

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ
ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ
ΠΡΟΓΡΑΜΜΑ ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ
«ΠΡΟΗΓΜΕΝΑ ΣΥΣΤΗΜΑΤΑ ΠΛΗΡΟΦΟΡΙΚΗΣ»

**Hermes.chorochronos.org: Μια Διαδικτυακή Πύλη για την
Ανάλυση Τροχιών Κινούμενων Αντικειμένων**

**Hermes.chorochronos.org: A Web Portal for the Analysis of
Moving Object Trajectories**

Μεταπτυχιακή Διατριβή

Καλλιτζάκη Σωτηρία

e-mail: skallitzaki@gmail.com

Επιβλέπων: Ν. Πελέκης, Λέκτορας

ΠΕΙΡΑΙΑΣ, ΟΚΤΩΒΡΙΟΣ 2014

Τριμελής Εξεταστική Επιτροπή

(υπογραφή)

(υπογραφή)

(υπογραφή)

Νικόλαος Πελέκης

Ιωάννης Θεοδορίδης

Άγγελος Πικράκης

Λέκτορας

Καθηγητής

Λέκτορας

Πρόλογος

Η ακόλουθη διατριβή πραγματοποιήθηκε στο πλαίσιο της εκπλήρωσης των υποχρεώσεων του Μεταπτυχιακού Προγράμματος «Προηγμένα Συστήματα Πληροφορικής» του Πανεπιστημίου Πειραιώς. Η μελέτη του αντικειμένου πραγματοποιήθηκε υπό την επίβλεψη του Λέκτορα κου Πελέκη.

Πειραιάς, Οκτώβριος 2014

Πανεπιστήμιο Πειραιώς

Ευχαριστίες

Ευχαριστώ τον επιβλέπων Λέκτορα κ. Πελέκη για την καθοδήγηση τους σχετικά με τον θέμα αυτής της εργασίας καθώς και τον Καθηγητή κο Θεοδωρίδη για τις παρατηρήσεις του στην υλοποίηση της εφαρμογής. Επιπλέον ευχαριστώ θερμά τον υποψήφιο Διδάκτορα Μάριο Βόντα για την υποστήριξη του σε όλα τα στάδια της διατριβής και την βοήθεια του σε τεχνικά όσο και θεωρητικά ζητήματα.

Πανεπιστήμιο Πειραιώς

Πίνακας Περιεχομένων

Περιεχόμενα εικόνων	7
1 Εισαγωγή	11
1.1 Υπόβαθρο και κίνητρο	11
2 Σχετικές Εργασίες	13
2.1 Πύλες Γεωγραφικών Δεδομένων (GeoPortals)	13
2.2 M-Atlas	15
Δυνατότητες M-Atlas.	15
Αρχιτεκτονική.	16
2.3 Move Mine	17
Αρχιτεκτονική συστήματος Move Mine.	18
2.4 Move Mine 2.0	19
Σχέση έλξης-αποφυγής.	19
Σχέση ακολούθησης.	19
Αρχιτεκτονική Συστήματος.	19
2.5 ChoroChronos	20
Αρχιτεκτονική.	21
2.6 HermesMOD (Βάση δεδομένων κινούμενων αντικειμένων Ερμής)	23
Σχήμα Βάσης Δεδομένων.	23
2.7 Μέθοδοι εξόρυξης γνώσης για τροχιές κινούμενων αντικειμένων	25
Βήματα διαδικασίας συσταδοποίησης.	25
Αλγόριθμοι συσταδοποίησης.	26
2.8 Μέτρα Ταξινόμησης Χρονοσειρών	31
Μετρικές Συναρτήσεις Απόστασης.	31
Συναρτήσεις ομοιότητας βασισμένες στην Ευκλείδεια απόσταση.	35
DISSIM.	35
Δυναμική Στρέβλωση Χρόνου (Dynamic Time Warping - DTW).	36
Μεγαλύτερη Κοινή Υπο-ακολουθία (LCSS).	37
Απόσταση Επεξεργασίας σε Πραγματικές Ακολουθίες (EDR).	38
Απόσταση Επεξεργασίας με Πραγματική Ποινή (ERP)	40
3 Hermes.ChoroChronos.org	42

3.1	Αρχιτεκτονική.....	42
3.2	Ανέβασμα αρχείων δεδομένων	43
	Βασικές διαδικασίες.....	44
3.3	Δυναμικά ερωτήματα με παραμέτρους εύρος τιμών.....	47
	Χωροχρονικά Κριτήρια (Spatio-Temporal)	47
	Χρονικά Κριτήρια (Temporal).....	49
	Χωρικά Κριτήρια (Spatial).....	50
	Συνάρτηση εκτέλεσης ερωτήματος στη βάση.....	51
3.4	Ερώτημα με παράμετρο συγκεκριμένη χρονική στιγμή.....	52
	Συνάρτηση εκτέλεσης ερωτήματος στη βάση.....	53
3.5	Συσταδοποίηση.....	54
	Μέτρα Ταξινόμησης Χρονοσειρών.....	54
	Βήματα εξαγωγής συστάδων.....	55
3.6	Κοινές διαδικασίες.....	59
	Οπτικοποίηση δεδομένων κίνησης.....	59
4	Μελλοντική επέκταση.....	59
4.1	Ανακατασκευή και μετάπτωση δεδομένων	60
	Εντοπισμός τροχιών.....	60
	Τμηματοποίηση τροχιάς.....	60
4.2	Συσχέτιση δεδομένων με σημασιολογική πληροφορία.....	60
	Περιοχές Ενδιαφέροντος.....	61
	Γραμμές Ενδιαφέροντος.....	61
	Σημεία Ενδιαφέροντος.....	61
5	Συμπεράσματα.....	62
6	Βιβλιογραφία.....	63
7	Συντομογραφίες.....	65

Περιεχόμενα εικόνων

Εικόνα 1. Υπηρεσίες Πύλης Χωρικών Δεδομένων.....	14
Εικόνα 2. Πύλη CrawDad	14
Εικόνα 3. Κατάλογος γεωγραφικών δεδομένων geocat.ch	15
Εικόνα 4. Αρχιτεκτονική M-Atlas Engine	16
Εικόνα 5. Αρχιτεκτονική συστήματος Move Mine	18
Εικόνα 6. Αρχιτεκτονική Συστήματος Move Mine 2.0[21]	20
Εικόνα 7. Αρχιτεκτονική Συστήματος ChoroChronos.org.....	22
Εικόνα 8. Βήματα διαδικασίας συσταδοποίησης	26
Εικόνα 9. Αλγόριθμος K-means.....	27
Εικόνα 10. Δείγμα που έχει τρέξει με τον OPTICS με αποτέλεσμα το διάγραμμα εγγύτητας[18]	29
Εικόνα 11. Διάγραμμα γειτνίασης.....	31
Εικόνα 12. Παράδειγμα απόστασης Μανχάταν	33
Εικόνα 13. Παράδειγμα υπολογισμού Ευκλείδειας απόστασης.....	34
Εικόνα 14. Παράδειγμα υπολογισμού απόστασης	35
Εικόνα 15. Ευκλείδεια απόσταση και τεχνική DTW	36
Εικόνα 16. Παράδειγμα μεθόδου LCSS [15].....	38
Εικόνα 17. Αρχιτεκτονική hermes.chorochronos.org	43
Εικόνα 18. Επιλογή ανεβάσματος αρχείου	43
Εικόνα 19. Οθόνη ανεβάσματος αρχείου	44
Εικόνα 20: Πίνακες παραγόμενοι από σύνολο δεδομένων	45
Εικόνα 21: Πίνακας συνόλων δεδομένων.....	45
Εικόνα 22: Πίνακας με ορισμένες στήλες από στατιστικά συνόλου δεδομένων	46
Εικόνα 23: Συσχέτιση ομάδων χρηστών με σύνολα δεδομένων.....	47
Εικόνα 24. Παράδειγμα δυναμικού ερωτήματος με χωροχρονικά κριτήρια	48
Εικόνα 25. Παράδειγμα αποτελέσματος δυναμικού ερωτήματος	48
Εικόνα 26. Παράδειγμα οπτικοποίησης αποτελέσματος δυναμικού ερωτήματος με χωροχρονικά κριτήρια	49
Εικόνα 27. Παράδειγμα δυναμικού ερωτήματος με χωρικά κριτήρια	49
Εικόνα 28. Παράδειγμα οπτικοποίησης αποτελέσματος δυναμικών ερωτημάτων με χωρικά κριτήρια.....	50
Εικόνα 29. Παράδειγμα δυναμικών ερωτημάτων με χρονικά κριτήρια	50
Εικόνα 30. Παράδειγμα οπτικοποίησης αποτελέσματος δυναμικών ερωτημάτων με χρονικά κριτήρια.....	51
Εικόνα 31. Παράδειγμα δυναμικού ερωτήματος με παράμετρο συγκεκριμένη χρονική στιγμή.....	52
Εικόνα 32. Παράδειγμα αποτελέσματος δυναμικού ερωτήματος με παράμετρο συγκεκριμένη χρονική στιγμή	53
Εικόνα 33. Οπτικοποίηση παραδείγματος δυναμικού ερωτήματος με παράμετρο συγκεκριμένη χρονική στιγμή	53
Εικόνα 34. Διάγραμμα γειτνίασης.....	57
Εικόνα 35. Παράμετρος ξ	58
Εικόνα 36. Οπτικοποίηση αποτελέσματος συσταδοποίησης	59

Πανεπιστήμιο Πειραιώς

Περίληψη

Στη διπλωματική εργασία γίνεται παρουσίαση των επεκτάσεων της πλατφόρμας επεξεργασίας χωροχρονικών δεδομένων Hermes.ChoroChronos.org. Οι επεκτάσεις αφορούν στην υποβοήθηση τεχνικών ανάλυσης δεδομένων κίνησης, όπως αυτά περιγράφονται από τις τροχιές των κινούμενων αντικειμένων. Αναλυτικότερα, στο **κεφάλαιο 2** παρουσιάζονται η πύλη χωροχρονικών δεδομένων ChoroChronos.org, μέρος του οποίου αποτελεί το Hermes.ChoroChronos.org, η βάση δεδομένων κινούμενων αντικειμένων ΕΡΜΗΣ και οι δυνατότητες της, συστήματα παρόμοια του Hermes.ChoroChronos.org που υπάρχουν, μέτρα ταξινόμησης χρονοσειρών και αλγορίθμους συσταδοποίησης τροχιών κινούμενων αντικειμένων. Στο **κεφάλαιο 3** παρουσιάζονται οι επεκτάσεις που έγιναν στο σύστημα Hermes.ChoroChronos.org. Συγκεκριμένα αναφέρονται οι αρχιτεκτονική του συστήματος, η δυνατότητα ανεβάσματος και οπτικοποίησης αρχείου δεδομένων τροχιών κινούμενων αντικειμένων, η δυνατότητα υλοποίησης δυναμικών ερωτημάτων στις τροχιές κινούμενων αντικειμένων καθώς και την δυνατότητα συσταδοποίησης των τροχιών κινούμενων αντικειμένων. Στο **κεφάλαιο 4** αναφέρονται μελλοντικές επεκτάσεις που θα μπορούσαν να προστεθούν στο ChoroChronos.org

Abstract

The object of this thesis is the extensions of spatiotemporal data processing platform Hermes.ChoroChronos.org. Extensions relate to assist technical analysis of traffic data, such as those described by trajectories of moving objects. Specifically, **Chapter 2** presents the gateway spatiotemporal data ChoroChronos.org, part of which is the Hermes.ChoroChronos.org, the moving object database HERMES and its capabilities, existing systems similar to Hermes.ChoroChronos.org, measures for time series classification and algorithms for moving object trajectories clustering. **Chapter 3** presents the extensions of the system Hermes.ChoroChronos.org. More specifically the architecture of the system, the selection of uploading dataset and visualizing of trajectory moving objects datasets, the ability of dynamic queries in trajectories of moving objects and the possibility of trajectories moving objects clustering. **Chapter 4** provides future extensions that could be added at ChoroChronos.org

1 Εισαγωγή

Η υπάρχουσα γεωγραφική πύλη δεδομένων ChoroChronos.org έχει ως σκοπό να υποστηρίξει την επιστημονική κοινότητα για την ανταλλαγή συνόλου χωροχρονικών δεδομένων και αλγορίθμων. Από την εμπειρία που αποκτήθηκε μέχρι τώρα από το ChoroChronos.org φάνηκε ότι μία σημαντική έλλειψη που υπάρχει από τη παρεχόμενη λειτουργικότητα του είναι η δυνατότητα εκτέλεσης δυναμικών ερωτημάτων στα δεδομένα που φιλοξενούνται στο portal (πχ εξαγωγή υποσυνόλου των δεδομένων, με βάση κάποια φίλτρα), καθώς και λειτουργιών ανάλυσης και εξόρυξης γνώσης από αυτά.

Στο πλαίσιο της συγκεκριμένης εργασίας επεκτάθηκαν οι υπάρχουσες λειτουργίες του geoportals choroChronos.org βελτιώνοντας τις τεχνικές οπτικοποίησης των δεδομένων, προσθέτοντας δυνατότητα εκτέλεσης δυναμικών ερωτημάτων και υποστηρίζοντας την συσταδοποίηση των δεδομένων με ποικιλία διαφορετικών κριτηρίων ομοιότητας. Για την αποτελεσματική υλοποίηση των ανωτέρω επιλέχθηκε το σύστημα MOD Hermes, το οποίο χρειάστηκε να τροποποιηθεί κατάλληλα προσθέτοντάς του λειτουργικότητα σε τομείς όπως μία βιβλιοθήκη συναρτήσεων αποστάσεων (similarity library) και αλγόριθμοι συσταδοποίησης τροχιών κινούμενων αντικειμένων.

Η καινοτόμος λειτουργικότητα που θα προσφέρει το ChoroChronos.org στη κοινότητα του θα επιτρέπει στους χρήστες από οποιοδήποτε υπολογιστή έχει πρόσβαση στο διαδίκτυο να δημιουργούν λογαριασμό στην εφαρμογή ChoroChronos.org και να χρησιμοποιούν τις επιλογές δυναμικής επεξεργασίας του συνόλου κινούμενων αντικειμένων που τους ενδιαφέρει χωρίς να χρειάζεται κάποια εγκατάσταση εφαρμογής ή βάσης τοπικά στον υπολογιστή.

1.1 Υπόβαθρο και κίνητρο

Η συλλογή γεωγραφικών και χωροχρονικών δεδομένων συνεχώς αναπτύσσεται λόγω της καθημερινής συλλογής δεδομένων από συσκευές κινητών τηλεφώνων και άλλες συσκευές εντοπισμού θέσης. Αυτά τα δεδομένα προκύπτουν από την αναγκαιότητα ανίχνευσης τέτοιων ασύρματων, κινητών συσκευών προκειμένου να υποστηριχθεί η αλληλεπίδραση με την υποδομή του δικτύου. Η πληροφορία που συλλέγεται από αυτές τις συσκευές δεν έχει καμία χρησιμότητα χωρίς την χρήση εργαλείων ανάλυσης που τα αποκρυπτογραφεί, προκειμένου να τους αποδοθεί νόημα. Συνεπώς η αναγκαιότητα μελέτης αυτών των δεδομένων και εύρεσης τρόπων εξαγωγής χρήσιμης πληροφορίας, έρχεται στο προσκήνιο.

Σε αυτό το σημείο προκύπτει το ερώτημα. Τι είδους ανάλυση μπορεί να υλοποιηθεί σε γεωγραφικά και χρονικά δεδομένα; Πώς θα αναπαρασταθούν στο χρήστη που θέλει να τα αναλύσει; Τι θα τον ενδιέφερε και πώς θα μπορούσε εύκολα να κάνει αυτή την ανάλυση; Πώς μπορούν να βρεθούν πρότυπα συμπεριφοράς γι' αυτά τα δεδομένα; Προκειμένου να καλυφθούν ορισμένα σημεία από τις ανωτέρω ανάγκες έχουν δημιουργηθεί πολλές πύλες χωρικών δεδομένων, όπου μπορούν να αποθηκευθούν σύνολα δεδομένων ή εργαλεία για οπτικοποίηση δεδομένων ή/και βελτιστοποίησης δεδομένων, ή αλγόριθμοι επεξεργασίας διαφόρων ειδών συνόλων δεδομένων. Παρόλο που μπορεί να βρεθεί ένα πλήθος εργαλείων, ο χρήστης υποχρεούται να συλλέξει

και να συγκρίνει τα διάφορα εργαλεία μεταξύ τους ώστε να εξάγει οποιοδήποτε είδος πληροφορίας, καθώς δεν υπάρχει διαδικτυακή πλατφόρμα που να παρέχει οπτικοποίηση και ανάλυση χωροχρονικών δεδομένων με εργαλεία και αλγορίθμους επεξεργασίας διαδικτυακά. Η πύλη χωροχρονικών δεδομένων ChoroChronos.org κατάλληλη για δεδομένα και διεργασίες κίνησης ήρθε για να συμπληρώσει αυτό το κενό. Σε αυτή τη πύλη αυτή τη στιγμή οι εγγεγραμμένοι χρήστες μπορούν να δωρίσουν χωρικά και χωροχρονικά δεδομένα καθώς και αλγορίθμους κατάλληλους για την επεξεργασία τους. Ο στόχος αυτής της εργασίας είναι η συμβολή στην εξέλιξη αυτής της πύλης σε διαδικτυακό σημείο υπηρεσιών επεξεργασίας χωροχρονικών δεδομένων. Στο ChoroChronos.org προστέθηκαν οι δυνατότητες εκτέλεσης ερωτημάτων δυναμικά πάνω στα σύνολα δεδομένων. Επιπλέον προστέθηκε και η δυνατότητα πιο σύνθετων λειτουργιών, όπως είναι η συσταδοποίηση.

Η διαδικασία της συσταδοποίησης απαιτεί υψηλή υπολογιστική ισχύ καθώς ο χρόνος που απαιτείται για την εκτέλεση του αλγορίθμου εξαρτάται κατά βάση από τον όγκο των δεδομένων. Γι' αυτό το λόγο η εκτέλεση της διαδικασίας συσταδοποίησης υλοποιείται από τη μηχανή βάσης χωροχρονικών δεδομένων, τον Hermes, καθώς λόγω της ευρετηρίασης που χρησιμοποιεί μπορεί να εκτελέσει σε πολύ μικρότερο υπολογιστικό χρόνο τα ερωτήματα που υλοποιούνται.

2 Σχετικές Εργασίες

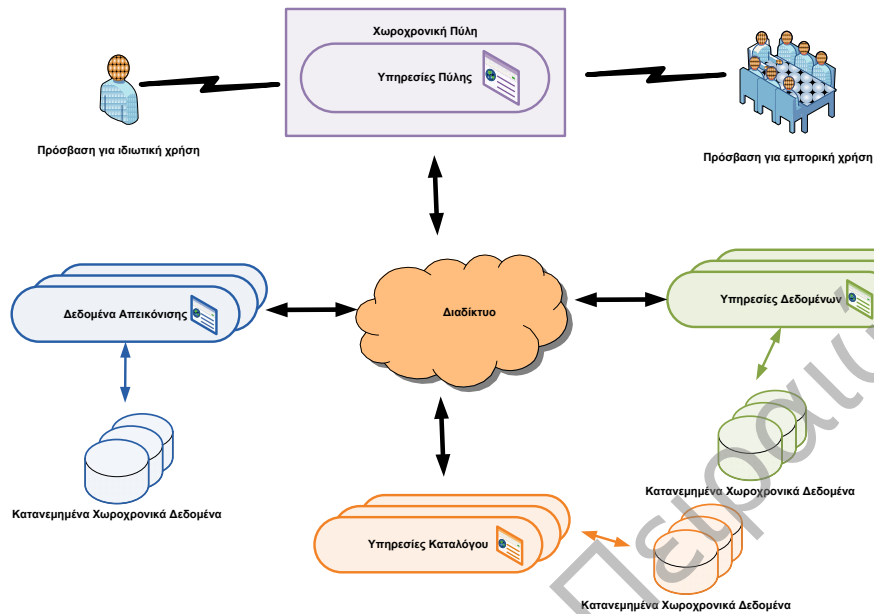
2.1 Πύλες Γεωγραφικών Δεδομένων (GeoPortals)

Στην εργασία του ο Tait (2005) [2] προσδιόρισε μία πύλη γεωγραφικών δεδομένων (geoportals) ως «ένα δικτυακό ιστότοπο που αποτελεί πύλη εισόδου σε γεωγραφικό περιεχόμενο που υπάρχει στο διαδίκτυο, ή πιο απλά, ένα δικτυακό ιστότοπο όπου μπορεί να βρεθεί γεωγραφικό περιεχόμενο». Οι πύλες γεωγραφικών δεδομένων κατηγοριοποιούνται με διάφορους τρόπους. Ενώ οι Maguire and Longley(2005) [3] χωρίζουν τις πύλες γεωγραφικών δεδομένων σε δύο κατηγορίες: πύλες με υπηρεσίες καταλόγου γεωγραφικών δεδομένων και πύλες με εφαρμογές για γεωγραφικά δεδομένα, οι Tang and Selwood (2005) διαχωρίζουν τις πύλες γεωγραφικών δεδομένων σε 3 κατηγορίες: πύλες υπηρεσίας καταλόγου, πύλες με εφαρμογές και επιχειρησιακές πύλες.

Οι υπηρεσίες που υποστηρίζει μία πύλη χωρικών δεδομένων, σύμφωνα με την Ανοιχτή Γεωχωρική Κοινοπραξία (Open Geospatial consortium -OGC) είναι οι ακόλουθες:

1. Υπηρεσίες Πύλης: Παροχή ενός μοναδικού σημείου προσπέλασης της πληροφορίας των χωρικών δεδομένων στην πύλη καθώς και υπηρεσίες διαχείρισης του περιεχομένου και των χρηστών της πύλης.
2. Υπηρεσίες Καταλόγου: Παρέχουν τον εντοπισμό χωρικών υπηρεσιών και πληροφοριών όπου και βρίσκονται και πληροφορίες σχετικά με τις υπηρεσίες και τις πληροφορίες που εντοπίζει ο χρήστης.
3. Υπηρεσίες Απεικόνισης: Χρησιμοποιούνται για την επεξεργασία της χωρικής πληροφορίας και την προετοιμασία της για παρουσίαση.
4. Υπηρεσίες Δεδομένων: Χρησιμοποιούνται για την παροχή χωρικού περιεχομένου και την επεξεργασία δεδομένων. [1]

Η αρχιτεκτονική μίας πύλης γεωγραφικών δεδομένων παρουσιάζεται στην εικόνα που ακολουθεί.



