

# ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ

## ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ

### ΤΟΜΕΑΣ ΤΕΧΝΟΛΟΓΙΑΣ ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΥΠΟΛΟΓΙΣΤΩΝ

#### ΤΕΧΝΗΤΗ ΝΟΗΜΟΣΥΝΗ

ΧΕΙΜΕΡΙΝΟ ΕΞΑΜΗΝΟ 2018

ΘΕΜΑ 2

#### Περιγραφή

Σε αυτή την εργασία καλείστε να βελτιώσετε την υπηρεσία εξυπηρέτησης πελατών ταξί που κατασκευάσατε στο πλαίσιο της 1ης εργασίας. Στην πρώτη εργασία υλοποιήσατε μια παραλλαγή του αλγορίθμου αναζήτησης  $A^*$  και κατασκευάσατε ένα πρόγραμμα Java το οποίο, δεδομένης της γεωγραφικής θέσης του πελάτη, των γεωγραφικών θέσεων των διαθέσιμων ταξί και έναν οδικό χάρτη υπολόγιζε το ταξί που θα εξυπηρετήσει τον πελάτη, επιλέγοντας το ταξί που βρίσκεται στη μικρότερη δυνατή απόσταση από τον πελάτη και προτείνοντας πιθανόν περισσότερες από μία ισοδύναμες διαδρομές μέχρι τον πελάτη. Στην παρούσα εργασία θα βελτιώσετε την υπηρεσία ώστε να λειτουργεί με τον εξής τρόπο:

1. Ο πελάτης πέρα από τη γεωγραφική θέση όπου βρίσκεται θα δηλώνει και τη γεωγραφική θέση του προορισμού του. Με βάση αυτή, η υπηρεσία θα υπολογίζει τη διαδρομή που θα ακολουθήσει το ταξί που τελικά θα επιλεγεί. Η επιλογή της διαδρομής θα μπορεί να γίνει με επιπλέον κριτήρια πέραν της ελάχιστης διανυόμενης απόστασης, όπως προτίμηση οδών ταχύτερης κυκλοφορίας, αποφυγή των οδών με υψηλή κίνηση κατά την ώρα που θα γίνει η μεταφορά, προτίμηση των φωτισμένων οδών αν η διαδρομή είναι νυχτερινή, κ.ά.
2. Η υπηρεσία θα κατατάσσει αρχικά τα διαθέσιμα ταξί που μπορούν να εξυπηρετήσουν τον πελάτη με βάση την απόστασή τους από τον πελάτη, προτιμώντας πιθανώς και στην περίπτωση αυτή οδούς ταχύτερης κυκλοφορίας, και αποφεύγοντας οδούς με υψηλή κίνηση κατά την ώρα που θα γίνει η μεταφορά. Η ικανότητα ενός ταξί να εξυπηρετήσει τον πελάτη θα καθορίζεται από στοιχεία, όπως ο αριθμός των ατόμων που μπορεί να μεταφέρει το ταξί σε σχέση με τον αριθμό που ζητεί ο πελάτης, η γλώσσα που μιλά ο οδηγός του ταξί σε σχέση με τη γλώσσα που δηλώνει ότι μιλά ο πελάτης, το αν το ταξί μπορεί να εξυπηρετήσει μεταφορές μεγάλων αποστάσεων, κλπ. Τα  $k$  πρώτα ταξί αυτής της κατάταξης (π.χ.  $k = 5$ ) θα παρουσιάζονται στον χρήστη, μαζί με το μήκος της διαδρομής ως τον πελάτη που έχει υπολογιστεί.
3. Στη συνέχεια, η υπηρεσία θα κατασκευάζει μια δεύτερη κατάταξη των  $k$  πρώτων ταξί, με βάση άλλα χαρακτηριστικά, όπως την αξιολόγηση που διαθέτει κάθε ταξί από τους χρήστες, την κατάσταση του οχήματος, την καταλληλότητά του για τη συγκεκριμένη μεταφορά, κ.ά. Η δεύτερη αυτή κατάταξη θα παρουσιάζεται στον πελάτη παράλληλα με την πρώτη, ώστε ο πελάτης να μπορεί να κάνει την τελική του επιλογή.

