.php**, για την εισαγωγή των δεδομένων στον πίνακα «damage_types3».
- **F3_create_and_insert_points_crowd_data.php**, για τη δημιουργία των τυχαίων σημείων των καταγραφών γύρω από τα δέκα αρχικά σημεία, τη διαμόρφωση του τελικού csv αρχείου με όλες τις πληροφορίες των καταγραφών, και την εισαγωγή όλων των πληροφοριών στον πίνακα «crowd_data3».
- **F3_kmeans.php**, για την υλοποίηση της ομαδοποίησης των καταγραφών των προβλημάτων με χρήση του αλγορίθμου `-kmeans` και την προσθήκη των πινάκων «kmeans_polygons3», «kmeans_points3» και «unique_cases3» στη βάση δεδομένων.
- **F3_kml_points_and_html.php**, για τον υπολογισμό των μοναδικών περιπτώσεων προβλημάτων και τη δημιουργία ενός kml αρχείου για κάθε τύπο προβλήματος, τη δημιουργία ενός kml αρχείου για κάθε τύπο προβλήματος για το σύνολο των δεδομένων, και τη δημιουργία ενός html αρχείου για την απεικόνιση των δεδομένων των kmls σε διαδικτυακούς χάρτες.

Τα αρχεία «F3_kmeans.php», «F3_kml_avg_points.php» και «F3_kml_all_data.php» υλοποιήθηκαν επτά φορές, για κάθε μία από τις οποίες αποδόθηκε διαφορετική τιμή στην παράμετρο της απόστασης κατά τη δημιουργία των πολυγώνων ομαδοποίησης, με αποτέλεσμα την εφαρμογή επτά διαφορετικών περιπτώσεων ομαδοποιήσεων για τα ίδια δεδομένα.

Η εισαγωγή των δεδομένων στον πίνακα «damage_types3» πραγματοποιήθηκε με άντληση της πληροφορίας από txt αρχεία, ενώ η εισαγωγή των δεδομένων στον πίνακα «crowd_data3» πραγματοποιήθηκε με άντληση της πληροφορίας από κατάλληλα διαμορφωμένο csv αρχείο.

Μετά την ολοκλήρωση του «τρεξίματος» των πέντε rhr αρχείων, προέκυψαν πέντε διαφορετικοί πίνακες με τα αντίστοιχα δεδομένα μέσα στη βάση δεδομένων του λογισμικού PostgreSQL. Επιπλέον, εξήχθησαν δύο kml αρχεία για κάθε τύπο προβλήματος με τα αποτελέσματα που προέκυψαν από την υλοποίηση του κώδικα.

Στην εικόνα που ακολουθεί παρουσιάζονται οι πίνακες και τα πεδία κάθε πίνακα που δημιουργήθηκαν στη βάση δεδομένων έπειτα από την υλοποίηση του αλγορίθμου.

| crowd_data3 | damage_types3 | kmeans_polygons3 | kmeans_points3 | unique_cases3 |
|--|--|--|---|--|
| <input type="checkbox"/> id | <input type="checkbox"/> t_code | <input type="checkbox"/> polygons_gm | <input type="checkbox"/> kmeans | <input type="checkbox"/> kmeans |
| <input type="checkbox"/> latitude | <input type="checkbox"/> description | <input type="checkbox"/> centroids_geom | <input type="checkbox"/> latitude | <input type="checkbox"/> uc_damage_types |
| <input type="checkbox"/> longitude | <input type="checkbox"/> colour | <input type="checkbox"/> centroids_latitude | <input type="checkbox"/> longitude | <input type="checkbox"/> avg_latitude |
| <input type="checkbox"/> name | <input type="checkbox"/> kml_unique_colour | <input type="checkbox"/> centroids_longitude | <input type="checkbox"/> severity | <input type="checkbox"/> avg_longitude |
| <input type="checkbox"/> surname | <input type="checkbox"/> kml_crowd_colour | | <input type="checkbox"/> type_of_damage | <input type="checkbox"/> avg_severity |
| <input type="checkbox"/> phone_number | <input type="checkbox"/> styleurl_unique | | <input type="checkbox"/> cd_id | <input type="checkbox"/> num_of_records |
| <input type="checkbox"/> email | <input type="checkbox"/> styleurl_crowd | | <input type="checkbox"/> id | <input type="checkbox"/> uc_id |
| <input type="checkbox"/> city | <input type="checkbox"/> icon | | <input type="checkbox"/> geom | <input type="checkbox"/> geom |
| <input type="checkbox"/> type_of_damage | | | | |
| <input type="checkbox"/> short_description | | | | |
| <input type="checkbox"/> date | | | | |
| <input type="checkbox"/> severity | | | | |
| <input type="checkbox"/> capture | | | | |
| <input type="checkbox"/> updated | | | | |
| <input type="checkbox"/> created | | | | |
| <input type="checkbox"/> project | | | | |
| <input type="checkbox"/> created_by | | | | |
| <input type="checkbox"/> updated_by | | | | |
| <input type="checkbox"/> geom | | | | |
| <input type="checkbox"/> unique_cases_id | | | | |

Εικόνα 6.34: Οι πίνακες που δημιουργήθηκαν στη βάση δεδομένων για την «Περίπτωση 3»

6.3.4 Απεικόνιση ακατέργαστων αποτελεσμάτων στο Quantum GIS

Στην παρούσα ενότητα παρουσιάζεται η πορεία των ακατέργαστων δεδομένων στο περιβάλλον του λογισμικού Quantum GIS, πριν την απεικόνισή τους στους τελικούς διαδικτυακούς χάρτες μέσω των kml αρχείων. Για την απεικόνιση των δεδομένων της βάσης δεδομένων στο Quantum GIS απαιτείται η υλοποίηση σύνδεσης με την συγκεκριμένη βάση δεδομένων όπως περιγράφηκε σε προηγούμενη ενότητα.

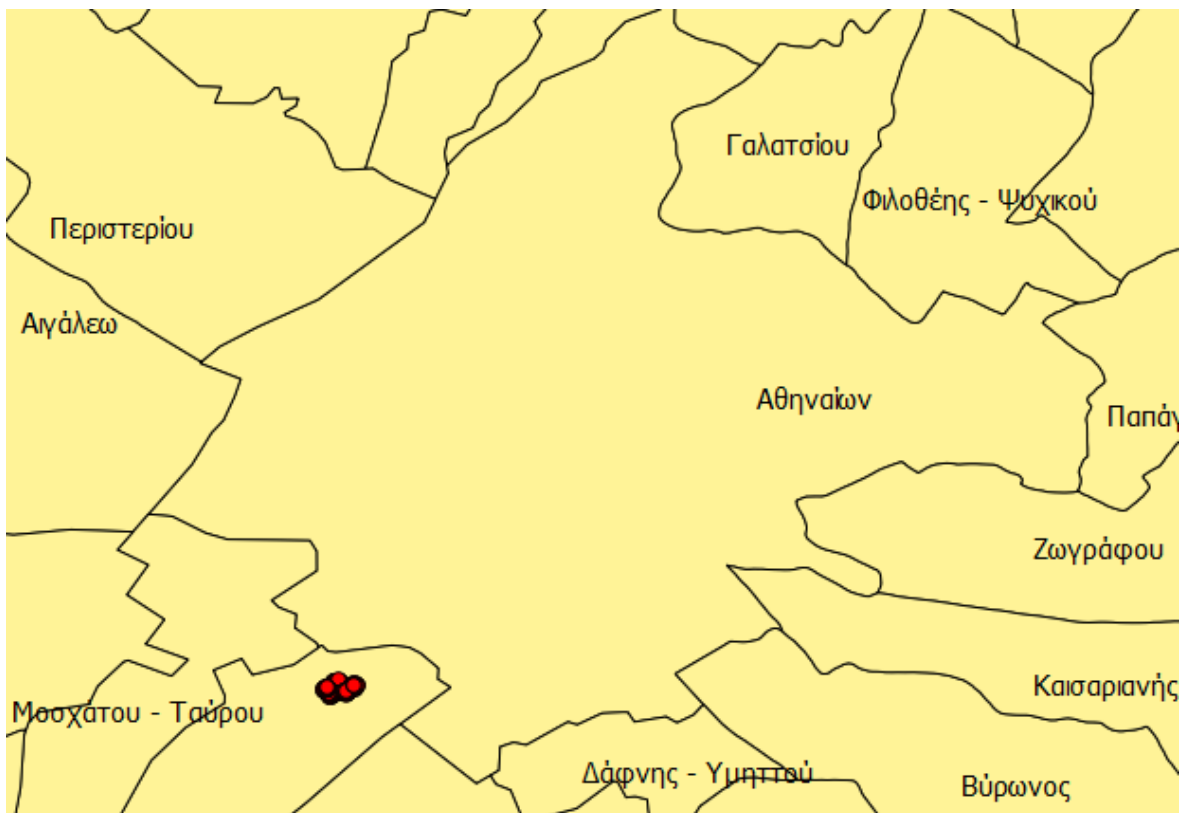
Όλοι οι χάρτες που παρουσιάζονται έχουν ως υπόβαθρο χάρτη των καλλικρατικών δήμων της Ελλάδος. Στους χάρτες απεικονίζονται τα δεδομένα με τη σειρά που υλοποιήθηκαν στη βάση δεδομένων. Πιο συγκεκριμένα απεικονίζονται:

- ✓ τα αρχικά δεδομένα όπως προκύπτουν από τις καταγραφές των προβλημάτων από το πλήθος,
- ✓ τα δεδομένα των καταγραφών που πραγματοποιήθηκαν για ένα συγκεκριμένο τύπο προβλήματος,
- ✓ τα πολύγωνα ομαδοποίησης που προέκυψαν από την υλοποίηση της πρώτης ομαδοποίησης των σημείων για την ορισμό των διανυσμάτων αρχικοποίησης,
- ✓ οι ομάδες σημείων των καταγραφών που προέκυψαν από την υλοποίηση της ομαδοποίησης με τον αλγόριθμο k-means,

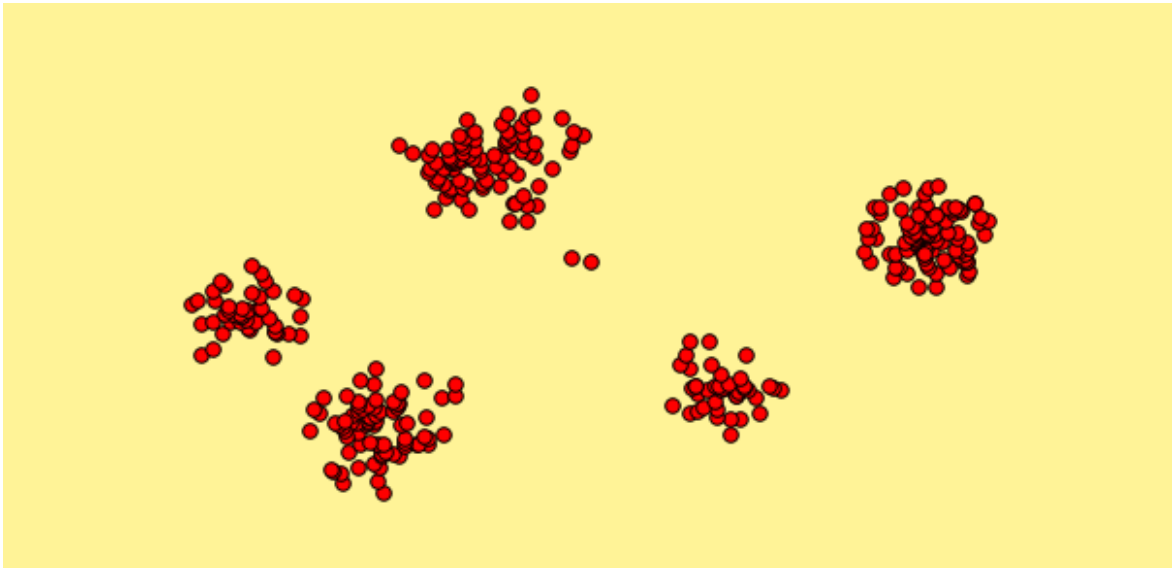
- ✓ και οι μοναδικές περιπτώσεις προβλημάτων που αντιπροσωπεύουν το σύνολο των καταγραφών για κάθε περίπτωση.

