

הפקולטה למדעי המחשבהטכניון – מכון טכנולוגי לישראל

סמסטר חורף התשע"ד

פרופ' רועי פרידמן

מרצה :

מר עומר כ"ץ

מתרגלים :

גבי אלה בולשינסקי

מערכות מסדי נתונים**236363**

מועד א' (י"ג באדר א' התשע"ד, 13 בפברואר 2014)

מס' סטודנט:**פירוט השאלות והניקוד:**

מס'	נושא	ניקוד
1	ERD	25
2	שאליות מידע	25
3	פירוטים ותלויות פונקציונליות	24
4	XML	18
5	NoSQL	8
סה"כ		100

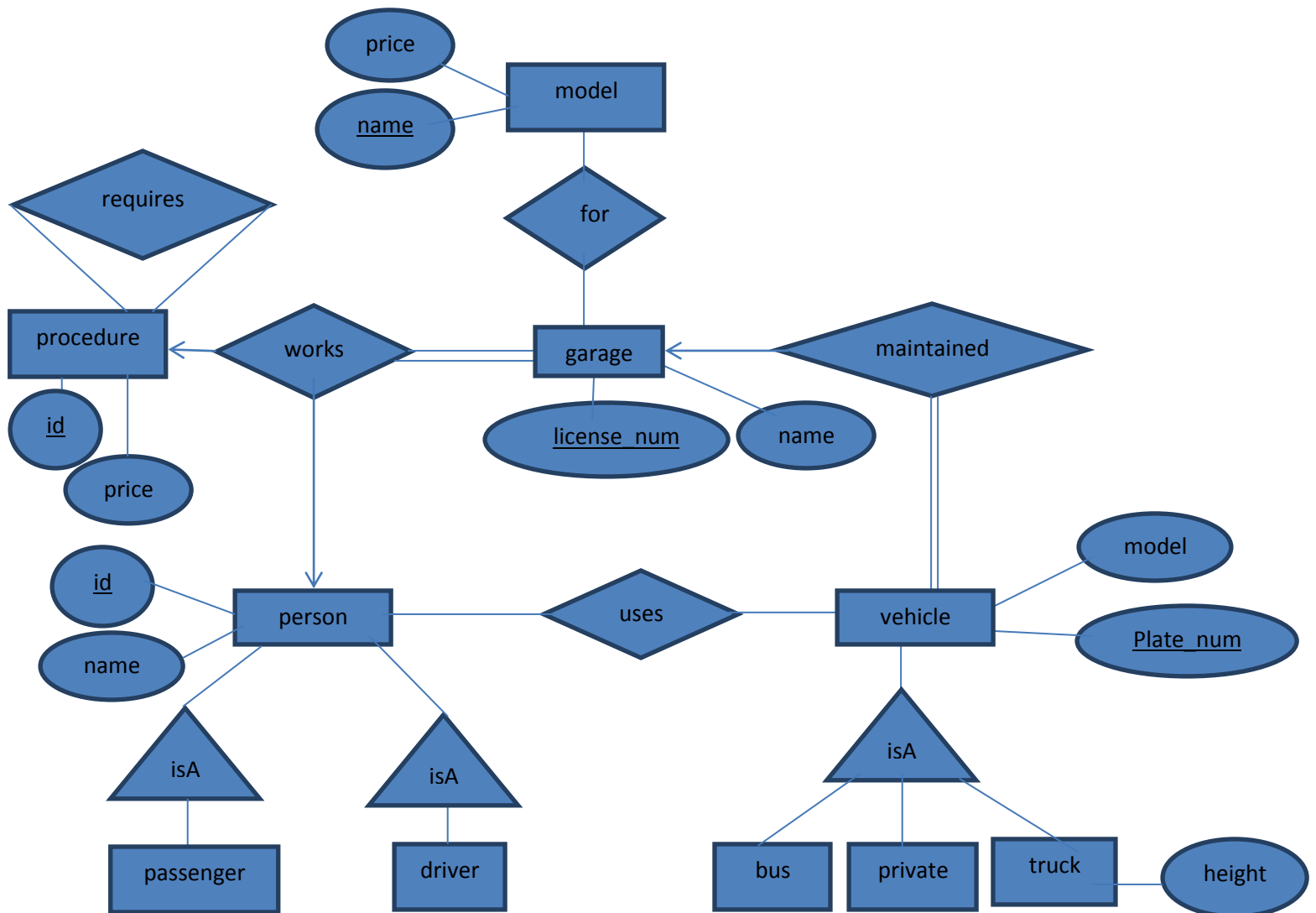
הנחיות לנבחנים

1. כתבו את התשובות אך ורק בטופס הבחינה, המחברת מיועדת לטייטה 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 חמש 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 שאלית.
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 כוללת 9 דפים (כולל דף זה), בהם **17 עמודים**. נא לוודא שיש בידכם את כל הטופס.
11. כאשר ניקוד 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 (25 נק')

