קידוד ואלגוריתמים לזכרונות

236379

פרויקט בנושא בעיית השחזור של מחרוזות

הפקולטה למדעי המחשב, טכניון

חורף, 2020-2021

מגישים:

אלין עאמר Aleen.amer@campus.technion.ac.il

זיו בן-ציון Ziv-ben-zion@campus.technion.ac.il

מבוא

רקע כללי ותיאור הבעיה

כיום, עם התפתחות עולמנו, נצברת כמות אדירה של מידע שנרצה לשמר כגון: תמונות, סרטוני וידאו, מוסיקה, מסמכים, שיחות, הודעות וכו'.., אחד הפתרונות הפוטנציאליים הוא אחסון מידע ב- DNA על ידי סנתוז וקריאתו על ידי פענוח ה-DNA, פתרון זה בלט מכיוון שהמידע בו ישמר לטווח ארוך.

יחד עם היתרונות של הפתרון המוצע לעיל, יש גם חסרונות, וזה בעת ביצוע הסנתוז או בעת הפענוח עלולים לשבש את המחרוזת המקורית וכתוצאה מכך, נוצר הרבה עותקים של אותה מחרוזת שמכילים שגיאות שאי אפשר להבחין כל כך ביניהם ונצטרך לשמר גם אותם, ופה מגיע אלגוריתם ה-clustering שבו אנו מחלקים את המחרוזות האלה לסטים שונים, שכל סט שייך למחרוזת מקורית אחת, וממנו צריך לשחזר את המחרוזת המקורית.

בעיית השחזור של המחרוזת מעותקים שלה שמכילים שגיאות, הוצגה לראשונה על ידי ולדימיר לוינשטיין והגדיר מרחק לוינשטין $d_L(x,y)$ בין שתי מחרוזות x,y כמספר מינימלי של הכנסות ומחיקות מ-x כדי להגיע ל-y. מרחק ℓ בין שתי מחרוזות, שהן מאותו אורך הוא חצי מרחק לוינשטיין, והוגדר ע"י

$$d_\ell(x,y)=\frac{1}{2}d_L(x,y)$$

כדורי ℓ ברדיוס r של מחרוזת בינארית באורך n, הוא כל המחרוזות באורך n שמתקבלות ע"י r מחיקות ו-r הכנסות, באופן פורמלי:

$$B_\ell(x;r)=\left\{y\in\ \Sigma_q^n\,\middle|\,d_\ell(x,y)\le r\right\}=\left\{y\in\ \Sigma_q^n\,\middle|\,d_L(x,y)\le 2r\right\}$$

לוינשטיין חקר ומצא שמספר מינימלי של עותקים שמכילים שגיאות של אותה מחרוזת שצריך כדי לשחזר אותה, הוא גדול יותר מהחיתוך המקסימלי בין שני כדורי לוינשטיין של כל שתי מילים בקוד.

בפרויקט הזה, נרצה לחקור ולעין בחיתוך בין כדורי ℓ של מחרוזות שונות, ברדיוס 1. בהינתן גודל המחרוזת n, נרצה לבדוק מספר מקסימלי של מחרוזות משותפות שמתקבלות ע"י הכנסה אחת ומחיקה אחת לשתי מחרוזות בינאריות שונות בארוך n.

$$max_{x,y\in\{0,1\}^n:x\neq y}|B_\ell(x;1)\cap B_\ell(y;1)|$$

הפתרון המוצע לבעיה

לפתרון הבעיה שתוארה לעיל החלטנו להשתמש קודם בסימולציה של הבעיה, ולפיה נאפין זוגות מחרוזות שנותנות חיתוך המקסימלי כדורי ℓ ברדיוס $r = 1$, ולקבל ערך מדויק עבור כל n, או לחסום מלמטה או מלמעלה הערך שנקבל.

סימולציה

בהתחלה נעשה חישוב ישיר ופשוט brute-force, כך שבהינתן n נעבור על כל זוגות של מחרוזות מאורך n, ואם זוג המחרוזות מקיים $d\ell(x,y) \leq 2$ נבדוק את חיתוכי הכדורים שלהם. הגענו ל- n=13 עם אותן תוצאות שנראה בהמשך. אך מכיוון שזמן ריצה של החישוב כזה היה ארוך מדי החלטנו לשפר זאת ע"י שימוש בטענה הבאה (ללא הוכחה):

טענה: יהיו $x, y \in \{0,1\}^n$ כך ש-$x \neq y$, אם מתקיים

$$max_{z,w\in\{0,1\}^n:z\neq w}|B\ell(z;1) \cap B\ell(w;1)| = |B\ell(x;1) \cap B\ell(y;1)|$$

אזי $d\ell(x,y) = 1$.

כלומר, ערך החיתוך המקסימלי יתקבל אך ורק על ידי זוגות שהן במרחק לוינשטיין 2.

<u>אלגוריתם הסימולציה:</u>

נעבור לכל 1≤ i ≤18:

- נגדיר max=0
- נבנה את כדורי ℓ של כל המחרוזות בארוך i, ברדיוס 1.
- עבור כל מחרוזת A באורך i:
  - עבור כל מחרוזת B בכדור של מחרוזת A:
    - בדוק את החיתוך של כדורים של המחרוזות A ו-B.
    - אם הערך גדול ממש מ-max, עדכן את max ושמור המחרוזות A ו-B.
    - אחרת, אם הערך שווה ל-max, הוסף את A ו-B כמחרוזות שנותנות את המקסימום.

הסימולציה עלתה ל-Github, וניתן למצוא אותה [פה](#).