Στην πρώτη εργασία, ως οδικός χάρτης σας δόθηκαν κάποια δεδομένα προερχόμενα από το OpenStreetMap, τα οποία αντιπροσώπευαν συγκεκριμένα γεωγραφικά σημεία τα οποία ανήκαν σε κάποιες γραμμές, οι οποίες για απλότητα και για να επικεντρωθείτε στην υλοποίηση του αλγορίθμου  $A^*$ , ερμηνεύτηκαν αδιάκριτα ως οδοί. Προφανώς, όπως πιθανώς να διαπιστώσατε, δεν αποτελούσαν όλα τα σημεία αυτά σημεία οδών, και ακόμη και αν αποτελούν σημεία οδών, οι διάφορες οδοί μπορεί να έχουν διαφορετικά χαρακτηριστικά τα οποία αγνοήθηκαν. Στην εργασία αυτή θα σας δοθεί ένα πιο πλούσιο σύνολο δεδομένων που περιέχει περισσότερες πληροφορίες σχετικά με το είδος των γεωγραφικών γραμμών με τις οποίες εργαστήκατε στην 1η εργασία. Η βελτιωμένη υπηρεσία θα πρέπει να χρησιμοποιεί αυτά τα δεδομένα ώστε να χρησιμοποιεί με μεγαλύτερη ακρίβεια τον οδικό χάρτη, και να μπορεί κατά το δυνατόν να διακρίνει τις γραμμές του χάρτη που είναι οδοί και επιτρέπουν την κίνηση οχημάτων επί αυτών από τις υπόλοιπες γραμμές του χάρτη που αντιπροσωπεύουν άλλα γεωγραφικά στοιχεία. Επιπλέον θα πρέπει να γίνεται σεβαστή η επιτρεπτή κατεύθυνση κίνησης σε κάθε οδό.

#### Δεδομένα

Για να υλοποιήσετε την υπηρεσία θα θεωρήσετε ότι σας δίνονται τα παρακάτω.

1. Ένα αρχείο `client.csv` που περιέχει πληροφορίες για μια αίτηση πελάτη. Το αρχείο έχει την εξής μορφή:

```
X, Y, X_dest, Y_dest, time, persons, language, luggage  
23.733912,37.975687,23.772518,38.012301,20:00,3,greek,1
```

όπου  $X$  και  $Y$  το γεωγραφικό μήκος και πλάτος της θέσης του πελάτη,  $X_{dest}$  και  $Y_{dest}$  το γεωγραφικό μήκος και πλάτος του προορισμού του,  $time$  η τρέχουσα ώρα,  $persons$  το πλήθος των ατόμων που θα πρέπει να μεταφέρει το ταξί,  $language$  η γλώσσα που μιλάει ο πελάτης και  $luggage$  το πλήθος των αποσκευών που θέλει να μεταφέρει.

2. Ένα αρχείο taxis.csv που περιέχει έναν κατάλογο με ταξί και διάφορες πληροφορίες. Το αρχείο έχει την εξής μορφή:

```
X,Y,id,available,capacity,languages,rating,long_distance,type
23.741587,37.984125,100,yes,1-2,greek,8.2,yes,subcompact
23.691133,37.967339,110,yes,1-4,english,8,6,yes,large
23.757838,37.962517,120,yes,1-4,greek,9.2,yes,compact
...
```

όπου X και Y είναι το γεωγραφικό μήκος πλάτος της θέση του ταξί, id το αναγνωριστικό του, available αν είναι διαθέσιμο ή όχι την τρέχουσα χρονική στιγμή, capacity το πλήθος των ατόμων που μπορεί να μεταφέρει, languages οι γλώσσες που μιλά ο οδηγός (διαχωρισμένες μεταξύ τους με το σύμβολο |), rating ο μέσος βαθμός από την αξιολόγηση των χρηστών (με άριστα το 10), long\_distance αν το ταξί διατίθεται για μετακίνηση σε μεγάλες αποστάσεις εκτός πόλης, και type ο τύπος του οχήματος.

