

**ΕΠΙΧΕΙΡΗΣΙΑΚΟ ΠΡΟΓΡΑΜΜΑ «Ανταγωνιστικότητα Επιχειρηματικότητα και
Καινοτομία»**

**ΑΞΟΝΑΣ ΠΡΟΤΕΡΑΙΟΤΗΤΑΣ 03 «Ανάπτυξη επιχειρηματικότητας με Τομεακές
προτεραιότητες»**

ΔΡΑΣΗ «ΕΡΕΥΝΩ – ΔΗΜΙΟΥΡΓΩ – ΚΑΙΝΟΤΟΜΩ»

**LinkGeoML: Αυτοματοποιημένη και ακριβής
διασύνδεση γεωχωρικών δεδομένων με τη
χρήση μεθόδων μηχανικής μάθησης**

ΚΩΔΙΚΟΣ ΟΠΣ «5030745»



ΤΙΤΛΟΣ ΠΑΡΑΔΟΤΕΟΥ

**Π1.1: «Προδιαγραφή περιπτώσεων χρήσης, βασικών δεικτών
απόδοσης και συνόλου αναφοράς αξιολόγησης»**

Πακέτο Εργασίας	ΠΕ1: Μοντελοποίηση και ανάλυση απαιτήσεων
Υπεύθυνος Φορέας	Ερατοσθένης ΑΕ
Είδος Παραδοτέου	Αναφορά
Ενδεικτικός Μήνας Παράδοσης	M12
Ημερομηνία Παράδοσης	8/7/2019 (M12)



Με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης

ΙΣΤΟΡΙΚΟ ΑΛΛΑΓΩΝ

Έκδοση	Ημερομηνία	Εργασίες	Συγγραφείς
0.1	22/02/2019	Δομή και πίνακας περιεχομένων του παραδοτέου	Γιώργος Γιαννόπουλος (ΑΘ.), Δημήτριος Σκούτας (ΑΘ.), Νώντας Τσάκωνας (ΕΡ.)
0.2	11/05/2019	Πρώτη έκδοση της περιγραφής των σεναρίων χρήσης	Νώντας Τσάκωνας (ΕΡ.), Μανώλης Αλεξανδράκης (ΕΡ.)
0.3	24/05/2019	Διάφορες προσθήκες και βελτιώσεις	Γιώργος Γιαννόπουλος (ΑΘ.), Δημήτριος Σκούτας (ΑΘ.)
0.4	03/06/2019	Διάφορες προσθήκες και βελτιώσεις	Νώντας Τσάκωνας (ΕΡ.), Ιωάννης Μαρακάκης (ΕΡ.), Μιχάλης Αεράκης (ΓΕ.)
0.5	20/06/2019	Εσωτερικά επιθεωρημένη έκδοση	Δημήτριος Σκούτας (ΑΘ.), Θοδωρής Δαλαμάγκας (ΑΘ.), Γιώργος Γιαννόπουλος (ΑΘ.)
1.0	05/07/2019	Τελική έκδοση	Δημήτριος Σκούτας (ΑΘ.), Γιώργος Γιαννόπουλος (ΑΘ.), Νώντας Τσάκωνας (ΕΡ.)

ΠΙΝΑΚΑΣ ΣΥΝΤΜΗΣΕΩΝ

ΕΓΣΑ87	Ελληνικό Γεωδαιτικό Σύστημα Αναφοράς 1987
ΕΣΥΕ	Εθνική Στατιστική Υπηρεσία της Ελλάδος
ΣΕ	Σημείο Ενδιαφέροντος
TK	Ταχυδρομικός Κώδικας
GIS	Geographic Information System – Γεωγραφικό Πληροφοριακό Σύστημα
KPI	Key Performance Indicator – Βασικός Δείκτης Απόδοσης
LPIS	Land Parcel Identification System
OSM	OpenStreetMap

ΠΙΝΑΚΑΣ ΠΕΡΙΕΧΟΜΕΝΩΝ

1. ΕΙΣΑΓΩΓΗ	5
2. ΣΕΝΑΡΙΑ ΧΡΗΣΗΣ	7
2.1. ΣΧ1: Διασύνδεση Χωρο-κειμενικών Οντοτήτων	7
2.1.1. Περιγραφή προβλήματος	7
2.1.2. Εμπορική εφαρμογή	8
2.1.3. Τρέχουσα κατάσταση	8
2.1.4. Στόχος του προβλήματος.....	8
2.1.5. Σύνολα δεδομένων	9
2.1.6. Βασικοί Δείκτες Απόδοσης	9
2.2. ΣΧ2: Κατηγοριοποίηση Σημείων Ενδιαφέροντος	11
2.2.1. Περιγραφή προβλήματος	11
2.2.2. Εμπορική εφαρμογή	11
2.2.3. Τρέχουσα κατάσταση	12
2.2.4. Στόχος του προβλήματος.....	12
2.2.5. Σύνολα δεδομένων	13
2.2.6. Βασικοί Δείκτες Απόδοσης	14
2.3. ΣΧ3: Γεωκωδικοποίηση Διευθύνσεων	16
2.3.1. Περιγραφή προβλήματος	16
2.3.2. Εμπορική εφαρμογή	16
2.3.3. Τρέχουσα κατάσταση	17
2.3.4. Στόχος του προβλήματος.....	18
2.3.5. Σύνολα δεδομένων	18
2.3.6. Βασικοί Δείκτες Απόδοσης	21
2.4. ΣΧ4: Ολοκλήρωση Γεωτεμαχίων/Δρόμων	22
2.4.1. Περιγραφή προβλήματος	22
2.4.2. Εμπορική εφαρμογή	27
2.4.3. Τρέχουσα κατάσταση	27
2.4.4. Στόχος του προβλήματος.....	27
2.4.5. Σύνολα δεδομένων	28
2.4.6. Βασικοί Δείκτες Απόδοσης	29
3. ΣΥΝΟΨΗ	30

ΠΕΡΙΛΗΨΗ

Το παρόν έγγραφο αποτελεί το Παραδοτέο 1.1 του έργου, το οποίο περιγράφει το αποτέλεσμα των εργασιών της ΥΕ 1.1. Αρχικά καθορίζονται και εξειδικεύονται τα σενάρια χρήσης των αναπτυσσόμενων μεθόδων, λαμβάνοντας υπόψη τις απαιτήσεις και τους βασικούς δείκτες απόδοσης των επιχειρηματικών εταίρων. Επίσης, περιγράφονται τα σύνολα των δεδομένων και τα εμπορικά σενάρια εφαρμογής, ως προς τη διασύνδεση οντοτήτων γεωχωρικών δεδομένων καθώς και οι ακριβείς απαιτήσεις και τα επιθυμητά αποτελέσματα.

Το κάθε σενάριο χρήσης που περιγράφεται στο παραδοτέο δομείται ως εξής:

- Προσδιορισμός του προβλήματος.
- Περιγραφή της εμπορικής εφαρμογής του προβλήματος για τους επιχειρηματικούς εταίρους.
- Ανάλυση της τρέχουσας κατάστασης στο επιχειρησιακό περιβάλλον των εταίρων, όσον αφορά το πρόβλημα.
- Επεξήγηση του στόχου του έργου και των αναπτυσσόμενων νέων μεθόδων για το πρόβλημα.
- Παρουσίαση των συνόλων δεδομένων που αφορούν το πρόβλημα.
- Παράθεση των βασικών δεικτών απόδοσης της εφαρμογής των νέων μεθόδων στο πρόβλημα.

Τα βασικά σενάρια που αναλύονται στις αντίστοιχες ενότητες είναι τα εξής:

- Διασύνδεση Χώρο-κειμενικών Οντοτήτων (Διευθύνσεων, Σημείων Ενδιαφέροντος (ΣΕ), Τοπωνυμίων),
- Κατηγοριοποίηση ΣΕ,
- Γεωκωδικοποίηση Διευθύνσεων,
- Ολοκλήρωση και Διασύνδεση Γεωτεμαχίων και Δρόμων.

1. Εισαγωγή

Στην επιχειρηματική τους δραστηριότητα οι εταιρείες ΕΡΑΤΟΣΘΕΝΗΣ και GEODATA διαχειρίζονται και επεξεργάζονται ένα μεγάλο πλήθος γεωχωρικών¹ δεδομένων. Τα δεδομένα αυτά είτε αποτελούν ένα άμεσα εμπορεύσιμο προϊόν είτε χρησιμοποιούνται για τις ανάγκες έργων και μελετών που ανατίθενται στις δύο εταιρίες.

Βασικό πρόβλημα και απαίτηση τέτοιων έργων είναι η διασύνδεση, κατηγοριοποίηση, ταξινόμηση και ολοκλήρωση γεωχωρικών δεδομένων. Η διασύνδεση δεδομένων από μόνη της μπορεί να αποτελεί το καθεαυτού πρόβλημα ενός έργου ή να απαιτείται ως μία αρχική διεργασία που συνεισφέρει στην τελική λύση (π.χ. στην ολοκλήρωση ή κατηγοριοποίηση των δεδομένων). Ενδεικτικά, παραθέτουμε τα παρακάτω δύο παραδείγματα:

- 1) Μία απαίτηση για «καθαρισμό» ενός μεγάλου αρχείου με στοιχεία διευθύνσεων (π.χ. οι διευθύνσεις ενός πελατολογίου) απαιτεί τη διασύνδεση όλων των εγγραφών του αρχείου μεταξύ τους, για την παραγωγή ενός «καθαρού» και μοναδικοποιημένου-ενοποιημένου αρχείου διευθύνσεων πελατών. Στο παράδειγμα αυτό η διασύνδεση δεδομένων είναι το βασικό πρόβλημα του έργου.
- 2) Μία απαίτηση για την συμπλήρωση ή διόρθωση μίας ιδιότητας (attribute) πεδίου κατηγοριοποίησης ενός συνόλου ΣΕ μπορεί να απαντηθεί μέσω της διασύνδεσης των σημείων αυτών με κάποιο άλλο υποσύνολο ΣΕ, όπου η ιδιότητα αυτή είναι διαθέσιμη. Στο παράδειγμα αυτό η διασύνδεση δεδομένων λειτουργεί ως συμπληρωματικό και επιβοηθητικό πρόβλημα του έργου.

Τα βασικά σενάρια που θα εξετάσουμε στα πλαίσια του παρόντος έργου και τα οποία εμπεριέχουν διαδικασία διασύνδεσης, είτε ως αυτοτελές πρόβλημα είτε ως μέρος ενός προβλήματος ολοκλήρωσης, κατηγοριοποίησης, επισημείωσης, ταξινόμησης ή επιλογής, είναι τα ακόλουθα:

- Η διασύνδεση Χώρο-κειμενικών Οντοτήτων (Διευθύνσεων, ΣΕ, Τοπωνυμίων),
- Η Κατηγοριοποίηση ΣΕ,
- Η Γεωκωδικοποίηση Διευθύνσεων,
- Η Ολοκλήρωση και Διασύνδεση Γεωτεμαχίων και Δρόμων.

