

סמסטר אביב התשע"ג

מרצח : פרופ' יוהן מקובסקי

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

מועד א' (ה' באב התשע"ג, 12 ביולי 2013)

מס' סטודנט:**פירוט השאלות והניקוד:**

מס'	נושא	ניקוד
1	ERD	26
2	שאליות מידע	31
3	פירוטים ותלויות פונקציונליות	27
4	XML	20
סה"כ		104

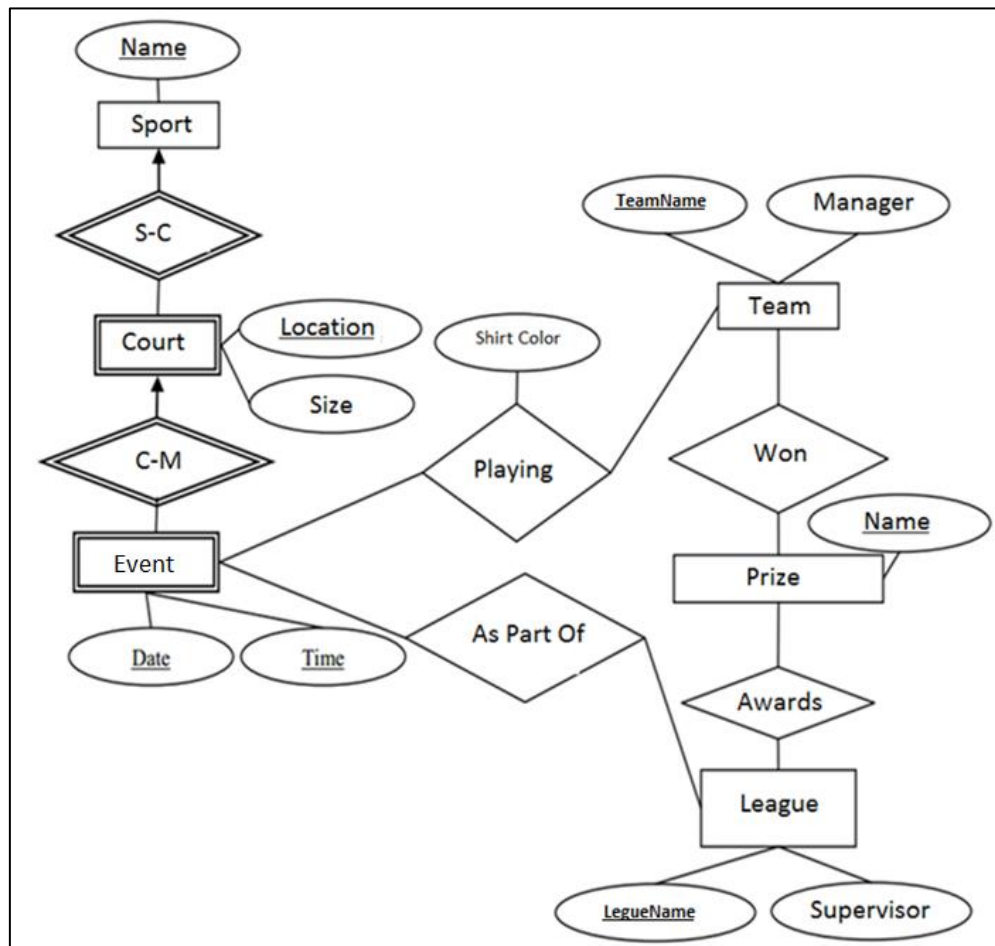
הנחיות לנבחנים

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 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 כוללת 9 דפים (כולל דף זה), **בהם 17 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 (26 נק')

נתונה דיאגרמת ה-ERD הבאה :



בדיאגרמה לפניכם מוצגת מערכת משחקי ספורט. טיפוס היישויות sport מייצג את ענפי הספורט השונים, Court את מגרשי הספורט, Event את משחקי הספורט (או אירועי הספורט). Event משחקות (Playing) קבוצות בצבע חולצה מסוים. משחק יכול להעשות במסגרת (As Part Of) ליגה מסוימת (League) ולכל ליגה יש מפקח. לכל קבוצה (Team) יש מנהל והיא יכולה לזכות בפרסים (Won, Prize) כאשר פרס יכול להיות משויך לליגה.

