

סמסטר חורף התשע"ו

פרופ' בני קימלפלד

מרצה :

מר רוני ליכר
גב' הדר פרנקל
מר נמרוד רייפר

מתרגלים :

מערכות מסדי נתונים**236363**

מועד א'

(כ"ח בשבט התשע"ו, 7 בפברואר 2016)

פירוט החלקים והניקוד:

מס'	נושא	מס' שאלות בחלק	מס' שאלות שיש לענות עליהן	ניקוד
1	שאלות מידע רלציוניות	4	3	34
2	תכן מסדי נתונים	3	3	33
3	מודלים לא רלציוניים	4	3	33
סה"כ		11	9	100

הנחיות לנבחנים

1. כתבו את התשובות אך ורק במחברת התשובות. טופס הבחינה לא ייבדק.
2. מותר לכתוב את התשובות בעפרון, באחריותכם לוודא שתשובתכם קריאה.
3. כל חומר עזר הכתוב על נייר בלבד מותר בשימוש.
4. אין לקבל או להעביר חומר עזר כלשהו בזמן הבחינה.
5. בבחינה שלוש שאלות עם אפשרויות בחירה בשאלות 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 שאלית.**
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 11 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 – 34 נק'

להלן מערכת המתארת אנשים, אירועים שהתרחשו, ותצפיות על אנשים שנצפו באירועים. לשם כך, המידע חולק לטבלאות בצורה הבאה:

אירוע, מוגדר ע"י מיקום, זמן ותיאור. $\text{Event}(\text{place}, \text{time}, \text{description})$

אדם, מוגדר ע"י ת.ז, גיל וגובה. $\text{Person}(\text{ID}, \text{age}, \text{height})$

קישור אדם לאירוע, מוגדר ע"י ת.ז, מיקום, זמן ופעולה שביצע. $\text{Connection}(\text{ID}, \text{place}, \text{time}, \text{action})$
הקו התחתון משמש, כרגיל, לסימון מפתח.

1. SQL, 15 נק' – שאלת חובה.

א. בסעיף זה עליכם לכתוב שאלות SQL המחזירות את כל האירועים אליהם נקשרו לפחות k אנשים (k הינו פרמטר של השאלתה, לדוגמא " $WHERE \text{age} < k$ ", אפשר להחליפו עם כל מספר טבעי אי שלילי).

עליכם לכתוב 2 שאלות בצורות שונות, כפי שיתואר בהמשך, אשר מחזירות את אותו הפלט.

1. 4 נק' - כתבו שאלת SQL המחזירה את הפלט ולא משתמשת ב-HAVING (ניתן להשתמש בקיון שאלות).

2. 4 נק' - כתבו שאלת SQL המחזירה את הפלט, משתמשת ב-HAVING בצורה הגיונית ואינה מקננת שאלות.

ב. 7 נק' - כתבו שאלת SQL המחזירה את כל האנשים שנצפו רוקדים זומבה (מתואר ע"י $\text{description} = \text{"Maman"}$ action="Zumba" בקישור) במהלך בחינה כלשהי בממ"ן (מתואר ע"י $\text{description} = \text{"Maman"}$ באירוע) והתקיים שהם היו הצעירים ביותר מבין הנבחנים באותה הבחינה.

- מספיק להחזיר את תעודת הזהות של אותם אנשים.

2. RA, 10 נק' – שאלת חובה.

א. 2 נק' - כתבו שאלת RA המוצאת את האירועים שלא השתתפו בהם אנשים כלל.

ב. 3 נק' - כתבו שאלת RA המחזירה את האנשים שהיו בכל אירוע שהתקיים בירושלים.

ג. 5 נק' – הוכיחו בצורה פורמלית כי פעולת ההטלה הינה בלתי תלויה באופרטורים מהקבוצה $\{\times, \setminus, \rho, \cup, \sigma\}$.

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

3. RC, 9 נק' – שאלת בחירה מבין 3 ו-4.

א. 6 נק' – כתבו שאילתת RC המחזירה את האנשים הצעירים ביותר מבין בעלי הגובה המקסימלי. על השאילתה להחזיר רשומות מהצורה (ID, age) .

