

מרצה : פרופ' יוהן מקובסקי
 מתרגלים : מר יבגני אברמוביץ'
 מר עדי עומרי

סמסטר אביב התשע"ב

מערכות מסדי נתונים
236363

מועד ב' (כ"ג בתשרי התשע"ג, 9 באוקטובר 2012)

מס' סטודנט:

פירוט השאלות והניקוד:

מס'	נושא	ניקוד
1	ERD	30
2	שאליות מידע	34
3	תלויות פונקציונליות	24
4	XML	15
סה"כ		103

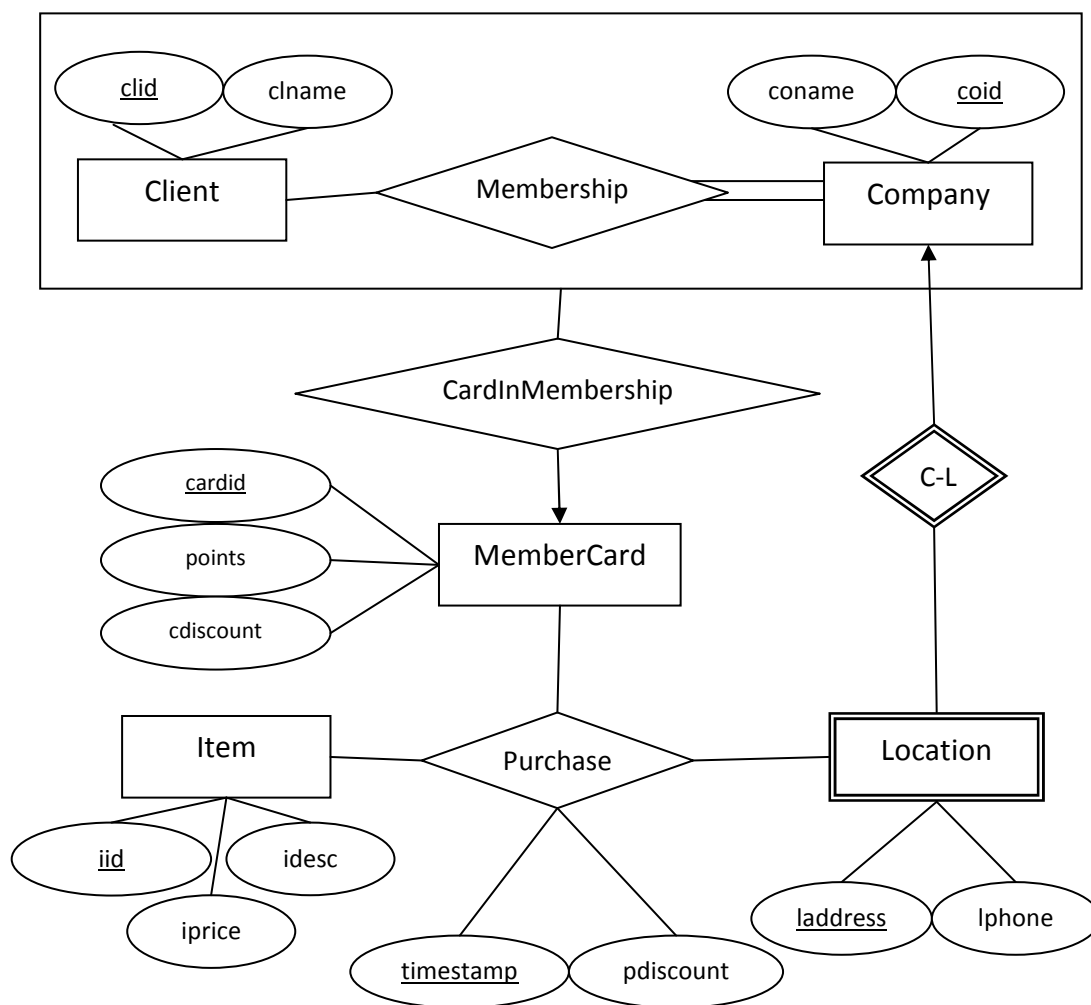
הנחיות לנבחנים

1. כתבו את התשובות אך ורק בטופס הבחינה, המחברת מיועדת לטיוטה בלבד.
2. מותר ומומלץ לכתוב את התשובות בעפרון.
3. בדף האחרון יש מקום נוסף לתשובות. אם צריך מקום נוסף לתשובות, השתמשו במקום זה תוך ציון הדבר ליד השאלה המקורית.
4. כל חומר עזר כתוב על נייר מותר בשימוש.
5. אין לקבל או להעביר חומר עזר כלשהו בזמן הבחינה.
6. בבחינה ארבע שאלות ללא בחירה. יש לענות עליהן במלואן.
7. בכל מקום שלא נאמר אחרת, יש לנמק את התשובות בקצרה. **תשובות לא מנומקות לא תתקבלנה, למעט במקומות שבהם אתם מתבקשים לכתוב שאלתה.**
8. יש להשתמש רק בסימנים או פונקציות שנלמדו בתרגול או בהרצאה או שמופיעות בשקפים של הקורס. כל שימוש בסימון שאינו כזה מחייב הסבר מלא של משמעות הסימון.
9. משך הבחינה שלוש שעות. תכננו את הזמן בהתאם. **לא תינתנה הארכות זמן במהלך המבחן.**
10. הבחינה (ללא דף הסריקה) כוללת 8 דפים (כולל דף זה), **בהם 15 עמודים**. נא לוודא שיש בידכם את כל הטופס.
11. הניקוד אינו נועד לשקף את קושי השאלה ולכן מומלץ לקרוא קודם את כל השאלות.

בהצלחה

שאלה 1 - ERD (30 נק')

נתונה דיאגרמה של מערכת מועדוני לקוחות של חברות.



שאלות:

א. (6 נק') הציגו את הטבלאות המתקבלות מתרגום ישיר של ה-ERD.

טבלה : Purchase
שדות : iid,cardid,laddress,coid,timestamp,pdiscount
מפתחות : { iid,cardid,laddress,coid,timestamp }
תלויות פונקציונליות : { iid,cardid,laddress,coid,timestamp }→pdiscount

טבלה : CardInMembership
שדות : clid,coid,cardid
מפתחות : { clid,coid }
תלויות פונקציונליות : { clid,coid }→cardid

ב. 12 נק') ענו על השאלות הבאות על ה-ERD. חובה לנמק (בקצרה)

i. 4 נק') האם ייתכנו שני סניפים (סניף=Location) של **חברות** (חברה=Company) **שונות** באותה הכתובת (address)?

כן. address לבדה אינה מפתח של location, לכן ייתכנו שתי רשומות עם אותו address ו-coid שונה.

ii. 4 נק') האם ייתכנו שני סניפים של **אותה חברה** באותה הכתובת (address)?

