

סמסטר אביב התשע"ד

פרופ' אלדר פישר

מרצה :

גב' אלה בולשינסקי
מר רוני ליכר

מתרגלים :

מערכות מסדי נתונים

236363

מועד א' (י"ח בתמוז התשע"ד, 16 ביולי 2014)

מס' סטודנט:**פירוט השאלות והניקוד:**

מס'	נושא	ניקוד
1	ERD	27
2	שאלות מידע	33
3	פירוקים ותלויות פונקציונליות	19
4	XML	21
סה"כ		100

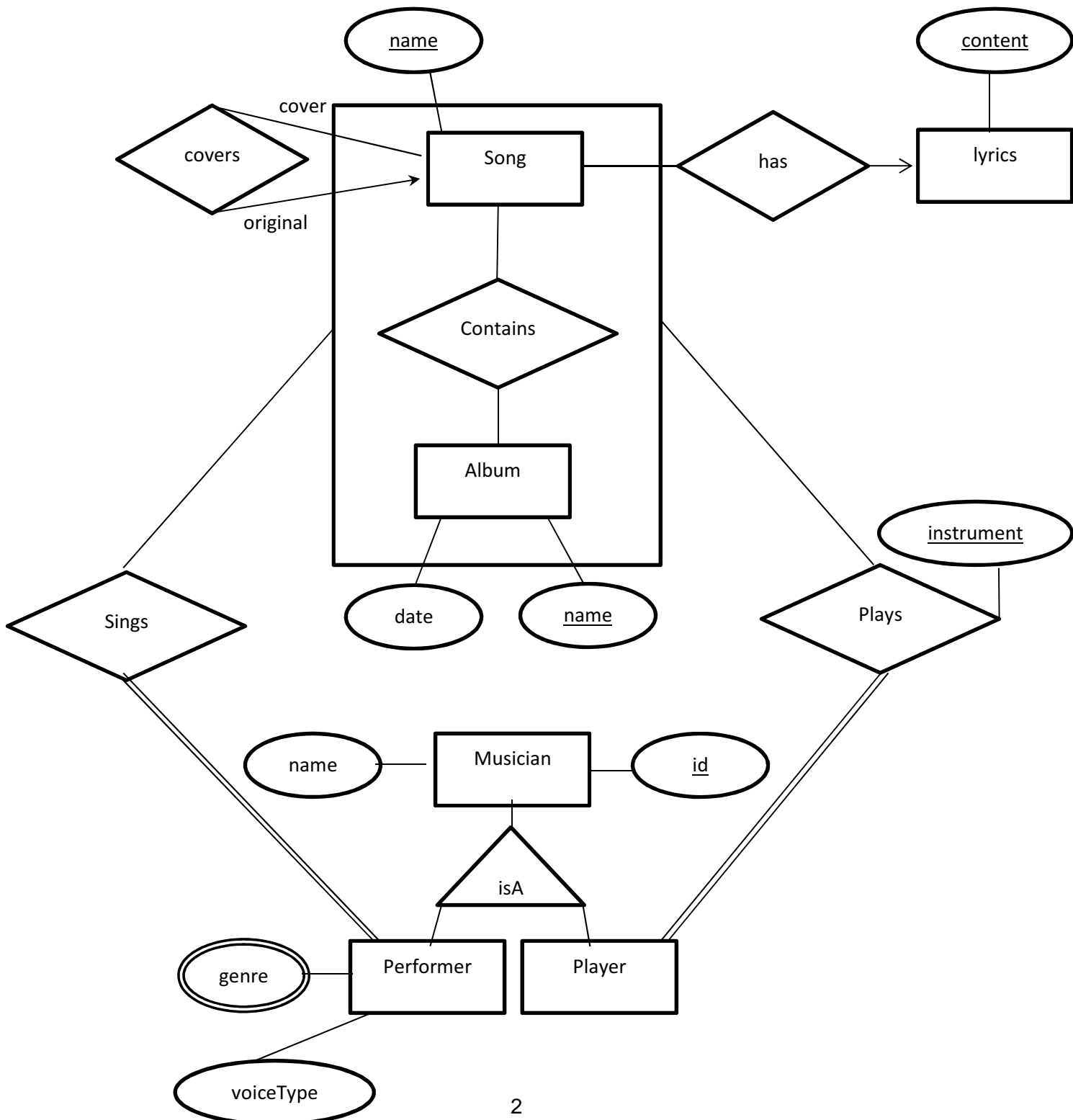
הנחיות לנבחנים

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 שאלתה.**
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 (27 נק'): :

