

סמסטר חורף התשע"ג

מרצה : פרופ' יוהן מקובסקי

מתרגלים : מר עומר כ"ץ  
מר דניאל גרפונקל**מערכות מסדי נתונים  
236363**

מועד ב' (ל' בניסן התשע"ג, 10 באפריל 2013)

**מס' סטודנט:****פירוט השאלות והניקוד:**

מס'	נושא	ניקוד
1	ERD	27
2	שאליות מידע	36
3	פירוקים ותלויות פונקציונליות	27
4	XML	14
סה"כ		104

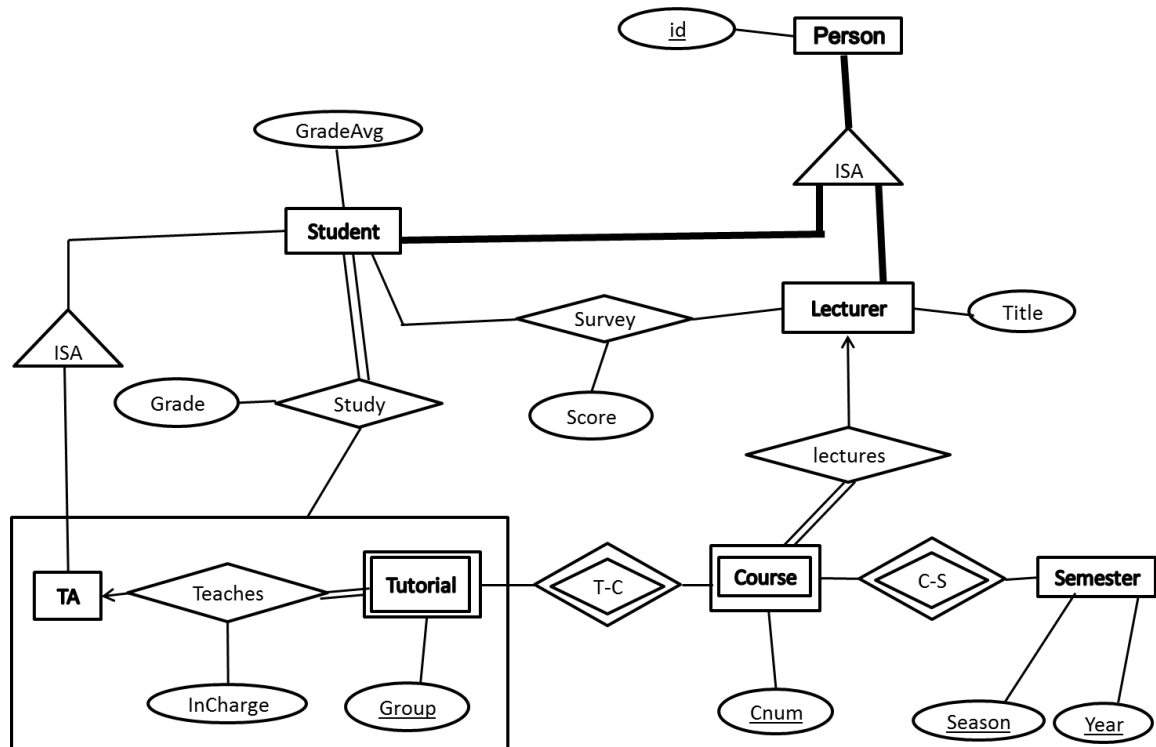
**הנחיות לנבחנים**

1. כתבו את התשובות אך ורק בטופס הבחינה, המחברת מיועדת לטייטה 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 שאלית.**
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 כוללת 8 דפים (כולל דף זה), **בהם 16 עמודים**. נא לוודא שיש בידכם את כל הטופס.
11. כאשר ניקוד תתי הסעיפים אינו מצוין, ניקוד הסעיף מתחלק שווה ביניהם.
12. הניקוד אינו נועד לשקף את קושי השאלה ולכן מומלץ לקרוא קודם את כל השאלות.

