

מרצה : פרופ' אלדר פישר סמסטר אביב התשע"ד

מתרגלים : גב' אלה בולשינסקי  
מר רוני ליכר

**מערכות מסדי נתונים**  
**236363**

מועד ב' (כ"ז באלול התשע"ד, 22 בספטמבר 2014)

<b>מס' סטודנט:</b>
--------------------

**פירוט השאלות והניקוד:**

מס'	נושא	ניקוד
1	ERD	26
2	שאליות מידע	26
3	פירוקים ותלויות פונקציונליות	20
4	XML	28
<b>סה"כ</b>		<b>100</b>

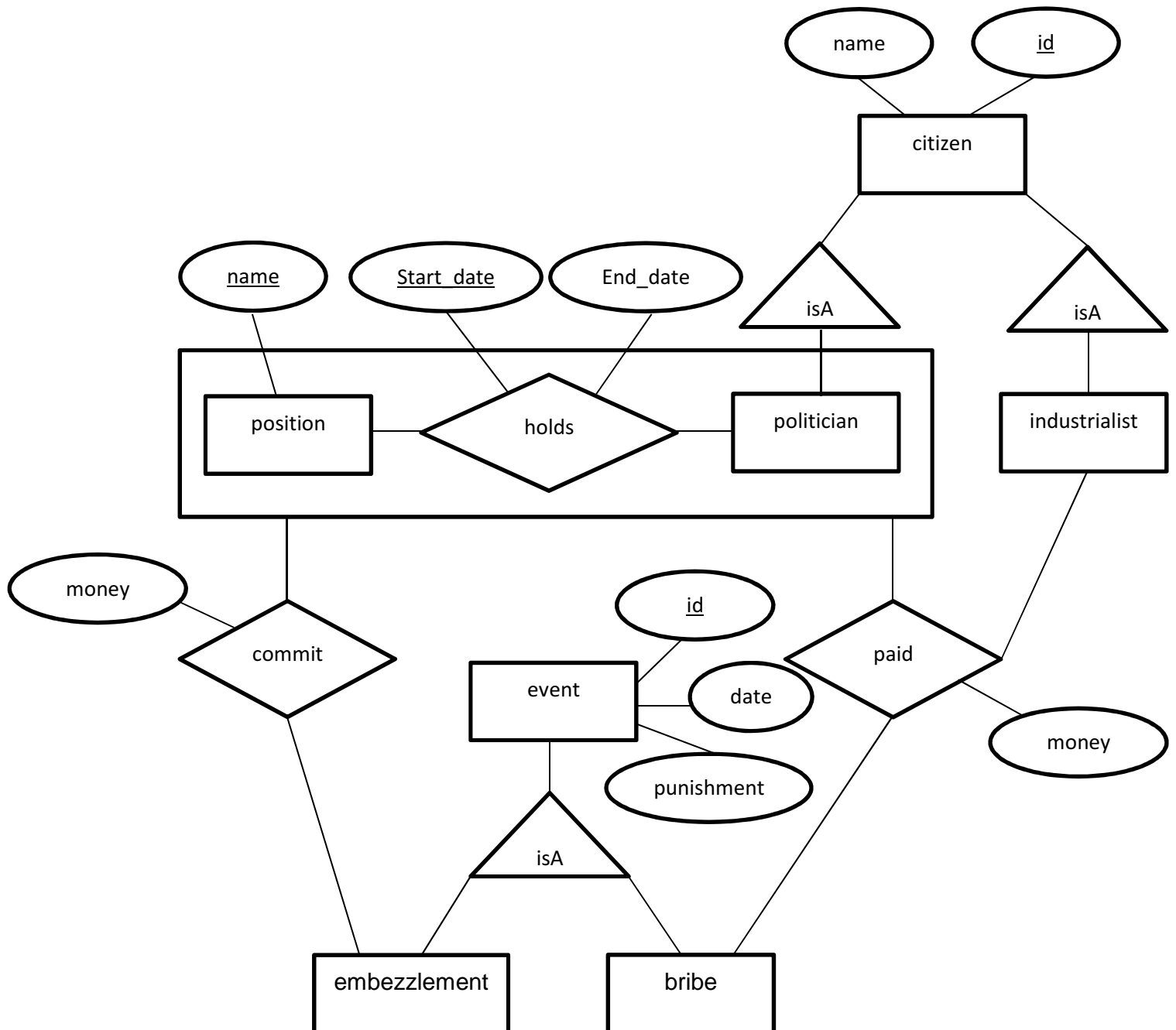
**הנחיות לנבחנים**

1. כתבו את התשובות אך ורק בטופס הבחינה, המחברת מיועדת לטיוטה 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 בלבד מותר בשימוש.
5. אין לקבל או להעביר חומר עזר כלשהו בזמן הבחינה.
6. בבחינה ארבע 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 8 דפים (כולל דף זה), **בהם 15 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 (26 נק'):

