



הטכניון - מכון טכנולוגי לישראל
הפקולטה למדעי המחשב

ד"ר אלדר פישר
סעאב מנסור
לינה זריבץ'

חורף תשס"ז
18, פברואר 2007

מערכות מסדי נתונים – 236363
מועד א'

הזמן: 3 שעות
במבחן זה 10 עמודים

שאלה	נקודות
שאלה 1 – ERD	12
שאלה 2 – שפות שאלתא	41
שאלה 3 – Design	26
שאלה 4 – XML	21
סה"כ	100

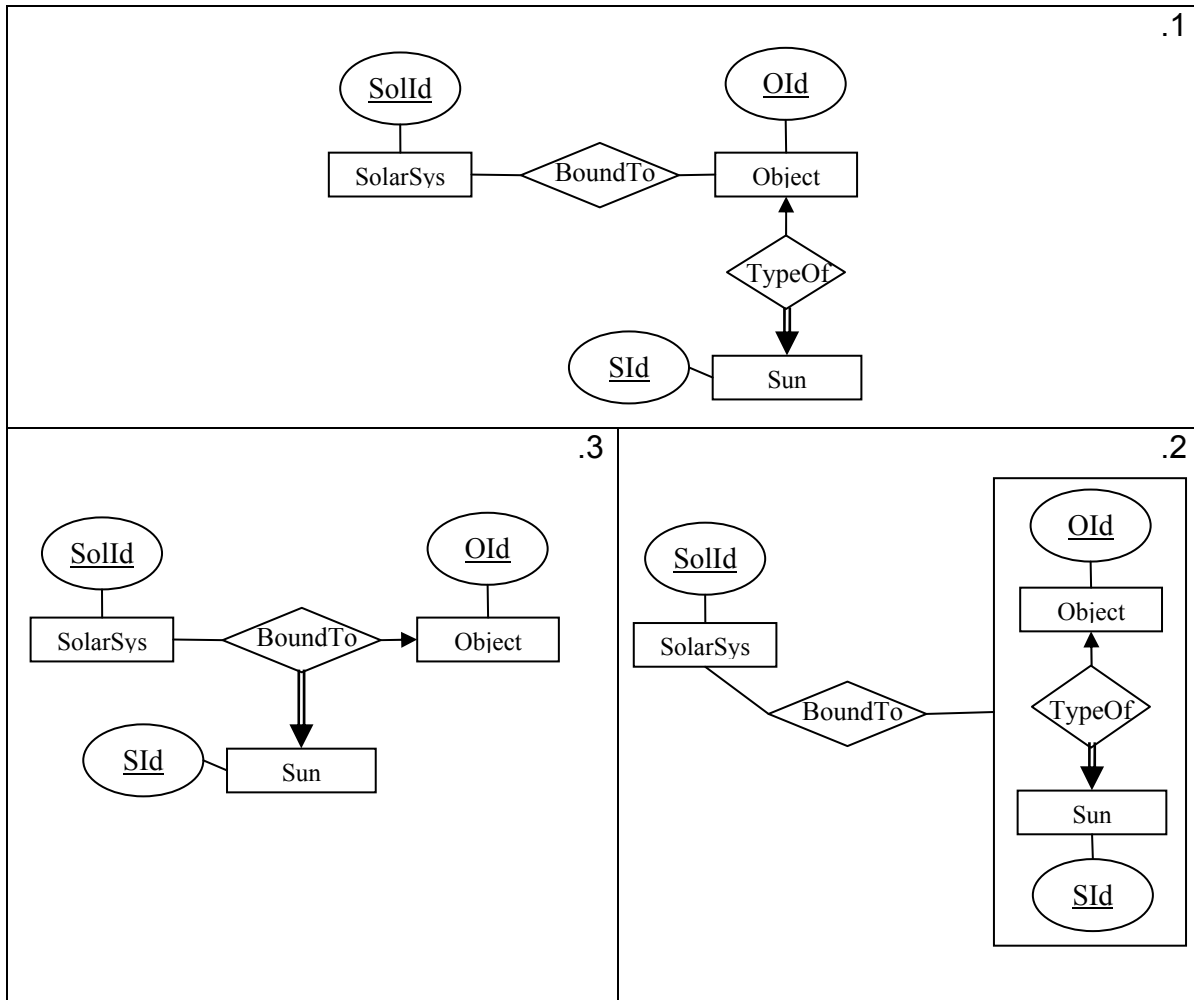
הנחיות:

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

בהצלחה

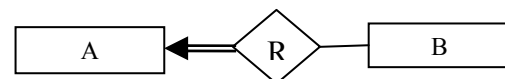
שאלה 1 – ERD (12 נק')

נתונות שלוש סכמות ה-ERD הבאות. הסכמות מנסות (לא תמיד בהצלחה) לתאר מסד נתונים שמכיל גרמי שמיים (Object), שמשות (Sun) שהן סוג של גרמי שמיים, והימצאות של גרמי שמיים במערכות שמש (SolarSys):



כתבו האם כל תוכן שניתן לייצוג בסכמה מספר 1, ניתן גם לייצוג בסכמה מספר 2 ומספר 3 ולהפך. אם התשובה היא לא, תארו דוגמה לתוכן שניתן לייצוג בסכמה אחת ולא ניתן לייצוג בסכמה השנייה.

משמעות הסימון



היא גם קו כפול בין A ו-R, וגם חץ בין R ו-A.

שימו לב שלא ניתן להשתמש בערכי NULL בטבלאות המתוארות ע"י ה-ERD.

סכמות	האם כל מסד שניתן לייצוג בסכמה א', ניתן לייצוג בסכמה ב'?	האם כל מסד שניתן לייצוג בסכמה א', ניתן לייצוג בסכמה ב'?
א': סכמה 1 ב': סכמה 2	לא. בסכמה 2 רק שמשות מקושרות למערכות שמש. למשל בסכמה 1 ניתן לבנות את המסד: SolarSys={so1}, Object={o1,o2}, Sun={s1}, TypeOf={(o1,s1)}, BoundTo={(o2,so1)} אך בסכמה 2, אם נרצה ש-o2 יהיה ב-BoundTo, הוא יצטרך להיות בקשר עם שמש כלשהי בגלל ההקבצה.	כן
א': סכמה 1 ב': סכמה 3	לא, למשל בסכמה 1 שמש לא חייבת להיות ביחס עם מערכת שמש, אבל בסכמה 3 היא תמיד נמצאת. לא, למשל בסכמה 3 שמשות שונות יכולות להיות ביחס עם אותו גרם שמימי, ע"י שינוי מערכת השמש, בסכמה 1 זה לא אפשרי. למשל בסכמה 3: SolarSys={so1}, Object={o1,o2}, Sun={s1,s2}, BoundTo={(o1,s1,so1), (o1,s2,so2)} בסכמה 1, צריך ש: TypeOf = {(o1,s1),(o1,s2)} אך זה מפר את התלות ש- Old→Sld	

שאלה 2 – שפות שאילתא (41 נק')

נתונות הסכמות הבאות:

Sun(SId, SName, SAstr)
Planet(PId, PName, PAstr)
SolSys(SId, PId)

הסכמה Sun מתארת את השמשות כאשר SId מייצג את המספר הקטלוגי, SName הוא השם ו-SAstr הוא שם האסטרונום המגלה.

הסכמה Planet מתארת כוכבי לכת ומשמעות השדות אנלוגית לאלו של סכמה Sun. הסכמה SolSys מתארת מערכות שמש, כלומר שייכות של כוכבי לכת לשמשות.

ניתן להניח שכל גרם שמימי (שמש או כוכב לכת) המופיע ב-SolSys, מופיע גם ברציה המתאימה (Sun או Planet).

א. (9 נק') כתבו שאילתא ב-SQL, אשר בודקת אם מתקיים שלכל אסטרונום שגילה גרם שמימי, הוא גם גילה את כל מערכת השמש שהגרם השמימי הזה שייך אליה. אם התנאי מתקיים יש להחזיר תוצאה ריקה, אחרת החזירו זוגות של אסטרונום ומערכת שמש שגילה חלק ממנה. ניתן להשתמש במבט עזר אחד לכל היותר.

```
CREATE VIEW Discovery as
SELECT DISTINCT mOld, Astr, SId
FROM (SELECT SId AS mOld, SAstr AS Astr FROM Sun
      UNION
      SELECT PId AS mOld, PAstr AS Astr FROM Planet ) H
,SolSys S
WHERE mOld = S.PId OR mOld=S.SId

SELECT Astr, Sid
FROM Discovery D
GROUP BY Astr, SId
HAVING COUNT(mOld) <> (SELECT COUNT(PId)+1 FROM SolSys
                        WHERE D.SId=SId)
```

עבור הסעיפים הבאים, נניח כי SId ו-PId הם מטיפוס integer וכי הערכים עבור שדות אלו ניתנו בהתאם לסדר כרונולוגי של גילוי שמשות וכוכבי לכת. בפרט, השמש/כוכב הלכת שהתגלה ראשון הוא בעל המספר הקטלוגי המינימאלי. עם זאת, לא מובטח שהערכים הנ"ל עבור שני גילויים עוקבים אכן נבדלים ביניהם ב-1.

