

סמסטר חורף התשע"ט

מרצה: פרופ' עודד שמואלי

מתרגלים: ענבר קסלסי

משה סבאג

אסף ישורון

מערכות מסד נתונים

236363

מועד ב'

3 במרץ 2019

פירוט החלקים והניקוד:

חלק	נושא	מס' שאלות בחלק	ניקוד
1	תכן מסדי נתונים	2	35
2	שאליות במודל היחסים	2	40
3	מודלים לא יחסיים	2	25

הנחיות לנבחנים

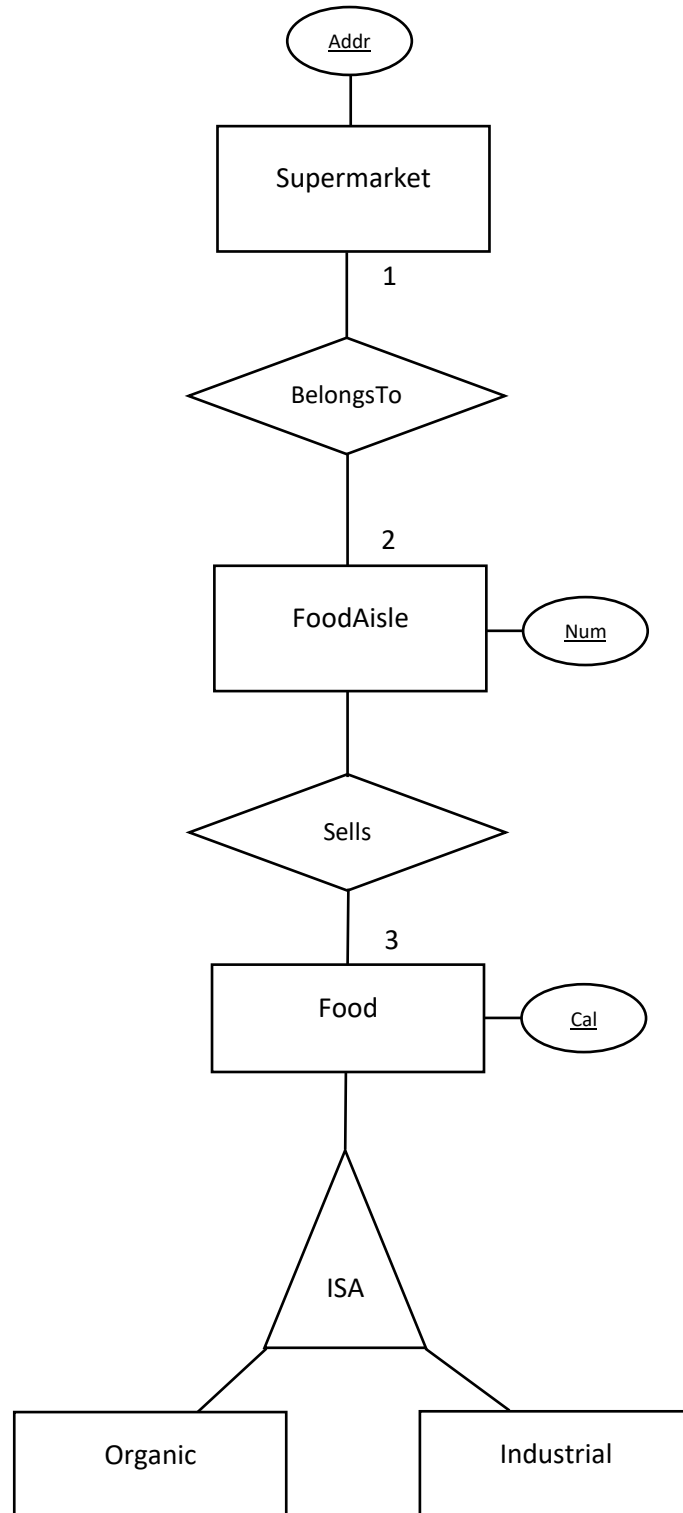
1. כתבו את התשובות אך ורק בטופס הבחינה ובמקום המיועד להן, מחברת הטייטה לא תיבדק.
2. כל חומר עזר הכתוב על נייר בלבד מותר בשימוש.
3. אין לקבל או להעביר חומר כלשהו בזמן הבחינה.
4. יש להשתמש רק בסימנים או פונקציות שנלמדו בתרגול או בהרצאה והמופיעים בשקפים של הקורס. כל שימוש בסימון שאינו כזה מחייב הסבר מלא של משמעות הסימון.
5. משך הבחינה הינו שלוש שעות, תכננו את הזמן בהתאם.
6. בבחינה 6 שאלות בשלושה חלקים. נא וודאו שיש בידכם את כל הטופס.
7. אין לכתוב בעפרון.