לצורך פתירת הסעיף אתם רשאים להשתמש גם ביחס $R_{\leq}(x, y) = \{(x, y) | x \leq y\}$.

ב. 3 נק' – באחת ההרצאות אלמה אמרה ליוסף:

"לתחשיב היחסים יש את אותו כוח הביטוי כמו לאלגברה הרלציונית, כפי שנלמד בהרצאות."

יוסף ענה לה שזה לא נכון, ונימק בדוגמא הבאה:

$$\{(x_1, \dots, x_n) | \neg R(x_1, \dots, x_n)\}$$

הסבר את המקור לסתירה בין טענות הסטודנטים, והסבר באיזה אופן אמירתה של אלמה תקפה.

4. Datalog, 9 נק' – שאלת בחירה מבין 3 ו-4.

א. 6 נק' – בסעיף זה נגדיר מסלול חברתי בין אדם p ל- p' אם קיים מסלול $p_1 - p_2 - \dots - p_n$ כך ש- $p = p_1$ ו- $p' = p_n$ ומתקיים התנאי הבא: עבור כל זוג אנשים p_i, p_{i+1} ולכל $1 \leq i \leq n - 1$ מתקיים שקיים אירוע אשר שניהם השתתפו בו.

כתבו תכנית $Datalog^{\neg}$ (Datalog עם שלילות) הניתנת לריבוד, אשר מחזירה את כל זוגות האנשים (ID_x, ID_y) אשר מתקיים שקיים ביניהם מסלול חברתי אך מתקיים שהם לא נכחו באף אירוע יחד.

בנוסף, הראו ריבוד חוקי לתוכנית.

ב. 3 נק' – נתונים חוקי ה-EDB בשני מסדי נתונים שונים:

DB Left: A(1,2).	DB Right: A(1,2). A(2,1).
---------------------	---------------------------------

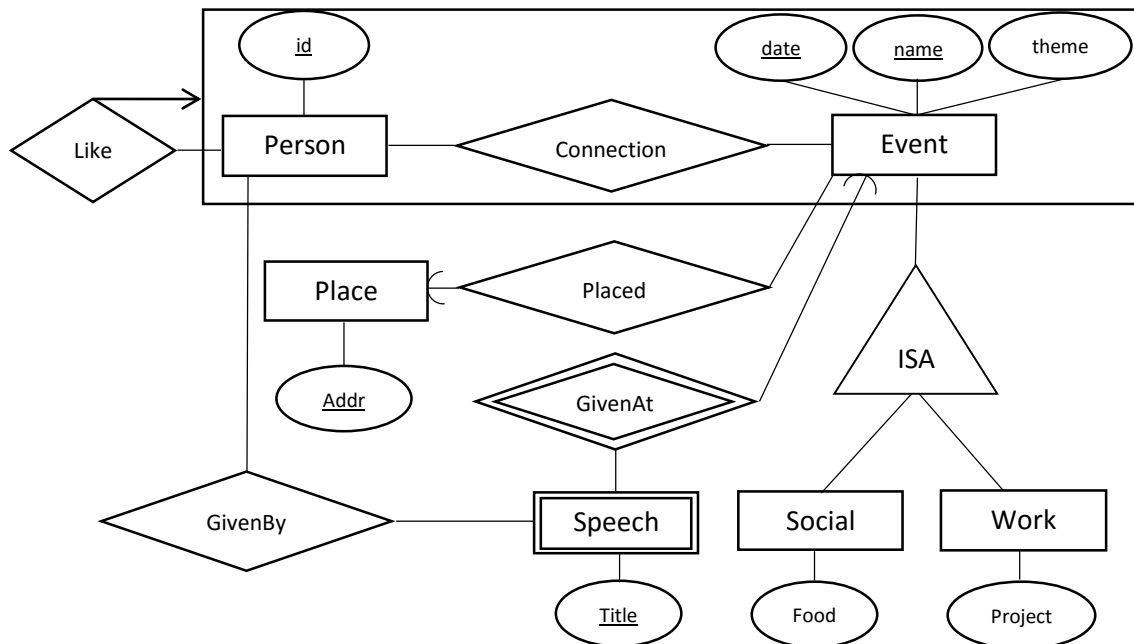
כתוב תכנית $Datalog^{\neg}$ (Datalog עם שלילות) אשר בהרצתה על DB Left תחזיר $B(1,2)$ ובהרצתה על DB Right תחזיר קבוצה ריקה.