Εικόνα 1. Υπηρεσίες Πύλης Χωρικών Δεδομένων

Οι πύλες γεωγραφικών δεδομένων συνήθως περιλαμβάνουν σύνολα δεδομένων διαφορετικών τύπων, από γεωγραφικά δεδομένα θέσης έως δεδομένα κίνησης βακτηριδίων. Επίσης προσφέρουν πρόσβαση σε εργαλεία επεξεργασίας συνόλου δεδομένων που μπορούν να χρησιμοποιηθούν ανεξάρτητα από το portal, εφόσον ο χρήστης τα λάβει στον υπολογιστή του.

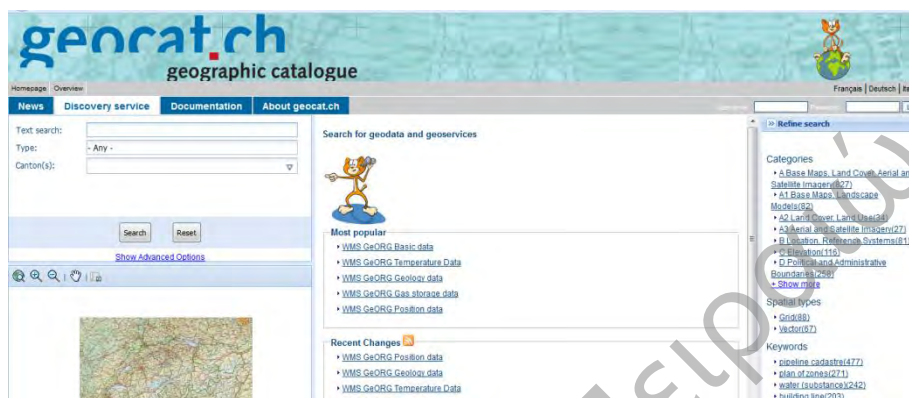
Ορισμένα χαρακτηριστικά παραδείγματα πύλων γεωγραφικών δεδομένων είναι τα <http://crawdad.cs.dartmouth.edu/> και <http://www.geocat.ch>.



Εικόνα 2. Πύλη CrawlDad

Το CRAWDAD αποτελεί μία κοινότητα πηγών για ταξινόμηση δεδομένων ασύρματων δικτύων στο Dartmouth. Στη πύλη αυτή έχει αναπτυχθεί μία εφαρμογή για τη ταξινόμηση των δεδομένων των ασύρματων δικτύων και εργαλείων για την συλλογή, ανωνυμοποίηση και ανάλυση των δεδομένων. Επίσης έχει συμπεριληφθεί μία δομή μεταδεδομένων για περαιτέρω πληροφορίες για κάθε σύνολο δεδομένων και κάθε εργαλείο που βρίσκεται σε αυτή τη πύλη. Τα μεταδεδομένα αποτελούνται από μία

περιγραφή των δεδομένων ή των εργαλείων, σχετικούς συγγραφείς και εργασίες καθώς και παλαιότερες εκδόσεις δεδομένων και εργαλείων. Η συνεισφορά των συνόλων δεδομένων προέρχεται από χρήστες του ιστοτόπου και η ομάδα του CrawDad τα ελέγχει και δημιουργεί μία περιγραφή των μεταδεδομένων.



Εικόνα 3. Κατάλογος γεωγραφικών δεδομένων geocat.ch

Το geocat.ch αποτελεί έναν κατάλογο γεωγραφικών δεδομένων και εργαλείων επεξεργασίας δεδομένων. Επίσης έχει συμπεριληφθεί μία δομή μεταδεδομένων για περαιτέρω πληροφορίες για κάθε σύνολο δεδομένων και κάθε εργαλείο που βρίσκεται σε αυτή τη πύλη. Τα δεδομένα και οι υπηρεσίες κατηγοριοποιούνται βάσει του είδους των δεδομένων που αντιπροσωπεύουν, με λέξεις κλειδιά που έχουν δοθεί, το τύπο των χωρικών δεδομένων (διανυσματικά ή σε grid), τον οργανισμό που τα έχει δωρίσει κ.α.

2.2 M-Atlas.

Το σύστημα M-Atlas, που αποτελεί συντομογραφία για το mobility atlas (άτλαντας κίνησης), είναι ένα σύστημα που μπορεί να χρησιμοποιηθεί για δημιουργία και περιήγηση σε έναν περιεκτικό κατάλογο συμπεριφορών κίνησης σε μία περιοχή. Ο βασικός σκοπός της εφαρμογής είναι η μετάβαση από τα δεδομένα που λαμβάνονται από τα GPS σε αναπαράσταση συμπεριφορών κίνησης.

Δυνατότητες M-Atlas.

Οι δυνατότητες της εφαρμογής M-Atlas είναι οι ακόλουθες:

1. Δυνατότητα οπτικοποίησης με ιστόγραμμα του συνόλου των κινούμενων οχημάτων κάθε ώρα της ημέρας.
2. Δυνατότητα χρωματικής αναπαράστασης στο χάρτη της πυκνότητας των κινήσεων για κάποια συγκεκριμένη ώρα. Οι συχνές περιοχές σημειώνονται με κόκκινο χρώμα, οι λιγότερο συχνά επισκεψιμες περιοχές με πράσινο και οι περιοχές που δεν έχουν επισκεψιμότητα, με βαθύ μπλε χρώμα.

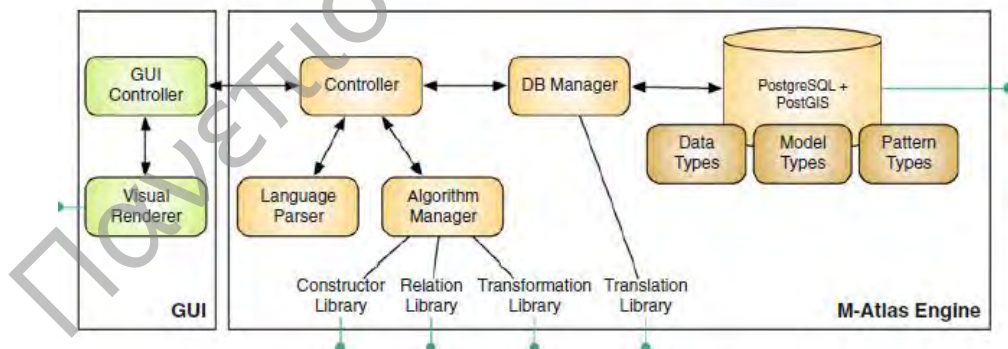
3. Δυνατότητα εξαγωγής στατιστικών στοιχείων με γραφικά μέσα για βασικά μεγέθη που περιγράφουν τα ταξίδια και αναπαριστώνται στα σύνολα δεδομένων των τροχιών κινούμενων αντικειμένων, όπως:
 - a. τα χιλιόμετρα των ταξιδιών,
 - b. τη διάρκεια των ταξιδιών,
 - c. το συσχετισμό του μήκους και της ταχύτητα των ταξιδιών,
 - d. τη γυροσκοπική ακτίνα του οχήματος (η μέση απόσταση ενός οχήματος από τη πιο πιθανή τοποθεσία του),
 - e. την πυκνότητα κινούμενων και στάσιμων οχημάτων στο χώρο και το χρόνο.
4. Ανάλυση της συμπεριφοράς κίνησης μέσω τεχνικών εξόρυξης γνώσης. Για αυτή την λειτουργία έχει υλοποιηθεί ο αλγόριθμος συσταδοποίησης TimeFocused-Optics[18] που βασίζεται στη πυκνότητα και χρησιμοποιεί διάφορες συναρτήσεις απόστασης ικανές να ομαδοποιήσουν τις τροχιές με διάφορους τρόπους ώστε να ικανοποιούνται οι ανάγκες των χρηστών.

Αρχιτεκτονική.

Αρχιτεκτονικά το M-Atlas έχει τρεις υψηλού επιπέδου συνιστώσες:

1. Μία βάση δεδομένων για τα δεδομένα των τροχιών, τα μοντέλα και τα πρότυπα
2. Μία χωροχρονική γλώσσα ερωτημάτων για δεδομένα τροχιών, μοντέλα και πρότυπα
3. Ένα σύνολο κατασκευαστών χωροχρονικών μοντέλων και προτύπων.

Αναλυτικότερα η δομή του M-Atlas φαίνεται στο ακόλουθο διάγραμμα [27]:



Εικόνα 4. Αρχιτεκτονική M-Atlas Engine

Για κάθε ερώτημα που υποβάλλεται στο σύστημα ακολουθείται η εξής διαδικασία:

- Ένα ερώτημα υποβάλλεται μέσω της γραφικής διεπαφής στη μονάδα επεξεργασίας (Controller Module), η οποία συντονίζει τις εργασίες που γίνονται

από όλες τις υπόλοιπες μονάδες. Ο αναλυτής της γλώσσας (Language Parser) αναλύει το εισερχόμενο ερώτημα.

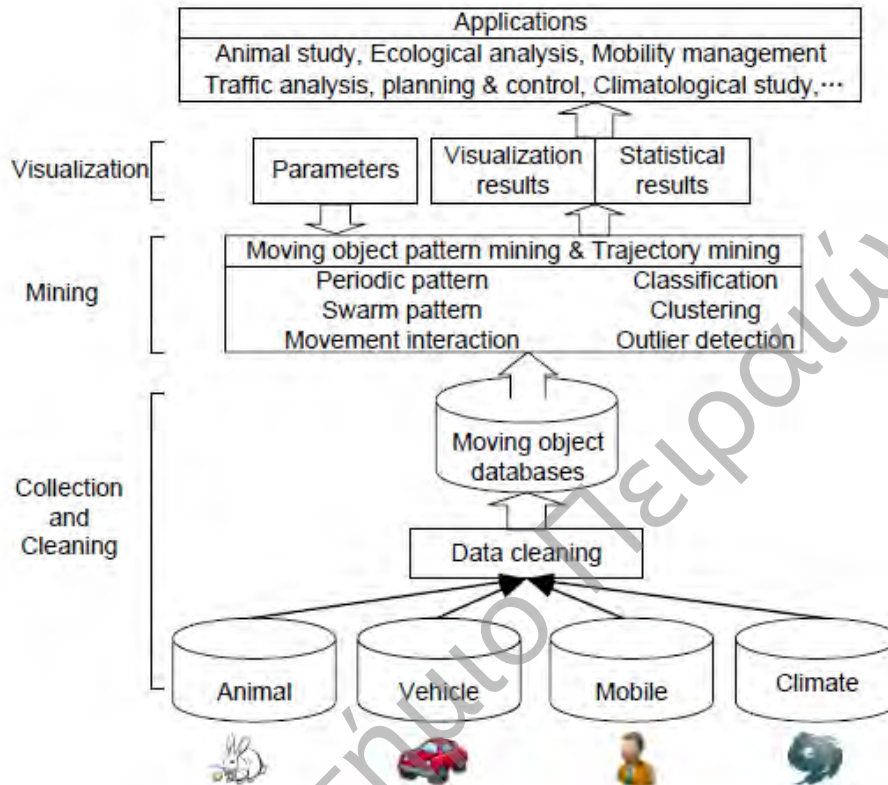
- Τα τυπικά ερωτήματα SQL αποστέλλονται άμεσα στο Διαχειριστή της βάση (Database Manager) και εκτελούνται από το σύστημα διαχείρισης της αντικειμενο-σχεσιακής βάσης δεδομένων.
- Τα υπόλοιπα ερωτήματα που γίνονται στη γλώσσα M-Atlas μεταφράζονται από τον αναλυτή γλώσσας σε ένα εκτελέσιμο πλάνο, που συνδυάζει ερωτήματα στη βάση δεδομένων και κλήσεις σε μεθόδους που παρέχονται από τον διαχειριστή αλγορίθμου (Algorithm Manager).
- Τα αποτελέσματα του ερωτήματος αποθηκεύονται στο σύστημα διαχείρισης της αντικειμενο-σχεσιακής βάσης δεδομένων και όταν ζητείται, προβάλλονται μέσω της μονάδας επεξεργασίας στην γραφική διεπαφή του χρήστη.

Οι ακίδες αναπαριστούν τα μοντέλα που μπορούν να επεκταθούν μέσω του συστήματος πρόσθετων μονάδων.

2.3 Move Mine

Το Move Mine έχει σχεδιαστεί για εξόρυξη δεδομένων από κινούμενα αντικείμενα. Ενσωματώνει πολλές διαδικασίες εξόρυξης δεδομένων, όπως πρότυπα εξόρυξης γνώσης από κινούμενα αντικείμενα και εξόρυξη γνώσης από τροχιές με σύγχρονες μεθόδους. Ενδεικτικά το Move Mine μπορεί να χρησιμοποιηθεί για την αυτόματη ανίχνευση της περιόδου κίνησης, για τον εντοπισμό προτύπων μαζικής κίνησης όπως τα κοπάδια και τα σμήνη, για την συσταδοποίηση, ταξινόμηση και ανίχνευση ακραίων τιμών των τροχιών προκειμένου να πραγματοποιηθεί ανάλυση των δεδομένων.

Αρχιτεκτονική συστήματος Move Mine.



Εικόνα 5. Αρχιτεκτονική συστήματος Move Mine

Η αρχιτεκτονική του συστήματος του Move Mine αποτελείται από τρία διακριτά επίπεδα:

- Συλλογή και καθαρισμός δεδομένων

Σε αυτό το επίπεδο γίνεται η συλλογή και η εκκαθάριση των δεδομένων των κινούμενων αντικειμένων.

- Εξόρυξη γνώσης

Ένα ευρύ φάσμα διαδικασιών εξόρυξης γνώσης από τα δεδομένα εφαρμόζεται στην βάση δεδομένων επιτρέποντας στους χρήστες να αναλύσουν τα δεδομένα από διάφορες πλευρές. Οι διαδικασίες αυτές συμπεριλαμβάνουν πρότυπα εντοπισμού περιόδου, πρότυπα εντοπισμού σμήνους, εντοπισμό αλληλεπίδρασης κίνησης, συσταδοποίηση τροχιών κινούμενων αντικειμένων, εντοπισμό ακραίων τιμών και ταξινόμηση.

- Οπτικοποίηση

Η οπτικοποίηση πραγματοποιείται σε επίπεδο δύο διαστάσεων ή ενσωματώνεται σε άλλα εργαλεία οπτικοποίησης, όπως είναι το Google Map. Επιπλέον όταν υπάρχει η δυνατότητα αναπαριστώνται οπτικά ορισμένα στατιστικά στοιχεία ώστε να παρέχετε στους χρήστες ευρύτερη γνώση σχετικά με τα αποτελέσματα.

2.4 Move Mine 2.0

Η εξέλιξη του Move Mine έχει ως σκοπό τον εντοπισμό σχέσεων των αντικειμένων από τα δεδομένα κίνησης. Αυτό υλοποιήθηκε προσθέτοντας νέες μεθόδους εξόρυξης μοτίβων δυναμικής συσχέτισης των δεδομένων. Οι μέθοδοι που προστέθηκαν επικεντρώνονται σε δύο είδη προτύπων συσχέτισης ανά ζεύγη: σχέσης έλξης-αποφυγής και σχέσης ακολούθησης.

Σχέση έλξης-αποφυγής

Στην βιβλιογραφία η μελέτη των σχέσεων των κινούμενων αντικειμένων έχει επικεντρωθεί κυρίως στις σχέσεις έλξης των αντικειμένων. Στο Move Mine συγκρίνονται δύο τροχιές κίνησης σε σχέση με την απόκλιση που έχει η πραγματική συχνότητα συνάντησης των τροχιών από την αναμενόμενη συχνότητα συνάντησης. Διαισθητικά εάν τα αντικείμενα συναντώνται περισσότερο ή λιγότερο από το αναμενόμενο, η σχέση είναι περισσότερο πιθανό να είναι έλξης ή αποφυγής αντίστοιχα.

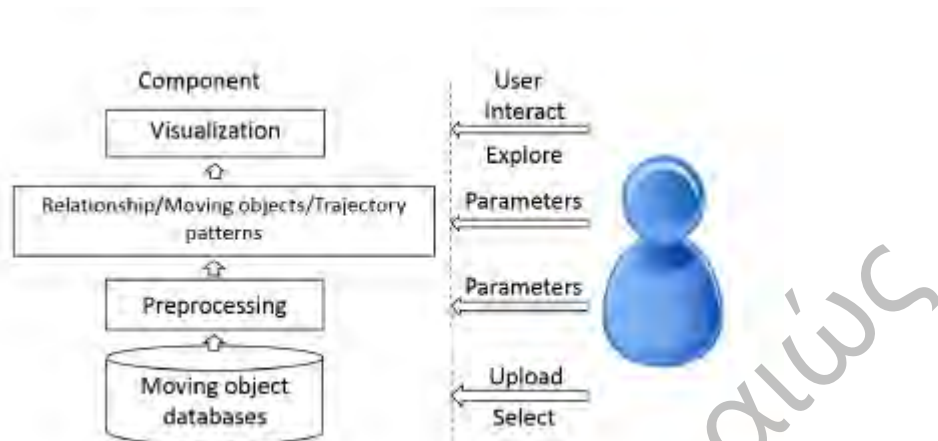
Σχέση ακολούθησης

Η σχέση που επιδιώκεται να αναπαρασταθεί είναι η σχέση που έχουν παραδείγματος χάριν ο αρχηγός μίας ομάδας με τα μέλη που τον ακολουθούν.

Τα ζητήματα που έχουν ληφθεί υπόψη στο Move Mine για τον εντοπισμό αυτής της σχέσης είναι:

- Το μέλος μπορεί να μην έχει ακριβώς την ίδια τροχιά με τον αρχηγό.
- Ο χρόνος από τον οποίο αποκλίνει το μέλος από τον αρχηγό είναι συνήθως άγνωστος και ποικίλει.
- Η σχέση μέλους –αρχηγού μπορεί να έχει μικρή διάρκεια.

Αρχιτεκτονική Συστήματος



Εικόνα 6. Αρχιτεκτονική Συστήματος Move Mine 2.0[21]

Η αρχιτεκτονική του Move Mine 2.0 είναι παρόμοια με αυτή του Move Mine. Πλέον τα επίπεδα είναι 4 μετά την προσθήκη του επιπέδου 1 και έχουν προστεθεί λειτουργικότητες στα υπάρχοντα επίπεδα.

Επίπεδο 1: Ο χρήστης μπορεί να κατεβάσει το σύνολο δεδομένων που τον ενδιαφέρει από τον εξυπηρετητή ή τοπικά από τον υπολογιστή του.

Επίπεδο 2: Ο χρήστης υλοποιεί προεπεξεργασία των δεδομένων χρησιμοποιώντας τις σχετικές παραμέτρους που του παρέχονται, όπως πχ το χρονικό διάστημα που τον ενδιαφέρει να μελετήσει.

Επίπεδο 3: Ο χρήστης επιλέγει τις διαδικασίες εξόρυξης γνώσης που επιθυμεί να εκτελεστούν.

Επίπεδο 4: Ο χρήστης μπορεί να οπτικοποιήσει τα δεδομένα του.

2.5 ChoroChronos

Η πύλη γεωγραφικών δεδομένων ChoroChronos.org παρουσιάστηκε το 2011 στην εργασία ChoroChronos.org: A GeoPortal for Movement Data and Processes [23].

Οι εργασίες που γίνονται σε αυτή είναι η αποθήκευση και η διαχείριση χωροχρονικού περιεχομένου. Συγκεκριμένα οι χρήστες μπορούν να δωρίσουν σύνολα χωροχρονικών δεδομένων ή αλγόριθμους επεξεργασίας τους. Για κάθε αρχείο που υπάρχει στο ChoroChronos.org υπάρχει και μία σειρά μεταδεδομένων που παρέχει τις πληροφορίες γι' αυτό, όπως τη πηγή απ' όπου λήφθηκε το αρχείο, μία περίληψη γι' αυτό, ένα σχετικό γράφημα, ημέρα εισαγωγής, συνολικές προβολές κα. Οι χρήστες του συστήματος για να μπορέσουν να δωρίσουν κάποιο αρχείο και να κάνουν πλήρη χρήση των δυνατοτήτων αυτής της πύλης πρέπει να έχουν κάνει εγγραφή. Για τις ανωτέρω εργασίες που αφορούν τα χωροχρονικά δεδομένα χρησιμοποιείται η μηχανή βάσης δεδομένων κινούμενων αντικειμένων Hermes (Ερμής).

