בעיית המחרוזות

1. <u>מופע</u>: קבוצת מילים שונות באורך 3, $W=\{w_1,\ldots,w_n\}$.
2. <u>פלט</u>: האם קיימת מחרוזת T באורך $n+2$ כך שתת-המחרוזות באורך 3 של T הן בדיוק W? (כן / לא)

אלגוריתם: רדוקציה למסלול המילטון

נתאר אלגוריתם מבוסס רדוקציה למסלול המילטון:

1. <u>ממיר קלט</u>:
   נבנה גרף $G=(V,E)$:
   - קדקודים: $V=W$
   - קשתות: יש קשת בין כל זוג מילים x,y מהצורה $x=abc$ וְ- $y=bcd$.
2. <u>נפעיל</u> אלגוריתם למציאת מסלול המילטון ב- G (מימוש: קופסה שחורה)
3. <u>ממיר פלט</u>: נחזיר "כן" אם הקופסה השחורה מצאה מסלול, וְ-"לא", אחרת.

הוכחת נכונות

1. <u>משפט</u>: האלגוריתם מחזיר "כן" אם"ם קיימת מחרוזת חוקית T עבור W.
2. <u>טענה 1</u>: אם קיימת מחרוזת חוקית T עבור W אז בגרף G יש מסלול המילטון.
3. <u>טענה 2</u>: אם בגרף יש מסלול המילטון אזי קיימת מחרוזת חוקית T עבור W.

הוכחת המשפט (על סמך טענות העזר)

1. <u>$\Leftarrow$</u>:
   אם האלגוריתם מחזיר "כן" אז לפי ממיר הפלט יש מסלול המילטון ב- G.
   לכן לפי טענה 2 יש מחרוזת חוקית T עבור W.
2. <u>$\Rightarrow$</u>:
   אם יש מחרוזת חוקית T עבור W אזי מטענה 1 יש מסלול המילטון ב- G, ואז לפי ממיר הפלט, האלגוריתם מחזיר "כן".

$\square$

הוכחת טענת עזר 1

נסמן את המחרוזת T ע"י:

$$T=a_1a_2\cdots a_na_{n+1}a_{n+2}$$

[יש במחרוזת $n+2$ אותיות; כל אחת מהאותיות אנו מסמנים ע"י a_i מתאים.]

לכל $1\le i\le n$ נגדיר את x_i להיות תת-המחרוזת $a_ia_{i+1}a_{i+2}$.

<u>טענה</u>: הסדרה $(x_1,x_2,\ldots,x_n)$ היא **מסלול המילטון** ב- G.

נוכיח ראשית שהיא **מסלול** ב- G, ואז שהיא מסלול **המילטון**.

1. <u>מסלול</u>:

T חוקית $\Leftarrow$ כל x_i היא מילה ב- $W \Leftarrow$ לפי ממיר הקלט, x_i קדקוד ב- V.
לכל זוג קדקודים עוקבים x_i, x_{i+1} מתקיים:

$$\begin{array}{ccccccc} x_i & = & a_i & a_{i+1} & a_{i+2} & \\ x_{i+1} & = & & a_{i+1} & a_{i+2} & a_{i+3} \end{array}$$

כלומר הסיפא של $x_i =$ הרישא של x_{i+1} ולכן, לפי ממיר הקלט, (x_i, x_{i+1}) קשת בגרף G.
דהיינו, $x_1, \ldots, x_n$ מסלול מכוון ב- G.

2. <u>המילטון</u>:
T מחרוזת חוקית, לכן כל מילה ב- W מופיעה כתת-מחרוזת x_i כלשהי של T.
כלומר, כל מילה ב- W מופיעה כקדקוד במסלול [[אותו המסלול שדיברנו עליו עד עכשיו]].
זה אומר שכל הקדקודים ב- V מופיעים במסלול, לכן מכך שיש n מילים שונות (n קדקודים ב- G), ויש בדיוק n קדקודים מסלול – נובע שכל קדקוד מופיע בדיוק פעם אחת במסלול.