Στο σημείο αυτό πρέπει να σημειωθούν οι εξής παρατηρήσεις:

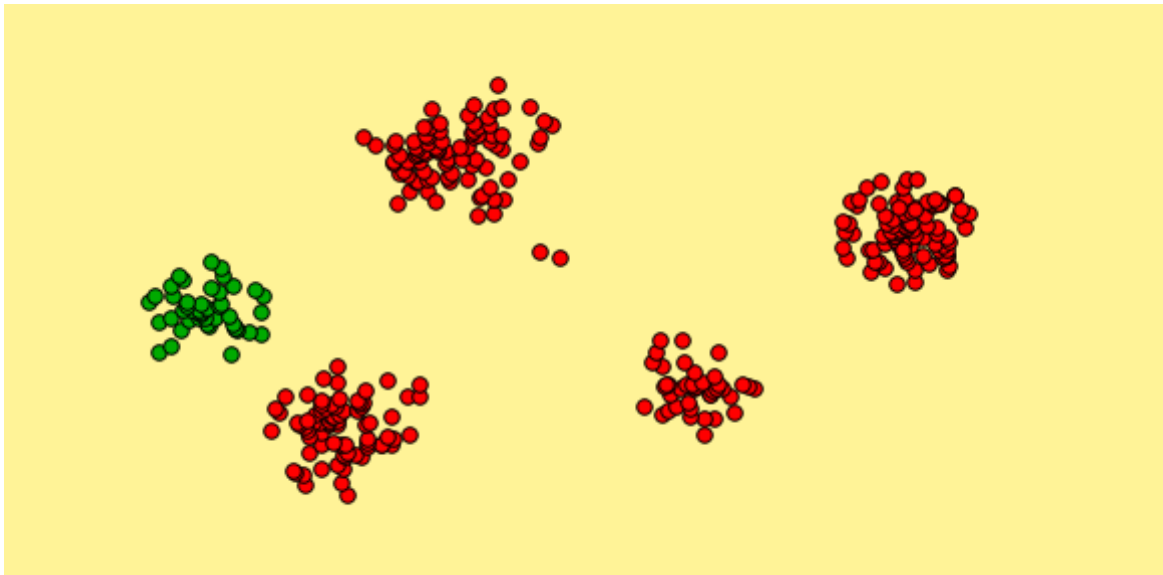
- Στις εικόνες 6.30 παρουσιάζεται το σύνολο των σημείων των καταγραφών σε σχέση με την έκταση του συνόλου των σημείων των καταγραφών των δύο προηγούμενων περιπτώσεων. Παρατηρούμε ότι η έκταση που καλύπτουν τα σημεία των δύο προηγούμενων περιπτώσεων είναι πολύ μεγάλη, ενώ η έκταση που καλύπτουν τα σημεία της τρίτης περίπτωσης περιορίζεται σε πολύ μικρή περιοχή. Η πραγματική εικόνα των σημείων των καταγραφών παρουσιάζεται στην εικόνα 6.31 σε μεγέθυνση.
- Στην εικόνα 6.32 παρουσιάζεται το σύνολο των καταγραφών για την περίπτωση προβλήματος «σπασμένο πεζοδρόμιο» που απεικονίζεται με πράσινο χρώμα, σε σχέση με τις καταγραφές για τα υπόλοιπα προβλήματα που διαφοροποιούνται με κόκκινο χρώμα.
- Στην εικόνα 6.33 παρουσιάζεται ο χάρτης των πολυγώνων ομαδοποίησης, καθώς επίσης και τα σημεία των καταγραφών που περικλείονται σε κάθε πολύγωνο και ανήκουν στην ομάδα που ορίζει το κάθε πολύγωνο. Τα πολύγωνα της αρχικής ομαδοποίησης προέκυψαν για μέγιστη απόσταση 20 μέτρα μεταξύ των σημείων μίας ομάδας.
- Στην εικόνα 6.35 κάθε ομάδα σημείων που προέκυψε από την ομαδοποίηση k-means απεικονίζεται με το ίδιο χρώμα.
- Τέλος, στην εικόνα 6.36 απεικονίζονται με μπλε χρώμα οι μέσοι όροι των σημείων των καταγραφών ανά περίπτωση προβλήματος, ενώ στην εικόνα 6.37 απεικονίζονται τα ίδια σημεία σε σχέση με τις αρχικές καταγραφές από τις οποίες προέκυψαν.



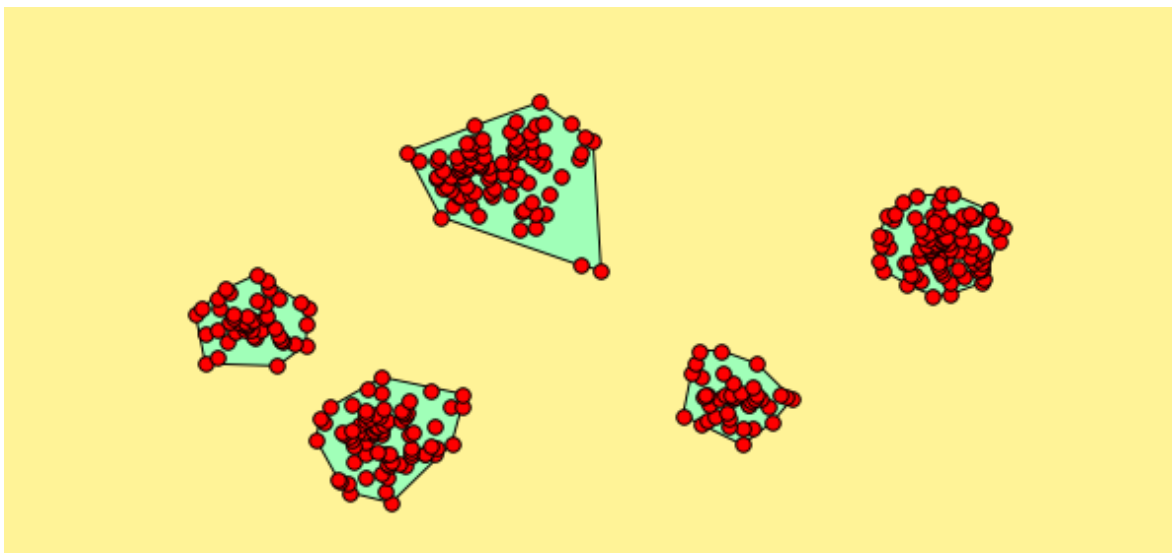
Εικόνα 6.35: Απεικόνιση του συνόλου των αρχικών δεδομένων των καταγραφών



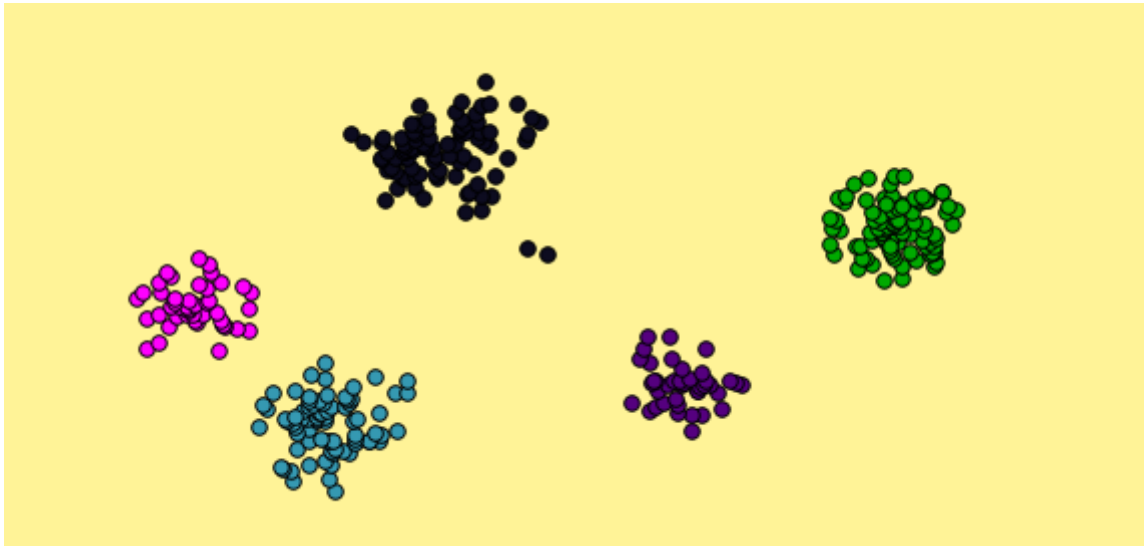
Εικόνα 6.36: Απεικόνιση του συνόλου των αρχικών δεδομένων των καταγραφών



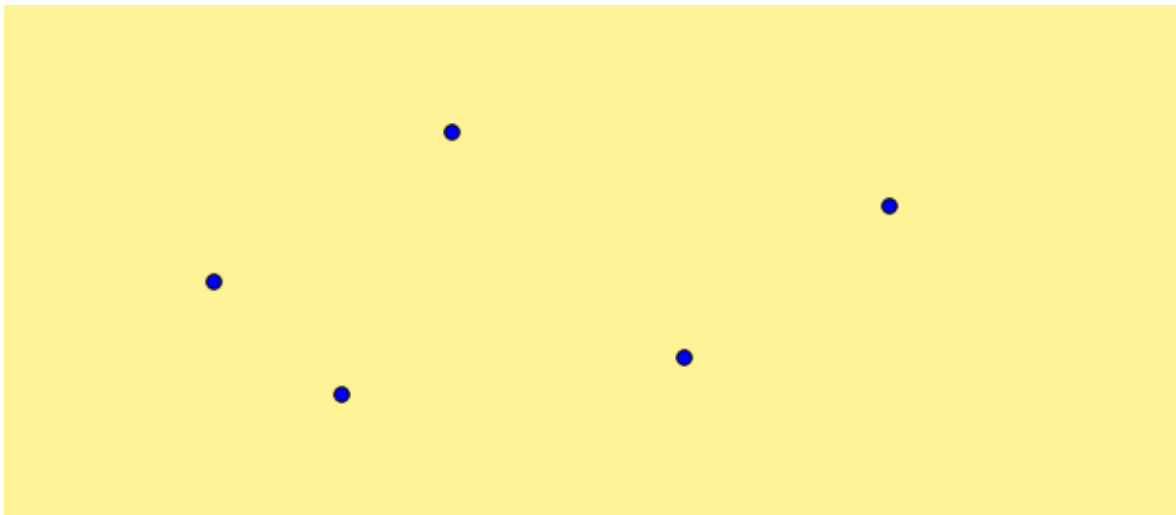
Εικόνα 6.37: Σύνολο των καταγραφών για τον τύπο προβλήματος «σπασμένο πεζοδρόμιο» (πράσινο χρώμα)



