

סמסטר חורף התשע"ו

פרופ' בני קימפלד

מרצה :

מר רוני ליכר  
גב' הדר פרנקל  
מר נמרוד רייפר

מתרגלים :

**מערכות מסדי נתונים****236363**

מועד ב'

(כ"ו באדר א' התשע"ו, 6 במרץ 2016)

**פירוט החלקים והניקוד:**

מס'	נושא	מס' שאלות בחלק	מס' שאלות שיש לענות עליהן	ניקוד
1	שאלות במודל היחסים	4	3	40
2	תכן מסדי נתונים	2	2	24
3	מודלים לא יחסיים	4	3	36
<b>סה"כ</b>		<b>10</b>	<b>8</b>	<b>100</b>

**הנחיות לנבחנים**

1. כתבו את התשובות אך ורק במחברת התשובות. טופס הבחינה לא ייבדק.
2. מותר לכתוב את התשובות בעפרון, באחריותכם לוודא שתשובתכם קריאות.
3. כל חומר עזר הכתוב על נייר בלבד מותר בשימוש.
4. אין לקבל או להעביר חומר עזר כלשהו בזמן הבחינה.
5. בבחינה שלושה חלקים עם אפשרויות בחירה בחלקים 1 ו-3. יש לוודא שסימנתם באופן ברור את השאלות שבחרתם לענות עליהם. במקרה של סימון לא ברור, השאלות הראשונות תיבדקנה.
6. בכל מקום שלא נאמר אחרת, יש לנמק את התשובות בקצרה. **תשובות לא מנומקות לא תתקבלנה, למעט במקומות שבהם אתם מתבקשים לכתוב שאלתה.**
7. 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.
8. משך הבחינה הינו שלוש שעות. תכננו את הזמן בהתאם. **לא תינתנה הארכות זמן במהלך המבחן.**
9. הבחינה כוללת 10 שאלות ב-3 חלקים. נא לוודא שיש בידכם את כל הטופס.
10. כאשר ניקוד תתי הסעיפים אינו מצוין, ניקוד הסעיף מתחלק שווה ביניהם.
11. הניקוד אינו נועד לשקף את קושי השאלה, ולכן מומלץ לקרוא קודם את כל השאלות.

**בהצלחה**

## חלק 1 – שאלות מידע רלציוניות – 40 נק'

בחלק זה עליכם להשתמש ביחסים הבאים (לא הוגדרו מפתחות):

**Address**(id, street, number, city, type)

אדם בעל המזהה id מתגורר ברחוב street מספר בית number בעיר city.  
במידה ו-type="rent" – הוא שוכר אותה ובמידה ו-type="owns" – הוא רכש את הדירה. דוגמא לרשומה:  
(123, "Hertzel", 10, "Haifa", "rent")

**Owns**(id, street, number, city)

אדם בעל המזהה id בעל הדירה ברחוב street מספר בית number בעיר city.

**MayorOf**(id, city)

אדם בעל המזהה id הוא ראש העיר city.

**Board**(id, neighborhood, city)

אדם בעל המזהה id משתתף בוועד של שכונה neighborhood בעיר city.

### 1. SQL, 16 נק' – שאלת חובה.

א. 9 נק' – בסעיף זה התעלמו מהיחסים אשר הוגדרו בתחילת החלק.  
הניחו כי **עיר** מיוצגת ע"י שם בלבד ו**שכונה** מיוצגת ע"י שם השכונה ושם העיר בה היא נמצאת. בנוסף, מסד הנתונים ריק, אינו מכיל אף טבלה. כל טבלה שתיצרו חייבת לקיים את הגבלות הייצוג.

עליכם לכתוב מספר פקודות SQL המבצעות את הפעולות הבאות. כתבו גם פקודות נוספות הנדרשות לאפשר הפעולות, אם כאלו קיימות.

- יצירת טבלת ערים.
- יצירת טבלת שכונות.
- הוספת שכונה.
- מחיקת העיר בה נמצאת השכונה שהוספתם.

