



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΕΚΔΟΣΗ ΗΛΕΚΤΡΟΝΙΚΩΝ ΕΠΙΧΕΙΡΗΣΕΩΝ ΚΑΙ ΔΙΟΙΚΗΤΙΚΩΝ ΥΠΗΡΕΣΙΩΝ
ΤΟΜΕΑΣ ΠΕΡΙΣΤΑΣΙΑΣ ΤΗΣ ΠΡΟΓΝΩΣΤΙΚΗΣ ΚΑΙ ΥΠΟΛΟΓΙΣΤΙΚΗΣ

Βάσεις Δεδομένων
Ακ. Έτος 2016-2017, 7ο Εξάμηνο Γενικών
Η. Καζάνης, Κ. Δόκα
Εξέταση Φεβρουαρίου
Χρόνος εξέτασης: 3 ώρες

Θέμα 1 (30%)

Θεωρήστε σχεσιακή βάση δεδομένων που περιγράφει τη βιομηχανία του κινηματογράφου.

- ACTOR (actorid, name)
- FILM (filmid, title, year, directorid)
- PERFORMANCE (actorid, filmid, character)
- DIRECTOR (directorid, name)

- (α) Γράψτε σε Σχεσιακή Αλγεβρα (Relational Algebra) τα παρακάτω ερωτήματα (queries).
- Βρες τα ονόματα όλων των ηθοποιών που έπαιξαν τον χαρακτήρα του Batman, μαζί με τον τίτλο και τη χρονολογία κυκλοφορίας των αντίστοιχων ταινιών.
 - Βρες τα ονόματα όλων των ταινιών που έχουν συμπρωταγωνιστήσει ο Ryan Gosling και η Emma Stone.
 - Βρες τα ονόματα των ηθοποιών που έχουν παίξει σε όλες τις ταινίες που έχει σκηνοθετήσει ο Quentin Tarantino.
 - Γράψτε σε SQL τα παρακάτω ερωτήματα (queries).
 - Βρες το πλήθος των ηθοποιών που έπαιξαν σε οποιαδήποτε ταινία ανά χρονιά.
 - Βρες τους τίτλους όλων των ταινιών για τις οποίες έχει γίνει remake, μαζί με το πλήθος των remakes. (hint: Οι ταινίες πρέπει να έχουν το ίδιο title αλλά διαφορετικό id)
 - Βρες τα ονόματα των ηθοποιών που έχουν παίξει σε τουλάχιστον μια ταινία που σκηνοθετήσει ο Tarantino.

Θέμα 2 (20%)

(α) Για τις παρακάτω σχέσεις και αντίστοιχες συνειρημένες εξαρτήσεις:

- $R(A, B, C, D, E) \quad F = \{A \rightarrow B, C \rightarrow D\}$
- $R(A, B, C, D) \quad F = \{A \rightarrow B, BC \rightarrow D, A \rightarrow C\}$

- απαντήστε στο ακόλουθο ερώτημα:
- Είναι το F μια ελάχιστη κανονική κάλυψη;
- Βρείτε όλα τα υποσύνολα βασικών της σχέσης R .
- Αναγνωρίστε αν το σύνολο γνωρισμάτων κανονίζει την κανονική μορφή 3CNF.
- Αν το σύνολο δεν είναι σε 3CNF, επισημαίνετε ποια ή ποιες εξαρτήσεις την παραβιάζουν και βρείτε μια απλοποιημένη σε σύνολο από 3CNF σχήματα που διατηρούν τις εξαρτήσεις και τη συνένοια χωρίς επώλες.

Ενα σχήμα είναι σε κανονική μορφή 3CNF εάν και μόνο εάν για κάθε $X \rightarrow Y$ που σχηματίζει ένα από των δύο συμβαίνει είτε το $X \rightarrow Y$ είναι τετριμμένη συναρτησιακή εξάρτηση ή το X είναι υπερασπύ.

(β) Θεωρήστε τα παρακάτω σύνολα συναρτησιακών εξαρτήσεων πάνω στη σχέση $R(A, B, C)$.

$$F1 = \{A \rightarrow B, B \rightarrow C\}$$

$$F2 = \{A \rightarrow B, A \rightarrow C\}$$

$$F3 = \{A \rightarrow B, AB \rightarrow C\}$$

Ποια από αυτά τα σύνολα είναι ισοδύναμα;

Ένας τρόπος για να δείχθει ότι 2 σύνολα εξαρτήσεων είναι ισοδύναμα είναι να υπολογιστούν τα κλειδιά των δύο συνόλων και να φανεί ότι είναι ίδια. Ένας ευκολότερος τρόπος είναι να χτίσουμε ορισμένοι μετασχηματισμοί χρησιμοποιώντας τους κανόνες Armstrong και να φανεί ότι τα δύο σύνολα εξαρτήσεων είναι ουσιαστικά τα ίδια.