3. Ένα αρχείο nodes.csv που περιέχει έναν κατάλογο με γεωγραφικές συντεταγμένες σημείων διάφορων γεωγραφικών γραμμών. Κάθε σημείο συνοδεύεται από ένα αναγνωριστικό, και η γραμμή επί της οποίας βρίσκεται το σημείο αναγνωρίζεται από ένα αναγνωριστικό γραμμής; συνεπώς όλα τα σημεία του καταλόγου με το ίδιο αναγνωριστικό γραμμής αντιστοιχούν σε σημεία της ίδιας γραμμής. Στο αρχείο, τα σημεία της ίδιας γραμμής δίνονται ακολουθιακά το ένα μετά το άλλο. Τα σημεία κάθε γραμμής ορίζουν ευθύγραμμα τμήματα που είτε προσεγγίζουν καμπύλες διαδρομές είτε αντιστοιχούν σε σημεία τομής με άλλες γραμμές. Συνεπώς ένα σημείο μπορεί να ανήκει σε περισσότερες από μία γραμμές. Το αρχείο έχει την εξής μορφή:

```
X,Y,line_id,node_id,name
23.7140723,37.950289,23181765,250675895,Θυατείρων
23.714813,37.9497098,23181765,1177176105,Θυατείρων
23.7188913,37.9491957,23181771,250925215,Εφέσσου
23.7184186,37.949578,23181771,250925214,Εφέσσου
...
```

όπου X και Y είναι το γεωγραφικό μήκος και πλάτος του σημείου, line\_id το αναγνωριστικό της γραμμής στην οποία ανήκει το σημείο, node\_id το αναγνωριστικό του σημείου και name το όνομα της γραμμής (δεν δίνεται πάντα). Το αρχείο αυτό είναι παρόμοιο με το αρχείο που σας δόθηκε στην 1η εργασία. Το καινούργιο στοιχείο που περιέχει είναι το node\_id.