[[הסבר: יש n מילים שונות ב- W, וכולן מופיעות כתתי-מילים ב- T.
לכן יש ב- T לפחות n תתי-מילים שונות.
אבל ב- T יש **בדיוק** n מילים, לכן כל תת-מילה מופיעה בדיוק פעם אחת.
כל תת-מילה של T מופיעה כקדקוד במסלול, שגם הוא באורך n, לכן יש לפחות n קדקודים שונים במסלול, ואז כל קדקוד מופיע בדיוק פעם אחת.]]

$\square$

הוכחת טענת עזר 2

נסמן את מסלול ההמילטון ע"י:

$$P = \langle y_1, y_2, \ldots, y_n \rangle$$

נסמן את המילה [[מ- W]] שמתאימה ל- y_1 ע"י: $\qquad y_1 = abc_1$
לכל מילה אחרת, נסמן את האות האחרונה ב- y_i ע"י c_i, כלומר: $\qquad y_i = **c_i$

<u>טענה</u>: המילה $T = \underbrace{abc_1c_2 \cdots c_n}_{n+2 \text{ letters}}$ היא מחרוזת חוקית עבור W.

1. <u>שלב I</u> – תת-המחרוזות של T הן בדיוק $\{y_1, \ldots, y_n\}$:
תת-המחרוזת הראשונה היא $y_1 = abc_1$.
ניזכר שלפי ממיר הקלט, הסיפא באורך 2 של $y_i =$ הרישא באורך 2 של y_{i+1}.
לכן $y_2 = bc_1c_2 \Leftarrow$ תת-המחרוזת השנייה היא y_2.
כעת ניתן להמשיך באינדוקציה ולהוכיח שתת-המחרוזת ה- i היא בדיוק $y_i = c_{i-2}c_{i-1}c_i$.

2. <u>שלב II</u> – המילים $\{y_1, \ldots, y_n\}$ הן בדיוק W:

P מסלול המילטון, לכן כל קדקוד מופיע בדיוק פעם אחת ב- $y_1,\dots,y_n$.
לפי ממיר הקלט, הקדקודים הם המילים W, ולכן כל מילה ב- W מופיעה בדיוק פעם אחת באוסף $\{y_1,\dots,y_n\}$.

□

שלב I + שלב II $\Leftarrow$ תת-המחרוזות באורך 3 של T הן בדיוק $W \Leftarrow T$ מחרוזת חוקית.

□

ניתוח זמן ריצה

3. ממיר קלט:
(בניית הגרף G) n קדקודים וְ- n^2 בדיקת קשתות $\Leftarrow$ $O(n^2)$.
4. מימוש של הקופסה השחורה:
[כיום מאמינים שהבעיה של מציאת מסלול המילטון היא בעיה קשה מאוד [[זה עוד לא הוכח חד-משמעית]], אשר ע"מ לפתור אותה דרוש אלגוריתם הפועל בזמן אקספוננציאלי לכל הפחות.]
$O(2^n)$ במקרה הכי גרוע.
5. ממיר פלט: $O(1)$.
6. סה"כ: $O(2^n)$.

בהוכחה זו ראינו כי:

המילטון $\le$ מחרוזות

בעיית מסלול אוילר

הגדרה [מסלול אוילר]

מסלול אוילר בגרף $G=(V,E)$ הינו מסלול המשתמש בכל קשת בגרף בדיוק פעם אחת (מסלול עם $|E|$ קשתות שעובר בכל הקשתות).

תזכורת [מתי יש מסלול אוילר]

בגרף מכוון $G=(V,E)$ קיים מסלול אוילר אם"ם קיימים קדקודים $s,t\in V$ כך ש:

7. לכל $v\in V\setminus\{s,t\}$, דרגת הכניסה של v = דרגת היציאה של v.
8. עבור s: דרגת היציאה = דרגת הכניסה + 1
9. עבור t: דרגת הכניסה = דרגת היציאה + 1

(בגרף לא מכוון התנאים הם שיהיו בדיוק $|V|-2$ קדקודים בעלי דרגה אי-זוגית.)

[[הערה: אם דרגת הכניסה = דרגת היציאה עבור **כל** הקדקודים בגרף (בלי 2 קדקודים מיוחדים) אז יש **מעגל אוילר**; בקורס זה אנו כנראה מבדילים בין מסלול למעגל, וכאשר אנו אומרים "מסלול" אנו מתכוונים למסלול שאינו מעגל.]]

