

Wstęp do informatyki

Wykład 3

Uniwersytet Wrocławski

2015

Dane przetwarzane przez „komputer”

Tydzień temu:

- Wejście, wyjście, wartości przetwarzane: liczby całkowite (dowolnie duże)

Dziś:

- Podstawowa jednostka informacji, binarna reprezentacja informacji;
- Cyfrowa/binarna reprezentacja liczb: naturalnych, całkowitych, rzeczywistych;
- Cyfrowa/binarna reprezentacja: znaków, obrazów, wideo, dźwięku,...

Podstawowa jednostka informacji

Bit:

- dwa stany (0 i 1)

Dlaczego tak „mało”:

- urządzenia cyfrowe („komputery”) powinny „automatycznie” i niezawodnie odróżniać wartości

Jak uzyskać więcej „stanów” – wartości?

- ciąg k bitów może przyjmować 2^k różnych wartości

Bity a komórki pamięci komputera

Maszyna RAM:

- komórka pamięci przechowuje dowolnie dużą liczbę całkowitą

Świat rzeczywisty:

- komórka pamięci to ciąg bitów o ustalonej długości – **bajt**;
- 1 bajt = 8 bitów (zazwyczaj...);
- formalnie: bajt to najmniejsza „adresowalna” jednostka pamięci komputera

Binarna reprezentacja liczb naturalnych

Binarna reprezentacja liczb naturalnych

Ciąg binarny = ciąg, którego elementami mogą być tylko zera i jedynki.

Binarna reprezentacja liczby naturalnej $x > 0$ to ciąg binarny $x_k \dots x_1 x_0$ taki, że:

$$x = \sum_{i=0}^k x_i \cdot 2^i$$

oraz $x_k = 1$.

Binarna reprezentacja zera = 0.

Binarna reprezentacja liczb naturalnych

Fakt

Każda liczba naturalna posiada dokładnie jedną reprezentację binarną.

Problem algorytmiczny

Wejście: a – liczba naturalna

Wyjście: $x_k \dots x_1 x_0$ – ciąg tworzący binarną reprezentację liczby a

Binarna reprezentacja liczb naturalnych

Algorytm

1. $i \leftarrow 0$
2. $x[0] \leftarrow a \bmod 2$
3. $a \leftarrow a / 2$
4. dopóki $a > 0$:
 - a) $i \leftarrow i + 1$
 - b) $x[i] \leftarrow a \bmod 2$
 - c) $a \leftarrow a / 2$
5. dopóki $i \geq 0$:
 - a) wypisz $x[i]$
 - b) $i \leftarrow i - 1$

Oznaczenia:

- $a \bmod b$ = reszta z dzielenia a przez b
- a/b = wynik dzielenia całkowitego a przez b (zaokrąglenie wyniku w dół do liczby całkowitej)

Binarna reprezentacja liczb rzeczywistych

Binarna reprezentacja liczb rzeczywistych

Binarna reprezentacja liczby rzeczywistej $a > 0$ to $x_k \dots x_1 x_0$, $x_{-1} x_{-2} x_{-3} x_{-4} \dots$ taka, że:

$$a = \sum_{i=0}^k x_i \cdot 2^i + \sum_{i=1}^{\infty} x_{-i} \cdot 2^{-i}$$

oraz $x_k = 1$, $x_i \in \{0, 1\}$ dla każdego i .

x_4	x_3	x_2	x_1	x_0	,	x_{-1}	x_{-2}	x_{-3}	x_{-4}	x_{-5}	...
2^4	2^3	2^2	2^1	2^0	,	2^{-1}	2^{-2}	2^{-3}	2^{-4}	2^{-5}	...

Binarna reprezentacja liczb rzeczywistych

Binarna reprezentacja liczby rzeczywistej $a > 0$ to $x_k \dots x_1 x_0$, $x_{-1} x_{-2} x_{-3} x_{-4} \dots$ taka, że:

$$a = \sum_{i=0}^k x_i \cdot 2^i + \sum_{i=1}^{\infty} x_{-i} \cdot 2^{-i}$$

oraz $x_k = 1$, $x_i \in \{0, 1\}$ dla każdego i .

Reprezentacja nieskończona?

- Jeśli $x_{-p} = 0$ dla każdego $p > j$: wartości $x_{-j-1}, x_{-j-2}, x_{-j-3}, \dots$ pomijamy!
- Analogia: dziesiętna reprezentacja $\frac{1}{2}, \frac{1}{3}, \dots$

Binarna reprezentacja liczb rzeczywistych

Reprezentacja nieskończona?

- Jeśli $x_p = 0$ dla każdego $p > j$: wartości $x_{j-1}, x_{j-2}, x_{j-3}, \dots$ pomijamy!
- Analogia: dziesiętna reprezentacja $\frac{1}{2}, \frac{1}{3}, \dots$

Fakt

Jeśli liczba a posiada skończoną reprezentację binarną, to reprezentacja ta jest jednoznaczna.

