

סמסטר אביב התשע"ג

מרצח : פרופ' יוהן מקובסקי

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

מועד ב' (ג' בחשוון התשע"ג, 7 באוקטובר 2013)

מס' סטודנט:**פירוט השאלות והניקוד:**

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

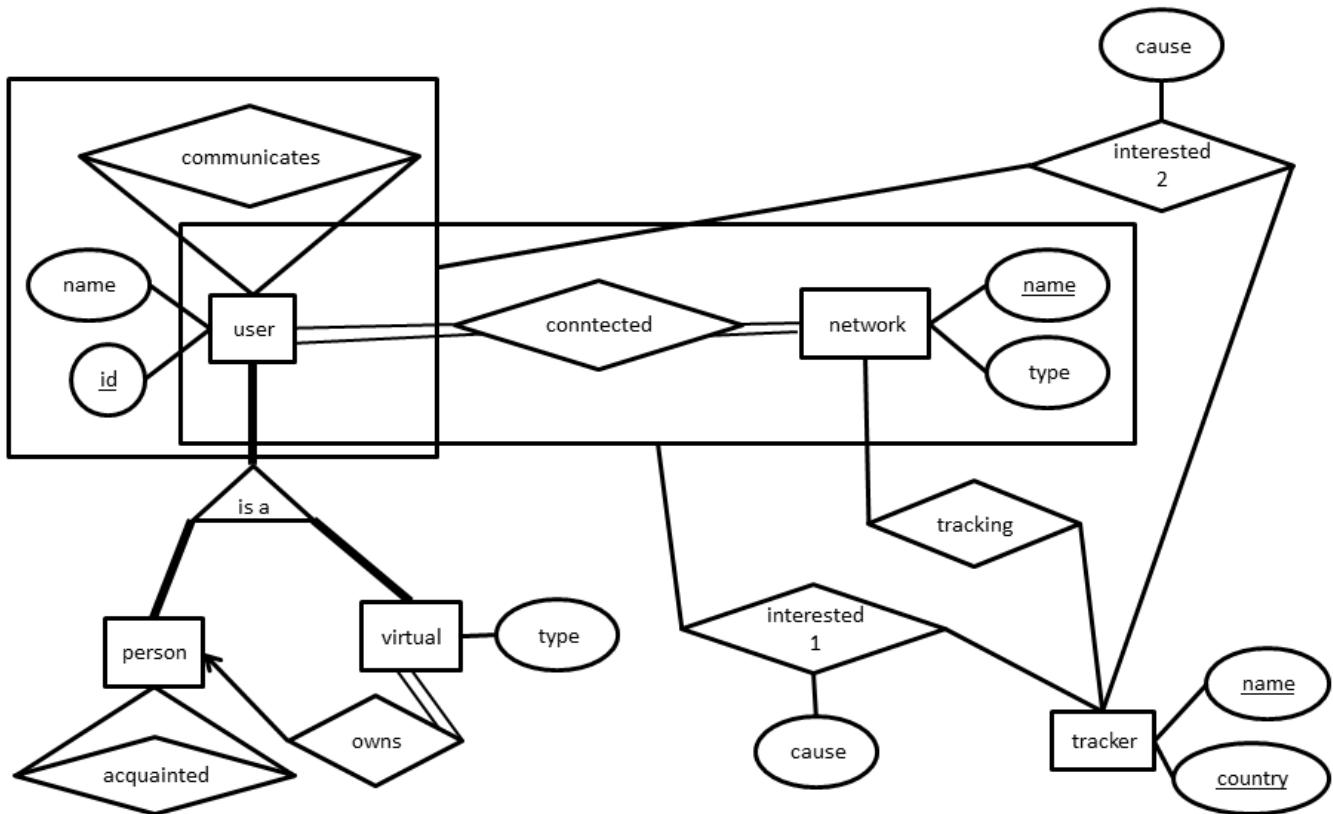
הנחיות לנבחנים

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 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 (26 נק')