בהצלחה!

חלק 1 - תכן מסדי נתונים 35 נק'

שאלה 1 - ERD, 8 נק'

התבוננו בתרשים ה-ERD הבא, המגדיר מערכת מסד נתונים של סופרמרקט (המספרים הם לצורך סימון בלבד והינם חסרי משמעות בתרשים ה-ERD):



לפניכם דרישות חדשות המחייבות שינוי בתרשים, ולאחר כל דרישה הצעות לשינוי בתרשים שיביאו, אולי, לקיום הדרישה. עבור כל שינוי מוצע הקיפו בעיגול את האפשרות הנכונה מבין האפשרויות הנתונות.

1. **דרישה:** בכל סופרמרקט חייבת להיות לפחות שורת מדפי מזון (FoodAisle) אחת ולא משנה לנו אם תהיה אפשרות שיהיו מספר שורות מדפי מזון (FoodAisles) או לא.

פתרון מוצע: נשנה את היחס $BelongsTo$ ל $Unique$ לכיוון $Supermarket$, כלומר נהפוך את קצה הקו המסומן ב-(1) לראש חץ עגול.

הקיפו את האפשרות הנכונה:

- א. הפתרון אינו מקיים את הדרישה, עדיין ייתכנו סופרמרקטים ללא שורות מדפי מזון.
- ב. הפתרון יביא לכך שלכל סופרמרקט תהיה שורת מדפי מזון אחת בדיוק, ולכן יקיים את הדרישה.
- ג. הפתרון יביא לכך שלכל סופרמרקט תהיה שורת מדפי מזון אחת או יותר, ולכן יקיים את הדרישה.
- ד. אף תשובה אינה נכונה.

פתרון מוצע אחר: נשנה את היחס $BelongsTo$ ל $Unique$ לכיוון $FoodAisle$, כלומר נהפוך את קצה הקו המסומן ב-(2) לראש חץ עגול.

הקיפו את האפשרות הנכונה:

- א. הפתרון אינו מקיים את הדרישה, עדיין ייתכנו סופרמרקטים ללא שורות מדפי מזון.
- ב. הפתרון יביא לכך שלכל סופרמרקט תהיה שורת מדפי מזון אחת בדיוק, ולכן יקיים את הדרישה.
- ג. הפתרון יביא לכך שלכל סופרמרקט תהיה שורת מדפי מזון אחת או יותר, ולכן יקיים את הדרישה.
- ד. אף תשובה אינה נכונה.

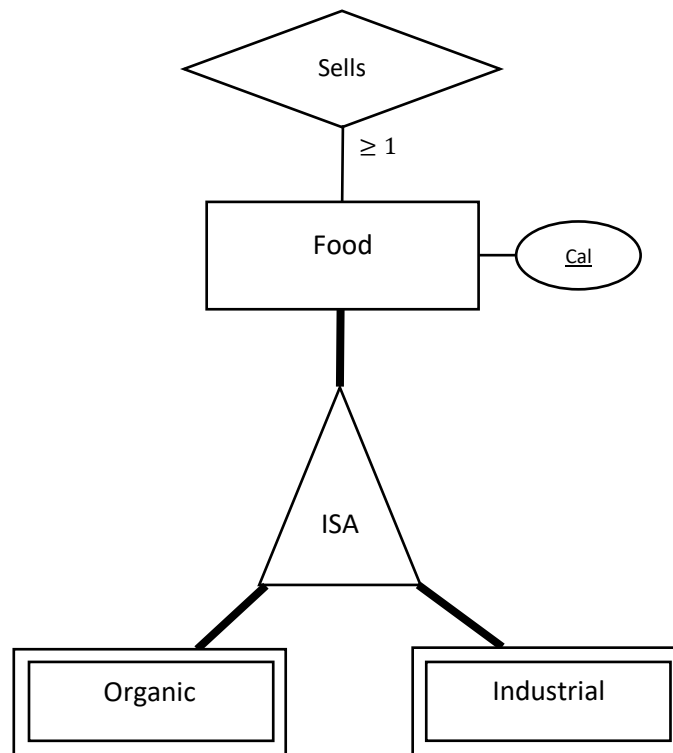
2. **דרישה:** בכל שורת מדפי מזון (FoodAisle) צריך להימכר (Sells) לפחות מוצר מזון אורגני (Organic) אחד או לפחות מוצר מזון תעשייתי (Industrial) אחד (ה"או" לא אקסלוסיבי, כלומר ייתכן שיימכרו גם מוצרי מזון אורגני וגם מוצרי מזון תעשייתי בשורה של מדפי מזון).

