

מסדי נתונים - תרגיל 5 - Design Theory

מיכל לוי - 313573743 , רועי בן יוסף - 307920116

31 בדצמבר 2020

שאלה 1

1.
ע"פ F שנתון:

$$F = \{ \\ ID \rightarrow \text{Name, Sex} \\ \text{Year, Season} \rightarrow \text{City} \\ ID, \text{Year, Season, City} \rightarrow \text{Name, Sex, Age, Height, Weight, Team, NOC} \\ \text{Event} \rightarrow \text{sport} \\ \text{Team} \rightarrow \text{NOC} \\ \text{NOC, Year} \rightarrow \text{Team} \\ ID, \text{Year, Season, Team, NOC, Event} \rightarrow \text{sport, Medal} \\ \}$$

נשתמש באלגוריתם שלמדנו בכיתה ע"מ למצוא את כל המפתחות:
ראשית נמצא מפתח כלשהו ע"י שימוש ב- minimize :
נראה כי הסגור של $\{ID, \text{Year, Season, City, Event}\}$ כולל בתוכו את כל היחס ולכן מפתח על.
הסרה של City , אינה משנה את הסגור מכיוון ש: $\text{Year, Season} \rightarrow \text{City}$.
הסרה של כל אחד מהאלמנטים האחרים ב $\{ID, \text{Year, Season, Event}\}$ דווקא כן משנה את הסגור,
ולכן הוא **מפתח**
בנוסף ניתן להבחין כי האלמנטים הללו לא נמצאים בצד ימין של אף יחס גרירה.
כלומר:

$$\forall X \rightarrow Y \in F \\ Y \cap \{ID, \text{Year, Season, Event}\} = \emptyset$$

לכן זהו המפתח **היחיד**

2.

מכיוון שכל אטריביוט בטבלה מכיל ערך אטומי, ומכיוון שאין אטריביוטים שחוזרים על עצמם, הטבלה בצורה נורמלית ראשונה.
ניתן לשלול התאמה ל- $BCNF$ ול- $3NF$ בעזרת יחס הגרירה $NOC \rightarrow Team$. זאת מכיוון שהיחס לא טריויאלי, $Team$ הוא אינו מפתח על, ו- NOC לא חלק מאף מפתח.

.

3.

נתון הפירוק:

$$\begin{aligned} R_1 &= (ID, Year, Season, Name, Sex, Age, Height, Weight) \\ R_2 &= (ID, Year, Season, City, Team, Event, Sport, Medal) \\ R_3 &= (Team, NOC) \end{aligned}$$

נשתמש באלגוריתם שלמדנו ע"מ לבדוק אם הפירוק הוא ללא אובדן:

בתחילת האלגוריתם														
תת יחס	ID	Name	Sex	Age	Height	Weight	Team	NOC	Year	Season	City	Sport	Event	Medal
R1	a1	a2	a3	a4	a5	a6	b1,7	b1,8	a9	a10	b1,11	b1,12	b1,13	b1,14
R2	a1	b2,2	b2,3	b2,4	b2,5	b2,6	a7	b2,8	a9	a10	a11	a12	a13	a14
R3	b3,1	b3,2	b3,3	b3,4	b3,5	b3,6	a7	a8	b3,9	b3,10	b3,11	b3,12	b3,13	b3,14
ID→Name,Sex														
תת יחס	ID	Name	Sex	Age	Height	Weight	Team	NOC	Year	Season	City	Sport	Event	Medal
R1	a1	a2	a3	a4	a5	a6	b1,7	b1,8	a9	a10	b1,11	b1,12	b1,13	b1,14
R2	a1	a2	a3	b2,4	b2,5	b2,6	a7	b2,8	a9	a10	a11	a12	a13	a14
R3	b3,1	b3,2	b3,3	b3,4	b3,5	b3,6	a7	a8	b3,9	b3,10	b3,11	b3,12	b3,13	b3,14
Year,Season → City														
תת יחס	ID	Name	Sex	Age	Height	Weight	Team	NOC	Year	Season	City	Sport	Event	Medal
R1	a1	a2	a3	a4	a5	a6	b1,7	b1,8	a9	a10	a11	b1,12	b1,13	b1,14
R2	a1	a2	a3	b2,4	b2,5	b2,6	a7	b2,8	a9	a10	a11	a12	a13	a14
R3	b3,1	b3,2	b3,3	b3,4	b3,5	b3,6	a7	a8	b3,9	b3,10	b3,11	b3,12	b3,13	b3,14
Team → NOC														
תת יחס	ID	Name	Sex	Age	Height	Weight	Team	NOC	Year	Season	City	Sport	Event	Medal
R1	a1	a2	a3	a4	a5	a6	b1,7	b1,8	a9	a10	a11	b1,12	b1,13	b1,14
R2	a1	a2	a3	b2,4	b2,5	b2,6	a7	a8	a9	a10	a11	a12	a13	a14
R3	b3,1	b3,2	b3,3	b3,4	b3,5	b3,6	a7	a8	b3,9	b3,10	b3,11	b3,12	b3,13	b3,14
ID,Year,Season,City → Name,Sex,Age,Height,Weight,Team,NOC														
תת יחס	ID	Name	Sex	Age	Height	Weight	Team	NOC	Year	Season	City	Sport	Event	Medal
R1	a1	a2	a3	a4	a5	a6	b1,7	b1,8	a9	a10	a11	b1,12	b1,13	b1,14
R2	a1	a2	a3	a4	a5	a6	a7	a8	a9	a10	a11	a12	a13	a14
R3	b3,1	b3,2	b3,3	b3,4	b3,5	b3,6	a7	a8	b3,9	b3,10	b3,11	b3,12	b3,13	b3,14

נבחין כי לאחר תיקון הסתירות בגרירה:

$$ID, Year, Season, City \rightarrow Name, Sex, Age, Height, Weight, Team, NOC$$

שורה $R2$ מכילה כולה ערכי a . לכן הפירוק חסר אובדן.

4.

נמצא אם קיימת צורה נורמלית כלשהי.

ראשית נמצא כיסוי מינימלי:

נתחיל עם F :

$$F = \{ \begin{array}{l} ID \rightarrow \text{Name, Sex} \\ \text{Year, Season} \rightarrow \text{City} \\ ID, \text{Year, Season, City} \rightarrow \text{Name, Sex, Age, Height, Weight, Team, NOC} \\ \text{Event} \rightarrow \text{sport} \\ \text{Team} \rightarrow \text{NOC} \\ \text{NOC, Year} \rightarrow \text{Team} \\ ID, \text{Year, Season, Team, NOC, Event} \rightarrow \text{sport, Medal} \end{array} \}$$

נפרק את צד ימין של כל גרירה:

$$G_1 = \{ \begin{array}{l} ID \rightarrow \text{Name} \\ ID \rightarrow \text{Sex} \\ \text{Year, Season} \rightarrow \text{City} \\ ID, \text{Year, Season, City} \rightarrow \text{Name} \\ ID, \text{Year, Season, City} \rightarrow \text{Sex} \\ ID, \text{Year, Season, City} \rightarrow \text{Age} \\ ID, \text{Year, Season, City} \rightarrow \text{Height} \\ ID, \text{Year, Season, City} \rightarrow \text{Weight} \\ ID, \text{Year, Season, City} \rightarrow \text{Team} \\ ID, \text{Year, Season, City} \rightarrow \text{NOC} \\ \text{Event} \rightarrow \text{sport} \\ \text{Team} \rightarrow \text{NOC} \\ \text{NOC, Year} \rightarrow \text{Team} \\ ID, \text{Year, Season, Team, NOC, Event} \rightarrow \text{sport} \\ ID, \text{Year, Season, Team, NOC, Event} \rightarrow \text{Medal} \end{array} \}$$

נסיר גרירות מיותרות:

$$\begin{aligned}
G_2 = \{ & \\
& ID \rightarrow Name \\
& ID \rightarrow Sex \\
& Year, Season \rightarrow City \\
& ID, Year, Season, City \rightarrow Age \\
& ID, Year, Season, City \rightarrow Height \\
& ID, Year, Season, City \rightarrow Weight \\
& ID, Year, Season, City \rightarrow Team \\
& Event \rightarrow sport \\
& Team \rightarrow NOC \\
& NOC, Year \rightarrow Team \\
& ID, Year, Season, Team, NOC, Event \rightarrow Medal \\
& \}
\end{aligned}$$