נתונה דיאגרמת ה-ERD הבאה :



בדיאגרמה לפניכם מוצגת מערכת מעקב אחר משתמשים. המערכת מכילה משתמשים user (שיכולים להיות אנשים person או משתמשים וירטואליים virtual שנמצאים בבעלות של משתמש אחר) ורשתות חברתיות network. משתמשים יכולים להיות מחוברים connected לרשת חברתית. משתמשים יכולים לתקשר אחד עם השני. שני אנשים יכולים להכיר acquainted אחד את השני מחוץ לרשתות החברתיות. בנוסף קיימים גופי tracker שיכולים לעקוב tracking אחר רשתות חברתיות ולהצהיר על עניין במשתמשים ברשת מסוימת interested 1 או בתקשורת בין שני משתמשים interested 2 (באמצעות הקשר communicates). נגדיר ששני משתמשים יכולים לתקשר ישירות אם הם חברים באותה רשת חברתית או שהם מכירים אחד את השני. שימו לב, הקשר communicates מייצג תקשורת בין שני משתמשים אך אינו הופך אותה לאפשרית

שאלות:

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

שם הטבלה: Owens
(1 נק') שדות: virtual.id, person.id
(1 נק') מפתחות: virtual.id
(2 נק') תלויות פונקציונליות: virtual.id → person.id

שם הטבלה : Interested1
(1 נק') שדות : tracker.name, country, id, network.name, cause
(1 נק') מפתחות : tracker.name, country, id, network.name
(2 נק') תלויות פונקציונליות : cause → {tracker.name, country, id, network.name}

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

i. (3 נק') האם ייתכן שמשתמש וירטואלי יהיה בבעלות של משתמש וירטואלי אחר?

לא, ה-ERD מצהיר במפורש שהקשר owns הוא בין משתמש וירטואלי למשתמש אנושי ולא יתכן שאותו משתמש יהיה גם אנושי וגם וירטואלי.