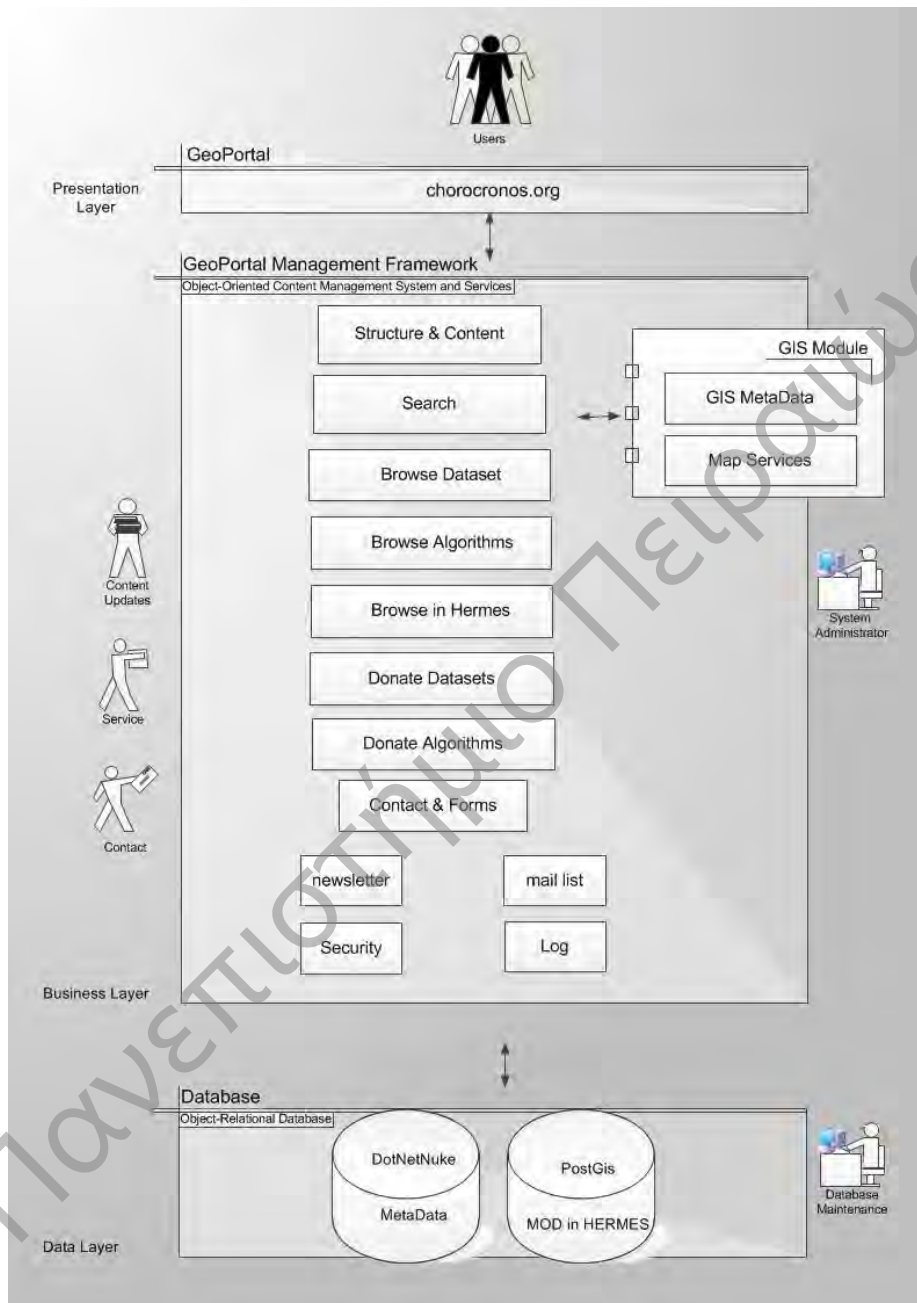
Αρχιτεκτονική

Για την υλοποίηση της εφαρμογής ώστε να πραγματοποιεί τις ανωτέρω εργασίες έχει χρησιμοποιηθεί αρχιτεκτονική τριών επιπέδων και την συνεργασία διαφορετικών υπολογιστικών συστημάτων όπως φαίνεται στην Εικόνα 7.

Στο **χαμηλότερο επίπεδο** (επίπεδο δεδομένων) βρίσκεται το σύνολο των δεδομένων και των διαδικασιών καθώς και τα κατάλληλα μεταδεδομένα. Για την αποθήκευση και την διαχείριση του χωροχρονικού περιεχομένου χρησιμοποιείται η βάση δεδομένων κινούμενων αντικειμένων Ερμής.

Στο **μεσαίο επίπεδο** (επίπεδο διαχείρισης) υλοποιείται η διαχείριση όλων των διαδικασιών στο στη γεωγραφική πύλη. Συγκεκριμένα υλοποιούνται οι διαδικασίες εισαγωγής, διαγραφής, ενημέρωσης και ανάκτησης σύμφωνα με τα κριτήρια επιλογής του συνόλου των δεδομένων και των διαδικασιών που βρίσκονται στον Ερμή.

Στο **υψηλό επίπεδο** (επίπεδο παρουσίασης) βρίσκεται η διεπαφή, μέσω της οποίας οι χρήστες μπορούν να προσπελάσουν τη πύλη και να καλέσουν τις διάφορες διαδικασίες.



Εικόνα 7. Αρχιτεκτονική Συστήματος ChoroChronos.org

2.6 HermesMOD (Βάση δεδομένων κινούμενων αντικειμένων Ερμής)

Το Hermes MOD αποτελεί ένα σύστημα διαχείρισης βάσης δεδομένων κινούμενων αντικειμένων (MOD). Ο Ερμής έχει υλοποιηθεί ως πρόσθετο σε δύο συστήματα διαχείρισης δεδομένων, της Oracle και τη PostgreSQL.

Στην βιβλιογραφία γίνεται αναφορά της αρχιτεκτονικής της υλοποίησης που έχει υλοποιηθεί στο Oracle DBMS [24][25] και ο Μάριος Βόντας στην μεταπτυχιακή του διατριβή [29] αναλύει την υλοποίηση που έχει γίνει στην PostgreSQL η οποία χρησιμοποιείται στο Hermes.ChoroChronos.org και η οποία αναλύεται στη συνέχεια.

Ο Ερμής βασίζεται στην υπάρχουσα λειτουργικότητα της PostgreSQL και την επεκτείνει ώστε να υποστηρίζει δεδομένα τροχιών. Στον πυρήνα του ΕΡΜΗΣ περιέχει ορισμένους τύπους δεδομένων χωρικούς και χρονικούς όπως επίσης και τύπους δεδομένων που αποτελούν ενοποίηση των χωρικών και χρονικών (πχ τύποι χωροχρονικών δεδομένων). Όλη η λειτουργικότητα του ΕΡΜΗΣ και τα χαρακτηριστικά του βασίζονται και χρησιμοποιούν το ανωτέρω μοντέλο. Στον ΕΡΜΗΣ μία τροχιά αποτελεί μία συνέχεια από δείγματα περιοχών με χρονική σήμανση (p_i , t_i) όπου p_i είναι ένα σημείο δύο διαστάσεων (x_i , y_i) και t_i είναι η καταγεγραμμένη χρονική στιγμή του σημείου p_i .

Ο ΕΡΜΗΣ παρέχει μία SQL διεπαφή αποτελούμενη από τύπους, συναρτήσεις και τελεστές που μπορεί να συνδυάσει ο χρήστης προκειμένου να κατασκευάσει δεδομένα και να υλοποιήσει υπολογισμούς πάνω σε αυτά. Αυτή η διεπαφή είναι προσβάσιμη από μία σειρά πρωτοκόλλων όπως τα JDBC, ODBC και πρακτικά από οποιοδήποτε πρωτόκολλο έχει τη δυνατότητα σύνδεσης με έναν τυπικό εξυπηρετητή PostgreSQL. Γι' αυτούς τους λόγους ο ΕΡΜΗΣ μπορεί να χρησιμοποιηθεί για το σχεδιασμό και την υλοποίηση μίας βάσης δεδομένων με σκοπό την διαχείριση χωροχρονικών δεδομένων καθώς και για την υλοποίηση οποιασδήποτε εφαρμογής για την διαχείριση των χωροχρονικών δεδομένων.

Σχήμα Βάσης Δεδομένων.

Η δομή της βάσης δεδομένων έχει σχεδιαστεί ώστε να μπορεί να φιλοξενήσει πολλαπλά σύνολα δεδομένων. Έχοντας ορισμένα τους τύπους δεδομένων ο βασικό χαρακτηριστικό που λείπει είναι οι μέθοδοι για την αποθήκευση των δομημένων δεδομένων στη βάση. Αυτό λύθηκε αναπτύσσοντας μία υποδομή μεταδεδομένων, ουσιαστικά ενός καταλόγου, που θα μας βοηθήσει στην φιλοξενία πολλαπλών συνόλων δεδομένων με διαφορετικά χαρακτηριστικά.

Αυτή η υποδομή μεταδεδομένων παίρνει τη μορφή πίνακα με το όνομα dataset. Κάθε σύνολο δεδομένων σε αυτό το πίνακα έχει ένα μοναδικό αναγνωριστικό και ένα μοναδικό μικρό όνομα "short name" πχ [1, "imis"] ή [2, "milan"]. Τα περισσότερα χωροχρονικά σύνολα δεδομένων είναι της μορφής [objectID, trajectoryID, t, lon, lat] όπου "objectID" είναι το αναγνωριστικό του αντικειμένου, "trajectoryID" είναι το αναγνωριστικό της τροχιάς του αντικειμένου, "t" είναι η ώρα στο σύστημα UTC (Co-

ordinated Universal Time) στην οποία έγινε η καταγραφή και “lon” και “lat” είναι οι μοίρες του γεωγραφικού πλάτους και γεωγραφικού μήκους στο WGS 84 Γεωγραφικό Σύστημα Συντεταγμένων. Τα “objectID” και “TrajectoryID” συνδιάζονται για να σχηματιστεί το μοναδικό αναγνωριστικό κάθε τροχιάς σε συγκεκριμένο σύνολο δεδομένων. [29]

Κατάλογος Μεταδεδομένων.

Ο πίνακας dataset που αναφέρθηκε προηγουμένως είναι ο κατάλογος μεταδεδομένων και κάθε γραμμή του αντιστοιχίζεται με ένα σύνολο δεδομένων. Η δομή του είναι η ακόλουθη:

- **id** είναι μία στήλη με μοναδικό αύξοντα ακέραιο αριθμό που αποτελεί το βασικό κλειδί του πίνακα. Κάθε σύνολο δεδομένων που φιλοξενείται του δίδεται ένα id.
- **name** είναι μία στήλη κειμένου που περιέχει ένα μοναδικό μικρό όνομα για κάθε σύνολο δεδομένων.
- **name_long** είναι μία στήλη κειμένου που περιέχει ένα όνομα φιλικό προς το χρήστη για κάθε σύνολο δεδομένων.
- **parent_dataset** είναι ένα ξένο κλειδί σε μία υπάρχουσα γραμμή στο πίνακα όπου, όταν δεν είναι NULL, υποδεικνύει μία σχέση γονιού-παιδιού μεταξύ των συνόλων δεδομένων.
- **parent_dataset_notes** είναι μία στήλη κειμένου που περιέχει σημειώσεις για την σχέση πατέρα-γονιού του συνόλου.
- **local_ref_poi** είναι μία στήλη PointLL που αποτελεί ένα σημείο αναφοράς για τη μετατροπή των συντεταγμένων.
- **SRID** είναι μία στήλη ακέραιου αριθμού που περιέχει το κωδικό EPSG του προβαλλόμενου συστήματος αναφοράς στο οποίο αποθηκεύεται το σύνολο δεδομένων στον EPMH.
- Επιπλέον υπάρχουν ορισμένα στατιστικά στοιχεία που αποθηκεύονται σε αυτό το πίνακα.
 - Όρια του συνόλου δεδομένων (tmin, tmax, lx, ly, hx, hy, llon, llat, hlon, hlat).
 - Μεσαίες τιμές του συνόλου δεδομένων (Centroid x, centroid y, centroid lon, centroid lat).
 - Πλήθος αντικειμένων/ τροχιών / σημείων / τμημάτων.
 - ελάχιστος/μέσος/ μέγιστος αριθμός σημείων ανά τροχιά/ διάρκεια τροχιάς / μήκος τροχιάς.
- **notes** είναι μία στήλη κειμένου που περιέχει σημειώσεις για το σύνολο δεδομένων. [29]

2.7 Μέθοδοι εξόρυξης γνώσης για τροχιές κινούμενων αντικειμένων

Ο μεγάλος όγκος δεδομένων των τροχιών κινούμενων αντικειμένων χρήζει τμηματοποίησης ώστε να εξαχθεί πληροφορία που θα μπορεί να οδηγήσει σε συγκεκριμένα συμπεράσματα. Μία βασική μέθοδος εξόρυξης γνώσης είναι η συσταδοποίηση. Το πρόβλημα της συσταδοποίησης σχετίζεται με την τμηματοποίηση ενός συνόλου δεδομένων σε συστάδες έτσι ώστε τα στοιχεία του συνόλου των δεδομένων που ανήκουν σε μία συστάδα να είναι περισσότερο όμοια μεταξύ τους από ότι είναι με τα στοιχεία οποιασδήποτε άλλης συστάδας.

Οι αλγόριθμοι συσταδοποίησης καλούνται να αντιμετωπίσουν ένα πλήθος ζητημάτων, ορισμένα από τα οποία είναι:

- **Ακραία σημεία:** σε κάθε πραγματικό σύνολο δεδομένων υπάρχουν τιμές που δεν ανταποκρίνονται σε υπαρκτά σημεία. Αυτό μπορεί να προκληθεί από παρεμβολές του σήματος ή
- **Άγνωστο πλήθος συστάδων:** δεν είναι εκ των προτέρων γνωστό το πλήθος των συστάδων που είναι επιθυμητό να σχηματιστεί, συνεπώς ο αλγόριθμος πρέπει να έχει τη δυνατότητα να εντοπίζει το σύνολο των συστάδων.
- **Δυναμικά Μεταβαλλόμενα Δεδομένα:** να γίνεται αλλαγή συστάδων με στην πορεία του χρόνου.
- **Κλιμάκωση:** να μπορεί να γίνει κλιμάκωση στον αριθμό των σημείων και των διαστάσεων των αντικειμένων.
- **Ερμηνεία και Αξιολόγηση των αποτελεσμάτων:** με πιο τρόπο θα ερμηνευτούν και θα αξιολογηθούν τα αποτελέσματα κάθε αλγορίθμου.

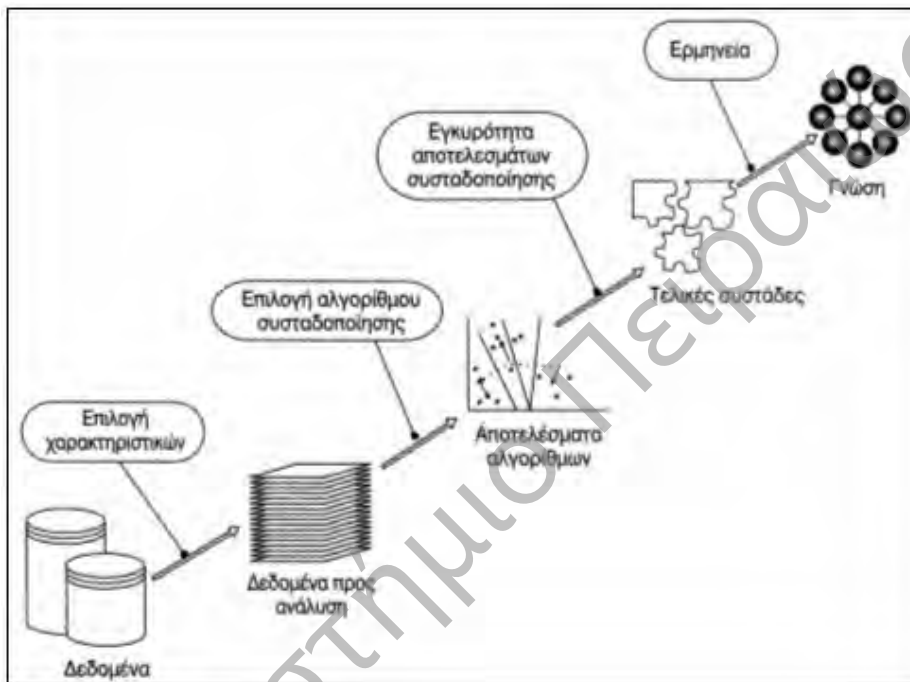
Βήματα διαδικασίας συσταδοποίησης.

Για την συσταδοποίηση δεδομένων η διαδικασία που ακολουθείται είναι η ακόλουθη:

- **Επιλογή χαρακτηριστικών γνωρισμάτων:** Επιλέγονται τα πιο αντιπροσωπευτικά γνωρίσματα στα οποία πρόκειται να εφαρμοστεί η συσταδοποίηση ώστε να επιτευχθεί η βέλτιστη ομοιογένεια σε κάθε συστάδα. Γι' αυτό το λόγο η προεπεξεργασία των δεδομένων πριν την εφαρμογή της διαδικασίας συσταδοποίησης είναι απαραίτητη.
- **Επιλογή αλγορίθμου συσταδοποίησης:** Σε αυτό το στάδιο γίνεται η επιλογή ενός αλγορίθμου που θα οδηγήσει σε ένα καλό σχήμα συσταδοποίησης για ένα σύνολο δεδομένων. Για τη επιλογή του αλγορίθμου χρησιμοποιείται το μέτρο γειννίας και το κριτήριο συσταδοποίησης τα οποία ορίζουν απόλυτα τον αλγόριθμο, καθώς επίσης και η δυνατότητά του να καθορίσει ένα σχήμα συσταδοποίησης που να προσαρμόζεται στο συγκεκριμένο σύνολο δεδομένων.
- **Επικύρωση αποτελεσμάτων:** σε αυτό το στάδιο γίνεται αξιολόγηση των αποτελεσμάτων του αλγορίθμου συσταδοποίησης σύμφωνα με κατάλληλα κριτήρια ορθότητας συσταδοποίησης και τεχνικές. Παράδειγμα ενός τέτοιου κριτηρίου είναι η σύγκριση των αποτελεσμάτων της ανάλυσης με κάποια ήδη γνωστά αποτελέσματα ή η σύγκριση των αποτελεσμάτων δύο διαφορετικών συσταδοποι-

ήσεων. Η ποιότητα της συσταδοποίησης εξαρτάται από την ομοιότητα και την μέθοδο υλοποίησης της συσταδοποίησης.

- **Ερμηνεία και παρουσίαση των αποτελεσμάτων:** Αποτελεί το τελευταίο στάδιο της διαδικασίας συσταδοποίησης, όπου εξάγεται γνώση από τις παραχθείσες συστάδες.



Εικόνα 8. Βήματα διαδικασίας συσταδοποίησης

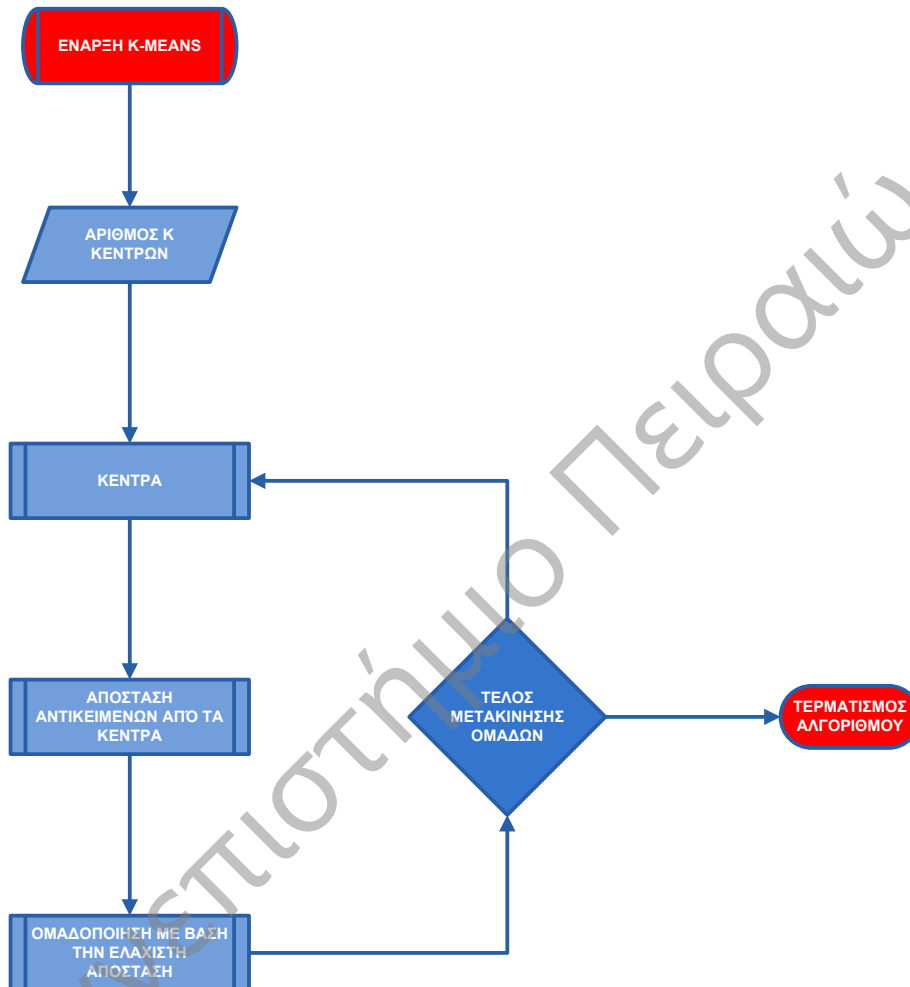
Αλγόριθμοι συσταδοποίησης.

Υπάρχει ένα μεγάλο πλήθος από αλγόριθμους συσταδοποίησης που έχουν προταθεί στη διεθνή βιβλιογραφία, και ο καθένας τους βασίζεται σε διαφορετική φιλοσοφία. Σύμφωνα με την βασική φιλοσοφία που ακολουθούν χωρίζονται σε διάφορες κατηγορίες συσταδοποίησης. Δύο πολύ σημαντικές κατηγορίες συσταδοποίησης που θα αναλύσουμε στη συνέχεια είναι η συσταδοποίηση βασισμένη στο κέντρο βάρους (Centroid-based) και η συσταδοποίηση βασισμένη στη πυκνότητα (Density – based).

Συσταδοποίηση βασισμένη στο κέντρο βάρους.

