

ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΗΛΕΚΤΡΟΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ ΥΗΕ ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ

ΣΥΣΤΗΜΑΤΑ ΚΑΙ ΤΕΧΝΟΛΟΓΙΕΣ ΓΩΝΣΗΣ 8° ΕΞΑΜΗΝΟ 2019-2020

ΕΡΓΑΣΤΗΡΙΑΚΟ ΘΕΜΑ

ΟΝΟΜΑΤΕΠΩΝΥΜΟ: ΓΙΑΝΝΙΟΣ ΓΕΩΡΓΙΟΣ-ΤΑΞΙΑΡΧΗΣ

ΑΡΙΘΜΟΣ ΜΗΤΡΩΟΥ: 031 16 156

Α. Περιγραφή Θέματος

Θεματικό Άξονα της Εργαστηριακού Θέματος αποτελεί η κατασκευή μιας Σημασιολογικής Βάσης Γνώσης η οποία περιλαμβάνει δεδομένα για μετακινήσεις Μέσων Μαζικής Μεταφοράς. Πιο ειδικά η οντολογία κατασκευάστηκε για τα Μέσα Μεταφοράς του ΟΑΣΑ, ωστόσο όμως οι Έννοιες, οι Ρόλοι και οι Ιδιότητες που χρησιμοποιήθηκαν ήταν γενικού χαρακτήρα, ώστε να μπορούν εύκολα να ενσωματωθούν και άλλα μέσα (πχ πλοία, αεροπλάνα, τρένα).

Β. Ρύθμιση Περιβάλλοντος

Για την επιτυχή εκτέλεση όλων των ζητούμενων ερωτημάτων, κρίθηκε αναγκαίο να διαμορφωθεί κατάλληλα το Περιβάλλον Εργασίας. Να σημειωθεί ότι το σύνολο της εργασίας πραγματοποιήθηκε σε Windows 10. Παρακάτω φαίνονται αναλυτικά τα Βήματα:

1. Εγκατάσταση Protégé v.5.5.0 για Windows 10:

Με τη βοήθεια του συγκεκριμένου προγράμματος, έγινε εφικτή η κατασκευή της Οντολογίας, δηλαδή ο ορισμός Εννοιών, Ρόλων και Ιδιοτήτων (Βλ Μέρος Γ.1)

2. Λήψη Δεδομένων για τις Μετακινήσεις Μέσων του ΟΑΣΑ:

Στη συνέχεια έγινε λήψη των ακόλουθων .txt (από τον παραπάνω υπερσύνδεσμο, και αποθηκεύτηκαν στο φάκελο εργασίας: C:\Users\USER\Desktop\gnosi\Thema\googletransit).Στις εικόνες φαίνονται αντιπροσωπευτικά δείγματα των επιμέρους αρχείων.:

routes.txt: Δεδομένα γραμμής και τύπου μέσου μεταφοράς

route_id,route_short_name,route_long_name,route_desc,route_type,route_color, route_text_color T3-20,T3,NEO ΦΑΛΗΡΟ - ΑΣΚΛΗΠΕΙΟ ΒΟΥΛΑ,,900,3FFF19,000000 T4-20,T4,"ΣΥΝΤΑΓΜΑ - NEO ΦΑΛΗΡΟ",,900,3FFF19,000000 T5-20,T5,"ΣΥΝΤΑΓΜΑ - ΒΟΥΛΑ",,900,3FFF19,000000 M2-20,M2,"ΑΝΘΟΥΠΟΛΗ - ΣΤ.ΕΛΛΗΝΙΚΟ",,1,D81Α13,000000

