

2^ο Σύνολο Ασκήσεων

Καταληκτική Ημερομηνία Παράδοσης: Τρίτη 6 Δεκεμβρίου 2016, στο μάθημα
Θεματική Ενότητα: Γλώσσες ερωτήσεων (Γυπικές Γλώσσες (Σχεσιακή Άλγεβρα, Σχεσιακός Λογισμός), SQL). Λογικός Σχεδιασμός

Για γενικές οδηγίες σχετικά με τις ασκήσεις, συμβουλευτείτε την ιστοσελίδα του μαθήματος.

Οι Ασκήσεις 1, 2, 3, 4 αναφέρονται σε μία σχεσιακή βάση δεδομένων για ένα κοινωνικό δίκτυο με το παρακάτω σχήμα.

USER(user-id, username, country, sex)
GROUP(group-name, user-id, date-of-joining)
POST(post-id, user-id, date-of-post, time-of-post)
POST-CONTENT(post-id, keyword)

Ο πίνακας (σχέση) USER έχει το id (user-id), όνομα (username) τη χώρα (country) και το φύλο (sex) του χρήστη. Δεν υπάρχουν χρήστες με το ίδιο id καθώς και χρήστες με το ίδιο username. Οι χρήστες συμμετέχουν σε ομάδες (group). Η σχετική πληροφορία βρίσκεται στον πίνακα GROUP όπου διατηρούμε το όνομα της ομάδας (group-name), το id του χρήστη (user-id) και την ημερομηνία της εγγραφής του στην ομάδα (date-of-joining). Οι χρήστες γράφουν μηνύματα που αποτελούνται από 0 έως 10 λέξεις (keywords). Η σχετική πληροφορία αποθηκεύεται στον πίνακα POST και POST-CONTENT. Στον πίνακα POST έχουμε το id του μηνύματος (post-id), το id του χρήστη που το έγραψε (user-id) και τον ακριβή χρόνο δημοσίευσης του μηνύματος - συγκεκριμένα, την ημερομηνία (date-of-post) και τη χρονική στιγμή (time-of-post), ενώ στον πίνακα POST-CONTENT έχουμε τις λέξεις (keyword) του μηνύματος.

Άσκηση 1 (σχεδιασμός)

- (α) Δώστε ένα κατάλληλο μοντέλο οντοτήτων/συσχετίσεων για το παραπάνω σχεσιακό σχήμα.
- (β) Δώστε τις (μη τετριμμένες) συναρτησιακές εξαρτήσεις που ισχύουν στον πίνακα USER.
- (γ) Υποθέστε ότι τα μηνύματα δεν έχουν id (δηλαδή, ότι δεν υπάρχει το γνώρισμα post-id). Ακόμα υποθέστε ότι ένας χρήστης μια συγκεκριμένη ημερομηνία και χρονική στιγμή μπορεί να γράψει μόνο ένα μήνυμα. Τροποποιήστε κατάλληλα το σχεσιακό σχήμα και το μοντέλο οντοτήτων/συσχετίσεων.

Άσκηση 2 (σχεσιακή άλγεβρα)

Διατυπώστε σε σχεσιακή άλγεβρα (χωρίς χρήση συναθροιστικών συναρτήσεων) ερωτήσεις που να δίνουν ως αποτέλεσμα:

- (α) Το περιεχόμενο (δηλαδή, τις λέξεις) όλων των μηνυμάτων που έχουν γράψει οι χρήστες που ανήκουν στην ομάδα με group-name 'book-lovers'.
- (β) Το περιεχόμενο (δηλαδή, τις λέξεις) όλων των μηνυμάτων που έχουν γράψει οι χρήστες που ανήκουν στην ομάδα με group-name 'book-lovers' αφού έγιναν μέλη αυτής της ομάδας (δηλαδή, συμπεριλάβετε μόνο τα μηνύματα που έγραψαν οι χρήστες σε ημερομηνία που είναι μεταγενέστερη της ημερομηνίας εγγραφής τους στην ομάδα).
- (γ) Το πιο καινούργιο μέλος της ομάδας 'book-lovers', δηλαδή, το μέλος με την πιο πρόσφατη ημερομηνία εγγραφής.
- (δ) Τα username των παθητικών χρηστών, δηλαδή των χρηστών που ούτε έχουν γράψει μήνυμα, ούτε συμμετέχουν σε ομάδα.
- (ε) Ζεύγη από user-id όμοιων χρηστών. Δύο χρήστες θεωρούνται όμοιοι αν ανήκουν σε τουλάχιστον δύο ίδιες ομάδες και έχουν γράψει μηνύματα που περιέχουν τουλάχιστον 3 ίδιες λέξεις. Κάθε ζεύγος πρέπει να εμφανίζεται στο αποτέλεσμα μόνο μια φορά, δηλαδή, αν εμφανίζεται το ζεύγος (id1, id2), να μην εμφανίζεται και το (id2, id1).

Άσκηση 3 (σχεσιακή λογισμός)

Διατυπώστε σε σχεσιακό λογισμό πλειάδων τις ερωτήσεις (β), (γ) και (ε) της Άσκησης 1. Για την ερώτηση (γ) δώστε δύο διατυπώσεις μία με χρήση του υπαρξιακού και μία με χρήση του καθολικού ποσοδεικτη.

Άσκηση 4 (SQL)

Τα παρακάτω πρέπει να υλοποιηθούν στη MySQL (ή στην SQLite).

