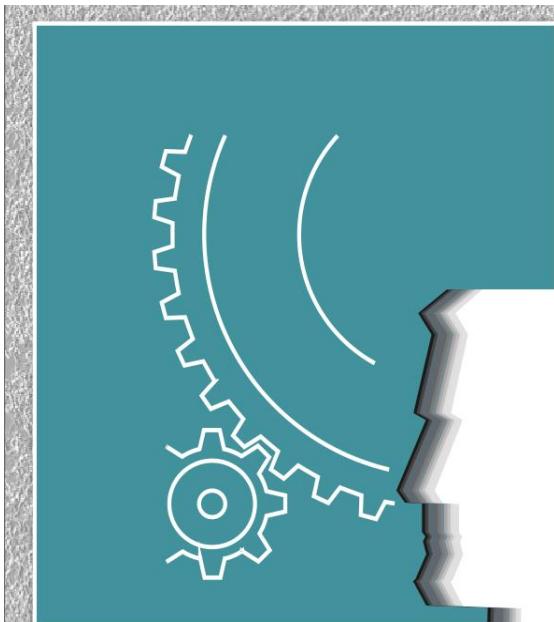


ΣΗΜΕΙΩΣΕΙΣ ΣΤΙΣ ΒΑΣΕΙΣ ΔΕΔΟΜΕΝΩΝ



ανάλυση
c o m p u t e r

ΕΡΓΑΣΤΗΡΙΟ ΕΛΕΥΘΕΡΩΝ ΣΠΟΥΔΩΝ
ΣΤΗΝ ΕΦΑΡΜΟΣΜΕΝΗ ΠΛΗΡΟΦΟΡΙΚΗ

...λυση
σπουδων

ΑΓ. ΑΝΔΡΕΟΥ 130-132 & ΓΟΥΝΑΡΗ, Τ.Κ. 26222, ΠΑΤΡΑ

ΤΗΛ: 310-097, 324-900 - ΤΗΛ/FAX: 313-547

E-MAIL: janelisj@otenet.gr, janelisk@otenet.gr

Computer – Ανάλυση

Βάσεις Δεδομένων- Computer Ανάλυση
Περιεχόμενα

1 ΘΕΜΑΤΑ ΜΕ ΔΙΑΓΡΑΜΜΑ ΟΝΤΟΤΗΤΩΝ ΣΥΣΧΕΤΙΣΕΩΝ ΚΑΙ ΣΧΕΣΙΑΚΟ ΜΟΝΤΕΛΟ.....	5
1.1 Θέμα 1 Σεπτέμβριος 2009	5
1.2 Θέμα 1 Φεβρουάριος 2013.....	7
1.3 Θέμα 1 Ιανουάριος 2014	9
1.4 Λυμένο Παράδειγμα 1.....	10
1.5 Λυμένο Παράδειγμα 2.....	11
1.6 Θέμα 1 Ιούλιος 2016.....	14
1.7 Θέμα 1 Φεβρουάριος 2017.....	16
1.8 Θέμα 1 Νοέμβριος 2018-Εργαστήριο	19
2 ΘΕΜΑΤΑ ΜΕ ΣΧΕΣΙΑΚΗ ΑΛΓΕΒΡΑ, QBE και SQL	28
2.1 Θέμα 3 Φεβρουάριος 2009.....	28
2.2 Θέμα 2 Φεβρουάριος 2013.....	31
2.3 Θέμα 5 Φεβρουάριος 2008.....	32
2.4 Θέμα 2 Φεβρουάριος 2010.....	35
2.5 Θέμα 4 Φεβρουάριος 2008.....	36
2.6 Θέμα 3 Ιανουάριος 2014	41
2.7 Θέμα 3 Ιούνιος 2014	44
2.8 Θέμα 2 Φεβρουάριος 2015.....	46
2.9 Θέμα 2 Ιούλιος 2015	48
2.10 Θέμα 3 Ιούλιος 2016	52
2.11 Θέμα 6 Σεπτέμβριος 2009	53
2.12 Θέμα 2 Σεπτέμβριος 2017	54
2.13 Θέμα 2 Ιούνιος 2017	57
2.14 Θέμα 2 Φεβρουάριος 2018.....	Σφάλμα! Δεν έχει οριστεί σελιδοδείκτης.
3 ΘΕΜΑΤΑ ΜΕ BTREES.....	66
3.1 Παραδείγματα Εισαγωγών και Διαγραφών σε BTree.....	66
3.2 Θέμα 7 Φεβρουάριος 2008.....	68
3.3 Θέμα 5 Φεβρουάριος 2012.....	69
3.4 Θέμα 8 Φεβρουάριος 2013.....	71
3.5 Θέμα 5 Φεβρουάριος 2013.....	73
3.6 Θέμα 5 Φεβρουάριος 2015 και Θέμα 5 Σεπτέμβριος 2017	75
4 ΘΕΜΑΤΑ ΜΕ ΕΥΡΕΤΗΡΙΟ	77
4.1 Υποδειγματική Άσκηση με Ευρετήριο.....	77
4.2 Θέμα 4 Φεβρουάριος 2009.....	84
4.3 ΘΕΜΑ 6 Φεβρουάριος 2008	86
4.4 ΘΕΜΑ 6 Φεβρουάριος 2010	87
4.5 Θέμα 5 –Σεπτέμβριος 2009	88
4.6 Θέμα 6 –Φεβρουάριος 2013	90
4.7 Θέμα 6 –Φεβρουάριος 2014	92
4.8 Θέμα 6 –Ιούνιος 2014	93
4.9 Θέμα 4 Φεβρουάριος 2015	95

5 ΣΥΝΑΡΤΗΣΙΑΚΕΣ ΕΞΑΡΤΗΣΕΙΣ, ΚΑΝΟΝΙΚΕΣ ΜΟΡΦΕΣ, ΕΛΑΧΙΣΤΗ ΚΑΛΥΨΗ, ΔΙΑΤΗΡΗΣΗ ΕΞΑΡΤΗΣΕΩΝ Κ.Λ.Π.	103
5.1 Συναρτησιακές Εξαρτήσεις (<i>Functional Dependencies-FDs</i>).....	103
5.2 Αξιώματα Armstrong.....	103
5.3 Κανόνες που προκύπτουν από τα Αξιώματα Armstrong	103
5.4 Κλειστότητα Συνόλου Συναρτησιακών Εξαρτήσεων F^+	105
5.5 Αλγόριθμος Υπολογισμού Κλειστότητας Συνόλου Συναρτησιακών Εξαρτήσεων.....	105
5.5.1 Παράδειγμα Υπολογισμού Κλειστότητας Συνόλου Συναρτησιακών Εξαρτήσεων.....	105
5.6 Κλειστότητα Γνωρίσματος A^+	106
5.7 Τετριμένες Εξαρτήσεις.....	106
5.8 Αλγόριθμος Υπολογισμού Κλειστότητας Γνωρίσματος X^+	106
5.8.1 Παράδειγμα 1 Υπολογισμού Κλειστότητας Γνωρισμάτων	107
5.8.2 Παράδειγμα 2 Υπολογισμού Κλειστότητας Γνωρισμάτων	108
5.9 Κανονική Κάλυψη του F	108
5.10 Ελάχιστη Κάλυψη του F	108
5.11 Αλγόριθμος Εύρεσης Κανονικής Κάλυψης.....	109
5.11.1 Πλεονάζοντα Γνωρίσματα.....	109
5.11.2 Παράδειγμα 1 Εύρεσης Ελάχιστης/Κανονικής Κάλυψης του συνόλου F	110
5.11.3 Παράδειγμα 2 Εύρεσης Ελάχιστης/Κανονικής Κάλυψης του F	112
5.11.4 Παράδειγμα 3 Εύρεσης Ελάχιστης/Κανονικής Κάλυψης του F	112
5.12 Κανονικές Μορφές	113
5.12.1 Παράδειγμα Κανονικοποίησεων.....	113
5.13 Διαχωρισμός (Αποσύνθεση) Χωρίς Απώλειες Συνδέσμου	118
5.13.1 Παράδειγμα 1 με Διαχωρισμό Χωρίς Απώλειες Συνδέσμου	118
5.13.2 Παράδειγμα 2 με Διαχωρισμό Χωρίς Απώλειες Συνδέσμου	118
5.14 Διατήρηση Εξαρτήσεων.....	119
5.14.1 Παράδειγμα με Διατήρηση Εξαρτήσεων.....	119
5.15 ΘΕΜΑΤΑ ΜΕ ΚΑΝΟΝΙΚΕΣ ΜΟΡΦΕΣ, ΣΥΝΑΡΤΗΣΙΑΚΕΣ ΕΞΑΡΤΗΣΕΙΣ, ΕΛΑΧΙΣΤΗ ΚΑΛΥΨΗ Κ.Λ.Π.	120
5.15.1 Θέμα 4 Σεπτέμβριος 2009	120
5.15.2 Θέμα 4 Φεβρουάριος 2013	121
5.15.3 Θέμα 6 Σεπτέμβριος 2008	124
5.15.4 Θέμα 3 Ιανουάριος 2014	124
5.15.5 Θέμα 3 Φεβρουάριος 2015	126
5.15.6 Θέμα 4- Ιανουάριος 2016	127
5.15.7 Θέμα 3 Φεβρουάριος 2017	128
5.15.8 Θέμα 2- Ιανουάριος 2014	129
5.15.9 Θέμα 2- Ιανουάριος 2016	130
5.15.10 Θέμα 2- Ιανουάριος 2013	130
5.15.11 Θέμα 2- Ιανουάριος 2018	131
6 R-δέντρα (R-trees).....	136
6.1 Χαρακτηριστικά R-Tree	136
6.2 Παράδειγμα 1 σε R-Tree.....	136
6.3 Παράδειγμα 2 σε R-Tree.....	137
6.4 Παράδειγμα 3 σε R-Tree.....	137
6.5 Αναζήτηση στο R-Tree.....	137
6.6 Εισαγωγή σε Tree – Επιλογή Φύλλου	138
7 ΘΕΜΑΤΑ ΘΕΩΡΙΑΣ.....	141
7.1 Τι γνωρίζετε για τους περιορισμούς ακεραιότητας;	141
7.2 Ποια τα είδη των κλειδιών;	141
7.3 Τι ονομάζεται Stored-Procedure και τι ονομάζεται Trigger- Θέμα 3 Σεπτέμβριος 2016	141

Βάσεις Δεδομένων- Computer Ανάλυση

7.4	Θέμα 3 Φεβρουάριος 2010.....	142
7.5	Θέμα 3 Φεβρουάριος 2013.....	143
7.6	Θέμα 3 Φεβρουάριος 2016.....	144

Computer Ανάλυση

1 ΘΕΜΑΤΑ ΜΕ ΔΙΑΓΡΑΜΜΑ ΟΝΤΟΤΗΤΩΝ ΣΥΣΧΕΤΙΣΕΩΝ ΚΑΙ ΣΧΕΣΙΑΚΟ ΜΟΝΤΕΛΟ

1.1 Θέμα 1 Σεπτέμβριος 2009

Θεωρείστε μια βάση δεδομένων η οποία θα διατηρεί πληροφορίες για τα δρομολόγια των υπεραστικών λεωφορείων ενός νομού της Ελλάδας. Συγκεκριμένα θα πρέπει να καταγράφεται ποιος είναι ο οδηγός του δρομολογίου (Όνομα, Επίθετο, AT, Ημερομηνία γέννησης, Ημερομηνία Πρόσληψης), πότε πραγματοποιήθηκε το δρομολόγιο (Ημερομηνία και Ωρα) και μεταξύ ποιων πόλεων. Επιπλέον για κάθε οδηγό θα πρέπει να δηλώνεται τουλάχιστον ένα τηλέφωνο επικοινωνίας, ενώ μια πόλη δεν μπορεί να προσδιορίζεται μόνο από το όνομα της καθώς μπορεί να υπάρχουν δύο πόλεις με το ίδιο όνομα. Τέλος υποθέστε ότι κάθε δρομολόγιο ξεκινάει από μια πόλη και καταλήγει σε μια άλλη χωρίς ενδιάμεσες στάσεις.

(i) Σχεδιάστε το διάγραμμα E-R για τη βάση που θα προκύψει με βάση την παραπάνω περιγραφή και διατυπώστε γραπτώς τις υπόθεσις που κάνετε

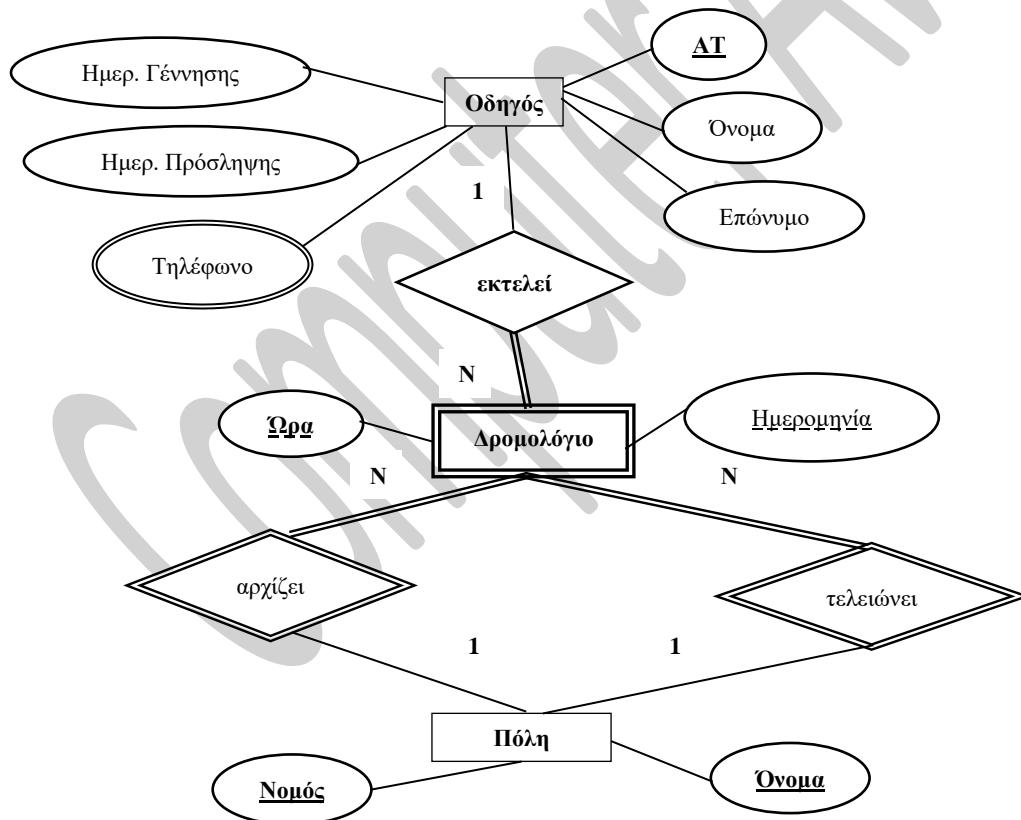
(ii) Μετατρέψτε το E-R διάγραμμα που σχεδιάσατε σε σχεσιακό όπου θα φαίνονται τα κλειδιά (ξένα, πρωτεύοντα κ.λ.π.) της κάθε σχέσης

(iii) Τροποποιήστε το E-R διάγραμμα που προέκυψε από το προηγούμενο ερώτημα ώστε στη βάση δεδομένων να είναι δυνατό να αποθηκευτεί και το πλήθος των επιβατών που επιβαίνουν στο συγκεκριμένο δρομολόγιο

Σημ. Να μη χρησιμοποιηθεί καμιά σχέση βαθμού μεγαλύτερου από 2

Απάντηση

(i) Διάγραμμα ER



ii) Σχετικό Μοντέλο

Οδηγός

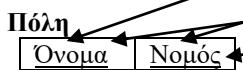
AT	Όνομα	Επώνυμο	Ημ. Γέννησης	Ημ. Πρόσληψης
----	-------	---------	--------------	---------------

Τηλέφωνα

Τηλέφωνο	AT
----------	----

Δρομολόγιο

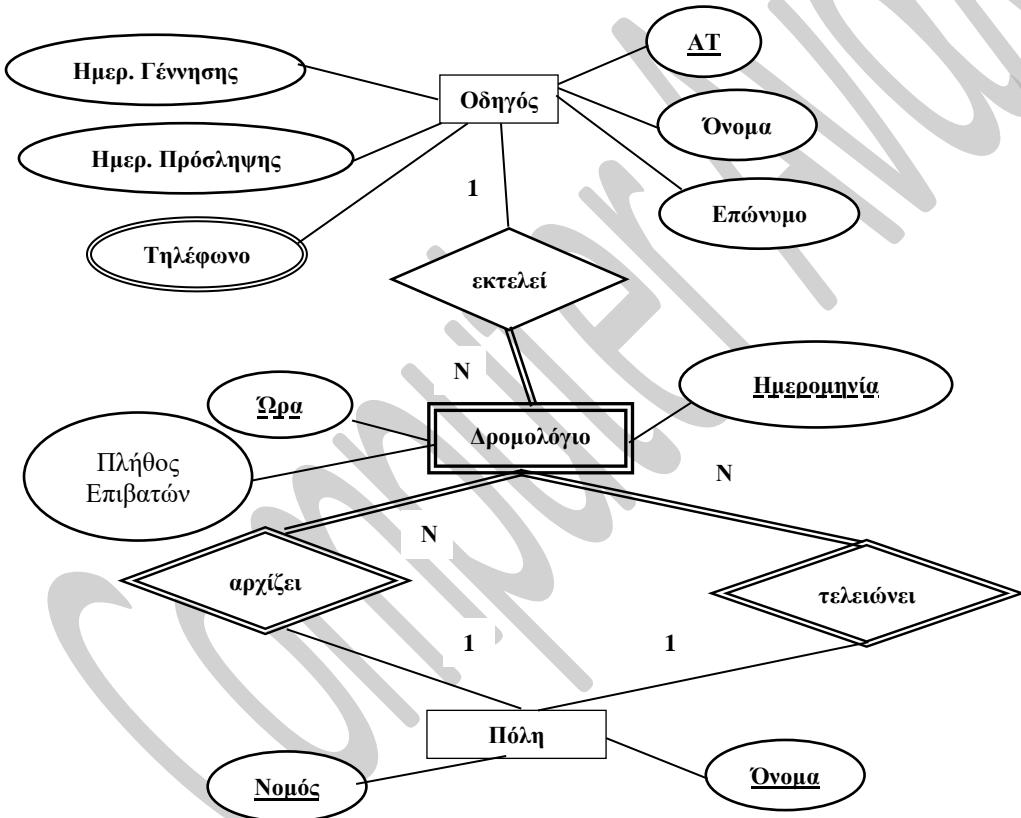
Ημερομηνία	Ωρα	AT	Όνομα Αφετηρίας	Νομός Αφετηρίας	Όνομα Προορισμού	Νομός Προορισμού
------------	-----	----	-----------------	-----------------	------------------	------------------



iii) Τροποποιημένο E-R

Απάντηση

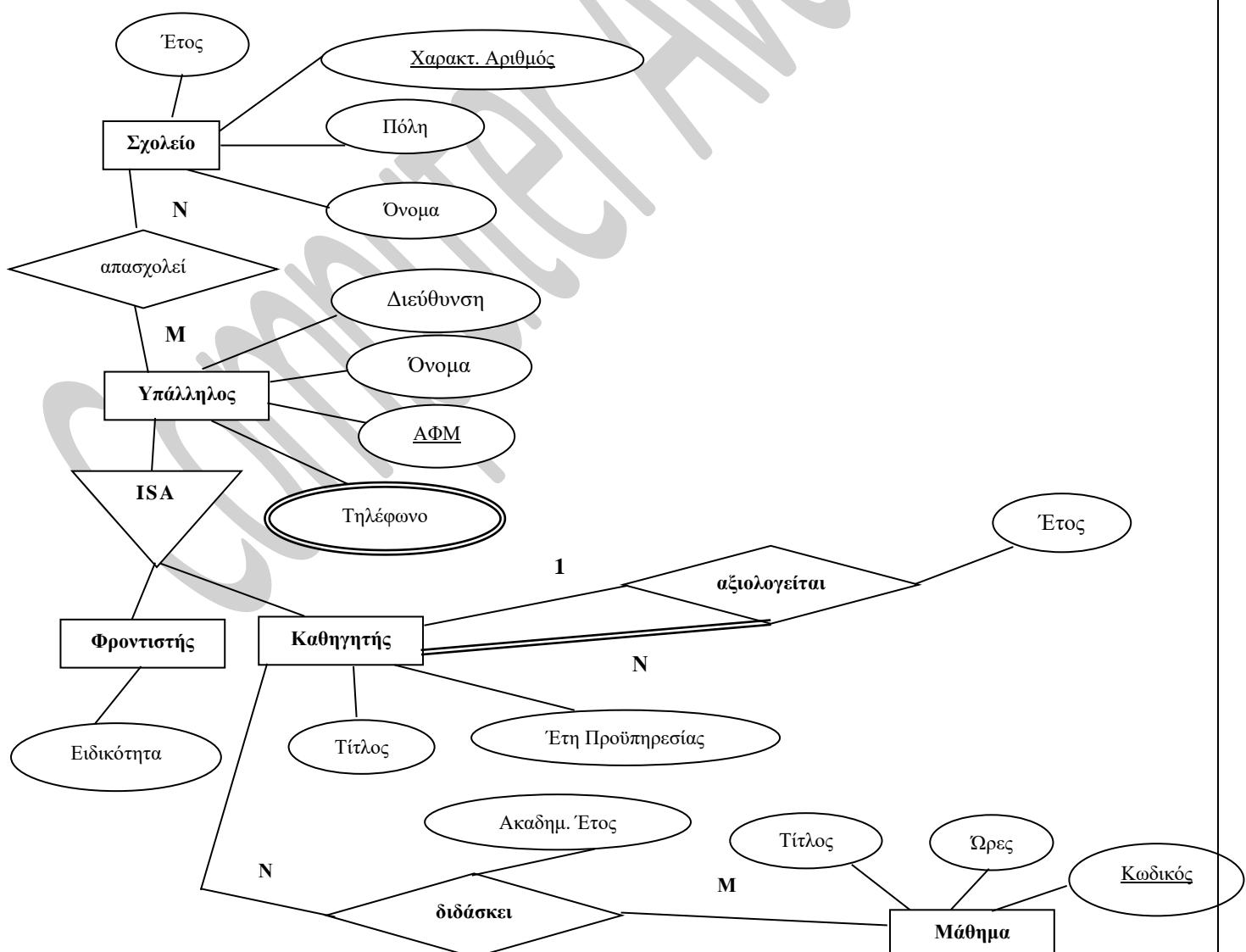
(i) Διάγραμμα ER



1.2 Θέμα 1 Φεβρουάριος 2013

Το Υπουργείο Παιδείας επιθυμεί να δημιουργήσει μια βάση με τις σχολικές μονάδες της Ελλάδας. Κάθε σχολείο έχει έναν χαρακτηριστικό αριθμό, ενώ επιθυμεί να διατηρεί πληροφορία για την πόλη στην οποία βρίσκεται και το όνομα του σχολείου. Σε κάθε σχολική μονάδα εργάζονται διάφοροι υπάλληλοι. Οι υπάλληλοι μπορεί να εργάζονται σε διαφορετική σχολική μονάδα ανά ακαδημαϊκό έτος. Για κάθε υπάλληλο διατηρείται πληροφορία σχετικά με το Α.Φ.Μ τους, το όνομα, την διεύθυνση κατοικίας τους και το/-α τηλέφωνο/-α τους. Οι υπάλληλοι μπορεί να είναι είτε φροντιστές της σχολικής μονάδας είτε καθηγητές. Οι φροντιστές μπορεί να είναι διαφόρων ειδικοτήτων όπως καθαρίστριες, κηπουροί, υδραυλικοί κτλ. Κάθε καθηγητής διαθέτει έναν τίτλο πτυχίου και κάποια έτη προϋπηρεσίας, ενώ αξιολογείται από κάποιο σύμβουλο καθηγητή. Σε κάθε ακαδημαϊκό έτος ορίζονται ποιοι θα είναι σύμβουλοι καθηγητές και ποιους καθηγητές θα αξιολογήσουν (ένας καθηγητής αξιολογείται μόνο από έναν σύμβουλο καθηγητή). Τέλος, κάθε καθηγητής διδάσκει κάθε ακαδημαϊκό έτος κάποιο/-α μαθήματα, για τα οποία επιθυμούμε να διατηρείται πληροφορία σχετικά με τον τίτλο του μαθήματος και τις ώρες διδασκαλίας (κάθε μάθημα έχει μοναδικό κωδικό).

- Σχεδιάστε το ER διάγραμμα σύμφωνα με τις προδιαγραφές που ορίζονται παραπάνω.
- Δημιουργήστε το αντίστοιχο σχεσιακό, με βάση το ER που σχεδιάσατε στο προηγούμενο ερώτημα.

Απάντηση**Διάγραμμα E-R**

Σχεσιακό Σχήμα

Σχολείο

<u>Χαρακτ. Αριθμός</u>	Πόλη	Όνομα
------------------------	------	-------

Απασχόληση

<u>Χαρακτ. Αριθμός</u>	<u>ΑΦΜ</u>	Έτος
------------------------	------------	------

Υπάλληλος

<u>ΑΦΜ</u>	Διεύθυνση	Όνομα
------------	-----------	-------

Τηλέφωνα

<u>Τηλέφωνο</u>	<u>ΑΦΜ Υπαλλήλου</u>
-----------------	----------------------

Φροντιστής

<u>ΑΦΜ</u>	Ειδικότητα
------------	------------

Καθηγητής

<u>ΑΦΜ</u>	Τίτλος	Έτη Προϋπηρεσίας	ΑΦΜ Αξιολογητή
------------	--------	------------------	----------------

Διδασκαλία

<u>Κωδικός</u>	<u>ΑΦΜ</u>	Ακαδημαϊκό Έτος
----------------	------------	-----------------

Μάθημα

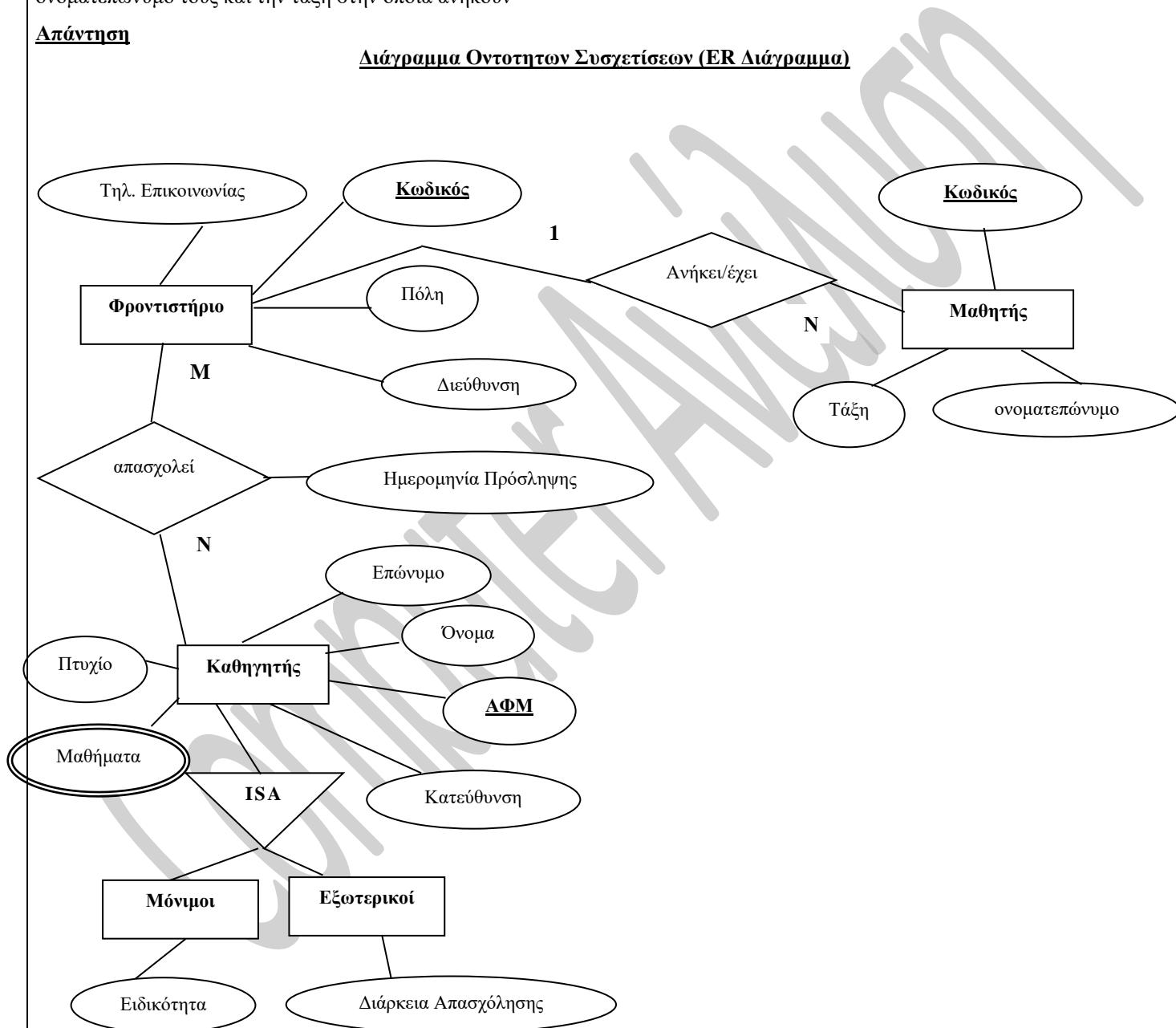
<u>Κωδικός</u>	Τίτλος	Ωρες
----------------	--------	------

1.3 Θέμα 1 Ιανουάριος 2014

Κατασκευάστε μια βάση δεδομένων η οποία θα διατηρεί πληροφορίες για ένα σύνολο φροντιστηρίων δευτεροβάθμιας εκπαίδευσης. Οι πληροφορίες που θέλουμε να παρακολουθήσουμε περιλαμβάνουν ένα μοναδικό κωδικό, την πόλη, τη διεύθυνση και ένα τηλέφωνο επικοινωνίας κάθε φροντιστηρίου. Στο κάθε φροντιστήριο ανήκει ένα σύνολο καθηγητών για τους οποίους θέλουμε να γνωρίζουμε το ΑΦΜ τους, το όνομα και το επώνυμο τους, το πτυχίο το οποίο έχουν, την κατεύθυνση στην οποία διδάσκουν και τα μαθήματα που διδάσκουν. Κάθε καθηγητής μπορεί να διδάσκει σε ένα ή περισσότερα φροντιστήρια της αλυσίδας ενώ επίσης θέλουμε να κρατάμε την ημερομηνία που ο κάθε καθηγητής προσλήφθηκε στο φροντιστήριο. Κάποιοι από τους καθηγητές είναι εξωτερικοί (εξ-κτακτοί) συνεργάτες και κάποιοι άλλοι είναι μόνιμοι. Για τους εξωτερικούς συνεργάτες θέλουμε επίσης να γνωρίζουμε τη διάρκεια απασχόλησης τους. Στο φροντιστήριο είναι γραμμένοι μαθητές για τους οποίους θέλουμε να διατηρούμε ένα μοναδικό κωδικό, το ονοματεπώνυμο τους και την τάξη στην οποία ανήκουν

Απάντηση

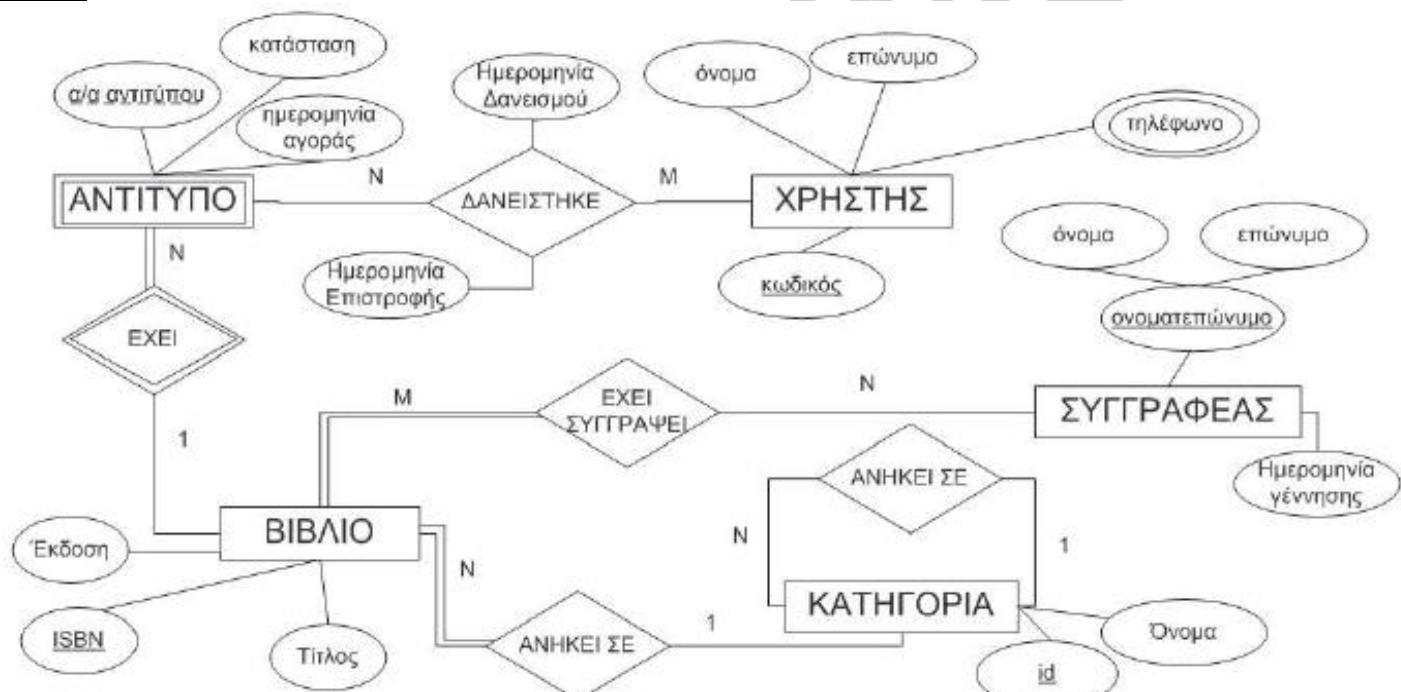
Διάγραμμα Οντοτητών Συσχετίσεων (ER Διάγραμμα)



1.4 Λυμένο Παράδειγμα 1

- Κείμενο Προδιαγραφών:
 - Υλοποίηση μιας ΒΔ για τον κατάλογο μιας Βιβλιοθήκης.
 - Η βιβλιοθήκη διατηρεί πληροφορίες για τα βιβλία που έχει στην κατοχή της καθώς επίσης και για συγγραφείς. Κάθε βιβλίο πρέπει να έχει γραφτεί από κάποιον συγγραφέα του οποίου τα στοιχεία διατηρεί η βιβλιοθήκη.
 - Η βιβλιοθήκη διατηρεί σύστημα ιεραρχικής κατηγοριοποίησης των βιβλίων της. Κάθε κατηγορία μπορεί να είναι υποκατηγορία μιας άλλης κτλ. Κάθε βιβλίο πρέπει να ανήκει σε κάποια κατηγορία.
 - Κάθε βιβλίο έχει μοναδικό ISBN, η βιβλιοθήκη διατηρεί όμως αντίτυπα από κάθε βιβλίο, τα οποία δανείζει σε χρήστες.
 - Οι χρήστες ανοίγουν λογαριασμό στη βιβλιοθήκη και λαμβάνουν έναν μοναδικό κωδικό.
 - Οι χρήστες μπορούν να δανειστούν βιβλία.

Απάντηση



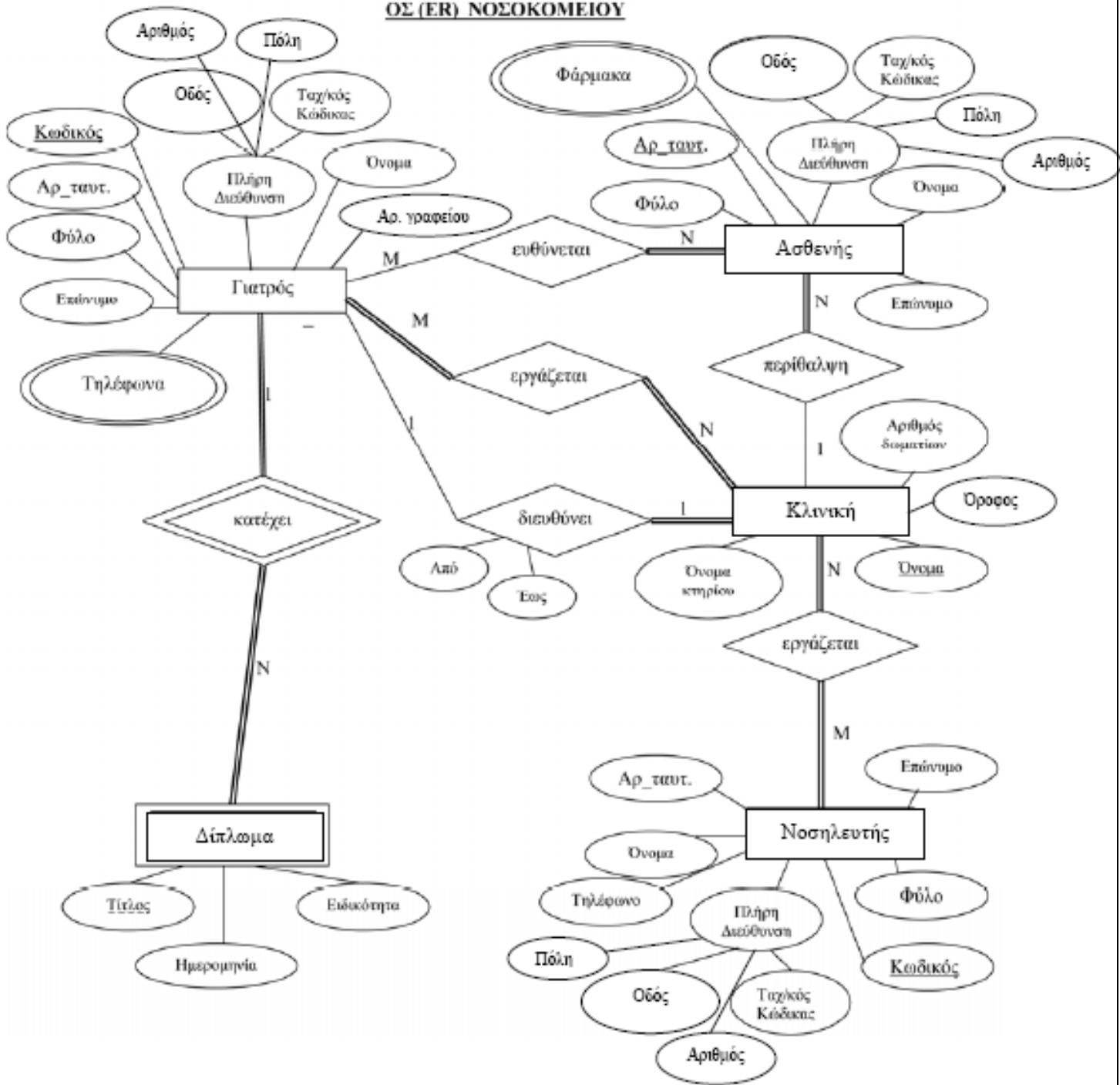
1.5 Λυμένο Παράδειγμα 2

Θέμα 1

Με βάση το κείμενο απαιτήσεων που ακολουθεί σχεδιάστε το ER διάγραμμα της βάσης δεδομένων ενός νοσοκομείου, καταγράφοντας τις παραδοχές που χρειάστηκε να γίνουν ώστε να καταλήξετε στη συγκεκριμένη σχεδίαση Στο νοσοκομείο υπάρχουν κλινικές, η καθεμιά με μοναδικό όνομα, με έναν αριθμό (πλήθος) δωματίων, στεγάζονται σε κάποιο κτίριο (θα καταγράφουμε το όνομα του κτιρίου και τον όροφο). Σε κάθε κλινική εργάζονται πολλοί γιατροί, ενώ κάθε γιατρός μπορεί να εργάζεται σε περισσότερες από μια κλινικές (σίγουρα όμως σε μία τουλάχιστον). Κάθε κλινική πρέπει να έχει διευθυντή έναν γιατρό και στο σύστημα θέλουμε επίσης να καταγράφουμε την ημερομηνία που κάποιος ξεκινάει τη θητεία του ως διευθυντής και την ημερομηνία που την ολοκληρώνει. Για τους γιατρούς η ΒΔ θα πρέπει να κρατάει στοιχεία για το όνομα και το επώνυμο, τον αριθμό ταυτότητας, το φύλο, την πλήρη διεύθυνση κατοικίας (οδός, αριθμός, πόλη και Τ.Κ.), τον αριθμό του γραφείου του στο νοσοκομείο, καθώς και ένα σύνολο τηλεφώνων επικοινωνίας. Κάθε γιατρός έχει έναν μοναδικό κωδικό στο σύστημα και επίσης πρέπει να κρατάμε πληροφορίες σχετικά με τα διπλώματά του (για κάθε δίπλωμα γνωρίζουμε τον τίτλο, την ημερομηνία απόκτησης και την ειδικότητα αλλά όπως μας πληροφόρησαν, είναι πιθανό δύο ή και περισσότεροι γιατροί να έχουν κάποιο δίπλωμα με κοινό τίτλο, ημερομηνία και ειδικότητα). Κάθε γιατρός είναι υπεύθυνος για έναν ή περισσότερους ασθενείς, αλλά ενδεχομένως να υπάρχουν περίοδοι που δεν έχει κάποιον ασθενή. Για τον ασθενή τα στοιχεία που χρειάζεται να αποθηκεύουμε στο σύστημα είναι το όνομα και το επώνυμο, τον αριθμό ταυτότητας, το φύλο, την πλήρη διεύθυνση κατοικίας (οδός, αριθμός, πόλη και Τ.Κ.), καθώς και τα ονόματα των φαρμάκων που παίρνει. Κάθε ασθενής πρέπει να είναι στην ευθύνη ενός ή και περισσότερων γιατρών και να περιθάλπεται σε μια κλινική του νοσοκομείου (γνωρίζουμε ότι κάθε νοσηλευτής μπορεί να εργάζεται σε πολλές κλινικές). Για τους νοσηλευτές θα πρέπει να καταγράφουμε πληροφορίες για το όνομα και το επώνυμο, τον αριθμό ταυτότητας, το φύλο, την πλήρη διεύθυνση κατοικίας (οδός, αριθμός, πόλη και Τ.Κ.), καθώς και ένα τηλέφωνο επικοινωνίας. Και για τους νοσηλευτές, το σύστημα αποδίδει έναν μοναδικό κωδικό σε κάθε έναν.

Απάντηση

Βάσεις Δεδομένων- Computer Ανάλυση

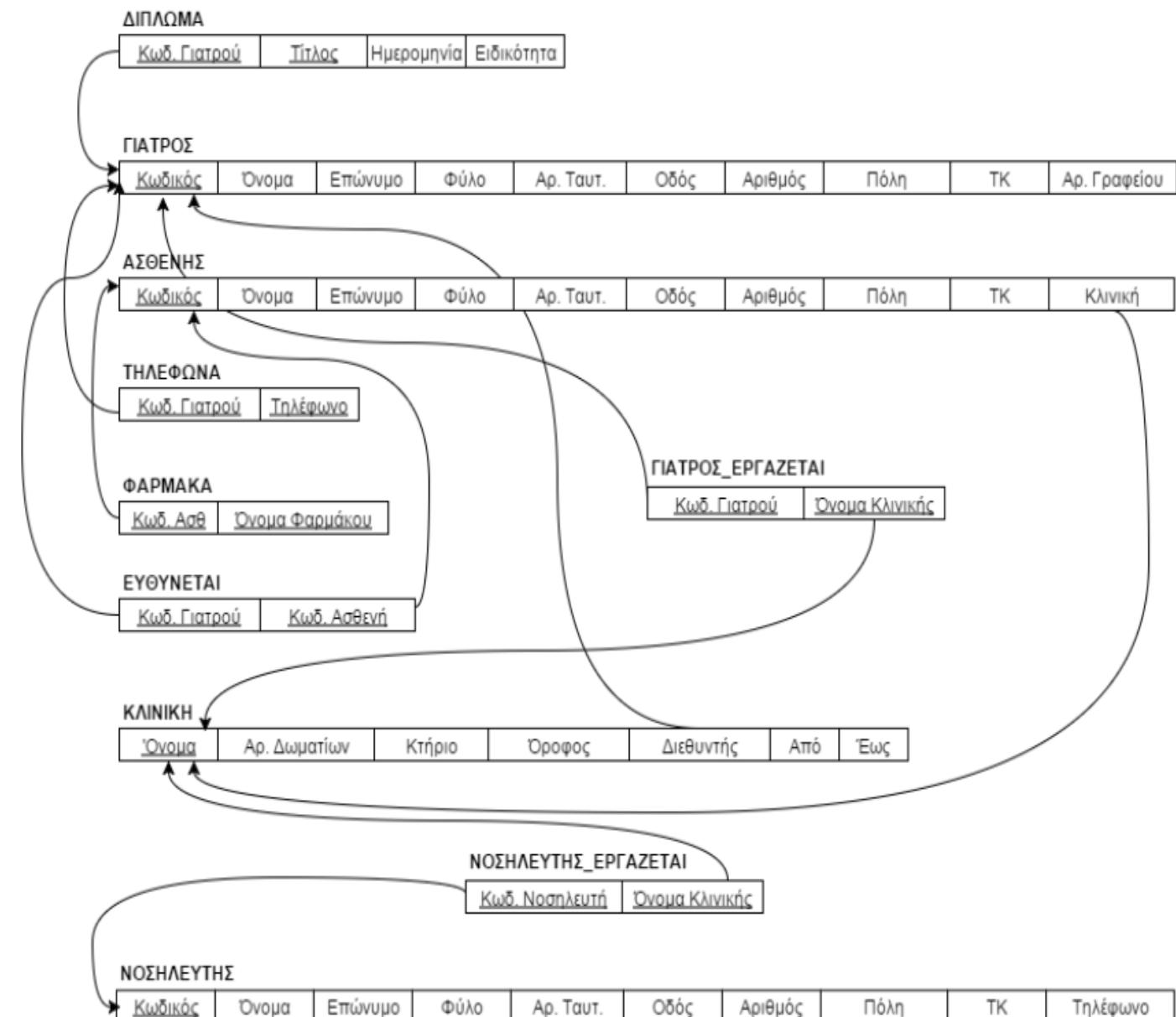


Βάσεις Δεδομένων- Computer Ανάλυση

Θέμα2

Με βάση το ER διάγραμμα του θέματος 1 κατασκευάστε το σχεσιακό διάγραμμα της ΒΔ του νοσοκομείου

Απάντηση:



1.6 Θέμα 1 Ιούλιος 2016

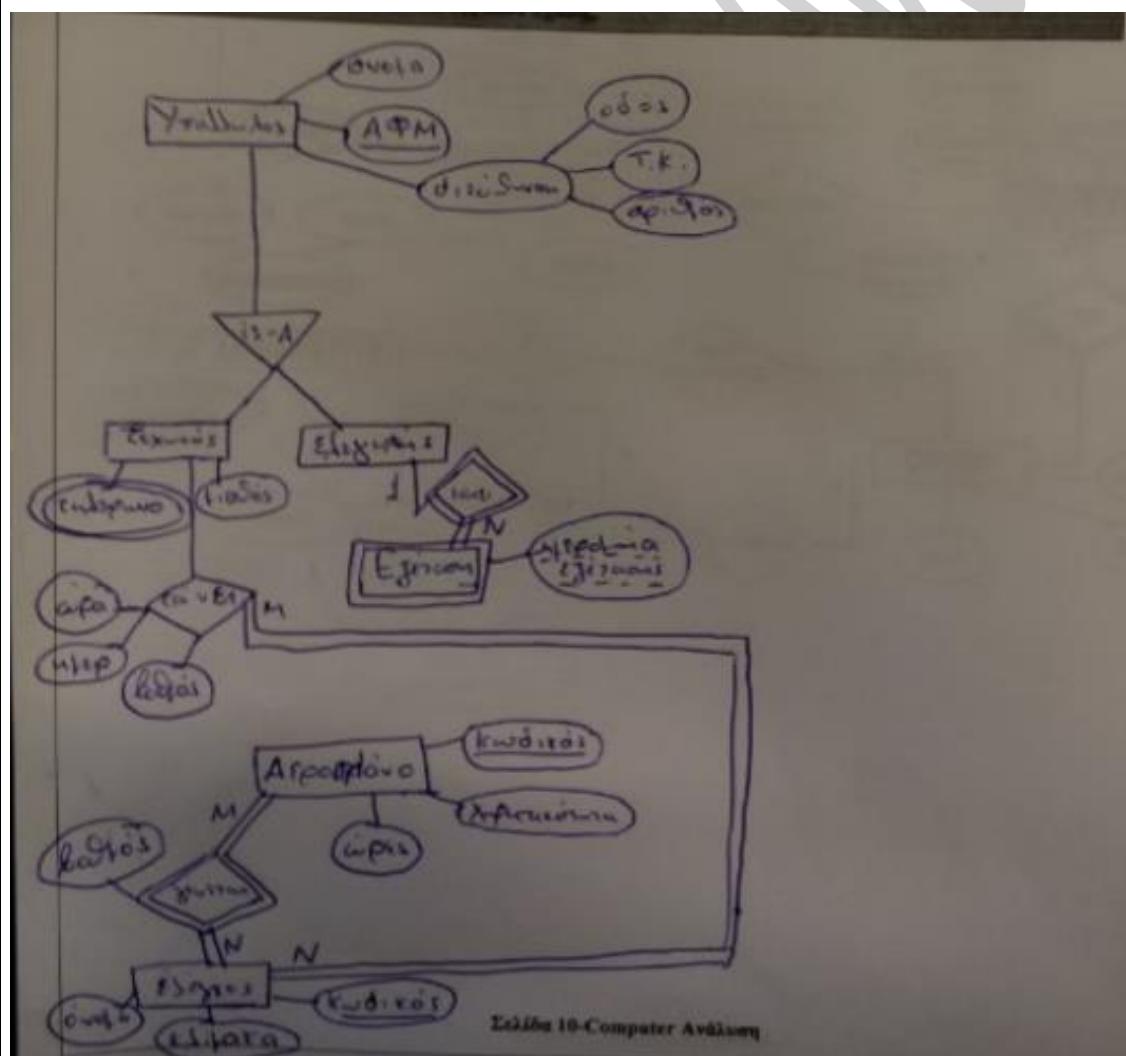
Θεωρήστε μία βάση δεδομένων η οποία διατηρεί πληροφορίες για ένα αεροδρόμιο. Στο αεροδρόμιο εργάζονται υπέλληπτοι, για τους οποίους διατηρείται πληροφορία σχετικά με το όνομά τους, το ΑΦΜ και την διεύθυνσή τους. Παρασυγκεκριμένα, για τους τεχνικούς δια πρέπει να αποθηκεύονται οι αριθμοί τηλεφώνων τους, τοπικά μοντέλο αεροπλάνων στο/α οποίο/α εξειδικεύονται καθώς και ο μισθός τους. Οι ελεγκτές εκείνων πρέπει να υποβάλλονται σε επήσεις απρικές εξετάσεις. Για κάθε δυνατή από αυτούς αποθηκεύονται την ημερομηνία της τελευταίας εξέτασης. Για κάθε αεροπλάνο δια πρέπει να αποθηκεύονται ο κωδικός του, ο αριθμός θέσων (χωρητικότητα) που διαθέτει και οι άρες πτήσεων που έχουν καταγραφεί. Στο αεροδρόμιο απαιτούνται περιοδικές έλεγχοι ώστε να εξασφαλίσουν ότι τα αεροπλάνα είναι σε κατάσταση να πετάξουν. Κάθε έλεγχος χαρακτηρίζεται από το όνομα του, τον κωδικό του και μια κλίμακα βαθμολόγησης. Για εναντίον δεδομένο έλεγχο ο οποίος πραγματοποιείται από έναν ή περισσότερους τεχνικούς δια πρέπει να διατηρείται η ημερομηνία διεξαγωγής του, οι άρες που διήρκεσε ο έλεγχος και ο βαθμός που έλαβε στο συγκεκριμένο τεστ το αεροπλάνο.

- (i) Σχεδιάστε ένα διάγραμμα E-R για την παρεπόμπη εφαρμογή και διατυπώστε γραπτώς τις υποθέσεις που κάνετε.

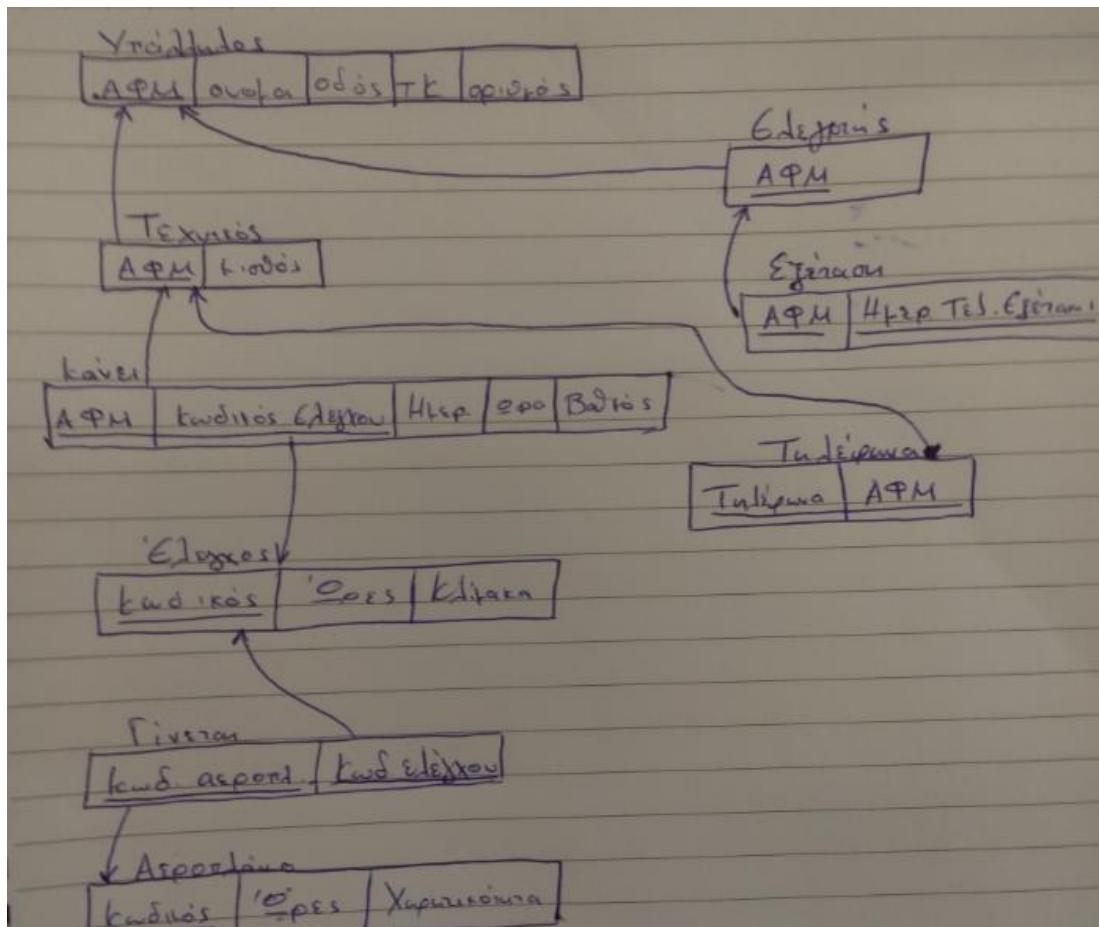
(ii) Μετατρέψτε το E-R διάγραμμα που σχεδιάσατε σε σχετικό όπου θα φαίνονται τα κλειδιά (ξένα, πρωτεύοντα, κτλ.) της κάθε σχέσης.