Τα παραπάνω σενάρια αποτελούν πραγματικά προβλήματα και απασχολούν έργα που υλοποιούν οι εταιρείες ΕΡΑΤΟΣΘΕΝΗΣ και GEODATA, όπως τεκμηριώνεται αναλυτικά στην επόμενη ενότητα, ξεχωριστά για το κάθε σενάριο χρήσης.

¹ Στην πλειονότητα των περιπτώσεων που αφορούν το έργο, αλλά και γενικότερα εφαρμογές διαχείρισης και ολοκλήρωσης γεωχωρικών δεδομένων, τα δεδομένα, πέρα από τις γεωχωρικές ιδιότητές τους, συνοδεύονται και από κάποια μορφή κειμενική πληροφορία (όνομα, όνομα ιδιοκτήτη, περιγραφή, κτλ.). Για αυτό, οι όροι «γεωχωρικό» και «χωρο-κειμενικό» χρησιμοποιούνται εναλλάξ στο παρόν παραδοτέο, έχοντας την ίδια σημασία.

Οι διαδικασίες της διασύνδεσης, κατηγοριοποίησης, ταξινόμησης και ολοκλήρωσης στα παραπάνω σενάρια, στην υπάρχουσα κατάσταση, γίνονται χειροκίνητα ή στην καλύτερη περίπτωση ημιαυτόματα με την χρήση διαφόρων GIS εργαλείων λογισμικού. Ωστόσο, αυτό είναι ανεπιθύμητο διότι έχει τις εξής αρνητικές συνέπειες:

- Μεγάλη καθυστέρηση στην υλοποίηση των έργων
- Σπατάλη ανθρώπινων πόρων
- Μειωμένη ποιότητα αποτελέσματος (παράγοντας ανθρώπινου λάθους)
- Αύξηση συνολικά του κόστους των έργων κάνοντας τα πολλές φορές απαγορευτικά για τους υποψήφιους πελάτες.

Δεδομένων των παραπάνω, η βασική απαίτηση που προκύπτει είναι η ανάπτυξη κατάλληλων εργαλείων μηχανικής μάθησης, τα οποία θα ενσωματωθούν στις παραπάνω επιχειρησιακές διαδικασίες (σενάρια), με σκοπό να αυτοματοποιήσουν τις διαδικασίες διασύνδεσης, κατηγοριοποίησης, ταξινόμησης και ολοκλήρωσης των δεδομένων και, κατ' επέκταση, να μειώσουν τους ανασταλτικούς παράγοντες των χειροκίνητων διαδικασιών που εφαρμόζονται έως σήμερα από τις δύο εταιρείες.

Συμπερασματικά, οι απαιτήσεις χρήσης, οι οποίες αφορούν το σύνολο των υφιστάμενων διαδικασιών και υπηρεσιών που φιλοδοξεί να επεκτείνει και να βελτιώσει το έργο, είναι δύο και σαφείς: *αυτοματοποίηση* και *υψηλή ακρίβεια* αποτελεσμάτων. Οι συγκεκριμένες απαιτήσεις εξειδικεύονται και ποσοτικοποιούνται περαιτέρω για κάθε σενάριο, με βάση τους επιμέρους βασικούς δείκτες απόδοσης (KPIs) και τις αντίστοιχες μετρικές αξιολόγησης ακρίβειας, για το εκάστοτε πρόβλημα μηχανικής μάθησης που εφαρμόζουμε.

Στην ακόλουθη ενότητα περιγράφονται αναλυτικότερα τα σενάρια χρήσης του έργου.

2. Σενάρια Χρήσης

2.1. ΣΧ1: Διασύνδεση Χωρο-κειμενικών Οντοτήτων

Πεδίο Εφαρμογής: Γεωκωδικοποίηση, Γεωχωρική Ανάλυση, Κτηματογράφηση

Εμπλεκόμενοι Φορείς: Ερατοσθένης, Geodata

2.1.1. Περιγραφή προβλήματος

Η διασύνδεση οντοτήτων συνίσταται στην αναγνώριση, από δύο ή περισσότερες πηγές (σύνολα) δεδομένων, των οντοτήτων εκείνων που είναι ίδιες, δηλαδή που αντιστοιχούν στις ίδιες οντότητες πραγματικού κόσμου. Στη διαχείριση γεωχωρικών δεδομένων, το πρόβλημα εξειδικεύεται στη διασύνδεση:

- Τοπωνυμίων
- Ταχυδρομικών Διευθύνσεων και
- Σημείων Ενδιαφέροντος (ΣΕ)

Οι παραπάνω οντότητες χαρακτηρίζονται ως χωρο-κειμενικές, διότι περιλαμβάνουν χαρακτηριστικά πεδία (attributes) τόσο περιγραφικά, όσο και γεωμετρίας. Κάθε οντότητα από τις παραπάνω παρουσιάζει διαφορετικά χαρακτηριστικά και χρειάζεται διαφορετικό και εξειδικευμένο χειρισμό στη διασύνδεση τους. Για το λόγο αυτό περιγράφονται στη συνέχεια χωριστά.

2.1.1.1. Διασύνδεση Τοπωνυμίων

Μια οντότητα Τοπωνυμίου περιλαμβάνει συνήθως τα παρακάτω χαρακτηριστικά (attributes):

- Όνομα Τοπωνυμίου
- Τυχόν Εναλλακτικά ονόματα Τοπωνυμίου
- Προαιρετικά, Γεωμετρία Τοπωνυμίου, σημειακή (x, y) ή πολυγωνική

Η διασύνδεση διαφορετικών πηγών δεδομένων Τοπωνυμίων δύναται να συμπληρώσει και να διορθώσει τα χαρακτηριστικά ενός τοπωνυμίου, για παράδειγμα εμπλουτίζοντας την περιγραφή του με εναλλακτικά ονόματα ή καθιστώντας ακριβέστερη τη θέση του.

2.1.1.2. Διασύνδεση Ταχυδρομικών Διευθύνσεων

Μια οντότητα διεύθυνσης χαρακτηρίζεται από τα παρακάτω πεδία (attributes):

- Οδός
- Αριθμός
- Ταχυδρομικός Κώδικας (ΤΚ)
- Πόλη
- Νομός
- Συντεταγμένες Θέσης (x, y)

Κάποια από τα παραπάνω πεδία μιας πηγής δεδομένων διευθύνσεων μπορεί να μην είναι γνωστά. Με την διασύνδεση διαφορετικών πηγών δεδομένων διευθύνσεων, δύναται να συμπληρωθούν ή να διορθωθούν κάποια από τα πεδία μιας διεύθυνσης και κατά αυτόν τον τρόπο να προκύψει μια ακριβέστερη και πληρέστερη διεύθυνση.

2.1.1.3. Διασύνδεση Σημείων Ενδιαφέροντος

Η οντότητα ΣΕ περιλαμβάνει συνήθως τα παρακάτω χαρακτηριστικά (attributes):

- Όνομα ΣΕ
- Κατηγορία
- Τυχόν Υποκατηγορίες
- Σημειακή θέση ΣΕ (x, y)

Η διασύνδεση διαφορετικών πηγών ΣΕ δύναται να συμπληρώσει και να διορθώσει τα χαρακτηριστικά ενός συνόλου ΣΕ.

2.1.2. Εμπορική εφαρμογή

Η εμπορική εφαρμογή του προβλήματος αφορά τη διασύνδεση διαφορετικών πηγών δεδομένων Τοπωνυμίων, Διευθύνσεων και ΣΕ με σκοπό τη συμπλήρωση και διόρθωση των δεδομένων του Πελάτη ή για την αξιοποίησή τους σε ευρύτερα έργα και μελέτες χωρικής ανάλυσης και Κτηματογράφησης, που αναλαμβάνουν οι εταιρείες GEODATA και ΕΡΑΤΟΣΘΕΝΗΣ.

Ενδεικτικά, τέτοια έργα αφορούν μελέτες χωρικής ανάλυσης μιας γεωγραφικής περιοχής που άπτονται διαφόρων τομέων, όπως η αγορά ακινήτων, οι ασφαλιστικές μελέτες, η αστική ανάπτυξη και κινητικότητα, εμπορικές μελέτες για πελατολόγια, νέες αγορές, κ.λ.π, περιβαλλοντολογικές μελέτες, κτηματογράφηση, κ.α.

Οι πελάτες των παραπάνω έργων είναι είτε δημόσιοι οργανισμοί (Δήμοι, Υπουργία, ΝΠΔΔ), είτε ιδιωτικοί φορείς (τράπεζες, ασφαλιστικές, τηλεπικοινωνιακές, μεσίτες, κ.α.).

2.1.3. Τρέχουσα κατάσταση

Προκειμένου να υλοποιήσουν τα παραπάνω έργα οι εταιρείες GEODATA και ΕΡΑΤΟΣΘΕΝΗΣ, επεξεργάζονται και διασυνδέουν μεγάλα σύνολα δεδομένων Τοπωνυμίων, Διευθύνσεων και ΣΕ των πελατών τους, με τα δικά τους αντίστοιχα σύνολα. Η επεξεργασία και διασύνδεση γίνεται χειροκίνητα, με την έννοια ότι κάθε γεωχωρική οντότητα εξετάζεται και συγκρίνεται με τις υπόλοιπες οντότητες από το χειριστή του συστήματος ξεχωριστά, με χρήση λογισμικών GIS όπως: QGIS και ArcGIS και εργαλείων σχεσιακών βάσεων Δεδομένων (PostgreSQL, Oracle). Αυτόματη ή ημιαυτόματη διαδικασία διασύνδεσης δεν εφαρμόζεται.

2.1.4. Στόχος του προβλήματος

Η παραπάνω διαδικασία διασύνδεσης διαφορετικών συνόλων δεδομένων γεωχωρικών οντοτήτων, αποτελεί μια κοπιώδη, αργή και ακριβή διαδικασία, η οποία γίνεται πανάκριβη και αποτρεπτική για τους πελάτες, ειδικά όταν πρόκειται για μεγάλα σύνολα δεδομένων. Επομένως, ο στόχος του προβλήματος είναι η ανάπτυξη κατάλληλων εργαλείων με χρήση αλγορίθμων μηχανικής μάθησης, τα οποία θα χρησιμοποιούνται για τη διασύνδεση

διαφορετικών πηγών δεδομένων Τοπωνυμίων/Διευθύνσεων/ΣΕ, ελαχιστοποιώντας την ανάγκη για χειροκίνητη επεξεργασία και διασύνδεση.

2.1.5. Σύνολα δεδομένων

Οι εταιρείες ΕΡΑΤΟΣΘΕΝΗΣ και GEODATA διατηρούν και επεξεργάζονται μεγάλα σύνολα δεδομένων και των τριών κατηγοριών (<http://www.eranet.gr/geodata/el/gproducts.html>), Τοπωνυμίων/Διευθύνσεων/ΣΕ τα οποία χρησιμοποιούν ως πηγές αναφοράς για την διασύνδεση και διόρθωση άλλων πηγών δεδομένων. Παράλληλα, εκμεταλλεύονται ανοικτά σύνολα δεδομένων, όπως το OpenStreetMap² (OSM) το οποίο αποτελεί μία εκτεταμένη πηγή ΣΕ και τοπωνυμίων, καθώς και το Geonames³, το οποίο επίσης αποτελεί μία ευρέως χρησιμοποιούμενη πηγή τοπωνυμίων.