פתרון מוצע: נכתוב $1 \geq$ מעל הקו המסומן ב-(3).

הקיפו בעיגול את האפשרות הנכונה:

- א. הפתרון אינו מקיים את הדרישה, עדיין ייתכנו שורות מדפי מזון שאינן מוכרות לא מוצר אורגני ולא מוצר תעשייתי.
- ב. הפתרון מביא לכך שבכל שורת מדפי מזון יימכר גם מוצר אורגני אחד לפחות וגם מוצר תעשייתי אחד לפחות, ולכן מקיים את הדרישה.
- ג. הפתרון מקיים את הדרישה מבלי לדרוש שבכל שורת מדפי מזון יימכר גם מוצר אורגני אחד לפחות וגם מוצר תעשייתי אחד לפחות.
- ד. אף תשובה אינה נכונה.

פתרון מוצע אחר: נכתוב " ≥ 1 " מעל הקו המסומן ב-(3), ובנוסף נקיף במרובע כפול כל אחת מסוגי הישויות Industrial ו-Organic, באופן הבא ונדגיש את הקווים הנוגעים ב ISA (שאר התרשים, שאינו מוצג בציר, נשאר כפי שהוא):



הקיפו בעיגול את האפשרות הנכונה:

- א. הפתרון אינו מקיים את הדרישה, עדיין ייתכנו שורות מדפי מזון שאינן מוכרות לא מוצר אורגני ולא מוצר תעשייתי.
- ב. הפתרון מביא לכך שבכל שורת מדפי מזון יימכר גם מוצר אורגני אחד לפחות וגם מוצר תעשייתי אחד לפחות, ולכן מקיים את הדרישה.
- ג. הפתרון מקיים את הדרישה מבלי לדרוש שבכל שורת מדפי מזון יימכר גם מוצר אורגני אחד לפחות וגם מוצר תעשייתי אחד לפחות.
- ד. אף תשובה אינה נכונה.

שאלה 2 - פירוקים ותלויות פונקציונליות, 27 נק'

א. נתונה סכמת הרלציה $R = (A, B, C, D, E)$ וקבוצת התלויות פונקציונליות הבאה:

$$F = \{ACB \rightarrow D, D \rightarrow A, D \rightarrow B, E \rightarrow A, C \rightarrow B\}$$

1. (2 נק') מהם המפתחות של R תחת F?

2. (4 נק') מצא פירוק 3NF עבור R, F ופרט את שלבי ריצת האלגוריתם:

פירוק:

ריצת אלגוריתם:

לסעיפים הבאים אין קשר ישיר לסעיף הקודם.

ב. (4 נק') הוכח/הפרך: בהינתן סכמת רלציה R וקבוצת תלויות פונקציונליות F, עבור כל פירוק של R, F כך ש:

- a. אף סכמת רלציה בפירוק אינה R עצמה.
- b. עבור כל סכמת רלציה בפירוק R_i מתקיים כי $F[R_i]$ מכילה לפחות תלות לא טריוויאלית אחת.
- c. הפירוק משמר מידע.
- d. הפירוק משמר תלויות.

אז הפירוק הוא ל-3NF.