Απάντηση

Διάγραμμα Οντοτητών Συγχετίσεων (ΕΡ Διάγραμμα)



**Βάσεις Δεδομένων- Computer Ανάλυση
Σχεσιακό Μοντέλο**



1.7 Θέμα 1 Φεβρουάριος 2017 - Θέμα 1 Ιανουάριος 2019 - Θέμα 1-Ιανουάριος 2020

Αναλυτήμενομενα σχεδιάστουμε την βάση δεδομένων για την αποθήκευση πληροφοριών που αφορούν μια αλυσίδα γυμναστηρίων τα οποία βρίσκονται στην ίδια περιοχή. Παρακάτω περιγράφουμε τις οντότητες και τις σχέσεις που πρέπει να αποθηκευτούν στην βάση δεδομένων:

- Κάθε γυμναστήριο έχει όνομα (μοναδικό), διεύθυνση, ΤΚ κι έναν ή περισσότερους τηλεφωνικούς αριθμούς.
- Έχει εργαζόμενους που καθορίζονται μοναδικά από το ΑΦΜ τους. Επιπλέον αποθηκεύουμε το όνομα, το επώνυμο και το τηλέφωνο επικοινωνίας.
- Κάθε εργαζόμενος εργάζεται σε ένα ή και περισσότερα γυμναστήρια. Για παράδειγμα ένας εργαζόμενος ενδέχεται να εργάζεται το πρωί στο ένα και το απόγευμα σε άλλο. Μας ενδιαφέρει να γνωρίζουμε το ποσοστό του χρόνου που απασχολείται σε κάθε γυμναστήριο.
- Μερικοί εργαζόμενοι έχουν ιδιαίτερες ευθύνες, είτε ως υπεύθυνοι γυμναστηρίου, είτε ως υπεύθυνοι υποδοχής, είτε ως προσωπικοί γυμναστές.
- Ως υπεύθυνοι γυμναστηρίου μπορεί να παρακολουθούν άνω του ενός γυμναστηρίου. Κάθε γυμναστήριο έχει μοναδικό υπεύθυνο.
- Στους προσωπικούς γυμναστές αποθηκεύουμε τις πιστοποιήσεις τους (πχ πιστοποίηση αεροβικής).
- Για τους πελάτες του κάθε γυμναστηρίου, αποθηκεύουμε έναν μοναδικό αριθμό ID, το ονοματεπώνυμό τους, το τηλέφωνό τους, την ηλικία τους, τον αριθμό της κάρτας εισόδου, την ημερομηνία λήξης της συνδρομής τους και μια σύντομη καταγραφή προβλημάτων υγείας. Ένας πελάτης μπορεί να πηγαίνει σε άνω του ενός γυμναστηρίου της αλυσίδας.
- Κάθε πελάτης μπορεί να δηλώσει φίλους του που δεν αποτελούν πελάτες της αλυσίδας με σκοπό να λάβει έκπτωση επί της συνδρομής του. Σε κάθε φίλο αποθηκεύουμε το όνομα του και την ηλικία του, τα οποία ως συνδυασμός είναι μοναδικός κωδικός και το email του.

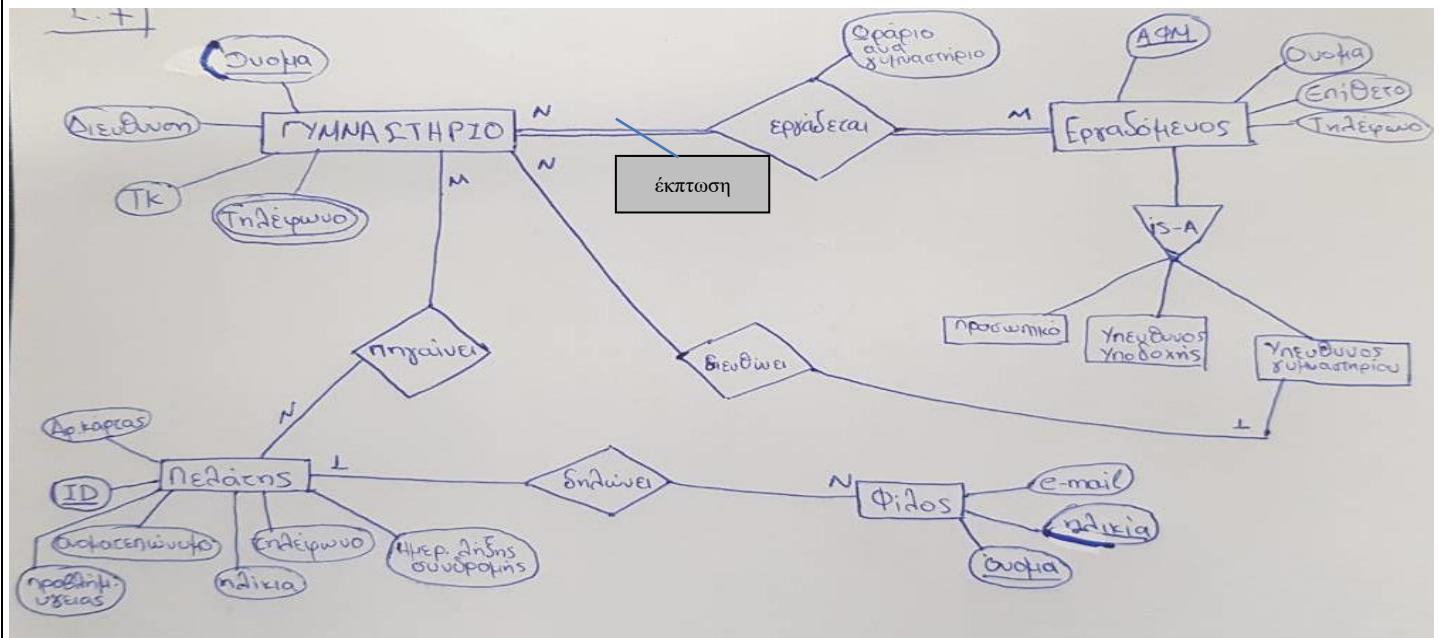
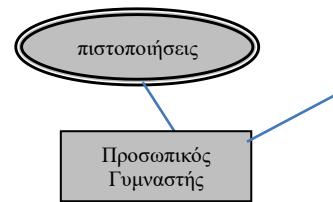
Πότε ένα σύνολο οντοτήτων έχει ολική συμμετοχή σε ένα σύνολο συσχετίσεων; Πως υποδηλώνεται σχεδιαστικά στο διάγραμμα E-R; Υπάρχει τέτοιο σύνολο οντοτήτων στην άνω ανάλυση κι αν ναι ποιο είναι. Αιτιολογείστε την απάντησή σας. (5 μονάδες)

Σχεδίαστε ένα διάγραμμα E-R για την παραπάνω εφαρμογή και διατυπώστε γραπτώς τις υποθέσεις που κάνετε (15 μονάδες)

Απάντηση

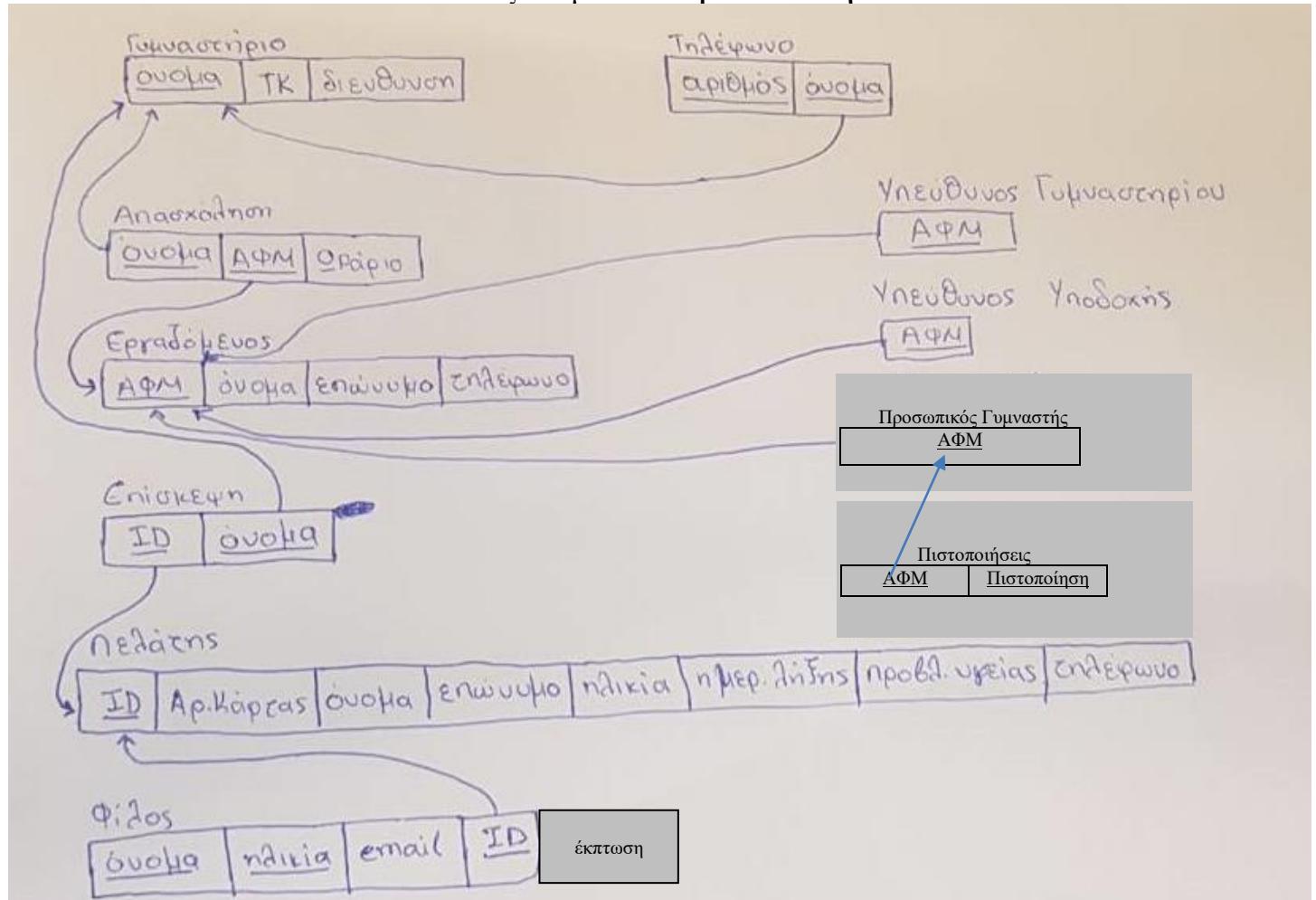
Ένα σύνολο οντοτήτων έχει ολική συμμετοχή σε ένα σύνολο συσχετίσεων όταν κάθε εγγραφή (πλειάδα) της οντότητας συμμετέχει υποχρεωτικά στη συσχέτιση αυτή. Αυτό υποδηλώνεται σχεδιαστικά με διπλή γραμμή από την οντότητα στη συσχέτιση. Στο ER που σχεδιάσαμε θεωρούμε ότι η οντότητα Εργαζόμενος έχει ολική συμμετοχή στη συσχέτιση Απασχολείται γιατί ένας εργαζόμενος πρέπει να εργάζεται σε γυμναστήριο αφού θεωρείται εργαζόμενος και κάθε γυμναστήριο σίγουρα απασχολεί κάποια ώτομα.

Διάγραμμα Οντοτητών Συσχετίσεων (ER Διάγραμμα)



Σχεσιακό Μοντέλο

Βάσεις Δεδομένων- Computer Ανάλυση



1.8 Θέμα 1 Νοέμβριος 2018-Εργαστήριο

Καλείστε να σχεδιάσετε το ER διάραμμα μιας βάσης δεδομένων για την υλοποίηση ενός διαδικτυακού συστήματος προσλήψεων προσωπικού (e-recruitment) το οποίο χρησιμοποιούν εταιρείες που αναρτούν διαθέσιμες θέσεις εργασίας και υποψήφιοι που αναζητούν εργασία. Για τη ΒΔ έχετε στη διάθεσή σας τις παρακάτω λειτουργικές προδιαγραφές:

Οι χρήστες του συστήματος διακρίνονται σε δύο κατηγορίες: υποψήφιοι (που κάνουν αιτήσεις για τις διαθέσιμες θέσεις εργασίας) και Υπεύθυνοι Προσλήψεων (που ανακοινώνουν τις θέσεις εργασίας). Για τους Χρήστες του συστήματος διατηρούμε πληροφορίες για το username (μοναδικό για κάθε χρήστη), το password, ένα email και την ημερομηνία εγγραφής. Για κάθε Υποψήφιο το σύστημα αποθηκεύει ένα βιογραφικό, τις ξένες γλώσσες που γνωρίζει (μία ή και περισσότερες), το όνομα του αρχείου με τις συστατικές επιστολές (θεωρούμε ότι κάθε υποψήφιος μπορεί να υποβάλει στο σύστημα ένα αρχείο με όλες τις συστατικές επιστολές του), καθώς και ένα κείμενο όπου αναφέρονται τυχόν πιστοποιήσεις που έχει.

Κάθε υποψήφιος διαθέτει ένα ή περισσότερα πτυχία και για καθένα από αυτά καταγράφουμε τον τίτλο του και το ίδρυμα που το απονέμει (ο συνδυασμός τίτλου και ιδρύματος θεωρείται μοναδικός), το έτος κτήσης και το βαθμό του Υποψήφιου στο πτυχίο. Ένα πτυχίο μπορεί να το κατέχουν περισσότεροι από ένας Υποψήφιοι, αλλά τουλάχιστον ένας. Οι υποψήφιοι μπορεί και να μην κατέχουν κάποιο πτυχίο.

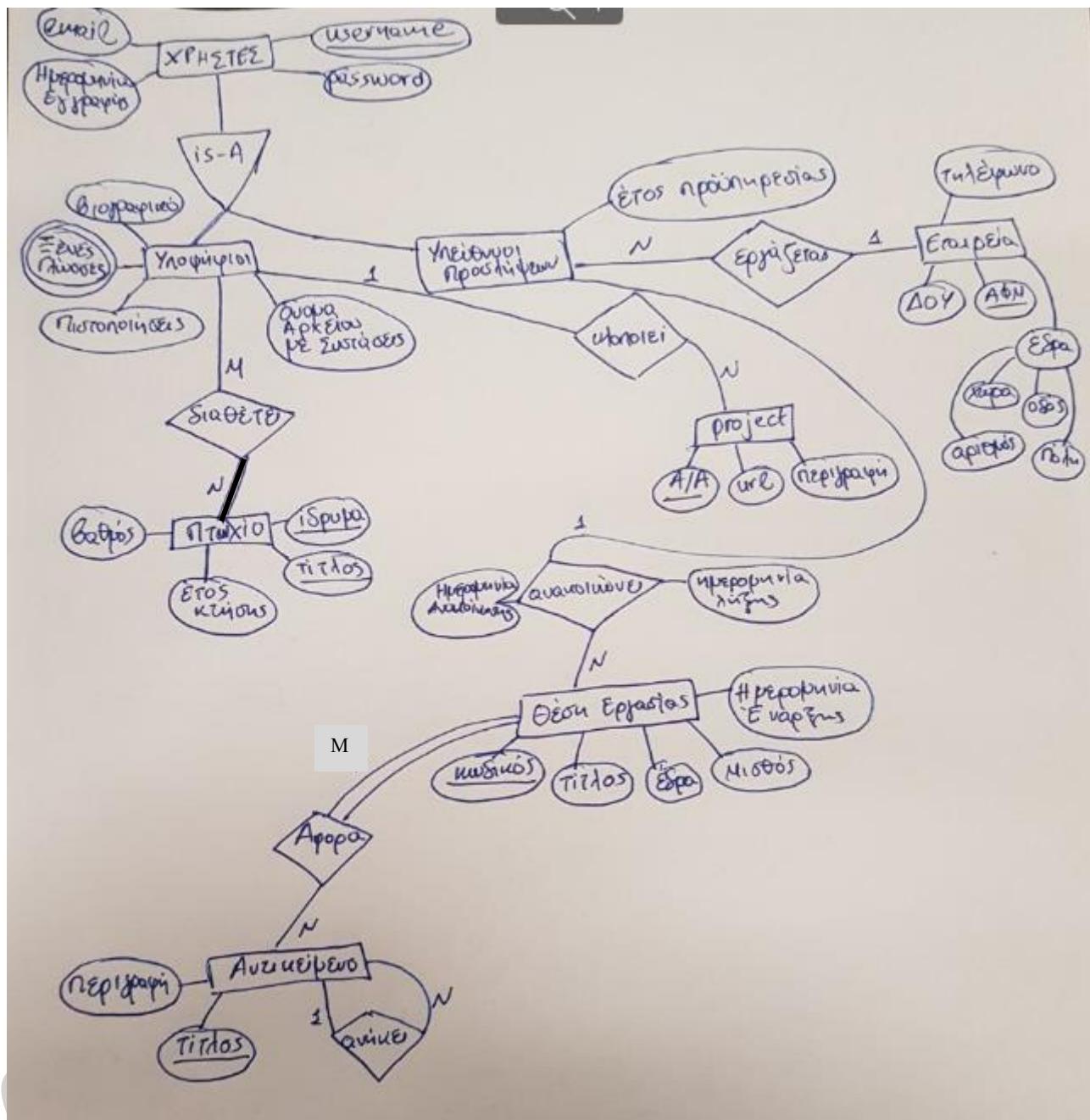
Επιπλέον οι υποψήφιοι μπορούν να καταγράφουν στο σύστημα τα projects που έχουν υλοποιήσει. Για κάθε project αποθηκεύεται μια περιγραφή της υλοποίησης, ένα url όπου είναι διαθέσιμος ο κώδικας και άλλες σχετικές πληροφορίες και κάθε project και έχει έναν αύξοντα αριθμό (a/a) μοναδικό για τον υποψήφιο που το έχει υλοποιήσει (το πρώτο project που υποβάλει ο υποψήφιος στο σύστημα έχει τον a/a 1, το δεύτερο το 2 κ.ο.κ. και αυτό ισχυει για κάθε υποψήφιο). Ένας υποψήφιος μπορεί να μην έχει κάποιο project.

Κάθε υπεύθ. προσλήψεων εργάζεται σε μία εταιρεία και το σύστημα αποθηκεύει για αυτόν τα έτη προϋπηρεσίας του (όχι στη συγκεκριμένη εταιρεία αλλά συνολικά ως υπευθύνου προσλήψεων). Για τις εταιρείες το σύστημα διατηρεί στοιχεία για το ΑΦΜ (μοναδικό), τη ΔΟΥ, το τηλέφωνο και την έδρα (οδός, αριθμός, πόλη και χώρα). Μια εταιρεία μπορεί να απασχολεί πολλούς υπευθύνους προσλήψεων. Κάθε ανακοίνωση θέσης εργασίας από έναν υπεύθυνο προσλήψεων συνοδεύεται από την ημερομηνία ανακοίνωσης και την ημερομηνία λήξης υποβολών.

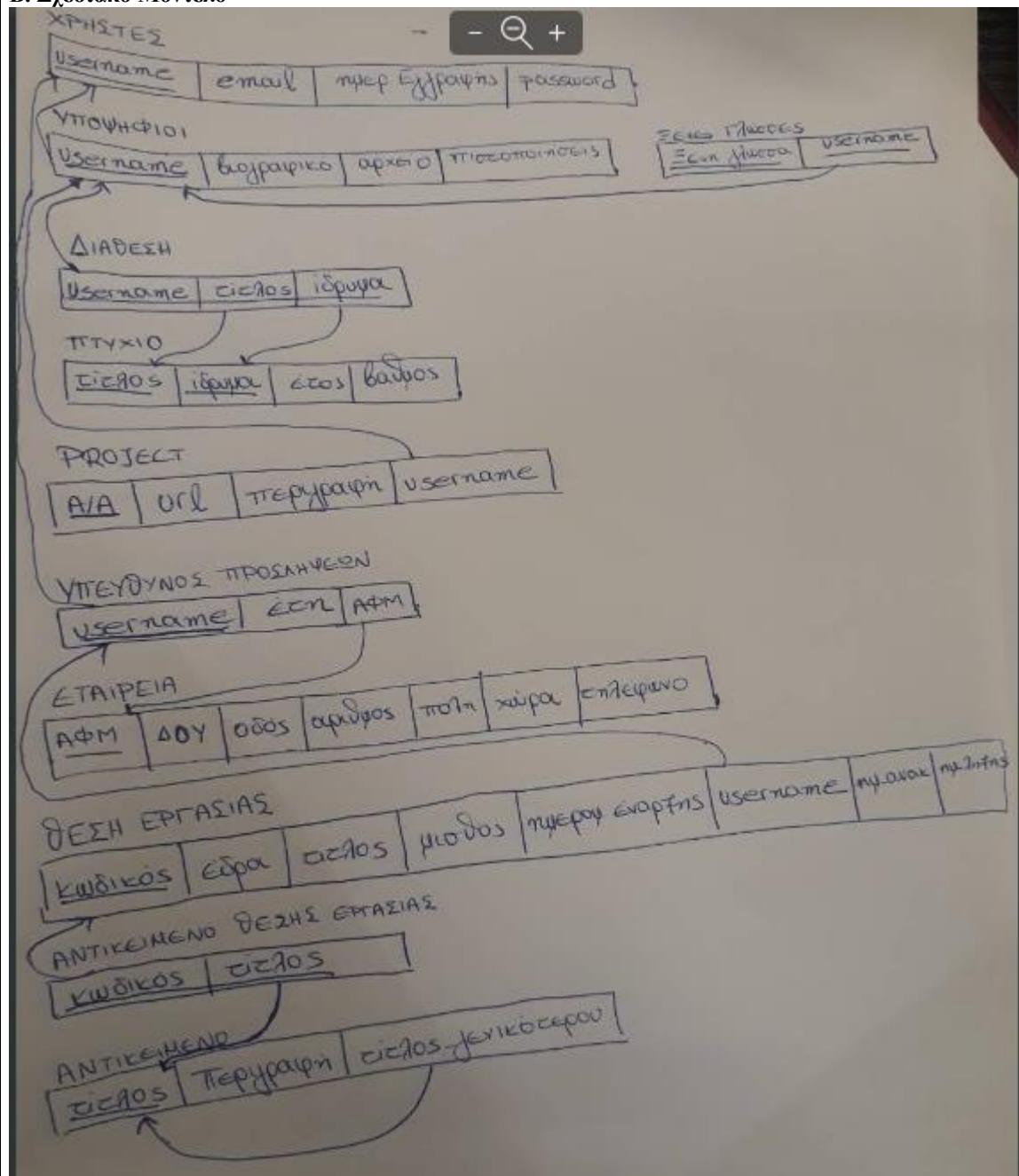
Για κάθε θέση εργασίας διατηρούμε έναν κωδικό (μοναδικό), τον τίτλο της, την έδρα όπου θα εργάζεται ο υποψήφιος που θα προσληφθεί, τον μισθό της και την ημερομηνία έναρξης εργασίας. Επιπλέον, κάθε θέση εργασίας αφορά ένα (τουλάχιστον) ή περισσότερα αντικείμενα (γνώσεις) για τα οποία αποθηκεύουμε τίτλο (μοναδικός) και περιγραφή. Κάθε γνωστικό αντικείμενο μπορεί να ανήκει σε ένα γενικότερο αντικείμενο (και κάθε αντικείμενο μπορεί να περιέχει πολλά πιο ειδικά αντικείμενα. Περισσότερες από μία θέσης εργασίας μπορούν να αφορούν το ίδιο αντικείμενο.

Σημείωση: Σε περίπτωση που χρειαστεί να κάνετε κάποιες παραδοχές στη σχεδίασή σας, φροντίστε να τις καταγράψετε με σαφήνεια.

A. Διάγραμμα Οντοτήτων-Συσχετίσεων (E-R)



B. Σχεσιακό Μοντέλο

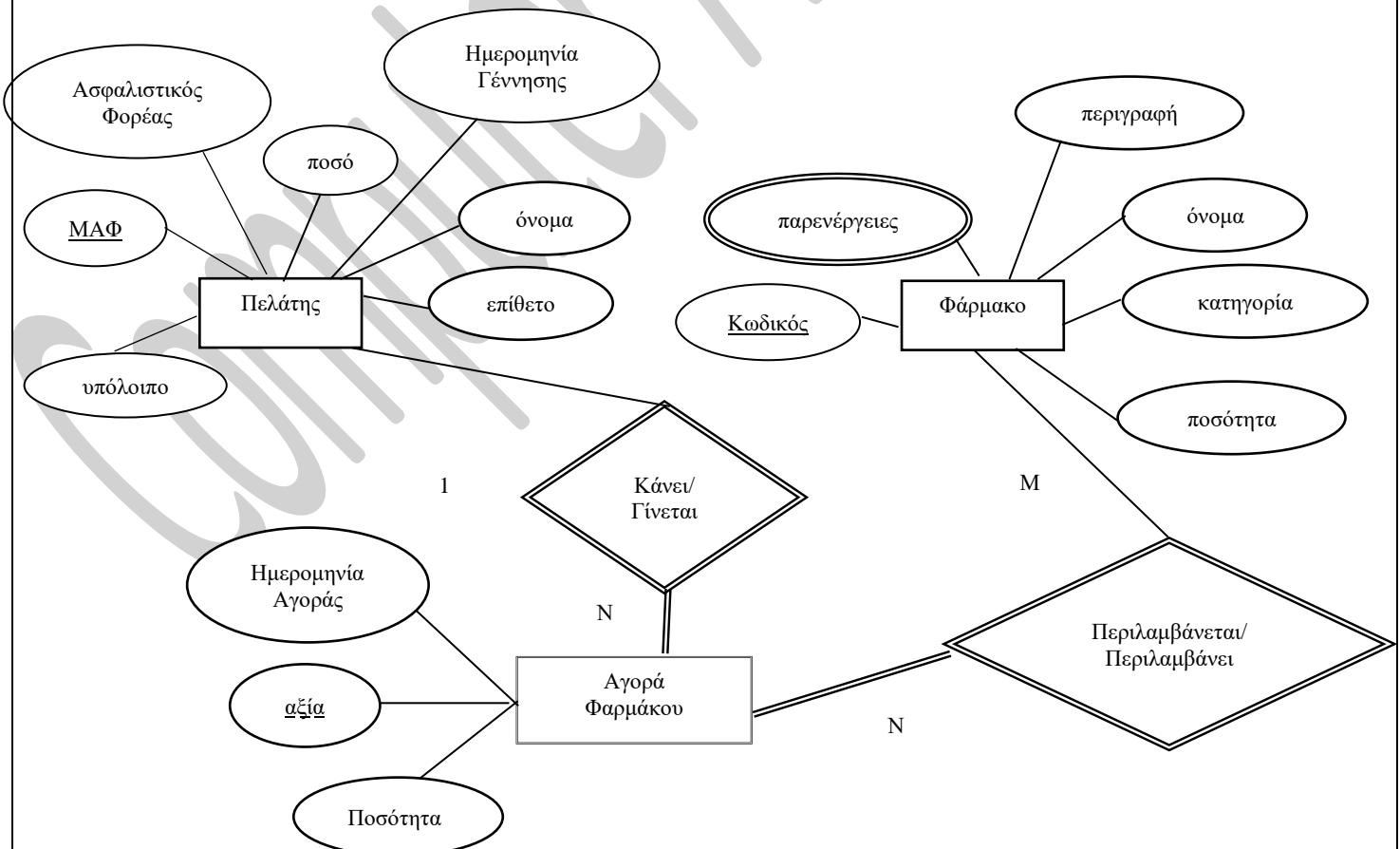


Θέμα 1 [30 μονάδες]

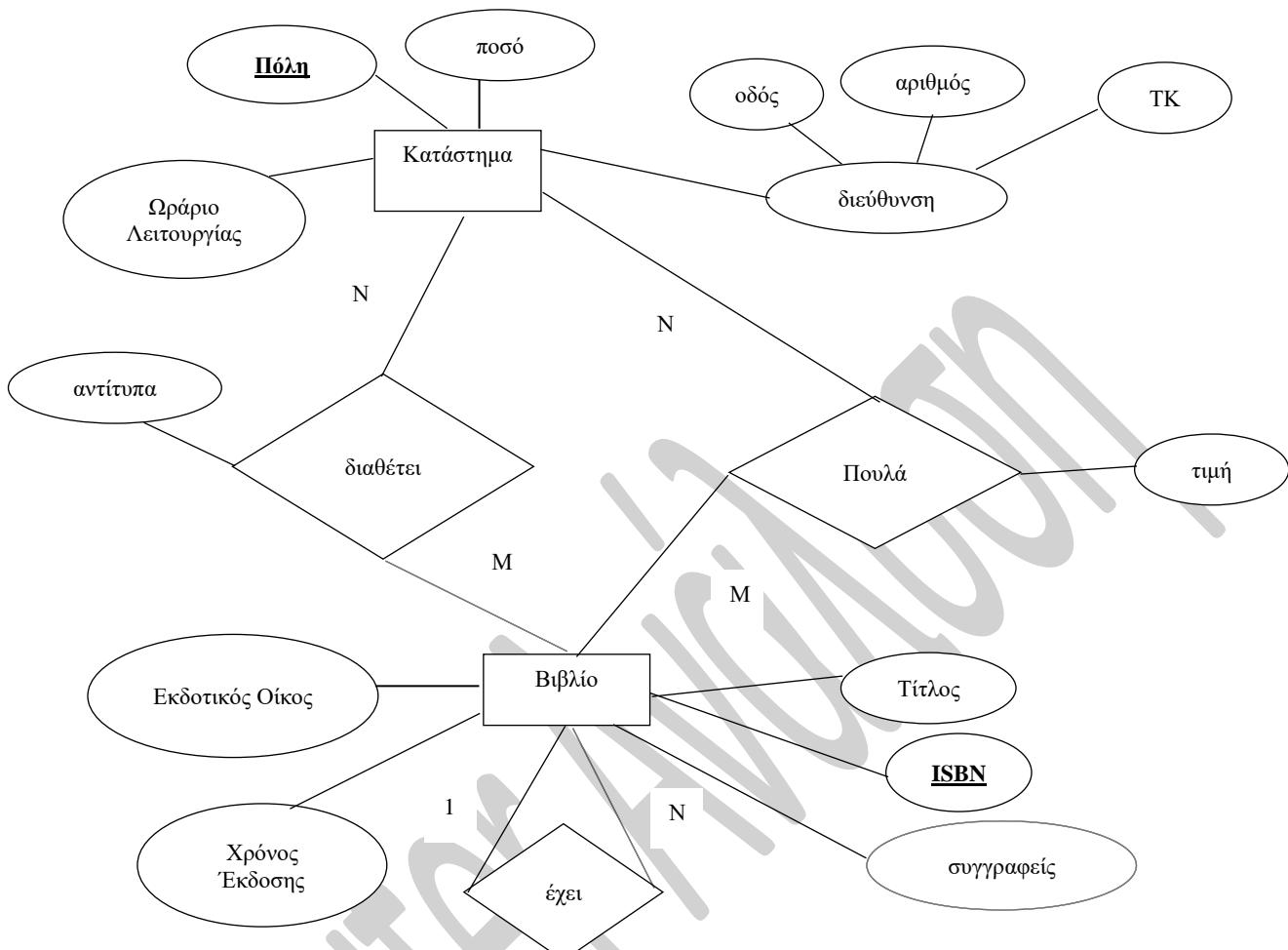
Θεωρήστε μία βάση δεδομένων η οποία διατηρεί πληροφορίες για ένα ερευνητικό κέντρο. Το κέντρο αυτό απασχολεί ερευνητές για τους οποίους πρέπει να αποθηκεύουμε το όνομα τους, το επώνυμο τους, την ειδίκευση που έχουν, τα ερευνητικά τους ενδιαφέροντα και τον μισθό που παίρνουν. Οι ερευνητές που απασχολούνται στο κέντρο μπορεί να είναι είτε βασικοί ερευνητές, είτε εξωτερικοί συνεργάτες. Για τους βασικούς ερευνητές πρέπει να γνωρίζουμε την βαθμίδα στην οποία ανήκουν. Για τους εξωτερικούς επιστημονικούς συνεργάτες μας ενδιαφέρει να αποθηκεύουμε το ίδρυμα/ερευνητικό κέντρο της κύριας απασχόλησής τους. Το ερευνητικό κέντρο αναλαμβάνει να υλοποιήσει έργα για τα οποία διατηρούμε πληροφορίες σχετικά με τον τίτλο, την διάρκειά τους, τον φορέα χρηματοδότησής τους και το ποσό προϋπολογισμού. Σε κάθε έργο πρέπει να ανατίθενται συμμετέχοντες βασικοί ερευνητές, οι οποίοι λαμβάνουν ένα χρηματικό bonus για την συμμετοχή τους. Επιπλέον, οι βασικοί ερευνητές μπορούν να συμμετέχουν σε περισσότερα από ένα έργο. Τα ερευνητικά έργα μπορούν να απασχολούν εξωτερικούς συνεργάτες, ενώ κάθε εξωτερικός επιστημονικός συνεργάτης πρέπει να απασχολείται σε ένα έργο για συγκεκριμένο χρονικό διάστημα, το οποίο αποθηκεύεται στην βάση. Τέλος, όλοι οι ερευνητές δημοσιεύουν επιστημονικά άρθρα (δημοσιεύσεις), για τα οποία διατηρούμε τον τίτλο τους, το συνέδριο/περιοδικό και το έτος δημοσιεύσης.

(i) Σχεδιάστε ένα διάγραμμα E-R για την παραπάνω εφαρμογή και διατυπώστε γραπτώς τις υποθέσεις που κάνατε

(ii) Μετατρέψτε το E-R διάγραμμα που σχεδιάσατε σε σχεσιακό όπου θα φαίνονται τα κλειδιά (ξένα, πρωτεύοντα, κτλ) της κάθε σχέσης

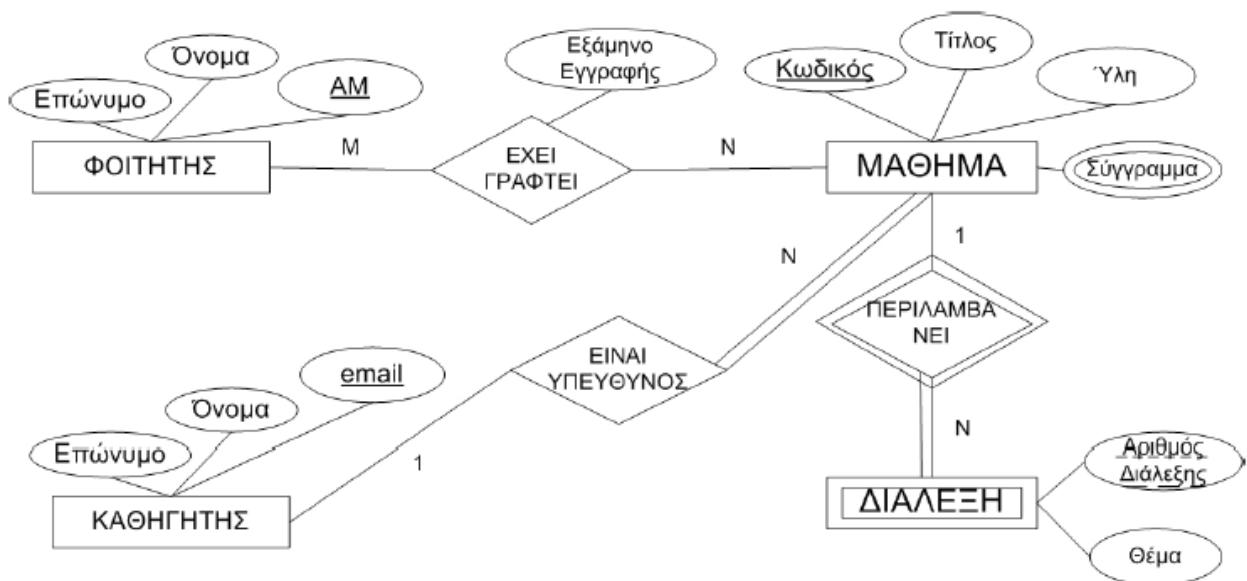
Παράδειγμα1-Φαρμακείο

Βάσεις Δεδομένων- Computer Ανάλυση
Παράδειγμα 2-Βιβλιοπωλείο





Παράδειγμα 1° - Πανεπιστήμιο



Καθηγητής

<u>email</u>	όνομα	επώνυμο
--------------	-------	---------

Μάθημα

Κωδικός τίτλος Ύλη email

Συγγράμματα

<u>Σύγγραμμα</u>	<u>Κωδικός</u>
------------------	----------------

Απόλεξη

Διαλέξης

Εγγονών

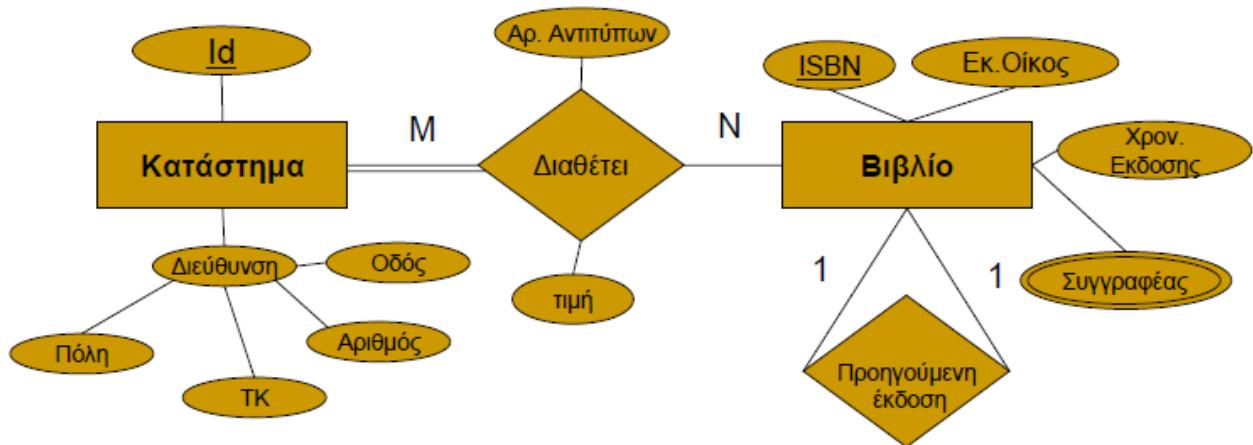
Εγγραφή

Φοιτητής

ΑΜ όνομα επώνυμο



Παράδειγμα 2° – Βιβλιοπωλείο



Κατάστημα

<u>id</u>	Πόλη	αριθμός	ΤΚ	οδός
-----------	------	---------	----	------

Διάθεση

<u>id</u>	<u>ISBN</u>	τιμή	Αριθμός αντιτύπων
-----------	-------------	------	-------------------

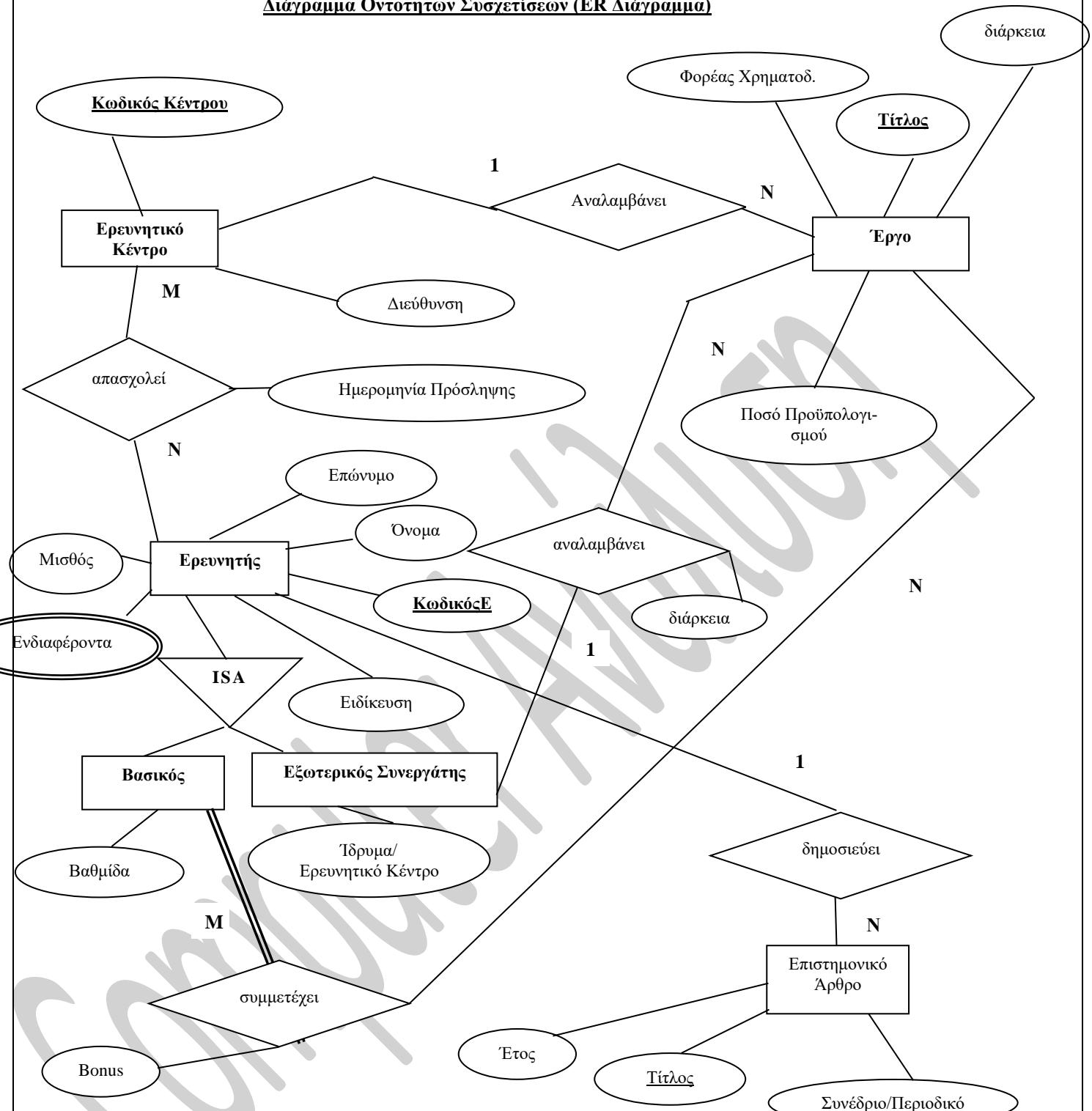
Βιβλίο

ISBN Εκδοτικός οίκος Χρονολογία έκδοσης ISBN επανέκδοσης

Συγγραφείς

Συγγραφέας ISBN

Βάσεις Δεδομένων- Computer Ανάλυση Διάγραμμα Οντοτητών Συγχετίσεων (ER Διάγραμμα)



Υποθέσεις

- Το ερευνητικό κέντρο είναι ισχυρή οντότητα με κλειδί Κωδικός_Κέντρου
 - Ο Ερευνητής είναι ισχυρή οντότητα με κλειδί ΚωδικόςΕ
 - Το Επιστημονικό Άρθρο είναι ισχυρή οντότητα με κλειδί ίτλος
 - Τα Ενδιαφέροντα είναι πλειότιμο γνώρισμα

Σχεδιακό Μοντέλο

Ερευνητικό Κέντρο

Κωδικός Κέντρου

Ερευνητής

Κωδικός_E	Όνομα	Επώνυμο	Μισθός	Ειδίκευση
-----------	-------	---------	--------	-----------

Βασικός <u>Κωδικός_E</u>	Βαθμίδα
-----------------------------	---------

Εξωτερικός Συνεργάτης <u>Κωδικός_E</u>	Βαθμίδα
---	---------

Συμμετέχει Έργο <u>Τίτλος</u>	Κωδικός_E	Bonus
----------------------------------	-----------	-------

Έργο <u>Τίτλος</u>	Διάρκεια	Φορέας Χρηματοδότησης	Ποσό Προϋπολογισμού
-----------------------	----------	-----------------------	---------------------

Αναλαμβάνει Έργο <u>Τίτλος</u>	Κωδικός_E	Διάρκεια
-----------------------------------	-----------	----------

Επιστημονικό Άρθρο <u>Τίτλος</u>	Έτος	Συνέδριο/Περιοδικό	Κωδικός_E
-------------------------------------	------	--------------------	-----------

Επιστημονικά Ενδιαφέροντα

Επιστημονικό Ενδιαφέρον	<u>Κωδικός_E</u>
-------------------------	------------------

2 ΘΕΜΑΤΑ ΜΕ ΣΧΕΣΙΑΚΗ ΑΛΓΕΒΡΑ, QBE και SQL

2.1 Θέμα 3 Φεβρουάριος 2009

Θεωρήστε τη σχέση **ΠΡΟΙΟΝ**

ΚΩΔ_Π:INT	ΟΝΟΜΑ: VARCHAR
1	ΟΜΠΡΕΛΑ
2	ΚΑΡΕΚΛΑ
3	ΠΕΤΣΕΤΑ

Θεωρήστε τη σχέση **ΠΕΛΑΤΗΣ**

ΑΜ:INT	ΟΝΟΜΑ: VARCHAR
1	ΜΑΡΙΑ
2	ΜΑΡΙΑ
3	ΓΙΑΝΝΗΣ

Θεωρήστε τη σχέση **ΑΓΟΡΑ**

ΚΩΔ_Π:INT	ΑΜ_Π: INT	ΠΟΣΟΤΗΤΑ: INT
1	1	1
2	1	2
1	2	2
1	3	1
2	3	3
3	1	3

Να εμφανίσετε το αποτέλεσμα εκτέλεσης καθεμίας από τις ακόλουθες εντολές για όσες εντολές είναι εφικτό να εκτελεστούν:

- α)ΠΟΝΟΜΑ(ΠΕΛΑΤΗΣ)
- β)ΠΟΝΟΜΑ(ΠΕΛΑΤΗΣ) \cup ΠΟΝΟΜΑ(ΠΡΟΙΟΝ)
- γ)σ ποσοτητα ≥ 2 (Π_{ΚΩΔ_Π,ΑΜ}(ΑΓΟΡΑ \bowtie ΠΡΟΙΟΝ))
- δ)ΠΙΕΛΑΤΗΣ-ΠΡΟΙΟΝ
- ε)Π_{ΚΩΔ_Π, ΑΜ_Π}(ΑΓΟΡΑ) \div Π_{ΚΩΔ_Π}(ΠΡΟΙΟΝ)

Απάντηση

- α)ΠΟΝΟΜΑ(ΠΕΛΑΤΗΣ)

ΟΝΟΜΑ
ΜΑΡΙΑ
ΓΙΑΝΝΗΣ

β)ΠΟΝΟΜΑ(ΠΕΛΑΤΗΣ) \cup ΠΟΝΟΜΑ(ΠΡΟΙΟΝ)

ONOMA
ΓΙΑΝΝΗΣ
ΚΑΡΕΚΛΑ
ΜΑΡΙΑ
ΟΜΠΡΕΛΑ
ΠΕΤΣΕΤΑ

γ)σποσοτητα $>=2$ (Πκωδ_π,αμ(ΑΓΟΡΑ \bowtie ΠΡΟΙΟΝ))

Πρώτα υπολογίζουμε τον πίνακα ΑΓΟΡΑ \bowtie ΠΡΟΙΟΝ που είναι ο ακόλουθος:

ΑΓΟΡΑ.ΚΩΔ_Π	ΑΜ_Π	ΠΟΣΟΤΗΤΑ	ΠΡΟΪΟΝ.ΚΩΔ_Π	ΟΝΟΜΑ
1	1	1	1	ΟΜΠΡΕΛΑ
2	1	2	2	ΚΑΡΕΚΛΑ
1	2	2	1	ΟΜΠΡΕΛΑ
1	3	1	1	ΟΜΠΡΕΛΑ
2	3	3	2	ΚΑΡΕΚΛΑ
3	1	3	3	ΠΕΤΣΕΤΑ

Μετά υπολογίζουμε τον πίνακα Πκωδ_π,αμ_π(ΑΓΟΡΑ \bowtie ΠΡΟΙΟΝ) που είναι ο ακόλουθος:

ΑΓΟΡΑ.ΚΩΔ_Π	ΑΜ_Π
1	1
2	1
1	2
1	3
2	3
3	1

Τέλος ο τελεστής σποσοτητα $>=2$ δεν μπορεί να εφαρμοστεί διότι στον τελικό πίνακα που έχουμε μπροστά μας δεν υπάρχει το πεδίο ΠΟΣΟΤΗΤΑ

δ)ΠΕΛΑΤΗΣ-ΠΡΟΙΟΝ

Ο τελεστής – επιλέγει τις εγγραφές που ανήκουν στον 1^ο πίνακα δηλ. στον πίνακα ΠΕΛΑΤΗΣ και όχι στο 2^ο δηλ. στον πίνακα ΠΡΟΙΟΝ

ΑΜ	ΟΝΟΜΑ
1	MAPRIA
2	MAPRIA
3	ΓΙΑΝΝΗΣ

ΠΡΟΣΟΧΗ: Η προϋπόθεση για να εφαρμόσουμε τους τελεστές ένωση, τομή και διαφορά είναι οι σχέσεις να είναι συμβατές ως προς την ένωση δηλ. να έχουν τον ίδιο αριθμό στηλών (γνωρισμάτων) και τα αντίστοιχα πεδία τους να είναι του ίδιου τύπου.

ε)Π_{ΚΩΔ_Π,ΑΜ_Π(ΑΓΟΡΑ)} ÷ Π_{ΚΩΔ_Π(ΠΡΟΙΟΝ)}

Πρώτα κατασκευάζουμε τον πίνακα που προκύπτει από τον τελεστή Π_{ΚΩΔ_Π, ΑΜ_Π(ΑΓΟΡΑ)}

ΚΩΔ_Π	ΑΜ_Π
1	1
2	1
1	2
1	3
2	3
3	1

Μετά υπολογίζουμε τον πίνακα που προκύπτει από τον τελεστή Π_{ΚΩΔ_Π(ΠΡΟΪΟΝ)}

ΚΩΔ_Π
1
2
3

Μετά υπολογίζουμε τον πίνακα που προκύπτει από τον τελεστή διαίρεσης

ΚΩΔ_Π	ΑΜ_Π	÷	ΚΩΔ_Π	=	ΑΜ_Π
1	1				
2	1				
1	2				
1	3				
2	3				
3	1				

Παρατήρηση

Για να εφαρμόσουμε τον τελεστή διαιρεσης πρέπει οι 2 σχέσεις να έχουν κοινή στήλη (κοινό γνώρισμα). Στο πηλίκο πηγαίνουν όλες οι στήλες του διαιρετέου εκτός της κοινής στήλης και όλες οι εγγραφές του διαιρετέου που έχουν όλες τις τιμές του διαιρέτη στο κοινό πεδίο. Εδώ πάνε στο πηλίκο οι τιμές του AM_P που έχουν και τις 3 τιμές του πεδίου KΩΔ_P.

2.2 Θέμα 2 Φεβρουάριος 2013

Δίνονται οι παρακάτω σχέσεις:

Πελάτης			
<u>Id</u>	Όνομα	Email	Τηλ.

Προϊόν			
<u>Id</u>	Όνομα	Κατηγορία	Τιμή

Αγορά		
<u>Id-Πελάτης</u>	<u>Id-Προϊόν</u>	Ημερομηνία

Να γράψετε σε **σχεσιακή άλγεβρα** και σε **SQL** τα παρακάτω ερωτήματα:

- Τα ονόματα των πελατών που έχουν αγοράσει κάποιο προϊόν
- Τα ονόματα των προϊόντων που έχουν αγοραστεί και κοστίζουν πάνω από X ευρώ
- Τα ονόματα και τις τιμές των προϊόντων που δεν έχουν πουληθεί
- Τα ονόματα και τα τηλέφωνα των πελατών που έχουν πραγματοποιήσει αγορές πάνω από Y ευρώ

Απάντηση**ΣΧΕΣΙΑΚΗ ΑΛΓΕΒΡΑ**

- .Πόνομα(Πελάτης $\bowtie_{id=id_πελάτης}$ Αγορά)
- .Πόνομα($\sigma_{τιμή > X}$ (Προϊόν $\bowtie_{id=id_προϊόν}$ Αγορά))
- .Πόνομα, τιμή (Προϊόν) – Πόνομα, τιμή (Προϊόν $\bowtie_{id=id_προϊόν}$ Αγορά)
- .Πελάτης.όνομα, τηλέφωνο($\sigma_{τιμή > Y}$ (Πελάτης $\bowtie_{id=id_πελάτης}$ Αγορά) $\bowtie_{id_προϊόν=id}$ Προϊόν)

SQL

a.

```
SELECT Πελάτης.όνομα
FROM Πελάτης INNER JOIN Αγορά ON id=id_πελάτης;
          ή εναλλακτικά
SELECT Πελάτης.όνομα
FROM Πελάτης, Αγορά
WHERE id=id_πελάτης;
```

b.

```
SELECT Προϊόν.όνομα
FROM Προϊόν INNER JOIN Αγορά ON id=id_προϊόν
WHERE Τιμή > X;
          ή εναλλακτικά
SELECT Προϊόν.όνομα
```

Βάσεις Δεδομένων- Computer Ανάλυση

FROM Προϊόν, Αγορά

WHERE id=id_προϊόν AND Τιμή>X;

c.