תוצאות שקיבלנו:

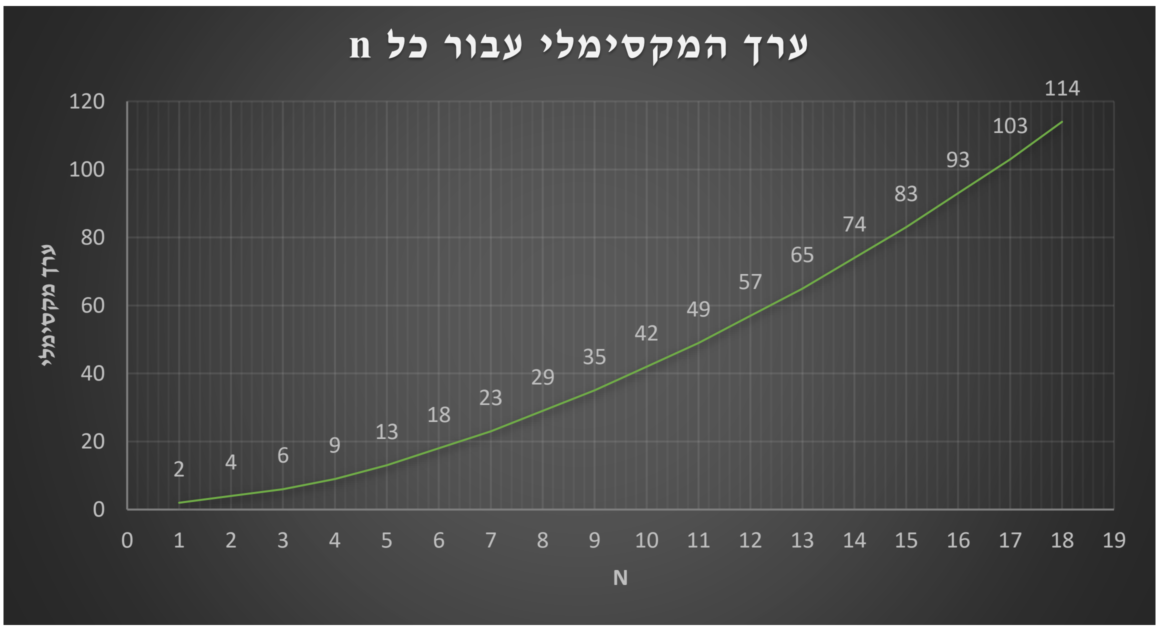
n	Largest intersection	Maximizing couples
1	2	1-0
2	4	10-01
3	6	010-001 100-010 101-010 101-011

		110-101
4	9	0010-0100 0110-0101 1001-0101 1010-0110 1010-1001 1101-1011
5	13	01001-00101 01101-01011 10100-10010 11010-10110
6	18	011010-010110 101001-100101
7	23	0101001-0100101 0110101-0101101 1010010-1001010 1011010-1010110
8	29	01010010-01001010 10110101-10101101
9	35	010100101-010010101 010110101-010101101 101010010-101001010 101101010-101011010
10	42	0101101010-0101011010 1010100101-1010010101
11	49	01010100101-01010010101 01011010101-01010110101 10101001010-10100101010 10101101010-10101011010
12	57	010101001010-010100101010 101011010101-101010110101
13	65	0101010010101-0101001010101 0101011010101-0101010110101 1010101001010-1010100101010 1010110101010-1010101101010
14	74	01010110101010-01010101101010 10101010010101-10101001010101
15	83	010101010010101-010101001010101 010101101010101-010101011010101 101010100101010-101010010101010 101010110101010-101010101101010
16	93	0101010100101010-0101010010101010 1010101101010101-1010101011010101
17	103	01010101001010101-01010100101010101 01010101101010101-01010101011010101 10101010100101010-10101010010101010 10101011010101010-10101010110101010
18	114	010101011010101010-010101010110101010 101010101001010101-101010100101010101

אפיון של זוגות המילים שממקסמות את גודל החיתוך:

יהיו $x, y \in \{0,1\}^n$ כך ש-$x \neq y$ והן נותנות את הערך החיתוך המקסימלי, אזי הן מקיימות:

- ❖ $d\ell(x,y) = 1$ – מספר המחיקות וההכנסות ל-x כדי להגיע ל-y הן 2.
- ❖ $r(y) = r(x) = n - 1$ – מספר הרצפים הזהים ב- x וב- y זהה, והוא שווה ל- n-1.
- ❖ $d_H(x,y) = 2$– מספר הביטים השונים בין y ו x − הן 2.
  והאינדקסים שבהם הביטים שונים ב-x ו-y הן סמוכים ובאמצע,
  - כך שעבור n זוגי יתקבלו ב- $\frac{n}{2}, \frac{n}{2} + 1$
  - ועבור n אי-זוגי יתקבלו ב- $\left\lfloor\frac{n}{2}\right\rfloor, \left\lfloor\frac{n}{2}\right\rfloor + 1,$ או ב- $\left\lceil\frac{n}{2}\right\rceil, \left\lceil\frac{n}{2}\right\rceil + 1$



קיבלנו, שעבור n>4 קיימת תלות בערך המקסימלי של n בערך המקסימלי של n-1.

הוכחת הערך שהתקבל:

1) סימונים והגדרות

- $F(n)$ תוגדר באמצעות נוסחת נסיגה:

$$F(n) = \begin{cases} F(n-1) + \left\lfloor \frac{n}{2} \right\rfloor + 2 & n > 4 \\ 9 & n = 4 \end{cases}$$

- $M(n) = max_{x,y \in \{0,1\}^n, x \neq y} |B_l(x;1) \cap B_l(y;1)|$
- $B_l(x) = B_l(x;1)$
- לכל $b \in \{0,1\}$:

$$\bar{b} = \begin{cases} 1, & b = 0 \\ 0, & b = 1 \end{cases}$$

- d_x: אינדקס הסרת (deletion) ביט מ-x.
- i_x: אינדקס הוספת (insertion) ביט ל-x. הביט נוסף לפני האינדקס i_x.
- m: ערך הביט שמוכנס ל-x.
- b: ערך הביט שמוכנס ל-y.
- לכל $x \in \{0,1\}^n$: $x = x_1 x_2 x_3 \dots x_n$

