



פרופסור עודד שמואלי
מר סעאב מנסור
גברת לינה זריבץ'

חורף תשס"ו
26, מרץ 2006

מערכות מסדי נתונים - 236363

בחן 2 - מועד ב' - פתרון

הזמן: 105 דקות (1:45 שעות)
במבחן זה 8 עמודים

שם פרטי: _____

שם משפחה: _____

מס' סטודנט: _____

פקולטה: _____

שאלה	נקודות	מתוך
שאלה 1	פירוקים	28
שאלה 2	צורות נורמליות	28
שאלה 3	XPath	12
שאלה 4	XQuery	16
שאלה 5	Datalog	16
סה"כ		100

הערות:

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

בהצלחה

שאלה 1 – פירוקים (28 נקודות)

נתונה סכמה $R(A,B,C,D,E,G)$ וקבוצת תלויות פונקציונליות
 $F = \{AB \rightarrow DE, DE \rightarrow C, AE \rightarrow BD, C \rightarrow D, D \rightarrow ACE\}$
שים לב שהאטריביוט G אינו מופיע באף תלות של F .

1. (5 נקודות) מצא כיסוי מינימאלי של F .

$$F_c = \{AB \rightarrow D, AE \rightarrow B, C \rightarrow D, D \rightarrow A, D \rightarrow C, D \rightarrow E\}$$

2. (5 נקודות) רשום את כל המפתחות הקבילים של R, F .

DG, CG, ABG, AEG

3. (6 נקודות) מצא עבור R, F הנתונים, פירוק משמר מידע ומשמר תלויות לסכמות ב-BCNF הכולל מספר מינימאלי של תתי-סכמות.

$ABCDE, DG$

4. (6 נקודות) האם קיים עבור R, F הנתונים, פירוק משמר מידע ומשמר תלויות לשתי תתי-סכמות הנמצאות ב-BCNF והמכילות לכל היותר 3 אטריביוטים? אם כן, מצא אותו. אם לא, נמק.

לא קיים פירוק כזה, כי אז החיתוך בין שתי הסכמות יהיה ריק ושימור המידע לא

יתקיים.

5. (6 נקודות) האם קיים עבור R, F הנתונים, פירוק משמר מידע ומשמר תלויות לארבע תתי-סכמות הנמצאות ב-BCNF והמכילות, כל אחת, לכל היותר 3 אטריביוטים? אם כן, מצא אותו. אם לא, נמק.

ABD, ABE, ACD, ABG או ABD, ABC, ABE, ABG

שאלה 2 – צורות נורמאליות (28 נקודות)

יהיו R_1, F_1 ו- R_2, F_2 סכמות הנמצאות ב-BCNF. בנוסף, ידוע כי $R_1 \subset R_2$ וגם $R_1 \neq R_2$ (R_1 מוכל **ממש** ב- R_2). כמו כן נתון כי הן F_1 והן F_2 מכילות לפחות תלות לא טריוויאלית אחת.

1. (7 נקודות) האם $(F_1 \cup F_2)$, R_2 בהכרח נמצאת ב-BCNF? אם כן, תן הוכחה פורמאלית. אם לא, תן דוגמא.

לא, דוגמא נגדית:

$R_1(A, C), F_1 = \{C \rightarrow A\}$

$R_2(A, B, C), F_2 = \{B \rightarrow C, B \rightarrow A\}$

2. (7 נקודות) האם R_2, F_1 יכולה להיות ב-BCNF? אם כן, תן דוגמא. אם לא, תן הוכחה פורמאלית.

R_2, F_1 לא יכולה להיות ב-BCNF, כיוון שכל מפתח-על חייב להכיל את $R_2 \setminus R_1$.

3. (7 נקודות) האם R_2, F_1 יכולה להיות ב-3NF? אם כן, תן דוגמא. אם לא, תן הוכחה פורמאלית.

$R_1(A,B), F_1 = \{A \rightarrow B, B \rightarrow A\}, R_2(A,B,C)$

4. (7 נקודות) האם $R_1, \pi_{R_1} F_2$ בהכרח נמצאת ב-BCNF? אם כן, תן הוכחה פורמאלית. אם לא, תן דוגמא.

כן. נסמן לשם נוחות את $\pi_{R_1} F_2$ ב- F .

