

תרגיל בית מס' 3

נושאי התרגיל: תלויות פונקציונליות, צורות נורמליות, MongoDB

- מתרגלת אחראית: שיר רוטמן.
- ההגשה **בזוגות ומוקלדת** בפורמט PDF בלבד, עד לתאריך 22/6/2023.
- שאלות על התרגיל יש לשאול בפורום הייעודי בפיאצה:
<https://piazza.com/technion.ac.il/spring2023/236363>
- שאלות אדמיניסטרטיביות יש להפנות לאופיר פדר.

פרטי המגשים:

ת.ז.	
1.	
2.	

שאלה 1 (25 נק')

נתונה סכמה $R = \{A, B, C, D, E, G, H\}$ וקבוצת תלויות פונקציונליות
 $F = \{AH \rightarrow B, AGH \rightarrow C, BD \rightarrow C, DH \rightarrow EC, C \rightarrow E, CG \rightarrow ABH\}$
בנוסף, נתון פירוק $X = \{R_1(B, C, D), R_2(A, C, E, H), R_3(C, D, G, H)\}$

- האם X ב-BCNF? הוכיחו/נמקו
- האם X ב-3NF? הוכיחו/נמקו
- הראו כיצד האלגוריתם לבדיקת שימור מידע שנלמד בכיתה פועל על פירוק זה, ובהתאם לתוצאה קבעו האם X משמר מידע או לא
- מצאו דוגמה למסד נתונים $r \models F$ שמקיים את המסקנה שהגעתם אליה בסעיף הקודם, והראו זאת עפ"י הגדרת שימור המידע (תזכורת: פירוק הוא משמר מידע אם לכל יחס r מעל R כך ש- $r \models F$, מתקיים $(\bigwedge_{i=1 \dots n} \pi_{R_i}(r) = r$
- האם X משמר תלויות? הראו ע"פ האלגוריתם שנלמד בכיתה

שאלה 2 (20 נק')

נתונה סכמה $R = \{A, B, C, D, E, G\}$ וקבוצת תלויות פונקציונליות
 $F = \{AC \rightarrow D, BD \rightarrow G, G \rightarrow ABCD\}$

- מצאו שני מפתחות קבילים שונים של R . עבור אחד מהם הראו כיצד הוא מתקבל מהרצת האלגוריתם למציאת מפתח.
- מצאו פירוק X משמר מידע של R , כך ש- X הוא מהצורה $X = \{R_1(\circ, \circ, \circ), R_2(\circ, \circ, \circ), R_3(\circ, \circ, \circ)\}$ ו- $R_1 \cap R_2 \cap R_3 = \emptyset$ והוכיחו ע"פ האלגוריתם שהוא אכן משמר מידע. אין צורך להראות כיצד מצאתם את הפירוק.
- שימו לב – פתרונות נכונים (במובן של שימור מידע) המשתמשים ב-2 תתי-סכמות יקבלו 60% מהניקוד ופתרונות נכונים המשתמשים ביותר תתי-סכמות או בתתי-סכמות גדולות יותר יקבלו 80% מהניקוד (פתרונות נכונים עם 3 תתי-סכמות קטנות יותר יקבלו ניקוד מלא)

הסעיפים הבאים בלתי תלויים בסעיפים הקודמים של השאלה.

תהי סכמה $R = (A_1, A_2, \dots, A_n)$ וקבוצת תלויות פונקציונליות F הכוללת תלות אחת בלבד f .

- הוכיחו/הפריכו: R היא ב-BCNF אם"מ לכל $i \in \{1, \dots, n\}$ מתקיים ש- A_i מופיע ב- f לפחות פעם אחת או ש- f טריוויאלית.
- הוכיחו/הפריכו: F היא כיסוי מינימלי של R אם"מ כל A_i מופיע ב- f בדיוק פעם אחת.

שאלה 3 (25 נק')

נתונה סכמה $R = \{A, B, C, D, E, G, H\}$ וקבוצת תלויות פונקציונליות $F = \{B \rightarrow G, BG \rightarrow E, C \rightarrow D, E \rightarrow CH, EH \rightarrow D, G \rightarrow B, G \rightarrow H\}$.

א. מצאו מפתח קביל (אין צורך להראות כיצד הגעתם אליו), והראו כיצד ניתן להסיק ממנו את כל הסכמה R בעזרת אקסיומות ארמסטרונג.

