

מערכות בסיסי נתונים

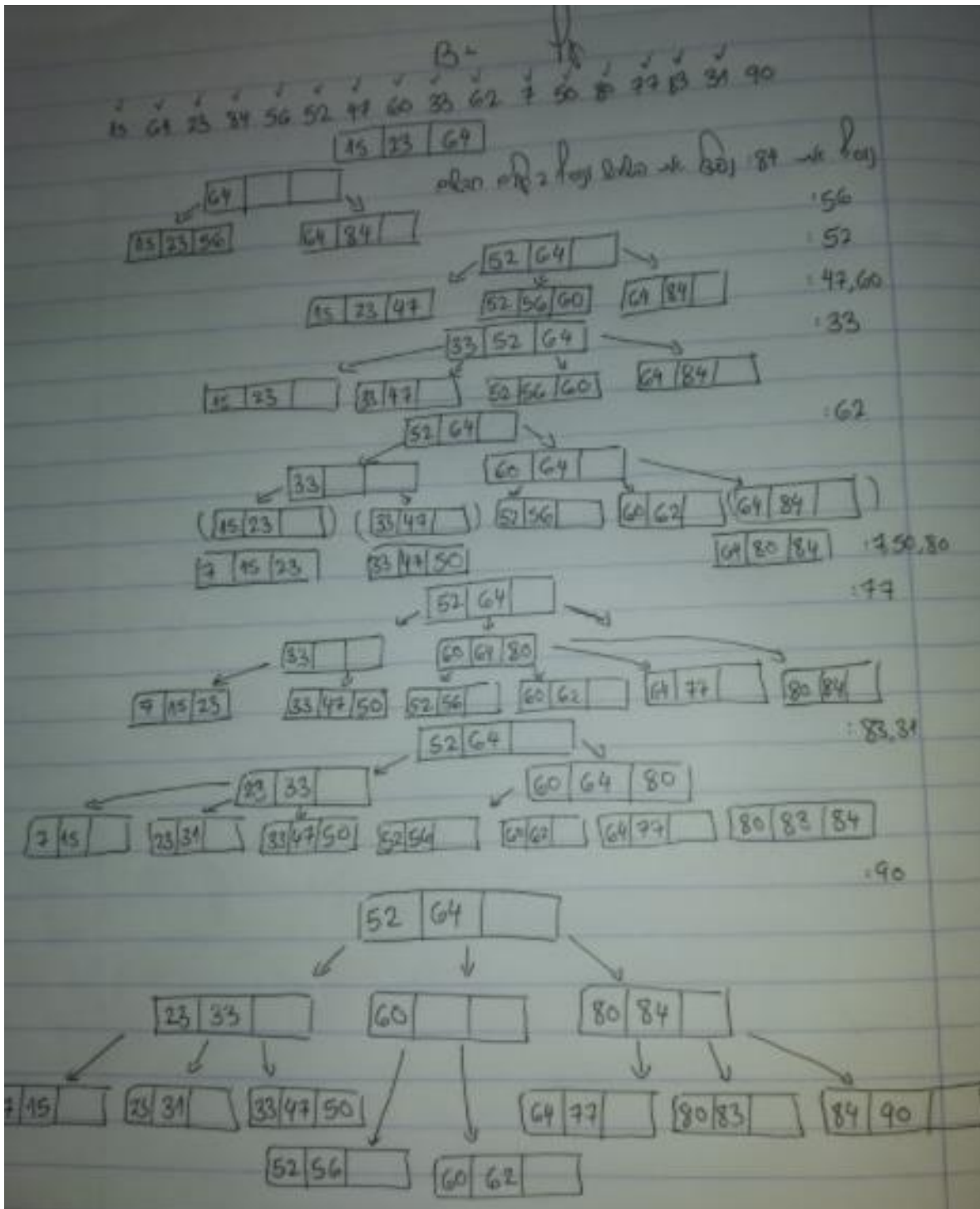
ממן 15

שאלה 1

א. עץ B+

נתון $n=4$, כלומר בכל צומת יכולים להיות בין 1 ל-3 ערכים ובין 2 ל-4 צמתים.