2) לכל $n \in \mathbb{N}$ זוגי נגדיר $x^*(n), y^*(n) \in \{0,1\}^n$ באופן הבא:

$$x^*(n)_i = \begin{cases} \overline{x^*(n)_{i-1}}, & i \neq \left\lfloor \frac{n}{2} \right\rfloor \\ x^*(n)_{i-1}, & i = \left\lfloor \frac{n}{2} \right\rfloor \end{cases}$$

$$y^*(n)_i = \begin{cases} \overline{x^*(n)_{i-1}}, & i \neq \left\lfloor \frac{n}{2} \right\rfloor \\ x^*(n)_{i-1}, & i = \left\lfloor \frac{n}{2} \right\rfloor + 2 \end{cases}$$

כאשר $x^*(n)_1 \in \{0,1\}$

לדוגמה:

$$x^*(8) = 01001010$$
$$y^*(8) = 01010010$$

לכל $n \in \mathbb{N}$ אי זוגי נגדיר $x^*(n), y^*(n) \in \{0,1\}^n$ באופן הבא:

$$x^*(n)_i = \begin{cases} \overline{x^*(n)_{i-1}}, & i \neq \left\lfloor \frac{n}{2} \right\rfloor + 1 \\ x^*(n)_{i-1}, & i = \left\lfloor \frac{n}{2} \right\rfloor + 1 \end{cases}$$

$$y^*(n)_i = \begin{cases} \overline{x^*(n)_{i-1}}, & i \neq \left\lfloor \frac{n}{2} \right\rfloor + 3 \\ x^*(n)_{i-1}, & i = \left\lfloor \frac{n}{2} \right\rfloor + 3 \end{cases}$$

כאשר $x^*(n)_1 \in \{0,1\}$

לדוגמה:

$$x^*(7) = 1011010$$
$$y^*(7) = 1010110$$

טענה 2: לכל $6 \leq n \in \mathbb{N}$ מתקיים $M(n) \geq F(n)$.

נוכיח טענה חזקה יותר: $|B_l(x^*(n)) \cap B_l(y^*(n))| \geq F(n)$.

נוכיח באינדוקציה על n:

בסיס: עבור $n = 6$ נקבל מחיפוש מחשב $|B_l(x^*(n)) \cap B_l(y^*(n))| = 18$, ואכן

$$F(n) = F(4) + 3 + 2 + 2 + 2 = 9 + 9 = 18 \geq 18$$

סגור: נניח כי הטענה נכונה עבור $n-1$ ונבחר $x = x^*(n), y = y^*(n)$

נפריד למקרים לפי זוגיות n:

<u>אם n זוגי:</u>

נסמן $x = x'x_n$, $y = y'y_n$ ויהי $w' \in B_l(x') \cap B_l(y')$.

w' נוצר מהוספה והסרה ל-x' לכן $w = w'x_n \in B_l(x)$ (ניתן לעשות את אותה הוספה והסרה).

w' נוצר גם מהוספה והסרה ל-y' ולכן $w = w'x_n = w'y_n \in B_l(y)$.

קיבלנו $w'x_n \in B_l(x) \cap B_l(y)$

מתקיים $\left\lfloor \frac{n}{2} \right\rfloor = \left\lfloor \frac{n-1}{2} \right\rfloor + 1$ ולכן $x' = x^*(n-1), y' = y^*(n-1)$.

מהנחת האינדוקציה מספר הוקטורים w הללו הוא לפחות $F(n-1)$.

נעשה $insertion$ ל-y לקבלת $v = y\overline{y_n}$.

יהי $1 \leq i \leq \left\lfloor \frac{n}{2} \right\rfloor$. נסמן ב-$v^i$ את הוקטור v לאחר הסרת הביט ה-i. לכל $1 \leq i \leq \left\lfloor \frac{n}{2} \right\rfloor$ $v_i \neq v_{i+1}$ ולכן $v^i \neq v^j$ לכל $1 \leq i < j \leq \left\lfloor \frac{n}{2} \right\rfloor + 1$.

קיבלנו $v^i = y_1 y_2 \dots y_{i-1} y_{i+1} \dots y_n \overline{y_n}$.

נשים לב שהמילה $v^i \neq w'$ לכל v^i (מכיוון שהביט האחרון שונה).

נסיר מ-x את הביט האחרון ונקבל את x'. לאחר מכן נוסיף את הביט $\overline{y_i}$ במיקום ה-i ונקבל:

$$x_1 x_2 \dots x_{i-1} \overline{y_i} x_i x_{i+1} \dots x_{n-1} =$$
$$y_1 y_2 \dots y_{i-1} y_{i+1} \dots\dots y_n \overline{y_n} = v^i$$

לכן לכל $1 \leq i \leq \left\lfloor \frac{n}{2} \right\rfloor + 1$: $v^i \in B_l(x) \cap B_l(y)$

כעת נוסיף ל-x' את $x_{\left\lfloor \frac{n}{2} \right\rfloor}$ במיקום ה-$\left\lfloor \frac{n}{2} \right\rfloor + 2$ ונקבל

$$u = x_1 \overline{x_1} x_1 \overline{x_1} \dots \overline{x_{\left\lfloor \frac{n}{2} \right\rfloor}} x_{\left\lfloor \frac{n}{2} \right\rfloor} x_{\left\lfloor \frac{n}{2} \right\rfloor} \overline{x_{\left\lfloor \frac{n}{2} \right\rfloor}} x_{\left\lfloor \frac{n}{2} \right\rfloor} x_{\left\lfloor \frac{n}{2} \right\rfloor} \overline{x_{\left\lfloor \frac{n}{2} \right\rfloor}} x_{\left\lfloor \frac{n}{2} \right\rfloor} \overline{x_{\left\lfloor \frac{n}{2} \right\rfloor}} x_{\left\lfloor \frac{n}{2} \right\rfloor} \overline{x_{\left\lfloor \frac{n}{2} \right\rfloor}} \dots$$

נוסיף ל-y' את $x_{\left\lfloor \frac{n}{2} \right\rfloor}$ במיקום ה-$\left\lfloor \frac{n}{2} \right\rfloor$ ונקבל גם כן את u, כאשר $u \neq v^i$ לכל $1 \leq i \leq \left\lfloor \frac{n}{2} \right\rfloor + 1$.

