# קודי הופמן ודחיסת נתונים

מפגש 6

• בעיית עץ הפורש המינימלי

בעיית עץ הפורש המינימליאלגוריתם של פרים

- בעיית עץ הפורש המינימלי
  - אלגוריתם של פרים
- האלגוריתם של קרוסקל

• דחיסת נתונים באמצעות קידוד הופמן

דחיסת נתונים באמצעות קידוד הופמןהצגת הבעיה

- דחיסת נתונים באמצעות קידוד הופמן
  - הצגת הבעיה
    - הגדרות

- דחיסת נתונים באמצעות קידוד הופמן
  - הצגת הבעיה
    - הגדרות
  - האלגוריתם של הופמן



## מטרות התקשורת

## מטרות התקשורת

תקשורת יעילה •

## מטרות התקשורת

- תקשורת יעילה •
- ניתנת לפענוח

$$f(space) = 4/24 \cdot$$

$$f(space) = 4/24 \cdot f(o) = 3/24 \cdot$$

$$f(o) = 3/24$$

$$f(space) = 4/24 \cdot f(o) = 3/24 \cdot f(i) = 2/24 \cdot f(i)$$

### i love algorithm course.

$$f(space) = 4/24 \cdot f(o) = 3/24 \cdot f(i) = 2/24 \cdot f(i)$$

6

0 - יהיה space יהיה •

- הקידוד של space יהיה
  - 1 הקידוד של ס יהיה

- 0 הקידוד של space הקידוד של
  - הקידוד של o יהיה 1
  - הקידוד של i יהיה 01

- הקידוד של space יהיה
  - הקידוד של ס יהיה 1
  - הקידוד של i יהיה 01
    - ...I")I •

- הקידוד של space יהיה
  - הקידוד של ס יהיה 1
  - הקידוד של i יהיה 01
    - ...I")I •

- 0 יהיה space יהיה
  - הקידוד של o יהיה 1
  - הקידוד של i יהיה 01
    - ...I")I •

נקבל: 01001101001011010010100100

- 0 הקידוד של space יהיה
  - הקידוד של o יהיה 1
  - הקידוד של i יהיה 01
    - ...I")I •

נקבל: 0100110100101010100101 איך אפשר לקרוא את זה?

א"ב שאנחנו רוצים לקודד - S •

- א"ב שאנחנו רוצים לקודד S
  - פונק' שכיחויות של הא"ב f

$$f:S \to (0,\infty)$$

- א"ב שאנחנו רוצים לקודד S
  - פונק' שכיחויות של הא"ב f

$$f:S \to (0,\infty)$$

קידוד תחיליות של א"ב -  $\gamma$  •

$$\gamma: \mathsf{S} \to \{\mathsf{0},\mathsf{1}\}^*$$

- א"ב שאנחנו רוצים לקודד S
  - פונק' שכיחויות של הא"ב f

$$f:S\to (0,\infty)$$

קידוד תחיליות של א"ב -  $\gamma$  •

$$\gamma: \mathsf{S} \to \{\mathsf{0},\mathsf{1}\}^*$$

.  $\gamma$  מספר ביטים ממוצע לאות בקידוד -  $ABL(\gamma)$ 

$$ABL(\gamma) = \sum_{x \in S} f(x) \cdot |\gamma(x)|$$

 $f:S o (0,\infty)$  ושכיחויות S הקלט: קבוצה •

- $f: \mathcal{S} o (0,\infty)$  ושכיחויות  $\mathcal{S}$  הקלט: קבוצה  $oldsymbol{\cdot}$
- $\gamma:S \to \{0,1\}^*$  מיפוי יפוי  $x,y \in S$  מיפוי מעקיים תנאי התחיליות:

9

- $f: \mathcal{S} o (0,\infty)$  ושכיחויות  $\mathcal{S}$  הקלט: קבוצה  $oldsymbol{\cdot}$
- $\gamma:S \to \{0,1\}^*$  מיפוי יפוי  $x,y \in S$  מיפוי מעקיים תנאי התחיליות:

9

## בעיית קוד תחיליות אופטימלי

- $f:S o (0,\infty)$  ושכיחויות S ושכיחויות
  - $\gamma: \mathit{S} 
    ightarrow \{0,1\}^*$  מיפוי  $\bullet$

כך שלכל  $x,y \in S$  שונים, מתקיים תנאי התחיליות:

. אף אחת מהמחרוזות הבינריות  $\gamma(x), \gamma(y)$  אינה החילית של אחת יאף אחת אף יאף אחת מהמחרוזות הבינריות הבינריות יאף אחת מהמחרוזות הבינריות הב