נניח בשלילה כי R_1, F אינה נמצאת ב-BCNF, כלומר קיימת תלות לא טריוויאלית

$X \rightarrow A$ כך ש- X אינו מפתח על ב- R_1, F . ז"א שקיים אטריביוט $B \in R_1$ כך ש- $B \notin X_F^+$.

מכאן נובע כי $B \notin X_{F_2}^+$ כי כל תלות ב- F מתקיימת גם ב- R_2, F_2 .

לכן $X \rightarrow A$ מתקיימת ב- R_2, F_2 למרות ש- X אינו מפתח ב- R_2, F_2 .

זו סתירה לעובדה כי R_2, F_2 נמצאת ב-BCNF.

שאלה 3 – XPath (12 נקודות)

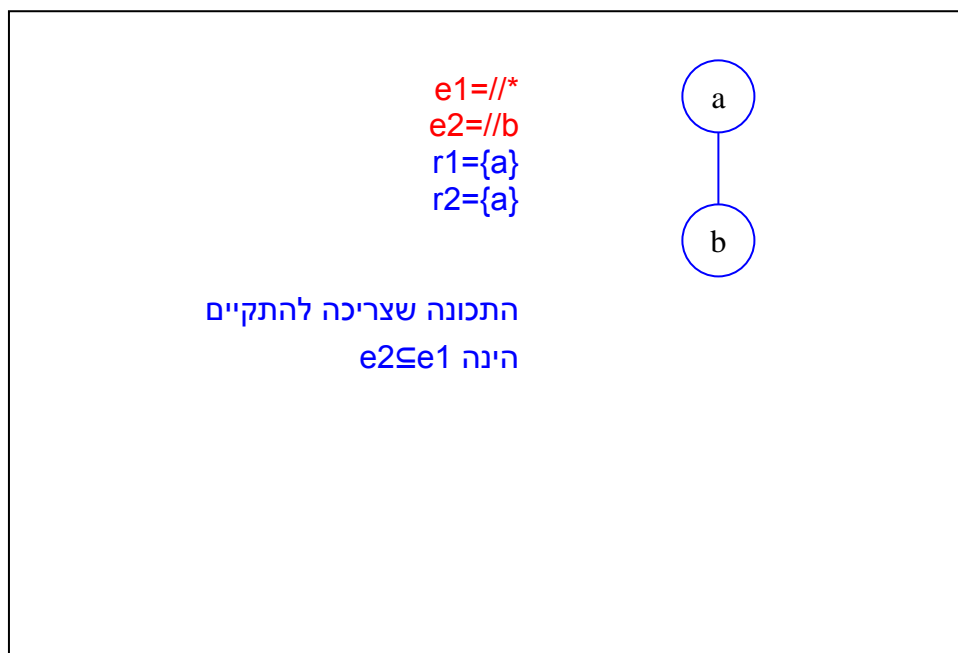
נתונים שני ביטויים ב- XPath, e_1, e_2 שמחזירים קבוצות של צמתים. בנוסף, נתונים שני ביטויי ה- XPath הבאים:

$r_1: e_1[\text{count}(. | e_2) \neq \text{count}(e_2)]$

$r_2: e_1[\text{count}(. | e_2) \neq \text{count}(e_2)] | e_2[\text{count}(. | e_1) \neq \text{count}(e_1)]$

א. (6 נקודות) צייר עץ שמייצג מסמך XML ומצא ביטויים ספציפיים e_1 ו- e_2 , כך שביטויי ה- XPath r_1, r_2 יחזירו תוצאות שוות. על e_1 ו- e_2 להחזיר קבוצות צמתים לא ריקות על העץ שציירת.

צינו איזו תכונה כללית e_1, e_2 צריכים לקיים על מנת שהביטויים r_2, r_1 יחזירו את אותם הערכים.



ב. (6 נקודות) הסבר בקצרה את משמעות ביטויי ה- XPath r_1, r_2 באופן כללי (לאו דווקא רק עבור הביטויים הספציפיים בסעיף א').

הביטוי הראשון מחשב את ההפרש $e_1 \setminus e_2$

הביטוי השני מחשב את ההפרש הסימטרי בין e_1 ל- e_2

שאלה 4 – XQuery (16 נקודות)

בסעיפים הבאים נתייחס לייצוג העצי של מסמך XML.