במדינת בנגיה ישנה שחיתות רבה בקרב פוליטיקאים. ה-ERD הבא מתאר פוליטיקאים (politician) אשר מחזיקים (holds) במשרה מסוימת (position) החל מתאריך start\_date ועד לתאריך end\_date. במסגרת תפקידם הם יכולים לקבל (paid) שוחד (bribe) ומבצעים (commit) עבירות מעילה (embezzlement). עבירות מעילה ושחוד הם מאורעות (event). השוחד ניתן ע"י תעשיינים (industrialist) לפוליטיקאים. תעשיינים ופוליטיקאים הם אזרחים (citizen) במדינת בנגיה. money הוא סכום כסף ששולם בעבירת המעילה או השוחד (בהתאמה).



- א. 16 נק' - ענו על השאלות הבאות על פי דיאגרמת ה-ERD. חובה לנמק (בקצרה):
- i. 4 נק' - האם יתכן שפוליטיקאי יאייש את אותה משרה במספר טווחי תאריכים? למשל יתחיל ב-2006 ויסיים ב-2007 ואז שוב יתחיל ב-2010 ויסיים ב-2015? אם לא, מדוע לא. אם כן, באיזה מקרה זה יתאפשר?

כן, כל עוד תאריכי ההתחלה שונים.

- ii. 4 נק' – האם באותו אירוע מעילה יכולים להופיע מספר פוליטיקאים?

כן, הקשר הוא רבים לרבים.

- iii. 4 נק' - האם יתכן שבמסגרת שני אירועי שוחד שונים יקבל אותו הפוליטיקאי מאותו התעשיין את אותו סכום כסף?

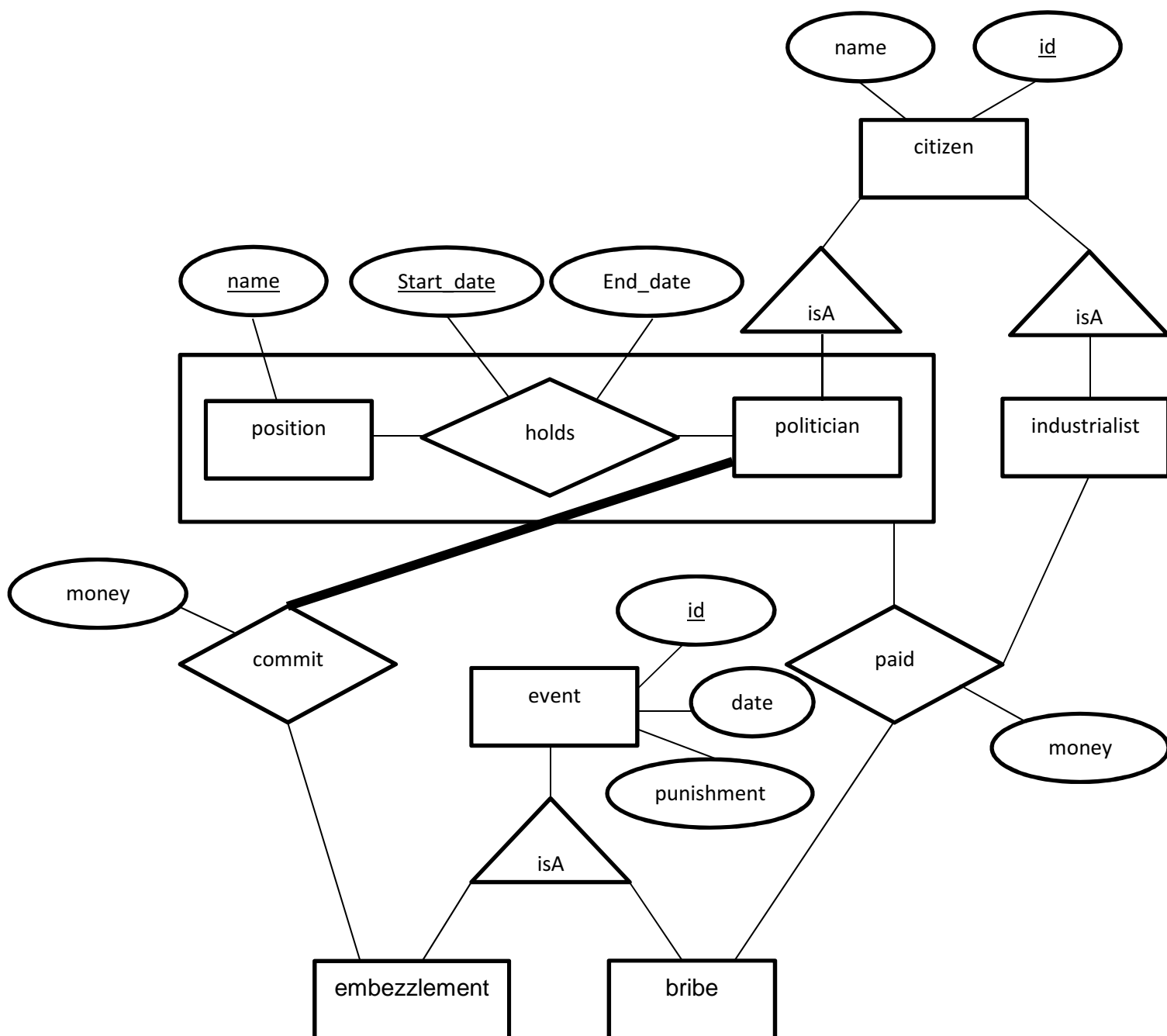
כן, היות ובטבלה Paid יש גם את `bribe.id`, `position.name`, `holds.start_date` כמפתחות אפשר להגדיר מספר אירועים של אותו תעשיין נותן אותו שוחד לאותו הפוליטיקאי.