ג. (2 נק') השתמש בחוקי ארמסטרונג (ניתן להשתמש בהרחבות לחוקים - Provable Rules) על מנת להראות כי:

$$\{A \rightarrow B, AB \rightarrow CD\} \models A \rightarrow C$$

ד. (5 נק') מצאו סכמת רלציה R , קבוצה של תלויות פונקציונליות F ופירוק של R כך ש:

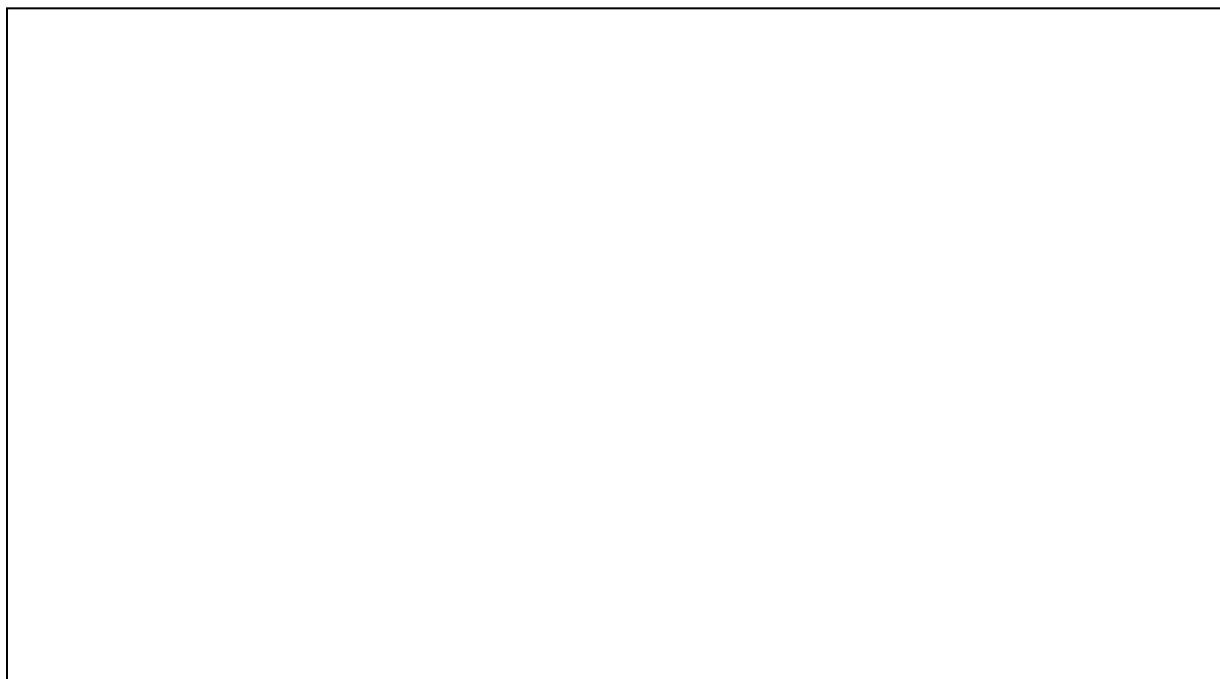
- a. F מהווה כיסוי מינימלי של עצמה.
- b. F מכילה לפחות תלות לא טריוויאלית אחת.
- c. הפירוק משמר תלויות.
- d. עבור כל תלות $X \rightarrow A$ ב- F , מתקיים כי XA לא מוכל באף סכמת רלציה בפירוק.

ה. (10 נק') נתונה סכמת הרלציה $R = (A, B, C, D, E, G)$ וקבוצת תלויות פונקציונליות F כלשהי המהווה כיסוי מינימלי של עצמה כך שעבור R, F מתקיים כי CD ו- AB הם המפתחות היחידים.

עבור הטענות הבאות כתבו האם הן חייבות להתקיים עבור F , לא ייתכן שיתקיימו עבור F , או עשויות (כלומר, ייתכן אך לא הכרחי) להתקיים עבור F ותנו הסבר או דוגמא משכנעים.