שאלות:

א. (8 נק') השלימו את הטבלאות הבאות המתקבלות מתרגום ישיר של ה-ERD.

שם הטבלה: Court
(1 נק') שדות: Name, Location, Size
(1 נק') מפתחות: Name, Location
(2 נק') תלויות פונקציונליות: $\{Name, Location\} \rightarrow Size$

שם הטבלה : Awards
(1 נק') שדות : Name, LeagueName
(1 נק') מפתחות : Name, LeagueName
(2 נק') תלויות פונקציונליות : אין

- ב. ענו על השאלות הבאות על פי דיאגרמת ה-ERD בלבד. חובה לנמק (בקצרה)
- i. (3 נק') האם ה-ERD מחייב שלא יהיו שתי קבוצות (Team) עם אותו צבע חולצה (Shirt Color) באותו משחק (Event)?

לא, המפתחות של Paying הם אלו שקובעים את צבע החולצה, ומזהה הקבוצה הוא אחד מהם. לכן עבור שתי קבוצות עם מזהים שונים יכול להיות אותו צבע חולצה באותו המשחק.

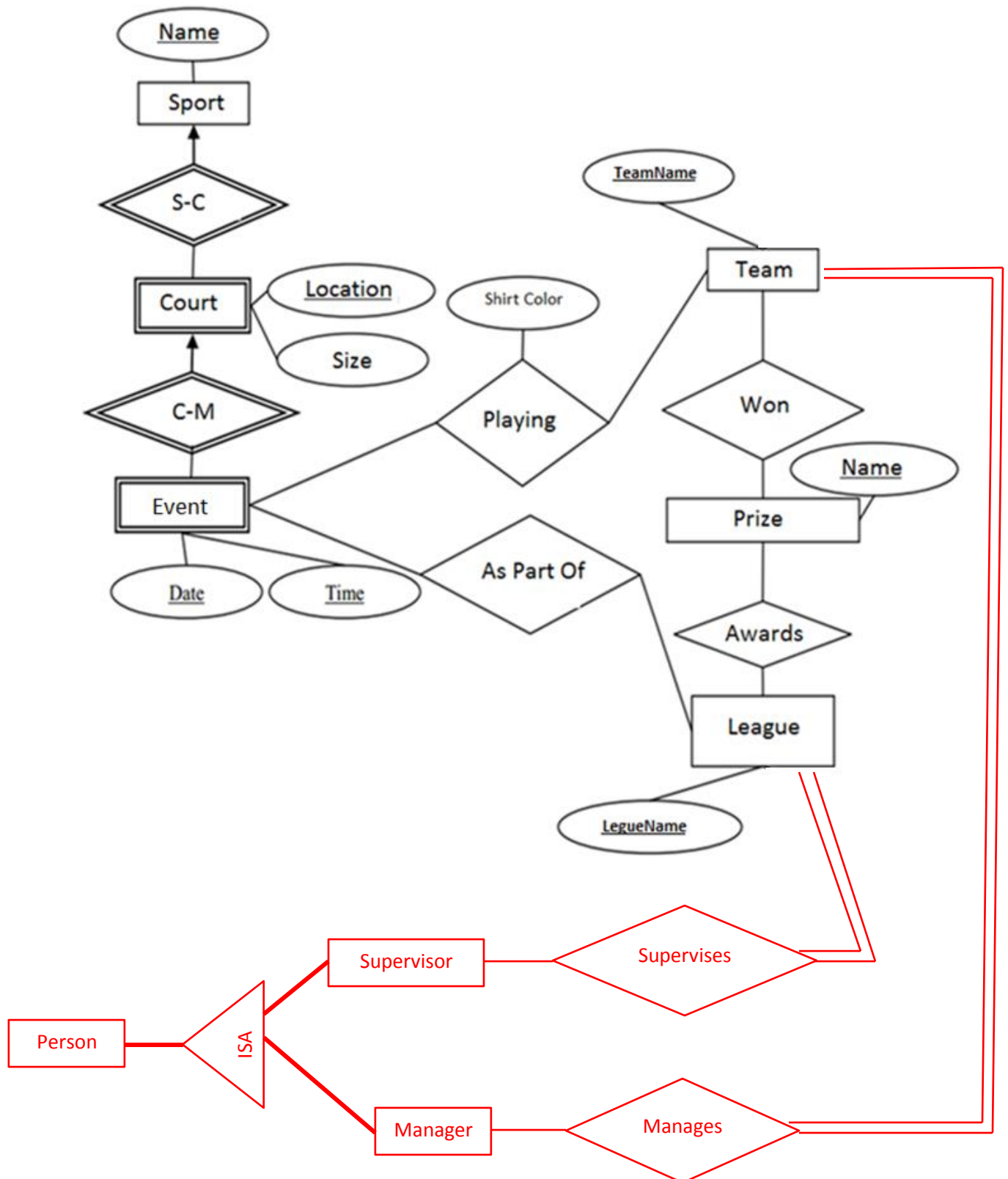
- ii. (3 נק') האם באותו מקום ובאותה השעה יכולים להיות שני משחקים?