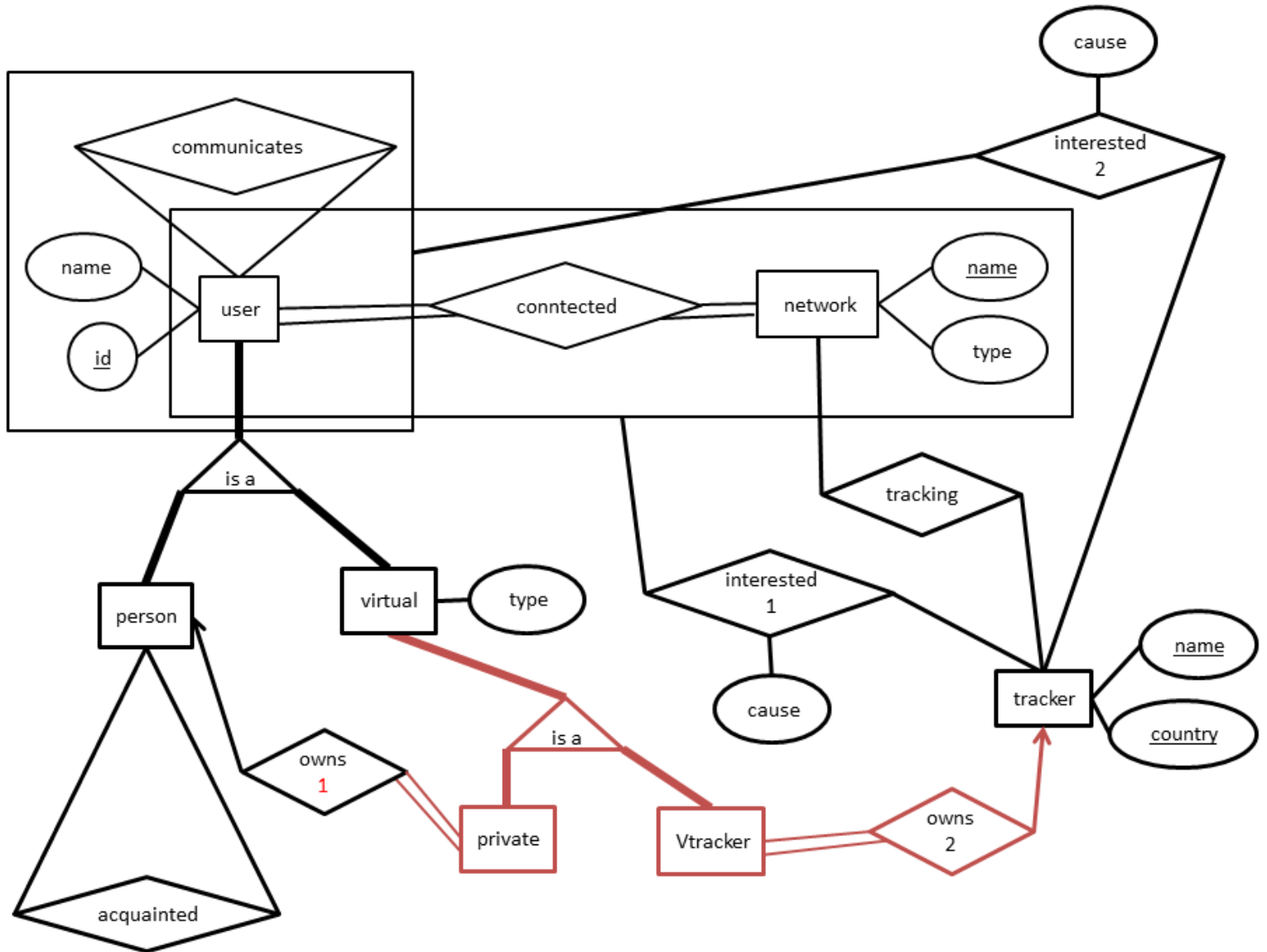
ii. (3 נק') האם ה-ERD מאפשר לשני משתמשים שלא מכירים אחד את השני לתקשר (ישירות או בעקיפין) גם אם אינם חברים באותה רשת חברתית?

כן, הם יכולים לתקשר דרך משתמש שלישי שיכול לתקשר עם שניהם והוא ישמש כמתווך.

iii. (3 נק') האם גוף ביון יכול לעקוב אחר רשת חברתית ריקה (שאף משתמש אינו מחובר אליה)?

לא, הקשר connected מיוצג עם קו עבה ולכן כל רשת חברתית צריכה להשתתף בו ולכן היא לא ריקה.

ג. 9 נק') גופי הביון רוצים לשפר את יכולות המעקב שלהם ולהשתמש במשתמשים מדומים (וירטואליים) לצורך התחזות. כיצד יש לשנות את הדיאגרמה הנתונה על מנת לאפשר לגופי ביון להחזיק משתמשים וירטואליים? שימו לב, משתמש וירטואלי בשימוש גוף ביון לא יכול להיות בבעלות פרטית וכך משתמש וירטואלי חייב להיות בבעלות כלשהי. שנו את הדיאגרמה בהתאם והסבירו.



שאלה 2 – שאילות מידע (31 נק')

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

- tracker(name, country)
- network(name, type)
- user(id, name)
- person(id)
- virtual(id, type)
- connected(name, id)
- tracking(tname, tcountry, nname)
- acquainted(id1, id2)
- communicates(id1, id2)

א. (9 נק') כתבו שאילתת SQL שמחזירה:

מזהי כל גופי הביון (tracker) שעוקבים (tracking) אחרי רשת חברתית (network) בעלת יותר מ-1000 משתמשים (user) שאינם וירטואליים (virtual).
מותר להשתמש במבטים רק אם משתמשים בהם יותר מפעם אחת.

```
SELECT name, country
FROM tracker, tracking
WHERE tracker.name = tracking.tname AND tracker.country = tracking.tcountry AND
      Tracking.nname in (
        SELECT name
        FROM network, connected
        WHERE network.name = connected.name AND connected.id NOT IN (
          SELECT id
          FROM virtual
        )
      )
GROUP BY network.name
HAVING COUNT(DISTINCT connected.id) > 1000
)
```