Ένας από τους πιο γνωστούς αλγόριθμους συσταδοποίησης βασισμένους στο κέντρο βάρους είναι ο K-means. Ο K-means βασίζεται στην διάσπαση. Χρησιμοποιώντας μία παράμετρο K διαιρεί n στοιχεία σε K συστάδες με μικρή ομοιότητα μεταξύ των συστάδων και ελαχιστοποιεί την συνολική απόσταση της κάθε συστάδας από το κέντρο της. Ο υπολογισμός ομοιότητας υλοποιείται από τη μέση τιμή του συνόλου

κάθε συστάδας. Το μέτρο ομοιότητας που χρησιμοποιεί ο αλγόριθμος είναι η Ευκλείδεια απόσταση μεταξύ των αντικειμένων. [19]



Εικόνα 9. Αλγόριθμος K-means

Το βασικό μειονέκτημα του αλγορίθμου συσταδοποίησης K-means είναι η αναγκαιότητα της τιμής K. Δεν υπάρχει κάποιο εφαρμόσιμο στοιχείο για την λήψη της απόφασης του πλήθους των συστάδων που θα παραχθούν. Ο K-means είναι ευαίσθητος στην αρχική τιμή και για διαφορετικές αρχικές τιμές μπορεί να παραχθούν διαφορετικές συστάδες.

Συσταδοποίηση βασισμένη στη πυκνότητα .

Αυτού του είδους οι αλγόριθμοι είναι ανθεκτικοί σε προβλήματα όπως ο θόρυβος και οι ακραίες τιμές, καθώς δεν επηρεάζουν αισθητά τη συνολική διανομή της πυκνότητας των δεδομένων και μπορούν να αναγνωρίσουν τις διαφορετικές συχνότητες στο σύνολο των σημείων. Αυτή είναι μία πολύ σημαντική ιδιότητα για τα πραγματικά σύνολα δεδομένων, επειδή έχουν μη ταξινομημένη πυκνότητα στο εσωτερικό της δομής της συστάδας και ακραίες τιμές. Γι' αυτό το λόγο μεγάλο σύνολο τοπικών πυκνοτήτων μπορεί να χρειαστούν ώστε να αναγνωριστούν οι συστάδες σε διαφορετικές περιοχές αντί για μία γενική παράμετρο πυκνότητας.

Ένας από τους σημαντικότερους αλγόριθμους που βασίζονται στη πυκνότητα είναι ο Optics. Η βασική ιδέα του αλγορίθμου συσταδοποίησης Optics είναι ότι για κάθε στοιχείο μίας συστάδας η γειτονιά του δοσμένης μίας ακτίνας ϵ , πρέπει να περιέχει τουλάχιστον ένα ελάχιστο πλήθος αντικειμένων MinPts. Οι τιμές ϵ και MinPts αντιπροσωπεύουν τις ελάχιστες τιμές εισόδου του αλγορίθμου. Με αυτό τον τρόπο ο Optics δημιουργεί μία ταξινόμηση σημείων που σχετίζεται με τις τιμές γεινιάσης. Επιπλέον είναι αξιοσημείωτο ότι το αποτέλεσμα του Optics είναι ένα διάγραμμα γεινιάσης (reachability plot), που αποτελεί μία διαισθητική, ανεξάρτητη από τα δεδομένα οπτικοποίηση της δομής των συστάδων δεδομένων. Αυτό αποδίδει πρόσθετη πληροφορία για καλύτερη κατανόηση των δεδομένων και επίσης χρησιμοποιούνται για την εκχώρηση κάθε στοιχείου στην αντίστοιχη συστάδα ή στο θόρυβο, αντίστοιχα.

Ο αλγόριθμος συσταδοποίησης Optics συγκριτικά με τον K-means έχει ένα πολύ βασικό πλεονέκτημα. Βρίσκει δυναμικά το πλήθος των συστάδων που πρέπει να δημιουργηθούν, δεν απαιτεί τον ορισμό τους εκ των προτέρων, όπως ο αλγόριθμος k-means. Επίσης ο αλγόριθμος k-means απαιτεί όλα τα αντικείμενα να ανήκουν σε κάποια συγκεκριμένη συστάδα. Αυτό δημιουργεί σοβαρά προβλήματα στη χρήση του σε πραγματικά δεδομένα. Γι' αυτούς τους λόγους θεωρήθηκε καταλληλότερος για την ομαδοποίηση τροχιών κινούμενων αντικειμένων ο Optics και έγινε χρήση του σε αυτή τη πλατφόρμα. Συγκεκριμένα γίνεται χρήση μίας υλοποίησης του Optics αλγορίθμου και μίας αυτοματοποιημένης τεχνικής παραγωγής μίας ιεραρχικής δομής συσταδοποίησης από τον αλγόριθμο Optics, όπως αυτά προτάθηκαν στην εργασία των Ankerst et al. [5]

Αλγόριθμος Optics

Σε αυτό το κεφάλαιο παρουσιάζονται συνοπτικά οι ορισμοί στους οποίους βασίζεται η λειτουργική διαδικασία του αλγορίθμου Optics.

Απόσταση Πυρήνα (core-distance)

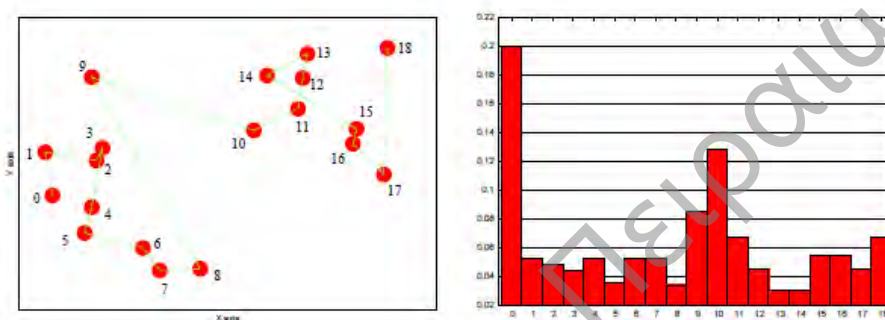
Έστω $p \in D$ ένα αντικείμενο σε ένα σύνολο δεδομένων D , ϵ μία τιμή απόστασης, $N_\epsilon(p)$ η ϵ -γειτονιά του p , MinPts ένας φυσικός αριθμός και MinPts-απόσταση (p) η απόσταση από το p στη MinPts-γειτονιά. Τότε η απόσταση πυρήνα του αντικειμένου p ορίζεται ως απόσταση πυρήνα $\alpha_{\epsilon, \text{MinPts}}(p) =$

$$\begin{cases} \text{απροσδιόριστο,} & \text{αν } N_\epsilon(p) < \text{MinPts} \\ \text{MinPts} - \text{απόσταση}(p), & \text{διαφορετικά} \end{cases}$$

Απόσταση γειτνίασης (Reachability-distance)

Έστω p και $o \in D$ αντικείμενα από ένα σύνολο δεδομένων D , $N_\varepsilon(o)$ η ε -γειτονιά του o και $MinPts$ ένας φυσικός αριθμός. Τότε η απόσταση γειτνίασης του p σε σχέση με το o ορίζεται ως απόσταση-γειτνίασης $_{\varepsilon, MinPts}(p, o) =$

$$\left\{ \begin{array}{l} \text{απροσδιόριστο, αν } |N_\varepsilon(o)| < MinPts \\ \max(\text{απόσταση} - \text{πυρήνα}(o), \text{απόσταση}(o, p)), \text{ διαφορετικά} \end{array} \right\}$$



Εικόνα 10. Δείγμα που έχει τρέξει με τον OPTICS με αποτέλεσμα το διάγραμμα εγγύτητας[18]

Η ταξινόμηση των συστάδων ενός συνόλου δεδομένων μπορεί να αναπαρασταθεί και να γίνει αντιληπτή γραφικά. Κατ' αρχή, μπορεί κάποιος να δει τη δομή συσταδοποίησης ενός συνόλου δεδομένων εάν οι τιμές της απόστασης εγγύτητας για κάθε αντικείμενο αναπαρασταθούν σε ταξινόμηση συστάδων.

Η απόσταση εγγύτητας του p σε σχέση με το o είναι η απόστασή τους. Όταν το p είναι πολύ κοντά στο o η απόσταση κανονικοποιείται σε μία κατάλληλη τιμή.

Τα βασικά βήματα του αλγορίθμου Optics είναι τα ακόλουθα:

- Αρχικά επιλέγεται ένα αντικείμενο p_0 με τυχαίο τρόπο.
- Στη συνέχεια για i επαναλήψεις, το επόμενο αντικείμενο p_i που επιλέγεται από το σύνολο D , είναι αυτό που έχει τη μικρότερη απόσταση γειτνίασης σε σχέση με όλα τα υπόλοιπα αντικείμενα του πυρήνα που έχουν ήδη ελεγχθεί.
- Η διαδικασία επαναλαμβάνεται έως ότου ελεγχθούν όλα τα αντικείμενα του συνόλου D .

Το σύνολο της διαδικασίας του αλγορίθμου μπορεί να αναπαρασταθεί σε ένα διάγραμμα γειτνίασης:

Ο οριζόντιος άξονας παρουσιάζει τα αντικείμενα ταξινομημένα με την σειρά που έγινε η επιλογή των στοιχείων $0, \dots, |D| - 1$, και στον οριζόντιο άξονα για κάθε i απόσταση γειτνίασης που αντιστοιχίζεται στο αντικείμενο p_i που αναπαριστάται διαγραμματικά. Η συχνότητα $(p_0, \dots, p_{|D|-1})$ καλείται και συσταδική αναταξινόμηση των αντικειμένων στο D .

Ο αλγόριθμος OPTICS παράγει την επαυξημένη ταξινόμηση της συστάδας λαμβάνοντας υπόψη την ταξινόμηση των σημείων, τις τιμές γειτνίασης και τις τιμές του πυρήνα, όπως αυτές έχουν περιγραφεί ανωτέρω. Στη συνέχεια παρουσιάζεται μία

τεχνική αυτοματοποιημένης ανάλυσης που χρησιμοποιεί μόνο την ταξινόμηση και τις τιμές γειτνίασης.

Τεχνικές αναγνώρισης της δομής της συστάδας.

Στην εφαρμογή χρησιμοποιείται μία αυτοματοποιημένη τεχνική για ανάλυση της ταξινόμησης των συστάδων και παραγωγή μίας ιεραρχικής δομής συσταδοποίησης. Κατά τη ταξινόμηση των σημείων, που δημιουργήθηκε από τον OPTICS, δημιουργούνται κοιλάδες που αναπαριστούν τις συστάδες και κορυφές που μπορεί να έχουν διάφορα ύψη. Για την ανίχνευση του σημείου αρχής και τέλους μίας συστάδας ορίζονται ορισμένες μεταβατικές περιοχές.

Ένα σημείο p καλείται ξ -απότομα προς τα πάνω εάν το επόμενο σημείο από αυτό είναι υψηλότερο από το p κατά ποσοστό $\xi\%$. Αντίστοιχα ένα σημείο p καλείται ξ -απότομα προς τα κάτω εάν το επόμενο σημείο από αυτό είναι χαμηλότερο από το p κατά ποσοστό $\xi\%$.

Μία περιοχή ονομάζεται ξ -απότομα ανοδική αν το πρώτο και το τελευταίο σημείο είναι ξ -απότομα προς τα πάνω και τα ενδιάμεσα σημεία έχουν το καθένα ύψος τουλάχιστον ίδιο με το προηγούμενο τους. Επίσης δεν πρέπει να περιέχει πάρα πολλά διαδοχικά σημεία που δεν είναι ανοδικά. Αντίστοιχα ορίζεται και η ξ -απότομα καθοδική περιοχή.

Πιο φορμαλιστικά, μία συστάδα ορίζεται ως εξής:

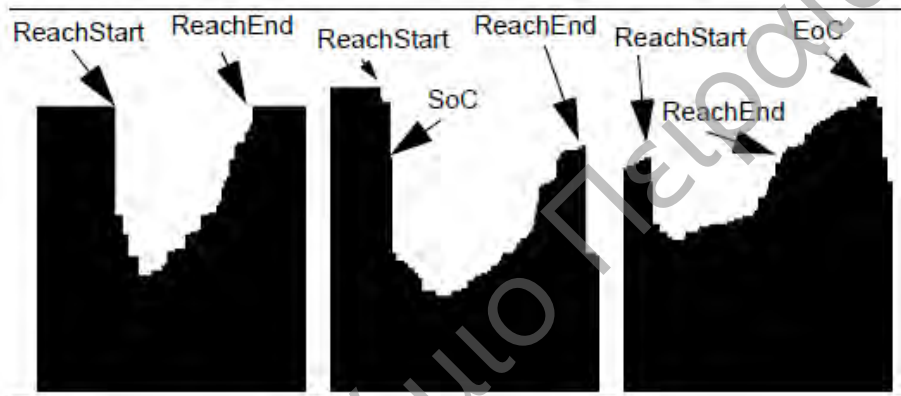
Ένα σύνολο σημείων $cluster_{\xi}(C) \Leftrightarrow \exists D = [s, e] \subseteq [1, n]$ καλείται ξ -συστάδα εάν ικανοποιεί τα ακόλουθα:

- 1) $DownArea_{\xi}(D) \wedge s \in D$
- 2) $UpArea_{\xi}(U) \wedge e \in U$
- 3) a) $e - s \geq MinPts$
b) $x, s_D < x < e_U: (r(x) \leq \min(r(s_D), r(e_U)) \times (1 - \xi))$
- 4) $(s, e) = \begin{cases} (\max\{x \in D \mid r(x) > r(e_U + 1)\}, e_U) & \text{if } r(s_D) \times (1 - \xi) \geq r(e_U + 1) \\ (s_D, \min\{x \in U \mid r(x) < r(s_D)\}) & \text{if } r(e_U + 1) \times (1 - \xi) \geq r(s_D) \\ (s_D, e_U) & \text{otherwise} \end{cases}$

- 1) Το σημείο έναρξης της συστάδας να ανήκει στη περιοχή ξ -απότομα καθοδική D ($DownArea_{\xi}(D)$)
- 2) Το σημείο τερματισμού της συστάδας να ανήκει στη περιοχή ξ -απότομα ανοδική U ($UpArea_{\xi}(U)$)
- 3) a) Η συστάδα να αποτελείται από τουλάχιστον $MinPts$ γείτονες
b) Η τιμή εγγύτητας όλων των σημείων της συστάδας να είναι τουλάχιστον $\xi\%$ χαμηλότερα από την τιμή εγγύτητας του πρώτου σημείου της περιοχής D και του σημείου που έπεται της περιοχής U
- 4) Μπορεί αν οριστεί το σημείο έναρξης και ολοκλήρωσης της συστάδας.

Αναλόγως την τιμή εγγύτητας του πρώτου σημείου της περιοχής D (ReachStart) και της τιμής εγγύτητας του σημείου που έπεται του τελευταίου σημείου της περιοχής U (ReachEnd), διακρίνουμε τρεις περιπτώσεις:

- Οι δύο αυτές τιμές να απέχουν το μέγιστο $\xi\%$: η συστάδα ξεκινάει από το πρώτο σημείο της D και ολοκληρώνεται στο τελευταίο σημείο της U.
- Η τιμή ReachStart να είναι μεγαλύτερη από τη ReachEnd περισσότερο από $\xi\%$: η συστάδα ξεκινάει από το σημείο της D που έχει κατά προσέγγιση την ίδια τιμή εγγύτητας με το σημείο ReachEnd.
- Η τιμή ReachEnd να είναι μεγαλύτερη από τη ReachStart περισσότερο από $\xi\%$: η συστάδα ξεκινάει από το σημείο της U που έχει κατά προσέγγιση την ίδια τιμή εγγύτητας με το σημείο ReachStart.



Εικόνα 11. Διάγραμμα γειτνίασης

2.8 Μέτρα Ταξινόμησης Χρονοσειρών

Για τον ορισμό των αποστάσεων μεταξύ των σημείων των συστάδων (τα οποία στην προκειμένη περίπτωση είναι τροχιές κινούμενων αντικειμένων) χρησιμοποιούνται διάφορες συναρτήσεις υπολογισμού απόστασης που παρουσιάζονται στη συνέχεια.

Μετρικές Συναρτήσεις Απόστασης.

Για τον υπολογισμό της ομοιότητας των τροχιών κινούμενων αντικειμένων απαιτείται ο υπολογισμός της απόστασης μεταξύ τους, σε διάφορες χρονικές στιγμές. Οι συναρτήσεις απόστασης που ικανοποιούν τη τριγωνική ανισότητα καλούνται «Μετρικές συναρτήσεις απόστασης»

Μετρική L_p .

Η L_p -μετρική d_{L_p} για δύο διανύσματα R, S και $1 \leq p \leq \infty$, είναι μία μετρική νόρμα του R^n που ορίζεται από το τύπο $\|r - s\|_p$, όπου η L_p -νόρμα $\|r\|_p$ ορίζεται από το τύπο:

$$\|r\|_p = \left(\sum_{i=1}^n |r_i|^p \right)^{\frac{1}{p}}$$

Συνεπώς η L_p -μετρική ορίζεται από το τύπο:

$$\left(\sum_{i=1}^{\infty} |r_i - s_i|^p \right)^{\frac{1}{p}}$$

Τα πλεονεκτήματα των L_p – μετρικών είναι ότι η πολυπλοκότητα του υπολογισμού είναι γραμμική, δεν απαιτούν ιδιαίτερες παραμέτρους και είναι ιδιαίτερα απλές στην εφαρμογή και την ευρετηρίαση τους.

Επιπλέον η Ευκλείδεια απόσταση είναι έχει εξίσου θεαματικά αποτελέσματα με πολύ πολυπλοκότερες προσεγγίσεις. [6] Τα βασικά μειονεκτήματα αυτών των μετρικών είναι η αναγκαιότητα να έχουν το ίδιο μήκος οι τροχιές, η ευαισθησία τους στο θόρυβο και στις χρονικές αποκλίσεις και η ανικανότητα χειρισμού της τοπικής χρονικής μετατόπισης. [7]

Οι μετρικές που χρησιμοποιούνται στην υλοποίηση αυτής της εργασίας είναι οι μετρικές L_1 (Μανχάταν), L_2 (Ευκλείδεια) and L_∞ (Τσεμπισέβ).

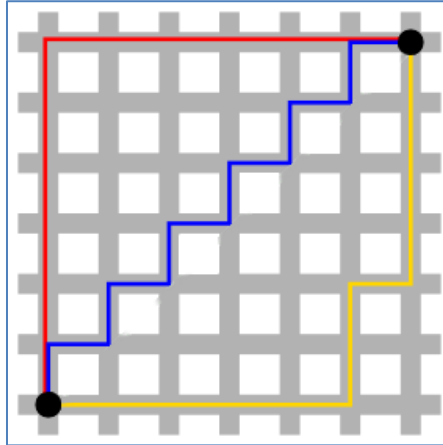
Μετρική L_1 (Manhattan).

Η Μανχάταν απόσταση $L_1(R,S)$, μεταξύ δύο διανυσμάτων R,S , σε ένα n -διαστάσεων πραγματικό διανυσματικό χώρο με σύστημα συγκεκριμένων Καρτεσιανών συντεταγμένων, είναι το άθροισμα του μήκους των προβολών του ευθύγραμμου τμήματος μεταξύ των σημείων των αξόνων των συντεταγμένων. Πιο επίσημα,

$$L_1(R,S) = \sum_{i=1}^n |r_{i,x} - s_{i,x}| + |r_{i,y} - s_{i,y}|$$

Η απόσταση Μανχάταν είναι περισσότερο αποτελεσματική με τη διαχείριση των ακραίων τιμών και του θορύβου συγκριτικά με την Ευκλείδεια απόσταση, επειδή χρησιμοποιείται η απόλυτη τιμή.

Μία γραφική αναπαράσταση του υπολογισμού της απόστασης είναι η ακόλουθη:



Εικόνα 12. Παράδειγμα απόστασης Μανχάταν

Μετρική L_2 (Ευκλείδεια).

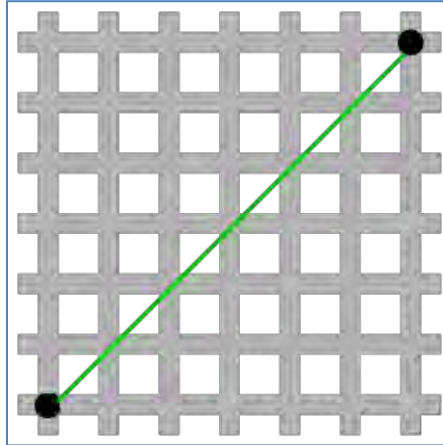
Η Ευκλείδεια απόσταση είναι μία από τις κοινότερες μετρικές και βασίζεται στο Πυθαγόρειο Θεώρημα. Είναι παραπλήσια της απόστασης Μανχάταν, με μόνη διαφορά τη χρήση των τετραγωνικών αποκλίσεων αντί της απόλυτης απόκλισης.

Ο μαθηματικός τύπος της Ευκλείδειας απόστασης προκύπτει από το τύπο της μετρικής L_p τοποθετώντας την τιμή 2 στο p και για δύο τροχιές κινούμενων αντικειμένων R, S είναι:

$$L_2(R, S) = \sqrt{\sum_{i=1}^n (r_{i,x} - s_{i,x})^2 + (r_{i,y} - s_{i,y})^2}$$

Όπου n είναι το πλήθος των σημείων.

Ένα γραφικό παράδειγμα του υπολογισμού της απόστασης είναι το ακόλουθο:



Εικόνα 13. Παράδειγμα υπολογισμού Ευκλείδειας απόστασης

Το μειονέκτημα της Ευκλείδειας απόστασης είναι ότι δεν μπορεί να χειριστεί την τοπική χρονική μετατόπιση και την διαφορά στα μήκη των τροχιών.

Μετρική L_∞ (Τσεμπισέβ).

Η απόσταση Τσεμπισέβ χρησιμοποιεί μόνο τη μεγαλύτερη απόκλιση από τις δύο τροχιές, αντί των αποκλίσεων από όλα τα σημεία, όπως γινόταν με τις προηγούμενες μετρικές. Η απόσταση Τσεμπισέβ έχει χρησιμοποιηθεί και στο παρελθόν για την εύρεση ομοιότητας δύο τροχιών, αλλά δεν μπορεί να χρησιμοποιηθεί για την ταξινόμηση τροχιών βασισμένη στην απόσταση, καθώς μπορεί να υπολογίσει μόνο τη μεγαλύτερη αλλά όχι τη συνολική απόσταση μεταξύ δύο τροχιών.