4. Ένα αρχείο lines.csv που περιέχει έναν κατάλογο με γεωγραφικές γραμμές, σημεία των οποίων αποτελούν οι κόμβοι που περιέχονται στο αρχείο nodes.csv. Κάθε γραμμή αναγνωρίζεται από το αναγνωριστικό γραμμής της και συνοδεύεται από διάφορα χαρακτηριστικά, συγκεκριμένα από τα highway, name, oneway, lit, lanes, maxspeed, railway, boundary, access, natural, barrier, tunnel, bridge, incline, waterway, busway, toll. Τα χαρακτηριστικά αυτά μπορεί να έχουν κάποια τιμή ή να είναι κενά. Η απουσία τιμής μπορεί να σημαίνει ότι το συγκεκριμένο χαρακτηριστικό δεν έχει νόημα για το συγκεκριμένο είδος γραμμής ή ότι δεν υπάρχει διαθέσιμη πληροφορία. Η παρουσία τιμής για κάποιο χαρακτηριστικό (ανεξάρτητα της συγκεκριμένης τιμής του), καθορίζει το είδος της γραμμής. Έτσι π.χ. αν το highway, railway, boundary, natural, barrier, waterway έχει τιμή σημαίνει ότι η γραμμή είναι οδός, σιδηροδρομική γραμμή, γεωγραφικό όριο, φυσικό τείχος, τείχος ή υδάτινη οδός αντίστοιχα. Οι συγκεκριμένες τιμές αυτών των πεδίων δίνουν πληροφορία για το είδος της οδού, σιδηροδρομικής γραμμής, γεωγραφικού ορίου, τείχους, φυσικού τοίχους ή υδάτινης οδού, αντίστοιχα. Από τα υπόλοιπα χαρακτηριστικά, το name δίνει το όνομα της γραμμής, το oneway πληροφορία για το αν η γραμμή είναι μονής ή διπλής κατεύθυνσης (αν η τιμή είναι yes τότε είναι μονής κατεύθυνσης και η φορά της οδού, όπως καθορίζεται από τη σειρά των σημείων από τα οποία αποτελείται, είναι αυτή που κωδικοποιείται στο αρχείο nodes.csv, αν η τιμή είναι -1 τότε είναι μονής κατεύθυνσης με φορά αντίθετη από αυτή που κωδικοποιείται στο αρχείο nodes.csv, ενώ αν είναι no είναι διπλής κατεύθυνσης), το lit προσδιορίζει αν υπάρχει φωτισμός, το lanes πόσες λωρίδες κυκλοφορίας υπάρχουν, το maxspeed το όριο ταχύτητας, το access τυχόν περιορισμούς πρόσβασης, το tunnel προσδιορίζει αν είναι τούνελ, το bridge προσδιορίζει αν είναι γέφυρα, το incline αν έχει αξιοσημείωτη κλίση, το busway αν διαθέτει λεωφορειόδρομο και το toll αν απαιτείται η καταβολή διόδων. Τα χαρακτηριστικά αυτά είναι μια επιλογή από το σύνολο των χαρακτηριστικών που διαθέτουν τα δεδομένα του OpenStreetMap. Αναλυτικές πληροφορίες για κάθε πεδίο και τις δυνατές τιμές του μπορείτε να αναζητήσετε στις ιστοσελίδες του τύπου [http://wiki.openstreetmap.org/wiki/Key:όνομα\\_χαρακτηριστικού](http://wiki.openstreetmap.org/wiki/Key:όνομα_χαρακτηριστικού), π.χ. <http://wiki.openstreetmap.org/wiki/Key:highway>, <http://wiki.openstreetmap.org/wiki/Key:lit>, κλπ. Το αρχείο έχει την εξής μορφή:

```
id, highway, name, oneway, lit, lanes, maxspeed, railway, boundary, access, natural, barrier, tunnel, bridge,
incline, waterway, busway, toll
5168803, primary, Κηφισίας, yes, yes, 3, 60, , , , , , , , , ,
23182600, secondary, Ελευθερίου Βενιζέλου, yes, yes, 2, 50, , , , , , yes, , ,
23185096, footway, , no, , , , , , , yes, , ,
23200035, Γραμμή1, , , , subway, , , , , yes, , , ,
...
```

όπου id είναι το αναγνωριστικό της γραμμής και τα υπόλοιπα πεδία είναι τα διάφορα χαρακτηριστικά της γραμμής. Όπως προαναφέρθηκε, σε πολλές πεδία δεν δίνεται τιμή είτε διότι το πεδίο δεν αφορά τη συγκεκριμένη γραμμή είτε διότι δεν υπάρχει πληροφορία.

5. Ένα αρχείο `traffic.csv` που περιέχει γενικά στοιχεία κίνησης για κάποιες οδούς ανά χρονικά διαστήματα. Το αρχείο έχει την εξής μορφή:

```
id, name, traffic_info
5168803, Κηφισίας, 09:00-11:00=high|13:00-15:00=medium|17:00-19:00=high
...
```

όπου `id` είναι το αναγνωριστικό της γραμμής, `name` το όνομά της και `info` οι υπάρχουσες πληροφορίες κίνησης. Μπορεί να παρέχονται πληροφορίες για περισσότερα του ενός χρονικά διαστήματα, οι οποίες διαχωρίζονται μεταξύ τους με το σύμβολο `|`. Για κάθε χρονικό διάστημα δίνεται το χρονικό εύρος του και το επίπεδο κίνησης (`high`, `medium`, `low`). Τα δυνατά χρονικά εύρη είναι τα δίωρα 00:00-02:00, 02:00-04:00, ..., 22:00-00:00.

