

2^ο Σύνολο Ασκήσεων

Καταληκτική Ημερομηνία Παράδοσης: Παρασκευή 17 Δεκεμβρίου 2021, 5μμ

Θεματική Ενότητα: Γλώσσες Ερωτήσεων: τυπικές γλώσσες (σχεσιακή άλγεβρα, σχεσιακός λογισμός), SQL.

Η παράδοση θα γίνει μέσω του ecourse
Παραδώστε ένα pdf αρχείο με τις απαντήσεις σας σε όλες τις ασκήσεις.

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

Οι ασκήσεις αφορούν δεδομένα από το YELP¹, που είναι ένα σύστημα για αξιολογήσεις επιχειρήσεων. Στην Εικόνα 1, βλέπετε το διάγραμμα του σχήματος όλης της βάσης δεδομένων του YELP. Εμείς θα χρησιμοποιήσουμε μόνο τους πίνακες: business, user, review, και friends. Επίσης θα χρησιμοποιήσουμε, ένα υποσύνολο των δεδομένων που αφορά αξιολογήσεις επιχειρήσεων της Βοστώνης το 2020. Οι πίνακες έχουν τα παρακάτω πεδία:

- Ο πίνακας **business** περιέχει τις παρακάτω πληροφορίες για τις επιχειρήσεις. Το αναγνωριστικό (id, που είναι μοναδικό), όνομα (name), γειτονιά (neighborhood), διεύθυνση (address), πόλη (city), πολιτεία (state), ταχυδρομικό κώδικα (postal_code), γεωγραφικό πλάτος και μήκος (latitude, longitude) μέσω αριθμού αστεριών (stars), αριθμό αξιολογήσεων (review_count), και αν η επιχείρηση είναι ανοικτή ή όχι (is_open).
- Ο πίνακας **user** περιέχει τις παρακάτω πληροφορίες για τους χρήστες. Το αναγνωριστικό (user_id, που είναι μοναδικό), όνομα (name), πλήθος αξιολογήσεων (review_count), την ημερομηνία και ώρα εγγραφής του χρήστη στο σύστημα (yelping_since), το μέσο αριθμό αστεριών (average_stars), τον αριθμό των αξιολογήσεων που ο χρήστης έχει χαρακτηρίσει χρήσιμες (useful), αστείες (funny) και cool, τον αριθμό των οπαδών που έχει ο χρήστης (fans) καθώς και τους αριθμούς από τις θετικές κριτικές (compliment) που έχει δεχθεί ο χρήστης σε διάφορα θέματα.
- Ο πίνακας **review** περιέχει τις παρακάτω πληροφορίες για τις αξιολογήσεις. Το αναγνωριστικό της αξιολόγησης (id), το αναγνωριστικό του χρήστη (user_id) που έκανε την αξιολόγηση, το αναγνωριστικό της επιχείρησης που αφορά η αξιολόγηση (business_id), τα αστέρια (stars), τον αριθμό των ψήφων που πήρε ως χρήσιμη (useful), αστεία (funny) και cool, το κείμενο (text), καθώς και την ημερομηνία που γράφτηκε (date).
- Ο πίνακας **friend** περιέχει τα id των χρηστών που είναι φίλοι (user_id, friend_user_id).

Σημειώστε ότι ο πίνακας review δεν περιέχει όλες τις αξιολογήσεις των επιχειρήσεων, οπότε οι τιμές του πεδίου review_count στους πίνακες business και user και των πεδίων stars στον πίνακα business και average_stars στον πίνακα user δεν είναι απαραίτητα ίσες αντίστοιχα με το πλήθος και το μέσο όρο του αριθμού των αστεριών των αξιολογήσεων της επιχείρησης ή του χρήστη που εμφανίζονται στον πίνακα review. Επίσης, οι τιμές κάποιων πεδίων στα δεδομένα έχουν null τιμές.

Άσκηση 1 (σχεσιακή άλγεβρα)

Διατυπώστε σε σχεσιακή άλγεβρα (χωρίς χρήση συναθροιστικών συναρτήσεων τις οποίες δεν καλύψαμε στο μάθημα) ερωτήσεις που δίνουν ως αποτέλεσμα:

(α) Τις επιχειρήσεις (το id και το όνομα τους) που έχουν περισσότερες από 30 αξιολογήσεις (σε αυτό το ερώτημα χρησιμοποιείστε το πεδίο review_count του πίνακα business), μέσο αριθμό αστεριών μεγαλύτερο του 4 (σε αυτό το ερώτημα χρησιμοποιείστε το πεδίο stars του πίνακα business), και όλες τους οι αξιολογήσεις του στον πίνακα review έχουν τουλάχιστον 3 αστέρια.