ב. (7 נק') נתונה השאילתה הבאה ב-RA :

$$R1 := \pi_{id1,id2}(\rho_{id \rightarrow id1}(connected) \bowtie \rho_{id \rightarrow id2}(connected))$$

$$R2 := (\rho_{name \rightarrow name1,id \rightarrow id1}(user) \times \rho_{name \rightarrow name2,id \rightarrow id2}(user))$$

$$Result := \pi_{name1,name2} \left(R2 \bowtie acquainted \bowtie ((\pi_{id1,id2} R2) \setminus R1) \right)$$

כתבו מה תהיה תוצאת השאילתה (כלומר, מה יהיה התוכן של הרלציה $Result$).

זוגות של שמות של משתמשים שהם מכרים אך אינם משתמשים באותה רשת חברתית

ג. (7 נק') נגדיר ששני משתמשים ($user$) שאינם חברים באותה רשת חברתית ($network$) (למשל משתמש א' חבר רק ברשת A ומשתמש ב' חבר רק ברשת B) יכולים לתקשר אחד עם השני דרך מתווך/פרוקסי עם קיים משתמש וירטואלי שחבר בשתי הרשתות (כלומר חבר גם בA וגם בB). כתבו תכנית Datalog שמוצאת את כל זוגות המזהים של משתמשים שיכולים לתקשר (ישירות או דרך מספר כלשהו של מתווכים). שימו לב, אין להתייחס בתשובתכם לאפשרות ששני משתמשים יכירו ($acquainted$) אחד השני ויתקשרו כך.

$conduit(N, N) \leftarrow connected(N, _)$

$conduit(N1, N2) \leftarrow$
 $conduit(N1, N3), connected(N3, Id), connected(N2, Id), virtual(Id, _)$

$result(Id1, Id2) \leftarrow$
 $connected(N1, Id1), connected(N2, Id2), conduit(N1, N2), user(Id1, _), user(Id2, _)$

ד. (8 נק') נתונה השאילתה הבאה בDRC בצורה הנורמלית SRNF:

$$\exists tn, tc \left(\neg \exists id1, id2 \left(\exists c (interested2(tn, tc, id1, id2, c)) \right. \right. \\ \left. \left. \wedge \neg \exists n \left((connected(n, id1) \vee connected(n, id2)) \wedge tracking(tn, tc, n) \right) \right) \right)$$

קבעו האם השאילתה הנתונה היא Safe Range.

הראו את הדרך לפי השלבים שנלמדו בכיתה והסבירו.

השלב הראשון הוא חישוב קבוצת המשתנים החופשיים.

ממבנה השאילתה ניתן לראות שכל המשתנים קשורים בכמת ולכן אין משתנים חופשיים ולכן $free(\psi) = \phi$

השלב השני הוא חישוב קבוצת המשתנים המוגבלים בתחום ($rr(\psi)$)

$$rr(interested2(tn, tc, id1, id2, c)) = \{tn, tc, id1, id2, c\}$$

$$rr(connected(n, id1)) = \{n, id1\}$$

$$rr(connected(n, id2)) = \{n, id2\}$$

$$rr(tracking(tn, tc, n)) = \{tn, tc, n\}$$

$$rr(connected(n, id1) \vee connected(n, id2)) = \{n\}$$

$$rr((connected(n, id1) \vee connected(n, id2)) \wedge tracking(tn, tc, n)) = \{tn, tc, n\}$$

$$rr(\exists n ((connected(n, id1) \vee connected(n, id2)) \wedge tracking(tn, tc, n))) = \{tn, tc\}$$