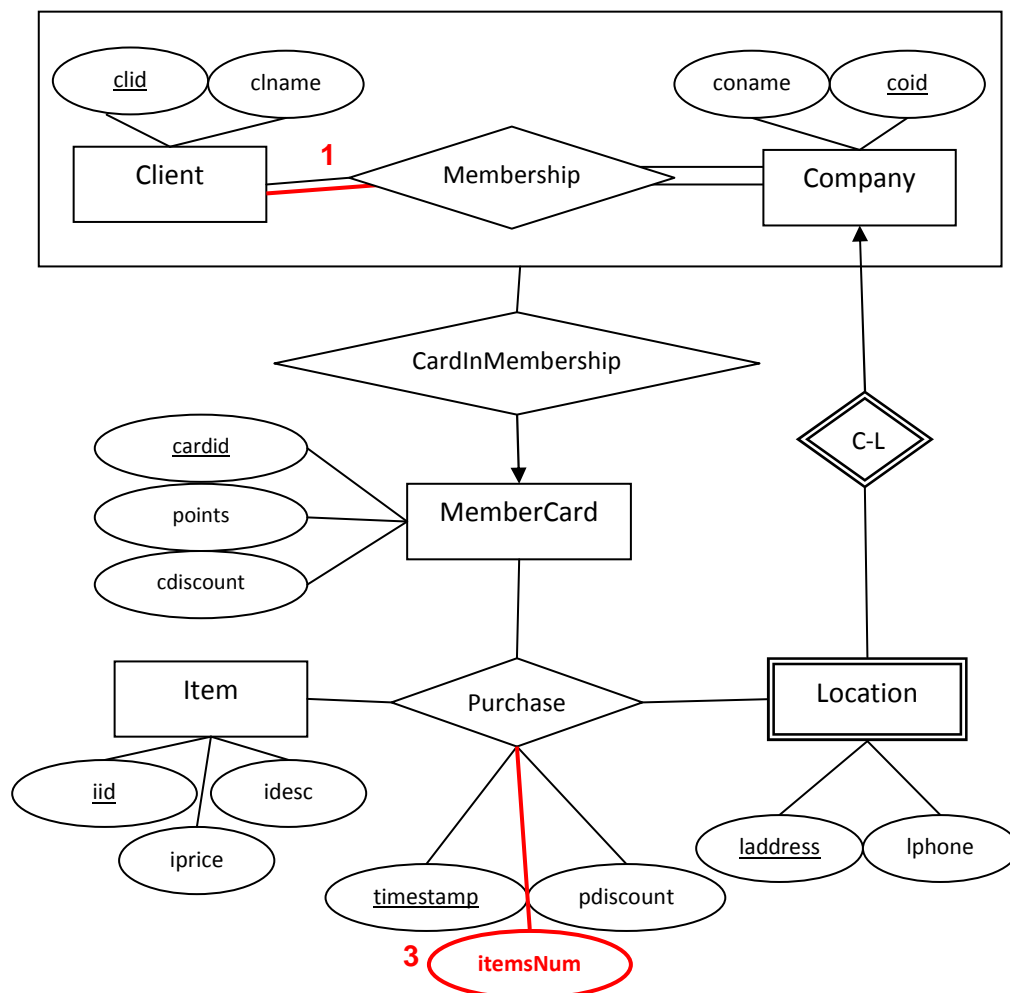
לא. אז יהיו שני ערכים ב-Location עם אותו זוג (coid,address) וזה לא ייתכן כי הזוג הוא מפתח.

iii. 4 נק') האם ניתן להימנע משימוש ביחסים טרינאריים (יחסים בין 3 ישויות) בדיאגרמה?