(α) Δημιουργείτε τους πίνακες. Ορίστε κατάλληλα πεδία ορισμού για τα γνωρίσματα τους, καθώς και τα κλειδιά, ξένα κλειδιά και όποιους άλλους περιορισμούς θεωρείστε απαραίτητους.

(β) Δώστε SQL ερωτήσεις που να δίνουν ως αποτέλεσμα τα παρακάτω:

- (i) Το ερώτημα (β) της Άσκησης 1.
- (ii) Το ερώτημα (δ) της Άσκησης 1.
- (iii) Το ερώτημα (ε) της Άσκησης 1.
- (iv) Για τα μηνύματα από χρήστες που ανήκουν στην ομάδα 'book-lovers', δώστε τις λέξεις (keywords) που εμφανίζονται σε τουλάχιστον 3 από αυτά τα μηνύματα σε ζεύγη ως εξής keyword1, count) όπου count είναι το πλήθος εμφανίσεων της λέξης keyword1 σε αυτά μηνύματα.
- (v) Τις 10 πιο δημοφιλείς λέξεις (δηλαδή τις λέξεις που εμφανίζονται στα περισσότερα μηνύματα).
- (vi) Για όλους τους χρήστες το πλήθος των μηνυμάτων που έχουν γράψει. Ειδικά, για τους χρήστες που δεν έχουν γράψει κάποιο μήνυμα να εμφανίζεται ο αριθμός 0.

(γ) Τα παρακάτω ερωτήματα αφορούν στην χρήση όψεων.

- (i) Ορίστε μια όψη GROUP-COUNT(group-name, count) που να περιέχει για κάθε ομάδα (group-name) το πλήθος των μηνυμάτων που έχουν γράψει τα μέλη της. Δείτε το περιεχόμενο αυτής της όψης, χρησιμοποιώντας το select * from GROUP-COUNT. Στη συνέχεια, εισάγετε ένα καινούργιο μήνυμα στους δύο σχετικούς πίνακες (POST και POST-CONTENT). Δείτε το περιεχόμενο της όψης μετά την εισαγωγή χρησιμοποιώντας πάλι το select * from GROUP-COUNT.
- (ii) Είναι η παραπάνω όψη τροποποιήσιμη (updatable). Εξηγήστε.

(δ) Κάντε την παρακάτω τροποποίηση

Για κάθε μήνυμα με ημερομηνία μέσα στο 2016, αν αυτό το μήνυμα έχει λιγότερο από 2 λέξεις, προσθέστε τη λέξη 'dbassgn2'.

Άσκηση 5 (λογικός σχεδιασμός)

(α) Έστω το σχεσιακό σχήμα $R(A, B, C, D, E)$ στο οποίο ισχύει το σύνολο συναρτησιακών εξαρτήσεων $F = A \rightarrow B, B \rightarrow C, D \rightarrow E$.

(i) Θεωρείστε τη διάσπαση του R σε $R_1(A, B, C)$ και $R_2(D, E)$. Για το καθένα από τα παρακάτω δείξτε αν ισχύει ή όχι.

(1) Η διάσπαση διατηρεί τις εξαρτήσεις (2) Η διάσπαση είναι χωρίς απώλειες στη συνένωση (3) Η R_1 και η R_2 είναι σε 3NF (4) Η R_1 και η R_2 είναι σε BCNF

(ii) Θεωρείστε τη διάσπαση του R σε $R_1(A, B, C)$, $R_2(A, D)$ και $R_3(D, E)$. Για το καθένα από τα παρακάτω εξηγήστε αν ισχύει ή όχι.

(1) Η διάσπαση διατηρεί τις εξαρτήσεις (2) Η διάσπαση είναι χωρίς απώλειες στη συνένωση (3) Η R_1 , R_2 και R_3 είναι σε 3NF (4) Η R_1 , η R_2 και R_3 είναι σε BCNF

Άσκηση 6 (σχεσιακή άλγεβρα και λογικός σχεδιασμός)

Θεωρείστε δύο σχέσεις R και S με σχήματα $R(A, B, C)$ και $S(A, D)$ αντίστοιχα. Υποθέστε ότι το γνώρισμα A της R είναι πρωτεύον κλειδί για την R και το γνώρισμα A της S είναι ξένο κλειδί που αναφέρεται στην R . Επίσης, υποθέστε ότι τα γνωρίσματα της σχέσης S δεν μπορούν να πάρουν null τιμές. Το * συμβολίζει τη φυσική συνένωση. Έστω ότι η R έχει r πλειάδες και η S έχει r πλειάδες.

(α) Ποιος είναι ο μικρότερος και ποιος ο μεγαλύτερος αριθμός πλειάδων που μπορεί να επιστρέψει καθένα από τις παρακάτω ερωτήσεις και γιατί:

- (i) $\pi_A(R) - \pi_A(S)$
- (ii) $\pi_A(S) - \pi_A(R)$
- (iii) $R * S$

(β) Ποια συναρτησιακή εξάρτηση πρέπει να ισχύει στην R ώστε το αποτέλεσμα των παρακάτω δύο ερωτήσεων (i) και (ii) να είναι πάντα (δηλαδή, σε οποιοδήποτε έγκυρο στιγμιότυπο της βάσης δεδομένων) το ίδιο και εξηγήστε γιατί

(i) $\pi_B(\sigma_{C=c_1}(R)) - \pi_B(\sigma_{C=c_2}(R))$ και (ii) $\pi_B(\sigma_{C=c_1}(R))$, όπου c_1, c_2 ανήκουν στο πεδίο τιμών του C .