

קודי הופמן ודחיסת נתונים

מפגש 6

- בעיית עץ הפורש המינימלי

- בעיית עץ הפורש המינימלי
- אלגוריתם של פרים

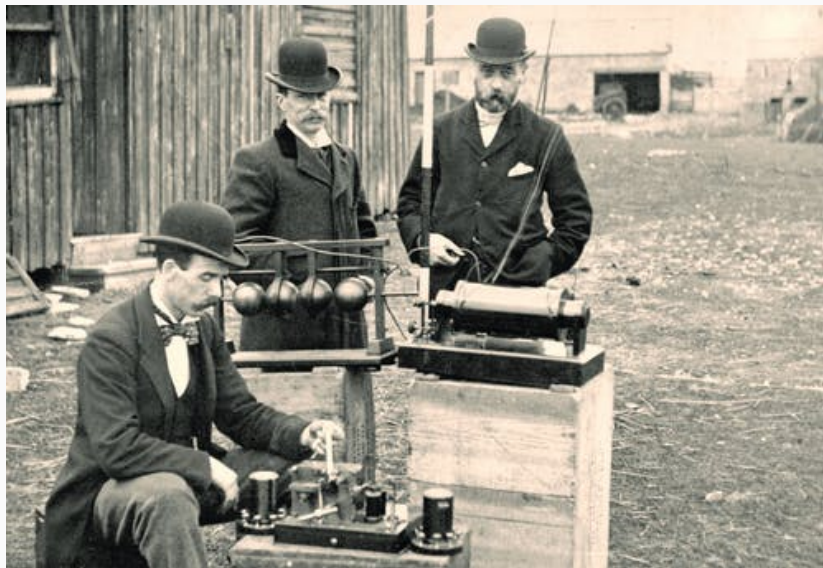
- בעיית עץ הפורש המינימלי
 - אלגוריתם של פרים
 - האלגוריתם של קרוסקל

- דחיסת נתונים באמצעות קידוד הופמן

- דחיסת נתונים באמצעות קידוד הופמן
- הצגת הבעיה

- דחיסת נתונים באמצעות קידוד הופמן
 - הצגת הבעיה
 - הגדרות

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



- תקשורת יעילה

- תקשורת יעילה
- ניתנת לפענוח

i love algorithm course.

i love algorithm course.

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

i love algorithm course.

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

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

i love algorithm course.

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

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

$$f(i) = 2/24 \cdot$$

i love algorithm course.

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

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

$$f(i) = 2/24 \cdot$$

$$\dots \cdot$$

- הקידוד של space יהיה - 0

- הקידוד של space יהיה - 0
- הקידוד של s יהיה - 1

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

- הקידוד של space יהיה - 0
- הקידוד של s יהיה - 1
- הקידוד של i יהיה - 01
- וכ"ו...

- הקידוד של space יהיה - 0
- הקידוד של s יהיה - 1
- הקידוד של i יהיה - 01
- וכ"ו...

- הקידוד של space יהיה - 0
- הקידוד של s יהיה - 1
- הקידוד של i יהיה - 01
- וכ"ו...

נקבל: 01001101001010110010100100

- הקידוד של space יהיה - 0
- הקידוד של s יהיה - 1
- הקידוד של i יהיה - 01
- וכ"ו...

נקבל: 01001101001010110010100100
איך אפשר לקרוא את זה?

