

סמסטר חורף התשע"ב

פרופ' ח' אלדר פישר

מרצה :

מר יבגני אברמוביץ'
מר דימה אלנבוגן

מתרגלים :

**מערכות מסדי נתונים
236363**

מועד א' (ה' באדר התשע"ב, 28 בפברואר 2012)

<u>מס' ת.ז.</u>

מספר השאלה	נקודות
1	/27
2	/20
3	/24
4	/29
סה"כ	/100

הנחיות לנבחנים

1. כתבו את התשובות אך ורק בטופס הבחינה, המחברת מיועדת לטיוטה 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 שאילתה.**
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 (ללא דף הסריקה) כוללת 12 דפים, כולל דף זה. נא לוודא שיש בידכם את כל הטופס.
11. כאשר הניקוד של 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 (27 נק')

נתונים מאפיינים של ממלכת גמדים:

לכל גמד יש שם ייחודי, שבט וגובה. כל גמד-כורה עובד במכרה יחיד וכרה עד היום כמות מסוימת, הנמדדת בקילוגרמים, של משאבים. לכל מכרה יש מיקום ייחודי (נתון ע"י תיאור), ומשאב אחד בדיוק שניתן לירות בו. לכל משאב יש שם ייחודי ומחיר לקילוגרם. מחיר המשאב אחיד בכל המכרות. גמד יכול להיות גמד-כורה ו/או גמד-מנהל. לכורה חייב להיות מנהל אחד לפחות, ולמנהל חייב להיות כורה אחד לפחות מנוהל על ידו.

א. (6 נק') ענו ונמקו בקצרה על השאלות הבאות:

i. האם יכול להיות כורה המנוהל ע"י שני מנהלים בו זמנית?

כן. מצוין שהכורה מנוהל ע"י אחד לפחות, אין מגבלה על מספר המנהלים.

ii. האם יכול להיות כורה שכרה שני משאבים שונים?

לא. כורה עובד במכרה יחיד ובמכרה יש משאב יחיד, ולכן הוא יכול לירות משאב אחד בלבד.

iii. האם יכול להיות מנהל שכוריו כרו שני משאבים שונים?

כן. מנהל יכול לנהל שני גמדים שונים שכל אחד עובד במכרה עם משאב אחר.

ב. 14 נק') שרטטו ERD המתאר את הממלכה.

התשובה – בנספח.

הקודים לטעויות אופייניות והעונשים בעבורן :

E1	קו היה צריך לצאת מישותליחס אחר	-2
E2	כדאי היה לאחד קשרים	-2
E3	סוג הקו שגוי	-2
E4	אטריביוט אינו אמור להיות מפתח	-2
E5	אטריביוט שייך לישותליחס אחר	-2
E6	אטריביוט שלא מתואר בטקסט השאלה	-1
E7	הקשר בצורה שגויה מאחד שני קשרים	-3
E8	קשר מערב ישויות לא רלוונטיות בהתבסס בטקסט השאלה	-3
E9	לא ציירת 	-1
E10	חסר אטריביוט	-3
E11	משולש ISA בכיוון לא נכון	-1
E12	צריכים להופיע שני משולשי ISA	-3
E13	קשר שלא הוגדר בטקסט השאלה	-3

ג. (7 נק') בכל אחד מהקשרים ששרטטתם : כתבו כיסוי לקבוצת כל התלויות הפונקציונאלית המתקיימות בו.

Manages: \emptyset

WorksIn: $\{ \text{gname} \rightarrow \{ \text{mined}, \text{mlocation} \} \}$

MineOf : $\{ \text{mlocation} \rightarrow \text{rname} \}$

שאלה 2 – שאלות מידע (20 נק')

נתונות הרלציות הבאות, אשר אינן בהכרח תואמות את התיאור וה-ERD שבניתם בשאלה הקודמת:

gnome(name, gtribe) – גמד ששמו name (הניחו שאין שתי שורות ב-gnome עם אותו שם גמד) הוא חבר השבט gtribe.

mine(mname, mresource, mKgPrice) – מכרה ששמו mname (הניחו שאין שתי שורות ב-mine עם אותו שם מכרה), המשאב שלו הוא mresource, והמחיר לק"ג של המשאב הוא mKgPrice.

works(gname, mname, amountKg) – גמד ששמו gname עובד במכרה (הניחו שאין שתי שורות ב-works עם אותו שם גמד) ששמו mname, וכרה עד עכשיו amountKg ק"ג מהמשאב של המכרה.

managedBy(ename, mname) – גמד ששמו ename מנהל ע"י גמד ששמו mname.