Ο μαθηματικός τύπος της απόστασης Τσεμπισέβ για τις τροχιές R, S είναι:

$$L_\infty(R, S) = \max_{i=1}^n (|r_{i,x} - s_{i,x}|, |r_{i,y} - s_{i,y}|)$$

Όπου n είναι το άθροισμα των σημείων των τροχιών.

Ένα γραφικό παράδειγμα του υπολογισμού της απόστασης είναι το ακόλουθο:

2	2	2	2	2
2	1	1	1	2
2	1	0	1	2
2	1	1	1	2
2	2	2	2	2

Εικόνα 14. Παράδειγμα υπολογισμού απόστασης

Συναρτήσεις ομοιότητας βασισμένες στην Ευκλείδεια απόσταση.

Έχοντας δύο τροχιές $p_1, p_2, p_3, \dots, p_n$ και $r_1, r_2, r_3, \dots, r_n$ η ευκλείδεια απόσταση μπορεί να χρησιμοποιηθεί για τις ακόλουθες συναρτήσεις ομοιότητας:

Starts only: Η απόσταση των τροχιών ορίζεται ως η Ευκλείδεια απόσταση των σημείων p_1 και r_1 έναρξης των τροχιών χωρίς να λαμβάνεται υπόψη ο χρόνος.

Ends only: Η απόσταση των τροχιών ορίζεται ως η Ευκλείδεια απόσταση των σημείων p_n και r_n τερματισμού των τροχιών χωρίς να λαμβάνεται υπόψη ο χρόνος.

Starts ends: Η απόσταση των τροχιών ορίζεται ως η μέση τιμή της Ευκλείδειας απόστασης των σημείων p_1 και r_1 έναρξης των τροχιών και των σημείων p_n και r_n τερματισμού των τροχιών χωρίς να λαμβάνεται υπόψη ο χρόνος.

DISSIM.

Οι ανωτέρω μετρικές αγνοούν την διάσταση του χρόνου της κίνησης. Αυτό έχει ως αποτέλεσμα να υπολογίζονται μόνο οι χωρικές ομοιότητες μεταξύ των τροχιών και θεωρείται πως οι τροχιές έχουν το ίδιο μήκος και το ίδιο ρυθμό δειγματοληψίας.

Στην εργασία του ο Φρέντζος και η σχετική ομάδα [9] προκειμένου να επιτύχουν τον υπολογισμό της ομοιότητας των χρονοσειρών με διαφορετικό ρυθμό δειγματοληψίας όρισαν την έννοια της χωροχρονικής *ανομοιότητας* μεταξύ δύο τροχιών R και S που και οι δύο υπάρχουν σε ένα ορισμένο χρονικό διάστημα $[t_1, t_m]$, έχοντας ορισμένη Ευκλείδεια απόσταση σε αυτό το χρόνο.

Επειδή ο ακριβής ορισμός της συνάρτησης είναι αριθμητικά πολύ δύσκολος ο υπολογισμός προτάθηκε το ακόλουθο για χρήση:

Λήμμα: Η τιμή ανομοιότητας μεταξύ δύο σημείων που κινούνται γραμμικά με το χρόνο, μπορεί να υπολογιστεί από την ακόλουθη έκφραση:

$$DISSIM(R, S) \approx \frac{1}{2} \sum_{k=1}^{n-1} ((D_{R,S}(t_k) + D_{R,S}(t_{k+1})) (t_{k+1} - t_k))$$

Με σφάλμα υπολογισμού, που εξαρτάται από τις τιμές t_k, t_{k+1} , που οριοθετείται από τον ακόλουθο τύπο:

$$E_{R,S} \leq \sum_{k=1}^{n-1} \begin{cases} \frac{(t_{k+1} - t_k)^3}{12} |D_{R,S}^{(2)}(-b/2a)|, & \text{εάν } t_k \leq -b/2a \leq t_{k+1} \\ \frac{(t_{k+1} - t_k)^3}{12} |D_{R,S}^{(2)}(t_{k+1})|, & \text{εάν } t_k \leq t_{k+1} \leq -b/2a \\ \frac{(t_{k+1} - t_k)^3}{12} |D_{R,S}^{(2)}(t_k)|, & \text{εάν } -b/2a \leq t_k \leq t_{k+1} \end{cases}$$

όπου :

$$D_{R,S} = \sqrt{at^2 + bt + c}$$

Με αυτό τον τρόπο ορίζει την Ευκλείδεια απόσταση μεταξύ δύο σημείων που κινούνται με γραμμικές συναρτήσεις χρόνου μεταξύ συνεχών χρονοσημάτων.

Δυναμική Στρέβλωση Χρόνου (Dynamic Time Warping - DTW).

Η τεχνική της Δυναμικής Στρέβλωσης Χρόνου (DTW) χρησιμοποιεί μία δυναμική προσέγγιση για την ευθυγράμμιση των χρονοσειρών με διαφορές στη χρονική διάρκεια και/ή ταχύτητα.

Αρχικά είχε προταθεί ως εργαλείο αναγνώρισης φωνής από τους Berndt και Clifford [10] στην κοινότητα εξόρυξης γνώσης, ώστε να επιτραπεί οι χρονικές σειρές να «τεντωθούν» ή να «συμπυκνωθούν» προκειμένου να επιτύχει καλύτερο ταίριασμα των χρονοσειρών. Το υπολογιστικό κόστος του DTW στα μεγάλα σύνολα δεδομένων είναι γραμμικό. Στην αρχική του εκδοχή ο αλγόριθμος υπολογισμού απόστασης DTW δεν περιέχει παραμέτρους, ωστόσο η επιβολή ενός προσωρινού περιορισμού στο μέγεθος του παραθύρου στρέβλωσης βελτιώνει την αποτελεσματικότητα του αλγορίθμου στους υπολογισμούς και την ακρίβεια του στην μέτρηση της ομοιότητας των χρονοσειρών, επειδή στην εκτεταμένη στρέβλωση υπάρχει μεγάλη πιθανότητα εσφαλμένου ταιριάσματος μεταξύ δύο χρονοσειρών καθώς και εμφάνιση ομοιότητας που δεν υπάρχει στην πραγματικότητα. Η στρέβλωση με περιορισμό χρησιμοποιείται για ανάπτυξη της χαμηλότερα ορισμένης απόστασης και για την ευρετηρίαση των χρονοσειρών βασισμένη στο DTW.

Στην ακόλουθη εικόνα αναπαριστάται οπτικά η διαφορά σύγκρισης δύο χρονοσειρών με την ευκλείδεια απόσταση και τη τεχνική DTW.



Εικόνα 15. Ευκλείδεια απόσταση και τεχνική DTW

Η Ευκλείδεια απόσταση κάνει αντιστοίχιση ένα προς ένα των δύο ακολουθιών, ενώ η τεχνική DTW επιτρέπει την μία προς πολλά αντιστοίχιση. Όπως μπορούμε να δούμε τη Ευκλείδεια απόσταση υπολογίζει εσφαλμένα ότι οι δύο σειρές έχουν μεγάλη ανομοιογένεια επειδή οι κορυφές τους εμφανίζονται σε διαφορετική χρονική στιγμή. Αυτό διορθώνεται από τη DTW μη γραμμική αντιστοίχιση (εικόνα 16). [11] DTW είναι μία βελτίωση των μετρικών συναρτήσεων, αλλά και αυτή είναι ευαίσθητη στο θόρυβο καθώς όλα τα σημεία των χρονοσειρών πρέπει να αντιστοιχηθούν, ακόμα και οι ακραίες τιμές.

Το υπολογιστικό κόστος του DTW είναι $O(n^2)$ [12] με τη χρήση όμως συγκεκριμένων τεχνικών οριοθέτησης των σημείων και επιτάχυνσης της αναζήτησης ομοιότητας επιτυγχάνεται η μείωση του κόστους στα μεγάλα σύνολα δεδομένων να είναι γραμμικό. [11]

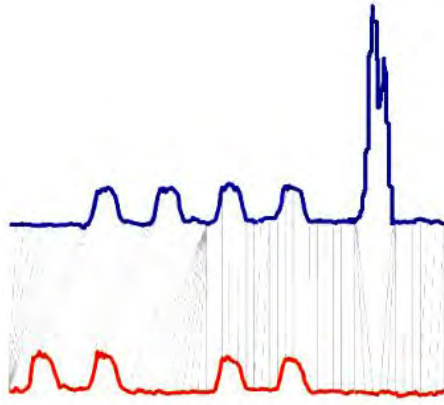
Η συνάρτηση DTW για δύο δισδιάστατες χρονίσεις R, S ορίζεται ως εξής [13]:

$$DTW(R, S) = L_p((r_{x,n}, r_{y,n}), (s_{x,m}, s_{y,m})) + \min \left\{ \begin{array}{l} DTW(R, \text{Head}(S)), \\ DTW(\text{Head}(R), S), \\ DTW(\text{Head}(R), \text{Head}(S)) \end{array} \right\}$$

όπου $\text{Head}(X) = ((x_{1,x}, x_{1,y}) \dots (x_{n-1,x}, x_{n-1,y}))$ και L_p είναι κάθε L_p μετρική.

Μεγαλύτερη Κοινή Υπο-ακολουθία (LCSS).

Η μέθοδος LCSS (Longest Common Subsequence) προσπαθεί να ταιριάζει δύο ακολουθίες επιτρέποντας το “τέντωμα” των ακολουθιών χωρίς την αναδιάταξη της συνέχειας των ακολουθιών, ενώ επιτρέπει κάποια σημεία της μίας ακολουθίας να μην αντιστοιχηθούν με σημεία της δεύτερης και αντιστρόφως. Το βασικό της πλεονέκτημα συγκριτικά με την Ευκλείδεια απόσταση και την DTW, είναι ότι η LCSS είναι πιο αποδοτική στο θόρυβο καθώς μπορεί να αγνοεί τα σημεία/τμήματα θορύβου, σε αντίθεση με τις πρώτες όπου προσπαθούν να ταιριάζουν όλα τα σημεία της μίας τροχιάς με την άλλη. [14]



εικόνα 16. Παράδειγμα μεθόδου LCSS [15]

Ο ορισμός που έχει δοθεί από τον Μ. Βλάχο και την ομάδα του [14] είναι:

Ορισμός: Δοσμένου ενός ακεραίου δ και ενός πραγματικού αριθμού $0 \leq \epsilon \leq 1$, ορίζεται $LCSS_{\delta, \epsilon}(R, S)$ ως εξής:

$$LCSS(R, S) = \begin{cases} 0, & \text{αν } R \text{ ή } S \text{ είναι κενό} \\ 1 + LCSS(\text{Head}(R), \text{Head}(S)), & \text{αν } |r_{n,x} - s_{m,x}| < \epsilon \\ & \text{και } |r_{n,y} - s_{m,y}| < \epsilon \\ & \text{και } |r - s| \leq \delta \\ \max\{LCSS(\text{Head}(R), S), LCSS(R, \text{Head}(S))\}, & \text{διαφορετικά} \end{cases}$$

Η σταθερά δ ελέγχει τη χρονική διαφορά που μπορεί να υπάρχει ώστε να ταιριάζει ένα σημείο από μία ακολουθία με ένα σημείο από την άλλη ακολουθία. Η σταθερά ϵ είναι η μέγιστη απόσταση που μπορεί να έχουν δύο σημεία ώστε να ταιριάζουν.

Η απόσταση μεταξύ δύο χρονοσειρών R, S με μήκος n, m μπορεί να οριστεί ως εξής:

$$D_{\delta, \epsilon}(R, S) = 1 - \frac{LCSS_{\delta, \epsilon}(R, S)}{\min(n, m)}$$

Απόσταση Επεξεργασίας σε Πραγματικές Ακολουθίες (EDR).

Η συνάρτηση EDR βασίζεται στην συνάρτηση επεξεργασία απόστασης (edit distance) που χρησιμοποιείται ευρέως στη βιοπληροφορική και την αναγνώριση φωνής, για τη μέτρηση ομοιότητας μεταξύ δύο συμβολοσειρών.[7] Η συνάρτηση EDR απομακρύνει τα σημεία του θορύβου υλοποιώντας κβαντισμό στην απόσ-

ταση μεταξύ ενός ζεύγους στοιχείων σε τιμές 0 ή 1. Αναζητώντας τον ελάχιστο αριθμό μεταβολών ώστε η μία τροχιά να μετατραπεί σε άλλη η συνάρτηση EDR έχει τη δυνατότητα να χειρίζεται τοπικές μετατοπίσεις χρόνου. Επιπλέον, θέτοντας ποινές σε τμήματα που δεν έχουν αντιστοιχηθεί βελτιώνεται η ακρίβεια της συνάρτησης. Καθώς οι τροχιές δεν είναι συμβολοσειρές, αλλά ζεύγη ακολουθιών αριθμητικών τιμών, ο L. Chen στην εργασία του [16] ορίζει την αντιστοίχιση μεταξύ δύο στοιχείων διαφορετικών τροχιών.

Ορισμός 1: Ένα ζεύγος διανυσμάτων στοιχείων τροχιάς r_i και s_j από δύο τροχιές R και S, αντίστοιχα ταιριάζουν όταν ισχύει το ακόλουθο:

$$match(r_i, s_j) = \begin{cases} 1, & \text{εάν } |r_{i,x} - s_{j,x}| \leq \epsilon \text{ και } |r_{i,y} - s_{j,y}| \leq \epsilon \\ 0, & \text{διαφορετικά} \end{cases}$$

όπου ϵ είναι το κατώφλι ταιριάσματος.

Ορισμός 2: Δοθέντων δύο τροχιών R και S με μήκος n και m αντίστοιχα, η απόσταση επεξεργασίας σε πραγματικές ακολουθίες (EDR) μεταξύ R και S είναι ο αριθμός ενεργειών εισαγωγής, διαγραφής ή αντικατάστασης που χρειάζονται ώστε να μετατραπεί η τροχιά R στην τροχιά S. Η συνάρτηση $EDR(R,S)$ ορίζεται ως εξής:

$$EDR(R, S) = \begin{cases} n, & \text{εάν } m = 0 \\ m, & \text{εάν } n = 0 \\ \min \left\{ \begin{array}{l} EDR(Rest(R), Rest(S)) + subcost, \\ EDR(Rest(R), S) + 1, \\ ERP(R, Rest(S)) + 1 \end{array} \right\}, & \text{διαφορετικά} \end{cases}$$

Όπου $Rest(R) = ((r_{2,x}, r_{2,y}) \dots (r_{n,x}, r_{n,y}))$ και το κόστος ανάθεσης είναι:

$$subcost = \begin{cases} 0, & \text{εάν } match(r_1, s_1) = 1 \\ 1, & \text{διαφορετικά} \end{cases}$$

Στον δεύτερο ορισμό το κόστος ενέργειας αντικατάστασης, εισαγωγής ή διαγραφής είναι μόνο 1 όπως ορίζεται και στον αυθεντικό ορισμό της συνάρτησης επεξεργασίας απόστασης (edit distance) [16].

Συγκριτικά με την Ευκλείδεια απόσταση, DTW, ERP και LCSS, η συνάρτηση EDR έχει τα ακόλουθα πλεονεκτήματα:

- Στην EDR, το κατώφλι αντιστοίχισης μειώνει τα φαινόμενα θορύβου κβαντίζοντας την απόσταση μεταξύ του ζεύγους δύο στοιχείων σε 2 τιμές, 0 ή 1, όπως και η συνάρτηση LCSS. Συνεπώς το φαινόμενο των ακραίων τιμών στην υπολογισμένη απόσταση είναι πολύ μικρότερη σε σχέση με τον υπολογισμό της συνάρτησης EDR από την απόσταση με την Ευκλείδεια απόσταση, τη συνάρτηση DTW και την ERP, που είναι όλες ευαίσθητες στο θόρυβο.

- Επιδιώκοντας τον ελάχιστο αριθμό ενεργειών επεξεργασίας που απαιτείται για την μετατροπή μίας τροχιάς σε άλλη, η συνάρτηση EDR έχει τη δυνατότητα χειρισμού της τοπικής χρονικής μεταβολής που δεν μπορεί να χειριστεί η Ευκλείδεια απόσταση.
- Σε αντίθεση με την συνάρτηση LCSS, η συνάρτηση EDR ορίζει ποινές για τα κενά που εντοπίζονται μεταξύ τμημάτων τροχιών που έχουν αντιστοιχηθεί, σύμφωνα με το μήκος των κενών, βελτιώνοντας την ακρίβεια σε σχέση με το αποτέλεσμα της συνάρτησης LCSS.

Απόσταση Επεξεργασίας με Πραγματική Ποινή (ERP)

Η συνάρτηση ERP είναι συνδυασμός των μετρικών συναρτήσεων απόστασης, που δεν μπορούν να υποστηρίξουν την χρονική μετατόπιση, και των μη μετρικών συναρτήσεων απόστασης, που υποστηρίζουν τη χρονική μετατόπιση. Η συνάρτηση ERP συνδυάζει τα πλεονεκτήματα των μετρικών και των μη μετρικών συναρτήσεων απόστασης καθώς υποστηρίζει την τοπική χρονική μετατόπιση και είναι και μετρική. [17] Η συνάρτηση ERP χρησιμοποιεί πραγματικές ποινές μεταξύ δύο μη κενών στοιχείων και μία σταθερή τιμή για τον υπολογισμό της απόστασης των κενών. Ο αλγόριθμος που χρησιμοποιεί είναι ο ακόλουθος:

$$dist_{erp}(r_i, s_i) = \begin{cases} |r_i - s_i| & \text{εάν } r_i, s_i \text{ δεν είναι κενά} \\ |r_i - g| & \text{αν } s_i \text{ είναι κενό} \\ |s_i - g| & \text{αν } r_i \text{ είναι κενό} \end{cases}$$

όπου g είναι μία σταθερή τιμή.

Βάσει του ανωτέρω αλγορίθμου ορίζεται η απόσταση ERP μεταξύ δύο χρονοσειρών ως εξής:

$$ERP(R, S) = \begin{cases} \sum_{i=1}^n |s_i - g|, & \text{εάν } m = 0 \\ \sum_{i=1}^n |r_i - g|, & \text{εάν } n = 0 \\ \min \left\{ \begin{array}{l} ERP(Rest(R), Rest(S)) + dist(t_1, s_1), \\ ERP(Rest(R), S) + dist(r_1, g), \\ ERP(R, Rest(S)) + dist(s_1, g) \end{array} \right\}, & \text{διαφορετικά} \end{cases}$$

Όπου $Rest(R) = ((r2, x, r2, y) \dots (m, x, m, y))$, και g είναι μία σταθερή τιμή.

Στην εφαρμογή ως σταθερή τιμή χρησιμοποιείται το σημείο $(0,0)$ όπως προτείνεται από τους συγγραφείς της μεθόδου. [17]

Στον ακόλουθο πίνακα συνοψίζονται τα χαρακτηριστικά των συναρτήσεων απόστασης όπως αυτά έχουν αναλυθεί [22]:

<i>Συναρτήσεις Απόστασης</i>	<i>Χρονική Μετατόπιση</i>	<i>Ανοχή στο θόρυβο</i>	<i>Μετρική</i>	<i>Υπολογιστικό Κόστος</i>
L ₁ (Manhattan)	-	-	√	O(n)
L ₂ (Euclidean)	-	-	√	O(n)
L _∞ (Maximum)	-	-	√	O(n)
DISSIM	√	-	√	
Dynamic Time Warping (DTW)	√	-	-	O(n ²)
Longest Common SubSequence (LCSS)	√	√	-	O(n ²)
ERP	√	-	√	O(n ²)
EDR	√	√	-	O(n*m)

Πίνακας 1. Χαρακτηριστικά μέτρων ταξινόμησης χρονοσειρών

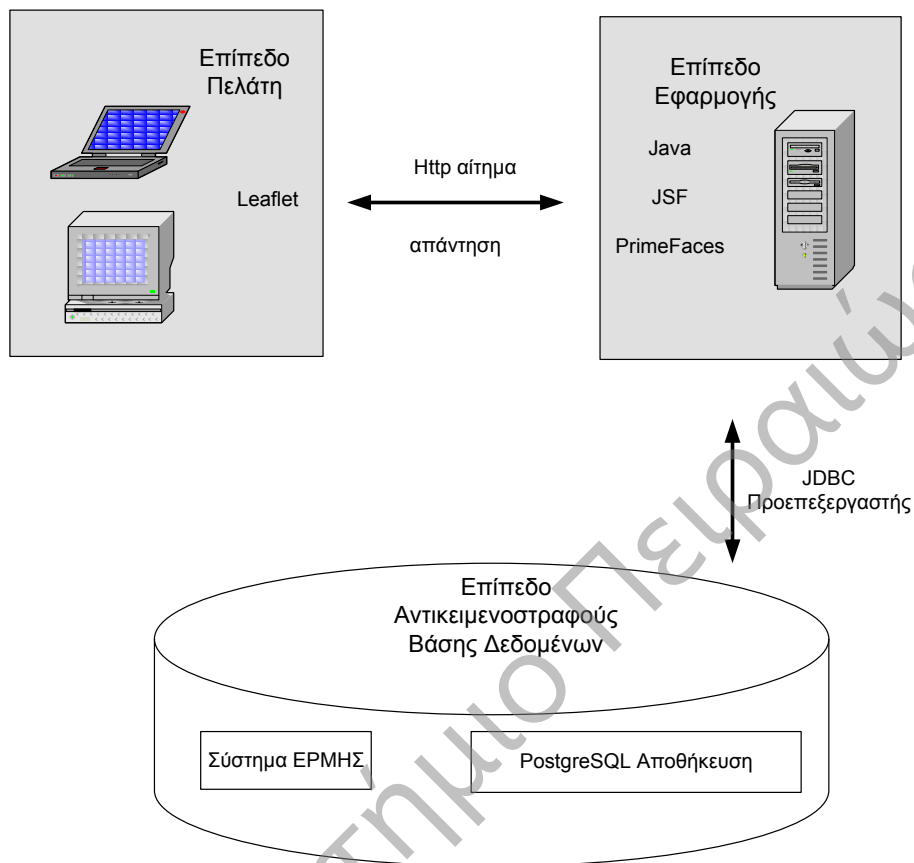
3 Hermes.ChoroChronos.org

Από τη χρήση του ChoroChronos.org εντοπίστηκε η ανάγκη τα δεδομένα που συλλέγονται να μπορούν να αναπαρασταθούν σε χάρτη αλλά και να γίνεται επεξεργασία τους σε πραγματικό χρόνο. Οι μέχρι τώρα εφαρμογές που υπάρχουν είναι παραθυρικές, οπότε απαιτείται ο χρήστης να εγκαταστήσει την σχετική εφαρμογή στον υπολογιστή του, να υλοποιήσει τις εργασίες σύνδεσης του σχετικού κάθε φορά εργαλείου με την βάση και αν τρέξει απαιτούμενα ερωτήματα στη βάση, ώστε να δημιουργηθούν οι βιβλιοθήκες που απαιτούνται. Το hermes.ChoroChronos.org αποτελεί μία διαδικτυακή εφαρμογή όπου ο χρήστης το μόνο που χρειάζεται για να επεξεργαστεί τα δεδομένα που τον ενδιαφέρουν είναι σύνδεση με το διαδίκτυο. Επιπλέον η διαδικτυακή αυτή εφαρμογή δίνει τη δυνατότητα στο χρήστη να μοιραστεί αυτά τα δεδομένα με άλλους χρήστες, μέσω των ομάδων χρηστών που μπορούν να δημιουργήσουν λειτουργικότητα που δεν παρέχεται από τα συστήματα Move Mine και M-Atlas.

