



236363 – מסדי נתונים, אביב 2023

תרגיל בית מס' 1

נושאי התרגיל: ERD, RA, RC, Datalog

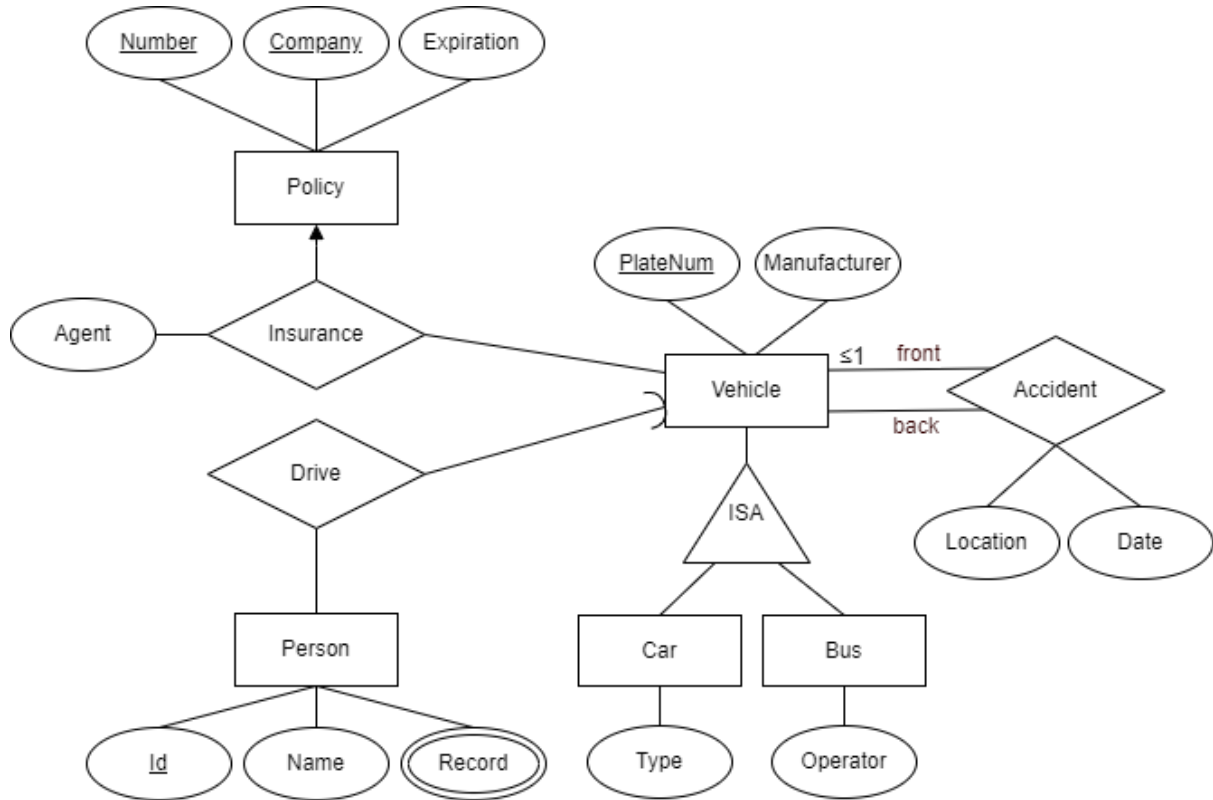
- מתרגלת אחראית: שיר רוטמן.
- ההגשה **בזוגות ומוקלדת** בפורמט PDF ב**לבד**, עד לתאריך 15/5/2023.
- שאלות על התרגיל יש לשאול בפורום הייעודי בפיאצה:
<https://piazza.com/technion.ac.il/spring2023/236363>
- שאלות אדמיניסטרטביות יש להפנות לאופיר פדר.