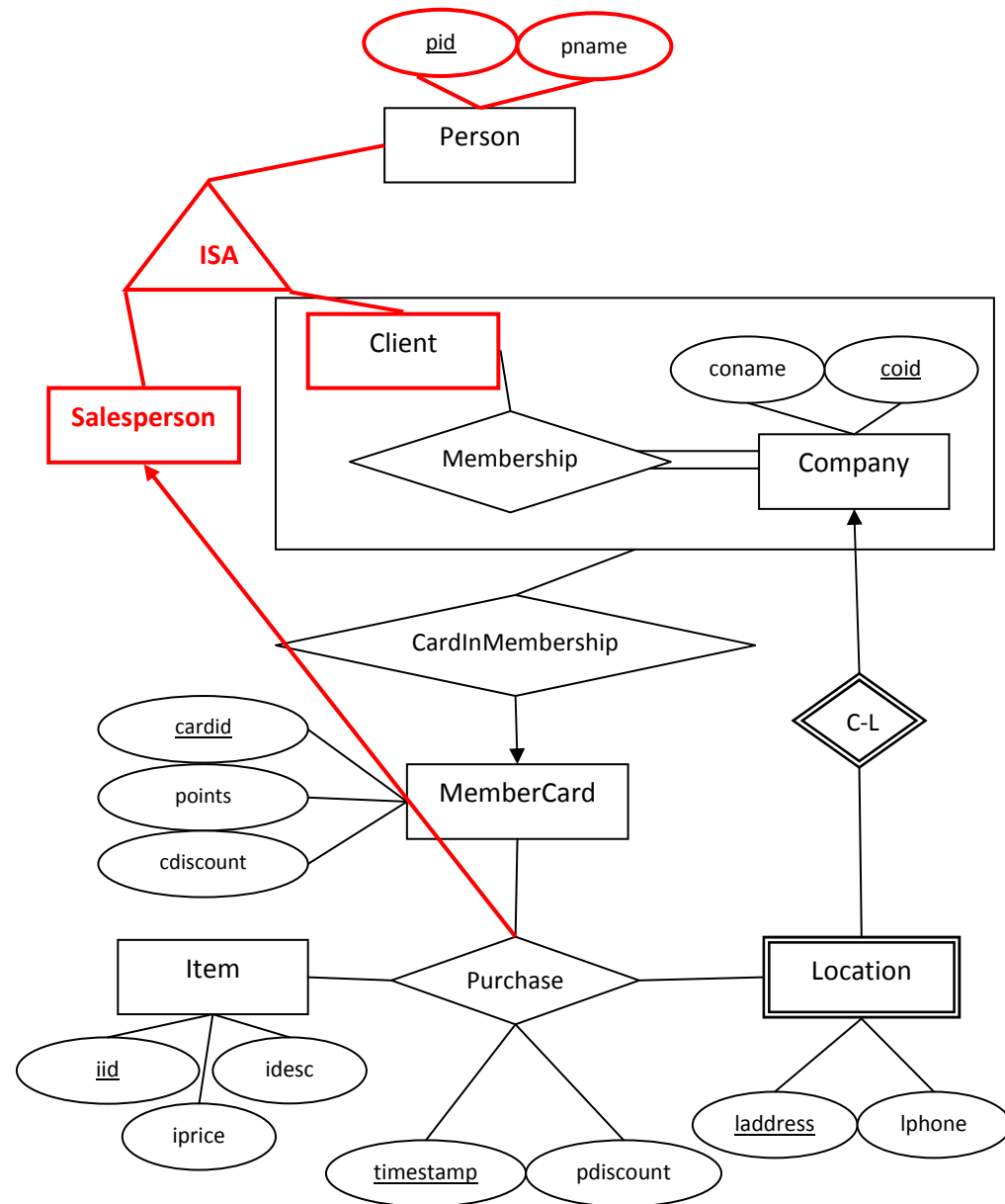
כן. את היחס Purchase (שהוא היחס הטרינארי היחיד) ניתן להמיר בשני יחסים, למשל יחס בין Item ל-Location, סביבו הקבצה ויחס נוסף בין הקבצה זו לבין MemberCard.