סה"כ ספרנו כבר $\left\lfloor\frac{n}{2}\right\rfloor + 2$ אפשרויות זרות נוספות ולכן

$$|B_l(x^*) \cap B_l(y^*)| \geq F(n-1) + \left\lfloor\frac{n}{2}\right\rfloor + 2$$

<u>אם n אי זוגי:</u>

נסמן $x = x_1x'$, $y = y_1y'$ ויהי $w' \in B_l(x') \cap B_l(y')$.

w' נוצר מהוספה והסרה ל-x' לכן $w = x_1w' \in B_l(x)$ (ניתן לעשות את אותה הוספה והסרה).

w' נוצר גם מהוספה והסרה ל-y' ולכן $w = x_1w' = y_1w' \in B_l(y)$.

קיבלנו $x_1w' \in B_l(x) \cap B_l(y)$

מתקיים $\left\lfloor\frac{n}{2}\right\rfloor - 1 = 1 + \left\lfloor\frac{n}{2}\right\rfloor = \left\lfloor\frac{n-1}{2}\right\rfloor$ ולכן $x' = x^*(n-1), y' = y^*(n-1)$.

מהנחת האינדוקציה מספר הוקטורים w הללו הוא לפחות $F(n-1)$.

נעשה $insertion$ ל-x לקבלת $v = \overline{x_1}x$.

יהי $\left\lfloor\frac{n}{2}\right\rfloor + 2 \leq i \leq n+1$. נסמן ב-$v^i$ את הוקטור v לאחר הסרת הביט ה-i. לכל $\left\lfloor\frac{n}{2}\right\rfloor + 2 \leq i \leq n+1$, $v_i \neq v_{i+1}$ ולכן $v^i \neq v^j$ לכל $\left\lfloor\frac{n}{2}\right\rfloor + 2 \leq i < j \leq n+1$.

קיבלנו $v^i = \overline{x_1}x_1x_2x_3 \dots x_{i-1}x_{i+1} \dots x_{n-1}x_n$.

נשים לב שהמילה $v^i \neq w'$ לכל v^i (מכיוון שהביט הראשון שונה).

נסיר מ-y את הביט הראשון ונקבל את y'. לאחר מכן נוסיף את הביט $\overline{x_i}$ במיקום ה-i ונקבל:

$$\begin{aligned} &y_2y_3 \dots y_{i-1}\overline{x_i}y_iy_{i+1} \dots y_{n-1}y_n = \\ &x_2x_3 \dots x_{i-1}\overline{x_i}x_ix_{i+1} \dots x_{n-1}x_n = \\ &\overline{x_1}x_1x_2x_3 \dots x_{i-1}x_{i+1} \dots x_{n-1}x_n = v^i \end{aligned}$$

לכן לכל $\left\lfloor\frac{n}{2}\right\rfloor + 2 \leq i \leq n+1$: $v^i \in B_l(x) \cap B_l(y)$

יש סה"כ $\left\lfloor\frac{n}{2}\right\rfloor + 1 = \frac{n+1}{2} = n - \frac{n-1}{2} = n - \left\lfloor\frac{n}{2}\right\rfloor = n + 1 - \left(\left\lfloor\frac{n}{2}\right\rfloor + 2\right) + 1$

כעת נוסיף ל-x' את $x_{\left\lfloor\frac{n}{2}\right\rfloor}$ במיקום ה-$\left\lfloor\frac{n}{2}\right\rfloor + 3$ ונקבל

$$u = x_1\overline{x_1}x_1\overline{x_1} \dots \overline{x_{\left\lfloor\frac{n}{2}\right\rfloor}}x_{\left\lfloor\frac{n}{2}\right\rfloor}x_{\left\lfloor\frac{n}{2}\right\rfloor}\overline{x_{\left\lfloor\frac{n}{2}\right\rfloor}}x_{\left\lfloor\frac{n}{2}\right\rfloor}x_{\left\lfloor\frac{n}{2}\right\rfloor}\overline{x_{\left\lfloor\frac{n}{2}\right\rfloor}}x_{\left\lfloor\frac{n}{2}\right\rfloor}\overline{x_{\left\lfloor\frac{n}{2}\right\rfloor}}x_{\left\lfloor\frac{n}{2}\right\rfloor}\overline{x_{\left\lfloor\frac{n}{2}\right\rfloor}} \dots$$

נוסיף ל-y' את $x_{\left\lfloor\frac{n}{2}\right\rfloor}$ במיקום ה-$\left\lfloor\frac{n}{2}\right\rfloor + 1$ ונקבל גם כן את u, כאשר $u \neq v^i$ לכל $\left\lfloor\frac{n}{2}\right\rfloor + 2 \leq i \leq n+11$.

סה"כ ספרנו כבר $\left\lfloor\frac{n}{2}\right\rfloor + 2$ אפשרויות זרות נוספות ולכן

$$|B_l(x^*) \cap B_l(y^*)| \geq F(n-1) + \left\lfloor\frac{n}{2}\right\rfloor + 2$$

הטענה מתקיימת עבור n זוגי ואי-זוגי ולכן:

$.M(n) \geq |B_l(x^*) \cap B_l(y^*)| \geq F(n)$

3) **טענה 3: יהיו $x, y \in \{0,1\}^n$ אזי עבור $x' = x_n x_{n-1} \dots x_1, y' = y_n y_{n-1} \dots y_1$**

$$|B_l(x) \cap B_l(y)| = |B_l(x') \cap B_l(y')|$$

יהי $w \in B_l(x) \cap B_l(y)$ שנוצר ע"י הסרת הביט d_x מ-x והכנסת הביט m במיקום i_x ונוצר גם ע"י הסרת הביט d_y מ-y והכנסת הביט b במיקום i_y:

$$w = x_1 x_2 \dots x_{d_x - 1} x_{d_x + 1} \dots x_{i_x - 1} m x_{i_x} x_{i_x + 1} \dots x_n$$

נסיר מ-x' את הביט עם אינדקס $n - d_x$:

$$x_n x_{n-1} \dots x_{d_x + 1} x_{d_x - 1} \dots x_2 x_1$$

נוסיף את הביט m בין i_x ל-i_{x-1}:

$$x_n x_{n-1} \dots x_{i_x + 1} x_{i_x} m x_{i_x - 1} \dots x_{d_x + 1} x_{d_x - 1} \dots x_2 x_1$$

קיבלנו את $w' = w_n w_{n-1} \dots w_2 w_1$ ולכן $w' \in B_l(x')$.

