תכנון אלגוריתמים		21.10.2020
	# הרצאה 2: רדוקציות	
מרצה: אורי שטמר		*Textbook: Cormen, Leiserson, Rivest, Stein. Introduction to Algorithms.*

אנחנו ממשיכים לנתח את האלגוריתם שהצגנו לבעיית תתי המחרוזות (האלגוריתם מבוסס רדוקציה לבעיית מסלול המילטון). בסוף השיעור הקודם הצגנו את המשפט הבא ואת 2 טענות העזר הבאות:

<u>משפט:</u> אם יש פתרון אז האלגוריתם יחזיר פתרון. אם אין פתרון אז האלגוריתם יחזיר שאין פתרון.

<u>טענה 1:</u> אם אין פתרון אז אין מסלול המילטון בגרף

<u>טענה 2:</u> אם יש פתרון לבעיה אז בגרף קיים מסלול המילטון. יתרה מכך, המחרוזת שממיר הפלט בונה ממסלול המילטון היא פתרון לבעיה.

ראינו איך המשפט נובע משתי טענות העזר. עכשיו אנחנו צריכים להוכיח את טענות העזר.

נתחיל מהוכחת טענה 1.

טענה שקולה לטענה 1: אם יש מסלול המילטון בגרף אזי קיים פתרון לבעיית המחרוזות

<u>הוכחת טענה 1 (ע"י הוכחת הטענה השקולה):</u>

נסתכל על מסלול המילטון בגרף: $P = (v_1, v_2, \ldots, v_n)$.
(מכיוון שזהו מסלול המילטון, כל צומת w_i מופיעה כאן <u>בדיוק</u> פעם אחת)
נריץ את ממיר הפלט על המסלול ונקבל מחרוזת T.

נראה ש- T הוא פתרון חוקי, מה שיוכיח את הטענה.

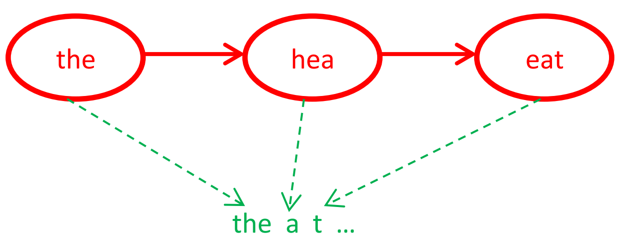
לצורך כך, תהי w_j מחרוזת קלט כלשהי. נרצה להראות ש- w_j היא תת מחרוזת של T.

ראשית, בגלל ש- P הוא מסלול המילטון, הקוד' w_j מופיע במסלול P (בדיוק פעם אחת). כלומר, קיים v_ℓ ב- P כך ש- $w_j = v_\ell$.

לכן, כדי להראות ש- w_j היא תת מחרוזת של T מספיק להראות שלכל קודקוד v_i במסלול P מתקיים שהמחרוזת שלו נמצאת ב- T.

למה שזה יהיה נכון?
נזכר איך ממיר הפלט בונה את המחרוזת T מתוך המסלול P:

לוקחים את 3 האותיות של הקוד' הראשון
ואח"כ את האות האחרונה של כל קוד' במסלול



בגלל שיש <u>קשתות</u> בגרף לאורך המסלול אז גם כשמהקוד' השני לקחנו רק את a, לפני כן יש לנו את he מהקוד' הקודם ולכן למרות שמהקוד' השני לקחנו רק את a עדיין המחרוזת של הקוד' השני נמצאת ב- T. זה נכון גם עבור שאר הקוד' במסלול, כלומר לכל קוד' במסלול P, המחרוזת שלו נמצאת ב- T.
(אפשר לפרמל זאת באינדוקציה.)

$\Leftarrow$ מסקנה: כל v_i היא תת מחרוזת של T ולכן אנחנו מקבלים של מחרוזת קלט w_j היא תת מחרוזת של T.

$\Leftarrow$ מכיוון שאנחנו מניחים (לשם פשטות) שהמחרוזות בקלט שונות זו מזו אנחנו מקבלים של- T יש <u>לפחות</u> n תתי מחרוזות שונות באורך שלוש – כל מחרוזות הקלט.

מכיוון ש- T היא באורך $n+2$, יש לה לכל היותר n תתי מחרוזות שונות באורך 3 ולכן אין ל- T תתי מחרוזות נוספות.