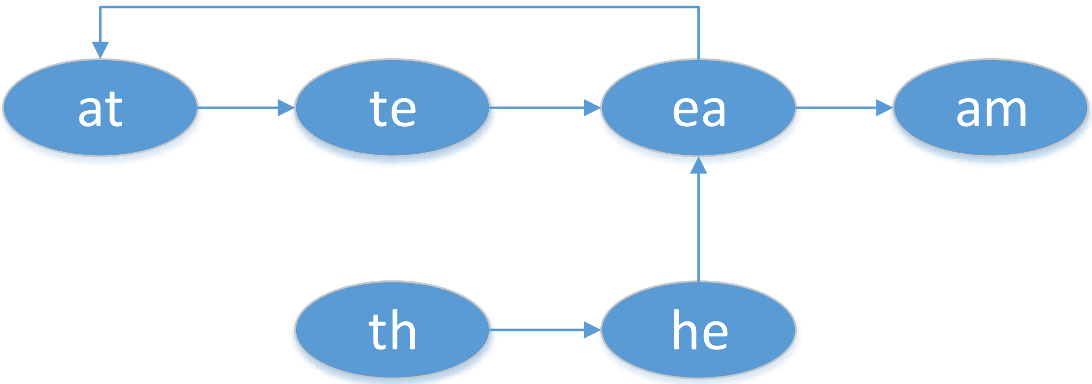
רדוקציה מבעיית המחרוזות לבעיית מסלול אוילר

1. <u>נפעיל ממיר קלט</u>:

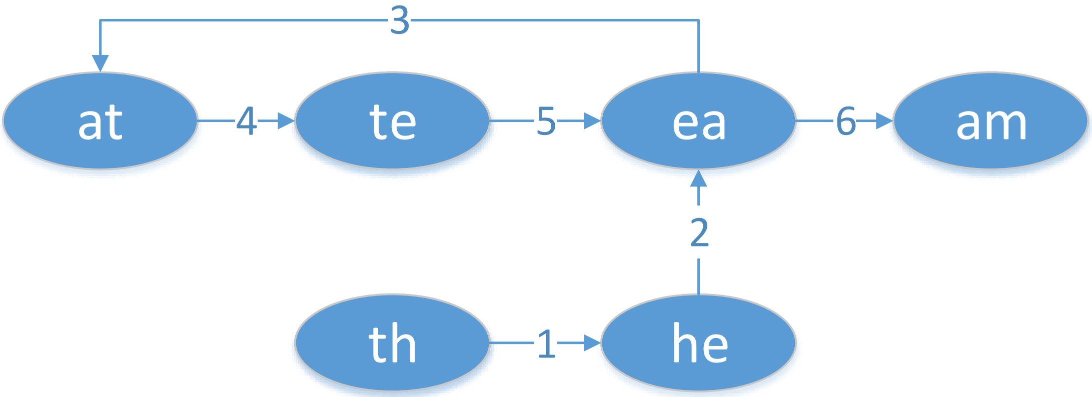
נבנה גרף מכוון $G=(V,E)$, כאשר:

- $E=W$
- לכל מילה $a_1a_2a_3 \in W$ נוסיף קדקודים a_1a_2, a_2a_3 (אם הם עוד לא קיימים), וקשת $a_1a_2a_3$ תעבור מ- a_1a_2 ל- a_2a_3.

דוגמה: $W=\{\text{ate, eat, tea, eam, the, hea}\}$



נמספר את הקשתות כדי להראות בגרף מסלול אוילר:



ואם נחבר את המילים של הקשתות לפי הסדר נקבל:

$T=\text{theateam}$

[[הערה: The fool has been pitied]]

2. <u>נפעיל אלגוריתם</u> לבדיקת קיום מסלול אוילר (מימוש הקופסה השחורה).

3. <u>נפעיל ממיר פלט</u>:
נחזיר אותה תשובה (כן / לא) כמו הקופסה השחורה.

הוכחת נכונות

- <u>משפט</u>: האלגוריתם מחזיר "כן" אם"ם קיימת מחרוזת חוקית T עבור W.
- <u>טענה 1</u>: אם קיימת מחרוזת חוקית T אזי יש מסלול אוילר בגרף שבנינו, $G=(V,E)$.
- <u>טענה 2</u>: אם יש מסלול אוילר ב- G אזי יש מחרוזת חוקית T עבור W.