Στο hermes.ChoroChronos.org ο χρήστης έχει τη δυνατότητα ανεβάσματος αρχείου στο HermesMOD και την άμεση εμφάνιση στατιστικών στοιχείων γι' αυτό, την εκτέλεση δυναμικών ερωτημάτων σε πραγματικό χρόνο και την οπτικοποίηση τους καθώς και η δυνατότητα υλοποίησης του αλγορίθμου συσταδοποίησης Optics. Για την οπτικοποίηση των δεδομένων που βρίσκονται στη βάση δεδομένων γίνεται χρήση της βιβλιοθήκης Leaflet. Η βιβλιοθήκη Leaflet είναι ανοικτού λογισμικού ανεπτυγμένη σε javascript και είναι κατάλληλη για την δημιουργία αλληλεπιδραστικών χαρτών ιδιαίτερα φιλικών προς το χρήστη.

3.1 Αρχιτεκτονική

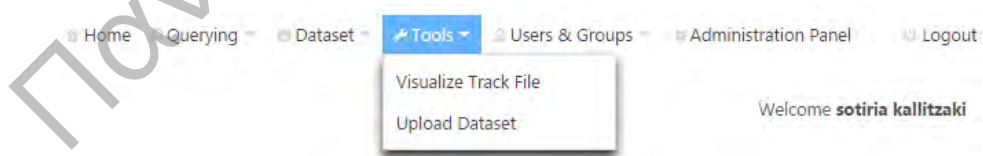
Η εφαρμογή αποτελείται από 3 επίπεδα. Το χαμηλότερο επίπεδο είναι το επίπεδο βάσης δεδομένων. Το ενδιάμεσο επίπεδο είναι το επίπεδο της εφαρμογής και το υψηλότερο επίπεδο είναι το επίπεδο του πελάτη. Η εφαρμογή Hermes.ChoroChronos.org έχει υλοποιηθεί με την γλώσσα προγραμματισμού Java, και τις βιβλιοθήκες JSF και PrimeFaces. Το επίπεδο της εφαρμογής συνδέεται με την βάση δεδομένων που χρησιμοποιεί το πρότυπο Hermes MOD. Στο επίπεδο του πελάτη με την χρήση της βιβλιοθήκης Leaflet γίνεται προβολή σε χάρτη.



Εικόνα 17. Αρχιτεκτονική hermes.choroChronos.org

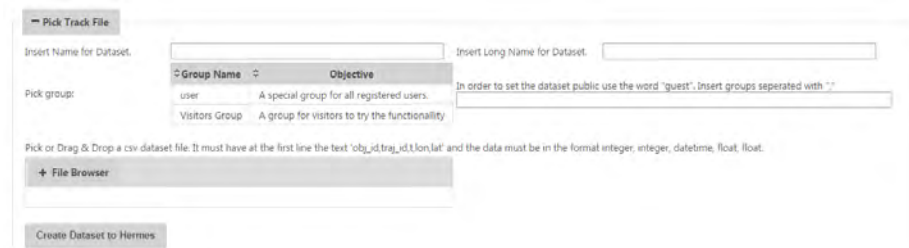
3.2 Ανέβασμα αρχείων δεδομένων

Επιλέγοντας ο χρήστης από το μενού Tools-> Upload Dataset



Εικόνα 18. Επιλογή ανεβάσματος αρχείου

Μπορεί να δει τη σχετική οθόνη με την δυνατότητα ανεβάσματος του αρχείου δεδομένων που τον ενδιαφέρει.



Εικόνα 19. Οθόνη ανεβάσματος αρχείου

Στην σελίδα Upload Dataset μπορεί:

- **Insert Name for Dataset:** να εισάγει το όνομα του συνόλου των δεδομένων, όπως αυτό θα αποθηκευθεί στη βάση,
- **Insert Long Name for Dataset:** να εισάγει το όνομα του συνόλου των δεδομένων όπως αυτό θα εμφανίζεται στη πύλη
- **Pick group:** να εισάγει τις ομάδες χρηστών που θα έχουν πρόσβαση σε αυτό το σύνολο δεδομένων
- **File Browser:** να επιλέξει το αρχείο με τα δεδομένα που τον ενδιαφέρει να ανεβάσει στη πύλη
- **Create Dataset to Hermes:** να εισάγει το συγκεκριμένο σύνολο δεδομένων στην πύλη.

Ολοκληρώνοντας την διαδικασία εισαγωγής, ο χρήστης μπορεί να εκτελέσει ερωτήματα στα δεδομένα που εισήγαγε και να χρησιμοποιήσει τις δυνατότητες που του δίνει το chorochronos.org και το MOD Hermes.

Για το ανέβασμα αρχείου δεδομένων απαιτείται το αρχείο να είναι σε μορφή txt ή csv, η πρώτη γραμμή να είναι η ακόλουθη:

obj_id, traj_id, t, lon, lat
όπου:

- **obj_id:** ο μοναδικός κωδικός του αντικειμένου σε μορφή ακεραίου
- **traj_id:** ο μοναδικός κωδικός κάθε τροχιάς σε μορφή ακεραίου
- **t:** η χρονική στιγμή στην οποία έγινε η καταγραφή του αντικειμένου στη μορφή ΕΕΕΕ-MM-HH ΩΩ:ΛΛ:ΔΔ
- γεωγραφικό μήκος σε μοίρες
- γεωγραφικό πλάτος σε μοίρες

Βασικές διαδικασίες.

Οι βασικές διαδικασίες που χρησιμοποιούνται σε αυτή τη σελίδα είναι:

GetFile.

Διαδικασία που ενεργοποιείται όταν επιλέγεται το σχετικό αρχείο. Καλεί την `copyFile` που αντιγράφει το αρχείο που θέλουμε σε σημείο που έχουμε επιλέξει.

hermesUpload.

Διαδικασία που καλείται ώστε το σύνολο των δεδομένων να εισαχθεί στον EPMH.
Συγκεκριμένα:

1. Δημιουργούνται οι πίνακες

- Όνομα συνόλου δεδομένων_seg
- Όνομα συνόλου δεδομένων_obj
- Όνομα συνόλου δεδομένων_traj

Αν παραδείγματος χάριν έχουμε ένα σύνολο δεδομένων με όνομα test θα δημιουργηθούν οι ακόλουθοι πίνακες:



Εικόνα 20: Πίνακες παραγόμενοι από σύνολο δεδομένων

- Πίνακας **_obj** έχει μία στήλη με το μοναδικό κωδικό του αντικειμένου
- Πίνακας **_traj** έχει μία στήλη με το μοναδικό κωδικό του αντικειμένου που είναι ξένο κλειδί στο πίνακα **_obj**, μία στήλη με το μοναδικό αριθμό της τροχιάς και μία στήλη με την τροχιά
- Πίνακας **_seg** έχει μία στήλη με το μοναδικό κωδικό του αντικειμένου που είναι ξένο κλειδί στο πίνακα **_obj**, μία στήλη με το μοναδικό αριθμό της τροχιάς που είναι ξένο κλειδί στο πίνακα **_traj**, μία στήλη με το μοναδικό κωδικό του τμήματος της τροχιάς και μία στήλη με την τροχιά

2. Ενημερώνεται ο πίνακας hdatasets με τα στοιχεία του συνόλου δεδομένων:

hdatasets
id
name
name_long
lrp
srid
segment_storage
trajectory_storage
subtrajectory_storage
semantics_enabled
partitioning_t_range
partitioning_x_range
partitioning_y_range
notes

εικόνα 21: Πίνακας συνόλων δεδομένων

Ο πίνακας hdatasets είναι μία βελτιωμένη εκδοχή του πίνακα dataset που αναφέρεται στο κεφάλαιο 2.6. Πλέον τα στατιστικά στοιχεία του συνόλου δεδομένων αποθηκεύονται στον πίνακα hdatasets_statistics.

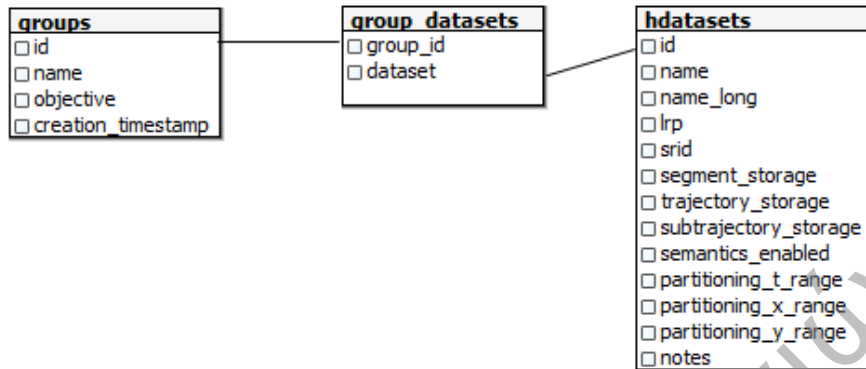
3. Ενημερώνεται ο πίνακας hdatasets_statistics με τα σχετικά στατιστικά στοιχεία του συνόλου δεδομένων.

hdatasets_statistics
<input type="checkbox"/> dataset
<input type="checkbox"/> nr_objects
<input type="checkbox"/> nr_trajectories
<input type="checkbox"/> nr_segments
<input type="checkbox"/> nr_points
<input type="checkbox"/> tmin
<input type="checkbox"/> tmax
<input type="checkbox"/> lx
<input type="checkbox"/> ly
<input type="checkbox"/> hx
<input type="checkbox"/> hy
<input type="checkbox"/> llon
<input type="checkbox"/> llat
<input type="checkbox"/> hlon
<input type="checkbox"/> hlat
<input type="checkbox"/> points_centroid_t
<input type="checkbox"/> points_centroid_x
<input type="checkbox"/> points_centroid_y
<input type="checkbox"/> points_centroid_lon
<input type="checkbox"/> points_centroid_lat
<input type="checkbox"/> trajectories_centroid_t
<input type="checkbox"/> trajectories_centroid_x
<input type="checkbox"/> trajectories_centroid_y
<input type="checkbox"/> trajectories_centroid_lon
<input type="checkbox"/> trajectories_centroid_lat
<input type="checkbox"/> duration
<input type="checkbox"/> area
<input type="checkbox"/> normalized_sampling_rate

εικόνα 22: Πίνακας με ορισμένες στήλες από στατιστικά συνόλου δεδομένων

Τα στατιστικά στοιχεία μπορούν να περιλαμβάνουν μέγιστες και τις ελάχιστες τιμές γεωγραφικού πλάτους, γεωγραφικού μήκους και χρονικών στιγμών τροχιών, μεσαίες τιμές τροχιών κα

4. ενημερώνεται ο πίνακας group_datasets με τη πληροφορία που σχετίζει το σύνολο δεδομένων με συγκεκριμένες ομάδες χρηστών.



εικόνα 23: Συσχέτιση ομάδων χρηστών με σύνολα δεδομένων

Ο πίνακας group_datasets συνδέει το πίνακα hdatasets που περιέχει τη πληροφορία των συνόλων δεδομένων, με το πίνακα groups που περιέχει τις ομάδες των χρηστών. Κάθε ομάδα χρηστών μπορεί να έχει πρόσβαση σε πάνω από ένα σύνολο δεδομένων και αντίστοιχα κάθε σύνολο δεδομένων μπορεί να είναι προσβάσιμο από περισσότερες από μία ομάδες χρηστών.

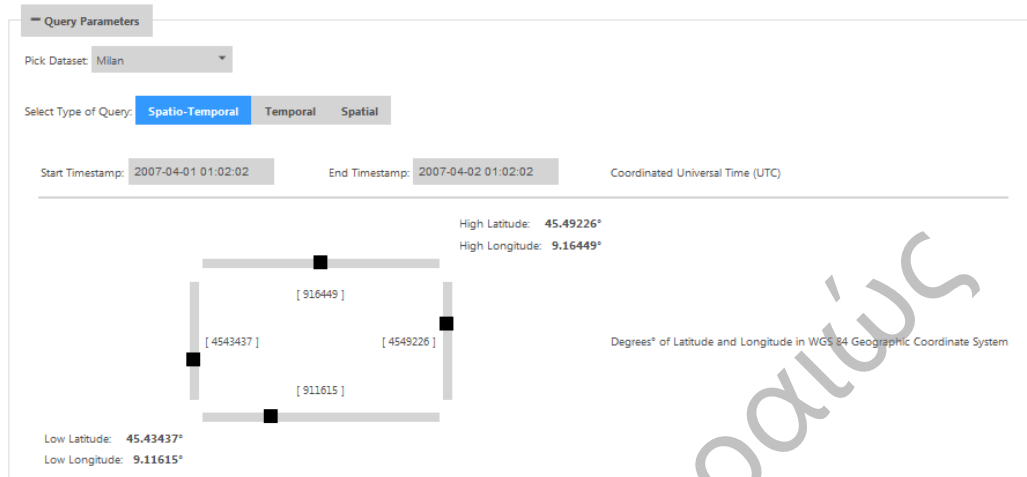
3.3 Δυναμικά ερωτήματα με παραμέτρους εύρος τιμών.

Προστέθηκε η δυνατότητα αναζήτησης πληροφορίας με συγκεκριμένα χωρικά ή/και χρονικά κριτήρια. Με την επιλογή του μενού Querying -> Window Query ο χρήστης μπορεί να επιλέξει το εύρος τιμών για το οποίο θέλει να ενημερωθεί. Μπορεί να επιλέξει αν θα κάνει ερώτημα στη βάση με κριτήρια χρονικά και χωρικά (Spatio-Temporal), μόνο χρονικά (Temporal) ή μόνο χωρικά (Spatial).

Η βασική κλάση στην οποία πραγματοποιούνται οι διαδικασίες διαχείρισης των δυναμικών ερωτημάτων εύρους τιμών είναι η WindowQuery. Για την βελτιστοποίηση του χρόνου απόκρισης των ερωτημάτων έχει γίνει χρήση του ευρετηρίου R-Tree που έχει υλοποιηθεί στο HERMES MOD πάνω στα segments (τμήματα τροχιάς). Η υλοποίηση αυτή έγινε χρησιμοποιώντας τον τελεστή &&.

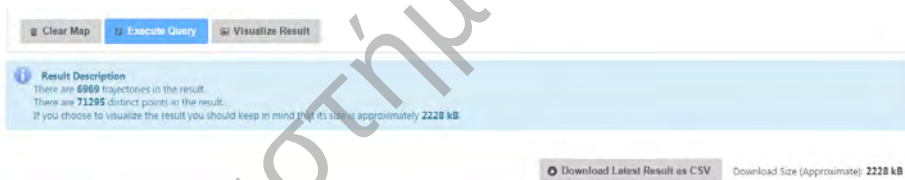
Χωροχρονικά Κριτήρια (Spatio-Temporal).

Με την επιλογή χωροχρονικών κριτηρίων ο χρήστης μπορεί να επιλέξει την χρονική στιγμή έναρξης και την χρονική στιγμή ολοκλήρωσης της αναζήτησης καθώς και την ελάχιστη και τη μέγιστη τιμή γεωγραφικού μήκους και γεωγραφικού πλάτους που τον ενδιαφέρει σε μίρες.



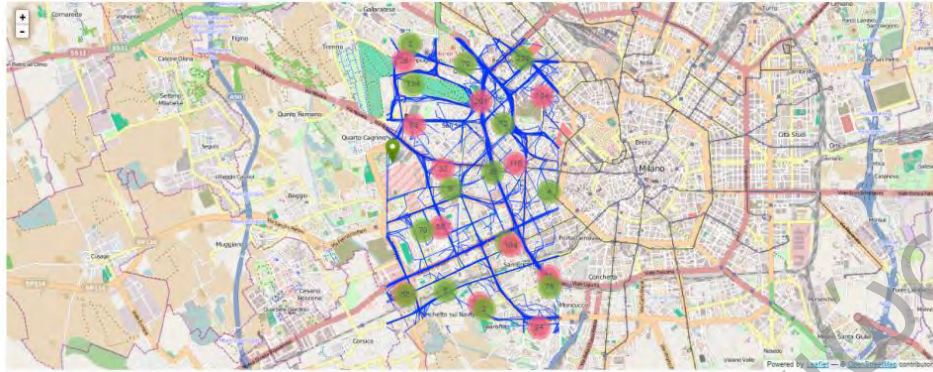
Εικόνα 24. Παράδειγμα δυναμικού ερωτήματος με χωροχρονικά κριτήρια

Στη συνέχεια επιλέγοντας την επιλογή εκτέλεσης του ερωτήματος, εμφανίζεται μία σύντομη περιγραφή του αποτελέσματος του ερωτήματος. Συγκεκριμένα εμφανίζεται το σύνολο των τροχιών κινούμενων αντικειμένων και το σύνολο των σημείων που αποτελούν αποτέλεσμα της συγκεκριμένης αναζήτησης, καθώς και το μέγεθος του αρχείου που θα οπτικοποιηθεί.



Εικόνα 25. Παράδειγμα αποτελέσματος δυναμικού ερωτήματος

Επιλέγοντας το κουμπί «Download Latest Result as CSV» ο χρήστης μπορεί να κατεβάσει τα αποτελέσματα του ερωτήματος. Επιλέγοντας την οπτικοποίηση του αποτελέσματος με το κουμπί “Visualize Result”, εμφανίζεται το αποτέλεσμα στο χάρτη.



Εικόνα 26. Παράδειγμα οπτικοποίησης αποτελέσματος δυναμικού ερωτήματος με χωροχρονικά κριτήρια

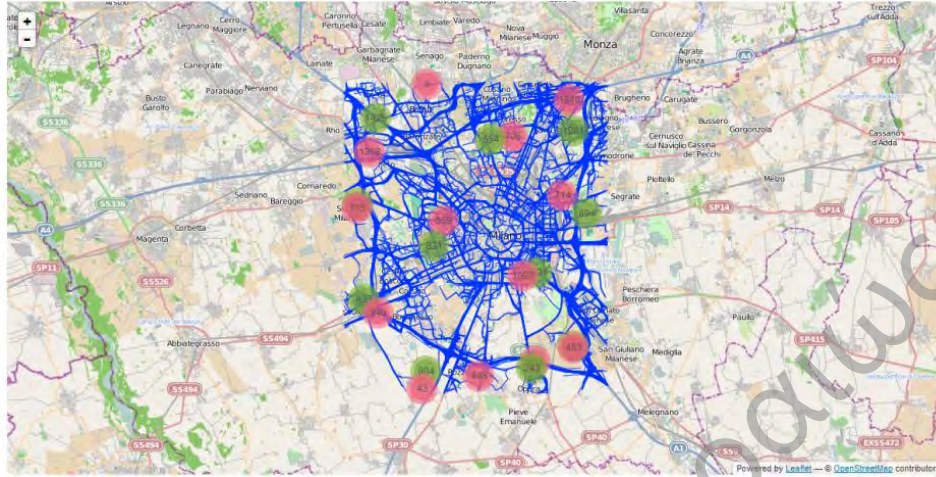
Κατά την οπτικοποίηση του αποτελέσματος παρατηρούμε ότι εκτός από τις τροχιές που αποτελούνται από πολλά σημεία αναπαράστασης πάνω στο χάρτη, προβάλλονται με πράσινο χρώμα τα σημεία έναρξης των τροχιών και με κόκκινο χρώμα τα σημεία τερματισμού των τροχιών κινούμενων αντικειμένων. Η τιμή που βρίσκεται εντός των σχετικών κύκλων αποτελούν το άθροισμα των αντίστοιχων σημείων.

Χρονικά Κριτήρια (Temporal).

Με την επιλογή χρονικών κριτηρίων ο χρήστης μπορεί να επιλέξει την χρονική στιγμή έναρξης και την χρονική στιγμή ολοκλήρωσης της αναζήτησης που τον ενδιαφέρει και στην συνέχεια να επιλέξει την εκτέλεση του ερωτήματος.