- iv. 4 נק' - האם יתכן שאדם יקבל שוחד מעצמו?

כן, אין מניעה שאזרח יהיה גם פוליטיקאי וגם תעשיין ואז אין מניעה שהוא יופיע עם עצמו בקשר `paid`.

ב. 10 נק' - בכל סעיף, תארו את ההבדל בנתונים שמאוחסנים במסד במידה ומבצעים את השינוי המוצע בסעיף (כל סעיף באופן בלתי תלוי בסעיף האחר). לנוחיותכם בכל סעיף מצורף ה-ERD **לאחר השינוי** (השינוי מוצג בקווים עבים רק לצורך זיהוי השינוי – אין לייחס לעובי משמעות מבחינת הדיאגרמה).

i. 5 נק' - מחברים את commit ישירות לpolitician ולא להקבצה על holds.



מקום עבור כתיבת התשובה נמצא בעמוד הבא.

מעילה עתה מבוצעת ע"י פוליטיקאי בלי קשר לתפקידיו (ויכולה להתבצע אף אם אין כאלו).

The diagram is an Entity-Relationship (ER) model. It features several entities and their relationships:

- citizen** (Entity): Attributes are name and id.
- position** (Entity): Attribute is name.
- politician** (Entity): No attributes are shown.
- event** (Entity): Attributes are id, date, and punishment.
- industrialist** (Entity): No attributes are shown.
- bribe** (Entity): No attributes are shown.
- embezzlement** (Entity): No attributes are shown.

Relationships and Constraints:

- holds** (Relationship): Connects **position** and **politician**. It is a many-to-many relationship with a thick line to **politician** and a double line to **holds**.
- isA** (Relationship): Connects **citizen** to **politician** and **event** to **bribe**. These are one-to-one relationships with double lines to both entities.
- to** (Relationship): Connects **politician** and **industrialist**. It is a many-to-many relationship with thick lines to both entities.
- paid** (Relationship): Connects **industrialist** and **bribe**. It is a many-to-many relationship with a thick line to **paid** and a double line to **bribe**. An attribute **money** is associated with this relationship.
- commit** (Relationship): Connects **event** and **embezzlement**. It is a many-to-many relationship with a thick line to **commit** and a double line to **embezzlement**.

6

ענה לא שומרים יותר תשלום של כל תעשיין לכל פוליטיקאי בתפקיד, אלא רק את התשלום הכולל של התעשיין באותו אירוע שוחד.