SELECT Προϊόν.Όνομα, Προϊόν.τιμή

FROM Προϊόν

WHERE id NOT IN (SELECT id_προϊόν **FROM** Αγορά);

d.

SELECT Πελάτης.όνομα, Τηλέφωνο

FROM Πελάτης **INNER JOIN** Αγορά **ON** Πελάτης.id=id_πελάτης **INNER JOIN** Προϊόν **ON** id_προϊόν=Προϊόν.id

WHERE Τιμή>Y;

ή

SELECT Πελάτης.όνομα, Τηλέφωνο

FROM Πελάτης, Αγορά, Προϊόν

WHERE Πελάτης.id=id_πελάτης **AND** id_προϊόν=Προϊόν.id **AND** Τιμή>Y;

2.3 Θέμα 5 Φεβρουάριος 2008

q	
A	B
1	13
2	33
3	23
4	4
5	44
6	25
7	78
8	12

r		
A	C	D
1	1	13
2	1	33
2	2	32
4	3	4
4	5	44
8	2	25

k		
A	E	D
1	3	12
1	6	11
4	1	31
2	1	32
8	1	25

Δώστε το αποτέλεσμα σε καθένα από τα παρακάτω ερωτήματα:

1. **SELECT A
FROM q
ORDER BY B ASC;**
2. **SELECT COUNT(*)
FROM q
WHERE B > 20**
3. **SELECT COUNT(DISTINCT(E))
FROM k
WHERE D >= (SELECT MAX(A)
FROM q);**
4. **SELECT q.A, SUM(k.D)
FROM q INNER JOIN k ON
q.A=k.A
WHERE q.B>10
GROUP BY q.A ;**
5. **SELECT k.A , COUNT(r.A)
FROM r INNER JOIN k ON r.A =
k.A AND r.D= k.D
GROUP BY k.A
HAVING SUM(r.D)>5;**
6. **SELECT COUNT(q1.B)
FROM q AS q1 INNER JOIN q AS
q2 ON q1.A=q2.A+1
WHERE q1.A>7;**

Απάντηση

1)

A
4
8
1
3
6
2
5
7

2)

count(*)
5

3) Πρώτα εκτελείται το υποερώτημα στην παρένθεση και η απάντηση του υποερωτήματος είναι:

max(A)
8

Μετά εκτελείται το εξωτερικό ερώτημα και επιλέγει τα μοναδικά Ε του πίνακα k που το D τους είναι ≥ 8 και η απάντηση είναι:

COUNT(DISTINCT(E))
3

4)

Πρώτα εκτελείται το τμήμα: q INNER JOIN k ON q.A=k.A

q.A	q.B	k.A	k.E	k.D
1	13	1	3	12
1	13	1	6	11
2	33	2	1	32
4	4	4	1	31
8	12	8	1	25

Μετά εκτελείται το τμήμα: WHERE q.B>10

q.A	q.B	k.A	k.E	k.D
1	13	1	3	12
1	13	1	6	11
2	33	2	1	32
8	12	8	1	25

Βάσεις Δεδομένων- Computer Ανάλυση

Μετά εκτελείται το τμήμα: **GROUP BY q.A**

q.A	q.B	k.A	k.E	k.D
1	13	1	3	12
1	13	1	6	11
2	33	2	1	32
8	12	8	1	25

Τέλος εκτελείται το τμήμα: **SELECT(q.A,SUM(k.D))**

A	SUM(k.D)
1	23
2	32
8	25

5)

Πρώτα εκτελείται το τμήμα: **r INNER JOIN k ON r.A=k.A AND r.D=k.D**

r.A	r.C	r.D	k.A	k.E	k.D
2	2	32	2	1	32
8	2	25	8	1	25

Μετά εκτελείται το τμήμα: **GROUP BY k.A**

r.A	r.C	r.D	k.A	k.E	k.D
2	2	32	2	1	32
8	2	25	8	1	25

Μετά εκτελείται το τμήμα: **HAVING SUM(r.D)>5**

r.A	r.C	r.D	k.A	k.E	k.D
2	2	32	2	1	32
8	2	25	8	1	25

Τέλος εκτελείται το τμήμα: **SELECT k.A,COUNT(r.A)**

k.A	COUNT(r.A)
2	1
8	1

6)

Πρώτα εκτελείται το τμήμα: **FROM q as q1 INNER JOIN q AS q2 ON q1.A=q2.A+1**

q1.A	q1.B	q2.A	q2.B
2	33	1	13
3	23	2	33
4	4	3	23
5	44	4	4
6	25	5	44
7	78	6	25
8	12	7	78

Μετά εκτελείται το τμήμα: **WHERE q1.A>7**

q1.A	q1.B	q2.A	q2.B
8	12	7	78

Τέλος εκτελείται το τμήμα: **SELECT count(q1.B)**

count(q1.B)
1

2.4 Θέμα 2 Φεβρουάριος 2010

Θεωρήστε το παρακάτω σχετικό σχήμα το οποίο διατηρεί πληροφορίες σχετικά με τα στοιχεία μιας εταιρίας:

Τοποθεσία (Τοπ_κωδικός, Όνομα)

Κατάστημα (Κατ_κωδικός, Όνομα, Τοπ_κωδικός)

Εργασία (Ερ_κωδικός, Επάγγελμα)

Υπάλληλος (Ερ_κωδικός, Όνομα, Επώνυμο, Μισθός, Κατ_κωδικός, Ερ_κωδικός, Διευθυντής_κωδικός)

Εκφράστε σε SQL τα παρακάτω ερωτήματα:

1. Βρείτε όλους τους υπαλλήλους που το επώνυμο τους ξεκινά από X
2. Βρείτε τους υπαλλήλους που παίρνουν το μεγαλύτερο μισθό από όλους τους υπαλλήλους στο κατάστημα με κωδικό 30
3. Βρείτε τα ονόματα και το πλήθος των καταστημάτων τα οποία έχουν περισσότερους από 5 υπαλλήλους και εμφανίστε τα σε φθίνουσα σειρά ως προς το όνομα του καταστήματος

Απάντηση

1.

SELECT *

FROM υπάλληλος

WHERE επώνυμο **LIKE** 'x%';

2.

SELECT *

FROM υπάλληλος

WHERE μισθός > **MAX ALL (SELECT** μισθός **FROM** Υπάλληλος **WHERE** Κατ_κωδικός=30);

Εναλλακτικά:

SELECT *

FROM υπάλληλος

WHERE μισθός > (**SELECT MAX(μισθός) FROM** Υπάλληλος **WHERE** Κατ_κωδικός=30);

3.

CREATE VIEW result **AS**

SELECT κατάστημα.όνομα, **COUNT** (ερ_κωδικός)

FROM υπάλληλος **INNER JOIN** κατάστημα **ON** κατάστημα.κατ_κωδικός = υπάλληλος.κατ_κωδ

GROUP BY κατ_κωδικός

HAVING COUNT (ερ_κωδικός)>5;

SELECT όνομα, **COUNT** (ερ_κωδικός)

FROM result

ORDER BY όνομα **DESC**;

2.5 Θέμα 4 Φεβρουάριος 2008

Υποθέστε το παρακάτω σχεσιακό σχήμα το οποίο περιγράφει μια τράπεζα.

- Παράρτημα(όνομα-παραρτήματος, πόλη-παραρτήματος, ενεργητικό)
- Πελάτης(όνομα-πελάτης, διεύθυνση-πελάτη, customer-only)
- Λογαριασμός(αριθμός-λογαριασμού, όνομα-παραρτήματος, υπόλοιπο)
- Δάνειο(αριθμός-δανείου, όνομα-παραρτήματος, ποσό)
- Καταθέτης(όνομα-πελάτη, αριθμός-λογαριασμού)
- Δανειολήπτης(όνομα-πελάτη, αριθμός-δανείου)

(Υπονοείται ότι τα γνωρίσματα με κοινό όνομα είναι ξένα κλειδιά)

Αναπτύξτε σε σχεσιακή άλγεβρα και σε SQL τα παρακάτω ερωτήματα:

1. Βρείτε όλα τα δάνεια πάνω από 1.200€

2. Βρείτε μόνο τον αριθμό δανείου για κάθε δάνειο πάνω από 1.200€

3. Βρείτε τα ονόματα των πελατών που έχουν κάποιο δάνειο, λογαριασμό ή και τα δύο

4. Βρείτε τα ονόματα των πελατών που έχουν κάποιο δάνειο στο κατάστημα της Πάτρας καθώς και το ποσό του δανείου

5. Για κάθε παράρτημα (εμφανίζοντας το όνομα του) να υπολογίσετε το πλήθος των πελατών που έχουν δάνειο σε αυτό

6. Για το παράρτημα της Πάτρας (εμφανίζοντας το όνομα του) να υπολογίσετε το συνολικό ποσό δανείων που έχει δώσει στους πελάτες του

7. Βρείτε το όνομα κάθε δανειολήπτη και το συνολικό ποσό των δανείων που έχει πάρει αλλά μόνο για τους δανειολήπτες που το συνολικό ποσό δανείων τους είναι μεγαλύτερο από 100.000€

8. Βρείτε τα ονόματα των πελατών που έχουν μόνο κάποιο δάνειο και όχι λογαριασμό

9. Βρείτε τα ονόματα, τις διευθύνσεις και το υπόλοιπο των πελατών για όσους πελάτες έχουν λογαριασμό πάνω από 500€

Απάντηση

1.

Παριθμός-δανείου, όνομα-παραρτήματος, ποσό($\sigma_{\text{ποσό} > 1200}$ (Δάνειο))

SELECT αριθμός-δανείου, όνομα-παραρτήματος, ποσό

FROM Δάνειο

Βάσεις Δεδομένων- Computer Ανάλυση

WHERE ποσό>1200;

Εναλλακτικά:

SELECT *

FROM Δάνειο

WHERE ποσό>1200;

2.

Παριθμός-δανείου(Σποσό>1200(Δάνειο))

SELECT αριθμός-δανείου

FROM Δάνειο

WHERE ποσό>1200;

3.

$\Delta\Delta \leftarrow \text{Δανειολήπτης} \bowtie_{\text{αριθμός-δανείου} = \text{αριθμός-δανείου}} \text{Δάνειο}$. Εναλλακτικά $\Delta\Delta \leftarrow \text{Δανειολήπτης}^* \text{Δάνειο}$ διότι το κοινό τους πεδίο έχει το ίδιο όνομα

Πόνομα-πελάτη($\Delta\Delta$)

$\text{ΚΛ} \leftarrow \text{Καταθέτης} \bowtie_{\text{αριθμός-λογαριασμού} = \text{αριθμός-λογαριασμού}} \text{Λογαριασμός}$ Εναλλακτικά $\text{ΚΛ} \leftarrow \text{Καταθέτης}^* \text{Λογαριασμός}$

Πόνομα-πελάτη(ΚΛ)

Πόνομα-πελάτη($\Delta\Delta$) \cup Πόνομα-πελάτη(ΚΛ)

SELECT όνομα-πελάτη

FROM Δανειολήπτης, Δάνειο

WHERE Δανειολήπτης.αριθμός-δανείου= Δάνειο.αριθμός-δανείου

Union

SELECT όνομα-πελάτη

FROM Καταθέτης, Λογαριασμός

WHERE Λογαριασμός.αριθμός-λογαριασμού= Καταθέτης.αριθμός-αριθμός-λογαριασμού

Εναλλακτικά:

SELECT όνομα-πελάτη

FROM Δανειολήπτης **INNER JOIN** Δάνειο ον Δανειολήπτης.αριθμός-δανείου= Δάνειο.αριθμός-δανείου

UNION

SELECT όνομα-πελάτη

FROM Καταθέτης **INNER JOIN** Λογ. ον Καταθέτης.αριθμός-λογαριασμού=Λογαριασμός.αριθμός-λογαριασμού;

Βάσεις Δεδομένων- Computer Ανάλυση

4.

ΠΔ \leftarrow Παράτημα $\bowtie_{\text{όνομα-παραρτήματος} = \text{όνομα-παραρτήματος}$ Δάνειο. (Β Τρόπος: ΠΔ \leftarrow Παράτημα*Δάνειο)

Πόνομα-πελάτη, ποσό(Σπόλη-παραρτήματος='Πάτρα'(ΠΔ))

SELECT όνομα-πελάτη

FROM Παράτημα, Δάνειο, Δανειολήπτης

WHERE Παράτημα.όνομα-παραρτήματος = Δάνειο.όνομα-παραρτήματος **AND** Δάνειο.αριθμός-δανείου = Δανειολήπτης.αριθμός-δανείου **AND** πόλη-παραρτήματος='Πάτρα';

Εναλλακτικά

SELECT όνομα-πελάτη

FROM Παράτημα **INNER** Δάνειο

WHERE Παράτημα.όνομα-παραρτήματος = Δάνειο.όνομα-παραρτήματος **AND** πόλη-παραρτήματος='Πάτρα';

5.

ΠΔ \leftarrow Παράτημα $\bowtie_{\text{όνομα-παραρτήματος} = \text{όνομα-παραρτήματος}$ Δάνειο. (Β Τρόπος: ΠΔ \leftarrow Παράτημα*Δάνειο)

ΠΔΔ \leftarrow ΠΔ $\bowtie_{\text{αριθμός-δανείου} = \text{αριθμός-δανείου}$ Δανειολήπτης. (Β τρόπος: ΠΔΔ \leftarrow ΠΔ*Δανειολήπτης)

Α(Όνομα-Παραρτήματος, Πλήθος-Πελατών) \leftarrow όνομα-παραρτήματος F_{count(όνομα-πελάτη)}(ΠΔΔ)

πόνομα-παραρτήματος, πλήθος-πελατών(Α)

SELECT όνομα-παραρτήματος, count(όνομα-πελάτη)

FROM Παράτημα, Δάνειο, Δανειολήπτης

WHERE Παράτημα.όνομα-παραρτήματος= Δάνειο.όνομα-παραρτήματος **AND** Δάνειο.αριθμός-δανείου= Δανειολήπτης.αριθμός-δανείου

GROUP BY όνομα-παραρτήματος

Εναλλακτικά

SELECT όνομα-παραρτήματος, COUNT(όνομα-πελάτη)

FROM Παράτημα **INNER JOIN** Δάνειο ον Παράτημα.όνομα-παραρτήματος= Δάνειο.όνομα-παραρτήματος **INNER JOIN** Δανειολήπτης ον Δάνειο.αριθμός-δανείου= Δανειολήπτης.αριθμός-δανείου

GROUP BY όνομα-παραρτήματος

6.

ΠΔ \leftarrow Παράτημα $\bowtie_{\text{όνομα-παραρτήματος} = \text{όνομα-παραρτήματος}$ Δάνειο $\bowtie_{\text{Δανειολήπτης}}$ (Β Τρόπος: ΠΔ \leftarrow Παράτημα*Δάνειο)

ΠΠ \leftarrow Πόνομα-πελάτη, ποσό(Σπόλη-παραρτήματος='Πάτρα'(ΠΔ))

Α(πόλη-παραρτήματος, συνολικό-Ποσό) \leftarrow πόλη-παραρτήματος F_{sum(ποσό)}(ΠΠ)

Βάσεις Δεδομένων- Computer Ανάλυση

Πόνομα-πελάτη, συνολικό-ποσό(Α)

SELECT πόλη-παραρτήματος, sum(ποσό) as συνολικό-ποσό

FROM Παράρτημα, Δάνειο

WHERE Παράρτημα.όνομα-παραρτήματος=Δάνειο.όνομα-παραρτήματος **AND** πόλη-παραρτήματος='Πάτρα'

GROUP BY πόλη-παραρτήματος;

Εναλλακτικά

SELECT όνομα-πελάτη, sum (ποσό) as συνολικό-ποσό

FROM Παράρτημα **INNER JOIN** Δάνειο ον Παράρτημα.όνομα-παραρτήματος=Δάνειο.όνομα-παραρτήματος **AND** πόλη-παραρτήματος='Πάτρα'

GROUP BY πόλη-παραρτήματος;

7.

ΔΔ \leftarrow Δανειολήπτης αριθμός-δανείου= αριθμός-δανείου Δάνειο. (Β Τρόπος: ΠΔ \leftarrow Δανειολήπτης*Δάνειο)

ΠΔ(όνομα-πελάτη, συνολικό-ποσό) \leftarrow όνομα-πελάτη F_{sum(ποσό)(ΔΔ)}

Πόνομα-πελάτη(Συνολικό-ποσό>10000 (ΠΔ))

SELECT όνομα-πελάτη, sum (ποσό) as συνολικό-ποσό

FROM Δανειολήπτης, Δάνειο

WHERE Δανειολήπτης.αριθμός-δανείου=Δάνειο. αριθμός-δανείου

GROUP BY όνομα-πελάτη

Having sum (ποσό)>10000

8.

ΚΛ \leftarrow Καταθέτης αριθμός-λογαριασμού= αριθμός-λογαριασμού Λογαριασμός.(Β τρόπος: ΚΛ \leftarrow Καταθέτης*Λογαριασμός)

ΚΑΤ \leftarrow Πόνομα-πελάτη(ΚΛ)

ΟΛΟΙ \leftarrow Πόνομα-πελάτη(Πελάτης)

Πόνομα-πελάτη(ΟΛΟΙ-ΚΑΤ)

SELECT όνομα-πελάτη

FROM Πελάτης

WHERE όνομα-πελάτη not in (**SELECT** όνομα-πελάτη **FROM** Καταθέτης **INNER JOIN** Λογαριασμός ον Καταθέτης.αριθμός-λογαριασμού= Λογαριασμός.αριθμός-λογαριασμού)

9.

Βάσεις Δεδομένων- Computer Ανάλυση

ΚΠ←Καταθέτης  όνομα-πελάτη= όνομα-πελάτη Πελάτης.(Β τρόπος: ΚΠ←Καταθέτης*Πελάτης)

ΚΠΛ←ΚΠ  αριθμός-λογαριασμού= αριθμός-λογαριασμού Λογαριασμός

πόνομα-πελάτη, διεύθυνση-πελάτη, υπόλοιπο ($\sigma_{\text{υπόλοιπο} > 500}$ (ΚΠΛ))

SELECT όνομα-πελάτη, διεύθυνση-πελάτη, υπόλοιπο

FROM Πελάτης, Καταθέτης, Λογαριασμός

WHERE Καταθέτης.όνομα-πελάτη=Πελάτης.όνομα-πελάτη **AND** Πελάτης.αριθμός-λογαριασμού= Λογαριασμός. αριθμός-λογαριασμού **AND** υπόλοιπο>500

Εναλλακτικά:

SELECT όνομα-πελάτη, διεύθυνση-πελάτη, υπόλοιπο

FROM Πελάτης, Καταθέτης, Λογαριασμός

WHERE Καταθέτης.όνομα-πελάτη=Πελάτης.όνομα-πελάτη **AND** Πελάτης.αριθμός-λογαριασμού= Λογαριασμός. αριθμός-λογαριασμού **AND** υπόλοιπο>500

2.6 Θέμα 3 Ιανουάριος 2014

Δίνεται το παρακάτω σχεσιακό σχήμα το οποίο διατηρεί πληροφορίες σχετικά με ένα ζαχαροπλαστείο:

- Ζαχαροπλαστείο(κωδικός, πόλη, διεύθυνση)
- Γλυκό(κωδικός, όνομα, είδος, τιμή)
- Διατηρεί(Γλυκό_κωδικός, Ζαχαροπλαστείο_κωδικός)
- Ζαχαροπλάστης(ΑΦΜ, ονοματεπώνυμο, μισθός)
- Παρασκευάζει(Γλυκό_κωδικός, Ζαχαροπλάστης_ΑΦΜ)

Εκφράστε τα ακόλουθα ερωτήματα σε Σχεσιακή Άλγεβρα (ΣΑ) και SQL όπως σας ζητείται:

- i. Βρείτε τα ονόματα των γλυκών που είναι είδους 'πάστα' και κοστίζουν έως και 3 ευρώ (ΣΑ) [2 μονάδες], SQL [2 μονάδες]
- ii. Βρείτε τους κωδικούς των ζαχαροπλαστείων της Πάτρας που έχουν γλυκά με όνομα 'σοκολατίνα' (ΣΑ) [3 μονάδες], SQL [3 μονάδες]
- iii. Βρείτε τον κωδικό ζαχαροπλαστείου και το μέσο μισθό των εργαζομένων ανά ζαχαροπλαστείο για όσα ζαχαροπλαστεία παρασκευάζουν γλυκά με τιμή ≤ 100 ευρώ (ΣΑ) [5 μονάδες]
- iv. Βρείτε τους κωδικούς των ζαχαροπλαστείων που παρασκευάζουν πάνω από 10 γλυκά (SQL) [5 μονάδες]

Απάντηση

i.

Σχεσιακή Άλγεβρα

$\pi_{\text{όνομα}}(\sigma_{\text{είδος} = \text{'πάστα'}} \text{ AND } \tauιμή \leq 3(\Gamma\text{λυκό}))$

SQL

SELECT όνομα

FROM Γλυκό

WHERE είδος **LIKE** 'Πάστα' **AND** τιμή <= 3;

QBE

ΓΑΥΚΟ			
Κωδικός	Όνομα	Είδος	Τιμή
	P.	_m	_n
Συνθήκες			
$_m = \text{'Πάστα'}$ and $_n \leq 3$			

ii.

Σχεσιακή Άλγεβρα

$\pi_{\text{ζαχαροπλαστείο.κωδικός}}(\text{Σόνομα}=\text{'σοκολατίνα'} \text{ AND } \text{πόλη}=\text{'Πάτρα'}) ((\text{Ζαχαροπλαστείο} \bowtie_{\text{κωδικός}=\text{Ζαχαροπλαστείο_κωδικός}} \Delta \text{ιατηρεί}) \bowtie_{\text{γλυκό_κωδικός}=\text{κωδικός}} \Gamma \text{λυκό})$

Εναλλακτικός τρόπος γραφής ερωτήματος σε Σχεσιακή Άλγεβρα

Αν τα κοινά πεδία είχαν το ίδιο όνομα θα έγραφα $\pi_{\text{ζαχαροπλαστείο.κωδικός}}(\text{Σόνομα}=\text{'σοκολατίνα'} \text{ AND } \text{πόλη}=\text{'Πάτρα'}) ((\text{Ζαχαροπλαστείο} * \Delta \text{ιατηρεί}) * \Gamma \text{λυκό})$

SQL

SELECT Ζαχαροπλαστείο.κωδικός

FROM Ζαχαροπλαστείο **INNER JOIN** Διατηρεί **ON** κωδικός= Ζαχαροπλαστείο_κωδικός **INNER JOIN** Γλυκό **ON** Γλυκό_κωδικός= κωδικός

WHERE όνομα **LIKE** 'σοκολατίνα' **AND** πόλη **LIKE** 'Πάτρα';

QBE

ΖΑΧΑΡΟΠΛΑΣΤΕΙΟ		
Κωδικός	Πόλη	Διεύθυνση
P.		
_m	_x	
Συνθήκες		
<u>_x='Πάτρα'</u>		

ΔΙΑΤΗΡΕΙ	
ζαχαροπλαστείο_κωδικός	γλυκό_κωδικός
<u>_m</u>	<u>_n</u>

ΓΛΥΚΟ			
κωδικός	Όνομα	Είδος	Τιμή
<u>_n</u>	<u>_y</u>		
Συνθήκες			
<u>_y='σοκολατίνα'</u>			

iii.

Σγειακή Αλγεβρα

A $\leftarrow \sigma_{\text{Τιμή} \leq 100} (((\text{Ζαχαροπλάστης} \bowtie_{\text{ΑΦΜ}=\text{Ζαχαροπλάστης}_\text{ΑΦΜ} \text{Παρασκευάζει}) \bowtie_{\text{γλυκό}_\text{κωδικός}=\text{Κωδικός}_\text{Γλυκό}) \bowtie_{\text{κωδικός}_\text{γλυκό}_\text{κωδικός}=\text{Διατηρεί})) \bowtie_{\text{Ζαχαροπλαστείο}_\text{κωδικός}=\text{κωδικός}_\text{Ζαχαροπλαστείο})$

B($\text{Ζαχ}_\text{Κωδ}$, $\text{Μέσος}_\text{Μισθός}$) $\leftarrow_{\text{ΖΑΧΑΡΟΠΛΑΣΤΕΙΟ.ΚΩΔΙΚΟΣ}} \text{F}_{\text{AVERAGE}}(\text{ΜΙΣΘΟΣ})(\text{A})$

$\Pi_{\text{Ζαχ}_\text{Κωδ}, \text{Μέσος}_\text{Μισθός}}(\text{B})$

SQL

SELECT Ζαχαροπλαστείο.κωδικός, **AVG**(Μισθός)

FROM Ζαχαροπλάστης **INNER JOIN** Παρασκευάζει **ON** ΑΦΜ=Ζαχαροπλάστης_ΑΦΜ **INNER JOIN** Γλυκό **ON** Γλυκό_κωδικός=κωδικός **INNER JOIN** Διατηρεί **ON** κωδικός=Γλυκό_κωδικός **INNER JOIN** Ζαχαροπλαστείο **ON** Ζαχαροπλαστείο_κωδικός=κωδικός

WHERE Τιμή≤100

GROYP BY Ζαχαροπλαστείο.κωδικός;

iv.

Σγειακή Αλγεβρα

A $\leftarrow ((\text{Ζαχαροπλαστείο} \bowtie_{\text{κωδικός}=\text{Ζαχαροπλαστείο}_\text{κωδικός} \text{Διατηρεί}) \bowtie_{\text{Γλυκό}_\text{κωδικός}=\text{κωδικός}_\text{Γλυκό}))$

B($\text{K}_\text{Z}, \Pi_\Gamma$) $\leftarrow_{\text{Ζαχαροπλαστείο.Κωδικός}} \text{F}_{\text{COUNT}}(\text{Γλυκό}_\text{κωδικός})(\text{A})$

$\Pi_{\text{K}_\text{Z}}(\sigma_{\Pi_\Gamma > 10}(\text{B}))$

Εναλλακτικός τρόπος γραφής Σγειακής Αλγεβρας

A(Κωδικός, Πλήθος) $\leftarrow_{\text{ΖΑΧΑΡΟΠΛΑΣΤΕΙΟ.ΚΩΔΙΚΟΣ}} \text{F}_{\text{COUNT}}(\text{Γλυκό}_\text{κωδικός}) ((\text{Ζαχαροπλαστείο}_\text{κωδικός} * \text{Διατηρεί}) * \text{Γλυκό})$

$\Pi_{\text{Κωδικός}}(\sigma_{\text{Πλήθος} > 10}(\text{A}))$

SQL

SELECT Ζαχαροπλαστείο.κωδικός

FROM Ζαχαροπλαστείο **INNER JOIN** Διατηρεί **ON** Ζαχαροπλαστείο.κωδικός=Διατηρεί.Ζαχαροπλαστείο_κωδικός **INNER JOIN** Γλυκό **ON** Διατηρεί.γλυκό_κωδικός=Γλυκό.κωδικός

GROUP BY Ζαχαροπλαστείο.κωδικός

HAVING COUNT (Γλυκό.Κωδικός)>10;

2.7 Θέμα 3 Ιούνιος 2014 και Ιούνιος 2020

Θεωρείστε το παρακάτω σχεσιακό σχήμα το οποίο διατηρεί πληροφορίες σχετικά με τις αεροπορικές πτήσεις μιας εταιρίας.

- Πτήση(πτ_κωδικός, προορισμός, αφετηρία, απόσταση, ώρα_αναχώρησης, ώρα_άφιξης, τιμή)
- Αεροσκάφος(α_κωδικός, τύπος, απόσταση_πλεύσης)
- Πιλότος(π_κωδικός, π_όνομα, μισθός)
- Πιστοποιημένος(π_κωδικός, α_κωδικός)

Κάθε πτήση έχει ένα μοναδικό κωδικό και κάθε αεροσκάφος έχει επίσης ένα μοναδικό κωδικό. Κάθε πιλότος είναι πιστοποιημένος για κάποιο/κάποια αεροσκάφος/η. Η απόσταση_πλεύσης ορίζεται ως η μέγιστη απόσταση που μπορεί να διανύσει ο συγκεκριμένος τύπος αεροσκάφους.

Εκφράστε τα ακόλουθα ερωτήματα στις γλώσσες επερωτήσεων Σχεσιακή Άλγεβρα (ΣΑ) και SQL “όπως σας ζητείται:

1. Βρείτε τον κωδικό πτήσης για τις πτήσεις που έχουν αφετηρία την Αθήνα και προορισμό τη Νέα Υόρκη
2. Βρείτε τα ονόματα των πιλότων που είναι πιστοποιημένοι για αεροσκάφος τύπου ‘Boeing’
3. Βρείτε το μέσο μισθό των πιλότων που είναι πιστοποιημένοι για αεροσκάφος τύπου ‘Airbus’ και το όνομα τους δεν ξεκινά από Α
4. Για κάθε πιλότο ο οποίος είναι πιστοποιημένος για περισσότερα από 5 αεροσκάφη, βρείτε τον κωδικό πιλότου και τη μέγιστη απόσταση_πλεύσης των αεροσκαφών για τα οποία αυτός/ή είναι πιστοποιημένος/η

Απάντηση

Ερώτημα 1

<p>Σχεσιακή Άλγεβρα</p> <p>Π_{πτ_κωδικός}(Σ_{προορισμός}=‘Νέα Υόρκη’ and Αφετηρία=‘Αθήνα’ (Πτήση))</p>
<p>SQL</p> <p>SELECT πτ_κωδικός FROM πτήση WHERE προορισμός LIKE “Νέα Υόρκη” AND Αφετηρία LIKE ‘Αθήνα’;</p>

Ερώτημα 2

Σχεσιακή Άλγεβρα

$\Pi_{\pi_όνομα}(\sigma_{τύπος='Boeing'} ((Πιλότος \bowtie_{\pi_κωδικός=\pi_κωδικός} Πιστοποιημένος) \bowtie_{\alpha_κωδικός=\alpha_κωδικός} Αεροσκάφος))$

SQL

SELECT Όνομα

FROM πιλότος **INNER JOIN** Πιστοποιημένος **ON** πιλότος.π_κωδικός=πιστοποιημένος.π_κωδικός **INNER JOIN** Αεροσκάφος **ON** Πιστοποιημένος.α_κωδικός=Αεροσκάφος.α_κωδικός

WHERE τύπος **LIKE** 'Boeing';

Ερώτημα 3

SQL

SELECT AVG(Μισθός)

FROM πιλότος **INNER JOIN** Πιστοποιημένος **ON** πιλότος.π_κωδικός=πιστοποιημένος.π_κωδικός **INNER JOIN** Αεροσκάφος **ON** Πιστοποιημένος.α_κωδικός=Αεροσκάφος.α_κωδικός

WHERE τύπος **LIKE** 'airbus **AND** όνομα **NOT LIKE** 'A%'

GROUP BY τύπος;

Ερώτημα 4

SQL

SELECT π_κωδικός, **MAX**(απόσταση_πλεύσης)

FROM πιλότος **INNER JOIN** Πιστοποιημένος **ON** πιλότος.π_κωδικός=πιστοποιημένος.π_κωδικός **INNER JOIN** Αεροσκάφος **ON** Πιστοποιημένος.α_κωδικός=Αεροσκάφος.α_κωδικός

GROUP BY π_κωδικός

HAVING COUNT(α_κωδικός)>5;

2.8 Θέμα 2 Φεβρουάριος 2015

Θεωρείστε το παρακάτω σχεσιακό σχήμα το οποίο διατηρεί πληροφορίες σχετικά με τις αεροπορικές πτήσεις μιας εταιρίας.

- Πτήση (**πτ_κωδικός**, προορισμός, αφετηρία, απόσταση, ώρα_αναχώρησης, ώρα_άφιξης, τιμή)
- Αεροσκάφος (**α_κωδικός**, τύπος, απόσταση_πλεύσης)
- Πιλότος (**π_κωδικός**, π_όνομα, μισθός)
- Πιστοποιημένος (**π_κωδικός**, **α_κωδικός**)

Κάθε πτήση έχει ένα μοναδικό κωδικό, κάθε αεροσκάφος έχει επίσης ένα μοναδικό κωδικό. Κάθε πιλότος είναι πιστοποιημένος για κάποιο (κάποια) αεροσκάφος (αεροσκάφη). Ως απόσταση πλεύσης ορίζεται η μέγιστη απόσταση που μπορεί να διανύσει ο συγκεκριμένος τύπος αεροσκάφους.

Εκφράστε τα ακόλουθα ερωτήματα στις γλώσσες επερωτήσεων Σχεσιακή Άλγεβρα, SQL και QBE όπως σας ζητείται:

1. Βρείτε τον κωδικό πτήσης για τις πτήσεις που έχουν αφετηρία την Αθήνα και προορισμό τη Νέα Υόρκη με ΣΑ, SQL, QBE
2. Βρείτε τα ονόματα των πιλότων που είναι πιστοποιημένοι για κάποιο αεροσκάφος τύπου 'Boeing' με ΣΑ, SQL, QBE
3. Βρείτε το μέσο μισθό των πιλότων που είναι πιστοποιημένοι για το αεροσκάφος τύπου 'Airbus' και το όνομα τους δεν ξεκινά από Α με ΣΑ, SQL
4. Για κάθε πιλότο ο οποίος είναι πιστοποιημένος για περισσότερα από 3 αεροσκάφη, βρείτε τον κωδικό πιλότου και τη μέγιστη απόσταση πλεύσης του αεροσκάφους για το οποίο αυτός (αυτή) είναι πιστοποιημένος (πιστοποιημένη) με SQL
5. Βρείτε τα ονόματα των πιλότων των οποίων ο μισθός είναι μικρότερος από την τιμή της πιο φθηνής διαδρομής από τη Θεσσαλονίκη στη Μόσχα με SQL

Απάντηση

Ερώτημα 1

Σχεσιακή Άλγεβρα

Π_{πτ_κωδικός}(Σ_{προορισμός='Νέα Υόρκη'} and Αφετηρία='Αθήνα' (Πτήση))

SQL

SELECT πτ_κωδικός

FROM πτήση

WHERE προορισμός **LIKE** 'Νέα Υόρκη' **AND** Αφετηρία **LIKE** 'Αθήνα';

QBE

Πτήση						
πτ_κωδικός	Προορισμός	Αφετηρία	Απόσταση	Ωρα Αναχώρησης	Ωρα Αφίξης	Τιμή
P._x	_m	_n				
Συνθήκες						
<u>_n='Αθήνα'</u> and <u>_m='Νέα Υόρκη'</u>						

Ερώτημα 2

Σχεσιακή Άλγεβρα

$\Pi_{\pi_όνομα}(\sigma_{\tauύπος='Boeing'}((Πιλότος \bowtie_{\pi_κωδικός=\pi_κωδικός} Πιστοποιημένος) \bowtie_{\alpha_κωδικός=\alpha_κωδικός} Αεροσκάφος))$
ή εναλλακτικά

$\Pi_{\pi_όνομα}(\sigma_{\tauύπος='Boeing'}((Πιλότος *Πιστοποιημένος)*Αεροσκάφος))$

SQL

SELECT $\pi_όνομα$

FROM πιλότος **INNER JOIN** Πιστοποιημένος **ON** πιλότος. $\pi_κωδικός$ =πιστοποιημένος. $\pi_κωδικός$ **INNER JOIN** Αεροσκάφος **ON** Πιστοποιημένος. $\alpha_κωδικός$ =Αεροσκάφος. $\alpha_κωδικός$ **WHERE** τύπος **LIKE** ‘Boeing’;

QBE

ΠΙΛΟΤΟΣ		
$\pi_κωδικός$	$\pi_όνομα$	μισθός
_x	P._y	

ΠΙΣΤΟΠΟΙΗΜΕΝΟΣ	
$\alpha_κωδικός$	$\pi_κωδικός$
_z	_x

ΑΕΡΟΣΚΑΦΟΣ		
$\alpha_κωδικός$	τύπος	Απόσταση_πλεύσης
_z	_n	
Συνθήκες		
$_n='boeing'$		

Ερώτημα 3

Σχεσιακή Άλγεβρα

$F_{avg}(\muσθός)(\sigma_{\tauύπος='Airbus'}(\text{and } \pi_όνομα \neq 'A'))((Πιλότος *Πιστοποιημένος)*Αεροσκάφος))$

SQL

SELECT AVG(Μισθός)

FROM πιλότος **INNER JOIN** Πιστοποιημένος **ON** πιλότος. $\pi_κωδικός$ =πιστοποιημένος. $\pi_κωδικός$ **INNER JOIN** Αεροσκάφος **ON** Πιστοποιημένος. $\alpha_κωδικός$ =Αεροσκάφος. $\alpha_κωδικός$

WHERE τύπος **LIKE** 'airbus' **AND** όνομα **NOT LIKE** 'A%';

Ερώτημα 4

SQL

SELECT π_κωδικός, **MAX**(απόσταση_πλεύσης)

FROM πιλότος **INNER JOIN** πιστοποιημένος **ON** πιλότος.π_κωδικός=πιστοποιημένος.π_κωδικός **INNER JOIN** Αεροσκάφος **ON** Πιστοποιημένος.α_κωδικός=Αεροσκάφος.α_κωδικός

GROUP BY π_κωδικός

HAVING COUNT (α_κωδικός)>3;

Ερώτημα 5

SELECT π_όνομα **FROM** Πιλότος **WHERE** μισθός<**MIN ALL** (**SELECT** τιμή **FROM** Πτήση **WHERE** αφετηρία **LIKE** 'Θεσσαλονίκη' **AND** προορισμός **LIKE** 'Μόσχα');

Εναλλακτικά:

SELECT π_όνομα **FROM** Πιλότος **WHERE** μισθός < (**SELECT MIN** (τιμή) **FROM** Πτήση **WHERE** αφετηρία **LIKE** 'Θεσσαλονίκη' **AND** προορισμός **LIKE** 'Μόσχα');

2.9 Θέμα 2 Ιούλιος 2015

Θεωρήστε τους παρακάτω πίνακες:

R		
A	B	C
3	5	2
3	7	1
1	NULL	11
2	NULL	17
15	15	2

S		
B	C	E
5	5	7
7	4	6
21	2	6
5	NULL	2

Q	
C	D
4	10
2	1
7	21
2	NULL

Βάσεις Δεδομένων- Computer Ανάλυση

Δώστε το αποτέλεσμα για καθένα από τα παρακάτω SQL ερωτήματα:

- 1) **SELECT * FROM R ORDER BY C desc;**
- 2) **SELECT MAX(C), COUNT(C) AS PARAMETER FROM Q**
- 3) **SELECT E FROM S RIGHT OUTER JOIN R ON S.B = R.B WHERE S.C IS NOT NULL**
- 4) **SELECT Q.C, COUNT (R.B) FROM R INNER JOIN Q ON R.C= Q.C GROUP BY R.A HAVING MAX(R.B)>5**
- 5) **SELECT E FROM S WHERE C IN (SELECT C FROM Q WHERE C>=3);**

Απάντηση

- 1) SELECT * FROM R ORDER BY C desc;

A	B	C
2	NULL	17
1	NULL	11
3	5	2
15	15	2
3	7	1

- 2) SELECT MAX(C), COUNT(C) AS PARAMETER FROM Q

MAX(C)	PARAMETER
7	4

Παρατηρήσεις

- To COUNT μετρά πλήθος
- To COUNT δεν μετρά το NULL
- To COUNT μετρά τα διπλότυπα
- To COUNT(DISTINCT(C)) μετρά τις μοναδικές τιμές

- 3) SELECT E FROM S RIGHT OUTER JOIN R ON S.B = R.B WHERE S.C IS NOT NULL

S RIGHT OUTER JOIN R ON S.B = R.B

S.B	S.C	S.E	R.A	R.B	R.C
5	5	7	3	5	2
5	NULL	2	3	5	2
7	4	6	3	7	1
NULL	NULL	NULL	1	NULL	11
NULL	NULL	NULL	2	NULL	17
NULL	NULL	NULL	15	15	2

WHERE S.C IS NOT NULL

S.B	S.C	S.E	R.A	R.B	R.C
5	5	7	3	5	2

7	4	6	3	7	1
---	---	---	---	---	---

SELECT E

E
7
6

Παρατήρηση

- To RIGHT JOIN και RIGHT OUTER JOIN είναι ίδια
- To LEFT JOIN και LEFT OUTER JOIN είναι ίδια
- Αν δινόταν η εντολή **SELECT E FROM S LEFT OUTER JOIN R ON S.B = R.B WHERE S.C IS NOT NULL**, η απάντηση θα ήταν:

S LEFT OUTER JOIN R ON S.B = R.B

S.B	S.C	S.E	R.A	R.B	R.C
5	5	7	3	5	2
7	4	6	NULL	NULL	NULL
21	2	6	NULL	NULL	NULL
2	NULL	17	NULL	NULL	NULL
15	15	2	NULL	NULL	NULL

WHERE S.C IS NOT NULL

S.B	S.C	S.E	R.A	R.B	R.C
5	5	7	3	5	2
7	4	6	NULL	NULL	NULL
21	2	6	NULL	NULL	NULL
15	15	2	NULL	NULL	NULL

SELECT E

S.E
7
6
6
2

ΠΡΟΣΟΧΗ: ΕΜΦΑΝΙΖΟΝΤΑΙ ΚΑΙ ΤΑ ΔΙΠΛΟΤΥΠΑ ΑΦΟΥ ΔΕΝ ΥΠΑΡΧΕΙ Η ΕΝΤΟΛΗ DISTINCT

4) SELECT Q.C, COUNT (R.B) FROM R INNER JOIN Q ON R.C=Q.C GROUP BY R.A HAVING MAX (R.B)>5

R INNER JOIN Q ON R.C= Q.C

R.A	R.B	R.C	Q.C	Q.D
3	5	2	2	1
3	5	2	2	NULL
15	15	2	2	NULL

Βάσεις Δεδομένων- Computer Ανάλυση

15	15	2	2	1
----	----	---	---	---

GROUP BY R.A

R.A	R.B	R.C	Q.C	Q.D
3	5	2	2	1
3	5	2	2	NULL
15	15	2	2	NULL
15	15	2	2	1

HAVING MAX (R.B)>5

R.A	R.B	R.C	Q.C	Q.D
15	15	2	2	NULL
15	15	2	2	1

SELECT Q.C, COUNT (R.B)

Q.C	COUNT (R.B)
2	2

5) SELECT E FROM S WHERE C IN (SELECT C FROM Q WHERE C>=3);

SELECT C FROM Q WHERE C>=3

C
4
7

SELECT E FROM S WHERE C IN (4, 7)

E
6

2.10 Θέμα 3 Ιούλιος 2016

Εστω μια βάση δεδομένων με τις παρακάτω σχέσεις:

- **Προϊόν** (ID, όνομα, τιμή, περιγραφή)
- **Πελάτης** (ID, όνομα, πόλη, τηλέφωνο)
- **Πώληση** (Πελάτης_ID, Προϊόν_ID, ημερομηνία, ποσότητα)

Να γράψετε σε SQL και σχεσιακή άλγεβρα τα παρακάτω ερωτήματα:

1. Τα ονόματα των προϊόντων που κοστίζουν από 10 έως και 30 Ευρώ
2. Τα ονόματα των πελατών που έχουν αγοράσει κάποιο προϊόν με ποσότητα άνω των 20 τεμαχίων
3. Τα ονόματα και τις τιμές των προϊόντων που δεν έχουν αγοραστεί μετά την ημερομηνία 25/03/2016
4. Τα ονόματα των πελατών που έχουν αγοράσει προϊόντα συνολικής αξίας άνω των 100 ευρώ

Απάντηση

- 1.

SQL: *SELECT* όνομα **FROM** Προϊόν **WHERE** τιμή **BETWEEN** 10 **AND** 30;

Σχεσιακή Άλγεβρα: Πόνομα ($\sigma_{\text{τιμή} \geq 10 \text{ and } \text{τιμή} \leq 30}$ (Προϊόν))

- 2.

SQL: *SELECT* όνομα **FROM** Πελάτης **INNER JOIN** Πώληση **ON** id=Πελάτης_id **WHERE** ποσότητα>20;

Σχεσιακή Άλγεβρα: Πόνομα($\sigma_{\text{ποσότητα} > 20}$ (Πελάτης $\bowtie_{\text{id}=\text{Πελάτης}_\text{id}$ Προϊόν) Πώληση))

- 3.

SQL: *SELECT* όνομα, τιμή **FROM** Προϊόν **WHERE** id **NOT IN** (*SELECT* id **FROM** Προϊόν **INNER JOIN** Πώληση **ON** id=Προϊόν_id **WHERE** ημερομηνία >='2016-3-25');

Σχεσιακή Άλγεβρα: Πόνομα, τιμή (Προϊόν) - Πόνομα, τιμή ($\sigma_{\text{ημερομηνία} \geq '2016-3-25'}$ (Προϊόν $\bowtie_{\text{id}=\text{Προϊόν}_\text{id}$ Πώληση)))

- 4.

SELECT πελάτης.όνομα **FROM** Πελάτης **INNER JOIN** Πώληση **ON** Πελάτης.id=Πώληση.Πελάτης_id **INNER JOIN** Προϊόν **ON** Προϊόν_id=Προϊόν.id **GROUP BY** πελάτης.id **HAVING SUM** (τιμή)>100;

Σχεσιακή Άλγεβρα:

A \leftarrow (Πελάτης $\bowtie_{\text{id}=\text{Πελάτης}_\text{id}$ Πώληση) $\bowtie_{\text{προϊόν}_\text{id}=\text{id}}$ ΠΡΟΙΟΝ

B (ONOMA, SUM(TIMH)) $\leftarrow_{\pi\text{ελάτης.όνομα}F \text{ SUM}(\text{τιμή})(A)$

Πόνομα($\sigma_{> \text{SUM}(\text{τιμή}) > 100}$ (B))

2.11 Θέμα 6 Σεπτέμβριος 2009

Χρησιμοποιώντας το παρακάτω σχήμα βάσης, να μετατρέψετε σε γλώσσα QBE κάθε ένα από τα παρακάτω ερωτήματα SQL:



- 1) select distinct ΟΝΟΜΑ from ΠΟΤΗΣ where ID>5;
- 2) select ΠΟΤΗΣ.ΟΝΟΜΑ from ΠΙΝΕΙ inner join ΠΟΤΗΣ on ΠΟΤΗΣ.ID=ΠΙΝΕΙ.ID where ΠΟΣΟΤΗΤΑ≥3;

Απάντηση

1)

ΠΟΤΗΣ	
ID	ΟΝΟΜΑ
>5	P._x

ή

ΠΟΤΗΣ	
ID	ΟΝΟΜΑ
>5	P.

ή

ΠΟΤΗΣ	
ID	ΟΝΟΜΑ
>5	P.UNQ_x

ή

ΠΟΤΗΣ	
ID	ΟΝΟΜΑ
_n	P._x
Συνθήκες	
_n>5	

Παρατήρηση

Το QBE κάνει αυτόματη απαλοιφή διπλοτύπων.

2)

ΠΟΤΗΣ	
ID	ΟΝΟΜΑ
_n	P.

ΠΙΝΕΙ		
ID	ΚΩΔ	ΠΟΣΟΤΗΤΑ
_n		>=3

ΠΟΤΗΣ	
ID	ΟΝΟΜΑ
_n	P.

ΠΙΝΕΙ		
ID	ΚΩΔ	ΠΟΣΟΤΗΤΑ
_n		_m
Συνθήκη		
$_m >= 3$		

2.12 Θέμα 2 Σεπτέμβριος 2017

Θεωρείστε το παρακάτω σχεσιακό σχήμα το οποίο αφορά πληρωμές για αγορές εξαρτημάτων με βάσει τους τιμοκαταλόγους εξαρτημάτων από διάφορους προμηθευτές.

Suppliers (sid: integer, sname: string, city: string, address: string)

Parts (pid: integer, pname: string, color: string)

Catalog (sid: integer, pid: integer, cost: real)

a) Να γράψετε 1) σε σχεσιακή άλγεβρα και 2) σε SQL τα παρακάτω ερωτήματα:

- Τα ονόματα των προμηθευτών που διαθέτουν εξαρτήματα σε χρώμα 'κόκκινο'
- Τα ονόματα και τις τιμές των εξαρτημάτων με χρώμα 'πράσινο' που διαθέτουν προμηθευτές από την 'Πάτρα'

b) Για την SQL έκφραση:

```
SELECT C1.pid
FROM Catalog C1
WHERE EXISTS (SELECT *
               FROM Catalog C2
              WHERE C1.pid=C2.pid);
```

Περιγράψτε με σαφήνεια το ερώτημα που εξυπηρετεί. Ποια είναι η αντίστοιχη έκφραση σε σχεσιακή άλγεβρα;

Απάντηση

a1.

Σχεσιακή Άλγεβρα

$\Pi_{sname}(\sigma_{color='κόκκινο'}(Suppliers * Catalog) * Parts)$

SQL

SELECT sname

FROM Suppliers, Catalog, Parts

WHERE Suppliers.sid = Catalog.sid **AND** Catalog.pid = Parts.pid **AND** color **LIKE** 'κόκκινο';

Βάσεις Δεδομένων- Computer Ανάλυση

Εναλλακτικά η προηγούμενη εντολή γράφεται ως εξής:

SELECT sname

FROM Suppliers **INNER JOIN** Catalog **ON** Suppliers.sid = Catalog.sid **INNER JOIN** Parts **ON** Catalog.pid = Parts.pid

WHERE color **LIKE** 'κόκκινο';

Computer Ανάλυση

a2.

Σχεσιακή Άλγεβρα

$\Pi_{\text{pname}, \text{cost}} (\sigma_{\text{color}=\text{'πράσινο'}} \text{ and } \text{city}=\text{'Πάτρα'}) ((\text{Suppliers} * \text{Catalog}) * \text{Parts})$

SQL

SELECT pname, cost

FROM Suppliers, Catalog, Parts

WHERE Suppliers.sid = Catalog.sid **AND** Catalog.pid = Parts.pid **AND** color **LIKE** 'πράσινο' **AND** city **LIKE** 'Πάτρα';

Εναλλακτικά η προηγούμενη εντολή γράφεται ως εξής:

SELECT pname, cost

FROM Suppliers **INNER JOIN** Parts **ON** Suppliers.sid = Catalog.sid **INNER JOIN** Parts **ON** Catalog.pid = Parts.pid

WHERE color **LIKE** 'πράσινο' **AND** city **LIKE** 'Πάτρα';

b) Αρχικά εκτελούμε το υποερώτημα που αφορά τη συσχέτιση του πίνακα Catalog με τον εαυτό του όπως στο παράδειγμα:

Catalog (sid: integer, pid: integer, cost: real)

C1 **INNER JOIN** C2 **ON** C1.pid=C2.pid

sid	pid	cost
100	1	20
200	1	15
200	2	30
300	3	40

C1.sid	C1.pid	C1.cost	C2.sid	C2.pid	C2.cost
100	1	20	100	1	20
100	1	20	200	1	15
200	1	15	100	1	20
200	1	15	200	1	15
200	2	30	200	2	30
300	3	40	300	3	40

Parts (pid: integer, pname: string, color: string)

pid	name	color
1	DVD	black
2	CD	red
3	Flash disk	black
4	Monitor	white
5	Printer	red

Συμπέρασμα: Το ερώτημα εμφανίζει κωδικούς εξαρτημάτων που προμηθεύονται από τουλάχιστον ένα προμηθευτή.

b) Η αντίστοιχη έκφραση σε σχεσιακή άλγεβρα είναι:

$\Pi_{\text{c1.pid}} (\rho_{\text{c1}}(\text{Catalog}) \bowtie_{\text{c1.pid}=\text{c2.pid}} \rho_{\text{c2}}(\text{Catalog}))$

2.13 Θέμα 2 Ιούνιος 2017

Θεωρείστε το παρακάτω σχεσιακό σχήμα το οποίο αφορά πληρωμές για αγορές εξαρτημάτων με βάσει τους τιμοκαταλόγους εξαρτημάτων από διάφορους προμηθευτές.

Suppliers (**sid: integer**, sname: string, city: string, address: string)

Parts (**pid: integer**, pname: string, color: string)

Catalog (**sid: integer, pid: integer**, cost: real)

a) Να γράψετε 1) σε σχεσιακή άλγεβρα και 2) σε SQL τα παρακάτω ερωτήματα:

- Τα ονόματα των προμηθευτών που διαθέτουν εξαρτήματα σε χρώμα 'κόκκινο'
- Τα ονόματα και τις τιμές των εξαρτημάτων με χρώμα 'πράσινο' που διαθέτουν προμηθευτές από την 'Πάτρα'

b) Για την SQL έκφραση:

SELECT C1.pid

FROM Catalog C1

WHERE EXISTS (SELECT *

FROM Catalog C2

WHERE C1.pid=C2.pid **AND** C1.sid≠C2.sid);

- Περιγράψτε με σαφήνεια το ερώτημα που εξυπηρετεί
- Ποια είναι η αντίστοιχη έκφραση σε σχεσιακή άλγεβρα

Απάντηση

a1.

Σχεσιακή Άλγεβρα

$\Pi_{sname}(\sigma_{color='κόκκινο'}((Suppliers * Catalog) * Parts))$

SQL

SELECT sname

FROM Suppliers, Catalog, Parts

WHERE Suppliers.sid = Catalog.sid **AND** Catalog.pid = Parts.pid **AND** color **LIKE** 'κόκκινο';

a2.

Σχεσιακή Άλγεβρα

$\Pi_{pname, cost}(\sigma_{color='πράσινο' \text{ and } city='Πάτρα'}((Suppliers * Catalog) * Parts))$

SQL

SELECT pname, cost

FROM Suppliers, Catalog, Parts

WHERE Suppliers.sid = Catalog.sid **AND** Catalog.pid = Parts.pid **AND** color **LIKE** 'πράσινο' **AND** city **LIKE** 'Πάτρα';

Απάντηση

a) Έστω για παράδειγμα ότι έχουμε τους ακόλουθους πίνακες Catalog (τυχαία παραδείγματα)

Πίνακας 1 - Catalog

sid	pid	cost
1	100	250
1	200	150
2	100	60
2	300	120
3	100	280
3	300	140

Πίνακας 2 - Catalog

sid	pid	cost
1	100	250
2	200	150
3	300	120

Αν κάνουμε την αυτοσυσχέτιση στον Πίνακα 1 προκύπτουν τα ακόλουθα αποτελέσματα:

c1.sid	c1.pid	c1.cost	c2.sid	c2.pid	c1.cost
1	100	250	2	100	60
1	100	250	3	100	280
2	100	60	1	100	250
2	6	120	3	100	280
3	100	280	1	100	250
3	100	280	2	100	60
3	300	140	2	300	120

Αν κάνουμε την αυτοσυσχέτιση στον Πίνακα 2 ΔΕΝ ΠΡΟΚΥΠΤΕΙ ΚΑΜΜΙΑ ΕΓΓΡΑΦΗ:

Συμπέρασμα: Το συνολικό ερώτημα εμφανίζει τα parts που προμηθεύονται από τουλάχιστον 2 διαφορετικούς προμηθευτές

b) Η αντίστοιχη έκφραση σε σχεσιακή άλγεβρα είναι:

$$\Pi_{c1.pid}(\rho_{c1}(\text{Catalog}) \bowtie_{c1.sid=c2.sid \text{ and } c1.pid=c2.pid} \rho_{c2}(\text{Catalog}))$$

2.14 Θέμα 3 Ιανουάριος 2019

Θεωρείστε το παρακάτω σχεσιακό σχήμα το οποίο αφορά πληρωμές για αγορές εξαρτημάτων με βάσει τους τιμοκαταλόγους εξαρτημάτων από διάφορους προμηθευτές.