$\Leftarrow$ מסקנה: מחרוזות הקלט הן <u>בדיוק</u> כל תתי המחרוזות של T באורך 3.

מ.ש.ל.

כעת אנחנו רוצים להוכיח את טענה 2. נסתכל שוב על טענה 2:

טענה 2 [חלק א]: אם יש פתרון לבעיה אז בגרף קיים מסלול המילטון.
טענה 2 [חלק ב]: יתרה מכך, המחרוזת שממיר הפלט בונה ממסלול המילטון היא פתרון לבעיה.

הוכחנו את טענה 1 ע"י כך שהראינו שאם מריצים את ממיר הפלט על מסלול המילטון בגרף אז ממיר הפלט מחזיר פתרון לבעיה. כלומר את [חלק ב] של טענה 2 כבר הוכחנו. לכן, כדי להוכיח את טענה 2 מספיק להראות שאם יש פתרון לבעיה אז בגרף קיים מסלול המילטון.

נראה זאת בדוגמה שלנו:

THEATEAM

THE → HEA → EAT → ATE → TEA → EAM

יש פה n קודקודים, כל קודקוד מופיע בדיוק פעם אחת וזהו מסלול המילטון (יש קשת מקוד' לקוד' כי 2 האותיות הראשונות הן 2 האותיות האחרונות).

<u>הוכחת טענה 2:</u>
לפי תנאי הטענה, קיים פתרון לבעיה. ניקח פתרון כנ"ל T.

נסמן ב- $x_1, \dots, x_n$ את תתי המחרוזות של T באורך 3 לפי סדר ההופעה. כלומר, x_i היא תת המחרוזת באורך 3 המתחילה בתו ה i של T.
(T באורך $n+2$ ולכן יש לו בדיוק n תתי מחרוזות באורך 3).

מכיוון ש- T הוא פתרון, כל x_i כנ"ל מתאימה לבדיוק אחת ממחרוזות הקלט (ולכן לאחד הקודקודים בגרף). בנוסף, עפ"י הגדרת $x_1, \dots, x_n$, שתי האותיות האחרונות ב- x_i הן שתי האותיות הראשונות ב- x_{i+1}.
כלומר, $x_1 \to x_2 \to \cdots \to x_n$ הוא מסלול באורך n בגרף.

כמו שאמרנו, כל x_i כנ"ל מתאימה לבדיוק אחת ממחרוזות הקלט, ולכן כל מחרוזת w_i מהקלט מופיעה בדיוק פעם אחת במסלול הנ"ל. לכן זהו מסלול המילטון.

מ.ש.ל. (זה מסיים את הוכחת נכונות הרדוקציה)

האלגוריתם:

קלט: $w_1, \ldots, w_n$

- ממיר הקלט: בנה גרף G מ- $w_1, \ldots, w_n$
- מצא מסלול המילטון בגרף G אם יש, אחרת החזר "אין מסלול"
- ממיר הפלט: בנה את המחרוזת T מתוך המסלול אם יש, ואחרת החזר שאין פתרון.

ניתוח זמן הריצה:

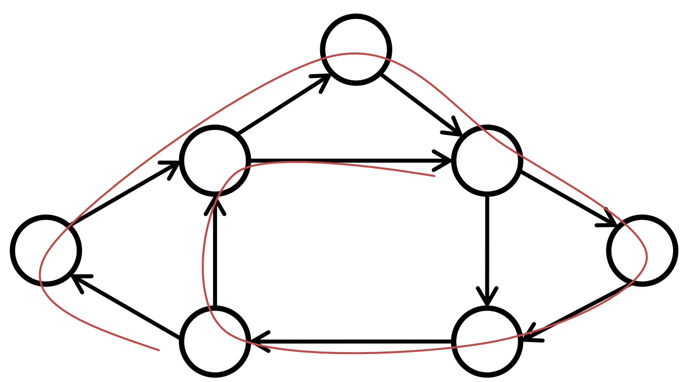
- בניית הגרף $O(n^2)$ כי לכל זוג $w_i w_j$ צריך לבדוק אם יש קשת מ- w_i ל- w_j
- **??? האלגוריתם הטוב ביותר הידוע הוא בסיבוכיות $2^{O(n)}$**
- $O(n)$ עבור בנית המחרוזות

מה קרה פה???
קיבלנו אלגוריתם עם זמן ריצה אקספוננציאלי! זה גרוע!
נסינו למצוא פתרון יעיל ומצאנו פתרון לא יעיל.
לבעיה המקורית שלנו (בעית תתי המחרוזות) כן קיים פתרון יעיל.
הגענו למצב הזה משום שעשינו רדוקציה מבעיה קלה יותר לבעיה קשה יותר.

