



**הטכניון - מכון טכנולוגי לישראל**  
**הפקולטה למדעי המחשב**

אביב תשס"ח  
8, ספטמבר 2008

ד"ר אמיר שפילקה  
נדב שרגאי

**מערכות מסדי נתונים – 236363**

**מועד א'**  
**הזמן: 3 שעות**  
במבחן זה 10 עמודים

**עדכונים שנכתבו על הלוח במהלך הבחינה כתובים באדום**

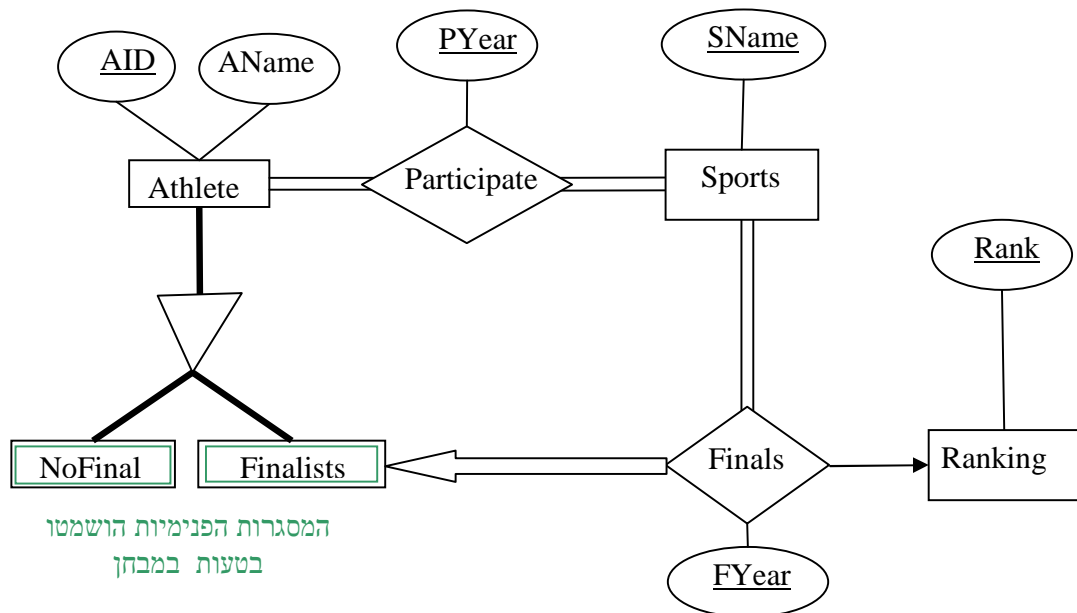
שאלה	נקודות
שאלה 1 – ERD	20
שאלה 2 – שפות שאילתא	36
שאלה 3 – פירוקים וצורות נורמליות	24
שאלה 4 – XML	20
סה"כ	100

**הנחיות:**

- יש לענות על כל השאלות **בטופס הבחינה** מחברות הטיוטה לא תיאספנה.
  - חומר עזר מותר: דף נוסחאות A3 יחיד מודפס משני הצדדים.
  - 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 (20 נק')

נתונה סכמת ה-ERD הבאה המתארת ספורטאים שהגיעו לגמרים באולימפיאדות ומיקומם הסופי.



הסבר לסכמה:

הסימון מתאר קו כפול (השתתפות מלאה) וחץ (חד ערכיות).

ישויות:

**Athlete** - ישות המתארת אתלטים. אתלט מאופיין ע"י מזהה ייחודי (**AID**) ושם (**AName**).  
 האתלטים יכולים להיות אתלטים שהגיעו לגמר במקצוע כלשהוא (**Finalists**) או אתלטים שלא הגיעו לאף גמר (**NoFinal**).  
**Ranking** - ישות המתארת את המיקום הסופי של האתלט בגמר. המיקום מאופיין ע"י הדירוג (**Rank**) שהוא שדה מספרי.  
**Sports** - ישות המכילה ענפי ספורט. ענף ספורט מאופיין ע"י שמו (**SName**).