Suppliers (scode: integer, sname: string, city: string, address: string, city: string)

Parts (PCODE: integer, pname: string, color: string)

Catalog (scode: integer, pcode: integer, cost: real)

a) Να γράψετε σε 1)SQL και 2) σχεσιακή άλγεβρα τα παρακάτω ερωτήματα:

i) Τα ονόματα των προμηθευτών που διαθέτουν εξαρτήματα με χρώμα 'κόκκινο'

ii) Τα ονόματα και τις τιμές των εξαρτημάτων με χρώμα 'κόκκινο' ή 'πράσινο' τα οποία διαθέτουν προμηθευτές από την 'Πάτρα'

b) Για την SQL έκφραση:

SELECT C1.pcode

FROM Catalog C1

WHERE EXISTS (SELECT C2.pcode

FROM Catalog C2

WHERE C1.pcode=C2.pcode AND C1.scode ≠ C2.scode);

i) Απαντήστε στο γραπτό σας, ποιο ερώτημα σε φυσική γλώσσα αφορά η παραπάνω έκφραση ξεκινώντας με το "Ποιοι είναι οι κωδικοί εξαρτημάτων στον τιμοκατάλογο που _____"

ii) Ποια είναι η αντίστοιχη έκφραση σε σχεσιακή άλγεβρα;

Απάντηση

a) i)

Σχεσιακή Άλγεβρα

$\Pi_{sname}(\sigma_{color='κόκκινο'}(Suppliers * Catalog) * Parts)$

SQL

SELECT sname

FROM Suppliers, Catalog, Parts

WHERE Suppliers.sid = Catalog.sid AND Catalog.pid = Parts.pid AND color **LIKE 'κόκκινο'**

a) ii)

Σχεσιακή Άλγεβρα

$\Pi_{pname, cost}(\sigma_{πόλη='Πάτρα' \text{ and } color='κόκκινο' \text{ or } color='πράσινο'}(Suppliers * Catalog) * Parts)$

SQL

Βάσεις Δεδομένων- Computer Ανάλυση

SELECT pname, cost

FROM Suppliers, Catalog, Parts

WHERE suppliers.scode = Catalog.sid **AND** catalog.pcode = parts.pcode **AND** Πόλη *like* 'Πάτρα'**AND** color **LIKE** 'κόκκινο' **OR** color **LIKE** 'πράσινο'

b) i)

Ποιοι είναι οι κωδικοί των εξαρτημάτων στον τιμοκατάλογο των εξαρτημάτων που προμηθεύονται από τουλάχιστον 2 διαφορετικούς προμηθευτές

b) ii)

Η έκφραση σε σχεσιακή άλγεβρα: $\Pi_{c1.pcode}(\rho_{c1}(\text{Catalog}) \bowtie_{c1.scode=c2.scode \text{ and } c1.pcode=c2.pcode} \rho_{c2}(\text{Catalog}))$

2.15 Θέμα2 Ιανουάριος 2020

Θεωρήστε τους παρακάτω πίνακες:

R			S		T		
A	B	C	F	B	D	A	F
32	2	4		16	'C'	2	11
65	88	2		16	'A'	65	16
65	2	5		23	'U'	9	NULL
7	3	4		11	'X'	66	9
9	12	2		5			
				34			
				9			

Δώστε το αποτέλεσμα για καθένα από τα παρακάτω ερωτήματα:

- 1) **SELECT B FROM R WHERE C<5 ORDER BY A DESC;**
- 2) **SELECT DISTINCT B FROM R LEFT OUTER JOIN S ON R.B=S.B WHERE R.C>3;**
- 3) **SELECT SUM(A) AS ATHROISMA, COUNT(DISTINCT B) AS PLITHOS FROM R;**
- 4) **Π_{D,A}(T)**
- 5) **Π_{C(σ_{B>2 AND B<15(R)}}**

Θεωρείστε το παρακάτω σχεσιακό σχήμα το οποίο αφορά πληροφορίες που διατηρεί μια εταιρεία για τις πωλήσεις των προϊόντων της στους πελάτες της

Προϊόν (ID, όνομα, τιμή, περιγραφή)

Πελάτης (ID, όνομα, τηλέφωνο)

Πώληση (Πελάτης ID, Προϊόν ID, ημερομηνία, ποσότητα)

Να γράψετε σε SQL τα παρακάτω ερωτήματα:

- 6) Τα ονόματα των πελατών που έχουν αγοράσει κάποιο προϊόν που κοστίζει πάνω από Y Ευρώ
- 7) Τα ονόματα και τις τιμές των προϊόντων που δεν έχουν αγοραστεί
- 8) Τα ονόματα των πελατών που έχουν αγοράσει προϊόντα συνολικής αξίας άνω των Z Ευρώ

Απάντηση

- 1) **SELECT B FROM R WHERE C<5 ORDER BY A DESC;**

B	A
88	65
2	32
12	9
3	7

Βάσεις Δεδομένων- Computer Ανάλυση

Το τελικό αποτέλεσμα είναι:

B
88
2
12
3

2) **SELECT DISTINCT B FROM R LEFT OUTER JOIN S ON R.B=S.B WHERE R.C>3;**

R LEFT OUTER JOIN S ON R.B=S.B

A	R.B	C	F	S.B
32	2	4	NULL	NULL
65	88	2	16	88
65	88	2	5	88
65	2	5	NULL	NULL
7	3	4	16	3
7	3	4	11	3
9	12	2	NULL	NULL

WHERE R.C>3

A	R.B	C	F	B
7	3	4	16	3
7	3	4	11	3
32	2	4	NULL	NULL
65	2	5	NULL	NULL

SELECT DISTINCT B

R.B
2
3

ΑΡΑ ΤΕΛΙΚΗ ΑΠΑΝΤΗΣΗ

R.B
2
3

3) **SELECT SUM(A) AS ATHROISMA, COUNT(DISTINCT B) AS PLITHOS FROM R;**

ATHROISMA	PLITHOS
178	4

Παρατήρηση

Η συνάρτηση count υπολογίζει πλήθος μετρώντας και τα διπλότυπα

Η συνάρτηση count distinct υπολογίζει πλήθος χωρίς τα διπλότυπα

Βάσεις Δεδομένων- Computer Ανάλυση
Υποθετική Ερώτηση

SELECT SUM(A) AS ATHROISMA, COUNT(DISTINCT B) AS PLITHOS FROM R GROUP BY A;

A	B	C
32	2	4
65	88	2
65	2	5
7	3	4
9	12	2

ATHROISMA	PLITHOS
32	1
130	1
7	1
9	1

4) $\Pi_{D,A}(T)$

D	A
'C'	2
'A'	65
'U'	9
'X'	66

5) $\Pi_C(\sigma_{B>2 \text{ AND } B<15}(R))$

C
2
4

6) Τα ονόματα των πελατών που έχουν αγοράσει κάποιο προϊόν που κοστίζει πάνω από Y Ευρώ

SELECT Πελάτης.όνομα

FROM Πελάτης **INNER JOIN** Πώληση **ON** Πελάτης.ID=πελάτης_ID **INNER JOIN** Προϊόν **ON** Προϊόν_ID=Προϊόν.ID

WHERE τιμή > Y

7) Τα ονόματα και τις τιμές των προϊόντων που δεν έχουν αγοραστεί

SELECT Προϊόν.Όνομα, τιμή

FROM Προϊόν

WHERE Προϊόν.id **NOT IN** (**SELECT** προϊόν_ID **FROM** Πώληση);

8) Τα ονόματα των πελατών που έχουν αγοράσει προϊόντα συνολικής αξίας άνω των Z Ευρώ

SELECT Πελάτης.Όνομα

FROM Προϊόν **INNER JOIN** Πώληση **ON** Προϊόν.ID=Πώληση.Προϊόν_ID **INNER JOIN** Πελάτης **ON** Πώληση.Πελάτης_ID=Πελάτης.ID

GROUP BY Πελάτης.ID

HAVING SUM(τιμή)>Z

2.16 Θέμα 2 Φεβρουάριος 2018

Έστω μια βάση δεδομένων με τις παρακάτω σχέσεις:

- **Προϊόν** (ID, όνομα, τιμή, περιγραφή)
- **Πελάτης** (ID, όνομα, πόλη, τηλέφωνο)
- **Πώληση** (Πελάτης_ID, Προϊόν_ID, ημερομηνία, ποσότητα)

Να γράψετε σε SQL και σχεσιακή άλγεβρα τα παρακάτω ερωτήματα:

1. Τα ονόματα των πελατών που έχουν πραγματοποιήσει αγορές
2. Τα ονόματα των προϊόντων που δεν έχουν πουληθεί
3. Τα ονόματα των πελατών που έχουν αγοράσει προϊόντα συνολικής αξίας άνω των Z ευρώ

Απάντηση

Σχεσιακή Άλγεβρα

1 .Πόνομα(**Πελάτης** $\bowtie_{id=\pi\epsilon\lambda\alpha\tauη\varsigma_ID}$ **Πώληση**)

2. Πόνομα(**Προϊόν**) – Πόνομα(**Προϊόν** $\bowtie_{id=\pi\roio\varsigma_ID}$ **Πώληση**)

3. $A \leftarrow ((\text{Πελάτης} \bowtie_{id=\pi\epsilon\lambda\alpha\tauη\varsigma_ID} \text{Πώληση}) \bowtie_{\pi\roio\varsigma_ID=id} \text{Προϊόν})$

$B(\text{πελάτης.όνομα}, \text{συνολική_τιμή}) \leftarrow_{\pi\epsilon\lambda\alpha\tauη\varsigma_όνομα} (\text{F}_{sum(\text{τιμή})}(A))$

$\pi_{\text{πελάτης.όνομα}}(\sigma_{\text{συνολική_τιμή}>Z}(B))$

SQL

a.

SELECT Πελάτης.όνομα

FROM Πελάτης **INNER JOIN** Πώληση **ON** id=πελάτης_ID;

β.

SELECT Όνομα

FROM Προϊόν

WHERE id **NOT IN** (**SELECT** προϊόν_ID **FROM** Πώληση);

SELECT Πελάτης.Όνομα

FROM Προϊόν, Πελάτης, Πώληση

Βάσεις Δεδομένων- Computer Ανάλυση

WHERE Προϊόν.ID= Πώληση.προϊόν_ID **AND** Πώληση.Πελάτης.ID= Πελάτης.ID

GROUP BY Πελάτης.ID

HAVING SUM(τιμή)>Z

Computer Analytics

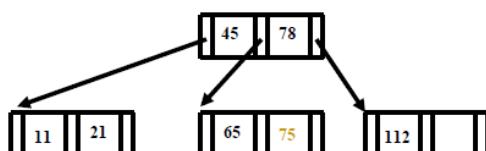
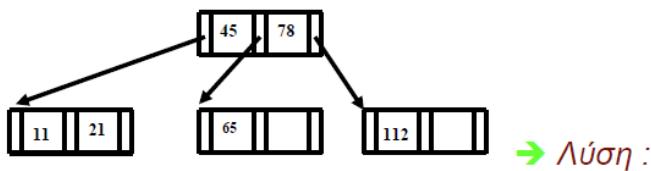
3 ΘΕΜΑΤΑ ΜΕ BTREES

Θεωρία για BTrees

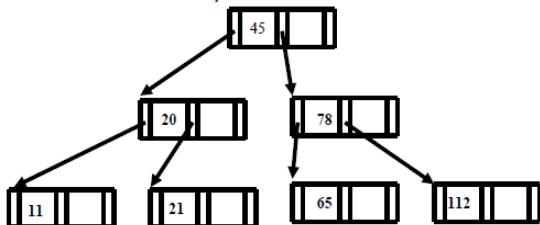
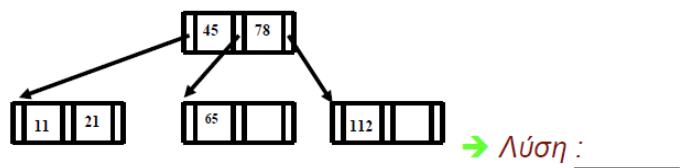
- Κάθε κόμβος έχει τόσα κλειδιά όσα τα παιδιά του μείον 1
- Ο τρόπος που τοποθετούνται τα κλειδιά είναι ΑΠ <Ρίζα <ΔΠ
- Όλα τα φύλλα βρίσκονται στο ίδιο επίπεδο

3.1 Παραδείγματα Εισαγωγών και Διαγραφών σε BTree

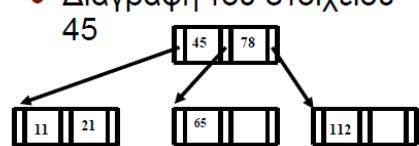
- Εισαγωγή του στοιχείου 75



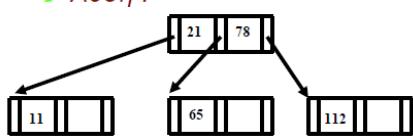
- Εισαγωγή του στοιχείου 20



- Διαγραφή του στοιχείου 45

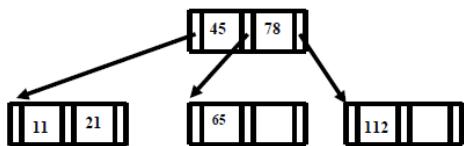


→ Λύση :

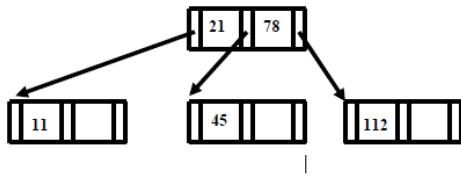


- Επέλεξε το μεγαλύτερο κλειδί από το αριστερότερο υπόδεντρο (ή το μικρότερο από το δεξί υπόδεντρο)

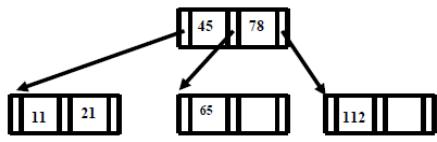
- Διαγραφή του στοιχείου 65



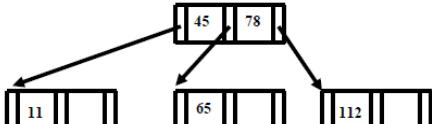
→ Λύση :



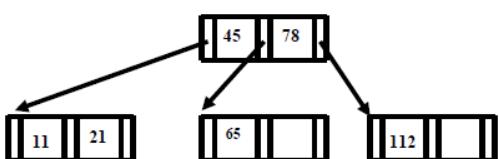
- Διαγραφή του στοιχείου 21



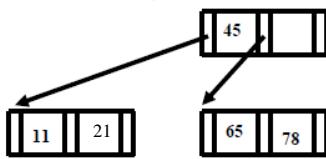
→ Λύση :



- Διαγραφή του στοιχείου 112



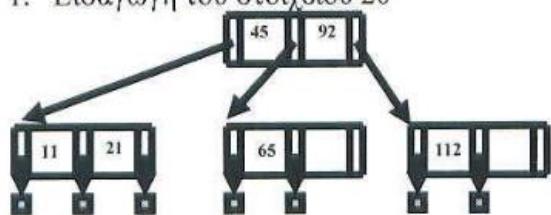
→ Λύση :



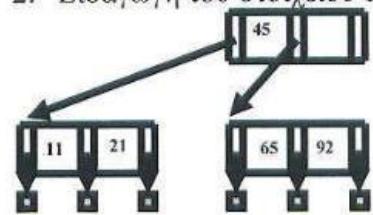
3.2 Θέμα 7 Φεβρουάριος 2008

Σε κάθε υποερώτημα της παρούσας άσκησης σας δίνεται ένα στιγμιότυπο της ανάπτυξης ενός B-tree. Εσείς καλείστε να σχεδιάσετε το δέντρο που θα προκύψει μετά από μία πράξη εισαγωγής ή διαγραφής.

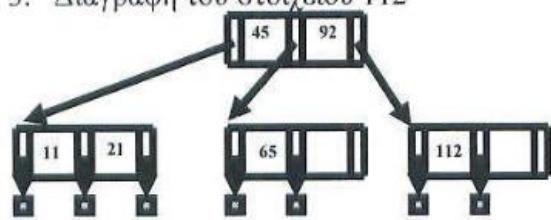
1. Εισαγωγή του στοιχείου 20



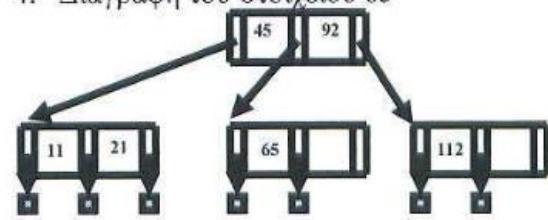
2. Εισαγωγή του στοιχείου 185



3. Διαγραφή του στοιχείου 112

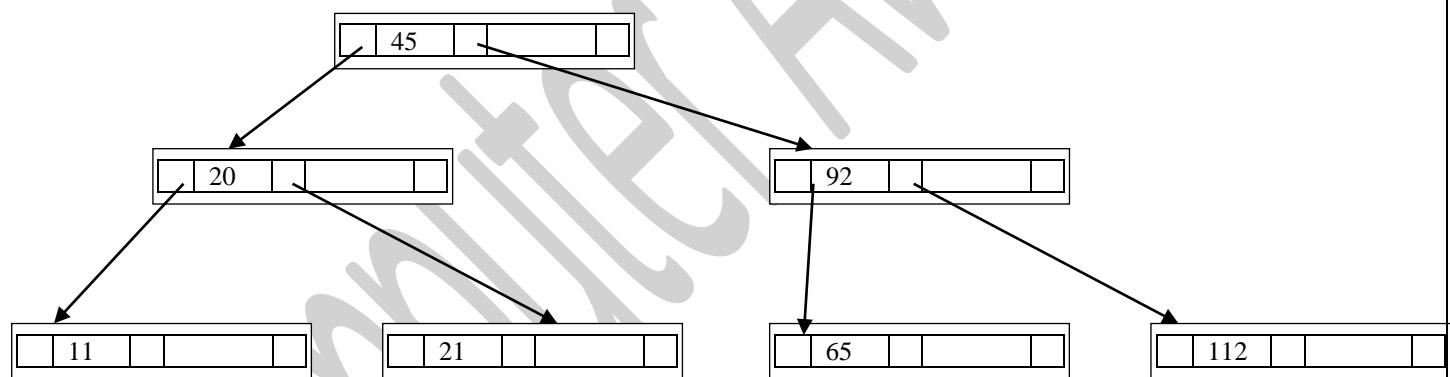


4. Διαγραφή του στοιχείου 65

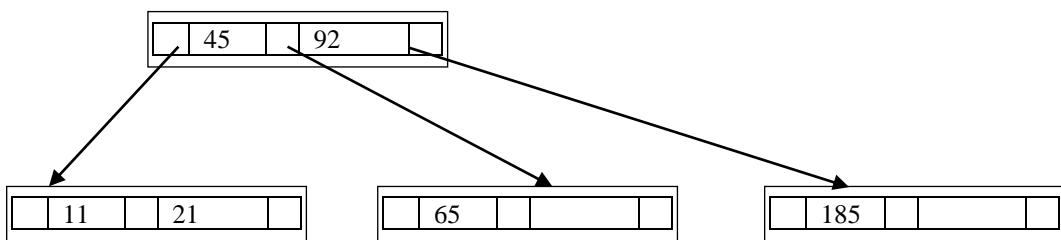


Απάντηση

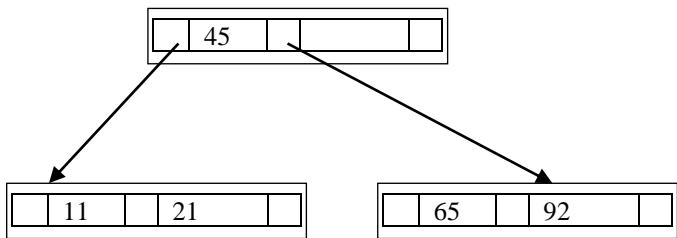
a)



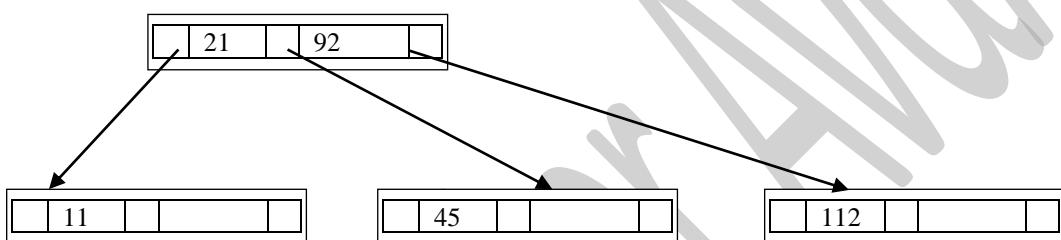
β)



γ)

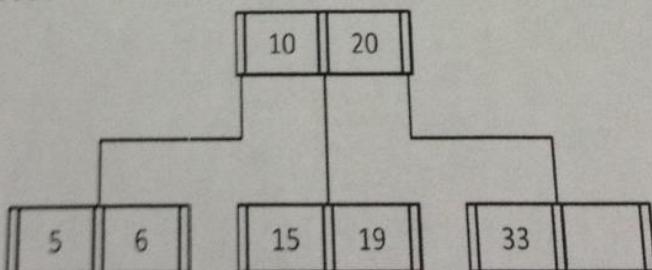


δ)



3.3 Θέμα 5 Φεβρουάριος 2012

Δίνεται το παρακάτω B-Tree

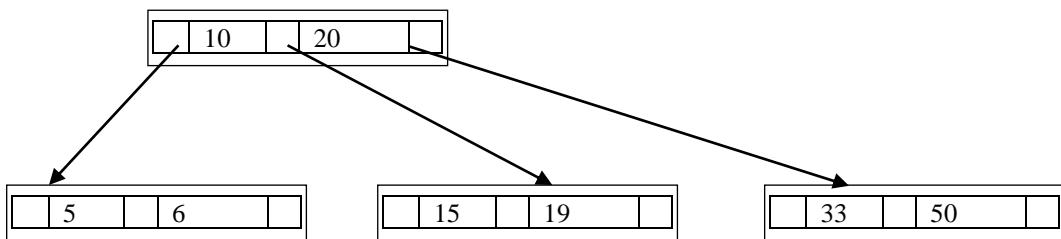


Παρουσιάστε πως τροποποιείται το δένδρο εφαρμόζοντας τις παρακάτω πράξεις (κάθ στο αρχικό):

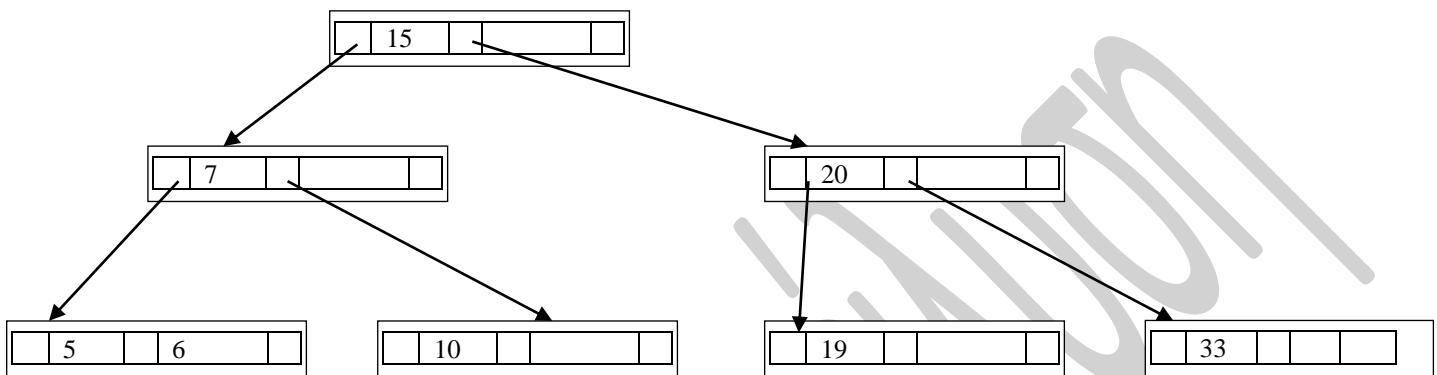
- Εισαγωγή του 50
- Εισαγωγή του 7
- Διαγραφή του 19
- Διαγραφή του 33
- Διαγραφή του 20

Απάντηση

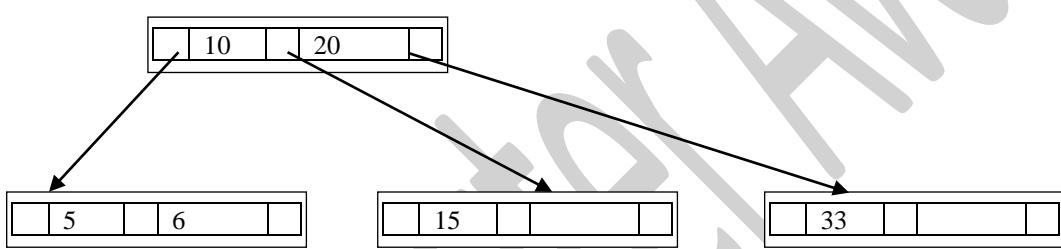
- Εισαγωγή 50



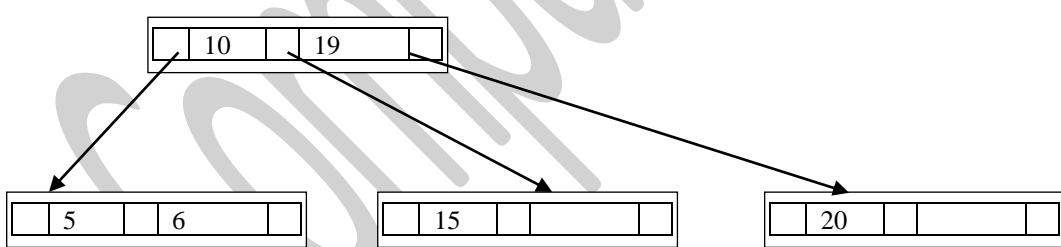
β) Εισαγωγή 7



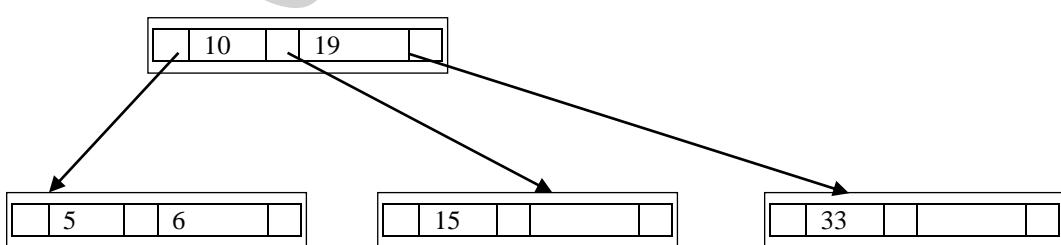
γ) Διαγραφή 19



δ) Διαγραφή 33

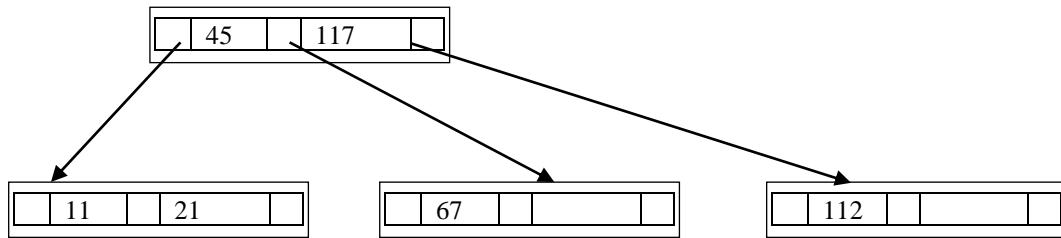


ε) Διαγραφή 20



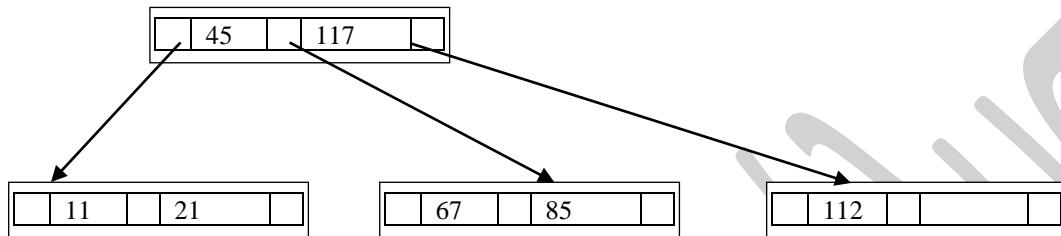
3.4 Θέμα 8 Φεβρουάριος 2013

α) Δίνεται το δέντρο

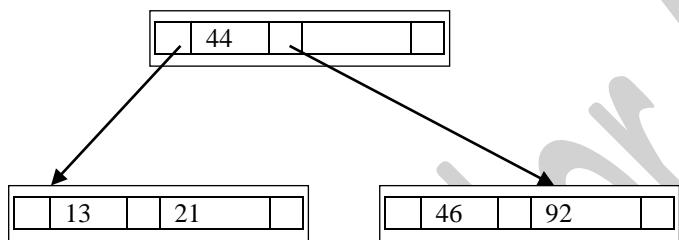


Να γίνει Εισαγωγή του 85

Απάντηση

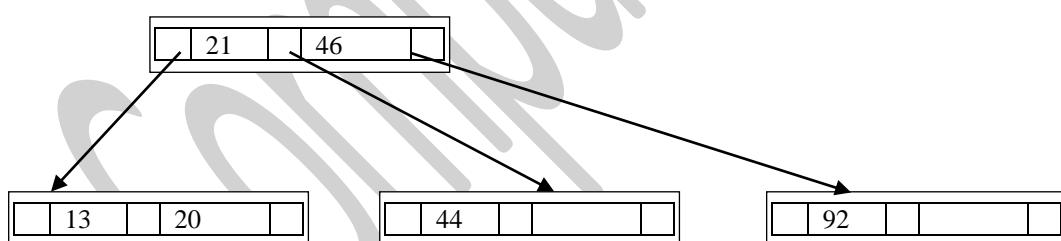


β) Δίνεται το Δέντρο

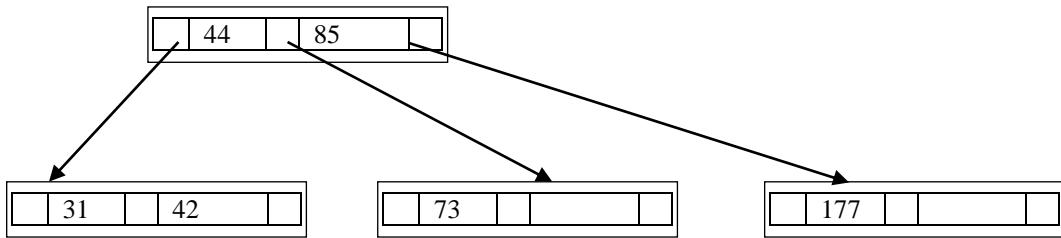


Να γίνει Εισαγωγή του 20

Απάντηση

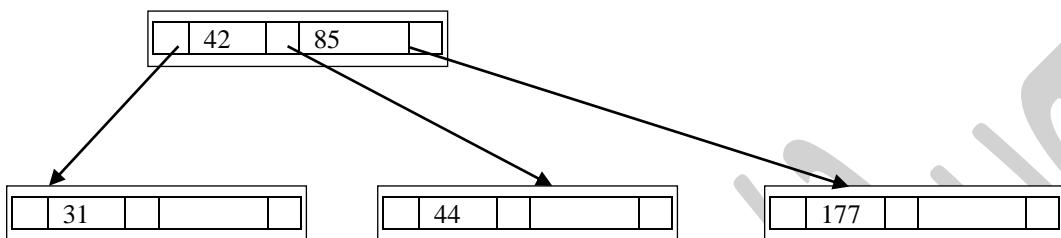


γ) Δίνεται το δέντρο:

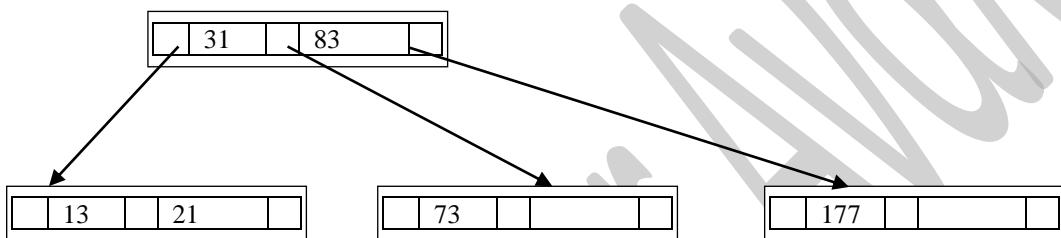


Να γίνει Διαγραφή του 73

Απάντηση

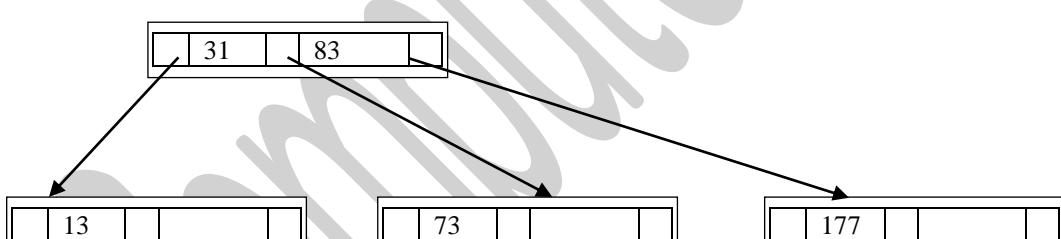


δ) Δίνεται το δέντρο:



Να γίνει διαγραφή του 21

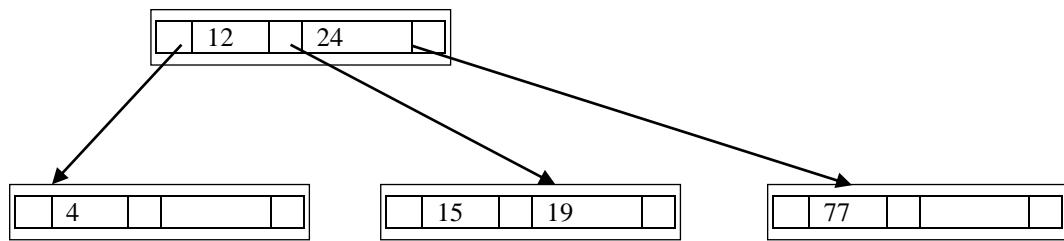
Απάντηση



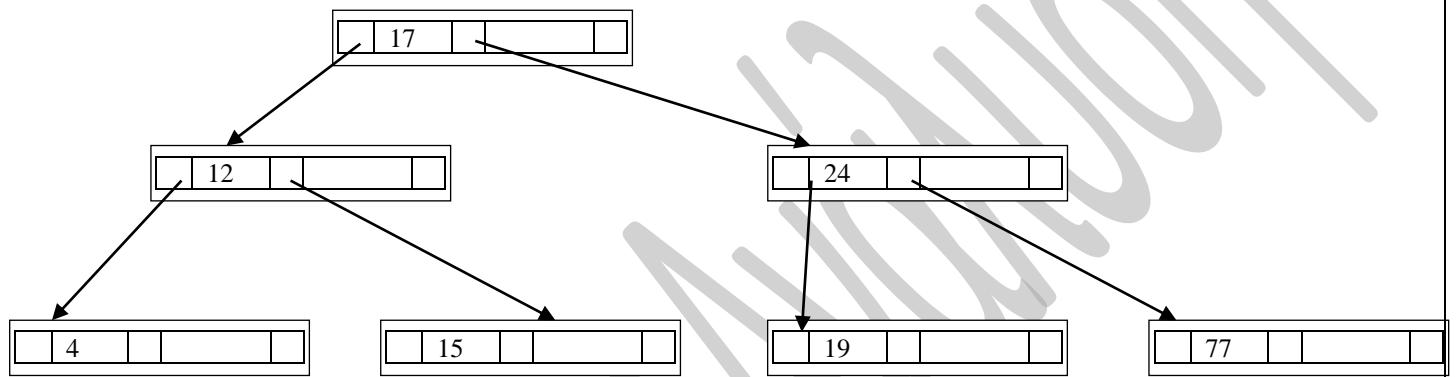
3.5 Θέμα 5 Φεβρουάριος 2013

Α ερώτημα

Αρχικό δέντρο

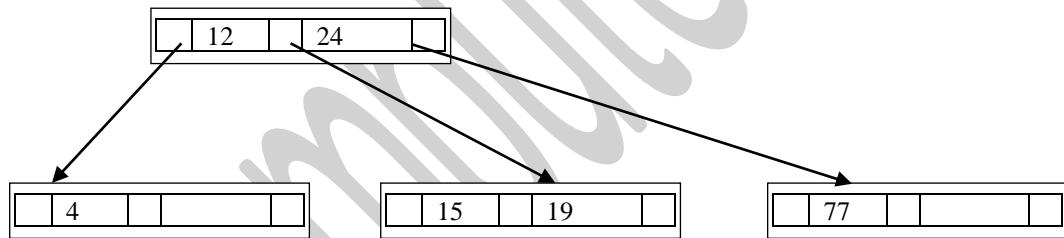


Εισαγωγή 17

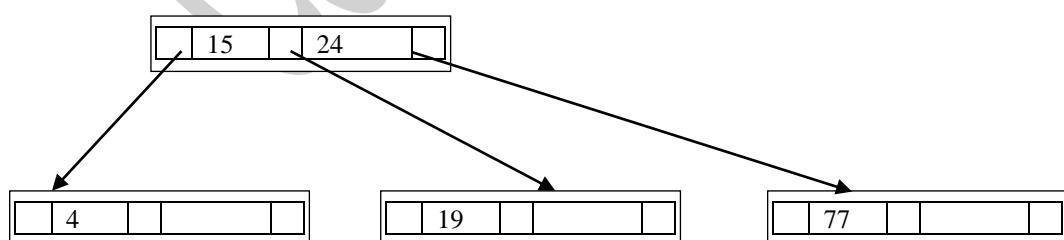


Β ερώτημα

Αρχικό δέντρο

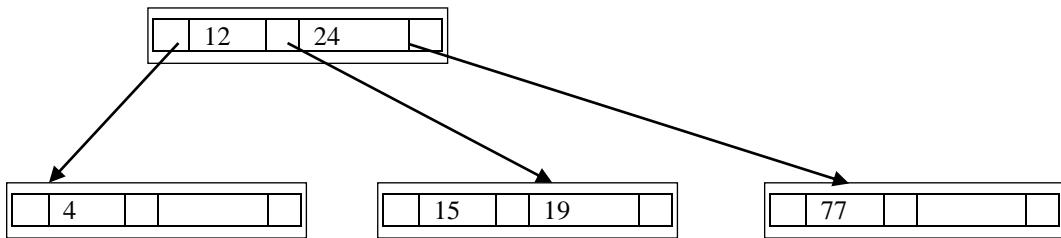


Διαγραφή 12

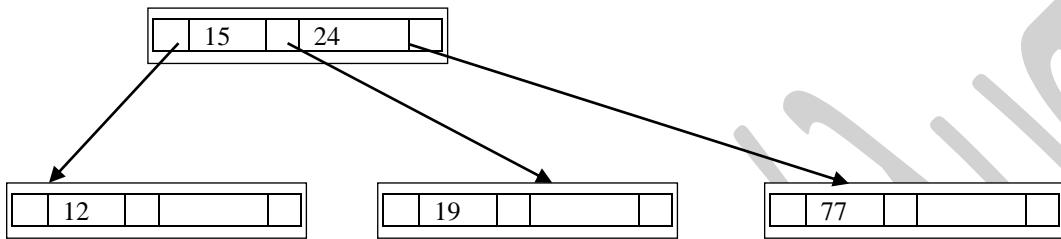


Γ ερώτημα

Αρχικό δέντρο

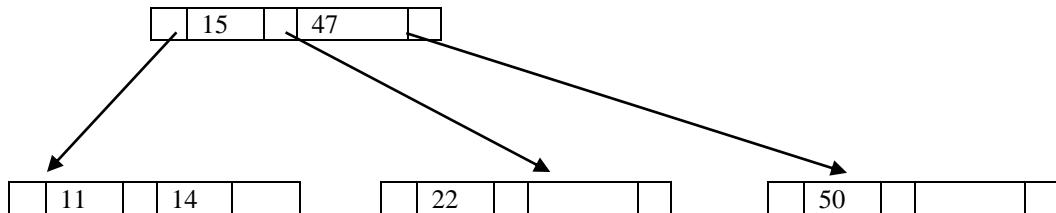


Διαγραφή 4



3.6 Θέμα 5 Φεβρουάριος 2015 και Θέμα 5 Σεπτέμβριος 2017 και Ιούνιος 2020

Υποθέστε το παρακάτω B-tree:

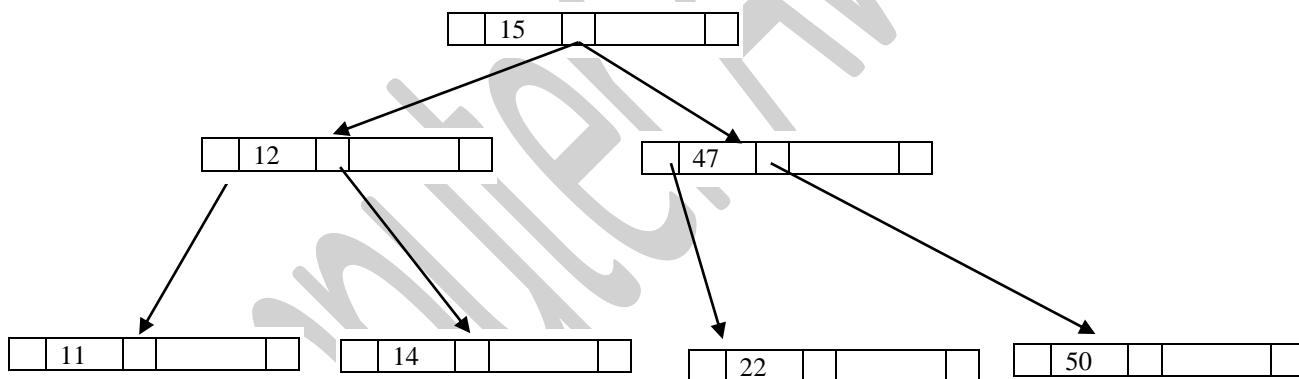


Παρουσιάστε πως τροποποιείται το δέντρο σε κάθε περίπτωση μετά την εφαρμογή των παρακάτω πράξεων (οι πράξεις εφαρμόζονται πάντα στο αρχικό δέντρο)

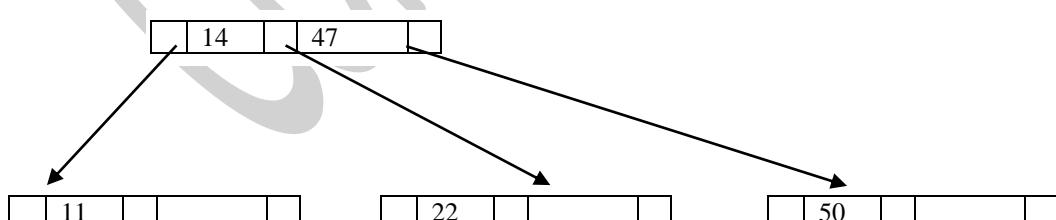
- 1) Εισαγωγή του στοιχείου 12
- 2) Διαγραφή του στοιχείου 15
- 3) Εισαγωγή του στοιχείου 61
- 4) Διαγραφή του στοιχείου 50
- 5) Διαγραφή του στοιχείου 22

Απάντηση

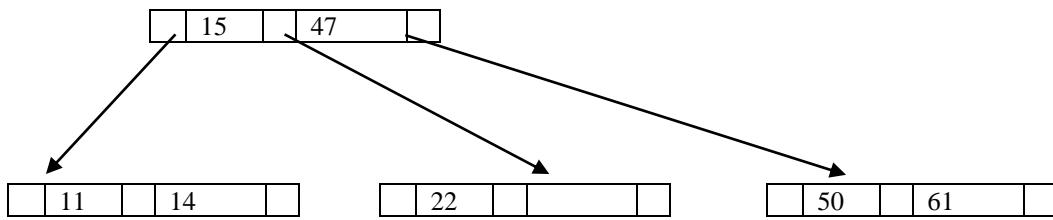
- 1) Εισαγωγή του 12



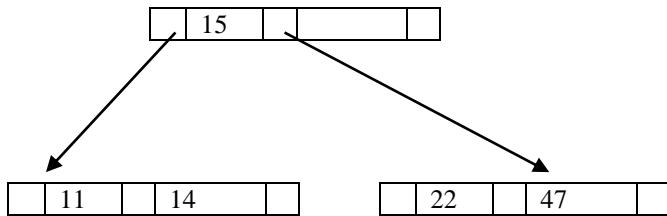
- 2) Διαγραφή του 15



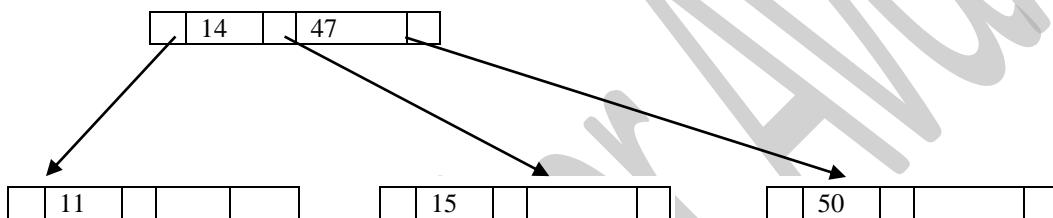
3) Εισαγωγή του 61



4) Διαγραφή του 50



5) Διαγραφή του 22



3.7 Θέμα 4 Ιανουάριος 2020

Θέμα 4: Ευρετηριοποίηση [15 μονάδες]

Υποθέστε το παρακάτω B-tree:

```

graph TD
    Root[15, 47] --> Child1[11, 14]
    Root --> Child2[22]
    Root --> Child3[50]
  
```

- Παρουσιάστε πως τροποποιείται το δέντρο σε κάθε περίπτωση μετά την εφαρμογή των παρακάτω πράξεων (οι πράξεις εφαρμόζονται πάντα στο αρχικό δέντρο): (10 μον.)

 - Εισαγωγή του στοιχείου 12
 - Διαγραφή του στοιχείου 15
 - Διαγραφή του στοιχείου 50

- Με βάση το παρακάτω σχήμα σχεδιάστε (επάνω στην κόλλα των θεμάτων) τα βέλτιστα ελάχιστα περιοριστικά ορθογώνια (*optimal minimum bounding rectangles*) που αντιστοιχούν σε R-tree με fan-out=4 προκειμένον να δεικτοδοτηθούν όλα τα δεδομένα-σημεία του σχήματος. Να δικαιολογήσετε την απάντηση σας. Σχεδιάστε το αντίστοιχο R-tree που προκύπτει. (5 μον.)

4 ΘΕΜΑΤΑ ΜΕ EYPETHΡΙΟ

4.1 Υποδειγματική Άσκηση με Ευρετήριο

Θεωρείστε ένα δίσκο με μέγεθος μπλοκ $B=512$ bytes, ένα Δείκτη Μπλοκ $P=6$ bytes και ένα Δείκτη Εγγραφής $Pr=7$ bytes. Ένα αρχείο έχει $r=30.000$ εγγραφές τύπου ERGAZOMENOS σταθερού μήκους. Η κάθε εγγραφή έχει τα επόμενα πεδία:

- ONOMA (30 bytes)
- AR_TAYT (9 bytes)
- KWD_TM (9 bytes)
- DIEYTH (40 bytes)
- THL (9 bytes)
- HM_GEN (8 bytes)
- FYLO (1 byte)
- KWD_ERG (4 bytes)
- MISTH (4 bytes)
- Χρησιμοποιείται ένα ακόμα byte ως σημάδι διαγραφής

Υποθέστε ότι το αρχείο είναι διατεταγμένο ως προς το πεδίο κλειδί AR_TAYT. Αν κατασκευάσουμε ένα πρωτεύον ευρετήριο στο πεδίο αυτό, να υπολογίσετε το επί τοις εκατό κέρδος στο μέσο πλήθος προσπελάσεων μπλοκ που απαιτούνται για την ανάκτηση μιας εγγραφής του αρχείου όταν δίνεται η τιμή του AR_TAYT. Να επαναλάβετε την άσκηση αν θεωρήστε ότι το αρχείο δεν είναι διατεταγμένο ως προς το πεδίο κλειδί.

Απάντηση

Μέγεθος Μπλοκ **B=512 bytes**

Μέγεθος Δείκτη Μπλοκ **P=6 bytes**

Μέγεθος Δείκτη Εγγραφής **Pr=7bytes**

Αριθμός Εγγραφών **r=30.000 εγγραφές**

Μέγεθος Εγγραφής **R=115 bytes (30+9+9+40+9+8+1+4+4+1)**

Θεωρούμε ότι το πρωτεύον κλειδί είναι το πεδίο AR_TAYT. Διακρίνουμε δύο περιπτώσεις:

- α) Το αρχείο δεδομένων είναι διατεταγμένο ως προς το πρωτεύον κλειδί και η αναζήτηση γίνεται με βάση το πρωτεύον κλειδί
- β) Το αρχείο δεδομένων είναι διατεταγμένο ως προς κάποιο πεδίο το οποίο δεν είναι κλειδί και η αναζήτηση γίνεται με βάση το πρωτεύον κλειδί

Απάντηση

Βάσεις Δεδομένων- Computer Ανάλυση
ΠΕΡΙΠΤΩΣΗ Α1: Αρχείο Διατεταγμένο ως προς το πεδίο Κλειδί AR_TAYT

- ✓ Το Αρχείο Δεδομένων είναι διατεταγμένο ως προς το πρωτεύον κλειδί, χωρίς ευρετήριο
- ✓ **Βήμα 1: Υπολογισμός bfr:** $bfr = \left\lceil \frac{B}{R} \right\rceil = \left\lceil \frac{512 \text{ bytes}}{115 \text{ bytes/record}} \right\rceil = \left\lfloor 4,45 \right\rfloor = 4 \text{ record/block}$ όπου **bfr=blocking factor** (παράγοντας ομαδοποίησης) και δείχνει πόσες εγγραφές τύπου ERGAZOMENOS χωράνε σε κάθε μπλοκ δίσκου
- ✓ **Βήμα 2 :Υπολογισμός blocks Αρχείου:** $b = \left\lceil \frac{r}{bfr} \right\rceil = \left\lceil \frac{30.000 \text{ records}}{4 \frac{\text{records}}{\text{block}}} \right\rceil = 7500 \text{ block}$ Το b δείχνει τον αριθμό των block που καταλαμβάνει το αρχείο Υπαλλήλων
- ✓ **Βήμα 3:** Κάνουμε Δυαδική Αναζήτηση για Εντοπισμό Υπαλλήλου με βάση τον AR_TAYT και οι προσπελάσεις είναι $\log_2 b = \log_2 7500 = 12,91 \approx 13$ προσπελάσεις ($\log_{10} 7500 / \log_{10} 2 = 0,301$)

ΠΕΡΙΠΤΩΣΗ Α2: Αρχείο Διατεταγμένο ως προς το πεδίο Κλειδί AR_TAYT με Πρωτεύον Ευρετήριο

- ✓ Το Πρωτεύον Ευρετήριο είναι ταξινομημένο ως προς το Πρωτεύον Κλειδί AR_TAYT. Το Πρωτεύον Ευρετήριο είναι **Αραιό** δηλ. περιέχει ένα δείκτης προς κάθε block (προς την 1^η εγγραφή κάθε block)
- ✓ **Το μέγεθος κάθε εγγραφής στο Αραιό Ευρετήριο είναι:** $R_i = 9 \text{ bytes (AP_TAYT)} + 6 \text{ bytes (Μέγεθος Δείκτη Block)} = 15 \text{ bytes}$
- ✓ **Βήμα 1:** Ο παράγοντας ομαδοποίησης είναι: $bfr_i = \left\lceil \frac{B}{R_i} \right\rceil = \left\lceil \frac{512}{15} \right\rceil = 34 \text{ records ευρετηρίου/block}$
- ✓ **Βήμα 2:** Το πλήθος blocks που καταλαμβάνει το ευρετήριο είναι:
$$b_i = \left\lceil \frac{r_i}{bfr_i} \right\rceil = \left\lceil \frac{7.500 \text{ records ευρετηρίου}}{34 \text{ records ευρετηρίου/block}} \right\rceil = 221 \text{ blocks}$$
- ✓ **Βήμα 3:** Κάνουμε απευθείας δυαδική αναζήτηση στο αραιό ευρετήριο το οποίο είναι ταξινομημένο ως προς το πεδίο κλειδί AR_TAYT και το πλήθος προσπελάσεων σε αυτό είναι $\log_2 b_i = \log_2 221 = 8$ προσπελάσεις
- **Βήμα 4:** Ο τελικός αριθμός προσπελάσεων είναι $8+1=9$ προσπελάσεις. Οι 8 προσπελάσεις γίνονται στο ευρετήριο για τον εντοπισμό του εργαζομένου και η επιπλέον προσπέλαση γίνεται για να πάμε στο αντίστοιχο block και να φορτώσουμε την εγγραφή στη μνήμη
- Το ποσοστό κέρδους σε σχέση με την αναζήτηση χωρίς ευρετήριο είναι $(13-9)/13 \bullet 100\% = 31\%$
- **Το συμπέρασμα είναι το ευρετήριο επιτρέπει το γρηγορότερο εντοπισμό εγγραφών**

ΑΤ	ΔΕΙΚΤΗΣ
1	
6	
16	
.....