$$rr(\neg \exists n ((connected(n, id1) \vee connected(n, id2)) \wedge tracking(tn, tc, n))) = \phi$$

$$rr(\exists c (interested2(tn, tc, id1, id2, c))) = \{tn, tc, id1, id2\}$$

$$rr(\exists c (interested2(tn, tc, id1, id2, c)) \wedge \neg \exists n ((connected(n, id1) \vee connected(n, id2)) \wedge tracking(tn, tc, n))) \\ = \{tn, tc, id1, id2\}$$

$$rr(\exists id1, id2 (\exists c (interested2(tn, tc, id1, id2, c))$$

$$\wedge \neg \exists n ((connected(n, id1) \vee connected(n, id2)) \wedge tracking(tn, tc, n)))) = \{tn, tc\}$$

$$rr(\neg \exists id1, id2 (\exists c (interested2(tn, tc, id1, id2, c))$$

$$\wedge \neg \exists n ((connected(n, id1) \vee connected(n, id2)) \wedge tracking(tn, tc, n)))) = \phi$$

$$rr(\exists tn, tc (\neg \exists id1, id2 (\exists c (interested2(tn, tc, id1, id2, c))$$

$$\wedge \neg \exists n ((connected(n, id1) \vee connected(n, id2)) \wedge tracking(tn, tc, n)))) = \perp$$

קיבלנו שלא מתקיים $free(\psi) = rr(\psi)$ ולכן השאילתה היא אינה Safe Range.

שאלה 3 – פירוקים ותלויות פונקציונליות (27 נק')

א. (9 נק') נתונה הטענה הבאה :

נתונה סכמה $R[U]$ בעלת קבוצת התכונות U וקבוצת תלויות פונקציונליות F , כאשר קבוצת התכונות $K \subseteq U$ היא מפתח של R , כלומר מתקיים $K \rightarrow U \in F^+$. נחלק את U לשתי תת קבוצות $U_1, U_2 \subseteq U$ כך שמתקיים $U_1 \cap U_2 = \emptyset, U_1 \cup U_2 = U$. $K \cap U_1 \neq \emptyset, K \cap U_2 \neq \emptyset$. הפירוק של U ל $R[U_1], R[U_2]$ ו $R[K]$ הינו משמר מידע.

הרלציה הבאה מהווה דוגמה נגדית לטענה:

A	B	C	D
1	2	3	4
7	2	5	9
8	6	3	4

מצאו פירוק U_1, U_2 וקבוצת תלויות F כך שהטבלה הנ"ל מפריכה את הטענה והסבירו.

$$U_1 = \{A, C\}$$

$$U_2 = \{B, D\}$$

$$F = \{BC \rightarrow A, C \rightarrow D\}$$

$$K = \{B, C\}$$

הסבר:

קל לראות שהתנאים לקיום הטענה מתקיימים. נבחן האם גם מסקנת הטענה מתקיימת:

$$R[U_1] = \begin{Bmatrix} 1 & 3 \\ 7 & 5 \\ 8 & 3 \end{Bmatrix}, R[U_2] = \begin{Bmatrix} 2 & 4 \\ 2 & 9 \\ 6 & 4 \end{Bmatrix}, R[K] = \begin{Bmatrix} 2 & 3 \\ 2 & 5 \\ 6 & 3 \end{Bmatrix}$$

הצירוף של הטבלאות הללו הוא:

$$R[U_1] \bowtie R[U_2] \bowtie R[K] = \begin{Bmatrix} 1 & 2 & 3 & 4 \\ 1 & 2 & 3 & 9 \\ 1 & 6 & 3 & 4 \\ 7 & 2 & 5 & 4 \\ 7 & 2 & 5 & 9 \\ 8 & 2 & 3 & 4 \\ 8 & 2 & 3 & 9 \\ 8 & 6 & 3 & 4 \end{Bmatrix}$$

מכיוון שהצירוף אינו זהה לרלציה המקורית אנו יודעים שהפירוק אינו משמר מידע ולכן הטענה אינה נכונה.

