תשובות למבחן מספר 28: סמסטר 2026א׳, 2 בפברואר 2026, שאלון 422, מועד 81

תשובה לשאלה 1:

שאלה 1 סעיף א׳:

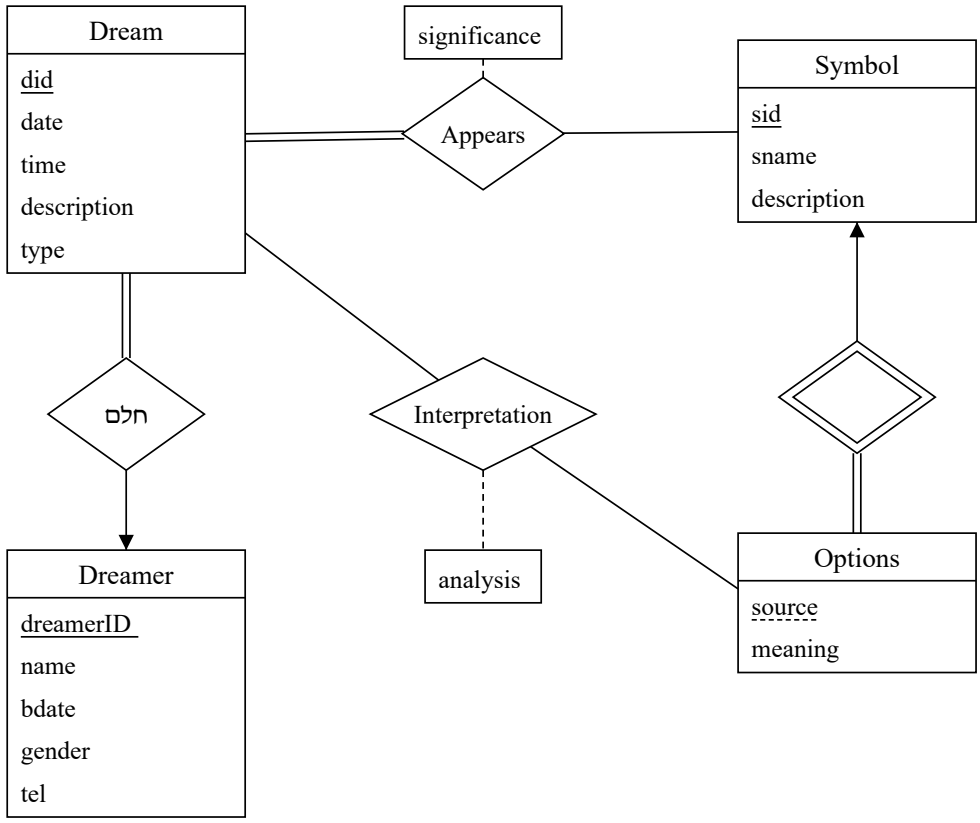
שלב ראשון – נכתוב טבלת עזר: לכל יחס בשאלון נבדוק מה הוא מכיל, וכך נדע איזה רכיב הוא בדיאגרמה.

	Interpretation	Options	Dreamer	Appears	Symbol	Dream
יחס מסוים מכיל שדה שמזהה באופן ייחודי רשומה בישות אחרת, כלומר יש בו מפתח זר המצביע לישות נפרדת.	Dream Options	Symbol		Dream Symbol		Dreamer
כאשר ליחס יש מזהה ייחודי משלו (ולא רק מפתחות מושאלים מישויות אחרות), הוא הופך להיות ישות עצמאית בפני עצמה.			X		X	X
יחס ששומר בתוכו מזהים של שתי ישויות שונות או יותר משמש כגשר המקשר בין הישויות הללו, ולכן הופך ליחס קישור.	X			X		
יחס המחזיק גם מזהה חלקי משלו וגם מזהה של ישות אחרת (שעליה הוא תלוי), מגדיר ישות חלשה שאינה עומדת בפני עצמה.		X				
מצב שבו קיימים מספר יחסים החולקים מזהה משותף אך נבדלים בתכונות נוספות, מעיד על מבנה היררכי של ירושה בין טיפוסים שונים.						

שלב שני – נבטא את האילוצים מהטקסט:

- Dream, Symbol ו-Dreamer הן ישויות. ב-Dream יש dreamerID – החולם של החלום, אחד לכל חלום. ״אין חלום ללא חולם״: קשר ״חלם״ עם קו כפול מ-Dream וחץ אל Dreamer.
- Appears מקשר בין Dream ל-Symbol (המפתח הוא שני המזהים), ו-significance היא תכונה של הקשר. ״בכל חלום יש סמלים״: קו כפול מ-Dream אל Appears.
- המפתח של Options הוא (sid, source), ו-sid הוא מזהה סמל. כלומר Options היא ישות חלשה של Symbol, עם המזהה הפנימי source (קו תחתון מרוסק), וקשר מזהה (מעוין כפול) עם חץ אל Symbol וקו כפול מ-Options (מדריך הלמידה, פרק 6, עמ׳ 155–157).
- המפתח של Interpretation הוא (did, sid, source) – המזהה של Dream והמזהה של Options. לכן Interpretation הוא קשר רבים-לרבים בין Dream ל-Options, ו-analysis היא תכונה של הקשר.