Αραιό ή Πρωτεύον Ευρετήριο

1	ΝΙΚΟΣ	ΓΕΩΡΓΙΟΥ	1000
2	ΜΑΡΙΑ	ΝΙΚΟΛΑΟΥ	2000
3	ΚΩΣΤΑΣ
4	ΓΙΑΝΝΗΣ

6	ΕΛΕΝΗ	ΓΕΩΡΓΙΟΥ	1000
8	ΔΗΜΟΣ	ΝΙΚΟΛΑΟΥ	2000
9	ΜΑΚΗΣ
15	ANNA

16	ΕΛΕΝΗ	ΓΕΩΡΓΙΟΥ	1000
18	ΔΗΜΟΣ	ΝΙΚΟΛΑΟΥ	2000
20	ΜΑΚΗΣ
21	ANNA

ΠΕΡΙΠΤΩΣΗ Α3: Αρχείο Διατεταγμένο ως προς το Πρωτεύον κλειδί με Πολυεπίπεδο Ευρετήριο

- Ο παράγοντας διακλάδωσης είναι $f_0=bfr := 34$ εγγραφές ευρετηρίου/block
- $b_1=b_i=221$ block. Το b_1 είναι ο αριθμός των block που καταλαμβάνει το ευρετήριο 1^ο επιπέδου
- $b_2 = \left\lceil \frac{b_1}{f_0} \right\rceil = \left\lceil \frac{221}{34} \right\rceil = 7$ block. Το b_2 είναι ο αριθμός των block που καταλαμβάνει το ευρετήριο 2^ο επιπέδου
- $b_3 = \left\lceil \frac{b_2}{f_0} \right\rceil = \left\lceil \frac{7}{34} \right\rceil = 1$ block. Το b_3 είναι ο αριθμός των block που καταλαμβάνει το ευρετήριο 3^ο επιπέδου
- Αριθμός Επιπέδων $t=3$ και απαιτούνται $3+1=4$ προσπελάσεις για τον εντοπισμό ενός υπαλλήλου
- Το ποσοστό κέρδους σε σχέση με την αναζήτηση με πρωτεύον ευρετήριο ενός επιπέδου είναι $(9-4)/9 \bullet 100\% = 55,5\%$

Βάσεις Δεδομένων- Computer Ανάλυση

ΑΤ	ΔΕΙΚΤΗΣ
1	
6	
.....

1	ΝΙΚΟΣ	ΓΕΩΡΓΙΟΥ	1000
2	ΜΑΡΙΑ	ΝΙΚΟΛΑΟΥ	2000
3	ΚΩΣΤΑΣ
4	ΓΙΑΝΝΗΣ

ΑΤ	ΔΕΙΚΤΗΣ
1	
6	
16	
.....

Πολυεπίπεδο
Ευρετήριο

ΑΤ	ΔΕΙΚΤΗΣ
16	
20	
.....

6	ΕΛΕΝΗ	ΓΕΩΡΓΙΟΥ	1000
8	ΔΗΜΟΣ	ΝΙΚΟΛΑΟΥ	2000
9	ΜΑΚΗΣ
15	ANNA

16	ΕΛΕΝΗ	ΓΕΩΡΓΙΟΥ	1000
18	ΔΗΜΟΣ	ΝΙΚΟΛΑΟΥ	2000
20	ΜΑΚΗΣ
21	ANNA

Βάσεις Δεδομένων- Computer Ανάλυση
ΠΕΡΙΠΤΩΣΗ Β1: Αρχείο ΟΧΙ Διατεταγμένο ως προς το Πεδίο Κλειδί

- Το Αρχείο Δεδομένων είναι διατεταγμένο ως προς κάποιο πεδίο που δεν είναι κλειδί, χωρίς ευρετήριο.
- **Βήμα 1: Υπολογισμός bfr:** $bfr = \left\lceil \frac{B}{R} \right\rceil = \left\lceil \frac{512 \text{ bytes}}{115 \text{ bytes/record}} \right\rceil = \left\lceil 4,45 \right\rceil = 4 \text{ record/block}$ όπου **bfr=blocking factor** (παράγοντας ομαδοποίησης) και δείχνει πόσες εγγραφές τύπου ERGAZOMENOS χωράνε σε κάθε μπλοκ δίσκου
- **Βήμα 2 :Υπολογισμός blocks Αρχείου:** $b = \left\lceil \frac{r}{bfr} \right\rceil = \left\lceil \frac{30.000 \text{ records}}{4 \frac{\text{records}}{\text{block}}} \right\rceil = 7500 \text{ block}$ Το b δείχνει τον αριθμό των block που καταλαμβάνει το αρχείο Υπαλλήλων
- **Βήμα 3:** Κάνουμε **σειριακή αναζήτηση** για εντοπισμό υπαλλήλου με βάση τον AR_TAYT και οι προσπελάσεις είναι $b/2=7.500/2=3.750$. Η σειριακή αναζήτηση γίνεται γιατί το αρχείο δεν είναι διατεταγμένο ως προς το κλειδί AR_TAYT και η αναζήτηση γίνεται με βάση το πεδίο AR_TAYT

ΠΕΡΙΠΤΩΣΗ Β2: Αρχείο Διατεταγμένο ως κάποιο πεδίο ΟΧΙ κλειδί με Δευτερεύον Ευρετήριο

- ✓ Το Δευτερεύον Ευρετήριο είναι ταξινομημένο ως προς το Πρωτεύον Κλειδί AR_TAYT. Το Δευτερεύον ευρετήριο είναι Πυκνό δηλ. περιέχει ένα ξεχωριστό Δείκτη προς κάθε Εγγραφή
- ✓ **Το μέγεθος κάθε εγγραφής στο Πυκνό Ευρετήριο είναι:** $R_i=9 \text{ bytes (AR_TAYT)}+7 \text{ bytes (Μέγεθος Δείκτη Εγγραφής)}=16 \text{ bytes}$
- ✓ **Βήμα 1:** Ο παράγοντας ομαδοποίησης είναι: $bfr_i = \left\lceil \frac{B}{R_i} \right\rceil = \left\lceil \frac{512}{16} \right\rceil = 32 \text{ records ευρετηρίου/block}$
- ✓ **Βήμα 2:** Το πλήθος blocks που καταλαμβάνει το ευρετήριο είναι:
$$b_i = \left\lceil \frac{r_i}{bfr_i} \right\rceil = \left\lceil \frac{30.000 \text{ records εργαζομένων}}{32 \text{ records ευρετηρίου/block}} \right\rceil = \lceil 937,5 \rceil = 938 \text{ block}$$
- ✓ **Βήμα 3:** Κάνουμε απευθείας δυαδική αναζήτηση στο πυκνό ευρετήριο το οποίο είναι ταξινομημένο ως προς τον AR_TAYT και το πλήθος αναζητήσεων είναι $\log_2 b_i = \log_2 938 = 10$ προσπελάσεις
- **Βήμα 4:** Ο τελικός αριθμός προσπελάσεων είναι $10+1=11 \text{ προσπελάσεις}$. Οι 10 προσπελάσεις γίνονται στο ευρετήριο για τον εντοπισμό του εργαζομένου και η επιπλέον προσπέλαση γίνεται για να πάμε στο αντίστοιχο block και να φορτώσουμε την εγγραφή του στη μνήμη
- Το ποσοστό κέρδους σε σχέση με την αναζήτηση χωρίς ευρετήριο είναι $(3.750-11)/3750 \bullet 100\% = 99,7\%$
- Το συμπέρασμα είναι το ευρετήριο επιτρέπει το γρηγορότερο εντοπισμό εγγραφών

ΑΤ	ΔΕΙΚΤΗΣ
1	
2	
3	
4	
5	
.....

Πυκνό ή Δευτερεύον
Ευρετήριο

1	ΝΙΚΟΣ	ΒΑΣΙΛΕΙΟΥ	1000
2	ΜΑΡΙΑ	ΓΕΩΡΓΙΟΥ	2000
12	ΚΩΣΤΑΣ
4	ΓΙΑΝΝΗΣ

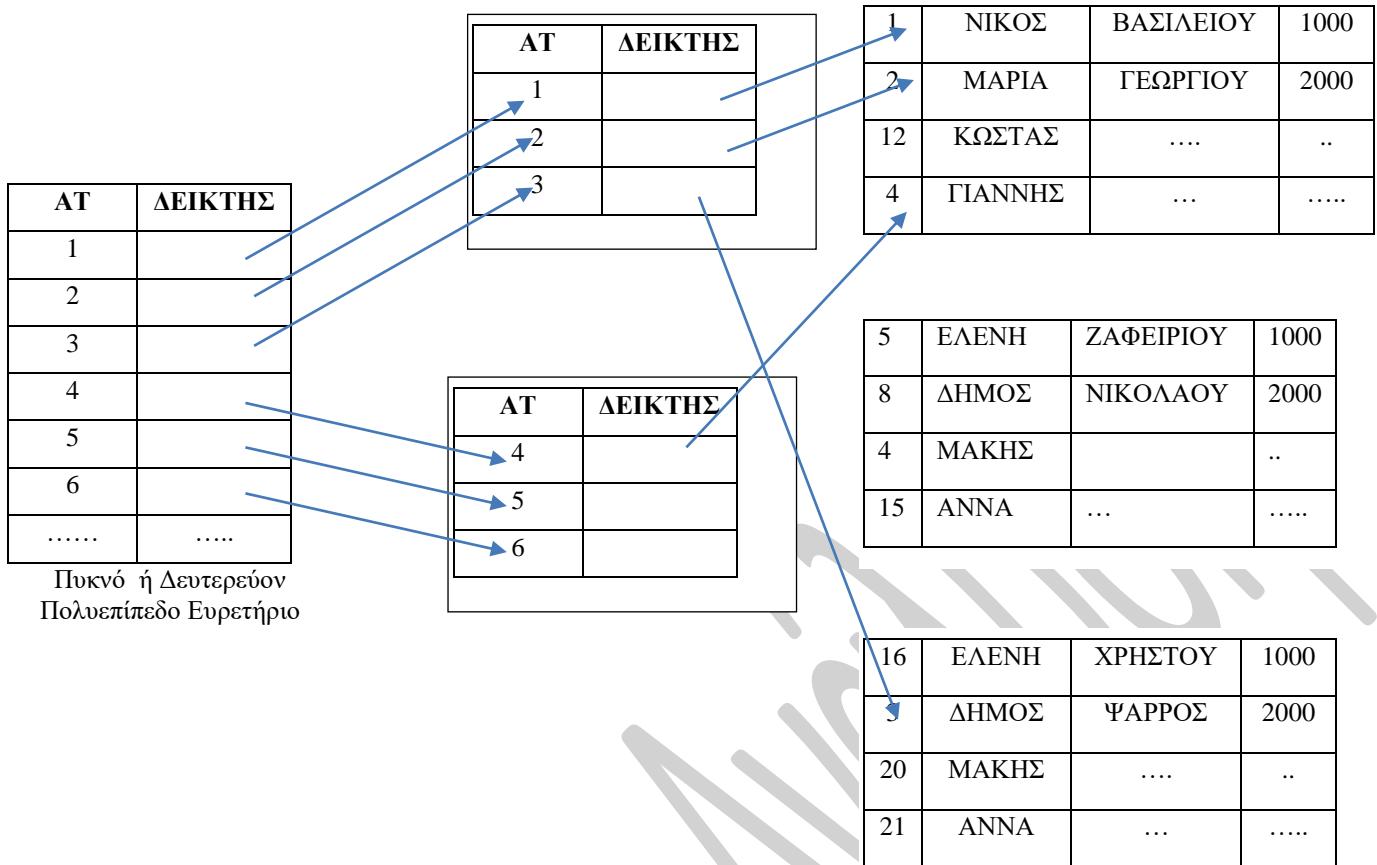
5	ΕΛΕΝΗ	ΖΑΦΕΙΡΙΟΥ	1000
8	ΔΗΜΟΣ	ΝΙΚΟΛΑΟΥ	2000
4	ΜΑΚΗΣ		..
15	ANNA

16	ΕΛΕΝΗ	ΧΡΗΣΤΟΥ	1000
3	ΔΗΜΟΣ	ΨΑΡΡΟΣ	2000
20	ΜΑΚΗΣ
21	ANNA

ΠΕΡΙΠΤΩΣΗ Β3: Αρχείο Διατεταγμένο ως προς το Κάποιο Πεδίο που ΔΕΝ είναι Κλειδί με Πολυεπίπεδο Ευρετήριο

- Ο παράγοντας διακλάδωσης είναι $f_0 =$ παράγοντας ομαδοποίησης ευρετηρίου = $bfr_i = 32$ εγγραφές ευρετηρίου/block
- $b_1 = b_i = 938$ block. Το b_1 είναι ο αριθμός των block που καταλαμβάνει το ευρετήριο 1^ο επιπέδου
- $b_2 = \left\lceil \frac{b_1}{f_0} \right\rceil = \left\lceil \frac{938}{32} \right\rceil = 30$ block. Το b_2 είναι ο αριθμός των block που καταλαμβάνει το ευρετήριο 2^ο επιπέδου
- $b_3 = \left\lceil \frac{b_2}{f_0} \right\rceil = \left\lceil \frac{30}{32} \right\rceil = 1$ block. Το b_3 είναι ο αριθμός των block που καταλαμβάνει το ευρετήριο 3^ο επιπέδου
- Αριθμός Επιπέδων $t=3$ και απαιτούνται $3+1=4$ προσπελάσεις
- Το ποσοστό κέρδους σε σχέση με την αναζήτηση με πρωτεύον ευρετήριο είναι $(11-4)/11 \cdot 100\% = 63,6\%$

Βάσεις Δεδομένων- Computer Ανάλυση



4.2 Θέμα 4 Φεβρουάριος 2009

Έστω ένα αρχείο με $r=32.768$ εγγραφές του τύπου Εκπαιδευτικός όπως παρουσιάζεται παρακάτω:

ΑΦΜ	ΟΝΟΜΑ	ΕΠΩΝΥΜΟ	ΒΑΘΜΙΔΑ	ΕΙΔΙΚΟΤΗΤΑ
4 bytes	18 bytes	18 bytes	12 bytes	11 bytes

Ένα επιπλέον byte χρησιμοποιείται ως σημάδι διαγραφής. Το μέγεθος μπλοκ του δίσκου είναι $B=512$ bytes.

Ερωτήσεις:

1. Υπολογίστε τον παράγοντα ομαδοποίησης bfr (blocking factor) για το συγκεκριμένο αρχείο
2. Ποιος είναι ο απαιτούμενος αριθμός block προκειμένου να αποθηκευτεί το σύνολο των εγγραφών του αρχείου, υποθέτοντας μη εκτεινόμενη οργάνωση
3. Έστω ότι επιθυμούμε να πραγματοποιήσουμε κάποια αναζήτηση. Ποιος ο απαιτούμενος αριθμός προσπελάσεων μπλοκ στο δίσκο για να βρεθεί η ζητούμενη εγγραφή αν το αρχείο είναι διατεταγμένο (μέση περίπτωση); Ποιος είναι ο αντίστοιχος αριθμός προσπελάσεων αν το αρχείο δεν είναι διατεταγμένο;
4. Ας υποθέσουμε ότι το αρχείο που περιέχει τις εγγραφές είναι διατεταγμένο ως προς το πεδίο ΟΝΟΜΑ. Έστω ότι χρησιμοποιούμε ευρετήριο στο πεδίο ΕΠΩΝΥΜΟ με μήκος δείκτη μπλοκ $P=14$ bytes. Ποιος είναι ο αντίστοιχος αριθμός προσπελάσεων μπλοκ στο δίσκο στη μέση περίπτωση για το παρακάτω SQL ερώτημα;

SELECT *

FROM Εκπαιδευτικός

WHERE Επώνυμο = ‘Βασιλείου’;

5. Ας υποθέσουμε ότι θέλουμε να μετατρέψουμε το ευρετήριο του ερωτήματος 4 σε ένα πολυεπίπεδο ευρετήριο. Πόσα επίπεδα απαιτούνται ώστε να δεικτοδοτηθεί το σύνολο των εγγραφών;

Απάντηση

- ✓ Το R είναι το μέγεθος κάθε εγγραφής και προκύπτει ως εξής: 4 bytes [ΑΦΜ]+18 bytes [ΟΝΟΜΑ] +18 bytes [ΕΠΩΝΥΜΟ]+12 bytes [ΒΑΘΜΙΔΑ]+11 Bytes [ΕΙΔΙΚΟΤΗΤΑ] +1 byte [ΣΗΜΑΔΙ ΕΓΓΡΑΦΗΣ] =64 bytes
- ✓ Το B είναι το μέγεθος block το οποίο δίνεται ως 512 bytes.

1. Ο παράγοντας ομαδοποίησης bfr (blocking factor) είναι $bfr = \left\lceil \frac{B}{R} \right\rceil = \left\lceil \frac{512 \text{ bytes}}{64 \text{ bytes}} \right\rceil = \frac{2^9}{2^6} = 2^3 = 8 \text{ record / block}$

2. Ο απαιτούμενος αριθμός block $b = \left\lceil \frac{r}{bfr} \right\rceil = \left\lceil \frac{32.768 \text{ record}}{8 \text{ record / block}} \right\rceil = 4.096 \text{ block}$ όπου r είναι το πλήθος των εγγραφών του αρχείου το οποίο δίνεται ως 32.768. Αρα όλο το αρχείο με εγγραφές του τύπου Εκπαιδευτικός καταλαμβάνει συνολικά 4.096 block

3.

- ✓ Αν το αρχείο είναι διατεταγμένο τότε εφαρμόζουμε δυαδική αναζήτηση σε αυτό για να εντοπίσουμε μια εγγραφή και ο αριθμός των προσπελάσεων είναι: $\log_2 4.096 = \log_2 2^{12} = 12$ προσπελάσεις
- ✓ Αν το αρχείο δεν είναι διατεταγμένο τότε εφαρμόζουμε σειριακή αναζήτηση σε αυτό για να εντοπίσουμε μια εγγραφή και ο αριθμός των προσπελάσεων είναι: $b/2=4.096/2=2.048$ προσπελάσεις

Βάσεις Δεδομένων- Computer Ανάλυση

4. Χρησιμοποιούμε πυκνό ευρετήριο γιατί το αρχείο είναι ταξινομημένο ως προς το πεδίο Όνομα και το ευρετήριο είναι ταξινομημένο ως προς το πεδίο Επώνυμο και εμείς κάνουμε αναζήτηση με βάση το Επώνυμο. Το ευρετήριο αυτό θα περιέχει ένα δείκτη προς κάθε εγγραφή του αρχείου.

- ✓ Το μέγεθος κάθε εγγραφής στο ευρετήριο είναι: $R_i = 18 \text{ bytes} [\text{Επώνυμο}] + 14 \text{ bytes} [\text{Δείκτης προς εγγραφή}] = 32 \text{ bytes}$
- ✓ Το B είναι το μέγεθος block το οποίο δίνεται ως 512 bytes.
- ✓ Ο παράγοντας ομαδοποίησης στο ευρετήριο είναι: $bfr_i = \left\lceil \frac{B}{R_i} \right\rceil = \left\lceil \frac{512}{32} \right\rceil = \frac{2^9}{2^5} = 2^4 = 16 \text{ εγγραφές ευρετηρίου/block}$
- ✓ Το πλήθος blocks που καταλαμβάνει το ευρετήριο είναι: $b_i = \left\lceil \frac{r}{bfr_i} \right\rceil = \left\lceil \frac{32.768}{16} \right\rceil = 2.048 \text{ block όπου } r \text{ το πλήθος εγγραφών του ευρετηρίου το οποίο δίνεται ως 32.768 εγγραφές}$
- ✓ Η αναζήτηση ενός εκπαιδευτικού με βάση το Επώνυμο γίνεται στο ευρετήριο το οποίο είναι ταξινομημένο ως προς το πεδίο Επώνυμο και ο αριθμός των προσπελάσεων για τον εντοπισμό του εφαρμόζοντας δυαδική αναζήτηση είναι: $\log_2 b_i = \log_2 2048 = \log 2^{11} = 11 \text{ προσπελάσεις} + 1 \text{ προσπέλαση για την ανάκτηση της εγγραφής από το δίσκο} = 12 \text{ προσπελάσεις.}$

5.

- ✓ Πρώτα υπολογίζουμε τον παράγοντα διακλάδωσης $f_0 = bfr_i = 16 \text{ records ευρετηρίου/block}$
- ✓ Αριθμός block ευρετηρίου 1^{ου} επιπέδου: $b_1 = b_i = 2.048 \text{ block}$
- ✓ Αριθμός block ευρετηρίου 2^{ου} επιπέδου: $b_2 = \left\lceil \frac{b_1}{f_0} \right\rceil = \left\lceil \frac{2.048}{16} \right\rceil = 128 \text{ Block}$
- ✓ Αριθμός block ευρετηρίου 3^{ου} επιπέδου: $b_3 = \left\lceil \frac{b_2}{f_0} \right\rceil = \left\lceil \frac{128}{16} \right\rceil = 8 \text{ Block}$
- ✓ Αριθμός block ευρετηρίου 4^{ου} επιπέδου: $b_4 = \left\lceil \frac{b_3}{f_0} \right\rceil = \left\lceil \frac{8}{16} \right\rceil = 1 \text{ Block}$
- ✓ Αριθμός Επιπέδων=4
- ✓ Αριθμός Προσπελάσεων για εντοπισμό εγγραφής=4+1=5 προσπελάσεις

Παρατήρηση

- **Μη εκτεινόμενη** (unspanned) οργάνωση: οι εγγραφές δεν επιτρέπεται να διασχίζουν τα όρια ενός block
- **Αχρησιμοποίητος χώρος:** $B - bfr \bullet R \text{ bytes ανά block}$
- **Πιο εύκολη η προσπέλαση**
- **Εκτεινόμενη** (spanned) οργάνωση: αποθήκευση μέρους μιας εγγραφής σε ένα block και το υπόλοιπο σε ένα άλλο block - δείκτης στο τέλος του πρώτου τμήματος δείχνει στο block που περιέχει το υπόλοιπο

4.3 ΘΕΜΑ 6 Φεβρουάριος 2008

Ένα αρχείο έχει $r=65536$ εγγραφές του τύπου φοιτητής με σταθερό μήκος. Κάθε εγγραφή έχει τα ακόλουθα πεδία:

- ΟΝΟΜΑ (30 bytes)
- AT(9 bytes) Πρωτεύον κλειδί
- Ημερ_Γεν(8 bytes)
- Φύλο (1 byte)
- Ένα επιπλέον byte χρησιμοποιείται ως σημάδι διαγραφής
- Κύρια κατεύθυνση(4 bytes)
- Δευτερεύουσα κατεύθυνση (4bytes)
- Κωδ_Πτυχίου(3 bytes)
- Έτος(4 bytes)

Το αρχείο είναι αποθηκευμένο σε δίσκο με τα εξής χαρακτηριστικά

- Μέγεθος μπλοκ $B=512$ bytes

Οι εγγραφές είναι ταξινομημένες στον δίσκο με βάση το πρωτεύων κλειδί.

Ερωτήσεις:

1. (5%) Υπολογίστε το μέγεθος εγγραφής R σε bytes, τον παράγοντα ομαδοποίησης bfr (blocking factor) και τον αριθμό των μπλοκ b , υποθέτοντας μη εκτεινόμενη οργάνωση κάθε δίσκου.
2. (5%) Εστω ότι θέλουμε να δημιουργήσουμε ένα ευρετήριο στο γνώρισμα Όνομα. Τι είδους ευρετήριο (πυκνό ή αραιό) θα επιλέγατε και γιατί;
3. (10%) Πόσες προσπελάσεις block απαιτούνται για να βρούμε την εγγραφή του φοιτητή με AT T342311 αν υποθέσουμε ότι:
 - a. Δεν χρησιμοποιούμε κανενάς είδους ευρετήριο.
 - b. Χρησιμοποιούμε το πρωτεύων ευρετήριο στο γνώρισμα AT με μήκος δείκτη block $P=7$ bytes.

Απάντηση

1.

Το R είναι τα μέγεθος κάθε εγγραφής και είναι ίσο με 64 bytes.

$$\text{Ο παράγοντας ομαδοποίησης } bfr = \left\lceil \frac{B}{R} \right\rceil = \left\lceil \frac{512}{64} \right\rceil = 8 \text{ Εγγραφές/Block}$$

$$\text{Το πλήθος blocks που δεσμεύει το αρχείο } b = \left\lceil \frac{r}{bfr} \right\rceil = \left\lceil \frac{65.536}{8} \right\rceil = 8.192 \text{ Blocks}$$

2. Το ευρετήριο θα είναι πυκνό διότι δημιουργείται σε διαφορετικό πεδίο από αυτό που είναι ταξινομημένο το αρχείο. Πιο συγκεκριμένα το αρχείο είναι ταξινομημένο ως προς το πεδίο AT, ενώ το ευρετήριο είναι ταξινομημένο ως προς το πεδίο Όνομα.

Παρατήρηση

Όταν το αρχείο και το ευρετήριο είναι ταξινομημένα ως προς το ίδιο πεδίο τότε το ευρετήριο είναι αραιό και όταν ταξινομημένα σε διαφορετικά πεδία τότε το ευρετήριο είναι πυκνό.

3.

a. Επειδή κάνουμε αναζήτηση εγγραφής ως προς τον AT και το αρχείο είναι ταξινομημένο ως προς τον AT θα εφαρμόσουμε δυαδική αναζήτηση και το πλήθος των προσπελάσεων είναι $\log_2 8.192 = 13$ προσπελάσεις

β. Αν δημιουργήσουμε αραιό ευρετήριο ως προς τον AT έχουμε τα εξής:

- ✓ Το μέγεθος κάθε εγγραφής στο ευρετήριο είναι: $R_i=9$ bytes [AT] + 7 bytes [Μέγεθος Δείκτη]=16 bytes
- ✓ Ο παράγοντας ομαδοποίησης είναι: $bfr_i = \left\lceil \frac{B}{R_i} \right\rceil = \left\lceil \frac{512}{16} \right\rceil = 32$ Εγγραφές Ευρετηρίου/Block
- ✓ Το πλήθος blocks που καταλαμβάνει το ευρετήριο είναι: $b_i = \left\lceil \frac{r}{bfr_i} \right\rceil = \left\lceil \frac{8.192}{32} \right\rceil = 256$ Block

- ✓ Ο αριθμός των προσπελάσεων στο ευρετήριο εφαρμόζοντας δυαδική αναζήτηση είναι: $\log_2 b_i = \log_2 256 = 8 + 1 = 9$ προσπελάσεις

4.4 ΘΕΜΑ 6 Φεβρουάριος 2010

Ένα αρχείο έχει $r=32.768$ εγγραφές του τύπου Αθλητής_Καλαθοσφαίρισης με σταθερό μήκος, Κάθε εγγραφή έχει τα ακόλουθα πεδία:

- ΚΩΔΙΚΟΣ (4 bytes)
- ΟΝΟΜΑ (9 bytes)
- ΕΠΩΝΥΜΟ (9 bytes)
- ΘΕΣΗ (8 bytes)
- ΗΛΙΚΙΑ (2 bytes)
- Ένα επιπλέον byte χρησιμοποιείται ως σημάδι διαγραφής

Το αρχείο είναι αποθηκευμένο σε δίσκο με τα εξής χαρακτηριστικά:

Μέγεθος μπλοκ $B=512$ bytes.

Οι εγγραφές είναι ταξινομημένες στο δίσκο με βάση το πρωτεύον κλειδί.

Ερωτήσεις:

1. Υπολογίστε τον παράγοντα ομαδοποίησης (blocking factor) για το συγκεκριμένο αρχείο.
2. Ποιος είναι ο απαιτούμενος αριθμός block προκειμένου να αποθηκευτεί το σύνολο των εγγραφών του αρχείου υποθέτοντας μη εκτεινόμενη οργάνωση;
3. Ας υποθέσουμε ότι το αρχείο μας είναι διατεταγμένο και επιθυμούμε να πραγματοποιήσουμε μια αναζήτηση. Ποιος ο απαιτούμενος αριθμός προσπελάσεων μπλοκ στο δίσκο για να βρεθεί η ζητούμενη εγγραφή για τη μέση περίπτωση τόσο για τη γραμμική όσο και για τη δυαδική αναζήτηση;
4. Ας υποθέσουμε ότι χρησιμοποιούμε δευτερεύον πυκνό ευρετήριο (ως προς το πρωτεύον κλειδί) με μήκος δείκτη μπλοκ $P=4$ bytes. Ποιος ο απαιτούμενος αριθμός προσπελάσεων μπλοκ στο δίσκο για να βρεθεί η ζητούμενη εγγραφή στη μέση περίπτωση;

Απάντηση

$$1. bfr = \left\lceil \frac{B}{R} \right\rceil = \left\lceil \frac{512}{33} \right\rceil = \left\lceil 15,51 \right\rceil = 15 \text{ εγγραφές/block}$$

$$2. b = \left\lceil \frac{r}{bfr} \right\rceil = \left\lceil \frac{32.768}{15} \right\rceil = \left\lceil 2.184,53 \right\rceil = 2.185 \text{ Blocks}$$

3. Δυαδική Αναζήτηση $\rightarrow \log_2 b = 11,13$ προσπελάσεις

Γραμμική Αναζήτηση $\rightarrow b/2 = 1024$ προσπελάσεις

4.

$$bfr_i = \left\lceil \frac{B}{R_i} \right\rceil = \left\lceil \frac{512}{8} \right\rceil = 64 \text{ εγγραφές/block}$$

$$b_i = \left\lceil \frac{r_i}{bfr_i} \right\rceil = \left\lceil \frac{32.768}{64} \right\rceil = 512 \text{ Blocks}$$

$\log_2 512 = 9 + 1 = 10$ προσπελάσεις

4.5 Θέμα 5 –Σεπτέμβριος 2009

Ένα αρχείο $r=65.536$ εγγραφές σταθερού μήκους του πίνακα CD. Κάθε εγγραφή έχει τα ακόλουθα:

- ΚΩΔ (8 bytes)
- ΤΙΤΛΟΣ (24 bytes)
- ΚΩΔ. ΣΥΝΘΕΤΗ (8 bytes)
- ΚΩΔ. ΕΡΜΗΝΕΥΤΗ (8 bytes)
- ΚΩΔ. ΕΤΑΙΡΕΙΑΣ ΠΑΡΑΓΩΓΗΣ (8 bytes)
- ΕΤΟΣ ΚΥΚΛΟΦΟΡΙΑΣ (7 bytes)
- Ένα επιπλέον byte χρησιμοποιείται ως σημάδι διαγραφής

Το αρχείο είναι αποθηκευμένο σε δίσκο με τα εξής χαρακτηριστικά:

Μέγεθος μπλοκ $B=1.024$ bytes

Σημείωση

Οι εγγραφές είναι ταξινομημένες στο δίσκο με βάση το πρωτεύον κλειδί.

Δεν υπάρχουν δύο εγγραφές με ίδια τιμή στο πεδίο ΤΙΤΛΟΣ.

Ερωτήσεις

1. Υπολογίστε τον παράγοντα ομαδοποίησης bfr (blocking factor) για το συγκεκριμένο αρχείο.
2. Ποιος είναι ο αριθμός block που απαιτείται προκειμένου να αποθηκευτεί το σύνολο των εγγραφών του αρχείου, υποθέτοντας μη εκτεινόμενη οργάνωση;
3. Εστω ότι δημιουργούμε ένα δευτερεύον ευρετήριο (ενός επιπέδου) στο γνώρισμα ΤΙΤΛΟΣ. Με δεδομένο ότι το μήκος δείκτη block που θα χρησιμοποιηθεί είναι 8 bytes, ποιο είναι το ελάχιστο πλήθος blocks που απαιτούνται ώστε να δεικτοδοτηθεί το σύνολο των εγγραφών;
4. Εστω το παρακάτω ερώτημα σε SQL: **SELECT * FROM CD WHERE TITLOΣ= "Somewhere far beyond"**
 - a. με χρήση δευτερεύοντος ευρετηρίου
 - b. χωρίς τη χρήση δευτερεύοντος ευρετηρίου, στη μέση περίπτωση

Απάντηση

$$1. \text{ bfr} = \left\lceil \frac{B}{R} \right\rceil = \left\lceil \frac{1024}{64} \right\rceil = 16 \text{ εγγραφές/block}$$

$$2. \text{ b} = \left\lceil \frac{r}{bfr} \right\rceil = \left\lceil \frac{65536}{16} \right\rceil = 4.096 \text{ Blocks}$$

3. Χρησιμοποιούμε πυκνό ευρετήριο.

- ✓ Το μέγεθος κάθε εγγραφής στο ευρετήριο είναι: $R_i=24+8=32$ bytes
- ✓ Ο παράγοντας ομαδοποίησης είναι: $bfr_i = \left\lceil \frac{B}{R_i} \right\rceil = \left\lceil \frac{1024}{32} \right\rceil = 32$ εγγραφές/block

Βάσεις Δεδομένων- Computer Ανάλυση

- ✓ Το πλήθος blocks που καταλαμβάνει το ευρετήριο είναι: $b_i = \left\lceil \frac{r_i}{bfr_i} \right\rceil = \left\lceil \frac{65536}{32} \right\rceil = 2.048$ blocks

4.

a. Κάνουμε δυαδική αναζήτηση στο δευτερεύον ευρετήριο το οποίο είναι ταξινομημένο ως προς τον τίτλο και το πλήθος αναζητήσεων είναι:

- ✓ Ο αριθμός των προσπελάσεων είναι: $\log_2 b_i = \log_2 2.048 = 11 + 1 = 12$ προσπελάσεις

b. Ψάχνουμε απευθείας στο αρχείο για να εντοπίσουμε την εγγραφή με βάση τον τίτλο της όπως φαίνεται από το ερώτημα της SQL. Επειδή το αρχείο των CD είναι ταξινομημένο ως προς το πεδίο ΤΙΤΛΟΣ θα εφαρμόσουμε δυαδική αναζήτηση. Άρα ο αριθμός των προσπελάσεων είναι:

$$\log_2 b = \log_2 4.096 = 12 \text{ προσπελάσεις}$$

Αν αναζήτηση γινόταν με βάση άλλο γνώρισμα τότε θα εφαρμόζαμε σειριακή αναζήτηση

4.6 Θέμα 6 –Φεβρουάριος 2013

Έστω ένα αρχείο με $r=65536$ εγγραφές του τύπου ΦΟΙΤΗΤΗΣ με σταθερό μήκος όπως παρουσιάζεται παρακάτω:

ΑΜ	ΟΝΟΜΑ	ΕΠΩΝΥΜΟ	ΤΗΛ.
4 bytes	49 bytes	69 bytes	4 bytes

Επίσης έχουμε:

- Ένα επιπλέον byte χρησιμοποιείται ως σημάδι διαγραφής.
- Το μέγεθος μπλοκ του δίσκου είναι $B=1024$ bytes

Ερωτήσεις

- a. Ποιος είναι ο απαιτούμενος αριθμός block προκειμένου να αποθηκευτεί το σύνολο των εγγραφών του αρχείου υποθέτοντας μη εκτεινόμενη οργάνωση;
- β. Έστω ότι επιθυμούμε να πραγματοποιήσουμε αναζήτηση με βάση ένα συγκεκριμένο τηλέφωνο.
- i. Ποιος είναι ο απαιτούμενος αριθμός προσπελάσεων block στο δίσκο για να βρεθεί η ζητούμενη εγγραφή αν το αρχείο είναι διατεταγμένο (μέση περίπτωση);
- ii. Ποιος είναι ο αντίστοιχος αριθμός προσπελάσεων αν το αρχείο δεν είναι διατεταγμένο;
- γ. Ας υποθέσουμε ότι έχουμε πυκνό ευρετήριο ενός επιπέδου στο πεδίο τηλέφωνο με μήκος δείκτη block $P=4$ bytes. Ποιος είναι ο απαιτούμενος αριθμός προσπελάσεων block στο δίσκο για πραγματοποιήσουμε αναζήτηση με βάση ένα συγκεκριμένο τηλέφωνο;
- δ. Ας υποθέσουμε ότι το αρχείο που περιέχει τις εγγραφές είναι διατεταγμένο ως προς το πεδίο Α.Μ. Έστω ότι χρησιμοποιούμε ευρετήριο πάνω στο πεδίο Α.Μ. με μήκος δείκτη μπλοκ στο δίσκο $P=4$ bytes. Ποιος είναι ο απαιτούμενος αριθμός προσπελάσεων μπλοκ στο δίσκο (στη μέση περίπτωση) για το παρακάτω ερώτημα SQL:

SELECT * FROM ΦΟΙΤΗΤΗΣ WHERE AM=10;

ε. Το αραιό ευρετήριο είναι πολυεπίπεδο. Βρείτε τον αριθμό επιπέδων και το μέσο αριθμό προσπελάσεων

Απάντηση

a.

- Το μέγεθος κάθε εγγραφής είναι $R=4+49+69+4+1=127$ bytes
- Πρώτα υπολογίζουμε τον παράγοντα ομαδοποίησης (blocking factor) που δείχνει πόσες εγγραφές του αρχείου χωράνε σε κάθε μπλοκ του δίσκου. Από τις συγκεκριμένες τιμές που δίνονται προκύπτει ότι
- Ο Παράγοντας Ομαδοποίησης blocking factor (bfr) είναι:

$$bfr = \left\lfloor \frac{B}{R} \right\rfloor = \left\lfloor \frac{1.024 \text{ bytes/block}}{127 \text{ bytes/record}} \right\rfloor = \left\lfloor 8,06 \right\rfloor = 8 \text{ record / block} \text{ εγγραφές/block}$$

- Το πλήθος των blocks για την αποθήκευση του αρχείου είναι: $b = \left\lfloor \frac{r}{bfr} \right\rfloor = \left\lfloor \frac{65.536 \text{ records}}{8 \text{ records/block}} \right\rfloor = 8.192 \text{ blocks block}$

β.

i. Αν το αρχείο είναι διατεταγμένο (στη μέση περίπτωση) ο μέσος αριθμός προσπελάσεων block στο δίσκο, εφαρμόζοντας τον αλγόριθμο δυαδικής αναζήτησης, είναι $\log_2 b = \log_2 8.192 = 13$ προσπελάσεις

ii. Αν το αρχείο δεν είναι διατεταγμένο (στη μέση περίπτωση) ο μέσος αριθμός προσπελάσεων block στο δίσκο, εφαρμόζοντας τον αλγόριθμο σειριακής αναζήτησης, είναι $\frac{b}{2} = \frac{8.192}{2} = 4.096$ προσπελάσεις

γ. Το ευρετήριο είναι πυκνό διότι το αρχείο είναι διατεταγμένο ως προς το πεδίο A.M. ενώ το ευρετήριο είναι διατεταγμένο ως προς το πεδίο τηλέφωνο, άρα το ευρετήριο θα έχει ένα δείκτη προς κάθε εγγραφή του αρχείου. Το μέγεθος κάθε καταχώρισης στο πυκνό ευρετήριο είναι: $R_i =$ μήκος δείκτη block (4 bytes) + Τηλέφωνο (4 bytes) = 8 bytes. Ο παράγοντας ομαδοποίησης είναι:

$$bfr_i = \left\lfloor \frac{B}{R_i} \right\rfloor = \left\lfloor \frac{1.024}{8} \right\rfloor = 256 \text{ εγγραφές ευρετηρίου σε block του δίσκου.}$$

Ο αριθμός εγγραφών στο πυκνό ευρετήριο είναι όσες

και οι εγγραφές του αρχείου δηλ. 65.536 εγγραφές διότι κάθε δείκτης στο πυκνό ευρετήριο είναι δείκτης προς μια εγγραφή. Ο αριθ-

$$\text{μός μπλοκ που δεσμεύει συνολικά το πυκνό ευρετήριο είναι: } b_i = \left\lceil \frac{65.536}{256} \right\rceil = 256 \text{ block.}$$

Στο πυκνό ευρετήριο εφαρμόζουμε

πάντα δυαδική αναζήτηση και οι προσπελάσεις σε αυτό είναι: $\log_2 b = \log_2 256 = 8$ προσπελάσεις. Ο τελικός αριθμός προσπελάσεων είναι: $8+1=9$ προσπελάσεις

δ. Το ευρετήριο είναι αραιό διότι το αρχείο είναι διατεταγμένο ως προς το πεδίο A.M. ενώ και το ευρετήριο είναι διατεταγμένο ως προς το ίδιο πεδίο, άρα το αραιό ευρετήριο θα έχει ένα δείκτη προς κάθε μπλοκ του αρχείου. Το μέγεθος κάθε καταχώρισης στο αραιό ευρετήριο είναι: $R_i =$ μήκος δείκτη block (4 bytes)+ AM (4 bytes) = 8 bytes. Ο παράγοντας ομαδοποίησης είναι:

$$bfr_i = \left\lfloor \frac{B}{R_i} \right\rfloor = \left\lfloor \frac{1.024}{8} \right\rfloor = 256 \text{ εγγραφές ευρετηρίου/block.}$$

Ο αριθμός εγγραφών στο αραιό ευρετήριο θα είναι όσα και τα μπλοκ

του αρχείου δηλ. 8.192 εγγραφές διότι κάθε δείκτης στο αραιό ευρετήριο θα είναι όπως αναφέραμε δείκτης προς ένα μπλοκ δίσκου.

$$\text{Ο αριθμός μπλοκ που δεσμεύει συνολικά το αραιό ευρετήριο είναι: } b_i = \left\lceil \frac{r_i}{bfr_i} \right\rceil = \left\lceil \frac{8.192}{256} \right\rceil = 32 \text{ block.}$$

Στο αραιό ευρετήριο εφαρμόζουμε πάντα δυαδική αναζήτηση και οι προσπελάσεις σε αυτό είναι: $\log_2 32 = \log_2 2^5 = 5$ προσπελάσεις. Ο τελικός αριθμός προσπελάσεων είναι: $5+1=6$ προσπελάσεις

ε.

- Ο παράγοντας διακλάδωσης είναι $f_0 = bfr_i = 256$
- $b_1 = b_i = 32$
- $b_2 = \left\lceil \frac{b_1}{f_0} \right\rceil = 1$
- **Άρα 2 επίπεδα και $2+1=3$ προσπελάσεις**

4.7 Θέμα 6 –Φεβρουάριος 2014

Ένα αρχείο έχει 2000 εγγραφές του τύπου Υπάλληλος και κάθε εγγραφή έχει τα ακόλουθα πεδία:

- ΑΦΜ (4 bytes)
- ΟΝΟΜΑ (20 bytes)
- ΕΠΩΝΥΜΟ (22 bytes)
- ΘΕΣΗ (14 bytes)
- ΗΛΙΚΙΑ (2 bytes)
- Ένα επιπλέον byte χρησιμοποιείται ως σημάδι διαγραφής

Το αρχείο είναι αποθηκευμένο σε δίσκο με τα εξής χαρακτηριστικά:

- Μέγεθος block 512 bytes

i. Έστω ότι το αρχείο είναι διατεταγμένο και επιθυμούμε να πραγματοποιήσουμε κάποια αναζήτηση. Ποιος είναι ο απαιτούμενος αριθμός προσπελάσεων μπλοκ στο δίσκο για να βρεθεί η ζητούμενη εγγραφή στη μέση περίπτωση; Ποιος είναι ο αντίστοιχος αριθμός προσπελάσεων αν τα αρχείο δεν είναι διατεταγμένο;

ii. Ας υποθέσουμε ότι το αρχείο που περιέχει τις εγγραφές είναι διατεταγμένο ως προς το πεδίο ΑΦΜ. Έστω ότι χρησιμοποιούμε ευρετήριο πάνω στο πεδίο ΕΠΩΝΥΜΟ με μήκος δείκτη μπλοκ $P=10$ bytes. Ποιος είναι ο απαιτούμενος αριθμός προσπελάσεων μπλοκ στο δίσκο (στη μέση περίπτωση) για την εκτέλεση του παρακάτω SQL ερωτήματος:

```
SELECT *
FROM Υπάλληλος
WHERE Επώνυμο =”Παπαδόπουλος”;
```

iii. Ας υποθέσουμε ότι θέλουμε να μετατρέψουμε το ευρετήριο του προηγούμενου ερωτήματος σε ένα πολυεπίπεδο ευρετήριο. Πόσα επίπεδα απαιτούνται ώστε να δεικτοδοτηθεί το σύνολο των εγγραφών και ποιος ο αριθμός των προσπελάσεων μπλοκ για την εύρεση μιας εγγραφής;

Απάντηση

i.

- Το μέγεθος κάθε εγγραφής είναι $R=4+20+22+14+2+1=63$ bytes
- Ο Παράγοντας Ομαδοποίησης (bfr) είναι: $bfr = \left\lceil \frac{B}{R} \right\rceil = \left\lceil \frac{512}{63} \right\rceil = 8$ εγγραφές/block
- Το πλήθος των blocks για την αποθήκευση του αρχείου είναι: $b = \left\lceil \frac{r}{bfr} \right\rceil = \left\lceil \frac{2000}{8} \right\rceil = 250$ block
- Αν το αρχείο είναι διατεταγμένο (στη μέση περίπτωση) ο μέσος αριθμός προσπελάσεων block στο δίσκο, εφαρμόζοντας τον αλγόριθμο δυαδικής αναζήτησης, είναι $\log_2 b = \log_2 250 = 8$ προσπελάσεις
- Αν το αρχείο δεν είναι διατεταγμένο (στη μέση περίπτωση) ο μέσος αριθμός προσπελάσεων block στο δίσκο, εφαρμόζοντας τον αλγόριθμο σειριακής αναζήτησης, είναι $\frac{b}{2} = \frac{250}{2} = 125$ προσπελάσεις

ii.

- Το μέγεθος κάθε εγγραφής στο ευρετήριο είναι $R_i=22+10=32$

Βάσεις Δεδομένων- Computer Ανάλυση

- Ο Παράγοντας Ομαδοποίησης (bfr_i) είναι: $bfr_i = \left\lceil \frac{B}{R_i} \right\rceil = \left\lceil \frac{512}{32} \right\rceil = 16$ εγγραφές ευρετηρίου/block
- Το πλήθος των blocks για την αποθήκευση του ευρετηρίου $b_i = \left\lceil \frac{r_i}{bfr_i} \right\rceil = \left\lceil \frac{2000}{16} \right\rceil = 125$
- Ο μέσος αριθμός προσπελάσεων block στο ευρετήριο είναι $\log_2 b_i = \log_2 125 = 7$. Σε αυτές προσθέτουμε μια ακόμα προσπέλαση και προκύπτουν $7+1=8$ προσπελάσεις

iii.

- Ο παράγοντας διακλάδωσης είναι $f_0 = bfr_i = 16$
- $b_1 = b_i = 125$
- $b_2 = \left\lceil \frac{b_1}{f_0} \right\rceil = 8$
- $b_3 = \left\lceil \frac{b_2}{f_0} \right\rceil = 1$
- **Άρα 3+1=4 προσπελάσεις**

4.8 Θέμα 6 –Ιούνιος 2014

Ένα αρχείο έχει 30.000 εγγραφές του τύπου Φοιτητής με σταθερό μήκος εγγραφής και κάθε εγγραφή έχει τα ακόλουθα πεδία:

ID	4 bytes
ΟΝΟΜΑ	27 bytes
ΕΠΩΝΥΜΟ	26 bytes
ΠΑΝΕΠΙΣΤΗΜΙΟ	22 bytes
ΤΜΗΜΑ	22 bytes

- Ένα επιπλέον byte χρησιμοποιείται ως σημάδι διαγραφής.
- Το μέγεθος μπλοκ του δίσκου είναι $B=1024$ bytes

Ερωτήσεις:

1. Υπολογίστε τον παράγοντα ομαδοποίησης bfr (blocking factor) για το συγκεκριμένο αρχείο
2. Ποιός είναι ο απαιτούμενος αριθμός block προκειμένου να αποθηκευτεί το σύνολο των εγγραφών του αρχείου υποθέτοντας μη εκτεινόμενη οργάνωση;
3. Εστω ότι επιθυμούμε να πραγματοποιήσουμε κάποια αναζήτηση. Ποιος είναι ο απαιτούμενος αριθμός προσπελάσεων block στο δίσκο για να βρεθεί η ζητούμενη εγγραφή αν το αρχείο είναι διατεταγμένο (μέση περίπτωση); Ποιος είναι ο αντίστοιχος αριθμός προσπελάσεων αν το αρχείο δεν είναι διατεταγμένο;
4. Ας υποθέσουμε ότι το αρχείο που περιέχει τις εγγραφές είναι διατεταγμένο ως προς το πεδίο ΕΠΩΝΥΜΟ. Εστω ότι χρησιμοποιούμε ευρετήριο πάνω στο πεδίο ΟΝΟΜΑ με μήκος δείκτη μπλοκ $P=15$ bytes. Ποιος είναι ο απαιτούμενος αριθμός προσπελάσεων μπλοκ στο δίσκο (στη μέση περίπτωση) για την εκτέλεση του παρακάτω SQL ερωτήματος:

```
SELECT *
FROM Φοιτητής
WHERE Όνομα = "Γιώργος";
```

Βάσεις Δεδομένων- Computer Ανάλυση

5. Ας υποθέσουμε ότι θέλουμε να μετατρέψουμε το ευρετήριο του προηγούμενου ερωτήματος 4 σε ένα πολυεπίπεδο ευρετήριο. Πόσα επίπεδα απαιτούνται ώστε να δεικτοδοτηθεί το σύνολο των εγγραφών;

Απάντηση

1.

a. Το μέγεθος κάθε εγγραφής είναι $R=4+27+26+22+22+1=102$ bytes

Πρώτα υπολογίζουμε τον παράγοντα ομαδοποίησης (blocking factor) που δείχνει πόσες εγγραφές του αρχείου χωράνε σε κάθε μπλοκ του δίσκου. Από τις συγκεκριμένες τιμές που δίνονται προκύπτει ότι

- Ο Παράγοντας Ομαδοποίησης blocking factor (bfr) είναι: $bfr = \left\lceil \frac{B}{R} \right\rceil = \left\lceil \frac{1.024}{102} \right\rceil = 10,039 = 10$ εγγραφές/block

2. Το πλήθος των blocks για την αποθήκευση του αρχείου είναι: $b = \left\lceil \frac{r}{bfr} \right\rceil = \left\lceil \frac{30000}{10} \right\rceil = 3000$ block

3.

i. Αν το αρχείο είναι διατεταγμένο (στη μέση περίπτωση) ο μέσος αριθμός προσπελάσεων block στο δίσκο, εφαρμόζοντας τον αλγόριθμο δυαδικής αναζήτησης, είναι $\log_2 b = \log_2 8.192 = 11,59$ προσπελάσεις

ii. Αν το αρχείο δεν είναι διατεταγμένο (στη μέση περίπτωση) ο μέσος αριθμός προσπελάσεων block στο δίσκο, εφαρμόζοντας τον αλγόριθμο σειριακής αναζήτησης, είναι $\frac{b}{2} = \frac{3000}{2} = 1500$ προσπελάσεις

4. Το ευρετήριο είναι αραιό διότι το αρχείο είναι διατεταγμένο ως προς το πεδίο Επώνυμο ενώ το ευρετήριο είναι διατεταγμένο ως προς το ONOMA, άρα το πυκνό ευρετήριο θα έχει ένα δείκτη προς κάθε εγγραφή του αρχείου. Το μέγεθος κάθε καταχώρισης στο πυκνό ευρετήριο είναι: $R_i =$ μήκος δείκτη block (15 bytes) + ONOMA (27 bytes) = 42 bytes. Ο παράγοντας ομαδοποίησης είναι:

$$bfr_i = \left\lceil \frac{B}{R_i} \right\rceil = \left\lceil \frac{1.024}{42} \right\rceil = 24$$
 εγγραφές ευρετηρίου/block. Ο αριθμός εγγραφών στο πυκνό ευρετήριο θα είναι όσες και οι εγγραφές του αρχείου δηλ. 30000 εγγραφές διότι κάθε δείκτης θα δείχνει σε μια εγγραφή. Ο αριθμός μπλοκ που δεσμεύει συνολικά το πυκνό ευρετήριο είναι: $b_i = \left\lceil \frac{r_i}{bfr_i} \right\rceil = \left\lceil \frac{30000}{24} \right\rceil = 1250$ block. Στο πυκνό ευρετήριο εφαρμόζουμε δυαδική αναζήτηση και οι προσπελάσεις σε αυτό είναι: $\log_2 1250 = 10,32$ προσπελάσεις. Ο τελικός αριθμός προσπελάσεων είναι: $10,32+1=11,32$ προσπελάσεις

5. Ο παράγοντας διακλάδωσης είναι $f_0=bfr_i=24$

$$b_1 = b_i = 1250$$

$$b_2 = \left\lceil \frac{b_1}{f_0} \right\rceil = \left\lceil \frac{1250}{24} \right\rceil = 53$$

$$b_3 = \left\lceil \frac{b_2}{f_0} \right\rceil = 3$$

$$b_4 = \left\lceil \frac{b_3}{f_0} \right\rceil = 1$$

Άρα $4+1=5$ προσπελάσεις

4.9 Θέμα 4 Φεβρουάριος 2015

Έστω ένα αρχείο με $r=32768$ εγγραφές του τύπου Εργαζόμενος όπως παρουσιάζεται παρακάτω:

ID	ΟΝΟΜΑ	ΕΠΩΝΥΜΟ	ΘΕΣΗ	ΤΜΗΜΑ
4 bytes	18 bytes	18 bytes	12 bytes	11 bytes

Ένα επιπλέον bytes ως χρησιμοποιείται ως σημάδι διαγραφής. Το μέγεθος μπλοκ του δίσκου είναι $B=1024$ bytes.