ב. (18 נק') תהי F קבוצת תלויות פונקציונליות

$$F = \{A \rightarrow BCD, B \rightarrow ACD, CD \rightarrow ABD\}$$

לכל אחת מקבוצות התלויות הבאות קבע האם היא כיסוי והאם היא מינימלית (כלומר מהווה כיסוי מינימלי של עצמה). אין צורך לנמק.

i. (4 נק') $G_1 = \{A \rightarrow BCD, B \rightarrow C, C \rightarrow A\}$

<u>תכונה</u>	<u>האם מתקיימת?</u>
כיסוי	לא
מינימלי	לא

ii. (4 נק') $G_2 = \{A \rightarrow B, A \rightarrow C, A \rightarrow D, CD \rightarrow A, B \rightarrow A, B \rightarrow C, B \rightarrow D\}$

<u>תכונה</u>	<u>האם מתקיימת?</u>
כיסוי	כן
מינימלי	לא

iii. (4 נק') $G_3 = \{A \rightarrow C, A \rightarrow D, B \rightarrow C, B \rightarrow D, CD \rightarrow A, CD \rightarrow B\}$

<u>תכונה</u>	<u>האם מתקיימת?</u>
כיסוי	כן
מינימלי	כן

iv. (6 נק') מצאו כיסוי מינימלי ל- F ששונה מ- G_1, G_2, G_3 .

פתרונות אפשריים:

- $\{A \rightarrow B, A \rightarrow C, A \rightarrow D, B \rightarrow A, CD \rightarrow A\}$
- $\{A \rightarrow B, B \rightarrow C, B \rightarrow D, CD \rightarrow A\}$
- $\{A \rightarrow B, B \rightarrow A, B \rightarrow C, B \rightarrow D, CD \rightarrow B\}$
- $\{A \rightarrow C, A \rightarrow DB \rightarrow A, CD \rightarrow B\}$

לכל הפתרונות הללו ניתן להגיע מהרצת האלגוריתם שהוצג בתרגולים (ההבדל ביניהם הוא סדר המעבר על התלויות).

נתון DTD הבא :

1. <!ELEMENT dict (letter*) >
2. <!ELEMENT letter (word?,letter*)>
3. <!ATTLIST letter
4. char CDATA #REQUIRED
5. >
6. <!ELEMENT word EMPTY>
7. <!ATTLIST word
8. fullword ID #REQUIRED
9. meaning CDATA #IMPLIED
10. synonyms IDREFS #IMPLIED
11. >

מסד הנתונים מתאר מילון.

אלמנט השורש של המסד הוא האלמנט dict.

המסד מכיל עץ כך שכל ענף בעץ הוא אות נוספת במילה. תחת לכל אות יכולה להיות מילה (כלומר המילה שמורכבת מהמסלול הנוכחי בעץ) ויכולות להיות אותיות נוספות.

כל מילה מכילה את המשמעות שלה והפניות למילים הנרדפות לה (synonyms).

להלן דוגמא למסמך XML שתקף ל-DTD הנתון :

```
<dict>
  <letter char="o">
    <letter char="n">
      <word fullword="on" synonyms=" " />
      <letter char="e">
        <word fullword="one" meaning="1" synonyms=" " />
      </letter>
    </letter>
  <letter char="r">
    <word fullword="or" meaning="disjunction" />
  </letter>
</dict>
```

שאלות:

- א. (6 נק') נגדיר מסלול מת במילון כמסלול שאינו מסתיים במילה (כלומר רצף אותיות כלשהו שאינו מהווה מילה). נרצה לשנות את ה DTD הנתון כך שלא יתכנו במילון מסלולים מתים (מבלי להפוך את המילון למילון ריק).
- שנו את ה DTD הנתון על מנת שיתמוך בשינויים הנדרשים. מותר להוסיף שורות חדשות. אם ברצונכם לשנות ו/או למחוק שורות קיימות ציינו את מספר השורה. הסבירו כיצד השינוי המוצע מונע קיום מסלולים מתים.