**בהצלחה**

## שאלה 1 – ERD (27 נק')

נתונה דיאגרמת ה-ERD הבאה :



דיאגרמת ה-ERD מתארת סטודנטים שלומדים קורסים שמתורגלים על ידי מתרגלים (שגם הם סטודנטים אך לא בהכרח למדו את הקורס בעצמם). כל קורס מועבר על ידי מרצה שמקבל ציון מהסטודנטים.

**שאלות:**

א. (10 נק') השלימו את הטבלאות הבאות המתקבלות מתרגום ישיר של ה-ERD.

שם הטבלה: <b>Study</b>
שדות: id, Group, Cnum, Season, Year, Grade
מפתחות: id, Group, Cnum, Season, Year
תלויות פונקציונליות: id, Group, Cnum, Season, Year -> Grade

שם הטבלה: <b>Survey</b>
שדות: StudentId, LecturerId, score
מפתחות: StudentId, LecturerId
תלויות פונקציונליות: StudentId, LecturerId -> score

ב. ענו על השאלות הבאות על פי דיאגרמת ה-ERD המקורית. חובה לנמק (בקצרה)

i. (3 נק') האם יתכן מתרגל (TA) שלא לומד (Study) אף קורס?

מכיוון שכל מתרגל הוא גם סטודנט וכל סטודנט חייב ללמוד לפחות קורס אחד, לא יתכן מתרגל שלא לומד.

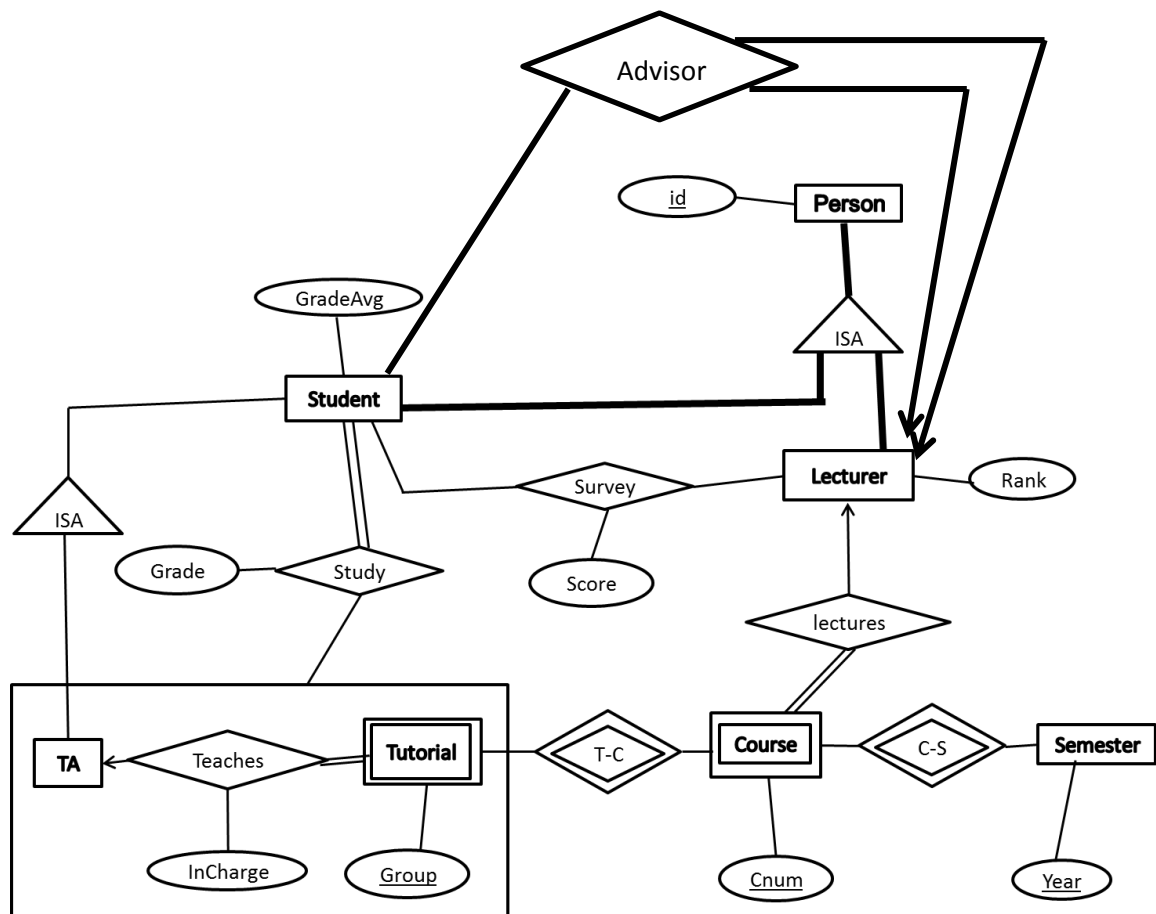
ii. (3 נק') כמה קורסים יכול מרצה (Lecturer) ללמוד (Study)?