הוכחת המשפט

זהה כמעט לחלוטין להוכחת המשפט ברדוקציה הקודמת.
אתם מוזמנים לנסות את זה בבית [[אם אפשר, אז בלי לשרוף שום דבר]].

הוכחת טענת עזר 1

תהי T מחרוזת חוקית; נסמנה: $T=a_1a_2\cdots a_na_{n+1}a_{n+2}$.
נגדיר לכל $i=1,\ldots,n+1$ את המילה $y_i=a_ia_{i+1}$.
<u>טענה</u>: הסדרה $\langle y_1,y_2,\ldots,y_{n+1}\rangle$ היא מסלול אוילר ב- G.
לשם ההוכחה יש להראות שהסדרה היא **מסלול**, וכן שהסדרה היא מסלול **אוילר**.

- <u>הסדרה היא מסלול</u>:
T מחרוזת חוקית, לכן כל תת-מחרוזת $a_ia_{i+1}a_{i+2}$ היא מילה ב- W.

לכן, לפי תיאור ממיר הקלט, הגרף G מכיל קדקודים $\overset{y_{i+1}}{\overbrace{a_{i+1}a_{i+2}}}$, $\overset{y_i}{a_ia_{i+1}}$ וכן קשת ביניהם.

כלומר y_i,y_{i+1} קדקודים ב- V והזוג (y_i,y_{i+1}) הוא קשת ב- E.

- <u>הסדרה היא מסלול אוילר</u>:
המחרוזת T היא חוקית, לכן אוסף תתי-המחרוזות $a_ia_{i+1}a_{i+2}$ הוא בדיוק W, ולפי מה שנאמר [בהוכחה הקודמת], זה אומר שאוסף הקשתות (y_i,y_{i+1}) הוא בדיוק W.
יש לנו בדיוק n קשתות במסלול, וכיוון שכל המילים שונות, זה אומר שהשתמשנו בכל קשת בגרף (שמתאימה למילה ב- W) בדיוק פעם אחת.
דהיינו, זהו מסלול אוילר ב- G.

□

הוכחת טענת עזר 2

[הסבר רעיוני]
כדי לעבור ממסלול אוילר $P=\langle x_1,\ldots,x_{n+1}\rangle$ למחרוזת, נסמן $x_1=ab_1$ ואת האות השנייה של x_i נסמן ב- b_i [[לכל $2\le i\le n+1$]].

צ"ל ש- $ab_1b_2\cdots b_{n+1}$ מחרוזת חוקית [וזאת עושים באופן מאוד דומה להוכחה המקבילה מהרדוקציה הקודמת].

ניתוח זמן ריצה

1. <u>ממיר הקלט</u>:
ע"מ להימנע מבניית עותקים מרובים של אותו הקדקוד, ניתן להשתמש ב-hash table [[נשמור בטבלה את כל המחרוזות שעבורם בנינו קדקוד, ונבדוק בעזרתה אם קדקוד קיים לפני שנבנה אותו]]. סיבוכיות: $O(n)$.
2. <u>בדיקת מסלול אוילר</u>: $O(|E|)=O(n)$.
3. <u>ממיר הפלט</u>: $O(1)$.

הערה חשובה

כאשר בודקים את זמן הריצה של הפעלת הקופסה השחורה, **<u>יש לשים לב</u>** לגודל הקלט אותו מכניסים לקופסה השחורה – זה לא תמיד יהיה באותו הגודל כמו הקלט המתקבל עבור הבעיה המקורית!

ע"מ להמחיש, נעשה ניסוי מחשבתי (Gedankenexperiment):

- נניח שגודל הגרף ברדוקציה הוא n^2
- נניח שלבדוק אם קיים מסלול אוילר לוקח זמן $O\left(|E|^{\frac{3}{2}}\right)$
- במקרה זה, סיבוכיות הפעלת הקופסה השחורה היא $O\left(\left(n^2\right)^{\frac{3}{2}}\right)=O\left(n^3\right)$
- [[ואם סתם היינו שמים " n " בסיבוכיות של הקופסה השחורה, היינו מקבלים $O\left(n^{\frac{3}{2}}\right)$]]