באופן דומה $w' \in B_l(y')$ ולכן $w' \in B_l(x') \cap B_l(y')$ ומתקיים

$$|B_l(x) \cap B_l(y)| \leq |B_l(x') \cap B_l(y')|$$

באופן סימטרי נקבל גם

$$|B_l(x) \cap B_l(y)| \geq |B_l(x') \cap B_l(y')|$$

ולכן

$$|B_l(x) \cap B_l(y)| = |B_l(x') \cap B_l(y')|$$

4) **טענה 4: לכל $n \geq 10$, $x, y \in \{0,1\}^n$ כך ש-$x_1 \neq y_1, x_n \neq y_n$: $|B_l(x) \cap B_l(y)| \leq F(n)$**

מקרה 1: נניח כי לכל $1 \leq i \leq n$ $x_i = y_i$.

יהי $w \in B_l(x) \cap B_l(y)$. מאחר ו $x_1 \neq y_1, x_n \neq y_n$ בהכרח מתקיים אחד מהבאים:

1. $d_x = 1, i_y = n$
2. $d_x = 1, d_y = n$
3. $i_x = 1, i_y = n$
4. $i_x = 1, d_y = n$,
5. $d_y = 1, i_x = n$
6. $d_y = 1, d_x = n$
7. $i_y = 1, i_x = n$,
8. $i_y = 1, d_x = n$
9. $i_x = 1, d_x = n$
10. $d_x = 1, i_x = n$
11. $i_y = 1, d_y = n$
12. $d_y = 1, i_y = n$

1.1) $d_x = 1, i_y = n$, נניח $i_x \leq d_y$:

$$w = x_2 x_3 \dots x_{i_x - 1} m x_{i_x} x_{i_x + 1} \dots \dots x_{d_y - 1} x_{d_y} x_{d_y + 1} \dots x_n$$
$$w = \overline{x_1} x_2 \dots x_{i_x - 2} x_{i_x - 1} x_{i_x} \dots \dots \dots x_{d_y - 1} x_{d_y + 1} \dots \ \overline{x_n} b$$

מתקיים

$$\overline{x_1} = x_2 = x_3 = \cdots = x_{i_x-1} = m$$

כלומר הערך של m נקבע לפי ערך הביטים שלפניו. מאחר והם כולם זהים אין משמעות למיקום ויש רק אפשרות אחת כזו במקרה 1.1.

כעת נניח $i_x > d_y$:

$$w = x_2x_3 \dots x_{d_y-1}x_{d_y}x_{d_y+1} \dots \dots x_{i_x-1}mx_{i_x}x_{i_x+1} \dots x_n$$
$$w = \overline{x_1}x_2 \dots x_{d_y-2}x_{d_y-1}x_{d_y+1} \dots x_{i_x-1}x_{i_x} \dots \dots \dots \overline{x_n}b$$

מתקיים

$$m = x_{i_x} = x_{i_x+1} = \cdots = x_{n-1} = \overline{x_n}$$

לכן, כמו קודם, גם פה יש אפשרות אחת בלבד.

1.2) $d_x = 1, d_y = n$: לאחר מחיקת הביט d_x יש $n+1$ אפשרויות ל$insertion$, כפי שראינו בהרצאה.

1.3) $i_x = 1, i_y = n$: נניח $d_x \le d_y$:

$$w = mx_1x_2 \dots x_{d_x-1}x_{d_x+1} \dots x_{d_y-1}x_{d_y}x_{d_y+1} \dots\ x_n$$
$$w = \overline{x_1}x_2x_3 \dots x_{d_x}x_{d_x+1} \dots \dots x_{d_y-1}x_{d_y+1} \dots \dots \overline{x_n}b$$

מתקיים

$$x_1 = x_2 = \cdots = x_{d_x-1} = x_{d_X}$$

כלומר כל הביטים זהים עד לביט שנחמק מx ולכן יש רק אפשרות אחת לקביעת אינדקס זה.

אם $d_y < d_x$ באופן סימטרי גם כן יש אפשרות אחת.

1.4) סימטרי ל1.1 – 2 אפשרויות.

סה״כ אפשרויות ב1.1-1.4: $2+2+2+n+1 = n+7$.

1.5-1.8) סימטרי ל1.1-1.4: $n+7$ אפשרויות.

1.9-1.12) כל מקרה מגדיר אפשרות אחת ולכן סה״כ 4 אפשרויות.

כלומר יש לכל היותר $2n+18$ וקטורי $w \in B_l(x) \cap B_l(y)$.

מקרה 2: נניח כי מקרה 1 לא מתקיים ויהי $1 < i < n$ מקסימלי עבורו $x_i \neq y_i$. נניח בה"כ $i \geq \left\lfloor \frac{n}{2} \right\rfloor$ (מטענה 3 ניתן להפוך את המילים ולקבל אותו גודל חיתוך ולכן ההנחה תקפה).

2.2) $d_x = 1, d_y = n$: אם $i_x \geq i$ יש לכל היותר $2 + \left\lfloor \frac{n}{2} \right\rfloor \geq 2 + n - i$ אפשרויות.

אם $i_x < i$ אז בהכרח $i_y \leq i$ (אחרת $x_i = y_i = \overline{x_i}$ בסתירה):
נניח $i_x \leq i_y$:

$$w = x_2 x_3 \dots x_{i_x - 1} m x_{i_x} \dots \dots$$
$$w = \overline{x_1} x_2 \dots x_{i_x - 2} x_{i_x - 1} \dots \dots$$

מתקיים

$$x_2 = x_3 = \cdots = x_{i_x - 2} = x_{i_x - 1} = m$$

כלומר יש אפשרות אחת.
אם $i_x > i_y$ באופן דומה גם כן יש אפשרות אחת.