אמנם המזהה id שמייצג גם את הסטודנטים וגם את המרצים נכלל ביחס study אך מכיוון שמזהי הסטודנטים ומזהי המרצים הם נפרדים (לא יתכנו מרצה וסטודנט בעלי אותו מזהה) ורק סטודנטים משתתפים ביחס study, מרצה לא יכול ללמוד קורסים, כלומר יכול ללמוד 0 קורסים.

iii. (3 נק') האם סטודנט (Student) יכול להירשם (Study) לקורס שאין לו תרגולים?

מכיוון שסטודנט נרשם לקבוצת תרגול לא יתכן שסטודנט ירשם לקורס שאין לו תרגולים.

ג. 8 נק') שנו את הדיאגרמה המקורית כך שלסטודנט (Student) יכול להיות מנחה (Advisor) מבין המרצים (Lecturer) ולכל סטודנט יש לכל היותר 2 מנחים.



## שאלה 2 – שאילות מידע (36 נק')

השאלות בסעיף זה מתייחסות לדיאגרמת ה-ERD שניתנה בשאלה 1 ולטבלאות שמהוות תרגום ישיר ממנה. ציינו את סדר העמודות בטבלאות בהן אתם משתמשים בתשובתיכם.

א. (6 נק') תרגמו את השאילתה הבאה לשפה טבעית:

```
SELECT l.id FROM Lecturer l, Survey s
WHERE l.id = s.lecturerId AND NOT EXIST (SELECT 1 FROM Lectures where lecturerId=l.id)
GROUP BY l.id
HAVING AVG(score)>=3
```

השאילתה מחזירה את מזהי כל המרצים שלא לימדו אף קורס וממוצע הציון שקיבלו במשאל הוא לפחות 3

ב. (6 נק') תרגמו את השאילתה הבאה לשפה טבעית:

$$\{Cnum \mid \exists year, season (Course(Cnum, year, season) \wedge \neg \exists group (Tutorial(group, Cnum, year, season)))\}$$

השאילתה מחזירה את מזהי כל הקורסים שאין להם קבוצת תרגול

ג. 6 נקי) שימו לב שהשאלתה הנתונה היא בSRNF. האם היא safe range? הוכיחו את תשובתכם.

נחשב את קבוצת המשתנים המתוחמים על פי השלבים שהוצגו בהרצאות והתרגולים:

$$rr(\text{Course}(\text{Cnum}, \text{year}, \text{season})) = \{\text{Cnum}, \text{year}, \text{season}\}$$

$$rr(\exists \text{year}, \text{season}(\text{Course}(\text{Cnum}, \text{year}, \text{season}))) = \{\text{Cnum}\}$$

$$rr(\text{Tutorial}(\text{group}, \text{Cnum}, \text{year}, \text{season})) = \{\text{group}, \text{Cnum}, \text{year}, \text{season}\}$$

$$rr(\exists \text{group}(\text{Tutorial}(\text{group}, \text{Cnum}, \text{year}, \text{season}))) = \{\text{Cnum}, \text{year}, \text{season}\}$$

$$rr(\neg \exists \text{group}(\text{Tutorial}(\text{group}, \text{Cnum}, \text{year}, \text{season}))) = \phi$$

$$rr(\psi) = rr((\exists \text{year}, \text{season}(\text{Course}(\text{Cnum}, \text{year}, \text{season})))$$

$$\wedge \neg \exists \text{group}(\text{Tutorial}(\text{group}, \text{Cnum}, \text{year}, \text{season})))) = \{\text{Cnum}\}$$

ממבנה השאלתה ניתן לראות שמתקיים  $free(\psi) = \{\text{Cnum}\}$  ולכן מתקיים  $rr(\psi) = free(\psi)$  ולכן השאלתה היא safe range.

