

הטכניון - מכון טכנולוגי לישראל
הפקולטה למדעי המחשב



אביב תשס"ה
8 ביולי 2005

דר' אלדר פישר
מר סעאב מנסור
גברת לינה זריבץ'

מערכות מסדי נתונים מועד א'

הזמן: 3 שעות
במבחן זה 11 עמודים

שם פרטי: _____

שם משפחה: _____

מס' סטודנט: _____

פקולטה: _____

שאלה	נקודות	מתוך
שאלה 1 – ERD + ODL		20
שאלה 2 – SQL		18
שאלה 3 – RA + RC		18
שאלה 4 – Design		25
שאלה 5 – XML		19
סה"כ		100

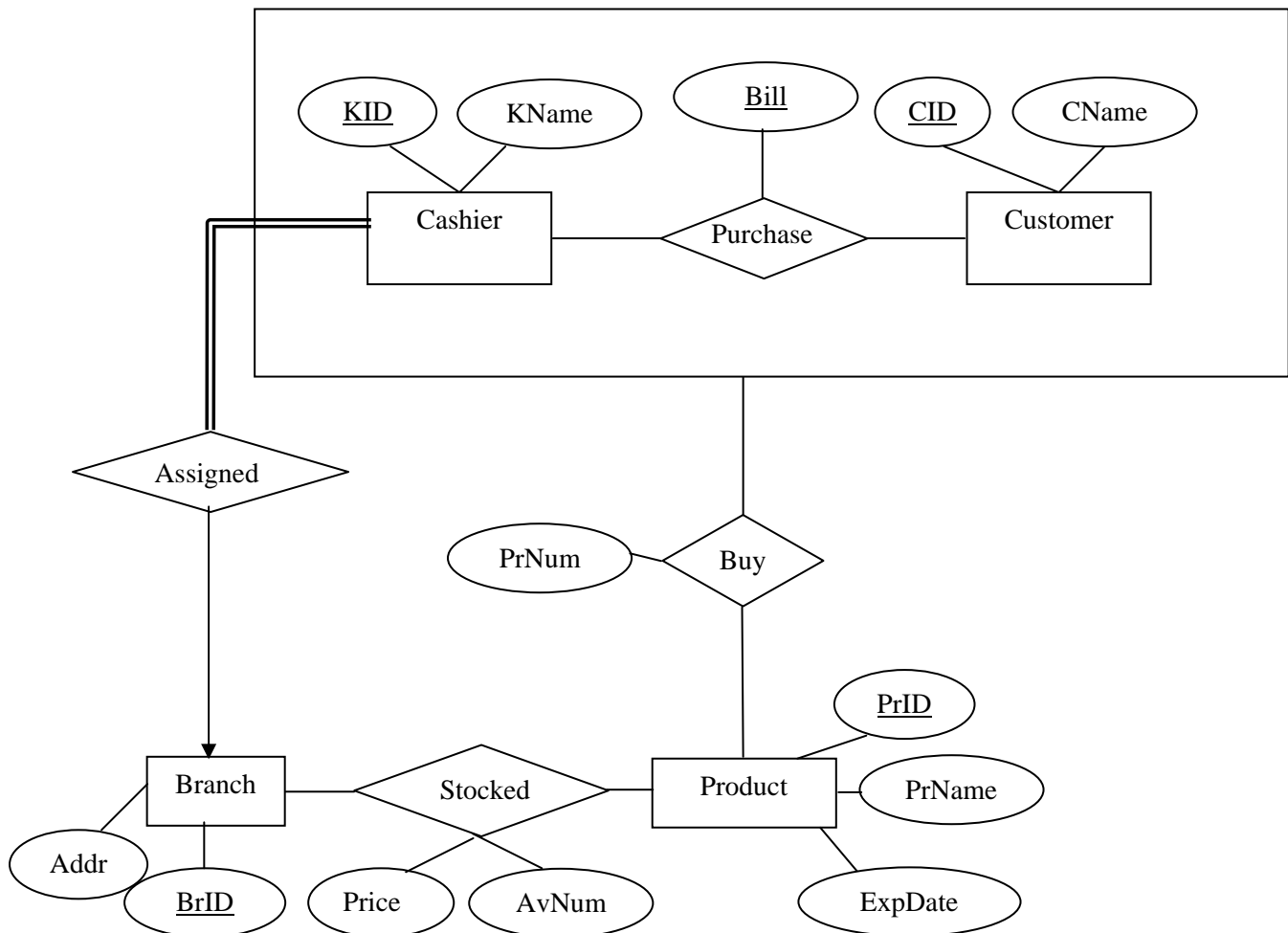
הערות:

- יש לענות על כל השאלות **בטופס הבחינה**.
- חומר עזר מותר: רק דברים שעשויים מנייר.
- אין להחזיק מכשיר אלקטרוני כלשהו.**
- קראו היטב את ההוראות שבתחילת כל שאלה ואת ההסברים לדיאגרמות.
- מומלץ שתכננו היטב את זמנכם, **לא תינתנה הארכות.**
- ערעורים יש להגיש תוך שבועיים ממועד פרסום התוצאות.
- לא יתקבלו ערעורים בנוסח "בדיקה מחמירה מדי".

בהצלחה

שאלה 1 – ERD + ODL (20 נק')

נתונה סכימת ה-ERD הבאה שמתארת מסד נתונים של רשת כל-בו:



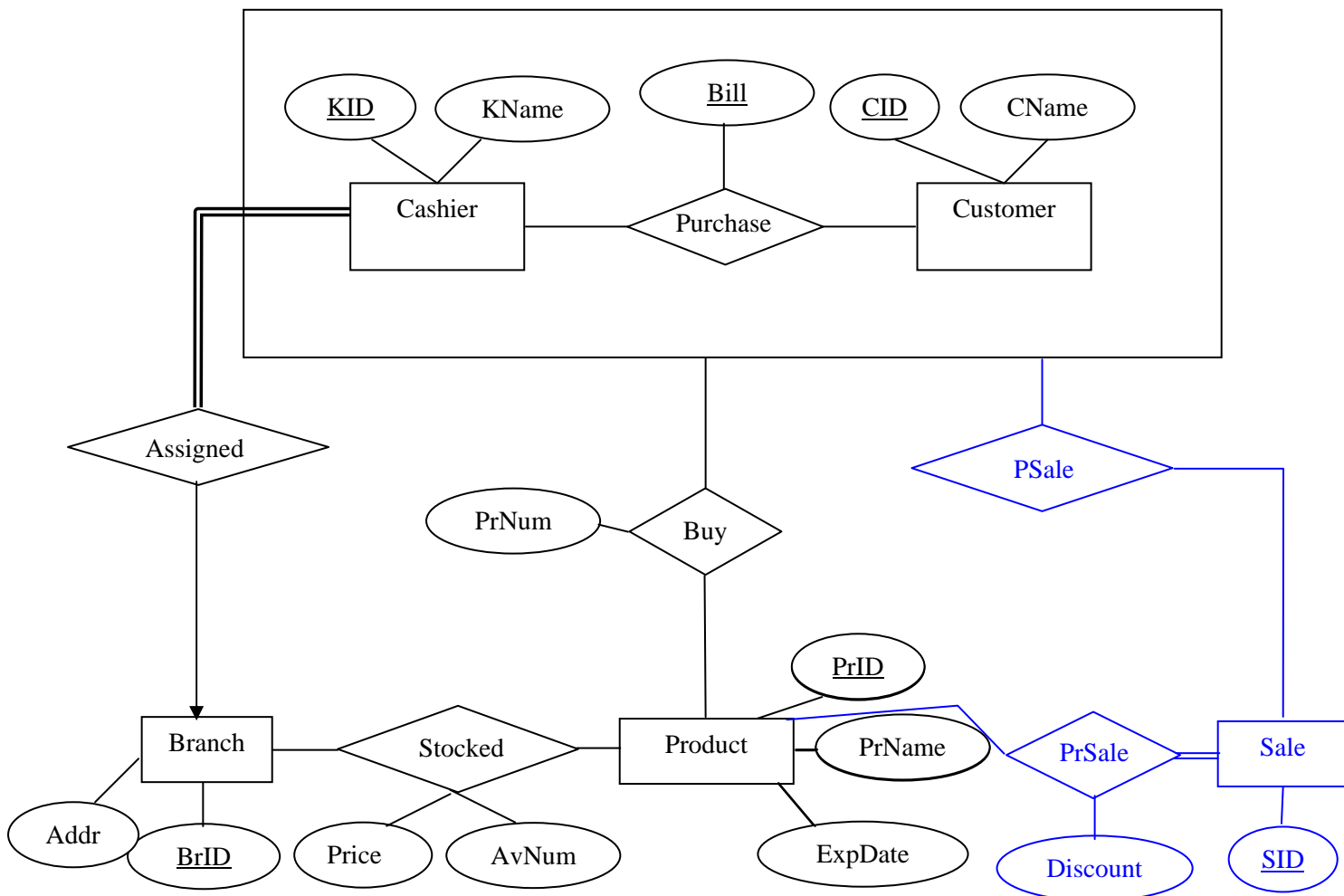
הסברים לדיאגרמה:

- Customer – מייצג לקוח של הכול-בו. לכל לקוח יש שם (CName) ומספר תעודת זהות (CID) ייחודי.
- Cashier – מייצג קופאיות שעובדות בכול-בו. לכל קופאית יש שם (KName) ומספר מזהה ייחודי (KID).
- Purchase – רכישה של לקוח אצל קופאית מסוימת. כל רכישה מאופיינת ע"י מספר חשבונית ייחודי (Bill).
- Product – ישות שמתארת מוצר. לכל מוצר יש שם שמתאר את המוצר (PrName), מספר מזהה ייחודי (PrID), והתאריך שבו פג התוקף של המוצר (ExpDate).
- Buy – יחס שמייצג קניות של מוצרים בכמות מסוימת (PrNum) שהתבצעו ברכישה (Purchase) בכול-בו.
- Branch – ישות המייצגת סניף של הכול-בו. כל סניף מאופיין ע"י מספר מזהה ייחודי (BrID) וכתובת (Addr).
- Stocked – יחס שמייצג קשר בין מוצר לסניף. הקשר בין מוצר לסניף מאופיין ע"י הכמות שבה זמין המוצר (AvNum) ואת מחירו (Price).
- Assigned – יחס שמייצג עבודה של קופאית בסניף מסוים. כל קופאית עובדת בסניף אחד בדיוק.

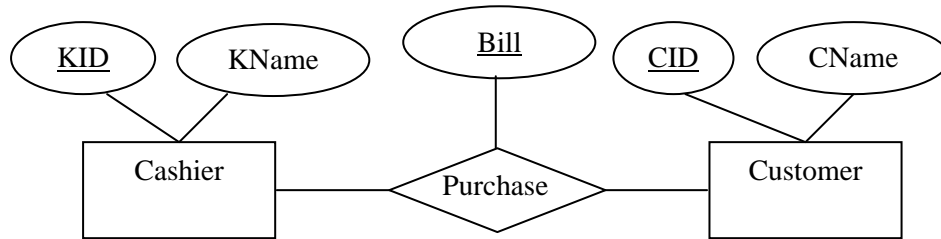
א. (10 נק') כל-בו "אבו מבצעים" החליט להשתמש בדיאגרמת ה-ERD הנ"ל, אך כיוון שזו רשת שידועה במבצעים מטורפים, הדיאגרמה אינה עונה על דרישות המערכת שלהם.

עדכנו את הדיאגרמה כך שתתמוך במוצרים במבצע. לכל מבצע יש קוד יחודי. כל מבצע מעניק הנחות (באחוזים) למוצר אחד או יותר (אותו מבצע יכול לתת אחוזי הנחה שונים למוצרים שונים). על כל קניה ניתן להפעיל מבצע או מספר מבצעים שיהיו תקפים עבור כל המוצרים הרלוונטיים שנקנו באותה קניה.