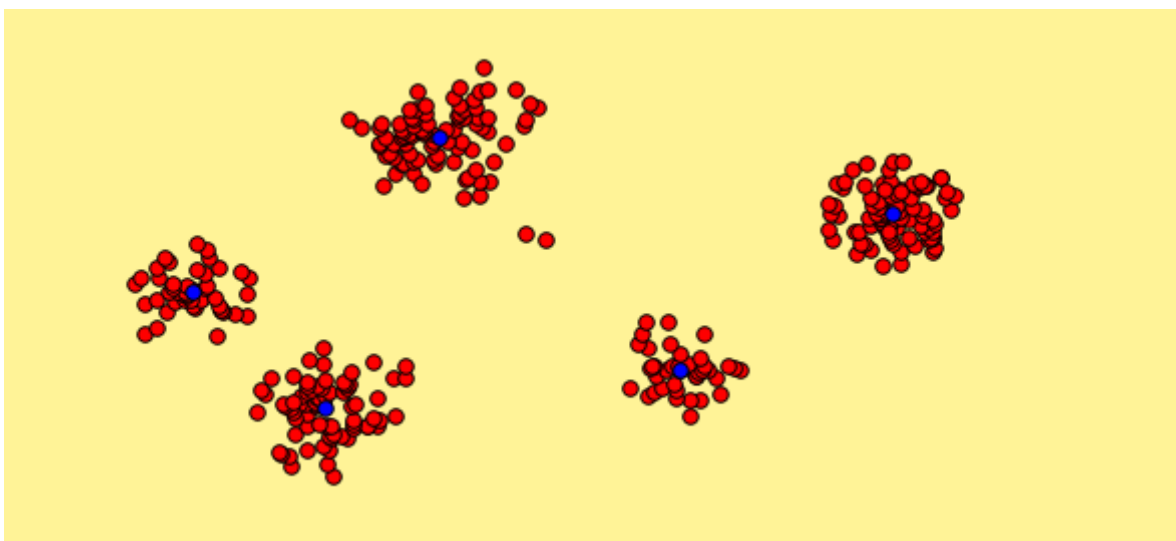
Εικόνα 6.38: Απεικόνιση πολυγώνων ομαδοποίησης και των σημείων των καταγραφών



Εικόνα 6.39: Χρωματική διαφοροποίηση των δεδομένων με βάση την ομαδοποίηση k-means



Εικόνα 6.40: Απεικόνιση των μοναδικών σημείων που αντιπροσωπεύουν το σύνολο των καταγραφών κάθε περίπτωσης



Εικόνα 6.41: Απεικόνιση των μοναδικών σημείων σε σχέση με το σύνολο των καταγραφών κάθε περίπτωσης

6.3.5 Τελικά αποτελέσματα


Στην ενότητα αυτή παρατίθενται τα αποτελέσματα του «τρεξίματος» της εφαρμογής για την τρίτη περίπτωση, όπως προέκυψαν από τη δημιουργία των kml αρχείων. Πραγματοποιήθηκαν επτά διαφορετικές ομαδοποιήσεις των σημείων των καταγραφών με βάση τη μέγιστη δυνατή απόσταση μεταξύ των σημείων κατά τη δημιουργία των πολυγώνων ομαδοποίησης.

Από την εφαρμογή του αλγορίθμου δημιουργήθηκαν τα εξής KML αρχεία:

- | | |
|------------------------|-------------------------|
| ➤ 1-CROWD-20130226.kml | ➤ 6-CROWD-20130226.kml |
| ➤ 1-UNIQ-20130226.kml | ➤ 6-UNIQ-20130226.kml |
| ➤ 2-CROWD-20130226.kml | ➤ 7-CROWD-20130226.kml |
| ➤ 2-UNIQ-20130226.kml | ➤ 7-UNIQ-20130226.kml |
| ➤ 3-CROWD-20130226.kml | ➤ 8-CROWD-20130226.kml |
| ➤ 3-UNIQ-20130226.kml | ➤ 8-UNIQ-20130226.kml |
| ➤ 4-CROWD-20130226.kml | ➤ 9-CROWD-20130226.kml |
| ➤ 4-UNIQ-20130226.kml | ➤ 9-UNIQ-20130226.kml |
| ➤ 5-CROWD-20130226.kml | ➤ 10-CROWD-20130226.kml |
| ➤ 5-UNIQ-20130226.kml | ➤ 10-UNIQ-20130226.kml |

Κάθε αριθμός στο όνομα των αρχείων υποδηλώνει τον τύπο προβλήματος που απεικονίζεται. Οι αντιστοιχίες των αριθμών και των τύπων προβλήματος παρουσιάζονται στη συνέχεια.

Εδώ πρέπει να σημειωθεί ότι επειδή ο αρχικός αριθμός σημείων που δημιουργήθηκαν ήταν 10, δηλαδή μικρότερος από τον αριθμό των τύπων προβλημάτων που δηλώθηκαν αρχικά, στην τρίτη περίπτωση που μελετάται είναι αναμενόμενο να μη συμμετέχουν και οι 15 αρχικοί τύποι προβλήματος στην υλοποίηση της εφαρμογής. Επειδή στη συγκεκριμένη περίπτωση μελετάται το ενδεχόμενο γεινίασης δύο ή και παραπάνω περιπτώσεων όμοιου τύπου προβλημάτων, κάποιες από τις κατηγορίες προβλημάτων επαναλαμβάνονται. Συγκεκριμένα, οι τύποι προβλήματος που χρησιμοποιούνται στην πρώτη περίπτωση εφαρμογής και παρουσιάζονται στην αριστερή στήλη του παρακάτω σχήματος μετατρέπονται στους δέκα επαναλαμβανόμενους ανά κατηγορίες τύπους που παρουσιάζονται στη δεξιά στήλη του σχήματος. Έτσι κάθε περίπτωση προβλήματος αντιστοιχίζεται με έναν τύπο προβλήματος της δεξιάς στήλης του σχήματος με τη συγκεκριμένη ιεραρχία. Επομένως όλα τα τυχαία σημεία που θα δημιουργηθούν γύρω από τα αρχικά σημεία 1, 2 και 3 θα αντιστοιχηθούν με το πρόβλημα «02-λακκούβα», ενώ τα τυχαία σημεία που θα δημιουργηθούν γύρω από τα αρχικά σημεία 6, 7, 8 και 9 θα αντιστοιχηθούν με το πρόβλημα «06-αλλοιωμένη ταμπέλα» σύμφωνα με τη δεξιά στήλη του σχήματος που ακολουθεί.

| | | |
|-------------------------|---|-----------------------|
| 01-ανώμαλο οδόστρωμα | | 02-λακκούβα |
| 02-λακκούβα | | 02-λακκούβα |
| 03-σπασμένο πεζοδρόμιο | | 02-λακκούβα |
| 04-χαλασμένο φανάρι | | 04-χαλασμένο φανάρι |
| 05-πεσμένο δέντρο |  | 05-πεσμένο δέντρο |
| 06-αλλοιωμένη ταμπέλα | | 06-αλλοιωμένη ταμπέλα |
| 07-υπερχείλιση φρεατίου | | 06-αλλοιωμένη ταμπέλα |
| 08-διαρροή αγωγού | | 06-αλλοιωμένη ταμπέλα |