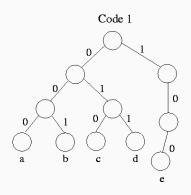
## בעיית קוד תחיליות אופטימלי

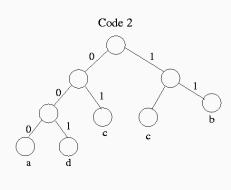
- $f:S o (0,\infty)$  ושכיחויות S ושכיחויות
  - $\gamma: \mathit{S} 
    ightarrow \{0,1\}^*$  מיפוי  $\bullet$

:טך שלכל אונים, מתקיים תנאי התחיליות מ $x,y \in S$ 

- - $ABL(\gamma) = \sum_{x \in S} f(x) |\gamma(x)|$  מטרה: למזער את הסכום •

תכנון האלגוריתם של הופמן





• משפט: קידוד הבנוי על בסיס עץ בינארי הוא קוד של תחיליות

- משפט: קידוד הבנוי על בסיס עץ בינארי הוא קוד של תחיליות
- **משפט:** עץ בינארי שמתאים לקוד אופטימלי של תחיליות הוא עץ מלא.

T ממוצע משוקלל של עומקי העלים בעץ - ABL(T)

$$ABL(T) = \sum_{x \in S} f(x) \cdot depth_{T}(x)$$

T ממוצע משוקלל של מחלים בעץ - ABL(T)

$$ABL(T) = \sum_{x \in S} f(x) \cdot depth_{T}(x)$$

 $f:S o (0,\infty)$  ושכיחויות S ושכיחויות •

.T ממוצע משוקלל של עומקי - ABL(T)

$$ABL(T) = \sum_{x \in S} f(x) \cdot depth_T(x)$$

- $f: \mathsf{S} o (0,\infty)$  ושכיחויות S אבוצה הקלט: הקלט:
  - S עץ בינארי T המכיל קבוצת עלים •

.T ממוצע משוקלל של עומקי העלים בעץ - ABL(T)

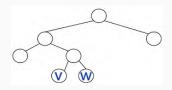
$$ABL(T) = \sum_{x \in S} f(x) \cdot depth_{T}(x)$$

- $f: \mathsf{S} o (0,\infty)$  ושכיחויות S אבוצה הקלט: הקלט:
  - S עץ בינארי T המכיל קבוצת עלים
    - ABL(T) מטרה: למזער את הסכום •

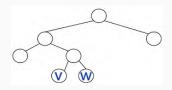
#### משפט מפתח

נביט בעץ בינארי  $T^*$  שמתאים לקוד אופטימלי של תחיליות. יהי v קודקוד עמוק ביותר בעץ. אז v עלה ויש לו שכן w שהינו גם עלה.

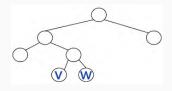
יש לנו שתי עלים אחים עמוקים ביותר - איזה אותיות הן ייצגו?



יש לנו שתי עלים אחים עמוקים ביותר - איזה אותיות הן ייצגו?

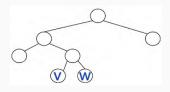


יש לנו שתי עלים אחים עמוקים ביותר - איזה אותיות הן ייצגו?

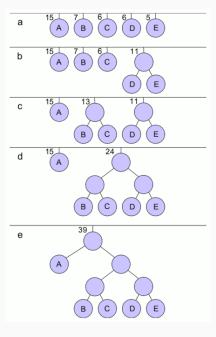


י טענה: נניח ש- u ו- v הם עלים ב-  $T^*$  עץ המותאם לקוד תחיליות אופטימלי. אם הם מקיימים depth(u) < depth(v), אזי התדירות של האות שמוצמדת ל- u גדולה מהתדירות של האות שמוצמדת ל- u

יש לנו שתי עלים אחים עמוקים ביותר - איזה אותיות הן ייצגו?



- י טענה: נניח ש- u ו- v הם עלים ב-  $T^*$  עץ המותאם לקוד תחיליות של אופטימלי. אם הם מקיימים depth(u) < depth(v), אזי התדירות של האות שמוצמדת ל- u גדולה מהתדירות של האות שמוצמדת ל- u
  - מסקנה: קיים קוד אופטימלי של תחיליות ועץ תואם  $T^*$  שבו שתי האותיות בעלות השכיחיות הנמוכות ביותר משוייכות לעלים שהם אחים ב- $T^*$ .



:S ליצירת קוד תחיליות לשפה עם א"ב

1. אם ב-S יש שתי אותיות או פחות:

:S ליצירת קוד תחיליות לשפה עם א"ב

1. אם ב-S יש שתי אותיות או פחות:

ליצירת קוד תחיליות לשפה עם א"ב S:

1. אם ב-S יש שתי אותיות או פחות:

• קבע את הראשונה בקידוד 0 ואת השניה בקידוד 1.