עדכנו את סכימת ה-ERD הבאה, כך שתעקוב גם אחר המבצעים והשימוש בהם.



ב. (10 נק') ייצגו את חלק ה-ERD הבא ב-ODL, תוך כדי התחשבות גם בקשרים עם שאר הישויות בדיאגרמה המקורית (אך ללא התוספות של הסעיף הקודם). אפשר להשתמש בנתון הנוסף שלכל קניה יש מספר חשבונית ייחודי.



Interface Cashier (extent Cashiers key KID)

```
{
    attribute integer KID;
    attribute string KName;
    relationship Set<Purchase> commitPurch inverse Purchase::commitBy;
    relationship Branch Assigned inverse Branch::brCashiers;
}
```

Interface Purchase (extent Purchases key Bill)

```
{
    attribute integer Bill;
    relationship Cashier commitBy inverse Cashier::commitPurch;
    relationship Customer theCustomer inverse Customer::custPurchases;
    relationship Buy prBought inverse Buy::buyPurchase;
}
```

Interface Customer (extent Customers key CID)

```
{
    attribute integer CID;
    attribute string CName;
    relationship Set<Purchase> custPurchases inverse Purchase:: theCustomer;
}
```

שאלה 2 – SQL (18 נק')

הערה: בכל הסעיפים הבאים הניחו תרגום סטנדרטי של סכמת ה-ERD משאלה 1 לסכמות רלציוניות:
Cashier, Purchase, Customer, Product, Buy, Branch, Stocked, Assigned

א. (10 נק') נתונה הרלציה NewBuy[BrID,PrID,PrNum] שמייצגת קנייה שעוד לא בוצעה בסניף מסוים, ז"א שכל השורות מכילות את אותו מספר סניף BrID, וכל שורה תייצג מוצר PrID וכמות PrNum. אם NewBuy ניתנת למימוש (המוצרים נמצאים במלאי בכמות המתאימה) עדכנו (באמצעות הוראת SQL) את היחס Stocked כך שיתאים לתוכן המסד לאחר ביצוע הקנייה ב-NewBuy, אחרת על המסד להישאר ללא שינוי.

UPDATE Stocked

SET AvNum = (SELECT AvNum-PrNum FROM NewBuy NB

WHERE NB.PrID = Stocked.PrID AND NB.BrID = Stocked.BrID)

WHERE 0 <= ALL (SELECT AvNum-PrNum FROM NewBuy NB LEFT OUTER JOIN Stocked S

ON (S.BrID = NB.BrID AND S.PrID = NB.PrID))

AND (PrID,BrID) IN (SELECT PrID,BrID FROM NewBuy)

הנחה: כל מוצר מופיע פעם אחת לכל היותר ב-NewBuy

ב. (8 נק') כתבו במילים פשוטות מה מבצעת השאילתא הבאה:

```
SELECT KID, MAX(MPCK)
FROM ( SELECT KID, CID, MAX(SumPrNum) AS MPCK
      FROM (SELECT KID, CID, Bill, SUM(PrNum) AS SumPrNum
            FROM Buy
            GROUP BY KID, CID, Bill ) SP
      GROUP BY KID, CID
      HAVING COUNT(*) >= ALL ( SELECT COUNT(*)
                              FROM Purchase
                              WHERE KID = SP.KID
                              GROUP BY KID, CID )
      ) Mystery
GROUP BY KID
```

Mystery מכילה לכל קופאית את כל הלקוחות אשר ביצעו אצלה את מספר הרכישות המקסימלי, ולכל לקוח

כנ"ל Mystery מחשבת את מספר המוצרים המקסימלי ברכישה אחת שהוא ביצע אצל קופאית זו.

השאילתא מחזירה לכל קופאית את מספר המוצרים המקסימלי שנקנו ברכישה אחת אשר התבצעה ע"י

אחד מהלקוחות עם מספר הרכישות המקסימלי אצלה.

