

Αναφορά Θέματος στο Μάθημα «Συστήματα και Τεχνολογίες Γνώσης»

Όνομα Φοιτητή: Βουλγαρίδης Ιωάννης

A.M. : 031 16151

Εξάμηνο : 8

Κατασκευή Οντολογίας

Για την εκπόνηση της εργασίας αναπτύχθηκε μια οντολογία χρησιμοποιώντας το γραφικό εργαλείο Protégé. Η οντολογία προέκυψε σε μεγάλο βαθμό από τα δεδομένα του ΟΑΣΑ που είχαμε στην διάθεση μας. Οι κλάσεις(έννοιες) που υλοποιήθηκαν είναι οι κλάσεις **Agency** (υποκλάση της βασικής κλάσης της owl) η οποία αναπαριστά των οργανισμό παροχής των συγκοινωνιακών υπηρεσιών και περιέχει μοναδικό στιγμιότυπο της τον ΟΑΣΑ, αφού εστιάζουμε μόνο στην πόλη της Αθήνας. Επίσης, η έννοια **Routes** είναι υποέννοια της owl:thing και περιέχει τα μέσα. Χωρίζεται σε επιμέρους υποέννοιες **Bus, Tram, Trolley, Metro** κλπ για να καλύψει όλο το εύρος των πιθανών μέσων μεταφοράς. Άλλες κλάσεις είναι επίσης η έννοια **Trips** η οποία περιέχει την έννοια των διαδρομών και είναι υποκλάση της owl:Thing. Επίσης υπάρχει και η **Station**, που περιλαμβάνει τους Σταθμούς. Οι έννοιες **StationArrival** και **StationDeparture** είναι κλάσεις που σχηματίζονται ουσιαστικά από τούπλες σταθμών και χρόνων (άφιξης και αναχώρησης αντίστοιχα) και είναι χρήσιμες έννοιες για τα ερωτήματα που εμπλέκουν χρόνο άφιξης σε κάποιο σταθμό.

Οι ρόλοι που αναπτύχθηκαν είναι **belongsAgency** (με range την έννοια Agency) η οποία δείχνει σε ποιον Οργανισμό ανήκει κάποιο μέσο, σταθμός ή διαδρομή, ο ρόλος **Serves** (domain: Route, Range: Trip) που δείχνει ποιο μέσο (Route) εκτελεί ποια διαδρομή (Trip) και η έννοια **hasStation** που συσχετίζει τα στιγμιότυπα των κλάσεων StationArrival και StationDeparture με τις στάσεις (Stations) στις οποίες αναφέρονται. Υπάρχουν επίσης οι ρόλοι **arrivesAt**, **departsAt** (domain: trip, range: stationArrival/stationDeparture). Οι ρόλοι αυτοί δείχνουν από ποίο σταθμό αναχώρησης/άφιξης περνάει ένα δρομολόγιο.

Τα data properties που χρειάστηκαν είναι το **Coordinates** που περιέχει τις γεωγραφικές συντεταγμένες ενός σταθμού, το **Name** που περιέχει ένα string και περιγράφει το όνομα ενός Σταθμού (Station), Μέσου (Route) ή διαδρομής (Trip). Η ιδιότητα **Description** είναι επίσης ένα string που περιέχει περιγραφή για κάποιο μέσο, σταθμό ή διαδρομή και τέλος η ιδιότητα **onTime** που αφορά τους σταθμούς άφιξης και προορισμού και δείχνει την ώρα άφιξης/αναχώρησης αντίστοιχα.

Μετατροπή δεδομένων σε RDF τριάδες και εισαγωγή

Το επόμενο στάδιο για την εργασία είναι η επεξεργασία των δεδομένων που κατεβάσαμε από .csv μορφή σε rdf τριάδες. Αυτό γίνεται με χρήση Python 3.8. Ο κώδικας υπάρχει στα αρχεία της εργασίας με το όνομα `cast_to_rdf.py`. Σημειώνεται εδώ ότι τα αρχεία έχουν .txt κατάληξη αλλά η δομή τους ακολουθεί το πρότυπο του csv και έτσι μπορούμε να τα διαβάσουμε σαν .csv. Η λειτουργία που επιτελεί είναι αρκετά απλή: Ανοίγουμε τα αρχεία που κατεβάσαμε και κάνουμε parsing σαν .csv απομονώνοντας τα πεδία. Ανάλογα με το αρχείο, σχηματίζουμε τις τριάδες συνδυάζοντας τα κατάλληλα πεδία και τα αποθηκεύουμε σε νέο αρχείο με ίδιο όνομα και κατάληξη .ttl.