agency.txt: Πληροφορίες Φορέα παροχής των δεδομένων

stops.txt: Δεδομένα στάσεων

```
stop_id,stop_code,stop_name,stop_desc,stop_lat,stop_lon,location_type 010001,010001,ΣΤΡΟΦΗ,Επί της ΕΛ.ΒΕΝΙΖΕΛΟΥ,37.9986082641367,23.6649846246733,0 010002,010002,ΚΟΛΩΝΕΣ,Επί της ΛΕΩΦ.ΕΛ.ΒΕΝΙΖΕΛΟΥ,37.996719436997,23.6631823242898,0 010003,010003,AΓ.ΒΑΡΒΑΡΑ,Επί της ΛΕΩΦ.ΕΛ.ΒΕΝΙΖΕΛΟΥ,37.9946947354979,23.6611529618238,0 010004,010004,ΔΗΜΑΡΧΕΙΟ,Επί της ΗΠΕΙΡΟΥ,37.9928326565131,23.6592710105915,0 010005,010005,ΚΡΗΤΗΣ,Επί της ΠΑΛΑΙΩΝ ΠΟΛΕΜΙΣΤΩΝ,37.9904261625749,23.6561503633519,0
```

stop_times.txt: Δεδομένα δρομολογίων – χρόνου διέλευσης

```
trip_id,arrival_time,departure_time,stop_id,stop_sequence,pickup_type,drop_off_type
9803638-THΛΕΜΑ-T3-Παρασκευή-02,05:30:00,05:30:00,400045,1,0,0
9803638-THΛΕΜΑ-T3-Παρασκευή-02,05:31:00,05:31:00,400044,2,0,0
9803638-THΛΕΜΑ-T3-Παρασκευή-02,05:33:00,35:33:00,320025,3,0,0
9803638-THΛΕΜΑ-T3-Παρασκευή-02,05:35:00,05:35:00,240119,4,0,0
9803638-THΛΕΜΑ-T3-Παρασκευή-02,05:36:00,05:36:00,240038,5,0,0
9803638-THΛΕΜΑ-T3-Παρασκευή-02,05:38:00,05:38:00,380117,6,0,0
```

trips.txt: Δεδομένα γραμμής – δρομολογίων

3. Εγκατάσταση OpenLink Virtuoso

Αφού εγκαταστάθηκε το Virtuoso για Windows10, τροποιήθηκαν κάποιες παράμετροι στο αρχείο virtuoso.ini που βρισκόταν στο ακόλουθο path :

"C:\Program Files\OpenLink Software\Virtuoso OpenSource 7.2\database"

Πιο ειδικά οι αλλαγές αφορούσαν τα εξής:

 Προστέθηκε στη Γ66, το ακόλουθο path, ώστε να επιτρέπεται το ανέβασμα των 3-αδων RDF στη Βάση Γνώσης, από τη συγκεκριμένη τοποθεσία:

```
SchedulerInterval = 10

DirsAllowed = ., ../vad, C:/Users/USER/Desktop/gnosi/Thema

ThreadCleanupInterval = 0

ThreadThreshold = 10
```

2. Αλλάξαμε τις default τιμές:

```
A\pió NumberOfBuffers = 10000 , MaxDirtyBuffers = 6000 \sigma\epsilon NumberOfBuffers = 340000 MaxDirtyBuffers = 250000,
```

ώστε να αυξηθεί η απόδοση.

Γ. Αξιολόγηση

1. Παρουσίαση Οντολογίας

Όπως προαναφέρθηκε η κατασκευή της Οντολογίας πραγματοποιήθηκε με τη χρήση του Protégé. Στο 1° Tab λοιπόν, (Active Ontology) δώσαμε το IRI της Οντολογίας μας, το οποίο ήταν:

http://www.semanticweb.org/user/ontologies/2020/6/transport

Όπως προαναφέρθηκε, με τη χρήση του Protégé, κατασκευάστηκε η Οντολογία. Στόχο αποτελούσε η δημιουργία ενός αρχείου .ttl, στο οποίο θα συμπεριλαμβάνονταν :

Α. Ονόματα Κλάσεων

Από το Tab Classes, δημιουργήσαμε τις ακόλουθες έννοιες:

- Company: Όνομα της εταιρίας Μεταφορών-Συγκοινωνιών (πχ ΟΑΣΑ, Aegean, ΟΣΕ κλπ.)
- Schedules: Γενική κλάση η οποία περιλαμβάνει Δρομολόγια για διάφορα Μέσα Μεταφοράς
 - Airplane
 - o Bus
 - Metro
 - Train
 - Tram

StartsFrom

- Trolley
- Station : Περιλαμβάνει τις Στάσεις
- StationEnd : Στάση και Χρόνος Άφιξης (Ζεύγος Τιμών)
- StationStart : Στάση και Χρόνος Αναχώρησης (Ζεύγος Τιμών)
- TripsInfo : Η κλάση με τις Διαδρομές

Β.Ιδιότητες Αντικειμένων (Object Properties)

Από το Tab Object Properties, δημιουργήσαμε τις ακόλουθες Ιδιότητες Αντικειμένων:

- HasStation
 : Συνδέει Station και (StationStart ή StationEnd)
- OfCompany
 Serves
 Suvδέει Schedule και Company
 Συνδέει Schedule και TripsInfo
 StopsAt
 Συνδέει TripsInfo και StationEnd

Γ. Ιδιότητες Αντικειμένων (Data Properties)

- Coordinates : Προσδιορίζει συντεταγμένες ενός Station
- Information : Πληροφορία για κάποιο StationStart ή StationEnd (π.χ. μέρα)

: Συνδέει TripsInfo και StationStart

Name : Όνομα Station ή διαδρομής (TripsInfo)
 ΤimeInfo : Χρόνος Άφιξης ή Χρόνος Αναχώρησης

Η οντολογία αποθηκεύτηκε με τη χρήση του Protégé σε αρχείο ontology.ttl στο φάκελο εργασίας μας C:\Users\USER\Desktop\gnosi\Thema

2. Μετατροπή Δεδομένων σε 3-αδες RDF

Στη συνέχεια υλοποιήθηκε κώδικας ώστε να μετατραπούν τα δεδομένα σε 3άδες RDF. Ο κώδικας Python που περιλαμβάνεται στο αρχείο **RdfConverter.py**, δημιούργησε 4 διαφορετικά αρχεία ttl (routes, stops, stop_times, trips). Στη συνέχεια (χειροκίνητα για το καθένα), αλλάξαμε την κωδικοποίηση από ANSI σε UTF-8, ώστε να αναγνωρίζονται οι ελληνικοί χαρακτήρες και προσθέσαμε όλες τις δηλώσεις για τα προθέματα που βρίσκονται στην αρχή του ontology.ttl

Η διαδικασία θα παρουσιαστεί για ένα τύπο Δεδομένων(Ομοίως για τα υπόλοιπα). Έστω το αρχείο stops.txt (input), δείγμα του οποίου φαίνεται και στην σελ. 2 της Αναφοράς και περιέχει δεδομένα της μορφής:

010001,010001,ΣΤΡΟΦΗ,Επί της ΕΛ.ΒΕΝΙΖΕΛΟΥ,37.9986082641367,23.6649846246733,0

Στην ουσία λοιπόν διαβάζαμε πληροφορία η οποία χωρίζονταν με κόμμα. Αρχικά λοιπόν (και μέχρι το πρώτο κόμμα) διαβάσαμε τον κωδικό της στάσης (010001), ο οποίος αποτέλεσε το υποκείμενο όλων των παραπάνω προτάσεων. (Τοποθετήθηκε στην αρχή του print(), της κάθε πρότασης). Στη συνέχεια, απομονώσαμε το όνομα «ΣΤΡΟΦΗ», και την Πληροφορία «Επί της ΕΛ.ΒΕΝΙΖΕΛΟΥ» που βρίσκονταν και αυτά εκατέρωθεν του «,».