בנוסף, האם ניתן לכתוב תכנית Datalog ללא שלילות אשר תספק את אותן התוצאות? אם כן כתבו את התכנית, אחרת נמקו מדוע לא ניתן.

חלק 2 – תכן מסדי נתונים - 33 נק'

5. ERD, 11 נק' – שאלת חובה.

להלן מערכת המתארת אנשים, אירועים שהתרחשו ותצפיות על אנשים שנצפו באירועים. לאירועים יש סיווגים שונים, אם הם חברתיים הם מכילים אוכל ואם הם קשורים לעבודה, מתואר שם פרויקט שקשור לאירוע. אירועים מקושרים למקום בו הם מתקיימים. יחס ה-like מתאר שאדם אוהב השתייכות של אדם לאירוע כלשהו. באירועים ניתנים נאומים ע"י אנשים.



- א. 2 נק' - תארו את הטבלאות הנוצרות ע"י Speech ו-GivenBy בצורה שמשמרת את תלויות הדיאגרמה בצורה המיטבית. סמנו מפתחות בקווים תחתונים.
- ב. 2 נק' - תארו את הטבלאות הנוצרות מתרגום ישיר של קשר ה-Like. סמנו מפתחות בקווים תחתונים.
- ג. 1 נק' – האם אדם יכול לנאום באירוע שהוא אינו מקושר אליו?
- ד. 1 נק' - האם יכולים להיות במסד שני נאומים עם אותה הכותרת?
- ה. 1 נק' - האם אירוע חברה שניתן באותו התאריך ותחת אותו השם יכול להכיל יותר מפרויקט אחד?
- ו. 2 נק' – במידה ובמאגר נתונים קיימים k אירועים, מהו המספר המינימלי והמקסימלי של מקומות ששמורים במערכת? הסבירו.
- ז. 2 נק' - תארו במילים כיצד יש לשנות את הדיאגרמה על מנת לתמוך בתוספת הבאה: לכל אדם נשמור עד 5 מקומות בהם היה גר (Place). - פתרונות מסורבלים יזכו לניקוד חלקי.

6. תלויות פונקציונאליות, 10 נק' – שאלת חובה.

א. 6 נק' – נתונה הסכמה $R(A, B, C, D, E)$ וקבוצת התלויות $F = \{A \rightarrow B, AC \rightarrow DB, C \rightarrow AB, B \rightarrow D, DE \rightarrow A\}$. הראו כיסוי מינימלי כלשהו עבור F (אין חובה להראות איך הגעתם אליו) והוכיחו כי זהו אכן כיסוי מינימלי.

ב. 4 נק' – תהיינה סכמה (U, F) ונתונה תת קבוצה W של U . בנוסף נתונה X תת קבוצה של W . הוכיחו או הפריכו כל אחת מהטענות הבאות :

1. אם X הוא מפתח של (U, F) אז X הוא מפתח של $(W, F[W])$.

2. אם X הוא מפתח של $(W, F[W])$ אז X הוא מפתח של (U, F) .

*בתרגול ניתן הסימון האלטרנטיבי $\pi_W F$ ל- $F[W]$.

7. פירוקים וצורות נורמליות 12 נק' – שאלת חובה.

א. 4 נק' – תהי (U, F) סכמה, ו U_1, U_2 פירוק של (U, F) . נתון כי (U, F) היא ב BCNF. הוכיחו/הפריכו: הסכמות $(U_1, F[U_1])$ ו- $(U_2, F[U_2])$ הינן ב-BCNF.