ב. (7 נק') בהינתן אסטרונום A נגיד שכוכבי לכת PId1 ו-PId2 השייכים לאותה מערכת שמש הינם גילויים עוקבים עבור A אם A גילה את שניהם וגם אחד מהכוכבים התגלה מיד אחרי הכוכב השני ללא גילוי ביניים אחרים של אותו האסטרונום. כתבו שאילתא ב-RA אשר מחזירה רלציה כך שלכל רשומה (pid1, pid2, astr) בה מתקיים כי pid1 ו-pid2 הם גילויים עוקבים של אסטרונום astr. נסמן רלציה זאת ב-SeqDisc(PId1, PId2, Astr).
 רמז - לשם פישוט הפתרון השתמשו ברלציה המוגדרת כדלקמן:

$$PAS(PId, PAstr, SId) = \pi_{PId, PAstr, SId} (Planet \bowtie SolSys)$$

הפתרון נתון ע"י הביטוי הבא:

$$\pi_{PId1, PId2, PAstr} \left(\sigma_{PId1 < PId2} \left(\left(\rho_{PId \rightarrow PId1} PAS \right) \bowtie \left(\rho_{PId \rightarrow PId2} PAS \right) \right) \right) \setminus \\ \pi_{PId1, PId2, PAstr} \left(\sigma_{PId1 < PId3 \wedge PId3 < PId2} \left(\left(\rho_{PId \rightarrow PId1, SId \rightarrow SId1} PAS \right) \bowtie \left(\rho_{PId \rightarrow PId2, SId \rightarrow SId2} PAS \right) \right) \bowtie \left(\rho_{PId \rightarrow PId3, SId \rightarrow SId3} PAS \right) \right)$$

ג. (7 נק') נגיד שאסטרונום a1 הוא מתחרה של a2 אם a2 היה ל-a2 לפחות זוג אחד של גילויים עוקבים של כוכבי לכת, וכן בין כל שני גילויים עוקבים של כוכבי לכת של a2 השייכים לאותה מערכת שמש יש ל-a1 גם גילוי של כוכב לכת השייך לאותה מערכת שמש.
 בנוסף לרלציות המקוריות, השתמשו ברלציה SeqDisc על מנת לממש שאילתא ב-DRC אשר מחזירה את כל זוגות האסטרונומים (a1, a2) אשר לגביהם מתקיים כי a1 הינו מתחרה של a2.

$$\{ \langle A1, A2 \rangle \mid \exists PId1, PId2 \text{ SeqDisc}(PId1, PId2, A2) \wedge \\ \forall PId', PId'', SId \left(\left(SolSys(SId, PId') \wedge SeqDisc(PId', PId'', A2) \right) \rightarrow \right. \\ \left. \exists PId3, PN3 \left(Planet(PId3, PN3, A1) \wedge SolSys(SId, PId3) \wedge (PId' < PId3) \wedge (PId3 < PId'') \right) \right) \}$$

ד. (6 נק') נתונה תוכנית Datalog המשתמשת בפרדיקטי sun, planet EDB ו-solsys ומשמעותם וסדר המשתנים בהם הוא בהתאם לסכמות שהוגדרו בתחילת השאלה.

$\text{not_x}(S, P) \leftarrow \text{solsys}(S, P), \text{solsys}(S, P'), (P < P')$.
 $x(S, P) \leftarrow \text{solsys}(S, P), \neg \text{not_x}(S, P)$.
 $\text{query}(A1, A2) \leftarrow \text{sun}(S, SN, A1), \text{planet}(P, PN, A2), x(S, P)$.

כתבו במילים פשוטות מה מחזירה התוצאה של השאילתא $\text{query}(X, Y)$?

השאילתא מחזירה עובדות $\text{query}(A1, A2)$ עבורן מתקיים ש- $A1$ ו- $A2$ הינם

אסטרונומים וגם קיימת שמש sid כך ש- $A1$ גילה אותה ואילו $A2$ גילה את כוכב

הלכת המאוחר ביותר במערכת השמש של sid.