פרטי המגישים:

ת.ז.	
1.	200819456
2.	313511602

שאלה 1 – ERD (35 נק')

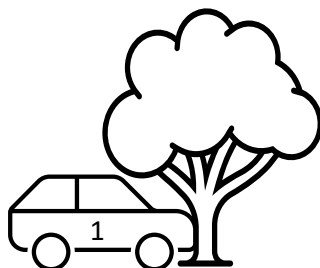
התבוננו בתרשים ה-ERD הבא שתוכנן עבור שמירת מידע על תאונות דרכים:



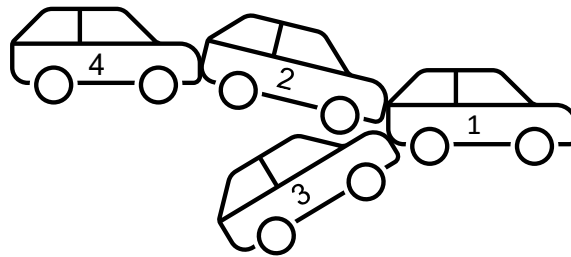
תיאור הישויות בתרשים:

- **Person** – אדם; Id הוא מספר תעודת הזהות שלו, Name הוא שמו ו-Record מחזיק את הסטוריית עבירות התנועה שלו.
- **Policy** – פוליסת ביטוח; Number הוא מספר הפוליסה, Company חברת הביטוח ו-Expiration תאריך פקיעת התוקף שלה.
- **Vehicle** – כלי רכב; PlateNum הוא מספר לוחית הרישוי שלו ו-Manufacturer היצרן.
- **Car** – מכונית; Type הוא סוג המכונית (פרטית, שכורה וכו').
- **Bus** – אוטובוס; Operator מחזיק את החברה שמפעילה אותו.
- front ו-back מציינים את הרכבים הקדמי והאחורי בהתאמה בעת התנגשות/תאונה.
- לוחית רישוי היא מספר חיובי, ובמידה ובתאונה מסויימת אין רכב קדמי, מס' הלוחית של הרכב הקדמי הוא 0 (ניתן להניח שקיים במסד הנתונים Vehicle שמספרו הוא 0).

לדוגמה, התאונה הבאה תיוצג ע"י רשומה יחידה בה הלוחית של הרכב האחורי היא 1 והלוחית של הרכב הקדמי היא 0:



ואילו התאונה הבאה תיוצג למשל ע"י 3 רשומות – אחת בה הלוחית הקדמית היא 1 והאחורית היא 2, אחת בה הלוחית הקדמית היא 1 והאחורית היא 3 ואחת בה הלוחית הקדמית היא 2 והאחורית היא 4:



א. תרגמו את תרשימים ה-ERD לטבלאות המתאימות על-פי הכללים שנלמדו בקורס. עבור כל טבלה, יש לרשום את סכמת הטבלה שתקבל בתרגום, כולל סימון מפתחות בקו תחתון וציון מפתחות זרים.

Policy(Number, Company, Expiration)
 Insurance(PlateNum, Agent, Number, Company) Number & Company are FK
 Person(Id, Name, PlateNum) PlateNum is FK
 Record(Id, Record)
 Vehicle(PlateNum, Manufacturer)
 Car(PlateNum, Type) PlateNum is FK
 Bus(PlateNum, Operator) PlateNum is FK
 Accident(PlateNum1, Location, Date, PlateNum2) PlateNum1 & PlateNum2 are FK

ב. בהתייחס לתרשימים ה-ERD הנתון, עבור כל אחת מהטענות הבאות קבעו האם היא נכונה או לא ונמקו:

1. ייתכן רכב ללא פוליסת ביטוח.

נכון, האילוץ הוא שלכל רכב יש לכל היותר חברת ביטוח אחת, יכול להיות שאין כלל.

2. לכל אוטובוס יש נהג (Person שמקיים את היחס Drive).

לא נכון, לפי אילוצי השרטוט, לכל נהג יש רכב אך לא בהכרח לכל רכב יש נהג.

3. ייתכן רכב עם יותר מנהג אחד.

נכון, אך ההיפך לא נכון, זאת כיוון שהחץ של unique reference מופנה לכיוון הרכב.

ג. כעת רוצים לשנות את התרשים כך שניתן יהיה לדעת גם מי השוטר שהגיע למקום התאונה ומתי. מה הם השינויים המינימליים שיש לעשות בתרשים על מנת להשיג זאת?