ב. 7 נק' – בהינתן היחסים שניתנו למעלה, כתבו שאלת SQL אשר מחזירה זוגות של אדם ועיר אשר מקיימים לפחות אחד מהתנאים הבאים:

1. האדם הוא ראש אותה עיר.
2. האדם גר עם עוד חמישה אנשים לפחות באותו הבית הממוקם בעיר.

### 2. RA, 12 נק' – שאלת חובה.

יש לענות על סעיפים א' וב' בהינתן היחסים שניתנו בתחילת החלק.

א. 4 נק' – כתבו שאלת RA כך שעבור כל אדם שיש בבעלותו לפחות שני בתים וכן הוא מתגורר באחד הבתים שבבעלותו, מחזירה את מספר הזיהוי שלו.

ב. 4 נק' – כתבו שאלת RA המחזירה את השכונות שקיים בוועד שלהן אדם שלא שוכר אף דירה.

ג. 4 נק' – תהיינה  $R_1, R_2, R_3$  סכמות ויהיו  $r_1, r_2, r_3$  יחסים מעל הסכמות הנ"ל בהתאמה. הוכיחו או הפריכו:

$$\pi_{R_2}(r_1 \bowtie r_2 \bowtie r_3) \subseteq r_2$$

שימו לב, יש לענות בדיוק על שאלה אחת מבין 3 (RC) ו-4 (Datalog).

3. RC, 12 נק' – שאלת בחירה מבין 3 ו-4.

בשאלה זו הניחו כי קיימים היחסים הבאים:  $R(A, B)$ ,  $S(B, C)$

א. כתוב שאילתת RC אשר מחזירה True אם ורק אם מתקיים  $\pi_B(R) \subseteq \pi_B(S)$ .

ב. עבור השאילתות הבאות, הסבר האם הן תלויות בתחום.

$$\{(x) | \forall y (S('Anna', y) \rightarrow R(y, x))\} \quad (a)$$

$$\{(x, z) | \exists y R(x, y) \wedge S(y, z)\} \quad (b)$$

4. Datalog, 12 נק' – שאלת בחירה מבין 3 ו-4.

א. להלן תכנית Datalog (הנח כי פרדיקט U קיים במסד הנתונים):

$$V(x, y) \leftarrow U(x, y)$$

$$S(x, y) \leftarrow V(y, x)$$

$$S(x, y) \leftarrow V(x, z), V(y, z)$$

האם התכנית רקורסיבית? נמק את תשובתך.

ב. בהרצאות נלמד על גרף תלויות של תכנית Datalog.

תזכורת: גרף תלויות הינו גרף מכון  $(V, E)$  כך ש:

$V$  – צמתים המייצגים פרדיקטים מסוג  $IDB$ .

$E$  – קשתות כאשר קשת  $R \rightarrow S$  נמצאת בקבוצה במידה וקיים כלל עם  $S$  בראש החוק (צד

שמאל) ו  $R$  בגוף החוק (צד ימין).

נסתכל על תכניות עם שלילה ( $Datalog^-$ ) ונוסיף להגדרה של גרף התלויות את התוספת הבאה:

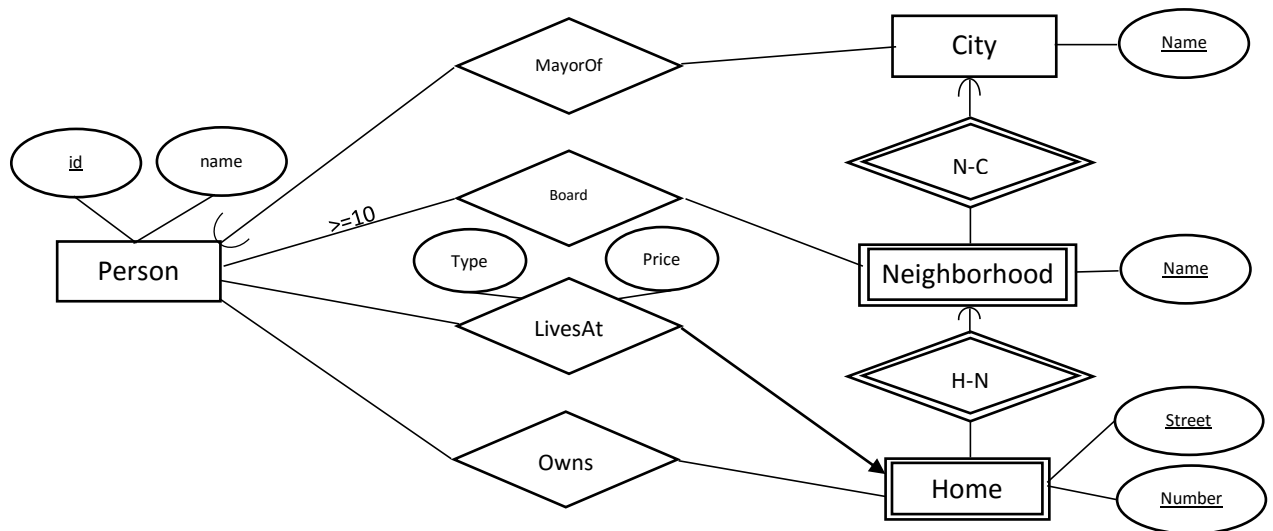
קשת  $R \rightarrow S$  בגרף התלויות תסומן כקשת שלילית אם מתקיים כי  $R$  נמצא בגוף החוק עם סימן שלילה.

הוכח את הטענה: תכנית  $Datalog^-$  אשר הגרף שלה מכיל מעגל עם קשת שלילית לא ניתנת לריבוד.

## חלק 2 – תכן מסדי נתונים - 24 נק'

### 5. ERD, 12 נק' – שאלת חובה.

להלן מערכת המתארת בתים, שכונות וערים. לערים יש ראשי ערים. לשכונות יש ועדים. בתים הם בבעלות אנשים ואנשים גרים בבתיים. אדם המתגורר בבית יכול לעשות זאת ע"י שכירה או רכישה (נקבע ע"י השדה type) ובהתאם מצוין הסכום ששולם (עבור שכירות – מחיר שכירות חודשי, ועבור רכישה – סכום הרכישה).



א. 3 נק' - תרגמו את היחס LivesAt לטבלה באופן אשר משמר את התלויות בצורה הטובה ביותר, סמנו מפתחות בקווים תחתונים.

ב. 3 נק' - לאילו מטיפוסי הישויות הבאים ניתן למצוא במערכת שתי ישויות בעלות אותו השם (name)?  
**הסבר בקצרה.**

- a. Neighborhood.
- b. City.
- c. Person.

ג. 2 נק' - מהו מספר הוועדים אשר אדם כלשהו יכול להשתתף בהם?

ד. 4 נק' – נרצה לשמור עבור השכרת בית, בין אחד לחמישה אנשים שהיו ערבים לשכירות הבית ע"י השוכר (לשוכרים שונים יתכנו ערבים שונים, גם אם הם שוכרים את אותו הבית). הסבירו כיצד ניתן לשנות ה-ERD כדי לייצג זאת באופן בו תלויות הדיאגרמה נשמרות בצורה הטובה ביותר.

6. תלויות פונקציונאליות, פירוקים וצורות נורמליות, 12 נק' – שאלת חובה.

א. 4 נק' - הוכיחו או הפריכו: כל פירוק  $3NF$  משמר תלויות.

ב. 4 נק' – תהי סכמה  $U = ABCD$  ותהיינה  $W_1, W_2$  תת סכמות של  $U$ , שונות זו מזו, כך שכל אחת מכילה שלוש תכונות. תהי  $F$  קבוצת תלויות פונקציונליות המכילה לפחות תלות אחת שאיננה טריוויאלית.

הוכיחו או הפריכו: הפירוק ל  $W_1, W_2$  משמר מידע.

ג. 4 נק' – נתונה סכמה וקבוצת תלויות:

$$U = ABCDE, F = \{A \rightarrow BC, AD \rightarrow E, EB \rightarrow C, C \rightarrow DE\}$$

ופירוק של הסכמה:

$$W_1 = ABE, W_2 = ACD, W_3 = DE$$

הראו ריצה של האלגוריתם לבדיקת שימור תלויות שנלמד בכיתה ע"מ לבדוק האם הפירוק הנ"ל משמר תלויות. קבעו ע"פ התוצאה אם הוא משמר תלויות או לא.