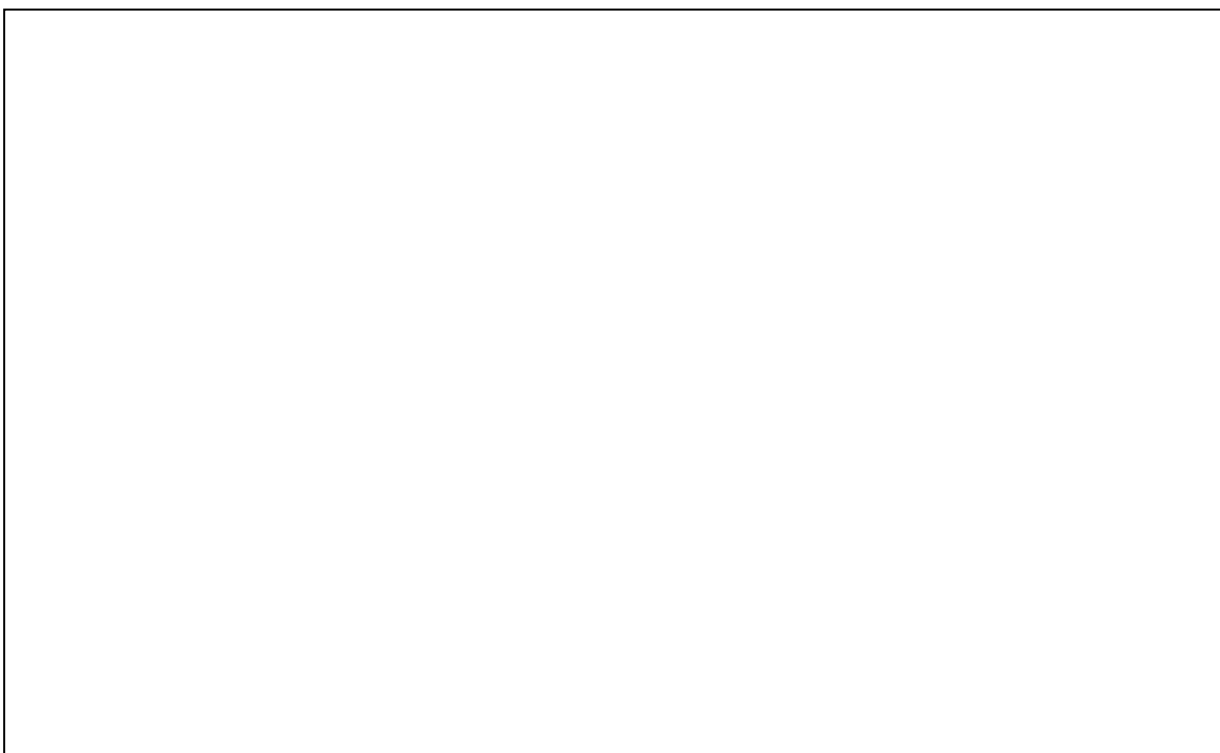
1. קיימת ב- F תלות המכילה את E בצד ימין שלה.

2. קיימת ב F תלות המכילה את E בצד שמאל שלה.



3. כל התלויות ב F מכילות אטריביוט אחד בדיוק בצד שמאל.

רמז: בסעיף זה ניתן להניח כי האטריביוטים EG מושמטים מ R כאשר יתר התנאים נשארים כפי שהיו.



חלק 2- שאלות מידע רלציוניות – 40 נק'

שאלה 3 - SQL, RA 21 נק'

קבוצת סטודנטים מהפקולטה למדעי המחשב החליטו להקים אתר אינטרנט בשם "Testock". האתר יכיל סריקות של מבחנים בקורסים שונים. כאשר לכל סריקה יהיה מוצמד הציון שלה והקורס אליו היא משויכת. סריקה מסוימת היא של בחינה אחת של סטודנט אחד. יכולות להופיע מספר סריקות שונות עבור אותו הקורס. לא ייתכן כי אותה בחינה תיסרק פעמיים. בנוסף, כל בחינה תופרד לשאלות שמהן היא מורכבת וכל שאלה בסריקה תישמר בנפרד.

במסד הנתונים קיימות הטבלאות הבאות:

Courses:

<u>cid</u>	Faculty
------------	---------

Scans:

<u>scanId</u>	cid	Grade
---------------	-----	-------

Questions:

<u>scanId</u>	<u>Quest_num</u>	Quest_grade
---------------	------------------	-------------

תוכן הטבלאות:

הטבלה Courses מחזיקה רשומות עם מספר קורס (cid), ושם הפקולטה אליה הקורס משויך (Faculty).

הטבלה Scans מחזיקה רשומות של סריקות. כל סריקה היא של בחינה כלשהי. הטבלה מכילה מזהה סריקה (scanId), מספר הזיהוי של הקורס אליו משויכת הבחינה בסריקה (cid) והציון (בין 0 ל-100) שקיבל הנבחן באותה בחינה שנסרקה (Grade).
בנוסף, cid הוא מפתח זר לטבלה Courses.