שאלה 3 – RA + RC (18 נק')

א. (7 נק') השאלה מתייחסת לדיאגרמת ה-ERD משאלה 1.

להלן תרגומה החלקי לטבלאות:

Buy(CID, KID, Bill, PrID, PrNum)

Assigned(KID, BrID)

Stocked(PrID, BrID, Price, AvNum)

הערה: לצורך שאלה זאת ניתן להניח כי המפתח של רלצית Buy הוא Bill ביחד עם PrID.

נתונה שאילתת ה-DRC הבאה:

$\{ \langle \text{PrID} \rangle \mid \exists \text{cid, kid, b, pnum, bid, bid', pr, pr', av, av'} \\ (\text{Buy}(\text{cid, kid, b, PrID, pnum}) \wedge (\text{b} = 100) \wedge \\ \text{Assigned}(\text{kid, bid}) \wedge \text{Stocked}(\text{PrID, bid, pr, av}) \wedge \\ \text{Stocked}(\text{PrID, bid', pr', av'}) \wedge (\text{bid'} \neq \text{bid}) \wedge (\text{pr'} < \text{pr})) \}$

1. (3 נק') כתבו במילים פשוטות מה משמעות השאילתא.

השאילתא מחזירה את כל המוצרים השייכים לחשבונית מספר 100 שניתן היה לרכושם במחיר יותר נמוך

בסניף אחר השונה מסניף הרכישה.

2. (4 נק') תרגמו את השאילתא הנ"ל לשאילתא באלגברה רלציונית.

$\text{BillProd}(\text{BrID, PrID, Price}) = \pi_{\text{BrID, PrID, Price}} (\sigma_{\text{Bill}=100} (\text{Buy} \bowtie \text{Stocked} \bowtie \text{Assigned}))$

$\pi_{1\text{PrID}} (\sigma_{\text{PrID1} = \text{PrID2} \wedge \text{Price1} > \text{Price2} \wedge \text{BrID1} \neq \text{BrID2}} \text{BillProd} \times \text{Stocked})$

ב. (5 נק') נתונה רלציה r מעל הסכמה R[A, B, C].

כתבו שאילתא ב-TRC המחזירה תוצאה לא ריקה אם ורק אם התלות הרב ערכית $A \twoheadrightarrow B$ מתקיימת ב-r.

$\forall t, u (t \in r \wedge u \in r \wedge (t[A] = u[A]) \rightarrow \exists v (v \in r \wedge (v[A] = t[A]) \wedge (v[B] = t[B]) \wedge (u[C] = v[C])))$

ג. (6 נק') נתונה רלציה r מעל הסכמה R[A, B, C, D, E] ורלציה s מעל הסכמה S[D, E] המוגדרות מעל תחום אינסופי. הגודל של r אינו מוגבל.

עבור הסעיפים הבאים סמנו את התשובה הנכונה ביותר ונמקו. לדוגמא, אם בחרתם בתשובה א', יש להסביר גם מדוע תשובות ב' ו-ג' אינן נכונות, וכו'.

תשובות ללא נימוק יפסלו ולא יזוכו בנקודות!

1. (3 נק') נניח כי A הינו על מפתח של R וכן $|s| > 1$.
מהו המספר המקסימלי של רשומות המוחזרות ע"י הביטוי $r \div s$?

א – 0 רשומות ב – רשומה אחת ג – לא מוגבל
נימוק:

התשובה היא א'.

נניח כי $\langle d_1, e_1 \rangle$ ו- $\langle d_2, e_2 \rangle$ הן שתי שורות שונות ב- s . לא ייתכן כי ב- r קיימות השורות מהצורה

$\langle a, b, c, d_1, e_1 \rangle$ ו- $\langle a, b, c, d_2, e_2 \rangle$ (עבור כל הערכים האפשריים של a, b, c), כיוון שאז זו תהיה סתירה לכך ש- A הינו על מפתח.