Ερωτήσεις:

1. Υπολογίστε τον παράγοντα ομαδοποίησης bfr (blocking factor) για το συγκεκριμένο αρχείο
2. Ποιος είναι ο απαιτούμενος αριθμός block προκειμένου να αποθηκευτεί το σύνολο των εγγραφών του αρχείου, υποθέτοντας μη εκτεινόμενη οργάνωση
3. Έστω ότι επιθυμούμε να πραγματοποιήσουμε κάποια αναζήτηση. Ποιος ο απαιτούμενος αριθμός προσπελάσεων μπλοκ στο δίσκο για να βρεθεί η ζητούμενη εγγραφή αν το αρχείο είναι διατεταγμένο (μέση περίπτωση); Ποιος είναι ο αντίστοιχος αριθμός προσπελάσεων αν το αρχείο δεν είναι διατεταγμένο;
4. Ας υποθέσουμε ότι το αρχείο που περιέχει τις εγγραφές είναι διατεταγμένο ως προς το πεδίο ΟΝΟΜΑ. Έστω ότι χρησιμοποιούμε ευρετήριο στο πεδίο ΕΠΩΝΥΜΟ με μήκος δείκτη μπλοκ $P=14$ bytes. Ποιος είναι ο αντίστοιχος αριθμός προσπελάσεων μπλοκ στο δίσκο στη μέση περίπτωση για το παρακάτω SQL ερώτημα;

SELECT *

FROM Εργαζόμενος

WHERE Επώνυμο = ‘Βασιλείου’;

5. Ας υποθέσουμε ότι θέλουμε να μετατρέψουμε το ευρετήριο το ερωτήματος 4 σε ένα πολυεπίπεδο ευρετήριο. Πόσα επίπεδα απαιτούνται ώστε να δεικτοδοτηθεί το σύνολο των εγγραφών;

Απάντηση

1. $bfr = \left\lceil \frac{B}{R} \right\rceil = \left\lceil \frac{1024}{64} \right\rceil = 16$ εγγραφές/block (Μέγεθος Εγγραφής=4+18+18+12+11+1)
2. $b = \left\lceil \frac{r}{bfr} \right\rceil = \left\lceil \frac{32768}{16} \right\rceil = 2048$ Blocks καταλαμβάνει το αρχείο
- 3.
- Αν το αρχείο είναι διατεταγμένο (στη μέση περίπτωση) ο μέσος αριθμός προσπελάσεων block στο δίσκο, εφαρμόζοντας τον **αλγόριθμο δυαδικής αναζήτησης**, είναι $\log_2 b = \log_2 2048 = \log_2 2^{11} = 11$ προσπελάσεις
- Αν το αρχείο δεν είναι διατεταγμένο (στη μέση περίπτωση) ο μέσος αριθμός προσπελάσεων block στο δίσκο, εφαρμόζοντας τον αλγόριθμο **σειριακής αναζήτησης**, είναι $\frac{b}{2} = \frac{2048}{2} = 1024$ προσπελάσεις
4. Το μέγεθος κάθε καταχώρισης στο πυκνό ευρετήριο είναι: $R_i =$ μήκος δείκτη block (14 bytes) + επώνυμο (18 bytes) = 32 bytes. Ο παράγοντας ομαδοποίησης είναι: $bfr_i = \left\lceil \frac{B}{R_i} \right\rceil = \left\lceil \frac{1.024}{32} \right\rceil = 32$ εγγραφές ευρετηρίου σε block του δίσκου. Ο αριθμός εγγραφών στο πυκνό ευρετήριο είναι όσες και οι εγγραφές του αρχείου δηλ. 32768 εγγραφές διότι κάθε δείκτης στο πυκνό ευρετήριο είναι δείκτης προς μια εγγραφή. Ο αριθμός μπλοκ που δεσμεύει συνολικά το πυκνό ευρετήριο είναι:

$$b_i = \left\lceil \frac{r_i}{bfr_i} \right\rceil = \left\lceil \frac{32768}{32} \right\rceil = 1024 \text{ block. Στο πυκνό ευρετήριο εφαρμόζουμε πάντα δυαδική αναζήτηση και οι προσπελάσεις σε}$$

αυτό είναι: } log₂b = log₂1024 = 10 προσπελάσεις. Ο τελικός **αριθμός προσπελάσεων είναι: 10+1=11 προσπελάσεις**

5. Ο παράγοντας διακλάδωσης είναι f₀=bfr_i=32

b₁ = b_i = 1024 είναι τα block του ευρετηρίου.

$$b_2 = \left\lceil \frac{b_1}{f_0} \right\rceil = \left\lceil \frac{1024}{32} \right\rceil = 32$$

$$b_3 = \left\lceil \frac{32}{f_0} \right\rceil = \left\lceil \frac{32}{32} \right\rceil = 1$$

Άρα έχουμε 3 επίπεδα και οι προσπελάσεις είναι **3+1=4 προσπελάσεις**

4.10 Θέμα 5 Φεβρουάριος 2017

Έστω ένα αρχείο με $r=16.384$ εγγραφές του τύπου ΦΟΙΤΗΤΗΣ με σταθερό μήκος. Κάθε εγγραφή έχει τα ακόλουθα πεδία:

ΚΩΔΙΚΟΣ	ΟΝΟΜΑΤΕΠΩΝΥΜΟ	ΔΙΕΥΘΥΝΣΗ	ΤΗΛΕΦΩΝΟ
6 bytes	60 bytes	30 bytes	10 bytes

Το αρχείο είναι αποθηκευμένο σε δίσκο με μέγεθος μπλοκ $B=512$ bytes.

Ερωτήσεις:

1. Υπολογίστε τον παράγοντα ομαδοποίησης bfr (blocking factor) για το συγκεκριμένο αρχείο
2. Ποιος είναι ο απαιτούμενος αριθμός block προκειμένου να αποθηκευτεί το σύνολο των εγγραφών του αρχείου, υποθέτοντας μη εκτεινόμενη οργάνωση
3. Ας υποθέσουμε ότι το αρχείο μας δεν είναι διατεταγμένο και επιθυμούμε να πραγματοποιήσουμε κάποια αναζήτηση. Ποιος ο απαιτούμενος αριθμός προσπελάσεων μπλοκ στο δίσκο για να βρεθεί η ζητούμενη εγγραφή για τη μέση περίπτωση;
4. Εάν το αρχείο μας είναι διατεταγμένο ως προς τον κωδικό και επιθυμούμε να πραγματοποιήσουμε αναζήτηση για συγκεκριμένο κωδικό φοιτητή. Ποιος ο απαιτούμενος αριθμός προσπελάσεων μπλοκ στο δίσκο για να βρεθεί η ζητούμενη εγγραφή;
5. Αν υποθέσουμε ότι δεικτοδοτούμε τα δεδομένα ως προς το ονοματεπώνυμο χρησιμοποιώντας ένα πολυεπίπεδο ευρετήριο με παράγοντα αναδίπλωσης bfr_i (που θα πρέπει να υπολογίσετε), πόσα επίπεδα απαιτούνται ώστε να δεικτοδοτηθεί το σύνολο των εγγραφών με δεδομένο ότι το μήκος δείκτη block που θα χρησιμοποιηθεί είναι 4 bytes;

Απάντηση

1.

- Το μέγεθος κάθε εγγραφής είναι $R=6+60+30+10=106$ bytes
- Ο Παράγοντας Ομαδοποίησης blocking factor (bfr) είναι: $bfr = \left\lceil \frac{B}{R} \right\rceil = \left\lceil \frac{512}{106} \right\rceil = \left\lfloor 4,83 \right\rfloor = 4$ εγγραφές/block

2. Το πλήθος των blocks για την αποθήκευση του αρχείου είναι: $b = \left\lceil \frac{r}{bfr} \right\rceil = \left\lceil \frac{16.384}{4} \right\rceil = 4.096$ block

3. Αν το αρχείο δεν είναι διατεταγμένο (στη μέση περίπτωση) ο μέσος αριθμός προσπελάσεων block στο δίσκο, εφαρμόζοντας τον αλγόριθμο σειριακής αναζήτησης, είναι $\frac{b}{2} = \frac{4.096}{2} = 2.048$ προσπελάσεις
4. Αν το αρχείο είναι διατεταγμένο (στη μέση περίπτωση) ο μέσος αριθμός προσπελάσεων block στο δίσκο, εφαρμόζοντας τον αλγόριθμο δυαδικής αναζήτησης, είναι $\log_2 4.096 = 12$ προσπελάσεις

5.

- Το μέγεθος κάθε καταχώρισης στο ευρετήριο είναι: $R_i = 60+4=64$ bytes. (Ονοματεπώνυμο και Δείκτης)

- Ο παράγοντας ομαδοποίησης είναι: $bfr_i = \left\lfloor \frac{B}{R_i} \right\rfloor = \left\lfloor \frac{512}{64} \right\rfloor = 8$ εγγραφές ευρετηρίου/block

- Ο αριθμός μπλοκ που δεσμεύει συνολικά το ευρετήριο είναι: $b_i = \left\lceil \frac{r_i}{bfr_i} \right\rceil = \left\lceil \frac{16.384}{8} \right\rceil = 2.048$ blocks ευρετηρίου

- Ο παράγοντας διακλάδωσης είναι $f_0 = bfr_i = 8$ εγγραφές ευρετηρίου/block

- Τα blocks του ευρετηρίου στο 1^o επίπεδο είναι $b_1 = b_i = 2.048$ blocks

- Τα blocks του ευρετηρίου στο 2^o επίπεδο είναι $b_2 = \left\lceil \frac{b_1}{f_0} \right\rceil = \left\lceil \frac{2.048}{8} \right\rceil = 256$

- Τα blocks του ευρετηρίου στο 3^o επίπεδο είναι $b_3 = \left\lceil \frac{b_2}{f_0} \right\rceil = \left\lceil \frac{256}{8} \right\rceil = 32$

- Τα blocks του ευρετηρίου στο 4^o επίπεδο είναι $b_4 = \left\lceil \frac{b_3}{f_0} \right\rceil = \left\lceil \frac{32}{8} \right\rceil = 4$

- Τα blocks του ευρετηρίου στο 4^o επίπεδο είναι $b_5 = \left\lceil \frac{b_4}{f_0} \right\rceil = \left\lceil \frac{4}{8} \right\rceil = 1$

- **Άρα απαιτούνται 5 επίπεδα και 6 προσπελάσεις**

4.11 Θέμα 5 Φεβρουάριος 2019

Έστω ένα αρχείο με $r=16.384$ εγγραφές του τύπου ΣΥΝΔΡΟΜΗΤΗΣ με σταθερό μήκος. Κάθε εγγραφή έχει τα ακόλουθα πεδία:

ΚΩΔΙΚΟΣ	ΟΝΟΜΑΤΕΠΩΝΥΜΟ	ΔΙΕΥΘΥΝΣΗ	ΤΗΛΕΦΩΝΟ
6 bytes	60 bytes	30 bytes	10 bytes

Το αρχείο είναι αποθηκευμένο σε δίσκο με μέγεθος μπλοκ $B=256$ bytes.

5.1. Ποιος είναι ο απαιτούμενος αριθμός block προκειμένου να αποθηκευτεί το σύνολο των εγγραφών του αρχείου, υποθέτοντας μη εκτεινόμενη οργάνωση

5.2. Ας υποθέσουμε ότι το αρχείο μας δεν είναι διατεταγμένο και επιθυμούμε να πραγματοποιήσουμε κάποια αναζήτηση. Ποιος ο απαιτούμενος αριθμός προσπελάσεων μπλοκ στο δίσκο για να βρεθεί η ζητούμενη εγγραφή για τη μέση περίπτωση;

5.3. Εάν το αρχείο μας είναι διατεταγμένο ως προς τον κωδικό και επιθυμούμε να πραγματοποιήσουμε αναζήτηση για συγκεκριμένο κωδικό συνδρομητή. Ποιος ο απαιτούμενος αριθμός προσπελάσεων μπλοκ στο δίσκο για να βρεθεί η ζητούμενη εγγραφή;

5.4. Αν υποθέσουμε ότι δεικτοδοτούμε τα δεδομένα ως προς το ονοματεπώνυμο χρησιμοποιώντας ένα πολυεπίπεδο ευρετήριο με παράγοντα διακλάδωσης (fan-out) bfr_i (που θα πρέπει να υπολογίσετε), πόσα επίπεδα απαιτούνται ώστε να δεικτοδοτηθεί το σύνολο των εγγραφών με δεδομένο ότι το μήκος δείκτη block που θα χρησιμοποιηθεί είναι 4 bytes;

Απάντηση

5.1. Μέγεθος εγγραφής = $6+60+30+10=106$ bytes

$$bfr = \left\lceil \frac{B}{R} \right\rceil = \left\lceil \frac{256}{106} \right\rceil = \left\lfloor 2,415 \right\rfloor = 2 \text{ εγγραφές/block}$$

$$b = \left\lceil \frac{r}{bfr} \right\rceil = \left\lceil \frac{16.384}{2} \right\rceil = 8.192 \text{ Blocks}$$

5.2. Εφαρμόζουμε σειριακή αναζήτηση και ο μέσος αριθμός προσπελάσεων θα είναι $b/2=4.096$ προσπελάσεις

5.3. Εφαρμόζουμε δυαδική αναζήτηση και ο μέσος αριθμός προσπελάσεων θα είναι $\log_2 b = \log_2 8.192 = 13$ προσπελάσεις

5.4. Θα χρειαστούμε πυκνό ευρετήριο γιατί το αρχείο είναι διατεταγμένο ως προς τον κωδικό και εμείς δεικτοδοτούμε το ευρετήριο ως προς το ονοματεπώνυμο

- ✓ Το μέγεθος κάθε εγγραφής στο ευρετήριο είναι: $R_i=60$ bytes μέγεθος ονοματεπώνυμου+4 bytes μέγεθος δείκτη=64 bytes
- ✓ Ο παράγοντας ομαδοποίησης είναι: $bfr_i = \left\lceil \frac{B}{R_i} \right\rceil = \left\lceil \frac{256}{64} \right\rceil = 4$ εγγραφές ευρετηρίου/block
- ✓ **Επειδή το αρχείο είναι διατεταγμένο ως προς τον Κωδικό ενώ το ευρετήριο ως προς το Ονοματεπώνυμο θα χρησιμοποιήσουμε πυκνό ευρετήριο**

- ✓ Το πλήθος blocks που καταλαμβάνει το ευρετήριο είναι: $b_i = \left\lceil \frac{r_i}{bfr_i} \right\rceil = \left\lceil \frac{16.384}{4} \right\rceil = 4.096$ blocks
- ✓ Ο παράγοντας διακλάδωσης είναι $f_0=bfr_i=4$ εγγραφές ευρετηρίου/block
- ✓ **$b_1=b_i=4096$ block.** Το b_1 είναι ο αριθμός των block που καταλαμβάνει το ευρετήριο 1^{ον} επιπέδου
- ✓ $b_2 = \left\lceil \frac{b_1}{f_0} \right\rceil = \left\lceil \frac{4096}{4} \right\rceil = 1024$ block καταλαμβάνει το ευρετήριο 2^{ον} επιπέδου
- ✓ $b_3 = \left\lceil \frac{b_2}{f_0} \right\rceil = \left\lceil \frac{1024}{4} \right\rceil = 256$ block καταλαμβάνει το ευρετήριο 3^{ον} επιπέδου
- ✓ $b_4 = \left\lceil \frac{b_3}{f_0} \right\rceil = \left\lceil \frac{256}{4} \right\rceil = 64$ block καταλαμβάνει το ευρετήριο 4^{ον} επιπέδου
- ✓ $b_5 = \left\lceil \frac{b_4}{f_0} \right\rceil = \left\lceil \frac{64}{4} \right\rceil = 16$ block καταλαμβάνει το ευρετήριο 5^{ον} επιπέδου
- ✓ $b_6 = \left\lceil \frac{b_5}{f_0} \right\rceil = \left\lceil \frac{16}{4} \right\rceil = 4$ block που καταλαμβάνει το ευρετήριο 6^{ον} επιπέδου
- ✓ $b_7 = \left\lceil \frac{b_{65}}{f_0} \right\rceil = \left\lceil \frac{4}{4} \right\rceil = 1$ block που καταλαμβάνει το ευρετήριο 7^{ον} επιπέδου
- ✓ Αριθμός Επιπέδων $t=7$

4.12 Θέμα 5 Ιανουάριος 2020

Θέμα 5: Οργάνωση Αρχείων [20 μονάδες]

Έστω ένα αρχείο με $r=65536$ εγγραφές του τύπου **Πελάτης** με σταθερό μήκος, όπως παρουσιάζεται παρακάτω:

Πελάτης				
Α.Φ.Μ.	Όνομα	Επώνυμο	Επάγγελμα	Διεύθυνση
10 bytes	14 bytes	16 bytes	10 bytes	13 bytes

Ένα επιπλέον byte χρησιμοποιείται ως σημάδι διαγραφής. Το μέγεθος μπλοκ του δίσκου είναι $B = 1024$ bytes.

1. Υπολογίστε τον παράγοντα ομαδοποίησης bfr (blocking factor) για το συγκεκριμένο αρχείο. (5 μον.)
2. Ποιος είναι ο απαιτούμενος αριθμός block προκειμένου να αποθηκευθεί το σύνολο των εγγραφών του αρχείου, υποθέτοντας μη εκτεινόμενη οργάνωση; (5 μον.)
3. Ας υποθέσουμε ότι χρησιμοποιούμε δευτερεύον πυκνό ευρετήριο ως προς το πεδίο Α.Φ.Μ. με μήκος δείκτη block $P=6$ bytes. Ποιος είναι ο απαιτούμενος αριθμός προσπελάσεων μπλοκς στον δίσκο για να βρεθεί η ζητούμενη εγγραφή στη μέση περίπτωση; (10 μον.)

Απάντηση

1.

- Το μέγεθος κάθε εγγραφής είναι $R=10+14+16+10+13+1=64$ bytes
- Ο Παράγοντας Ομαδοποίησης blocking factor (bfr) είναι: $bfr = \left\lceil \frac{B}{R} \right\rceil = \left\lceil \frac{1024}{64} \right\rceil = 16$ εγγραφές/block

2. Το πλήθος των blocks για την αποθήκευση του αρχείου είναι: $b = \left\lceil \frac{r}{bfr} \right\rceil = \left\lceil \frac{65.536}{16} \right\rceil = 4.096$ block

3.

- Το μέγεθος κάθε καταχώρισης στο ευρετήριο είναι: $R_i = 10+6=16$ bytes. (ΑΦΜ και Δείκτης)
- Ο παράγοντας ομαδοποίησης είναι: $bfr_i = \left\lceil \frac{B}{R_i} \right\rceil = \left\lceil \frac{1024}{16} \right\rceil = 64$ εγγραφές ευρετηρίου/block
- Ο αριθμός μπλοκ που δεσμεύει συνολικά το ευρετήριο είναι: $b_i = \left\lceil \frac{r_i}{bfr_i} \right\rceil = \left\lceil \frac{65.536}{64} \right\rceil = 1.024$ blocks ευρετηρίου
- Στο πυκνό ευρετήριο εφαρμόζουμε πάντα δυαδική αναζήτηση και οι προσπελάσεις σε αυτό είναι: $\log_2 b = \log_2 1024 = 10$ προσπελάσεις. Ο τελικός αριθμός προσπελάσεων είναι: **10+1=11 προσπελάσεις**

Computer Ανάλυση

5 ΣΥΝΑΡΤΗΣΙΑΚΕΣ ΕΞΑΡΤΗΣΕΙΣ, ΚΑΝΟΝΙΚΕΣ ΜΟΡΦΕΣ, ΕΛΑΧΙΣΤΗ ΚΑΛΥΨΗ, ΔΙΑΤΗΡΗΣΗ ΕΞΑΡΤΗΣΕΩΝ Κ.Λ.Π.

5.1 Συναρτησιακές Εξαρτήσεις (Functional Dependencies-FDs)

- Για την αντιμετώπιση του πλεονασμού και της ασυνέπειας στην αξιολόγηση ενός σχήματος βάσεων δεδομένων χρησιμοποιούμε τις συναρτησιακές εξαρτήσεις. Η συναρτησιακή εξάρτηση (functional dependency) είναι ένας είδος περιορισμού ακεραιότητας ο οποίος γενικεύει την έννοια του κλειδιού
- Μια Συναρτησιακή Εξάρτηση είναι ένας περιορισμός μεταξύ δυο ομάδων γνωρίσματων μιας βάσης δεδομένων.
 - Κάντε την παραδοχή ότι όλα τα γνωρίσματα μιας βάσης αποθηκεύονται σε ένα καθολικό πίνακα $R(A_1, A_2, \dots, A_n)$
- Έστω ότι $R=(A_1, A_2, \dots, A_n)$ και ότι $X \subseteq R$ και $Y \subseteq R$, τότε η Συναρτησιακή Εξάρτηση $X \rightarrow Y$, υποδηλώνει ότι μια ανάθεση τιμών στο σύνολο X προσδιορίζει μοναδικά το σύνολο Y .
 - Δηλαδή, εάν $t1[X]=t2[X]$, τότε $t1[Y]=t2[Y]$
 - Επομένως είναι μια γενίκευση της έννοιας του κλειδιού.
- Θα αναφερόμαστε στα FDs ως ακολούθως:
 - Το X προσδιορίζει συναρτησιακά το Y
 - Το Y είναι συναρτησιακά **εξαρτώμενο** από το X

Παράδειγμα

$X \rightarrow Y$

1. Ο Κωδικός προσδιορίζει το Επώνυμο και το Όνομα δηλ. Κωδικός → Επώνυμο Όνομα
2. Το Όνομα, Επώνυμο και Πατρώνυμο προσδιορίζουν μοναδικά την Ηλικία δηλ. Όνομα, Επώνυμο, Πατρώνυμο → Ηλικία

5.2 Αξιώματα Armstrong

Ανακλαστικότητα: $Y \subseteq X \Rightarrow X \rightarrow Y$

Επαυξητικότητα: $X \rightarrow Y \Rightarrow XW \rightarrow YW$

Μεταβατικότητα: $\begin{matrix} X \rightarrow Y \\ Y \rightarrow Z \end{matrix} \Rightarrow X \rightarrow Z$

Ανακλαστικότητα → Ένα υπερσύνολο γνωρίσματων προσδιορίζει συναρτησιακά ένα οποιοδήποτε υποσύνολο αυτού

5.3 Κανόνες που προκύπτουν από τα Αξιώματα Armstrong

A. Διασπαστικός Κανόνας (Decomposition Rule)

Av $X \rightarrow YZ$ τότε $X \rightarrow Y$ και $X \rightarrow Z$

Απόδειξη

Για την απόδειξη $X \rightarrow Y$

1. $X \rightarrow YZ$ (δίνεται)
2. $YZ \rightarrow Y$ (Ανακλαστικότητα διότι $Y \subseteq YZ$)
3. $X \rightarrow Y$ (Μεταβατικότητα από 1 και 2)

Για την απόδειξη $X \rightarrow Z$

1. $X \rightarrow YZ$ (δίνεται)
2. $YZ \rightarrow Z$ (Ανακλαστικότητα διότι $Z \subseteq YZ$)

3. $X \rightarrow Z$ (Μεταβατικότητα από 1 και 2)

Β. Κανόνας Σύνθεσης (Composition Rule)

Av $X \rightarrow Y$ και $A \rightarrow B$ τότε $XA \rightarrow YB$

Απόδειξη

1. $X \rightarrow Y$ (δίνεται)
2. $A \rightarrow B$ (δίνεται)
3. $XA \rightarrow YA$ (Επαύξηση κανόνα 1 με A)
4. $XA \rightarrow Y$ (Διάσπαση 3)
5. $XA \rightarrow XB$ (Επαύξηση κανόνα 2 με X)
6. $XA \rightarrow B$ (Διάσπαση 5)
7. $XA \rightarrow YB$ (Ενωτικός Κανόνας 4 και 6)

Γ. Ενωτικός Κανόνας (Union Rule)

Av $X \rightarrow Y$ και $X \rightarrow Z$ τότε $X \rightarrow YZ$

Απόδειξη

1. $X \rightarrow Y$ (δίνεται)
2. $X \rightarrow Z$ (δίνεται)
3. $X \rightarrow XX$
4. $XX \rightarrow XY$ (επαύξηση 1 με το X)
5. $XY \rightarrow YZ$ (επαύξηση 2 με το Y)
6. $X \rightarrow XY$ (Μεταβατικότητα 3 και 4)
7. $X \rightarrow YZ$ (Μεταβατικότητα 4 και 5)

Δ. Ψευδο-μεταβατικός Κανόνας

Av $X \rightarrow Y$ και $YZ \rightarrow W$ τότε $XZ \rightarrow W$

Απόδειξη

1. $X \rightarrow Y$ (δίνεται)
2. $YZ \rightarrow W$ (δίνεται)
3. $XZ \rightarrow YZ$ (επαύξηση 1 με Z)
4. $XZ \rightarrow W$ (Μεταβατικότητα 3 και 2)

5.4 Κλειστότητα Συνόλου Συναρτησιακών Εξαρτήσεων F^+ OXI

- Δοθέντος ενός συνόλου F από ένα σύνολο συναρτησιακών εξαρτήσεων (ΣE) το F^+ είναι το σύνολο όλων των συναγόμενων ΣE

Παράδειγμα

Παίρνει ($\Delta F M$, $A M$, βαθμός, όνομα, διεύθυνση)

$$\begin{array}{l} \Delta F M, A M \rightarrow \text{βαθμός} \\ \Delta F M \rightarrow \text{όνομα, διεύθυνση} \end{array} \quad \left. \right\} F$$

$$\begin{array}{l} \Delta F M, A M \rightarrow \text{βαθμός} \\ \Delta F M \rightarrow \text{όνομα, διεύθυνση} \\ \Delta F M \rightarrow \Delta F M \\ \Delta F M, A M \rightarrow \text{διεύθυνση} \\ A M, \text{διεύθυνση} \rightarrow A M \\ \dots \end{array} \quad \left. \right\} F^+$$

5.5 Αλγόριθμος Υπολογισμού Κλειστότητας Συνόλου Συναρτησιακών Εξαρτήσεων - OXI

$$F^+ = F$$

repeat

for κάθε εξάρτηση f του F^+
εφάρμοσε ανκλαστική και αυξητική ιδιότητα
πρόσθεσε τις παραγόμενες εξαρτήσεις στο F^+
for κάθε ζευγάρι εξαρτήσεων f_1, f_2 που ανήκει στο F^+
if μπορεί να εφαρμοστεί προσεταιριστικότητα
 then πρόσθεσε την παραγόμενη εξάρτηση
στο F^+
until F^+ δεν αλλάζει περαιτέρω

5.5.1 Παράδειγμα Υπολογισμού Κλειστότητας Συνόλου Συναρτησιακών Εξαρτήσεων - OXI

■ Δίνεται

$$\begin{aligned} R &= (A, B, C, G, H, I) \\ F &= \{ \begin{aligned} A &\rightarrow B \\ &A \rightarrow C \\ &CG \rightarrow H \\ &CG \rightarrow I \\ &B \rightarrow H \end{aligned} \} \end{aligned}$$

■ Μερικά μέλη του F^+ .

Απάντηση

Από τις Σ.Ε. $A \rightarrow B$ και $B \rightarrow H$ εφαρμόζοντας μεταβατικότητα $A \rightarrow H$ προστίθεται στο F^+

Από την Σ.Ε. $A \rightarrow C$ εφαρμόζοντας επαύξηση με G προκύπτει $AG \rightarrow CG$. Δίνεται $CG \rightarrow I$. Εφαρμόζοντας μεταβατικότητα προκύπτει η Σ.Ε. $AG \rightarrow I$ την οποία προσθέτουμε στο F^+

$$CG \rightarrow HI \text{ (Ιδιότητα Ένωσης)} : CG \rightarrow H \text{ (αύξηση)} \Rightarrow$$

$$\Rightarrow (CG)CG \rightarrow CGH \text{ δηλ. } CG \rightarrow CGH \text{ (1)}$$

$$\text{Από } CG \rightarrow I \text{ (αύξηση)} \Rightarrow CGH \rightarrow HI \text{ (2)}$$

Από (1) και (2) εφαρμόζωντας προσεταιριστικότητα παίρνουμε το αποτέλεσμα.

5.6 Κλειστότητα Γνωρίσματος A+

- Διθέντος ενός συνόλου F από ένα σύνολο FD συναρτησιακών εξαρτήσεων το A^+ είναι το σύνολο όλων των γνωρίσματων που καθορίζονται (εξαρτώνται) από το A

Παράδειγμα

Πιάρνει (ΑΦΜ, ΑΜ, βαθμός, όνομα, διεύθυνση)

$$\begin{array}{l} \text{ΑΦΜ, ΑΜ} \rightarrow \text{βαθμός} \\ \text{ΑΦΜ} \rightarrow \text{όνομα, διεύθυνση} \\ \{ \text{ΑΦΜ} \}^+ = ?? \end{array} \quad \left. \right\} F$$

Απάντηση

- Αρχικά η κλειστότητα $\{ \text{ΑΦΜ} \}^+ = \{ \text{ΑΦΜ} \}$
- Από την Σ.Ε. $\text{ΑΦΜ} \rightarrow \text{όνομα, διεύθυνση}$ η κλειστότητα του γνωρίσματος ΑΦΜ δηλ. $\{ \text{ΑΦΜ} \}^+$ επαυξάνεται με τα γνωρίσματα όνομα και διεύθυνση. Άρα $\{ \text{ΑΦΜ} \}^+ = \{ \text{ΑΦΜ} \} \cup \{ \text{όνομα, διεύθυνση} \} = \{ \text{ΑΦΜ, όνομα, διεύθυνση} \}$
- Από τις υπάρχουσες εξαρτήσεις το $\{ \text{ΑΦΜ} \}^+ + \underline{\text{δεν μπορεί να επαυξηθεί}}$ άρα η τελική μορφή του $\{ \text{ΑΦΜ} \}^+ = \{ \text{ΑΦΜ, όνομα, διεύθυνση} \}$
- Επειδή $\{ \text{ΑΦΜ} \}^+$ ΔΕΝ ΠΕΡΙΛΑΜΒΑΝΕΙ ΟΛΑ ΤΑ ΓΝΩΡΙΣΜΑΤΑ ΤΟΥ ΠΙΝΑΚΑ 'Πιάρνει' ΔΕΝ ΑΠΟΤΕΛΕΙ ΥΠΟΨΗΦΙΟ ΚΛΕΙΔΙ ΤΟ ΑΦΜ

Παρατήρηση

Αν $A^+ = \{ \text{όλα τα γνωρίσματα του πίνακα} \}$
τότε ' A' υποψήφιο κλειδί

5.7 Τετριμμένες Εξαρτήσεις - OXI

Η συναρτησιακή εξάρτηση $\alpha \rightarrow \beta$ είναι τετριμμένη αν $\beta \subseteq \alpha$.

Παραδείγματα:

Όνομα, Επμο \rightarrow Επμο

Μισθός \rightarrow Μισθός

5.8 Αλγόριθμος Υπολογισμού Κλειστότητας Γνωρίσματος X+ - OXI

- 1) $X^+ := X$
- 2) repeat
- 3) $old_X^+ := X^+$
- 4) for each FD $Y \rightarrow Z$ in F do
- 5) if $Y \subseteq X^+$ then $X^+ := X^+ \cup Z$
- 6) until ($old_X^+ == X^+$)

Εναλλακτική Διατύπωση Αλγόριθμου Υπολογισμού Κλειστότητας Γνωρίσματος - OXI

- Δίνεται πίνακας (σχέση) R, ένα σύνολο από Σ.Ε. και ένα σύνολο πεδίων X του R
- Κλειστότητα του X ή X^+ είναι το σύνολο όλων των πεδίων Y του R για τα οποία ισχύει $X \rightarrow Y$
- Αλγόριθμος: Ξεκινούμε θέτοντας $X^+ = X$ και μετά για κάθε Σ.Ε. $Y \rightarrow Z$ με $Y \subseteq X$ προσθέτουμε το Z στο X^+ . Συνεχίζουμε μέχρι να μην μπορεί να προστεθεί κάτι στο X^+

5.8.1 Παράδειγμα 1 Υπολογισμού Κλειστότητας Γνωρισμάτων

$F = \{ SSN \rightarrow Ename,$
 $Pnumber \rightarrow \{ Pname, Plocation \}$
 $\{ Ssn, Pnumber \} \rightarrow Hours \}$

Να υπολογιστούν οι ακόλουθες κλειστότητες:

- 1) $\{ Ssn \}^+ = \{ Ssn, Ename \}$
- 2) $\{ Pnumber \}^+ = \{ Pnumber, Pname, Plocation \}$
- 3) $\{ Ssn, Pnumber \}^+ = \{ Ssn, Pnumber, Ename, Pname, Plocation, Hours \}$

Απόδειξη 3^{ου}

- Αρχικά η κλειστότητα είναι $\{ SSN, Pnumber \}^+ = \{ SSN, Pnumber \}$
- Από τη Σ.Ε. $SSN \rightarrow Ename$ η κλειστότητα $\{ SSN, Pnumber \}^+$ επαυξάνεται κατά το γνώρισμα $Ename$ δηλ. $\{ SSN, Pnumber \}^+ = \{ SSN, Pnumber \} \cup \{ Ename \} = \{ SSN, Pnumber, Ename \}$
- Από τη Σ.Ε. $SSN, Pnumber \rightarrow Hours$ η κλειστότητα $\{ SSN, Pnumber, Ename \}^+$ επαυξάνεται κατά το γνώρισμα $Hours$ δηλ. $\{ SSN, Pnumber \}^+ = \{ SSN, Pnumber, Ename \} \cup \{ Hours \} = \{ SSN, Pnumber, Ename, Hours \}$
- Από τη Σ.Ε. $Pnumber \rightarrow Pname, Plocation$ η κλειστότητα $\{ SSN, Pnumber, Ename, Hours \}^+$ επαυξάνεται κατά τα γνωρίσματα $Pname, Plocation$ δηλ. $\{ SSN, Pnumber, Ename, Hours \}^+ = \{ SSN, Pnumber, Ename, Hours \} \cup \{ Pname, Plocation \} = \{ SSN, Pnumber, Ename, Hours, Pname, Plocation \}$
- Παρατηρούμε ότι η κλειστότητα των γνωρισμάτων $\{ SSN, Pnumber \}$ περιλαμβάνει όλα τα γνωρίσματα. Άρα ο συνδυασμός $SSN, Pnumber$ αποτελεί ένα υποψήφιο κλειδί του πίνακα (της σχέσης)

Συμπέρασμα

- Το SSN δεν αποτελεί υποψήφιο κλειδί
- Το $Pnumber$ δεν αποτελεί υποψήφιο κλειδί
- Το $SSN, Pnumber$ αποτελεί υποψήφιο κλειδί γιατί περιλαμβάνει όλα τα γνωρίσματα της σχέσης

5.8.2 Παράδειγμα 2 Υπολογισμού Κλειστότητας Γνωρισμάτων

- $R = \{A, B, C, G, H, I\}$

$$\begin{aligned} F &= \{ A \rightarrow B \\ &\quad A \rightarrow C \\ &\quad CG \rightarrow H \\ &\quad CG \rightarrow I \\ &\quad B \rightarrow H \} \end{aligned}$$

- $(AG)^+ ?$

- 1. result = AG

- 2. result = ABCG ($A \rightarrow C$ and $A \rightarrow B$)
- 3. result = ABCGH ($CG \rightarrow H$ and $CG \subseteq AGBC$)
- 4. result = ABCGHI ($CG \rightarrow I$ and $CG \subseteq AGBCH$)

- Is (AG) a candidate key ?

- 1. It is a super key.

- 2. $(A^+) = BC, (G^+) = G$.

YES.

5.9 Κανονική Κάλυψη του F

Μια κανονική κάλυψη του F είναι ένα σύνολο από συναρτησιακές εξαρτήσεις F_c τέτοιο ώστε

- Το F συνεπάγεται λογικά όλες τις εξαρτήσεις του F_c
- Το F_c συνεπάγεται λογικά όλες τις εξαρτήσεις του F
- Δεν υπάρχει συναρτησιακή εξάρτηση στο F_c που να περιέχει εξωτερικό γνώρισμα
- Η αριστερή πλευρά κάθε συναρτησιακής εξάρτησης στο F_c είναι μοναδική

5.10 Ελάχιστη Κάλυψη του F

Δοθέντος ενός συνόλου F από συναρτησιακές εξαρτήσεις ελάχιστο ή κανονικό κάλυμμα είναι το ελάχιστο σύνολο ισοδύναμων εξαρτήσεων. Το χρειαζόμαστε γιατί μας διευκολύνει στην εύρεση υποψήφιων κλειδιών

5.11 Ασκήσεις με Κανονική Κάλυψη

- Παραδείγματα:

-Στο $F=\{A \rightarrow B, B \rightarrow C, A \rightarrow C\}$ η $A \rightarrow C$ πλεονάζει.

Απαλοιφή πλεονάζοντος γνωρίσματος από δεξιά μιας Σ.Ε.

-Στο $F=\{A \rightarrow B, B \rightarrow C, A \rightarrow CD\}$ η $A \rightarrow CD$ μπορεί να απλοποιηθεί στην $A \rightarrow D$. Στην περίπτωση αυτή θα λέμε ότι το γνώρισμα C είναι πλεονάζον στην $A \rightarrow CD$.

Ο λόγος είναι ο εξής:

Από την $A \rightarrow CD$ προκύπτουν οι Σ.Ε. $A \rightarrow C$ και $A \rightarrow D$ από το διασπαστικό κανόνα.

Δίνονται επίσης οι Σ.Ε. $A \rightarrow B$ και $B \rightarrow C$.

Από αυτές με το μεταβοτικό κανόνα προκύπτει ότι $A \rightarrow C$, οπότε αφού η $A \rightarrow C$ μπορεί να προκύψει αυτό σημαίνει ότι είναι περιττή και το γνώρισμα C στο οποίο καταλήγει απαλείφεται. Συνεπώς η $A \rightarrow CD$ απλοποιείται στην $A \rightarrow D$.

Η κανονική κάλυψη είναι $\{A \rightarrow B, B \rightarrow C, A \rightarrow D\}$

Απαλοιφή πλεονάζοντος γνωρίσματος από δεξιά μιας Σ.Ε. -ΟΞΙ

Παράδειγμα: Έστω $F = \{A \rightarrow C, AB \rightarrow CD\}$ και θέλουμε να δούμε αν το C είναι πλεονάζον στην $AB \rightarrow CD$

Έστω ότι το C είναι πλεονάζον και το απαλείφουμε. Προκύπτει το σύνολο $F = \{A \rightarrow C, AB \rightarrow D\}$

Από τη Σ.Ε. $A \rightarrow C$ προκύπτει ότι $A \rightarrow AC$ και $AB \rightarrow ACB$ που είναι ίδιο με το $AB \rightarrow ABC$ (1)

Από τη Σ.Ε. $AB \rightarrow D$ προκύπτει ότι $ABC \rightarrow DC$ που είναι ίδιο με το $ABC \rightarrow CD$ (2)

Με μεταβατικό κανόνα από (1) και (2) προκύπτει ότι $AB \rightarrow CD$ οπότε αφού προκύπτει και πάλι η αρχική Σ.Ε. αυτό σημαίνει ότι το C είναι πλεονάζον γνώρισμα

Απαλοιφή πλεονάζοντος γνωρίσματος από αριστερά μιας Σ.Ε.

-Στο $F = \{A \rightarrow B, B \rightarrow C, AC \rightarrow D\}$ η $AC \rightarrow D$ μπορεί να απλοποιηθεί στην $A \rightarrow D$. Ομοίως το γνώρισμα C είναι πλεονάζον στην $AC \rightarrow D$.

Ο λόγος είναι ο εξής:

Από τις Σ.Ε. $A \rightarrow B$ και $B \rightarrow C$ προκύπτει $A \rightarrow C$.

Εμπειρικός Κανόνας: Στην εξάρτηση $AC \rightarrow D$ το A είδαμε ότι προσδιορίζει το C, άρα το γνώρισμα C είναι πλεονάζον και απαλείφεται οπότε η Σ.Ε. $AC \rightarrow D$ απαλείφεται στην $A \rightarrow D$.

Η κανονική κάλυψη είναι $\{A \rightarrow B, B \rightarrow C, A \rightarrow D\}$

Απαλοιφή πλεονάζοντος γνωρίσματος από αριστερά μιας Σ.Ε.

Παράδειγμα: Έστω $F = \{A \rightarrow C, AB \rightarrow C\}$ και θέλουμε να δούμε αν το B είναι πλεονάζον στην $AB \rightarrow C$

Αφού μόνο τον το A προσδιορίζει το C το B είναι πλεονάζον γνώρισμα και απαλείφεται. Άρα $\{A \rightarrow C, A \rightarrow C\}$. Διαγράφεται η μια διπλή εξάρτηση και προκύπτει ότι η κανονική κάλυψη είναι $\{A \rightarrow C\}$

5.12 Αλγόριθμος Εύρεσης Κανονικής Κάλυψης

Αλγόριθμος Υπολογισμού F_c

repeat

Αντικατέστησε συν. εξαρτήσεις του F της μορφής
 $\alpha_1 \rightarrow \beta_1$ και $\alpha_1 \rightarrow \beta_2$ με $\alpha_1 \rightarrow \beta_1 \beta_2$ (κανόνας ένωσης).

Βρες μια συν. εξάρτηση $\alpha \rightarrow \beta$ που έχει πλεονάζον γνώρισμα είται στο α είται στο β .

Σβήσε πλεονάζοντα γνωρίσματα από το $\alpha \rightarrow \beta$.

until F does not change

5.12.1 Πλεονάζοντα Γνωρίσματα - ΟΞΙ

Έστω $\alpha \rightarrow \beta$ συναρτησιακή εξάρτηση που ανήκει στο F.

- Το γνώρισμα A είναι πλεονάζον στο α αν $A \in \alpha$ και το F παράγει το $(F - \{\alpha \rightarrow \beta\}) \cup \{(\alpha - A) \rightarrow \beta\}$.

Παράδειγμα: Έστω $F = \{A \rightarrow C, AB \rightarrow C\}$ και θέλουμε να δούμε αν το B είναι πλεονάζον στην $AB \rightarrow C$

Σύμφωνα με τον ορισμό θα έπρεπε το F να παράγει το:

$$(F - \{AB \rightarrow C\}) \cup \{(AB - B) \rightarrow C\} = \{A \rightarrow C\} \cup \{A \rightarrow C\} = \{A \rightarrow C\}$$

που προφανώς ισχύει.

Βάσεις Δεδομένων- Computer Ανάλυση

- Το γνώρισμα A είναι πλεονάζον στο β αν $A \in \beta$ και το $(F - \{\alpha \rightarrow \beta\}) \cup \{\alpha \rightarrow (\beta - A)\}$ παράγει το F .

Παράδειγμα: Έστω $F = \{A \rightarrow C, AB \rightarrow CD\}$ και θέλουμε να δούμε αν το C είναι πλεονάζον στην $AB \rightarrow CD$

Θα πρέπει το $\{A \rightarrow C, AB \rightarrow D\}$ να παράγει το F που ισχύει αφού $A \rightarrow C \Rightarrow AB \rightarrow ABC$ (1)

$AB \rightarrow D \Rightarrow ABC \rightarrow CD$ (2)

(1,2 με προσεταιριστική δίνει $AB \rightarrow CD$)

Γενικότερα αν $\alpha \rightarrow \beta$ και $\gamma \rightarrow \delta$ τότε $\alpha\gamma \rightarrow \beta\delta$

5.12.2 Παράδειγμα 1 Εύρεσης Ελάχιστης/Κανονικής Κάλυψης του συνόλου F

Δίνεται το ακόλουθο σύνολο Σ.Ε.

$AB \rightarrow C$ (1)

$A \rightarrow BC$ (2)

$B \rightarrow C$ (3)

$A \rightarrow B$ (4)

Να υπολογιστεί το ελάχιστο κάλυμμα του συνόλου των Σ.Ε.

Απάντηση

Σύμφωνα με το διασπαστικό Κανόνα η Σ.Ε. $A \rightarrow BC$ μπορεί να γραφεί ως $A \rightarrow B$ και $A \rightarrow C$ και προκύπτει το ακόλουθο σύνολο συναρτησιακών εξαρτήσεων:

$AB \rightarrow C$ (1)

$A \rightarrow B$ (2)

$A \rightarrow C$ (3)

$B \rightarrow C$ (4)

$A \rightarrow B$ (5)

Η εξάρτηση $A \rightarrow B$ γράφεται μόνο μια φορά και προκύπτει το ακόλουθο σύνολο συναρτησιακών εξαρτήσεων:

$AB \rightarrow C$ (1)

$A \rightarrow B$ (2)

$A \rightarrow C$ (3)

$B \rightarrow C$ (4)

Επειδή $AB \rightarrow C$ και $B \rightarrow C$ προκύπτει ότι το A είναι πλεονάζων γνώρισμα και απαλείφεται διότι αρκεί μόνο το γνώρισμα B για να προσδιορίσει το C . Οπότε η εξάρτηση $AB \rightarrow C$ απλοποιείται σε $B \rightarrow C$

$B \rightarrow C$ (1)

$A \rightarrow B$ (2)

$A \rightarrow C$ (3)

$B \rightarrow C$ (4)

Βάσεις Δεδομένων- Computer Ανάλυση

Η εξάρτηση $B \rightarrow C$ γράφεται μόνο μια φορά και προκύπτει το ακόλουθο σύνολο συναρτησιακών εξαρτήσεων:

$A \rightarrow B$ (1)

$A \rightarrow C$ (2)

$B \rightarrow C$ (3)

Με χρήση μεταβατικού κανόνα από τις Σ.Ε. $A \rightarrow B$ και $B \rightarrow C$ προκύπτει $A \rightarrow C$. Άρα διαγράφουμε την εξάρτηση $A \rightarrow C$ και προκύπτει το ακόλουθο τελικό σύνολο συναρτησιακών εξαρτήσεων:

$A \rightarrow B$ (1)

$B \rightarrow C$ (2)

ΑΠΟΤΕΛΕΣΜΑ	
ΠΡΙΝ	META
$AB \rightarrow C$ (1)	$A \rightarrow B$ (2')
$A \rightarrow BC$ (2)	$B \rightarrow C$ (3)
$B \rightarrow C$ (3)	
$A \rightarrow B$ (4)	

5.12.3 Παράδειγμα 2 Εύρεσης Ελάχιστης/Κανονικής Κάλυψης του F

- Δίνεται το σύνολο λειτουργικών εξαρτήσεων F για το σχήμα (Α,Β,Γ):
 - $A \rightarrow BC$
 - $B \rightarrow C$
 - $A \rightarrow B$
 - $AB \rightarrow C$

Να υπολογιστεί η κανονική κάλυψη F_c για το F

Απάντηση

Βήμα 1- Εξετάζουμε όλες τις ΣΕ με δύο γνωρίσματα στο αριστερό τους μέλος (για απαλοιφή κάποιου γνωρίσματος).

- Πρέπει να εξετάσουμε αν η ΣΕ $AB \rightarrow C$ έχει πλεονάζον γνώρισμα στο αριστερό της μέλος
- Από τις ΣΕ $A \rightarrow B$ και $B \rightarrow C$ προκύπτει ότι $A \rightarrow C$. Άρα η ΣΕ $AB \rightarrow C$ μπορεί να απλοποιηθεί σε $A \rightarrow C$. Το B είναι πλεονάζων γνώρισμα
- Το νέο σύνολο ΣΕ που προκύπτει είναι: $F' = \{A \rightarrow BC, B \rightarrow C, A \rightarrow B, A \rightarrow C\}$

Βήμα 2- Εξετάζουμε όλες τις ΣΕ με δύο γνωρίσματα στο δεξιό τους μέλος (για απαλοιφή κάποιου γνωρίσματος)

- Επειδή $A \rightarrow B$ και $A \rightarrow C$ προκύπτει ότι $A \rightarrow BC$ (ενωτικός κανόνας). Άρα η ΣΕ $A \rightarrow BC$ μπορεί να απαλειφθεί (Προκύπτει από τις προηγούμενες)
- Το νέο σύνολο ΣΕ που προκύπτει είναι: $F' = \{B \rightarrow C, A \rightarrow B, A \rightarrow C\}$

Βήμα 3- Απαλοιφή για πλεονάζουσες εξαρτήσεις

Εξετάζουμε αν υπάρχουν πλεονάζουσες ΣΕ στο F'

- Επειδή $A \rightarrow B$ και $B \rightarrow C$ προκύπτει ότι $A \rightarrow C$. Άρα η ΣΕ $A \rightarrow C$ μπορεί να απαλειφθεί (Προκύπτει από τις προηγούμενες)
- Η Ελάχιστη Κάλυψη (Κάλυμμα) είναι $F^c = \{A \rightarrow B, B \rightarrow C\}$

Συμπέρασμα: Όλες οι Σ.Ε. που υπονοούνται απαλείφονται

5.12.4 Παράδειγμα 3 Εύρεσης Ελάχιστης/Κανονικής Κάλυψης του F

Έστω το σύνολο των ΣΕ $F = \{A \rightarrow BC, B \rightarrow AC, C \rightarrow AB\}$. Να υπολογιστεί η κανονική/ελάχιστη κάλυψη F_c για το F.

Απάντηση

Βήμα 1

Το Βήμα αυτό παραλείπεται διότι δεν υπάρχουν ΣΕ με δύο γνωρίσματα στο αριστερό τους μέλος

Βήμα 2

Σύμφωνα με το Διασπαστικό Κανόνα $X \rightarrow YZ$ έπεται ότι $X \rightarrow Y$ και $X \rightarrow Z$ και προκύπτουν τα εξής:

- Από την ΣΕ $A \rightarrow BC$ συνάγεται ότι $A \rightarrow B$ και $A \rightarrow C$
- Από την ΣΕ $B \rightarrow AC$ συνάγεται ότι $B \rightarrow A$ και $B \rightarrow C$
- Από την ΣΕ $C \rightarrow AB$ συνάγεται ότι $C \rightarrow A$ και $C \rightarrow B$

Βάσεις Δεδομένων- Computer Ανάλυση

- Επειδή $B \rightarrow C$ προκύπτει ότι στη ΣΕ $A \rightarrow BC$ το C είναι πλεονάζον γνώρισμα και απαλείφεται. Το σύνολο F μειώνεται στο $F_c = \{A \rightarrow B, B \rightarrow AC, C \rightarrow AB\}$
- Επειδή $A \rightarrow C$ προκύπτει ότι στη Σ.Ε. $B \rightarrow AC$, το C πλεονάζον γνώρισμα και απαλείφεται. Το σύνολο F μειώνεται στο $F_c = \{A \rightarrow B, B \rightarrow A, C \rightarrow AB\}$
- Επειδή $A \rightarrow B$ προκύπτει ότι στη Σ.Ε. $C \rightarrow AB$, το B είναι πλεονάζον γνώρισμα και απαλείφεται. Το σύνολο F μειώνεται στο τελικό $F_c = \{A \rightarrow B, B \rightarrow A, C \rightarrow A\}$

5.13 Κανονικές Μορφές

- 1 NF (Normal Form)
- 2 NF
- BCNF (Boyce-Codd NF)
- 3NF
- Ένα σχήμα σχέσης R είναι σε 1^η κανονική μορφή (σε 1 NF) αν δεν περιέχει πλειότιμα γνωρίσματα
- Ένα σχήμα σχέσης R είναι σε 2^η κανονική μορφή (σε 2 NF) αν είναι σε 1 NF και το κλειδί είναι είτε είναι απλό γνώρισμα είτε αν το κλειδί είναι σύνθετο γνώρισμα, τότε τα υπόλοιπα γνωρίσματα που δεν ανήκουν στο κλειδί θα πρέπει να προσδιορίζονται συναρτησιακά από όλο το κλειδί και όχι μόνο από ένα μέρος του
- Ένα σχήμα σχέσης R είναι σε 3^η κανονική μορφή (3 NF) αν είναι σε 2 NF και ΔΕΝ περιέχει μεταβατικές εξαρτήσεις. Μεταβατική εξάρτηση είναι η εξάρτηση όπου ένα γνώρισμα που δεν ανήκει στο κλειδί προσδιορίζει κάποιο άλλο γνώρισμα (ή κάποια άλλα γνωρίσματα)
- Ένα σχήμα σχέσης R είναι σε BCNF αν είναι σε 3NF και επιπλέον σε κάθε ΣΕ της μορφής $X \rightarrow A$ της R το X είναι υπερκλειδί της R

5.13.1 Παραδείγματα Κανονικοποίησεων

5.13.1.1 Κανονικοποίηση σε 1 NF

ΟΝΟΜΑ_ΤΜ	ΚΩΔ_ΤΜ	ΔΙΕΥΘΥΝΤΗΣ	ΤΟΠΟΘΕΣΙΕΣ_ΤΜ
Έρευνα	5	12345678	{Αθήνα, Πάτρα, Τρίπολη}
Διοίκηση	4	23456789	{Τρίπολη}
Πωλήσεις	1	34567890	{Αθήνα}

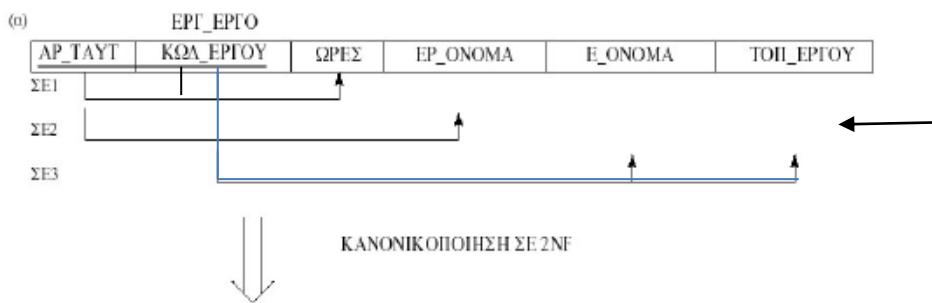
- Ισχύει $\text{ΟΝΟΜΑ_ΤΜ} \rightarrow \text{ΤΟΠΟΘΕΣΙΕΣ_ΤΜ}$ αλλά δεν έχουμε 1NF.

Η σχέση δεν είναι σε 1 NF διότι περιέχει το πλειότιμο γνώρισμα ΤΟΠΟΘΕΣΙΕΣ_ΤΜ

ΟΝΟΜΑ_ΤΜ	ΚΩΔ_ΤΜ	ΔΙΕΥΘΥΝΤΗΣ	ΤΟΠΟΘΕΣΙΕΣ_ΤΜ
Έρευνα	5	12345678	Αθήνα
Έρευνα	5	12345678	Πάτρα
Έρευνα	5	12345678	Τρίπολη
Διοίκηση	4	23456789	Τρίπολη
Πωλήσεις	1	34567890	Αθήνα

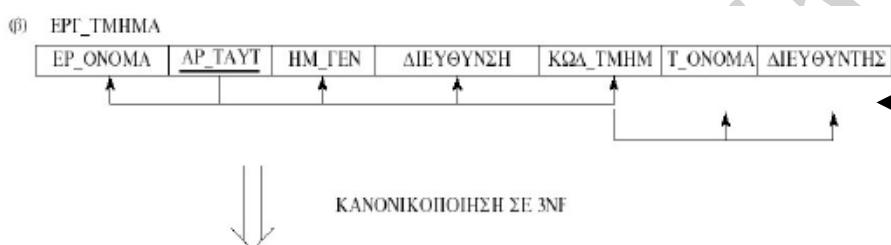
Η κανονικοποίηση γίνεται δημιουργώντας ξεχωριστή εγγραφή για κάθε τιμή του πλειότιμου γνωρίσματος. Η νέα σχέση που προκύπτει είναι σε 1 NF

5.13.1.2 Κανονικοποίηση σε 2 NF



Η σχέση αυτή δεν είναι σε 2NF διότι η ΣΕ2 και ΣΕ3 παραβιάζουν την 2NF. Συγκεκριμένα τα γνωρίσματα ΕΡ_ΟΝΟΜΑ, Ε_ΟΝΟΜΑ και ΤΟΠ_ΕΡΓΟΥ δεν προσδιορίζονται από όλο το κλειδί αλλά μόνο από ένα μέρος του.