יחסים:

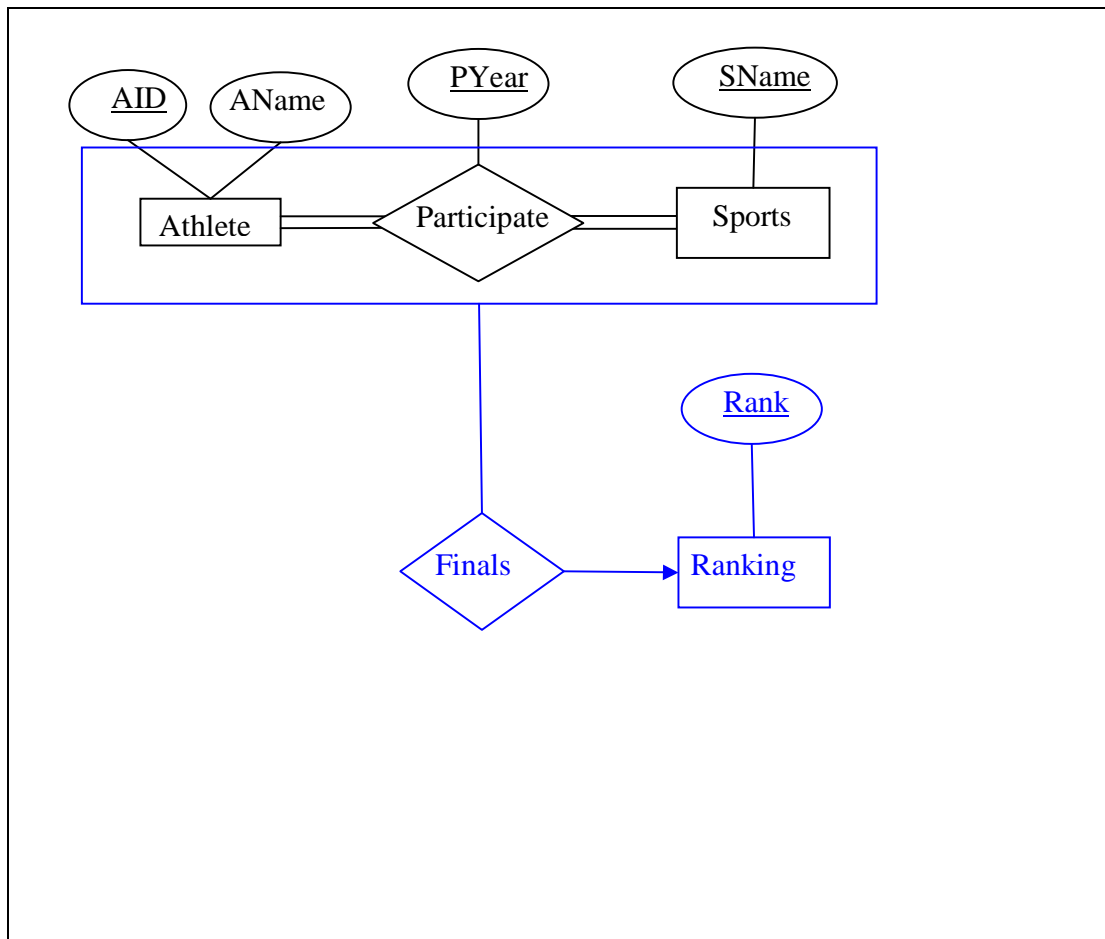
**Participate** - יחס המתאר התחרות של אתלט במקצוע מסוים בשנה מסוימת (**PYear**).  
**Finals** - יחס המתאר השתתפות של אתלט בגמר במקצוע מסוים בשנה מסוימת (**FYear**) ומיקומו הסופי.

א. (8 נק') מלאו את הטבלה הבאה שמתייחסת לשדות ולמפתחות (הקבילים) של יחסים בסכמת ה-ERD.

Participate	
שדות	AID, PYear, SName
מפתחות	AID, PYear, SName

Finals	
שדות	SName, AID, Rank, Fyear
מפתחות	SName, Rank, FYear - ו SName, AID, FYear

ב. (6 נק') שרטטו ERD כך שנתונים על גמר ודירוג סופי יינתנו רק לספורטאים שהתחרו באותה שנה (לפי Participate).



