

תרגיל 2 : אלגברה רלציונית וSQL

תאריך הגשה : 23: 55, 22.11.20.

הוראות הגשה :

בתרגיל זה אתם נדרשים להגיש קובץ zip בודד שיכלול את הקבצים הבאים :

- ex2.pdf עם התשובות לשאלות בחלק של אלגברה רלציונית.
- q1.sql
- q2.sql
- q3.sql
- q4.sql
- q5.sql
- q6.sql
- README שמכיל שורה בודדת ובו ה-login של הסטודנט שמגיש את התרגיל. אם התרגיל מוגש בזוגות, על שורה זאת להכיל את שני ה-login מופרדים בפסיק.

שימו לב :

- נא לקרוא על הדרישות המנהליות של הקורס בלינק באתר הקורס כדי למלא את ההוראות להגשה של קבצים שרוקים!
- תרגיל מוקלד יזכה ב- 2 נקודות בונוס!

אלגברה רלציונית (50 נקודות) :

שאלה א' (36 נקודות) :

נתונים היחסים הבאים מתוך מסד נתונים של IMDb :

Movies (movieId, title, rating, year, duration, genre)

Actors (actorId, name, byear, dyear)

PlaysIn (movieId, actorId, character)

הערות :

- לכל סרט (Movie) יש מספר מזהה (movieId), כותרת (title), דירוג הסרט (rating), שנת יציאה (year), משך הסרט בדקות (duration) וז'אנר (genre).
- לכל שחקן (Actor) יש מספר מזהה (actorId), שם (name), שנת לידה (byear), ושנת פטירה (dyear). עבור שחקן חי, dyear הוא null.
- לכל משחק בסרט (PlaysIn) יש מספר מזהה של סרט ושחקן (movieId, actorId), ואת שם הדמות שהשחקן שיחק (character).

כתבו את השאילתות הבאות באלגברה רלציונית :

1. החזר את שמות כל השחקנים ששיחקו בסרט דמות ששמה "George".
2. החזר את המספר המזהה וכותרת של כל הסרטים ששיחקו בהם שחקנים שמתו בשנת היציאה של הסרט, וגם הסרט הוא בז'אנר 'Documentary' או 'Drama'.

3. החזר את המספר המזהה של כל השחקנים שגם השתתפו בסרט ששיחק בו השחקן 'Charles Chaplin', וגם השתתפו בסרט כלשהו באורך של יותר מ-90 דקות.
4. החזר מספר מזהה של שחקנים ששיחקו בלפחות סרט אחד, וכל הסרטים בהם שחקו קבלו ערך rating גבוה מ-7.
5. החזר את השמות של שחקנים ששיחקו בכל הסרטים ששיחק בהם השחקן עם id=100. (לא כולל השחקן עם id=100)
6. החזר את שמות הסרטים ששיחקו בהם בדיוק 2 שחקנים מעל גיל 70 (היו בני יותר מ-70 כשיצא הסרט). שימו לב שניתן לכתוב בתנאי בחירה של אלגברה רלציונית ביטוי אלגברי, למשל, אם $R(A,B)$ יחס, ניתן לכתוב ביטוי כגון $\sigma_{A>B}R$.

שאלה ב' (14 נקודות):

- נתונים היחסים: $R(A,B)$, $S(B,C,D)$
לכל אחד מהביטויים הבאים:
- אם הביטויים שקולים, הוכיחו זאת.
 - אחרת, אם יש הכלה בכיוון אחד בלבד, הוכיחו את כיוון ההכלה והראו דוגמה נגדית לכיוון השני.
 - אם אין הכלה באף כיוון, הראו דוגמאות (ש) שמראות זאת.

$$1. \pi_{A,D}(R \bowtie \pi_{B,D}\sigma_{C=8}(S)) \equiv? \pi_{A,D}\sigma_{C=8}(R \bowtie S)$$

$$2. \pi_{A,D}(R \bowtie S) \equiv? \pi_A R \bowtie \pi_D S$$

שאלות SQL (50 נקודות):

בכל התשובות לשאלות בחלק זה:

- השתמשו ב-SELECT DISTINCT כדי למנוע כפילויות בתשובות (אם כפילויות עלולות להווצר בתשובה).
 - **שימו לב:** בכל סעיף כתוב באיזה סדר למיין את התוצאות וכן את שמות העמודות בתוצאה.
- באתר הקורס יש קובץ create.sql המכיל הגדרות עבור הטבלאות וקובץ drop.sql המכיל פקודות המוחקות את הטבלאות. כמו כן, נתונים הקבצים:

- actors.csv
- movies.csv
- playsIn.csv

הקבצים מכילים מידע חלקי אך אמיתי אודות סרטים ושחקנים מהאתר IMDb. את המידע המלא ניתן למצוא ב-<https://www.imdb.com/interfaces>
הקבצים שלנו מכילים מידע על כ-10,000 סרטים, שעבר "ניקוי" והתאמה לצורך התרגיל.

ניתן למצוא את הקבצים גם במערכת המחשבים במעבדה בתיקיה:

~ db/data/

ניתן להעתיק אותם לתיקיה שלכם.

על מנת לבדוק את התרגיל שלכם, יש ליצור את הטבלאות בעזרת create.sql, ולטעון לתוכן נתונים בעזרת הפקודות

cat movies.csv | psql -h dbcourse public -c "copy movies from STDIN DELIMITER ',' CSV HEADER"

cat actors.csv | psql -h dbcourse public -c "copy actors from STDIN DELIMITER ',' CSV HEADER"

cat playsIn.csv | psql -h dbcourse public -c "copy playsIn from STDIN DELIMITER ',' CSV HEADER"

כתבו את השאילתות הבאות בSQL. שם הקובץ שבו צריכה להופיע התשובה לכל שאלה נמצא בתחילת השאלה. שימו לב ששאלות 1-6 דומות לשאלות בחלק של אלגברה, אך ישנם שינויים שיש לשים לב אליהם בשאלות 1,5,6.

1. (q1.sql) מצא את שמות כל השחקנים ששיחקו בסרט דמות ששמה מתחיל ב- "George", והחזר את שמות השחקנים ושם הדמות ששיחקו.

יש להחזיר טבלה עם העמודות name, character ממיון ראשוני לפי actorName ומיון שניוני לפי character.

2. (q2.sql) החזר את המספר המזהה וכותרת של כל הסרטים ששיחקו בהם שחקנים שמתו בשנת היציאה של הסרט, וגם הסרט הוא בז'אנר 'Documentary' או 'Drama'.

יש להחזיר טבלה עם העמודות movieid, title ממיון ראשוני לפי movieid ומיון שניוני לפי title.

3. (q3.sql) החזר את המספר המזהה של כל השחקנים שהשתתפו בסרט ששיחק בו השחקן 'Charles Chaplin', וגם השתתפו בסרט כלשהו באורך של יותר מ-90 דקות.

יש להחזיר טבלה עם העמודה actorId ממיון לפי actorId.

4. (q4.sql) החזר מספר מזהה של שחקנים ששיחקו רק בסרטים עם rating גבוה מ-7, ושיחקו לפחות בסרט אחד.

יש להחזיר טבלה עם העמודה actorId ממיון לפי actorId.

5. (q5.sql) החזר זוגות של actorid1, actorid2 כך ש:

- actorid2 משחק לפחות בסרט אחד
- השחקן עם מספר מזהה actorid1 משחק (לפחות) בכל הסרטים שמשחק בהם השחקן עם מספר מזהה actorid2
- actorid2 ≠ actorid1

יש להחזיר טבלה עם העמודות הנקראות actorid1, actorid2 ממיון לפי actorid1 ומיון שניוני לפי actorid2. יש להחזיר כל זוג פעם אחת בלבד.

6. (q6.sql) החזר מספר מזהה וכותרת של כל הסרטים שבהם יש לפחות שני שחקנים שהפרש גילם הוא יותר מ-60.

יש להחזיר טבלה עם העמודות movieid, title ממיון ראשוני לפי movieid ומיון שניוני לפי title.

בהצלחה!