

Οργάνωση Βάσεων Βιοϊατρικών Δεδομένων Εξόρυξη Γνώσης Βιοϊατρικών Δεδομένων

Σεμινάριο 7: Αλγόριθμοι για επεξεργασία ερωτήσεων και βελτιστοποίηση

**Ευάγγελος Καρκαλέτσης,
Αναστασία Κριθαρά, Γεώργιος Πετάσης**

Εργαστήριο Τεχνολογίας Γνώσεων & Λογισμικού,

Ινστιτούτο Πληροφορικής & Τηλεπικοινωνιών, Ε.Κ.Ε.Φ.Ε. “Δημόκριτος”

Τηλ.: 210-6503197, Fax: 210-6532175, {vangelis, akrihara, petasis}@iit.demokritos.gr

Ακαδημαϊκό Έτος: 2016 – 2017

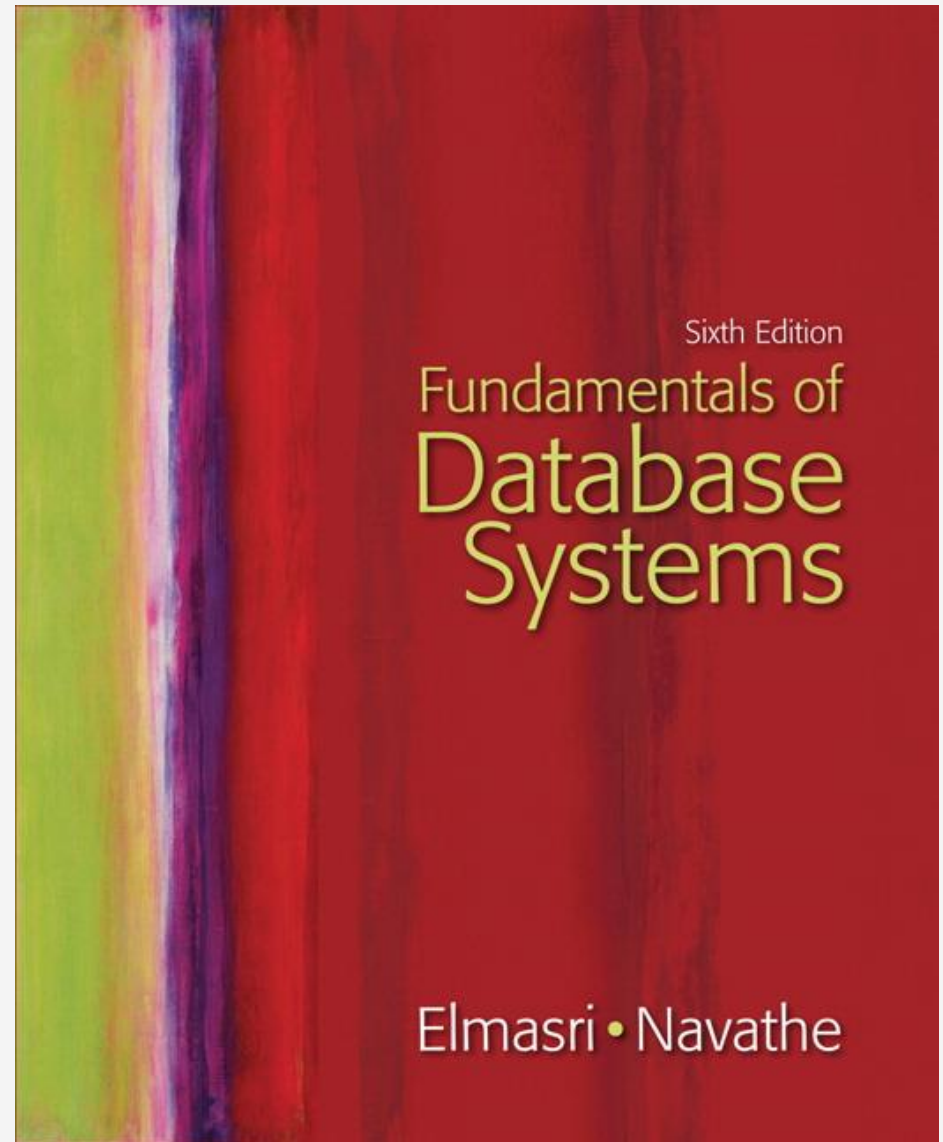
Οι διαφάνειες αυτού του σεμιναρίου βασίζονται
σε διαφάνειες των:

Δ. Γουνόπουλος, Ι. Ιωαννίδης

<http://eclass.uoa.gr/modules/document/document.php?course=D463&openDir=/4c2b32a5i4iq>

ΚΕΦΑΛΑΙΟ 19*

Αλγόριθμοι Επεξεργασίας και
Βελτιστοποίησης Ερωτήσεων



Copyright © 2011 Ramez Elmasri and Shamkant B. Navathe, Ελληνική Έκδοση Δίαυλος

Θα μιλήσουμε για

- Εισαγωγή στη βελτιστοποίηση ερωτήσεων
- Μετάφραση SQL ερωτήσεων σε Σχεσιακή Άλγεβρα
- Αλγόριθμοι Εξωτερικής Ταξινόμησης
- Αλγόριθμοι για τις πράξεις ΕΠΙΛΟΓΗ και ΣΥΝΕΝΩΣΗ
- Αλγόριθμοι για ΠΡΟΒΟΛΗ και συνόλων
- Υλοποίηση Συναθροιστικών πράξεων και εξωτερικών συνενώσεων

Εισαγωγή στη βελτιστοποίηση ερωτήσεων

- Βελτιστοποίηση ερωτήσεων
 - Η διαδικασία επιλογής κατάλληλης στρατηγικής εκτέλεσης μιας ερώτησης*
- Δύο εσωτερικές αναπαραστάσεις μιας ερώτησης:
 - Δέντρο ερώτησης
 - Γράφος ερώτησης

*εύλογα αποδοτική στρατηγική

Εισαγωγή στη βελτιστοποίηση ερωτήσεων

Ερώτηση σε γλώσσα υψηλού επιπέδου

ΛΕΚΤΙΚΗ ΑΝΑΛΥΣΗ,
ΣΥΝΤΑΚΤΙΚΗ ΑΝΑΛΥΣΗ
ΚΑΙ ΕΛΕΓΧΟΣ ΕΓΚΥΡΟΤΗΤΑΣ

Ερώτηση σε ενδιάμεση μορφή

ΒΕΛΤΙΣΤΟΠΟΙΗΤΗΣ
ΕΡΩΤΗΣΕΩΝ

Σχέδιο εκτέλεσης

ΓΕΝΝΗΤΟΡΑΣ ΚΩΔΙΚΑ
ΕΡΩΤΗΣΕΩΝ

Κώδικας για εκτέλεση της ερώτησης

ΥΠΟΣΥΣΤΗΜΑ
ΕΚΤΕΛΕΣΗΣ ΕΡΩΤΗΣΕΩΝ

Αποτέλεσμα ερώτησης

Ο κώδικας μπορεί να:

- Εκτελείται άμεσα (λειτουργία διερμηνέας)
- Αποθηκεύεται και να εκτελείται αργότερα, όταν είναι απαραίτητο (λειτουργία μεταγλώττισης)

Μετάφραση SQL Ερωτήσεων σε Σχεσιακή Άλγεβρα

- **Μπλοκ ερώτησης:**
 - Η βασική μονάδα που μπορεί να μεταφρασθεί σε αλγεβρικούς τελεστές και να βελτιστοποιηθεί.
- Ένα μπλοκ ερώτησης περιέχει μια SELECT-FROM-WHERE έκφραση, καθώς και προτάσεις GROUP BY και HAVING αν αποτελούν τμήμα του μπλοκ.
- **Οι εμφωλευμένες ερωτήσεις** μέσα σε μια ερώτηση θεωρούνται ξεχωριστά μπλοκ ερωτήσεων.
- Οι συναθροιστικοί τελεστές (MAX, MIN, SUM και COUNT) στην SQL πρέπει να περιλαμβάνονται στην επεκτεταμένη άλγεβρα.

Αλγόριθμοι για Εξωτερική Ταξινόμηση (1)

- **Εξωτερική Ταξινόμηση:**

- Αναφέρεται σε αλγορίθμους ταξινόμησης που είναι κατάλληλοι για μεγάλα αρχεία εγγράφων αποθηκευμένα στο δίσκο που δεν μπορούν να χωρέσουν ολόκληρα στην κύρια μνήμη, όπως τα περισσότερα αρχεία των βάσεων δεδομένων.

- **Στρατηγική ταξινόμησης-συγχώνευσης:**

- Ξεκινά με την ταξινόμηση μικρών υποαρχείων (**σειρές ή runs**) του κυρίως αρχείου και στη συνέχεια με συγχώνευση των ταξινομημένων σειρών, δημιουργώντας μεγαλύτερα ταξινομημένα υποαρχεία, που γίνεται με την σειρά συγχώνευση και αυτών.

Αλγόριθμοι για Εξωτερική Ταξινόμηση (2)

- Φάση ταξινόμησης:

$$n_R = \left\lceil \left(\frac{b}{n_B} \right) \right\rceil$$

- Φάση συγχώνευσης:

$$d_M = \text{Min}(n_B - 1, n_R); n_P = \lceil (\log_{d_M}(n_R)) \rceil$$

- n_R : πλήθος των αρχικών σειρών
- b : πλήθος των μπλοκ του αρχείου
- n_B : διαθέσιμος χώρος στο μπάφερ
- d_M : βαθμός συγχώνευσης
 - πλήθος σειρών που μπορούν να συγχωνευθούν σε ένα πέρασμα
 - Ένα μπάφερ μπλοκ για κάθε τρέξιμο της συνένωσης, και ένα μπλοκ για να καταχωρείται ένα μπλοκ αποτελέσματος
- n_P : αριθμός περασμάτων

Αλγόριθμοι για Εξωτερική Ταξινόμηση (3)

```
set   i ← 1;
      j ← b; {το μέγεθος του αρχείου σε μπλοκ}
      k ← nb; {το μέγεθος του μπάφερ σε μπλοκ}
      m ← ⌈(j/k)⌉;
{φάση της ταξινόμησης}
while (i ≤ m)
  do {
    διάβασε τα επόμενα k μπλοκ του αρχείου στο μπάφερ ή αν απομένουν λιγότερα από k
    μπλοκ διάβασε τα μπλοκ που απομένουν
    ταξινόμησε τις εγγραφές στο μπάφερ και να γραφούν σαν προσωρινό υποαρχείο;
    i ← i + 1;
  }
{Φάση της συγχώνευσης: συγχώνευση των υποαρχείων μέχρι να μείνει 1}
set   i ← 1;
      p ← ⌈logk-1 m⌉;
      j ← m;
while (i ≤ p)
  do {
    n ← 1;
    q ← ⌈(j/(k-1))⌉; {πλήθος υποαρχείων που θα γραφούν στο πέραςμα αυτό}
    while (n ≤ q)
      do {
        διάβασε τα επόμενα k-1 υποαρχεία (από το προηγούμενο πέραςμα) ένα μπλοκ τη φορά;
        συνένωση και γράψιμο σαν νέο υποαρχείο;
        n ← n + 1;
      }
    j ← q;
    i ← i + 1;
  }
```

Αλγόριθμοι για Εξωτερική Ταξινόμηση (4)

- Παράδειγμα:
 - $b = 1024$ μπλοκ το μέγεθος του αρχείου
 - $n_B = 5$ μπλοκ
 - n_P : αριθμός περασμάτων
- Φάση ταξινόμησης:
 - $n_R = \left\lceil \left(\frac{b}{n_B} \right) \right\rceil = \frac{1024}{5} = 205$ ταξινομημένες σειρές (υποαρχεία στο δίσκο)
- Φάση συγχώνευσης:
 - $d_M = \text{Min}(n_B - 1, n_R) = \text{Min}((5 - 1), 205) = 4$ σειρές σε κάθε πέρασμα
 - Πρώτο πέρασμα: $\frac{205}{4} = 52$
 - Δεύτερο πέρασμα: $\frac{52}{4} = 13$
 - Τρίτο πέρασμα: $\frac{13}{4} = 4$
 - Τέταρτο πέρασμα: $\frac{4}{4} = 1$
 - $n_P = \lceil (\log_{d_M}(n_R)) \rceil = \lceil (\log_4(205)) \rceil = 4$ περάσματα

Αλγόριθμοι για Εξωτερική Ταξινόμηση (5)

- Απόδοση αλγορίθμου
 - πλήθος των μπλοκ που διαβάζονται και γράφονται στο δίσκο:

$$(2 * b) + (2 * (b * (\log_{d_M} n_R)))$$

- Πλήθος των μπλοκ προσπελάσεων για τη φάση της ταξινόμησης, δύο προσπελάσεις σε κάθε μπλοκ του αρχείου, μια για διάβασμα στη μνήμη και μια για γράψιμο των εγγραφών πίσω στο δίσκο, μετά την ταξινόμηση.
- Το πλήθος των μπλοκ προσπελάσεων για τη φάση της συγχώνευσης. Σε κάθε πέρασμα διαβάζονται και γράφονται ένα πλήθος μπλοκ του δίσκου που είναι περίπου όσο το πλήθος των μπλοκ του αρχικού αρχείου b . Το πλήθος των περασμάτων είναι $\log_{d_M} n_R$.

Αλγόριθμοι για τις πράξεις ΕΠΙΛΟΓΗΣ και ΣΥΓΧΩΝΕΥΣΗΣ (1)

- Υλοποίηση της πράξης ΕΠΙΛΟΓΗ (SELECT)
- Παραδείγματα:
 - (ΠΡ1): $\sigma_{AP_TAYT='123456789'}$ (ΕΡΓΑΖΟΜΕΝΟΣ)
 - (ΠΡ2): $\sigma_{K\Omega\Delta_TMHM>5}$ (ΤΜΗΜΑ)
 - (ΠΡ3): $\sigma_{API\Theta_T=5}$ (ΕΡΓΑΖΟΜΕΝΟΣ)
 - (ΠΡ4): $\sigma_{API\Theta_T=5 \text{ AND } ΜΙΣΘΟΣ>30000 \text{ AND } ΦΥΛΛΟ='F'}$ (ΕΡΓΑΖΟΜΕΝΟΣ)
 - (ΠΡ5): $\sigma_{E_APTAYT='123456789' \text{ AND } K_EPΓO=10}$ (ΑΠΑΣΧΟΛΗΣΗ)

Αλγόριθμοι για τις πράξεις ΕΠΙΛΟΓΗΣ και ΣΥΓΧΩΝΕΥΣΗΣ (2)

- Υλοποίηση της πράξης ΕΠΙΛΟΓΗ (SELECT)
- Μέθοδοι αναζήτησης για απλή επιλογή:
 - **A1 Σειριακή αναζήτηση** (εξαντλητική αναζήτηση):
 - Ανάκτηση κάθε εγγραφής του αρχείου και έλεγχος αν οι τιμές των γνωρισμάτων ικανοποιούν τη συνθήκη επιλογής.
 - **A2 Δυαδική αναζήτηση**:
 - Αν η συνθήκη επιλογής περιλαμβάνει μια σύγκριση ισότητας σε ένα γνώρισμα-κλειδί πάνω στο οποίο το αρχείο είναι ταξινομημένο, μπορεί να χρησιμοποιηθεί δυαδική αναζήτηση (η οποία είναι πιο αποτελεσματική από την A1).