הדיאגרמה הבאה מתארת כלי רכב (vehicle) שיכולים להיות משאיות (truck), מכוניות פרטיות (private) ואוטובוסים (bus), את האנשים שנוהגים בהם (driver) או משתמשים בהם כנוסעים (passenger). כמו כן, כלי רכב מטופלים (maintained) במוסכים (garage) בהם מתבצעים טיפולים לרכבים (procedure) ע"י אנשים שעובדים (works) במוסך. בנוסף הדיאגרמה מתארת את דגמי המכוניות (model) המטופלות בכל מוסך.



שאלות:

- א. (10 נק') השלימו את הטבלאות הבאות המתקבלות מתרגום ישיר של ה-ERD.
שימו לב: במידה ויש יותר מאפשרות אחת באחת השורות שעליכם להשלים ציינו את כל האפשרויות.

שם הטבלה: works
(2 נק') שדות: procedure.id, person.id, license_num
(2 נק') מפתחות: license_num או procedure.id, license_num, person.id
(1 נק') תלויות פונקציונליות: procedure.id -> person.id, license_num וגם person.id -> procedure.id, license_num

שם הטבלה: truck
(2 נק') שדות: height, plate_num
(2 נק') מפתחות: plate_num
(1 נק') תלויות פונקציונליות: plate_num -> height