2. <!ELEMENT letter ((word,letter*)|(letter+))>

לאחר התיקון המוצע, האופציה היחידה לסיום מסלול היא באלמנט מסוג word. אם לא הגענו לאלמנט word נהיה מחויבים להמשיך את המסלול על ידי הוספת לפחות אות אחת.

- ב. (5 נק') נתונה שאילתת ה XPath1.0 הבאה:
- ```
/*/*/*letter[@char="a"]//letter[@char="z"]//word [count(id(@synonyms)) > 0] /@fullword
```
- כתבו בשפה טבעית מה מחשבת השאילתה הנתונה

השאילתה מחזירה את מזהי כל המילים במילון שהאות השלישית בהן היא האות a, הן מכילות את האות z במקום כלשהו לאחר האות a (כלומר אחרי האות השלישית) ויש להן לפחות מילה נרדפת אחת.

ג. 9 נק') כתבו תכנית XQuery שמקבלת מסמך XML בשם dict.xml התקף ל DTD הנתון בתחילת השאלה ומייצרת ממנו מסמך חדש. המסמך החדש יכיל לכל אות שמסיימת מילה (כלומר האות האחרונה במילה) שקיימת במסמך המקורי אלמנט שיכיל תת אלמנט עבור כל אחת מהמילים שמסתיימות באות זו. תת האלמנט עבור כל מילה יכיל את המילה עצמה (כתכונה), את משמעותה (אם קיימת) ואת כל המילים הנרדפות לה. יש למיין את האותיות, המילים והמילים הנרדפות בסדר לקסיקוגרפי עולה.

על הפלט שלכם להיות תקף ל DTD הבא :

**DTD:**

```
<!ELEMENT dict2 (letter2*)>
<!ELEMENT letter2 (word2+)>
<!ATTLIST letter2
 char CDATA #REQUIRED
>
<!ELEMENT word2 (meaning?,synonym*)>
<!ATTLIST word2
 fullword CDATA #REQUIRED
>
<!ELEMENT synonym (PCDATA)>
<!ELEMENT meaning (PCDATA)>
```

העמוד הבא מכיל קלט ופלט לדוגמה.

## בהינתן הקלט הבא:

```
<dict>
 <letter char="a">
 <letter char="a">
 <word fullword="aa" meaning="mean" synonyms="ba" />
 </letter>
 <letter char="b">
 <word fullword="ab" meaning="mean2" synonyms="" />
 </letter>
 </letter>
 <letter char="b">
 <letter char="a">
 <word fullword="ba" synonyms="aa" />
 </letter>
 </letter>
</dict>
```

## נצפה לפלט הבא:

```
<dict2>
 <letter2 char="a">
 <word2 fullword="aa">
 <meaning>mean</meaning>
 <synonym>ba</ synonym >
 </word2>
 <word2 fullword="ba">
 <synonym>aa</ synonym >
 </word2>
 </letter2>
 <letter2 char="b">
 <word2 fullword="ab">
 <meaning>mean2</meaning>
 </word2>
 </letter2>
</dict2>
```

```

<dict2>{
 for $char in distinct-values(doc(dict.xml)//letter[word]/@char)
 let $words := doc(dict.xml)//word[../@char=$char]
 order by $char
 return
 <letter2 char={$char}>{
 for $word in $words
 let $synonyms := id($word/@synonyms)/@fullword
 order by $word/@fullword
 return
 <word2 fullword={$word/@fullword}>
 if($word/@meaning) then
 <meaning>{$word/@meaning}</meaning>
 {
 for $s in $synonyms
 order by $s
 return
 <synonym>{$s }</synonym>
 }
 </word2>
 </letter2>
 }
}</dict2>

```

### 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: \_\_\_\_\_