Αλγόριθμοι για τις πράξεις ΕΠΙΛΟΓΗΣ και ΣΥΓΧΩΝΕΥΣΗΣ (3)

- Υλοποίηση της πράξης ΕΠΙΛΟΓΗ (SELECT)
- Μέθοδοι αναζήτησης για απλή επιλογή (συν.):
 - **A3 Χρήση πρωτεύοντος ευρετηρίου (ή κλειδιού κατακερματισμού)**
 - Αν η συνθήκη επιλογής περιλαμβάνει σύγκριση ισότητας σε ένα γνώρισμα κλειδί με πρωτεύον ευρετήριο (ή με κλειδί κατακερματισμού – hash key), χρησιμοποιούμε το πρωτεύον ευρετήριο (ή το κλειδί κατακερματισμού) για την ανάκτηση της εγγραφής.
 - **A4 Χρήση ενός πρωτεύοντος ευρετηρίου για ανάκτηση πολλών εγγραφών**
 - Αν η συνθήκη σύγκρισης είναι $>$, $\geq$, $<$, ή $\leq$ σε ένα πεδίο-κλειδί με πρωτεύον ευρετήριο, χρησιμοποιούμε το ευρετήριο για να βρούμε την εγγραφή που ικανοποιεί την αντίστοιχη συνθήκη, και στη συνέχεια ανακτούμε όλες τις εγγραφές που ακολουθούν στο ταξινομημένο αρχείο.

Αλγόριθμοι για τις πράξεις ΕΠΙΛΟΓΗΣ και ΣΥΓΧΩΝΕΥΣΗΣ (4)

- Υλοποίηση της πράξης ΕΠΙΛΟΓΗ (SELECT)
- Μέθοδοι αναζήτησης για απλή επιλογή (συν.):
 - **A5 Χρήση ευρετηρίων συστάδων για ανάκτηση πολλών εγγραφών:**
 - Αν η συνθήκη επιλογής περιλαμβάνει μια συνθήκη ισότητας σε **ένα γνώρισμα που δεν είναι κλειδί** αλλά έχει οριστεί σε αυτό ένα ευρετήριο συστάδων, χρησιμοποιούμε το ευρετήριο συστάδων για την ανάκτηση όλων των εγγραφών που ικανοποιούν τη συνθήκη επιλογής.
 - **A6 Χρήση δευτερεύοντος ευρετηρίου (B⁺ Δέντρου) σε μια συνθήκη ισότητας:**
 - Αυτή η μέθοδος αναζήτησης μπορεί να χρησιμοποιηθεί για να ανακτηθεί μια μοναδική εγγραφή, αν το πεδίο είναι κλειδί ή για να ανακτηθούν πολλαπλές εγγραφές αν το πεδίο δεν είναι κλειδί.
 - Επιπλέον, μπορεί να χρησιμοποιηθεί για ανάκτηση εγγράφων, για συνθήκες που περιλαμβάνουν τους τελεστές >, >=, <, or <=.

Αλγόριθμοι για τις πράξεις ΕΠΙΛΟΓΗΣ και ΣΥΓΧΩΝΕΥΣΗΣ (5)

- Υλοποίηση της πράξης ΕΠΙΛΟΓΗ (SELECT)
- Μέθοδοι αναζήτησης για πολύπλοκες επιλογές:
 - **A7 Συζευκτική επιλογή με χρήση ατομικού ευρετηρίου:**
 - Αν ένα γνώρισμα που περιλαμβάνεται σε οποιαδήποτε **απλή συνθήκη** της συζευκτικής συνθήκης έχει μια διαδρομή προσπέλασης που επιτρέπει τη χρήση μιας από τις μεθόδους A2 έως A6, χρησιμοποιούμε αυτή τη συνθήκη για την ανάκτηση των εγγράφων και στη συνέχεια ελέγχουμε αν κάθε ανακτηθείσα εγγραφή ικανοποιεί τις υπόλοιπες απλές συνθήκες της συζευκτικής συνθήκης
 - **A8 Συζευκτική επιλογή με χρήση σύνθετου ευρετηρίου:**
 - Αν δύο ή περισσότερα γνωρίσματα συμμετέχουν σε συνθήκες ισότητας στη συζευκτική συνθήκη και υπάρχει σύνθετο ευρετήριο στο συνδυασμό πεδίων, μπορούμε να χρησιμοποιήσουμε άμεσα το ευρετήριο.

Αλγόριθμοι για τις πράξεις ΕΠΙΛΟΓΗΣ και ΣΥΓΧΩΝΕΥΣΗΣ (6)

- Υλοποίηση της πράξης ΕΠΙΛΟΓΗ (SELECT)
- Μέθοδοι αναζήτησης για πολύπλοκες επιλογές (συν.):
 - **A9 Συζευκτική επιλογή με τομή των δεικτών εγγράφων:**
 - Αν υπάρχουν δευτερεύοντα ευρετήρια σε όλα (ή σε μερικά) πεδία που ενέχονται σε συνθήκες σύγκρουσης ισότητας της συζευκτικής συνθήκης και αν τα ευρετήρια περιλαμβάνουν δείκτες εγγραφών (αντί δείκτες μπλοκ).
 - Κάθε ευρετήριο μπορεί να χρησιμοποιηθεί για ανάκτηση ενός **συνόλου δεικτών εγγραφών** που ικανοποιούν τη μεμονωμένη απλή συνθήκη.
 - Η **τομή** αυτών των συνόλων δεικτών δίνει τους δείκτες εγγραφών που ικανοποιούν τη συζευκτική συνθήκη, οι οποίοι χρησιμοποιούνται για την άμεση ανάκτηση αυτών των εγγράφων.
 - Αν μερικές μόνο από τις συνθήκες έχουν δευτερεύοντα ευρετήρια, κάθε ανακτώμενη εγγραφή ελέγχεται για να διαπιστωθεί αν ικανοποιεί και τις υπόλοιπες συνθήκες.

Αλγόριθμοι για τις πράξεις ΕΠΙΛΟΓΗΣ και ΣΥΓΧΩΝΕΥΣΗΣ (7)

- Υλοποίηση της πράξης ΕΠΙΛΟΓΗ (SELECT) :
 - Όταν η επιλογή προσδιορίζεται από **μια απλή συνθήκη**, μπορούμε να ελέγχουμε μόνον αν υπάρχει κάποια διαδρομή προσπέλασης στο γνώρισμα που ενέχεται στη συνθήκη.
 - Αν υπάρχει διαδρομή προσπέλασης χρησιμοποιείται η μέθοδος που αντιστοιχεί σ' αυτή τη διαδρομή προσπέλασης, διαφορετικά χρησιμοποιείται η σειριακή αναζήτηση A1 (βλ. ΠΡ1, ΠΡ2 και ΠΡ3)
 - Για **συζευκτικές συνθήκες επιλογής**, όταν υπάρχουν διαδρομές προσπέλασης για περισσότερα από ένα από τα γνωρίσματα που ενέχονται στις συνθήκες, ο βελτιστοποιητής πρέπει να επιλέξει τη διαδρομή προσπέλασης που ανακτά τις λιγότερες εγγραφές κατά τον πιο αποδοτικό τρόπο.
 - **Διαζευκτικές συνθήκες επιλογής**

Αλγόριθμοι για τις πράξεις ΕΠΙΛΟΓΗΣ και ΣΥΓΧΩΝΕΥΣΗΣ (8)

- Υλοποίηση της πράξης ΣΥΝΕΝΩΣΗ (JOIN) :
 - ΣΥΝΕΝΩΣΗ ΙΣΟΤΗΤΑΣ (EQUIJOIN), ΦΥΣΙΚΗ ΣΥΝΕΝΩΣΗ (NATURAL JOIN):
 - Συνένωση δύο αρχείων
 - e.g. $R \bowtie_{A=B} S$
 - Πολλαπλές συνενώσεις: συνενώσεις περισσότερων από δύο αρχείων
 - e.g. $R \bowtie_{A=B} S \bowtie_{C=D} T$
- Παραδείγματα:
 - (ΠΡ6): ΕΡΓΑΖΟΜΕΝΟΣ $\bowtie_{\text{ΑΡΙΘ}_T=\text{ΚΩΔ_ΤΜΗΜ}} \text{ΤΜΗΜΑ}$
 - (ΠΡ7): ΤΜΗΜΑ $\bowtie_{\text{ΔΙΕΥΘΥΝΤΗΣ}=\text{ΑΡ_ΤΑΥΤ}} \text{ΕΡΓΑΖΟΜΕΝΟΣ}$

Αλγόριθμοι για τις πράξεις ΕΠΙΛΟΓΗΣ και ΣΥΓΧΩΝΕΥΣΗΣ (9)

- Υλοποίηση της πράξης ΣΥΝΕΝΩΣΗ (JOIN) :
- Μέθοδοι υλοποίησης συνενώσεων:
 - Σ1 **Εμφωλευμένοι βρόγχοι** (εξαντλητική αναζήτηση):
 - Για κάθε εγγραφή t της σχέσης R (εξωτερικός βρόγχος), ανακτάται κάθε εγγραφή s από τη σχέση S (εσωτερικός βρόγχος) και ελέγχεται αν οι δύο εγγραφές ικανοποιούν τη συνθήκη συνένωσης $t[A] = s[B]$.
 - Σ2 **Συνένωση απλού βρόγχου** (χρήση μιας δομής προσπέλασης για ανάκτηση των εγγραφών που ικανοποιούν τη συνθήκη):
 - Αν υπάρχει ευρετήριο (ή κλειδί κατακερματισμού) για ένα από τα δύο γνωρίσματα της συνένωσης — έστω το B της σχέσης S — ανακτώνται όλες οι εγγραφές t της σχέσης R , μία προς μία, και χρησιμοποιείται η δομή προσπέλασης για να ανακτηθούν άμεσα όλες οι εγγραφές s της σχέσης S που ικανοποιούν τη συνθήκη $s[B] = t[A]$.

Αλγόριθμοι για τις πράξεις ΕΠΙΛΟΓΗΣ και ΣΥΓΧΩΝΕΥΣΗΣ (10)

- Υλοποίηση της πράξης ΣΥΝΕΝΩΣΗ (JOIN) :
- Μέθοδοι υλοποίησης συνενώσεων:
 - **Σ3 Συνένωση ταξινόμησης-συγχώνευσης:**
 - Αν οι εγγραφές των R και S είναι φυσικά ταξινομημένες σύμφωνα με τις τιμές των γνωρισμάτων συνένωσης A και B αντίστοιχα, μπορούμε να υλοποιήσουμε τη συνένωση με τον πιο αποδοτικό τρόπο.
 - Και τα δύο αρχεία σαρώνονται σύμφωνα με τη διάταξη των γνωρισμάτων συνένωσης, και επιλέγεται ο συνδυασμός των εγγραφών με τις ίσες τιμές στα A και B.
 - Με αυτή τη μέθοδο οι εγγραφές κάθε αρχείου σαρώνονται μόνο μια φορά, προκειμένου να ταιριάξουν με αυτές του άλλου αρχείου— εκτός αν και τα A και B δεν είναι γνωρίσματα-κλειδιά, στην οποία περίπτωση η μέθοδος πρέπει να τροποποιηθεί ελαφρά.

Αλγόριθμοι για τις πράξεις ΕΠΙΛΟΓΗΣ και ΣΥΓΧΩΝΕΥΣΗΣ (11)

- Υλοποίηση της πράξης ΣΥΝΕΝΩΣΗ (JOIN) :
- Μέθοδοι υλοποίησης συνενώσεων:
 - **Σ4 Συνένωση με κατακερματισμό:**
 - Οι εγγραφές των αρχείων R και S διαμερίζονται σε μικρότερα αρχεία. Η διαμέριση κάθε αρχείου γίνεται με χρήσης της ίδιας συνάρτησης κατακερματισμού ή στο γνώρισμα συνένωσης A της R και B της S.
 - *Φάση διαμέρισης:* Ένα πέρασμα από το αρχείο με τις λιγότερες εγγραφές (έστω το R) απεικονίζει τις εγγραφές της R στις διάφορες διαμερίσεις της R.
 - *Φάση αναζήτησης:* Το άλλο αρχείο (S) διατρέχεται μόνο μία φορά με απεικόνιση κάθε εγγραφής του στον κατάλληλο κάδο, όπου η εγγραφή συνδυάζεται με όλες τις εγγραφές του αρχείου R που ταιριάζουν με αυτή.