Για παράδειγμα, από το αρχείο `routes.txt`:

```
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
```

Και στο αντίστοιχο `routes.ttl`:

```
:rT3 rdf:type owl:NamedIndividual .
:rT3 rdf:type :Tram .
:rT3 :belongsAgency :OASA .
:rT3 :Name "NEO ΦΑΛΗΡΟ - ΑΣΚΛΗΠΕΙΟ ΒΟΥΛΑ" .
```

Επίσης, από το αρχείο `stop.txt`:

```
Αρχείο Επεξεργασία Μορφή Προβολή Βοήθεια
stop_id,stop_code,stop_name,stop_desc,stop_lat,stop_lon,location_type
010001,010001,ΣΤΡΟΦΗ,Επί της ΕΛ.BENIZEΛΟΥ,37.9986082641367,23.6649846246733,0
```

Και στο αντίστοιχο `stops.ttl`:

```
:s10001 rdf:type owl:NamedIndividual .
:s10001 rdf:type :Station .
:s10001 :belongsAgency :OASA .
:s10001 :Coordinates "POINT(37.998608264136706 23.6649846246733)"^^vrd:Geometry .
:s10001 :Name "ΣΤΡΟΦΗ" .
:s10001 :Description "Επί της ΕΛ.BENIZEΛΟΥ" .
```

Σημειώνεται εδώ ότι σε ότι αφορά πεδία id προσαρτάται ένα επιπλέον γράμμα

Αφού σχηματιστούν οι τριάδες, πρέπει να φορτωθούν στο virtuoso όλα τα αρχεία που παράχθηκαν. Για τον λόγο, αυτό στο αρχείο virtuoso.ini εντοπίζουμε την παράμετρο DirsAllowed και προσθέτουμε το μονοπάτι C:/Users/user/Desktop/stg/ που περιέχει τα αρχεία. Στην συνέχεια με δικαιώματα διαχειριστή εκτελούμε από τον φάκελο στον οποίο είναι εγκατεστημένο το virtuoso (bin) **εντολή virtuoso +foreground +configfile ../database/virtuoso.ini** για να εκκινήσει ο server.

SPARQL Ερωτήματα

1. Κατάλογος όλων των λεωφορείων της αθήνας

```
select distinct ?route
where{
  ?route rdf:type transport:Bus. }
```

Η «απάντηση» στο παραπάνω ερώτημα είναι:

route
http://www.semanticweb.org/johnvoul/ontologies/2020/6/transport#r021
http://www.semanticweb.org/johnvoul/ontologies/2020/6/transport#r022
http://www.semanticweb.org/johnvoul/ontologies/2020/6/transport#r024
http://www.semanticweb.org/johnvoul/ontologies/2020/6/transport#r025
http://www.semanticweb.org/johnvoul/ontologies/2020/6/transport#r026
http://www.semanticweb.org/johnvoul/ontologies/2020/6/transport#r027
http://www.semanticweb.org/johnvoul/ontologies/2020/6/transport#r032
http://www.semanticweb.org/johnvoul/ontologies/2020/6/transport#r035
http://www.semanticweb.org/johnvoul/ontologies/2020/6/transport#r036
http://www.semanticweb.org/johnvoul/ontologies/2020/6/transport#r040
http://www.semanticweb.org/johnvoul/ontologies/2020/6/transport#r046
http://www.semanticweb.org/johnvoul/ontologies/2020/6/transport#r049
http://www.semanticweb.org/johnvoul/ontologies/2020/6/transport#r051
http://www.semanticweb.org/johnvoul/ontologies/2020/6/transport#r054
http://www.semanticweb.org/johnvoul/ontologies/2020/6/transport#r057
http://www.semanticweb.org/johnvoul/ontologies/2020/6/transport#r060
http://www.semanticweb.org/johnvoul/ontologies/2020/6/transport#r100
http://www.semanticweb.org/johnvoul/ontologies/2020/6/transport#r101
http://www.semanticweb.org/johnvoul/ontologies/2020/6/transport#r106
http://www.semanticweb.org/johnvoul/ontologies/2020/6/transport#r109
http://www.semanticweb.org/johnvoul/ontologies/2020/6/transport#r112
http://www.semanticweb.org/johnvoul/ontologies/2020/6/transport#r120
http://www.semanticweb.org/johnvoul/ontologies/2020/6/transport#r122
http://www.semanticweb.org/johnvoul/ontologies/2020/6/transport#r123
http://www.semanticweb.org/johnvoul/ontologies/2020/6/transport#r126
http://www.semanticweb.org/johnvoul/ontologies/2020/6/transport#r130
http://www.semanticweb.org/johnvoul/ontologies/2020/6/transport#r131
http://www.semanticweb.org/johnvoul/ontologies/2020/6/transport#r135

2. Κατάλογος λεωφορείων που παίρνου από την στάση «8^η Ζωγράφου»

```

select distinct ?route
where {
    ?stationarr rdf:type transport:StationArrival .
    ?stationarr transport:hasStation ?station .
    ?station transport:Name "8η ΖΩΓΡΑΦΟΥ".
    ?trip transport:arrivesAt ?stationarr .
    ?route transport:Serves ?trip . }

```

Η απάντηση σε αυτό το ερώτημα είναι:

route
http://www.semanticweb.org/johnvoul/ontologies/2020/6/transport#r608
http://www.semanticweb.org/johnvoul/ontologies/2020/6/transport#r230
http://www.semanticweb.org/johnvoul/ontologies/2020/6/transport#r235

Όπου πράγματι τα λεωφορεία 608,230,235 είναι τα μοναδικά που εξυπηρετούν αυτή την στάση.