בקורס "מבנים בדידים וקומבינטוריקה" ראיתם 2 בעיות דומות – בעית מסלול המילטון ובעית מסלול אוילר – שלאחת מהן ידוע פתרון יעיל ולאחת לא. לכן נרצה עכשיו לעשות רדוקציה מבעית תתי המחרוזות לבעיית מסלול אוילר.

בעיית מסלול אוילר:

קלט: גרף מכוון $G = (V, E)$
פלט: מסלול שעובר בדיוק פעם אחת דרך כל <u>קשת</u> בגרף, או מחזיר שאין מסלול כנ"ל.



בקורס "מבנים בדידים וקומבינטוריקה" ראיתם את המשפט הבא:

משפט אוילר:

בגרף מכוון וקשיר יש מסלול אוילר אם ורק אם אחד משני התנאים הבאים מתקיים:

(1) לכל קוד' בגרף מתקיים שדרגת הכניסה שלו שווה לדרגת היציאה שלו (ואז יש מעגל אוילר)
(2) לכל הצמתים פרט לשני צמתים בגרף דרגת הכניסה שווה לדרגת היציאה, עבור צומת אחד דרגת היציאה גדולה בדיוק באחד מדרגת הכניסה ועבור צומת אחרת דרגת הכניסה גדולה בדיוק באחד מדרגת היציאה.

- קיים אלגוריתם בזמן $O(|E|)$ שמוצא מסלול אוילר אם קיים.

אנחנו רוצים להראות רדוקציה מבעיית תתי המחרוזות לבעיית מסלול אוילר.
כעת כל קשת צריכה להתאים למחרוזת בקלט

תאור פורמלי של הרדוקציה:

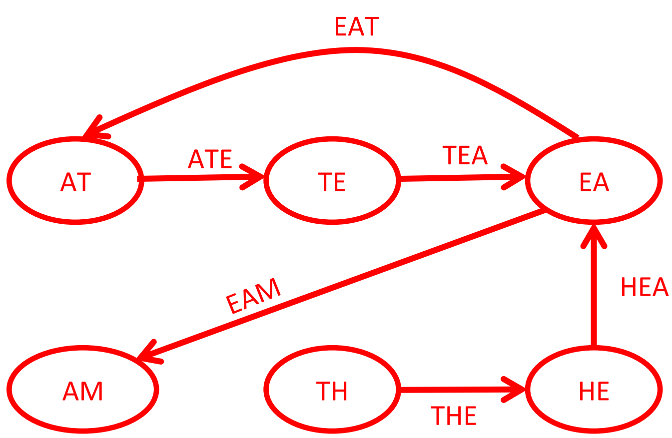
ממיר קלט: בהינתן קלט $w_1, \ldots, w_n$ נבנה גרף מכוון $G = (V, E)$ כאשר:

* צומת $v \in V$ מתאימה למחרוזת באורך 2 שהיא או שתי אותיות ראשונות במילת קלט או שני אותיות אחרונות.
  הערה: כל מחרוזת כנ"ל תופיע פעם אחת בלבד.
  למשל, בדוגמה שלנו עם הקלט ATE, EAT, TEA, EAM, THE, HEA נקבל שהמחרוזת EA תופיע רק פעם אחת בתור קודקוד, למרות שהיא מופיעה 4 פעמים כתת מחרוזת של מילת קלט.

* עבור כל מחרוזת w_i תהייה קשת בין הצומת המתאימה לשתי האותיות הראשונות של w_i לבין הצומת המתאימה לשתי האותיות האחרונות.

בדוגמה שלנו:

ATE, EAT, TEA, EAM, THE, HEA



ממיר פלט:
בהינתן מסלול אוילר P נבנה מחרוזת T שמתחילה בשתי האותיות של הקודקוד הראשון במסלול, אח"כ האות האחרונה של הקודקוד השני במסלול, אח"כ האות האחרונה של הקודקוד השלישי וכן הלאה עד לאות האחרונה של הקודקוד האחרון במסלול.