ד. 6 נקי) כתבו שאילתת SQL המחזירה:

כל המתרגלים (TA) והקורסים (Course) שהם תרגלו (Teaches) לפחות 2 סמסטרים ואת הציון (Grade) שלהם כאשר הם למדו (Study) את הקורס (אם קיים ציון כזה, אחרת null)

```
SELECT t.id, t.Cnum, s.grade
FROM ( SELECT id, Cnum
      FROM ( SELECT DISTINCT id, Cnum, season, year FROM Teaches)
      GROUP BY id, Cnum
      HAVING COUNT(*) >= 2) t
LEFT OUTER JOIN Study ON (t.id=study.id AND t.Cnum=Study.Cnum);
```

ה. 6 נק') כתבו תכנית Datalog עם שלילות שמחזירה :

ממוצע הציונים (GradeAvg) של הסטודנטים (Student) שמתרגלים (Teaches) קורס (Course) שהם למדו (Stdudy) ולא תרגלו את עצמם או אף אחד שתרגל אותם או שתרגל מישהו... שתרגל מישהו שתרגל אותם. (יש להחזיר את הממוצע של כל אחד מהסטודנטים, לא לחשב ממוצע של כולם)

```
transitiveTeaches(x, y)
← Teaches(x, group, Cnum, season, year), Study(y, group, Cnum, season, year, _)

transitiveTeaches(x, y) ← transitiveTeaches(x, z), transitiveTeaches(z, y)

result(avg) ← student(id, avg), TA(id, avg), Teaches(id, _ Cnum, _ _),
Study(id, _ Cnum, _ _ _), ¬transitiveTeaches(id, id)
```

ו. 6 נק') כתבו שאילתת RA שמחזירה :

כל המרצים (Lecturer) שלימדו קורס (Course) בו כל הסטודנטים (Student) קיבלו ציון (Grade) מעל 80.

```
BelowEighty :=  $\pi_{cNum}(\sigma_{grade \leq 80}(Study))$ 
AllAboveEighty :=  $\pi_{cNum}(Study) \setminus BelowEighty$ 
Result :=  $\pi_{id}(Lectures \times AllAboveEighty)$ 
```

התקבל גם פתרון שמפריד בין קורסים על פי הסמסטר:

```
BelowEighty :=  $\pi_{cNum, seson, year}(\sigma_{grade \leq 80}(Study))$ 
AllAboveEighty :=  $\pi_{cNum, season, year}(Study) \setminus BelowEighty$ 
Result :=  $\pi_{id}(Lectures \times AllAboveEighty)$ 
```

### שאלה 3 – פירוקים ותלויות פונקציונליות (27 נק')

נתונה רלציה  $R[U]$  וקבוצת תלויות  $F$  כך שהרלציה היא ב-BCNF. נבחר אטריביוט  $A \notin U$  כן ש  $A \notin U$ . נרחיב את  $R[U]$  ל- $S[AU]$ .

א. (5 נק') נרחיב את קבוצת התלויות  $F$  ל- $F_1 = F \cup \{A \rightarrow U\} \cup \{K \rightarrow A \mid K \text{ is superkey of } R\}$ . הוכיחו ש- $S[AU]$  היא ב-BCNF עבור  $F_1$ .

על מנת שסכמה תהיה ב-BCNF צריך להתקיים שכל אחת מהתלויות בה היא או טריוויאלית או שצד שמאל שלה מהווה מפתח של הסכמה.

מכיוון שהוספנו את התלות  $A \rightarrow U$  מתקיים ש  $A$  הוא מפתח ולכן תלות חדשה זו היא בסדר.