### חלק 3 – מודלים לא רלציוניים – 36 נק'

שימו לב, יש לענות בדיוק על שלוש שאלות מבין 7 (XML), 8 (Neo4j), 9 (MongoDB) ו-10 (RDF).

7. XML – 12 נק' - שאלת בחירה מבין 7, 8, 9, ו-10.

להלן מסמך DTD לייצוג מידע על פעילות בפורום הכוללת שליחת הודעות בקטגוריות שונות וניהול קטגוריות ע"י מנהלים (moderator).

```
<!ELEMENT forum (category+)>
<!ELEMENT category ((moderator,message*)|category+)>
<!ATTLIST category serial ID #REQUIRED
                  name CDATA #REQUIRED
                  active (yes|no) "yes">
<!ELEMENT message (#PCDATA)>
<!ATTLIST message title CDATA #IMPLIED
                  author CDATA #IMPLIED>
<!ELEMENT moderator EMPTY>
<!ATTLIST moderator email CDATA #REQUIRED
                  name CDATA #IMPLIED>
```

forum.dtd

```
<!DOCTYPE forum SYSTEM "forum.dtd">
<forum>
  <category serial="c23" name="CS Courses">
    <moderator email="mira@technion.ac.il" name="Mira"/>
    <message title="Welcome" author="Mira">
      Welcome students!
    </message>
  </category>
</forum>
```

forum.xml

א. 4 נק' - לכל אחת מהטענות הבאות הסבר האם היא מתקיימת **בכל מסמך XML חוקי** (ע"פ ה-DTD) ונמק:

א. קיימת לפחות הודעה אחת בפורום.

ב. השאילתה הבאה מחזירה תשובה לא ריקה: `//@email`.

ב. 2 נק' - האם מחיקת השורה הראשונה ("`<!DOCTYPE...>`") במסמך ה-XML הנ"ל יכולה לשנות את התשובה לשאילתת XPath כלשהי? אם כן, הראה שאילתה כזו, ואם לא הסבר מדוע לא קיימת שאילתה כזו.

ג. 6 נק' - כתוב שאילתות XPath 2.0 עבור המשימות הבאות:

א. מצאו את כל הקטגוריות הפעילות (`active="yes"`) המכילות לפחות קטגוריה אחת, ומתקיים שכל קטגוריה המקוננת תחתן (תת-קטגוריה, תת-קטגוריה של תת-קטגוריה, וכו') אינה פעילה (`active="no"`).

ב. מצאו את כל ההודעות בעלות כותרת ייחודית בפורום.

## 8. Neo4j – 12 נק' - שאלת בחירה מבין 7, 8, 9, ו-10.

נתון מסד נתונים גרפי המכיל צומת מסוג אחד:

Person
Id

צומת זה יכול להיות מקושר ע"י סוג הקשר (relationship) הבא:

BoughtFrom: קשר המקשר בין אדם (person) ואדם (person) ומכיל את התכונות price ו-product. ומציין שהאדם רכש מוצר כלשהו במחיר כלשהו.

א. 5 נק' - כתבו שאילתת Cypher המוצאת מעגלי קניה. מעגל קניה יתקיים עבור מוצר כלשהו שנמכר ע"י אדם X לאדם אחר ולאחר שרשרת מכירות של אותו המוצר, נמכר בחזרה ל-X. עליכם להחזיר זוגות של אדם ומוצר שהיו חלק ממעגל קניה.  
\* שימו לב, לאורך כל המעגל, מדובר על אותו המוצר.

ב. 7 נק' – כתבו שאילתת Cypher המוצאת אנשים הנמצאים במרחק של 5 מכירות **לכל היותר** מאנשים שמכרו **נעליים** ורק נעליים (הבהרה: מכרו מספר חיובי כלשהו, גדול שווה ל-1, של זוגות נעליים, ולא מכרו מוצרים אחרים).