5.13.1.3 Κανονικοποίηση σε 3NF



Κάθε ΣΕ μετατρέπεται σε ξεχωριστή σχέση. Η κάθε σχέση που προκύπτει είναι σε 2 NF



Η σχέση αυτή δεν είναι σε 3NF διότι το γνώρισμα ΚΩΔ_ΤΜΗΜ που δεν αποτελεί μέρος του κλειδιού προσδιορίζει συναρτησιακά τα γνωρίσματα Τ_ΟΝΟΜΑ και ΔΙΕΥΘΥΝΤΗΣ. Η εξάρτηση αυτή ονομάζεται μεταβατική

Κάθε μεταβατική εξάρτηση μετατρέπεται σε ξεχωριστή σχέση. Η κάθε σχέση που προκύπτει είναι σε 3 NF

Παράδειγμα σχέσης που δεν είναι σε 3NF

Δίνεται η σχέση EMP_DEPT(firstName, employeeNumber, dateOfBirth, address, departmentNumber, departmentName). Το υπόψηφο κλειδί είναι το **employeeNumber**.

Ας θεωρήσουμε τις ακόλουθες Σ.Ε.

- employeeNumber → firstName, dateOfBirth, address, departmentNumber
- departmentNumber → departmentName

Η σχέση EMP_DEPT ΔΕΝ είναι σε 3NF διότι η 2^η Σ.Ε. είναι μεταβατική

Διόρθωση

EMP_DEPT(firstName, employeeNumber, dateOfBirth, address, departmentNumber).

NEW_DEPT(departmentNumber, departmentName)

Παράδειγμα Κανονικοποίησης σε 3NF

Φοιτητής (αριθ_φοιτητή, όνομα_φοιτητή)

Τάξη (αριθ_μαθήματος, ώρα_μαθήματος, αριθ_κτιρίου, όνομα_κτιρίου)

Φοιτητής_Τάξη (αριθ_φοιτητή, αριθ_μαθήματος)

Στο προηγούμενο σχήμα έχουμε τις εξαρτήσεις:

- αριθ_μαθήματος → αριθ_κτιρίου
- αριθ_κτιρίου → όνομα_κτιρίου

που αποτελούν μία μεταβατική εξάρτηση

Η κανονικοποίηση σε 3NF δίνει το ακόλουθο σχήμα:

Φοιτητής (αριθ_φοιτητή, όνομα_φοιτητή)

Τάξη (αριθ_μαθήματος, ώρα_μαθήματος, αριθ_κτιρίου)

Κτίριο (αριθ_κτιρίου, όνομα_κτιρίου)

Φοιτητής_Τάξη (αριθ_φοιτητή, αριθ_μαθήματος)

5.13.1.4 Ασκηση Κανονικοποίησης σε 2NF και 3 NF με εύρεση υποψήφιου Κλειδιού

- Θεωρείστε τη σχέση $R = \{A, B, C, D, E, F, G, H, I, J\}$ και το σύνολο των συναρτησιακών εξαρτήσεων
 $F = \{\{A\} \rightarrow \{D, E\}, \{A, B\} \rightarrow \{H, J\}, \{E\} \rightarrow \{F\}, \{B\} \rightarrow \{C\}\}$.
 - 3a. Να βρείτε ένα πρωτεύον κλειδί για την R
 - 3b. Να κανονικοποιηθεί η R σε 2NF
 - 3c. Να κανονικοποιηθεί η R σε 3NF

Λύση:

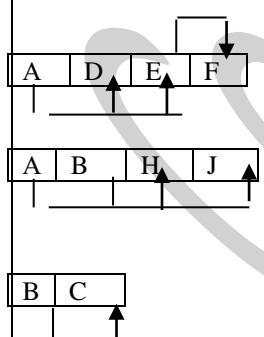
- 3a. Προφανές Υπερκλειδί για την R είναι το $K := \{A, B, C, D, E, F, G, H, I, J\}$
 Δεν μπορούμε να αφαιρέσουμε τα A, B, G, I
 Τα D, E μπορούν να αφαιρεθούν διότι $A \rightarrow D, E$
 Τα H, J μπορούν να αφαιρεθούν διότι $A, B \rightarrow H, J$
 Επίσης το F διότι $E \rightarrow F$ και το C διότι $B \rightarrow C$
 Η κλειστότητα $(A, B, G, I)^+$ με βάση την F είναι $\{A, B, C, D, E, F, G, H, I, J\}$ και επομένως το $\{A, B, G, I\}$ είναι ένα υποψήφιο κλειδί για την R
- 3b. Κανονικοποίηση σε 2NF
 $R1(A, D, E, F), R2(A, B, H, J) \text{ και } R3(B, C)$
- 3c. Κανονικοποίηση σε 3NF
 $R1(A, D, E), R2(E, F), R3(A, B, H, J) \text{ και } R4(B, C)$

Παρατηρήσεις για την Κανονικοποίηση στο παράδειγμα αυτό

Για Κανονικοποίηση σε 2 NF μετατρέπω κάθε Σ.Ε. που δίνεται:

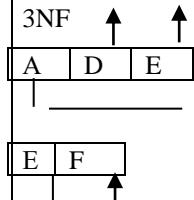
- $A \rightarrow D, E$
- $A, B \rightarrow H, J$
- $E \rightarrow F$
- $B \rightarrow C$

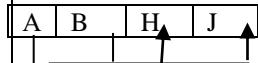
σε σχέση (πίνακα). Στη 2NF μπορώ να έχω μεταβατικές εξαρτήσεις άρα:



- Οι σχέσεις σε 2NF είναι οι εξής: $R1(A, D, E, F), R2(A, B, H, J), R3(B, C)$ και $R4(G, I)$

Για Κανονικοποίηση σε 3 NF απαλείφω τη μεταβατική εξάρτηση $E \rightarrow F$ από τη πρώτη σχέση και προκύπτουν οι τελικές σχέσεις σε 3NF





- Οι σχέσεις σε 3NF είναι οι εξής: R1(A, D, E), R2(E, F) και R3(A, B, H, J), R4(B, C), R5 (G, I)

5.13.1.5 Κανονικοποίηση σε 3 NF με διαφορετικό τρόπο

Δίνεται η σχέση $R=(A, B, C, D, E, F, G)$ και το ακόλουθο ελάχιστο κάλυμμα: $F_c=\{A\rightarrow CEF, D\rightarrow G, C\rightarrow D\}$. Δίνεται επίσης ότι το μοναδικό υποψήφιο κλειδί είναι το AB.

Για κανονικοποίηση σε 3NF, η αρχική σχέση R διασπάται σε κάθε μια από τις Σ.Ε. του F_c δίνοντας τους εξής πίνακες:

- $R1=(\underline{A}, C, E, F)$
- $R2=(\underline{D}, G)$
- $R3=(\underline{C}, D)$

Εφόσον το μοναδικό υποψήφιο κλειδί δεν υπάρχει σε κάποιον από τους πίνακες που προέκυψαν το προσθέτουμε ως ξεχωριστή σχέση. Η τελική διάσπαση σε 3NF είναι:

- $R1=(\underline{A}, C, E, F)$
- $R2=(\underline{D}, G)$
- $R3=(\underline{C}, D)$
- $R4=(\underline{A}, \underline{B})$

Η κάθε μια από τις σχέσεις R1, R2, R3, R4 είναι σε 3NF

5.13.1.6 Μέθοδοι Κανονικοποίησης σε 3NF

1) Απαλείφουμε τις μεταβατικές εξαρτήσεις από τις σχέσεις που είναι σε 2NF δημιουργώντας για κάθε μεταβατική εξάρτηση ένα ξεχωριστό πίνακα (σχέση)

2) Υπολογίζουμε το ελάχιστο κάλυμμα F_c από το σύνολο F των συναρτησιακών εξαρτήσεων. Στη συνέχεια δημιουργούμε για κάθε εξάρτηση του ελάχιστου καλύμματος F_c μια ξεχωριστή σχέση (ένα ξεχωριστό πίνακα). Αν το υποψήφιο κλειδί δεν συμπεριλαμβάνεται σε κάποια σχέση τότε το βάζουμε μόνο του σε μια νέα σχέση. Η καθεμία από αυτές τις σχέσεις είναι σε 3NF.

5.13.1.7 Κανονικοποίηση σε BCNF

Μια σχέση είναι σε BCNF αν για κάθε Σ.Ε. $A\rightarrow B$, το A είναι υπερκλειδί στην R . Κάθε σχέση σε BCNF είναι και σε 3NF αλλά το αντίθετο δεν ισχύει. Συνεπώς μια σχέση R είναι σε BCNF αν σε κάθε συναρτησιακή εξάρτηση $A\rightarrow B$ το A είναι υπερκλειδί. Αν σε μία σχέση R η εξάρτηση $A\rightarrow B$ παραβιάζει τη BCNF, τότε διασπάμε τη σχέση R σε δύο σχέσεις:

- R_1 με πεδία τα $A\cup B$ και κλειδί το A και
- R_2 με πεδία τα $A\cup(R-A-B)$ και κλειδί όλα τα πεδία

Παράδειγμα 1 Κανονικοποίηση σε BCNF

Στη σχέση $R=(\text{Φοιτητής}, \text{Μάθημα Διδάσκων})$ η εξάρτηση Διδάσκων \rightarrow Μάθημα παραβιάζει τη BCNF διότι ο Διδάσκων δεν είναι κλειδί. Η κανονικοποίηση σε BCNF γίνεται ως εξής:

- $R1=(\text{Διδάσκων}, \text{Μάθημα})$ με κλειδί το Διδάσκων
- $R2=(\text{Διδάσκων}, \text{Φοιτητής})$ με κλειδί όλα τα γνωρίσματα
- Καθεμία από τις σχέσεις $R1$ και $R2$ είναι σε BCNF.

Παράδειγμα 2 Κανονικοποίηση σε BCNF

Δίνεται η σχέση STUDENT_COURSE(studentNumber, socialSecurityNumber, courseNumber). Ένας φοιτητής μπορεί να παρακολουθεί πολλά courses και ένα course μπορεί να έχει πολλούς φοιτητές. Τα υποψήφια κλειδιά είναι:

1. socialSecurityNumber, courseNumber
2. studentNumber, courseNumber

Δίνονται οι ακόλουθες Σ.Ε.

1. studentNumber \rightarrow socialSecurityNumber
2. socialSecurityNumber \rightarrow studentNumber

Μπορούμε να συμπεράνουμε ότι η σχέση STUDENT_COURSE ΔΕΝ είναι σε BCNF επειδή το studentNumber δεν είναι κλειδί ή υπερκλειδί στη σχέση STUDENT_COURSE. Η κανονικοποίηση σε BCNF γίνεται ως εξής:

- $R_1=(\text{studentNumber}, \text{socialSecurityNumber})$ με κλειδί το studentNumber
- $R_2=(\text{studentNumber}, \text{courseNumber})$ με κλειδί όλα τα γνωρίσματα
- Καθεμία από τις σχέσεις $R1$ και $R2$ είναι σε BCNF.

5.14 Διαχωρισμός (Αποσύνθεση) Χωρίς Απώλειες Συνδέσμου

Μια διάσπαση της σχέσης R στις σχέσεις R₁ και R₂ είναι Χωρίς Απώλειες Συνδέσμου (lossless join decomposition) αν

$$R_1 \cap R_2 \rightarrow R_1 \text{ είτε αν } R_1 \cap R_2 \rightarrow R_2$$

δηλ. αν τα γνωρίσματα της R₁ ∩ R₂ είναι κλειδί είτε για την R₁ ή για την R₂.

5.14.1 Παράδειγμα 1 με Διαχωρισμό Χωρίς Απώλειες Συνδέσμου

- 1: Υποθέστε ότι αποσυνθέτουμε το σχήμα
R=(A,B,C,D,E) σε

- R₁ (A,B,C)
- R₂ (A,D,E)

Δείξτε ότι αυτή η αποσύνθεση είναι η αποσύνθεση χωρίς απώλειες συνδέσμου, αν ισχύει το παρακάτω σύνολο F από λειτουργικές εξαρτήσεις:

$$\begin{aligned} A &\rightarrow BC \\ CD &\rightarrow E \\ B &\rightarrow D \\ E &\rightarrow A \end{aligned}$$

Απάντηση

- Η τομή των σχέσεων R₁ =(A, B, C) και R₂ =(A, D E) είναι το γνώρισμα A δηλ. R₁ ∩ R₂ = {A}. Εξετάζουμε αν το γνώρισμα αποτελεί κλειδί είτε για την R₁ ή την R₂
- Λόγω της ΣΕ A → BC το A είναι κλειδί για το σχήμα σχέσης R₁
- Άρα η αποσύνθεση είναι χωρίς απώλειες συνδέσμου

5.14.2 Παράδειγμα 2 με Διαχωρισμό Χωρίς Απώλειες Συνδέσμου

Να εξετάσετε αν η αποσύνθεση της σχέσης R στις σχέσεις R₁ και R₂ είναι χωρίς απώλειες συνδέσμου.

Παράδειγμα: R = {Τίτλος, Έτος, Διάρκεια, Είδος, Όνομα-Ηθοποιού, Διεύθυνση, Έτος-Γέννησης}

Τίτλος Έτος → Διάρκεια	R ₁ ∩ R ₂ = {Τίτλος, Έτος}
Τίτλος Έτος → Είδος	<u>Υπερκλειδί</u> για την R ₁ άρα
Όνομα Ηθοποιού → Διεύθυνση	αποσύνθεση χωρίς απώλειες συνδέσμου
Όνομα-Ηθοποιού → Έτος Γέννησης	

$$R_1 = \{\text{Τίτλος, Έτος, Διάρκεια, Είδος}\}$$

$$R_2 = \{\text{Τίτλος, Έτος, Όνομα-Ηθοποιού, Διεύθυνση, Έτος-Γέννησης}\}$$

5.15 Διατήρηση Εξαρτήσεων

Μια αποσύνθεση της R σε R₁ και R₂ διατηρεί τις εξαρτήσεις αν οι υπάρχουσες εξαρτήσεις δεν εκτείνονται στις R₁ και R₂

5.15.1 Παράδειγμα με Διατήρηση Εξαρτήσεων

- το σχήμα σχέσης R(A, B, C, D)
- το σύνολο συναρτησιακών εξαρτήσεων
 - A → C
 - B → C
 - B → A
- Έστω η αποσύνθεση S(A, C) και T(A, B, D)

Να εξετάσετε αν η αποσύνθεση διατηρεί τις εξαρτήσεις.

Απάντηση

Παρατηρούμε ότι:

- η εξάρτηση A → C βρίσκεται στη σχέση S(A, C)
- η εξάρτηση B → C εκτείνεται και στις δύο σχέσεις S(A, C) και T(A, B, D)
- Άρα η αποσύνθεση δεν διατηρεί τις εξαρτήσεις

5.16 ΘΕΜΑΤΑ ΜΕ ΚΑΝΟΝΙΚΕΣ ΜΟΡΦΕΣ, ΣΥΝΑΡΤΗΣΙΑΚΕΣ ΕΞΑΡΤΗΣΕΙΣ, ΕΛΑΧΙΣΤΗ ΚΑΛΥΨΗ Κ.Λ.Π.**5.16.1 Θέμα 4 Σεπτέμβριος 2009**

Θεωρείστε το σχήμα:

AM
ONOMA_ΦΟΙΤΗΤΗ
ΕΤΟΣ
ΕΞΑΜΗΝΟ
ΚΩΔ_ΜΑΘ
ONOMA_ΜΑΘΗΜΑΤΟΣ
ΒΑΘΜΟΣ

και το ακόλουθο σύνολο συναρτησιακών εξαρτήσεων F:

- $AM \rightarrow ONOMA_ΦΟΙΤΗΤΗ$
- $AM, KΩΔ_ΜΑΘΗΜΑΤΟΣ, ΕΞΑΜΗΝΟ, ΕΤΟΣ \rightarrow ΒΑΘΜΟΣ$
- $KΩΔ_ΜΑΘΗΜΑΤΟΣ \rightarrow ONOMA_ΜΑΘΗΜΑΤΟΣ$

i) Να διασπαστεί το παραπάνω σχήμα ώστε κάθε σχήμα που θα προκύψει να πληροί την BCNF σε σχέση με το σύνολο συναρτησιακών εξαρτήσεων F. Εξηγήστε σύντομα γιατί τα σχήματα που προτείνετε είναι σε BCNF.

ii) Το σχήμα που έχει προκύψει πληροί την 3NF; Δικαιολογήστε.

Απάντηση**Θεωρία BCNF**

Αν σε μία σχέση R η εξάρτηση $A \rightarrow B$ παραβιάζει τη BCNF, τότε διασπούμε τη σχέση (πίνακα) R σε δύο επιμέρους σχέσεις (πίνακες) ως εξής:

- R_1 με πεδία τα $A \cup B$ και κλειδί το A και
 - R_2 με πεδία τα $A \cup (R - A - B)$ και κλειδί όλα τα πεδία
- i)
- Στην εξάρτηση $AM \rightarrow ONOMA_ΦΟΙΤΗΤΗ$ το AM δεν είναι κλειδί διότι η κλειστότητα $\{AM\}^+ = \{AM, ONOMA_ΦΟΙΤΗΤΗ\}$ δεν περιλαμβάνει όλα τα γνωρίσματα της αρχικής σχέσης. Άρα η εξάρτηση $AM \rightarrow ONOMA_ΦΟΙΤΗΤΗ$ παραβιάζει τη BCNF. Οπότε η αρχική σχέση διασπάται σε $R_1 = \{AM, ONOMA_ΦΟΙΤΗΤΗ\}$ και $R_2 = \{KΩΔ_ΜΑΘΗΜΑΤΟΣ, ΕΞΑΜΗΝΟ, ΕΤΟΣ, ΒΑΘΜΟΣ, ONOMA_ΜΑΘΗΜΑΤΟΣ\}$
 - Στην εξάρτηση $KΩΔ_ΜΑΘΗΜΑΤΟΣ \rightarrow ONOMA_ΜΑΘΗΜΑΤΟΣ$ το πεδίο $KΩΔ_ΜΑΘΗΜΑΤΟΣ$ δεν είναι κλειδί διότι η κλειστότητα $\{KΩΔ_ΜΑΘΗΜΑΤΟΣ\}^+ = \{KΩΔ_ΜΑΘΗΜΑΤΟΣ, ONOMA_ΜΑΘΗΜΑΤΟΣ\}$ δεν περιλαμβάνει όλα τα γνωρίσματα της σχέσης R2. Άρα η $KΩΔ_ΜΑΘΗΜΑΤΟΣ \rightarrow ONOMA_ΜΑΘΗΜΑΤΟΣ$ παραβιάζει τη BCNF. Συνεπώς η σχέση R2 διασπάται σε $R_3 = \{KΩΔ_ΜΑΘΗΜΑΤΟΣ, ONOMA_ΜΑΘΗΜΑΤΟΣ\}$ και σε $R_4 = \{KΩΔ_ΜΑΘΗΜΑΤΟΣ, AM, ΕΞΑΜΗΝΟ, ΕΤΟΣ, ΒΑΘΜΟΣ\}$

Βάσεις Δεδομένων- Computer Ανάλυση

- Στην εξάρτηση ΑΜ, ΚΩΔ_ΜΑΘΗΜΑΤΟΣ, ΕΞΑΜΗΝΟ, ΕΤΟΣ → ΒΑΘΜΟΣ τα πεδία ΑΜ, ΚΩΔ_ΜΑΘΗΜΑΤΟΣ, ΕΞΑΜΗΝΟ, ΕΤΟΣ αποτελούν κλειδί διότι η κλειστότητα {ΑΜ, ΚΩΔ_ΜΑΘΗΜΑΤΟΣ, ΕΞΑΜΗΝΟ, ΕΤΟΣ}+= {ΑΜ, ΚΩΔ_ΜΑΘΗΜΑΤΟΣ, ΕΞΑΜΗΝΟ, ΕΤΟΣ, ΒΑΘΜΟΣ} ΔΕΝ περιλαμβάνει όλα τα γνωρίσματα της σχέσης R οπότε η εξάρτηση αυτή παραβιάζει τη BCNF και χρειάζεται περαιτέρω διάσπαση σε $R_5=\{\underline{\text{ΑΜ, ΚΩΔ ΜΑΘΗΜΑΤΟΣ, ΕΞΑΜΗΝΟ, ΕΤΟΣ, ΒΑΘΜΟΣ}}\}$ και $R_6=\{\underline{\text{ΑΜ, ΚΩΔ ΜΑΘΗΜΑΤΟΣ, ΟΝΟΜΑ ΦΟΙΤΗΤΗ, ΟΝΟΜΑ ΜΑΘΗΜΑΤΟΣ}}\}$

- Άρα η τελική αποσύνθεση σε BCNF είναι οι σχέσεις:

$$R_1=\{\underline{\text{ΑΜ, ΟΝΟΜΑ ΦΟΙΤΗΤΗ}}\}$$

$$R_2=\{\underline{\text{ΑΜ, ΕΞΑΜΗΝΟ, ΕΤΟΣ, ΒΑΘΜΟΣ, ΚΩΔ ΜΑΘΗΜΑΤΟΣ, ΟΝΟΜΑ ΜΑΘΗΜΑΤΟΣ}}\}$$

$$R_3=\{\underline{\text{ΚΩΔ ΜΑΘΗΜΑΤΟΣ, ΟΝΟΜΑ ΜΑΘΗΜΑΤΟΣ}}\}$$

$$R_4=\{\underline{\text{ΑΜ, ΕΞΑΜΗΝΟ, ΕΤΟΣ, ΒΑΘΜΟΣ, ΚΩΔ ΜΑΘΗΜΑΤΟΣ, ΟΝΟΜΑ ΦΟΙΤΗΤΗ}}\}$$

$$R_5=\{\underline{\text{ΚΩΔ ΜΑΘΗΜΑΤΟΣ, ΑΜ, ΕΞΑΜΗΝΟ, ΕΤΟΣ, ΒΑΘΜΟΣ}}\}$$

$$R_6=\{\underline{\text{ΑΜ, ΚΩΔ ΜΑΘΗΜΑΤΟΣ, ΕΞΑΜΗΝΟ, ΕΤΟΣ ΟΝΟΜΑ ΦΟΙΤΗΤΗ, ΟΝΟΜΑ ΜΑΘΗΜΑΤΟΣ}}\}$$

ii) Το σχήμα που έχει προκύψει πληροί την 3NF αφού όλες οι σχέσεις δεν περιέχουν μεταβατικές εξαρτήσεις.

5.16.2 Θέμα 4 Φεβρουάριος 2013

Θεωρήστε το σχήμα $R=(A, B, C, D, E)$ και το ακόλουθο σύνολο συναρτησιακών εξαρτήσεων F:

- $A \rightarrow B$
- $BC \rightarrow E$
- $ED \rightarrow A$

Ερωτήσεις:

- Υπολογίστε την κλειστότητα των συνόλων των γνωρισμάτων AC, CE, CDE, BCD και ACD. Ποια από τα παραπάνω σύνολα γνωρισμάτων αποτελούν υποψήφια κλειδιά του παραπάνω σχήματος; Δικαιολογήστε την απάντηση σας.
- Έστω το σχήμα R διασπάται σε $R_1(A, B, D)$ και $R_2(B, C, E)$
 - Αποτελεί ο παραπάνω διαχωρισμός ένα διαχωρισμό χωρίς απώλειες συνδέσμου;
 - Διατηρούνται οι εξαρτήσεις στον παραπάνω διαχωρισμό;

Απάντηση

a.

$$\text{Υπολογισμός κλειστότητας } \{A, C\}^+$$

- $\{A, C\}^+ = \{A, C\} \rightarrow$ στην αρχή η κλειστότητα περιλαμβάνει τα ίδια τα γνωρίσματα
- $A \rightarrow B$ άρα $\{A, C\}^+ = \{A, C\} \cup \{B\} = \{A, C, B\}$
- $BC \rightarrow E$ άρα $\{A, C\}^+ = \{A, C, B\} \cup \{E\} = \{A, B, C, E\}$

Το $\{A, C\}^+$ δεν μπορεί να αλλάξει και ΔΕΝ περιλαμβάνει όλα τα γνωρίσματα του σχήματος. Συνεπώς το AC ΔΕΝ αποτελεί υποψήφιο κλειδί.

Υπολογισμός κλειστότητας $\{C, E\}^+$

$$1. \{C, E\}^+ = \{C, E\}$$

Το $\{C, E\}^+$ δεν μπορεί να αλλάξει και ΔΕΝ περιλαμβάνει όλα τα γνωρίσματα του σχήματος. Συνεπώς το CE ΔΕΝ αποτελεί υποψήφιο κλειδί

Υπολογισμός κλειστότητας $\{C, D, E\}^+$

$$1. \{C, D, E\}^+ = CDE$$

$$2. ED \rightarrow A \text{ áρα } \{C, D, E\}^+ = CDE \cup A$$

$$3. A \rightarrow B \text{ áρα } \{C, D, E\}^+ = CDEA \cup B$$

Άρα το $\{C, D, E\}^+$ περιλαμβάνει όλα τα γνωρίσματα του σχήματος και συνεπώς το CDE αποτελεί υποψήφιο κλειδί

Υπολογισμός κλειστότητας $\{B, C, D\}^+$

$$1. \{B, C, D\}^+ = BCD$$

$$2. BC \rightarrow E \text{ áρα } \{B, C, D\}^+ = BCD \cup E$$

$$3. ED \rightarrow A \text{ áρα } \{B, C, D\}^+ = BCDE \cup A$$

Άρα το $\{B, C, D\}^+$ περιλαμβάνει όλα τα γνωρίσματα του σχήματος και συνεπώς το BCD αποτελεί υποψήφιο κλειδί

Υπολογισμός κλειστότητας $\{A, C, D\}^+$

$$1. \{A, C, D\}^+ = ACD$$

$$2. A \rightarrow B \text{ áρα } \{A, C, D\}^+ = ACD \cup B$$

$$3. BC \rightarrow E \text{ áρα } \{A, C, D\}^+ = ACDBE$$

Άρα το $\{A, C, D\}^+$ περιλαμβάνει όλα τα γνωρίσματα του σχήματος και συνεπώς το ACD αποτελεί υποψήφιο κλειδί

b.

i. Ο διαχωρισμός ενός σχήματος R σε δύο σχήματα $R_1(A, B, D)$ και $R_2(B, C, E)$ είναι χωρίς απώλειες (lossless join) αν ισχύει ένα από τα ακόλουθα:

$$R_1 \cap R_2 \rightarrow R_1 \text{ ή } R_1 \cap R_2 \rightarrow R_2$$

Οι δοθείσες συναρτησιακές εξαρτήσεις είναι:

- $A \rightarrow B$
- $BC \rightarrow E$
- $ED \rightarrow A$

Προκύπτει ότι $R_1 \cap R_2 = \{B\}$. Το B δεν προσδιορίζει συναρτησιακά ούτε την $R_1 = \{A, B, D\}$ ούτε την $R_2 = \{B, C, E\}$ και συνεπώς ο διαχωρισμός δεν είναι χωρίς απώλειες (είναι με απώλειες)

Δίνεται το σχήμα σχέσης R το οποίο διασπάται σε $R_1=\{A, B, D\}$ και $R_2=\{B, C, E\}$. Η διάσπαση αυτή διατηρεί τις εξαρτήσεις αν κάθε προϋπάρχουσα συναρτησιακή εξάρτηση ανήκει αποκλειστικά σε μια σχέση (δηλ. δεν εκτείνεται σε δύο σχέσεις). Οι δοθείσες συναρτησιακές εξαρτήσεις είναι:

- $A \rightarrow B$
- $BC \rightarrow E$
- $ED \rightarrow A$
- Παρατηρούμε ότι η εξάρτηση $ED \rightarrow A$ εκτείνεται και στις δύο σχέσεις R_1 και R_2 , άρα οι εξαρτήσεις δεν διατηρούνται

Υπενθύμιση

- ✓ Ένας διαχωρισμός ενός σχήματος R σε 2 σχήματα R_1 και R_2 είναι χωρίς απώλειες συνδέσμου (lossless JOIN decomposition) αν ισχύει ένα από τα ακόλουθα: $R_1 \cap R_2 \rightarrow R_1 \wedge R_1 \cap R_2 \rightarrow R_2$
- ✓ Δίνεται ένα σχήμα σχέσης R το οποίο διασπάται σε R_1, R_2, R_3, \dots κ.λ.π. Η διάσπαση αυτή διατηρεί τις εξαρτήσεις αν κάθε προϋπάρχουσα συναρτησιακή εξάρτηση ανήκει αποκλειστικά σε μια σχέση.

5.16.3 Θέμα 6 Σεπτέμβριος 2008

Θεωρήστε το σχήμα $R=(A, B, C, D, E)$ και το ακόλουθο σύνολο συναρτησιακών εξαρτήσεων F :

1. $A \rightarrow BC$
2. $CD \rightarrow E$
3. $B \rightarrow D$
4. $E \rightarrow B$

Με βάση τα παραπάνω, μπορεί να θεωρηθεί ο συνδυασμός γνωρισμάτων CD ως υποψήφιο κλειδί; Δικαιολογήστε την άποψη σας

Απάντηση

Υπολογισμός κλειστότητας $\{C, D\}^+$

1. $\{C, D\}^+ = CD$
2. $CD \rightarrow E$ áρα $\{C, D\}^+ = CD \cup E$
3. $E \rightarrow B$ áρα $\{C, D\}^+ = CDE \cup B$

Παρατηρούμε ότι το $\{C, D\}^+$ δεν μπορεί να αλλάξει και δεν περιλαμβάνει όλα τα γνωρίσματα του σχήματος, áρα το CD δεν αποτελεί υποψήφιο κλειδί

5.16.4 Θέμα 3 Ιανουάριος 2014 και Ιανουάριος 2020

Θεωρήστε το σχήμα σχέσης $R(A, B, C, D, E)$ και το ακόλουθο σύνολο συναρτησιακών εξαρτήσεων F :

- AC \rightarrow D
B \rightarrow CE
C \rightarrow A
D \rightarrow B

- a. Αποτελεί το γνώρισμα D υποψήφιο κλειδί; Δικαιολογήστε την άποψη σας
- b. Είναι ο διαχωρισμός σε $R_1=(A, B, D)$ και $R_2=(A, B, C, E)$ χωρίς απώλειες συνδέσμου (lossless-join decomposition); Δικαιολογήστε την άποψη σας. Διατηρούνται οι εξαρτήσεις;
- c. Αποσυνθέστε τη σχέση R ώστε οι σχέσεις που θα προκύψουν να βρίσκονται σε 3NF (3^η κανονική μορφή)

Απάντηση

A Ερώτημα

Υπολογισμός κλειστότητας $\{D\}^+$

1. $\{D\}^+ = D$
2. $D \rightarrow B$ áρα $\{D\}^+ = D \cup B = DB$
3. $B \rightarrow CE$ áρα $\{D\}^+ = DB \cup CE = DBCE$
4. $C \rightarrow A$ áρα $\{D\}^+ = DBCE \cup A = DBCEA$

Άρα το $\{D\}^+$ συμπεριλαμβάνει όλα τα γνωρίσματα του σχήματος R και συνεπώς το D αποτελεί υποψήφιο κλειδί

Β Ερώτημα

$R_1 = \{A, B, D\}$ και $R_2 = \{A, B, C, E\}$. Η $R_1 \cap R_2 = \{A, B\}$ δεν προσδιορίζει συναρτησιακά ούτε το R_1 ούτε το R_2 διότι τα γνωρίσματα A και B δεν αποτελούν κλειδί ούτε για τη σχέση R_1 ούτε για τη σχέση R_2 άρα ο διαχωρισμός δεν είναι χωρίς απώλειες συνδέσμου.

Επίσης οι εξαρτήσεις δεν διατηρούνται διότι η εξάρτηση $AC \rightarrow D$ εκτείνεται σε δύο πίνακες.

Γ Ερώτημα

Για κανονικοποίηση σε 3NF υπολογίζουμε την ελάχιστη κάλυψη F_c από το σύνολο F των συναρτησιακών εξαρτήσεων. Μετά δημιουργούμε για κάθε εξάρτηση της ελάχιστης κάλυψη F_c μια ξεχωριστή σχέση (ένα ξεχωριστό πίνακα). Αν το υποψήφιο κλειδί δεν συμπεριλαμβάνεται σε κάποια σχέση τότε το βάζουμε μόνο του σε μια νέα σχέση

Αρχικό Σύνολο Σ.Ε.

$AC \rightarrow D$

$B \rightarrow CE$

$C \rightarrow A$

$D \rightarrow B$

- Δεν υπάρχουν εξαρτήσεις που να έχουν κοινό αριστερό μέλος
- Πρέπει να απαλείψουμε τα περιττά γνωρίσματα (πιθανά περιττά γνωρίσματα είναι όσα είναι διπλά είτε αριστερά είτε δεξιά των εξαρτήσεων)
 - Δίνεται $C \rightarrow A$. Επίσης $AC \rightarrow D$, άρα $C \rightarrow D$ από μόνο του οπότε το A είναι πλεονάζων γνώρισμα και απαλείφεται. Η ελάχιστη κάλυψη F_c γίνεται $F_c = \{C \rightarrow D, B \rightarrow CE, C \rightarrow A, D \rightarrow B\}$
 - Οι Σ.Ε. με το ίδιο αριστερό μέλος συνενώνονται σε μια άρα $F_c = \{C \rightarrow DA, B \rightarrow CE, D \rightarrow B\}$
 - Αυτό είναι το ελάχιστο κάλυμμα. Διασπάμε κάθε Σ.Ε. του F_c ως ξεχωριστή σχέση. Η καθεμία σχέση είναι σε 3NF και παρατηρούμε ότι το υποψήφιο κλειδί που είναι το D περιέχεται σε μια από αυτές τις σχέσεις
 - $R_1 = \{\underline{C}, D, A\}$
 - $R_2 = \{\underline{B}, C, E\}$
 - $R_3 = \{\underline{D}, B\}$
 - Παρατηρούμε ότι δεν λείπουν γνωρίσματα άρα οι σχέσεις R_1, R_2, R_3 είναι σε 3NF

5.16.5 Θέμα 3 Φεβρουάριος 2015 και Ιούνιος 2020

Θεωρείστε το σχήμα R(A, B, C, D, E) και το ακόλουθο σύνολο συναρτησιακών εξαρτήσεων F:

A→DE

BD→C

D→B

C→A

1) Να αποδειχθεί ότι ο διαχωρισμός σε (A, B, C) και (A, D, B, E) είναι ένας διαχωρισμός χωρίς απώλειες συνδέσμου (lossless-join)

2) Αποσυνθέστε την παραπάνω σχέση ώστε οι σχέσεις που θα προκύψουν βρίσκονται σε 3 NF (3^η Κανονική Μορφή)

Απάντηση

1) Η $R_1 \cap R_2 = \{A, B\}$

- Το $\{A, B\}$ δεν αποτελεί κλειδί ούτε για την R_1 ούτε για την R_2
- Άρα ο διαχωρισμός της R σε R_1 και R_2 δεν είναι χωρίς απώλειες συνδέσμου

2)

Υπολογίζουμε την ελάχιστη κάλυψη F_c

Αρχικό Σύνολο Σ.Ε.

A→DE

BD→C

D→B

C→A

- Δεν υπάρχουν εξαρτήσεις που να έχουν κοινό αριστερό μέλος
- Πρέπει να απαλείψουμε τα περιττά γνωρίσματα (υποψήφια είναι όσα είναι διπλά είτε αριστερά είτε δεξιά των εξαρτήσεων)
- Στη Σ.Ε. $BD \rightarrow C$ το B είναι πλεονάζων γνώρισμα και αφαιρείται διότι το $D \rightarrow B$ οπότε αρκεί μόνο το D για να προσδιορίσει το C
 - Η ελάχιστη κάλυψη F_c γίνεται $F_c = \{A \rightarrow DE, D \rightarrow C, D \rightarrow B, C \rightarrow A\}$
 - Οι εξαρτήσεις $D \rightarrow C$ και $D \rightarrow B$ έχουν κοινό αριστερό μέλος άρα ενοποιούνται σε μια εξάρτηση. Συνεπός η ελάχιστη κάλυψη F_c γίνεται: $F_c = \{A \rightarrow DE, C \rightarrow A, D \rightarrow CB\}$

Δημιουργούμε για κάθε Σ.Ε. του F_c μια σχέση. Άρα προκύπτουν οι σχέσεις σε 3NF:

➤ $R_1 = \{\underline{A}, D, E\}$

➤ $R_2 = \{\underline{C}, A\}$

➤ $R_3 = \{\underline{D}, B, C\}$

➤ Παρατηρούμε ότι δεν λείπουν γνωρίσματα άρα οι σχέσεις R1, R2, R3 είναι σε 3NF

Παρατήρηση

Βάσεις Δεδομένων- Computer Ανάλυση

Αν έλειπε κάποιο γνώρισμα γιατί έχει απαλειφθεί ως πλεονάζων, τότε εξετάζουμε αν αυτό αποτελεί υποψήφιο κλειδί με την κλειστότητα και αν ναι το βάζουμε σε ξεχωριστό πίνακα

5.16.6 Θέμα 4- Ιανουάριος 2016

Θεωρείστε το σχήμα $R(A, B, C, D, E)$ με τις ακόλουθες συναρτησιακές εξαρτήσεις:

$$\{A\} \rightarrow \{BE\}$$

$$\{C\} \rightarrow \{D\}$$

Εστω ότι διασπάμε το σχήμα R σε $R_1(A, B, E)$, $R_2(A, C)$, $R_3(C, D)$

1) Αποδείξτε ότι ο παραπάνω **διαχωρισμός είναι χωρίς απώλειες** (lossless join)

2) Διατηρεί ο παραπάνω διαχωρισμός τις εξαρτήσεις;

3) Ποια **κανονική μορφή** ακολουθούν τα σχήματα R_1 , R_2 και R_3 ;

Απάντηση

1) Για να είναι διαχωρισμός χωρίς απώλειες συνδέσμου πρέπει να ικανοποιείται ένα από τα ακόλουθα:

- Το $R_1 \cap R_2 \rightarrow R_1$ ή $R_1 \cap R_2 \rightarrow R_2$. Το $R_1 \cap R_2 = \{A\}$ το οποίο είναι κλειδί στο R_1 . Άρα η $R_1 \cap R_2$ προσδιορίζει συναρτησιακά το R_1
- Το $R_2 \cap R_3 \rightarrow R_2$ ή $R_2 \cap R_3 \rightarrow R_3$. Το $R_2 \cap R_3 = \{C\}$ το οποίο είναι κλειδί στο R_3 . Άρα η $R_2 \cap R_3$ προσδιορίζει συναρτησιακά το R_3
- Άρα είναι **lossless join διαχωρισμός**

2) Ο παραπάνω διαχωρισμός δεν διατηρεί τις εξαρτήσεις διότι η $A \rightarrow BE$ εκτείνεται σε 2 σχέσεις

3)

- Στη σχέση R_1 το A είναι κλειδί άρα είναι σε BCNF
- Στη σχέση R_2 δεν υπάρχουν εξαρτήσεις, το κλειδί είναι το σύνολο γνωρισμάτων A, C και είναι σε BCNF
- Στη σχέση R_3 το γνώρισμα C είναι κλειδί άρα είναι και αυτή σε BCNF

5.16.7 Θέμα 3 Φεβρουάριος 2017

Για την αντιμετώπιση του πλεονασμού και της ασυνέπειας στην αξιολόγηση ενός σχήματος βάσεων δεδομένων χρησιμοποιούμε τις συναρτησιακές εξαρτήσεις. Η συναρτησιακή εξάρτηση (functional dependency) είναι ένας είδος περιορισμού ακεραιότητας ο οποίος γενικεύει την έννοια του κλειδιού. Απαντήστε τα παρακάτω ερωτήματα:

1. Αποδείξτε τον ψευδομεταβατικό κανόνα (Pseudo-transitivity: $X \rightarrow Y$ και $Y \rightarrow Z$ τότε $X \rightarrow Z$)
2. Τι ορίζουμε ως ελάχιστο κάλυμμα F_c ενός συνόλου συναρτησιακών εξαρτήσεων F ; Είναι μοναδικό; Γιατί το χρειαζόμαστε;
3. Για το σχήμα $r=(A, B, C, D)$ υπολογίστε το ελάχιστο κάλυμμα για το παρακάτω σύνολο συναρτησιακών εξαρτήσεων:

$$A \rightarrow D, B \rightarrow CD, AC \rightarrow D$$

Απάντηση

1. Απόδειξη Ψευδομεταβατικού Κανόνα

- Θα δείξουμε ότι αν $X \rightarrow Y$ και $Y \rightarrow Z$ τότε $X \rightarrow Z$
- Δίνεται $X \rightarrow Y$
- Προκύπτει $XW \rightarrow WY$ από κανόνα επαύξησης και ενωτικό κανόνα
- Δίνεται $WY \rightarrow Z$
- Άρα από μεταβατικό κανόνα προκύπτει ότι $XW \rightarrow Z$

2. Ορισμός για το Ελάχιστο Κάλυμμα (Ελάχιστη Κάλυψη)

- Δοθέντος ενός συνόλου F από Συναρτησιακές Εξαρτήσεις F_c (Ελάχιστο Κάλυμμα) είναι το ελάχιστο σύνολο ισοδύναμων συναρτησιακών εξαρτήσεων
- Το χρειαζόμαστε γιατί μας διευκολύνει στον υπολογισμό των Υποψήφιων Κλειδιών
- Το ελάχιστο Κάλυμμα ΔΕΝ ΕΙΝΑΙ ΜΟΝΑΔΙΚΟ

3. Αλγόριθμος Υπολογισμού Ελάχιστης Κάλυψης

Βήμα 1

Σε κάθε συναρτησιακή εξάρτηση (ΣE) που έχει στο αριστερό της μέρος δύο γνωρίσματα εξετάζουμε αν μπορεί να απαλειφθεί κάποιο από αυτά τα γνωρίσματα και αν ναι τότε το απαλείφουμε.

Βήμα 2

Εξετάζουμε αν υπάρχουν **πλεονάζουσες συναρτησιακές εξαρτήσεις** στο σύνολο των συναρτησιακών εξαρτήσεων και αν ναι τις απαλείφουμε. Οι τρεις $\Sigma.E.$ που δίνονται είναι οι

- $A \rightarrow D$
- $B \rightarrow CD$
- $AC \rightarrow D$

Βάσεις Δεδομένων- Computer Ανάλυση

Η εξάρτηση $AC \rightarrow D$ μπορεί να απλοποιηθεί στην $A \rightarrow D$ (Το C είναι περιττό γνώρισμα αφού λόγω της Σ.Ε. $A \rightarrow D$ που δίνεται αρκεί μόνο το A για να προσδιορίσει το D)

- $A \rightarrow D$
- $B \rightarrow CD$
- $A \rightarrow D$

Οι ίδιες εξαρτήσεις απαλείφονται σε μια. Άρα προκύπτει το σύνολο Σ.Ε.

- $A \rightarrow D$
- $B \rightarrow CD$

Το τελικό σύνολο Σ.Ε. που αποτελεί και ελάχιστο κάλυμμα είναι το $F_c = \{A \rightarrow D, B \rightarrow CD\}$

5.16.8 Θέμα 2- Ιανουάριος 2014

Να απαντήσετε στις παρακάτω ερωτήσεις:

- i. Έστω ένα σύνολο συναρτησιακών εξαρτήσεων F και μια εξάρτηση $a \rightarrow b$ στο F. Πότε λέμε ότι μια ιδιότητα A της παραπάνω εξάρτησης είναι εξωτερική (πλεονάζουσα);
- ii. Δώστε τον αλγόριθμο υπολογισμού μιας κανονικής κάλυψης F_c για το F

Απάντηση

- i.
 - Ένα χαρακτηριστικό μίας συναρτησιακής εξάρτησης ονομάζεται **εξωτερικό ή πλεονάζον** (extraneous) αν είναι δυνατόν να απαλειφθεί χωρίς να αλλάξει η κλειστότητα του συνόλου των εξαρτήσεων
 - Έστω η εξάρτηση $X \rightarrow Y$ του συνόλου F. Ορίζουμε ότι ένα χαρακτηριστικό A είναι **πλεονάζον** αν:
 1. $A \subset X$ και το σύνολο F είναι ισοδύναμο με το $(F - \{X \rightarrow Y\}) \cup \{(X - A) \rightarrow Y\}$ ή
 2. $A \subset Y$ και το σύνολο F είναι ισοδύναμο με το $(F - \{X \rightarrow Y\}) \cup \{X \rightarrow (Y - A)\}$.

Παρατήρηση: Ένα γνώρισμα είναι περιττό (εξωτερικό) αν ισχύει η κλειστότητα τόσο πριν όσο και μετά την απαλοιφή του

ii.

Αλγόριθμος Υπολογισμού Κανονικής Κάλυψης

- **$F_c = F$**
- **Repeat**
 - Χρησιμοποιήστε τον κανόνα ένωσης για να αντικαταστήσετε τις εξαρτήσεις του F_c της μορφής $\alpha_1 \rightarrow \beta_1$ και $\alpha_1 \rightarrow \beta_2$ με $\alpha_1 \rightarrow \beta_1 \beta_2$
 - Βρείτε μια λειτουργική εξάρτηση $\alpha \rightarrow \beta$ στο F_c με μια εξωτερική ιδιότητα στο α ή στο β
 - Αν βρεθεί μια εξωτερική ιδιότητα, διαγράψτε την από το $\alpha \rightarrow \beta$
- **Until** το F_c να μην αλλάζει

5.16.9 Θέμα 2- Ιανουάριος 2016

Να βρεθεί η κανονική κάλυψη του F για τα ακόλουθα σύνολα:

$$R = (A, B, C)$$

$$F = \{A \rightarrow BC$$

$$B \rightarrow C$$

$$A \rightarrow B$$

$$AB \rightarrow C\}$$

Απάντηση

- Όλες οι εξαρτήσεις που έχουν το ίδιο αριστερό μέρος ενώνονται σε μια. Άρα οι Σ.Ε. $A \rightarrow B$ και $A \rightarrow BC$ ενώνονται στην $A \rightarrow BC$ και η κανονική κάλυψη είναι:
 - $F = \{A \rightarrow BC, B \rightarrow C, AB \rightarrow C\}$
- Το A είναι πλεονάζον γνώρισμα στην Σ.Ε. $AB \rightarrow C$ λόγω της $B \rightarrow C$ (Το B αρκεί μόνο του για να προσδιορίσει το C). Η κανονική κάλυψη είναι:
 - $F_c = \{A \rightarrow BC, B \rightarrow C, B \rightarrow C\}$
- Οι ίδιες εξαρτήσεις ενώνονται σε μια. Η κανονική κάλυψη είναι:
 - $F_c = \{A \rightarrow BC, B \rightarrow C\}$
- Το C είναι πλεονάζον γνώρισμα στο $A \rightarrow BC$ και απαλείφεται γιατί όταν σε μια εξάρτηση της μορφής $X \rightarrow YZ$ ισχύει ότι $Y \rightarrow Z$ το Z είναι περιττό γνώρισμα. Η κανονική κάλυψη είναι:
 - $F_c = \{A \rightarrow B, B \rightarrow C\}$

5.16.10 Θέμα 2- Ιανουάριος 2013

Να βρεθεί η κανονική κάλυψη του F για τα ακόλουθα σύνολα:

$$R = (A, B, C, D, E)$$

$$F = \{A \rightarrow B$$

$$AB \rightarrow C$$

$$D \rightarrow AC\}$$

Απάντηση

Το B είναι πλεονάζον γνώρισμα στην $AB \rightarrow C$ και απαλείφεται γιατί όταν σε μια εξάρτηση της μορφής $XY \rightarrow Z$ ισχύει ότι $X \rightarrow Y$ το Y είναι περιττό γνώρισμα. Οι νέες Σ.Ε. που προκύπτουν είναι:

- $A \rightarrow B$
- $A \rightarrow C$
- $D \rightarrow AC$

Το C είναι πλεονάζον γνώρισμα στην $D \rightarrow AC$ και απαλείφεται γιατί όταν σε μια εξάρτηση της μορφής $X \rightarrow YZ$ και $Y \rightarrow Z$ το Z είναι περιττό γνώρισμα. Οι νέες Σ.Ε. που προκύπτουν είναι:

- $A \rightarrow B$
- $A \rightarrow C$
- $D \rightarrow A$

Τέλος όλες οι εξαρτήσεις που έχουν το ίδιο αριστερό μέρος ενώνονται σε μια. Άρα η κανονική κάλυψη του F είναι:

- $F_c = \{A \rightarrow BC, D \rightarrow A\}$

5.16.11 Θέμα 2- Ιανουάριος 2018

Θεωρείστε το σχήμα σχέσης $R(A, B, C, D, E, F, G)$ με τις ακόλουθες συναρτησιακές εξαρτήσεις:

- $AE \rightarrow B$
- $BE \rightarrow D$
- $A \rightarrow CD$
- $F \rightarrow C$
- $E \rightarrow G$

1. Έστω ο διαχωρισμός σε $R_1=(A, D, C)$, $R_2=(A, B, E, F)$ και $R_3=(E, G)$. Είναι ο παραπάνω διαχωρισμός χωρίς απώλειες (lossless join); Δικαιολογήστε την απάντηση σας;

2. Αποτελεί ο συνδυασμός γνωρισμάτων (A, E) υποψήφιο κλειδί; Δικαιολογήστε την άποψη σας;

3. Ποιο είναι το πρωτεύον κλειδί της παραπάνω σχέσης;

4. Τι σημαίνει Κλειστότητα Συνόλου Συναρτησιακών Εξαρτήσεων F^+ ;

5. Απόδειξη Ψευδο-μεταβατικού Κανόνα

Απάντηση

1. Επειδή ο διαχωρισμός γίνεται σε 3 σχέσεις πρέπει η τομή των σχέσεων ανά δύο να αποτελεί κλειδί σε μια τουλάχιστον από αυτές Εξετάζουμε τα ακόλουθα:

- Το $R_1 \cap R_2 \rightarrow R_1 \wedge R_1 \cap R_2 \rightarrow R_2$. Το $R_1 \cap R_2 = \{A\}$ το οποίο είναι κλειδί στη σχέση R_1 λόγω της εξαρτησης $A \rightarrow CD$. Άρα η τομή $R_1 \cap R_2$ προσδιορίζει συναρτησιακά το R_1 δηλ. $R_1 \cap R_2 \rightarrow R_1$
- Το $R_2 \cap R_3 \rightarrow R_2 \wedge R_2 \cap R_3 \rightarrow R_3$. Το $R_2 \cap R_3 = \{E\}$ το οποίο είναι κλειδί στη σχέση R_3 λόγω της εξαρτησης $E \rightarrow G$. Άρα η τομή $R_2 \cap R_3$ προσδιορίζει συναρτησιακά το R_3 δηλ. $R_2 \cap R_3 \rightarrow R_3$
- Συνεπώς ο διαχωρισμός είναι χωρίς απώλειες συνδέσμου (lossless join διαχωρισμός)

2. Το υποψήφιο κλειδί υπολογίζεται ως εξής:

Υπολογισμός Υποψήφιου Κλειδιού

- $\{A, E\}^+ = \{A, E\}$
- Λόγω της εξαρτησης $A \rightarrow CD$ προκύπτει ότι $\{A, E\}^+ = \{A, E, C, D\}$
- Λόγω της εξαρτησης $E \rightarrow G$ προκύπτει ότι $\{A, E\}^+ = \{A, E, C, D, G\}$
- Λόγω της εξαρτησης $AE \rightarrow B$ προκύπτει ότι $\{A, E\}^+ = \{A, E, C, D, G, B\}$
- Η κλειστότητα δεν περιλαμβάνει όλα τα γνωρίσματα της R άρα το $\{A, E\}$ δεν αποτελεί υποψήφιο κλειδί

3. Το πρωτεύον κλειδί υπολογίζεται ως εξής:

Υπολογισμός Πρωτεύοντος Κλειδιού

- Από τις Σ.Ε. που δίνονται διαγράφουμε όλα τα γνωρίσματα στο δεξιό τους μέρος και όχι ταυτόχρονα και στο αριστερό τους μέρος
 - AE → B
 - BE → D
 - A → CD
 - F → C
 - E → G
- Τα C, D μπορούν να αφαιρεθούν επειδή είναι στο δεξιό μέρος της Σ.Ε. A → CD
- Το C επίσης αφαιρείται επειδή είναι στο δεξιό μέρος της Σ.Ε. F → C
- Το D επίσης αφαιρείται επειδή είναι στο δεξιό μέρος της Σ.Ε. BE → D
- Το G αφαιρείται επειδή είναι στο δεξιό μέρος της Σ.Ε. E → G
- Το B βρίσκεται και στο δεξιό μέρος της Σ.Ε. AE → B αλλά και στο αριστερό μέρος της Σ.Ε. BE → D γιατό δεν διαγράφεται
- Απομένουν τα γνωρίσματα A, B, E, F
- Η κλειστότητα $(A, B, E, F)^+$ με βάση τις Σ.Ε. της F είναι $\{A, B, E, F, C, G, D\}$ και συνεπώς το (A, B, E, F) είναι πρωτεύον κλειδί

4. Η Κλειστότητα ενός Συνόλου Συναρτησιακών Εξαρτήσεων F^+ διατυπώνεται ως εξής: **Δοθέντος ενός συνόλου Σ.Ε. F, η Κλειστότητα που συμβολίζεται ως F^+ είναι το σύνολο όλων των συναγόμενων Σ.Ε.**

5. Ο Ψευδο-μεταβατικός Κανόνας διατυπώνεται ως εξής: Av X → Y και YZ → W τότε XZ → W

Απόδειξη Ψευδο-μεταβατικού Κανόνα

1. X → Y (δίνεται)
2. YZ → W (δίνεται)
3. XZ → YZ (επαύξηση 1 με Z)
4. XZ → W (Μεταβατικότητα 3 και 2)

5.16.12 Θέμα 3- Ιανουάριος 2019

Θέμα 3 [20 μονάδες]

Για την αντιμετώπιση του πλεονασμού και της ασυνέπειας στην αξιολόγηση ενός σχήματος δεδομένων χρησιμοποιούμε τις συναρτησιακές εξαρτήσεις. Η συναρτησιακή εξαρτήση (*dependency*) είναι ένα είδος περιορισμού ακεραιότητας ο οποίος γενικεύει την εννοια του κλειδαρίου.

3.1 Περιγράψτε τις συνθήκες ώστε ένα σχήμα R με σύνολο συναρτησιακών εξαρτήσεων 1 κανονικοποιημένο σε 3η κανονική μορφή, (5 μονάδες)

3.2 Τι ορίζουμε ελάχιστο κάλυμμα Fc ενός συνόλου συναρτησιακών εξαρτήσεων F; Ειναι γιατί το χρειαζόμαστε; (5 μονάδες)

3. Θεωρήστε το σχήμα R(A, B, C, D, E, G) με τις ακόλουθες συναρτησιακές εξαρτήσεις $AB \rightarrow C$, $AG \rightarrow E$, $B \rightarrow D$, $E \rightarrow G$ οι οποίες αποτελούν κι ελάχιστο κάλυμμα. στα ότι διασπάμε το σχήμα R σε R1(A, B, C, D), R2 (A, E, G).

