



הטכניון - מכון טכנולוגי לישראל
הפקולטה למדעי המחשב

דר' אמיר שפילקה
סעאב מנסור
ארינה לווה

אביב תשס"ז
5, אוקטובר 2007

מערכות מסדי נתונים – 236363

מועד א'
הזמן: 3 שעות
במבחן זה 13 עמודים

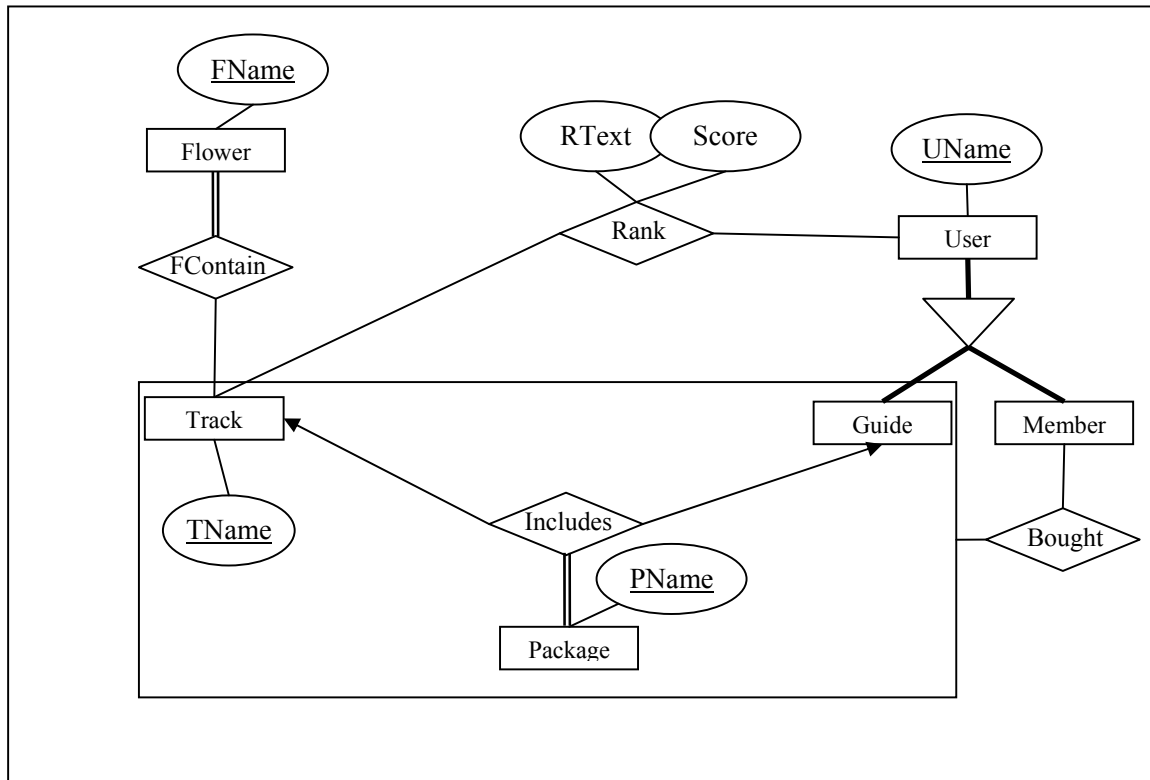
שאלה	נקודות
שאלה 1 – ERD	14
שאלה 2 – שפות שאילתא	32
שאלה 3 – Design	23
שאלה 4 – XML	21
שאלה 5 – DATALOG	10
סה"כ	100

הנחיות:

- 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 ואת ההסברים לסכמות.
 - 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 – ERD (14 נק')

נתונה סכמת ה-ERD הבאה שמתארת מסד נתונים של אתר מסלולי טיולים:



הסבר לסכמה:

ישויות:

User – ישות המתארת משתמשים במערכת. לכל משתמש יש שם ייחודי (UName). קיימים שני סוגים של משתמשים:

- **Member** – מתאר חברים רשומים באתר.
- **Guide** – מתאר מדריכים רשומים באתר.