לכל תלות  $K \rightarrow A$  מתקיים  $K \rightarrow U$  מכיוון ש  $A$  היה מפתח ב- $U$  המקורית ולכן לאחר הוספת התלות החדשה  $K$  הוא עדיין מפתח ולכן התלויות הללו בסדר.

לכל תלות ב- $F$  המקורית, אם התלות היא טריוויאלית אז היא עדיין טריוויאלית, אחרת צד שמאל של התלות הוא מפתח ולפי התלות החדשה  $K \rightarrow A$  צד שמאל נשאר מפתח ולכן גם תלויות אלה בסדר.

כל התלויות ב- $F_1$  מקיימות את התנאים עובר ב-BCNF ולכן הסכמה היא ב-BCNF.

ב. (5 נק') נרחיב כעת את קבוצת התלויות  $F$  ל- $F_2 = F \cup \{A \rightarrow U\}$ . הוכיחו או הפריכו: האם  $S[AU]$  היא ב-BCNF עבור  $F_2$ .

תהי תלות  $x \rightarrow y \in F$ .

מכיוון ש  $A \notin U$  מתקיים גם  $x \notin A$  ו  $y \notin A$ .

האטריביוט החדש  $A$  לא מופיע בצד ימין של אף אחת מהתלויות הקיימות או התלות החדשה (מכיוון ש  $A \notin U$ ). כתוצאה מכך ומכיוון ש  $x \notin A$  נקבל ש  $x$  אינו מפתח ולכן תלות זו לא מקיימת את תנאי ב-BCNF ולכן הסכמה אינה ב-BCNF.

ג. (5 נק') האם וכיצד תשובתכם לסעיף ב' הייתה משתנה אם  $R[U]$  לא הייתה ב-BCNF עבור  $F$ . הוכיחו או הפריכו בתנאים החדשים: האם  $S[AU]$  היא ב-BCNF עבור  $F_2$ .

התשובה אינה משתנה.

אם  $R$  אינה ב-BCNF אזי קיימת תלות  $x \rightarrow y \in F$  שאינה תלות טריוויאלית ו  $x$  אינו מפתח של  $R$ .

כפי שהוסבר בסעיף ב' לאחר הוספת האטריביוט החדש והתלות החדשה לא מתקיים  $x \rightarrow A \in F_2$  ולכן  $x$  עדיין אינו מפתח (לא ניתן להסיק ממנו את  $A$  בנוסף לאטריביוט שלא היה ניתן להסיק ממנו ב- $F$  המקורית).

מכאן שהתלות  $x \rightarrow y$  אינה מקיימת את תנאי ב-BCNF ולכן הסכמה אינה ב-BCNF.



נתונה הסכמה  $R[NCPA]$  (N = Name, C = City, P = Phone, A = AreaCode) ונתונה קבוצת התלויות הפונקציונליות

$$F = \left\{ \begin{array}{l} NC \rightarrow PA \\ C \rightarrow A \\ A \rightarrow C \end{array} \right\}$$

ד. (3 נק') מצאו את כל המפתחות של  $R[NCPA]$  על פי  $F$ .

המפתחות של  $R$  על פי  $F$  הם:  $\{NC, NA, NCA, NCP, NAP, NCPA\}$

ה. (3 נק') הראו ש  $R[NCPA]$  אינה ב  $BCNF$  עבור  $F$ .

מכיוון שלפי הסעיף הקודם לא מתקיים  $C$  הוא מפתח ולא מתקיים ש  $A$  הוא מפתח אנו מקבלים שהתלויות  $C \rightarrow A$  ו  $A \rightarrow C$  אינן מקיימות את תנאי  $BCNF$  ולכן הסכמה אינה ב  $BCNF$ .

ו. (4 נק') מצאו פירוק לBCNF עבור  $R[NCPA]$  ו  $F$ .

נפרק לפי האלגוריתם הראשון שהוצג בתרגול.

לפי הסעיף הקודם הסכמה R המקורית אינה בBCNF ולכן תנאי העצירה אינו מתקיים.

התלות  $C \rightarrow A$  מפרה את תנאי BCNF ולכן נבחר לפרק על פיה.

מתקבל הפירוק  $\{R_1[NPC], R_2[CA]\}$ .