2.1) $d_x = 1, i_y = n$: אם $i_x \geq i$ זה כבר נספר ב-2.2.
אם $i_x < i$ אז בהכרח $d_y \leq i$ (אחרת $x = \overline{x_i}$ בסתירה) וגם אם $i_x < d_y$ נקבל $x_2 = \cdots = x_{i_x - 1} = m$ ולכן זה נספר ב2.2.

אם $d_y < i_x$:

$$w = x_2 x_3 \dots x_{d_y - 1} x_{d_y} x_{d_y + 1} \dots$$
$$w = \overline{x_1} x_2 \dots x_{d_y - 2} x_{d_y - 1} x_{d_y + 1} \dots$$

כלומר $x_2 = x_3 = \cdots = x_{d_y - 1} = x_{d_y}$ ולכן יש אפשרות אחת.

2.3) $i_x = 1, i_y = n$: אם $d_x \leq i$ בהכרח $d_y < i$ (אחרת $x_i = x_i$) ונניח גם $d_x \leq d_y$:

$$w = m x_1 x_2 \dots x_{d_x - 1} x_{d_x + 1} \dots$$
$$w = x_1 x_2 x_3 \dots x_{d_x} x_{d_x + 1} \dots$$

מתקיים $x_1 = x_2 = x_3 = \cdots = x_{d_x - 1} = x_{d_x}$ לכן יש אפשרות אחת לd_x.
אם $d_y < d_x$ באופן דומה יש אפשרות אחת.
אם $d_x > i$ יש לכל היותר $\left\lfloor \frac{n}{2} \right\rfloor \geq n - i$ אפשרויות.

2.4) סימטרי ל2.1 – אפשרות אחת.

ס"הכ ב 2.1-2.4 קיבלנו לכל היותר $n + 7 \geq 1 + 2 + \left\lfloor \frac{n}{2} \right\rfloor + 1 + \left\lfloor \frac{n}{2} \right\rfloor + 3$ אפשרויות.

2.5-2.8) סימטרי ל 2.1-2.4: לכל היותר $n+7$ אפשרויות.

2.9-2.12) כל אחד מגדיר אפשרות אחת

ספרנו לכל היותר $n+7+n+7+4=2n+18$ אפשרויות, כמו במקרה 1.

נראה כעת כי לכל $n\geq 10$ מתקיים $F(n)\geq 2n+18$ ומכך ינבע כי $|B_l(x)\cap B_l(y)|\leq F(n)$.

נראה זאת באינדוקציה על n:

בסיס: עבור $n=10$: $F(n)=42\geq 2n+18=38$.

סגור: נניח נכונות עבור $n-1$:

$$F(n)=F(n-1)+\left\lfloor\frac{n}{2}\right\rfloor+2\geq 2(n-1)+18+\left\lfloor\frac{n}{2}\right\rfloor+2=2n+18+\left\lfloor\frac{n}{2}\right\rfloor \geq 2n+18$$

5) **טענה 5: לכל $\boldsymbol{n\geq 12, x,y\in\{0,1\}^n}$ כך ש-$\boldsymbol{x_n\neq y_n}$ או $\boldsymbol{x_1\neq y_1}$ מתקיים $\boldsymbol{|B_l(x)\cap B_l(y)|\leq F(n)}$**

נניח בה״כ $x_n\neq y_n$ (ניתן להניח לפי טענה 3).

אם $x_1\neq y_1$ הטענה מתקיימת לפי טענה 4. לכן נניח $x_1=y_1$.

נוכיח את הטענה באינדוקציה על n.

בסיס: מחיפוש מחשב נקבל כי עבור $n=10$ $|B_l(x)\cap B_l(y)|\leq 42=F(n)$ לכל $x,y\in\{0,1\}^n$.

סגור: נניח את נכונות הטענה עבור $n-1$.

תהי $w\in B_l(x)\cap B_l(y)$. נניח בה״כ $x_1=y_1=0$ ונסמן $x=0x', y=0y'$.

5.1) אם $w_1=0$:

נסמן $w=0w'$ ונפצל למקרים הבאים:

5.1.1) אם $i_x>1, d_x>1$ הרי שניתן לעשות אותה הוספה והכנסה על x',y' ולכן $w'\in B_l(x')\cap B_l(y')$.

5.1.2) אם $i_x>1, d_x=1$ הרי שבהכרח $w_1=w_2=0$ ולכן $d_x=1$ שקול ל$d_x=2$ וכמו במקרה הקודם $w'\in B_l(x')\cap B_l(y')$

5.1.3) אם $d_x=i_x=1$ ניתן להגיע מ-x' ל-w' ללא הכנסות ומחיקות כלומר $x'=w'$.

$x_1=w_1$ ולכן גם $x=w$. כלומר ניתן להגיע מ-y ל-x ע״י הכנסה ומחיקה אחת כאשר $x_1=y_1$ ולכן $w'\in B_l(y')\Rightarrow w'\in B_l(x')\cap B_l(y')$

5.1.4) אם $i_x=1, d_x>1$: האות שנוספה היא בהכרח 0 כי $w_1=0$, כלומר: $w=0x_1\ldots$ כאשר $x_1=0$ ולכן ניתן היה להוסיף את האות 0 גם עבור $i_x=2$ ולקבול אותה מילה. כמו במקרה 1.1 מתקיים גם כאן $w'\in B_l(x')\cap B_l(y')$.

לכן, אם $w_1=0$ אזי $w'\in B_l(x')\cap B_l(y')$ ומהנחת האינדוקציה נקבל

$$|\{0w':0w'\in B_l(x)\cap B_l(y)\}|\leq |B_l(x')\cap B_l(y')|\leq F(n-1)$$

5.2) כעת נניח $w_1 = 1$. $x_1 = y_1 = 0$ ולכן חייב להתקיים אחד מהבאים:

1. $d_x = d_y = 1$ 2. $d_x = 1, i_y = 1$ 3. $i_x = 1, d_y = 1$

4. $i_y = 1, i_x = 1$

5.2.1) אם $d_x = d_y = 1$:

$x_n \neq y_n$ ולכן בהכרח $i_x = n$ או $i_y = n$ ⟸ 2 אפשרויות לכל היותר.