נקבל אפוא את התשובה הסופית:



בדיקה: אם נמיר את הדיאגרמה לטבלאות, נקבל בדיוק את ששת היחסים שבשאלון – Dream מקבל את dreamerID מהקשר חלם, Options מקבל את sid מהבעלים שלו (Symbol), ו-Interpretation מקבל את המפתח של Dream (did) ואת המפתח של Options (sid, source).

שאלה 1 סעיף ב' תת-סעיף 1:

נכון, הקשר Interpretation מחבר חלום ל-Options – משמעות של סמל **לפי מקור מסוים**. לאותו סמל יש כמה ישויות Options (אחת לכל מקור), ולכן באותו חלום אפשר לקשר את הסמל לכמה Options שונים – ניתוח שונה לפי כל מקור.

שאלה 1 סעיף ב' תת-סעיף 2:

לא נכון, בין Symbol לבין הקשר המזהה יש קו יחיד – השתתפות חלקית. כלומר יכול להיות סמל שאין לו אף Options (אף משמעות), ובוודאי אין דרישה ל-2 לפחות.

שאלה 1 סעיף ב' תת-סעיף 3:

לא נכון, בין Dreamer לבין הקשר חלם יש קו יחיד – השתתפות חלקית. כלומר יכול להיות חולם שאין לו אף חלום.

שאלה 1 סעיף ב' תת-סעיף 4:

נכון, מהקשר חלם יוצא חץ אל Dreamer – לכל חלום **לכל היותר** חולם אחד, ובין Dream לקשר יש קו כפול – לכל חלום **לפחות** חולם אחד. יחד: בדיוק חולם אחד.

שאלה 1 סעיף ב' תת-סעיף 5:

לא נכון, Options היא ישות חלשה של Symbol שהמזהה הפנימי שלה הוא source בלבד. כלומר כל ישות Options מזוהה על ידי הסמל והמקור, ולכן לסמל מסוים יש לכל היותר ישות Options אחת לכל מקור – ובה משמעות (meaning) אחת.

תשובה לשאלה 2:

שאלה 2 סעיף א':

שלב ראשון – החולמת: מגדר אישה, וגיל 80 לפחות במובן של השאלה: השנה הנוכחית פחות שנת הלידה גדול או שווה ל-80 (אז השנה ימלאו לה 80 או שכבר מלאו). כדי שהשאילתה תהיה נכונה לכל תאריך, משתמשים ב-CURRENT_DATE, ואת השנה של תאריך נקבל מ-date_part, כמו במדריך הלמידה:

```
date_part('year', CURRENT_DATE) - date_part('year', r.bdate) >= 80
```

שלב שני – שני חלומות של אותה חולמת: נצרף את dream פעמיים (d1, d2) – מכפלה של dreamer ושני עותקים של dream, עם התנאים `d1.dreamerID = r.dreamerID` ו-`d2.dreamerID = r.dreamerID` ב-WHERE – וניקח זוגות שההפרש בין התאריכים שלהם הוא שבוע לכל היותר, לכל כיוון: `d1.date - d2.date <= 7` וגם `d2.date - d1.date <= 7`.

שלב שלישי – כל זוג פעם אחת: בלי תנאי נוסף כל זוג יופיע פעמיים – (א, ב) וגם (ב, א) – וגם חלום עם עצמו (הפרש 0). התנאי `d1.did < d2.did` משאיר כל זוג פעם אחת, ורק חלומות שונים.

> **שימו לב:** את המילים "בהפרש של שבוע" פירשנו כ"עד שבוע". אם הכוונה להפרש של שבוע בדיוק, מחליפים את התנאי ב-`(d1.date - d2.date = 7 OR d2.date - d1.date = 7)`. ערך המגדר בטבלה לא נתון בשאלון, ולכן כתבנו 'אישה'.

נקבל אפוא את התשובה הסופית:

```
SELECT r.dreamerID, d1.did AS did1, d2.did AS did2
FROM dreamer r, dream d1, dream d2
WHERE d1.dreamerID = r.dreamerID
  AND d2.dreamerID = r.dreamerID
  AND r.gender = 'אישה'
  AND date_part('year', CURRENT_DATE) - date_part('year', r.bdate) >= 80
  AND d1.did < d2.did
  AND d1.date - d2.date <= 7
  AND d2.date - d1.date <= 7;
```

שאלה 2 סעיף ב':

שלב ראשון – סמלים עם יותר מ-5 משמעויות: כל שורה ב-options היא משמעות אחת של סמל (לפי מקור אחד). נקבץ לפי sid ונשאיר את הסמלים שיש להם יותר מ-5 שורות:

```
SELECT sid FROM options GROUP BY sid HAVING COUNT(*) > 5
```

שלב שני – חלום שאין בו אף סמל כזה: NOT EXISTS – אין ב-appears שורה של החלום עם סמל מהרשימה של השלב הראשון.