Αλγόριθμοι για τις πράξεις ΕΠΙΛΟΓΗΣ και ΣΥΓΧΩΝΕΥΣΗΣ (12)

(α) ταξινόμησε τις πλειάδες της R στο γνώρισμα A; (* υποθέτουμε ότι η R έχει n πλειάδες (εγγραφές)*)
ταξινόμησε τις πλειάδες της S στο γνώρισμα B; (* υποθέτουμε ότι η S έχει m πλειάδες (εγγραφές)*)
θέσε $i \leftarrow 1, j \leftarrow 1$;
while ($i \leq n$) and ($j \leq m$)
do { if $R(i)[A] > S(j)[B]$
 then θέσε $j \leftarrow j+1$
elseif $R(i)[A] < S(j)[B]$
 then θέσε $i \leftarrow i+1$
else { (* $R(i)[A]=S(j)[B]$, επομένως γράφουμε στην έξοδο μια πλειάδα *)
 γράφουμε το συνδυασμό πλειάδων $\langle R(i), S(j) \rangle$ στη σχέση T;
 (* γράφουμε στην έξοδο άλλες πλειάδες που ταιριάζουν με την $R(i)$, αν υπάρχουν *)
 θέσε $l \leftarrow j+1$;
 while ($l \leq m$) and ($R(i)[A]=S(l)[B]$)
 do { γράφουμε το συνδυασμό πλειάδων $\langle R(i), S(l) \rangle$ στη σχέση T;
 θέσε $l \leftarrow l+1$
 }
 (* γράφουμε στην έξοδο άλλες πλειάδες που ταιριάζουν με την $S(j)$, αν υπάρχουν *)
 θέσε $k \leftarrow i+1$;
 while ($k \leq n$) and ($R(k)[A]=S(j)[B]$)
 do { γράφουμε το συνδυασμό πλειάδων $\langle R(k), S(j) \rangle$ στη σχέση T;
 θέσε $k \leftarrow k+1$
 }
 θέσε $i \leftarrow k, j \leftarrow l$
}

}

$T \leftarrow R \bowtie_{A=B} S$

Αλγόριθμοι για τις πράξεις ΕΠΙΛΟΓΗΣ και ΣΥΓΧΩΝΕΥΣΗΣ (17)

- Υλοποίηση της πράξης ΣΥΝΕΝΩΣΗ (JOIN) :
- Παράγοντες που επηρεάζουν την απόδοση συνένωσης:
 - Διαθέσιμος μπάφερ χώρου
 - Παράγοντας επιλεξιμότητας συνένωσης ενός αρχείου με ένα άλλο, σε σχέση με μια συνθήκη ισότητας
 - Ποσοστό των εγγραφών του ενός αρχείου που θα συνενωθούν με εγγραφές άλλου αρχείου
 - Επιλογή εσωτερικής VS εξωτερικής σχέσης

Αλγόριθμοι για τις πράξεις ΕΠΙΛΟΓΗΣ και ΣΥΓΧΩΝΕΥΣΗΣ (18)

- Υλοποίηση της πράξης ΣΥΝΕΝΩΣΗ (JOIN) :
- Άλλες μέθοδοι υλοποίησης συνενώσεων:
- Συνένωση κατακερματισμού με διαμέριση
 - Φάση διαμέρισης:
 - Κάθε αρχείο (R και S) χωρίζονται σε M σύνολα με τη χρήση της ίδιας συνάρτησης κατακερματισμού για τη διαμέριση των δύο αρχείων στο γνώρισμα συνένωσης (A και B):
 - $R_1, R_2, R_3, \dots, R_m$ και $S_1, S_2, S_3, \dots, S_m$
 - Το ελάχιστο πλήθος μπλοκ ενδιάμεσης μνήμης που απαιτούνται για τη φάση της διαμέρισης είναι: $M+1$.
 - Για κάθε διαμέριση ένα υποαρχείο δίσκου δημιουργείται για να αποθηκεύει αυτή τη διαμέριση.
 - Φάση συνένωσης:
 - Απαιτούνται M επαναλήψεις, μία για κάθε διαμερισμένο αρχείο.
 - Κατά την i επανάληψη, γίνεται συνένωση των διαμερίσεων R_i και S_i

Αλγόριθμοι για τις πράξεις ΕΠΙΛΟΓΗΣ και ΣΥΓΧΩΝΕΥΣΗΣ (19)

- Υλοποίηση της πράξης ΣΥΝΕΝΩΣΗ (JOIN) :
- Συνένωση κατακερματισμού με διαμέριση:
 - Έστω ότι το R_i είναι μικρότερο από το S_i
 1. Αντιγραφή των εγγραφών του R_i σε μπάφερ της μνήμης.
 2. Διάβασε όλα τα μπλοκ του S_i -ένα τη φορά- και κάθε εγγραφή χρησιμοποιείται για **αναζήτηση** εγγραφών με την ίδια τιμή στο γνώρισμα συνένωσης με τη διαμέριση R_i .
 3. Γίνεται συνένωση των εγγραφών που ταιριάζουν και καταχωρείται στο αρχείο του αποτελέσματος.

Αλγόριθμοι για τις πράξεις ΕΠΙΛΟΓΗΣ και ΣΥΓΧΩΝΕΥΣΗΣ (20)

- Υλοποίηση της πράξης ΣΥΝΕΝΩΣΗ (JOIN) :
- Κόστος συνένωσης κατακερματισμού με διαμέριση:
 1. Κάθε εγγραφή διαβάζεται μια φορά, και γράφεται πίσω στο δίσκο μία φορά κατά τη φάση της διαμέρισης:
$$(b_R + b_S), (b_R + b_S)$$
 2. Κατά τη φάση της συνένωσης, κάθε εγγραφή διαβάζεται δεύτερη φορά για να εκτελεστεί η συνένωση:
$$(b_R + b_S)$$
 3. Εγγραφή του αρχείου του αποτελέσματος:
$$b_{RES}$$
- Total Cost:
$$- 3 * (b_R + b_S) + b_{RES}$$

Αλγόριθμοι για τις πράξεις ΕΠΙΛΟΓΗΣ και ΣΥΓΧΩΝΕΥΣΗΣ (21)

- Υλοποίηση της πράξης ΣΥΝΕΝΩΣΗ (JOIN) :
- **Υβριδικός αλγόριθμος συνένωσης κατακερματισμού:**
 - Παραλλαγή της συνένωσης κατακερματισμού με διαμέριση:
 - Η *φάση συνένωσης* μιας από τις διαμερίσεις περιλαμβάνεται στη *φάση διαμέρισης*.
 - **Φάση διαμέρισης:**
 - Διαμερίζεται το μικρότερο από τα δύο αρχεία – ένα μπλοκ για κάθε μία από τις $M-1$ διαμερίσεις, και τα υπόλοιπα μπλοκ στη πρώτη διαμέριση.
 - Διαμερίζονται οι εγγραφές του δεύτερου αρχείου της συνένωσης – το μεγαλύτερο αρχείο.
 - **Φάση συνένωσης:**
 - Απαιτούνται $M-1$ επαναλήψεις για τις διαμερίσεις $R_1, R_2, R_3, \dots, R_m$ και $S_1, S_2, S_3, \dots, S_m$. R_1 και S_1 συνενώνονται κατά τη διαμέριση του S_1 , και τα αποτελέσματα της συνένωσης R_1 και S_1 γράφονται στο δίσκο στο τέλος της φάσης διαμέρισης.

- **Αλγόριθμοι για την πράξη ΠΡΟΒΟΛΗ:**

- Αν η <λίστα γνωρισμάτων> περιλαμβάνει ένα κλειδί της σχέσης R:
 - Το αποτέλεσμα της προβολής θα έχει το ίδιο πλήθος πλειάδων με τη σχέση R.
 - Σε κάθε πλειάδα του αποτελέσματος θα εμφανίζονται μόνο τα γνωρίσματα που περιλαμβάνονται στη <λίστα γνωρισμάτων>.
- Αν η <λίστα γνωρισμάτων> δεν περιλαμβάνει ένα κλειδί της σχέσης R:
 - Ταξινόμηση του αποτελέσματος και απαλοιφή των διπλότυπων πλειάδων (οι οποίες θα εμφανίζονται σε διαδοχικές θέσεις μετά την ταξινόμηση).
 - Κατακερματισμός για την απαλοιφή των διπλότυπων: Κατά την εισαγωγή μιας εγγραφής στον κατάλληλο κάδο κατακερματισμού, αν πρόκειται για διπλότυπη πλειάδα, δεν εισάγεται.

Αλγόριθμοι για τις πράξεις ΕΠΙΛΟΓΗΣ και ΣΥΓΧΩΝΕΥΣΗΣ (13)

(β) καταχώρησε μια πλειάδα $t[<\text{λίστα γνωρισμάτων}>]$ στη σχέση T' για κάθε πλειάδα t της R ;
(* η σχέση T' περιέχει το αποτέλεσμα της προβολής πριν την απομάκρυνση των διπλοτύπων *)
if η $<\text{λίστα γνωρισμάτων}>$ περιλαμβάνει ένα κλειδί της R
then $T \leftarrow T'$
else { ταξινόμησε τις πλειάδες της T' ;
θέσε $i \leftarrow 1, j \leftarrow 2$;
while $i \leq n$
do { γράφουμε την πλειάδα $T'[i]$ στη σχέση T ;
while $T'[i] = T'[j]$ and $j \leq n$ do $j \leftarrow j+1$; (* απομάκρυνση διπλοτύπων *)
 $i \leftarrow j; j \leftarrow i+1$
}
}
(* η σχέση T περιέχει το αποτέλεσμα της προβολής μετά την απομάκρυνση διπλοτύπων *)

$$T \leftarrow \pi_{<\text{λίστα γνωρισμάτων}>}(R)$$

- **Αλγόριθμοι για τις ΣΥΝΟΛΟΘΕΩΡΗΤΙΚΕΣ πράξεις:**
 - ΕΝΩΣΗ (UNION)
 - ΤΟΜΗ (INTERSECTION)
 - ΣΥΝΟΛΟΘΕΩΡΗΤΙΚΗ ΔΙΑΦΟΡΑ (SET DIFFERENCE)
 - ΚΑΡΤΕΣΙΑΝΟ ΓΙΝΟΜΕΝΟ (CARTESIAN PRODUCT)

- **Αλγόριθμοι για τις ΣΥΝΟΛΟΘΕΩΡΗΤΙΚΕΣ πράξεις:**
- Το **ΚΑΡΤΕΣΙΑΝΟ ΓΙΝΟΜΕΝΟ** των σχέσεων R και S περιλαμβάνει μια εγγραφή για κάθε συνδυασμό εγγραφών από τις σχέσεις R και S . Τα γνωρίσματα του αποτελέσματος περιλαμβάνουν όλα τα γνωρίσματα των R και S .
- **Ανάλυση κόστους** του ΚΑΡΤΕΣΙΑΝΟΥ ΓΙΝΟΜΕΝΟΥ
 - Αν η σχέση R έχει n εγγραφές και j γνωρίσματα, και η σχέση S έχει m εγγραφές και k γνωρίσματα, το αποτέλεσμα θα έχει $n * m$ εγγραφές και $j * k$ γνωρίσματα.
- Η πράξη ΚΑΡΤΕΣΙΑΝΟΥ ΓΙΝΟΜΕΝΟΥ είναι **πολύ ακριβή** και είναι σημαντικό να αποφεύγεται και να αντικαθίσταται από άλλες ισοδύναμες πράξεις.

Αλγόριθμοι για τις πράξεις ΠΡΟΒΟΛΗΣ και ΣΥΝΟΛΟΘΕΩΡΗΤΙΚΕΣ (4)

- **Αλγόριθμοι για τις ΣΥΝΟΛΟΘΕΩΡΗΤΙΚΕΣ πράξεις:**
- **ΕΝΩΣΗ (UNION)**
 - Ταξινόμηση των δύο σχέσεων στο ίδιο γνώρισμα.
 - Ταυτόχρονη σάρωση και συγχώνευση των δύο ταξινομημένων αρχείων, και όταν η ίδια πλειάδα εμφανίζεται και στα δύο αρχεία, καταχωρείται μόνο μια φορά το αποτέλεσμα.
- **ΤΟΜΗ (INTERSECTION)**
 - Ταξινόμηση των δύο σχέσεων στο ίδιο γνώρισμα.
 - Ταυτόχρονη σάρωση και συγχώνευση των δύο ταξινομημένων αρχείων, και όταν η ίδια πλειάδα εμφανίζεται και στα δύο αρχεία, καταχωρείται το αποτέλεσμα.
- **ΣΥΝΟΛΟΘΕΩΡΗΤΙΚΗ ΔΙΑΦΟΡΑ R-S (SET DIFFERENCE)**
 - Ταξινόμηση των δύο σχέσεων στο ίδιο γνώρισμα.
 - Ταυτόχρονη σάρωση και συγχώνευση των δύο ταξινομημένων αρχείων, και όταν η ίδια πλειάδα εμφανίζεται μόνο στη σχέση R (κι όχι στην S), καταχωρείται το αποτέλεσμα.

Αλγόριθμοι για τις πράξεις ΕΠΙΛΟΓΗΣ και ΣΥΓΧΩΝΕΥΣΗΣ (14)

```
(γ) ταξινόμησε τις πλειάδες των σχέσεων R και S χρησιμοποιώντας τα ίδια γνωρίσματα μοναδικής ταξινόμησης;  
  θέσε  $i \leftarrow 1, j \leftarrow 1$ ;  
  while ( $i \leq n$ ) and ( $j \leq m$ )  
  do { if  $R(i) > S(j)$   
      then { εγγραφή της πλειάδας  $S(j)$  στη σχέση T  
            θέσε  $j \leftarrow j+1$   
          }  
      elseif  $R(i) < S(j)$   
      then { εγγραφή της πλειάδας  $R(i)$  στη σχέση T  
            θέσε  $i \leftarrow i+1$   
          }  
      else θέσε  $j \leftarrow j+1$  (*  $R(i)=S(j)$  , ώστε να υπερπηδήσουμε μία από τις διπλότυπες πλειάδες *)  
    }  
  if ( $i \leq n$ ) then πρόσθεσε τις πλειάδες  $R(i)$  έως  $R(n)$  στη σχέση T;  
  if ( $j \leq m$ ) then πρόσθεσε τις πλειάδες  $S(j)$  έως  $S(m)$  στη σχέση T.
```

$T \leftarrow R \cup S$

Αλγόριθμοι για τις πράξεις ΕΠΙΛΟΓΗΣ και ΣΥΓΧΩΝΕΥΣΗΣ (15)

```
(δ) ταξινόμησε τις πλειάδες των σχέσεων R και S χρησιμοποιώντας τα ίδια γνωρίσματα μοναδικής ταξινόμησης;
  θέσε  $i \leftarrow 1, j \leftarrow 1$ ;
  while ( $i \leq n$ ) and ( $j \leq m$ )
  do   { if  $R(i) > S(j)$ 
        then   θέσε  $j \leftarrow j+1$ 
        elseif  $R(i) < S(j)$ 
        then   θέσε  $i \leftarrow i+1$ 
        else   { εγγραφή της πλειάδας  $R(i)$  στη σχέση T; (*  $R(i)=S(j)$  , επομένως γράφουμε την πλειάδα στην έξοδο *)
                  θέσε  $i \leftarrow i+1, j \leftarrow j+1$ 
                }
      }
}
```

$$T \leftarrow R \cap S$$

Αλγόριθμοι για τις πράξεις ΕΠΙΛΟΓΗΣ και ΣΥΓΧΩΝΕΥΣΗΣ (16)

```
(ε) ταξινόμησε τις πλειάδες των σχέσεων R και S χρησιμοποιώντας τα ίδια γνωρίσματα μοναδικής ταξινόμησης;
  θέσε  $i \leftarrow 1, j \leftarrow 1$ ;
  while ( $i \leq n$ ) and ( $j \leq m$ )
  do   { if  $R(i) > S(j)$ 
        then   θέσε  $j \leftarrow j+1$ 
        elseif  $R(i) < S(j)$ 
        then   { εγγραφή της πλειάδας  $R(i)$  στη σχέση T; (* η πλειάδα  $R(i)$  δεν έχει αντίστοιχη  $S(j)$ , επομένως γράφουμε την
                   $R(i)$  στην έξοδο *)
                  θέσε  $i \leftarrow i+1$ 
                }
        else   θέσε  $i \leftarrow i+1, j \leftarrow j+1$ 
      }
  if ( $i \leq n$ ) then πρόσθεσε τις πλειάδες  $R(i)$  έως  $R(n)$  στη σχέση T;
```

T ← R – S

Υλοποίηση ΣΥΝΑΘΡΟΙΣΤΙΚΩΝ πράξεων και ΕΞΩΤΕΡΙΚΩΝ ΣΥΝΕΝΩΣΕΩΝ (1)

- Υλοποίηση ΣΥΝΑΘΡΟΙΣΤΙΚΩΝ Πράξεων:
- ΣΥΝΑΘΡΟΙΣΤΙΚΕΣ συναρτήσεις:
 - MIN, MAX, SUM, COUNT and AVERAGE
- Επιλογές για την υλοποίηση των παραπάνω τελεστών:
 - Σάρωση του πίνακα
 - Χρήση κατάλληλου ευρετηρίου (αν υπάρχει)
- Παράδειγμα:
 - **SELECT MAX** (ΜΙΣΘΟΣ)
 - **FROM** ΕΡΓΑΖΟΜΕΝΟΣ;
- Αν υπάρχει ένα (με αύξουσα σειρά) ευρετήριο στο ΜΙΣΘΟΣ για τη σχέση ΕΡΓΑΖΟΜΕΝΟΣ, τότε ο βελτιστοποιητής μπορεί να αποφασίσει να χρησιμοποιήσει το ευρετήριο για να ψάξει για τη μεγαλύτερη τιμή, ακολουθώντας τον πιο δεξιό δείκτη σε κάθε κόμβο ευρετηρίου από τη ρίζα στο πιο δεξιό φύλλο.
- Αντίστοιχα για τη συνάρτηση MIN, ακολουθείται ο πιο αριστερός δείκτης από τη ρίζα στο πιο αριστερό φύλλο.