ה- ERD הבא מייצג מסד נתונים של אלבומי מוזיקה (album), כאשר אלבום מכיל שירים (song). מסד הנתונים שומר גם נתונים עבור גרסות כיסוי לשירים, כלומר, קאברים (covers). ניתן לקשר את השירים למילות השיר שמיוצגות ע"י טיפוס היישויות lyrics. מסד הנתונים שומר גם נתונים אודות מוזיקאים- musicians (ובפרט גם על נגנים- players וזמרים- performers). כאשר שירים באלבום מבוצעים (sings) ע"י זמרים ומנגנים (plays) בהם נגנים. זמרים יכולים לשיר בסגנונות שונים (genre), רשימת הסגנונות יכולה להיות מאוד גדולה.



א. פרטו את כל הטבלאות (יתכנו יותר מ-1), השדות והמפתחות בהן, והתלויית הפונקציונליות שיווצרו בהן, עבור טיפוס הישות הבאים:
 i. 4 נק' - Plays

שם הטבלה: Plays
שדות: id,instrument,song.name,album.name
מפתח: id,instrument,song.name,album.name
תלויית פונקציונליות: רק טריוויאליות

ii. 4 נק' - Performer

יווצרו שתי טבלאות (טבלה נפרדת עבור התכונה הרב ערכית):

עבור הטבלה הראשונה:

Fields: Id, voiceType

Key: Id

FD: $Id \rightarrow voiceType$

עבור הטבלה השנייה:

Fields: Id,genre

Key:Id,genre

FD:Only trivial.

ב. בהתבסס על דיאגרמת ה-ERD הנתונה בתחילת שאלה 1 ענו על השאלות הבאות ונמקו בקצרה
i. 4 נק' - האם שיר שהוא קאבר של שיר אחר (שאינו הוא עצמו) לא-יכול/יכול/חייב להיות
עם שם אחר?

חייב. היות וכל קאבר הוא בעצם גם שיר (ולכן יחודיות במפתח תופסת ביחס לשירים כולל קאברים).

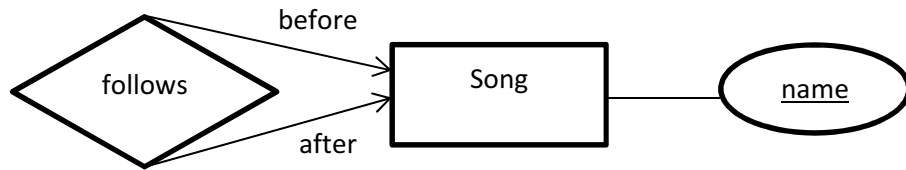
ii. 4 נק' - האם שיר שהוא קאבר של שיר אחר לא-יכול/יכול/חייב להיות עם מילים אחרות?

קאבר הוא גם שיר (כמו קודם) ולכן יכול להיות מחובר למילים שונות, אבל יכול גם זהות (אין אילוץ
שוני בחיבור ע"י טיפוס קשר)

iii. 4 נק' - האם אותו אדם לא-יכול/יכול/חייב לשיר וגם לנגן באותו שיר?

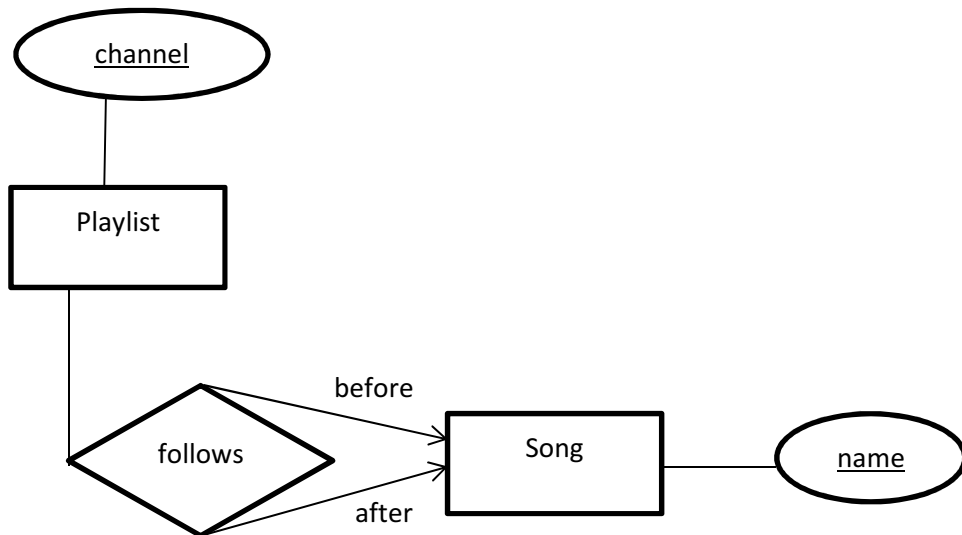
לא יכול. היות ואין אדם שהוא גם זמר וגם נגן (מהגדרת קשר ה-isA)