כן, לדוגמא אם מדובר משחקים של ענפי ספורט שונים.

- iii. (3 נק') האם קבוצה יכולה לזכות בגביע ששייך לליגה שהיא כלל לא השתתפה במשחקה?

כן, אין שום הגבלה על כך.

ג. 9 נק') כיצד צריך להוסיף לדיאגרמה את manager עבור Team ואת supervisor עבור League כך שכל התנאים על supervisor/manager בדיאגרמה המקורית ישמרו וגם לא יתכן שאותו אדם יהיה גם supervisor וגם manager? שנו את הדיאגרמה בהתאם.
(התכונות manager וsupervisor שהופיעו בדיאגרמה המקורית נמחקו על מנת לאפשר לכן להוסיף את הנתונים הללו בדרך לבחירתכם)



שאלה 2 – שאילתות מידע (31 נק')

השאלות בסעיף זה מתייחסות לדיאגרמת ה-ERD שניתנה בשאלה 1.
להלן התרגום הישיר של חלק מהטבלאות ב-ERD. יתכן ולא תזדקקו לכל הטבלאות הללו על מנת לפתור את השאלות ויתכן שתזדקקו לטבלאות נוספות. אם תשתמשו בטבלאות נוספות לציין את התרגום הישיר של טבלאות הללו:

- Team(TeamName,Manager)
- Event(Date,Time,Location,Name)
- Playing(Date,Time,Location,Name,TeamName,ShirtColor)
- AsPartOf(Date,Time,Location,Name,LeagueName)
- Won(TeamName,Name)

א. (9 נק') כתבו שאילתת SQL שמחזירה:

**מבין הקבוצות שלבשו חולצות בשני צבעים שונים בדיוק כך שאחד מהם הוא אדום, מצאו את שמות המנהלים של הקבוצות שזכו בהכי הרבה פרסים.
מותר להשתמש במבטים רק אם משתמשים בהם יותר מפעם אחת.**

```
CREATE VIEW Amount AS
SELECT TeamName, COUNT(Name) AS Num
FROM Won
WHERE TeamName IN (
    SELECT DISTINCT TeamName
    FROM Playing
    WHERE shirt color = 'RED'
) AND TeamName IN (
    SELECT DISTINCT TeamName FROM Playing
    GROUP BY TeamName
    HAVING COUNT(DISTINCT shirt color) = 2
)
GROUP BY TeamName;
```

```
SELECT manager
FROM Amount a, Team
WHERE NOT EXISTS (
    SELECT *
    FROM Amount b
    WHERE b.Num>a.Num
) AND Team.TeamName=a.TeamName;
```

ב. (7 נק') כתבו שאילת RA שמחזירה:
שמות הפרסים בהם זכו הקבוצות ששיחקו רק בצבע חולצה אדום.