ג. (6 נק') ניתן העתק של הדיאגרמה. עבור כל אחד מהתנאים הבאים, סמנו בטבלה אם הוא מתקיים. אם לא, שנו את הדיאגרמה. לכל שינוי שאתם מבצעים בדיאגרמה, ציינו את התנאי שלקיומו השינוי נדרש.

תנאי	מתקיים (כן/לא)
1. לא ייתכן לקוח (Client) שאינו חבר במועדון לקוחות (לא יופיע ב-Membership).	לא
2. ייתכנו כמה לקוחות שיהיו חברים במועדון לקוחות של אותה חברה (Company) עם אותו כרטיס חבר (MemberCard).	כן
3. ייתכנו כמה פריטים (פריט=Item) ברכישה אחת (Purchase). (הבהרה בזמן המבחן: כמה פריטים מאותו סוג (אותו Item))	לא



ד. (6 נק') ניתן העתק של הדיאגרמה. שנו אותה כך שלכל רכישה (Purchase) יישמר מוכר יחיד (Salesperson). למוכר יישמרו כל הפרטים של לקוח (Client), אך הוא לא יוכל להיות חבר במועדון לקוחות של אף חברה (להופיע ב-Membership). אין להוסיף מגבלות מיותרות מעבר למה שצוין, ואין להוסיף את השינויים מסעיף ג'.



ניתן גם לחבר את Purchase ל-Salesperson ע"י הקבצה.

שאלה 2 – שאילתות מידע (34 נק')

נתונות הסכמות הבאות (כולל סדר המשתנים עבור DRC ודטלוג), אשר אינן בהכרח תואמות את ה-ERD מהשאלה הקודמת, ואינן בהכרח מהוות סכמות אופטימאליות עבור המסד:

$\text{client}(\text{cid}, \text{cname}, \text{caddress})$ – לקוח בעל מזהה cid (ייחודי), שם cname וכתובת caddress.

$\text{membercard}(\text{mcid}, \text{cid}, \text{companyid}, \text{mcpoints}, \text{mcdiscout})$ – כרטיס חבר בעל מזהה mcid (ייחודי) של לקוח בעל מזהה cid בחברה בעלת מזהה companyid ומספר נקודות mcpoints בכרטיס. הכרטיס מקנה הנחת כרטיס של mcdiscout.