הטבלה Questions מחלקת כל סריקה לשאלות מהן הסריקה מורכבת ומחזיקה רשומות של שאלות השייכות לסריקות. הטבלה מכילה מספר הזיהוי של הסריקה (scanId) מספר השאלה (Quest_num) והניקוד (כלומר הציון) שאותה שאלה קיבלה בסריקה (Quest_grade).
המוסכמה: לכל סריקה, Quest_num הוא בסדר עולה מ-1 עד k כאשר k הוא מספר השאלות שהיו באותה בחינה. בנוסף, scanId הוא מפתח זר לטבלה Scans.

שימו לב כי מפתחות מסומנים בקו תחתון בטבלאות.

Courses:

<u>cid</u>	Faculty
236363	CS
236510	CS
044105	EE

Scans:

<u>scanId</u>	cid	Grade
1	236363	100
2	236363	90
3	044105	80

Questions:

<u>scanId</u>	<u>Quest_num</u>	Quest_grade
1	1	50
1	2	50
2	1	40
2	2	41
3	1	30
3	2	20
3	3	30

דודו, סטודנט למדעי המחשב בטכניון, השתמש באתר כדי להתכונן למבחן בממ"ן. כאשר פתר את אחד המועדים הקודמים באתר שם לב כי יש חוסר התאמה בין הציון שכתוב עבור סריקת המבחן בטבלה Scans ובין סכום הנקודות שקיבלו השאלות של אותה סריקה בטבלה Questions. לדוגמא: עבור סריקה בעלת מזהה scanId = 2 הופיע כי הציון שלה 90 בטבלה Scans ואילו בטבלה Questions סכום הניקוד של השאלות השייכות לאותה סריקה היה 81.

1. (4 נק') כתבו שאילתת SQL המחזירה את כל מזהי הסריקות (scanId) שהציון שלהן (Grade) בטבלה Scans שווה לסכום הנקודות שקיבלו השאלות השייכות לאותה סריקה בטבלה Questions.

מעתה ניתן להניח כי עבור כל סריקה הניקוד בטבלה Scans וסכימת הנקודות בטבלה Questions זהים.

2. (6 נק') כתבו שאילתת SQL המחזירה את כל הפקולטות (Faculty) כך שלכל קורס המשויך לפקולטה קיימת סריקה שקיבלה ציון 100, כלומר עבורה Grade=100 (שימו לב כי יתכנו מספר סריקות שונות לאותו הקורס). המשיכו את כתיבת השאילתה שתחילתה נתונה.

```
SELECT Faculty
FROM Courses AS Co
GROUP BY Faculty
HAVING
```

3. (6 נק')

א. כתבו שאילתה באלגברה רלציונית (RA) שתוצאתה תהיה מזהי הקורסים (cid) שלא קיימת סריקה השייכת לאותו קורס עבורה הציון 100.

ב. כתבו את השאילתה מסעיף 2, המחזירה את כל הפקולטות (Faculty) כך שלכל קורס המשויך לפקולטה קיימת סריקה שקיבלה ציון 100, כלומר עבורה Grade=100, באלגברה רלציונית (RA).

4. (5 נק') מה מחזירה שאילתת ה-RA הבאה? יש להקיף בעיגול את התשובה הנכונה.

$$\pi_{Faculty}(Courses \bowtie (\pi_{cid}(Scans) \setminus \pi_{cid}(\sigma_{grade < 90}(Scans))))$$

- א. את כל הפקולטות שלכל קורס השייך אליהן קיימת סריקה עם ציון גבוה או שווה ל-90.
- ב. את כל הפקולטות שקיים קורס השייך אליהן שקיימת לפחות סריקה אחת השייכת לאותו קורס עם ציון קטן מ-90.
- ג. את כל הפקולטות שקיים קורס השייך אליהן עבורו קיימת לפחות סריקה אחת עם ציון גבוה או שווה ל-90.
- ד. את כל הפקולטות שלכל קורס השייך אליהן כל הסריקות של מבחנים הן בעלות ציון גבוה או שווה ל-90.
- ה. את כל הפקולטות שקיים קורס השייך אליהן כך שכל הסריקות של מבחנים של אותו קורס הן בעלות ציון גבוה או שווה ל-90.
- ו. את כל הפקולטות שמופיעות בטבלת הקורסים כך שקיים קורס של אותה פקולטה בטבלת הקורסים.
- ז. אף תשובה מהנ"ל אינה נכונה.