האלגוריתם שמבוסס על הרדוקציה:

(1) הרץ את ממיר הקלט, כלומר בהינתן $w_1, \ldots, w_n$ בנה את הגרף G המתואר בממיר הקלט.
(2) מצא מסלול אוילר בגרף, או החזר שאין מסלול כזה.
(3) אם קיבלת מסלול אז בנה מחרוזת עפ"י ממיר הפלט, אחרת החזר שאין פתרון.

הוכחת נכונות:

משפט: האלגוריתם מחזיר פתרון אם קיים ואחרת מחזיר שאין פתרון

<u>טענה 1:</u> אם אין פתרון אז אין מסלול אויילר בגרף

<u>טענה 2:</u> אם יש פתרון אז יש מסלול אוילר בגרף, וממיר הפלט בונה פתרון מתוך המסלול.

<u>נראה קודם כי 2 הטענות גוררות את המשפט:</u>

* אם יש פתרון, אז עפ"י טענה 2 יש מסלול אוילר בגרף ולכן האלגוריתם שמוצא מסלול אוילר ימצא מסלול אוילר בגרף. עפ"י טענה 2, המחרוזת שנבנה ממנה היא פתרון.
* אם אין פתרון אז עפ"י טענה 1 אין מסלול אוילר. לכן האלגוריתם עבור בעיית מסלול אוילר יחזיר שאין מסלול ולכן האלגוריתם שבנינו יחזיר שאין פתרון.

מ.ש.ל.

כמו עם הרדוקציה הקודמת שראינו, יהיה לנו נח להסתכל על הטענה הבאה ששקולה לטענה 1:

טענה שקולה לטענה 1: אם יש מסלול אוילר בגרף אזי קיים פתרון לבעיית המחרוזות

<u>הוכחת טענה 1 (ע"י הוכחת הטענה השקולה):</u>

נסתכל על מסלול אוילר בגרף: $P = (v_1, v_2, \dots, v_{n+1})$.
(שימו לב: בגרף G בנינו קשת עבור כל מילת קלט ולכן ישנן n קשתות ולכן מסלול אוילר מכיל בדיוק $n+1$ קוד')
נריץ את ממיר הפלט על המסלול ונקבל מחרוזת T. נראה ש- T הוא פתרון חוקי, מה שיוכיח את הטענה.

סימון: עבור מחרוזת A באורך 2 נסמן ב [A] את התו השני של A.

בעזרת הסימון הזה, אנחנו יכולים לרשום את T בצורה הבאה:

$$T = v_1, [v_2], [v_3], \dots, [v_{n+1}]$$

נשים לב ש- T היא מחרוזת באורך $n+2$.

מכיוון ש- P הוא מסלול אוילר, אנחנו יודעים שהוא עובר דרך כל קשת בגרף (בדיוק פעם אחת). תהי w_j מילת קלט מסויימת ויהי i כך שהקשת שמתאימה ל- w_j במסלול P היא (v_i, v_{i+1}). לכן, מאופן בניית הגרף, מתקיים $w_j = v_i, [v_{i+1}]$.
אם $i = 1$ אזי אנחנו מיד רואים ש- w_j היא תת מחרוזת של $T = v_1, [v_2], \dots$.
אם $i > 1$ אז נסתכל על $A = [v_{i-1}], [v_i], [v_{i+1}]$, שהיא תת מחרוזת של T. מכיוון שיש בגרף קשת מ- v_{i-1} ל- v_i, מאופן בניית G, אנחנו יודעים שהאות השניה של v_{i-1} שווה לאות הראשונה של v_i. לכן $A = v_i, [v_{i+1}] = w_j$ ולכן w_j היא תת מחרוזת של T.

$\Leftarrow$ מסקנה: כל מחרוזת קלט היא תת מחרוזת של T.

$\Leftarrow$ מכיוון שאנחנו מניחים (לשם פשטות) שהמחרוזות בקלט שונות זו מזו אנחנו מקבלים של- T יש <u>לפחות</u> n תתי מחרוזות שונות באורך שלוש – כל מחרוזות הקלט.