הסעיפים הבאים הינם בלתי תלויים בסעיפים הקודמים.

ה. (12 נק') נתון ש- R היא רלציה עם התכונות A, B, C כאשר A, B מפתח ויש לה $n > 1$ שורות, וכן ש- S היא רלציה עם התכונות A, B, D כאשר D מפתח ויש לה $m > 1$ שורות.

בכל אחד מהסעיפים הבאים כתבו מה מספר השורות המינימלי שיכול להיות ומספר השורות המקסימלי שיכול להיות בכל אחת מתוצאות הביטויים הרלציונים הבאים, ונמקו בקצרה.

ביטוי	מינימום שורות	מקסימום שורות	נימוק קצר
$\pi_{A,B} R$	n	n	A, B מפתח ולכן לכל שורה ב- R יהיה ערך יחודי עבור ההטלה על A, B ומכאן שלא יהיו כפילויות למחוק בהטלה.
$R \bowtie S$	0	m	כל שורה ב- S יכולה להתאים לשורה אחת לכל היותר בצירוף (כי הערך המתאים על A, B יכול להיות לכל היותר בשורה אחת ב- R), אבל יתכנו גם שורות ב- S שאין עבורן שורות בצירוף בכלל.
$S \div \pi_{A,B} R$	0	0	נתון ש- $n > 1$ אשר זהו מספר השורות בהטלה של R . מצד שני, D מפתח ב- S ולכן עבור ערך נתון של D לא יתכנו שיש שורות המסכימות איתו ועם כל שורה בהטלה של R (כי לכל ערך כזה יש רק שורה אחת ב- S).

שאלה 3 – Design (26 נק')

א. (17 נק') נתונה הסכמה $R(A,B,C,D,E,G,H)$ וקבוצת התלויות הפונקציונליות

$$F = \{ DE \rightarrow H, BC \rightarrow A, ABH \rightarrow G, A \rightarrow DEHC, E \rightarrow H, AB \rightarrow GH \}$$

1. (6 נק') מצאו כיסוי מינימאלי של F .

$$F_c = \{ BC \rightarrow A, A \rightarrow D, A \rightarrow E, A \rightarrow C, E \rightarrow H, AB \rightarrow G \}$$

2. (4 נק') רשמו את כל המפתחות הקבילים של R, F .

$$BC, AB$$

3. (7 נק') מצאו פירוק ב-3NF המשמר מידע ותלויות המכיל מספר קטן ככל האפשר של תתי-סכמות.

$$ADEC, EH, ABCG$$

הסעיפים הבאים הינם בלתי תלויים בסעיף א'.

ב. (9 נק') בשני הסעיפים הבאים הניחו ש- Γ היא רלציה עם קבוצת התכונות U , ש- V היא תת קבוצה של U וש- X ו- Y הן תתי קבוצות זרות של V שאיחודן לא שווה ל- V .
1. (4 נק') אם $\pi_V \Gamma$ מקיים את $X \rightarrow Y$, האם גם Γ מקיים את $X \rightarrow Y$? הוכיחו אם כן, כתבו דוגמא נגדית אם לא.

כן: אחרת, אם היו ברלציה שתי שורות המסכימות על X ולא מסכימות על Y ,

אז ההטלות המתאימות שלהן על U (שמכילה את את X ו- Y) גם היו מסכימות על X

ולא מסכימות על Y , סתירה להנחה שההטלה מקיימת את התלות.

2. (5 נק') אם $\pi_V \Gamma$ מקיים את $X \rightarrow Y$, האם גם Γ מקיים את $X \rightarrow Y$? הוכיחו אם כן, כתבו דוגמא נגדית אם לא.

דוגמא נגדית: ניקח רלציה עם העמודות "חיה צבע התנהגות", ושלוש שורות בהתאמה:

"סוס לבן צוהל", "סוס לבן דוהר", "חתול שחור מגרגר".

ההטלה על "צבע" ו-"התנהגות" היא בעלת תלות רב ערכית מ-"צבע" ל-"התנהגות"

(זה קורה בכל רלציה עם שתי תכונות) אולם ברלציה המקורית אין תלות כזו.