### מתייחס לדיאגרמה המקורית

ג. (6 נק') נתון כי לאחר המרת הדיאגרמה לטבלאות בטבלה Finalists יש  $f$  שורות, בטבלת Ranking יש  $r$  שורות, בטבלה Sports יש  $s$  שורות. בנוסף נתון שב Finals יש נתונים רק משתי אולימפיאדות. מהו המספר המקסימאלי של שורות בטבלה Finals כתלות ב- $f, r, s$ ?

$$\min(2*f*s, 2*r*s)$$

### שאלה 2 – שפות שאילתא (36 נק')

$$R=R(A,B,C)$$

א. (6 נק') נתונים הביטויים הבאים:

$$1. (\pi_A(R) \times \pi_B(R) \times \pi_C(R)) \div \pi_A(R)$$

$$2. [(\pi_A(R) \times \pi_{B,C}(R)) \cup (\pi_{A,B}(R) \times \pi_C(R)) \cup (\pi_{A,C}(R) \times \pi_B(R))] \div \pi_A(R)$$

3. בחרו את האפשרות הנכונה ונמקו את תשובתכם:

- i. הביטויים שקולים
- ii. התוצאה של ביטוי 1 תמיד מוכלת בתוצאה של ביטוי 2
- iii. התוצאה של ביטוי 2 תמיד מוכלת בתוצאה של ביטוי 1
- iv. אף אחד משני הביטויים אינו מוכל בביטוי השני (התוצאה של 1 לא תמיד מוכלת בתוצאה של 2 ולהפך).

כל ביטוי מהצורה  $\pi_{X,Y}(R)$  עבור  $X$  ו- $Y$  ב- $\{A,B,C\}$  מוכל ב- $\pi_X(R) \times \pi_Y(R)$  ולכן הביטוי בשורה 2 מוכל בביטוי משורה 1. כמו כן הדוגמא בה  $R$  מכיל רק את השורות  $(0,0,0)$ ,  $(1,1,1)$ ,  $(2,2,2)$  מראה כי שורה 1 לא מוכלת בשורה 2.  
נשים לב כי ברגע שיחס  $S$  מוכל ב- $T$  אז לכל יחס  $R$  מתקיים ש- $S \div R$  מוכל ב- $T \div R$  (או שווה לו).

ב. (8 נק') תהי  $R(A)$  רלציה כך שב- $A$  רשומים מספרים. הראו כי לכל קבוע  $k$  ניתן לרשום ב- $RA$  שאלתא המחזירה את  $k$  האיברים הגדולים ב- $R$ .

$$X = \sigma_{A1 > A2, A2 > A3 \dots AK > A(K+1)} R \times R \times \dots \times R \quad \text{k+1 times}$$

$$R \setminus \pi_{A(K+1)} X$$

עמודה  $A(k+1)$  מכילה את כל האיברים שיש  $k$  איברים גדולים מהם, לכן היא מכילה את כל האיברים פרט ל- $k$  האיברים הכי גדולים.

**הערה:** בסעיפים הבאים הניחו תרגום סטנדרטי של סכמת ה-ERD המקורית משאלה 1 לסכמות רלציוניות:  $S$ =Sports,  $R$ =Rank,  $A$ =Athlete,  $Fsts$ =Finalists,  $NF$ = NoFinal,  $F$ =Finals,  $P$ =Participate. בשאלות העוסקות ב-DRC הניחו סדר אלפביתי על התכונות של הרלציות.