ב. הוכח/הפוך; במקרה של הפרכה, ציינו דוגמה נגדית, כלומר יחס המספק את F אך לא את התלות.

$$F \vdash BG \rightarrow EGH \quad .1$$

$$F \vdash CH \rightarrow AE \quad .2$$

כעת נתונים גם הפירוקים הבאים:

$$X_1 = \{R_1(A, B, G), R_2(B, C, E), R_3(D, E, H)\},$$

$$X_2 = \{R_1(A, G), R_2(B, E, G), R_3(C, D), R_4(D, E), R_5(G, H)\},$$

$$X_3 = \{R_1(A, G), R_2(B, G), R_3(C, D, G), R_4(D, E), R_5(G, H)\}$$

ג. האם X_1 משמר מידע? הוכיחו את תשובתכם.

ד. סמנו ✓ או ✗ בטבלה הבאה בהתאם לפירוקים הנתונים:

	X_2	X_3
BCNF		
3NF		

במידה וסימנתם ✗ באחד התאים, ציינו הפרה של הצורה הנורמלית (תלות שלא מקיימת את התנאים)

שאלה 4 (10 נק')

בהרצאה ראינו כי אקסיומות ארמסטרונג – רפלקסיביות, הכללה וטרנזיטיביות – מהוות מערכת הוכחה נאותה ושלמה.

הסעיפים הבאים מתייחסים למערכת אקסיומות קטנה יותר המורכבת מרפלקסיביות וטרנזיטיביות (ללא הכללה).

א. האם המערכת נאותה? הסבירו בקצרה.

ב. האם המערכת שלמה? אם כן, הסבירו איך תשתנה ההוכחה מההרצאה. אם לא, הוכיחו בעזרת דוגמה נגדית.

שאלה 5 – MongoDB (20 נק')

כידוע לכם, לאחרונה פרצה שביתה במוסדות האקדמיים בישראל על רקע פערים בהסכמי השכר של הסגל הבכיר.

בשאלה זו נבחן מסד נתונים המתאר את חברי הסגל וחלק מהנתונים שנלקחים בחשבון במהלך המו"מ. מסד הנתונים מכיל אוסף (collection) יחיד, SeniorStaff, כאשר כל מסמך (document) באוסף הוא מהצורה הבאה:

```
{
  _id: <ObjectId>,
  staff_member_name: <string>,
  staff_member_id: <string>,
  university: <string>,
  study: <string>,
  salary: <int>,
  estimated_retirement_salary: <int>,
  seniority_year: <int>,
  estimated_retirement_year: <int>
}
```

דוגמה למסמך אפשרי:

```
{
  _id: ObjectId("63d536a10b25ad09196b08e7"),
  staff_member_name: "Ted Mosby",
  staff_member_id: "240707089",
  university: "Tel-Aviv",
  study: "architecture",
  salary: 33,000,
  estimated_retirement_salary: 29,000,
  seniority_year: 2013,
  estimated_retirement_year: 2034
}
```

- א. כתבו שאילתת MongoDB המחזירה את שמות האוניברסיטאות שיש בהן חברי סגל חדשים, כלומר חברי סגל שהפכו לבכירים בשנת 2023.
- ב. כתבו שאילתת MongoDB המחזירה עבור כל אוניברסיטה את מספר חברי הסגל עם ותק של פחות מ-10 שנים (seniority_year אחרי 2013).
- ג. כתבו שאילתת MongoDB המחזירה עבור כל אוניברסיטה וכל שנה בה יש חבר סגל שצפוי לצאת לפנסיה, את ההפרש בין סכום המשכורות הנוכחי של חברי הסגל באוניברסיטה שצפויים לפרוש באותה שנה לסכום המשוערך של המשכורות של אותם חברי סגל בעת הפרישה. מיינו את התוצאות לפי שם האוניברסיטה, בסדר עולה.
- ד. כעת נבדוק באילו אוניברסיטאות המשכורות המשולמות גבוהות משמעותית מהמשכורות המשוערכות לאחר פרישה לגמלאות (כלומר צפוי "חסכון" ניכר לאחר פרישת חברי סגל). נאמר שחבר סגל הוא "יקר" אם המשכורת הנוכחית שלו גבוהה בלפחות 50% מהמשכורת הצפויה שלו בפנסיה. לדוגמה, טד מוסבי (שמתואר במסמך למעלה) הוא לא חבר סגל יקר היות ש- $1.14 \approx \frac{33000}{29000}$, כלומר המשכורת הנוכחית שלו גבוהה ב-14% בלבד מהמשכורת שלו לאחר הפרישה.
- השתמשו ב-MapReduce על מנת להחזיר עבור כל אוניברסיטה את הערך 1 אם יש לה יותר מ-20 חברי סגל יקרים שצפויים לפרוש בין השנים 2025 ל-2030 (כולל) ו-0 אחרת. בסעיף זה ניתן להשתמש בתנאי if של javascript (הסינטקס זהה לתנאי if של C). בנוסף, במידת הצורך ניתן להשתמש באופרטורי ההשוואה המופיעים ב**תיעוד של MongoDB**.