א. (4 נקודות) עץ ייקרא בינארי מלא אם לכל צומת בעץ, מלבד השורש, יש שניים או אפס בנים. כתוב שאילתא ב-XPath שבודקת אם העץ הוא בינארי מלא. אם כן, השאילתא תחזיר קבוצה ריקה אחרת תוחזר קבוצה לא ריקה.

```
//*[count(*)!=2 and count(*)!=0]
```

ב. (12 נקודות) עץ ייקרא בינארי שלם אם הוא בינארי מלא, וכל העלים נמצאים במרחק שווה מהשורש. כתוב שאילתא ב-XQuery שבודקת אם העץ שמיוצג ע"י המסמך tree.xml הוא בינארי שלם. אם כן, השאילתא תחזיר קבוצה ריקה, אחרת תוחזר קבוצה לא ריקה.

```
xquery version "1.0";

let $root := doc("tree.xml")/*
return
if ($root/descendant-or-self::*[count(*)!=2 and count(*)!=0]) then
  <false/>
else (
  let $depth := count($root/descendant-or-self::*[last()]/ancestor::*)
  for $v in $root//*[not(*)]
  return
  if (count($v/ancestor::*)!= $depth) then
    <false/>
  else ()
)
```

שאלה 5 – DATALOG (16 נקודות)

בשאלה זו נעסוק במערכת ייצור. במערכת שני סוגי חלקים – חלקים בסיסיים וחלקים מורכבים. לכל חלק מסוג בסיסי יש מחיר. מחירו של חלק מסוג מורכב שווה לסכום המחירים של כל החלקים הבסיסיים המרכיבים אותו, רקורסיבית.

פרדיקט ה- $\text{basic}(\text{ID}, \text{COST})$ מתאר חלקים בסיסיים. לכל חלק מסוג בסיסי ישנו מציין ייחודי (ID) ומחיר (COST).

פרדיקט ה- $\text{composite}(\text{ID}, \text{S1}, \text{S2}, \text{S3})$ מתאר חלקים מורכבים. לכל חלק מסוג מורכב ישנו מציין ייחודי (ID) ושלושה תתי-חלקים מהם הוא מורכב. כל אחד מתתי-החלקים יכול להיות מסוג בסיסי או מסוג מורכב.

דוגמא 1:

נניח שהמסד (EDB) מורכב מהעובדות הבאות:

$\text{basic}(a, 5)$.
 $\text{basic}(b, 10)$.
 $\text{composite}(c, a, a, b)$.
 $\text{composite}(d, c, c, b)$.

אז מחיר חלק c הנו 20 ומחיר חלק d הנו 50.

א. (10 נקודות) כתוב פרדיקט בשם $\text{bad}(X)$ המחזיר חלקים המכילים (באופן ישיר או עקיף) את עצמם. לדוגמא, אם נשנה את ה-EDB בדוגמא 1 ע"י הוספת העובדות: $\text{composite}(f, c, e, b)$. $\text{composite}(e, c, b, f)$. אזי $\text{bad}(X)$ יהא $\{e, f\}$.

$\text{bad}(X) \leftarrow \text{containedin}(X, X)$.

$\text{containedin}(X, Y) \leftarrow \text{containedin}(X, Z), \text{containedin}(Z, Y)$. /*X is contained in Y */

$\text{containedin}(X, Y) \leftarrow \text{composite}(Y, X, U, V)$.

$\text{containedin}(X, Y) \leftarrow \text{composite}(Y, U, X, V)$.

$\text{containedin}(X, Y) \leftarrow \text{composite}(Y, U, V, X)$.

ב. (6 נקודות) כתוב פרדיקט בשם $\text{total}(\text{Part}, \text{Cost})$ המחשב, בהינתן ID של חלק (בסיסי או מורכב) את מחירו. בדוגמא הקודמת, אחת העובדות המחושבות הנה: $\text{total}(d, 50)$.

ניתן להשתמש בפרדיקט הבנוי לשפה +, לדוגמא: $+(2, 5, 7)$.

$\text{total}(\text{Part}, \text{Cost}) \leftarrow \text{basic}(\text{Part}, \text{Cost})$.

$\text{total}(\text{Part}, \text{Cost}) \leftarrow \text{composite}(\text{Part}, U, V, W), \text{total}(U, C1), \text{total}(V, C2),$
 $\text{total}(W, C3), +(C1, C2, T1), +(T1, C3, \text{Cost})$.