2. (3 נק') נניח כי A הינו על מפתח של R וכן $|s| = 1$.
מהו המספר המקסימלי של רשומות המוחזרות ע"י הביטוי $r \div s$?

א – 0 רשומות ב – רשומה אחת ג – לא מוגבל
נימוק:

התשובה היא ג'.

הסבר א': כאשר s מכילה רק שורה אחת, הביטוי $r \div s$ שקול ל- r .
הסבר ב': נניח כי התחום הוא כל המספרים השלמים ו- s מכילה את השורה היחידה $\langle 0, 0 \rangle$. r מכילה שורות מהצורה $\langle i, i, 0, 0 \rangle$ עבור $1 \leq i \leq k$. הביטוי $r \div s$ יחזיר רלציה עם השורות $\langle i, i \rangle$ עבור $1 \leq i \leq k$.

שאלה 4 – Design (25 נק')

א. (18 נק') נתונה הסכמה $R[A,B,C,D,E]$ עם התלויות הפונקציונליות $F = \{AB \rightarrow C, AD \rightarrow CB, B \rightarrow D\}$.

1. (4 נק') כתבו כיסוי מינימלי של F .

$$\begin{aligned} \text{או } F_c &= \{AD \rightarrow C, B \rightarrow D, AD \rightarrow B\} \\ F_c &= \{AB \rightarrow C, B \rightarrow D, AD \rightarrow B\} \end{aligned}$$

2. (4 נק') מהם כל המפתחות הקבילים של R ?

ADE, ABE

3. (3 נק') האם R, F הם ב-3NF?

לא, למשל בשל התלות $AD \rightarrow C$.

4. (6 נק') האם קיים פירוק משמר מידע של R המכיל לא יותר משתי סכמות כך שכל הסכמות בפירוק ב-BCNF? אם כן, כתבו אותו. אם לא, נמקו!

לא קיים פירוק כזה.

הוכחה א' – הבעייתיות הינה באטריביוט E:

נתבונן בשני מקרים אפשריים:

- הסכמה שמכילה את האטריביוט E הינה בעלת קבוצת התלויות הפונקציונליות שאינה ריקה. במקרה זה הסכמה בוודאי לא תהיה ב-BCNF, כיוון ש- E לא נגרר ע"י אף תלות.
- הסכמה שמכילה את E הינה בעלת קבוצת תלויות ריקה. כיוון שהפירוק נדרש להיות משמר מידע האפשרויות היחידות לפירוק כזה הן $\{ADE, ABCD\}$ ו- $\{ABE, ABCD\}$. אך גם במקרים אלו הסכמה $ABCD$ אינה ב-BCNF.

הוכחה ב' – הבעייתיות הינה בתלות $B \rightarrow D$:

שני האטריביוטים D - B אינם יכולים להופיע בסכמה בעלת 3 אטריביוטים או יותר, כיוון שאז B לא יהיה על מפתח. לכן קיימות שתי אפשרויות:

- B ו- D יופיעו ביחד. הפירוק היחיד המשמר מידע הינו $\{BD, ABCE\}$, אך הסכמה $ABCE$ איננה ב-BCNF.
- B ו- D יופיעו לחוד, למשל B יופיע בסכמה אחת ו- D בסכמה השנייה. צריך לבחור את האטריביוטים שנמצאים בחיתוך – המועמדים הינם האיברים של קבוצת החזקה של $\{A, C, E\}$. אך נשים לב שאף מועמד אינו גורר את B ואינו גורר את D , לכן הפירוק בו B ו- D מופיעים לחוד בהכרח לא משמר מידע.

ב. (8 נק') נתונה סכמה R הנמצאת ב-3NF יחסית לקבוצת תלויות F אך לא ב-BCNF. הוכיחו כי ב-R יש לפחות שני מפתחות קבילים שונים.

לפי הנתון נובע, כי קיימת תלות $X \rightarrow A$ כך ש-X אינו על מפתח ו-A שייך למפתח קביל כלשהו (נסמנו K).

$K \cup X$ הינו על מפתח של R, כיוון ש- $X \rightarrow A$ ו- $K \rightarrow A$.