• S - א"ב שאנחנו רוצים לקודד

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

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

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

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

- γ - קידוד תחיליות של א"ב

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

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

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

- γ - קידוד תחילויות של א"ב

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

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

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

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

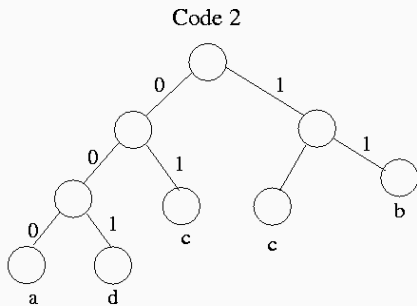
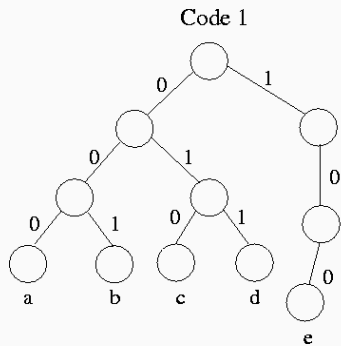
- **הקלט:** קבוצה S ושכיחויות $f : S \rightarrow (0, \infty)$
- **הפלט:** מיפוי $\gamma : S \rightarrow \{0, 1\}^*$
כך שלכל $x, y \in S$ שונים, מתקיים תנאי התחיליות:

- **הקלט:** קבוצה S ושכיחויות $f : S \rightarrow (0, \infty)$
- **הפלט:** מיפוי $\gamma : S \rightarrow \{0, 1\}^*$
כך שלכל $x, y \in S$ שונים, מתקיים תנאי התחיליות:

- **הקלט:** קבוצה S ושכיחויות $f : S \rightarrow (0, \infty)$
- **הפלט:** מיפוי $\gamma : S \rightarrow \{0, 1\}^*$
כך שלכל $x, y \in S$ שונים, מתקיים תנאי התחיליות:
 - אף אחת מהמחרוזות הבינריות $\gamma(x), \gamma(y)$ אינה תחילית של האחרת.

- **הקלט:** קבוצה S ושכיחויות $f : S \rightarrow (0, \infty)$
- **הפלט:** מיפוי $\gamma : S \rightarrow \{0, 1\}^*$
כך שלכל $x, y \in S$ שונים, מתקיים תנאי התחיליות:
 - אף אחת מהמחרוזות הבינריות $\gamma(x), \gamma(y)$ אינה תחילית של האחרת.
- **מטרה:** למזער את הסכום $ABL(\gamma) = \sum_{x \in S} f(x) |\gamma(x)|$

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.

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

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

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

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

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

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

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

- **הקלט:** קבוצה S ושכיחויות $f : S \rightarrow (0, \infty)$
- **הפלט:** עץ בינארי T המכיל קבוצת עלים S

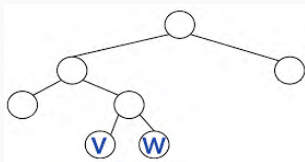
הגדרה: $ABL(T)$ - ממוצע משוקלל של עומקי העלים בעץ T .

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

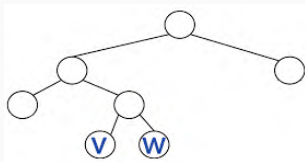
- **הקלט:** קבוצה S ושכיחויות $f : S \rightarrow (0, \infty)$
- **הפלט:** עץ בינארי T המכיל קבוצת עלים S
- **מטרה:** למזער את הסכום $ABL(T)$

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.

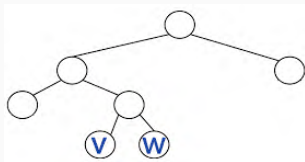
- יש לנו שתי עלים אחים עמוקים ביותר - איזה אותיות הן ייצגו?



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

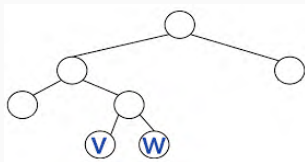


- יש לנו שתי עלים אחים עמוקים ביותר - איזה אותיות הן ייצגו?

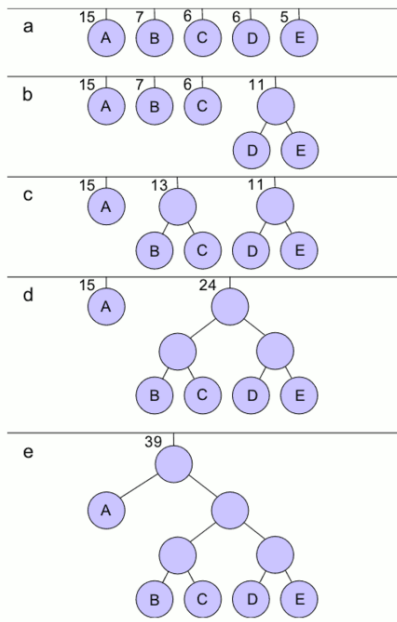


- טענה:** נניח ש- u ו- v הם עלים ב- T^* עץ המותאם לקוד תחיליות אופטימלי. אם הם מקיימים $depth(u) < depth(v)$, אזי התדירות של האות שמוצמדת ל- u גדולה מהתדירות של האות שמוצמדת ל- v .