Εικόνα 27. Παράδειγμα δυναμικού ερωτήματος με χωρικά κριτήρια

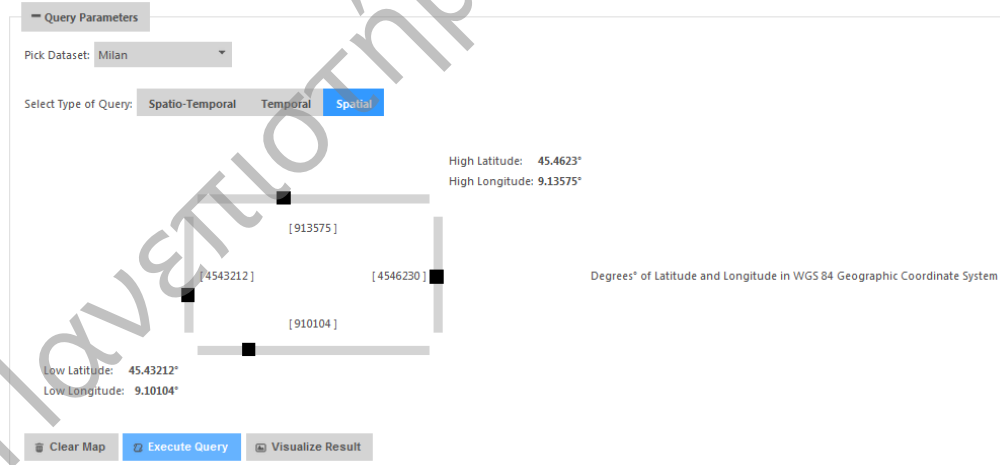
Το αποτέλεσμα του ανωτέρω ερωτήματος μπορούμε να το δούμε στην εικόνα που ακολουθεί.



Εικόνα 28. Παράδειγμα οπτικοποίησης αποτελέσματος δυναμικών ερωτημάτων με χωρικά κριτήρια

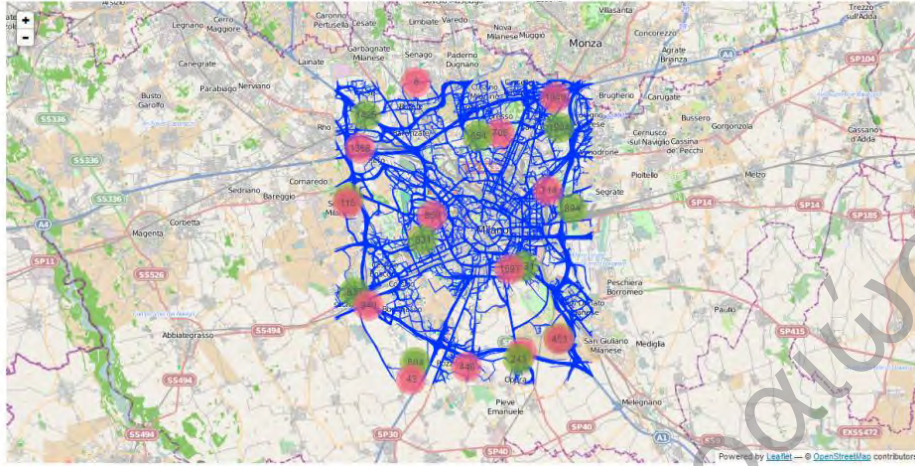
Χωρικά Κριτήρια (Spatial).

Με την επιλογή χωρικών κριτηρίων ο χρήστης μπορεί να επιλέξει την ελάχιστη και τη μέγιστη τιμή γεωγραφικού μήκους και γεωγραφικού πλάτους που τον ενδιαφέρει σε μίρες.



Εικόνα 29. Παράδειγμα δυναμικών ερωτημάτων με χρονικά κριτήρια

Η οπτικοποίηση του αποτελέσματος παρουσιάζεται στην ακόλουθη εικόνα:



Εικόνα 30. Παράδειγμα οπτικοποίησης αποτελέσματος δυναμικών ερωτημάτων με χρονικά κριτήρια

Συνάρτηση εκτέλεσης ερωτήματος στη βάση.

Έχει ως εισόδους:

1. Το είδος του ερωτήματος που έχει επιλεγθεί, δηλαδή αν το ερώτημα αφορά κριτήρια χωρικά, χρονικά και χωροχρονικά.
2. Η μέγιστη και η ελάχιστη χρονική στιγμή αναζήτησης.
3. Η μέγιστη και η ελάχιστη τιμή γεωγραφικού πλάτους αναζήτησης.
4. Η μέγιστη και η ελάχιστη τιμή γεωγραφικού μήκους αναζήτησης.
5. Το σύνολο δεδομένων πάνω στο οποίο εφαρμόζεται το ερώτημα.

Με την εκτέλεση του ερωτήματος εισάγονται τα επιλεγμένα σημεία σε έναν προσωρινό πίνακα *hermesui_pos* και ενημερώνονται το σύνολο των τροχιών, το σύνολο των σημείων και το μέγεθος του υποσυνόλου που δημιουργείται. Για την εκτέλεση του ερωτήματος γίνεται μετατροπή του των συντεταγμένων σε μέτρα, ώστε να μπορεί να γίνει η εκτέλεση του ερωτήματος από το σύστημα EPMHΣ και μετατροπή του χρόνου σε σύστημα που δεν λαμβάνει υπόψη την χρονική ζώνη. Για την εύρεση των τροχιών ελέγχονται τα τμήματα της τροχιάς (segments) που επικαλύπτουν τη περιοχή που δημιουργείται από τα γεωγραφικά σημεία και τα χρονικά σημεία που έχουν επιλεγθεί. Για την εύρεση αυτής της επικάλυψης χρησιμοποιείται ο τελεστής του EPMH «&&» και

- για την περιοχή που δημιουργείται από το μέγιστο και ελάχιστο γεωγραφικό πλάτος οι συναρτήσεις

$\text{BoxSP}(\text{PointSP}(X \text{ ελάχιστου σημείου}, Y \text{ ελάχιστου σημείου}), \text{PointSP}(X \text{ μέγιστου σημείου}, Y \text{ μέγιστου σημείου})),$

- για την περιοχή που δημιουργείται από την μέγιστη και ελάχιστη χρονική στιγμή, οι συναρτήσεις

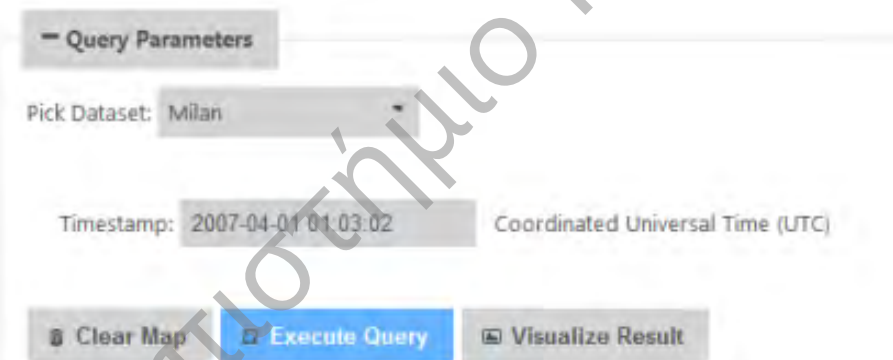
Period(ελάχιστη χρονική στιγμή, μέγιστη χρονική στιγμή)

- Για την περιοχή που δημιουργείται από το μέγιστο και ελάχιστο γεωγραφικό πλάτος και την μέγιστη και ελάχιστη χρονική στιγμή

BoxST(Period(ελάχιστη χρονική στιγμή, μέγιστη χρονική στιγμή), BoxSP(PointSP(X ελάχιστου σημείου, Y ελάχιστου σημείου), PointSP(X μέγιστου σημείου, Y μέγιστου σημείου)))

3.4 Ερώτημα με παράμετρο συγκεκριμένη χρονική στιγμή.

Με την επιλογή του μενού Querying -> TimeSlice Query ο χρήστης μπορεί να επιλέξει μία συγκεκριμένη χρονική στιγμή για την οποία θέλει να ενημερωθεί. Αυτή τη χρονική στιγμή μπορεί να την επιλέξει από τις διαθέσιμες χρονικές στιγμές του συνόλου δεδομένων στο οποίο γίνεται το ερώτημα.



Εικόνα 31. Παράδειγμα δυναμικού ερωτήματος με παράμετρο συγκεκριμένη χρονική στιγμή

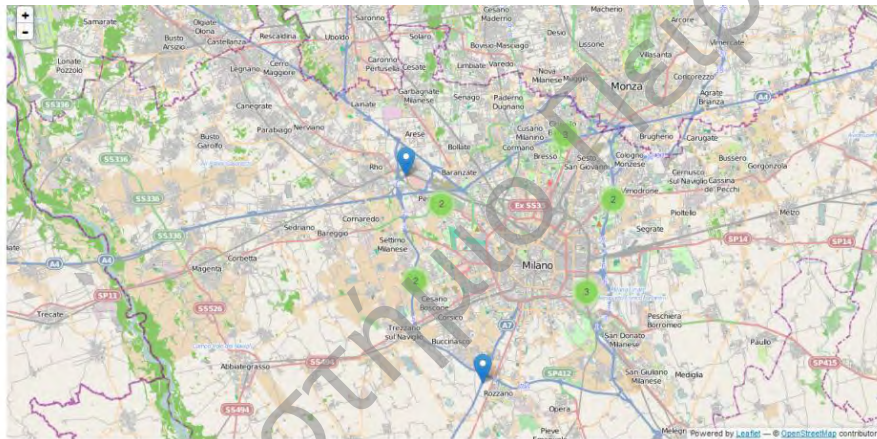
Επιλέγοντας το κουμπί *Execute Query* προβάλλεται στο χρήστη το σύνολο των σημείων που αποτελούν το αποτέλεσμα καθώς και το μέγεθος περίπου του αρχείου που θα οπτικοποιηθεί. Ανάλογος του μεγέθους είναι και ο χρόνος οπτικοποίησης. Ο χρήστης μπορεί και να κατεβάσει το σχετικό αρχείο επιλέγοντας «Download Latest Result as CSV»

Result Description
 There are 11 trajectory points in the result.
 If you choose to visualize the result you should keep in mind that its size is approximately 352 bytes

[Download Latest Result as CSV](#) Download Size (Approximate): 352 bytes

Εικόνα 32. Παράδειγμα αποτελέσματος δυναμικού ερωτήματος με παράμετρο συγκεκριμένη χρονική στιγμή

Η οπτικοποίηση του αποτελέσματος του ανωτέρω ερωτήματος εμφανίζεται στην ακόλουθη εικόνα. Πλέον γίνεται αναπαράσταση μεμονωμένων σημείων, καθώς σε μία συγκεκριμένη χρονική στιγμή κάθε τροχιά κινούμενου αντικειμένου αντιπροσωπεύεται από ένα μόνο γεωγραφικό σημείο. Οι αριθμοί που βρίσκονται μέσα στους χρωματιστούς κύκλους δείχνουν πόσα σημεία βρίσκονται στη συγκεκριμένη περιοχή. Επιλέγοντας τη κάθε περιοχή μπορούμε να δούμε τα σημεία αναλυτικότερα.



Εικόνα 33. Οπτικοποίηση παραδείγματος δυναμικού ερωτήματος με παράμετρο συγκεκριμένη χρονική στιγμή

Συνάρτηση εκτέλεσης ερωτήματος στη βάση.

Έχει ως εισόδους:

1. Τη χρονική στιγμή αναζήτησης σημείων.
2. Το σύνολο δεδομένων πάνω στο οποίο εφαρμόζεται το ερώτημα.

Με την εκτέλεση του ερωτήματος εισάγονται τα επιλεγμένα σημεία σε έναν προσωρινό πίνακα `hermesui_ros` και ενημερώνονται το σύνολο των τροχιών, το σύνολο των σημείων και το μέγεθος του υποσυνόλου που δημιουργείται.

Για την εκτέλεση του ερωτήματος γίνεται μετατροπή του χρόνου σε σύστημα που δεν λαμβάνει υπόψη την χρονική ζώνη. Για την εύρεση των τροχιών ελέγχονται τα τμήματα της τροχιάς (segments) που τέμνονται από την επιλεγμένη χρονική στιγμή. Για

την εύρεση αυτής της τομής χρησιμοποιείται ο τελεστής του EPMH «~» που χρησιμοποιεί την ευρετηρίαση του TB-tree.

3.5 Συσταδοποίηση

Με την επιλογή Quering -> T-Optics Clustering ο χρήστης έχει τις ακόλουθες δυνατότητες:

- δυνατότητα επιλογής μεταξύ των συναρτήσεων Euclidean, Manhattan, Chebyshev, DTW, LCSS, EDR και ERP, Euclidean Start, Euclidean End, Euclidean Start End για την ταξινόμηση των τροχιών καθώς και εισαγωγής των παραμέτρων που χρησιμοποιούνται για την κάθε συνάρτηση απόσταση
- προβολή της ταξινόμησης (ordering) των τροχιών
- εμφάνιση λίστας με τις εξαγόμενες συστάδες με δυνατότητα επιλογής των συστάδων που θα προβληθούν στο χάρτη
- οπτικοποίηση των δεδομένων των επιλεγμένων συστάδων
- δυνατότητα εξαγωγής σε μορφή csv του πίνακα των αποτελεσμάτων απόστασης όλων των τροχιών μεταξύ τους ενεργοποιώντας την επιλογή «Compute and Download Distance Matrix»

Ο πίνακας των αποτελεσμάτων απόστασης όλων των τροχιών μεταξύ τους εξάγεται με την μορφή

- objfrom_id
- trajfrom_id
- objto_id
- trajto_id
- traj_distance

Με αυτό το τρόπο δίνεται η δυνατότητα στο χρήστη να φιλτράρει ή να ταξινόμηση τα δεδομένα με τα κριτήρια που τον ενδιαφέρουν.

Μέτρα Ταξινόμησης Χρονοσειρών.

Οι συναρτήσεις απόστασης που χρησιμοποιούνται είναι μετρικές και μη μετρικές και βασίζονται στις συναρτήσεις που αναλύθηκαν στο κεφάλαιο 2.8 *Μέτρα Ταξινόμησης Χρονοσειρών*. Αναλυτικά:

- **Euclidean**: Χρησιμοποιείται η Ευκλείδεια απόσταση.
- **Euclidean Start**: Χρησιμοποιείται η Ευκλείδεια απόσταση μόνο στα σημεία αρχής των τροχιών.
- **Euclidean End**: Χρησιμοποιείται η Ευκλείδεια απόσταση μόνο στα σημεία τέλους των τροχιών.
- **Euclidean Start End**: Χρησιμοποιείται η Ευκλείδεια απόσταση μόνο στα σημεία αρχής και τέλους των τροχιών και λαμβάνεται η μέση τιμή ως απόσταση των τροχιών.

- **Manhattan:** Χρησιμοποιείται η απόσταση Μανχάταν.
- **Chebyshev:** Χρησιμοποιείται η απόσταση Chebyshev.
- **DTW:** Χρησιμοποιείται η τεχνική της Δυναμικής Στρέβλωσης Χρόνου. Αυτή η τεχνική δέχεται ως παραμέτρους το εύρος δυναμικής στρέβλωσης και τις μετρικές συναρτήσεις Ευκλείδεια και Μανχάταν.
- **LCSS:** Χρησιμοποιείται η μέθοδος της Μεγαλύτερης Κοινής Υπο-ακολουθίας. Οι παράμετροι που δέχεται είναι το χρονικό εύρος και το χωρικό εύρος. Συγκεκριμένα:
 - δ (time threshold): ακέραιος αριθμός.
 - δ perc: Επιπλέον παρέχεται και η δυνατότητα να εισάγει ο χρήστης το δ με ποσοστό. Η τιμή είναι δεκαδική και μεταξύ του 0 και του 1. Για την ενεργοποίηση αυτής της δυνατότητας πρέπει η ακέραιη τιμή δ να έχει τιμή μικρότερη του 0.
 - ϵ (distance threshold): η μέγιστη απόσταση που μπορεί να έχουν δύο σημεία ώστε να ταιριάζουν. Η τιμή αυτή μπορεί να είναι δεκαδική.
- **EDR:** Χρησιμοποιείται η Απόσταση Επεξεργασίας σε Πραγματικές Ακολουθίες. Οι παράμετροι που δέχεται είναι το κατώφλι ταιριάσματος ϵ . Η τιμή που δέχεται είναι ένας δεκαδικός αριθμός.
- **ERP:** Χρησιμοποιείται η Απόσταση Επεξεργασίας με Πραγματική Ποινή. Δέχεται ως παραμέτρους την Ευκλείδεια απόσταση και την απόσταση Μανχάταν.

Βήματα εξαγωγής συστάδων.

Για την εξαγωγή συστάδων τα βήματα είναι τα ακόλουθα:

1. Επιλογή του dataset που μας ενδιαφέρει
2. Επιλογή της συνάρτησης απόστασης με την οποία θέλουμε να γίνει η ταξινόμηση των τροχιών και εισαγωγή των παραμέτρων που απαιτούνται.
3. Επιλογή Ταξινόμησης Τροχιών

Επιλέγοντας την ταξινόμηση τροχιών καλούμε τον αλγόριθμο T-Optics. Συγκεκριμένα για την υλοποίηση της ταξινόμησης των τροχιών σύμφωνα με την απόσταση γειτνίασης καλείται στην εφαρμογή το ακόλουθο PL/SQL σενάριο:

```
SELECT TOptics_Traj ('"+ userBean.getDataset().getName()
+"_traj',
                    ' "
                    +
clusterParams.getDistanceFuntionNameForOptics(clusterPara
ms.getDistanceFunction())
                    + " ', "
                    +
clusterParams.getMinNeighbours()
                    + " ");
```

Η συνάρτηση TOptics_Traj δέχεται ως εισόδους:

- το πίνακα που περιέχει τις τροχιές που μας ενδιαφέρουν
- την διαδικασία εύρεσης γειτνίασης που μας ενδιαφέρει

- τον αριθμό των ελάχιστων γειτόνων που μπορεί να έχει μία συστάδα.

Για την διαδικασία εύρεσης γειτνίασης ισχύουν τα ακόλουθα:

Για κάθε μία από τις συναρτήσεις απόστασης υλοποιήθηκε διαδικασία αντίστοιχη της ακόλουθης που υπολογίζει την απόσταση γειτνίασης των τροχιών κινούμενων αντικειμένων. Κάθε τροχιά κινούμενου αντικειμένου του dataset ελέγχεται αν τέμνεται χρονικά με κάποια άλλη τροχιά. Στην περίπτωση που υπάρχει χρονική αλληλοκάλυψη υπολογίζεται η απόσταση των δύο τροχιών για την κοινή τους περίοδο. Ο λόγος για τον οποίο έχει γίνει η ακόλουθη τροποποίηση είναι ότι οι μετρικές συναρτήσεις για να υπολογίσουν την απόσταση μεταξύ δύο τροχιών πρέπει οι τροχιές να ξεκινάνε και να τελειώσουν ίδιες χρονικές στιγμές. Εάν η απόσταση είναι μη κενή τότε εισάγεται στο πίνακα TOptics_Traj_neighbors η απόσταση μεταξύ των τροχιών. Στην περίπτωση που η απόσταση δεν είναι μετρική, τότε παραβλέπεται η διαδικασία εύρεσης της χρονικής αλληλοκάλυψης των τροχιών. Η συνάρτηση που ακολουθεί αποτελεί τμήμα του EPMH που υλοποιήθηκε για την απόσταση γειτνίασης με την χρήση της Ευκλείδειας συνάρτησης.

```
CREATE OR REPLACE FUNCTION
he_TOptics_Traj_NN_Euclidean(DB text, p_obj integer,
p_traj integer) RETURNS void AS $$
DECLARE
    rec record;
    distance double precision;
    common_period Period;
    p Trajectory;
BEGIN
EXECUTE
'
SELECT traj
FROM ' || quote_ident(DB) || '
WHERE (obj_id, traj_id) = ($1, $2);
'
INTO p
USING p_obj, p_traj;
FOR rec IN
EXECUTE
'
SELECT obj_id, traj_id, traj
FROM ' || quote_ident(DB) || '
WHERE (obj_id, traj_id) <> ($1, $2)
AND traj && BoxST($3);
'
USING p_obj, p_traj, p
```



```

LOOP
  common_period := intersection(p::Period,
rec.traj::Period);
  IF common_period IS NULL THEN
    RETURN;
  END IF;

  distance := Euclidean(atPeriod(p, common_period)
atPeriod(rec.traj, common_period));

  IF distance IS NOT NULL THEN
    INSERT INTO TOptics_Traj_neighbors(obj_id, traj_id,
distance) VALUES (rec.obj_id, rec.traj_id, distance);
  END IF;
END LOOP;
END;
$$ LANGUAGE plpgsql STRICT;

```

Η διαδικασία ταξινόμησης υλοποιείται πατώντας το κουμπί Create Ordering και το αποτέλεσμα εμφανίζεται στο reachability plot. Στην εικόνα που ακολουθεί μπορούμε να δούμε το αποτέλεσμα για την επιλογή υλοποίησης της ταξινόμησης για την Ευκλείδεια απόσταση με ελάχιστο αριθμό γειτόνων 30.



Εικόνα 34. Διάγραμμα γειτνίασης

4. Εισαγωγή του ξ

Ο χρήστης στη συνέχεια εισάγει τη τιμή στη παράμετρο ξ για την οποία θέλει να εξαχθούν οι συστάδες. Η τιμή του ξ μπορεί να είναι μεταξύ του μηδέν και του 1.

5. Επιλογή της εξαγωγής συστάδων

Η διαδικασία που καλείται όταν επιλέγεται η εξαγωγή των συστάδων (Create Clusters) είναι η ακόλουθη:

```
SELECT TOptics_Traj_extractClusters(" +  
clusterParams.getKsi() + ");
```

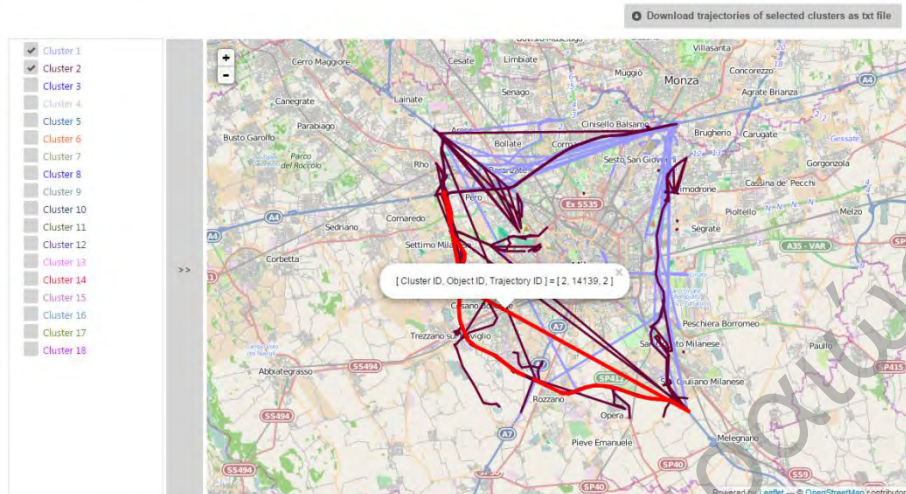
Η οποία αποτελεί διαδικασία του EPMH για την αυτοματοποιημένη εξαγωγή των συστάδων και χρησιμοποιεί την παράμετρο ξ και τον πίνακα των αποστάσεων γειτνίσεως. Όταν η διαδικασία ολοκληρωθεί δημιουργείται μία λίστα με όλα τις συστάδες.