2.1.6. Βασικοί Δείκτες Απόδοσης

Στον παρακάτω πίνακα παρουσιάζονται οι εκτιμώμενοι βασικοί δείκτες απόδοσης (KPIs) για το σενάριο της διασύνδεσης γεωχωρικών οντοτήτων διαφόρων ειδών (Τοπωνύμια/Διευθύνσεις/ΣΕ). Ο πρώτος δείκτης αφορά το ποσοστό της διαδικασίας διασύνδεσης που επιτελείται με χειροκίνητο, έναντι αυτοματοποιημένου τρόπου. Ο δεύτερος δείκτης αφορά την ανθρωπο-προσπάθεια που απαιτείται να ξοδευτεί, συνολικά ανά διαδικασία-εργασία διασύνδεσης, ενώ ο τρίτος δείκτης αφορά τη συνολική χρονική διάρκεια της εκάστοτε διαδικασίας. Ο τέταρτος δείκτης αφορά στην ακρίβεια διασύνδεσης που επιτυγχάνεται από τις μεθόδους αυτοματοποιημένης διασύνδεσης που θα υλοποιηθούν.

Σε κάθε δείκτη απόδοσης, αναθέτουμε μία ενδεικτική *επιθυμητή τιμή* την οποία στοχεύουμε να ικανοποιούν οι μέθοδοι που θα υλοποιηθούν μέχρι τη λήξη του έργου. Σημειώνουμε ότι διαφορετικά υπο-σενάρια διασύνδεσης παρουσιάζουν διαφορετικές απαιτήσεις και δυσκολίες, αφού αφορούν και αρκετά ετερογενή σύνολα δεδομένων (Τοπωνύμια/Διευθύνσεις/ΣΕ). Οπότε, οι θεωρημένες επιθυμητές τιμές αποτελούν, σε αυτό το στάδιο του έργου, ενδεικτικά μεγέθη, τα οποία δύναται να αναθεωρηθούν κατά τη διάρκεια της τελικής αξιολόγησης των υλοποιημένων μεθόδων.

α/α	Δείκτης	Επιθυμητή τιμή
1	Ποσοστό μείωσης χειροκίνητης διασύνδεσης σε σχέση με την υπάρχουσα κατάσταση	60%
2	Ποσοστό μείωσης ανθρωπο-προσπάθειας διασύνδεσης σε σχέση με την υπάρχουσα κατάσταση	50%
3	Ποσοστό μείωσης της χρονικής διάρκειας της διαδικασίας διασύνδεσης σε σχέση με την υπάρχουσα κατάσταση	50%

² <https://www.openstreetmap.org>

³ <https://www.geonames.org/>

4	Ποσοστό ακρίβειας αυτοματοποιημένης διασύνδεσης	80%
---	---	-----

Πίνακας 1: Βασικοί Δείκτες Απόδοσης στη Διασύνδεση

2.2. ΣΧ2: Κατηγοριοποίηση Σημείων Ενδιαφέροντος

Πεδίο Εφαρμογής: Γεωκωδικοποίηση, Γεωχωρική Ανάλυση

Εμπλεκόμενοι Φορείς: Geodata, Ερατοσθένης

2.2.1. Περιγραφή προβλήματος

Η Συλλογή και κατηγοριοποίηση ΣΕ αποτελεί βασικό επιχειρησιακό αντικείμενο των εταιρειών GEODATA και ΕΡΑΤΟΣΘΕΝΗΣ. Ως ΣΕ μπορεί να χαρακτηριστεί οποιοδήποτε σημείο στο γεωγραφικό χώρο το οποίο μπορεί να αναπαρασταθεί με βάση τις συντεταγμένες του (x, y), το όνομά του, καθώς και διάφορα πεδία ιδιοτήτων (attributes) και κατηγοριών (categories), που το χαρακτηρίζουν και το ταυτοποιούν. Ο καθορισμός των κατηγοριών (κατηγοριοποίηση), σε ένα σύνολο ΣΕ, έχει να κάνει συνήθως με την επιχειρησιακή τους χρήση και τις διάφορες γεωχωρικές αναλύσεις που χρησιμοποιούνται στα πλαίσια μιας μελέτης ή ενός γενικότερου έργου. Συνήθως, τηρείται μια βασική (master) έκδοση ΣΕ με κάποια σταθερά πεδία ιδιοτήτων ή/και κατηγοριών, η οποία εμπλουτίζεται, ανάλογα με τις ανάγκες και τις απαιτήσεις που προκύπτουν, μέσα από τα διάφορα έργα και μελέτες που αναλαμβάνουν να διεκπεραιώσουν οι δύο εταιρείες.

Συνοπτικά οι εργασίες που επιτελούνται σε σχέση με τα ΣΕ, είναι:

- Συλλογή, είτε με αποτύπωση στο πεδίο (με χρήση GPS), είτε με χρήση έτοιμων ανοιχτών πηγών δεδομένων.
- Επεξεργασία, όσον αφορά τη διόρθωση των υφιστάμενων πεδίων κατηγοριών τους, αλλά και εμπλουτισμός με επιπλέον πεδία για τις ειδικές ανάγκες ενός έργου. Συνήθως, σε αυτήν την περίπτωση, έχουμε να κάνουμε με βάσεις δεδομένων ΣΕ πελατών που απαιτούν διόρθωση και εμπλουτισμό των πεδίων κατηγοριών με νέα.

Το επιχειρησιακό πρόβλημα που προκύπτει για τις εταιρείες, όσον αφορά την επεξεργασία των κατηγοριών των ΣΕ, έχει να κάνει με την ανάπτυξη ενός εργαλείου το οποίο θα υποβοηθά και θα ελαχιστοποιεί την ανθρώπινη παρέμβαση, στη διόρθωση και ανάπτυξη νέων πεδίων κατηγοριοποίησης σε ένα σύνολο ΣΕ.

2.2.2. Εμπορική εφαρμογή

Η εμπορική εφαρμογή του προβλήματος αφορά την κατηγοριοποίηση (σε διάφορες κατηγορίες ενδιαφέροντος) μεγάλων συνόλων δεδομένων ΣΕ, για την αξιοποίησή τους σε ευρύτερα έργα και μελέτες χωρικής ανάλυσης που αναλαμβάνουν οι εταιρείες GEODATA και ΕΡΑΤΟΣΘΕΝΗΣ.

Τέτοια έργα, ενδεικτικά, είναι μελέτες και έργα χωρικής ανάλυσης μιας γεωγραφικής περιοχής που άπτονται των τομέων της αγοράς ακινήτων, των ασφαλιστικών μελετών, έργων αστικής ανάπτυξης και κινητικότητας, εμπορικών μελετών (πελατολόγια, νέες αγορές, κ.λ.π), περιβαλλοντολογικών μελετών κ.α.

Οι πελάτες των παραπάνω έργων είναι είτε δημόσιοι οργανισμοί (Δήμοι, Υπουργία, ΝΠΔΔ), είτε ιδιωτικοί φορείς (τράπεζες, ασφαλιστικές, τηλεπικοινωνιακές, real estate, κ.α.).

2.2.3. Τρέχουσα κατάσταση

Προκειμένου να υλοποιηθούν τα παραπάνω έργα οι εταιρείες GEODATA και ΕΡΑΤΟΣΘΕΝΗΣ συλλέγουν και επεξεργάζονται μεγάλα σύνολα ΣΕ. Η συλλογή και επεξεργασία γίνεται με χρήση κινητών συνεργείων με συστήματα εντοπισμού θέσης (GPS) και με αξιοποίηση διάφορων ανοιχτών πηγών δεδομένων.

Η επεξεργασία των ΣΕ αφορά την κατηγοριοποίησή τους σε διάφορες κατηγορίες και υποκατηγορίες ανάλογα με τις απαιτήσεις του έργου και του πελάτη. Παραδείγματος χάρη, τα ΣΕ του παρακάτω πίνακα χαρακτηρίζονται από το όνομα (name) και την γεωγραφική τους θέση (x,y) σε ΕΓΣΑ87 και ζητείται να κατηγοριοποιηθούν σε τρία (3) επίπεδα κατηγοριοποίησης, με αυξανόμενη εξειδίκευση των κατηγοριών: theme, class_name και subclass_n.

poi_id	name	theme	class_name	subclass_n	x	y
1	ΑΒ ΒΑΣΙΛΟΠΟΥΛΟΣ	Αγαθά - υπηρεσίες	Καταστήματα τροφίμων	Σουπερμάρκετ	4.813.199.089	42.082.857.605
2	NEOGENESIS ΕΠΕ	Περίθαλψη	Κλινικές	Ιδιωτικές κλινικές	4.813.530.814	42.083.222.381
3	ΦΑΡΜΑΚΕΙΟ ΣΠΑΝΟΥΔΗ ΣΟΦΙΑ & ΣΙΑ ΟΕ	Περίθαλψη	Φαρμακεία	_	4.811.643.399	42.083.806.565
4	ΚΕΝΤΡΟ ΑΝΘΡΩΠΙΝΗΣ ΑΝΑΠΑΡΑΓΩΓΗΣ	Περίθαλψη	Κλινικές	Ιδιωτικές κλινικές	4.813.580.307	42.083.939.571
5	DOMINO'S PIZZA	Φαγητό και ποτό	Εστιατόρια	Πιτσαρίες	4.813.567.993	42.084.420.986
6	REINA	Διασκέδαση	Κλαμπ	_	4.813.567.993	42.084.420.986
7	BP	Αυτοκίνηση	Πρατήρια βενζίνης	_	4.810.307.661	42.085.045.856
8	PARKING	Αυτοκίνηση	Χώροι στάθμευσης	_	4.813.134.896	42.085.280.574
9	ΖΥΘΟΣ	Φαγητό και ποτό	Εστιατόρια	Μπυραρίες	4.813.736.034	42.085.381.299
10	NAUTICA	Αγαθά - υπηρεσίες	Καταστήματα ένδυσης	_	4.809.604.393	42.085.478.786
11	MEDICANALYSIS ΕΠΕ	Περίθαλψη	Κλινικές	Ιδιωτικές κλινικές	4.812.475.374	4.208.573.429

Πίνακας 2: Ενδεικτικά δεδομένα Σημείων Ενδιαφέροντος

Η παραπάνω διαδικασία κατηγοριοποίησης γίνεται χειροκίνητα και με έρευνα (με καταγραφή πεδίου ή από τρίτες πηγές πληροφορίας) για τα διάφορα ΣΕ με χρήση λογισμικών GIS όπως: QGIS και ArcGIS. Δηλαδή δεν υπάρχει υλοποιημένη αυτή τη στιγμή κάποια διαδικασία ημιαυτόματης υποβοήθησης ή πλήρους αυτοματοποιημένης κατηγοριοποίησης των διαθέσιμων ΣΕ.