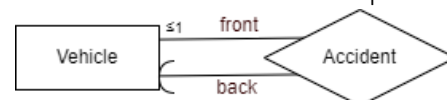
ניתן להפוך את היחס *Accident* לישות ולחבר אותה ל*Person* באמצעות היחס *CopArrived* בעל התכונה *Time*



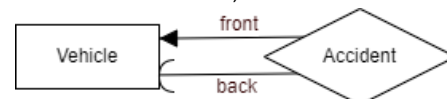
ד. עלה הצורך לשמור את הנתונים בצורה שתאפשר להסיק את הסדר הכרונולוגי בו הרכבים התנגשו בכל תאונה (למשל עבור תאונות שרשרת או תאונות מהצורה שתוארה בדוגמה השניה*), גם במחיר של אובדן המידע על מי התנגש במי**.

* בדוגמה הנתונה, לא ניתן לדעת למשל האם רכב 3 התנגש ברכב 1 לפני שרכב 4 התנגש ברכב 2 או להיפך, או בכלל מי היה הראשון שהתנגש ברכב 1.

** באותה דוגמה, כעת נסכים לקבל מצב בו לא נדע בין מי למי היו ההתנגשויות, אלא נתייחס למכוניות 1-4 כמעורבות באותה תאונה. הניחו שלא ייתכנו שתי תאונות שונות באותו מקום ובאותו תאריך. לשם כך התייעצו עם האחים אליס, בוב וצ'רלי ממך, בוגרי הקורס מסדי נתונים. צ'רלי טוען שהשינוי הבא יענה על הצורך הזה:



בוב טוען שאם נוריד את אילוץ הדדגה שבתרשים הנתון, נוכל גם להימנע מהצורך לאכוף ולתחזק אותו וגם לקיים את הדרישה הנוכחית, ומציע את השינוי הבא:



אליס טוענת ששניהם טועים. מי מהאחים צודק? הסבירו.

אליס צודקת.

הפתרון של צ'רלי אינו נכון כיוון שבמקרה של שני רכבים שעשו תאונה כשהם מאחורי רכב משותף (כמו בדוגמא 2) לא נוכל לשמור זאת בטבלה. הפתרון של בוב אינו נכון כיוון שנוצרת בעיה בתרגום לטבלאות, לא ברור איך להתייחס ל*Accident*.



שאלה 2 – RA (10 נק')

הוכיחו כי אופרטור ההטלה π אינו תלוי באופרטורים $\sigma, \rho, \cup, \times, \setminus$:

- בהינתן הסכמה $R(A,B)$ נראה כי אין שאלתה מעל $\{\sigma, \rho, \cup, \times, \setminus\}$ השקולה ל- $\pi_A(R)$.
נניח שקיימת φ כך ש: $\varphi \equiv \pi_A(R)$ כאשר φ מעל $\{\sigma, \rho, \cup, \times, \setminus\}$.
נראה באינדוקציה כי עבור כל φ כנ"ל מתקיימת התכונה: **מספר ה-attributes הוא לפחות 2**
בסיס: 1. $\varphi = R$, מתקיים כי ב- R יש שתי עמודות בסכמה.
2. $\varphi = \{< A: a, B: b >\}$, יש שני attributes שהם A ו- B .
צעד: בהינתן φ ו- ψ המקיימות את התכונה
1. $\sigma(\varphi) \subseteq R$ מקיימת את התכונה כיוון ש- $\sigma(\varphi) \subseteq R$.
2. $\rho(\varphi)$ מקיימת את התכונה כיוון ששינוי שם של עמודה אינו משנה את מספר העמודות בסכמה.
3. $\varphi - \psi$ מקיימת את התכונה כיוון שהתוצאה מוכלת ב- φ .
4. $\varphi \times \psi$ מקיימת את התכונה כיוון שמספר העמודות בסכמת התוצאה הוא סכום מספר העמודות בסכמות של φ ו- ψ .
5. $\varphi \cup \psi$ מקיימת את התכונה כיוון שהפעולה אינה משנה את הסכמה (שמשותפת לשניהן).
ראינו כי עבור כל φ מעל $\{\sigma, \rho, \cup, \times, \setminus\}$ מתקיימת התכונה, נשים לב כי עבור $\pi_A(R)$ התכונה אינה מתקיימת כי הסכמה של התוצאה היא בעלת עמודה אחת בלבד A . כלומר לא קיימת $\varphi \equiv \pi_A(R)$ כאשר φ מעל $\{\sigma, \rho, \cup, \times, \setminus\}$.