א. (12 נקודות) **גמד מנהל זוטרי מצטיין** הינו גמד שיש תחתיו 5 עובדים או פחות, כך שעובדיו כרו הכי הרבה משאבים (מבין אלה שיש תחתם 5 גמדים או פחות). כתבו שאלתת SQL המוצאת את שבטו של המנהל המצטיין.

אם יש כמה מנהלים מצטיינים בצבעים שונים, יש להחזיר את כל הצבעים.

מותר להשתמש במבטים, אבל רק אם הם אינם מיותרים. מבט יחשב למיותר אם שאר הביטוי משתמש בו לא יותר מפעם אחת.

```
CREATE VIEW workers_mined AS
```

```
SELECT SUM (works.amountKg) AS mined, gnome.gtribe AS tribe
```

```
FROM managedBy, works, gnome
```

```
WHERE managedBy.ename = works.gname AND managedBy.mname =  
gnome. name
```

```
GROUP BY manages.mname, gnome.gtribe
```

```
HAVING (COUNT(*) ≤ 5)
```

```
SELECT tribe
```

```
FROM workers_mined wm1
```

```
WHERE NOT EXISTS (SELECT 1 FROM workers_mined wm2 WHERE  
wm1.mined < wm2.mined)
```

הקודים לטעויות אופייניות והעונשים בעבורן :

F0	שימוש שגוי ב-MAX היה צריך להשתמש בתת-שאלתא	-7
F1	שימוש בתת-שאלתא במקום $HAVING count(*) < 5$	-1
F2	אין קישור בין מנהל לעובדיו	-3
F3	MAX(SELECT ...)	-3
F4	mname לא חד-משמעי (מופיע בשתי טבלאות רלוונטיות)	-2
F5	שימוש ב-LEFT OUTER JOIN עלול ליצור ערך NULL ב-SUM וואו MAX	-1
F6	חסרה בדיקה שהמנהל זוטר	-3
F7	טבלה מיותרת שמשתתפת ב-joint	-1
F8	אין קישור בין המנהל לשבטו	-1
F9	שדה פלט לא מופיע תחת GROUP BY	-2
F10	אסור להשתמש ב-alias חיצוני בתוך HAVING	-2
F11	אסור להשתמש ב-alias חיצוני בתוך תת-שאלתא	-5..-7
F12	חייבים לתת שם לתת-שאלתא	-1

כמעט כולם חילקו את המשימה ל-3 שלבים :

1. חישוב לכל מנהל זוטר כמה כרו עובדיו ;

2. מציאת המנהל הזוטר המצטיין ;

3. מציאת שבטו.

סך כל הורדות עבור השלב מס' 1 הוגבל ב-4 נק'. סך כל הורדות עבור השלב מס' 2 הוגבל ב-7 נק'. סך כל הורדות עבור השלב מס' 3 הוגבל ב-1 נק'. אף על פי כן הטעות F9 גררה 2 נק' ללא קשר באיזה השלב היא נעשתה.

לסטודנטים שלא הצלחנו להבחין בפתרונם את השלבים הנ"ל הנקודות הורדו מכלל 12 הנק' של משקל הסעיף.

ב. (8 נקודות) – כתבו שאילתת DRC המוצאת מנהלים המקיימים את התנאי הבא: לכל משאב (שנכרה באיזשהו מכרה) יש לפחות שני עובדים של המנהל שכורים אותו.

מותר להגדיר קבוצות עזר באמצעות שאילתות DRC בלבד.

אסור שהשאילתה תהיה תלוית תחום (domain-dependent). אין צורך להוכיח ששאילתה אינה תלוית תחום.

Must add an addition protection of c: (ממנהל משהו)

$$\{m \mid \exists c(\text{managedBy}(c,m)) \wedge \forall r(\exists n \exists p(\text{mine}(n,r,p)) \rightarrow \\ \exists w_1 \exists w_2 (w_1 \neq w_2 \wedge \text{managedBy}(w_1,m) \wedge \text{managedBy}(w_2,m) \wedge \\ \exists p_1 \exists p_2 \exists m_1 \exists m_2 \exists a_1 \exists a_2 (\text{works}(w_1,m_1,p_1) \wedge \text{mine}(m_1,r,p_1) \\ \wedge \text{works}(w_2,m_2,p_2) \wedge \text{mine}(m_2,r,p_2))))\}$$

שאלה 3 - נושאים מעורבים (24 נק')

הערה: אין תלות בין סעיפי השאלה (א', ב' ו-ג').