3. Τι ώρα περνάει το πρώτο δρομολόγιο από την στάση «8^η Ζωγράφου» και ποίο λεωφορείο εκτελεί το δρομολόγιο;

```
select distinct ?route , ?time
where{
    ?station rdf:type transport:Station.
    ?station transport:Name "8η ΖΩΓΡΑΦΟΥ".
    ?statArr transport:hasStation ?station.
    ?statArr transport:OnTime ?time.
    ?trip transport:arrivesAt ?statArr.
    ?route transport:Serves ?trip
}
order by asc 2 1
limit 1
```

Η απάντηση για το παραπάνω ερώτημα είναι:

route	time
http://www.semanticweb.org/johnvoul/ontologies/2020/6/transport#r230	"05:03:00"

4. Ποιες είναι οι συντεταγμένες των στάσεων με όνομα «8^η Ζωγράφου»;

```
select distinct ?coords, ?desc
where{
    ?station rdf:type transport:Station.
    ?station transport:Name "8η ΖΩΓΡΑΦΟΥ".
    ?station transport:Coordinates ?coords.
    ?station transport:Description ?desc. }
```

Η απάντηση σε αυτό το ερώτημα είναι:

coords	desc
"POINT(37.975664888452 23.776478337453)"^^<http://www.openlinksw.com/schemas/virtrdf#Geometry>	"Επί της ΚΟΥΣΙΔΟΥ"
"POINT(37.976367542418 23.776294028466)"^^<http://www.openlinksw.com/schemas/virtrdf#Geometry>	"Επί της ΛΕΩΦ.ΣΤΡΧΟΥ ΠΑΠΑΓΟΥ"
"POINT(37.974028606055 23.773887342054)"^^<http://www.openlinksw.com/schemas/virtrdf#Geometry>	"Επί της ΓΕΩΡ.ΖΩΓΡΑΦΟΥ"
"POINT(37.974704195718 23.773691709752)"^^<http://www.openlinksw.com/schemas/virtrdf#Geometry>	"Επί της ΓΕΩΡ.ΖΩΓΡΑΦΟΥ"

5. Ποιος είναι ο πλησιέστερος σταθμός μετρό από την στάση «8^η ΖΩΓΡΑΦΟΥ»;

```
select ?name, ?route ,(bif:st_distance (?coords,?statCoords))
where{
    ?station rdf:type transport:Station.
    ?station transport:Name "8η ΖΩΓΡΑΦΟΥ" .
    ?station transport:Coordinates ?statCoords.
    ?route rdf:type transport:Metro.
    ?route transport:Serves ?trip.
    ?trip transport:arrivesAt ?stat.
    ?stat transport:hasStation ?station2.
    ?station2 transport:Coordinates ?coords.
    ?station2 transport:Name ?name.
}
order by asc 3
limit 1
```

Και η απάντηση στο παραπάνω ερώτημα είναι:

name	route	callret-2
"ΣΤ.ΚΑΤΕΧΑΚΗ"	http://www.semanticweb.org/johnvoul/ontologies/2020/6/transport#rM3	1.70861

Προβλήματα – Περιορισμοί

Η παραπάνω οντολογία παράγει κάποια ικανοποιητικά αποτελέσματα αλλά παρουσιάζονται επίσης αρκετά και σημαντικά προβλήματα. Αρχικά, το πρώτο πρόβλημα είναι η παρουσία ελληνικών χαρακτήρων. Χρειάστηκε προσοχή ώστε όλα τα αρχεία να ανοίγονται με utf-8 – και αυτά στα οποία έγραφε το python script και αυτά από τα οποία διάβαζε . Σε αντίθετη περίπτωση, τα ερωτήματα επέστρεφαν τις σωστές απαντήσεις συν κάποιες επιπλέον με μη αλφαριθμητικούς χαρακτήρες στα πεδία Name και Description που ήταν strings. Επίσης, η ανάπτυξη και ο έλεγχος ορθής λειτουργίας της εφαρμογής έγινε σε Windows10 και συνεπώς δεν είναι γνωστό στην παρούσα φάση (κατά την οποία συντάχθηκε η αναφορά) αν χρειάζονται επιπλέον ρυθμίσεις συστήματος ή αλλαγές στον κώδικα για να λειτουργήσει. Άλλος ουσιαστικός περιορισμός της εφαρμογής είναι η δυσκολία στην συντήρηση και την ενημέρωση των δρομολογίων που πρέπει να γίνεται χειροκίνητα ξανατρέχοντας το python script με ανανεωμένα .csv αρχεία.