Πανεπιστημιο Ιωαννινων
Τμήμα Πληροφορικής

Βάσεις Δεδομένων
Ακαδημαϊκό Έτος 2004-2005

ΤΕΛΙΚΟ ΔΙΑΓΩΝΙΣΜΑ
ΕΞΕΤΑΣΤΙΚΗ ΙΑΝΟΥΑΡΙΟΥ

ΠΡΟΣΟΧΗ!! Διάρκεια 3 ώρες. Απαντήσεις χωρίς εξήγηση δε θα βαθμολογηθούν. Χρησιμοποιείστε το χρόνο σας συνετά. Καλή Επιτυχία.

Θέμα 1 [Μονάδες 25] Αναγνωρίζοντας τις γνώσεις και τις ικανότητές σας στις βάσεις δεδομένων, ο διευθυντής του μεγαλύτερου ιδιωτικού τηλεοπτικού καναλιού σας προσλαμβάνει στο Τμήμα Προγραμματισμού. Η πρώτη σας αποστολή είναι ο σχεδιασμός μιας βάσης δεδομένων για την υποστήριξη του προγραμματισμού ελληνικών κωμικών σειρών. Στη βάση δεδομένων διατηρείται πληροφορία για τις κωμικές σειρές όλων των καναλιών, τους ηθοποιούς που παίζουν σε αυτές και τα επεισόδια τους. Συγκεκριμένα:

- Για κάθε κωμική σειρά, έχουμε το όνομά της, το κανάλι που την προβάλει, το χρόνο παραγωγής και το κόστος παραγωγής. Δεν υπάρχουν κωμικές σειρές με το ίδιο όνομα και χρόνο παραγωγής.
- Για κάθε επεισόδιο, έχουμε τον αριθμό επεισοδίου, την ακροαματικότητά του και την ημερομηνία προβολής. Ο αριθμός επεισοδίου είναι μοναδικός για κάθε κωμική σειρά αλλά μπορεί να υπάρχουν κωμικές σειρές με τον ίδιο αριθμό επεισοδίου.
- Για τους ηθοποιούς, έχουμε το όνομα, την ημερομηνία γέννησης, και το φύλο τους. Δεν υπάρχουν ηθοποιοί με το ίδιο όνομα.
- Τέλος, ποιοι ηθοποιοί παίζουν σε κάθε σειρά και το ρόλο τους στη σειρά (π.χ., "Πόπωτας").

(α) Δώστε ένα κατάλληλο διάγραμμα Οντοτήτων/Συσχετίσεων (Ο/Σ). Μη ξεχάστε να εκφράσετε όλους τους περιορισμούς (π.χ., κλειδιά, πληθικότητα, τύπο συμμετοχών κλπ).

(β) Δώστε ένα κατάλληλο σχεσιακό μοντέλο. Μη ξεχάστε να εκφράσετε όλους τους περιορισμούς (π.χ., κλειδιά, ξένα κλειδιά κλπ).

(γ) Δώστε τους ορισμούς των πινάκων του σχήματος του ερωτήματος (β) σε SQL. Ορίστε κατάλληλα πεδία ορισμού για κάθε γνώρισμα καθώς και όλους τους περιορισμούς.

(δ) Δείξτε τι αλλάζει στο σχεδιασμό των (α) και (β) αν θέλουμε να έχουμε την πληροφορία ποιος ηθοποιός εμφανίζεται σε κάθε *επεισόδιο* μιας σειράς και το ρόλο του στο επεισόδιο.

(ε) Εκφράστε χρησιμοποιώντας την κατάλληλη συναρτησιακή εξάρτηση τον παρακάτω περιορισμό: "Ένα ρόλο σε μια κωμική σειρά τον παίζει μόνο ένας ηθοποιός".

Θέμα 2 [Μονάδες 25]

(α) Θεωρείστε το παρακάτω σχεσιακό σχήμα:

Βιβλίο($\underline{\text{ISBN}}$, Τίτλος, Εκδότης)
Συγγραφέας($\underline{\text{Όνομα}}$, Έτος-Γέννησης)
Έγραψε($\underline{\text{ISBN-Βιβλίου}}$, $\underline{\text{Όνομα-Συγγραφέα}}$)

Δώστε τις παρακάτω ερωτήσεις

(α) Σε σχεσιακή άλγεβρα, όλους τους συγγραφείς που έγραψαν μόνο βιβλία με εκδότη το Δίαυλο.
(β) Σε σχεσιακή άλγεβρα, τους συγγραφείς που έχουν γράψει τουλάχιστον ένα βιβλίο με τον ίδιο τίτλο αλλά διαφορετικό εκδότη.
(γ) Σε SQL, τον μεγαλύτερο σε ηλικία συγγραφέα που έγραψε τουλάχιστον δύο βιβλία με εκδόθηκαν από το Δίαυλο.
(δ) Σε σχεσιακό λογισμό πλειάδων, όλους τους συγγραφείς που έχουν γράψει βιβλία με τουλάχιστον δύο διαφορετικούς εκδότες.