נקטין את צד שמאל של הגרירה במידת האפשר:

$$\begin{aligned}
G_3 = \{ & \\
& ID \rightarrow Name \\
& ID \rightarrow Sex \\
& Year, Season \rightarrow City \\
& ID, Year, Season \rightarrow Age \\
& ID, Year, Season \rightarrow Height \\
& ID, Year, Season \rightarrow Weight \\
& ID, Year, Season \rightarrow Team \\
& Event \rightarrow sport \\
& Team \rightarrow NOC \\
& NOC, Year \rightarrow Team \\
& ID, Year, Season, Event \rightarrow Medal \\
& \}
\end{aligned}$$

וקיבלנו כיסוי מינימלי.

5.

האלגוריתם למציאת פירוק מתאים ל- $3NF$ הוא מציאת כיסוי מינימלי, הוספת כל היחסים שבכיסוי כפירוק, והוספת מפתח אם נדרש. לא נדרש להוסיף מפתח מכיוון שהמפתח היחיד של הטבלה (שאותו מצאנו בסעיף 1) כבר מופיע כחלק מהכיסוי המינימלי: $ID, Year, Season, Event$. לכן הפירוק ל- $3NF$ הוא:

$R_1 = (\text{ID}, \text{Name})$
 $R_2 = (\text{ID}, \text{Sex})$
 $R_3 = (\text{Year}, \text{Season}, \text{City})$
 $R_4 = (\text{ID}, \text{Year}, \text{Season}, \text{Age})$
 $R_5 = (\text{ID}, \text{Year}, \text{Season}, \text{Height})$
 $R_6 = (\text{ID}, \text{Year}, \text{Season}, \text{Weight})$
 $R_7 = (\text{ID}, \text{Year}, \text{Season}, \text{Team})$
 $R_8 = (\text{Event}, \text{sport})$
 $R_9 = (\text{Team}, \text{NOC})$
 $R_{10} = (\text{NOC}, \text{Year}, \text{Team})$
 $R_{11} = (\text{ID}, \text{Year}, \text{Season}, \text{Event}, \text{Medal})$

נשים לב כי $R_9 \subset R_{10}$ ולכן נוותר על R_9 ונקבל:

$R_1 = (\text{ID}, \text{Name})$
 $R_2 = (\text{ID}, \text{Sex})$
 $R_3 = (\text{Year}, \text{Season}, \text{City})$
 $R_4 = (\text{ID}, \text{Year}, \text{Season}, \text{Age})$
 $R_5 = (\text{ID}, \text{Year}, \text{Season}, \text{Height})$
 $R_6 = (\text{ID}, \text{Year}, \text{Season}, \text{Weight})$
 $R_7 = (\text{ID}, \text{Year}, \text{Season}, \text{Team})$
 $R_8 = (\text{Event}, \text{sport})$
 $R_9 = (\text{NOC}, \text{Year}, \text{Team})$
 $R_{10} = (\text{ID}, \text{Year}, \text{Season}, \text{Event}, \text{Medal})$

בפירוק $3NF$ הצורה הנורמלית של כל טבלה היא $3NF$ או $BCNF$.
 במקרה זה R_1, R_2, R_8 הן ב- $BCNF$ מהיותן יחסים עם 2 אטריוביטים.
 R_3 הוא גם כן ב- $BCNF$ מכיוון שהתלות היחידה הלא טריוויאלית ב- F_{R_3} היא $\text{Year}, \text{Season} \rightarrow \text{city}$
 ובה צד שמאל הוא מפתח על.
 כמו כן R_4, \dots, R_7 הן גם ב- $BCNF$ מכיוון שהיחסים הלא טריוויאליים היחידים ב- F_{R_4}, \dots, F_{R_7} הם:

$\text{ID}, \text{Year}, \text{Season} \rightarrow \text{Age}$
 $\text{ID}, \text{Year}, \text{Season} \rightarrow \text{Height}$
 $\text{ID}, \text{Year}, \text{Season} \rightarrow \text{Weight}$
 $\text{ID}, \text{Year}, \text{Season} \rightarrow \text{Team}$

בהתאמה, בהם צד שמאל הוא מפתח על.
 מאותה סיבה, R_{10} הוא גם ב- $BCNF$, התלות הלא טריוויאלית היחידה ב- $F_{R_{10}}$ היא

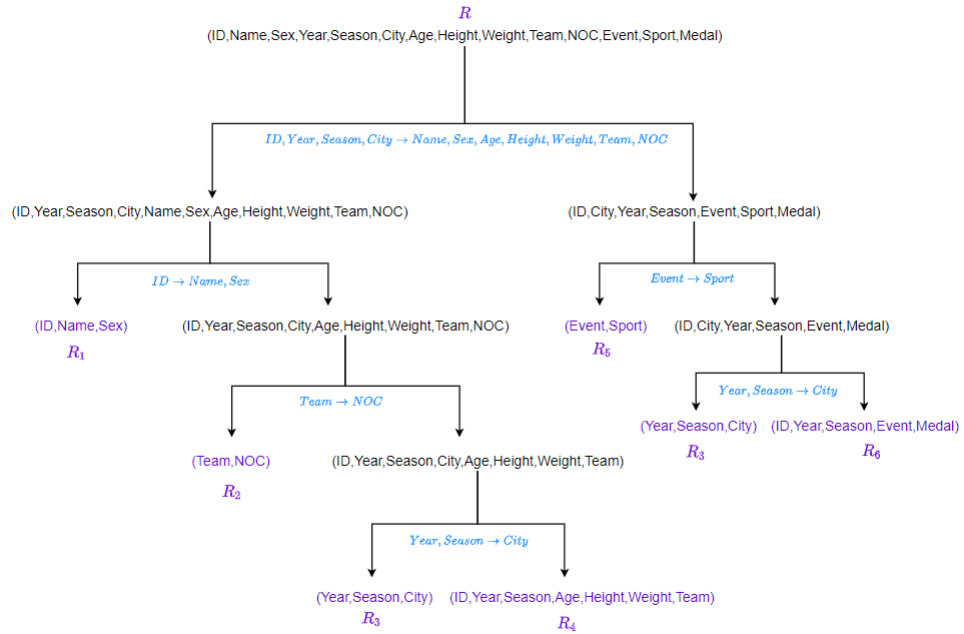
$\text{ID}, \text{Year}, \text{Season}, \text{Event} \rightarrow \text{Medal}$

בה צד שמאל הוא מפתח על.
רק R_8 היא ב- $3NF$ ולא ב- $BCNF$ מכיוון שהתלות

$$Team \rightarrow NOC$$

הנמצאת ב- F_{R_8} אינה טריוויאלית וצד שמאל בה הוא לא מפתח על.
.6

נריך את האלגוריתם למציאת פירוק $BCNF$.
מצורף פירוט של האלגוריתם בדיאגרמה, כשבכחול מסומנת הגרירה שהפרה את $BCNF$ באותו השלב ולפיה פירקנו, ובסגול מסומנים היחסים הסופיים אליהם פירקנו את הטבלה.



היחסים שהתקבלו בסוף הם:

$$\begin{aligned} R_1 &= (ID, Name, Sex) \\ R_2 &= (Team, NOC) \\ R_3 &= (Year, Season, City) \\ R_4 &= (ID, Year, Season, Height, Weight, Team) \\ R_5 &= (Event, Sport) \\ R_6 &= (ID, Year, Season, Event, Medal) \end{aligned}$$