Υλοποίηση ΣΥΝΑΘΡΟΙΣΤΙΚΩΝ πράξεων και ΕΞΩΤΕΡΙΚΩΝ ΣΥΝΕΝΩΣΕΩΝ (2)

- Υλοποίηση ΣΥΝΑΘΡΟΙΣΤΙΚΩΝ Πράξεων (συνέχεια):
- **SUM, COUNT και AVG**
 - Για ένα **πυκνό ευρετήριο** (υπάρχει καταχώρηση για κάθε εγγραφή στο κυρίως αρχείο):
 - Οι αντίστοιχοι υπολογισμοί εφαρμόζονται στις τιμές του ευρετηρίου.
- Για ένα **μη-πυκνό ευρετήριο**:
 - Πρέπει να χρησιμοποιηθεί το πραγματικό πλήθος των εγγραφών που αντιστοιχούν σε κάθε καταχώρηση του ευρετηρίου
- **GROUP BY**: η συνάρτηση συνάθροισης πρέπει να εφαρμοστεί ξεχωριστά σε κάθε ομάδα από πλειάδες.
 - Διαμερισμός του πίνακα σε υποσύνολα από πλειάδες, όπου κάθε διαμέριση έχει την ίδια τιμή για τα γνωρίσματα ομαδοποίησης.
 - Η συνάρτηση συνάθροισης εφαρμόζεται για τις πλειάδες σε κάθε ομάδα.
- Αν υπάρχει ένα ευρετήριο συστάδα στα γνωρίσματα ομαδοποίησης;

Υλοποίηση ΣΥΝΑΘΡΟΙΣΤΙΚΩΝ πράξεων και ΕΞΩΤΕΡΙΚΩΝ ΣΥΝΕΝΩΣΕΩΝ (3)

- Υλοποίηση της ΕΞΩΤΕΡΙΚΗΣ ΣΥΝΕΝΩΣΗΣ:
- **Πράξεις Εξωτερικής συνένωσης:**
 - Αριστερή εξωτερική συνένωση (Left outer join)
 - Δεξιά εξωτερική συνένωση (Right outer join)
 - Πλήρη εξωτερική συνένωση (Full outer join)
- Η πλήρης εξωτερική συνένωση δίνει αποτέλεσμα που αντιστοιχεί στην ένωση των αποτελεσμάτων της αριστερής και της δεξιάς εξωτερικής συνένωσης.
- Παράδειγμα:

```
SELECT      ΕΠΙΘΕΤΟ, ΟΝΟΜΑ, Τ_ΟΝΟΜΑ  
FROM        (ΕΡΓΑΖΟΜΕΝΟΣ LEFT OUTER JOIN ΤΜΗΜΑ ON  
              ΑΡΙΘ_Τ = ΚΩΔ_ΤΜΗΜ);
```
- Σημείωση: Το αποτέλεσμα αυτής της ερώτησης είναι ένας πίνακας με ονόματα εργαζομένων και το αντίστοιχο τμήμα τους. Μοιάζει με το αποτέλεσμα μιας κανονικής (εσωτερικής) συνένωσης, με εξαίρεση ότι αν μια πλειάδα ΕΡΓΑΖΟΜΕΝΟΣ δεν έχει ένα αντίστοιχο τμήμα, το όνομα του εργαζόμενου θα εμφανιστεί στον πίνακα του αποτελέσματος, αλλά για αυτή την πλειάδα το όνομα του τμήματος στο αποτέλεσμα της ερώτησης θα είναι null.

Υλοποίηση ΣΥΝΑΘΡΟΙΣΤΙΚΩΝ πράξεων και ΕΞΩΤΕΡΙΚΩΝ ΣΥΝΕΝΩΣΕΩΝ (4)

- Υλοποίηση της ΕΞΩΤΕΡΙΚΗΣ ΣΥΝΕΝΩΣΗΣ:
- **Τροποποιώντας έναν από τους αλγορίθμους συνένωσης:**
 - Συνένωση εμφωλευμένων βρόγχων ή απλής συνένωσης με επαναλήψεις. Για παράδειγμα:
 - Για μια αριστερή εξωτερική συνένωση, χρησιμοποιούμε την αριστερή σχέση στην εξωτερική επανάληψη ή απλή επανάληψη επειδή κάθε πλειάδα αυτής πρέπει να εμφανιστεί στο αποτέλεσμα.
 - Αν υπάρχουν πλειάδες από την άλλη σχέση που ταιριάζουν, παράγονται οι πλειάδες συνένωσης και αποθηκεύονται στο αποτέλεσμα.
 - Αν δεν βρεθεί πλειάδα που να ταιριάζει, η πλειάδα συμπεριλαμβάνεται στο αποτέλεσμα, αλλά συμπληρώνεται με τιμές null.

Υλοποίηση ΣΥΝΑΘΡΟΙΣΤΙΚΩΝ πράξεων και ΕΞΩΤΕΡΙΚΩΝ ΣΥΝΕΝΩΣΕΩΝ (5)

- Υλοποίηση της ΕΞΩΤΕΡΙΚΗΣ ΣΥΝΕΝΩΣΗΣ:
- Εκτέλεση συνδυασμού πράξεων της σχεσιακής άλγεβρας.
- Υλοποίηση της προηγούμενης αριστερής εξωτερικής συνένωσης:
 1. Υπολόγισε την (εσωτερική) συνένωση των πινάκων ΕΡΓΑΖΟΜΕΝΟΣ και ΤΜΗΜΑ
 - $T1 \leftarrow \pi_{\text{ΕΠΙΘΕΤΟ, ΟΝΟΜΑ, ΤΟΝΟΜΑ}}(\text{ΕΡΓΑΖΟΜΕΝΟΣ} \bowtie_{\text{ΑΡΙΘ_Τ=ΚΩΔ_ΤΜΗΜ}} \text{ΤΜΗΜΑ})$
 2. Βρες τις πλειάδες της ΕΡΓΑΖΟΜΕΝΟΣ που δεν εμφανίζεται στο αποτέλεσμα της συνένωσης (εσωτερικής)
 - $T2 \leftarrow \pi_{\text{ΕΠΙΘΕΤΟ, ΟΝΟΜΑ}}(\text{ΕΡΓΑΖΟΜΕΝΟΣ}) - \pi_{\text{ΕΠΙΘΕΤΟ, ΟΝΟΜΑ}}(T1)$
 3. Συμπλήρωσε κάθε πλειάδα της T2 με null τιμές στο πεδίο T_ΟΝΟΜΑ.
 - $T2 \leftarrow T2 \times \text{null}$
 4. Εφάρμοσε την πράξη της συνένωσης των T1 και T2 για να παραχθεί το αποτέλεσμα της αριστερής εξωτερικής συνένωσης.
 - $\text{ΑΠΟΤΕΛΕΣΜΑ} \leftarrow T1 \cup T2$
- Το κόστος αυτού του υπολογισμού της εξωτερικής συνένωσης, είναι το άθροισμα του κόστους των επιμέρους βημάτων (εσωτερική συνένωση, προβολές και ένωση).

Σύνοψη κεφαλαίου

- Εισαγωγή στη βελτιστοποίηση ερωτήσεων
- Μετάφραση SQL ερωτήσεων σε Σχεσιακή Άλγεβρα
- Αλγόριθμοι Εξωτερικής Ταξινόμησης
- Αλγόριθμοι για τις πράξεις ΕΠΙΛΟΓΗ και ΣΥΝΕΝΩΣΗ
- Αλγόριθμοι για ΠΡΟΒΟΛΗ και συνόλων
- Υλοποίηση Συναθροιστικών πράξεων και εξωτερικών συνενώσεων

Οργάνωση Βάσεων Βιοϊατρικών Δεδομένων Εξόρυξη Γνώσης Βιοϊατρικών Δεδομένων

Σεμινάριο 7, μέρος 2^ο: Αλγόριθμοι για επεξεργασία ερωτήσεων και βελτιστοποίηση

**Ευάγγελος Καρκαλέτσης,
Αναστασία Κριθαρά, Γεώργιος Πετάσης**

Εργαστήριο Τεχνολογίας Γνώσεων & Λογισμικού,

Ινστιτούτο Πληροφορικής & Τηλεπικοινωνιών, Ε.Κ.Ε.Φ.Ε. “Δημόκριτος”

Τηλ.: 210-6503197, Fax: 210-6532175, {vangelis, akrihara, petasis}@iit.demokritos.gr

Ακαδημαϊκό Έτος: 2016 – 2017

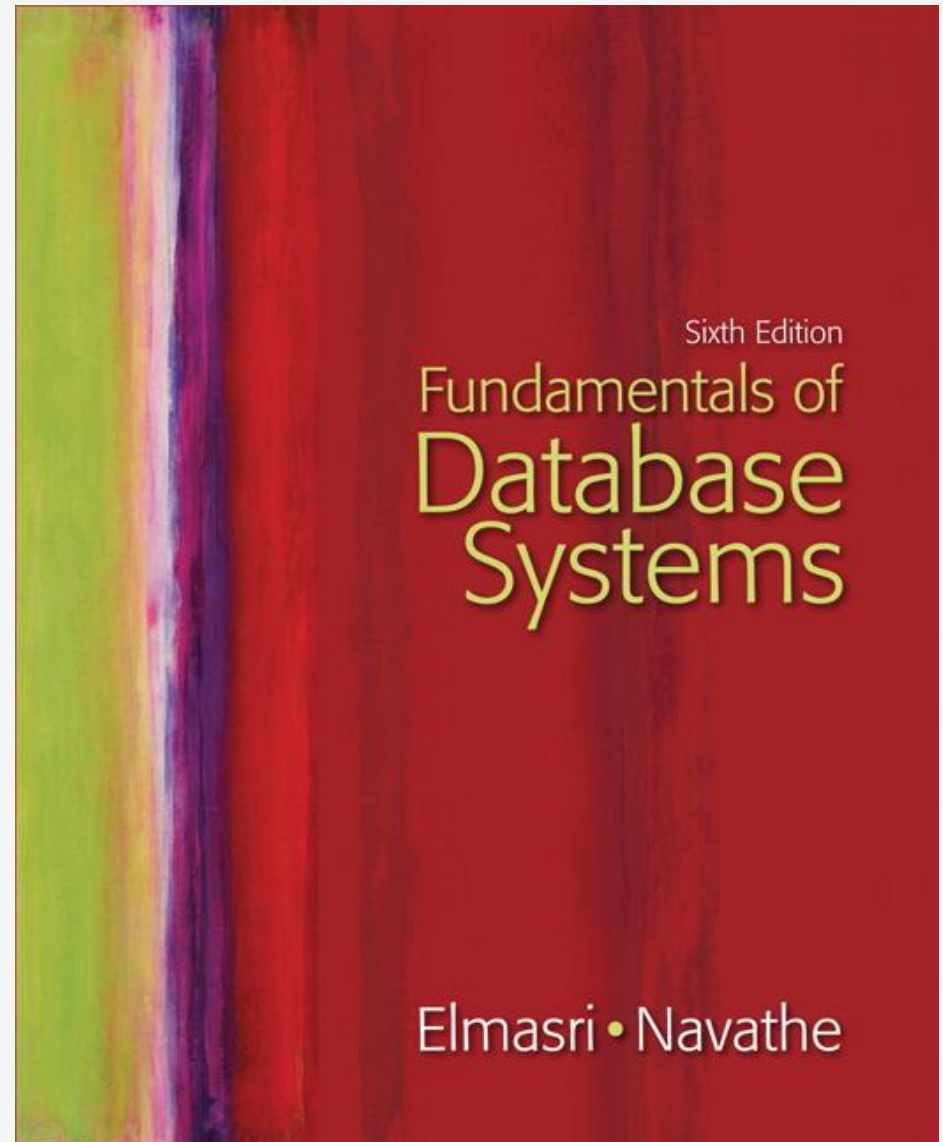
Οι διαφάνειες αυτού του σεμιναρίου βασίζονται
σε διαφάνειες των:

Δ. Γουνόπουλος, Ι. Ιωαννίδης

<http://eclass.uoa.gr/modules/document/document.php?course=D463&openDir=/4c2b32a5i4iq>

ΚΕΦΑΛΑΙΟ 19*

Αλγόριθμοι Επεξεργασίας και
Βελτιστοποίησης Ερωτήσεων



Copyright © 2011 Ramez Elmasri and Shamkant B. Navathe, Ελληνική Έκδοση Δίαυλος

Θα μιλήσουμε για

- Συνδυασμός Πράξεων με χρήση σωλήνωσης
- Χρήση ευριστικών μεθόδων στη βελτιστοποίηση ερωτήσεων
- Χρήση επιλεξιμότητας και εκτιμήσεων κόστους στη βελτιστοποίηση ερωτήσεων
- Σημασιολογική βελτιστοποίηση ερωτήσεων

Συνδυασμός πράξεων χρησιμοποιώντας σωλήνωση (1)

- Μία ερώτηση μεταφράζεται σε μια ακολουθία από σχεσιακές πράξεις.
 - Η εκτέλεση μία προς μία των πράξεων, δημιουργεί τόσα προσωρινά αρχεία όσες και οι πράξεις.
 - Η δημιουργία και αποθήκευση μεγάλων προσωρινών αρχείων στο δίσκο είναι χρονοβόρα και ακριβή.
- Εναλλακτική προσέγγιση:
 - **Ελάττωση του πλήθους των προσωρινών αρχείων**
 - Οι πλειάδες του αποτελέσματος μιας προηγούμενης πράξης προωθούνται ως είσοδος στην επόμενη πράξη σε μια **ροή (ή σωλήνωση)** αμέσως μόλις παράγονται.