ב. 3 נק' – הוכיחו שפירוק ל 3NF בעזרת האלגוריתם שראיתם בכיתה הוא משמר תלויות.

ג. 5 נק' – נתונה הסכמה (U, F) , עבור $U = \{A, B, C, D\}$ ו- $F = \{A \rightarrow B, B \rightarrow C, D \rightarrow A\}$, ופירוק $\{AB, BC, CD\}$. עליכם להראות הפעלה של האלגוריתם הנלמד בהרצאה לבדיקה האם הפירוק משמר מידע, ולהחליט בהתאם האם הפירוק משמר מידע או לא.

חלק 3 – מודלים לא רלציוניים - 33 נק'

שימו לב, יש לענות בדיוק על שלוש שאלות מבין 8 (XML), 9 (Neo4j), 10 (MongoDB) ו-11 (RDF).

8. XML – 11 נק' - שאלת בחירה מבין 8, 9, 10 ו-11.

נתון ה-DTD הבא המתאר כוח אדם בחברה כלשהי עם צומת מסמך hr:

```
<!DOCTYPE hr [  
  <!ELEMENT hr (employee*)>  
  <!ELEMENT employee (name,((officeNumber,phone,fax)|(phone,fax?)))>  
    <!ATTLIST employee id ID #REQUIRED>  
    <!ATTLIST employee company CDATA "Technion">  
    <!ATTLIST employee manager IDREF #IMPLIED>  
  <!ELEMENT name (#PCDATA)>  
  <!ELEMENT officeNumber (#PCDATA)>  
  <!ELEMENT phone (#PCDATA)>  
  <!ELEMENT fax (#PCDATA)>  
>]
```

א. 4 נק' – עבור כל אחד ממסמכי ה-XML הבאים, החליטו אם הם מקיימים את כל תנאי ה-DTD. במידה ולא, הראו תנאי אחד שמופר.

a)	<pre><employee id="e2" company="ASAT"> <name>Anna</name> <phone>0529516462</phone> </employee></pre>
b)	<pre><hr> <employee id="e2" company="ASAT" company="Technion"> <name>Anna</name> <phone>0529516462</phone> </employee> <employee id="e3"> <name>Amir</name> <officeNumber>132</officeNumber> <phone>I forgot my phone number, sorry</phone> </employee> </hr></pre>
c)	<pre><hr> <employee id="e2" company="ASAT"> <name>Anna</name> <fax>0529516463</fax> <phone>0529516462</phone> </employee> </hr></pre>

d)	<pre> <hr> <employee id="e2" company="ASAT"> <name>Anna</name> <phone>0529516462</phone> </employee> <employee id="e3" manager="e3"> <name>Amir</name> <officeNumber>132</officeNumber> <phone>0454221891</phone> <fax>134</fax> </employee> </hr> </pre>
----	---

ב. 4 נק' – כתבו מסמך XML יחיד אשר מקיים את ה-DTD ובנוסף מתקיים שעבור כל אחת משאילות ה-XPath הבאות קיימים במסמך לפחות 2 עובדים (employee) כך שאחד מוחזר ואחד לא. במידה ועבור שאילתה כלשהי (או שאילות כלשהן) לא יתכן מסמך כזה, נמקו מדוע. על המסמך לקיים את התנאי עבור השאילות שמאפשרות זאת.

- a) //employee[./@manager]
- b) //employee[//@manager]
- c) //employee[@company="ASAT"]
- d) //employee[@company]

ג. 2 נק' – כתבו שאילתת XPath 2.0 אשר מחזירה את כל המנהלים אשר מקיימים את התנאי הבא: לכל עובד ישיר של המנהל/ת, קיים משרד.

הבהרה: יש להחזיר אלמנטים ולא מזהים.

ד. 1 נק' – הציעו שינוי ל-DTD אשר יאפשר היררכית ניהול ברורה: לכל עובד קיים בדיוק מנהל אחד ולא קיימים מעגלים בשרשרת הניהול (כלומר, מתקבלת היררכית ניהול בצורת יער, קבוצת עצים).

