

ΘΕΜΑΤΑ ΔΙΠΛΩΜΑΤΙΚΩΝ ΕΡΓΑΣΙΩΝ
Εργ. Συστημάτων Βάσεων Γνώσεων & Δεδομένων

**ΑΞΙΟΛΟΓΗΣΗ ΣΥΣΤΗΜΑΤΩΝ ΔΙΑΧΕΙΡΙΣΗΣ RDF ΔΕΔΟΜΕΝΩΝ
ΣΕ ΕΦΑΡΜΟΓΕΣ ΔΙΑΣΥΝΔΕΔΕΜΕΝΩΝ ΔΕΔΟΜΕΝΩΝ (LINKED DATA)**

ΠΛΗΡΟΦΟΡΙΕΣ: Γιάννης Λιαγούρης, 210 7721436, liagos [ατ] dlab.ntua.gr

Νίκος Μπικάκης, 210 772 1402, bikakis [ατ] dlab.ntua.gr

ΠΕΡΙΛΗΨΗ: Σκοπός της εργασίας είναι η πειραματική αξιολόγηση συστημάτων διαχείρισης RDF δεδομένων (RDF Stores) με χρήση διασυνδεδεμένων δεδομένων (Linked Data) από πραγματικές εφαρμογές.

ΑΤΟΜΑ: 1-2

ΠΛΑΤΦΟΡΜΑ ΕΡΓΑΣΙΑΣ: Java/C++

ΣΥΝΤΟΜΗ ΠΕΡΙΓΡΑΦΗ: Το μοντέλο RDF (Resource Description Framework) [1] σχεδιάστηκε αρχικά για τον τυπικό προσδιορισμό μεταδεδομένων σχετικών με τους πόρους του παγκόσμιου ιστού, αλλά στην πραγματικότητα μπορεί να αναπαραστήσει οποιοδήποτε είδους δεδομένα. Η δομική μονάδα του RDF είναι η τριάδα <υποκείμενο, κατηγορήμα, αντικείμενο> (<subject, predicate, object>) που ονομάζεται και δήλωση (statement). Δεδομένου ότι το υποκείμενο μιας δήλωσης μπορεί να αποτελεί αντικείμενο κάποιας άλλης, πολλές τέτοιες RDF δηλώσεις συνθέτουν έναν κατευθυνόμενο γράφο με κόμβους subject και object και ακμές (predicates) που αντιστοιχούν στις σχέσεις μεταξύ τους.

Καθοριστικό ρόλο στην προσπάθεια διασύνδεσης δεδομένων κατέχει το *Linked Data Project (LDP)* [3]. Στο LDP, εκατοντάδες σύνολα δεδομένων (datasets) από διάφορους οργανισμούς δημοσιεύονται και συνδέονται μεταξύ τους ακολουθώντας το μοντέλο RDF. Ενδεικτικό της απήχησης του LDP είναι το γεγονός ότι ο όγκος των δεδομένων που ήδη φιλοξενεί αγγίζει τα 25 εκατ. RDF τριάδων, με σύνολα δεδομένων που φέρουν πάνω από 400 εκατ. διασυνδέσεις. Προφανώς, όσο ο όγκος των δεδομένων αυτών μεγαλώνει, τόσο οι τεχνικές προκλήσεις που αφορούν στην αποδοτική διαχείρισή τους αυξάνονται. Αν και μέχρι στιγμής έχουν αναπτυχθεί αρκετά συστήματα διαχείρισης RDF δεδομένων, τα πλεονεκτήματα της κάθε προσέγγισης δεν είναι εντελώς ξεκάθαρα, καθότι η βιβλιογραφία στερείται μιας πλήρους πειραματικής αξιολόγησης όλων των διαφορετικών προσεγγίσεων.

Σκοπός της παρούσας εργασίας είναι η πειραματική αξιολόγηση των συστημάτων διαχείρισης διασυνδεδεμένων RDF δεδομένων, με έμφαση σε συστήματα που προσανατολίζονται στην αποδοτική αποτίμηση ερωτημάτων στο σκληρό δίσκο. Επομένως, η εργασία θα ασχοληθεί με τα εξής θέματα:

- Μελέτη και εξοικείωση με τις γλώσσες περιγραφής (μετα)δεδομένων σε Linked Data περιβάλλον (RDF[1] και OWL [2]).
- Μελέτη και συγκριτική παρουσίαση των υπάρχοντων συστημάτων διαχείρισης RDF δεδομένων (OWLIM, Sesame, Virtuoso, Kowari, Jena, RDF-3X, YARS2, iStore, κ.λπ.)
- Ορισμός πειραματικής διαδικασίας αξιολόγησης (benchmarking) και διεξαγωγή πειραμάτων σε πραγματικά δεδομένα.

ΣΧΕΤΙΚΟ ΥΛΙΚΟ:

[1] *Resource Description Framework (RDF)*, http://www.w3schools.com/rdf/rdf_intro.asp

[2] *Web Ontology Language (OWL)*, http://www.w3schools.com/rdf/rdf_owl.asp

[3] *Linked Data - Connect Distributed Data Across the Web*, <http://linkeddata.org/>