Track – ישות המתארת מסלולים. לכל מסלול נשמר שם ייחודי (TName).

Flower – מתארת פרחים שיכולים להימצא במסלולים. לכל פרח נשמר שם ייחודי (FName).

Package – ישות המתארת חבילות טיולים. לכל חבילת טיול נשמר שם ייחודי (PName).

יחסים:

Rank – יחס המתאר מתן דירוג למסלול ע"י משתמש. הדירוג כולל טקסט (RText) וציון (Score).

FContains – יחס המתאר הימצאות של פרח במסלול.

Includes – יחס המתאר שחבילה מכילה מדריך ומסלול.

א. (6 נק') מלאו את הטבלה הבאה שמתייחסת לשדות ולמפתחות (הקבילים) של יחסים בסכמת ה-ERD. שימו לב כי אם ישנם מספר אפשרויות עבור המפתחות, יש לרשום את כל האפשרויות.

Rank	
שדות	UName, TName, RText, Score
מפתחות	UName, TName
Includes	
שדות	GUName, PName, TName
Bought	
מפתחות	{PName, TName, MUName} או {PName, GUName, MUName}

ב. (8 נק') בניח שנתון כי מספר הישויות מסוג Package הוא n_p , מסוג Track הוא n_t ומסוג Guide הוא n_g , כמו כן נתון כי $n_p, n_t, n_g > 0$. תנו חסמי מינימום ומקסימום על מספר הקשרים מסוג Includes. נמקו בקצרה.

מינימום: n_p

מקסימום: $\min(n_p \cdot n_t, n_p \cdot n_g)$

שאלה 2 – שפות שאילתא (32 נק')

הערה: בסעיפים הבאים הניחו תרגום סטנדרטי של סכמת ה-ERD משאלה 1 לסכמות רלציוניות: User, Member, Guide, Track, Flower, Package, Rank, FContains, Includes.

א. (8 נק') נגדיר דירוג של מסלול כממוצע כל הציונים שניתנו לאותו מסלול ע"י משתמשי המערכת. נגדיר דירוג של פרח, כממוצע הדירוגים (Score) של כל המסלולים בהם הפרח נמצא. כתבו שאילתא ב-SQL אשר מחזירה את הפרחים בעלי הדירוג הגבוה ביותר. ניתן להשתמש בשני מבטים לכל היותר. ניתן להניח כי לכל מסלול יש דירוג אחד לפחות.

```
CREATE VIEW ARank AS
SELECT TName,AVG(Score) AS AR
FROM Rank
GROUP BY TName

CREATE VIEW Help AS
SELECT FName, AVG(AR) AS AvgSc
FROM FContains F, ARank R
WHERE F.TName=R.TName
GROUP BY FName

SELECT FName FROM Help
WHERE AvgSc = SELECT MAX(AvgSc) FROM Help
```

- ב. (8 נק') כתבו שאילתא ב-RA אשר מחזירה את המסלולים, כך שקיים עבורם איזשהו מדריך שמקיים את כל הדרישות הבאות:
- המדריך נתן למסלול את הציון הגבוה ביותר מבין הציונים שניתנו ע"י אותו המדריך
 - המדריך דירג לפחות שלושה מסלולים
 - לא כל המסלולים המדורגים על ידי המדריך קיבלו את אותו הציון.
- ניתן להיעזר בחישובי ביניים, כל עוד מסבירים אותם.