ג. 7 נק' - ללא קשר ל-ERD שנתון בתחילת שאלה 1, נתון ה-ERD הבא:



באמצעות ה-ERD הנתון ניתן להגדיר סדר השמעה בין שירים (סדר זה לא בהכרח מגדיר סדר טוב על הקבוצה). כאשר כל שיר מזוהה על ידי שמו.

נגדיר Playlist כקבוצת שירים. רוצים לשנות את ה-ERD כך שייצג playlist-ים ושיגדיר סדר בין השירים ב-playlist-ים (במקום סדר בין שירים כלשהם, כפי שנתון בדוגמה לעיל). כל שיר יכול להופיע בכמה playlist-ים שונים. כל playlist מזוהה ע"י שדה channel (ערוץ). הציעו שינוי ל-ERD שבאמצעותו ניתן יהיה להגדיר סדר בין שירים ב-playlist-ים שונים.



התקבלו גם פתרונות שדיברו על סדר בין שירים ברשימות שונות, ז"א פתרונות בהן נעשתה אגריגציה על קשר מתאים בין שיר לרשימת השמעה ולזו חיברו את רלציית העוקב.

שאלה 2 – שאליות מידע (33 נק'):

להלן סכמה אפשרית לחלק ממסד הנתונים הנתון ב-ERD בתחילת שאלה 1.
סכמה זו אינה בהכרח תרגום ישיר של ERD זה.

Album(date,aName)

Contains(sName,aName)

Covers(original, cover)

Song(name)

Has(content,name)

שימו לב: בשאלות SQL מותר ליצור מבטים רק אם משתמשים בהם יותר מפעם אחת.

א. 8 נק' - לכל שיר S נסמן ב- S_c את מספר הקאברים שנעשו לו.
לכל שיר S נסמן ב- S_{cc} את מספר הקאברים שנעשו לקאברים שלו.
כתבו שאלית SQL שמחזירה שירים שמתקיים עבורם $S_c < S_{cc}$, כלומר נעשו לקאברים שלו יותר קאברים מאשר שנעשו לו ישירות.

```
SELECT original
FROM covers C1
GROUP BY original
HAVING count(*) <
      (SELECT count(*)
       FROM covers C2, covers C3
       WHERE C1.original = C2.original
             And C2.cover = C3.original)
```

ב. 7 נק' - נדרג כל שיר ע"פ מספר האלבומים שמכילים קאבר לשיר.
כתבו שאילתת SQL שמחזירה את כל השירים שמדורגים מעל הממוצע ע"פ מדרג זה.

```
CREATE VIEW albums_num AS
SELECT name, count (DISTINCT aName) as C
FROM song LEFT OUTER JOIN
      (covers LEFT OUTER JOIN contains ON (cover = sName))
      ON (name = original)
GROUP BY name

SELECT name
FROM albums_num
WHERE C > (SELECT AVG(C)
           FROM albums_num)
```

ג. 7 נק' - כתבו שאילתת TRC שמחזירה את שמות כל האלבומים אשר לאף שיר באלבום לא קיים קאבר (הקאבר לא חייב להיות באותו אלבום).

$$\{t[name] \mid \exists u[date, aName] (u \in Album \wedge u[aName] = t[name] \wedge \forall r[sName, aName] ((r \in contains \wedge r[aName] = t[name]) \rightarrow \neg \exists s[original, cover] (s \in covers \wedge s[original] = r[sName])))\}$$

ד. 5 נק' - כתבו שאילתת RA שתחזיר את כל הזוגות של שם שיר ושם קאבר שהם באותו אלבום.