15, 64, 23, 84, 56, 52, 47, 60, 33, 62, 7, 50, 80, 77, 83, 31, 90



ע"מ שהעץ יגדל ברמה אחת, נתחיל מהעלים ונוסיף עליהם ערכים. כך נגרום לפיצול עד השורש.

כאשר השורש יפוצל העץ יגדל ברמה.

מספר הערכים המינימלי להוספת רמה הוא 7:

- הוספת ערך לעלה הכי מלא- 33,47,50. דבר זה יגרום לפיצול העלה ולהוספת עלה נוסף.

- הוספת 2 ערכים לעלה עם 2 ערכים תחת הצומת השמאלית. דבר זה יגרום לפיצול הצומת 23,33.

- הוספת 2 ערכים לעלה 84,90. דבר זה יגרום להוספת עלה נוסף בצד הימני.

- הוספת 2 ערכים לעלה 80,83. דבר זה יגרום ללצורך להוספת צומת נוספת אך מכיון שיש כבר 4 צמתים שהם המקסימום נצטרך לפצל את השורש.

האלגוריתם הוא הוספה בעלים הכי מלאים שיגרמו במהירות הגדולה ביותר לפיצול צומת שלהם.

ג. מקסימום ומינימום בעץ עם 4 רמות, $n=15$

מקסימום: לשורש יש לכל היותר 15 בנים, לכל אחד מהם 15 בנים, ולכל אחד מהם 15 בנים לכן לכל היותר יהיו $15^3 = 3,375$ עלים. בכל אחד מהם יהיו לכל היותר 14 ערכים, לכן נקבל $15^3 \times 14 = 47,250$ ערכים.

מינימום: לשורש יש לכל הפחות 2 בנים, לכל אחד מהם 8 בנים ולהם 8 בנים לכן נקבל $2 \times 8^2 = 128$ עלים לכל הפחות. בכל אחד מהעלים יהיו לכל הפחות 7 ערכים, לכן סה"כ נקבל $2 \times 8^2 \times 7 = 896$ ערכים לכל הפחות.

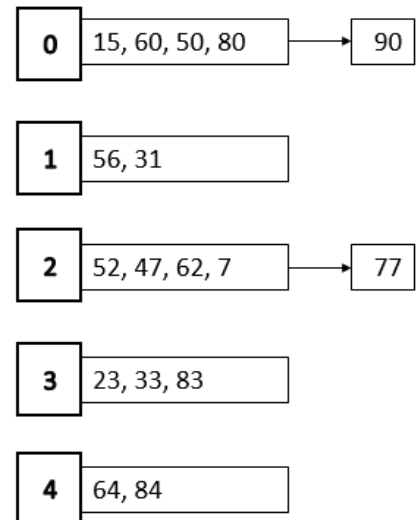
שאלה 2

ראשית נציג את טבלת ערכי הגיבוב עבור הערכים של הסעיפים הבאים:

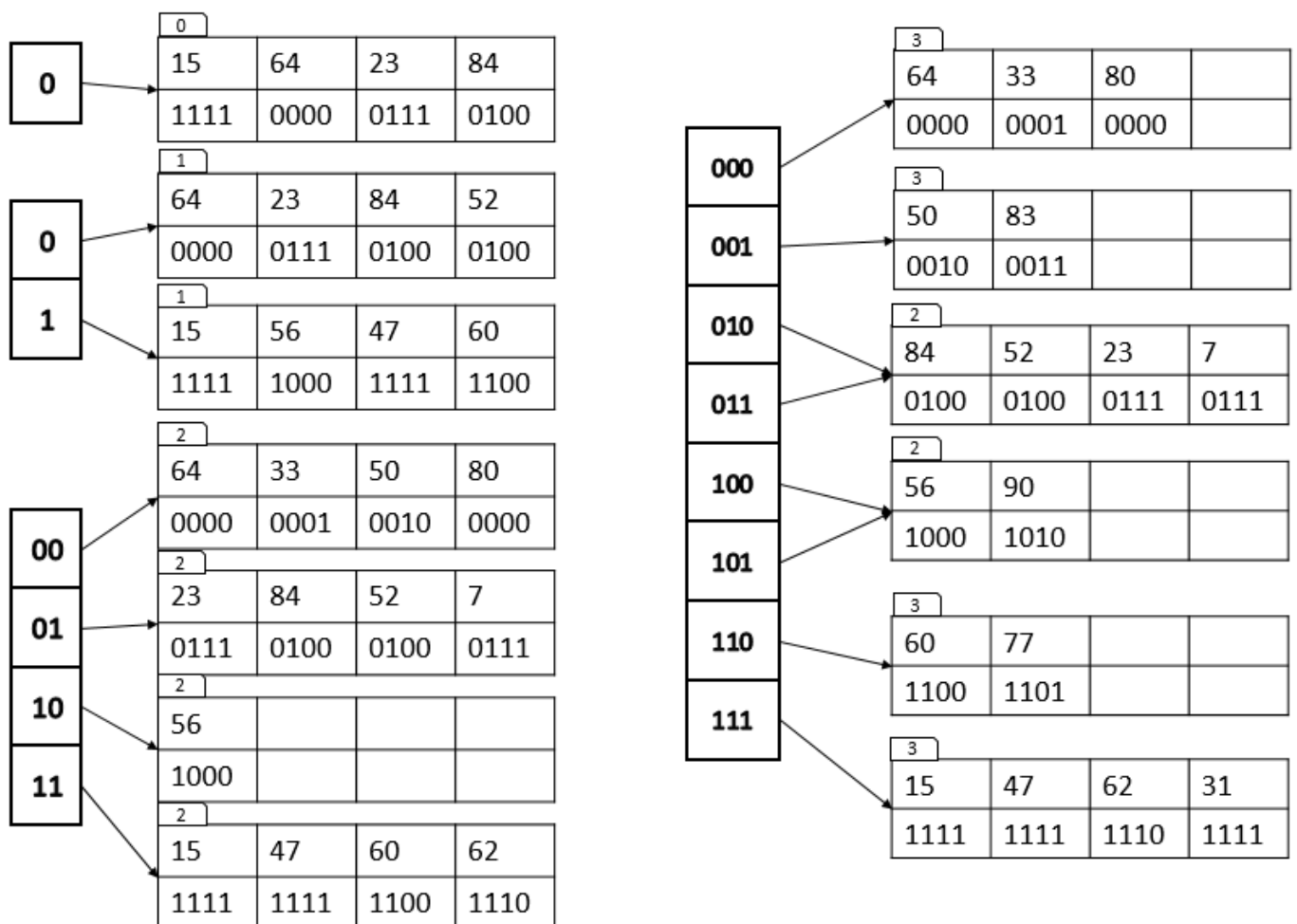
X	X mod 5	X mod 16	ייצוג בינרי- X mod 16
15	0	15	1111
64	4	0	0000
23	3	7	0111
84	4	4	0100
56	1	8	1000
52	2	4	0100
47	2	15	1111
60	0	12	1100
33	3	1	0001
62	2	14	1110
7	2	7	0111
50	0	2	0010
80	0	0	0000
77	2	13	1101
83	3	3	0011

1111	15	1	31
1010	10	0	90

א. מבנה גיבוב סטטי



ב. גיבוב מתרחב



ג. כמה שורות לכל הפחות

כדי לייצג 300 סלים, יש צורך ב-300 כניסות במדריך לפחות, אחת לכל סל. מספר הכניסות במדריך הוא תמיד חזקה של 2, ולכן מספר הכניסות המינימלי הוא 512.

שאלה 3

הערכת מספר השורות:

$$q(E,F) \text{ ו- } s(C,D,E), r(A,B,C)$$

ביחס r 1000 שורות; ביחס s 1500 שורות וביחס q 750 שורות

$$V(E,q)=750 \text{ ו- } V(E,s)=400, V(C,s)=600, V(C,r)=500, V(B,r)=800, V(A,r)=1000$$

ביחס r התכונה A היא המפתח מכיון שעבורה יש 1000 ערכים ייחודיים, כמספר השורות ביחס לעומת B ו- C .

