

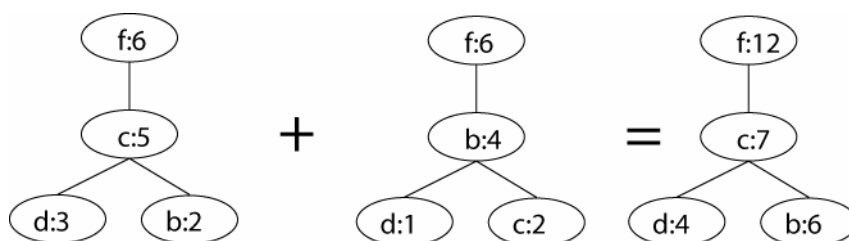
ΒΕΛΤΙΣΤΟΠΟΙΗΣΗ ΜΕΘΟΔΩΝ ΕΞΟΡΥΞΗΣ ΓΝΩΣΗΣ

ΠΛΗΡΟΦΟΡΙΕΣ: *M.Τερροβίτης, 210 772 1402, mter@dblab.ece.ntua.gr*

ΠΕΡΙΛΗΨΗ: Η διπλωματική εργασία στοχεύει στην υλοποίηση διαφόρων αλγόριθμων για την ανανέωση της δομής FP-tree, και της βελτιστοποίησης της εξόρυξης συχνών συνόλων αντικειμένων (frequent itemsets) από μία βάση δεδομένων.

ΑΤΟΜΑ: 1-2

ΠΛΑΤΦΟΡΜΑ ΕΡΓΑΣΙΑΣ: C/C++/Linux



ΣΥΝΤΟΜΗ ΠΕΡΙΓΡΑΦΗ: Ο τομέας εξόρυξης γνώσης (Data Mining) είναι ένας από τους τομείς που θέτουν από τις μεγαλύτερες ερευνητικές προκλήσεις στο χώρο των βάσεων δεδομένων, και είναι ιδιαίτερα δημοφιλής ερευνητικά. Από τις σημαντικότερες μορφές γνώσης που εξάγονται από τα δεδομένα είναι τα **σύνολα συχνών αντικειμένων** (frequent itemsets). Για να γίνει αποδοτικότερη η εξόρυξη τέτοιων συνόλων έχει προταθεί η κατασκευή του **δέντρου συχνών αντικειμένων** (FP-tree) από τα δεδομένα μίας βάσης. Καθώς όμως η κατασκευή του είναι από μόνη της μία δαπανηρή εργασία, προκύπτει το ερώτημα του πως είναι δυνατόν να ανανεώνουμε το δέντρο όταν αλλάζουν τα δεδομένα, καθώς και το πως είναι δυνατόν να εκμεταλλευτούμε υπάρχοντα δέντρα όταν διαφορετικά σύνολα δεδομένων ενοποιούνται. Στα πλαίσια αυτής της εργασίας θα πρέπει να γίνουν οι εξής ενέργειες:

- Να υλοποιηθούν τελεστές ένωσης και διαφοράς για τα δέντρα συχνών αντικειμένων.
- Να υλοποιηθούν γνωστές μέθοδοι ανανέωσης των δέντρων.
- Να συγκριθούν μεταξύ τους, καθώς και με την κατασκευή του δέντρου από την αρχή.

Ο βασικός κώδικας της δημιουργίας του FP-tree μπορεί να βρεθεί έτοιμος, οπότε και το βάρος θα είναι στην υλοποίηση των διαδικασιών ανανέωσης. Τέλος θα μελετηθεί και η συνολική συμπεριφορά του δέντρου ανάλογα με τις στατιστικές κατανομές των αντικειμένων της βάσης.