2.2.4. Στόχος του προβλήματος

Η παραπάνω χειροκίνητη κατηγοριοποίηση ενός συνόλου ΣΕ, αποτελεί μια κοπιώδη και ακριβή (ειδικά όταν απαιτεί καταγραφές πεδίου) διαδικασία, η οποία γίνεται πανάκριβη και αποτρεπτική για τους πελάτες, όταν πρόκειται για μεγάλα σύνολα ΣΕ και κυρίως όταν

απαιτείται προσθήκη νέων πεδίων κατηγοριών και υποκατηγοριών, που πιθανόν να απαιτεί ένα έργο.

Άρα ο στόχος του προβλήματος είναι η ανάπτυξη κατάλληλου εργαλείου, το οποίο θα επιτρέπει την επιλογή και εκπαίδευση (training) κατάλληλων αλγορίθμων μηχανικής μάθησης σε ένα αξιόπιστο και μικρό σύνολο ΣΕ (δεδομένα εκπαίδευσης), το οποίο θα έχει προκύψει από επιτόπια αποτύπωση ή από χειροκίνητη διερεύνηση και επιβεβαίωση της αξιοπιστίας του. Στη συνέχεια, οι εκπαιδευμένοι αλγόριθμοι θα μπορούν να εφαρμοστούν σε μεγαλύτερα σύνολα δεδομένων και να επισημειώσουν (κατηγοριοποιήσουν), αυτόματα ή ημιαυτόματα (επιλογή από προτεινόμενες τιμές), νέα ΣΕ με κατηγορίες και υποκατηγορίες.

Το παραπάνω εργαλείο θα ενσωματωθεί στα GIS εργαλεία επεξεργασίας ΣΕ με σκοπό την αυτόματη ή ημιαυτόματη κατηγοριοποίηση νέων συνόλων δεδομένων ΣΕ.

2.2.5. Σύνολα δεδομένων

Ενδεικτικά σύνολα ΣΕ τα οποία συλλέγουν, τηρούν και επεξεργάζονται οι εταιρείες GEODATA και ΕΡΑΤΟΣΘΕΝΗΣ είναι διαθέσιμα στην ακόλουθη ιστοσελίδα: <http://www.eranet.gr/geodata/el/gpoi.html>. Μία ενδεικτική κατηγοριοποίηση που χρησιμοποιείται, η οποία περιέχει δύο επίπεδα κατηγοριών, παρουσιάζεται στον ακόλουθο πίνακα:

1.Περίθαλψη	8.Εκπαίδευση
1.1.Ιδρύματα	8.1.Πρωτοβάθμια
1.2.Κλινικές	8.2.Δευτεροβάθμια
1.3.Νοσοκομεία	8.3.Τριτοβάθμια
	8.4.Εκπαιδευτήρια
2.Αθλητισμός & Αναψυχή	8.5.Παιδικοί σταθμοί
2.1.Αθλητικές εγκαταστάσεις	
2.2.Αθλητικοί όμιλοι	9.Επιχειρήσεις
2.3.Δημόσιοι χώροι	9.1.Δημόσιες επιχειρήσεις
2.4.Χώροι αναψυχής	9.2.Επιμελητήρια
	9.3.Επιχειρήσεις
3.Αυτοκίνηση	9.4.Τράπεζες
3.1.Εμπόριο αυτοκινήτων	
3.2.Ενοικιάσεις αυτοκινήτων	10.Αγαθά – Υπηρεσίες
3.3.Συνεργεία αυτοκινήτων	10.1.Εμπορικά κέντρα
3.4.Χώροι στάθμευσης	10.2.Εμπορικές υπηρεσίες
3.5.Πρατήρια βενζίνης	10.3.Καταστήματα ένδυσης
	10.4.Καταστήματα τροφίμων
4.Μετακίνηση	10.5.Καταστήματα
4.1.Αεροδρόμια	10.6.Περίπτερα

4.2.Σταθμοί μετρό	
4.3.Σταθμοί τράμ	
4.4.Σταθμοί προαστιακού	
4.5.Σταθμοί ΟΣΕ	
4.6.Σταθμοί ΚΤΕΛ	
5.Δημόσιος Τομέας	
5.1.Αστυνομία	
5.2.Δημόσιες Υπηρεσίες	
5.3.Κρατικά κτήρια (Δικαστήρια, Στρατόπεδα, Τελωνεία, Φυλακές)	
5.4.Λιμεναρχεία	
5.5.Πυροσβεστική	
5.6.Πρεσβείες	
5.7.Ταχυδρομεία	
6.Φαγητό – Ποτό	
6.1.Γρήγορο φαγητό	
6.2.Εστιατόρια	
6.3.Καφετέριες	
6.4.Ταβέρνες	
7.Διασκέδαση	
7.1.Ζωντανή μουσική	
7.2.Κέντρα διασκέδασης	
7.3.Κλαμπ	
7.4.Μπαρ	
7.5.Καζίνο	
	11.Πολιτισμός
	11.1.Θέατρα
	11.2.Κινηματογράφοι
	11.3.Πολιτιστικά κέντρα
	11.4.Συναυλιακοί χώροι
	11.5.Χώροι τέχνης
	11.6.Βιβλιοθήκες
	12.Τουρισμός Αξιοθέατα
	12.1.Αξιοθέατα
	12.2.Αρχαιολογικοί Χώροι
	12.3.Καταλύματα
	12.4.Μουσεία
	12.5.Υπηρεσίες Τουρισμού
	12.6.Παραλίες
	13.Διάφορα
	13.1.Θρησκευτικοί χώροι
	13.2.Χώροι εκδηλώσεων

Πίνακας 3: Κατηγοριοποίηση Σημείων Ενδιαφέροντος

2.2.6. Βασικοί Δείκτες Απόδοσης

Στον παρακάτω πίνακα παρουσιάζονται οι εκτιμώμενοι βασικοί δείκτες απόδοσης (KPIs) για το σενάριο της κατηγοριοποίησης ΣΕ. Ο πρώτος δείκτης αφορά το ποσοστό της διαδικασίας κατηγοριοποίησης που επιτελείται με χειροκίνητο, έναντι αυτοματοποιημένου τρόπου. Ο δεύτερος δείκτης αφορά την ανθρωπο-προσπάθεια που απαιτείται να ξοδευτεί, συνολικά ανά διαδικασία-εργασία κατηγοριοποίησης, ενώ ο τρίτος δείκτης αφορά τη συνολική χρονική διάρκεια της εκάστοτε διαδικασίας. Ο τέταρτος δείκτης αφορά στην

ακρίβεια κατηγοριοποίησης που επιτυγχάνεται από τις μεθόδους αυτοματοποιημένης κατηγοριοποίησης που θα υλοποιηθούν.

Σε κάθε δείκτη απόδοσης, αναθέτουμε μία ενδεικτική επιθυμητή τιμή την οποία στοχεύουμε να ικανοποιούν οι μέθοδοι που θα υλοποιηθούν, μέχρι τη λήξη του έργου.

α/α	Δείκτης	Επιθυμητή τιμή
1	Ποσοστό μείωσης χειροκίνητης κατηγοριοποίησης σε σχέση με την υπάρχουσα κατάσταση	60%
2	Ποσοστό μείωσης ανθρωπο-προσπάθειας κατηγοριοποίησης σε σχέση με την υπάρχουσα κατάσταση	40%
3	Ποσοστό μείωσης της διάρκειας κατηγοριοποίησης σε σχέση με την υπάρχουσα κατάσταση	60%
4	Ποσοστό ακρίβειας κατηγοριοποίησης	70%

Πίνακας 4: Βασικοί Δείκτες Απόδοσης στην Κατηγοριοποίηση

2.3. ΣΧ3: Γεωκωδικοποίηση Διευθύνσεων

Πεδίο Εφαρμογής: Γεωκωδικοποίηση, Γεωχωρική Ανάλυση

Εμπλεκόμενοι Φορείς: Geodata, Ερατοσθένης

2.3.1. Περιγραφή προβλήματος

Παρεμφερές πρόβλημα της κατηγοριοποίησης ΣΕ αποτελεί η διαδικασία της Γεωκωδικοποίησης Διευθύνσεων. Με τον όρο Γεωκωδικοποίηση (Geocoding) Διευθύνσεων εννοούμε την αντιστοίχιση κάθε διεύθυνσης με ένα σημείο συντεταγμένων (x,y). Μια διεύθυνση στην αρχική της μορφή περιγράφεται από όλα ή μερικά πεδία πληροφορίας, από τα παρακάτω:

- Διοικητική διαίρεση: Νομός, Δήμος, Δημοτικό διαμέρισμα (οικισμός)
- Τοπωνύμιο Περιοχής (location)
- Ταχυδρομικός Κωδικός (TK), Οδός, Αριθμός (από – έως)

Οι εταιρείες GEODATA – ΕΡΑΤΟΣΘΕΝΗΣ έχουν αναπτύξει εργαλεία λογισμικού για την αυτόματη ή ημιαυτόματη γεωκωδικοποίηση μεγάλων αρχείων διευθύνσεων, που έχουν να κάνουν κυρίως με μεγάλα πελατολόγια πελατών τους.

Τα εργαλεία αυτά χρησιμοποιούν μια βάση αναφοράς (master database) περιγραφικών στοιχείων διευθύνσεων (Διοικητικές διαιρέσεις, τοπωνύμια, TK, ονόματα οδών), προκειμένου, σε πρώτη φάση, να κάνουν καθαρισμό των περιγραφικών στοιχείων μιας δοθείσας διεύθυνσης και, σε δεύτερη φάση, να προχωρήσουν στη γεωκωδικοποίησή τους. Η γεωκωδικοποίηση πραγματοποιείται είτε με μεθοδολογία «γραμμικής παρεμβολής» με τη χρήση του γεωγραφικού επίπεδου γραμμών οδικού δικτύου, είτε με διασύνδεση με πίνακα προκαθορισμένων σημειακών διευθύνσεων με συντεταγμένες (x, y).

Το πρόβλημα που προκύπτει από την παραπάνω διαδικασία γεωκωδικοποίησης έχει να κάνει με την αξιοπιστία του τελικού αποτελέσματος, καταρχάς στον καθαρισμό των περιγραφικών στοιχείων μιας διεύθυνσης και κατόπιν στη γεωμετρική ακρίβεια των συντεταγμένων (x, y) που της αποδίδεται. Η ακρίβεια αυτή ελέγχεται ως προς την αξιοπιστία της, διότι:

- Πιθανόν να υπάρχουν σφάλματα στη master database και στα γεωγραφικά επίπεδα αναφοράς,
- Πιθανόν να προκύπτουν σφάλματα καθαρισμού δεδομένων
- Πιθανόν να προκύπτουν σφάλματα εξαιτίας του χρησιμοποιούμενου αλγορίθμου γεωκωδικοποίησης, που βασίζεται στην μέθοδο της «απλής παρεμβολής» με βάση την απόσταση για γεωκωδικοποίηση επί των γραμμών οδικών δικτύου και στη μέθοδο της εύρεσης του κεντροβαρούς για γεωκωδικοποίηση επί πολυγώνων τοπωνυμίων, διοικητικών υποδιαιρέσεων και TK (όταν υπάρχει έλλειψη οδού και αριθμού στην δοθείσα διεύθυνση).