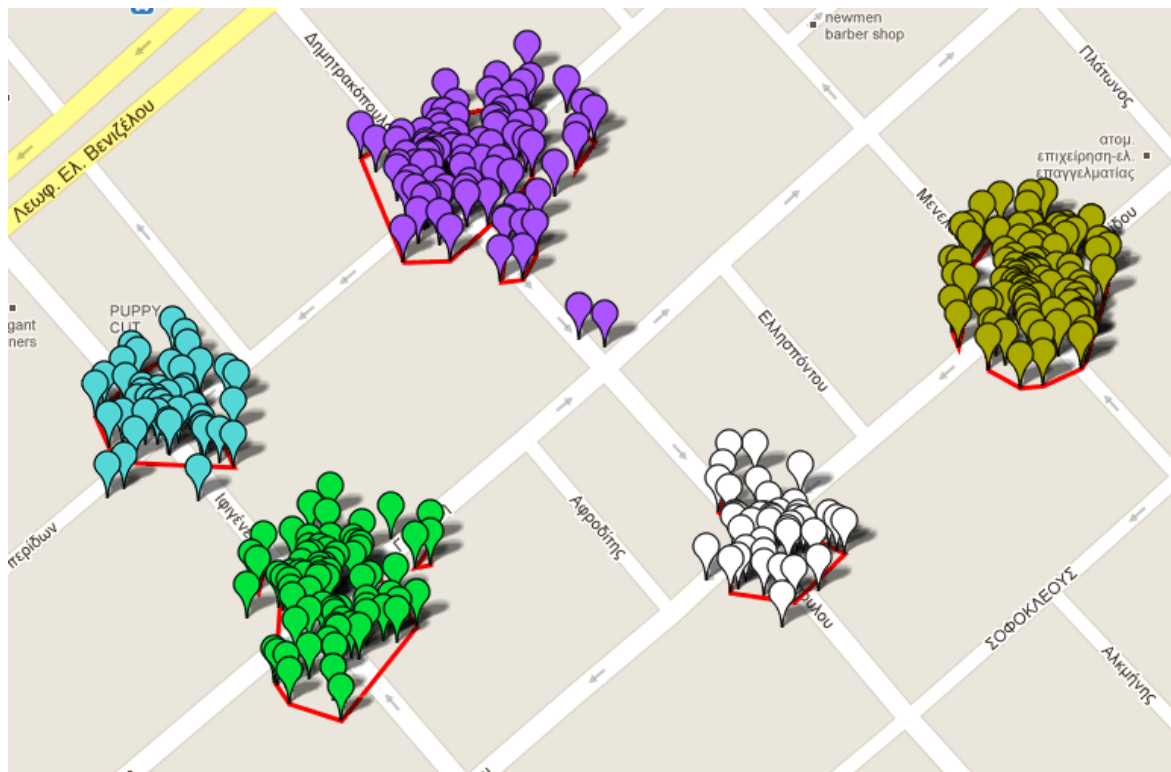
09-καμένη λάμπα
10-σπασμένο παγκάκι

06-αλλοιωμένη ταμπέλα
10-σπασμένο παγκάκι

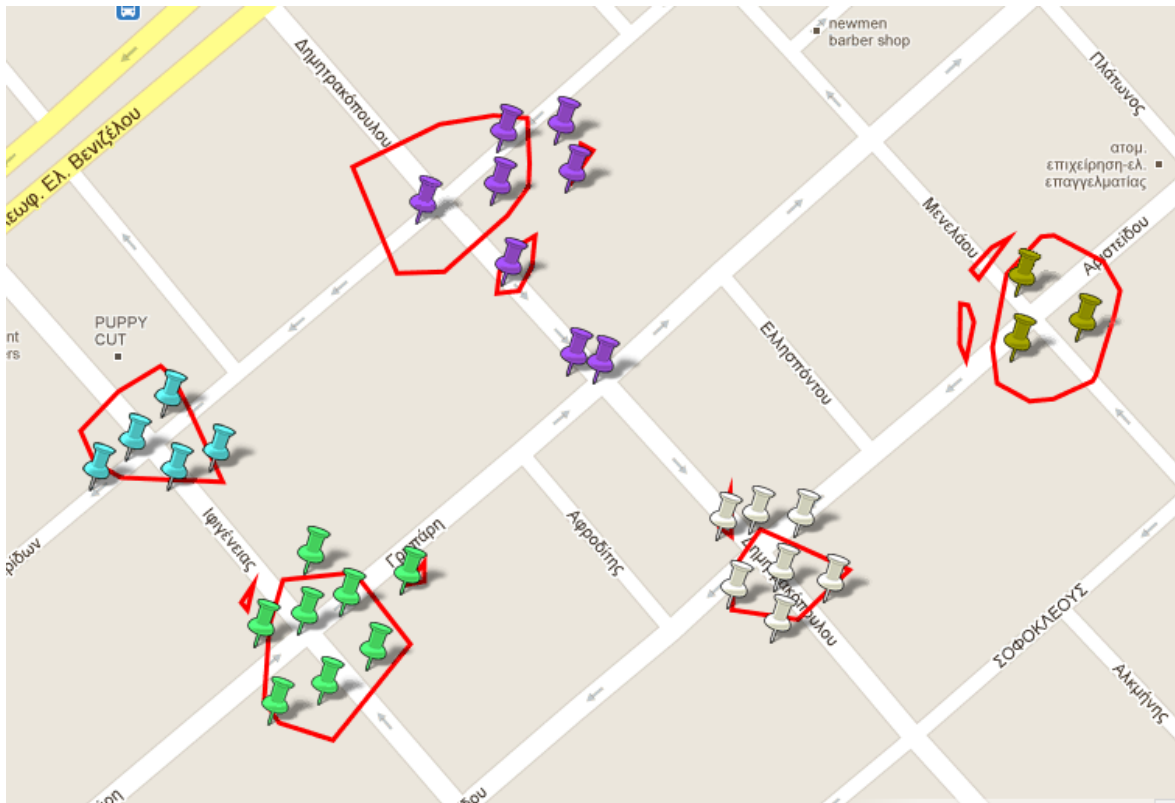
Όπως και στις δύο προηγούμενες περιπτώσεις εφαρμογής, έτσι και σε αυτήν την περίπτωση, δημιουργήθηκε ένα αρχείο CROWD για την απεικόνιση του συνόλου των καταγραφών ανά πρόβλημα για κάθε έναν από τους παραπάνω τύπους προβλημάτων, ενώ για την απεικόνιση των μέσων σημείων κάθε ομάδας καταγραφών ανά πρόβλημα δημιουργήθηκε από ένα αρχείο UNIQ.

Στη συνέχεια απεικονίστηκε το σύνολο των αποτελεσμάτων στο περιβάλλον του Google Maps αλλά και στην προσωπική ιστοσελίδα για όλες τις περιπτώσεις ομαδοποιήσεων. Παρουσιάζονται οι χάρτες του συνόλου των καταγραφών, καθώς επίσης και των μοναδικών περιπτώσεων που αντιστοιχούν στη μέση τιμή των καταγραφών για κάθε περίπτωση προβλήματος. Οι τιμές της μέγιστης δυνατής απόστασης μεταξύ των σημείων κατά τη δημιουργία των πολυγώνων ομαδοποίησης κυμαίνονται από 10 έως 40 μέτρα. Σε όλους τους χάρτες παρουσιάζονται οι ακόλουθοι τύποι προβλημάτων με τα αντίστοιχα χρώματα:

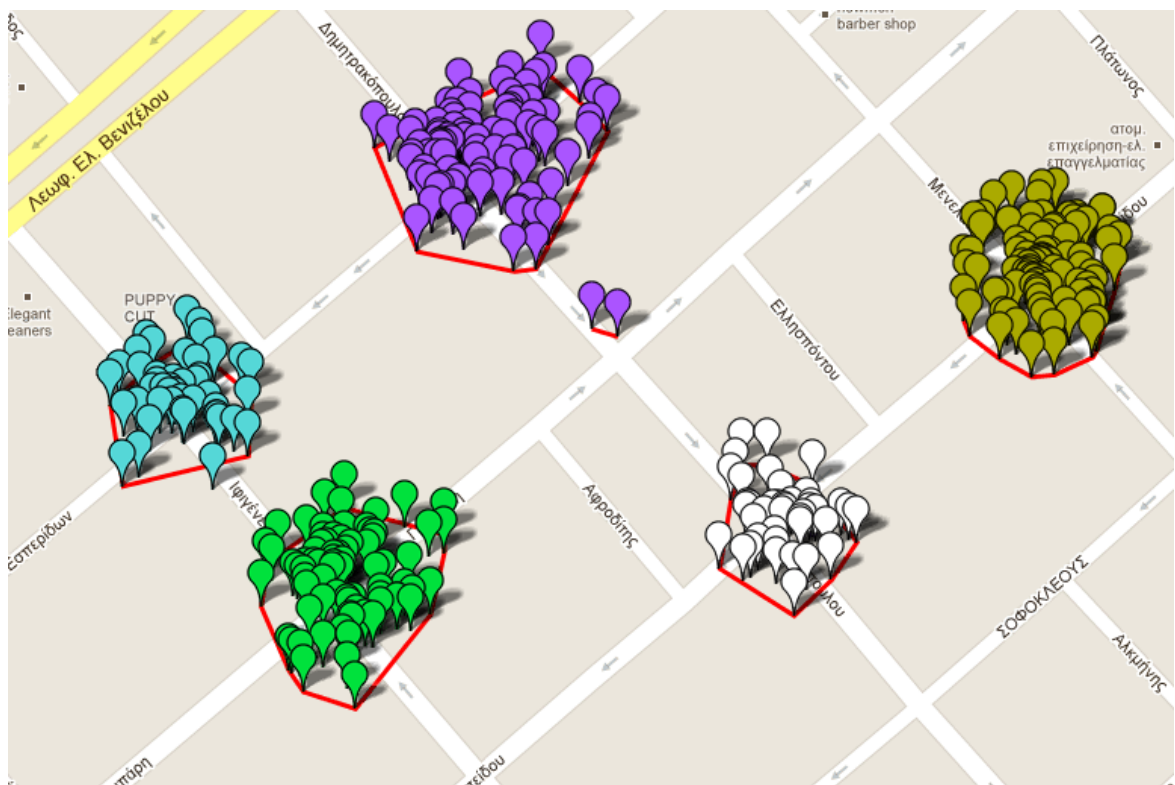
| | | | |
|--|---------------------|--|--------------------|
|  | λακούβα |  | αλλοιωμένη ταμπέλα |
|  | σπασμένο πεζοδρόμιο |  | σπασμένο παγκάκι |
|  | πεσμένο δέντρο |  | όρια πολυγώνου |



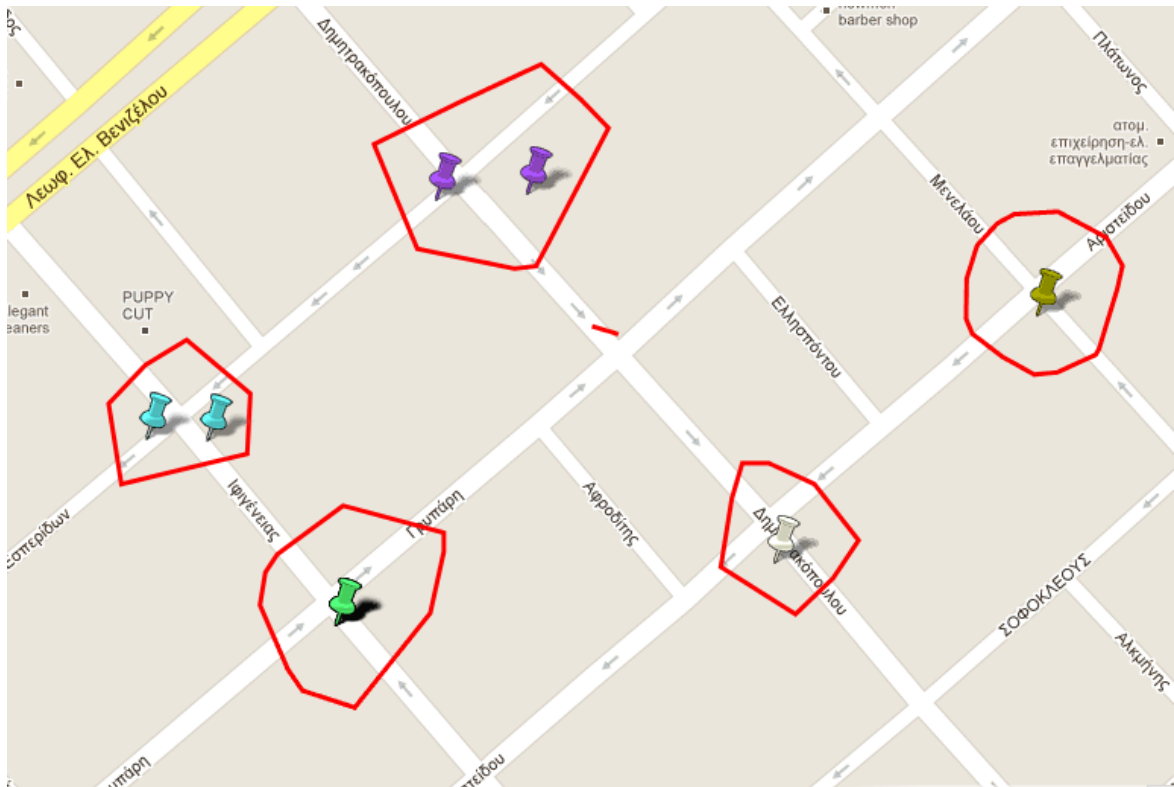
Εικόνα 6.42: Απεικόνιση του συνόλου των μοναδικών περιπτώσεων ανά τύπο προβλήματος για απόσταση ομαδοποίησης 10 m στο περιβάλλον του Google Maps για την «Περίπτωση 3»



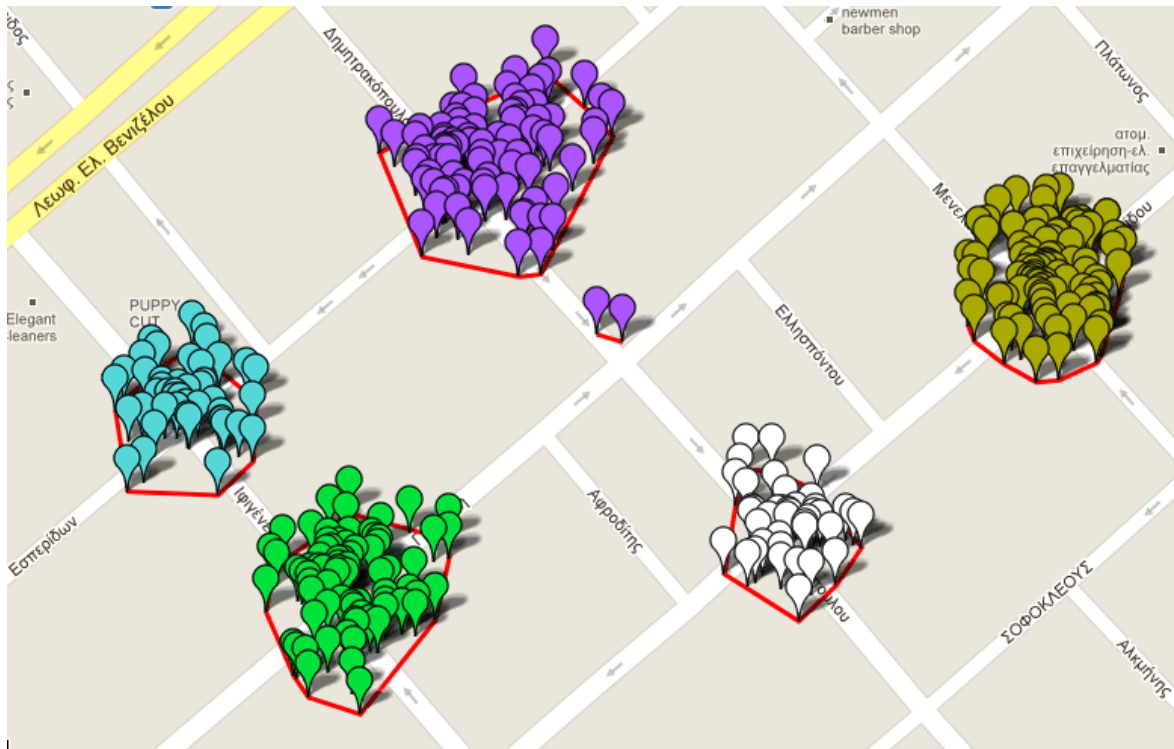
Εικόνα 6.43: Απεικόνιση του συνόλου των μοναδικών περιπτώσεων ανά τύπο προβλήματος για απόσταση ομαδοποίησης 10 m στο περιβάλλον του Google Maps για την «Περίπτωση 3»