שאלה 3 – RA, RC (30 נק')

בשאלה זו נסתכל על מסד הנתונים שמתואר בשאלה הראשונה (ERD). לכל אורך השאלה ניתן להניח שמסד הנתונים אינו מכיל ערכי null ושכל תאונה מעורבים בדיוק שני רכבים, קדמי ואחורי.

1. בסעיף זה ניתן להניח כי לא קיימות פוליסות ביטוח שתאריך פקיעתן עבר.

א. כתבו שאילתה ב-RA המחזירה את שמות האנשים הנוהגים ברכב שמבוטח בחברה BIG (Company = 'BIG').

$$\pi_{Name} \sigma_{Company=BIG} Person \triangleright \triangleleft Insurance$$

ב. כתבו שאילתה מקבילה ב-RC.

$$\{name \mid \exists i, p : Person(i, name, p) \wedge (\exists m, a, n : Vehicle(p, m) \wedge Insurance(p, a, n, BIG))\}$$

2.

א. כתבו שאילתה ב-RA המחזירה את הסוכנים שלא מבטחים **מכוניות** שהיו מעורבות בתאונה.

$$\pi_{Agent} Insurance - \pi_{Agent} (Insurance \triangleright \triangleleft (\pi_{PlateNum} Vehicle \triangleright \triangleleft \pi_{PlateNum} Accident))$$

ב. כתבו שאילתה מקבילה ב-RC.

$$\{a \mid \exists p, n, c : Insurance(p, a, n, c) \rightarrow \forall l, d, p' : \neg Accident(p, l, d, p') \wedge \neg Accident(p', l, d, p)\}$$

3.

א. כתבו שאילתה ב-RA שמחזירה את התאריכים בהם היתה בדיוק תאונת דרכים אחת.

$$\pi_{Date} Accident - \pi_{Date} \sigma_{PlateNum1 \neq PlateNum_1, PlateNum1 \neq PlateNum2} (Accident \triangleright \triangleleft \rho_{PlateNum1/PlateNum_1} Accident)$$

ב. כתבו שאילתה מקבילה ב-RC.

$$\{d \mid \exists p, l, p' : Accident(p, l, d, p') \rightarrow \forall q, l', q' : \neg Accident(q, l', d, q')\}$$

4. נתונה שאילתת ה-RA הבאה:

$$\pi_{\text{Agent}} \left(\text{Insurance} \bowtie \left(\left(\pi_{\text{Company, Manufacturer}} (\text{Insurance} \bowtie \text{Vehicle}) \right) \div \pi_{\text{Manufacturer Vehicle}} \right) \right)$$

א. הסבירו בקצרה ובמילים פשוטות מה השאילה מחזירה.

השאילתה מחזירה את הסוכן של חברות הביטוח המבטחות את כל יצרני הרכבים שעשו תאונות

ב. האם ניתן לכתוב את השאילתה הזו אך ורק עם האופרטורים $\{ \sigma, \pi, \rho, \times, \cup \}$? אם כן, הראו כיצד. אם לא, הסבירו מדוע.

לא ניתן, השאילתה דורשת שימוש באופרטור ההפרש ולא ניתן לכתוב אותה בלי האופרטור ה"ל" (כפי שראינו בשאלה 2)

5. דן, סטודנט בקורס מסדי נתונים ואחיהם הצעיר של אליס, בוב וצ'רלי, ניסח את שאילתת ה-RC הבאה במטרה להחזיר את הסוכנים שעובדים עם כלל חברות הביטוח:

{agent :

$\forall \text{company} \exists \text{plateNum, number} [\text{Insurance}(\text{plateNum, number, company, agent})]$

האם השאילתה של דן תלויה בתחום? אם כן, הראו דוגמה המוכיחה זאת ותקנו את השאילתה כך שלא תהיה תלויה בתחום. אם לא, נסחו את השאילתה ב-RA.

השאילתה תלויה בתחום.