2.3.2. Εμπορική εφαρμογή

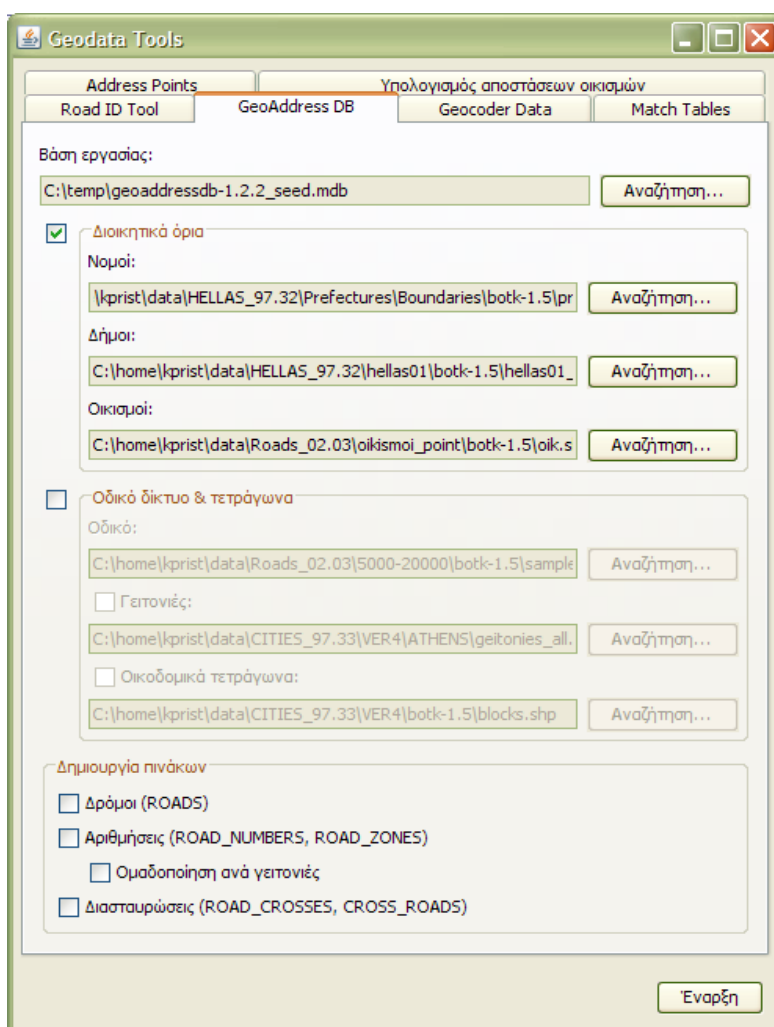
Η εμπορική εφαρμογή του προβλήματος αφορά τον αυτόματο ή ημιαυτόματο έλεγχο και διόρθωση των αποτελεσμάτων της γεωκωδικοποίησης διευθύνσεων στα έργα που

αναλαμβάνουν οι εταιρείες GEODATA και ΕΡΑΤΟΣΘΕΝΗΣ και αφορούν γεωκωδικοποίηση μεγάλων αρχείων διευθύνσεων.

Τέτοια έργα αφορούν συνήθως σε καθαρισμό και γεωκωδικοποίηση μεγάλων αρχείων διευθύνσεων πελατολογίων. Οι πελάτες των παραπάνω έργων είναι είτε δημόσιοι οργανισμοί (Δήμοι, Υπουργία, ΝΠΔΔ), είτε ιδιωτικοί φορείς (τράπεζες, ασφαλιστικές, τηλεπικοινωνιακές, αγορά ακινήτων, κ.α.).

2.3.3. Τρέχουσα κατάσταση

Η GEODATA και η ΕΡΑΤΟΣΘΕΝΗΣ χρησιμοποιούν για τις ανάγκες της γεωκωδικοποίησης την εφαρμογή Geodata Tools, που έχει αναπτυχθεί εσωτερικά με συνεργασία των δύο εταιρειών. Η εφαρμογή αποτελείται από ένα ενιαίο περιβάλλον, όπου παρέχονται εργαλεία για χρήση σε διάφορες εργασίες.



Εικόνα 1: Διεπαφή της εφαρμογής γεωκωδικοποίησης των εταιριών Ερατοσθένης/Geodata

Συνοπτικά η εφαρμογή Geodata tools συνίσταται στα ακόλουθα εργαλεία:

Road ID Tool

Το εργαλείο αυτό δέχεται ως είσοδο ένα γραμμικό shapefile, στο οποίο πρέπει να υπάρχουν:

- ένα αριθμητικό πεδίο που έχει ρόλο αναγνωριστικού του τόξου (π.χ. USERID)
- ένα πεδίο που αντιστοιχεί στο όνομα του δρόμου στον οποίο ανήκει το τόξο

Το εργαλείο αποδίδει ένα ROAD_ID για κάθε τόξο, έτσι ώστε για όλα τα τόξα που έχουν κοινό ROAD_ID να ισχύουν τα εξής:

- έχουν κοινή τιμή στο πεδίο ονόματος δρόμου
- το σύνολό τους αποτελεί ένα συνδεδεμένο κομμάτι δικτύου

GeoAddress DB

Με το εργαλείο αυτό παράγουμε τους πίνακες που συγκροτούν τη βάση οδών και αριθμήσεων ανά TK και τετράγωνο ΕΣΥΕ. Το εργαλείο αυτό χρησιμοποιείται επίσης και για τη δημιουργία της βάσης διευθύνσεων της εφαρμογής γεωκωδικοποίησης, όπως περιγράφεται σε επόμενο εργαλείο.

Geocoder Data

Το εργαλείο αυτό αναλαμβάνει να παράγει ένα πλήρες σύνολο δεδομένων για τη χρήση του στην εφαρμογή γεωκωδικοποίησης. Το αποτέλεσμα της διαδικασίας αποθηκεύεται σε κατάλογο που καθορίζεται από το χρήστη μέσω του εργαλείου.

Στα αρχεία εισόδου περιλαμβάνεται η βάση που δημιουργήθηκε από την διαδικασία που περιγράφηκε στα πλαίσια του εργαλείου GeoAddress DB. Θα πρέπει η βάση αυτή να περιλαμβάνει όλους τους πίνακες των διοικητικών ορίων, να έχει γίνει σε αυτήν μεταφορά των δεδομένων οδικού δικτύου, καθώς και να έχει δημιουργηθεί ο πίνακας ROADS.

Τα αρχεία shape file που θα επιλέξει ο χρήστης πρέπει να είναι ακριβώς τα ίδια αρχεία που χρησιμοποιήθηκαν κατά τη δημιουργία της βάσης εισόδου. Τα αρχεία εξόδου πρέπει να βρίσκονται στον υποκατάλογο "data" του καταλόγου εγκατάστασης της εφαρμογής γεωκωδικοποίησης.

2.3.4. Στόχος του προβλήματος

Η ανάπτυξη μεθόδων με χρήση τεχνολογιών μηχανικής μάθησης, που θα εξασφαλίζουν μεγαλύτερη ακρίβεια και εμπιστοσύνη στα αποτελέσματα της γεωκωδικοποίησης, είναι ο ζητούμενος στόχος του παρόντος προβλήματος.

Τα εργαλεία αυτά θα δρουν συμπληρωματικά ή και ενσωματωμένα στα υπάρχοντα εργαλεία και τεχνικές Γεωκωδικοποίησης των εταιρειών Geodata και Ερατοσθένης.

Συμπληρωματικά, με την έννοια ότι θα έρχονται να διορθώσουν το αποτέλεσμα της γεωκωδικοποίησης των υπάρχοντων εργαλείων και ενσωματωμένα, με την έννοια ότι θα αλλάξουν την διαδικασία της γεωκωδικοποίησης που εφαρμόζεται μέχρι τώρα. Συγκεκριμένα, προκειμένου να βελτιωθεί η ποιότητα γεωκωδικοποίησης, οι αλγόριθμοι θα διασυνδέουν και θα συγχωνεύουν επιμέρους αποτελέσματα γεωκωδικοποίησης από διαφορετικές πηγές, προκειμένου να επιλέγεται, από το εκπαιδευμένο μοντέλο μηχανικής μάθησης, η βέλτιστη κάθε φορά γεωκωδικοποίηση.

2.3.5. Σύνολα δεδομένων

Η Βάση δεδομένων εφαρμογής γεωκωδικοποίησης περιλαμβάνει τους εξής πίνακες:

Πίνακας «PREFECTURES»

Πεδίο	Περιγραφή
PREF_ID	Αναγνωριστικό Νομού
GREEKNAME	Όνομα Νομού

Πίνακας «MUNICIPALITIES»

Πεδίο	Περιγραφή
MUN_ID	Αναγνωριστικό Δήμου
GREEKNAME	Όνομα Δήμου
PREF_ID	Αναγνωριστικό Νομού για το Δήμο

Πίνακας «SETTLEMENTS»

Πεδίο	Περιγραφή
SETTL_ID	Αναγνωριστικό Πόλης / Οικισμού
GREEKNAME	Όνομα Πόλης / Οικισμού
MUN_ID	Αναγνωριστικό Δήμου για την Πόλη / Οικισμό
PREF_ID	Αναγνωριστικό Νομού για την Πόλη / Οικισμό
FID	Αναγνωριστικό εγγραφής στο shapefile οικισμών

Πίνακας «SETTL_POST_CODES»

Πεδίο	Περιγραφή
SETTL_ID	Αναγνωριστικό Πόλης / Οικισμού
POST_CODE	Ταχυδρομικός Κωδικός

Πίνακας «POST_CODES»

Πεδίο	Περιγραφή
POST_CODE	Ταχυδρομικός Κωδικός

Πίνακας «ROADS»

Πεδίο	Περιγραφή
ROAD_ID	Αναγνωριστικό οδού
GREEKNAME	Όνομα οδού
ALTNAME	Παλιό / εναλλακτικό όνομα οδού

<i>LABELNAME</i>	Όνομα οδού (σύντομο)
------------------	----------------------

Πίνακας «SEGMENTS»

Πεδίο	Περιγραφή
SEGMENT_ID	Αναγνωριστικό τόξου
SIDE	Πλευρά οδού ('L' για αριστερά, 'R' για δεξιά)
ROAD_ID	Αναγνωριστικό οδού
MUN_ID	Αναγνωριστικό Δήμου
POST_CODE	Ταχυδρομικός Κωδικός
NUM_FROM	Αρχή αρίθμησης της οδού για τα αντίστοιχα: πλευρά, ΤΚ και δήμο
NUM_TO	Τέλος αρίθμησης της οδού για τα αντίστοιχα: πλευρά, ΤΚ και δήμο
NUM_PARITY	0 – δεν έχουμε αρίθμηση 1 – περιττή αρίθμηση 2 – άρτια αρίθμηση 3 – μικτή / συνεχής αρίθμηση
FID	Αναγνωριστικό εγγραφής στο shape file τόξων
NODE_FROM	Αναγνωριστικό Κόμβου αρχής
NODE_TO	Αναγνωριστικό Κόμβου τέλους

Πίνακας «NODES»

Πεδίο	Περιγραφή
NODE	Αναγνωριστικό κόμβου
SEGMENT_ID	Αναγνωριστικό τόξου
RATIO	Άκρη τόξου (0 για αρχή, 1 για τέλος)
ROAD_ID	Αναγνωριστικό οδού
FID	Αναγνωριστικό εγγραφής στο shape file τόξων

Η βάση συνοδεύεται από δύο αρχεία shaperefile, όπου αποθηκεύεται μόνο η γεωμετρία (για λόγους απόδοσης στην ανάγνωσή τους):

- ένα αρχείο με τα τόξα (roads.shp)
- ένα αρχείο με τους οικισμούς (settlements.shp)

Τα πεδία FID στους πίνακες της βάσης αναφέρονται στον αύξοντα αριθμό της αντίστοιχης εγγραφής του shape file. Τα αρχεία αυτά είναι απαραίτητα για το χωρικό εντοπισμό των διευθύνσεων – σημείων.