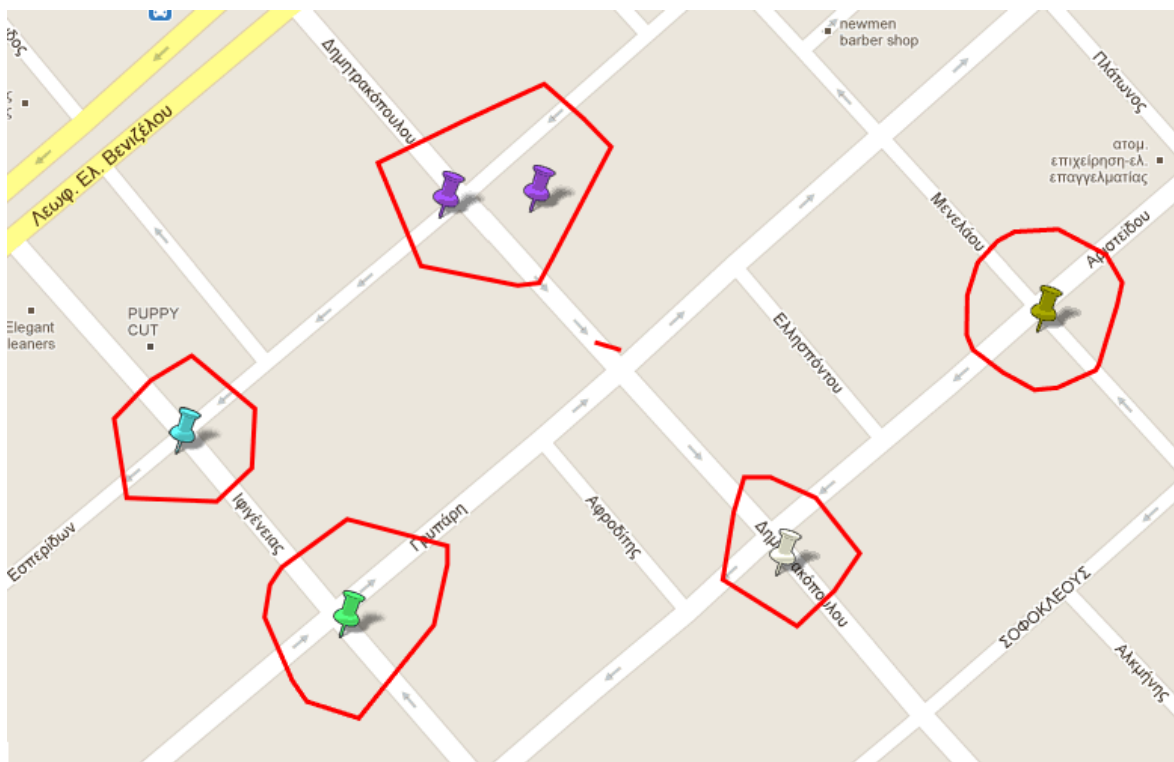
Εικόνα 6.44: Απεικόνιση του συνόλου των μοναδικών περιπτώσεων ανά τύπο προβλήματος για απόσταση ομαδοποίησης 15 m στο περιβάλλον του Google Maps για την «Περίπτωση 3»



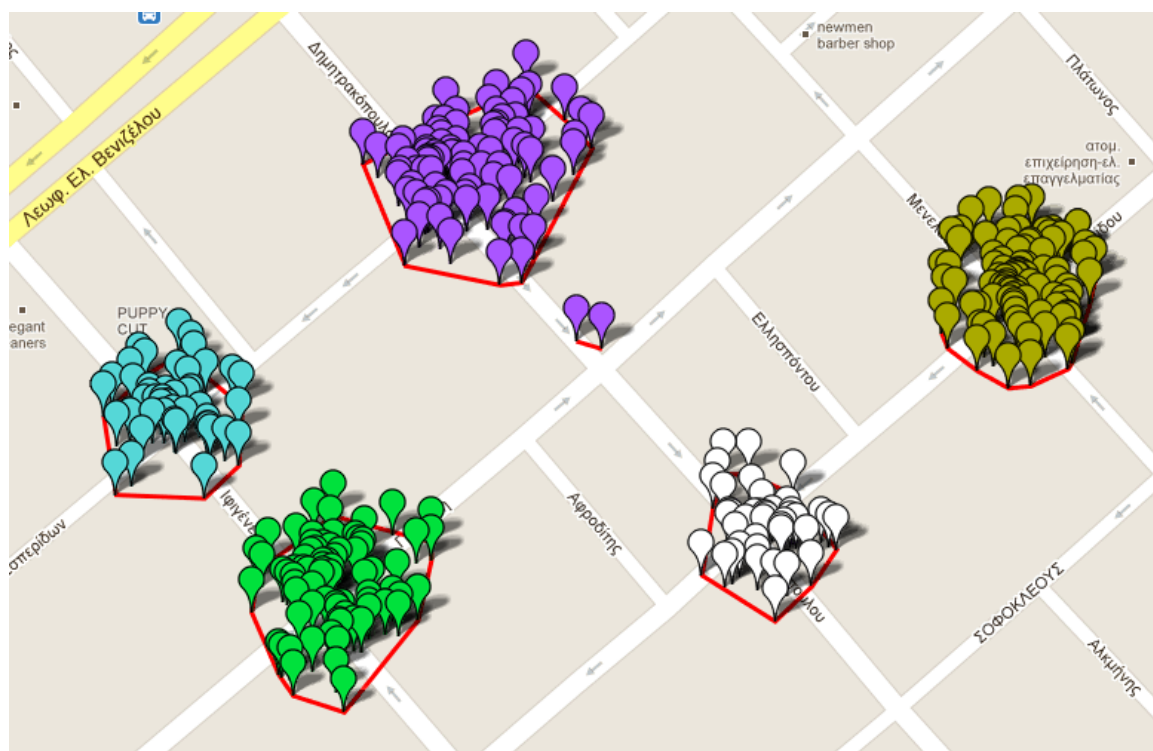
Εικόνα 6.45: Απεικόνιση του συνόλου των μοναδικών περιπτώσεων ανά τύπο προβλήματος για απόσταση ομαδοποίησης 15 m στο περιβάλλον του Google Maps για την «Περίπτωση 3»



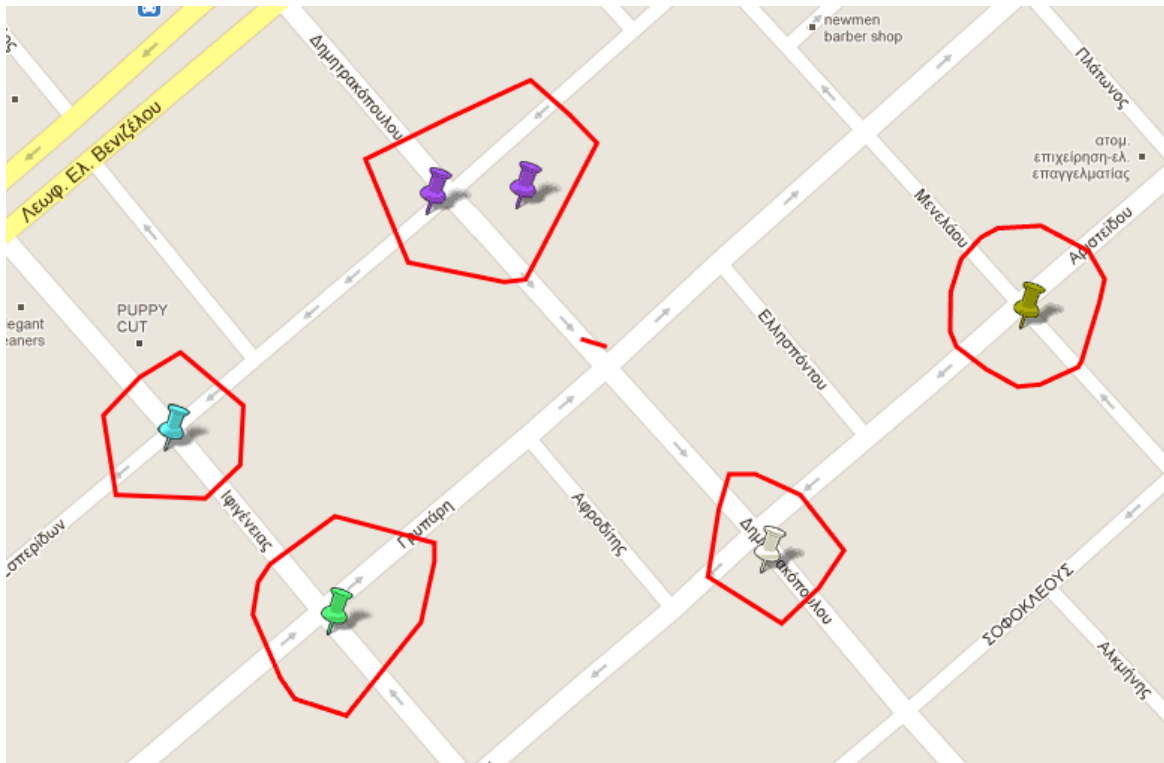
Εικόνα 6.46: Απεικόνιση του συνόλου των μοναδικών περιπτώσεων ανά τύπο προβλήματος για απόσταση ομαδοποίησης 20 m στο περιβάλλον του Google Maps για την «Περίπτωση 3»



Εικόνα 6.47: Απεικόνιση του συνόλου των μοναδικών περιπτώσεων ανά τύπο προβλήματος για απόσταση ομαδοποίησης 20 m στο περιβάλλον του Google Maps για την «Περίπτωση 3»