כעת כל אחת מתתי הסכמות היא בBCNF ולכן הפירוק כולו בBCNF.

(נדרש להראות שכל אחת מתתי הסכמות היא בBCNF, נראה זאת בסעיף c2)

ז. הראו שהפירוק שמצאתם הוא :

a. (2 נק') משמר מידע.

נשתמש במשפט: פירוק של סכמה  $R$  ל  $\{R_1, R_2\}$  הוא משמר מידע אם"ם מתקיים  
 $R_1 \cap R_2 \rightarrow R_2$  או  $R_1 \cap R_2 \rightarrow R_1$ .

$$R_1 \cap R_2 = \{NPC\} \cap \{CA\} = \{C\}$$

לפי התלות  $C \rightarrow A$  מתקיים  $R_2 = \{CA\} = C^+$  ולכן  $C \rightarrow R_2$  ולכן תנאי המשפט מתקיימים  
ומכאן שהפירוק משמר מידע.

b. (2 נק') משמר תלויות.

התלויות  $C \rightarrow A$  ו  $A \rightarrow C$  מוכלות ב  $R_2$  ולכן בהכרח נשמרות.

נותר להראות שהתלות  $NC \rightarrow PA$  נשמרת.

על מנת שתלות זו תישמר צריך להתקיים  $NC \rightarrow PA \in (\pi_{R_1}F \cup \pi_{R_2}F)^+$ .

$$\pi_{R_1}F = \{NC \rightarrow P\}, \pi_{R_2}F = \{A \rightarrow C, C \rightarrow A\}$$

$$\pi_{R_1}F \cup \pi_{R_2}F = \{NC \rightarrow P, A \rightarrow C, C \rightarrow A\}$$

על מנת לבדוק האם מתקיים  $NC \rightarrow PA \in (\pi_{R_1}F \cup \pi_{R_2}F)^+$  נבדוק האם מתקיים

$$PA \in NC_{\pi_{R_1}F \cup \pi_{R_2}F}^+$$

$$NC_{\pi_{R_1}F \cup \pi_{R_2}F}^+ = \{NCPA\} \quad (\text{לפי } C \rightarrow A \text{ ו } NC \rightarrow P)$$

מכיוון שמתקיים  $PA \in NC_{\pi_{R_1}F \cup \pi_{R_2}F}^+$  התלות נשמרת ומכאן שכל התלויות נשמרות ולכן הפירוק משמר תלויות.

c. (2 נק')  $BCNF$  (ללא התבססות על האלגוריתם ליצירת הפירוק).

כפי שהראינו בסעיף הקודם:

$$\pi_{R_1}F = \{NC \rightarrow P\}$$

$$\pi_{R_2}F = \{A \rightarrow C, C \rightarrow A\}$$

נבחן את  $R_1$ :

$$NC_{\pi_{R_1}F}^+ = \{NCP\} = R_1$$

מכאן ש  $NC$  הוא מפתח של  $R_1$  ולכן  $R_1$  ב  $BCNF$ .

נבחן את  $R_2$ :

$$A_{\pi_{R_2}F}^+ = \{AC\} = R_2$$

$$C_{\pi_{R_2}F}^+ = \{AC\} = R_2$$

מכאן ש  $A$  הוא מפתח של  $R_2$  וגם  $C$  הוא מפתח של  $R_2$  ולכן  $R_2$  ב  $BCNF$ .

מכיוון שכל אחת מתת הסכמות היא ב  $BCNF$  הפירוק כולו הוא ב  $BCNF$ .

## שאלה 4 – XML (14 נק')

נתון DTD עבור מסד הנתונים של ספר מתכונים :

```
<!ELEMENT cookbook (recipe*)>
<!ELEMENT recipe (ingredient+, step*)>
<!-- ATTLIST recipe
      name ID # REQUIRED
-->
<!ELEMENT ingredient EMPTY>
<!-- ATTLIST ingredient
      ingredientIndex ID #REQUIRED
      name CDATA #REQUIRED
      amount CDATA #REQUIRED
-->
<!-- ELEMENT step (#PCDATA)
-->
<!-- ATTLIST step
      ingredientsUsed idrefs #IMPLIED
      optional (no | yes) "no"
-->
```

שאלות:

א. ענו על השאלות הבאות לגבי ה-DTD ונמקו בקצרה :

i. (1 נק') האם יתכן מתכון בלי שלבים?