Συνδυασμός πράξεων χρησιμοποιώντας σωλήνωση (2)

- Παράδειγμα:
 - Αντί να υλοποιηθεί ξεχωριστά μία συνένωση, μπορεί να συνδυαστεί με δύο πράξεις επιλογής στα αρχεία εισόδου και μία τελική πράξη προβολής στο αρχείο του αποτελέσματος.
- Δυναμική δημιουργία του κώδικα για υλοποίηση πολλαπλών πράξεων.
- Οι πλειάδες του αποτελέσματος κάθε επιλογής προωθούνται ως είσοδος στον αλγόριθμο συνένωσης.
- Σωληνωτή εκτέλεση ή επεξεργασία βασισμένη στη ροή

Χρησιμοποίηση ευριστικών μεθόδων στη βελτιστοποίηση ερωτήσεων (1)

- Τεχνική βελτιστοποίησης με χρήση ευριστικών κανόνων
 1. Σε μια ερώτηση υψηλού επιπέδου ο συντακτικός αναλυτής δημιουργεί μια *αρχική εσωτερική αναπαράσταση*
 2. Βελτιστοποίηση της αναπαράστασης σύμφωνα με ευριστικούς κανόνες
 3. Δημιουργείται ένα σχέδιο εκτέλεσης ομάδων πράξεων, με βάση τις διαθέσιμες διαδρομές προσπέλασης στα αρχεία.
- Ένας από τους **κύριους ευριστικούς κανόνες** είναι να εκτελούνται οι πράξεις που ελαττώνουν το μέγεθος των προσωρινών αρχείων.
 - Παράδειγμα: εκτέλεση των πράξεων ΕΠΙΛΟΓΗ και ΠΡΟΒΟΛΗ πριν από την εκτέλεση των πράξεων συνένωσης και των άλλων δυαδικών πράξεων.

Χρησιμοποίηση ευριστικών μεθόδων στη βελτιστοποίηση ερωτήσεων (2)

- **Δέντρο ερώτησης:**

- Μία δενδρική δομή που αντιστοιχεί σε μια έκφραση της σχεσιακής άλγεβρας.
- Οι σχέσεις εισόδου αναπαρίστανται ως *κόμβοι φύλλα* του δένδρου
- Οι πράξεις σχεσιακής άλγεβρας ως *εσωτερικοί κόμβοι*.
- Μια εκτέλεση του δένδρου ερώτησης συνίσταται στην *εκτέλεση των πράξεων κάθε εσωτερικού κόμβου*, όταν οι σχέσεις εισόδου είναι διαθέσιμες και στη συνέχεια αντικατάσταση του εσωτερικού κόμβου από τη σχέση η οποία προκύπτει από την εκτέλεση της πράξης.

Χρησιμοποίηση ευριστικών μεθόδων στη βελτιστοποίηση ερωτήσεων (3)

- **Γράφος ερώτησης:**

- Μία γραφική δομή που αντιστοιχεί σε μια έκφραση του σχεσιακού λογισμού.
- Οι σχέσεις στην ερώτηση αναπαρίστανται από **κόμβους σχέσεων** (απλοί κύκλοι).
- Οι σταθερές τιμές αναπαρίστανται με **σταθερούς κόμβους** (διπλοί κύκλοι).
- Οι πράξεις αναπαρίστανται με **πλευρές** του γράφου.
- *Δεν δείχνει την σειρά εκτέλεσης των πράξεων.*
- Υπάρχει *μόνο ένας* γράφος που αντιστοιχεί σε κάθε ερώτηση.

Χρησιμοποίηση ευριστικών μεθόδων στη βελτιστοποίηση ερωτήσεων (4)

- Παράδειγμα:
 - Για κάθε έργο στην τοποθεσία 'Stafford', να ανακτηθεί ο κωδικός έργου, ο κωδικός του τμήματος που το ελέγχει και το επίθετο, η διεύθυνση και η ημερομηνία γέννησης του διευθυντή του τμήματος.

- Σχεσιακή άλγεβρα:

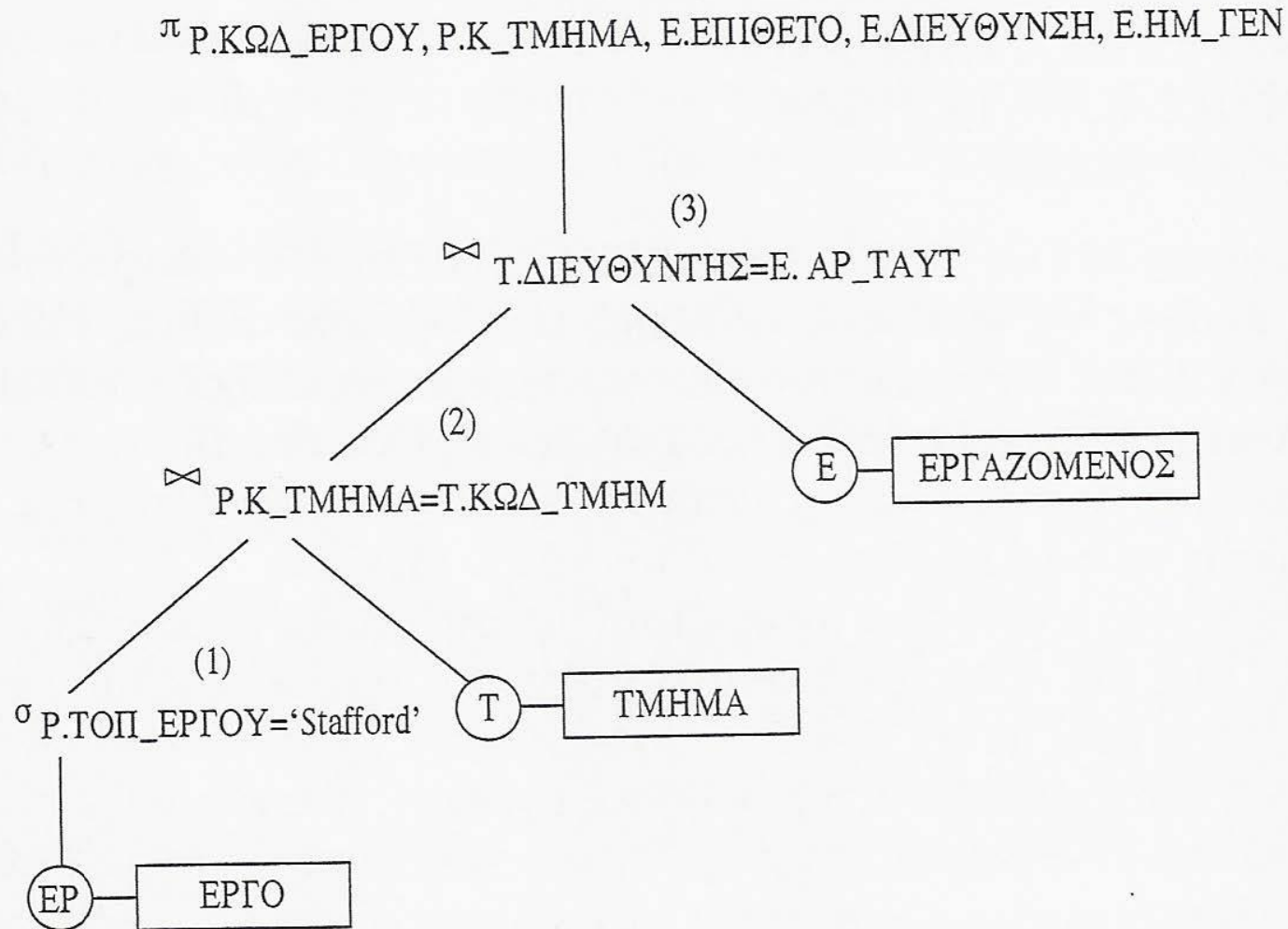
$$\pi_{KΩΔ_{ΕΡΓΟΥ}, KΩΔ_{ΤΜΗΜ}, ΕΠΙΘΕΤΟ, ΔΙΕΥΘΥΝΣΗ, ΗΜΕΡ_{ΓΕΝ}}(((\sigma_{ΤΟΠ_{ΕΡΓΟΥ}='Stanford'}(ΕΡΓΟ)) \bowtie_{K_{ΤΜΗΜΑ}=KΩΔ_{ΤΜΗΜ}}(ΤΜΗΜΑ)) \bowtie_{ΔΙΕΥΘΥΝΤΗΣ=ΑΡ_{ΤΑΥΤ}} ΕΡΓΑΖΟΜΕΝΟΣ))$$

- Ερώτηση SQL :

```
SELECT  P.KΩΔ_ΕΡΓΟΥ, P.K_ΤΜΗΜΑ, Ε.ΕΠΙΘΕΤΟ,  
        Ε.ΔΙΕΥΘΥΝΣΗ, Ε.ΗΜ_ΓΕΝ  
FROM    ΕΡΓΟ AS P, ΤΜΗΜΑ AS T, ΕΡΓΑΖΟΜΕΝΟΣ AS E  
WHERE   R.K_ΤΜΗΜΑ=T.KΩΔ_ΤΜΗΜ AND  
        T.ΔΙΕΥΘΥΝΤΗΣ=E.ΑΡ_ΤΑΥΤ AND  
        P.ΤΟΠ_ΕΡΓΟΥ='STAFFORD';
```

Χρησιμοποίηση ευριστικών μεθόδων στη βελτιστοποίηση ερωτήσεων (5)

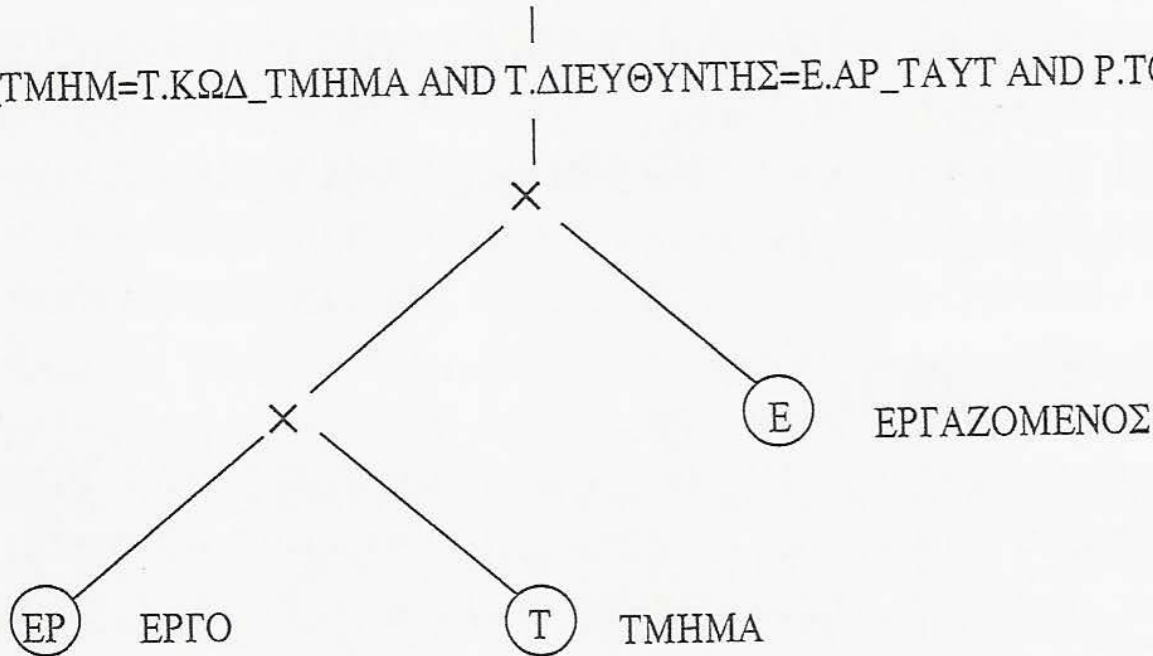
(α)



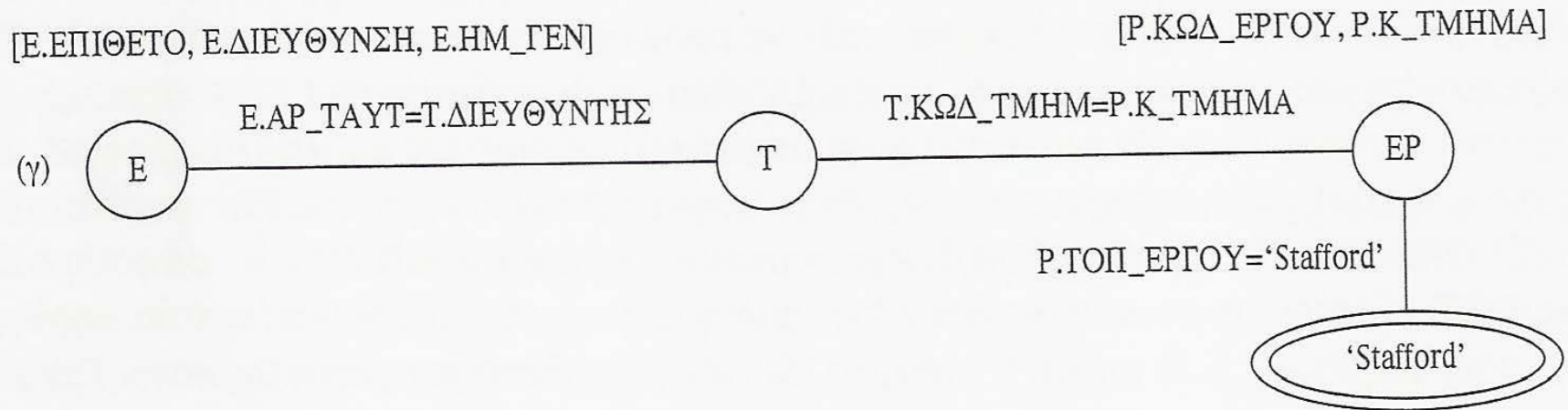
Χρησιμοποίηση ευριστικών μεθόδων στη βελτιστοποίηση ερωτήσεων (6)