$$A_1(\text{Guide}, \text{Tn}, \text{Scr}) = \pi_{(\text{Uname}, \text{Tname}, \text{score})} \text{Guide} \bowtie \text{Rank}$$

$$A_2(\text{Guide1}, \text{Guide2}, \text{Guide3}, \text{Tn1}, \text{Tn2}, \text{Tn3}, \text{Scr1}, \text{Scr2}, \text{Scr3}) =$$

$$(A_1 \bowtie A_1 \bowtie A_1)_{(\text{same Guide AND Tn1, Tn2, Tn3 are different AND Scr1} > \text{Scr2})}$$

$$A_3(\text{Guide}, \text{Track}, \text{Score}) = \pi_{(\text{Guide1}, \text{Tn1}, \text{Scr1})} A_2$$

$$A_4(\text{Guide}, \text{Track}, \text{Score}) = \pi_{(\text{Guide1}, \text{Tn2}, \text{Scr2})} A_2$$

$$A_5(\text{Guide}, \text{Track}, \text{Score}) = A_3 \setminus A_4$$

$$\text{Answer (Track)} = \pi_{\text{Track}} A_5$$

ג. (8 נק') נאמר כי משתמש (User) אוהב פרח מסוים אם הוא דירג את כל המסלולים בהם הפרח מופיע עם ציון גדול מ-8. נגיד שפרח נמצא בחבילה אם הפרח נמצא במסלול שכלול בחבילה. כתבו שאילתא ב-DRC שמחזירה לכל משתמש את כל חבילות הטיולים בהם יש רק פרחים שהמשתמש אוהב. (ניתן להיעזר בחישובי ביניים). הניחו כי הסדר של התכונות ברלציות הינו אלפביתי.

פרחים שמשתמש אוהב:

$A_1(\text{UName}, \text{FName}) = \{ \langle \text{UName}, \text{FName} \rangle : \forall \text{Trk} [\text{FContain}(\text{Trk}, \text{FName}) \Rightarrow (\text{Rank}(\text{UName}, \text{Trk}, \text{Score}, _) \wedge \text{Score} > 8)] \}$
פרחים שנמצאים במסלול שנמצא בחבילה:

$A_2(\text{PName}, \text{Fname}) = \{ \langle \text{PName}, \text{FName} \rangle : \exists \text{Trk} [\text{FContain}(\text{Trk}, \text{FName}) \text{ AND Includes}(\text{PName}, \text{Trk}, _)] \}$

תשובה:

$\text{Answer}(\text{UName}, \text{PName}) = \{ \langle \text{UName}, \text{PName} \rangle : \forall \text{Fname} [A_2(\text{PName}, \text{Fname}) \Rightarrow A_1(\text{UName}, \text{FName})] \}$

הסעיף הבא הינו בלתי תלוי בסעיפים הקודמים.

ד. (8 נק') הרלציה $Temp(rID, name, \underline{date}, high, low)$ מתארת אזור בעל מזהה rID , שם אזור $name$, וטמפרטורות מקסימאליות ומינימאליות בתאריכים שונים (שימו לב כי המפתח לרלציה הוא $\{rID, date\}$). הסבירו במלים פשוטות מה מחושב בשאילתא הבאה (אין להתייחס לרלציות העזר שחושבו במהלך השאילתא).

$$\begin{aligned} T_1(rID, date, high) &:= \pi_{rID, date, high}(Temp) \\ T_2(rID, date, low) &:= \pi_{rID, date, low}(Temp) \\ T_3(rID) &:= \rho_{rID1 \rightarrow rID} \pi_{rID1}(T_1 \bowtie high1 < high2 Temp) \\ T_4(rID) &:= \rho_{rID1 \rightarrow rID} \pi_{rID1}(T_2 \bowtie low1 > low2 Temp) \\ T_5(rID) &:= \pi_{rID}(Temp) \setminus T_3 \\ T_6(rID) &:= \pi_{rID}(Temp) \setminus T_4 \\ Answer(name) &:= \pi_{name}(Temp \bowtie (T_5 \cup T_6)) \end{aligned}$$

שמות האזורים שהשיגו (מתישהו) את הטמפרטורה הגבוהה ביותר או הנמוכה ביותר.