שאלה 4 – XML (21 נק')

נתון ה-DTD הבא עבור קטלוג של גרמי שמיים (צומת המסמך הוא catalog):

```
<!ELEMENT catalog (sun)*>
<!ELEMENT sun (name,mass,obs,planet*)>
<!ATTLIST sun cat ID #REQUIRED type CDATA #REQUIRED>
<!ELEMENT planet (name,mass,obs)>
<!ATTLIST planet cat ID #REQUIRED>
<!ELEMENT name (#PCDATA)>
<!ELEMENT mass EMPTY>
<!ATTLIST mass min CDATA #IMPLIED max CDATA #IMPLIED>
<!ELEMENT obs (#PCDATA)>
```

הערות על החלקים הפחות מובנים מאליהם:

- תכונות cat אמורות להכיל את המספר הקטלוגי של גרם השמיים.
- אלמנט mass מכיל חסמים עליונים ותחתונים ידועים על המסה. אם אחד מהם (או שניהם) חסר אז החסם הנ"ל אינו ידוע. נניח שיחידות המסה הן כאלה ש-"1" פירושו 10^{24} קילוגרם.
- אלמנט obs מכיל את שם האסטרונום שגילה את גרם השמיים. לכל גרם שמיים שומרים את שם המגלה, אין כזה דבר "מגלה של מערכת שמש שלמה".

א. (6 נק') כתבו שאילתת **XPath 1.0** אשר מחזיר את כל כוכבי הלכת אשר לפי מה שידוע עליהם יכולה להיות להם מסה של "12". למשל, כוכב עם מינימום של 11 ומקסימום של 13 הוא כוכב כזה, אבל גם כוכב ללא נתוני מסה בכלל הוא כזה.

הדרך הכי קצרה לכתוב את השאילתה (ע"י שימוש בהתנהגות של השוואות עם קבוצות ריקות) היא הדרך הבאה:

```
//planet[not(mass/@min>12) and not(mass/@max<12)]
```

ב. (8 נק') כתבו שאילתת **XQuery** אשר מחזירה רשימה של אסטרונומים וגרמי השמיים שהם גילו. לצורך זה אפשר להניח ששם קובץ המקור הוא cat.xml. על רשימת הפלט להיות מסמך XML המציית ל-DTD הבא (עם צומת מסמך list).

```
<!ELEMENT list (obs)*>
<!ELEMENT obs (name,sun*,planet*)>
<!ELEMENT name (#PCDATA)>
<!ELEMENT sun EMPTY>
<!ATTLIST sun cat ID #REQUIRED>
<!ELEMENT planet EMPTY>
<!ATTLIST planet cat ID #REQUIRED>
```

```
<list>{
  for $a in distinct-values(doc("cat.xml")//obs)
  let $ss := doc("cat.xml")//sun[obs=$a],
      $ps := doc("cat.xml")//planet[obs=$a]
  return
  <obs>{
    <name>{$a}</name>,
    (for $s in $ss return <sun cat="{ $s/@cat }"/>),
    (for $p in $ps return <planet cat="{ $p/@cat }"/>)
  }</obs>
}</list>
```

הערה: הביטוי

```
<sun>{$s/@cat}</sun>
```

שקול ל-

```
<sun cat="{ $s/@cat }"/>
```

ג. (7 נק') מסתבר עתה שיש גם מערכות שמש בינאריות, שבהן יש שתי שמשות יחד עם כוכבי לכת משותפים. תקנו עתה את ה-DTD של קטלוג גרמי השמיים כך שיהיה מסודר לפי מערכות שמש המכילות שמש אחת או שתיים כל אחת.

לנוחותכם, מצורף להלן ה-DTD המקורי. מותר להשאיר או לתקן ו/או למחוק שורות ישנות, וכן מותר להוסיף שורות חדשות.

```
<!ELEMENT catalog (sun)*>
<!ELEMENT catalog (solarsys)*>
<!ELEMENT solarsys (sun, sun?, planet*)>

<!ELEMENT sun (name,mass,obs,planet*)>
<!ELEMENT sun (name,mass,obs)>

<!ATTLIST sun cat ID #REQUIRED type CDATA #REQUIRED>

<!ELEMENT planet (name,mass,obs)>

<!ATTLIST planet cat ID #REQUIRED>

<!ELEMENT name (#PCDATA)>

<!ELEMENT mass EMPTY>

<!ATTLIST mass min CDATA #IMPLIED max CDATA #IMPLIED>

<!ELEMENT obs (#PCDATA)>
```