$$PlayedOnlyRed := \pi_{TeamName}(\sigma_{shirtcolor="RED"}Played) \setminus \pi_{TeamName}(\sigma_{shirtcolor \neq "RED"}Played)$$

$$Result := \pi_{name}(Won \times PlayedOnlyRed)$$

ג. (7 נק') נגדיר קבוצה כ-RedPlayed אם הקבוצה שיחקה במשחק שאינו משחק ליגה עם צבע חולצה אדום או ששיחקה במשחק שאינו משחק ליגה נגד קבוצה שהיא RedPlayed.
כתבו תכנית Datalog עם שלילות שמחשבת את קבוצת ה-RedPlayed.

$$LeagueGames(d, t, l, s) \leftarrow AsPartOf(d, t, l, s, _).$$

$$NotLeagueGames(d, t, l, s) \leftarrow Match(d, t, l, s), \neg LeagueGames(d, t, l, s).$$

$$RedPlayed(n) \leftarrow Team(n, _), Playing(d, t, l, s, n, c), c = "RED", NotLeagueGames(d, t, l, s).$$

$$RedPlayed(n) \leftarrow Team(n, _), RedPlayed(n'),$$

$$NotLeagueGames(d, t, l, s),$$

$$Playing(d, t, l, s, n, _), Playing(d, t, l, s, n', _).$$

ד. (8 נק') נתונה השאילתה הבאה בDRC :

$$\{< Name > | \exists team (Won(team, Name) \wedge \\ \neg \exists d1, t1, l1, s1, c1 (\neg (c1 = "RED") \wedge Match(d1, t1, l1, s1) \wedge Played(d1, t1, l1, s1, team, c1)) \wedge \\ \exists d2, t2, l2, s2, c2 (c2 = "RED" \wedge Played(d2, t2, l2, s2, team, c2)))\}$$

קבעו האם השאילתה הנתונה היא Safe Range.
(הראו את הדרך לפי השלבים שנלמדו בכיתה והסבירו)

השלב הראשון שעלינו לעשות הוא להמיר את השאילתה הנתונה לצורה הנורמלית SRNF.
נשים לב שהשאילתה הנתונה כבר נמצאת בצורה הנורמלית ולכן אין צורך לעשות בשלב זה כלום.
(אם ננסה להמיר אותה לSRNF נראה שאין שינויים שעלינו לבצע ובכך ניתן לאמת שהשאילתה אכן בSRNF)

השלב השני הוא חישוב קבוצת המשתנים החופשיים.
ממבנה השאילתה ניתן לראות שהמשתנה היחיד שאינו קשור בכמת (ולכן הוא חופשי) הוא המשתנה Name ולכן $free(\psi) = \{Name\}$
השלב השלישי הוא חישוב קבוצת המשתנים המוגבלים בתחום ($rr(\psi)$)

$$\begin{aligned} rr(Won(team, Name)) &= \{team, Name\} \\ rr(c1 = "RED") &= \{c1\} \\ rr(\neg(c1 = RED)) &= \phi \\ rr(Mathc(d1, t1, l1, s1)) &= \{d1, t1, l1, s1\} \\ rr(Played(d1, t1, l1, s1, team, c1)) &= \{d1, t1, l1, s1, team, c1\} \\ rr(\neg(c1 = "RED") \wedge Mathc(d1, t1, l1, s1) \wedge Played(d1, t1, l1, s1, team, c1)) &= \{d1, t1, l1, s1, team, c1\} \\ rr(\exists d1, t1, l1, s1, c1 (\neg(c1 = "RED") \wedge Mathc(d1, t1, l1, s1) \wedge Played(d1, t1, l1, s1, team, c1))) &= \{team\} \\ rr(\neg \exists d1, t1, l1, s1, c1 (\neg(c1 = "RED") \wedge Mathc(d1, t1, l1, s1) \wedge Played(d1, t1, l1, s1, team, c1))) &= \phi \\ rr(c2 = "RED") &= \{c2\} \\ rr(Played(d2, t2, l2, s2, team, c2)) &= \{d2, t2, l2, s2, team, c2\} \\ rr(c2 = "RED" \wedge Played(d2, t2, l2, s2, team, c2)) &= \{d2, t2, l2, s2, team, c2\} \\ rr(\exists d2, t2, l2, s2, c2 (c2 = "RED" \wedge Played(d2, t2, l2, s2, team, c2))) &= \{team\} \\ rr(Won(team, Name) \wedge \neg \exists d1, t1, l1, s1, c1 (\neg(c1 = "RED") \wedge Mathc(d1, t1, l1, s1) \wedge Played(d1, t1, l1, s1, team, c1)) \\ \wedge \exists d2, t2, l2, s2, c2 (c2 = "RED" \wedge Played(d2, t2, l2, s2, team, c2))) &= \{team, Name\} \\ rr(\exists team (Won(team, Name) \\ \wedge \neg \exists d1, t1, l1, s1, c1 (\neg(c1 = "RED") \wedge Mathc(d1, t1, l1, s1) \wedge Played(d1, t1, l1, s1, team, c1)) \\ \wedge \exists d2, t2, l2, s2, c2 (c2 = "RED" \wedge Played(d2, t2, l2, s2, team, c2)))) &= \{Name\} = rr(\psi) \end{aligned}$$