Πενθθυμίζονται ορισμένοι κανόνες που ισχύουν:

- Αν Y υποσύνολο του X , τότε $X \rightarrow Y$ (ανακλαστικότητα)
- Αν $X \rightarrow Y$ ισχύει, τότε $XZ \rightarrow YZ$ (προσαύξηση)
- Αν $X \rightarrow Y$ και $Y \rightarrow Z$, τότε $X \rightarrow Z$ (μεταβατικότητα)
- Αν $X \rightarrow Y$ και $X \rightarrow Z$, τότε $X \rightarrow YZ$ (ένωση)
- Αν $X \rightarrow YZ$, τότε $X \rightarrow Y$ και $X \rightarrow Z$ (αποσύνθεση)
- Αν $X \rightarrow Y$ και $ZY \rightarrow P$, τότε $XZ \rightarrow P$ (ψευδομεταβατικότητα)

Θέμα 3 (20%)

(α) Θεωρήστε τις σχέσεις

$R1(A, B, C)$

$R2(C, D, E)$

$R3(E, F)$

Εκ πρώτης όρας κλειδιά τα A, C και E αντίστοιχα. Η $R1$ έχει 1000, η $R2$ έχει 1500 και η $R3$ έχει 750 tuples. Δώστε μια εκτίμηση για το μέγεθος του αποτελέσματος της συνένωσης $R1 \bowtie R2 \bowtie R3$.

και δώστε μια αποδοτική στρατηγική για τον υπολογισμό της (τι είδους ευρετήριο θα κατασκευάζατε και σε ποια πεδία, π.αλγόριθμο θα εφαρμόζατε και πώς θα εκμεταλλευόσασταν τους δείκτες που δημιουργήσατε).

(β) Θεωρήστε τη σχέση

EMPLOYEE(eid, age, salary)

και υποθέστε ότι το μοναδικό ευρετήριο που υπάρχει είναι ένα ευρετήριο B+-tree στο γνώρισμα age. Ποιος θα ήταν ο καλύτερος τρόπος επίλυσης των παρακάτω ερωτημάτων που περιλαμβάνουν άρνηση (negation);

- i. $\sigma_{\text{age} < 40}(\text{EMPLOYEE})$
- ii. $\sigma_{\text{age} = 40}(\text{EMPLOYEE})$
- iii. $\sigma_{\text{age} < 40 \vee \text{salary} < 50000}(\text{EMPLOYEE})$

Θέμα 4 (30%)

Πάντα ονειρευόσασταν να ασχοληθείτε με την τέχνη, αλλά σας κέρδισε η επιστήμη (των υπολογιστών). Η παλιά αγάπη όμως ξύπνησε πάλι μέσα σας όταν σας ζητήθηκε να σχεδιάσετε μια βάση που κρατά πληροφορίες για το μουσείο τέχνης CoolArt. Το μουσείο έχει μια συλλογή από έργα τέχνης. Κάθε έργο τέχνης έχει μοναδικό id, έναν δημιουργό, μια χρονολογία κι έναν τίτλο. Οι δημιουργοί των έργων τέχνης είναι καλλιτέχνες με γνωστό όνομα, εθνικότητα και χρονολογία γέννησης.

Τα έργα τέχνης κατηγοριοποιούνται με βάση τον τύπο τους. Υπάρχουν 2 βασικοί τύποι: πίνακες και γλυπτά. Κάθε πίνακας χαρακτηρίζεται από το είδος του και την τεχνοτροπία του. Κάθε γλυπτό χαρακτηρίζεται από το υλικό του, το ύψος και το βάρος του.

Διάφορες εκθέσεις φιλοξενούνται στο μουσείο, καθεμιά από τις οποίες έχει όνομα, ημερομηνία έναρξης και λήξης. Οι εκθέσεις περιλαμβάνουν έργα τέχνης που εκτίθενται κατά τη διάρκειά τους.

(α) Σχεδιάστε ένα πλήρες διάγραμμα οντοτήτων-συσχετίσεων (E-R diagram), για τη Βάση Δεδομένων. Στο σχεδιασμό μπορείτε να κάνετε δικές σας (λογικές) υποθέσεις/παραδοχές, οι οποίες πρέπει να γράφονται ρητά. Το διάγραμμα πρέπει να είναι πλήρες υπό την έννοια ότι όλες οι ιδιότητες των οντοτήτων-συσχετίσεων πρέπει να δείχνονται (λόγοι πληθικότητας - cardinality ratio, περιορισμοί συμμετοχής - participation constraints, περιορισμοί κάλυψης - disjointness constraints, περιορισμοί πληρότητας - completeness constraints, κλειδιά, κλπ).

(β) Δώστε την αντίστοιχη περιγραφή (schema) ολόκληρης της Βάσης Δεδομένων στο σχεσιακό μοντέλο.

Σημείωση: Μπορείτε είτε να μετατρέψετε το E-R διάγραμμα του (α) σε ανάλογη σχεσιακή περιγραφή είτε να παρουσιάσετε το σχεσιακό μοντέλο απευθείας με τα στοιχεία περιγραφής που δίνονται στην εκφώνηση.