(β) π P.KΩΔ_ΕΡΓΟΥ, P.K_TMΗΜΑ, E.ΕΠΙΘΕΤΟ, E.ΔΙΕΥΘΥΝΣΗ, E.HM_GEN

σ P.K_TMΗΜ=T.KΩΔ_TMΗΜΑ AND T.ΔΙΕΥΘΥΝΤΗΣ=E.ΑΡ_ΤΑΥΤ AND P.ΤΟΠ_ΕΡΓΟΥ='Stafford'



Χρησιμοποίηση ευριστικών μεθόδων στη βελτιστοποίηση ερωτήσεων (7)



Χρησιμοποίηση ευριστικών μεθόδων στη βελτιστοποίηση ερωτήσεων (8)

- Ευριστική βελτιστοποίηση δένδρων ερωτήσεων:
 - Πολλές διαφορετικές εκφράσεις της σχεσιακής άλγεβρας – και επομένως πολλά διαφορετικά δέντρα ερωτήσεων – μπορούν είναι ισοδύναμες.
 - Ο ευριστικός βελτιστοποιητής μετασχηματίζει το αρχικό δέντρο ερώτησης σε ένα τελικό δένδρο ερώτησης, το οποίο μπορεί να εκτελεστεί με αποδοτικό τρόπο.
 - Παράδειγμα:

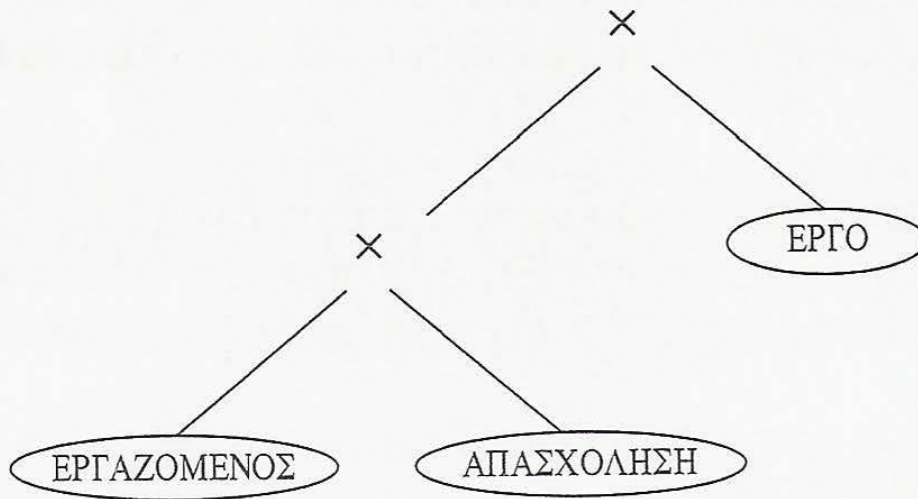
E: SELECT	ΕΠΙΘΕΤΟ
FROM	ΕΡΓΑΖΟΜΕΝΟΣ, ΑΠΑΣΧΟΛΗΣΗ, ΕΡΓΟ
WHERE	E_ONOMA='AQUARIUS' AND
	KΩΔ_ΕΡΓΟΥ=Κ_ΕΡΓΟ AND
	E_ΑΡΤΑΥΤ=ΑΡ_ΤΑΥΤ AND
	ΗΜ_ΓΕΝ > '1957-12-31';

Χρησιμοποίηση ευριστικών μεθόδων στη βελτιστοποίηση ερωτήσεων (9)

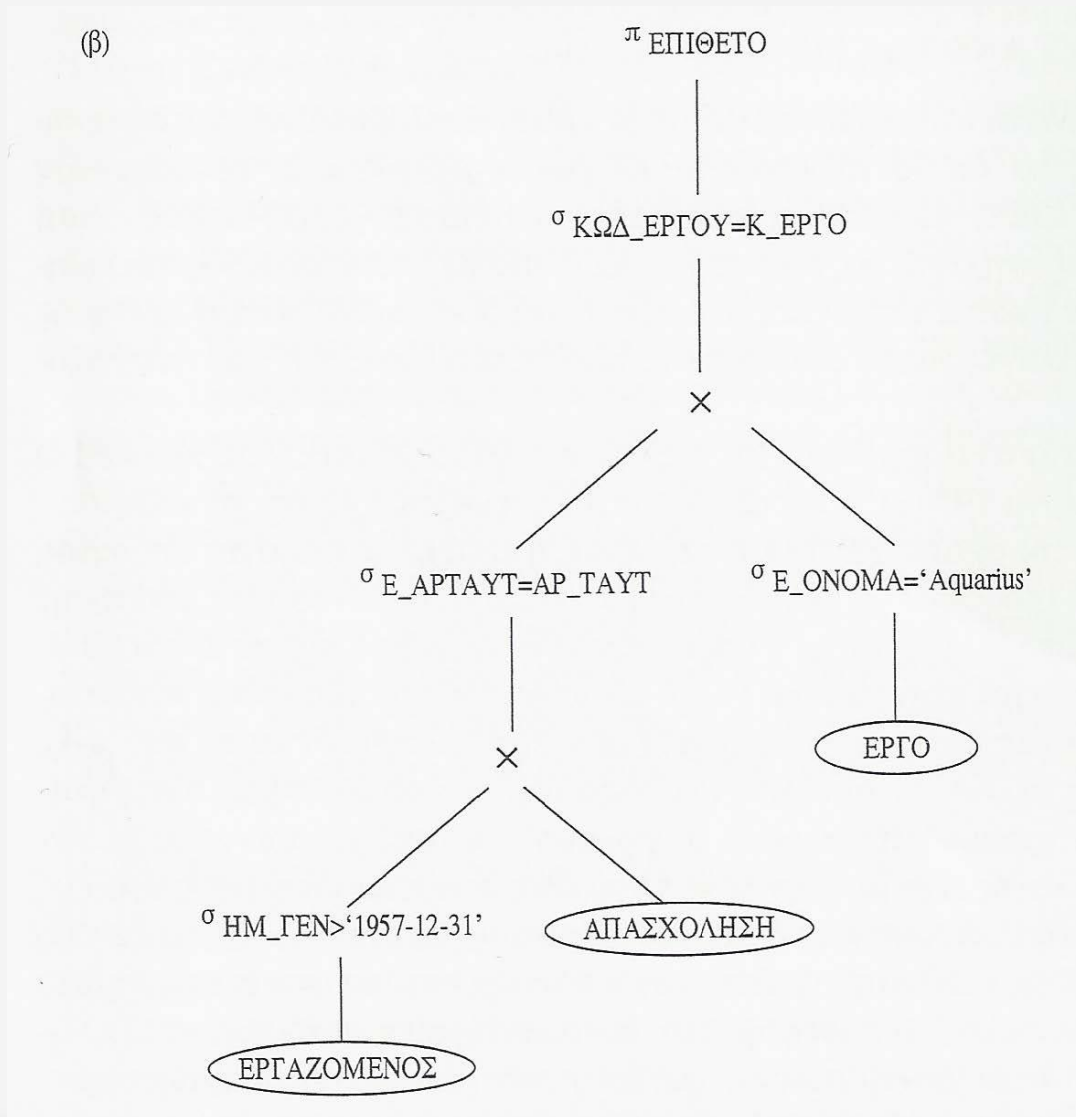
(α)

π ΕΠΙΘΕΤΟ

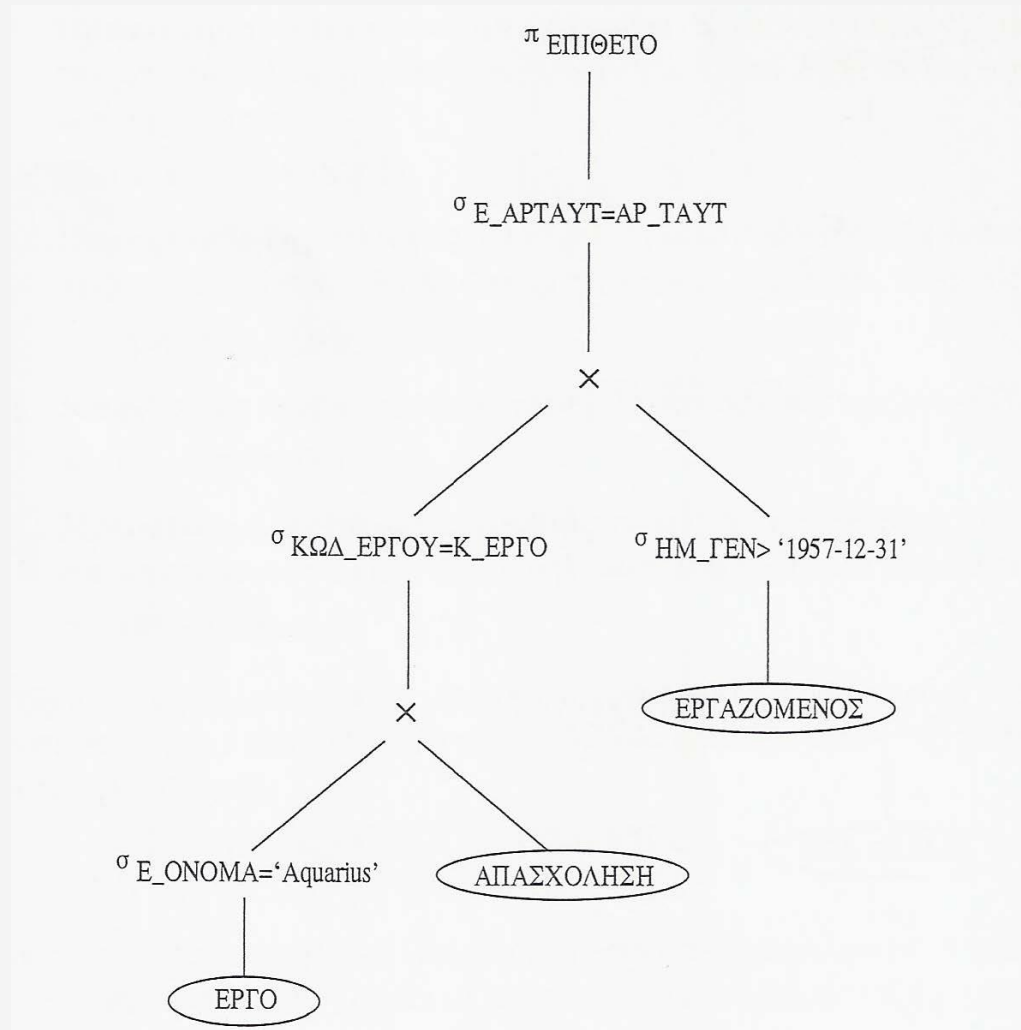
σ E_ONOMA='Aquarius' AND ΚΩΔ_ΕΡΓΟΥ=K_ΕΡΓΟ AND Ε_ΑΡΤΑΥΤ=ΑΡ_ΤΑΥΤ AND ΗΜ_ΓΕΝ>'1957-12-31'



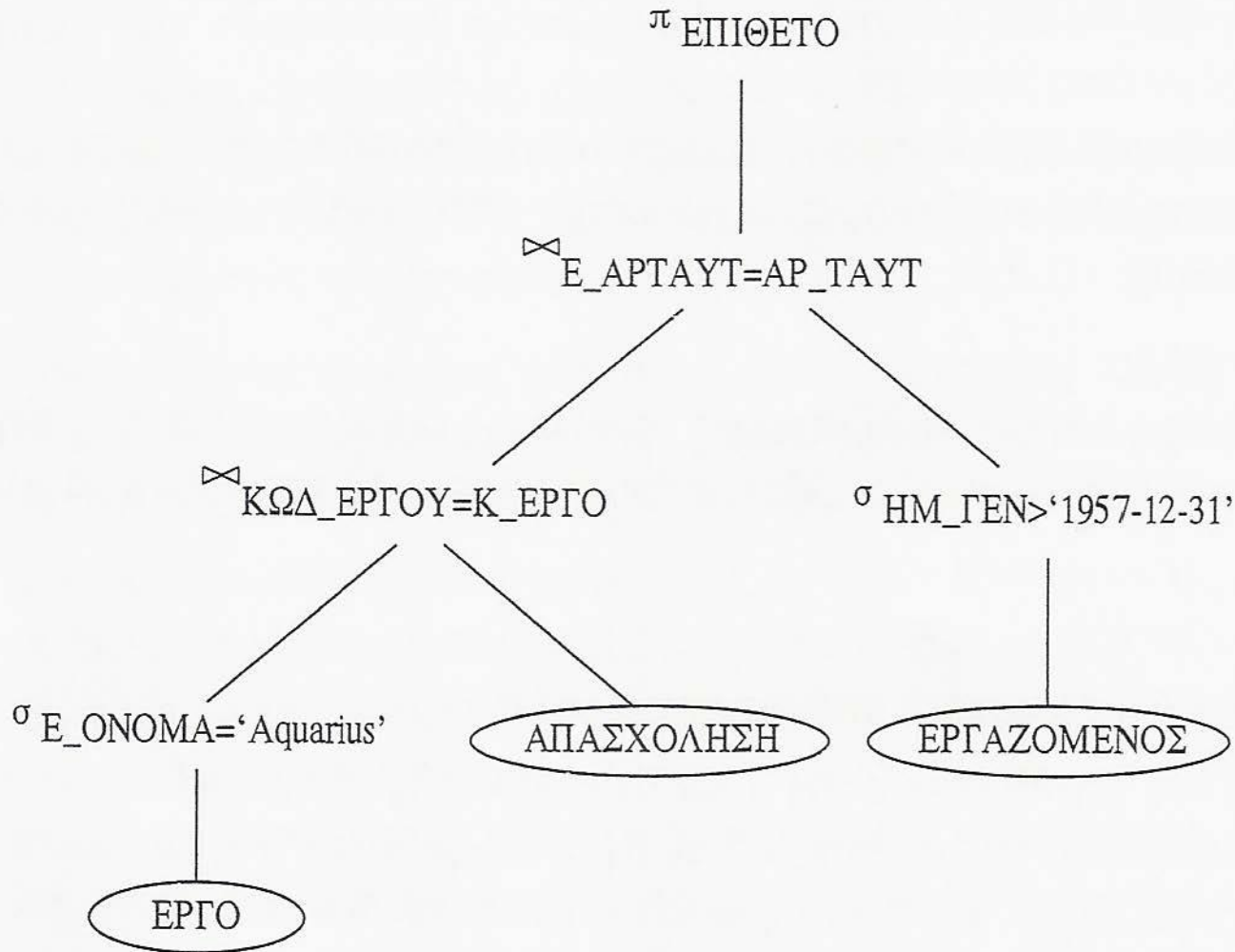
Χρησιμοποίηση ευριστικών μεθόδων στη βελτιστοποίηση ερωτήσεων (10)



