

15th Hellenic Data Management Symposium (HDMS'2017), August 2017, Athens, Greece

(based on Michalis Mountantonakis, Yannis Tzitzikas: [On Measuring the Lattice of Commonalities Among Several Linked Datasets](#). PVLDB 9(12): 1101-1112, 2016, http://users.ics.forth.gr/~tzitzik/publications/Tzitzikas_2016_VLDB.pdf)

Μετρώντας το Πλέγμα των Κοινών Στοιχείων μεταξύ πολλών Διασυνδεδεμένων Συνολοδεδομένων

Μιχάλης Μουνταντωνάκης και Γιάννης Τζίτζικας

Ινστιτούτο Πληροφορικής, Ίδρυμα Τεχνολογίας Έρευνας, και

Τμήμα Επιστήμης Υπολογιστών, Πανεπιστήμιο Κρήτης

Ένα μεγάλο πλήθος συνολοδεδομένων έχει δημοσιευτεί σύμφωνα με τους κανόνες των Συνδεδεμένων Δεδομένων (Linked Data), στο εξής πηγές, και ο αριθμός και το μέγεθος τους διαρκώς αυξάνεται. Παρ' όλο που ο κύριος στόχος τους είναι η διασύνδεση και η ολοκλήρωση, αυτή τη στιγμή είναι δύσκολο να εκτιμήσει κανείς πόσο συνδεδεμένες είναι οι πηγές του νέφους Ανοικτών Διασυνδεδεμένων Δεδομένων (LOD Cloud). Μετρήσεις (και ευρετήρια) που περιλαμβάνουν περισσότερες από δύο πηγές δεν είναι διαθέσιμες παρ' όλο που είναι σημαντικές: (α) για την εύρεση όλης της πληροφορίας που αφορά σε ένα συγκεκριμένο URI (ή σύνολο από URIs) και την προέλευσή της, (β) για να βοηθήσει την ανακάλυψη και επιλογή πηγής, (γ) για την εκτίμηση της ποιότητας του βαθμού διασύνδεσης μεταξύ οποιουδήποτε συνόλου πηγών και την παρακολούθησή της εξέλιξης της στο χρόνο, (δ) για την κατασκευή οπτικοποιήσεων που προσφέρουν πλουσιότερη επισκόπηση. Επειδή θα ήταν απαγορευτικά ακριβός ο υπολογισμός όλων αυτών των μετρήσεων ακολουθώντας μία απλοϊκή μέθοδο, σε αυτό το άρθρο παρουσιάζουμε ευρετήρια (και τους αλγόριθμους κατασκευής τους) που μπορούν να επιταχύνουν όλη τη διαδικασία. Συνοπτικά, παρουσιάζουμε i) ένα ευρετήριο για τα προθέματα των URIs ii) έναν κατάλογο για σχέσεις ισοδυναμίας που λαμβάνει υπ' όψιν του το συμμετρικό και μεταβατικό κλείσιμο των συσχετίσεων ισοδυναμίας που εμφανίζονται στα σύνολα δεδομένων, iii) ένα σημασιολογικό ευρετήριο στοιχείων (που αξιοποιεί τα προαναφερθέντα ευρετήρια) και τέλος iv) δύο αυξητικούς αλγορίθμους που επιταχύνουν τον υπολογισμό του πλέγματος (lattice) των κοινών στοιχείων ενός συνόλου πηγών. Αξιολογούμε την επιτάχυνση που επιτυγχάνεται με τα προτεινόμενα ευρετήρια και αλγορίθμους μέσω συγκριτικών αποτελεσμάτων και τέλος εκθέτουμε τα αποτελέσματα μετρήσεων που δεν έχουν γίνει στο παρελθόν για τη συνδεσιμότητα των πηγών που ανήκουν στο νέφος Ανοικτών Διασυνδεδεμένων Δεδομένων.