מכיוון ששלבים במתכון מוגדרת באמצעות `step*`, יתכן מתכון ללא שלבים

ii. (2 נק') האם יתכן שמתכון יכיל שני מצרכים זהים (בעלי אותו שם)? אם כן, הסבר מדוע והאם ניתן לתקן זאת? אם לא, הסבר מדוע והאם ניתן לתקן זאת?

שם המצרך מוגדר כ-CDATA ולכן אין מניעה שיהיו שני מצרכים בעלי אותו שם, בפרט באותו מתכון.

ניתן לתקן זאת על ידי הפיכת שם המצרך ל-ID וכך לא יהיו שני מצרכים עם אותו שם, בפרט באותו מתכון.

iii. (1 נק') האם יתכן ספר מתכונים לא ריק שלא משתמש באף מצרך?

אם ספר המתכונים אינו ריק אזי יש בו לפחות מתכון אחד.  
כל מתכון מכיל לפחות מצרך אחד מכיוון שמצרכים במתכון מוגדרים באמצעות ingredient+.  
לכן לא יתכן ספר מתכונים לא ריק שלא משתמש באף מצרך.

iv. (1 נק') האם יתכן מתכון שמשמש במתכון אחר כמצרך?

המצרכים בהם משתמש מתכון מזוהים באמצעות השדה ingredientsUsed ועל פי id של המצרכים.  
מכיוון שגם למצרך וגם למתכון יש שדה מסוג id, יתכן שהשדה ingredientsUsed יפנה למתכון ומכאן שיתכן שמתכון ישתמש במתכון אחר כמצרך.

ב. (4 נק') כתבו שאילתת XPath1.0 המחזירה את כל המתכונים שכוללים שלבים לא אופציונליים שאינם משתמשים באף מצרך.

```
//recipe[step[@optimal="no" and count(id(@ingredientsUsed))=0]]
```

- ג. ענו על השאלה הבאה על פי ה DTD המקורי שניתן בתחילת השאלה ללא השינויים מסעיף א'.
- i. (3 נק') כתבו תוכנית XQuery המקבלת מסמך recipes.xml התקף עבור ה-DTD המקורי הנתון בתחילת השאלה ומדפיסה מסמך המפרט לכל מתכון את השלבים הדרושים להכנתו ותחת כל שלב את שמות המצרכים הדרושים לו. אלמנט השורש של המסמך החדש יקרא recipes. יש ליצור אלמנט עבור כל מתכון ולכלול באלמנט זה את שם המתכון.

```
<recipes>
{
  for $recipe in doc("recipes.xml")/cookbook/recipe
  let $steps := $recipe/step
  return
    <recipe name="{ $recipe/name }">
      {
        for $step in $steps
        return
          <step optional="{ $step/@optional }">
            {
              <content>{ $step/text() }</content>
              for $ingredient in id($step/@ingredientsUsed)
              return
                { $ ingredient }
            }
          </step>
        }
      </recipe>
}
</recipes>
```

- ii. (2 נק') כתבו DTD המתאים לפלט התוכנית המבוקשת.

```
<!ELEMENT recipes (recipe*)>
<!ELEMENT recipe (step*)>
<!ATTLIST recipe
  name ID # REQUIRED
>
<!ELEMENT step (content, ingredient*)>
<!ATTLIST step
  optional (no | yes) "no"
>
<!ELEMENT content (#PCDATA)>
<!ELEMENT ingredient EMPTY>
<!ATTLIST ingredient
  ingredientIndex ID #REQUIRED
  name CDATA #REQUIRED
  amount CDATA #REQUIRED
>
```

### מקום נוסף לתשובות

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

שאלה: \_\_\_\_\_ סעיף: \_\_\_\_\_

שאלה: \_\_\_\_\_ סעיף: \_\_\_\_\_

שאלה: \_\_\_\_\_  
סעיף: \_\_\_\_\_