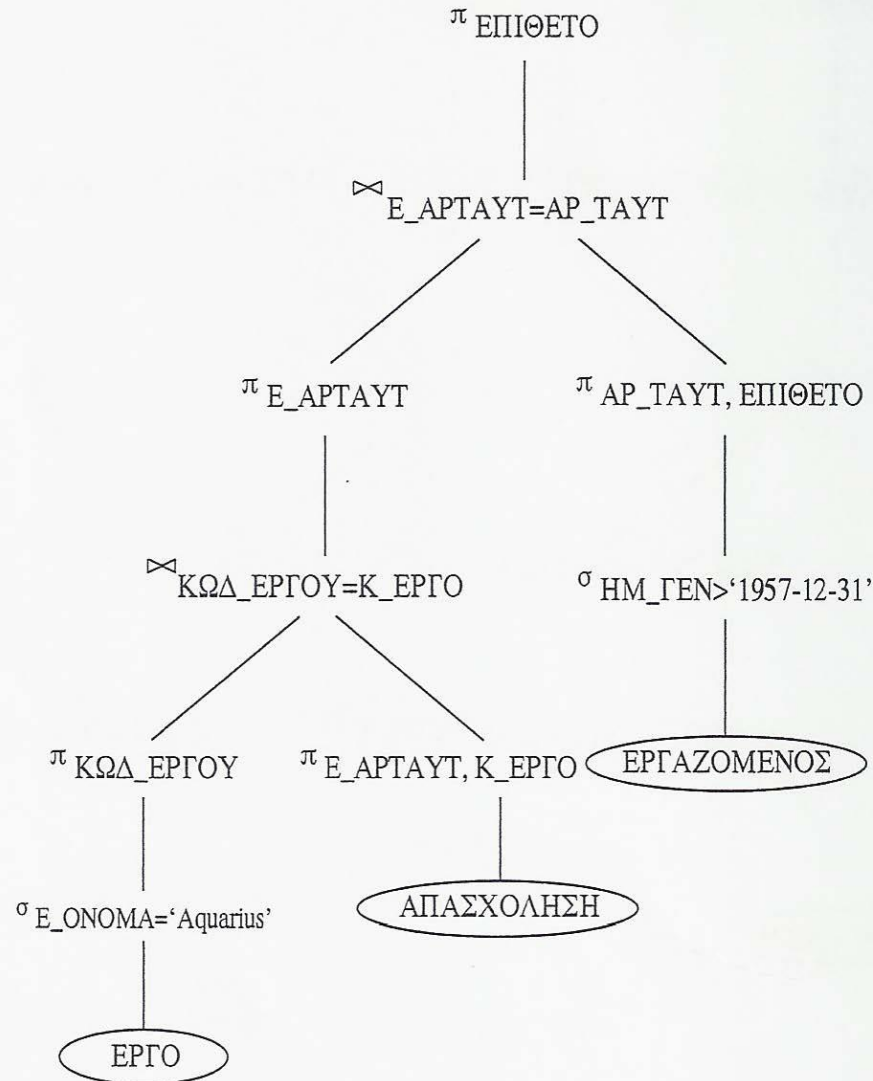
Χρησιμοποίηση ευριστικών μεθόδων στη βελτιστοποίηση ερωτήσεων (11)



Χρησιμοποίηση ευριστικών μεθόδων στη βελτιστοποίηση ερωτήσεων (12)



Χρησιμοποίηση ευριστικών μεθόδων στη βελτιστοποίηση ερωτήσεων (13)



Χρησιμοποίηση ευριστικών μεθόδων στη βελτιστοποίηση ερωτήσεων (14)

- Γενικοί κανόνες μετασχηματισμού για πράξεις της σχεσιακής άλγεβρας:
 1. **Διάδοση (cascade) της επιλογής (σ):** μία συζευκτική συνθήκη επιλογής μπορεί να διασπαστεί σε μια διαδοχή (ακολουθία) πράξεων επιλογής:
 - $\sigma_{c_1 \text{ AND } c_2 \text{ AND } \dots \text{ AND } c_n}(R) \equiv \sigma_{c_1}(\sigma_{c_2}(\dots(\sigma_{c_n}(R)) \dots))$
 2. **Αντιμεταθετικότητα της σ :** η επιλογή είναι αντιμεταθετική πράξη
 - $\sigma_{c_1}(\sigma_{c_2}(R)) \equiv \sigma_{c_2}(\sigma_{c_1}(R))$
 3. **Διάδοση της προβολής (π):** σε μια διαδοχή (ακολουθία) από πράξεις προβολής, μπορούν να αγνοηθούν όλες εκτός από την τελευταία:
 - $\pi_{\text{λιστα}_1}(\pi_{\text{λιστα}_2}(\dots(\pi_{\text{λιστα}_n}(R)) \dots)) \equiv \pi_{\text{λιστα}_1}(R)$
 4. **Αντιμετάθεση της σ με την π :** αν η συνθήκη επιλογής c περιλαμβάνει μόνο τα γνωρίσματα $A_1, A_2, \dots, A_n$ της λίστας προβολής, οι δύο πράξεις μπορούν να αντιμετατεθούν:
 - $\pi_{A_1, \dots, A_n}(\sigma_c(R)) \equiv \sigma_c(\pi_{A_1, \dots, A_n}(R))$

Χρησιμοποίηση ευριστικών μεθόδων στη βελτιστοποίηση ερωτήσεων (15)

- Γενικοί κανόνες μετασχηματισμού για πράξεις της σχεσιακής άλγεβρας (συνέχεια):

5. Αντιμεταθετικότητα της $\bowtie$ (και της $\times$): η πράξη $\bowtie$ είναι αντιμεταθετική, όπως και η πράξη $\times$:

$$- R \bowtie_c S \equiv S \bowtie_c R; R \times S \equiv S \times R$$

6. Αντιμεταθετικότητα της σ με την $\bowtie$ (ή την $\times$): αν το σύνολο των γνωρισμάτων της συνθήκης επιλογής c περιλαμβάνει μόνο γνωρίσματα μιας από τις σχέσεις συνένωσης – έστω της R – οι δύο πράξεις μπορούν να αντιμεταθούν ως ακολούθως:

$$- \sigma_c(R \bowtie S) \equiv (\sigma_c(R)) \bowtie S$$

– Εναλλακτικά, αν η συνθήκη επιλογής c μπορεί να γραφεί ως $(c_1 \text{ and } c_2)$, όπου η συνθήκη c_1 περιλαμβάνει μόνο γνωρίσματα της R και η συνθήκη c_2 μόνο γνωρίσματα της S , οι πράξεις αντιμετατίθενται ως ακολούθως:

$$- \sigma_c(R \bowtie S) \equiv (\sigma_{c_1}(R)) \bowtie (\sigma_{c_2}(S))$$

Χρησιμοποίηση ευριστικών μεθόδων στη βελτιστοποίηση ερωτήσεων (16)

- Γενικοί κανόνες μετασχηματισμού για πράξεις της σχεσιακής άλγεβρας (συνέχεια):

7. Αντιμεταθετικότητα των $\bowtie$ (ή $\times$): ας υποθέσουμε ότι η λίστα προβολής είναι $L = \{A_1, \dots, A_n, B_1, \dots, B_m\}$, όπου τα $A_1, \dots, A_n$ είναι γνωρίσματα της σχέσης R και $B_1, \dots, B_m$ της σχέσης S . Αν η συνθήκη συνένωσης c περιλαμβάνει μόνο γνωρίσματα που ανήκουν στην L , οι δύο πράξεις μπορούν να αντιμετωπιστούν ως ακολούθως:

$$- \pi_L(R \bowtie_c S) \equiv (\pi_{A_1, \dots, A_n}(R)) \bowtie_c (\pi_{B_1, \dots, B_m}(S))$$

- Αν η συνθήκη συνένωσης c περιλαμβάνει επιπλέον γνωρίσματα που δεν βρίσκονται στην L , αυτά πρέπει να προστεθούν στη λίστα προβολής και απαιτείται μια τελική πράξη προβολής.

Χρησιμοποίηση ευριστικών μεθόδων στη βελτιστοποίηση ερωτήσεων (17)

- Γενικοί κανόνες μετασχηματισμού για πράξεις της σχεσιακής άλγεβρας (συνέχεια):
- 8. Αντιμεταθετικότητα των συνολοθεωρητικών πράξεων:** οι συνολοθεωρητικές πράξεις $\cup$ και $\cap$ είναι αντιμεταθετικές, όχι όμως και η διαφορά ($-$).
 - 9. Προσεταιριστικότητα των $\bowtie$, $\times$, $\cup$ και $\cap$:** Κάθε μία από αυτές τις πράξεις είναι προσεταιριστική. Δηλαδή, αν το θ παριστάνει μία από τις τέσσερις αυτές τις πράξεις, τότε έχουμε:
$$- (R \theta S) \theta T \equiv R \theta (S \theta T)$$
 - 10. Επιμερισμός της επιλογής (σ) με τις συνολοθεωρητικές πράξεις:** η πράξη σ είναι επιμεριστική ως προς τις $\cup$, $\cap$ και $-$. Αν το θ παριστάνει μια από αυτές τις πράξεις, τότε έχουμε:
$$- \sigma_c(R \theta S) \equiv (\sigma_c(R)) \theta (\sigma_c(S))$$

Χρησιμοποίηση ευριστικών μεθόδων στη βελτιστοποίηση ερωτήσεων (18)

- Γενικοί κανόνες μετασχηματισμού για πράξεις της σχεσιακής άλγεβρας (συνέχεια):

11. Η πράξη της προβολής είναι επιμεριστική ως προς την ένωση $\cup$:

$$- \pi_L(R \cup S) \equiv (\pi_L(R)) \cup (\pi_L(S))$$

12. Μετατροπή μιας $(\sigma, \times)$ ακολουθίας σε $\bowtie$: Αν η συνθήκη c μιας επιλογής σ που ακολουθείται από ένα $\times$ αντιστοιχεί σε μια συνθήκη συνένωσης, η ακολουθία $(\sigma, \times)$ μετατρέπεται σε $\bowtie$ όπως:

$$- ((\sigma_c(R \times S)) \equiv (R \bowtie_c S))$$

- Άλλοι πιθανοί μετασχηματισμοί

Χρησιμοποίηση ευριστικών μεθόδων στη βελτιστοποίηση ερωτήσεων (19)

- Περιγραφή ενός ευριστικού αλγεβρικού αλγορίθμου βελτιστοποίησης:
 1. Χρησιμοποιώντας τον κανόνα 1, διασπώνται οι πράξεις επιλογής με συζευκτικές συνθήκες σε ακολουθίες πράξεων επιλογής.
 2. Χρησιμοποιώντας τους κανόνες 2, 4, 6, και 10 που αφορούν τη αντιμεταθετικότητα της επιλογής με άλλες πράξεις, μετατοπίζεται κάθε πράξη επιλογής όσο πιο κάτω επιτρέπεται από τα γνωρίσματα που περιλαμβάνονται στη συνθήκη επιλογής.
 3. Χρησιμοποιώντας τους κανόνες 5 και 9 που αφορούν την αντιμεταθετικότητα και την προσεταιριστικότητα των δυαδικών πράξεων, επαναδιευθετούνται οι κόμβοι-φύλλα έτσι που να εκτελούνται πρώτα οι κόμβοι-φύλλα με τις σχέσεις που έχουν τις πιο περιοριστικές πράξεις επιλογής.

Χρησιμοποίηση ευριστικών μεθόδων στη βελτιστοποίηση ερωτήσεων (20)

- Περιγραφή ενός ευριστικού αλγεβρικού αλγορίθμου βελτιστοποίησης (συνέχεια):
- 4. Χρησιμοποιώντας τον κανόνα 12, συνδυασμός μιας πράξης Καρτεσιανού γινομένου με μια πράξη επιλογής που ακολουθεί σε μια συνένωση, αν η συνθήκη αναπαριστά μια συνθήκη συνένωσης.
- 5. Χρησιμοποιώντας τους κανόνες 3, 4, 7, και 11 που αφορούν τη διάδοση της προβολής και την επιμεριστικότητα της προβολής με άλλες πράξεις, οι λίστες γνωρισμάτων προβολής διασπώνται και μετατοπίζονται όσο το δυνατόν πιο κάτω στο δένδρο ερώτησης. Όπου είναι απαραίτητο, δημιουργούνται νέοι κόμβοι πράξεων προβολής.
- 6. Εντοπισμός υποδέντρων που αναπαριστούν ομάδες πράξεων οι οποίες μπορούν να εκτελεσθούν από ένα μόνον αλγόριθμο.

Χρησιμοποίηση ευριστικών μεθόδων στη βελτιστοποίηση ερωτήσεων (21)

- **Περίληψη της ευριστικής αλγεβρικής βελτιστοποίησης:**

1. Η κύρια ευριστική μέθοδος είναι να εφαρμόζονται πρώτα οι πράξεις που ελαττώνουν το μέγεθος των ενδιάμεσων αποτελεσμάτων.
2. Εκτέλεση των πράξεων της επιλογής όσο το δυνατόν πιο γρήγορα, έτσι ώστε να ελαττωθεί το πλήθος των πλειάδων και η εκτέλεση των προβολών όσο το δυνατόν πιο γρήγορα, προκειμένου να ελαττωθεί το πλήθος των γνωρισμάτων. (Αυτό επιτυγχάνεται μετατοπίζοντας τις πράξεις επιλογής και προβολής όσο το δυνατόν πιο χαμηλά στο δένδρο.)
3. Οι πιο περιοριστικές πράξεις επιλογής και συνένωσης πρέπει να εκτελούνται πριν από άλλες παρόμοιες πράξεις. (Αυτό επιτυγχάνεται αναδιατάσσοντας μεταξύ τους κόμβους-φύλλα του δένδρου αποφεύγοντας Καρτεσιανά γινόμενα και προσαρμόζοντας κατάλληλα το υπόλοιπο δέντρο.)

