



Politechnika Wroclawska

Struktury danych
i złożoność obliczeniowa
Wykład 4.

Prof. dr hab. inż. Jan Magott



Kodowanie danych wejściowych

Sposób kodowania danych wejściowych ma wpływ na rozmiar problemu (długość danych wejściowych)

Czy każdy z następujących sposobów kodowania liczb:

- Dziesiętny,
- Binarny,
- Jedynkowy

jest „właściwy” ?



Alfabet jest skończonym zbiorem symboli.

Alfabet oznaczamy symbolem Σ .

Przykłady

Alfabet dziesiętny $\Sigma = \{0,1,2, \dots, 8,9, \sqcup\}$, gdzie \sqcup jest separatorem.

Alfabet dwójkowy $\Sigma = \{0,1, \sqcup\}$.

Alfabet jedynkowy $\Sigma = \{1, \sqcup\}$.



Słowem alfabetu Σ nazywamy skończony ciąg symboli alfabetu $\Sigma \setminus \{\sqcup\}$.

Szczególnym przypadkiem jest słowo puste czyli nie zawierające żadnego symbolu

Przykłady

Słowa dziesiętne: 536, 1429, 374502

Słowa dwójkowe: 10, 0, 1001011

Słowa jedynkowe: 111, 11111, 1, 11111111



Słownikiem alfabetu Σ jest zbiór wszystkich słów alfabetu $\Sigma \setminus \{\sqcup\}$.

Przykłady

Słownik dziesiętny:

$\{0, 1, 2, \dots, 9, 10, 11, \dots, 99, 100, 101, \dots\}$

Słownik dwójkowy:

$\{0, 1, 10, 11, 100, 101, \dots\}$

Słownik jedynekowy:

$\{1, 11, 111, 1111, \dots\}$



Językiem alfabetu Σ jest skończony ciąg słów tego alfabetu oddzielonych separatorami \sqcup .

Przykłady

Język dziesiętny: $402 \sqcup 31 \sqcup 58$

Język dwójkowy: $1011 \sqcup 101111 \sqcup 11 \sqcup 1000$

Język jedynkowy: $11 \sqcup 1111 \sqcup 1 \sqcup 111 \sqcup 111$



Reguła kodowania

1. Kolejność danych musi być określona
2. Każda dana jest reprezentowana jednym słowem alfabetu Σ
3. Musi istnieć metoda wyodrębniania poszczególnych danych
4. Dane o równych wartościach kodowane są tym samym słowem



Instancja I problemu podziału zbioru

Dane:

- $X = \{x_1, \dots, x_i, \dots, x_k\} = \{3, 11, 7, 23, 18\}$ - zbiór pięciu elementów,
- $B = 31, \quad \sum_{i=1}^5 x_i = 2 \cdot B.$

Pytanie:

Czy istnieje podzbiór $X_1 \subset X$ taki, że $\sum_{x_i \in X_1} x_i = B$?



Zakodujemy dane instancji I w następujący sposób:

$$x_1 \sqcup x_2 \sqcup x_3 \sqcup x_4 \sqcup x_5 \sqcup B \sqcup \sqcup$$

Kodowanie w alfabecie dziesiętnym

$$3 \sqcup 11 \sqcup 7 \sqcup 23 \sqcup 18 \sqcup 31 \sqcup \sqcup$$

Długość łańcucha kodującego instancję I wynosi

$$N_{10}(I) = 17$$



Zakodujemy dane instancji I w następujący sposób:

$$x_1 \sqcup x_2 \sqcup x_3 \sqcup x_4 \sqcup x_5 \sqcup B \sqcup \sqcup$$

Kodowanie w alfabecie dwójkowym

$$11 \sqcup 1011 \sqcup 111 \sqcup 10111 \sqcup 10010 \sqcup 11111 \sqcup \sqcup$$