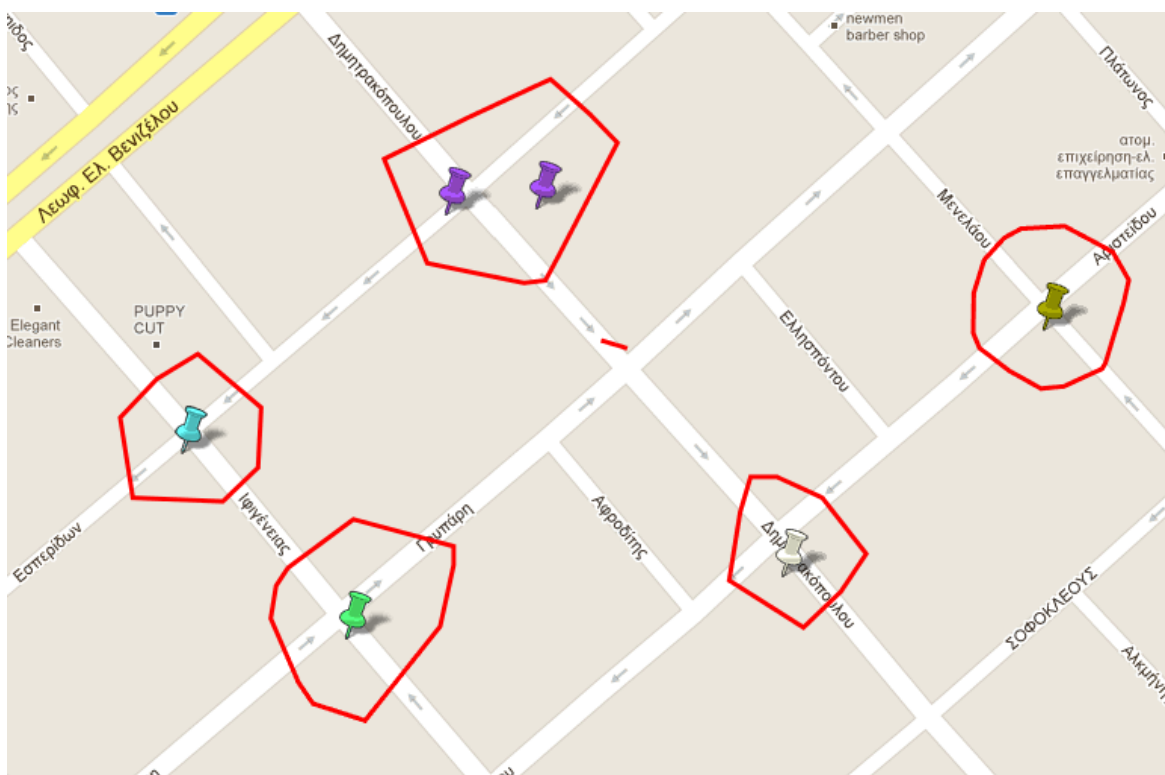
Εικόνα 6.48: Απεικόνιση του συνόλου των μοναδικών περιπτώσεων ανά τύπο προβλήματος για απόσταση ομαδοποίησης 25 m στο περιβάλλον του Google Maps για την «Περίπτωση 3»



Εικόνα 6.49: Απεικόνιση του συνόλου των μοναδικών περιπτώσεων ανά τύπο προβλήματος για απόσταση ομαδοποίησης 25 m στο περιβάλλον του Google Maps για την «Περίπτωση 3»



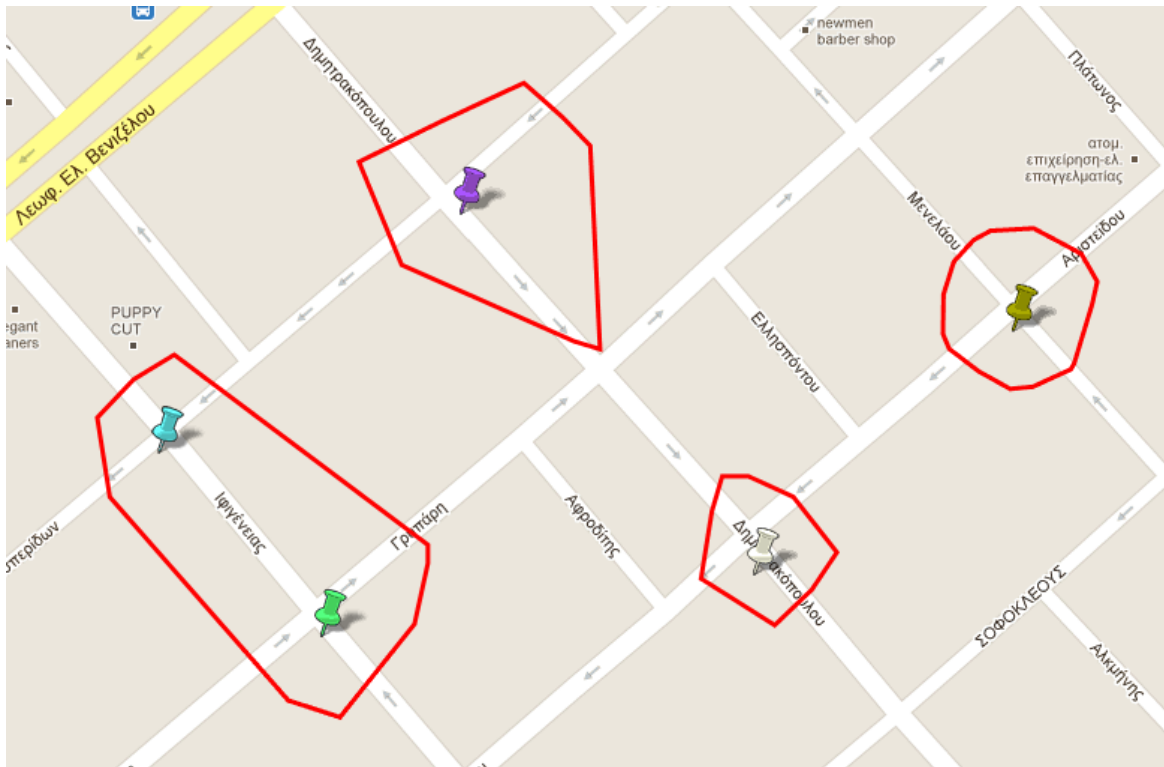
Εικόνα 6.50: Απεικόνιση του συνόλου των μοναδικών περιπτώσεων ανά τύπο προβλήματος για απόσταση ομαδοποίησης 30 m στο περιβάλλον του Google Maps για την «Περίπτωση 3»



Εικόνα 6.51: Απεικόνιση του συνόλου των μοναδικών περιπτώσεων ανά τύπο προβλήματος για απόσταση ομαδοποίησης 30 m στο περιβάλλον του Google Maps για την «Περίπτωση 3»



Εικόνα 6.52: Απεικόνιση του συνόλου των μοναδικών περιπτώσεων ανά τύπο προβλήματος για απόσταση ομαδοποίησης 35 m στο περιβάλλον του Google Maps για την «Περίπτωση 3»



Εικόνα 6.55: Απεικόνιση του συνόλου των μοναδικών περιπτώσεων ανά τύπο προβλήματος για απόσταση ομαδοποίησης 40 m στο περιβάλλον του Google Maps για την «Περίπτωση 3»

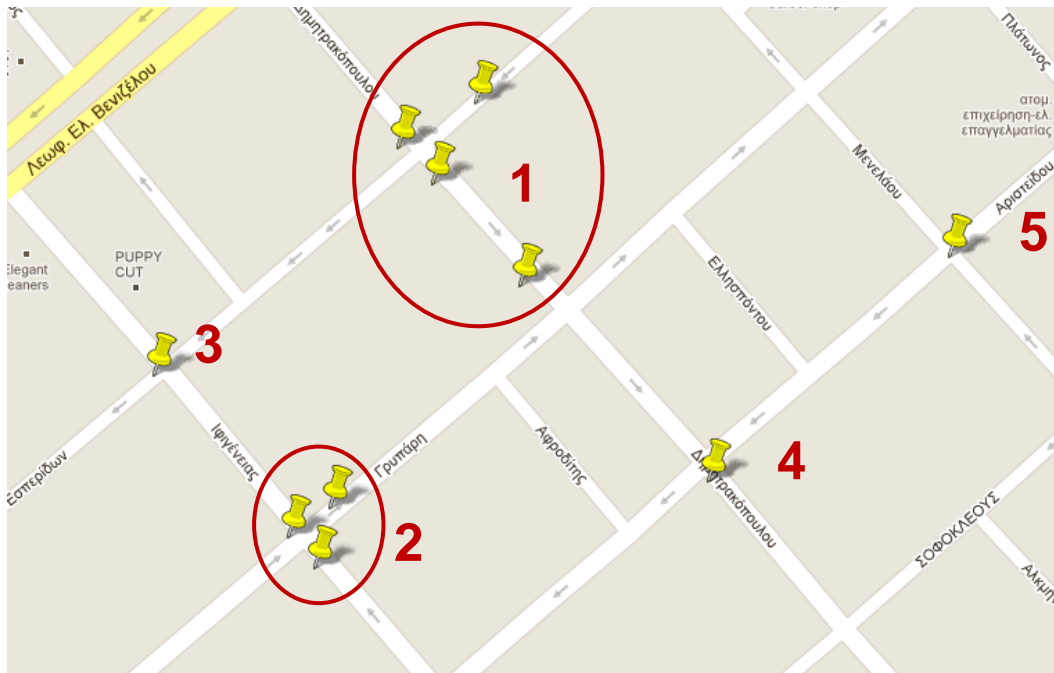
Η κόκκινη γραμμή στις παραπάνω απεικονίσεις οριοθετεί τα πολύγωνα που δημιουργήθηκαν από την πρώτη ομαδοποίηση που εφαρμόστηκε για την εύρεση των σημείων αρχικοποίησης. Τα χρώματα των σημείων προέκυψαν με βάση τη δεύτερη ομαδοποίηση των δεδομένων με τον αλγόριθμο k-means.

Παρατηρούμε ότι στην περίπτωση που δύο περιπτώσεις προβλημάτων όμοιου τύπου βρίσκονται πολύ κοντά, και στην περίπτωση που μελετάμε σε απόσταση μικρότερη των 50 μέτρων που είναι το άθροισμα των ακτινών της περιοχής των καταγραφών γύρω από δύο προβλήματα, τότε ο αλγόριθμος ομαδοποίησης δε λειτουργεί σωστά και δίνει εσφαλμένα αποτελέσματα. Πιο συγκεκριμένα παρατηρούμε τα εξής:

- ✓ Για το πρόβλημα «αλλοιωμένη ταμπέλα» υπάρχουν 4 διαφορετικές περιπτώσεις σε μικρή απόσταση μεταξύ τους όπως φαίνεται στον αρχικό χάρτη της επιλογής των 10 σημείων με αριθμό 1. Οι τρεις περιπτώσεις βρίσκονται σε σχετικά μικρότερη απόσταση μεταξύ τους με αποτέλεσμα να επικρατεί σύγχυση των σημείων των καταγραφών και ο διαχωρισμός σε ομάδες να είναι μη ρεαλιστικός. Η τετάρτη περίπτωση αν και βρίσκεται σε μεγαλύτερη απόσταση, επίσης δεν ομαδοποιείται σωστά παρόλο που για τη συγκεκριμένη περίπτωση έχουν γίνει μόλις δύο καταγραφές οι οποίες μάλιστα δε συγχέονται με τις υπόλοιπες καταγραφές. Τέλος, στο χάρτη μοναδικών περιπτώσεων θα έπρεπε να εμφανίζονται τέσσερις περιπτώσεις για το πρόβλημα «αλλοιωμένη ταμπέλα» κάτι το οποίο δε συμβαίνει σε καμία από τις ομαδοποιήσεις.
- ✓ Για το πρόβλημα «λακκούβα» υπάρχουν 3 διαφορετικές περιπτώσεις επίσης σε μικρή απόσταση μεταξύ τους όπως φαίνεται στο χάρτη με αριθμό 2. Ωστόσο σε καμία από τις επτά ομαδοποιήσεις που πραγματοποιήθηκαν με διαφορετική απόσταση