שלב שלישי – שם החולם: נצרף את dreamer – מכפלה עם התנאי r.dreamerID = d.dreamerID ב-WHERE.

נקבל אפוא את התשובה הסופית:

```
SELECT d.did, r.name
FROM dream d, dreamer r
WHERE r.dreamerID = d.dreamerID
  AND NOT EXISTS (SELECT *
                  FROM appears a
                  WHERE a.did = d.did
                    AND a.sid IN (SELECT o.sid
                                  FROM options o
                                  GROUP BY o.sid
                                  HAVING COUNT(*) > 5));
```

תשובה לשאלה 3:

״כל הניתוחים מאותו מקור״ שקול ל״**אין** שני ניתוחים של החלום ממקורות שונים״. לכן נמצא את החלומות שכן יש להם שני ניתוחים ממקורות שונים, ונוריד אותם.

שלב ראשון – הזוגות (חלום, מקור):

$$\mathrm{I} \leftarrow \pi_{\mathrm{did,\ source}}\,(\mathrm{Interpretation})$$

שלב שני – חלומות עם שני מקורות שונים: מכפלה של I עם עצמו (בשכפול לשני שמות, A ו-B), אותו חלום ומקורות שונים:

$$\mathrm{Two} \leftarrow \pi_{\mathrm{A.did}}\,(\sigma_{\mathrm{A.did}=\mathrm{B.did}\,\wedge\,\mathrm{A.source}\neq\mathrm{B.source}}\,(\rho_{\mathrm{A}}\,(\mathrm{I})\times\rho_{\mathrm{B}}\,(\mathrm{I})))$$

שלב שלישי – התוצאה: חלומות שיש להם ניתוח ($\pi_{\mathrm{did}}(I)$), פחות החלומות מהשלב השני, ולבסוף מצרפים את Dream לתיאור.

> **שימו לב:** חלום שאין לו אף ניתוח מקיים את התנאי ״באופן ריק״, אבל אין בו ניתוחים ״מאותו מקור״ – ולכן מתחילים מ-$\pi_{\mathrm{did}}(I)$ ולא מכל החלומות.

נקבל אפוא את התשובה הסופית:

$$\mathrm{I} \leftarrow \pi_{\mathrm{did,\ source}}\,(\mathrm{Interpretation})$$

$$\mathrm{Two} \leftarrow \pi_{\mathrm{A.did}}\,(\sigma_{\mathrm{A.did}=\mathrm{B.did}\,\wedge\,\mathrm{A.source}\neq\mathrm{B.source}}\,(\rho_{\mathrm{A}}\,(\mathrm{I})\times\rho_{\mathrm{B}}\,(\mathrm{I})))$$

$$\pi_{\mathrm{did,\ description}}\,(\mathrm{Dream}\bowtie(\pi_{\mathrm{did}}\,(\mathrm{I})-\mathrm{Two}))$$

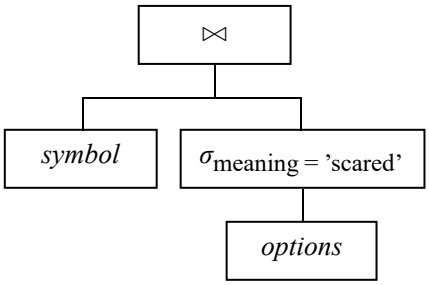
תשובה לשאלה 4:

שאלה 4 סעיף א' תת-סעיף 1:

האלגוריתם להערכת גודל התוצאה (מדריך הלמידה, פרק 11, עמ' 289–296):