ליצירת קוד תחיליות לשפה עם א"ב S:

1. אם ב-S יש שתי אותיות או פחות:

• קבע את הראשונה בקידוד 0 ואת השניה בקידוד 1.

2. אחרת

ליצירת קוד תחיליות לשפה עם א"ב S:

1. אם ב-S יש שתי אותיות או פחות:

• קבע את הראשונה בקידוד 0 ואת השניה בקידוד 1.

2. אחרת

- 1. אם ב-S יש שתי אותיות או פחות:
- קבע את הראשונה בקידוד 0 ואת השניה בקידוד 1.
  - 2. אחרת
- .S- שני האיברים עם התדירות הנמוכה ביותר בx,y ויהיו, x,y

- 1. אם ב-S יש שתי אותיות או פחות:
- קבע את הראשונה בקידוד 0 ואת השניה בקידוד 1.
  - 2. אחרת
- .S-שני האיברים עם התדירות הנמוכה ביותר בx,y יהיו, 2.1

- 1. אם ב-S יש שתי אותיות או פחות:
- קבע את הראשונה בקידוד 0 ואת השניה בקידוד 1.
  - 2. אחרת
- .S-שני האיברים עם התדירות הנמוכה ביותר בx,y יהיו, 2.1
- בעלת א"ב חדש S' ע"י מחיקת x,y והחלפתם באות דשה S' צור א"ב באלת באר באות ל"ג בעלי א"ב באלי א"ב
  - T' עם עץ S'ל ל- $\gamma'$  עם עץ גור בצורה רקורסיבית קוד תחיליות א

- 1. אם ב-S יש שתי אותיות או פחות:
- קבע את הראשונה בקידוד 0 ואת השניה בקידוד 1.
  - 2. אחרת
- .S- שני האיברים עם התדירות הנמוכה ביותר בx,y ויהיו, x,y
- בעלת א"ב חדש w ע"י מחיקת x,y והחלפתם באות דעש x,y ע"י מחיקת באות באות באר באר בעלת f(x)+f(y)
  - T' עם עץ א' ל-'S, עם עץ איר בצורה רקורסיבית קוד תחיליות איר בצורה רקורסיבית 2.3
    - בצורה הבאה: S הגדר קוד תחיליות עבור

- 1. אם ב-S יש שתי אותיות או פחות:
- קבע את הראשונה בקידוד 0 ואת השניה בקידוד 1.
  - 2. אחרת
- .S- שני האיברים עם התדירות הנמוכה ביותר בx,y ויהיו, x,y
- בעלת א"ב חדש w ע"י מחיקת x,y והחלפתם באות דעש x,y ע"י מחיקת באות באות באר באר בעלת f(x)+f(y)
  - T' עם עץ א' ל-'S, עם עץ איר בצורה רקורסיבית קוד תחיליות איר בצורה רקורסיבית 2.3
    - בצורה הבאה: S הגדר קוד תחיליות עבור

- 1. אם ב-S יש שתי אותיות או פחות:
- קבע את הראשונה בקידוד 0 ואת השניה בקידוד 1.
  - 2. אחרת
- .S-שני האיברים עם התדירות הנמוכה ביותר בx,y יהיו, 2.1
- בעלת א"ב חדש w ע"י מחיקת x,y והחלפתם באות דעש x,y ע"י מחיקת באות באות באר באר בעלת f(x)+f(y)
  - T' עם עץ א' ל-'S', עם עץ אין צור בצורה רקורסיבית קוד תחיליות א
    - בצורה הבאה: S הגדר קוד תחיליות עבור
      - T' התחל עם •

- 1. אם ב-S יש שתי אותיות או פחות:
- קבע את הראשונה בקידוד 0 ואת השניה בקידוד 1.
  - 2. אחרת
- S-טותר ב-ותר הנמוכה ביותר ב-x,y יהיו, 2.1
- - T' עם עץ S'ל ל- $\gamma'$  עם עץ גור בצורה רקורסיבית קוד תחיליות
    - בצורה הבאה: S הגדר קוד תחיליות עבור 2.4
      - T' התחל עם •
  - x,y-ט והוסף לו שתי ילדים המתיוגים שמתיוג ב-w

הוכחת נכונות

### תרגיל 4.20 מהמדריך

בהינתן קבוצת תווים S שהשכיחויות שלהם הן  $(0,\infty)$ , פתחו אלגוריתם ליצירת קוד תחיליות  $S \to \{0,1,2\}$  שימזער את אלגוריתם ליצירת קוד תחיליות את הנכונות ונתחו את הסיבוכיות. כדי לקצר את  $\sum_{x \in S} f(x) | \gamma(x) |$  התשובה במקצת, אתם יכולים להניח ש־S = |S| ואי־זוגי.