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



- 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- v .
- מסקנה:** קיים קוד אופטימלי של תחיליות ועץ תואם T^* שבו שתי האותיות בעלות השכיחות הנמוכות ביותר משויכות לעלים שהם אחים ב- T^* .



ליצירת קוד תחיליות לשפה עם א"ב S :

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:

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 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 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

2.1 יהיו, x, y שני האיברים עם התדירות הנמוכה ביותר ב- S .

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

2.1 יהיו, x, y שני האיברים עם התדירות הנמוכה ביותר ב- S .

2.2 צור א"ב חדש S' ע"י מחיקת x, y והחלפתם באות חדשה w בעלת

התדירות $f(x) + f(y)$.

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

2.1 יהיו, x, y שני האיברים עם התדירות הנמוכה ביותר ב- S .

2.2 צור א"ב חדש S' ע"י מחיקת x, y והחלפתם באות חדשה w בעלת

התדירות $f(x) + f(y)$.

2.3 צור בצורה רקורסיבית קוד תחיליות γ' ל- S' , עם עץ T'

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

2.1 יהיו, x, y שני האיברים עם התדירות הנמוכה ביותר ב- S .

2.2 צור א"ב חדש S' ע"י מחיקת x, y והחלפתם באות חדשה w בעלת

התדירות $f(x) + f(y)$.

2.3 צור בצורה רקורסיבית קוד תחיליות γ' ל- S' , עם עץ T'

2.4 הגדר קוד תחיליות עבור S : בצורה הבאה

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

2.1 יהיו, x, y שני האיברים עם התדירות הנמוכה ביותר ב- S .

2.2 צור א"ב חדש S' ע"י מחיקת x, y והחלפתם באות חדשה w בעלת

התדירות $f(x) + f(y)$.

2.3 צור בצורה רקורסיבית קוד תחיליות γ' ל- S' , עם עץ T'

2.4 הגדר קוד תחיליות עבור S : בצורה הבאה

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

2.1 יהיו, x, y שני האיברים עם התדירות הנמוכה ביותר ב- S .

2.2 צור א"ב חדש S' ע"י מחיקת x, y והחלפתם באות חדשה w בעלת

התדירות $f(x) + f(y)$.

2.3 צור בצורה רקורסיבית קוד תחיליות γ' ל- S' , עם עץ T'

2.4 הגדר קוד תחיליות עבור S : בצורה הבאה

- התחל עם T'

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

2.1 יהיו, x, y שני האיברים עם התדירות הנמוכה ביותר ב- S .

2.2 צור א"ב חדש S' ע"י מחיקת x, y והחלפתם באות חדשה w בעלת

התדירות $f(x) + f(y)$.

2.3 צור בצורה רקורסיבית קוד תחיליות γ' ל- S' , עם עץ T'

2.4 הגדר קוד תחיליות עבור S : בצורה הבאה

- התחל עם T'

- קח את העלה שמתיוג ב- w והוסף לו שתי ילדים המתיוגים כ- x, y .

הוכחת נכונות

בהינתן קבוצת תווים S שהשכיחויות שלהם הן $f : S \rightarrow (0, \infty)$, פתחו אלגוריתם ליצירת קוד תחיליות $\gamma : S \rightarrow \{0, 1, 2\}$ שימזער את $\sum_{x \in S} f(x) |\gamma(x)|$. הוכיחו את הנכונות ונתחו את הסיבוכיות. כדי לקצר את התשובה במקצת, אתם יכולים להניח ש- $|S| \geq 3$ ואיזוגי.