Η εφαρμογή παραγωγής της βάσης δεδομένων βάζει τις κατάλληλες τιμές για τα FID. Η βάση που προκύπτει είναι συμβατή **μόνο** με τα αρχεία από τα οποία προήλθε, και θα παραμένει συμβατή με αυτά, αρκεί να μην αλλάξει η σειρά αποθήκευσης των εγγραφών. Αυτό εξασφαλίζεται αν η μόνη επεξεργασία που θα υποστούν τα shape file είναι η προσθήκη πεδίων, η αφαίρεση πεδίων, ή το update των τιμών τους, **όχι** όμως η αφαίρεση ή προσθήκη εγγραφών.

Επιπρόσθετα, στα δεδομένα της εφαρμογής περιλαμβάνονται shape file για την απεικόνιση στον χάρτη:

- blocks.shp
- municipalities.shp
- prefectures.shp

2.3.6. Βασικοί Δείκτες Απόδοσης

Στον παρακάτω πίνακα παρουσιάζεται βασικός δείκτης απόδοσης (KPIs) για το σενάριο της γεωκωδικοποίησης διευθύνσεων. Σε αυτό το σενάριο χρήσης υπάρχει υφιστάμενο σύστημα γεωκωδικοποίησης των εταιριών GEODATA και ΕΡΑΤΟΣΘΕΝΗΣ, το οποίο μπορεί να παράξει με μαζικό και αυτοματοποιημένο τρόπο προτάσεις γεωκωδικοποίησης (συντεταγμένες (x,y)) για ένα σύνολο διευθύνσεων. Οι μέθοδοι που θα υλοποιηθούν θα λαμβάνουν υπόψη τις συγκεκριμένες προτάσεις, καθώς και προτάσεις συντεταγμένων από δύο επιπλέον, ανοικτά συστήματα γεωκωδικοποίησης (ArcGIS⁴, OpenStreetMap⁵) και λαμβάνουν μία απόφαση ως προς το προτιμότερο ζεύγος συντεταγμένων που θα προτείνουν για την εκάστοτε διεύθυνση. Δεδομένου λοιπόν ότι η υπάρχουσα διαδικασία είναι ήδη αυτοματοποιημένη σε μεγάλο βαθμό, οι προτεινόμενες μέθοδοι στοχεύουν αποκλειστικά στην αύξηση της ποιότητας, δηλαδή της ακρίβειας γεωκωδικοποίησης. Οπότε, σε αυτό το σενάριο υπάρχει μόνο ένας βασικός δείκτης, ο οποίος θα μετρά την ποσοστιαία βελτίωση της ακρίβειας.

α/α	Δείκτης	Εκτιμώμενη Τιμή
1	Ποσοστό βελτίωσης ακρίβειας γεωκωδικοποίησης	30%

Πίνακας 5: Βασικός Δείκτης Απόδοσης στη Γεωκωδικοποίηση

⁴ <https://geocode.arcgis.com/arcgis/>

⁵ <https://wiki.openstreetmap.org/wiki/Nominatim>

2.4. ΣΧ4: Ολοκλήρωση Γεωτεμαχίων/Δρόμων

Πεδίο Εφαρμογής: Κτηματογράφηση

Εμπλεκόμενοι Φορείς: Ερατοσθένης

2.4.1. Περιγραφή προβλήματος

Στις μελέτες κτηματογράφησης, βασική οντότητα των παραδοτέων είναι τα γεωτεμάχια. Με τον όρο γεωτεμάχια εννοούμε το σύνολο των πολυγώνων που απαρτίζουν μια γεωγραφική περιοχή. Τα πολύγωνα αυτά, για να είναι αποδεκτά σε μια μελέτη κτηματογράφησης, πρέπει να πληρούν ορισμένα βασικά χαρακτηριστικά. Συγκεκριμένα, τοπολογικά θα πρέπει να καλύπτουν πλήρως την υπό κτηματογράφηση περιοχή χωρίς κενά (holes) και επικαλύψεις μεταξύ τους (not overlap). Επίσης τα πολύγωνα αυτά θα πρέπει να είναι απλά και όχι νησίδες (multipolygon). Όσον αφορά τα περιγραφικά τους χαρακτηριστικά (attributes), τα βασικά είναι: το είδος Γεωτεμαχίου, η προέλευση, το εμβαδόν και η διεύθυνση.

Το PST είναι το παραδοτέο αρχείο της γεωμετρίας των γεωτεμαχίων (shapefile) μιας μελέτης κτηματογράφησης. Το PST, κατά τη διάρκεια μιας μελέτης κτηματογράφησης, παράγεται σε διάφορες εκδόσεις (παραδόσεις):

- Το προκαταρκτικό PST, που αποτελεί την αρχική διαμόρφωση από τις διάφορες πηγές δεδομένων (που περιγράφονται στην Ενότητα 2.4.5). Επί του προκαταρκτικού PST γίνονται δηλώσεις ιδιοκτησίας από τους πολίτες.
- Ενδιάμεσες εκδόσεις PST (στο 35%, στο 70% και στο 100% των δηλώσεων) μετά την εφαρμογή των δηλώσεων.
- Τελικό PST μετά την ανάρτηση (δημοσιότητα) και την εξέταση των ενστάσεων από τους πολίτες.

Στους επόμενους πίνακες περιγράφεται αναλυτικά η δομή του αρχείου γεωτεμαχίων (PST), σε μια μελέτη κτηματογράφησης.

Όνομα	K	Y	Τύπος	Πλάτος	Περιγραφή	Εύρος τιμών - παρατηρήσεις
FID	O	N	NUMBER	LONG	Μοναδικός αριθμός που δίδεται αυτόματα από το σύστημα (PK)	
KAEK	N	N	CHAR	12	Κωδικός Αριθμός Εθνικού Κτηματολογίου. Μοναδικός αύξων αριθμός (UK)	Συμπληρώνεται με βάση τον Πίνακα 2-1
PROP_TYPE	O	N	CHAR	4	Είδος γεωτεμαχίου	Παίρνει τιμές σύμφωνα με τον Πίνακα κωδικοποίησης PROP_TYPE

ORI_TYPE	O	N	NUMBER/ SHORT INTEGE R	1	Είδος Προέλευσης του γεωτεμαχίου.	Παίρνει τιμές με βάση τα παρακάτω: 1: Κτηματογράφηση 2: Πράξη εφαρμογής 3: Αναδασμός 4: Διανομή 5: Πράξη Καθορισμού αιγιαλού 6: Απαλλοτρίωση
ORI_CODE	O	O	CHARACTER	100	Κωδικός του γεωτεμαχίου στη Διοικητική Πράξη.	Συμπληρώνεται στην περίπτωση που η τιμή του ORI_TYPE είναι διάφορη του 1.
DEC_ID	O	O	CHARACTER	254	Συμπληρώνεται ο αύξων αριθμός (FID) που προκύπτει από τον Πίνακα DBOUND και αφορά στο όριο της διοικητικής πράξης στην οποία ανήκει το γεωτεμάχιο.	Συμπληρώνεται στην περίπτωση που η τιμή του ORI_TYPE είναι διάφορη του 1.
ADDRESS	O	N	CHARACTER	254	Τρέχουσα διεύθυνση κύριας οδού μπροστά από το γεωτεμάχιο. Ακολουθεί τους κανόνες ονοματολογίας που αναφέρονται στην παρ 1.2.3.	Στην περίπτωση που το γεωτεμάχιο έχει πρόσωπο σε παραπάνω από έναν δρόμους, συμπληρώνεται η διεύθυνση με την οποία έχει την μεγαλύτερη επαφή.
NUM	O	N	CHARACTER	9	Τρέχουσα αρίθμηση κύριας οδού μπροστά από το γεωτεμάχιο. Σε περίπτωση μη αριθμών γίνεται χρήση κεφαλαίων Ελληνικών χαρακτήρων. Στην περίπτωση που η αρίθμηση επεκτείνεται σε παραπάνω από έναν αριθμούς αναγράφεται το εύρος χωριζόμενο με παύλα (π.χ. 32-34)	Στην περίπτωση που το γεωτεμάχιο έχει πρόσωπο σε παραπάνω από έναν δρόμους, συμπληρώνεται ο αριθμός της οδού με την οποία έχει την μεγαλύτερη επαφή.
AREA	O	N	NUMBER	DOUBLE	Καταμετρηθέν εμβαδόν γεωτεμαχίου σε μ ²	Όπως προκύπτει από το Γεωγραφικό Σύστημα Πληροφοριών του αναδόχου.
SHAPE	O	N	BINARY		Γεωμετρία	
LEN	O	N	NUMBER	DOUBLE	Περίμετρος σε μ	

Πίνακας 6: Δομή πίνακα Γεωτεμαχίων (PST shape file)

Αντικείμενο	Είδος στοιχείου	Περιεχόμενη πληροφορία	PROP_TYPE
Γεωτεμάχιο	Πολύγωνο	Πολύγωνο γεωτεμαχίου	0101
Μεταλλείο	Πολύγωνο	Πολύγωνο μεταλλείου	0601
Δρόμος	Πολύγωνο	Πολύγωνο ειδικής έκτασης	0701
Ποτάμι – Ρέμα	Πολύγωνο		0702
Αιγιαλός	Πολύγωνο		0703
Παραλία	Πολύγωνο		0704
Λίμνη	Πολύγωνο		0705
Παρόχθια ζώνη	Πολύγωνο		0707

Πίνακας 7: Κωδικοί Γεωτεμαχίων (PROP_TYPE)

Πολυγωνικά Ιδιοκτησιακά	Ειδικά Αντικείμενα	Είδος στοιχείου	Περιεχόμενη πληροφορία	PROP_TYPE
		Πολύγωνο	Πολύγωνο πηγαδότοπου	0501
		Πολύγωνο	Πολύγωνο αντλιοστασίου	0507
		Πολύγωνο	Πολύγωνο υδραγωγείου	0508
	Ερημονησίδες (ιδιοκτησίες του Υπουργείου Οικονομικών)	Πολύγωνο	Πολύγωνο ερημονησίδας	0509
	Μύλοι	Πολύγωνο	Πολύγωνο μύλου	0510

Πίνακας 8: Κωδικοί Ειδικών ιδιοκτησιακών αντικειμένων που καταχωρίζονται στον PST

Επομένως, βασικό πρόβλημα της κτηματογράφησης είναι η σωστή και ακριβής, με βάση τις προδιαγραφές, αποτύπωση των γεωτεμαχίων. Τα ποιοτικά κριτήρια της αποτύπωσης των γεωτεμαχίων είναι:

- Γεωμετρική ακρίβεια αποτύπωσης:
- Γεωμετρική Συμβατότητα Γεωτεμαχίων:
 - Συμβατότητα σχετικής θέσης και σχήματος
 - Συμβατότητα εμβαδού.