## שאלה 2 – שאילות מידע (26 נק'):

נתונות הרלציות הבאות, המתייחסות לדיגארמת ה-ERD בשאלה 1, אך אינן בהכרח מהוות תרגום ישיר של הדיאגרמה. היעזרו בהן על מנת לפתור את הסעיפים א' ו-ב':

Citizen( <u>id</u> ,name)	Politian( <u>id</u> )
Event( <u>event_id</u> ,date,punishment)	industrialist( <u>id</u> )
Holds( <u>id</u> , <u>position_name</u> ,start_date,end_date)	bribe( <u>event_id</u> )
Commits( <u>citizen_id</u> , <u>embezzlement_id</u> , <u>position_name</u> ,money)	Embezzlement( <u>event_id</u> )
Position( <u>name</u> )	

א. 10 נק' - מכיוון שגם מנהלי מסד הנתונים הם עובדי מדינת בניה, אף אחד לא היה מופתע במיוחד כשחלק מתעודות הזהות של הפוליטיקאים (citizen\_id) ברלציה commit נעלמו והוחלפו בערכי NULL. **כתבו שאילתת SQL שיוצרת טבלה חדשה suspicious(event,id).** על הטבלה החדשה להכיל **כל זוג של מזהה אירוע של מעילה שהופיעה עם NULL בתעודת הזהות של הפוליטיקאי (שדה event\_id) יחד עם ת.ז. של הפוליטיקאי שהיה פעיל באותו תפקיד בתאריך ביצוע המעילה (שדה id).** הערה שניתנה במהלך הבחינה: יש להחליף את המשפט המחוק ב-"כל זוג של מזהה אירוע (embazzlement\_id) שהופיע ב-commits עם NULL בתעודת הזהות של פוליטיקאי"

```
CREATE Table suspicious(event,id) AS
SELECT event_id, id
FROM commit, holds, event
WHERE commit.citizen_id is NULL
      AND commit.position_name=holds.position_name
      AND holds.start_date<=event.date
      AND holds.end_date>=event.date
      AND event_id=embezzlement_id
```



ב. 8 נק' - כתבו שאילתת DRC בלתי תלויה בתחום (אין צורך להוכיח שהיא אינה תלויה בתחום)  
אשר תחזיר את מזהי ה-event-ים אשר מקיימים:

- זהו מזהה של מעילה.
- המזהה העוקב והמזהה הקודם לו, הם מזהי שוחד.

שימו לב: אין להניח שהמזהים הם בעלי מספרים עוקבים.

\* יש להניח שקיימים רק אירועי מעילה ושוחד.

$\{<id> \mid \text{embezzlement}(id)$

$\wedge \exists id1, id2 (\text{bribe}(id1) \wedge \text{bribe}(id2) \wedge id1 < id \wedge id < id2$

$\wedge \neg \exists id3 (\text{embezzlement}(id3) \wedge id3 \neq id \wedge id1 < id3 \wedge id3 < id2) \}$

ג. 8 נק' - בלי קשר לרלציות הנתונות בראש שאלה זו, כתבו תוכנית Datalog  $\neg$  לא  
רקורסיבית שמממשת אופרטור חילוק (של RA) של שתי רלציות  $R[A,B]$ ,  $S[B]$ . כך שמתקיים:  
 $Result(A) = R(A, B) \div S(B)$

$hatala(A) \leftarrow r(A, B)$

$not\_wanted(A) \leftarrow hatala(A), s(B), \neg r(A, B)$

$result(A) \leftarrow hatala(A), \neg not\_wanted(A)$

### שאלה 3 - פירוקים ותלויות פונקציונליות (20 נק'):

א. 8 נק' - עבור רלציה עם תכונות ABCD נתון שהפירוק AB, AC, AD משמר מידע ביחס לקבוצת תלויות פונקציונליות נתונה. הראו ש-A קובע (עם תלות פונקציונלית) לפחות שתיים מהתכונות האחרות. כלומר,  $A_F^+$  מכיל את A ועוד 2 תכונות לפחות.

נתבונן בצירוף הטבעי המתקבל מהפירוק, יתכנו 3 אפשרויות:

$$AB \triangleright \triangleleft AC \triangleright \triangleleft AD = ABC \triangleright \triangleleft AD$$

$$AB \triangleright \triangleleft AC \triangleright \triangleleft AD = ABD \triangleright \triangleleft AC$$

$$AB \triangleright \triangleleft AC \triangleright \triangleleft AD = ACD \triangleright \triangleleft AB$$

ללא הגבלת הכלליות נסתכל על האפשרות הראשונה:

היות ומתקיים שימור מידע מתקיים גם שימור מידע בפירוק לשתי רלציות. לכן מתקיים ש-

$A \rightarrow D$  או  $A \rightarrow BC$  (לפי המשפט לפירוק רלציה לשתיים) אם מתקיים  $A \rightarrow BC$  סיימנו. אחרת, נניח בשלילה שלא מתקיים  $A \rightarrow B$  וכן לא מתקיים  $A \rightarrow C$  אזי הפירוק של ABC ל-AB ו-AC לא משמר מידע בסתירה להנחת השלילה.