$\text{purchase}(\text{pid}, \text{iid}, \text{mcid}, \text{pdiscout})$ – רכישה בעלת מזהה pid (ייחודי) של פריט iid עם הצגת כרטיס mcid בעל מזהה mcid. ברכישה ניתנה הנחת רכישה של pdiscout.

$\text{item}(\text{iid}, \text{iprice}, \text{idesc})$ – פריט בעל מזהה iid (ייחודי), מחיר iprice, ותיאור idesc.

א. (7 נק') כתבו תוכנית דטלוג עם שלילות המגדירה את הפרדיקט holdsInactiveCard(D) שערכו true אם ורק אם D הוא מזהה של לקוח שמחזיק בכרטיס חבר כך שלא בוצעה אף רכישה עם הצגת הכרטיס. שימוש לא בטוח בשלילה עלול להוביל לציון 0 בסעיף.

$\text{activeCard}(C) \leftarrow \text{purchase}(X, Y, C, Z).$
 $\text{holdsInactiveCard} \rightarrow \text{membercard}(C, D, X, Y, Z), \neg \text{activeCard}(C).$

ב. (6 נק') כתבו שאילתת RA שעבור תוכן מסד נתונים עם k רשומות ב-item, תחזיר מזהי כרטיסי חבר כך שכרטיס יוחזר אם ורק אם הוא הוצג ברכישה של $k-1$ פריטים שונים בדיוק. שימו לב, k אינו קבוע, השאילתה צריכה לעבוד עבור כל k !

$\text{notPurchased} = (\pi_{\text{iid}}(\text{item}) \times \pi_{\text{cardid}}(\text{purchase})) \setminus \pi_{\text{iid}, \text{cardid}}(\text{purchase})$
 $\pi_{\text{cardid}} ($
 $\pi_{\text{cardid}, \text{iid}}(\text{notPurchased}) \setminus$
 $\pi_{\text{cardid1}, \text{iid1}}(\sigma_{\text{cardid1}=\text{cardid2} \wedge \text{iid1} \neq \text{iid2}}(\text{notPurchased} \times \text{notPurchased}))$
 $)$

ג. SQL (8 נק') ענו על תתי הסעיפים הבאים. מותר להשתמש במבטים (view=מבט) **נוספים**.

i. (4 נק') צרו מבט purchaseprice(cid,pid,price) המחשב לכל לקוח ורכישה שבוצעה עם כרטיס שברשותו את המחיר ששולם ברכישה, כאשר בכל רכישה הופעלו שתי הנחות, הנחת הרכישה והנחת הכרטיס.

מחיר אחרי ההנחה מחושב כ-price*discount.

```
CREATE VIEW purchaseprice AS
SELECT membercard.cid, purchase.pid,
       item.price * purchase.pdiscount * membercard.discount AS price
FROM membercard,purchase,item
WHERE membercard.mcid = purchase.mcid AND
       purchase.iid = item.iid
```

ii. (4 נק') צרו מבט totalprice(cid,sum) המחשב את **סכום כל המחירים** שאדם שילם, בתנאי שמספר רכישותיו קטן מ-100. יש להחזיר את הסכום גם עבור אנשים שלא ביצעו אף רכישה.

```
SELECT client.cid, SUM (purchaseprice.price)
FROM client LEFT OUTER JOIN purchaseprice ON
       client.cid = purchaseprice.cid
GROUP BY client.cid
HAVING COUNT(purchaseprice.pid) < 100
```

ד. (5 נק') תרגמו בתרגום ישיר את השאילתה הבאה ל-DRC.
שאילתה שאינה תרגום ישיר לא תתקבל.

$$\pi_{\text{item1.iid}}(\sigma_{\text{item1.iprice} < \text{item2.iprice}}(\text{item} \times \text{item}))$$

$$\{x | \exists y \exists z \exists u \exists v \exists w (\text{item}(x, y, z) \wedge \text{item}(u, v, w) \wedge x < u)\}$$

ה. (8 נק') כתבו שתי דוגמאות לשאילתות DRC. הנכם רשאים להשתמש בנוסחאות האטומיות הבאות בלבד: $x=y$, $\text{item}(x, y, z)$. (כמובן עם שמות משתנים כרצונכם)

i. (4 נק') שאילתה **תלוית-תחום**, המחזירה תת קבוצה של item לכל תחום.