3.3a. Είναι ο παραπάνω διαχωρισμός χωρίς απώλειες συνένωσης (*lossless-join*); Δικαιούστε την απάντησή σας. (5 μονάδες)

3.3b. Διατηρεί ο παραπάνω διαχωρισμός τις συναρτησιακές εξαρτήσεις; Δικαιούστε απάντησή σας. (5 μονάδες)

Απάντηση

3.1 Για να είναι σε 3NF πρέπει να είναι σε 2η κανονική μορφή και να μην έχει μεταβατικές εξαρτήσεις. Αυτό σημαίνει ότι ΔΕΝ πρέπει ένα γνώρισμα που δεν αποτελεί μέρος του κλειδιού να προσδιορίζει συναρτησιακά άλλα γνωρίσματα

3.2 Ελάχιστη Κάλυψη του F

Δοθέντος ενός συνόλου F από συναρτησιακές εξαρτήσεις ελάχιστο ή κανονικό κάλυμμα είναι το ελάχιστο σύνολο ισοδύναμων εξαρτήσεων. Το χρειαζόμαστε γιατί μας διευκολύνει στην εύρεση υποψήφιων κλειδιών

3.3.

α Για να είναι διαχωρισμός χωρίς απώλειες συνδέσμου πρέπει να ικανοποιείται ένα από τα ακόλουθα:

- Το $R_1 \cap R_2 \rightarrow R_1 \text{ ή } R_1 \cap R_2 \rightarrow R_2$. Το $R_1 \cap R_2 = \{A\}$ το οποίο ΔΕΝ είναι κλειδί στο R_1 ΟΥΤΕ στο R_2
- Άρα ο διαχωρισμός ΔΕΝ ΕΙΝΑΙ ΧΩΡΙΣ ΑΠΩΛΕΙΕΣ ΣΥΝΔΕΣΜΟΥ

3.3.b Ο παραπάνω διαχωρισμός διατηρεί τις εξαρτήσεις διότι δεν υπάρχει Σ.Ε. που να εκτείνεται σε 2 σχέσεις

Πιθανό Θέμα

1. Τι σημαίνει Κλειστότητα Συνόλου γνωρίσματος A^+ ;

2. Να διατυπώσετε τον αλγόριθμο εύρεσης της κλειστότητας γνωρίσματος.

Απάντηση

1. Δοθέντος ενός συνόλου F από ένα σύνολο συναρτησιακών εξαρτήσεων, το A^+ είναι το σύνολο όλων των γνωρισμάτων που εξαρτώνται από το A

2. Ο αλγόριθμος εύρεσης της κλειστότητας ενός γνωρίσματος διατυπώνεται ως εξής:

- Δίνεται πίνακας (σχέση) R, ένα σύνολο από Σ.Ε. και ένα σύνολο πεδίων X του R
- Κλειστότητα του X ή X^+ είναι το σύνολο όλων των πεδίων Y του R για τα οποία ισχύει $X \rightarrow Y$
- **Αλγόριθμος:** Ξεκινούμε θέτοντας $X^+ = X$ και για κάθε Σ.Ε. $Y \rightarrow Z$ με $Y \subseteq X$ προσθέτουμε το Z στο X^+ . Συνεχίζουμε μέχρι να μην μπορεί να προστεθεί κάτι στο X^+

6 Θέμα 4 Ιανουάριος 2019 - Ασκηση με Κατακερματισμό

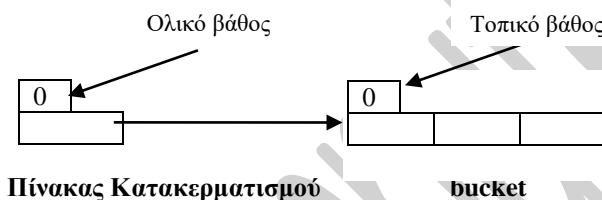
Έστω ότι διαθέτουμε τις παρακάτω εγγραφές (A..J) που προέκυψαν έπειτα από την εφαρμογή μιας συνάρτησης κατακερματισμού

Τιμή κλειδιού κατακερματισμού	
A	000110
B	111100
C	010111
D	010000
E	101001
F	010111
G	101001
H	011010
I	011010
J	001110

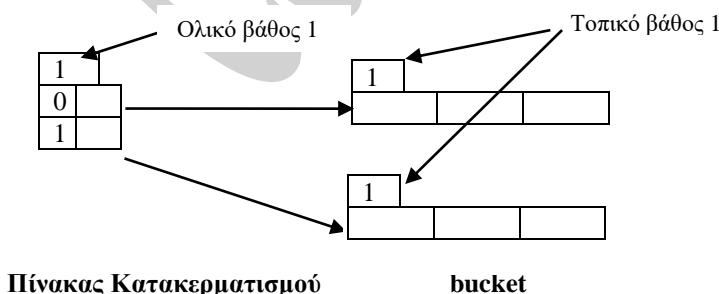
Θεωρείστε ότι εφαρμόζουμε επεκτατό κατακερματισμό όπου οι κάδοι μπορούν να διατηρούν 3 εγγραφές ο καθένας. Αρχικά η δομή είναι άδεια. Δείξτε τη μορφή που θα πάρει η δομή κατά την εισαγωγή των παραπάνω εγγραφών (με τη σειρά που σας δίνονται) εμφανίζοντας και τα αποτελέσματα. Για την ανάθεση των εγγραφών χρησιμοποιήστε τα πιο σημαντικά (αριστερότερα) bits. Σε περίπτωση υπερχείλισης το μέγεθος του καταλόγου διπλασιάζεται.

Απάντηση

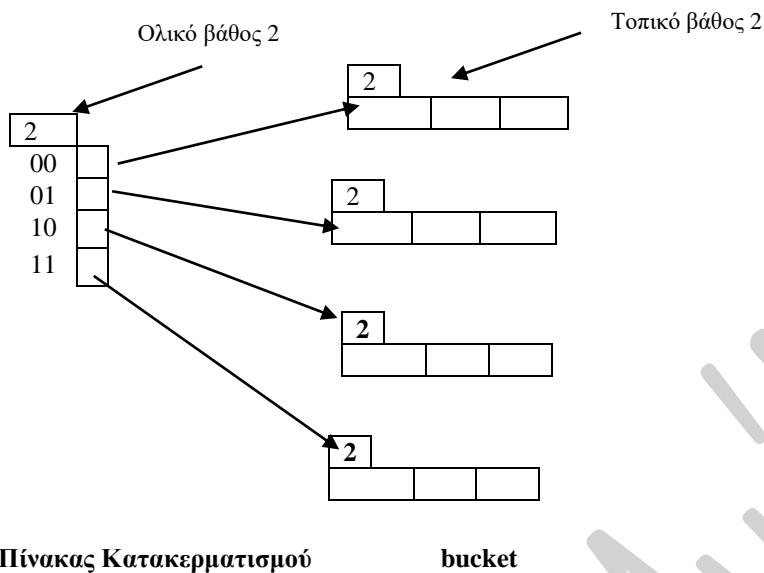
Με 0 bit στον πίνακα κατακερματισμού έχουμε το πολύ 1 bucket. Οι δοθείσες τιμές δεν χωράνε σε 1 bucket γιατί είναι προφανώς περισσότερες από 3. Διπλασιάζεται ο πίνακας κατακερματισμού και αυξάνεται το ολικό βάθος κατά ένα και αυξάνεται και το τοπικό βάθος κάθε bucket κατά 1.



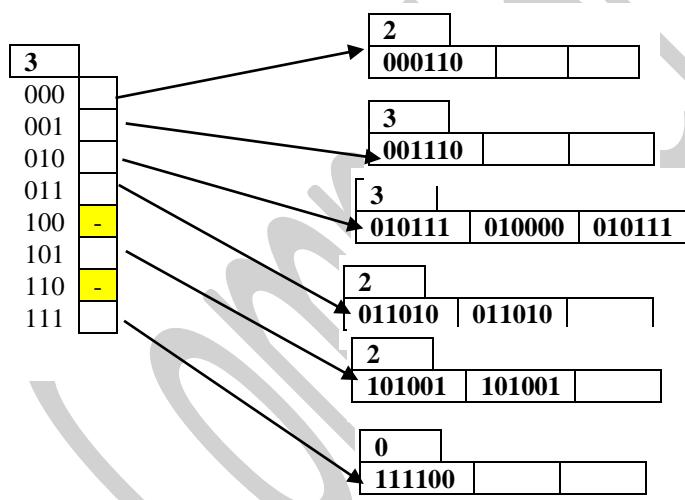
Με 1 bit στον πίνακα κατακερματισμού έχουμε το πολύ 2 buckets. Οι δοθείσες τιμές δεν χωράνε στα 2 buckets γιατί υπάρχουν περισσότερες από 3 τιμές που αρχίζουν με 0. Διπλασιάζεται ο πίνακας κατακερματισμού και αυξάνεται το ολικό βάθος κατά ένα και αυξάνεται και το τοπικό βάθος του bucket που διπλασιάζεται κατά 1.



Με 2 bit στον πίνακα κατακερματισμού έχουμε το πολύ 4 buckets. Με βάση τις δοθείσες τιμές δεν χωράνε σε 4 buckets γιατί υπάρχουν περισσότερες από 3 τιμές που αρχίζουν από 01 και δεν χωράνε στο bucket που δείχνει το 01. Άρα το bucket αυτό διπλασιάζεται.



Με 3 bit στον πίνακα κατακερματισμού έχουμε το πολύ 8 buckets. Οι δοθείσες τιμές χωράνε σε 4 buckets σύμφωνα με τον ακόλουθο τρόπο:



7 R-δέντρα (R-trees)

Η δομή δεδομένων R-Tree μπορεί να θεωρηθεί ως μια πολυδιάστατη έκδοση του B-Tree. Ας θεωρήσουμε ένα σύνολο με αντικείμενα στο R^2 (επίπεδο) στο οποίο θέλουμε να δημιουργήσουμε μια δομή ευρετηρίου (index structure). Αυτά τα αντικείμενα θα μπορούσαν να είναι πολύγωνα, γραμμές ή σύνολα σημείων (ένα ή περισσότερα σημεία). Αντί να θέσουμε στο ευρετήριο τα ίδια τα αντικείμενα, μπορούμε εναλλακτικά να τα αναπαραστήσουμε από το **ελάχιστο περιβάλλον ορθογώνιο ή MBR** (Minimum Bounded Rectangle). Το ευρετήριο βασίζεται σε αυτά τα MBR. Έτσι θα θεωρήσουμε ότι τα αντικείμενα που είναι αποθηκευμένα στο R-Tree είναι ορθογώνια στο R^2 αν και στην πραγματικότητα μπορούν να είναι πιο σύνθετα αντικείμενα σε οποιαδήποτε διάσταση.

Τα φύλλα του R-Tree αποθηκεύουν τα ακριβή MBR ή τα πλαίσια οριοθέτησης των επιμέρους γεωμετρικών αντικειμένων που περιλαμβάνουν μαζί με ένα δείκτη στην αντίστοιχη θέση αποθήκευσης. Τα φύλλα στο Tree περιέχουν εγγραφές της μορφής (MBR, object-id) ενώ οι εσωτερικοί κόμβοι περιέχουν εγγραφές της μορφής (MBR, child-ptr) όπου το child-ptr είναι δείκτης σε παιδί. Οι εσωτερικοί κόμβοι του δέντρου αποθηκεύουν τα ακριβή MBR των αντικειμένων ή είναι συλλογές αντικειμένων.

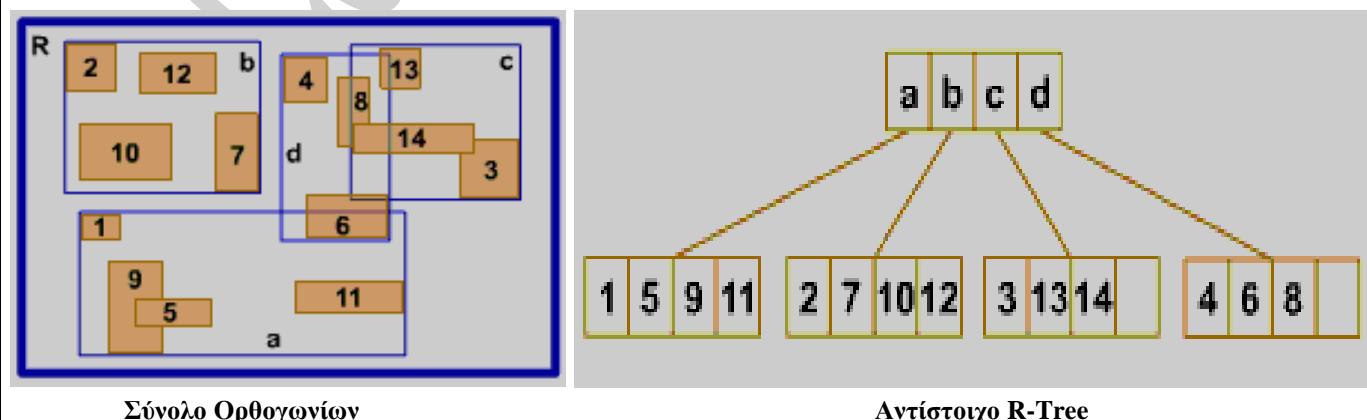
Το δέντρο κατασκευάζεται ιεραρχικά με την ομαδοποίηση των φύλλων σε μεγαλύτερα πλαίσια υψηλότερου επιπέδου τα οποία μπορεί να ομαδοποιηθούν σε ακόμη μεγαλύτερα πλαίσια στο αμέσως υψηλότερο επίπεδο.

Χαρακτηριστικά R-Tree

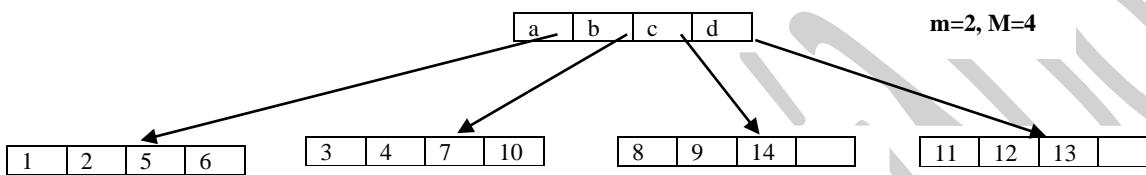
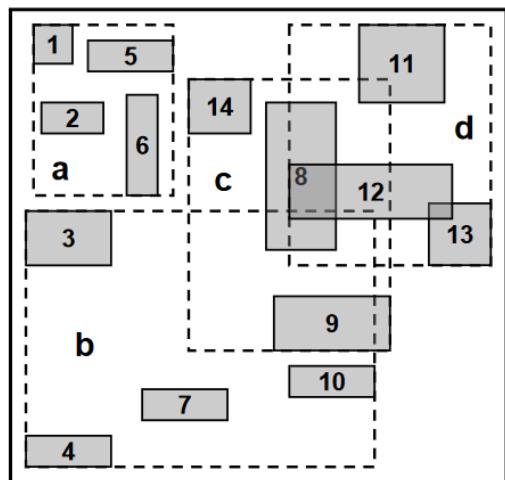
- Ο αριθμός των στοιχείων σε ένα εσωτερικό κόμβο του δέντρου (εκτός της ρίζας) είναι μεταξύ m και M όπου $m \in [0, M]$ και το M δίνεται από τον τύπο $M = \left\lfloor \frac{\text{size}(P)}{\text{size}(E)} \right\rfloor$ όπου P σελίδα δίσκου (disk page) και E στοιχείο (entry)
- Ο αριθμός των στοιχείων σε ένα φύλλο είναι M**
- Η ρίζα έχει τουλάχιστον 2 στοιχεία
- Όλα τα φύλλα είναι στο ίδιο επίπεδο
- Κάθε εσωτερικός κόμβος περιβάλλει τα παιδιά του και έχει μέγεθος m**
- Τα φύλλα περιλαμβάνουν τα πραγματικά αντικείμενα

Ένα σύνολο ορθογωνίων και το αντίστοιχο R-δέντρο φαίνονται στο επόμενο σχήμα:

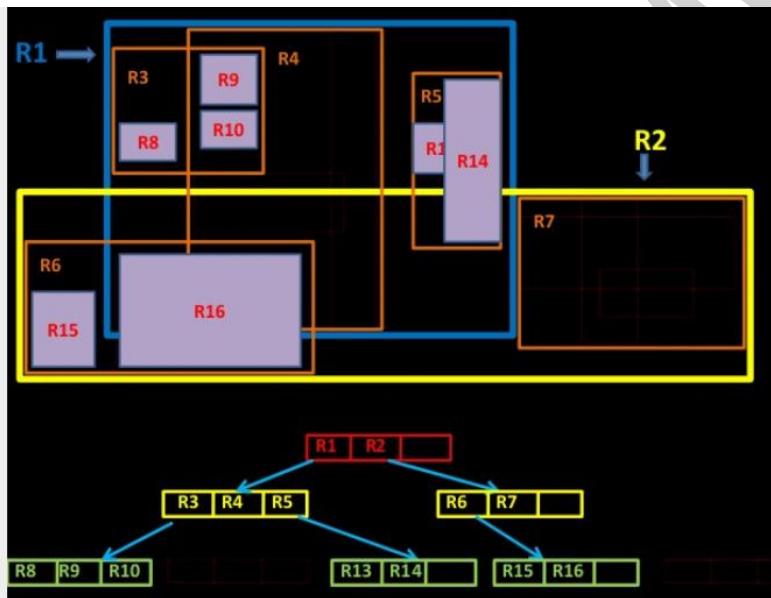
7.1 Παράδειγμα 1 σε R-Tree



7.2 Παράδειγμα 2 σε R-Tree



7.3 Παράδειγμα 3 σε R-Tree



7.4 Αναζήτηση στο R-Tree

Δοθέντος ενός σημείου q εντόπισε όλα τα mbb (minimum bounded boxes ή minimum bounded rectangles) που περιέχουν το q . Η επόμενη διαδικασία εκτελείται αναδρομικά ξεκινώντας από τη ρίζα

result = \emptyset

για ένα κόμβο N

αν N φύλλο τότε result = result $\cup \{N\}$

αλλιώς //αν το N είναι εσωτερικός κόμβος

για κάθε παιδί N' του N

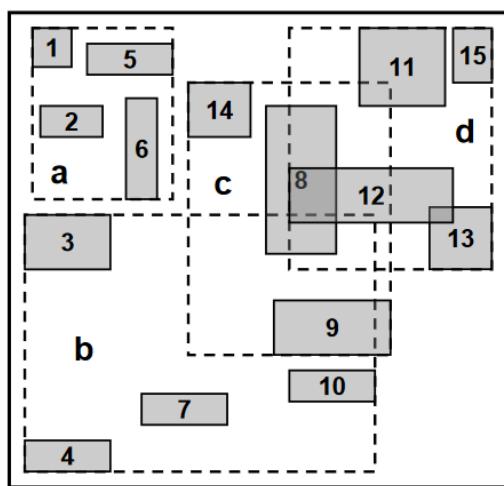
αν το ορθογώνιο του N' περιλαμβάνει το q τότε αναδρομικά εξετάζεται το N'

Η πολυπλοκότητα της αναζήτησης είναι $O(\log_m N)$ αν τα mbb δεν επικαλύπτονται και $O(N)$ αν επικαλύπτονται

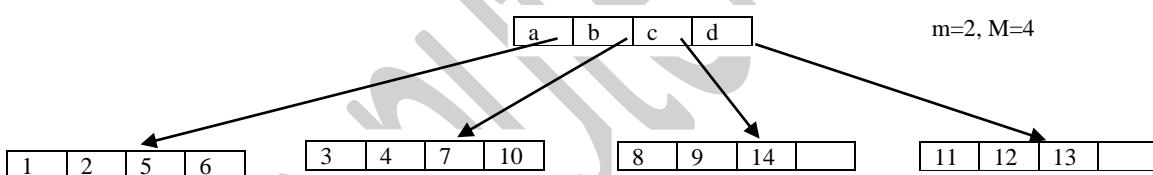
7.5 Εισαγωγή σε Tree – Επίλογη Φύλλου

- Διασχίζουμε το R-Tree από πάνω προς τα κάτω αρχίζοντας από τη ρίζα σε κάθε επίπεδο
- Αν υπάρχει κόμβος που το ορθογώνιο του περικλείει το νέο αντικείμενο που πρόκειται να εισαχθεί τότε ψάχνουμε το υποδέντρο του, αλλιώς αν δεν υπάρχει τέτοιο ορθογώνιο ψάχνουμε ένα κόμβο που το ορθογώνιο του είναι ελάχιστο και μετά ψάχνουμε το υποδέντρο αυτού
- Η διαδικασία ολοκληρώνεται μέχρι να φτάσουμε σε φύλλο
- Αν το φύλλο δεν είναι γεμάτο εισάγεται ένα στοιχείο της μορφής [mbb, oid] αλλιώς το φύλλο διασπάται και ενημερώνεται το μονοπάτι προς τη ρίζα αν χρειάζεται

Παράδειγμα Εισαγωγή νέου στοιχείου 15 (m=2, M=4)

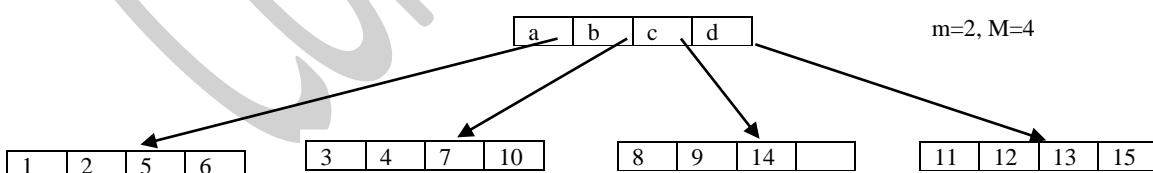


R-Tree πριν την εισαγωγή του 15



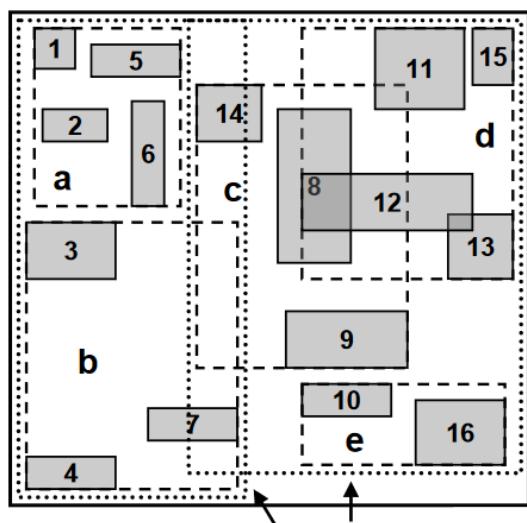
Επειδή το 15 περιέχεται στο ορθογώνιο d το βάζουμε ως παιδί του ορθογωνίου d (αφού υπάρχει κενός χώρος)

R-Tree μετά την εισαγωγή του 15

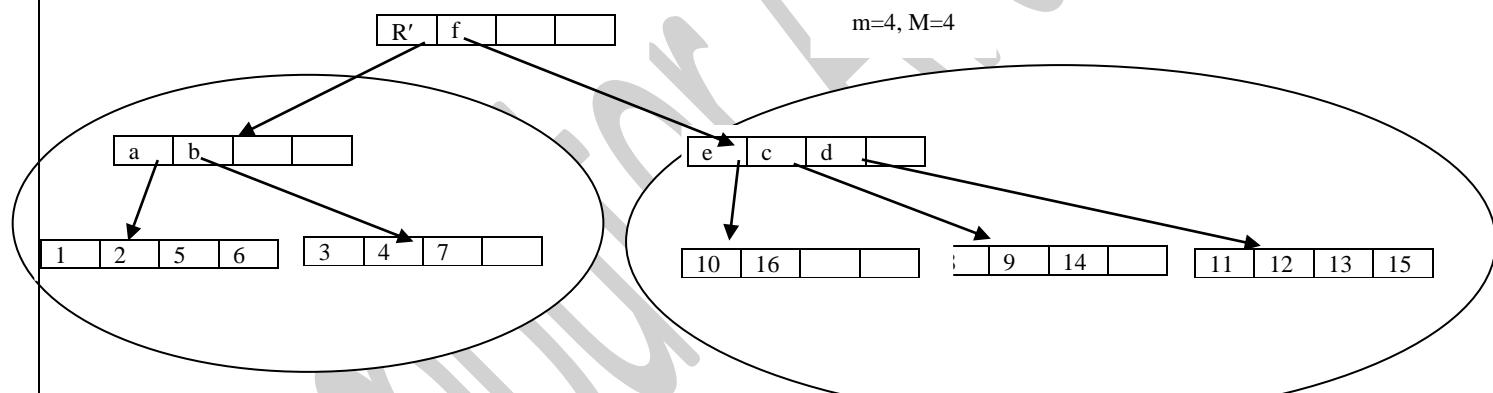


Βάσεις Δεδομένων- Computer Ανάλυση
Παράδειγμα Εισαγωγή νέου στοιχείου 16 (m=2, M=4)

Το 16 τοποθετείται σε σημείο που δεν υπάρχει ορθογώνιο. Πρέπει να προεκτείνουμε το ορθογώνιο b. Όμως το b είναι γεμάτο. Κάνουμε διάσπαση του b. Το κοντινότερο (πλησιέστερο) στοιχείο του b προς το 16 είναι το 10. Άρα δημιουργούμε μια νέα ορθογώνια περιοχή (mbr) με όνομα e που περιέχει το 10 και το 16. Το δέντρο ανεβαίνει επίπεδο. Δημιουργούμε 2 νέες ορθογώνιες περιοχές με ονόματα R' και f ενώ το αρχικό ορθογώνιο μετονομάζεται σε R'.

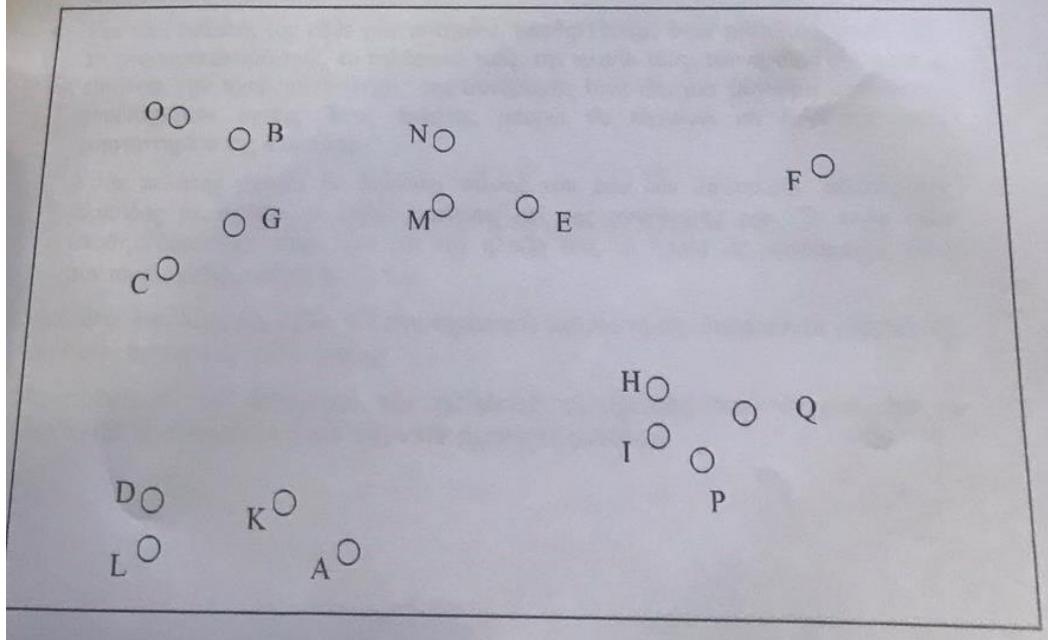


$m=4, M=4$



7.6 Θέμα 4 Ιανουάριος 2020

- Με βάση το παρακάτω σχήμα σχεδιάστε (επάνω στην κόλλα των θεμάτων) τα βέλτιστα ελάχιστα περιοριστικά ορθογώνια (*optimal minimum bounding rectangles*) που αντιστοιχούν σε R-tree με fan-out=4 προκειμένου να δεικτοδοτηθούν όλα τα δεδομένα-σημεία του σχήματος. Να δικαιολογήσετε την απάντηση σας. Σχεδιάστε το αντίστοιχο R-tree που προκύπτει. (5 μον.)



8 ΘΕΜΑΤΑ ΘΕΩΡΙΑΣ

8.1 Τι γνωρίζετε για τους περιορισμούς ακεραιότητας;

Απάντηση

- **Περιορισμός Πεδίου Ορισμού:** Η τιμή κάθε γνωρίσματος A πρέπει να είναι μία ατομική τιμή από το πεδίο ορισμού αυτού του γνωρίσματος dom(A)
- **Περιορισμός Κλειδιού:** όλες οι πλειάδες σε μία σχέση πρέπει να είναι διαφορετικές
- **Περιορισμός Ακεραιότητας Οντοτήτων:** Δε μπορεί η τιμή του πρωτεύοντος κλειδιού να είναι null
- **Περιορισμός Αναφορικής Ακεραιότητας:** Όταν μια πλειάδα μιας σχέσης S, αναφέρεται σε μια άλλη , τότε αυτή η άλλη πρέπει να υπάρχει.
- **Περιορισμός Σημασιολογικής Ακεραιότητας :**
Λογικοί περιορισμοί που ισχύουν στον πραγματικό κόσμο.

8.2 Ποια τα είδη των κλειδιών;

Απάντηση

- **Υπέρ-κλειδί SK :** ένα υποσύνολο των γνωρισμάτων μιας σχέσης η οποία προσδιορίζει μια συγκεκριμένη πλειάδα
- **Υποψήφιο κλειδί :** ένα ελάχιστο υπερκλειδί – αν αφαιρεθεί οποιοδήποτε γνώρισμα που σύνολο γνωρισμάτων που θα προκύψει παύει να θεωρείται κλειδί
- **Πρωτεύον κλειδί :** ένα από τα υποψήφια κλειδιά της σχέσης
- **Ξένο κλειδί :** ένα σύνολο γνωρισμάτων μιας σχέσης R₁ που αναφέρεται σε μία σχέση R₂ που ικανοποιεί τους κανόνες
 - Τα FKs έχουν το ίδιο πεδίο ορισμού με τα PKs στα οποία αναφέρονται
 - Μία τιμή του FK στην πλειάδα t1 της παρούσας κατάστασης r1(R₁)
 - είτε εμφανίζεται ως τιμή του PK σε κάποια πλειάδα t2 της παρούσας κατάστασης r2(R₂)
 - είτε είναι NULL

Διαφορά Κλειδιού με υπερκλειδί

- Το K είναι υπερκλειδί της σχέσης R αν και μόνο αν K→R
- Το K είναι υποψήφιο κλειδί της R αν και μόνο αν K→R και κανένα γνήσιο υποσύνολο του K δεν προσδιορίζει την R δηλ. δεν ισχύει ότι a∈K, a→R

8.3 Τι ονομάζεται Stored-Procedure και τι ονομάζεται Trigger- Θέμα 3 Σεπτέμβριος 2016

Απάντηση

- Μια stored procedure είναι μια υπορουτίνα, ένα πρόγραμμα
- Δημιουργείται και αποθηκεύεται στη βάση δεδομένων, δηλαδή στον server.
- Μπορεί να κληθεί και να εκτελεστεί από clients ή εφαρμογές που χρησιμοποιούν τη βάση.
- Έχει πρόσβαση στα δεδομένα της βάσης.
- Γράφεται σε μια ειδική γλώσσα η οποία ξαρτάται από το RDBMS που χρησιμοποιείται.

- Ένας trigger είναι ένα κομμάτι κώδικα, μια ρουτίνα
- Συνδέεται με ένα συγκεκριμένο πίνακα και καλείται όταν συμβεί ένα γεγονός στον πίνακα
- Συχνές χρήσεις:
 - Έλεγχος ορθότητας των δεδομένων που εισάγονται στον πίνακα.
 - Υλοποίηση λογικών περιορισμών.
 - Υπολογισμός παραγόμενων τιμών.
 - Συντήρηση logs πρόσβασης και αλλαγών στον πίνακα.
- Η υλοποίηση με triggers κοστίζει αρκετά, καθώς καθυστερούν την προσπέλαση στον πίνακα.

8.4 Θέμα 3 Φεβρουάριος 2010

Θέμα 3: Γενικές Ερωτήσεις [10 μονάδες]

- Ποια από τις παρακάτω προτάσεις είναι σωστή;
 - (Α) Για την αποθήκευση μιας νέας εγγραφής σε ένα αρχείο σωρού με δυναμικά μεταβαλλόμενο μήκος εγγραφών, αρκεί να βρούμε μια εγγραφή στο αρχείο με το bit ελέγχου ίσο με 1. Αν δεν βρεθεί τέτοια εγγραφή, τότε αποθηκεύουμε τη νέα εγγραφή στο τέλος του αρχείου.
 - (Β) Στα αρχεία εγγραφών σταθερού μήκους δεν υπάρχει ποτέ λόγος να εκτελείται η λειτουργία της αναδιάρθρωσης.
 - (Γ) Όταν ένα πεδίο μιας εγγραφής σε ένα αρχείο εγγραφών σταθερού μήκους είναι προαιρετικό, γίνεται εξουκονόμηση χώρου όταν το πεδίο αυτό δεν είναι συμπληρωμένο.
 - (Δ) Το πεδίο «Αριθμός_τηλεφώνου» σε ένα αρχείο με δυναμικά μεταβαλλόμενο μήκος εγγραφών μπορεί να λάβει περισσότερες της μιας τιμές.
- Αρχείο εγγραφών είναι διατεταγμένο με βάση το πεδίο «Κωδικός_αριθμός», το οποίο είναι θετικός αριθμός, διαφορετικός για κάθε εγγραφή του αρχείου. Το αρχείο αποτελείται από 1000 εγγραφές, ενώ κάθε εγγραφή καταλαμβάνει χώρο 130 bytes. Το σύστημα αποθήκευσης χρησιμοποιεί block μεγέθους 512 bytes. Κάθε εγγραφή αποθηκεύεται ολόκληρη σε ένα block (δηλ. δεν μπορεί να επιμεριστεί μεταξύ των τέλους ενός block και της αρχής ενός άλλου). Πόσες εγγραφές θα έχει ένας κατάλογος ενός επιπέδου (με δείκτες προς block) για τη δεικτοδότηση του παραπάνω αρχείου;
 - (Α) 250
 - (Β) 1000
 - (Γ) 334
 - (Δ) 1002

Απάντηση

1.
 - (Α) Λάθος διότι το bit ελέγχου ίσο με 1 έχει η μόνο τελευταία κενή εγγραφή. Αν δεν βρεθεί τέτοια εγγραφή τότε η στοίβα είναι γεμάτη κενών.
 - (Β) Λάθος διότι όταν υπάρχουν κενές εγγραφές θέλουμε μέσω αναδιάταξης να τις μεταφέρουμε στο τέλος του δίσκου.
 - (Γ) Σωστό διότι αν το πεδίο είναι προαιρετικό και δεν είναι συμπληρωμένο με τιμή τότε δεν δεσμεύεται χώρος γιαυτό.
 - (Δ) Σωστό διότι μπορούμε σε ένα πεδίο με μεταβαλλόμενο μέγεθος μπορούμε να καταχωρίσουμε όσες τιμές θέλουμε γίνεται σπατάλη χώρου όταν ορισμένες εγγραφές δεν έχουν τιμές για όλο το φυσικό χώρο που διατίθεται γιαυτές
- 2.

$$bfr = \left\lfloor \frac{B}{R} \right\rfloor = \left\lfloor \frac{512}{130} \right\rfloor = \left\lfloor 3,9 \right\rfloor = 3 \text{ Records/block}$$

$$b = \left\lceil \frac{r}{bfr} \right\rceil = \left\lceil \frac{1000}{3} \right\rceil = \lceil 333,33 \rceil = 334 \text{ Block}$$

Θα χρειαστούμε ένα δείκτη προς το κάθε block άρα συνολικά θα χρειαστούμε 334 δείκτες. Σωστή Απάντηση (Γ)

8.5 Θέμα 3 Φεβρουάριος 2013

Είμαι 3 [15 μονάδες]

a. Να δώσετε σύντομη απάντηση στις παρακάτω ερωτήσεις:

- Στο μοντέλο οντότητων – συσχετίσεων (ER) τι είναι η α) ασθενής οντότητα και β) το πλειότιμο γνώρισμα?
- Τι είναι α) το υπερκλειδί, β) το υποψήφιο κλειδί και γ) το πρωτεύον κλειδί?
- Για τι είδους ερωτήματα θα διαλέγατε ευρετήρια κατακερματισμού αντί για B-Trees? Να αιτιολογήσετε την απάντησή σας.

b. Να απαντήσετε τις παρακάτω ερωτήσεις με Σ(σωστό) ή Λ(λάθος) και να αιτιολογήσετε την απάντησή σας:

- Το αποτέλεσμα του ερωτήματος **Πλίστα_γνωρισμάτων...** (R)
- Έχει πάντα λιγότερες πλειάδες από την σχέση R.
Τα αραιά ευρετήρια δημιουργούνται πάνω στο γνώρισμα βάση του οποίου είναι ταξινομημένες οι εγγραφές στο αρχείο.

Απάντηση

a.

i.

α) Ασθενής Οντότητα λέγεται η οντότητα που δεν έχει γνώρισμα-κλειδί. Προσδιορίζεται από τη σχέση της (προσδιορίζουσα συσχέτιση) με άλλες οντότητες σε συνδυασμό με τις τιμές κάποιων γνωρισμάτων τους (μερικά κλειδιά). Ένας μη ισχυρός τύπος οντοτήτων έχει πάντα περιορισμό ολικής συμμετοχής ως προς την προσδιορίζουσα συσχέτισή του, δηλαδή η σχέση είναι **υποχρεωτική-πρέπει**

β) Πλειότιμο γνώρισμα ονομάζεται αυτό στο οποίο μια οντότητα έχει πολλές τιμές.

ii.

- **Υπέρ-κλειδί SK** : ένα υποσύνολο των γνωρισμάτων μιας σχέσης η οποία προσδιορίζει μια συγκεκριμένη πλειάδα
- **Υποψήφιο κλειδί** : ένα ελάχιστο υπερκλειδί – αν αφαιρεθεί οποιοδήποτε γνώρισμα που σύνολο γνωρισμάτων που θα προκύψει παύει να θεωρείται κλειδί
- **Πρωτεύον κλειδί** : ένα από τα υποψήφια κλειδιά της σχέσης
- **Ξένο κλειδί** : ένα σύνολο γνωρισμάτων μίας σχέσης R_1 που αναφέρεται σε μία σχέση R_2 που ικανοποιεί τους κανόνες
 - Τα FKs έχουν το ίδιο πεδίο ορισμού με τα PKs στα οποία αναφέρονται
 - Μία τιμή του FK στην πλειάδα t1 της παρούσας κατάστασης $r1(R_1)$
 - είτε εμφανίζεται ως τιμή του PK σε κάποια πλειάδα t2 της παρούσας κατάστασης $r2(R_2)$
 - είτε είναι **NULL**

iii. Ένα ευρετήριο κατακερματισμού χρησιμοποιείται για όλα τα ερωτήματα εκτός από:

- Ερωτήματα Διαστημάτων
- Προσεγγιστικά Ερωτήματα

b.

i. Λάθος διότι μόνο αν η σχέση που προκύπτει έχει διπλότυπα τότε μόνο το αποτέλεσμα του ερωτήματος **Πλίστα_γνωρισμάτων(R)** έχει λιγότερες πλειάδες από τη σχέση R.

ii. Σωστό, ένα αραιό ευρετήριο περιλαμβάνει το γνώρισμα σύμφωνα με το οποίο ταξινομούνται οι εγγραφές του αρχείου και αυτό δεν είναι το πρωτεύον κλειδί.

8.6 Θέμα 3 Φεβρουάριος 2016

Ποια η διαφορά μεταξύ κανονικής και ελάχιστης κάλυψης;

Απάντηση

- Μια κανονική κάλυψη του F είναι ένα σύνολο από συναρτησιακές εξαρτήσεις F_c τέτοιο ώστε
 - Το F συνεπάγεται λογικά όλες τις εξαρτήσεις του F_c
 - Το F_c συνεπάγεται λογικά όλες τις εξαρτήσεις του F
 - Δεν υπάρχει συναρτησιακή εξάρτηση στο F_c που να περιέχει εξωτερικό γνώρισμα
 - Η αριστερή πλευρά κάθε συναρτησιακής εξάρτησης στο F_c είναι μοναδική

Δοθέντος ενός F από συναρτησιακές εξαρτήσεις (ΣE) **ελάχιστη κάλυψη** F_c , ονομάζεται το ελάχιστο σύνολο ισοδύναμων συναρτησιακών εξαρτήσεων.

Η διαφορά μεταξύ της κανονικής κάλυψης και της ελάχιστης κάλυψης είναι ότι στην πρώτη επιτρέπεται να έχουμε παραπάνω από ένα χαρακτηριστικά στο αριστερό μέλος ενώ στην δεύτερη επιτρέπεται αυστηρά ένα.

Θέμα 5: Κατακερματισμός [10 μονάδες]

Έστω ότι διαθέτουμε τις παρακάτω εγγραφές (A,...,J) και τις τιμές που προέκυψαν έπειτα από συνάρτησης κατακερματισμού.

	Τιμή κλειδιού Κατακερματισμού
A	000110
B	111100
C	010111
D	010000
E	101001
F	010111
G	101001
H	011010
I	011010
J	001110

Θέμα 3 [10 μονάδες]

Θεωρήστε το σχήμα $R=(A,B,C,D,E)$ και το ακόλουθο σύνολο συναρτησιακών εξαρτήσεων F:

$A \rightarrow DE$

$BD \rightarrow C$

$D \rightarrow B$

$C \rightarrow A$

3.1 Να αποδειχθεί ότι ο διαχωρισμός σε (A,B,C) και (A,D,B,E) είναι ένας διαχωρισμός χωρίς απώλειες συνδέσμου (lossless-join). (5 μονάδες)

3.2 Αποσυνθέστε την παραπάνω σχέση ώστε οι σχέσεις που θα προκύψουν να βρίσκονται σε 3rd NF (3η κανονική μορφή) κι αιτιολογήστε την απάντησή σας. (5 μονάδες)

Θέμα 4 [10 μονάδες]

4.1 Αποδείξτε τον ψευδομεταβατικό κανόνα (Pseudo-transitivity: $X \rightarrow Y$ και $YW \rightarrow Z$ τότε $XW \rightarrow Z$) με την χρήση των αξιωμάτων του Armstrong (5 μονάδες)

4.2 Υπολογίστε το ελάχιστο κάλυμμα για το παρακάτω σύνολο συναρτησιακών εξαρτήσεων:
 $AB \rightarrow C, A \rightarrow BC, B \rightarrow C, A \rightarrow B$ (5 μονάδες)

Εξωτερικά Γνωρίσματα

- Αν F σύνολο συναρτησιακών εξαρτήσεων που περιέχει εξάρτηση $\alpha \rightarrow \beta$ τότε
 - Το γνώρισμα A είναι εξωτερικό στο α αν $A \in \alpha$
το F συνεπάγεται λογικά το $(F - \{\alpha \rightarrow \beta\}) \cup \{\alpha - A \rightarrow \beta\}$.
 - Το γνώρισμα A είναι εξωτερικό στο β αν $A \in \beta$
και το $(F - \{\alpha \rightarrow \beta\}) \cup \{\alpha \rightarrow (\beta - A)\}$ συνεπάγεται λογικά το F

Παράδειγμα 1

Έστω $F = \{A \rightarrow C, AB \rightarrow C\}$

Το B είναι εξωτερικό γνώρισμα στο $AB \rightarrow C$ διότι το $F = \{A \rightarrow C, AB \rightarrow C\}$ συνεπάγεται λογικά το $(F - \{AB \rightarrow C\}) = A \rightarrow C \cup (AB - B) = A \rightarrow C$

Παράδειγμα 2

Έστω $F = \{A \rightarrow C, AB \rightarrow CD\}$

Το C είναι εξωτερικό στο $AB \rightarrow CD$ διότι $C \in CD$ και το $(F - \{AB \rightarrow CD\}) = A \rightarrow C \cup AB \rightarrow D = \{A \rightarrow C, AB \rightarrow D\}$ συνεπάγεται λογικά το $F = \{A \rightarrow C, AB \rightarrow CD\}$ ακόμα και μετά τη διαγραφή του C .

Κανονική Κάλυψη

- Μια κανονική κάλυψη του F είναι ένα σύνολο από συναρτησιακές εξαρτήσεις F_c τέτοιο ώστε
 - Το F συνεπάγεται λογικά όλες τις εξαρτήσεις του F_c
 - Το F_c συνεπάγεται λογικά όλες τις εξαρτήσεις του F
 - Δεν υπάρχει συναρτησιακή εξάρτηση στο F_c που να περιέχει εξωτερικό γνώρισμα
 - Η αριστερή πλευρά κάθε συναρτησιακής εξάρτησης στο F_c είναι μοναδική

Αλγόριθμος Εύρεσης Κανονικής Κάλυψης

$F_c = F$

repeat

 Χρησιμοποίησε ενωτικό κανόνα (union rule) για να αντικαταστήσεις όλες τις Σ.Ε. στο F της μορφής $a_1 \rightarrow b_1$ και $a_1 \rightarrow b_2$ με τη Σ.Ε.

 της μορφής $a_1 \rightarrow b_1 b_2$

 Βρες μια Σ.Ε. $a \rightarrow b$ με εξωτερικό γνώρισμα είτε στο a είτε στο b . Αν βρεθεί εξωτερικό γνώρισμα τότε διέγραψε το από τη Σ.Ε.

$a \rightarrow b$

until F_c παραμείνει το ίδιο

Παράδειγμα Εφαρμογής Αλγόριθμου Εύρεσης Κανονικής Κάλυψης

Δίνεται $R = (A, B, C)$ και οι Σ.Ε. $F = \{A \rightarrow BC, B \rightarrow C, A \rightarrow B, AB \rightarrow C\}$. Να βρεθεί η Κανονική Κάλυψη F_c

Απάντηση

Συνδυάζουμε τις Σ.Ε. $F = \{A \rightarrow BC$ και $A \rightarrow B$ στην Σ.Ε. $A \rightarrow B$. Άρα η $F_c = \{A \rightarrow BC, B \rightarrow C, A \rightarrow B, AB \rightarrow C\}$

Το A είναι εξωτερικό γνώρισμα στη Σ.Ε. $AB \rightarrow C$. Εξετάζουμε αν το αποτέλεσμα της διαγραφής του A από την $AB \rightarrow C$ υπονοείται από τις άλλες εξαρτήσεις. Η απάντηση είναι ότι η Σ.Ε. $B \rightarrow C$ υπάρχει

Το C είναι εξωτερικό γνώρισμα στη Σ.Ε. $AB \rightarrow C$. Εξετάζουμε αν το αποτέλεσμα της διαγραφής του C από την $A \rightarrow BC$ υπονοείται από τις άλλες εξαρτήσεις. Η απάντηση είναι από $A \rightarrow B$ και $B \rightarrow C$ προκύπτει η Σ.Ε. $A \rightarrow B$

Η ελάχιστη κάλυψη είναι: $F_c = \{A \rightarrow B$ και $B \rightarrow C\}$

Κανονική Κάλυψη του F

- Κανονική κάλυψη συνόλου συναρτησιακών εξαρτήσεων F : Το μικρότερο σετ εξαρτήσεων του F από όπου μπορεί να προκύψει όλο το F .

- Παραδείγματα:

-Στο $F=\{A \rightarrow B, B \rightarrow C, A \rightarrow C\}$ η $A \rightarrow C$ πλεονάζει.

-Στο $F=\{A \rightarrow B, B \rightarrow C, A \rightarrow CD\}$ η $A \rightarrow CD$ μπορεί να απλοποιηθεί στην $A \rightarrow D$. Στην περίπτωση αυτή θα λέμε ότι το γνώρισμα C είναι πλεονάζον στην $A \rightarrow CD$.

Κανονική Κάλυψη του F

-Στο $F=\{A \rightarrow B, B \rightarrow C, AC \rightarrow D\}$ η $AC \rightarrow D$ μπορεί να απλοποιηθεί στην $A \rightarrow D$. Ομοίως το γνώρισμα C είναι πλεονάζον στην $AC \rightarrow D$.

Κανονική κάλυψη του F είναι ένα σύνολο συναρτησιακών εξαρτήσεων F_c τέτοιο ώστε:

- Το F μπορεί να παράγει το F_c και αντίστροφα.
- Δεν υπάρχει πλεονάζον γνώρισμα σε καμία συναρτησιακή εξάρτηση του F_c .
- Τα αριστερά μέρη (ορίζουσες) των εξαρτήσεων του F_c διαφέρουν μεταξύ τους.

Κανονική Κάλυψη του F

Αλγόριθμος Υπολογισμού F_c

repeat

Αντικατέστησε συν. εξαρτήσεις του F της μορφής $\alpha_1 \rightarrow \beta_1$ και $\alpha_1 \rightarrow \beta_2$ με $\alpha_1 \rightarrow \beta_1 \beta_2$ (κανόνας ένωσης).

Βρες μια συν. εξάρτηση $\alpha \rightarrow \beta$ που έχει πλεονάζον γνώρισμα είται στο α είται στο β .

Σβήσε πλεονάζοντα γνωρίσματα από το $\alpha \rightarrow \beta$.

until F does not change

Πλεονάζοντα Γνωρίσματα

Έστω $\alpha \rightarrow \beta$ συναρτησιακή εξάρτηση που ανήκει στο F .

- Το γνώρισμα A είναι πλεονάζον στο α αν $A \in \alpha$ και το F παράγει το $(F - \{\alpha \rightarrow \beta\}) \cup \{(\alpha - A) \rightarrow \beta\}$.

Παράδειγμα: Έστω $F = \{A \rightarrow C, AB \rightarrow C\}$ και θέλουμε να δούμε αν το B είναι πλεονάζον στην $AB \rightarrow C$

Σύμφωνα με τον ορισμό θα έπρεπε το F να παράγει το:

$$(F - \{AB \rightarrow C\}) \cup \{(AB - B) \rightarrow C\} = \{A \rightarrow C\} \cup \{A \rightarrow C\} = \{A \rightarrow C\}$$

που προφανώς ισχύει.

- Το γνώρισμα A είναι πλεονάζον στο β αν $A \in \beta$ και το $(F - \{\alpha \rightarrow \beta\}) \cup \{\alpha \rightarrow (\beta - A)\}$ παράγει το F .

Παράδειγμα: Έστω $F = \{A \rightarrow C, AB \rightarrow CD\}$ και θέλουμε να δούμε αν το C είναι πλεονάζον στην $AB \rightarrow CD$

Θα πρέπει το $\{A \rightarrow C, AB \rightarrow D\}$ να παράγει το F που ισχύει αφού $A \rightarrow C \Rightarrow AB \rightarrow ABC$ (1)
 $AB \rightarrow D \Rightarrow ABC \rightarrow CD$ (2)
(1,2 με προσεταιριστική δίνει $AB \rightarrow CD$)

Γενικότερα αν $\alpha \rightarrow \beta$ και $\gamma \rightarrow \delta$ τότε $\alpha\gamma \rightarrow \beta\delta$

Πλεονάζοντα Γνωρίσματα

- Για να ελέγξουμε αν το $A \in \alpha$ είναι πλεονάζον στο α : Υπολογίζουμε $(\alpha - \{A\})^+$ και κοιτάμε αν περιέχει το A . Αν ναι το A είναι πλεονάζον.
- Για να ελέγξουμε αν το $A \in \beta$ είναι πλεονάζον στο β : Υπολογίζουμε το α^+ χρησιμοποιώντας τις εξαρτήσεις του $(F - \{\alpha \rightarrow \beta\}) \cup \{\alpha \rightarrow (\beta - A)\}$, κοιτάμε αν περιέχει το A . Αν ναι το A είναι πλεονάζον

Για να βρούμε το ελάχιστο κάλυμμα απλοποιούμε επαναληπτικά το σύνολο των Σ.Ε. εφαρμόζοντας τις ακόλουθες μεθόδους:

Κανόνας 1: Απλοποιούμε μια Σ.Ε. με τον Κανόνα της Ένωσης (Union Rule). Έστω X, Y, Z σύνολα γνωρισμάτων. Av $X \rightarrow Y$ και $X \rightarrow Z$ τότε $X \rightarrow YZ$

Κανόνας 2: Απλοποιούμε μια Σ.Ε. απλοποιώντας το αριστερό μέλος της. Έστω X, Y σύνολα γνωρισμάτων και B ένα μοναδικό γνώρισμα ΠΟΥ ΔΕΝ ανήκει στο X . Έστω F η Σ.Ε. $XB \rightarrow Y$ και H η Σ.Ε. $X \rightarrow Y$. Av $F \Rightarrow X \rightarrow Y$ δηλ. μπορούμε να αντικαταστήσουμε την F από την H . Με άλλα λόγια αν $Y \subseteq X^+ F$ μπορούμε να αντικαταστήσουμε την F από την H .

Για παράδειγμα έστω F το ακόλουθο σύνολο Σ.Ε.

- $AB \rightarrow C$
- $A \rightarrow B$

Τότε $A^+_F = ABC$. Επειδή $Y \subseteq X^+_F$ μπορούμε να γράψουμε:

- $A \rightarrow C$
- $A \rightarrow B$

Το γνώρισμα B στην $AB \rightarrow C$ είναι πλεονάζων

Κανόνας 3: Απλοποιούμε μια Σ.Ε. απλοποιώντας το δεξιό μέλος της. Έστω X, Y σύνολα γνωρισμάτων και C ένα μοναδικό γνώρισμα ΠΟΥ ΔΕΝ ανήκει στο Y . Έστω F η Σ.Ε. $X \rightarrow YC$ και H η Σ.Ε. $X \rightarrow Y$. Av $H \Rightarrow X \rightarrow YC$ δηλ. μπορούμε να αντικαταστήσουμε την F από την H . Με άλλα λόγια αν $YC \subseteq X^+_H$ μπορούμε να αντικαταστήσουμε την F από την H .

Για παράδειγμα έστω F το ακόλουθο σύνολο Σ.Ε.

- $A \rightarrow BC$
- $B \rightarrow C$

Επειδή $A^+_H = ABC$. Επειδή $BC \subseteq X^+_H$ μπορούμε να αντικαταστήσουμε F με την H .