שאלה 3 – תלויות וצורות נורמליות (23 נק')

א. (4 נק') אוסף תכונות X הוא סגור, בהינתן F , אם $X = X_F^+$. נניח כי $R(A, B, C)$ היא סכמה ו- F אוסף תלויות כך שתתי הקבוצות הסגורות היחידות ב- R הן ϕ (הקבוצה הריקה) ו- $\{A, B, C\}$. הוכיחו כי התלות $A \rightarrow B$ נובעת מ- F .

אם לא מתקיימת התלות $A \rightarrow B$ אז B לא בסגור של A , ולכן הסגור של A (שהוא קבוצה סגורה)

אינו הקבוצה הריקה (כי A בתוכו) ואינו הכל (לא מכיל את B).

ב. (8 נק') הוכיחו כי מערכת כללי ההיסק הבאים היא נאותה ושלמה:

אם $Y \subseteq X$ אז $X \rightarrow Y$

אם $X \rightarrow Y$ אז לכל Z מתקיים $XZ \rightarrow YZ$

אם $X \rightarrow Y$ ו- $YZ \rightarrow W$ אז $XZ \rightarrow W$

נראה נאותות ע"י כך שנסיק את התכונה השלישית מאקסיומות ארמסטרונג, ונראה שלמות ע"י

כך שנגרור את אקסיומות ארמסטרונג (נשים לב ששתי התכונות הראשונות משותפות). ואכן, אם

$X \rightarrow Y$ ו- $YZ \rightarrow W$ אזי מתקיים $XZ \rightarrow YZ$ ומטרנזיטיביות (ארמסטרונג) נקבל $XZ \rightarrow W$. בכיוון

השני, כאשר Z הקבוצה הריקה אנו מקבלים את תכונת הטרנזיטיביות.

ג. (11 נק') נתון היחס $R(A,B,C)$ עם התלויות $A \rightarrow B$ ו- $B \rightarrow C$.

1. (3 נק') הראו כי היחס אינו ב- BCNF.

B אינו מפתח, אך התלות $B \rightarrow C$ מופיעה.

2. (3 נק') הוסיפו תלות נוספת כך שהיחס יהיה ב-BCNF.

אם נוסיף את התלות $C \rightarrow A$ אז כל אחד מהשדות הוא מפתח ולכן הפרוק ב-BCNF.

3. (5 נק') האם ניתן להוסיף תלות נוספת (לשתי התלויות המקוריות) כך שיתקבל יחס ב-

3NF שאינו ב-BCNF? (הוכיחו את טענתכם).

לא ניתן. נראה זאת ע"י כך שנעבור על כל התלויות שאינן נובעות מהתלויות הקיימות.

$B \rightarrow A$ או $C \rightarrow A$, כמו קודם נשים לב שנקבל פירוק ב-BCNF כי בכל מקרה A ו-B מפתחות.

$C \rightarrow B$, נותן פירוק שאינו 3NF כי B ו-C לא חלק ממפתח קביל.

$BC \rightarrow A$ גורר ש- $B \rightarrow A$ והפירוק ב-BCNF.

שאלה 4 – XML (21 נק')

נתון קטע ה-DTD הבא:

```
<!ELEMENT db (track|reco)*>
<!ELEMENT track (start,end,track*)>
<!ATTLIST track tid ID #REQUIRED>
<!ELEMENT reco (user,rtext,score)>
<!ATTLIST reco tid IDREF #REQUIRED>
<!ELEMENT start (#PCDATA)>
<!ELEMENT end (#PCDATA)>
<!ELEMENT user (#PCDATA)>
<!ELEMENT rtext (#PCDATA)>
<!ELEMENT score (#PCDATA)>
```

כאשר track מסמל מסלולי-טיול שכוללים מיקום התחלה, סיום, ומסלולי טיול שמרכיבים את המסלול. reco מסמל המלצות על מסלולים ע"י משתמש, עם ציון (score) וטקסט (rtext).

