**Μελέτη Περίπτωσης: Αναζήτηση και αξιοποίηση των Ανοιχτών Συνδεδεμένων Δεδομένων του LOD Syndesis Dataset από τον Ιστότοπο Datahub.**

Η παρούσα ερευνητική μελέτη, που σχετίζεται με την αναζήτηση Ανοιχτών Συνδεδεμένων Δεδομένων του LOD Syndesis Dataset από τον Ιστότοπο Datahub, και ακολούθως με τη συλλογή και την αξιοποίησή τους, συντάχθηκε από τους μεταπτυχιακούς φοιτητές : Κιόρτση Ευάγγελο και Κυριαζόπουλο Κωνσταντίνο στα πλαίσια του μαθήματος «Τεχνολογίες Ιστού και Ανάλυση Δεδομένων Ιστού» του Προγράμματος Μεταπτυχιακών Σπουδών στην Εφαρμοσμένη Πληροφορική, του Πανεπιστημίου Μακεδονίας στην κατεύθυνση Ηλεκτρονικό Επιχειρείν και Τεχνολογίες Καινοτομίας. Επιβλέπων καθηγητής είναι ο κ. Ταμπούρης Ευθύμιος.

Η παρουσίαση της συγκεκριμένης μελέτης έχει ως στόχο να πληροφορηθούν οι αναγνώστες τι είναι τα Ανοιχτά Συνδεδεμένα Δεδομένα (Linked Open Data), να κατανοήσουν την αξία και την χρησιμότητας τους και να εξοικειωθούν, στον βαθμό που είναι εφικτό, με ορολογίες που αφορούν τα εν λόγω δεδομένα. Η παρουσίαση της περίπτωσης (showcase) που ακολουθεί αφορά το σετ δεδομένων (ή Πύλη δεδομένων) ‘LOD Syndesis Dataset’ (με URL: <https://datahub.io/dataset/connectivity-of-lod-datasets>) και τα επί μέρους δεδομένα που περιλαμβάνονται στο παραπάνω σετ ανακτήθηκαν από την βάση Datahub χρησιμοποιώντας την γλώσσα SPARQL η οποία αποτελεί μια γλώσσα ερωτημάτων που έχει σχεδιαστεί με στόχο να ανακτά ο εκάστοτε ενδιαφερόμενος δεδομένα από σχεσιακές βάσεις δεδομένων. Επίσης, γίνεται γραφική αναπαράσταση των αποτελεσμάτων από δύο ερωτήματα σε γλώσσα SPARQL, συνοδευόμενα από τις αντίστοιχες αναλύσεις. Τόσο για τις γραφικές οπτικοποιήσεις όσο και για τις αναλύσεις αυτών χρησιμοποιήθηκαν τα λογισμικά Tableau, Exploratory, RapidMiner και Sgvizler.

Η Πύλη ‘Lod Syndesis Dataset’ περιλαμβάνει τα αποτελέσματα μετρήσεων που διεξήχθηκαν για ένα μεγάλο αριθμό Ανοιχτών Συνδεδεμένων Δεδομένων (Linked Open Data). Συγκεκριμένα, έχουν συλλεχθεί 302 RDF datasets και έχουν γίνει μετρήσεις όσον αφορά τον πλήθος των κοινών αντικειμένων (common Real World Objects - RWO) που μοιράζονται οποιαδήποτε από τα παραπάνω 302 datasets μεταξύ τους. Ο αριθμός των κοινών αντικειμένων (commonRWO) είναι στην ουσία ο αριθμός των κλάσεων (classes) που έχουν ισοδύναμα URIs. Για να γίνει αυτό πιο κατανοητό μπορούμε να αναφέρουμε το παράδειγμα με το αντικείμενο (Object): ‘Αριστοτέλης’ (Aristotle). Όταν θα ζητήσουμε από την Πύλη (μέσω της διατύπωσης του κατάλληλου ερωτήματος με την χρήση της γλώσσας SPARQL στο αντίστοιχο, βεβαίως, πλαίσιο ερωτημάτων/εντολών ή SPARQL Endpoint δηλαδή στο πλαίσιο όπου μπορούμε να πληκτρολογήσουμε τα ερωτήματα που επιθυμούμε) να μας επιστρέψει τα dataset εκείνα από το σύνολο των 302 που περιλαμβάνονται, που έχουν ως κοινό αντικείμενο τον ‘Αριστοτέλη’ τότε θα δούμε πως μας γυρνάει ένα πλήθος από datasets.

Επιπλέον, η Πύλη δεδομένων περιέχει τον αριθμό των κοινών αντικειμένων των 302 datasets ανά ζευγάρια (Pairs), ανά τριάδες (Triads), ανά τετράδες (Quads), ανά πεντάδες (Quinters) και ανά εξάδες (Hexads) δηλαδή συνδυασμούς μεταξύ 2, 3, 4,5 και 6 datasets, αντίστοιχα. Επιπροσθέτως, τα ζευγάρια των Common Real World Objects (CRWO) είναι τύπου rdf:type και έχουν 56.952 τριπλέτες, 1 κλάση, 7 properties, 7119 Υποκείμενα και 1106 Αντικείμενα. Οι τριάδες των Common Real World Objects (CRWO) είναι τύπου rdf:type και έχουν 778860 τριπλέτες, 1 κλάση, 7 properties, 86644 Υποκείμενα και 2091 Αντικείμενα. Οι τετράδες των Common Real World Objects (CRWO) είναι τύπου rdf:type και έχουν 8993130 τριπλέτες, 1 κλάση, 7 properties, 900932 Υποκείμενα και 3259 Αντικείμενα. Οι πεντάδες των Common Real World Objects (CRWO) είναι τύπου rdf:type και έχουν 10906576 τριπλέτες, 1 κλάση, 7 properties, 993490 Υποκείμενα και 3832 Αντικείμενα. Οι εξάδες των Common Real World Objects (CRWO) είναι τύπου rdf:type και έχουν 8109960 τριπλέτες, 1 κλάση, 7 properties, 677046 Υποκείμενα και 3735 Αντικείμενα. Τέλος, τα Datasets Used που χρησιμοποιήθηκαν σε πειράματα και είναι τύπου rdf:type και έχουν 1208τριπλέτες, 1 κλάση, 4 properties, 302 Υποκείμενα και 608 Αντικείμενα.