*הבהרה שניתנה בזמן הבחינה: לאחר "שם קאבר" יש להוסיף את המילה "שלו".

$$\pi_{original,cover}((\rho_{sName \rightarrow original} Contains \bowtie \rho_{sName \rightarrow cover} Contains) \bowtie Covers)$$

ה. 6 נק' - שאלה זו מתייחסת ל-ERD שסופק בסעיף 1.ג. להלן הפרדיקטים הנובעים מהדיאגרמה:

follows(Before,After)

song(Name)

כתבו תכנית Datalog עם שלילות אשר מוודאת שהשירים המוגדרים במסד הנתונים מהווים מעגל יחיד של שירים. צריך לבדוק גם קונסיסטנטיות עם התלויות הפונקציונאליות הנובעות מהדיאגרמה (ז"א את תנאי חד הערכיות ב-follows) וגם שאין למשל שני מעגלים או רק מסלול מכוון.

על התוכנית להגדיר את הפרדיקט invalid אשר מקיים שהמסד לא תקין אם מחזיר true ותקין אם מחזיר false.

יתקבלו גם תשובות עם פרדיקט invalid לא ריק עבור מסד נתונים לא תקין ו-invalid ריק עבור מסד נתונים תקין.

$$ancestor(X,Y) \leftarrow follows(X,Y).$$

$$ancestor(X,Y) \leftarrow ancestor(X,Z), ancestor(Z,Y).$$

$$invalid \leftarrow follows(X1,Y), follows(X2,Y), X1 \neq X2.$$

$$invalid \leftarrow follows(X,Y1), follows(X,Y2), Y1 \neq Y2.$$

$$invalid \leftarrow song(X1), song(X2), \neg ancestor(X1,X2).$$

שאלה 3 - פירוקים ותלויות פונקציונליות (19 נק'):

א. 7 נק' - נתונה הסכמה $R(A,B,C,D,E)$ וקבוצת התלויות

$$F = \{BC \rightarrow D, B \rightarrow A, A \rightarrow D, C \rightarrow E\}$$

טענה: הפירוק $\rho = \{R_1(A, B, C), R_2(A, D), R_3(D, E)\}$ לא משמר תלויות.

הראו דוגמא לרלציות r_1, r_2, r_3 שמקיימות את ההיטלים המתאימים של F אך לא מתקיים שימור תלויות עבור הצירוף הטבעי של הרלציות. נמקו מדוע הדוגמא שאתם מציגים מקיימת את הנדרש ואיזו תלות לא נשמרת בה.

התלות היחידה שלא נשמרת בפירוק היא $C \rightarrow E$ לכן, נראה דוגמא בה התלות הנ"ל לא נשמרת.

A	B	C
1	3	5
2	4	5

A	D
1	6
2	7

D	E
6	8
7	9

בדוגמא זו כל רלציה מקיימת את התלויות בהיטל של F עליה, אך ה-join הוא:

A	B	C	D	E
1	3	5	6	8
2	4	5	7	9

והתלות $C \rightarrow E$ לא נשמרת.

ב. 7 נק' - נתונה הסכמה $R(A,B,C,D,E)$ וקבוצת התלויות

$$F = \{E \rightarrow C, D \rightarrow A, AE \rightarrow D, AB \twoheadrightarrow C\}$$

שימו לב שהתלות האחרונה בקבוצת התלויות היא רב ערכית.

האם הפירוק הבא של R הוא משמר מידע בהנתן F , הוכיחו.

$$R_1(A, B, C), R_2(A, D), R_3(B, D, E)$$

מהתלות הרב ערכית נקבל שהפירוק $R_1(A,B,C), R_2'(A,B,D,E)$ הוא משמר מידע.

ענה נפרק את R_2' ל- (A,D) ו- (B,E,D)

מהמשפט לפירוק לשתי רלציות נקבל שהיות ו- $D \rightarrow A \in F$ הפירוק הנ"ל משמר מידע.

לכן, נקבל בסה"כ שהפירוק הנתון משמר מידע.

ג. 5 נק' – תנו דוגמא לקבוצת תלויות פונקציונאליות מעל רלציה $R(A,B,C)$ אשר נמצאת ב- BCNF ובכל זאת יש לה פירוק משמר מידע לא טריביאלי. נמקו.

$$F = \{A \rightarrow B, A \rightarrow C\}$$

$$\rho = \{R1(A, B), R2(A, C)\}$$

$$R1 \cap R2 = A \rightarrow AB = R1$$

A הוא על מפתח של הסכמה היות ומתקיים $A_F^+ = R$. לכן, כל תלות פונקציונאלית ב-F מקיימת שהצד השמאלי שלה הוא על מפתח. לכן הסכמה היא ב- BCNF.