נגדיר את התחומים $D1 = \{ 'a' \}, D2 = \{ 'a', 'b' \}$

כאשר ב-DB קיימת השורה $\text{Insurance}('a', 'a', 'a', 'a')$ בלבד

השאילתה בתחום הראשון תחזיר את את 'a' כסוכן מתאים אך בתחום השני הוא כבר לא מתאים כיוון שהסוכן ה"ל" לא עובד עם חברת הביטוח 'b' שנמצאת בתחום.



שאלה 4 – Datalog (25 נק')

נתונים הפרדיקטים הבאים (EDB) המייצגים מאגר מידע על סרטי קולנוע:

1. $Movies(Title, Director, Actor)$ – הסרטים שבמאגר; בכל רשומה יש את שם הסרט, במאי הסרט ושחקן המשחק בו.
2. $Cinemas(Theater, Address, Phone)$ – בתי קולנוע; בכל רשומה יש את שם הקולנוע, כתובתו ומספר הטלפון שלו.
3. $Schedule(Theater, Title, Time)$ – לוח הקרנות; בכל רשומה יש שם של בית קולנוע, סרט שמוקרן בו ומועד ההקרנה.

לאורך השאלה, הניחו שלכל סרט יש במאי אחד בלבד.

בסעיפים הבאים תיידרשו לכתוב תכניות Datalog, במידת הצורך עם שלילה. במידה ולתכנית שכתבתם נדרש ריבוד, ציינו זאת וסמנו כל רובד בנפרד, כך שמספר השכבות בריבוד יהיה מינימלי.

בנוסף, במידת הצורך, בכל סעיף ניתן להשתמש בפרדיקטים שהוגדרו בסעיפים קודמים (IDB).

א. כתבו תכנית Datalog המחזירה את השמות והכתובות של בתי הקולנוע שמקרנים סרטים בבימויו של טרנטינו.

```
TarantinoMovie(T) ← Movies(T, "Tarantino", a)
TarantinoTheater(T, A) ← Schedule(T, t, t'), Cinemas(T, A, p), TarantinoMovie(t)
```

ב. כתבו תכנית Datalog המחזירה את הסרטים שטרנטינו ביים אך לא שיחק בהם.

```
TarantinoPlay(M) ← Movies(M, d, "Tarantino")
TarantinoNotPlay(M) ← TarantinoMovie(M), ¬TarantinoPlay(M)
```



ג. כתבו תכנית Datalog המחזירה את הסרטים בהם כל השחקנים שיחקו באיזשהו סרט שביים טרנטינו.

```
PlayTarantinoMovie(A) ← Movies(t, "Tarantino", A)
WithTarantinoActor(M) ← Movies(M, d, a), PlayTarantinoMovie(a)
```



ד. כתבו תכנית Datalog המחזירה את הבמאים שכל אחד מהשחקנים הקיימים שיחק באיזשהו סרט שלהם.

```
DidntWorkWithDirector(A, D) ←  
Movies(t, D, a), Movies(t', d, A), ¬Movies(t', D, A)  
WorkedWithAll(D) ←  
Movies(t, D, a), Movies(t', D, a'), ¬DidntWorkWithDirector(a', D)
```

ה. כתבו תכנית Datalog המחזירה את כל הזוגות של שחקנים שמשחקים בדיוק באותם סרטים.

```
Separated(A, A') ← Movies(t, d, A), ¬Movies(t, d, A')  
WorkOnlyTogether(A, A') ←  
Movies(t, d, A), Movies(t, d, A'), ¬Separated(A, A'), ¬Separated(A', A)
```

ו. נגיד ששחקן א' מקושר לשחקן ב' במידה ושיחק איתו סרט או ששיחק באותו סרט עם שחקן אחר שמקושר אליו. פורמלית, אם נגדיר גרף בו הצמתים הם שחקנים והקשתות הן סרטים בהם שיחקו יחד, שחקנים מקושרים הם אלה שיש מסלול פשוט ביניהם. כתבו תכנית Datalog המחזירה את כל הזוגות של שחקנים שאינם מקושרים.

```
WorkedTogether(A, A') ← Movies(t, d, A), Movies(t, d, A')  
Path(A, A') ← WorkedTogether(A, A')  
Path(A, A') ← Path(A, I), WorkedTogether(I, A')  
NotConnected(A, A') ← ¬Path(A, A')
```