Βασικές πηγές δεδομένων για την εξαγωγή του τελικού PST γεωτεμαχίων, σε μια μελέτη κτηματογράφησης, είναι:

- Οι διοικητικές πράξεις των:
 - διανομών,
 - αναδασμών,
 - πράξεων εφαρμογής
 - απαλλοτριώσεων και
 - καθορισμού αιγιαλού
- Το Land Parcel Identification System (LPIS), που έχει προκύψει από νεότερες ψηφιοποιήσεις του ΥπΑΑΤ και λειτουργεί ως βάση αναγνώρισης των αγροτεμαχίων, για την πληρωμή των κρατικών ενισχύσεων προς τους Αγρότες.

- Αμπελουργικό ελαιοκομικό μητρώο
- Σύγχρονες Ορθοεικόνες (LSO25) και λήψης από drone, για την υποβοήθηση της οριοθέτησης

Τις τρεις (3) παραπάνω πρωτογενείς πηγές δεδομένων, έρχονται να διορθώσουν οι δηλώσεις ιδιοκτησίας των πολιτών οι οποίες σε πολλές περιπτώσεις περιλαμβάνουν και τοπογραφικά διαγράμματα. Στο συγκεκριμένο πρόβλημα της αποτύπωσης του «σωστού» PST, διακρίνουμε προς διερεύνηση δυο (2) υποπεριπτώσεις:

- Την αρχική επιλογή της ορθής τιμής του πεδίου ORI_TYPE (της επικρατέστερης προέλευσης) επί του τελικού PST.
- Σε δεύτερο χρόνο την ορθή επιλογή της ορθότερης γεωμετρίας για τα γεωτεμάχια (PST) από τις διαφορετικές πρωτογενείς πηγές, για την εξαγωγή του προκαταρκτικού υποβάθρου, επί του οποίου θα υποβάλλουν δηλώσεις ιδιοκτησίας οι πολίτες,

2.4.1.1. Επιλογή της τιμής του πεδίου ORI_TYPE (της επικρατέστερης προέλευσης) επί του τελικού PST

Στη συγκεκριμένη υποπερίπτωση, το πρόβλημα έχει να κάνει με την επιλογή της επικρατέστερης προέλευσης για κάθε οντότητα (γεωτεμάχιο) του τελικού PST, σε σχέση με τις αρχικές πηγές δεδομένων. Δηλαδή, μετά την επεξεργασία στα διάφορα στάδια της κτηματογράφησης, η εύρεση της επικρατέστερης αρχικής πηγής δεδομένων, βάση της οποίας παρήχθη η τελική γεωμετρία μιας οντότητας (γεωτεμαχίου) του τελικού PST.

Οι τιμές που παίρνει το πεδίο της επικρατέστερης προέλευσης (ORI_TYPE) σε μια μελέτη κτηματογράφησης είναι:

1. Κτηματογράφηση (δηλαδή, δεν ελήφθη υπόψη καμιά αρχική πηγή δεδομένων από τις επόμενες)
2. Πράξη εφαρμογής
3. Αναδασμός
4. Διανομή
5. Πράξη καθορισμού αιγιαλού
6. Απαλλοτρίωση

Η συμπλήρωση του πεδίου ORI_TYPE είναι σημαντική στην τελική μορφή του PST, προκειμένου να διασυνδέεται το κάθε γεωτεμάχιο (ή όχι) με τις διοικητικές πράξεις της περιοχής και κατά κάποιο τρόπο να αιτιολογείται η ύπαρξή του.

2.4.1.2. Επιλογή ορθότερης γεωμετρίας για το προκαταρκτικό PST

Στην συγκεκριμένη υποπερίπτωση του προβλήματος, αυτό που εξετάζουμε είναι η επιλογή της καταλληλότερης γεωμετρίας από τις προαναφερόμενες πηγές δεδομένων, προκειμένου να παραχθεί το προκαταρκτικό PST, επί του οποίου θα γίνουν οι δηλώσεις ιδιοκτησίας.

Είναι πολύ σημαντικό να έχουμε ένα προκαταρκτικό PST που να προσομοιάζει όσο το δυνατόν στην πραγματικότητα, γιατί σε διαφορετική περίπτωση δεν θα μπορούν να αναγνωριστούν τα γεωτεμάχια από τους ενδιαφερόμενους πολίτες κατά την υποβολή της δήλωσής τους.

Η κάθε οντότητα (γεωτεμάχιο) του PST είναι μία συγκεκριμένη πολυγωνική γεωμετρία η οποία έχει ένα περιορισμένο αριθμό από μεταδεδομένα, όπως:

- Στοιχεία ιδιοκτητή (Ονομαστικά στοιχεία, στοιχεία διεύθυνσης, ΑΦΜ, τηλέφωνο)
- Στοιχεία εμπράγματου Δικαιώματος (είδος και ποσοτικά στοιχεία δικαιώματος, αιτία κτήσης)
- Στοιχεία εγγράφων (τίτλων) που στοιχειοθετούν το δικαίωμα.

Ανάλογα με την περίπτωση χρήσης (τοποθεσία, διαθεσιμότητα συνόλων δεδομένων), ένα προκαταρκτικό PST για μία συγκεκριμένη περιοχή ενδέχεται να παράγεται εξετάζοντας και συνδυάζοντας έναν αριθμό από διαφορετικά σύνολα δεδομένων/layers (διανομές – αναδασμοί, LPIS, ελαιοκομικό – αμπελουργικό, ορθοεικόνες), τα οποία διαφέρουν αρκετά ως προς:

- τη δεκαετία παραγωγής τους και, κατ' επέκταση, το πόσο επικαιροποιημένα είναι
- την ακρίβειά τους (π.χ. κάποια σύνολα δεδομένων επικεντρώνονται μόνο στα καλλιεργήσιμα κομμάτια γης, ενώ άλλα στις συνολικές εκτάσεις).

Με βάση τα αρχικά σύνολα, χάρτες με ορθοεικόνες και διάφορους κανόνες προτεραιότητας, ο χρήστης της Ερατοσθένης παράγει το προκαταρκτικό PST.

2.4.1.3. Ολοκλήρωση και Διασύνδεση Δρόμων

Βασικό πρόβλημα στην διαδικασία της κτηματογράφησης μια αγροτικής περιοχής (όχι δηλαδή αστικής), που μπορεί όμως να βρει εφαρμογή και σε άλλες επιχειρησιακές δραστηριότητες, είναι η αναγνώριση και η οριοθέτηση των οδών (δρόμων) σε πολυγωνική μορφή, πάνω σε ένα γεωγραφικό σύστημα γεωτεμαχίων (parcels).

Στο επιχειρησιακό περιβάλλον της κτηματογράφησης, ένας δρόμος δεν αρκεί να οριοθετηθεί με την απεικόνισή του, μέσω της διανυσματοποίησης, πάνω σε ένα υπόβαθρο Ορθοφωτοχάρτη υψηλής ακρίβειας (αυτό πάντα είναι σχετικό), αλλά θα πρέπει να προσομοιάζει και στο σχήμα και τη θέση που ορίζουν οι σχετικές διοικητικές πράξεις καθορισμού του. Στην συγκεκριμένη περίπτωση, για τη χώρα μας, και σε αγροτικό περιβάλλον, οι διοικητικές πράξεις που καθορίζουν τα όρια των οδών είναι οι διανομές και αναδασμοί του Υπουργείου Αγροτικής Ανάπτυξης και Τροφίμων (ΥΠΑΑΤ), οι οποίες παίζουν και πρωτεύοντα ρόλο στη γεωμετρική αποτύπωση των δρόμων (αλλά και των γεωτεμαχίων). Επιπλέον δεδομένα που μπορεί να υπάρχουν σε μια περιοχή είναι οι πράξεις εφαρμογής, οι απαλλοτριώσεις και οι πράξεις καθορισμού αιγιαλού.

Οι διανομές – αναδασμοί περιλαμβάνουν χάρτες αποτύπωσης των γεωτεμαχίων καθώς και πίνακες ιδιοκτησίας (των πρώτων ιδιοκτητών).

Το πρόβλημα με τους χάρτες των διανομών και των αναδασμών είναι ότι, λόγω παλαιότητας, έχουν πρόβλημα ακρίβειας (θέσης και σχήματος) στην αποτύπωση που απαιτούν οι προδιαγραφές του σύγχρονου Κτηματολογίου, καθώς επίσης και λόγω της παρόδου πολλών ετών έχουν προκύψει νέα δεδομένα, από επιγενόμενες πράξεις (νόμιμες και μη), που έχουν διαμορφώσει αλλιώς την πραγματικότητα, όσον αφορά το σχήμα των γεωτεμαχίων και των δρόμων.

Για το λόγο αυτό, στη διαδικασία της κτηματογράφησης και ειδικότερα στην οριοθέτηση των οδών σε αγροτικό περιβάλλον, χρησιμοποιούνται και νεώτερες πηγές γεωχωρικών δεδομένων, που συσχετίζονται οπτικά, με επίθεση πάνω στους χάρτες των διανομών –

αναδασμών, προκειμένου οι δρόμοι να αποτυπωθούν με μεγαλύτερη ακρίβεια, χωρίς να αλλοιώνεται η ουσία της διοικητικής πράξης.

Τέτοιες πηγές πιο σύγχρονων δεδομένων είναι και το Land Parcel Identification System (LPIS), που έχει προκύψει από νεότερες ψηφιοποιήσεις του ΥπΑΑΤ και λειτουργεί ως βάση αναγνώρισης των αγροτεμαχίων, για την πληρωμή των κρατικών ενισχύσεων προς τους Αγρότες.

Το LPIS λοιπόν, στα πλαίσια της κτηματογράφησης, συσχετίζεται οπτικά με τους αναδασμούς και τις διανομές πάνω σε ένα υπόβαθρο ορθοφωτοχάρτη ακρίβειας 1:5.000, προκειμένου ο μελετητής να αποτυπώσει τα όρια των γεωτεμαχίων και των δρόμων.

2.4.2. Εμπορική εφαρμογή

Η εμπορική εφαρμογή του προβλήματος για την ΕΡΑΤΟΣΘΕΝΗΣ έχει να κάνει με τις μελέτες κτηματογράφησης, αποτύπωσης οδικού δικτύου (σε πολυγωνική μορφή), καθώς και με τα έργα που απαιτούν διασύνδεση διαφορετικών πηγών πολυγωνικών οντοτήτων, με σκοπό την εξαγωγή συνόλων δεδομένων πολυγώνων που θα απαντούν στις ζητούμενες προδιαγραφές του έργου.