פתרון אפשרי נוסף הוא לנסות להריץ את האלגוריתם לשימור מידע בפירוק ל-2 רלציות או יותר.

נניח בשלילה ש-A לא קובע שתיים מהתכונות האחרות. אזי נסתכל על שתי אפשרויות.

א. A לא קובע אף תכונה. ניצור את הטבלה עבור האלגוריתם לבדיקת שימור המידע:

	A	B	C	D
T1	a	B	C1	D1
T2	a	B2	c	D2
t3	a	B3	C3	d

היות ולא מתקיים A גורר תכונה כלשהי, לא ניתן לבצע שינויים נוספים עבור הטבלה ומתקיים שאין שימור תלויות בסתירה.

ב. נניח ש-A רק תכונה אחת. בלי הגבלת הכלליות נניח שזו B. אזי נוכל להפעיל את a גורר b על הטבלה ולהוריד את האינדקסים מכל אותיות ה-B. עתה לא ניתן להפעיל A גורר C או D, וכן לא ניתן להפעיל B גורר C או D, היות ומכך ינבע מטרנזיטיביות ש-A גורר C או D בניגוד להנחת השלילה. אזי גם כאן לא מתקיים שימור מידע בשלילה.

במידה ו-A גורר לפחות שתי תכונות, בלי הגבלת הכלליות B ו-C מתקיים שימור מידע שכן נוכל להסתכל על הפירוק:

$$AB, ACD$$

שישמר מידע. לאחריו לפרק את ACD ל-AC ו-AD שגם הוא ישמר מידע (לפי אלגוריתם לשימור מידע בפירוק ל-2 סכמות)

ב. 12 נק' תהי סכמה  $R(A_1, A_2, \dots, A_n)$  עם  $n > 3$  וקבוצת תלויות מינימאלית  $F_c$  הכוללת תלות אחת בלבד.

1. 6 נק' - הוכיחו ש-  $R$  היא ב-BCNF אם  $F_c$  היא מהצורה  $R \setminus A_i \rightarrow A_i$  עבור  $i$  כלשהו.

$R$  היא ב-BCNF אזי  $F_c$  היא מהצורה:  $R \setminus A_i \rightarrow A_i$

על מנת ש- $F_c$  תהיה מינימאלית התלות חייבת להכיל רק תכונה אחת בצידה הימני ללא הגבלת הכלליות נניח כי זוהי התכונה  $A_i$  ל- $i$  כלשהו. כדי ש- $R$  תהיה ב-BCNF הצד השמאלי של התלות צריך להיות על מפתח. לכל  $j$  כך ש- $i$  שונה מ- $j$  מתקיים ש-  $R \setminus A_i \not\rightarrow A_j$  שכן לא קיימת תלות ב- $F$  עבורה  $A_j$  נמצאת בצידה הימני. לכן, לכל  $j$ ,  $R \setminus A_i \rightarrow A_j$  אינו על מפתח.

$F_c$  היא מהצורה:  $R \setminus A_i \rightarrow A_i$  אזי  $R$  היא ב-BCNF:

מתקיים ש- $R \setminus A_i$  הוא על מפתח שכן  $\{R \setminus A_i\}^+ = R$ , זוהי התלות הפונקציונאלית היחידה ב- $F_c$  לכן  $R$  היא ב-BCNF.

2. 6 נק' - האם קיים ל- $R$  פירוק לא טריוויאלי משמר מידע לשתי תתי סכמות? הוכיחו את תשובתכם.

**הערה שניתנה במהלך הבחינה:  $F_c$  היא מהצורה שמוכיחים בתת סעיף הקודם.**

נניח בשלילה שקיים פירוק כנ"ל לשתי סכמות  $R_1$  ו- $R_2$ . לכן מתקיים ש-

$$R_1 \cap R_2 \rightarrow R_1$$

OR

$$R_1 \cap R_2 \rightarrow R_2$$

היות ו- $F$  כוללת רק תלות אחת לא טריוויאלית בהכרח מתקיים ש-  $R_1 \cap R_2 = R \setminus A_i$

אך אם שתי תתי הסכמות מכילות את כל התכונות פרט לתכונה אחת אז מתקיים ש- $R_1 = R$  או  $R_2 = R$  ואז הפירוק לא חוקי. אחרת,  $A_i$  לא מוכלת באף אחת מתתי הסכמות וגם זה סותר את הגדרת פירוק.