5.2.2) אם $d_X = 1, i_y = 1$: $x_n \neq y_n$ ולכן בהכרח $d_y = n$ או $i_x = n$ ⟸ 2 אפשרויות לכל היותר.

5.2.3) סימטרי ל-2.2.

5.2.4) סימטרי ל-2.1.

ספרנו סה״כ 8 אפשרויות עבור מקרה 5.2.

מ5.1 ו5.2 נקבל כי $|B_l(x) \cap B_l(y)| \leq F(n-1) + 8$

נוכיח ש-$F(n) \geq F(n-1) + 8$ ומכך ינבע $|B_l(x) \cap B_l(y)| \leq F(n)$.

$n \geq 12$ ולכן

$$F(n) = F(n-1) + \left\lfloor \frac{n}{2} \right\rfloor + 2 \geq F(n-1) + 6 + 2 \geq F(n-1) + 8$$

6) **טענה 6: לכל $\boldsymbol{x, y \in \{0,1\}^n}$, $\boldsymbol{x \neq y, n \geq 12}$ מתקיים**

$$|\boldsymbol{B_l(x) \cap B_l(y)}| \leq \boldsymbol{F(n)}$$

אם $x_n \neq y_n$ או $x_1 \neq y_1$ הטענה מתקיימת לפי טענה 5.

אחרת, $x_n = y_n$ וגם $x_1 = y_1$.

$x \neq y$ ולכן נוכל לסמן בi את האינדקס המינימלי עבורו $x_i \neq y_i$, כאשר $1 < i < n$. נניח בה״כ $i \leq \left\lfloor \frac{n}{2} \right\rfloor$ (ניתן להניח לפי טענה 3).

נוכיח כעת את הטענה באינדוקציה על n:

בסיס: עבור $n = 12$ נקבל בעזרת חיפוש מחשב כי לכל $x \neq y \in \{0,1\}^n$ הטענה מתקיימת.

סגור: נניח נכונות עבור $n-1$ ויהיו $w \in B_l(x) \cap B_l(y), x \neq y \in \{0,1\}^n$.

נסמן $w = w'w_n, y = y'y_n, x = x'x_n$ ונניח בה״כ $x_n = y_n = 0$.

אם $w_n = 0$ אז באופן דומה להוכחת 5.1 נקבל:

$$|\{w'0 : w'0 \in B_l(x) \cap B_l(y)\}| \leq |B_l(x') \cap B_l(y')| \leq F(n-1)$$

נניח כעת $w_n = 1$.

$x_n = y_n = 0 \neq w_n$ ולכן חייב להתקיים אחד מהבאים:

1) $d_x = d_y = n$ 2) $i_y = n, d_x = n$ 3) $d_y = n, i_x = n$

4) $i_y = n, i_x = n$

נניח בה"כ $x_{i-1} = \overline{x_i}$

6.1) אם $d_x = d_y = n$:
$x_i \neq y_i$ ולכן 6.1.1 או 6.1.2 חייבים להתקיים :

6.1.1) $i_y > i, i_x \leq i$:

$$w = x_1 x_2 \dots x_{i_x-1} m x_{i_x} x_{i_x+1} \dots x_{i-1} x_i x_{i+1} \dots$$

$$w = x_1 x_2 \dots x_{i_x-1} x_{i_x} x_{i_x+1} \dots x_{i-1} \overline{x_i} y_{i+1} y_{i+2} \dots$$

מתקיים $m = x_{i_x} = x_{i_x+1} = \cdots = x_{i-1} = \overline{x_i}$
לכן יש אפשרות יחידה כזו ובה $i_x = i, m = \overline{x_i}$

6.1.2) $i_x > i, i_y \leq i$:

$$w = x_1 x_2 \dots x_{i_y-1} x_{i_y} x_{i_y+1} \dots \ x_{i-1} x_i x_{i+1} \dots$$
$$w = x_1 x_2 \dots x_{i_y-1} b x_{i_y} \dots \dots \dots x_{i-2} x_{i-1} \overline{x_i} \dots$$

מתקיים $b = x_{i_y} = x_{i_y+1} = \cdots = x_{i-1} = x_i = \overline{x_{i+1}}$
כלומר ערך b יחיד ולכן האפשרות היחידה היא $i_y = i, b = x_i$

6.2) אם $i_y = n, d_x = n$:
$x_i \neq y_i$ ולכן חייב להתקיים אחד מהבאים :

1) $d_y \leq i, i_x > i$ 2) $d_y \geq i, i_x \leq i$ 3) $d_y \leq i, i_x \leq i$

6.2.1) $i_x > i, d_x = n$ ולכן מקרה זה כבר נספר ב6.1.2.

6.2.2) אם $d_y \geq i, i_x \leq i$:

$$w = x_1 x_2 \dots x_{i_x-1} m x_{i_x} x_{i_x+1} \dots \dots x_{i-1} x_i x_{i+1} \dots$$
$$w = x_1 x_2 \dots x_{i_x-1} x_{i_x} x_{i_x+1} \dots x_{i-1} \overline{x_i} y_{i+1} \dots \dots \dots$$

מתקיים $m = x_{i_x} = x_{i_x+1} = \cdots = x_{i-2} = x_{i-1} = \overline{x_i}$
אפשרות זו כבר נספרה ב6.1.1.