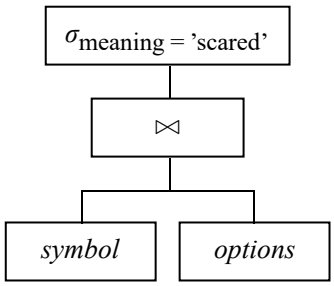
1. **מוציאים החוצה את התנאים הבדידים** – תנאי בחירה שנוגעים ליחס אחד בלבד; מפעילים אותם בסוף (כללי השקילות, עמ' 296).
2. **הופכים את המכפלות לצירופים** – תנאי שמשווה בין שני יחסים הופך את המכפלה לצירוף (עמ' 296):
   - מצרפים יחס אחד בכל פעם – כל צירוף חדש מצרף יחס **לתוצאה הקודמת**.
   - **אין** תנאי שמקשר בין שני היחסים (ואין תכונה משותפת) – זו מכפלה: $n_r \cdot n_s$ (עמ' 291, 294).
   - צירוף לפי תכונות בשמות שונים ($A = D$) או לפי זוג תכונות – אותה נוסחה, עם התכונה (או הזוג) שבכל צד.
3. **בכל צירוף – נוסחה אחת** (עמ' 294–295): מספר השורות בתוצאה = $\frac{n_r \cdot n_s}{\max(V(A,r),\ V(A,s))}$. מאיפה לוקחים את ה-$V$:
   - A **מפתח** של r – אז $V(A,r) = n_r$ (כל שורה עם ערך אחר).
   - בתוצאה של צירוף קודם – לכל עמודה נשאר ה-V שהיה לה ביחס שממנו היא באה.
   - V לא נתון, ו-A מפנה למפתח של יחס אחר – אז $V(A) \le n$ של היחס ההוא, וה-max הוא ה-n שלו.
   - (כש-A מפתח של צד אחד ו-V שלה בצד השני לא גדול יותר – יוצא בדיוק **מספר השורות בצד השני**.)
4. **מחלקים בתנאים הבדידים** (עמ' 289–290):
   - תנאי $A = c$ משאיר $\frac{1}{V(A,r)}$ מהשורות (ואם A מפתח – שורה אחת).
   - תנאי טווח – החלק היחסי מהטווח: שנה אחת מתוך 5 שנות נתונים משאירה $\frac{1}{5}$.
   - כמה תנאים עם "וגם" – מכפילים את החלקים.
5. **ההטלה** משאירה מכל שורה רק את העמודות המבוקשות, ובתוצאה אין שורות חוזרות (עמ' 30):
   - הטלה על תכונה אחת A של יחס – $V(A,r)$ שורות, אחת לכל ערך (עמ' 285).
   - אם העמודות שנשארות לא כוללות מפתח – **התשובה היא המספר שחושב**, ומוסיפים הסבר: שורות זהות מתאחדות, ולכן אחרי ביטול הכפילויות יישארו לכל היותר מספר הערכים השונים שיכולים להישאר בעמודות האלה.
6. **פעולות על קבוצות** (עמ' 291):
   - $r \cap s$ – צירוף לפי כל העמודות, באותה נוסחה של צעד 3.
   - $r - s$ – n_r פחות גודל החיתוך.
   - $r \cup s$ – לכל היותר $n_r + n_s$.

עץ הביטוי של השאילתה כפי שהיא נתונה (כמו העצים במדריך, פרק 11, עמ' 277–278): כל צומת הוא פעולה, והעלים הם היחסים:



שלב ראשון (צעד 1) – מוציאים החוצה את התנאי הבדיד: $\text{meaning} = \text{'scared'}$ – על options בלבד.

העץ אחרי צעד 1 – לפי כלל השקילות 6 (עמ' 296), בחירה שנוגעת ליחס אחד עוברת מעל הצירוף:



שלב שני (צעדים 2–3) – $\text{options} \bowtie \text{symbol}$ לפי sid: נציב בנוסחה. sid הוא המפתח של **symbol**, ולכן $V(\text{sid}, \text{symbol}) = n_{\text{symbol}} = 500$; ו-$V(\text{sid}, \text{options}) = 500$ נתון:

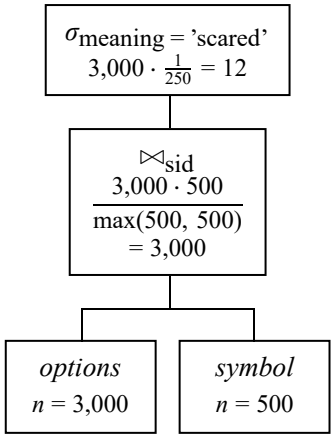
$$|\text{options} \bowtie \text{symbol}| = \frac{3{,}000 \cdot 500}{\max(500,\ 500)} = 3{,}000$$

כלומר כל שורה של options (משמעות של סמל לפי מקור) מצטרפת לסמל אחד בדיוק – הסמל שלה – ויוצא בדיוק מספר השורות של options.

שלב שלישי (צעד 4) – מחלקים בתנאי הבדיד: $V(\text{meaning}, \text{options}) = 250$, ולכן נשאר $\frac{1}{250}$ מהשורות:

$$\frac{3{,}000}{250} = 12$$

העץ אחרי צעדים 2–4 – בכל צומת מספר השורות שלו, מלמטה למעלה:



נקבל אפוא את התשובה הסופית:

הערכת מספר השורות בתוצאה: כ-12 שורות – המשמעויות 'scared' (מכל המקורות) עם פרטי הסמל שלהן.

שאלה 4 סעיף א' תת-סעיף 2:

שלב ראשון – בכמה חלומות מופיע כל סמל: כל שורה ב-appears היא הופעה של סמל אחד בחלום אחד. ב-appears יש 100,000 שורות, ומופיעים בהן $V(\text{sid}, \text{appears}) = 400$ סמלים שונים. בהתפלגות אחידה (מדריך הלמידה, פרק 11, עמ' 290):

$$\frac{n_{\text{appears}}}{V(\text{sid}, \text{appears})} = \frac{100{,}000}{400} = 250$$

נקבל אפוא את התשובה הסופית:

כל סמל (מבין הסמלים שמופיעים בחלומות) מופיע בממוצע ב-250 חלומות.