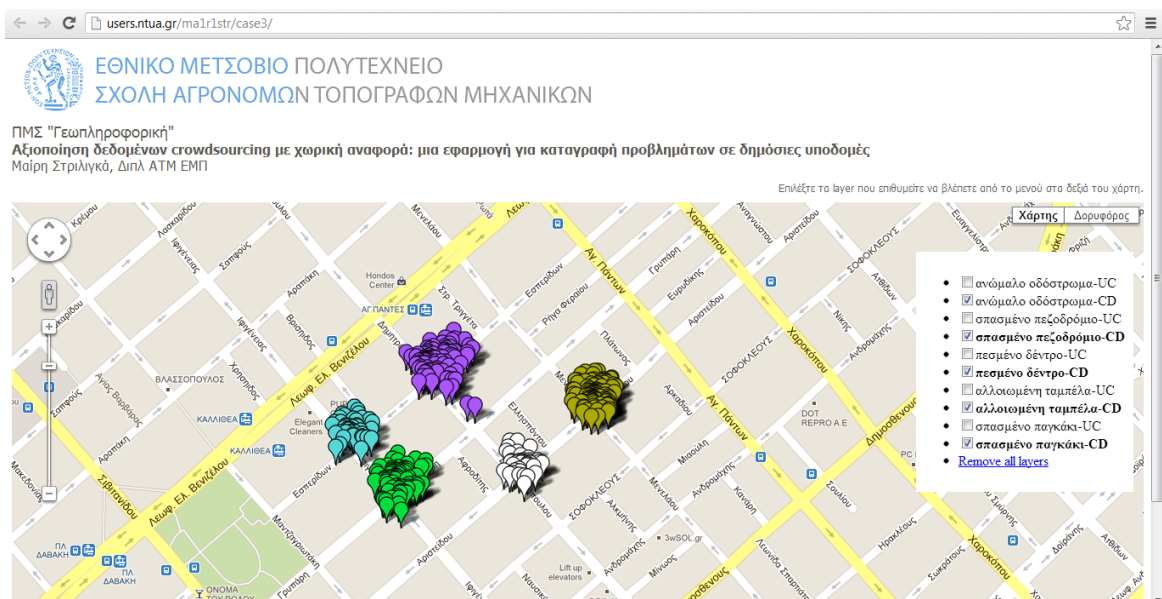
ομαδοποίησης δεν ήταν ικανοποιητικά τα αποτελέσματα. Στο χάρτη μοναδικών περιπτώσεων αντίστοιχα θα έπρεπε να εμφανίζονται τρεις περιπτώσεις για το πρόβλημα «λακούβα» κάτι το οποίο δε συμβαίνει, αλλά αντιθέτως είτε προκύπτουν παραπάνω ομάδες είτε πολύ λιγότερες από τις πραγματικές.

- ✓ Οι υπόλοιπες τρεις περιπτώσεις προβλημάτων με αριθμό 2,3 και 4 στο χάρτη, ομαδοποιήθηκαν σωστά καθώς βρίσκονται σε εύλογη μεταξύ τους απόσταση και ανήκουν σε διαφορετικούς τύπους προβλημάτων.

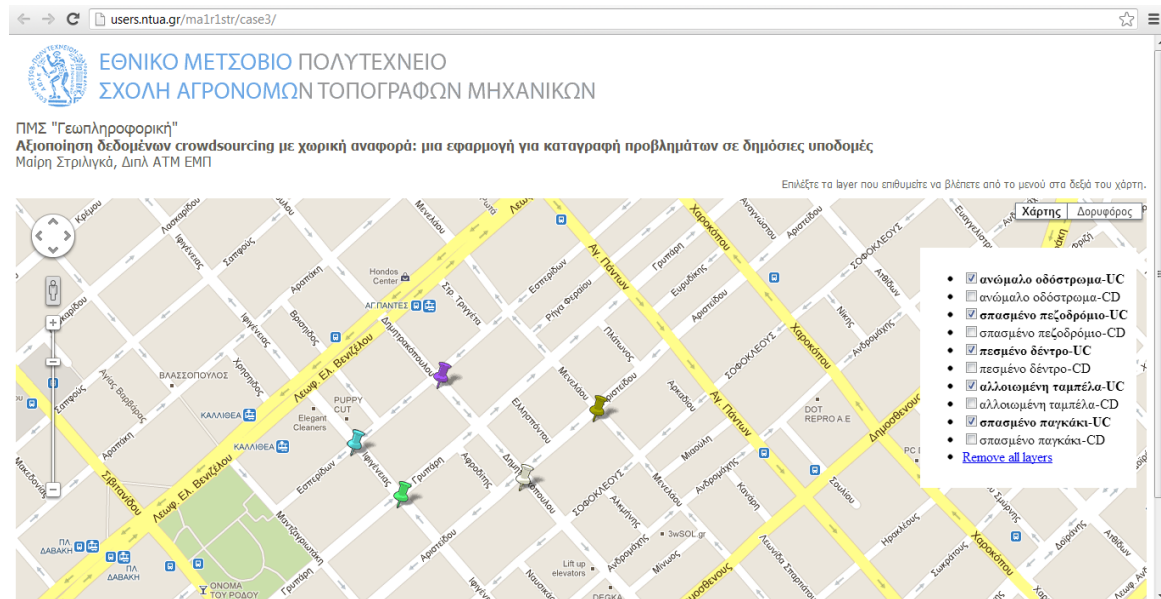


Εικόνα 6.56: Περιπτώσεις προβλημάτων με όμοιο τύπο προβλήματος ανά αριθμό

Στη συνέχεια απεικονίζονται τα αποτελέσματα της υλοποίησης για την περίπτωση που μελετάμε σε χάρτες στην προσωπική ιστοσελίδα που δημιουργήθηκε για το σκοπό αυτό. Παρουσιάζονται ενδεικτικά αποτελέσματα για απόσταση ομαδοποίησης 20 μέτρα.



Εικόνα 6.57: Απεικόνιση του συνόλου των μοναδικών περιπτώσεων ανά τύπο προβλήματος για απόσταση ομαδοποίησης 20 m σε προσωπική ιστοσελίδα για την «Περίπτωση 3»



Εικόνα 6.58: Απεικόνιση του συνόλου των μοναδικών περιπτώσεων ανά τύπο προβλήματος για απόσταση ομαδοποίησης 20 m σε προσωπική ιστοσελίδα για την «Περίπτωση 3»

7

ΣΥΜΠΕΡΑΣΜΑΤΑ ΚΑΙ ΠΡΟΤΑΣΕΙΣ

Κατά τη διάρκεια της εκπόνησης της μεταπτυχιακής εργασίας, συναντήθηκαν αρκετές δυσκολίες, κάποιες εκ των οποίων αντιμετωπίστηκαν αποτελεσματικά ενώ άλλες έθεσαν περιορισμούς στην υλοποίηση της εφαρμογής για την καταγραφή των προβλημάτων σε σχέση με προβλήματα σε δημόσιες υποδομές. Τα προβλήματα και οι περιορισμοί, καθώς επίσης και τα αποτελέσματα που παρήχθησαν διαμορφώνουν μία σειρά συμπερασμάτων αλλά και προτάσεων για περαιτέρω βελτίωση.

7.1 Συμπεράσματα

Μέσα από την εκπόνηση της μεταπτυχιακής εργασίας διαμορφώθηκε ένα σύνολο συμπερασμάτων που αφορά τόσο στις τεχνικές που χρησιμοποιήθηκαν όσο και στην εφαρμογή αυτή καθαυτή και τα αποτελέσματα από την υλοποίησή της. Τα συμπεράσματα που εξήχθησαν περιγράφονται στη συνέχεια:

- Το crowdsourcing αποτελεί μία αναπτυσσόμενη τεχνολογία η οποία μπορεί να εφαρμοστεί σε μια ευρεία ποικιλία προβλημάτων και εγείρει πολλές ενδιαφέρουσες τεχνικές και κοινωνικές προκλήσεις. Το crowdsourcing εάν αξιοποιηθεί σωστά μπορεί να προσφέρει γνώση και λύσεις στα προβλήματα του κοινωνικού συνόλου. Η εφαρμογή που υλοποιήθηκε στην παρούσα εργασία θα μπορούσε να αποτελέσει ένα εργαλείο διαχείρισης προβλημάτων για την ίδια την κοινωνία και τις δημόσιες αρχές.
- Όσοι περισσότεροι χρήστες συμμετέχουν στο crowdsourcing της καταγραφής προβλημάτων, τόσο πιο αξιόπιστα θα είναι τα αποτελέσματα. Η πιθανότητα κακόβουλης καταγραφής ενός προβλήματος μειώνεται όταν για το συγκεκριμένο πρόβλημα έχει πραγματοποιηθεί μεγάλος αριθμός καταγραφών από διαφορετικά άτομα.
- Ωστόσο η χρήση του crowdsourcing θα πρέπει να γίνεται πάντα με προσοχή, διότι δεν υπάρχει πραγματικός έλεγχος του πλήθους που συμμετέχει στο crowdsourcing και ελλοχεύει ο κίνδυνος της αλλοίωσης της εγκυρότητας των αποτελεσμάτων για μεμονωμένες περιπτώσεις καταγραφών προβλημάτων.
- Δίνοντας τα κατάλληλα κίνητρα στο πλήθος, αυξάνεται η συμμετοχικότητα των πολιτών στην επίλυση των προβλημάτων του τόπου τους και κατά συνέπεια συμβάλλουν στην αναβάθμιση του επιπέδου διαβίωσης. Με έξυπνη και συνετή χρήση του crowdsourcing μπορεί να δοθεί λύση και σε πολλά άλλα προβλήματα που αφορούν μία κοινωνία σε συλλογικό αλλά και σε ατομικό επίπεδο.

- Ο αλγόριθμος που χρησιμοποιήθηκε για την αρχική ομαδοποίηση των δεδομένων των καταγραφών σε μοναδικές περιπτώσεις με τη δημιουργία πολυγώνων, παρατηρήθηκε ότι δε λειτουργεί σωστά στην περίπτωση που τα σημεία βρίσκονται σε πολύ μικρή απόσταση μεταξύ τους, με αποτέλεσμα να απαιτείται η εύρεση εναλλακτικής λύσης όπως η αναζήτηση κάποιου άλλου αλγορίθμου ομαδοποίησης που θα ταιριάζει καλύτερα στο δεδομένα των προβληματικών περιπτώσεων και η αντικατάσταση του αλγορίθμου αυτού με μία εναλλακτική μέθοδο αναζήτησης των διανυσμάτων αρχικοποίησης που απαιτούνται για την υλοποίηση του αλγορίθμου k-means.
- Η συλλογή των δεδομένων θα μπορούσε να πραγματοποιηθεί και με άλλες μεθόδους, ωστόσο το αντικείμενο της εργασίας δεν ήταν η δημιουργία μίας εφαρμογής συλλογής δεδομένων αλλά η ανάδειξη τρόπων αξιοποίησης του crowdsourcing και η ένταξή του σε μία εφαρμογή με χωρική αναφορά.
- Οι χάρτες που προέκυψαν επιβεβαιώνουν ότι η υλοποίηση της εφαρμογής είναι έγκυρη και αξιόπιστη με εξαίρεση την περίπτωση γειννίασης των δεδομένων όπως ήδη αναφέρθηκε.