6.2.3) אם $d_y \leq i, i_x \leq i$:

$w = x_1 x_2 \dots x_{i_x-1} m x_{i_x} x_{i_x+1} \dots \dots\ \ x_{d_y-1} x_{d_y} x_{d_y+1} \dots \dots x_{i-1} x_i x_{i+1} \dots$

$w = x_1 x_2 \dots x_{i_x-1} x_{i_x} x_{i_x+1} \dots x_{d_{y-1}} x_{d_y+1} x_{d_y+2} \dots x_{i-1} \overline{x_i} y_{i+1} \dots \dots \dots$

מתקיים $x_{i-3} = x_{i-1},\ x_{i-2} = \overline{x_i}$

מאחר ו-$x_{i-1} = \overline{x_i}$ נקבל $x_{i-3} = x_{i-1} = \overline{x_i} = x_{i-2}$

וסה"כ $m = x_{i_x} = x_{i_x+1} = \dots = x_{i-2} = x_{i-1} = \overline{x_i}$

אפשרות זו כבר נספרה במקרה 6.2.2.

6.3) אם $d_y = n, i_x = n$:

$x_i \neq y_i$ ולכן חייב להתקיים אחד מהבאים :

1) $d_x \geq i, i_y \leq i$ 2) $d_x \leq i$

6.3.1) אם $d_x \geq i, i_y \leq i$:

$w = x_1 x_2 \dots x_{i_y-1} x_{i_y} x_{i_y+1} \dots\ x_{i-1} x_i x_{i+1} \dots$

$w = x_1 x_2 \dots x_{i_y-1} b x_{i_y} \dots \dots \dots x_{i-2} x_{i-1} \overline{x_i} \dots$

מתקיים $b = x_{i_y} = x_{i_y+1} = \dots = x_{i-1} = x_i$

זו אפשרות יחידה $i_y = i, b = x_i$, אך כבר נספרה במקרה 6.1.2.

6.3.2) אם $d_x \leq i$ יש מקסימום $i \leq \left\lfloor \frac{n}{2} \right\rfloor$ אפשרויות.

6.4) אם $i_y = n, i_x = n$:

$x_i \neq y_i$ אזי אחד מהבאים חייב להתקיים :

1) $d_x \leq i$ 2) $i_y \leq i, d_x > i$

4.1) אם $d_x \leq i$ מקרה כבר נספר ב6.3.2.

4.2) אם $i_y \leq i, d_x > i$:

$w = x_1 x_2 \dots x_{i_y-1} x_{i_y} x_{i_y+1} \dots\ x_{i-1} x_i x_{i+1} \dots$

$w = x_1 x_2 \dots x_{i_y-1} b x_{i_y} \dots \dots \dots x_{i-2} x_{i-1} \overline{x_i} \dots$

כאשר מקרה זה כבר נספר ב-6.3.1.

סה"כ אפשרויות עבור $w_n = 1$: לכל היותר $\left\lfloor \frac{n}{2} \right\rfloor + 2$

סה"כ אפשרויות עבור $w_n = 0$: לכל היותר $F(n-1)$

לכן קיבלנו $|B_l(x) \cap B_l(y)| \leq F(n-1) + \left\lfloor \frac{n}{2} \right\rfloor + 2 = F(n)$

7) טענה 7: לכל $n \geq 4$

$$\boldsymbol{F(n) = \begin{cases} \frac{1}{4}(n^2 + 8n - 13), & n\ odd \\ \frac{n^2}{4} + 2n - 3, & n\ even \end{cases}}$$

$$F(n) = F(n-1) + \left\lfloor \frac{n}{2} \right\rfloor + 2 = F(n-2) + \left\lfloor \frac{n-1}{2} \right\rfloor + \left\lfloor \frac{n}{2} \right\rfloor + 2 * 2 = \cdots$$

$$= F\big(n-(n-4)\big) + \left\lfloor \frac{n-(n-5)}{2} \right\rfloor + \left\lfloor \frac{n-(n-6)}{2} \right\rfloor + \cdots + \left\lfloor \frac{n}{2} \right\rfloor + 2(n-4) =$$

$$= F(4) + \left\lfloor \frac{5}{2} \right\rfloor + \left\lfloor \frac{6}{2} \right\rfloor + \cdots + \left\lfloor \frac{n}{2} \right\rfloor + 2n - 8 =$$

$$= F(4) + 2 + 3 + 3 + 4 + \cdots + \left\lfloor \frac{n}{2} \right\rfloor + 2n - 8$$

עבור n אי זוגי:

$$F(n) = F(4) + 2 + 2\left(\frac{\frac{n-5}{2}\left(3 + \frac{n-1}{2}\right)}{2}\right) + 2n - 8 =$$

$$= F(4) + 2 + \frac{3n-15}{2} + \frac{n^2 - 6n + 5}{4} + 2n - 8 =$$

$$= 9 + 2 + \frac{3n}{2} - \frac{15}{2} + \frac{n^2}{4} - \frac{6n}{4} + \frac{5}{4} + 2n - 8 = \frac{n^2}{4} + 2n - \frac{13}{4}$$

$$= \frac{1}{4}(n^2 + 8n - 13)$$

עבור n זוגי:

$$F(n) = F(4) + 2 + 2\left(\frac{\frac{n-6}{2}\left(3 + \frac{n-2}{2}\right)}{2}\right) + \frac{n}{2} + 2n - 8 =$$

$$= F(4) + 2 + \frac{3n-18}{2} + \frac{n^2 - 8n + 12}{4} + \frac{n}{2} + 2n - 8 =$$

$$= 9 + 2 + \frac{3n}{2} - \frac{18}{2} + \frac{n^2}{4} - \frac{8n}{4} + \frac{12}{4} + \frac{n}{2} + 2n - 8 = \frac{n^2}{4} + 2n - 3$$

8) סיכום

מטענה 6 קיבלנו שלכל $n \geq 12$

$$M(n) \leq F(n)$$

ומטענה 2 לכל $n \geq 6$

$$M(n) \geq F(n)$$

לכן סה"כ לכל $n \geq 12$:

$$M(n) = F(n)$$

מחיפוש מחשב נקבל גם כי לכל $4 \leq n < 12$:

$$M(n) = F(n)$$

ומטענה 7 יש לנו נוסחה סגורה ל-$F(n)$.

עבור $n < 4$ ניעזר שוב בחיפוש מחשב ונקבל:

$$M(1) = 2, \qquad M(2) = 4, \qquad M(3) = 6$$