ג. (8 נק') נאמר כי שני ספורטאים הם יריבים ספורטיביים אם יש ענף ספורט בו שניהם התחרו יחדיו לפחות פעם אחת ובכל שנה בה שניהם התחרו בענף הספורט הנ"ל הם סיימו במקומות 1 ו-2 בגמר. כתבו שאלתת DRC המחזירה את כל השלשות  $(SName, AID1, AID2)$  כאשר  $AID1$  ו- $AID2$  הם מזהים של יריבים בענף הספורט  $SName$  ו- $AID1 < AID2$ .

$$\{ \langle SName, AID1, AID2 \rangle \mid AID1 < AID2 \wedge \forall y ((P(AID1, y, SName) \wedge P(AID2, y, SName)) \rightarrow ((F(AID1, y, 1, SName) \wedge F(AID2, y, 2, SName)) \vee (F(AID1, y, 2, SName) \wedge F(AID2, y, 1, SName)))) \wedge \exists y (P(AID1, y, SName) \wedge P(AID2, y, SName)) \}$$

ד. (8 נק') נאמר כי ספורטאי הצטיין בשנה מסוימת אם הוא הגיע לגמר ברוב המקצועות בהם הוא התחרה באותה השנה. כתבו מבט excelled ב-SQL המחזיר את כל הזוגות (AID,Year) עבור כל מזהה של אתלט AID ושנה בה הצטיין.

```
CREATE VIEW excelled AS
SELECT DISTINCT f1.AID, f1.FYear as year
FROM Finals f1
WHERE (SELECT COUNT(*)
      FROM Finals f2
      WHERE f2.AID = f1.AID AND f2.FYear = f1.FYear
) > (SELECT COUNT(*)
     FROM Participate p
     WHERE p.AID = f1.AID AND p.PYear = f1.FYear
)/2.0
```

ה. (6 נק') הסבירו מה עושה שאילתת ה-SQL הבאה. הניחו כי excelled(AID,Year) הוא המבט שחושב בסעיף הקודם.

```
SELECT AVG(t.times)
FROM (SELECT COUNT(*) as times
      FROM excelled e
      WHERE year BETWEEN 1940 AND 2008
      GROUP BY AID
      HAVING COUNT(*) > 1
) t
```

השאילתא תחזיר את מספר הפעמים הממוצע שאתלט הצטיין בין השנים 1940 ל-2008 תוך התחשבות רק באתלטים שהצטיינו לפחות פעמיים בשנים האלו.

### שאלה 3 – פירוקים וצורות נורמליות (24 נק')

א. (6 נק') לכל אחד מהסעיפים הבאים אמרו האם הוא נכון או לא. נמקו!

i. (3 נק') כל פירוק BCNF הוא משמר מידע

לא נכון.  
נפרק את הסכמה  $R(A,B,C,D)$  ל- $R_1(A,B)$  ו- $R_2(C,D)$ .  $R_1$  ו- $R_2$  הם ב-BCNF כי יש להם רק שתי תכונות אך הפירוק כולו אינו משמר מידע כי החיתוך הוא קבוצה ריקה ומן הסתם לא  $R_1$  ולא  $R_2$  תלויים פונקציונלית בקבוצה ריקה.

ii. (3 נק') כל פירוק 3NF משמר תלויות

לא נכון.  
ניקח את הדוגמא מסעיף i ונוסיף לה ש- $R$  מקיים את התלות  $B \rightarrow C$ . הפירוק הוא ב-BCNF ולכן גם ב-3NF אבל התלות  $B \rightarrow C$  אינה נשמרת.

ב. (6 נק') יהי  $R(A,B,C,D)$  יחס עם קבוצת תלויות  $F$ . נאמר שקבוצת תכונות  $X$  היא סגורה אם  $X_F^+ = X$ . נתון כי רק הקבוצות  $\emptyset$ ,  $\{A,B\}$  ו- $\{A,B,C,D\}$  הן סגורות ב- $R$ . תנו כיסוי ל- $F$ . נמקו!