> **שימו לב:** הממוצע הוא על 400 הסמלים שמופיעים בחלומות. אם מחשבים ממוצע על כל 500 הסמלים (כולל 100 שלא מופיעים כלל), מתקבל $\frac{100{,}000}{500} = 200$.

שלב שני – לכמה סמלים כל מקור נותן משמעות: המפתח של options הוא (sid, source), ולכן כל שורה היא סמל אחר עבור אותו מקור. ב-options 3,000 שורות ו-$V(\text{source}, \text{options}) = 10$ מקורות:

$$\frac{n_{\text{options}}}{V(\text{source}, \text{options})} = \frac{3{,}000}{10} = 300$$

נקבל אפוא את התשובה הסופית:

כל מקור נותן בממוצע משמעות ל-300 סמלים.

שאלה 4 סעיף ב':

עבור $n = 5$ מתקיים: 3-5 מצביעים, 2-4 ערכים. מספר המצביעים בכל צומת נע בין $\lceil n/2 \rceil$ ל-n, ולכן מספר הערכים נע בין $\lceil n/2 \rceil - 1$ ל-$n - 1$; רק בשורש מותר שיהיו פחות (מדריך הלמידה, פרק 10, עמ' 251–252 – במדריך אותו מינימום לעלים ולצמתים פנימיים).

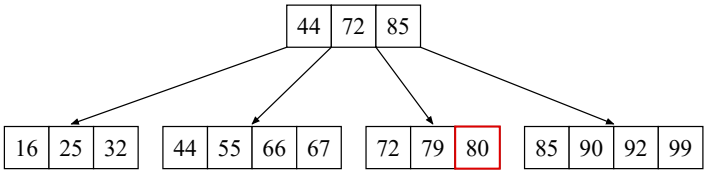
פיצול עלה (מדריך הלמידה, פרק 10, עמ' 259): בעלה מלא יש $n-1$ ערכים, ועם הערך החדש יש n ערכים שמפזרים בין שני עלים. כאשר n אי-זוגי מקבלים "חצי גדול" ו"חצי קטן": $\lceil n/2 \rceil$ ערכים עוברים לעלה אחד והשאר לשני – נבחר ש-$\lceil 5/2 \rceil = 3$ ערכים יעברו לעלה השמאלי, והשאר לימני, ונפעל כך לאורך כל השאלה. כדי לשמור על חוקי העץ, **מעתיקים** לצומת ההורה את הערך הקטן ביותר בעלה הימני: המצביע משמאלו מוביל לערכים הקטנים ממנו, והמצביע מימינו לערכים הגדולים או השווים לו.

פיצול צומת פנימי (מדריך הלמידה, פרק 10, עמ' 261): הערכים בצמתים הפנימיים משמשים לניווט בלבד, ולכן לא מעתיקים אותם: את הערך האמצעי **מעבירים** לצומת ההורה, את $\lfloor (n-1)/2 \rfloor = 2$ הערכים הקטנים ממנו לצומת השמאלי, ואת $\lceil (n-1)/2 \rceil = 2$ הגדולים ממנו לצומת הימני.

פיצול השורש (מדריך הלמידה, פרק 10, עמ' 259, 261): כשהשורש מתפצל, נוצר שורש חדש עם ערך יחיד ושני מצביעים לשני הצמתים שנוצרו.

הוספת 80:

הערך 80 גדול מ-72 אבל קטן מ-85, לכן הוא צריך להתווסף לעלה השלישי משמאל שמכיל את הערכים: 72, 79.
לאחר הוספת 80 נקבל: 72, 79, 80 – זה נותן 3 ערכים שזה תקין (לא חורג מהמקסימום של 4 ערכים).



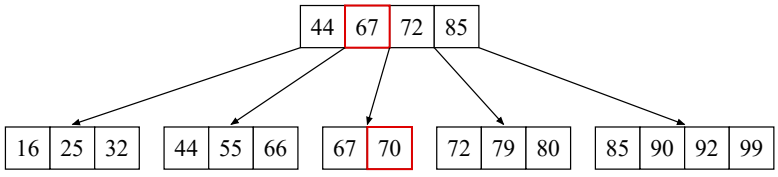
הוספת 70:

הערך 70 גדול מ-44 אבל קטן מ-72, לכן הוא צריך להתווסף לעלה השני משמאל שמכיל את הערכים:
44, 55, 66, 67. לאחר הוספת 70 נקבל: 44, 55, 66, 67, 70 – זה נותן 5 ערכים שזה גדול מהמקסימום של 4, לכן נצטרך לפצל את העלה.

פיצול העלה: 44, 55, 66, 67, 70

- עלה שמאלי (3 ערכים): 44, 55, 66
- עלה ימני (2 ערכים): 67, 70
- הערך שעולה למעלה הוא 67 – הערך הקטן ביותר של העלה הימני, שמועתק לצומת ההורה ונשאר גם בעלה (עמ' 259)

לאחר הוספת 67 לצומת הפנימי (במקרה זה השורש) נקבל: 44, 67, 72, 85 – זה נותן 4 ערכים שזה תקין.