שאלה 4 – Datalog, DRC , 19 נק'

נשתמש בסכמת מסד הנתונים של השאלה הקודמת. כעת נניח וקיימת בנוסף גם הטבלה הבאה:

Teaches:

<u>lecturer</u>	<u>cid</u>
-----------------	------------

כאשר lecturer מכיל את שם מרצה הקורס. Teaches הוא היחס בו המרצה ששמו ב- lecturer מלמד את הקורס בעל המזהה ב- cid. שימו לב שאת אותו הקורס (cid) ייתכן שמלמדים מספר מרצים.

ניתן להניח כי לא קיימים מרצים נוספים במערכת מלבד אלו שב Teaches.

א. (5 נק') כתבו תכנית DATALOG, מרובדת אם צריך, המגדירה את היחס $lecturerWithSacns(x)$ המכילה את שמות המרצים אשר עבור כל הקורסים אשר הם מלמדים קיימת סריקה כלשהי במערכת. הציגו ריבוד רלוונטי במידת הצורך.

ב. עבור סעיף זה נתון גם היחס הדו מקומי $R_<(a, b)$ אשר מתקיים עבור כל a, b אשר עבורם $a < b$ (כאשר a, b הם מחרוזות אז ההשוואה היא לקסיקוגרפית).

נתונה התוכנית הבאה

$$Friends(x, y) \leftarrow Teaches(x, c), Teaches(y, c), R_<(x, y)$$

$$Res(x, y) \leftarrow \neg Friends(x, y), Teaches(x, c_1), Teaches(y, c_2), R_<(x, y)$$

1. (2 נק') מה יהיה פלט התוכנית המרובדת עבור הפרדיקט Res ? בחר את התשובה הנכונה.

- א. זוגות של מרצים כך שהימני גדול מהשמאלי וקיים קורס ששני המרצים מלמדים אותו.
- ב. זוגות של מרצים כך שהימני קטן מהשמאלי וקיים קורס ששני המרצים מלמדים אותו.
- ג. זוגות של מרצים כך שהימני גדול מהשמאלי ולא קיים קורס ששני המרצים מלמדים אותו.
- ד. זוגות של מרצים כך שהימני קטן מהשמאלי ולא קיים קורס ששני המרצים מלמדים אותו.
- ה. אף תשובה אינה נכונה.

כעת, נתון ה EDB הבא:

Courses:

<u>cid</u>	Faculty
236363	CS
236501	CS

Teaches:

<u>lid</u>	<u>cid</u>
Oded	236363
Benny	236363
Shaul	236501

$R_<$: as defined, **Scans:** empty, **Questions:** empty

2. (3 נק') ציין 3 פתרונות אפשריים (מודלים מינימליים) של התוכנית הלא מרובדת.

ג. (4 נק') כתבו שאילתת DRC המחזירה סריקות של קורסים אותם מלמד מרצה אחד בדיוק וגם מרצה זה אינו "Benny".

ד. (5 נק') האם שאילתת ה-DRC הבאה היא Domain Independent? אם כן הסבר. אם לא, תן דוגמא קונקרטית שתשכנע בכך.

$$\{l, c: \left(\exists g, s \left((\neg \text{Scan}(s, c, g)) \rightarrow \text{Teaches}(l, c) \right) \right)\}$$

חלק 3- מודלים לא רלציוניים – 25 נק'

שאלה 5 XML, 17 נק'

נתון ה DTD הבא (כאשר erd הוא אלמנט השורש) בתוך מסמך שיכונה מעתה מסמך הדוגמה.

הניחו שההפרדה לשורות היא לצרכי תצוגה בלבד.

```
<?xml version="1.0" standalone="yes"?>
<!DOCTYPE erd [
    <!ELEMENT erd (entities, relationships)>
    <!ELEMENT entities (ent*)>
    <!ELEMENT relationships (rel*)>
        <!ELEMENT ent ( attr+)>
    <!ELEMENT rel ( participant+)>
    <!ELEMENT participant (#PCDATA)>
        <!ELEMENT attr (#PCDATA)>
    <!-- ATTLIST ent
        name ID #REQUIRED
    -->
    <!-- ATTLIST rel
        name CDATA #IMPLIED
        mult (mm | mo | om | oo | ref | weak) "mm"
    -->
    <!-- ATTLIST participant
        entID IDREF #REQUIRED
    -->
-->
]>
<erd>
    <entities>
        <ent name="person">
            <attr Name />
            <attr Height />
        </ent>
        <ent name="employer">
            <attr Company />
            <attr HQ />
            <attr Sector />
        </ent>
    </entities>
    <relationships >
        <rel name="EMPLOYMENT" mult= "mo" >
            <participant entID="person"></participant>
            <participant entID="employer"></participant>
        </rel>
    </relationships>
</erd>
```