$$\{x, y, z | \text{item}(x, y, z) \wedge \neg \exists z' (\neg \text{item}(x, y, z'))\}$$

ניקה $\text{item} = \{(1, 1, 1)\}$, עבור $D = \{1\}$ התוצאה היא $\{1, 1, 1\}$ ואילו לכל תחום אחר (שבהכרח מכיל את 1) התוצאה ריקה. בכל מקרה תוצאת השאילתה היא תת קבוצה של item כנדרש.

ii. (4 נק') שאילתה שאינה **תלוית תחום**, אך גם אינה שאילתת טווח-בטוח (Safe Range). יש לכתוב את השאילתה ב-SRNF.

$$\{x, y, z | \exists y' (\neg \text{item}(x, y', z)) \wedge \text{item}(x, y, z) \wedge \neg \text{item}(x, y, z)\}$$

(התוצאה היא קבוצה ריקה לכל תחום, אך יש \perp בחישוב)

או, יותר פשוט:

$$\{x, y, z | \exists w (\text{item}(x, y, z))\}$$

גם כאן יש \perp בחישוב...

שאלה 3 - תלויות פונקציונאליות (24 נק')

בשאלה זו, גודל של רלציה $|r|$ מוגדר כמספר ה-n-יות כפול הגודל של כל n-יה ("מספר השדות בטבלה").

בהינתן קבוצת תלויות F, פירוק מצמצם נפח אחסון (storage-saving) הוא פירוק משמר מידע של סכמה U ל- U_1, U_2, \dots, U_n , כך שקיימת רלציה $r[U] \models F$ (תחת הסכמה U) עבורה $\sum_i |\pi_{U_i}(r)| < |r|$.

א. נתונה $U = (C, S, Z)$ וקבוצת התלויות $F = \{CS \rightarrow Z, Z \rightarrow C\}$.

i. (4 נק') הראו פירוק מצמצם נפח אחסון של הסכמה. אין צורך לנמק.

$\{\{Z,C\}, \{Z,S\}\}$

ii. (4 נק') עבור הפירוק שמצאתם, הראו רלציה r, המוכיחה שהפירוק הוא אכן מצמצם נפח אחסון. על r לקיים את אי השוויון $\sum_i |\pi_{U_i}(r)| < \frac{3}{4} |r|$.

Z	C	S
1	2	3
1	2	4
1	2	5
1	2	6
1	2	7
1	2	8
1	2	9
1	2	10
1	2	11

$$20 = \sum_i |\pi_{U_i}(r)| < \frac{3}{4} |r| = \frac{81}{4}$$

בשאלה זו, גודל של רלציה r מוגדר כמספר ה-n-יות כפול הגודל של כל n-יה ("מספר השדות בטבלה").

בהינתן קבוצת תלויות F , פירוק מצמצם נפח אחסון (storage-saving) הוא פירוק משמר מידע של סכמה U ל- U_1, U_2, \dots, U_n , כך שקיימת רלציה $r[U] \models F$ (תחת הסכמה U) עבורה $\sum_i |\pi_{U_i}(r)| < |r|$.

ב. (4 נק') נתונה $U = (C, S, Z)$ וקבוצת התלויות $F = \{CS \rightarrow Z\}$. הוכיחו כי לא קיים לה פירוק מצמצם נפח אחסון.

פירוק שמכיל את U לא מצמצם נפח אחסון

פירוק שלא מכיל את U בהכרח אינו משמר מידע, ולכן גם לא מצמצם נפח אחסון.

ג. נתונה הסכמה $U = (A, B, C, D)$ וקבוצת התלויות $F = \{A \rightarrow BC, A \rightarrow D\}$.

i. (4 נק') האם הסכמה ב-BCNF? הוכיחו.

כן. A הוא מפתח של הסכמה (מראים זאת למשל ע"י חישוב A_F^+) ושתי התלויות ב- F הינן תלויות במפתח זה.

ii. (4 נק') נתון הפירוק $\{U_1, U_2\}$, כאשר $U_1=(A,C), U_2=(A,B,D)$. האם הפירוק משמר מידע? הוכיחו.

כן. $F=U_1 \cap U_2 \rightarrow U_1=A \rightarrow C$, מראים זאת למשל ע"י חישוב A_F^+ .

iii. (4 נק') האם הפירוק מתת הסעיף הקודם (ii) מצמצם נפח אחסון? הוכיחו.