ביחס q התכונה E היא המפתח מאותה סיבה כנ"ל.

$$A. \Pi_{AC}(r) \bowtie \Pi_{CE}(s)$$

ההטלה מ- r תחזיר לכל היותר 1000 שורות, וההטלה מ- s תחזיר לכל היותר 1500 שורות. מכיון שהצירוף לא לפי מפתח של אף אחד מהיחסים, לקבל בממוצע 2500 שורות.

$$B. \Pi_B(r) \cap \Pi_E(q)$$

ההטלה על B מ- r תחזיר 800 שורות, וההטלה על E מ- q תחזיר 750 שורות. החיתוך לא יכול להכיל יותר מ-750 שורות, לכן נקבל 750 שורות לכל היותר.

$$G. \Pi_{EC}(\sigma_{C=8}(s) \bowtie q)$$

ביחס q יש 750 שורות. ביחס s יש 1500 שורות, בו מספר הערכים הייחודיים של C הוא 600, לכן נעריך את מספר השורות שיש בהן ערך ספציפי בעמוד C כ- $1500/600=2.5=3$. הצירוף מתבצע על שדה E המשותף לשני היחסים ומפתח של q , לכן נעריך את יחס התוצאה כ-3. ההטלה של EC על יחס התוצאה יהיה 3 שורות בממוצע.

שאלה 4

$$\Pi_{manager.mid, email}(\sigma_{manager.mid=restaurant.mid \wedge manager.city="שבע"}(manager \times (restaurant \bowtie (\Pi_{lno}(restaurant) - \Pi_{lno}(inspection)))))$$

א. מה מוצאת השאילתה

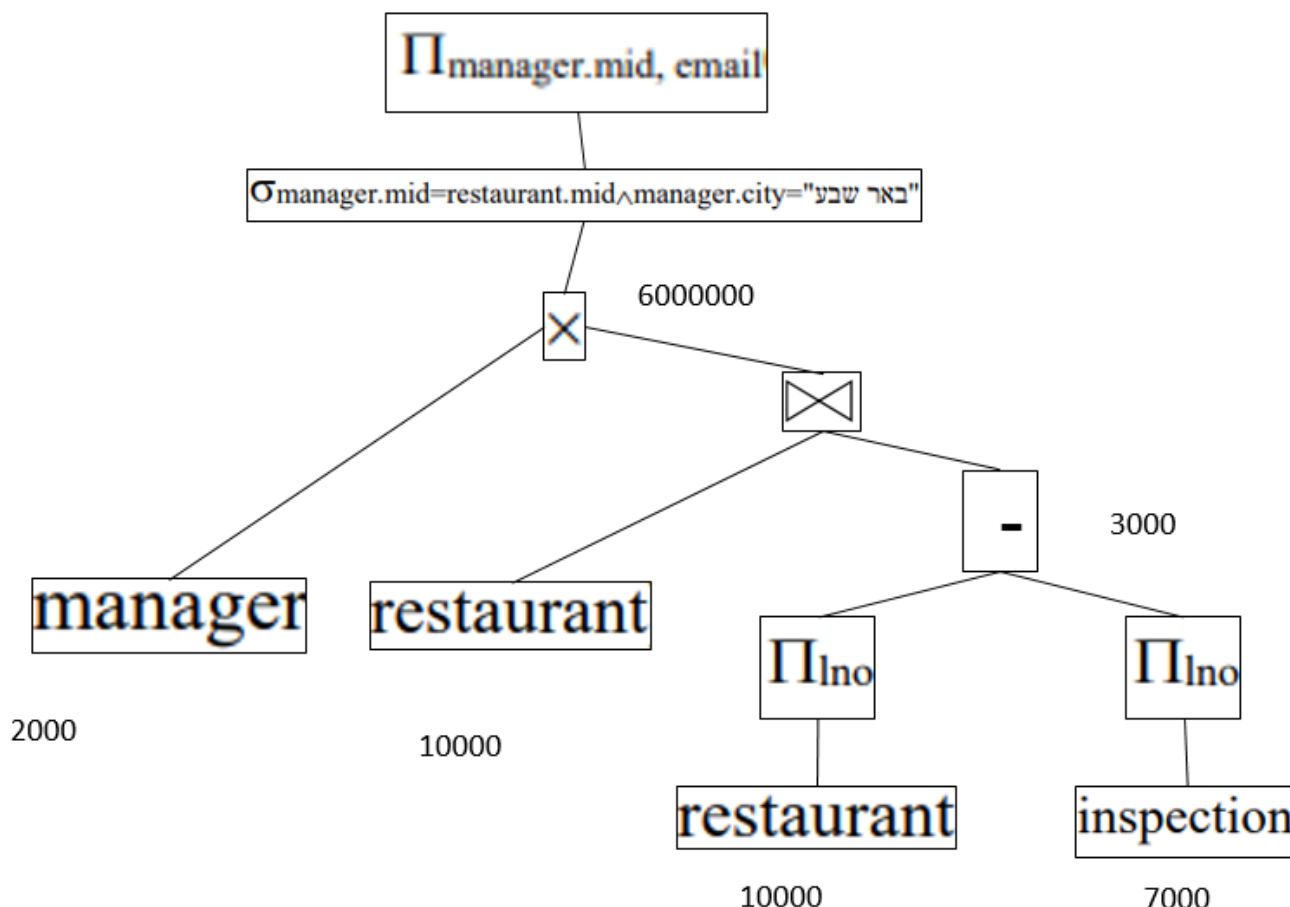
$$(\Pi_{lno}(restaurant) - \Pi_{lno}(inspection))$$