7.2 Προτάσεις

Η εφαρμογή που παρουσιάζεται στην παρούσα εργασία αποτελεί ένα ενδεικτικό παράδειγμα αξιοποίησης μίας crowdsourcing εφαρμογής για την ανάδειξη προβλημάτων σε δημόσιες υποδομές στα πλαίσια του μεταπτυχιακού προγράμματος και επιδέχεται επιπλέον τροποποιήσεις για την προσαρμογή της στις απαιτήσεις του εκάστοτε διαχειριστή της εφαρμογής και τη βελτιστοποίηση της αποδοτικότητας και της εικόνας του συστήματος. Οι προτάσεις για βελτίωση και περεταίρω αξιοποίηση της εφαρμογής παρουσιάζονται τη συνέχεια:

- Η καταγραφή των προβλημάτων πραγματοποιήθηκε με τη βοήθεια της crowdsourcing εφαρμογής Fulcrum App. Χρησιμοποιήθηκαν έτοιμα εργαλεία της εφαρμογής για τη δημιουργία της φόρμας καταγραφής προβλημάτων. Μία πρόταση για βελτίωση της εφαρμογής είναι η υλοποίηση εξ αρχής και εξ ολοκλήρου μίας crowdsourcing εφαρμογής προσαρμοσμένης αποκλειστικά στις δικές μας απαιτήσεις για την καταγραφή των προβλημάτων. Με αυτόν τον τρόπο η συλλογή των δεδομένων των καταγραφών δε θα εξαρτάται από τους περιορισμούς που θέτει μία τυποποιημένη έτοιμη εφαρμογή.
- Όσον αφορά στα δεδομένα που καταγράφονται από το πλήθος στην παρούσα εφαρμογή, ο χρήστης έχει τη δυνατότητα να επιλέξει μεταξύ ενός περιορισμένου αριθμού προκαθορισμένων ενδεικτικών προβλημάτων. Η ιδανική λύση θα ήταν να μπορεί το πλήθος να καταγράψει χωρίς περιορισμούς οποιοδήποτε πρόβλημα συναντάει σε δημόσιες υποδομές και στη συνέχεια να πραγματοποιείται τυποποίηση των δεδομένων και αρχειοθέτηση στη βάση δεδομένων.
- Επιπλέον προτείνεται ακόμα μεγαλύτερη αυτοματοποίηση στα στάδια υλοποίησης της εφαρμογής. Για παράδειγμα, η λήψη των δεδομένων των καταγραφών από την εφαρμογή Fulcrum θα μπορούσε να γίνεται αυτόματα και όχι χειροκίνητα όπως και έγινε στις περιπτώσεις που μελετήθηκαν.

- Απαραίτητη για τη βελτίωση της απόδοσης της εφαρμογής είναι επίσης η εύρεση λύσης για την ομαδοποίηση δεδομένων που βρίσκονται πολύ κοντά μεταξύ τους έτσι ώστε να μη συγχέονται οι καταγραφές διαφορετικών περιπτώσεων προβλημάτων όμοιου τύπου. Θα πρέπει να αναζητηθούν άλλες μέθοδοι εύρεσης των διανυσμάτων αρχικοποίησης για την ομαδοποίηση με τον αλγόριθμο k-means, οι οποίες θα προσαρμόζονται τοπικά καλύτερα στα δεδομένα, καθώς η ομαδοποίηση με τη δημιουργία πολυγώνων με βάση την απόσταση μεταξύ των σημείων δε δίνει ικανοποιητικά αποτελέσματα, με αποτέλεσμα εκ των προτέρων να γνωρίζουμε ότι κατ' επέκταση ούτε ο αλγόριθμος k-means θα δώσει την αναμενόμενη ομαδοποίηση.
- Επιπλέον μία αξιόλογη ιδέα θα ήταν η αξιοποίηση της δυνατότητας του Google Earth Live Streaming της εφαρμογής Fulcrum για τη δημιουργία δυναμικών χαρτών οι οποίοι θα ενημερώνονται ανά τακτά χρονικά διαστήματα με νέα δεδομένα καταγραφών.
- Τέλος η εφαρμογή αυτή μπορεί να αξιοποιηθεί πρακτικά από τις τοπικές αρχές για την ανάδειξη των προβλημάτων που αφορούν σε δημόσιες υποδομές. Οι δημόσιες αρχές θα εξοικονομήσουν χρόνο και χρήμα, καθώς οι καταγραφές θα πραγματοποιούνται από τους πολίτες. Έτσι οι αρχές θα έχουν τον πλήρη έλεγχο των προβλημάτων της πόλης με αποτέλεσμα την καλύτερη ιεράρχηση και διαχείριση των προβλημάτων αυτών, ενώ παράλληλα θα κινητοποιηθεί ο πολίτης και θα συμβάλλει δραστικά και άμεσα στην αναβάθμιση του βιοτικού επιπέδου της πόλης.

ΒΙΒΛΙΟΓΡΑΦΙΑ

- [1] Jeff Howe (2006), *The Rise of Crowdsourcing*, Wired Magazine.
- [2] Andrew Hudson-Smith, Michael Batty Andrew Crooks, and Richard Milton (2008), Mapping for the Masses: Accessing Web 2.0 through Crowdsourcing, *UCL Working Papers Series*, Centre for Advanced Spatial Analysis, University College London, UK.
- [3] AnHai Doan, Raghu Ramakrishnan, and Alon Halevy (2011), Crowdsourcing Systems on the World-Wide Web, *Communications of the acm*, pp.86-96.
- [4] Brabham, Daren (2008). Crowdsourcing as a Model for Problem Solving: An Introduction and Cases, *Convergence: The International Journal of Research into New Media Technologies*, London, Los Angeles, New Delhi and Singapore vol. 14 (1), pp.75–90
- [5] David Arthur, Sergei Vassilvitskii (2006). How Slow is the k-Means Method, Sedona, Arizona, USA, pp. 144-153.
- [6] Tapas Kanungo, David M. Mount, Nathan S. Netanyahu, Christine D. Piatko, Ruth Silverman, and Angela Y. Wu (2002). An Efficient k-Means Clustering Algorithm: Analysis and Implementation, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp.881-892.
- [7] Tapas Kanungo, David M. Mount, Ruth Silverman, Nathan S. Netanyahu, Angela Y. Wu, Christine Piatko (2000). The Analysis of a Simple k-Means Clustering Algorithm, *Computational Geometry*, Hong Kong China, pp.100-109.
- [8] Hamerly, G. and Elkan, C. (2002). Alternatives to the k-means algorithm that find better clusterings. *Proceedings of the eleventh international conference on Information and knowledge management (CIKM)*.
- [9] The PostgreSQL Global Development Group (2013). Appendix D. SQL Conformance, *PostgreSQL 9.1.8 Documentation*.
- [10] L. Rowe and M. Stonebraker (1987). *The POSTGRES data model*, Proc. VLDB Conference.
- [11] M. Stonebraker and L. Rowe (1986). *The design of POSTGRES*, Proc. ACM-SIGMOD Conference on Management of Data.
- [12] Steve Suehring (2010) . *JavaScript: Step by Step*, Second Edition, USA.
- [13] John Pollock (2010). *JavaScript: A beginner's guide*, Third Edition, USA.
- [14] the PHP Documentation Group (2013), Introduction: What can PHP do, *PHP Manual*.

- [15] Ullman, Larry (2005), *Εισαγωγή στην PHP για τον παγκόσμιο ιστό*, εκδόσεις Κλειδάριθμος.
- [16] Melonie, Julie C (2008), *Μάθετε PHP, MySQL και Apache*, εκδόσεις Γκιούρδας Μ.
- [17] Josie, Wernecke (2007), *The KML Handbook*, εκδόσεις Pearson Education (US)
- [18] Gray, James (2008), Getting Started With Quantum GIS, *Linux Journal*.

URL's

- [1] Crowdsourcing, *Wikipedia*.
<<http://en.wikipedia.org/wiki/Crowdsourcing>>
- [2] List of crowdsourcing projects, *Wikipedia*.
<http://en.wikipedia.org/wiki/List_of_crowdsourcing_projects>
- [3] What is Crowdsourcing, *Daily Crowdsourc*e.
<<http://dailycrowdsourc.com/crowdsourcing-basics/what-is-crowdsourcing>>
- [4] The Appeal of Crowdsourcing, *Crowdsourcing Org the industry website*.
<<http://www.crowdsourcing.org/document/the-appeal-of-crowdsourcing/7208>>
- [5] Spatial Clustering with PostGIS, *Geographic Information Systems*.
<<http://gis.stackexchange.com/questions/11567/spatial-clustering-with-postgis>>
- [6] Kmeans 1.0.0, *PostgreSQL Extension Network*.
<<http://pgxn.org/dist/kmeans/doc/kmeans.html>>
- [7] Eight Principles to successful Crowdsourcing, *Chaordix*.
<<http://www.chaordix.com/2009/08/eight-principles-to-successful-crowdsourcing>>
- [8] Stefan Lindegaard, 40 Examples of open Innovation and Crowdsourcing, *15INNO*.
<<http://www.15inno.com/2012/08/09/oicrowdexamples>>
- [9] Unsupervised classification algorithms, *Yale University*.
<http://www.yale.edu/ceo/Projects/swap/landcover/Unsupervised_classification.htm>
- [10] Fuzzy C-Means Clustering, *A Tutorial on Clustering Algorithms*.
<http://home.dei.polimi.it/matteucc/Clustering/tutorial_html/cmeans.html>
- [11] Fuzzy C-means Clustering algorithm, *Data Clustering Algorithms*.
<<https://sites.google.com/site/dataclusteringalgorithms/fuzzy-c-means-clustering-algorithm>>
- [12] SAS/STAT® 9.2 User's Guide, Second Edition.
<http://support.sas.com/documentation/cdl/en/statug/63033/HTML/default/viewer.htm#statug_fastclus_sect002.htm>

- [13] Clustering as a Mixture of Gaussians, *A Tutorial on Clustering Algorithms*.
<http://home.dei.polimi.it/matteucc/Clustering/tutorial_html/mixture.html>
- [14] K-means++, *Wikipedia*.
<<http://en.wikipedia.org/wiki/K-means%2B%2B>>
- [15] Renato Cordeiro de Amorim, Minkowski Weighted K-Means, *Data Mining Research*.
<<http://renatocamorim.com/2012/11/21/minkowski-weighted-k-means>>
- [16] *PHP*
<<http://www.php.net>>
- [17] HTML/CSS, JavaScript, SQL, PHP, *W3schools*.
<<http://www.w3schools.com>>
- [18] Keyhole Markup Language, *Wikipedia*.
<http://en.wikipedia.org/wiki/Keyhole_Markup_Language>
- [19] David Sklar and Adam Trachtenberg, What is PHP, *TechRepublic*
<<http://www.techrepublic.com/article/what-is-php/5074693>>
- [20] What is PHP, *Software Projects*
<<http://www.softwareprojects.org/php-what-is-01.htm>>
- [21] What is PHP, *wiseGEEK*.
<<http://www.wisegeek.com/what-is-php.htm>>
- [22] What PHP can do for you, *Peachpit*.
<<http://www.peachpit.com/articles/article.aspx?p=22290>>