

Project: Μελέτη-Υλοποίηση και Πειραματική Αξιολόγηση DHTs με Εφαρμογές τους

Υπεύθυνοι Καθηγητές: Σπύρος Σιούτας (Καθηγητής ΤΜΗΥΠ).

Στόχος: Στόχος του project είναι η υλοποίηση αποκεντρωμένου ευρετηρίου (decentralized Index) βασισμένο στον κατακερματισμό σε Περιβάλλον Προγραμματισμού της αρεσκείας σας, με προτίμηση τη γλώσσα python και η πειραματική αξιολόγησή τους με data sets συνθετικά (artificial synthetic-data sets) ή ακόμα και πραγματικά (rea-data sets) στις βασικές πράξεις: Build, Insert key, Delete key, Update key, Lookup key, node join (insert node), node leave or failure (delete node), range queries, similarity queries, κ.τ.λ..

ΖΗΤΟΥΜΕΝΟ (60%-70%):

Lookup and Range Queries σε Σύνολα Κειμένων: Υλοποιείτε CHORD και Pastry DISTRIBUTED INDEX οι κόμβοι των οποίων θα αποθηκεύουν ένα σύνολο δεδομένων (real data set) από κείμενα που αφορούν επιστήμονες επιστήμης υπολογιστών και θα αντληθούν μέσω ενός προσαρμοσμένου web crawler από το κάτωθι URL: https://en.wikipedia.org/wiki/List_of_computer_scientists. Για τη συγκεκριμένη εφαρμογή, μας ενδιαφέρουν τα κάτωθι τρία (3) πεδία: **(Surname:String, #Awards:Integer, Education:text-vector)**. Τα αποκεντρωμένα p2p συστήματα CHORD και PASTRY, θα αποθηκεύσουν τα παραπάνω δεδομένα σε (key, value) μορφή με key το πεδίο Education και value τα εναπομείναντα δεδομένα, προκειμένου να εντοπίζουν τους επιστήμονες που έκαναν βασικές σπουδές σε συγκεκριμένο τμήμα (π.χ. CEID, NTUA, MIT,...). Στη συνέχεια ο εμπλεκόμενος κόμβος θα αναζητάει τοπικά εκείνους τους επιστήμονες για τους οποίους ο αριθμός βραβείων που έχουν αποσπάσει να είναι μεγαλύτερος από ένα user_defined threshold (με ή χωρίς ευρετήριο, δεν παίζει ρόλο). Π.χ. σκεφτείτε ερωτήματα της μορφής: **«Βρείτε τους επιστήμονες της επιστήμης υπολογιστών από τη ΒΔ Wikipedia που έκαναν βασικές σπουδές στο ceid και έχουν αποσπάσει > 4 βραβεία»**. Να συγκριθούν πειραματικά τα 2 κατακερματισμένα πρωτόκολλα ως προς τη βασική πράξη lookup (lookup_chord + local range query VS lookup_Pastry +local range query). Τι παρατηρείτε? Επίσης να συγκρίνετε πειραματικά και τις πράξεις node join και node leave σε κάθε ένα από τα δύο πρωτόκολλα. Εδώ τι παρατηρείτε?

Προαπαιτούμενες Γνώσεις: Δομές Δεδομένων, Αλγόριθμοι και Πολυπλοκότητα, Πολυδιάστατες Δομές Δεδομένων, Βάσεις Δεδομένων, Αντικειμενοστραφής Προγραμματισμός, Συναρτησιακός Προγραμματισμός (Functional Programming).

Παραπομπές:

[https://en.wikipedia.org/wiki/Chord_\(peer-to-peer\)](https://en.wikipedia.org/wiki/Chord_(peer-to-peer))

[https://en.wikipedia.org/wiki/Pastry_\(DHT\)](https://en.wikipedia.org/wiki/Pastry_(DHT))

ΥΛΙΚΟ ΤΟΥ ΜΑΘΗΜΑΤΟΣ ΣΤΟ eClass.

(***)ΠΑΡΑΔΟΤΕΑ ΚΑΙ ΤΡΟΠΟΣ ΕΞΕΤΑΣΗΣ: DEMO ΠΑΡΟΥΣΙΑΣΗ ΚΑΙ ΠΑΡΑΔΟΣΗ TECHNICAL REPORT ΣΕ ΕΝΤΥΠΗ ΚΑΙ ΨΗΦΙΑΚΗ ΜΟΡΦΗ ΓΙΑ ΑΝΑΡΤΗΣΗ ΣΕ ΣΧΕΤΙΚΗ ΙΣΤΟΣΕΛΙΔΑ ΤΟΥ ΜΑΘΗΜΑΤΟΣ [ΕΞΕΤΑΣΤΙΚΗ ΠΕΡΙΟΔΟΣ ΙΑΝΟΥΑΡΙΟΥ – ΦΕΒΡΟΥΑΡΙΟΥ]. ΑΡΙΘΜΟΣ ΜΕΛΩΝ ΑΝΑ ΟΜΑΔΑ: 3-4.