המסעדות עבורן לא בוצעה בדיקה.

$$restaurant \bowtie (\Pi_{lno}(restaurant) - \Pi_{lno}(inspection))$$

פרטים על המסעדה הנ"ל.

השאילתה תשלוף את מזהה המנהלת והמייל שלה של מנהלת שגרה בבאר שבע, והמסעדות שלה לא עברו בדיקה.



ג. כמה שורות בממוצע יהיו בביטוי

מיוצגות 10,000 מסעדות ביחס restaurant, 2,000 מנהלות ביחס manager, 300,000 בדיקות תברואתיות, ו-50 פקחים.

(כמו כן ניתן לראות ששדה שם של מסעדה הוא ייחודי לגבי כל המסעדות במערכת)

מכיון ששדות lno, idate הן מפתחות ביחס inspection, יש לכל היותר $7,000 * 2,500 = 17,500,000$ צירופים שונים. אך ביחס זה יש רק 300,000 שורות, לכן היא לא מכילה את כל הצירופים האפשריים.

- Restaurant: 10,000 (נתון)

- Manager: 2,000 (נתון)

- Lno (restaurant): 10,000 (נתון), Lno (inspection): 7,000 (נתון). לכן $10,000 - 7,000 = 3,000$.

- מכיון שהתכונה המשותפת עבור הצירוף הטבעי היא מפתח של היחס restaurant, תוצאת הצירוף מכילה לכל היותר כמספר השורות ביחס השני- יחס התוצאה שהתקבל בשלב הקודם. לכן בעת יש לנו 3,000.

- ע"מ לחשב את המכפלה הקרטזית, מכיון שיש לנו את התנאי על המזהים לכן ניקח את השורות של היחס בו הוא מפתח זר- נשארונו עם 3,000.

- בעת יש לבדוק את תנאי הבחירה. יש 200 מנהלות למסעדה, לכן החישוב יהיה:

$$\frac{3,000}{200} = 15$$

- $10 = 2000 / 200$, ממוצע מנהלות שגרות בבאר שבע.

- בהנחה שההתפלגות אחידה ואין תלות בין הנתונים, מספר השורות בממוצע יהיה 10- כלומר, 10 מנהלות שגרות בבאר שבע ומסעדתן עדיין לא עברה בדיקה.

inspection (lno, idate, iid, stime, passed, violations)

$$N_{\text{restaurant}} = V(\text{lno}, \text{restaurant}) = V(\text{rname}, \text{restaurant}) = 10,000$$

$$N_{\text{manager}} = V(\text{mid}, \text{manager}) = V(\text{mid}, \text{restaurant}) = 2,000$$

$$N_{\text{inspection}} = 300,000$$

$$V(\text{lno}, \text{inspection}) = 7,000$$

$$V(\text{iid}, \text{inspection}) = 50$$

$$V(\text{idate}, \text{inspection}) = 2,500$$

$$V(\text{city}, \text{restaurant}) = 300$$

$$V(\text{city}, \text{manager}) = 200$$

$$V(\text{email}, \text{manager}) = 1500$$