## שאלה 4 – XML (28 נק'):

במדינת בנייה הפוליטיקאים אוהבים גם לנאום. עבור הנאומים הוחלט שמסמך XML יתאים יותר מאשר מסד נתונים רלציוני. נתון ה- DTD הבא, כאשר db הוא צומת המסמך.

1. <!ELEMENT db (politician)\*>
2. <!ELEMENT politician (speech)\*>
3. <!ATTLIST politician name ID #REQUIRED
4. gender (M|F) #IMPLIED>
5. <!ELEMENT speech (#PCDATA|action|emph)\*>
6. <!ATTLIST speech title ID #IMPLIED>
7. <!ELEMENT action (#PCDATA)>
8. <!ATTLIST action who IDREF #REQUIRED
9. what (shout|throw|other) #REQUIRED>
10. <!ELEMENT emph (#PCDATA)>

הסבר: במאגר (db) שמורים נתונים אודות פוליטיקאים (Politician) והנאומים אותם הם נושאים (speech). לכל פוליטיקאי נשמר גם שמו (name) ומינו (gender). לנאום נשמרת כותרת הנאום (title), ובנוסף, פעולות שהתבצעו במהלך הנאום (action), הדגשות (emph) ומלל הנאום. עבור פעולות נשמר מי ביצע אותם (who) ומה התבצע בהן (what). סוגי הפעולות האפשריים הם צעקה, זריקת חפץ או אחר.

הערה: נניח שהמצביעים "נכונים". כלומר, המצביע who אכן מצביע לפוליטיקאי.

א. 12 נק' - ענו על הסעיפים הבאים. נמקו בקצרה:

- i. 4 נק' - האם פוליטיקאי יכול לזרוק עגבניה (או אוביקט משמיץ אחר) במסגרת נאום שלו עצמו?

כן. אין מניעה שה-idref ב-action יצביע לפוליטיקאי שנושא את הנאום.

- ii. 4 נק' - האם שני פוליטיקאים יכולים לנאום נאומים בעלי כותרת זהה?

לא. הכותרת היא מסוג ID. היות ונאומים נשמרים בנפרד עבור כל פוליטיקאי לא יתכן שפוליטיקאים יחלקו נאום ולכן לא יתכן שתהיה להם אותה כותרת.

- iii. 4 נק' - האם שני פוליטיקאים יכולים לנאום אותו מלל נאום בדיוק (להעתיק אחד מהשני)?

כן. אין מניעה לשים אותו תוכן ב-PCDATA.

ב. 8 נק'- כתבו שאילתת XPath 2.0 אשר מחזירה, בלי כפילויות, את כל הפוליטיקאים שזרקו לפחות פעם אחת משהו בנאום של פוליטיקאי בן אותו מגדר. אפשר להחזיר את צמתי ה-politician המתאימים או לחילופין את שמות הפוליטיקאים המתאימים. פוליטיקאי שלא צוין עבורו מגדר נחשב בן מגדר אחר מכל שאר הפוליטיקאים במאגר.

```
distinct-values(for $t in
  //action[@what="throw" and id(@who)/@gender=../@gender]/@who
  return id($t))
```

ג. 8 נק'- כתבו שאילתת XPath 2.0 שמחזירה לכל נאום בעל כותרת את הכותרת ומספר הפוליטיקאים המשתתפים – אלו הם הנאום עצמו וכל מי שביצע פעולה במהלך הנאום. על התוצאה להיות סדרה של זוגות של כותרת (אפשר את צומת הכותרת המתאים או את תרגומו למחרוזת) ולאחריה מספר מתאים. אין לספור כפילויות (גם לא כפילות של נאום שגם מבצע אירוע בנאום של עצמו).

```
for $s in //speech[@title] return (@title,count(distinct-values(../@name, ../@who)))
```

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

שאלה: \_\_\_\_\_ סעיף: \_\_\_\_\_

שאלה: \_\_\_\_\_ סעיף: \_\_\_\_\_

שאלה: \_\_\_\_\_ סעיף: \_\_\_\_\_