.7

הטבלה לא משמרת תלות. לדוגמה יחס הגרירה:

$$Z = NOC, Year \rightarrow Team = Y$$

מקיים לכל R_i $Z \cap R_i = \emptyset, \{Year\}, \{NOC\}$ והסגור של אף אחת מהקבוצות האלה אינו מאפשר הרחבה של הקבוצות. כלומר:

$$\emptyset^+, \{Year\}^+, \{NOC\}^+ = \emptyset, \{Year\}, \{NOC\}$$

שאלה 2

א. מכיוון שלספרים שונים בהכרח יש שמות שונים, ניתן להסיק ששם הספר גורר את הדברים הבאים:

$$\begin{aligned} Name &\rightarrow Author \\ Name &\rightarrow User_Rating \\ Name &\rightarrow Reviews \\ Name &\rightarrow Price \\ Name &\rightarrow Genre \end{aligned}$$

ב. מכיוון שאי אפשר להסיק את $Year$ מ- $Name$, ואפשר להסיק את כל שאר השדות מ- $Name$, המפתח היחיד יהיה:

$$Name, Year$$

ג. מכיוון שאין אטריביוטים שחוזרים על עצמם, ומכיוון שכל אטריביוט מכיל ערך אטומי, הטבלה היא מצורה נומרלית ראשונה. בנוסף ניתן לראות שלא מתאים ל- $BCNF$ ול- $3NF$ מכיוון שיחס הגרירה $Name \rightarrow Author$ אינו טריוויאלי, $Name$ אינו מפתח על, ו- $Author$ אינו חלק מאף מפתח. ד. נוסף את השאילתה ואת התוצאות: שאילתה:

```

SELECT DISTINCT B1.name, B1.Author, B1.Year
FROM bestsellers B1 INNER JOIN bestsellers B2 ON (B1.Name = B2.Name)
WHERE B1.Author <> B2.Author or
B1.User_Rating <> B2.User_Rating or
B1.Reviews <> B2.Reviews or
B1.Price <> B2.Price or
B1.Genre <> B2.Genre

ORDER BY B1.name,B1.Year ASC

```

תוצאות:

name	author	year
Gone Girl	Gillian Flynn	2012
Gone Girl	Gillian Flynn	2013
Gone Girl	Gillian Flynn	2014
Quiet: The Power of Introverts in a World That Can't Stop Talking	Susan Cain	2012
Quiet: The Power of Introverts in a World That Can't Stop Talking	Susan Cain	2013
The 7 Habits of Highly Effective People: Powerful Lessons in Personal Change	Stephen R. Covey	2009
The 7 Habits of Highly Effective People: Powerful Lessons in Personal Change	Stephen R. Covey	2011
The 7 Habits of Highly Effective People: Powerful Lessons in Personal Change	Stephen R. Covey	2012
The 7 Habits of Highly Effective People: Powerful Lessons in Personal Change	Stephen R. Covey	2013
The 7 Habits of Highly Effective People: Powerful Lessons in Personal Change	Stephen R. Covey	2015
The 7 Habits of Highly Effective People: Powerful Lessons in Personal Change	Stephen R. Covey	2016
The 7 Habits of Highly Effective People: Powerful Lessons in Personal Change	Stephen R. Covey	2017
The Fault in Our Stars	John Green	2012
The Fault in Our Stars	John Green	2013
The Fault in Our Stars	John Green	2014
The Girl on the Train	Paula Hawkins	2015
The Girl on the Train	Paula Hawkins	2016
The Help	Kathryn Stockett	2009
The Help	Kathryn Stockett	2010
The Help	Kathryn Stockett	2011
The Immortal Life of Henrietta Lacks	Rebecca Skloot	2010
The Immortal Life of Henrietta Lacks	Rebecca Skloot	2011
The Immortal Life of Henrietta Lacks	Rebecca Skloot	2012
To Kill a Mockingbird	Harper Lee	2013
To Kill a Mockingbird	Harper Lee	2014
To Kill a Mockingbird	Harper Lee	2015
To Kill a Mockingbird	Harper Lee	2016
To Kill a Mockingbird	Harper Lee	2019
Unbroken: A World War II Story of Survival, Resilience, and Redemption	Laura Hillenbrand	2010
Unbroken: A World War II Story of Survival, Resilience, and Redemption	Laura Hillenbrand	2011
Unbroken: A World War II Story of Survival, Resilience, and Redemption	Laura Hillenbrand	2012
Unbroken: A World War II Story of Survival, Resilience, and Redemption	Laura Hillenbrand	2014

ה. מסקירת השאלתה השלמה (המציגה את כל העמודות) התלויות הבאות אינן נשמרות:

$Name \rightarrow User_Rating$

$Name \rightarrow Reviews$

$Name \rightarrow Price$

התלויות האחרות נשמרות.