הוספת 95:

הערך 95 גדול מ-85, לכן הוא צריך להתווסף לעלה הימני ביותר שמכיל את הערכים: 85, 90, 92, 99. לאחר הוספת 95 נקבל: 85, 90, 92, 95, 99 – זה נותן 5 ערכים שזה גדול מהמקסימום של 4, לכן נצטרך לפצל את העלה.

פיצול העלה: 85, 90, 92, 95, 99

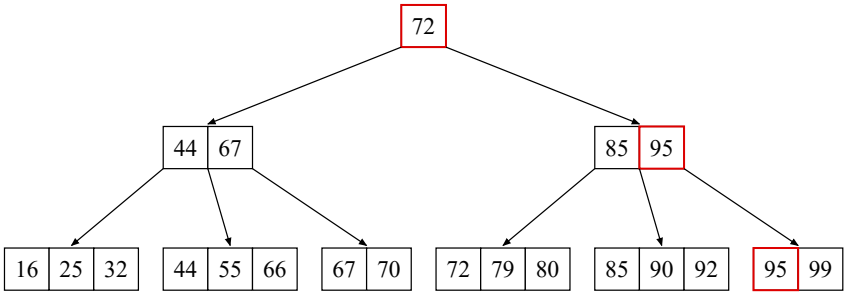
- עלה שמאלי (3 ערכים): 85, 90, 92
- עלה ימני (2 ערכים): 95, 99
- הערך שעולה למעלה הוא 95 – הערך הקטן ביותר של העלה הימני, שמועתק לצומת ההורה ונשאר גם בעלה (עמ' 259)

לאחר הוספת 95 לצומת הפנימי (במקרה זה השורש) נקבל: 44, 67, 72, 85, 95 – זה נותן 5 ערכים שזה גדול מהמקסימום של 4, לכן נצטרך לפצל גם את הצומת הפנימי.

פיצול הצומת הפנימי: 44, 67, 72, 85, 95

- צומת שמאלי (2 ערכים): 44, 67
- צומת ימני (2 ערכים): 85, 95
- הערך האמצעי 72 עובר לצומת ההורה, ולא נשאר באף אחד משני הצמתים (עמ' 261)

השורש התפצל, ולכן נוצר שורש חדש עם הערך היחיד 72 ושני מצביעים לשני הצמתים, והעץ גדל ברמה אחת (עמ' 259, 261).



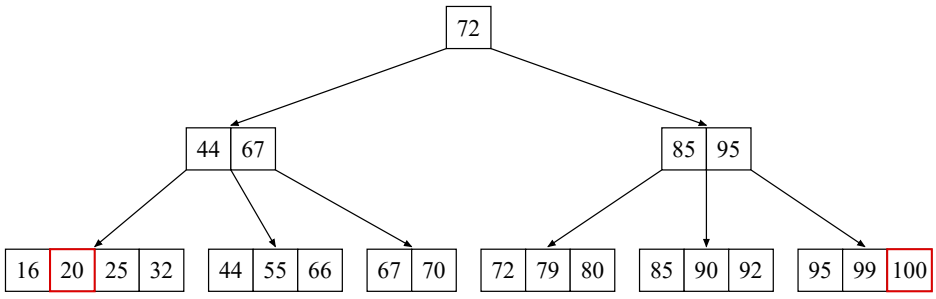
הוספת 20:

הערך 20 קטן מ-44, לכן הוא צריך להתווסף לעלה הראשון משמאל שמכיל את הערכים: 16, 25, 32. לאחר הוספת 20 נקבל: 16, 20, 25, 32 – זה נותן 4 ערכים שזה תקין (לא חורג מהמקסימום של 4 ערכים).

הוספת 100:

הערך 100 גדול מ-95, לכן הוא צריך להתווסף לעלה הימני ביותר שמכיל את הערכים: 95 ,99. לאחר הוספת 100 נקבל: 95, 99, 100 – זה נותן 3 ערכים שזה תקין (לא חורג מהמקסימום של 4 ערכים).

נקבל אפוא את התשובה הסופית:



תשובה לשאלה 5:

שלב ראשון – הנוסחה (מדריך הלמידה, פרק 8, עמ' 221): ערך הדירוג של דף j:

$$P[j] = \frac{\delta}{N} + (1-\delta) \cdot \sum_{i=1}^{N} T[i,j] \cdot P[i]$$

כאשר $N = 5$, ו-$T[i,j] = \frac{1}{i\text{-מספר הקישורים שיוצאים מ}}$ אם i מצביע ל-j. לכן $\frac{\delta}{N} = \frac{0.15}{5} = 0.03$ ו-$1 - \delta = 0.85$.

שלב שני – מספר הקישורים שיוצאים מכל דף: מ-A יוצאים 2 (ל-C, D), מ-B 1, מ-C 1, מ-D 1, ומ-E 3 (ל-A, B, D).

שלב שלישי – מי מצביע לכל דף, והצבה: בכל איטרציה מציבים את ערכי P **מהשלב הקודם** – הערכים שבשאלה.