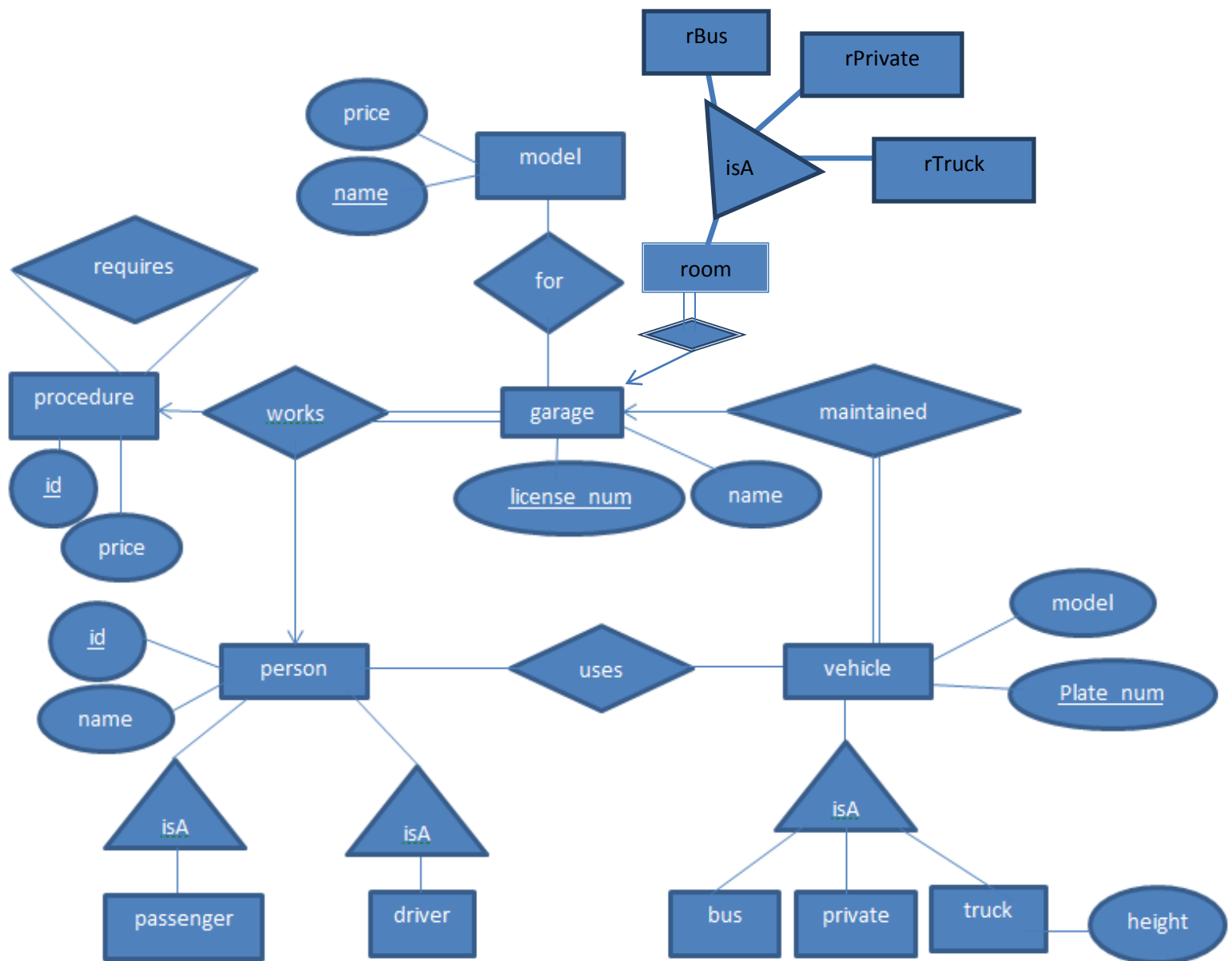
- ב. לכל אחד מהבאים, עליכם לציין האם הוא מתקיים במסד הנתונים (בהתבסס על תרשים ה-ERD) **ואם לא, הציעו תיקון (ניתן לתארו במילים) ל-ERD כך שהאילוץ יתקיים:**
- i. (3 נק') בכל מוסך (garage) מטופל (maintained) רכב אחד לפחות.

לא מתקיים. יש להוסיף קו נוסף לחץ בין maintained ל-garage

- ii. (3 נק') יתכן נהג באוטובוס שהוא נוסע עבור רכב פרטי כלשהו.

נכון. Person יכול להיות גם driver וגם passenger ולהופיע יותר מפעם אחת בקשר uses.

ג. 9 נק') בכל מוסך יש חדרי טיפולים שממוספרים מ-0 ועד מספר חדרי הטיפולים הקיימים במוסך. לכל מוסך המספר הסידורי של החדר מהווה את הזיהוי של החדר במוסך. כאשר בנוסף כל חדר הוא אחד מהבאים: חדר טיפולים לרכבים פרטיים, חדר טיפולים לאוטובוסים או חדר טיפולים למשאיות. הציעו שינוי לדיאגרמה הנתונה על מנת לייצג את החדרים במוסכים. שימו לב, **אין צורך** לקשר בין החדרים לרכבים המטופלים בהם ובין החדרים לטיפולים שמבוצעים בהם.