נשתמש בפירוק ל- $3NF$, מכיוון שהוא ללא אובדן, ומשמר תלות.

כיסוי מינימלי כבר מצאנו, ולכן עבור התלויות

$Name \rightarrow Author$

$Name \rightarrow User_Rating$

$Name \rightarrow Reviews$

$Name \rightarrow Price$

$Name \rightarrow Genre$

נוסיף את היחסים:

$$\begin{aligned}
R_1 &= (Name, Author) \\
R_2 &= (Name, User_Rating) \\
R_3 &= (Name, Reviews) \\
R_4 &= (Name, Price) \\
R_5 &= (Name, Genre)
\end{aligned}$$

כאשר $Name$ יהיה המפתח בכל תת יחס.
 בנוסף, מכיוון שאנו עדיין צריכים טבלת מפתח, נוסיף את היחס:

$$R_6 = (Name, Year)$$

נשים לב שמכיוון שבכל אחד מהיחסים R_1, \dots, R_6 הוא המפתח, לא מופרת אף תלות פונקציונלית.

שאלה 3

1. ניתן את הדוגמה הבאה:

$$R = (A, B, C)$$

$$F = (A \rightarrow B)$$

$$R_1 = (A, B), R_2 = (C)$$

הפירוק משמר את התלות הפונקציונלית היחידה, אך נאבד מידע.
 ניתן לוודא ע"י הרצת האלגוריתם שלמדנו:

	A	B	C
AB	a1	a2	b1,3
C	b2,1	b2.2	a3

2.

ניתן את הדוגמה הבאה:

$$R = (A, B, C)$$

$$F = (B \rightarrow C, A \rightarrow C)$$

$$R_1 = (A, B), R_2 = (B, C)$$

הפירוק הוא ללא אובדן:

	A	B	C
AB	a1	a2	a3
BC	b2,1	a2	a3

אך התלות הפונקציונלית $A \rightarrow C$ לא נשמרת, מכיוון ש:

$$A^+ \cap R_1 = A$$

$$A^+ \cap R_2 = \emptyset$$

כלומר לעולם לא נוכל להרחיב את A ל- AC .

3.

נכון.

בשיעור הראשון שאם יחס הוא ב- $BCNF$ אז הוא בהכרח גם ב- $3NF$.

נוכיח למקרה זה גם את הצד השני:

נניח שהיחס הוא ב- $3NF$

ע"מ שיחס יהיה ב- $3NF$ חייב להתקיים לפחות 1 מהדברים הבאים לכל $X \rightarrow Y \in F$:

1. X הוא מפתח על

2. לכל $A \in Y$:

א. או ש $A \in X$ ולכן טריוויאלי

ב. או ש- A הוא חלק ממפתח.

.

1. במידה ו- X הוא מפתח על, הטבלה היא ב- $BCNF$ ואם כן, סיימנו.

2. במידה ולכל $A \in Y$ מתקיים $A \in X$ היחס הוא טריוויאלי ולכן הטבלה ב- $BCNF$ ולכן סיימנו

נניח בשלילה שהיחס לא טריוויאלי, כלומר A לא מוכל ב- X ולכן חייב להיות חלק ממפתח. מהגדרת F ידוע כי:

$$B^+ \subseteq \{B, A\}$$

$$C^+ \subseteq \{C, A\}$$

$$D^+ \subseteq \{D, A\}$$

$$E^+ \subseteq \{E, A\}$$

לכן במידה ו- $F = \emptyset$, היחס ב- $BCNF$ (באופן ריק) ובמידה ו- $F \neq \emptyset$, המפתח הוא $BCDE$, בסתירה לכך ש- A הוא חלק מהמפתח. כלומר תנאי 2.ב לא מתקיים, ולכן תנאי 1 או 2.א מתקיים. ולכן היחס הוא ב- $BCNF$