קיבלנו שמתקיים $free(\psi) = \{Name\} = rr(\psi)$ ולכן השאילתה היא אכן Safe Range.

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

א. הוכח או הפרך את הטענות הבאות:

i. (4 נק') כל רלציה r בעלת בדיוק 2 אטריביוטים היא ב־BCNF.

הטענה נכונה.

אם לא קיימות תלויות לרלציה אזי מתקיים באופן ריק שהרלציה היא ב־BCNF.

אם קיימות תלויות וכולן טריוויאליות אזי מתקיים באופן ריק שהרלציה היא ב־BCNF.

אם קיימות תלויות לא טריוויאליות:

מכיוון שיש רק 2 אטריביוטים כל תלות כזו חייבת להיות מהצורה $X \rightarrow Y$ כך ש $X \cap Y = \emptyset$ ש $|X| = 1, |Y| = 1$. כתוצאה מכך בהכרח מתקיים ש X הוא מפתח של הרלציה ולכן היא ב־BCNF.

ii. (4 נק') לכל רלציה r בעלת בדיוק 3 אטריביוטים וקבוצת תלויות פונקציונליות F כך שכל

תלות בה היא מהצורה $X \rightarrow Y$ כך ש $X \cap Y = \emptyset$ ש $|X| = 2, |Y| = 1$, r היא ב־BCNF.

הטענה נכונה.

אם לא קיימות תלויות לרלציה אזי מתקיים באופן ריק שהרלציה היא ב־BCNF.

אם קיימות תלויות וכולן טריוויאליות אזי מתקיים באופן ריק שהרלציה היא ב־BCNF.

אם קיימות תלויות לא טריוויאליות:

מכיוון שיש רק 3 אטריביוטים וכל תלות היא מהצורה $X \rightarrow Y$ כך ש $X \cap Y = \emptyset$ ש $|X| = 2, |Y| = 1$, בהכרח מתקיים ש X הוא מפתח של הרלציה ולכן היא ב־BCNF.

iii. (4 נק') לכל רלציה r בעלת בדיוק 4 אטריביוטים וקבוצת תלויות פונקציונליות F כך שכל

תלות בה היא מהצורה $X \rightarrow Y$ כך ש $|X| = 2, |Y| = 1$, r היא ב־BCNF.

הטענה לא נכונה.

תהי סכמה $R[A, B, C, D]$ וקבוצת תלויות פונקציונליות $F = \{AB \rightarrow C\}$.

המפתח היחיד של סכמה זו הוא ABD . AB אינו מפתח ולכן הסכמה אינה ב־BCNF.

ב. תהי F קבוצת תלויות פונקציונליות

$$F = \{EA_0 \rightarrow B, EA_1 \rightarrow B, EA_2 \rightarrow B, EA_0 \rightarrow A_1, EA_1 \rightarrow A_0\}$$

i. (4 נק') האם F היא כיסוי מינימלי של עצמה? הסבר.

F אינה כיסוי מינימלי של עצמה.