## Ζητούμενα

Χρησιμοποιώντας τα δεδομένα που περιγράφηκαν παραπάνω καλείστε να τροποποιήσετε το πρόγραμμα που κατασκευάσατε στην πρώτη εργασία ώστε να υλοποιείται η υπηρεσία που περιγράφηκε παραπάνω. Ο κορμός του προγράμματος που θα κατασκευάσετε, δηλαδή ό,τι αφορά την προεπεξεργασία των δεδομένων, τον γενικό έλεγχο ροής και την παραλλαγή του αλγορίθμου  $A^*$ , θα είναι σε Java. Για ό,τι αφορά όμως τη γνώση σχετικά με τον κόσμο που διαθέτει το πρόγραμμα, θα χρησιμοποιήσετε λογικό προγραμματισμό και συγκεκριμένα την Prolog.

Αυτό σημαίνει ότι θα πρέπει κατ' αρχάς να μετασχηματίσετε τα διάφορα δεδομένα που σας δίνονται σε ένα σύνολο γεγονότων που θα μπορέσουν να χρησιμοποιηθούν από ένα απλό πρόγραμμα Prolog. Στη συνέχεια θα επαυξήσετε το σύνολο των γεγονότων με κάποιους απλούς λογικούς κανόνες, ώστε να μπορούν να συνάγονται χρήσιμα συμπεράσματα. Θα θεωρήσετε ότι το πρόγραμμα Java που θα κατασκευάσετε δεν θα διαθέτει καμία γνώση σχετικά με τις ιδιότητες των σημείων και των γραμμών του χάρτη και των ταξί, εκτός ίσως από τα αναγνωριστικά τους και τις γεωγραφικές συντεταγμένες τους. Για να λάβει οποιαδήποτε άλλη χρήσιμη πληροφορία θα πρέπει να απευθύνει κάποιο ερώτημα προς το πρόγραμμα Prolog.

Συνεπώς το πρόγραμμα Prolog θα πρέπει να παρέχει στο πρόγραμμα Java και στην παραλλαγή του αλγορίθμου  $A^*$  πληροφορίες σχετικά π.χ. με την προσβασιμότητα κάθε σημείου για το ταξί, για το ποιο σημείο θα πρέπει να προτιμηθεί σε κάθε μετάβαση κλπ. Θα μπορούσατε π.χ. να κατασκευάσετε κάποιους κανόνες της μορφής `canMoveFromTo(X, Y) :- ..., priority(R, T, Z) :- ...` για να αποφασίζει ο αλγόριθμος προς τα που θα κινηθεί (στο συγκεκριμένο παράδειγμα αν μπορεί να μεταβεί από το σημείο  $X$  στο  $Y$  και κάποια τιμή προτεραιότητας  $Z$  για την οδό  $R$  κατά το χρονικό διάστημα  $T$ , αξιοποιώντας π.χ. τις τιμές των χαρακτηριστικών `oneway` και `highway`). Δεδομένου ότι το μόνο τροποποιήσιμο μέρος της παραλλαγής του αλγορίθμου  $A^*$  είναι (πέραν του μεγέθους του μετώπου αναζήτησης) οι συναρτήσεις κόστους, οι πληροφορίες που θα λαμβάνει από το πρόγραμμα Prolog θα πρέπει προφανώς να χρησιμοποιούνται από αυτές, ώστε το κόστος να μην είναι απλώς κάποια γεωγραφική απόσταση, αλλά μια σταθμισμένη απόσταση. Ως βασικό κορμό του προγράμματος Prolog θα πρέπει να κατασκευάσετε ένα σύνολο γεγονότων `belongsTo(X, L)` για τους κόμβους  $X$  που ανήκουν στη γραμμή  $L$  και `next(X, Y)` για τους κόμβους προς τους οποίους είναι δυνατή η μετάβαση  $Y$  από το κόμβο  $X$  επειδή ανήκουν στην ίδια γραμμή και η κίνηση είναι συμβατή με την πληροφορία κατεύθυνσης για τη συγκεκριμένη γραμμή (όπως προκύπτει από το χαρακτηριστικό `oneway`). Τα υπόλοιπα χαρακτηριστικά που σας δίνονται στο αρχείο `lines.csv` και `traffic.csv` μπορείτε να τα χρησιμοποιήσετε κατά βούληση, για να επιτύχετε τον καλύτερο ορισμό προτιμώμενων διαδρομών. Θα πρέπει επίσης να επιλέξετε πώς θα χειριστείτε περιπτώσεις όπου η πληροφορία είναι ελλιπής. Με αντίστοιχο τρόπο θα πρέπει να μοντελοποιήσετε τα δεδομένα που αφορούν τα χαρακτηριστικά των ταξί.