שאלה 2 – שאילתות מידע (25 נק')

השאלות בסעיף זה מתייחסות לדיאגרמת ה-ERD שניתנה בשאלה 2.

להלן התרגום הישיר של חלק מהטבלאות ב-ERD. יתכן ולא תזדקקו לכל הטבלאות הללו על מנת לפתור את השאלות ויתכן שתזדקקו לטבלאות נוספות. אם תשתמשו בטבלאות נוספות יש לציין את התרגום הישיר של טבלאות הללו:

Requires(<u>id1</u> , <u>id2</u>)	for(<u>license_num</u> , <u>name</u>)
Procedure(<u>id</u> , <u>price</u>)	maintained(<u>plate_num</u> , license_num)
Garage(<u>license_num</u> , <u>name</u>)	model(<u>name</u> , <u>price</u>)
Driver(<u>id</u> , <u>name</u>)	uses(<u>person_id</u> , <u>plate_num</u>)
passenger(<u>id</u> , <u>name</u>)	bus(<u>plate_num</u> , <u>model</u>)
vehicle(<u>plate_num</u> , <u>model</u>)	

Requires(id1,id2) אם פרוצדורה בעלת מזהה id1 דורשת פרוצדורה בעלת מזהה id2.

א. (9 נק') נגדיר "מדד יוקרתיות" של מוסך כעלות ממוצעת (מחיר ממוצע) של דגמי המכוניות בהם מטפל המוסך. הניחו שלכל המוסכים יש "מדד יוקרתיות" שונה.
כתבו שאילתת SQL שתחזיר את:
עשרת המוסכים בעלי "מדד היוקרתיות" הגבוה ביותר.
מותר להשתמש במבטים רק אם משתמשים בהם יותר מפעם אחת.
אין לכלול בתוצאה מוסכים שלא מטפלים באף דגם של מכונית.

```
CREATE VIEW garage_price AS
SELECT garage. license_num, AVG(price) AS avg_price
FROM garage,for,model
WHERE garage. license_num=for. license_num
AND for.name= model.name
GROUP BY garage. license_num

SELECT G. license_num
FROM garage_price G
WHERE 10>(SELECT count(garage_price .license_num)
FROM garage_price
WHERE garage_price.price>G.price)
```

- ב. (8 נק') מוסכים מגדירים נהגי אוטובוסים מצטיינים כנהגים שנוהגים בכל האוטובוסים שמטופלים בהם ולא נוסעים (בתור נוסעים) באוטובוסים שמטופלים במוסכים אחרים. כתבו שאילתת RA שתחזיר:
- למוסך שה- license_num שלו הוא '12345' את מספרי ת.ז. של הנהגים המצטיינים עבורו.**
- כלומר, נהגים שנוהגים בכל האוטובוסים שמטופלים במוסך בעל license_num שמספרו '12345' ולא נוסעים (בתור נוסעים) באוטובוסים שמטופלים במוסכים אחרים.

$$\begin{aligned}
 relevantBuses &= \pi_{plate_num} \left(bus \triangleright \triangleleft \sigma_{license_num='12345'} maintained \right) \\
 irrelevantBuses &= \pi_{plate_num} \left(bus \triangleright \triangleleft \sigma_{license_num \neq '12345'} maintained \right) \\
 drivers &= \pi_{id, plate_num} \left(driver \triangleright \triangleleft \rho_{person_id \rightarrow id} uses \triangleright \triangleleft relevantBuses \right) \div relevantBuses \\
 passengers &= \pi_{id} \left(passenger \triangleright \triangleleft \rho_{person_id \rightarrow id} uses \triangleright \triangleleft irrelevantBuses \right) \\
 result &= \pi_{id} \left(drivers \right) \setminus passengers
 \end{aligned}$$

ג. (8 נק') כתבו שאילתא ב-safeDRC שתחזיר את:

כל מזהי הפרוצדורות שלא דורשות אף פרוצדורה קודמת או שמחירן קטן מ-100 ₪.
הראו שהשאילתא בטוחה.