2.4.3. Τρέχουσα κατάσταση

Η επιλογή της καταλληλότερης γεωμετρίας μιας οντότητας γεωτεμαχίου γίνεται σήμερα αποκλειστικά χειροκίνητα, με οπτική επισκόπηση, επιλογή και διόρθωση από τον (εξειδικευμένο) χρήστη της γεωμετρίας πολυγώνου από κάποια από τις διαθέσιμες πηγές δεδομένων (αναδασμοί – διανομές, LPIS). Η όλη διαδικασία γίνεται με εργαλεία GIS (QGIS, Autocad), με επίθεση των υποβάθρων των διανομών, αναδασμών και LPIS, πάνω σε υπόβαθρο ορθοφωτοχάρτη που παρέχει η εταιρία Ελληνικό Κτηματολόγιο Α.Ε. Επιπλέον, όπως προαναφέραμε, η διαδικασία αποτύπωσης των δρόμων στην κτηματογράφηση για τις αγροτικές περιοχές γίνεται με επίθεση των χαρτών διανομών – αναδασμών και LPIS, πάνω σε υπόβαθρο ορθοφωτοχάρτη και με οπτική αναγνώριση ο χρήστης αποφασίζει για την οριοθέτηση των δρόμων.

Οι διαδικασίες αυτές είναι αρκετά χρονοβόρες και η ακρίβειά τους εναπόκειται στην ικανότητα του χειριστή να συνεκτιμήσει τα στοιχεία για να αποδώσει τα ορθά όρια των πολυγωνικών γεωτεμαχίων και τη θέση του πολυγωνικού δρόμου. Σαν αποτέλεσμα λοιπόν έχει:

- Σημαντική καθυστέρηση για στην υλοποίηση των προκαταρκτικών πολυγώνων PST και των ορίων των δρόμων μια μεγάλης υπό κτηματογράφηση περιοχής και κατασπατάληση αρκετών ανθρωποωρών.
- Αρκετά σφάλματα στο τελικό αποτέλεσμα (λόγω μειωμένης ικανότητας του χρήστη στο να αποδώσει σωστά τα όρια του δρόμου).

2.4.4. Στόχος του προβλήματος

Στόχος της έρευνας στις δύο πρώτες υποπεριπτώσεις του προβλήματος είναι:

- Η αυτόματη ή ημιαυτόματη επιλογή της επικρατέστερης προέλευσης (ORI_TYPE) για την διασύνδεση της τελική γεωμετρίας ενός γεωτεμαχίου με την αρχική πηγή προέλευσης.
- Η αυτόματη ή ημιαυτόματη αναγνώριση της καταλληλότερης αρχικής γεωμετρίας (προκαταρκτικό PST), πάνω στην οποία θα βασιστεί ο χρήστης για να σχεδιάσει την τελική γεωμετρία της ιδιοκτησίας (τελικό PST),

Στην πρώτη υποπερίπτωση, το σύστημα θα πρέπει να παράγει μία ταξινόμηση των τιμών ORI_TYPE για κάθε τελική γεωμετρία του PST και ο χρήστης θα μπορεί να επιλέξει μια.

Στη δεύτερη υποπερίπτωση, παρόμοια, το σύστημα θα πρέπει να παράγει μια ταξινόμηση των αρχικών γεωμετριών (από τις διάφορες πηγές δεδομένων που προαναφέραμε), που θα αντιπροσωπεύει τη σχετική καταλληλότητά τους και θα δίνει στο χρήστη την δυνατότητα να επιλέξει μία.

Η ανάγκη που προκύπτει στο πλαίσιο της τρίτης υποπερίπτωσης είναι η ανάπτυξη ενός εργαλείου βασισμένου σε τεχνολογίες μηχανικής μάθησης για την υποβοήθηση της παραπάνω διαδικασίας αναγνώρισης των ορίων των οδών, καθώς και νέων οδών, μέσω της αυτόματης διασύνδεσης των διαφορετικών πηγών δεδομένων (Αναδασμοί, διανομές και LPIS). Πιο συγκεκριμένα, το εργαλείο αυτό θα πρέπει να συσχετίζει τις διανομές – αναδασμούς με το LPIS και με βάση κάποιες παραμέτρους διασύνδεσης, να επιλέγει αυτόματα ή ημιαυτόματα τα πολύγωνα των δρόμων, που εμφανίζονται να είναι διαφορετικά στη μία πηγή δεδομένων από την άλλη, έτσι ώστε να επισημειώνονται ως νέα κομμάτια δρόμων, και να εμπλουτίζουν τα υπάρχοντα σύνολα δεδομένων δρόμων. Κατά τον τρόπο αυτό θα αυξάνεται η ποιότητα της διαδικασίας αποτύπωσης των τελικών δρόμων.

2.4.5. Σύνολα δεδομένων

Τα σύνολα των δεδομένων που χρησιμοποιούνται στα τρία υποπροβλήματα είναι, κατά περίπτωση (ανάλογα δηλαδή με την περιοχή κτηματογράφησης μπορεί να υπάρχουν ή να μην υπάρχουν κάποια από αυτά), τα παρακάτω:

- Οι διοικητικές πράξεις των:
 - διανομών,
 - αναδασμών,
 - πράξεων εφαρμογής
 - απαλλοτριώσεων και
 - καθορισμού αιγιαλού
- Το Land Parcel Identification System (LPIS), που έχει προκύψει από νεότερες ψηφιοποιήσεις του ΥπΑΑΤ και λειτουργεί ως βάση αναγνώρισης των αγροτεμαχίων, για την πληρωμή των κρατικών ενισχύσεων προς τους Αγρότες.
- Αμπελουργικό ελαιοκομικό μητρώο
- Σύγχρονες Ορθοεικόνες (LSO25) και λήψης από drone, για την υποβοήθηση της οριοθέτησης

Τα δεδομένα αυτά είναι προσβάσιμα για τους μελετητές στα πλαίσια των έργων της Κτηματογράφησης και παρέχονται από τον οργανισμό του Εθνικού Κτηματολογίου για το

σκοπό αυτό. Η γεωγραφική απεικόνιση των δεδομένων αυτών είναι σε σύστημα συντεταγμένων ΕΓΣΑ87 και σε format shapefile.

2.4.6. Βασικοί Δείκτες Απόδοσης

Στον παρακάτω πίνακα παρουσιάζονται οι εκτιμώμενοι βασικοί δείκτες απόδοσης (KPIs) για το σενάριο της κατηγοριοποίησης ΣΕ. Ο πρώτος δείκτης αφορά το ποσοστό της διαδικασίας ολοκλήρωσης γεωτεμαχίων που επιτελείται με χειροκίνητο, έναντι αυτοματοποιημένου τρόπου. Ο δεύτερος δείκτης αφορά την ανθρωπο-προσπάθεια που απαιτείται να ξοδευτεί, συνολικά ανά διαδικασία-εργασία ολοκλήρωσης, ενώ ο τρίτος δείκτης αφορά τη συνολική χρονική διάρκεια της εκάστοτε διαδικασίας. Ο τέταρτος δείκτης αφορά στην ακρίβεια κατηγοριοποίησης που επιτυγχάνεται από τις μεθόδους αυτοματοποιημένης ολοκλήρωσης που θα υλοποιηθούν, είτε σε σχέση με το επιλεχθέν ORI_TYPE, είτε σε σχέση με την επιλεχθείσα αρχική γεωμετρία για το προκαταρκτικό PST. Τέλος, ο πέμπτος δείκτης αφορά την ακρίβεια αναγνώρισης νέων δρόμων, σε σχέση με το πόσοι από τους αναγνωριζόμενους είναι πραγματικά νέοι δρόμοι, ή ήδη υπάρχοντες στο σύνολο δεδομένων βάσης του προβλήματος

Σε κάθε δείκτη απόδοσης, αναθέτουμε μία ενδεικτική *επιθυμητή τιμή* την οποία στοχεύουμε να ικανοποιούν οι μέθοδοι που θα υλοποιηθούν, μέχρι τη λήξη του έργου.

α/α	Δείκτης	Εκτιμώμενη Τιμή
1	Ποσοστό μείωσης χειροκίνητης ολοκλήρωσης γεωτεμαχίων σε σχέση με τη υπάρχουσα κατάσταση	40%
2	Ποσοστό μείωσης ανθρωπο-προσπάθειας ολοκλήρωσης γεωτεμαχίων σε σχέση με την υπάρχουσα κατάσταση	50%
3	Ποσοστό μείωσης της διάρκειας ολοκλήρωσης γεωτεμαχίων σε σχέση με την υπάρχουσα κατάσταση	50%
4	Ποσοστό ακρίβειας ολοκλήρωσης γεωτεμαχίων	60%
5	Ποσοστό ακρίβειας διασύνδεσης δρόμων - εύρεσης νέων δρόμων	60%

Πίνακας 9: Βασικοί Δείκτες Απόδοσης στην Ολοκλήρωση

3. Σύνοψη

Στο Παραδοτέο 1.1 παρουσιάσαμε τα βασικά σενάρια χρήσης τα οποία χρησιμοποιούνται στο έργο LinkGeoML για την ανάπτυξη και αποτίμηση μεθόδων (χαρακτηριστικών εκπαίδευσης, αλγορίθμων και διαδικασιών) μηχανικής μάθησης για την αυτοματοποιημένη και ακριβή διασύνδεση, επισημείωση και ολοκλήρωση χωρο-κειμενικών δεδομένων. Σε αυτά, η διασύνδεση διαφόρων τύπων χωρο-κειμενικών δεδομένων υπεισέρχεται ως το βασικό πρόβλημα, ή ως συμπληρωματικό-επιβοηθητικό πρόβλημα.

Τα τέσσερα σενάρια που παρουσιάστηκαν, με κάποια από τα οποία να αναλύονται σε περισσότερα, επιμέρους υποσενάρια, σχετίζονται όλα με πραγματικές διαδικασίες και ανάγκες των εταιριών ΕΡΑΤΟΣΘΕΝΗΣ και GEODATA. Για καθένα από τα τέσσερα σενάρια, αναλύθηκε η εμπορική σημασία, η τρέχουσα κατάσταση, ο στόχος όσον αφορά τη βελτίωση της τρέχουσας κατάστασης, καθώς και ενδεικτικοί βασικοί δείκτες απόδοσης και τα σύνολα δεδομένων που θα χρησιμοποιηθούν, προκειμένου να αξιολογηθεί η συνεισφορά του έργου ως προς τη βελτίωση της τρέχουσας κατάστασης.

Τα σενάρια καλύπτουν μία ευρεία γκάμα μοντέλων και χαρακτηριστικών μηχανικής μάθησης, όπως περιγράφονται αναλυτικά στα Παραδοτέα Π1.2 και Π2.1. Για αυτό το λόγο, αποτελούν τη βάση της οργάνωσης της επιτελούμενης δουλειάς του έργου, και θα χρησιμοποιούνται ως αναφορά σε αρκετά από τα επόμενα παραδοτέα του έργου.