נבחן לדוגמה את התלות $EA_0 \rightarrow B$. תלות זו היא מיותרת. נראה זאת על ידי כך שנסיק אותה משאר התלויות הקיימות בקבוצה. לצורך כך נשאל האם מתקיים $B \in (EA_0)^+$.

$$(EA_0)^+ = \{EA_0\}$$

לפי התלות $EA_0 \rightarrow A_1$:

$$(EA_0)^+ = \{EA_0A_1\}$$

לפי התלות $EA_1 \rightarrow B$:

$$(EA_0)^+ = \{EA_0A_1B\}$$

מכאן שאכן מתקיים $B \in (EA_0)^+$, כלומר התלות מיותרת ולכן F אינה כיסוי מינימלי של עצמה.

ii. (4 נק') מצא את כל המפתחות הקבילים של F .

המפתחות הקבילים הם:

- EA_0A_2
- EA_1A_2

ג. (7 נק') בהינתן כיסוי מינימלי לקבוצת תלויות, האם הפירוק שיוצר אלגוריתם הפירוק לצורה הנורמלית $3NF$ (תוך שימוש בכיסוי הנתון) הוא יחיד?
אם כן, הסבירו מדוע. אם לא, הסבירו באילו מקרים יכולים להיות מספר פירוקים.

הפירוק שיוצר האלגוריתם אינו יחיד.

אם לקבוצת התלויות מספר מפתחות קבילים ואלגוריתם הפירוק אינו יוצר סכמה שכוללת מפתח עלינו להוסיף סכמה חדשה שת כלול מפתח. במצב זה כל מפתח שנבחר יניב פירוק אחר.

שאלה 4 – XML (! נק')

נתון DTD הבא :

1. <!ELEMENT postDB (branch*,client*,mail*) >
2. <!ELEMENT branch (city)>
3. <!ATTLIST branch
4. type CDATA #REQUIRED
5. mails IDREFS #REQUIRED
6. >
7. <!ELEMENT client (city)>
8. <!ATTLIST client
9. name ID #REQUIRED
10. >
11. <!ELEMENT city (#PCDATA)>
12. <!ELEMENT mail (date, content)>
13. <!ATTLIST mail
14. TrackingNumber ID #REQUIRED
15. type (letter|package) "letter"
16. sender IDREF #REQUIRED
17. recipient IDREF #REQUIRED
18. >
19. <!ELEMENT content (#PCDATA)>
20. <!ELEMENT date (year,month,day)>
21. <!ELEMENT year (#PCDATA)>
22. <!ELEMENT month (#PCDATA)>
23. <!ELEMENT day (#PCDATA)>

מסד הנתונים מתאר מידע על מערכת חלוקת דואר.

אלמנט השורש של המסד הוא האלמנט postDB.

המסד מכיל נתונים על סניפים, לקוחות ודואר.

כל סניף מכיל את העיר בה הוא נמצא, הפניה לכל פרטי הדואר שברשותו וטיפוס. הטיפוס של כל סניף הוא מספר שלם גדול או שווה ל1. ככל שהמספר יותר נמוך כך הסניף יותר מרכזי.

כל לקוח מיוצג על ידי שמו והעיר בא הוא רשום. כל לקוח יכול להיות רשום לכל היותר בעיר אחת.

פרטי הדואר מכילים תאריך שליח ה (שמורכב מיום, חודש ושנה) ותוכן. לכל פריט דואר מוצמדים גם הפניות ללקוח השולח והלקוח המקבל, טיפוס פריט הדואר (חבילה או מכתב), מספר מעקב.

שאלות:

א. (6 נק') מערכת חלוקת הדואר מתחדשת ואנו מעוניינים לאפשר לאותו לקוח (client, זהות הלקוח נקבעת לפי namen) להירשם במספר ערים (city). כמו כן אנו מעוניינים לאפשר לכל לקוח להפנות (forward) את כל הדואר שלו ללקוח אחר. שימו לב, שבמקרה שלקוח הפנה את הדואר שלו ללקוח אחר שרשום במספר ערים, נרצה לדעת לאיזה מן הערים הדואר מופנה. שנו את ה DTD הנתון על מנת שיתמוך בשינויים הנדרשים. הסבירו כיצד המידע החדש ייוצג באמצעות השינויים שהצעתם. מותר לעשות כל שינוי שרוצים. אם ברצונכם לשנות שורות קיימות ציינו את מספר השורה.