א. (8 נק') נתונה טבלה Tbl(Name, Val). הטיפוס של Name הוא מחרוזת שאורכה מוגבל ב-64 תווים. הטיפוס של Val הוא REAL. הערה/הנחיה: אף אחד מהערכים בטבלה לא יכול להכיל ערך NULL.

i. (3 נק') כתבו ביטוי SQL היוצר את הטבלה Tbl על סמך כל העובדות הנ"ל.

```
CREATE TABLE Tbl (  
    Name VARCHAR(64) NOT NULL,  
    Val REAL NOT NULL);
```

ii. (5 נק') בסעיף זה הנכם מתבקשים לכתוב שאילתת SQL שמחשבת את הערך השני בגודלו מכל הערכים השונים של העמודה Val (כלומר, תוך השמטת כפילויות). במידה ויש בטבלה פחות משני ערכים שונים של Val, על השאילתא להחזיר NULL. אין ליצור טבלאות זמניות או מבטים.

```
SELECT MAX(*) FROM  
(SELECT Val FROM Tbl EXCEPT  
    SELECT MAX(Val) FROM Tbl);
```

ב. (8 נק') נתונה רלציה r מעל סכמה R(A,B,C). לאחר פירוק הסכמה לתתי-סכמות S(A,B) ו-T(B,C) התקבלו רלציות t ו-s בהתאמה.

i. (2 נק') נמקו בקצרה מדוע על סמך הנתונים הנ"ל אי-אפשר באופן כללי לדעת האם הפירוק של R ל-S ו-T משמר מידע.

ללא F נתונה, אי אפשר לדעת אם B קובע את A או C או לא קובע אף אחד מהם.

ii. (6 נק') נסמן ב- r' את הרלציה שעבורה מתקיימים כל אחד משני התנאים הבאים:

- לאחר הפירוק מ- r' היו מתקבלות בדיוק אותן הרלציות s ו- t בהתאמה;
- הפירוק היה משמר את התוכן של r' .

הביעו את r' כתלות ב- r בלבד (ללא שימוש ב- s או t) כביטוי באלגברה רלציונית (RA)

$$\pi_{A,B}(r) \bowtie \pi_{B,C}(r)$$

ג. (8 נק') בשאלה זאת תכתבו תוכנית Datalog (ללא שלילות) בטוחה.

הניחו כי בהינתן סכמה R וקבוצת תלויות פונקציונליות לא טריביאליות F מעל הסכמה נתונים מראש הפרדיקטים המפורשים הבאים:

- $dep1(A,C)$ שמתקיים אם $A \rightarrow C \in F$
- $dep2(A,B,C)$ שמתקיים אם $AB \rightarrow C \in F$
- $inR(A)$ שמתקיים אם $A \in R$

בהסתמך על הפרדיקטים המפורשים עליכם לכתוב את הכללים לפרדיקט סתום $aClose(A,B)$ שמתקיים אם $B \in A_F^+$.

$aClose(A, A) \leftarrow inR(A).$
 $aClose(A, B) \leftarrow aClose(A, C), dep1(C, B).$
 $aClose(A, B) \leftarrow aClose(A, C), aClose(A, D), dep2(C, D, B).$

שאלה 4 - XML (29 נק')

נתון ה-DTD הבא, המתאר דו"חות הניתנים ע"י המנהלים על העובדים. אלמנט המסמך הוא file.