Χρησιμοποίηση ευριστικών μεθόδων στη βελτιστοποίηση ερωτήσεων (22)

- **Σχέδια εκτέλεσης ερωτήσεων**

- Ένα σχέδιο εκτέλεσης για μια έκφραση της σχεσιακής άλγεβρας που παριστάνεται σαν δένδρο ερωτήσεων περιλαμβάνει πληροφορίες για τις διαθέσιμες μεθόδους προσπέλασης για κάθε σχέση καθώς και για τους αλγορίθμους που μπορούν να χρησιμοποιηθούν για την υπολογισμό των σχεσιακών τελεστών που παριστάνονται στο δένδρο.
- **Υλοποιημένο υπολογισμό:** το αποτέλεσμα μιας πράξης αποθηκεύεται σε μια προσωρινή σχέση.
- **Σωληνωτό υπολογισμό:** καθώς παράγονται οι πλειάδες του αποτελέσματος μιας πράξης, στέλνονται άμεσα στην επόμενη πράξη της ακολουθίας της ερώτησης.

Χρήση της επιλεξιμότητας και της εκτίμησης κόστους στη βελτιστοποίηση ερωτήσεων (1)

- **Βελτιστοποίηση ερωτήσεων που βασίζεται στο κόστος:**
 - Εκτίμηση και σύγκριση του κόστους εκτέλεσης μιας ερώτησης χρησιμοποιώντας διαφορετικές στρατηγικές εκτέλεσης, και επιλογή της στρατηγικής με το *μικρότερο εκτιμώμενο κόστος*.
- Ζητήματα
 - Συνάρτηση κόστους
 - Πλήθος των στρατηγικών εκτέλεσης που θα εξεταστούν

Χρήση της επιλεξιμότητας και της εκτίμησης κόστους στη βελτιστοποίηση ερωτήσεων (2)

- Παράγοντες κόστους εκτέλεσης ερωτήσεων
 1. Κόστος προσπέλασης στη δευτερεύουσα μνήμη
 2. Κόστος αποθήκευσης
 3. Υπολογιστικό κόστος
 4. Κόστος χρήσης της μνήμης
 5. Επικοινωνιακό κόστος
- Σημείωση: Διαφορετικά συστήματα βάσεων δεδομένων δίνεται διαφορετική έμφαση στους διαφορετικούς παράγοντες κόστους.

Χρήση της επιλεξιμότητας και της εκτίμησης κόστους στη βελτιστοποίηση ερωτήσεων (3)

- Πληροφορίες καταλόγου που χρησιμοποιούνται στις συναρτήσεις κόστους:
 - Πληροφορίες για το μέγεθος κάθε αρχείου
 - Το πλήθος εγγραφών (πλειάδες) (r),
 - Το (μέσο) μέγεθος εγγραφής (R),
 - Το πλήθος των μπλοκ (b),
 - Ο παράγων σχηματισμού μπλοκ (blocking factor) (bfr)
- Πληροφορίες για τα ευρετήρια και τα ευρετηριασμένα γνωρίσματα κάθε αρχείου
 - Το πλήθος των επιπέδων (x) κάθε πολυεπίπεδου ευρετηρίου
 - Το πλήθος μπλοκ του πρώτου επιπέδου ευρετηρίου (b_{I1})
 - Το πλήθος των διακριτών τιμών (d) ενός γνωρίσματος
 - Η επιλεξιμότητα (sl) ενός γνωρίσματος
 - Πληθάριθμος επιλογής ενός γνωρίσματος ($s = sl * r$)

Χρήση της επιλεξιμότητας και της εκτίμησης κόστους στη βελτιστοποίηση ερωτήσεων (4)

Παράδειγμα:

- Έστω ένα γνώρισμα κλειδί:
 - των διακριτών τιμών (d) ενός γνωρίσματος;
 - Η επιλεξιμότητας (sl) ενός γνωρίσματος;
 - Πληθάριθμος επιλογής ενός γνωρίσματος ($s = sl * r$);

$$\sqrt{d} = r$$

$$\sqrt{sl} = 1/r$$

$$\sqrt{s} = 1$$

Χρήση της επιλεξιμότητας και της εκτίμησης κόστους στη βελτιστοποίηση ερωτήσεων (5)

- Παραδείγματα συναρτήσεων κόστους για την πράξη της επιλογής (Κόστος μεθόδου E_i : $C_{Ei} \leftarrow$ μεταφορές μπλοκ)
 - **E1. Μέθοδος σειριακής αναζήτησης**
 - $C_{E1a} = b$
 - Για συνθήκη ισότητας σε ένα κλειδί $C_{E1\beta} = \left(\frac{b}{2}\right)$ αν υπάρχει εγγραφή, αλλιώς $C_{E1\beta} = b$
 - **E2. Δυαδική αναζήτηση**
 - $C_{E2} = \log_2 b + \left\lceil \left(\frac{s}{bfr}\right) \right\rceil - 1$
 - Για συνθήκη ισότητας σε ένα κλειδί $C_{E2} = \log_2 b$, αφού $s = 1$

Χρήση της επιλεξιμότητας και της εκτίμησης κόστους στη βελτιστοποίηση ερωτήσεων (6)

- Παραδείγματα συναρτήσεων κόστους για την πράξη της επιλογής (συνέχεια)
 - Ε3α. Χρήση πρωτεύοντος ευρετηρίου για την ανάκτηση μιας εγγραφής
 - $C_{E3a} = \chi + 1$
 - Ε3β. Χρήση κλειδιού κατακερματισμού για την ανάκτηση μιας εγγραφής
 - $C_{E3b} = 1$

Χρήση της επιλεξιμότητας και της εκτίμησης κόστους στη βελτιστοποίηση ερωτήσεων (7)

- Παραδείγματα συναρτήσεων κόστους για την πράξη της επιλογής (συνέχεια)
 - **E4. Χρήση ευρετηρίου διάταξης για την ανάκτηση πολλαπλών εγγραφών**
 - $C_{E4} = \chi + (\frac{b}{2})$
 - **E5. Χρήση ευρετηρίου συστάδων για την ανάκτηση πολλών εγγραφών**
 - $C_{E5} = x + \left\lceil (\frac{s}{bfr}) \right\rceil$
 - **E6. Χρήση δευτερεύοντος ευρετηρίου (B⁺-δέντρου)**
 - Για σύγκριση ισότητας: $C_{E6\alpha} = \chi + s$
 - Για σύγκριση $>, \geq, <$ ή $\leq$: $C_{E6\beta} = \chi + (b_{I1}/2) + (\frac{r}{2})$

Χρήση της επιλεξιμότητας και της εκτίμησης κόστους στη βελτιστοποίηση ερωτήσεων (8)

- Παραδείγματα συναρτήσεων κόστους για την πράξη της επιλογής (συνέχεια)

– E7. Συζευκτική επιλογή

- Χρήση της E1 ή μιας από τις μεθόδους E2 έως E6.
- Στη τελευταία περίπτωση χρησιμοποιούμε μια συνθήκη για την ανάκτηση των εγγραφών και στη συνέχεια ελέγχουμε αν η εγγραφή που προσκομίσθηκε στη μνήμη, ικανοποιεί τις υπόλοιπες συνθήκες της σύζευξης.

– E8. Συζευκτική επιλογή με χρήση σύνθετου ευρετηρίου

- Παρομοίως με τις E3α, E5 και E6α, ανάλογα με τον τύπο του ευρετηρίου

Χρήση της επιλεξιμότητας και της εκτίμησης κόστους στη βελτιστοποίηση ερωτήσεων (9)

- Χρειαζόμαστε μια εκτίμηση του μεγέθους του αρχείου που προκύπτει ως αποτέλεσμα της πράξης συνένωσης
- Παραδείγματα χρήσης των συναρτήσεων κόστους για συνένωση
 - Επιλεξιμότητα συνένωσης (js)
 - $js = \frac{|R \bowtie_c S|}{|R \times S|} = \frac{|R \bowtie_c S|}{||R| * |S|}$
 - Αν δεν υπάρχει συνθήκη συνένωσης c , τότε $js = 1$;
 - Αν η συνθήκη συνένωσης δεν ικανοποιείται από καμία πλειάδα των σχέσεων, τότε $js = 0$;
 - Γενικά, $0 \leq js \leq 1$;
 - Μέγεθος του αρχείου του αποτελέσματος της συνένωσης
 - $|R \bowtie_c S| = js * |R| * |S|$

Χρήση της επιλεξιμότητας και της εκτίμησης κόστους στη βελτιστοποίηση ερωτήσεων (10)

- Παραδείγματα χρήσης των συναρτήσεων κόστους για συνένωση (συνέχεια)
 - **Σ1. Εμφωλευμένοι βρόχοι**
 - Χρήση της σχέσης R για τον εξωτερικό βρόχο:
 - $C_{\Sigma 1} = b_R + (b_R * b_s) + (\frac{js * |R| * |S|}{bfr_{RS}})$
 - **Σ2. Συνένωση με απλό βρόχο επανάληψης (χρήση μιας δομής προσπέλασης για ανάκτηση των εγγραφών που ικανοποιούν τη συνθήκη)**
 - Αν υπάρχει ευρετήριο για το γνώρισμα συνένωσης B της σχέσης S, με χ_B επίπεδα ευρετηρίου, μπορούμε να ανακτήσουμε κάθε εγγραφή της σχέσης R και στη συνέχεια να χρησιμοποιήσουμε το ευρετήριο για την ανάκτηση των εγγραφών t της σχέσης S που ικανοποιούν τη $t[B] = s[A]$.
 - Το κόστος εξαρτάται από το είδος του ευρετηρίου.

Χρήση της επιλεξιμότητας και της εκτίμησης κόστους στη βελτιστοποίηση ερωτήσεων (11)

- Παραδείγματα χρήσης των συναρτήσεων κόστους για συνένωση (συνέχεια)
 - **Σ2. Συνένωση με απλό βρόχο επανάληψης (συνέχεια)**
 - Για ένα δευτερεύον ευρετήριο:
 - $C_{\Sigma 2\alpha} = b_R + (|R| * (x_B + s_B)) + (\frac{js * |R| * |S|}{bfr_{RS}})$
 - Για ένα ευρετήριο συστάδων:
 - $C_{\Sigma 2\beta} = b_R + \left(|R| * (x_B + \frac{s_B}{bfr_B})\right) + (\frac{js * |R| * |S|}{bfr_{RS}})$
 - Για ένα πρωτεύον ευρετήριο:
 - $C_{\Sigma 2\gamma} = b_R + (|R| * (x_B + 1)) + (\frac{js * |R| * |S|}{bfr_{RS}})$
 - Για ένα κλειδί κατακερματισμού σε ένα από τα δύο γνωρίσματα συνένωσης – B ή R
 - $C_{\Sigma 2\delta} = b_R + (|R| * h) + (\frac{js * |R| * |S|}{bfr_{RS}})$
 - Όπου $h \geq 1$, ο μέσος αριθμός προσπελάσεων μπλοκ που απαιτούνται για ανάκτηση μιας εγγραφής, όταν δίνεται η τιμή του κλειδιού κατακερματισμού για αυτή την τιμή

Χρήση της επιλεξιμότητας και της εκτίμησης κόστους στη βελτιστοποίηση ερωτήσεων (12)

- Παραδείγματα χρήσης των συναρτήσεων κόστους για συνένωση (συνέχεια)

– Σ3. Συνένωση ταξινόμησης- συγχώνευσης

- $C_{\Sigma 3\alpha} = C_S + b_R + b_S + \left(\frac{js^*|R|^*|S|}{bfr_{RS}}\right)$

- C_S : κόστος ταξινόμησης.

Χρήση της επιλεξιμότητας και της εκτίμησης κόστους στη βελτιστοποίηση ερωτήσεων (13)

- **Ερωτήσεις πολλαπλών σχέσεων και διάταξη συνενώσεων**
 - Μία ερώτηση με συνενώσεις η σχέσεων θα έχει $n-1$ πράξεις συνένωσης, και επομένως μπορεί να έχει ένα μεγάλο πλήθος από διαφορετικές διατάξεις συνενώσεων, όταν χρησιμοποιούμε αλγεβρικούς μετασχηματισμούς.
 - Οι βελτιστοποιητές ερωτήσεων περιορίζουν τη δομή ενός δένδρου ερωτήσεων στα δένδρα σε βάθος προς τα αριστερά (ή σε βάθος προς τα δεξιά).
- **Βαθύ προς τα αριστερά δένδρο**
 - Ένα δυαδικό δένδρο, όπου το δεξί παιδί κάθε μη τερματικού κόμβου είναι πάντα μια βασική σχέση.
 - Προσφέρονται για σωλήνωση
 - Επιτρέπει στον βελτιστοποιητή να χρησιμοποιήσει όλα τα μονοπάτια προσπέλασης στη βασική σχέση (το δεξί παιδί) για την εκτέλεση της συνένωσης.

Σημασιολογική βελτιστοποίηση ερωτήσεων

- Σημασιολογική βελτιστοποίηση ερωτήσεων
 - Χρησιμοποιεί περιορισμούς που ορίζονται στο σχήμα της βάσης δεδομένων, προκειμένου να μετασχηματίσει την ερώτηση σε μία άλλη, η οποία θα εκτελείται πιο αποδοτικά
- Θεωρήστε την ερώτηση SQL:
SELECT E.ΕΠΙΘΕΤΟ, Μ.ΕΠΙΘΕΤΟ
FROM ΕΡΓΑΖΟΜΕΝΟΣ **AS** Ε, ΕΡΓΑΖΟΜΕΝΟΣ **AS** Μ
WHERE Ε.ΠΡΟΪΣΤΑΜΕΝΟΣ=Μ.ΑΡ_ΤΑΥΤ **AND**
Ε.ΜΙΣΘΟΣ>Μ.ΜΙΣΘΟΣ;
- Ας υποθέσουμε ότι στο σχήμα της βάσης δεδομένων είχαμε ορίσει τον περιορισμό ότι ένας εργαζόμενος δεν μπορεί να κερδίσει περισσότερα από τον άμεσο προϊστάμενό του.
- Αν ο σημασιολογικός βελτιστοποιητής ελέγξει για την ύπαρξη αυτού του περιορισμού, δεν χρειάζεται να εκτελέσει αυτή την ερώτηση, διότι γνωρίζει ότι το αποτέλεσμα της είναι το κενό σύνολο.

Σύνοψη κεφαλαίου

- Συνδυασμός Πράξεων με χρήση σωλήνωσης
- Χρήση ευριστικών μεθόδων στη βελτιστοποίηση ερωτήσεων
- Χρήση επιλεξιμότητας και εκτιμήσεων κόστους στη βελτιστοποίηση ερωτήσεων
- Σημασιολογική βελτιστοποίηση ερωτήσεων