Długość łańcucha kodującego instancję I wynosi

$$N_2(I) = 31$$



Zakodujemy dane instancji I w następujący sposób:

$$x_1 \sqcup x_2 \sqcup x_3 \sqcup x_4 \sqcup x_5 \sqcup B \sqcup \sqcup$$

Kodowanie w alfabecie jedynkowym

$$111 \sqcup 1111111111111111 \sqcup 11111111 \sqcup \dots 111111111111 \sqcup \sqcup$$

Długość łańcucha kodującego instancję I wynosi

$$N_1(I) = 100$$



Problem podziału zbioru

Dane:

- $X = \{x_1, \dots, x_i, \dots, x_k\}$ - zbiór k elementów $x_i \in N_+$,
gdzie $N_+ = \{1, 2, \dots\}$,
- $B \in N_+$,
- $\sum_{i=1}^k x_i = 2B$.

Pytanie:

Czy istnieje podzbiór $X_1 \subset X$ taki, że

$$\sum_{x_i \in X_1} x_i = B ?$$



Zakodujmy dane Problemu podziału w następujący sposób:

$$x_1 \sqcup x_2 \sqcup \dots \sqcup x_n \sqcup B \sqcup \sqcup$$

Długość łańcucha kodującego w alfabecie dziesiętnym

$$\begin{aligned} N_{10}(I) &\approx [\log x_1] + 1 + [\log x_2] + 1 + \dots + [\log x_n] + 1 + [\log B] + 2 = \\ &\sum_{i=1}^n [\log x_i] + n + [\log B] + 2 \\ &\leq (n + 1) \cdot \max_{j \in \{1, n\}} \{[\log x_i] + 1, [\log B]\} + 2 \\ &\approx O(n \cdot \max_{j \in \{1, n\}} \{[\log x_i] + 1, [\log B]\}) \end{aligned}$$

$[x]$ - najmniejsza liczba całkowita nie mniejsza niż x



Zakodujemy dane Problemu podziału w następujący sposób:

$$x_1 \sqcup x_2 \sqcup \dots \sqcup x_n \sqcup B \sqcup \sqcup$$

Długość łańcucha kodującego w alfabecie dwójkowym

$$\begin{aligned} N_2(I) &\approx [\log_2 x_1] + 1 + [\log_2 x_2] + 1 + \dots + [\log_2 x_n] + 1 + [\log_2 B] + 2 = \\ &\sum_{i=1}^n [\log_2 x_i] + n + [\log_2 B] + 2 \leq \\ &\leq (n + 1) \cdot \max_{j \in \{1, n\}} \{[\log_2 x_i] + 1, [\log_2 B]\} + 2 \\ &\approx O(n \cdot \max_{j \in \{1, n\}} \{[\log_2 x_i] + 1, [\log_2 B]\}) \end{aligned}$$

$\lceil x \rceil$ - najmniejsza liczba całkowita nie mniejsza niż x



Zakodujemy dane Problemu podziału w następujący sposób:

$$x_1 \sqcup x_2 \sqcup \dots \sqcup x_n \sqcup B \sqcup \sqcup$$

Długość łańcucha kodującego w alfabecie jedynkowym

$$N_1(I) = x_1 + 1 + x_2 + 1 + \dots + x_n + 1 + B + 2 =$$

$$\sum_{i=1}^n x_i + n + B + 2$$

$$x_i \approx 10^{\log_{10} x_i}$$



„Rozsądna i zwarta” reguła kodowania to reguła, która nie powoduje wykładniczego wzrostu rozmiaru danych wejściowych kodowanej instancji w stosunku do innych reguł kodowania.

„Rozsądne i zwarte” reguły kodowania instancji I dające łańcuchy o długościach $N_n(I)$, $N_m(I)$ spełniają relacje:

$$N_n(I) \leq p(N_m(I)) \text{ i } N_m(I) \leq p'(N_n(I)) \text{ dla wielomianów } p, p'.$$

Jeśli ten wymóg nie jest spełniony i np. $N_n(I) \leq k^{N_m(I)}$ dla pewnej stałej $k > 1$, to pierwsza jest „nierozsądna” (powoduje wykładniczy wzrost rozmiaru).

Zatem odrzucamy kodowanie danych wejściowych w alfabecie jedynkowym.