מכיוון ש- T היא באורך $n+2$, יש לה לכל היותר n תתי מחרוזות שונות באורך 3 ולכן אין ל- T תתי מחרוזות נוספות.

$\Leftarrow$ מסקנה: מחרוזות הקלט הן <u>בדיוק</u> כל תתי המחרוזות של T באורך 3.

מ.ש.ל.

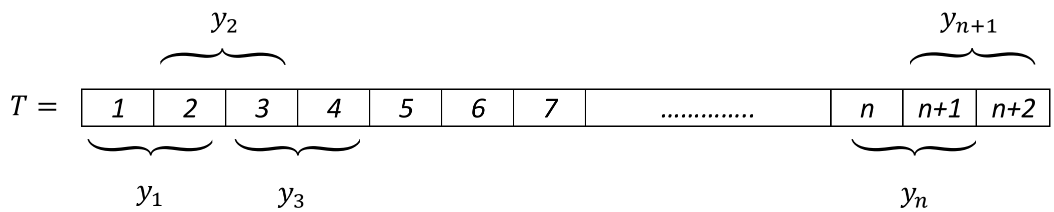
כמו עם הרדוקציה הקודמת שראינו, לאור טענה 1, כדי להוכיח א טענה 2 מספיק להוכיח את

__**טענה 2 [חלק א]: אם יש פתרון אז יש מסלול אוילר בגרף**__

__**הוכחה:**__
תהי T מחרוזת באורך $n+2$ המהווה פתרון עבור מחרוזות הקלט $w_1, w_2, \dots, w_n$.
נבנה מ- T מסלול אוילר בגרף.

עבור $i = 1,2, \dots, n+1$ נסמן ב- y_i את תת המחרוזת באורך 2 המתחילה בתו ה i של T.

בציור:



נראה כי $P = (y_1, y_2, \dots, y_{n+1})$ הוא <u>מסלול אוילר</u> בגרף.

(א) נראה שלכל i מתקיים ש- y_i הוא <u>קודקוד</u> בגרף:
תהי x_i תת המחרוזת באורך 3 המתחילה בתו ה i של T. מכיוון ש- T היא פתרון חוקי, אזי x_i היא אחת ממחרוזות הקלט. לכן בגרף קיים קודקוד y_i (=שתי האותיות הראשונות של x_i)
(צריך להתייחס בנפרד ל- y_{n+1} אבל זה דומה)

(ב) נראה שלכל i מתקיים ש- (y_i, y_{i+1}) היא קשת בגרף:
שוב, x_i היא מחרוזת קלט ומתקיים ששתי האותיות הראשונות הן y_i ושתי האותיות האחרונות שלה הן y_{i+1}. לכן (y_i, y_{i+1}) היא קשת בגרף.

מ (א)+(ב) אנחנו מקבלים ש- P הוא <u>מסלול</u> בגרף. מדוע הוא מסלול <u>אוילר</u>?

(ג) נראה ש- P הוא מסלול אוילר:
- לכל תת מחרוזת באורך 3 (סימנו כ- x_i) התאמנו קשת במסלול.
- מכיוון ש- T היא פתרון חוקי, כל תתי המחרוזות באורך 3 <u>שונות</u> זו מזו (כיוון שהן מתאימות למחרוזות הקלט שהן שונות זו מזו).
- לכן במסלול P יש n קשתות שונות זו מזו, כי אם $(y_i, y_{i+1}) = (y_j, y_{j+1})$ אזי $x_i = x_j$
- מכיוון שבגרף G יש בדיוק n קשתות, נובע כי כל אחת מהן מופיעה ב- P בדיוק פעם אחת
- לכן P הוא מסלול אוילר.

__**זמן ריצת האלגוריתם:**__
* ממיר הקלט: צריך לבנות את הגרף. עבור כל זוג אותיות (ראשונות או אחרונות) צריך לבדוק אם הופיעו בעבר (ואם לא אז ליצור קודקוד חדש עבורן). לשם כך נמיין את המחרוזות (באורך 2) עפ"י סדר לקסיקוגרפי. סה"כ זמן ריצה $O(n \cdot \log n)$.
הערה: ניתן להראות מימוש יעיל יותר.
* מציאת מסלול אוילר $O(|E|) = O(n)$
* ממיר הפלט $O(n)$

סה"כ זמן ריצה $O(n \cdot \log n)$.