Για να αποφύγετε τεχνικές δυσκολίες και να μπορέσετε να υλοποιήσετε ομαλά την επικοινωνία Java - Prolog, σας προτείνεται να χρησιμοποιήσετε το σύστημα JIProlog (την έκδοση <https://github.com/jiprolog/jiprolog/releases/download/4.1.4.1/JIProlog-4.1.4.1.zip>, από το οποίο θα χρειαστείτε το αρχείο `jipconsole.jar`). Για διευκόλυνσή σας, σας δίνεται το αρχείο `prolog.pl` με ένα στοιχειώδες πρόγραμμα Prolog με περιπτώσεις που πιθανόν να χρειαστείτε, και το `PrologExample.java` που επιδεικνύει πώς μπορεί να χρησιμοποιηθεί η μηχανή του JIProlog μέσα από ένα πρόγραμμα Java.

Αφού κατασκευάσετε το σύστημα, καλείστε να το εκτελέσετε για τον συγκεκριμένο χάρτη, τοποθεσία πελάτη και τοποθεσίες ταξί που σας δίνονται, αλλά και για μια ακόμη διαφορετική τοποθεσία πελάτη και τοποθεσίες ταξί τις οποίες θα επιλέξετε εσείς. Για την οπτική απεικόνιση των διαδρομών στους χάρτες, μπορείτε να χρησιμοποιήσετε την υπηρεσία My Maps της Google όπως σας προτάθηκε στην 1η εργασία.

Για την εργασία αυτή θα πρέπει να παραδώσετε:

1. Μια συνοπτική αναφορά, όπου θα περιγράφετε τον γενικό σχεδιασμό του συστήματος και τον τρόπο μοντελοποίησης των δεδομένων. Ειδικότερα, θα πρέπει να αναφερθείτε στις τροποποιήσεις που χρειάστηκε να κάνετε στο πρόγραμμα σε σχέση με την πρώτη εργασία, να εξηγήσετε τι αντιπροσωπεύουν τα κατηγορήματα που χρησιμοποιούνται στο πρόγραμμα Prolog και να αναπτύξετε τη λογική πάνω στην οποία βασίσατε τους κανόνες του λογικού προγράμματος που κατασκευάσατε.

Στην αναφορά θα πρέπει να συμπεριλάβετε και τα αποτελέσματα, δηλαδή τους χάρτες με απεικονισμένες τις διαδρομές των ταξί προς τον πελάτη, τη διαδρομή από το σημείο παραλαβής του πελάτη ως τον προορισμό του, και τις δύο κατατάξεις των διαθέσιμων ταξί που παράγει το πρόγραμμα.

2. Τον πηγαίο κώδικα του προγράμματος Java.
3. Ένα ή περισσότερα αρχεία με τα γεγονότα Prolog που κατασκευάσατε.
4. Ένα αρχείο με τους κανόνες Prolog που κατασκευάσατε.

5. Το αρχείο KML που υπολογίστηκαν για τα δεδομένα που σας δόθηκαν (ένα για κάθε μέγεθος μετώπου αναζήτησης).

6. Τα αρχεία .csv για τα υπόλοιπα πειράματά σας και τα αντίστοιχα αρχεία KML που προέκυψαν.

Προσέξτε ότι σας δίνεται μεγάλη ευελιξία ως προς το ποια δεδομένα και πώς θα τα χρησιμοποιήσετε. Το ζητούμενο είναι να υλοποιήσετε τις βασικές παραμέτρους της υπηρεσίας και να αξιοποιήσετε όσο περισσότερα δεδομένα θεωρείτε ότι μπορούν να συνεισφέρουν στην καλύτερη επίτευξη των στόχων της υπηρεσίας.