כל על מפתח מכיל איזשהו מפתח קביל ולכן גם $K \cup X$ מכיל מפתח קביל K' . בהכרח השונה מ-K כיוון

שהאטריביוט A שייך ל-K ואינו שייך ל- K' .

מש"ל.

שאלה 5 – XML (19 נק')

נתון ה-DTD הבא שמטרתו לתאר טקסט רב-לשוני:

```
<!ELEMENT text (#PCDATA | text)*>
<!ATTLIST text lang CDATA #REQUIRED>
```

קטע מסמך לדוגמא:

```
<text lang="en">
Both <text lang="de">du</text> and <text lang="de">Sie</text> can be translated to the English
word you, but use of <text lang="de">du</text> is considered disrespectful to strangers.
</text>
```

א. (5 נק') בניח עתה שאנו רוצים לשלב גם קטעים מודגשים במסמך, לפי הדוגמא הבאה:

```
<text lang="en">
Both <text lang="de">du</text> and <text lang="de">Sie</text> can be translated to the English
word <b lang="en">you</b>, but
<b lang="en">
use of <b lang="de">du</b> is considered disrespectful to strangers.
</b>
</text>
```

הכללים יהיו: לקטעים מודגשים גם יהיה ציון שפה. קטעים מודגשים יוכלו להכיל קטעים מודגשים אחרים (לשינוי שפה), קטעים רגילים יוכלו להכיל קטעים רגילים ומודגשים, אולם קטעים מודגשים לא יוכלו להכיל קטעים רגילים.

כתבו DTD אשר יהיה מתאים לקטעים רב לשוניים עם הדגשות כגון קטע הדוגמא.

```
<!ELEMENT text (#PCDATA | text | bold)*>
```

```
<!ATTLIST text lang CDATA #REQUIRED>
```

```
<!ELEMENT bold (#PCDATA | bold)*>
```

```
<!ATTLIST bold lang CDATA #REQUIRED>
```

ב. (7 נק') שאלה זו מתייחסת ל-DTD אשר הופיע לפני סעיף א'. כתבו שאילתת XPath אשר עבור מסמך טקסט רב-לשוני, מוצאת את כל הצמתים באנגלית (עם הערך en בתכונה lang), אשר הטקסט בתוכם (לא כולל צמתי אלמנט עם תויות text שהם בנים שלהם, יש לזכור כי צמתי הבנים עלולים להתייחס כבר לשפה אחרת מאנגלית) מכיל את המחרוזת tier.

```
//text[@lang="en"][contains(text(),"tier")]
```

ג. (7 נק') נתון המסמך relation.xml המתאר רלציה R[A,B,C] ע"פ הדוגמא הבאה:

```
<?xml version="1.0"?>
<rel>
  <tup><a>...</a><b>...</b><c>...</c></tup>
  <tup><a>...</a><b>...</b><c>...</c></tup>
  .....
</rel>
```

כתבו שאילתת XQuery אשר תחזיר סדרה ריקה אם הפירוק של R ל-S[A,B], T[B,C] הוא משמר מידע, ותחזיר סדרה לא ריקה אחרת.

```
for $t1 in doc("relation.xml")//tup
for $t2 in doc("relation.xml")//tup
where ($t1/b = $t2/b) and
  not(for $t3 in doc("relation.xml")//tup
    where ($t3/b= $t2/b) and ($t3/a=$t1/a) and ($t3/c= $t2/c)
    return $t3)
return ($t1,$t2)
```

הסבר: שלושת השורות הראשונות מבצעות מעין צירוף טבעי בין רלציות T ו-S. התנאי השני ב-where בודק האם כל שורה שנוצרת ע"י הצירוף הטבעי הנ"ל שייכת לרלציה המקורית כשמטרתו בעצם למצוא הפרות לשימור מידע. אם הפרה כזאת נמצאה אזי מוחזרות שתי הרשומות מתוך R האחראיות לכך ואם לא נמצאה אף הפרה - לא יוחזר דבר מהשאילתא.