9. Neo4j – 11 נק' - שאלת בחירה מבין 8, 9, 10 ו-11.

נתון מסד הנתונים הגרפי עבור אירועים המכיל את הצמתים הבאים:

person	event	place
id	time	city
age (in years)	description	address
height		

צמתים אלו יכולים להיות מקושרים ע"י שני סוגי קשרים (relationship):

Connection: קשר המקשר בין אדם (person) ואירוע (event) ומכיל את התכונה action.

Placed: קשר המקשר בין אדם (person) ומיקום (place) ואינו מכילו תכונות כלל.

א. 5 נק' - כתבו שאלתת Cypher המוצאת אנשים שהשתתפו באירועים (היו קשורים ל-event עם action והערך attends) והם היו בטווח של עד 5 שנים מממוצע הגילאים של המשתתפים באותו האירוע.

ב. 5 נק' – נגדיר שזוג אנשים מכירים זה את זה אם השתתפו באותו אירוע. משה (אדם בעל ת.ז. 123) מעוניין להכיר אנשים שהוא מכיר באופן טרנזיטיבי כך שבינו לבין האנשים הללו קיימים **לפחות 4 אנשי ביניים**. בנוסף משה ואותו איש המתאים להגדרה, גרים בערים שונות.
הבהרה: משה אינו מעוניין להכיר אנשים שהוא מכיר באמצעות פחות מ-4 אנשי ביניים.

ג. 1 נק' - תוך התייחסות לתשובתכם מסעיף ב', הסבירו כיצד תשובתכם מראה את היתרון של מסדי נתונים גרפים על SQL?

10. MongoDB – 11 נק' - שאלת בחירה מבין 8, 9, 10 ו-11.

לרשותך בסיס הנתונים 'db' המכיל אוסף בודד המתאר שחקנים הנקרא 'players'. כל מסמך באוסף 'palyers' הוא מהצורה הבאה:

```
{
  "_id" : <ObjectId>
  "name" : <string>
  "nationality" : <string>
  "height" : <int>,
  "weight" : <int>,
  "games" : [
    {
      "opponent" : <string>
      "duration" : <int>,
      "result" : <string>
    },
    ...
    {
      "opponent" : <string>
      "duration" : <int>,
      "result" : <string>
    }
  ]
}
```

להלן פירוט הסמנטי של כל אחד מן השדות:

- '_id' - מתאר את מזהה המסמך.
 - 'name' - שם השחקן.
 - 'nationality' - אזרחות השחקן.
 - 'height' - גובה השחקן בס"מ.
 - 'weight' - משקל השחקן בק"ג.
 - 'games' - מערך של משחקי השחקן כאשר כל איבר במערך בעל הפירוט הסמנטי הבא:
 - 'opponent' - היריב מולו שיחק השחקן.
 - 'duration' - משך הזמן שארך המשחק בדקות.
 - 'result' - תוצאת המשחק עבור השחקן. לתוצאת המשחק ביחס לשחקן שני ערכים אפשריים. 'L' המתאר הפסד, 'W' עבור ניצחון.
- להלן דוגמא של מסמך הקיים במאגר 'players'.

```
{
  "_id" : ObjectId("5691048096b75aa53104b939"),
  "name" : "Roger Federer",
  "nationality" : "SWZ",
  "height" : 175,
  "weight" : 71,
  "games" : [
    {
      "opponent" : "Novak Djokovic",
      "duration" : 160,
      "result" : "L"
    },
    {
      "opponent" : "Rafael Nadal",
      "duration" : 200,
      "result" : "W"
    }
  ]
}
```

מן המסמך בדוגמא, נסיק כי הוא מתאר את השחקן ששמו 'Roger Federer', בעל אזרחות שווייצרית על פי הסימון 'SWZ', גובהו של השחקן 175 ס"מ ומשקלו 71 ק"ג. עוד נסיק כי השחקן שיחק שני משחקים, מול השחקן 'Novak Djokovic', במשחק הראשון הוא הפסיד והמשחק ארך 160 דקות. במשחק השני הוא ניצח והמשחק ארך 200 דקות כאשר יריבו הפעם היה 'Rafael Nadal'.