לא.

נשים לב שכל ערכי A הם שונים – כי אם יש שתי שורות t_1, t_2 כך ש- $t_1[A]=t_2[A]$, אז בהכרח אחת משאר התכונות שונה:

אם C שונה, אז $r \neq A \rightarrow C$ בסתירה לכך ש- A הוא מפתח. מאוד דומה לגבי שאר התכונות.

עבור n שורות ב- r יתקיים $|r|=4n$.

ב- $\pi_{U_1}(r)$ יש n שורות כי כל ה- A ים שונים, לכן $|\pi_{U_1}(r)|=2n$

ב- $\pi_{U_2}(r)$ יש n שורות כי כל ה- A ים שונים, לכן $|\pi_{U_2}(r)|=3n$

$\sum_i |\pi_{U_i}(r)|=5n \geq 4n=|r|$, כלומר הפירוק אינו מצמצם נפח אחסון.

שאלה 4 - XML (15 נק')

מסד נתונים של מערכת מועדוני לקוחות תקף עבור ה-DTD הבא :

```
<!ELEMENT db (client*,company*)>
<!ELEMENT client EMPTY>
<!ATTLIST client
    cid ID #REQUIRED>
<!ELEMENT company (member*)>
<!ATTLIST company
    name CDATA #REQUIRED>
>
<!ELEMENT member (discount+)>
<!ATTLIST member
    clientid IDREF #REQUIRED
    points CDATA #REQUIRED>
>
<!ELEMENT discount #PCDATA>
```

שאלות:

א. (4 נק') נתון שיש $k > 0$ אלמנטי discount במסמך XML התקף עבור ה-DTD.

i. (2 נק') תנו חסם עליון (אם קיים) ותחתון (אם קיים) למספר אלמנטי ה-member במסמך כתלות ב-k.

יש לפחות אלמנט member אחד ולכל היותר k כאלה.

ii. (2 נק') תנו חסם עליון (אם קיים) ותחתון (אם קיים) למספר אלמנטי ה-client במסמך כתלות ב-k.

חייב להיות לפחות client אחד (כי חייב להיות member אחד לפחות, ה-clientid שלו חייב להצביע לאיזשהו ID ו-ID יש רק ב-client)
החסם העליון אינו תלוי ב-k.

ב. (7 נק') כתבו שאילתת XPath 1.0 המחזירה את החברה שכל הלקוחות (client=חבר) שמקיימת את התנאי. (member=חבר) במועדון הלקוחות שלה. אם יש כמה כאלה, יש להחזיר את הראשונה במסמך.

הבהרה בזמן המבחן: אין להניח שלקוח לא יכול להופיע יותר מפעם אחת כ-member בחברה.

```

/db/company[count(id(member/@clientid)) = count(/db/client)]

[not(preceding-sibling::company

[count(id(member/@clientid)) = count(/db/client)])]

או:

/db/company[count(id(member/@clientid)) = count(/db/client)][1]

```

ג. (4 נק') כתבו תוכנית XQuery העוברת על קובץ discounts.xml המתאים ל-DTD הנ"ל ומחזירה לכל הנחה (לכל אחד מערכי הטקסט המופיעים בתוך איזשהו אלמנט discount) את החברות שבהן היא ניתנת.

על פלט התוכנית להיות תקף עבור ה-DTD:

```

<!ELEMENT discounts (discount*)>
<!ELEMENT discount (company*)>
<!ATTLIST discount
    description CDATA #REQUIRED
>
<!ELEMENT company EMPTY>
<!ATTLIST company
    name CDATA #REQUIRED
>

```

```

<discounts>

{for $d in distinct-values(doc("discounts.xml")/db/company/member/discount)

return

<discount description="{ $d }">

    {for $c in doc("discounts.xml")/db/company

        [member/discount/text() = $d]

        return <company name="{ $c/@name }">}}

</discount>

} </discounts>

```

מקום נוסף לתשובות

אם אתם משתמשים בדף זה, ציינו זאת ליד השאלה/השאלות המקוריות, וציינו כאן את מספרי השאלה/השאלות.

שאלה: _____ סעיף: _____

שאלה: _____ סעיף: _____

שאלה: _____ סעיף: _____