Εικόνα 35. Παράμετρος ξ

6. Οπτικοποίηση και εξαγωγή αρχείου συστάδων

Για κάθε συστάδα που εξάγεται από την προηγούμενη διαδικασία δημιουργείται ένα αντικείμενο σε λίστα που έχει το μοναδικό κωδικό της συστάδας και ένα μοναδικό χρώμα. Ο χρήστης έχει τη δυνατότητα να επιλέξει τις συστάδες που τον ενδιαφέρουν και να τις προβάλει στο χάρτη επιλέγοντας το κουμπί >>. Κάθε τροχιά που εμφανίζεται στο χάρτη έχει το χρώμα αντιστοιχεί στο σχετικό αντικείμενο της λίστας. Η διαδικασία που καλείται κατά την επιλογή της προβολής είναι η visualizeTrajectories. Σε αυτή τη συνάρτηση καλούνται με PL/SQL ερώτημα τα στοιχεία των συστάδων που έχουν επιλεγεί και αποθηκεύονται σε μία μεταβλητή τύπου string με τη μορφή: κωδικός συστάδας, κωδικός αντικειμένου, κωδικός τροχιάς, πίνακας [γεωγραφικό μήκος, γεωγραφικό πλάτος], χρώμα αναπαράστασης συστάδας. Αυτές οι πληροφορίες μεταφέρονται στην σελίδα και με την χρήση της τεχνολογίας JSON και της βιβλιοθήκης Leaflet κάθε τροχιά χρωματίζεται αντίστοιχα με το τη συστάδα που αναπαριστά και σε κάθε τροχιά εισάγεται η πληροφορία του κωδικού της συστάδας, του κωδικού του αντικειμένου και του κωδικού της τροχιάς. Ο χρήστης επιλέγοντας μία τροχιά στο χάρτη, αυτή χρωματίζεται με το κόκκινο χρώμα και εμφανίζεται ο μοναδικός κωδικός της συστάδας, ο μοναδικός κωδικός του αντικειμένου και ο μοναδικός κωδικός της τροχιάς.



Εικόνα 36. Οπτικοποίηση αποτελέσματος συσταδοποίησης

Επιπλέον ο χρήστης έχει τη δυνατότητα τις τροχιές που υπάρχουν σε αυτές τις συστάδες να τις εξάγει σε αρχείο csv επιλέγοντας το κουμπί «Download trajectories of selected clusters as txt file».

3.6 Κοινές διαδικασίες

Οπτικοποίηση δεδομένων κίνησης.

Για την οπτικοποίηση των τροχιών χρησιμοποιείται η διαδικασία *visualizeTrajectories* όπου γίνεται αναζήτηση των σημείων των τροχιών που υπάρχουν στον βοηθητικό πίνακα *hermesui_pros* και γίνεται αποθήκευση των στοιχείων «*id αντικειμένου*», «*id τροχιάς*», μήτρα με «*γεωγραφικό πλάτος*» και «*γεωγραφικό μήκος*» κάθε σημείου της τροχιάς σε μία μεταβλητή και η τιμή αυτής της μεταβλητής μεταφέρεται στο αρχείο *html*. Με την χρήση της τεχνολογίας *JSON* γίνεται μεταγλώττιση της τιμής της μεταβλητής στην *JavaScript* και με την χρήση της βιβλιοθήκης *Leaflet* προστίθενται στα επίπεδα του χάρτη τα σημεία που οπτικοποιούνται καθώς και η πληροφορία «*id αντικειμένου*», «*id τροχιάς*», «*γεωγραφικό πλάτος*» και «*γεωγραφικό μήκος*» στα σημεία αρχής και τέλους κάθε τροχιάς. Η συγκεκριμένη διαδικασία χρησιμοποιείται από τη σελίδα *WindowQuery* και *TimeSliceQuery*.

4 Μελλοντική επέκταση

Το γεωγραφικό portal *chorochronos.org* μπορεί μελλοντικά να επεκτείνει τις δυνατότητες του ώστε να παρέχει ανακατασκευή και μετάπτωση των δεδομένων από τη σχεσιακή βάση στην αντίστοιχη γεωγραφική καθώς και συσχέτιση των δεδομένων με σημασιολογική πληροφορία. Με την ανακατασκευή των δεδομένων θα μπορούν να εισάγονται ακατέργαστα δεδομένα σε ποικίλες μορφές, και θα εμφανίζονται εξομα-

λυμένες τροχιές κίνησης απαλλαγμένες από θόρυβο και ασυνέχειες. Με την συσχέτιση των δεδομένων με σημασιολογικής πληροφορία θα μπορούν εξάγονται στοιχεία για την κίνηση ή στάση του αντικειμένου που θα συνδέονται με το οδικό δίκτυο αλλά και σχετική σημασιολογική πληροφορία για τη θέση του.

4.1 Ανακατασκευή και μετάπτωση δεδομένων

Για την ανακατασκευή και μετάπτωση των δεδομένων που βρίσκονται στη σχεσιακή βάση θα μπορούσε να υλοποιηθεί μία διαδικασία «καθαρισμού» και βελτιστοποίησης ώστε να τοποθετηθούν μεταγενέστερα με νέα πλέον μορφή στη χωροχρονική βάση δεδομένων του Hermes-MOD. Για την διαχείριση των outliers θα μπορούσαν να χρησιμοποιηθούν τεχνικές κατωφλίου με γνώμονα την ταχύτητα και την απόσταση. Η τεχνική που θα μπορούσε να χρησιμοποιηθεί είναι η ακόλουθη:

Εντοπισμός τροχιών.

Δοσμένης μίας μέγιστης απόστασης και μίας μέγιστης χρονικής διάρκειας, δύο σημεία που θα ξεπερνούν αυτά τα μεγέθη, θα αποτελούν το σημείο τερματισμού μίας τροχιάς και το σημείο έναρξης της επόμενης.

Τμηματοποίηση τροχιάς.

Στα δεδομένα που λαμβάνονται από τη σχεσιακή βάση θα υλοποιείται επίσης εύρεση των διακριτών τροχιών κίνησης και τμηματοποίηση αυτής. Κατά την τμηματοποίηση η τροχιά θα χωρίζεται σε διακριτά, μη επικαλυπτόμενα χρονικά, επεισόδια (episodes) [8]. Η τροχιά θα συσχετιστεί με έννοιες όπως αρχή, τέλος, στάση, κίνηση. Με τον χωρισμό της τροχιάς σε επεισόδια επιτυγχάνεται συμπίεση των δεδομένων και εξαγωγή πληροφορίας, όπως η ταχύτητα κίνησης και η κατεύθυνση.

Επιπλέον μπορεί στη συνέχεια η κίνηση ή η στάση να συσχετιστούν με περαιτέρω σημασιολογική πληροφορία.

Γι αυτό το χαρακτηρισμό της κίνησης, θα υπάρχει ένα πεδίο MOVEMENT που θα παίρνει ως τιμές [STOP] ή [MOVE] για τα σημεία που αποτελούν στάση και τμήματα τροχιάς κίνησης.

Ο εντοπισμός των τροχιών θα υλοποιείται με τον αλγόριθμο T-OPTICS που αναφέρεται στην εργασία [4]. Η μέθοδος αυτή χρησιμοποιεί ένα χρονικό κατώφλι για τον εντοπισμό των στάσεων και της κίνησης μέσα σε μία τροχιά. Η στάση αποτελεί υποτροχιά που περιορίζεται χωρικά σε μία συγκεκριμένη περιοχή και χαρακτηρίζεται από πολύ χαμηλή ταχύτητα. Ως κίνηση [MOVE] χαρακτηρίζεται ότι δεν αποτελεί στάση [STOP].

4.2 Συσχέτιση δεδομένων με σημασιολογική πληροφορία

Στις μελλοντικές επεκτάσεις του chorochronos.org θα μπορούσε να συμπεριληφθεί σημασιολογική πληροφορία για τις στάσεις που πραγματοποιεί το κινούμενο αντικείμενο, πχ για κινήσεις ανθρώπων οι στάσεις θα μπορούσαν να είναι το σπίτι, γραφείο, βιβλιοθήκη κα. Σημασιολογικές περιοχές είναι ένα σύνολο γεωγραφικών δεδομένων που έχουν κάποια σημασία. Για την συσχέτιση των τροχιών με τη σημασιολογική

πληροφορία θα μπορούσε να χρησιμοποιηθεί ο αλγόριθμος που προτείνεται στην εργασία [20]. Η σημασιολογική πληροφορία μπορεί να χωριστή σε περιοχές ενδιαφέροντος (Regions of Interest), διαδρομές-γραμμές ενδιαφέροντος (Lines of Interest) και σημεία ενδιαφέροντος (Points of Interest).

Προτείνεται η σημασιολογική ανάλυση μίας τροχιάς να πραγματοποιηθεί σε 3 επίπεδα. Σε κάθε επίπεδο θα μπορούσε να τοποθετηθεί πληροφορία για τα ακόλουθα:

Περιοχές Ενδιαφέροντος.

Το πρώτο επίπεδο θα μπορούσε να συσχετιστεί με τις σημασιολογικές περιοχές. Για την επίτευξη του αρκεί να υλοποιηθεί γεωγραφική σύνδεση (join) μεταξύ της τροχιάς και της περιοχής ενδιαφέροντος με τη χρήση του ελάχιστου περικλείον τετραγώνου της τροχιάς ή το κέντρο της, και την γεωγραφική περιοχή. Έπειτα μπορούν να συνδεθούν τα σημασιολογικά στοιχεία της περιοχής με την τροχιά.

Γραμμές Ενδιαφέροντος.

Για τον εντοπισμό των γραμμών ενδιαφέροντος προτείνεται να γίνεται εντοπισμός του σωστού τμήματος του οδικού δικτύου και σύνδεση με τον τρόπο μετακίνησης (πχ με μετρό, ποδήλατο κ.α.). Η συσχέτιση μπορεί να πραγματοποιηθεί για τα τμήματα των τροχιών που δηλώνεται κίνηση.

Αναλυτικά προτείνεται η ακόλουθη διαδικασία:

1. επιλογή υποψήφιου τμήματος οδικού δικτύου
2. υπολογισμός απόστασης σημείου –τμήματος
3. κανονικοποίηση της τοπικής απόστασης
4. υπολογισμός κανονικοποιημένης γενικής απόστασης
5. Ταίριασμα σημείων με τμήματα δικτύου
6. Σύνδεση με μεταφορικό μέσο, βασισμένο σε στοιχεία GPS και πληροφορίες για το είδος του τμήματος του οδικού δικτύου (πχ ποδηλατόδρομος).

Σημεία Ενδιαφέροντος.

Για την εύρεση της κατηγορίας του σημείου ενδιαφέροντος μπορεί να χρησιμοποιηθεί ένας αλγόριθμος βασισμένος στο Μαρκοβιανό μοντέλο (HMM) όπου αφού βρεθούν οι σχετικές πιθανότητες μοντελοποιείται το σημείο ενδιαφέροντος με τη χρήση Γκαουσιανής κατανομής και τεχνικών διακριτοποίησης και γειτνίασης για την βελτίωση της αποτελεσματικότητας του αλγορίθμου.

5 Συμπεράσματα

Η εξαγωγή πληροφορίας από τον μεγάλο όγκο δεδομένων που συλλέγονται καθημερινά σχετικά με την κίνηση μπορεί να γίνει με διάφορους τρόπους, αναλόγως το είδος την πληροφορίας που θέλουμε να ληφθεί. Λόγω του μεγέθους των δεδομένων απαιτείται από το επίπεδο της βάσης δεδομένων να υπάρχουν ειδικές δομές για την υποστήριξη τους. Το Hermes MOD έχει κατασκευαστεί γι' αυτή ακριβώς τη λειτουργία. Η διαδικτυακή εφαρμογή hermes.ChoroChronos.org χρησιμοποιεί την υποδομή του Hermes MOD για την ευκολότερη και αποδοτικότερη υλοποίηση των ερωτημάτων που γίνονται προς τη βάση.

Οι χρήστες μπορούν μέσω διαδικτύου, και χωρίς να χρειάζεται να υπάρχει κάποια εγκατάσταση στον υπολογιστή τους, να ανεβάσουν σε ειδικά διαμορφωμένη βάση που χρησιμοποιεί το Hermes MOD το σύνολο των δεδομένων που τους ενδιαφέρουν. Στη συνέχεια μπορούν να κάνουν δυναμικά ερωτήματα σε αυτό το σύνολο των δεδομένων ώστε να εξάγουν σχετική πληροφορία. Τα ερωτήματα μπορούν αν έχουν χωρικά ή χρονικά κριτήρια, ή και τα δύο. Το αποτέλεσμα των ερωτημάτων μπορεί να οπτικοποιηθεί σε χάρτη ή να εξαχθεί σε σχετικό αρχείο.

Ο χρήστης μπορεί και να εφαρμόσει συναρτήσεις απόστασης στις διάφορες τροχιές και να εξάγει το αποτέλεσμα τους σε αρχείο. Επιπλέον μπορεί για διάφορες συναρτήσεις απόστασης να εξάγει διάγραμμα γειτνίασης. Η ταξινόμηση των τροχιών γίνεται με την χρήση του αλγορίθμου T-Optics. Στη συνέχεια ο χρήστης έχει τη δυνατότητα να επιλέξει για διάφορες τιμές της ελάχιστης ποσοστιαίας διαφοράς της απόστασης γειτνίασης που μπορεί να έχουν δύο διαδοχικές τροχιές να εξάγει συστάδες.

Μελλοντικά μπορούν να υλοποιηθούν και διαδικασίες που να βοηθούν στον καθαρισμό των δεδομένων από το θόρυβο και να συσχετίζουν τα δεδομένα με σημασιολογική πληροφορία.

6 Βιβλιογραφία

- [1] Halil A., Çetin C. (2007), Geoportals and their Role in Spatial Data Infrastructures, www.researchgate.net/publication/228948121_Geoportals_and_their_role_in_spatial_data_infrastructures/file/79e415023b0d3f015c.doc
- [2] Tait M. G. (2005), Implementing Geoportals: Applications of Distributed GIS, *Computers, Environment and Urban Systems*, 29 (2005) 33–47.
- [3] Maguire, D. J., & Longley, P. A. (2005). The Emergence of Geoportals and Their Role in Spatial Data Infrastructures, *Computers, Environment and Urban Systems*, 29 (2005) 3–14.
- [4] Max Zimmermann, Thomas Kirste and Myra Spiliopoulou: Finding Stops in Error-Prone Trajectories of Moving Objects with Time-Based Clustering, *Intelligent Interactive Assistance and Mobile Multimedia Computing Communications in Computer and Information Science*, 2009, Volume 53, Part 8, 275-286
- [5] Mihael Ankerst, Markus M. Breunig, Hans-Peter Kriegel, Jörg Sander (1999). "OPTICS: Ordering Points to Identify the Clustering Structure". *ACM SIGMOD international conference on Management of data*. ACM Press. pp. 49–60.
- [6] H. Ding, G. Trajcevski, P. Scheuermann, X. Wang, and E. Keogh. Querying and Mining of Time Series Data: Experimental Comparison of Representations and Distance Measures. In *VLDB*, 2008
- [7] Chen, M. T. Oszu, and V. Oria. Robust and fast similarity search for moving object trajectories. In *SIGMOD Conference*, 2005
- [8] Zhixian Yan, Christine Parent, Stefano Spaccapietra, Dipanjan Chakraborty: A Hybrid Model and Computing Platform for Spatio-semantic Trajectories. *ESWC (1) 2010*: 60-75
- [9] E. Frenzos, K. Gratsias, and Y. Theodoridis. Index-based most similar trajectory search. In *ICDE*, 2007.
- [10] D. J. Berndt and J. Clifford. Using dynamic time warping to find patterns in time series. In *KDD Workshop*, 1994.
- [11] T Rakthanmanon, B Campana, A Mueen, G Batista, B Westover, Q Zhu, J Zakaria, E Keogh. Searching and mining trillions of time series subsequences under dynamic time warping. In *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*, p. 262-270, 2012
- [12] L. Chen, M. T. Oszu, and V. Oria. Robust and fast similarity search for moving object trajectories. In *SIGMOD Conference*, 2005.
- [13] M. Vlachos, M. Hadjieleftheriou, D. Gunopulos, and E. J. Keogh. Indexing Multidimensional Time-Series. *VLDBJ.*, 15(1), 2006.
- [14] M. Vlachos, D. Gunopulos, and G. Kollios. Discovering similar multidimensional trajectories. In *ICDE*, 2002.
- [15] http://perso.rd.francetelecom.fr/lemaire/CIDN/2013_AliceMarascu_compressed.pdf
- [16] L. Chen, M. T. Oszu, and V. Oria. Robust and fast similarity search for moving object trajectories. In *SIGMOD Conference*, 2005.

- [17] L. Chen and R. T. Ng. On the marriage of lp-norms and edit distance. In VLDB, 2004.
- [18] Mirco Nanni and Dino Pedreschi, Time-focused clustering of trajectories of moving objects, Journal of Intelligent Information Systems archive, Pages 267 – 289, November 2006
- [19] C. Zhang et al. / Journal of Information & Computational Science 10: 1 (2013) 193-199
- [20] Zhixian Yan, Dipanjan Chakraborty, Christine Parent, Stefano Spaccapietra, Karl Aberer: SeMiTri: a framework for semantic annotation of heterogeneous trajectories. EDBT 2011: 259-270
- [21] Fei Wuy, Tobias Kin, Hou Leiz, Zhenhui Liy, Jiawei Hanz, MoveMine 2.0: Mining Object Relationships from Movement Data, 40th International Conference on Very Large Data Bases, September 1st 5th, 2014, Hangzhou, China. Proceedings of the VLDB Endowment, Vol. 7, No. 13
- [22] Πτυχιακή Εργασία Φοιτητή Γιάννη Κωστή, Μέτρα Ταξινόμησης Χρονοσειρών, Πανεπιστήμιο Πειραιά
- [23] N. Pelekis, E. Stefanakis, I. Kopanakis, C. Zotali, M. Voudas, and Y. Theodoridis. "ChoroChronos.org: A GeoPortal for Movement Data and Processes", in the proceedings of the 10th International Conference on Spatial Information Theory (COSIT'11), Belfast, Maine, USA, 2011
- [24] N. Pelekis, E. Frenzos, N. Giatrakos, Y. Theodoridis: "HERMES: A Trajectory DB Engine for Mobility-Centric Applications", Int'l Journal of Knowledge-based Organizations (IJKBO), 5(3), 2015. IGI Global.
- [25] N. Pelekis, Y. Theodoridis, S. Vosinakis and T. Panayiotopoulos "Hermes - A Framework for Location-Based Data Management", In the Proceedings of the 10th International Conference on Extending Database Technology (EDBT06), LNCS 3896, pp. 1130-1134, Munich, Germany, 2006. Springer.
- [26] N. Pelekis. STAU: A spatio-temporal extension to ORACLE DBMS, PhD Thesis, UMIST, 2002.
- [27] Fosca Giannotti, Mirco Nanni, Dino Pedreschi, Fabio Pinelli, Chiara Renso, Salvatore Rinzivillo, Roberto Trasarti (2011), Unveiling the complexity of human mobility by querying and mining massive trajectory data, The VLDB Journal—The International Journal on Very Large Data Bases (Pages 695-719)
- [28] Zhenhui Li, Ming Ji, Jae-Gil Lee, Lu-An Tang, Yintao Yu, Jiawei Han, Roland Kays, MoveMine: Mining Moving Object Databases, SIGMOD'10, June 6–11, 2010, Indianapolis, Indiana, USA.
- [29] Marios Vontas, (2013), Msc Thesis Hermes – Building an Efficient Moving Object Database Engine, University of Piraeus

7 Συντομογραφίες

GPS: Global Positioning System (Παγκόσμιο Σύστημα Συντεταγμένων)

DISSIM: spatiotemporal dissimilarity algorithm (Χωροχρονικός αλγόριθμος ανομοιοότητας)

MOD: Mobile Object Database (Βάση δεδομένων Κινούμενων αντικειμένων)

DBMS: Database Management System (Σύστημα Διαχείρισης Βάσης Δεδομένων)

HermesMOD: Hermes MOD (Βάση δεδομένων κινούμενων αντικειμένων Ερμής)

TAU-TLL: TAU Temporal Literal Library Data Cartridge (Χρονική Βιβλιοθήκη Δεδομένων)

DDL: Data Definition Language (Γλώσσα Ορισμού Δεδομένων)

JSP: Java Server Pages (Σελίδες Εξυπηρετητή Java)

PL/SQL: Procedural Language/Structured Query Language (Διαδικαστική Γλώσσα/Δομημένο Ερώτημα Γλώσσας)

JDBC: Java Database Connectivity (Συνδεσιμότητα της Java με την βάση)

ODBC: Open Database Connectivity (Ανοικτή συνδεσιμότητα με την βάση)

ORDBMS: Object Oriented Database Management System (Σύστημα Διαχείρισης αντικειμενοστρεφούς βάσης)

JSON: JavaScript Object Notation (Αντικειμενοσυμβολισμός στη JavaScript)

CSV: Comma-separated values (τιμές που διαχωρίζονται με κόμμα)

UTC: Coordinated Universal Time (Διεθνές σύστημα ώρας)