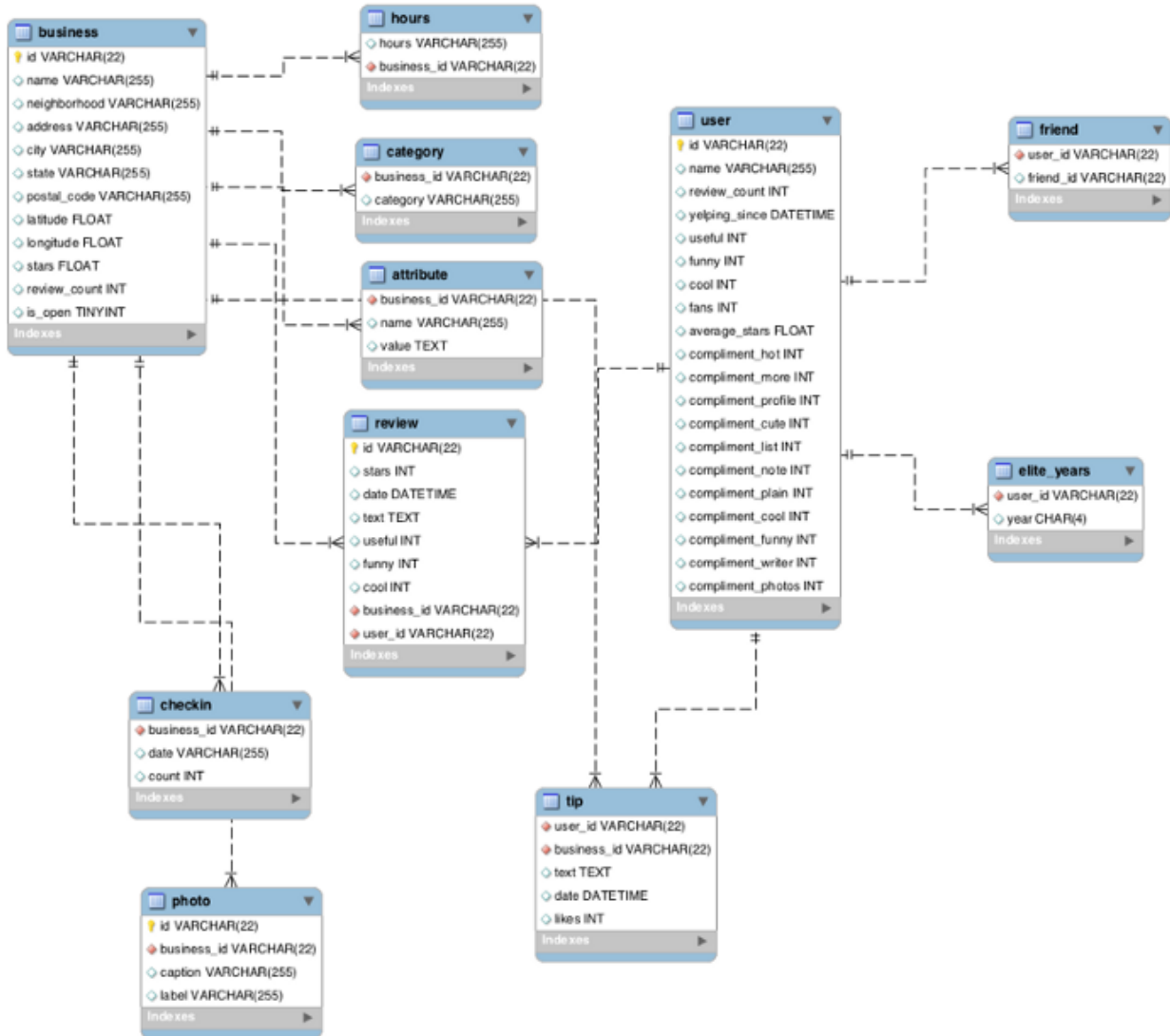
¹ <https://www.yelp.com/dataset>

(β) Ζεύγη φίλων (ζεύγη από user_id) που έχουν αξιολογήσει τουλάχιστον μια ίδια επιχείρηση. Αν εμφανίζεται το ζεύγος (id1, id2) να μην εμφανίζεται το συμμετρικό του (id2, id1) και επίσης να μην εμφανίζονται ταυτοτικά ζεύγη δηλαδή ζεύγη (id1, id2) όπου id2 = id1.