א. (5 נק') ב-DTD הנוכחי לא ניתן להשתמש באותו מסלול-טיול עם מזהה (tid) מסוים להרכבת מספר מסלולי-טיול (אותו tid שמזהה מסלול, לא יכול להופיע במספר מסלולים שונים). הצע שינוי ל-DTD שיאפשר זאת.

נשנה את track כך שיכיל התחלה וסיום ונוריד את track*, ונוסיף תכונת IDREFS מסוג

IMPLIED עבור המסלולים שמרכיבים אותו.

ב. (8 נק') כתבו שאילתא ב-XPATH 1.0 אשר בודקת לכל מסלול, אם תת המסלולים המרכיבים אותו באופן ישיר מתחברים למסלול עצמו. ז"א תת-המסלול הראשון מתחיל באותו מקום כמו מסלול האב, תת-המסלול האחרון מסתיים באותו מקום, ומקום סיום בן כלשהו יהיה התחלת הבן שאחריו. השאילתא תחזיר את כל המסלולים שעונים על דרישה זו. שימו לב כי מסלולים בסיסיים (מסלולים שלא מכילים תת-מסלולים) מקיימים את הדרישה.

```
//track[start = ./track[1]/start and end = ./track[last()]/end and not(./track[end !=  
following-sibling::*[1]/start]) ] | //track[not(track)]
```

ג. (8 נק') כתבו שאילתא ב-XQUERY, המחזירה את מסלולי האב (מסלולי-טיול שאינם בנים של מסלול-טיול) ממוינים בסדר יורד לפי ממוצע ציוניהם. שימו לב כי מסלול-טיול ללא דירוג ייחשב כבעל דירוג 0. הניחו כי שם מסמך העבודה הוא "1.xml".

```
<res>
{
for $t in doc("1.xml")/db/track
let $sum := sum(doc("1.xml")//reco[@tid=$t/@tid]/score),
    $count := count(doc("1.xml")//reco[@tid=$t/@tid])
order by (if ($count!=0) then $sum/$count else 0) descending
return <track>{$t/@tid}</track>
}
</res>
```

שאלה 5 – DATALOG (10 נק')

בהינתן ה-EDB ים $\text{track}(\text{Name}, \text{Start}, \text{End})$ ו- $\text{rank}(\text{Name}, \text{Score})$ (הרלציה rank מכילה לכל מסלול-טיול את הציון שלו). נגדיר:

- **מסלול** הוא שילוב של מסלולי-טיול כך שסיום של מסלול-טיול אחד הוא ההתחלה של הבא אחריו.
- **מסלול** הוא מעניין אם הוא מכיל לכל היותר מסלול-טיול אחד עם ציון קטן מ-5.

כתבו IDB $\text{shvil}(\text{Name})$ שיכיל את כל מסלולי-הטיול שנמצאים על **מסלול מעניין** כלשהו המחובר בין HERMON ל- EILAT. שימו לב שניתן ללכת על מסלול-טיול רק מ-Start ל-End.

```
path0(X,Y) ← track(N,X,Y), rank(N,S), S>=5.  
path0(X,Y) ← track(N,X,Z), rank(N,S), S>=5, path0(Z,Y).  
path1(X,Y) ← path0(X,Y).  
path1(X,Y) ← track(N,X,Y), rank(N,S), S<5.  
path1(X,Y) ← track(N,X,Z), rank(N,S), S>=5, path1(Z,Y).  
path1(X,Y) ← track(N,X,Z), rank(N,S), S<5, path0(Z,Y).  
  
shvil(Name) ← track(Name,X,Y), rank(Name,S), S>=5, path0("HERMON",X),  
path1(Y,"EILAT").  
shvil(Name) ← track(Name,X,Y), rank(Name,S), S>=5, path1("HERMON",X),  
path0(Y,"EILAT").  
shvil(Name) ← track(Name,X,Y), path0("HERMON",X), path0(Y,"EILAT").
```