ענה על השאלות הבאות ללא שימוש בקוד JavaScript.

א. 3 נק' - לכל אזרחות החזר את כמות השחקנים במאגר המשחקים תחת אזרחות זו. יש להחזיר רשימה אשר תכיל מסמך לכל אזרחות ואת כמות השחקנים בעלי אותה האזרחות. את הרשימה יש למיין בסדר יורד, כלומר מהגבוה לנמוך לפי ערך סך כל השחקנים.

ב. 4 נק' - כתוב הסבר מפורט על השאילתה הבאה תוך התייחסות לנקודות הבאות:

- פרט את השלבים בביצוע השאילתה תוך הסבר התוצאות המוחזרות לאחר כל אחד מן השלבים.
- בנוסף, הסבר מה מבצעת פעולת ה-'unwind' באופן כללי.

```
db.getCollection('players').aggregate([
  { $unwind: '$games' },
  { $project: { 'name':1, 'nationality':1, 'points': {
    $cond: { if: { $eq: [ "$games.result", 'W' ] }, then: 3, else: 1 }
  }
  },
  },
  { $group: {
    '_id': '$nationality',
    'a': { $avg: '$points' }
  }
  },
  { $sort: {'a':-1}
  },
  { $limit:1 }
])
```

ג. 4 נק' – כתוב שאילתת MongoDB אשר לכל nationality מחזירה את כמות המשחקים שאורכם פחות מ-100 דקות שבהם שחקן מ-nationality זה ניצח.

11. RDF – 11 נק' - שאלת בחירה מבין 8, 9, 10 ו-11.

נתון גרף ה-RDF הבא אשר נסמנו ב-G.

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

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

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

ex:Ahuva	ex:follows	ex:Basma
ex:Ahuva	ex:follows	ex:Denis
ex:Ahuva	rdf:type	dbo:Artist
ex:Basma	ex:follows	ex:Chen
ex:Basma	rdf:type	dbo:Dancer
ex:Chen	ex:tracks	ex:Erin
ex:Denis	rdf:type	ex:Member
ex:follows	rdfs:domain	ex:Member
ex:follows	rdfs:range	ex:Member
ex:tracks	rdfs:subPropertyOf	ex:follows

א. 2 נק' - הסבירו את ההבדל בין הסמנטיקה של פעולת ה-MINUS ב-SQL לבין הסמנטיקה של פעולת ה-MINUS ב-SPARQL.

ב. 6 נק' (2, 2, 2) - בהינתן הגרף G שתואר, כתבו את הפלט עבור הרצת שאילתות ה-SPARQL הבאות ע"י מנוע שאינו תומך ב-RDFS. כלומר, חשבו את $P(G)$ עבור כל אחת מהתבניות a, b ו- c .

a)	SELECT DISTINCT ?p WHERE { { ?p1 ex:follows ?p. OPTIONAL { ?p ex:follows ?p2. } }. ?p2 rdf:type ex:Member. }
b)	SELECT DISTINCT ?p ?t WHERE { { ?p1 ex:follows ?p. ?p1 rdfs:type dbo:Artist. OPTIONAL { ?p rdfs:type ?t. } } MINUS { ex:Ahuva rdfs:type ?t. } }

c)	SELECT DISTINCT ?p1 ?p2 WHERE { ?p1 ex:follows ?p2. MINUS { ?p2 ex:follows ?p3. ?p3 rdf:type ex:Member. } }
----	---

ג. 3 נק' - בהינתן הגרף G שתואר, כתבו את הפלט עבור הרצת שאילתת a - ע"י מנוע שתומר ב-RDFS. כלומר, חשבו את $P^{RDFS}(G)$ עבור התבנית a שבסעיף הקודם. הסבירו את ההבדלים.