(γ) Τις επιχειρήσεις (το id τους) που έχουν δεχθεί και θετικές αξιολογήσεις (δηλαδή, αξιολογήσεις με 4 ή 5 αστέρια) και αρνητικές αξιολογήσεις (δηλαδή, αξιολογήσεις με 1 ή 2 αστέρια).

(δ) Για το χρήστη με id = 59181, τις επιχειρήσεις (το id τους) που αυτός έχει αξιολογήσει με 5 αστέρια και κάθε φίλος του που τις έχει αξιολογήσει τους έχει δώσει λιγότερα από 3 αστέρια.

(ε) Τη μικρότερη ημερομηνία που κάποιος χρήστης εγγράφηκε στο σύστημα.



Εικόνα 1. Το σχήμα της βάσης δεδομένων για τις Ασκήσεις 1, 2 και 3. Θα χρησιμοποιήσουμε μόνο τους πίνακες: business, user, review, και friend.

Άσκηση 2 (σχεσιακός λογισμός)

Διατυπώστε σε σχεσιακό λογισμό:

(α) Την ερώτηση (β) της Άσκησης 1.

(β) Την ερώτηση (δ) της Άσκησης 1. Δώστε δύο διαφορετικές ερωτήσεις μία με χρήση του υπαρχιακού και μία με χρήση του καθολικού ποσοδείκτη.

Άσκηση 3 (SQL)

Τα παρακάτω πρέπει να υλοποιηθούν στη MySQL.

(α) Δημιουργείστε τους πίνακες και φορτώστε τα δεδομένα από τη σελίδα του μαθήματος.

(β) Διατυπώστε SQL ερωτήσεις που να δίνουν ως αποτέλεσμα:

(i) Την ερώτηση (α) της Άσκησης 1.

(ii) Για κάθε επιχείρηση (το id της) που έχει λάβει τουλάχιστον 3 αξιολογήσεις με 5 ή 4 αστέρια, το πλήθος (αθροιστικά) αυτών των αξιολογήσεων.

(iii) Το ποσοστό των επιχειρήσεων για τις οποίες οι θετικές αξιολογήσεις (αξιολογήσεις με 4 ή 5 αστέρια) είναι περισσότερες από τις αρνητικές αξιολογήσεις (αξιολογήσεις με 1 ή 2 αστέρια).

(iv) Για κάθε επιχείρηση (το id της), το μέσο αριθμό αστεριών των αξιολογήσεων που έχει δεχθεί λαμβάνοντας υπόψη μόνο τις αξιολογήσεις που έχουν τουλάχιστον 2 ψήφους ως χρήσιμες. Για τις επιχειρήσεις, για τις οποίες δεν υπάρχουν τέτοιες αξιολογήσεις να τυπώνεται ο αριθμός 0.

(v) Το πλήθος των χρηστών που έχουν δώσει περισσότερες αξιολογήσεις με 5 αστέρια από ό,τι αθροιστικά οι φίλοι τους.

(γ)

(i) Εισάγετε την τιμή “MA” στο πεδίο state του πίνακα business.

(ii) Εισάγετε ένα νέο χρήστη στον πίνακα user. Στη συνέχεια εισάγετε μια αξιολόγηση αυτού του χρήστη με 3 αστέρια για κάθε επιχείρηση που είναι ανοικτή.

(δ) [προαιρετικό +15 μονάδες] Στόχος της άσκησης είναι να χρησιμοποιήσετε SQL από Java ή Python. Για να συνδεθείτε (connect) με τη βάση δεδομένων, χρησιμοποιείστε τον κατάλληλο driver.

Γράψτε ένα πρόγραμμα που:

(i) Ο χρήστης θα δίνει ως είσοδο το όνομα ενός πίνακα, το όνομα ενός πεδίου του και μια τιμή για αυτό το πεδίο και θα τυπώνει στην οθόνη τις πλειάδες του πίνακα που έχουν την συγκεκριμένη τιμή σε αυτό πεδίο.

(ii) Θα τυπώνει για κάθε επιχείρηση το πλήθος των αξιολογήσεων για κάθε αριθμό αστεριών. Συγκεκριμένα, τα αποτελέσματα θα είναι μια ακολουθία από γραμμές. Κάθε γραμμή θα ξεκινά με το id και το όνομα της επιχείρησης και θα ακολουθούν πέντε ζεύγη (ένα για κάθε αριθμό αστεριών) της μορφής (αριθμός αστεριών, πλήθος αξιολογήσεων της επιχείρησης με αυτό τον αριθμό αστεριών), π.χ., ένα ζεύγος (5, 15) σημαίνει 15 αξιολογήσεις με 5 αστέρια.