הפירוק המוצע משמר מידע היות ומתקיים:

$$R1 \cap R2 = A \rightarrow AB = R1$$

שאלה 4 – XML (21 נק'):

נתון ה- DTD הבא עבור מסמכים עם צומת מסמך songbase:

1. <!ELEMENT songbase (musician+,song+)>
2. <!ELEMENT musician (name,instrument+)>
3. <!ATTLIST musician
4. id ID #REQUIRED>
5. <!ELEMENT name (#PCDATA)>
6. <!ELEMENT instrument (#PCDATA)>
7. <!ELEMENT song (name,lyrics)>
8. <!ATTLIST song
9. performers IDREFS #REQUIRED>
10. <!ELEMENT lyrics (verse,chorus)*>
11. <!ELEMENT verse (line)+>
12. <!ELEMENT line (#PCDATA)>
13. <!ELEMENT chorus (line)+>

הסבר: במאגר (songbase) יש אומנים (musician) ושירים (song). לאומן כותבים באיזה כלים (instrument) הוא יכול לנגן. לשיר יש מצביעים לאומנים המשתתפים, ופרט לשמו יש לו גם מילים (lyrics). המילים יכולות להכיל רצף של בית (verse)-פזמון (chorus) מספר כל שהוא של פעמים (לא חייבות להיות אותן מילים ב"פזמון" כל פעם), יכול להיות שלשיר שאין מילים (אז הצומת המתאים חסר בנים). לבתים ופזמונים יש חלוקה לשורות (line) שבהן הטקסט.

א. 5 נק' - רוצים שלשירים נציין באילו כלים השתמשו המבצעים שלהם. זאת באופן שלא יהיה אפשרי לרשום ביצוע בכלי שהמבצע לא יכול לנגן. איך ניתן לשנות את ה-DTD כך שיתמוך בשינויים הנדרשים? מותר להוסיף שורות חדשות. אם ברצונכם לשנות ו/או למחוק שורה קיימת, ציינו את מספר השורה.

מסירים את תכונה ה-ID של musician (מוחקים את שורות 3-4), במקומה מוסיפים תכונת ID לאלמנט instrument. לא צריך לשנות כלום בתכונות של song. שינוי זה גורם למצביעים להצביע עתה לכלי המשויך לאומן שלו (דרך אבהות).
התקבלו גם פתרונות שלא הסירו את תכונת ה-ID של musician.

בשאלות הבאות הניחו שמסמך ההקשר מציית ל-DTD המקורי מלמעלה. כל השאילות הדרושות הן ב-XPath 2.0.

ב. 8 נק' - השלימו את השאילתה הבאה שמחזירה את שמות השירים שבהם כל הפזמונים זהים לפזמון הראשון שלהם (כולל החלוקה שלהם לשורות).

אתם רשאים לכתוב את השאילתה כולה (ולא להשלים את השאילתא הנתונה). במידה ובחרתם לעשות זאת, אנא ציינו זאת.

```
//song[
    every $f in lyrics/chorus satisfies (count($f/line)=count(lyrics/chorus[1]/line))
    and ( _____
    _____
    _____
)]/name
```

every \$k in (1 to count(\$f/line)) satisfies \$f/line[\$k]=lyrics/chorus[1]/line[\$k]

ב- XPath אופרטור ה"וגם" הוא חלק מביטוי ה"לכל" לפניו ולכן אפשר להשתמש שם במשתנה \$f. עם זאת קיבלנו גם תשובות שמניחות שה"וגם" הוא מחוץ לביטוי ה"לכל" ומחבר את כל הביטוי הנ"ל עם הביטוי בסוגריים.

ג. 8 נק' - כתבו שאילתה שמחזירה לכל שורה שמופיעה במספר שירים שונים את השורה ולאחריה את שמות השירים בהם מופיעה (על השאילתה להחזיר סידרה שהיא שרשור של הנ"ל לכל שורה כזו). אסור שתהיינה כפילויות בשורות המוחזרות או בשירים המוחזרים לכל שורה. מותר להחזיר צמתים מבלי לתרגמם לערך-מחרוזת.

הערה: מותר להניח שאין שני שירים עם אותו שם

```
for $f in distinct-values(//line)
    return
        if (count (//song[./line = $f]) = 1)
        then ()
        else ($f, //song[./line = $f]/name)
```

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

שאלה: _____ סעיף: _____

שאלה: _____ סעיף: _____

שאלה: _____ סעיף: _____