## 9. MongoDB – 12 נק' - שאלת בחירה מבין 7, 8, 9, ו-10.

לרשותך אוסף Tourism המכיל פרטים אודות ערים ברחבי העולם, וכמות המלונות באותן ערים.

כל אלמנט באוסף Tourism מכיל מסמכים מהצורה הבאה:

```
{
  "_id": "10720",
  "city": "HAIFA",
  "country": "ISRAEL",
  "hotels": 50,
}
```

מסמך באוסף Tourism מתאר עיר (city) אשר נמצאת במדינה (country) ולעיר כמות מלונות המתואר באמצעות השדה hotels. \_id מתאר את מיקוד העיר. הניחו כי לא קיימות שתי ערים באוסף בעלי אותו המיקוד.

א. 4 נק' - תרגמו את שאילתת ה-SQL הבאה לשאילתת MongoDB.

```
SELECT country, SUM(hotels) AS CountryHotels
FROM Tourism
GROUP BY country
HAVING CountryHotels >= (800)
```

ב. 5 נק' - לכל מדינה החזירו את ממוצע מספר המלונות (hotels) לעיר, שם העיר בעלת מספר המלונות הגבוה ביותר ואת מספר המלונות עבור אותה עיר. בנוסף, לכל מדינה החזירו שדה נוסף level המתאר את רמת התירות באותה מדינה. ה-level של מדינה יהיה 1 אם ממוצע מספר המלונות לעיר של המדינה הוא לפחות 1,000, ו-0 אחרת.

ג. 3 נק' - ג'קי, מהנדס בכיר האחראי על אוסף Tourism, קיבל הנחיה מהמנהל שלו לשנות האוסף כך שיהיה ניתן להסיר מדינה וכל עריה מן המאגר ע"י מחיקה של מסמך יחיד. הצע שינוי מבני לאוסף Tourism אשר יאפשר לג'קי להיענות להנחית המנהל.



## 10. RDF – 12 נק' - שאלת בחירה מבין 7, 8, 9, ו-10.

נתון גרף ה-RDF הבא אשר נסמנו ב-G.

@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#>.

@prefix ex: <http://example.maman.cs.technion/>.

@prefix dbo: <http://dbpedia.org/ontology/>.

ex:BestFish	ex:servesDessert	ex:Cheesecake
ex:BestSalads	ex:servesDessert	ex:Knafeh
ex:BestSalads	dbo:city	ex:Haifa
ex:serves	rdfs:domain	dbo:restaurant
ex:serves	rdfs:range	dbo:food
ex:servesDessert	rdfs:subPropertyOf	ex:serves

א. 3 נק' – בהינתן שתי קבוצות ההשמות הבאות:

$A = \{ (?r \rightarrow \text{ex:BestFish}, ?d \rightarrow \text{ex:Cheesecake}) , (?r \rightarrow \text{ex:BestSalads}) \}$

$B = \{ (?s \rightarrow \text{ex:BestFish}, ?d \rightarrow \text{ex:Cheesecake}) , (?r \rightarrow \text{ex:BestFish}, ?d \rightarrow \text{ex:Knafeh}) \}$

שימו לב, המשתנים הם  $?r, ?s$  ו- $?d$ .

כתבו את תוצאת הצירוף הטבעי של A ו-B, כלומר, תוצאת  $A \bowtie B$ .

ב. 6 נק' - בהינתן הגרף G שתואר, כתבו את הפלטים עבור הרצת שאילתת ה-SPARQL הבאה ע"י מנוע

שאינו תומך ב-RDFS וע"י מנוע שתומך ב-RDFS.

```
SELECT ?r ?d {
  ?r ex:servesDessert ?d.
  MINUS {
    ?s dbo:city ex:Haifa.
    OPTIONAL {
      ?s ex:servesDessert ?d.
      ?s rdf:type dbo:restaurant.
    }
  }
}
```

ג. 3 נק' – הסבר במילים את משמעות השאילתה הבאה:

```
SELECT ?r ?stars {
  ?r rdf:type dbo:restaurant.
  ?g ex:mentions ?r.
  GRAPH ?g {
    ?r ex:awarded ?stars.
  }
}
```