- ל-C מצביעים A (מתוך 2) ו-D (מתוך 1):

$$0.03 + 0.85 \cdot \left(\frac{P[A]}{2} + \frac{P[D]}{1}\right) = 0.03 + 0.85 \cdot \left(\frac{0.2}{2} + 0.45\right) = 0.03 + 0.85 \cdot 0.55 = 0.4975 \approx 0.498$$

- ל-D מצביעים A (מתוך 2), B (מתוך 1) ו-E (מתוך 3):

$$P[D] = 0.03 + 0.85 \cdot \left(\frac{P[A]}{2} + \frac{P[B]}{1} + \frac{P[E]}{3}\right) = 0.03 + 0.85 \cdot (0.1 + 0.3 + 0.0667) = 0.4267 \approx 0.427$$

- ל-E מצביע רק C (מתוך 1):

$$P[E] = 0.03 + 0.85 \cdot \frac{P[C]}{1} = 0.03 + 0.85 \cdot 0.4 = 0.370$$

נקבל אפוא את התשובה הסופית:

$$P[C] = 0.498 \qquad P[D] = 0.427 \qquad P[E] = 0.370$$

נקבל אפוא את התשובה הסופית:

הדף שידורג הגבוה ביותר מבין C, D, E הוא **C** (0.498).

תשובה לשאלה 6:

שאלה 6 סעיף א':

בדיקת עודפות מצד ימין: בכל שורה מורידים תכונה אחת מצד ימין של התלות, ומחשבים את הסגור של צד שמאל בעזרת שאר התלויות. אם התכונה עדיין נמצאת בסגור – היא מיותרת, ומוחקים אותה.

תלות	תכונה	סגור	שינוי
$A \to BC$	B	$A^+ = AC$	
$A \to BC$	C	$A^+ = AB$	
$BCD \to E$	E	$BCD^+ = BCDA$	
$CD \to A$	A	$CD^+ = CD$	
$DE \to ABC$	A	$DE^+ = DEBCA$	$DE \to BC$
$DE \to BC$ עדכון בעקבות השינוי	B	$DE^+ = DECAB$	$DE \to C$
$DE \to C$	C	$DE^+ = DE$	

אוסף תלויות עדכני: $F = \{A \to BC,\ BCD \to E,\ CD \to A,\ DE \to C\}$

בדיקת עודפות מצד שמאל: בכל שורה מורידים תכונה אחת מצד שמאל של התלות, ומחשבים את הסגור של מה שנשאר. אם הסגור עדיין מכיל את צד ימין – התכונה מיותרת, ומוחקים אותה.

תלות	תכונה	סגור	שינוי
$BCD \to E$	B	$CD^+ = CDABE$	$CD \to E$
$CD \to E$ עדכון בעקבות השינוי	C	$D^+ = D$	
$CD \to E$ עדכון בעקבות השינוי	D	$C^+ = C$	
$CD \to A$	C	$D^+ = D$	
$CD \to A$	D	$C^+ = C$	
$DE \to C$	D	$E^+ = E$	
$DE \to C$	E	$D^+ = D$	

אוסף תלויות עדכני: $F = \{A \to BC,\ CD \to E,\ CD \to A,\ DE \to C\}$

בדיקת עודפות מצד ימין לאחר איחוד: תלויות עם אותו צד שמאל מאחדים לתלות אחת, ובודקים שוב את צד ימין של התלות המאוחדת.

תלות	תכונה	סגור	שינוי
$CD \to EA$	E	$CD^+ = CDAB$	
$CD \to EA$	A	$CD^+ = CDE$	

נקבל אפוא את הכיסוי הקנוני:

$$F_C = \{A \to BC,\ CD \to AE,\ DE \to C\}$$

שאלה 6 סעיף ב':

שלב ראשון – חלק מכל מפתח, נבדוק אילו תכונות לא מופיעות בשום צד ימין של תלות: אי אפשר לקבל אותן מתכונות אחרות, ולכן הן חייבות להיות בכל מפתח.

$$D$$

שלב שני – לא חלק ממפתח, נבדוק אילו תכונות מופיעות בצד ימין ולא בשום צד שמאל של תלות: הן נובעות מתכונות אחרות ולא עוזרות לקבל אף תכונה, ולכן הן לא יהיו באף מפתח.

$$B$$

שלב שלישי – נחשב את הסגורים של התכונות שכן יכולות להיות חלק ממפתח: מתחילים מתכונות שלב ראשון ומוסיפים בכל פעם עוד תכונות; קבוצה שהסגור שלה הוא כל R היא מפתח, ולא בודקים קבוצות שמכילות מפתח שכבר נמצא.

תכונה	סגור	מפתח
D	$D^+ = D$	
AD	$AD^+ = ADBCE$	כן
CD	$CD^+ = CDAEB$	כן
DE	$DE^+ = DECAB$	כן

נקבל אפוא את המפתחות הקבילים:

$$AD,\ CD,\ DE$$

שאלה 6 סעיף ג':

שלב ראשון – התכונות הראשיות: מסעיף ב' המפתחות הקבילים הם AD, CD, DE. לכן התכונות שמוכלות במפתח קביל כלשהו הן A, C, D, E, ורק B אינה מוכלת באף מפתח.