Με αυτόν τον απλό τρόπο (διαβάζοντας τιμές μεταξύ κομμάτων), και αποθηκεύοντας σε αρχείο με χρήση print(), καταφέραμε να μετατρέψουμε τα δεδομένα σε 3-αδες RDF

3. Παρουσίαση 3-άδων RDF

Παρακάτω παρουσιάζονται δείγματα εισόδων καθώς και εξόδων που προέκυψαν με τη χρήση κώδικα RdfConverter.py)

A.stops.txt (INPUT, Δεδομένα)

010001,010001,ΣΤΡΟΦΗ,Επί της ΕΛ.ΒΕΝΙΖΕΛΟΥ,37.9986082641367,23.6649846246733,0

A.stops.ttl (OUTPUT, Τριάδες RDF)

:s10001 rdf:type owl:NamedIndividual .

:s10001 rdf:type :Station .

:s10001 :OfCompany :OAΣA .

:s10001 :Coordinates "POINT(37.998608264136706 23.6649846246733)"^^vrdf:Geometry .

:s10001 :Name "ΣΤΡΟΦΗ" .

:s10001 :Information "Επί της ΕΛ.ΒΕΝΙΖΕΛΟΥ".

B.trips.txt (INPUT, Δεδομένα)

Τ3-20,ΤΗΛΕΜΑ-Τ3-Παρασκευή-02,9803638-ΤΗΛΕΜΑ-Τ3-Παρασκευή-02,ΝΕΟ ΦΑΛΗΡΟ - ΑΣΚΛΗΠΕΙΟ ΒΟΥΛΑ,0,2184458,

B.trips.ttl (OUTPUT, Τριάδες RDF)

```
:t9803638 rdf:type owl:NamedIndividual .
```

:t9803638 rdf:type :TripsInfo .

:t9803638:OfCompany:OAΣA.

:t9803638 :Name "NEO ΦΑΛΗΡΟ - ΑΣΚΛΗΠΕΙΟ ΒΟΥΛΑ" .

:rT3:Serves:t9803638.

C.stop_times.txt (INPUT, Δεδομένα)

9803638-ΤΗΛΕΜΑ-Τ3-Παρασκευή-02,05:30:00,05:30:00,400045,1,0,0

<u>C. stop_times.ttl (OUTPUT, Τριάδες RDF)</u>

:As0 rdf:type owl:NamedIndividual .

:As0 rdf:type :StationEnd .

:As0 :OfCompany :OAΣA .

:As0 :TimeInfo "05:30:00".

:As0 :Information "T3-Παρασκευή".

:As0:hasStation:s400045.

:t9803638 :StopsAt :As0 .

:Ds0 rdf:type owl:NamedIndividual .

:Ds0 rdf:type :StationStart .

:Ds0 :OfCompany :OAΣA .

:Ds0 :TimeInfo "05:30:00".

:Ds0 :Information "T3-Παρασκευή".

:Ds0:hasStation:s400045.

:t9803638 :StartsFrom :Ds0 .

.....

D.routes.txt (INPUT, Δεδομένα)

T3-20,T3,NEO ΦΑΛΗΡΟ - ΑΣΚΛΗΠΕΙΟ ΒΟΥΛΑ,,900,3FFF19,000000

D.routes.ttl (OUTPUT,Τριάδες RDF)

:rT3 rdf:type owl:NamedIndividual.

:rT3 rdf:type :Tram .

:rT3:OfCompany:OAΣA.

:rT3 :Name "NEO ΦΑΛΗΡΟ - ΑΣΚΛΗΠΕΙΟ ΒΟΥΛΑ" .

4. Ερωτήματα SPARQL

Τέλος για να τεθούν ερωτήματα SQL, ήταν αναγκαίο να ανέβουν τα δεδομένα στο server. Πρώτα από όλα ελέγχθηκε αν είναι αρχικοποιημένη η βάση. (Θα πρέπει στο C:\Program Files\OpenLink Software\Virtuoso OpenSource 7.2\database να βρίσκεται μόνο το αρχείο virtuoso.ini. Σε αντίθετη περίπτωση είναι καλό να διαγραφούν τα υπόλοιπα αρχεία αυτού του φακέλου (εκτός βέβαια από το .ini).