Θέμα 3 [Μονάδες 15] Θεωρείστε το σχήμα σχέσης $R(A, B, C, D)$ με το παρακάτω σύνολο συναρτησιακών εξαρτήσεων $\mathrm{F} = \{AB \to C,\ AD \to B,\ AC \to B,\ CD \to A\}$.

(α) Θεωρείστε τη διάσπαση της R σε $R_1(B, C, D)$ και R_2. Η εξάρτηση $CD \to B$ ανήκει (ισχύει) στην προβολή της F στο R_1 ή όχι; Εξηγείστε την απάντησή σας.
(β) Δώστε μια διάσπαση της R που να έχει απώλειες στη συνένωση και μια που να μην έχει. Εξηγείστε την απάντησή σας.
(γ) Δώστε ένα πιθανό στιγμιότυπο της R στο οποίο να μην ισχύει η $C \to A$ (αλλά φυσικά να ισχύει το F).

Θέμα 4 [Μονάδες 23] Έστω ένα B+ δέντρο με τάξη φύλλου 40, τάξη εσωτερικών κόμβων 35 και 4 επίπεδα (συμπεριλαμβανομένου του επιπέδου της ρίζας και των φύλλων) στο πεδίο A μιας σχέσης R. Το μέγεθος του πεδίου ευρετηριοποίησης είναι 6 bytes και του δείκτη (δέντρου, block και εγγραφής) 7 bytes, ο παράγοντας ομαδοποίησης του αρχείου δεδομένων 30 και το μέγεθος του block 1024 bytes.

(α) Υπολογίστε το μέγιστο μέγεθος (σε blocks) ενός αρχείου που μπορεί να ευρετηριοποιηθεί από αυτό το δέντρο για ένα αρχείο που είναι ταξινομημένο ως προς το πεδίο ευρετηριοποίησης και το πεδίο ευρετηριοποίησης είναι κλειδί.

(β) Υπολογίστε το μέγιστο μέγεθος (σε blocks) ενός αρχείου που μπορεί να ευρετηριοποιηθεί από αυτό το δέντρο όταν το αρχείο δεν είναι ταξινομημένο ως προς το πεδίο ευρετηριοποίησης, αλλά το πεδίο ευρετηριοποίησης είναι κλειδί.

(γ) Περιγράψτε έναν αλγόριθμο για τον υπολογισμό μιας ερώτησης $\sigma_{a1<A<a2}(R)$ για τις περιπτώσεις (α) και (β). Υπολογίστε το κόστος για κάθε περίπτωση αν η ερώτηση ικανοποιείται από το 1% των εγγραφών του αρχείου.

Θέμα 5 [Μονάδες 12] Θεωρείστε δύο σχέσεις R και S αποθηκευμένες σε δύο αρχεία. Η σχέση R έχει 45.000 πλειάδες και το αρχείο που είναι αποθηκευμένη 3.000 blocks. Η σχέση S έχει 100.000 πλειάδες και το αρχείο που είναι αποθηκευμένη 30.000 blocks. Θέλουμε να υπολογίσουμε τη συνένωση $R \bowtie_{R.A=S.B} S$. Ο παράγοντας ομαδοποίησης για το αποτέλεσμα είναι 35.

(α) Δώστε το μέγιστο και το ελάχιστο μέγεθος (σε blocks) του αποτελέσματος της συνένωσης όταν:
(i) το γνώρισμα A είναι κλειδί για την R και ξένο κλειδί για την S (το B αναφέρεται στο A),
(ii) το γνώρισμα A είναι κλειδί για την R και το B κλειδί για την S.

(β) Θεωρείστε ότι η σχέση R είναι ταξινομημένη ως προς το γνώρισμα A. Για τις παρακάτω περιπτώσεις εξηγείστε αν το αποτέλεσμα που προκύπτει είναι ταξινομημένο ως προς το γνώρισμα της συνένωσης ή όχι.
(i) Η σχέση R είναι η εσωτερική σχέση στον εμφωλευμένο βρόγχο (loop) υπολογισμού της συνένωσης
(ii) Η σχέση R είναι η εξωτερική σχέση στον εμφωλευμένο βρόγχο (loop) υπολογισμού της συνένωσης.