שלב שני – בדיקת התלות $A \to BC$: $A^+ = ABC \neq R$, ולכן A אינו מפתח-על – התלות מפרה את $BCNF$ (מדריך הלמידה, פרק 7, עמ' 201). כדי שתעמוד ב-$3NF$, כל תכונה ב-$\{B, C\} - \{A\}$ צריכה להיות מוכלת במפתח קביל (עמ' 206). C מוכלת במפתח CD, אבל B **אינה מוכלת באף מפתח קביל**.

נקבל אפוא:

התלות $A \to B$ מפרה גם את $BCNF$ וגם את $3NF$ (A אינו מפתח, ו-B אינה מוכלת באף מפתח קביל). לכן R **אינה** ב-$BCNF$ **ואינה** ב-$3NF$ – אף אחת מהן, ולכן היא ב-$1NF$.

לפי מדריך הלמידה, יחס שאינו ב-$3NF$ (ולכן גם לא ב-$BCNF$) נמצא ב-$1NF$ – הצורה הנורמלית הראשונה, שבה נמצא כל יחס במודל היחסים: כל הערכים אטומיים (עמ' 174). הצורה $2NF$ רק מוזכרת במדריך, ו"אין לה חשיבות מעשית" (עמ' 210).

שאלה 6 סעיף ד':

שלב ראשון – הפירוק ל-3NF מהכיסוי של סעיף א': לכל תלות ב- $F_C = \{A \to BC,\ CD \to AE,\ DE \to C\}$ יחס עם התכונות שלה (מדריך הלמידה, פרק 7, עמ' 208):

$$(A,B,C) \qquad (A,C,D,E) \qquad (C,D,E) \subseteq (A,C,D,E)$$

ב-(A,C,D,E) יש את התלות $A \to C$ (A אינו מפתח של היחס), ו-C מוכלת במפתח CD – כלומר היחס רק ב-$3NF$. השאלה מבקשת את **הצורה הגבוהה ביותר האפשרית**, ולכן נבדוק אם יש פירוק משמר מידע ומשמר תלויות שכולו ב-$BCNF$.

שלב שני – כיסוי קנוני אחר: הכיסוי הקנוני אינו יחיד, ומכיסוי שונה מתקבל פירוק שונה (מדריך הלמידה, פרק 7, עמ' 210). נחליף את $CD \to AE$ ב-$CD \to E$, ואת $DE \to C$ ב-$DE \to A$:

$$F_C' = \{A \to BC,\ CD \to E,\ DE \to A\}$$

F_C' שקול ל-F_C: מ-F_C' נקבל $CD \to E$ ואז $DE \to A$, כלומר $CD \to A$; ומ-$DE \to A$ ו-$A \to C$ נקבל $DE \to C$. בכיוון השני, מ-F_C: $CD \to E$ מתקבל מ-$CD \to AE$, ו-$DE \to C$ יחד עם $CD \to A$ נותן $DE \to A$. אין בו תכונות עודפות, ואין שני צדדי שמאל זהים – זהו כיסוי קנוני.

שלב שלישי – הפירוק לפי F_C': יחס לכל תלות: (A,B,C), (C,D,E), (A,D,E). היחס (A,D,E) מכיל את המפתח AD של R, ולכן אין צורך ביחס של מפתח, והפירוק משמר מידע. כל תלות של F_C' נמצאת בשלמותה ביחס אחד, ולכן הפירוק משמר תלויות.

יחסים	אוסף תלויות	מפתחות קבילים	צורה נורמלית
$R_1 = (A,B,C)$	$F_1 = \{A \to BC\}$	$K_1 = A$	$BCNF$
$R_2 = (C,D,E)$	$F_2 = \{CD \to E,\ DE \to C\}$	$K_2 = CD,\ DE$	$BCNF$
$R_3 = (A,D,E)$	$F_3 = \{DE \to A,\ AD \to E\}$	$K_3 = AD,\ DE$	$BCNF$

נקבל אפוא:

$$R_1 = (A,B,C), \quad R_2 = (C,D,E), \quad R_3 = (A,D,E) \Rightarrow BCNF$$

בכל אחד משלושת היחסים, כל התלויות הלא טריוויאליות הן תלויות במפתח של היחס, ולכן שלושתם ב-$BCNF$ – הצורה הגבוהה ביותר.

שימו לב: פירוק ל-$BCNF$ לפי אלגוריתם הפירוק (פיצול לפי $A \to BC$ ל-(A,B,C) ו-(A,D,E)) היה מאבד את התלות $CD \to AE$. לכן בודקים קודם את הפירוק ל-$3NF$, ומנסים כיסוי קנוני אחר כדי להגיע ל-$BCNF$ בלי לאבד תלויות.