$$\left\{ \langle id \rangle \vee \left(\left(\left(\exists p (procedure(id, p) \wedge p < 100) \right) \wedge \left(\neg \exists id1, p1 (procedure(id1, p1) \wedge requires(id, id1)) \right) \right) \right) \right\}$$

$procedure(id, p) \in \text{safe atomic}, p < 100 \text{ atomic}$

$\Rightarrow (procedure(id, p) \wedge p < 100)$ safe as id and p are bounded in $procedure$

$\Rightarrow \exists p (procedure(id, p) \wedge p < 100)$ safe

$(procedure(id1, p1) \wedge requires(id, id1))$ safe as it is \wedge between two safe atomic formulas

$\exists id1, p1 (procedure(id1, p1) \wedge requires(id, id1))$ is safe

$(\exists p (procedure(id, p)))$ is safe

$$\left(\left(\exists p (procedure(id, p)) \right) \wedge \left(\neg \exists id1, p1 (procedure(id1, p1) \wedge requires(id, id1)) \right) \right)$$

safe as it is \wedge between one safe formula and a negation of a safe formula

and the only free variable id is bounded in $procedure$.

eventually, \vee is performed on two safe formulas which

share the same free variable id

שאלה 3 – פירוקים ותלויות פונקציונליות (24 נק')

א. (8 נק') עבור סכמה רלציונית R בעלת $N \geq 4$ תכונות וקבוצת תלויות פונקציונלית F המכילה שתי תלויות פונקציונליות לא טריוויאליות, מהו הגודל המינימאלי והמקסימאלי של המפתחות הקבילים של R עבור F כלשהי? הוכיחו. שימו לב עליכם להוכיח ל- R כללי - $R(A_1, A_2, \dots, A_n)$.

גודל מינימאלי: כמובן שלא יתכן מפתח קביל בגודל 0 שכן הסגור של קבוצה ריקה הוא קבוצה ריקה. נראה שיתכן מפתח קביל בגודל 1: לדוגמא עבור $R(A_1, A_2, \dots, A_n)$

$F = \{A_1 \rightarrow A_2, A_1 \rightarrow A_3 \dots A_n\}$ יש להראות ש- A_1 מפתח קביל.

גודל מקסימאלי: לא יתכן מפתח קביל בגודל N שכן קיימת ב- F תלות לא טריוויאלית $X \rightarrow Y$,

ולכן $R \setminus Y \rightarrow R$. נראה שיתכן מפתח קביל בגודל $N-1$: לדוגמא $R(A_1, A_2, \dots, A_n)$ ו- $F = \{A_1 \rightarrow A_2, A_3 \rightarrow A_2\}$ מפתחות קביל עבור R ו- F הוא $R \setminus A_2$ והוא מגודל $N-1$.

ב. (8 נק') האם תשובתך מסעיף א' תשתנה אם נתון ש- F היא מינימאלית?

גודל מקסימאלי לא ישתנה (אותה הוכחה).

גודל מינימאלי הוא $N-2$ נשים לב שאם F מינימאלית שתי התלויות הן מהצורה $X \rightarrow A_i$ ו- $Y \rightarrow A_j$. אזי כל אטריביוט פרט ל- A_i ו- A_j חייב להופיע בכל מפתח קביל, אזי מפתח קביל הוא מגודל $N-2$ לפחות. עבור $R(A_1, \dots, A_n)$ ו- $F = \{A_1 \rightarrow A_2, A_1 \rightarrow A_3\}$ נקבל שמפתח קביל הוא $R \setminus \{A_2, A_3\}$ וגודלו הוא $N-2$.

ג. (8 נק') בהינתן כיסוי מינימאלי G נגדיר כיסוי מינימאלי מאוחד באופן הבא:

$$G' = \{X \rightarrow Y \mid Y = \{A \mid X \rightarrow A \in G\}\}$$

בהינתן כיסוי מינימאלי מאוחד בעל n תלויות, מה הגודל המינימאלי האפשרי עבור פירוק ל-3NF?