הכיסוי הבא הוא כיסוי מינימאלי:  
 $F_c = \{A \rightarrow B, B \rightarrow A, C \rightarrow D, D \rightarrow C, C \rightarrow A\}$   
הסבר:  
מאחר ו- $\{A,B\}$  סגורה ואף תת קבוצה שלה אינה סגורה הרי ש- $A_F^+ = B_F^+ = \{A,B\}$ . מכאן מקבלים  $A \rightarrow B, B \rightarrow A$ . מאחר ו- $\{A,B,C,D\}$  סגורה ואף תת קבוצה שלה שמכילה את  $C$  או  $D$  אינה סגורה הרי ש- $C_F^+ = D_F^+ = \{A,B,C,D\}$  כלומר  $C$  ו- $D$  מפתחות. מכאן נובעות שאר התלויות.  
נשים לב שאם הייתה קיימת תלות נוספת בכיסוי המינימאלי הרי שהיא חייבת להיות עם תת קבוצה של  $\{A,B\}$  באגף שמאל ו- $C$  או  $D$  באגף ימין, אבל אז  $\{A,B\}$  לא הייתה קבוצה סגורה בסתירה.

ג. (6 נק') נתונה סכמה R הנמצאת ב- 3NF אך לא ב- BCNF. הוכיחו כי ל-R שני מפתחות קבילים שונים.

R ב-3NF אך לא ב-BCNF. מכאן נקבל שיש תלות פונקציונלית לא טריביאלית  $X \rightarrow A$  כך ש-A חלק ממפתח קביל ו-X אינו על-מפתח. נאמר ש-A הוא חלק מהמפתח הקביל Y. מכאן שהקבוצה  $(Y \setminus \{A\}) \cup X$  היא על-מפתח. X אינו מכיל את A ולכן גם  $(Y \setminus \{A\}) \cup X$  אינו מכיל את A משמע שאם ניקח את על-המפתח הזה ונביא אותו למצב מינימלי כדי לקבל מפתח קביל הוא לא יכיל את A ולכן יהיה מפתח קביל שונה מ-Y.

ד. (6 נק') הראו כי הרלציה R(A,B,C,D,E,H) נמצאת ב- BCNF עבור קבוצת התלויות  $F = \{A \rightarrow BC, CD \rightarrow EH, AB \rightarrow D, C \rightarrow AH\}$

קודם כל נמצא כיסוי מינימלי

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

כפי שניתן לראות כל התלויות בכיסוי המינימלי הן תלויות פונקציונליות בהן צד שמאל הוא A או C אשר שניהם למעשה מפתחות קבילים של הרלציה. מכאן נקבל שכל תלות לא טריביאלית  $X \rightarrow Y$  המוכלת ב-  $F_c^+ = F^+$  מקיימת ש-X מכיל את A או C ולכן על-מפתח. בפרט נובע כי R היא ב-BCNF.



## שאלה 4 – XML (20 נק')

נתון ה-DTD הבא המתאר קובץ document (שהוא צומת המסמך) המכיל מאמרים (articles) שונים. כל מאמר מכיל טקסט ו/או ציטוטים ממאמרים אחרים.

```
<!ELEMENT document (article)* >
<!ELEMENT article (#PCDATA|citation)* >
<!ATTLIST article aid ID #REQUIRED >
<!ELEMENT citation (#PCDATA) >
<!ATTLIST citation ref IDREF #REQUIRED>
```

**מאמר אחד לפחות ציטט לפחות פעמיים.**

א. (7 נק') כתבו שאילתת xpath המחזירה את כל המאמרים שמאמר אחר ציטט אותם לפחות פעמיים.

`id(//citation/@ref[.=../following-sibling::*/@ref])`

ב. (7 נק') כתבו שאילתת xpath המחזירה את כל המאמרים המצטטים מאמר המצטט אותם (ז.א כל מאמר A כך שיש B המצוטט ב-A והמצטט את A).

`//article[id(//@ref)//@ref=@aid]`

ג. (6 נק') שנו את ה-DTD כך שכל מאמר יוכל להכיל גם הערות שוליים (footnotes). הערת footnote יכולה להכיל טקסט והפניה למאמר (לכל היותר למאמר אחד). לא ניתן לצטט footnote.

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

```
<!ELEMENT document (article)* >
```

```
<del>!ELEMENT article (#PCDATA|citation)*>
```

```
<!ELEMENT article (#PCDATA|citation|footnote)* >
```

```
<!ATTLIST article aid ID #REQUIRED >
```

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

```
<!ATTLIST citation ref IDREF #REQUIRED>
```

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

```
<!ATTLIST footnote ref IDREF #IMPLIED>
```