7. <! ELEMENT client (city+, forward?)>

<!ELEMENT forward (city)>

<!ATTLIST forward
to IDREF #REQUIRED
>

ב. (5 נק') נתונה שאילתת ה XPath1.0 הבאה :

//client[@name = //@recipient][not(city = //branch/city)]/@name

כתבו בשפה טבעית מה מחשבת השאילתה הנתונה

השאילתה מחזירה את שמות כל הלקוחות ששלחו אליהם פריט דואר אבל לא קיים אף סניף בעיר מגוריהם.

ג. 9 נק') כתבו תכנית XQuery שמקבלת מסמך XML בשם post.xml התקף DTD הנתון בתחילת השאלה ומייצרת ממנו דו"ח שנתי. הדו"ח יכיל לכל שנה (שקיימת במסמך) אלמנט שיכיל תת אלמנט עבור חבילות (package) ותת אלמנט עבור מכתבים (letter). תת האלמנטים יכילו את מספר פריטי הדואר מהסוג המתאים שנשלחו ואת נתוני פריטי הדואר (שמות השולח, חודש השליחה) (מתוך התאריך) ומספר המעקב של פריט הדואר). יש למיין את השנים בסדר עולה. על הפלט שלכם להיות תקף ל DTD הבא :

DTD:

```
<!ELEMENT report (yearly*)>
<!ELEMENT yearly (packages?,letters?)>
<!ATTLIST yearly
    year CDATA #REQUIRED
>
<!ELEMENT packages (mail+)>
<!ATTLIST packages
    num CDATA #REQUIRED
>
<!ELEMENT letters (mail+)>
<!ATTLIST letters
    num CDATA #REQUIRED
>
<!ELEMENT mail EMPTY>
<! ATTLIST mail
    tracking ID #REQUIRED
    sender IDREF # REQUIRED
    month CDATA #REQUIRED
>
```

העמוד הבא מכיל קלט ופלט לדוגמה.

שימו לב לטיפול נכון במקרה בו אין חבילות או אין מכתבים.

בהינתן הקלט הבא:

```
<postDB>
  <client name="moshe"><city>Haifa</city></client>
  <client name="david"><city>Haifa</city></client>
  <mail TrackingNumber="1" type="package" sender="moshe" recipient="david">
    <content>Something</content>
    <date><year>2011</year><month>12</month><day>31</day></date>
  </mail>
  <mail TrackingNumber="1" type="letter" sender="david" recipient="moshe">
    <content>Thanks</content>
    <date><year>2012</year><month>1</month><day>1</day></date>
  </mail>
</postDB>
```

נצפה לפלט הבא:

```
<report>
  <yearly year="2011">
    <packages num="1">
      <mail tracking ="1" sender="moshe" recipient="david"
        month="12" day="31" />
    </packages>
  </yearly>
  <yearly year="2012">
    <letters num="1">
      <mail tracking ="1" sender="david" recipient="moshe"
        month="1" day="1" />
    </letters>
  </yearly>
</report>
```

```

<report>{
  for $year in distinct-values(doc(post.xml)//year)
  let $packages := doc(post.xml)//mail[@type = "package"][date/year = $year]
  let $letters := doc(post.xml)//mail[@type = "letter"][date/year = $year]
  return
    <yearly year={$year}>
      if($packages) then
        <packages num={count($packages)}>
          for $p in $packages
          return
            <mail tracking={$p/@ TrackingNumber }
              sender={$p/@sender}
              month={$p/date/month}
            />
          </packages>
        if($letters) then
          < letters num={count($packages)}>
            for $l in $ letters
            return
              <mail tracking={$l/@ TrackingNumber }
                sender={$l/@sender}
                month={$l/date/month}
              />
            </ letters >
          </yearly>
    }</report>

```

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: _____