הגודל המינימאלי הוא 1.

בהינתן סכמה $R(A_1, A_2, \dots, A_n)$, כיסוי מינימאלי מאוחד אפשרי בעל n תלויות הוא:

$$G' = \{A_1 \rightarrow A_2 A_3 \dots A_n, A_2 \rightarrow A_1, A_3 \rightarrow A_1, \dots, A_n \rightarrow A_1\}$$

נשים לב שבמקרה זה האלגוריתם ליצירת פירוק ל-3NF יצור את n תת הסכמות הבאות:

- $\{A_1, A_2, A_3, \dots, A_n\}$
- $\{A_2, A_1\}$
- $\{A_3, A_1\}$
- ...
- $\{A_n, A_1\}$

מכיוון שכל תת הסכמות מוכלות בתת הסכמה הראשונה, נוכל לוותר על תת הסכמות הללו ולהישאר עם פירוק שמכיל תת סכמה בודדת.

שאלה 4 – XML (18 נק')

נתון ה-DTD:

1. <!ELEMENT registered_garages (garage*) >
2. <!ELEMENT garage (procedure+,car*) >
3. <!ATTLIST garage
4. License_num ID #REQUIRED
5. Name CDATA #REQUIRED>
- 6.
7. <!ELEMENT procedure EMPTY>
8. <!ATTLIST procedure
9. id ID #REQUIRED
10. required_procedures IDREFS #IMPLIED
11. price CDATA #REQUIRED>
- 12.
13. <!ELEMENT car EMPTY>
14. <!ATTLIST car
15. plate_num ID #REQUIRED
16. type CDATA #REQUIRED>

מסד הנתונים מתאר נתונים על מוסכים מורשים, הפרוצדורות שמתבצעות בהם ועל הרכבים המטופלים בהם.

אלמנט השורש במסד הוא אלמנט registered_garages.

לכל מוסך (garage) השמור במסד נשמור את מספר הרישיון שלו License_num ואת שמו Name. בנוסף בכל מוסך מטופלות מכוניות (car) עבורן נשמור את מספר לוחית הזיהוי שלהן (plate_num) ואת סוג הרכב (type).

בכל מוסך גם מתבצעים טיפולים (procedure) שעבור כל אחת מהפרוצדורות נשמור את מזהה הפרוצדורה id, פרוצדורות קודמות אותן יש לבצע לפניה (required_procedures) ואת מחירה (price).

שאלות:

- א. (6 נק') רוצים להוסיף ל-DTD ייצוג לעובדי המוסכים (שיכלול את מספרי הזהות שלהם) כך שכל עובד יכול לעבוד במוסך אחד בלבד ובכל מוסך חייב להיות לפחות עובד אחד. כמו כן, לכל פרוצדורה נרצה להוסיף מידע על העובדים שיודעים לבצע אותה (עובדים אלו לא בהכרח עובדים במוסך הרלוונטי). שנו את ה-DTD על מנת לייצג את התוספת המתוארת.

2. <!ELEMENT garage (procedure+,car*,worker+) >

11. price CDATA #REQUIRED

12. workers IDREFS #REQUIRED>

17. <!ELEMENT worker EMPTY>

18. <!ATTLIST worker

19. IDNum ID #REQUIRED>

- ב. (6 נק') כתבו שאילתת XPath שמחזירה:

את מזהי הטיפולים שעבורם מחירי כל אחת מהפרוצדורות הדרושות (required_procedured) גבוהה ממחיר הפרוצדורה עצמה.