Έπειτα, ανοίξαμε το cmd (**με Δικαιώματα Διαχειριστή**) στο φάκελο bin της Εγκατάστασης και εκκινήσαμε τον server με την εντολή:

Virtuoso-t +foreground +configfile ../database/virtuoso.ini¹

Μήνυμα Επιτυχούς Εκκίνησης (στο Τερματικό):

Server online at 1111

Στη συνέχεια, πραγματοποιήθηκε είσοδος στο <u>Virtuoso Conductor</u>, με στοιχεία:

Account: dba,

Password: dba

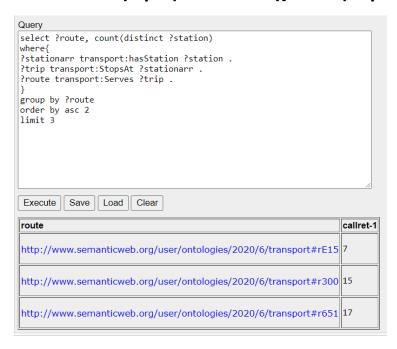
,ώστε στο Tab Namespaces του Linked Data, να προστεθεί το prefix: transport και το URI:

Σε άλλο τερματικό, στον υποφάκελο: C:\Users\USER\Desktop\gnosi\Thema, με τη χρήστη της εντολής **isql** πραγματοποιήθηκε είσοδος στο περιβάλλον του database. Στη συνέχεια φορτώσαμε τα αρχεία .ttl πληκτρολογώντας τις εντολές:

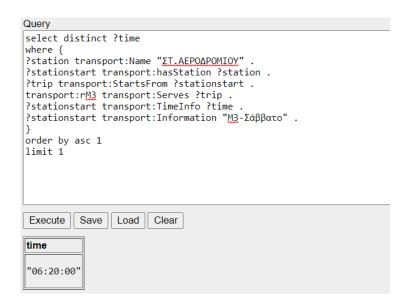
¹ Προσοχή!!!! Με copy-paste η εντολή δεν έβγαζε επιθυμητό αποτέλεσμα καθώς το τερματικό δεν αναγνώριζε το σύμβολο της «-».

Τέλος μεταβήκαμε στο tab **Linked Data** και πιο ειδικά στο tab **SPARQL**, όπου κατασκευάσαμε ερωτήματα SPARQL:

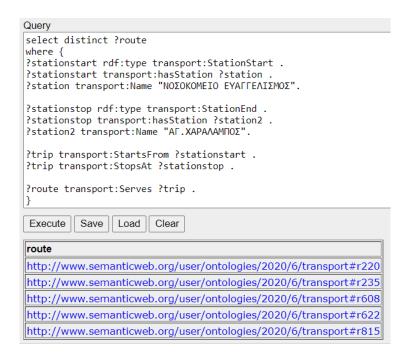
Α. Εμφάνιση (3) Μέσων Μεταφοράς τα οποία έχουν τις λιγότερες στάσεις



Β.Πότε φεύγει το πρώτο Μετρό από το «ΑΕΡΟΔΡΟΜΙΟ», την ημέρα του ΣΑΒΒΑΤΟΥ



Γ. Ποια Μέσα Μεταφοράς εξυπηρετούν τη διαδρομή «ΝΟΣΟΚΟΜΕΙΟ ΕΥΑΓΓΕΛΙΣΜΟΣ»-«ΑΓ.ΧΑΡΑΛΑΜΠΟΣ»



5. Προβλήματα – Περιορισμοί

Στη συνέχεια παρατίθενται κάποια Προβλήματα –Περιορισμοί που εντοπίστηκαν κατά τη διαδικασία κατασκευής της Βάσης Γνώσης:

- Πρώτον έγινε αντιληπτό ότι, κάποιες στάσεις, αν και αναφέρονταν στην ίδια περιοχή (παρόμοιες συντεταγμένες), είχαν διαφορετικό όνομα. Παραδείγματος χάριν, ο Σταθμός της Γραμμής 3 ονομαζόταν «ΣΤ.ΔΟΥΚ.ΠΛΑΚΕΝΤΙΑΣ», ενώ η αντίστοιχη στάση του λεωφορείου «ΣΤ.Δ.ΠΛΑΚΕΝΤΙΑΣ-ΓΕΦΥΡΑ». Θα ήταν επιθυμητό να υπήρχε μια γενική ονομασία, που να αναφέρεται στην περιοχή, ώστε να καθίσταται πιο εύκολη η αναζήτηση.
- Δεύτερον, ένα πρόβλημα που προέκυψε αφορούσε την χρήση ελληνικών χαρακτήρων έναντι αγγλικών. Για παράδειγμα στο δεύτερο SPARQL ερώτημα, όπου αναζητούσαμε το πρώτο δρομολόγιο για τη γραμμή Μετρο 3, πληκτρολογήσαμε transport:M3, όπου το «M» έπρεπε να είναι στα αγγλικά. Αν πληκτρολογούσαμε M3 με ελληνικούς χαρακτήρες, τότε δεν θα εμφανίζει έξοδο το SPARQL ερώτημα, με αποτέλεσμα ο χρήστης να έχει τη Ψευδαίσθηση ότι δεν υπάρχει απάντηση για το ερώτημά του. Θα ήταν επιθυμητό τα δεδομένα μας, να είχαν άλλα ονόματα ώστε να διαφαίνεται η ύπαρξη ελληνικών αγγλικών χαρακτήρων (π.χ. ΜΕΤRO3)

- Τρίτον, οφείλουμε να παραδεχτούμε ότι το αρχείο stop_times.ttl (που προέκυψε από το μετασχηματισμό του stop_times.txt), ήταν αδύνατο να ανοίξει με απλό Σημειωματάριο καθώς είχε μεγάλο όγκο δεδομένων, ενώ η παραμικρή ενέργεια στο αρχείο αυτό, επιβάρυνε τον Η/Υ.
- Τέταρτον , ήρθαμε αντιμέτωποι με ένα πρόβλημα που αφορούσε τη κωδικοποίηση UTF-8. Τα αρχεία που δημιούργησε ο κώδικας RdfConverter.py, (Στην ουσία οι 3άδες RDF), είχαν κωδικοποίηση ANSI με αποτέλεσμα όταν ανεβάσαμε τα .ttl αρχεία στον server να μην αναγνωρίζει τους ελληνικούς χαρακτήρες. Όταν λοιπόν θέταμε ένα ερώτημα που περιλάμβανε ελληνικούς χαρακτήρες σε κάποια 3-άδα (π.χ. «ΑΓ.ΧΑΡΑΛΑΜΠΟΣ»), αυτή δεν γινόταν κατανοητή από το Σύστημα Γνώσης, με αποτέλεσμα να μην εμφανίζεται καμία έξοδος. Και πως έγινε αντιληπτό; Στην ουσία ζητήσαμε από το σύστημα να εκτυπώσει κάποια ονόματα σταθμών με βάση συντεταγμένες. Το αποτέλεσμα ήταν απροσδιόριστοι χαρακτήρες στην έξοδο («????????») του SPARQL ερωτήματος. Οπότε αλλάξαμε τη κωδικοποίηση από ANSI σε UTF-8 και επαναλάβαμε το ανέβασμα των .ttl σε Server.

Παρόλα αυτά, βρέθηκαν λύσεις και συνεχίστηκε απρόσκοπτα η διαδικασία κατασκευής Βάσης Γνώσης, γεγονός που αποτυπώθηκε στην επιτυχή υποβολή ερωτημάτων SPARQL.

ΤΕΛΟΣ ΘΕΜΑΤΟΣ.