הבהרה בזמן המבחן: הניחו כי התאריך הוא מספר שלם.

```
<!ELEMENT file (review)*>
<!ATTLIST file updated CDATA #REQUIRED>
<!ELEMENT review (manager, worker*, rtext)>
<!ATTLIST review date CDATA #REQUIRED
            id ID#REQUIRED>
<!ELEMENT manager (#PCDATA)>
<!ELEMENT worker (#PCDATA)>
<!ELEMENT rtext (#PCDATA|emph|point)*>
<!ELEMENT emph (#PCDATA)>
<!ELEMENT point EMPTY>
<!ATTLIST point to IDREF #REQUIRED>
```

הסבר ל-DTD:

הקובץ מכיל דו"חות (review) שנכתבו ע"י מנהלים באתר. לכל דו"ח מצוין המנהל (manager), העובדים שאליהם הוא מתייחס (צומת worker לכל אחד), וטקסט הדו"ח. הטקסט יכול להכיל הדגשות (emph) ומצביעים לדו"חות אחרים (point). לכל דו"ח מזהה ייחודי שיכול לשמש להצבעות עליו. כמו כן שומרים תאריך עדכון אחרון של הקובץ (תכונה updated של הצומת file) ותאריך כתיבה של כל דו"ח (תכונה date של צומת הדו"ח). אין רישום נפרד של פרטי העובדים והמנהלים פרט למה שמופיע בדו"חות.

א. (8 נק') ענו ונמקו בקצרה על השאלות הבאות באשר ל-DTD.

i. האם יתכן שמנהל ייתן דו"ח בלי לציין עובדים?

כן. כוכבית(*) מציינת 0 עובדים או יותר, לכן ייתכן דו"ח המציין 0 עובדים.

ii. האם יתכן שאותו מנהל יתן דו"ח על אותו עובד יותר מפעם אחת?

כן. יתכנו שני דו"חות שבהם מופיעים אותו מנהל ואותו עובד (למעשה אין שום דרך ב-DTD לדרוש שתוכן צמתיים אלו יהיה שונה בין דו"חות).

iii. האם יתכן שדו"ח יכלול הצבעה על עצמו?

כן. תכונה מטיפוס IDREF יכולה להצביע לכל ID בקובץ, בפרט גם ל-ID של הדו"ח עצמו.

iv. האם תתכן הצבעה באמצע קטע מודגש?

לא. קטע מודגש מכיל צומת טקסט בלבד ולא ייתכנו בו אלמנטים כלשהם, בפרט לא אלמנטים עם תווית point.

ב. 9 נק') דו"ח יקרא "תמוה" אם התאריך שלו הוא אחרי תאריך עדכון הקובץ (התכונה של file), או לפני תאריך של דו"ח כל שהוא שהוא מצביע עליו. כתבו שאילתת XPath1.0 המחזירה את קבוצת צמתי ה-review של כל הדוחות התמוהים.

/file/review[date>/file/@updated or date<id(point/@to)/@date]

ג. (12 נק') נתונה שאילתת ה-XQuery1.0 הבאה (file.xml הוא הקובץ מהשאלות הקודמות) :

```
document {<list updated="{ doc("file.xml")/file/@updated }">
for $w in distinct-values(doc("file.xml")//worker)
order by $w
return (<worker><name>$w/text()</name>
      for $r in doc("file.xml")
      where $r/worker=$w
      order by $r/@date
      return ($r/manager,<rev>($r/rtext/text() union $r/rtext/emph)</rev>)
</worker>)
</list>}
```

i. כתבו במילים מה השאילתה מחזירה. יש לציין את כל הפרטים הרלוונטים.

לכל עובד שיש עליו דוחות יהיה את ריכוז הדוחות עליו (דו"ח שמציין מספר עובדים יופיע בריכוז של כל אחד מהם). העובדים ממויינים לפי שמם ולכל עובד הדוחות ממויינים לפי תאריך (אם כי פרט לתאריך עדכון הקובץ לא שומרים את התאריכים עצמם). לכל דו"ח שמם את המנהל שחיבר אותו ואת הטקסט כולל ההדגשות אך ללא המצביעים.

ii. כתבו DTD המתאים למסמך שהשאילתה מחזירה. יש להשתמש במירב האילוצים אשר עדיין תמיד מתקיימים בפלט השאילתה עבור מסמך file.xml תקין. עבור צמתים המכילים בן #PCDATA בלבד אפשר לציין זאת במרוכז במקום לכתוב שורה לכל אחד מהם.

```
<!ELEMENT list (worker*)>
<!ATTLIST file updated CDATA #REQUIRED>
<!ELEMENT worker (name,(manager,rev)+)>
<!ELEMENT rev (#PCDATA|emph)*>
מכילים בן טקסט בלבד : name, manager, emph
```

נספח:

תשובה לסעיף 1 א':