*אין צורך לבדוק יותר מרמה אחת (כלומר, פרוצדורות דרושות לפרוצדורות הדרושות)

```
//procedure[every $p in id(./@required_procedured) satisfies
./@price<$p/@price]/@id
```

ג. (6 נק') כתבו שאילתת XPath שמחזירה:

רשימה של שמות המוסכים ולכל מוסך את רשימת של מספרי לוחיות הזיהוי של המכוניות המטופלות בו.

אם שני מוסכים בעלי אותו השם הם יופיעו בנפרד (כלומר, אין לאחד את מספרי המכוניות בהם לאותה רשימה). לדוגמא, עבור קובץ XML נתון:

```
<?xml version="1.0" encoding="utf-8" ?>
<!DOCTYPE registered_garages SYSTEM "garage.dtd">
<registered_garages>
  <garage License_num='a12345' Name='best garage'>
    <car plate_num='n12-234-56' type='Ford' />
  </garage>
  <garage License_num='a12346' Name='2nd garage'>
    <car plate_num='n11-234-56' type='Nisan' />
    <car plate_num='n10-234-56' type='Shkoda' />
  </garage>
  <garage License_num='a12347' Name='2nd garage'>
    <car plate_num='n01-234-56' type='Nisan' />
  </garage>
</registered_garages>
```

תתקבל התשובה:

```
best garage
n12-234-56
2nd garage
n11-234-56
n10-234-56
2nd garage
n01-234-56
```

```
for $a in //garage
return ($a/@Name, $a/car/@plate_num)
```

שאלה 5 – NoSQL (8 נק')

נתון מסד נתונים גרפי Neo4j המכיל צמתים משלושת הסוגים הבאים (לכל צומת label יחיד):

Student	Lecturer	Course
Name	Name	Name
ID	ID	Catalogue_Number
Address		Syllabus

הקשתות בגרף הן מהסוגים הבאים (לכל קשת label יחיד):

Studies: מחבר בין Student לבין Course ומכיל את התכונות semester ו-grade.

Teaches: מחבר בין Lecturer לבין Course ומכיל את התכונות semester ו-classroom.

שימו לב: תשובות ארוכות ומסורבלות יתר על המידה עלולות לגרום להורדת נקודות.

- א. (4 נק') נגדיר פונקציית מרחק בין שני סטודנטים שונים כדלקמן: (i) הסטודנטים A ו-B במרחק 1 אם הם למדו קורס משותף. (ii) הסטודנטים A ו-B במרחק $n > 1$ אם n הוא המספר הקטן ביותר כך שקיים סטודנט C כך ש-A במרחק $n-1$ מ-C ו-C במרחק 1 מ-B. (iii) אם לא קיים n כזה, נגדיר את המרחק להיות 0.
- כתבו שאילתת Cypher שמחזירה:

המרחק בין שני סטודנטים שתעודת הזהות שלהם (שנניח כי היא יחידה) היא 12345 ו-67890.
 רמז: מומלץ לצייר (בדפי הטיוטא) מבנה טיפוסי של הגרף. בהינתן מסלול ב-Cypher, הפונקציה length מחזירה את אורכו. במידה והמסלול לא קיים, הפונקציה מחזירה 0.

```
MATCH (s1:Student {ID:'12345'}),(s2:Student {ID:'67890'}),p=shortestPath((s1)-[:Studies*]-(s2))
RETURN length(p)/2
```

(4 נק') כתבו שאילתת Cypher שמחזירה :

שמות כל המרצים שלימדו לפחות 3 מקצועות. (ניתן להניח שאין כפילויות בגרף)

```
MATCH (l:Lecturer)-[:Teaches]->(c:Course)
WITH l, count(c) as numcourses
WHERE numcourses>=3
RETURN l.name
```

מקום נוסף לתשובות

אם אתם משתמשים בדף זה, ציינו זאת ליד השאלה/השאלות המקוריות, וציינו כאן את מספר/י השאלה/השאלות.

שאלה: _____ סעיף: _____

שאלה: _____ סעיף: _____

שאלה: _____ סעיף: _____