א. (3 נקודות) עבור הביטוי

```
/**[name(.) = 'ent']
```

מה בדיוק יחזיר הביטוי על מסמך הדוגמה? ציין גם האם יוחזרו אלמנטים, אטריביוטים או משהו אחר?

ב. (4 נקודות) עבור הביטוי

```
/**/id(@*)/*[ /**/name(.) != 'attr']
```

מה בדיוק יחזיר הביטוי על מסמך הדוגמה? ציין גם האם יוחזרו אלמנטים, אטריביוטים או משהו אחר?

ג. (4 נקודות) עבור הביטוי

```
/**/id(@*)/*[name(.) != 'attr']
```

מה בדיוק יחזיר הביטוי על מסמך הדוגמה? ציין גם האם יוחזרו אלמנטים, אטריביוטים או משהו אחר?

ד. (4 נקודות) עבור הביטוי

`//ent/count(attr)`

מה בדיוק יחזיר הביטוי על מסמך הדוגמה? ציין גם האם יוחזרו אלמנטים, אטריביוטים או משהו אחר?

ה. (2 נקודות) נתונים שני ביטויים ב- XPATH:

a. `//rel/participant/@entID`

b. `//*[@entID [//participant/@entID]]`

האם לכל מסמך המספק את ה-DTD שני הביטויים הנ"ל מחזירים תשובות זהות?

סמן כן לא

נימוק קצר:

שאלה 6 Neo4j , 8 נק'

בעקבות המלצות הסטודנטים בקורס, האתר Maman Confession (שהופיע בשאלת ה-SQL במועד א') החליט לעבור משימוש במסד נתונים טבלאי (רלציוני) המתושאל באמצעות שאילתות SQL למסד נתונים גרפי (Neo4j) המתושאל באמצעות שאילתות בשפת Cypher.

האתר Maman Confession הינו אתר אליו יכולים להירשם משתמשים ולפרסם וידויים. בנוסף, משתמשים יכולים לסמן בלייק (כלומר לסמן שאהבו) וידויים שלהם ושל אחרים.

בגרף יש צמתים משני סוגים (השורה הראשונה ב- **bold** מייצגת את התגית (label) של הצומת ושאר השורות את השדות (attributes) של הצומת:

Confession
Cid
Content

User
Uid
Name
Faculty

עבור צומת עם label מסוג Confession מתקיים כי Cid מייצג את המזהה של הווידוי ו- Content מכיל את תוכן הווידוי.

עבור צומת עם label מסוג User מתקיים כי Uid מייצג את המזהה של ה- User (משתמש), Name מייצג את שם ה- User ו- Faculty מייצג את הפקולטה של ה- User.

הקשתות בגרף הן מהסוגים הבאים:

Wrote: מחבר בין User לבין Confession ומסמן שהמשתמש כתב את הווידוי.

Liked: מחבר בין User לבין Confession ומסמן שהמשתמש סימן בלייק שאהב את הווידוי.

א. (3 נק') כתבו שאילתת Cypher המחזירה זוגות צמתים של משתמשים a,b כך ש-a סימן בלייק וידוי שכתב b וכמו כן b סימן בלייק וידוי שכתב a.

ב. (5 נק') כתבו שאילתת Cypher המחזירה צמתים של משתמשים שלא כתבו אף וידוי וגם שסימנו בלייק את כל הווידויים בגרף המכילים את המחרוזת 'Troll'.

מקום לתשובות נוספות:

<u>חלק:</u>	<u>שאלה:</u>	<u>סעיף:</u>
<u>חלק:</u>	<u>שאלה:</u>	<u>סעיף:</u>
<u>חלק:</u>	<u>שאלה:</u>	<u>סעיף:</u>
<u>חלק:</u>	<u>שאלה:</u>	<u>סעיף:</u>