(inaczej: każda liczba posiada co najwyżej jedną skończoną reprezentację binarną)

Binarna reprezentacja liczb rzeczywistych

Problem algorytmiczny

Wejście:

a – liczba rzeczywista z przedziału $(0;1)$;

n – dodatnia liczba naturalna

Wyjście:

$x_{-1}x_{-2}\dots x_{-n+1}x_{-n}$ – n pierwszych bitów ułamkowej części w binarnej reprezentacji liczby a
(czyli $a = 0,x_{-1}x_{-2}\dots x_{-n+1}x_{-n}\dots$)

Binarna reprezentacja liczb rzeczywistych

Algorytm

1. $i \leftarrow 1$
2. dopóki ($a \neq 0$) oraz ($i \leq n$):
 1. $a \leftarrow 2 \cdot a$
 2. jeśli ($a \geq 1$):
 - a) wypisz 1 (czyli $x_{-i} \leftarrow 1$)
 - b) $a \leftarrow a - 1$
3. w przeciwnym przypadku:
 - a) wypisz 0 (czyli $x_{-i} \leftarrow 0$)
4. $i \leftarrow i + 1$

Binarna/cyfrowa reprezentacja liczb całkowitych

Cyfrowa reprezentacja liczb całkowitych

Reprezentacje liczb całkowitych :

- znak-moduł
- kod uzupełnień do 1
- kod uzupełnień do 2

Wspólna cecha powyższych reprezentacji liczb całkowitych:

- k – ustalona długość reprezentacji (liczba bitów)
- najstarszy bit (skrajnie lewy) pozwala określić znak liczby

Reprezentacja znak-moduł

Wartość opisywana ciągiem $x_{k-1} x_{k-2} \dots x_1 x_0$

– wartość bezwzględna odpowiada binarnej reprezentacji $x_{k-2} \dots x_1 x_0$ czyli

$$\sum_{i=0}^{k-2} x_i \cdot 2^i$$

- znak liczby:
 - dodatnia gdy $x_{k-1} = 0$,
 - ujemna gdy $x_{k-1} = 1$

Kod uzupełnień do 1 (U1)

Wartość opisywana ciągiem $x_{k-1} x_{k-2} \dots x_1 x_0$

- Jeśli $x_{k-1}=0$: „standardowa” reprezentacja liczby dodatniej

$$\sum_{i=0}^{k-2} x_i \cdot 2^i$$

- Jeśli $x_{k-1}=1$: liczba ujemna, bity wartości bezwzględnej zanegowane

$$-\sum_{i=0}^{k-2} (1 - x_i) \cdot 2^i$$

Kod uzupełnień do 2 (U2)

Wartość opisywana ciągiem $x_{k-1} x_{k-2} \dots x_1 x_0$ jest równa

$$-x_{k-1} \cdot 2^{k-1} + \sum_{i=0}^{k-2} x_i \cdot 2^i$$

Zakres reprezentowanych liczb:

Kodowanie	Zakres	Najmniejsza	Największa	zero
Znak-moduł	$[-(2^{k-1} - 1), 2^{k-1} - 1]$	1 11...11	011...11	00...00 10...00
U1	$[-(2^{k-1} - 1), 2^{k-1} - 1]$	1 00...00	011...11	00...00 11...11
U2	$[-2^{k-1}, 2^{k-1} - 1]$	1 00...00	011...11	00..00

Reprezentacje liczb całkowitych

Przykład. $k=8$

Liczba	Znak-moduł	U1	U2
0	00000000 10000000	00000000 11111111	00000000
109	01101101	01101101	01101101
-109	11101101	10010010	10010011

Kod uzupełnień do 2 (U2) – liczby przeciwne

Wejście:

k – liczba naturalna,

$x = x_{k-1}, x_{k-2}, \dots, x_1, x_0$ – reprezentacja liczby a w U2 na k bitach

Wyjście: y – reprezentacja liczby $-a$ w U2 na k bitach

Algorytm – ogólnie:

1. zaneguj wszystkie bity w ciągu x (negowanie: $1 \rightarrow 0$, $0 \rightarrow 1$);
2. do zanegowanego ciągu (traktowanego jako liczba w zapisie binarnym) dodaj 1.

Kod uzupełnień do 2 (U2) – liczby przeciwne

Wejście:

k – liczba naturalna

$x = x_{k-1}, x_{k-2}, \dots, x_1, x_0$ – reprezentacja liczby a w U2 na k bitach

Wyjście: y – reprezentacja liczby $-a$ w U2 na k bitach

Algorytm:

1. Jeśli $x = 100\dots 0$:
zwróć „Brak reprezentacji”
2. $x' \leftarrow x$ z zanegowanym każdym bitem
3. $x'' \leftarrow$ binarna reprezentacja $x' + 1$
4. Jeśli $\text{długość}(x'') > k$:
 $y \leftarrow x''$ bez skrajnie lewego bitu
wpp $y \leftarrow x''$
4. Zwróć y

Kod uzupełnień do 2 (U2)

Problem

Wejście:

x - liczba całkowita,
k – dodatnia liczba
naturalna

Wyjście:

zapis liczby x w
kodzie U2 na k
bitach

Algorytm – idea:

1. Jeśli x poza przedziałem $[-2^{k-1}, 2^{k-1} - 1]$: liczba poza zakresem; zakończ
2. Jeśli $x \geq 0$: zwróć binarną (k-bitową) reprezentację x
3. Jeśli $x < 0$:
 - Zaneguj wszystkie bity binarnej (k-bitowej) reprezentacji -x
 - Dodaj 1
 - Zwróć wynik dodawania

Kod uzupełnień do 2 (U2)

Problem

Wejście:

x - liczba całkowita,
 k – dodatnia liczba naturalna

Wyjście:

zapis liczby x w kodzie U2 na k bitach

Uwaga:

$\text{binarna}(x)$ zwraca zapis binarny liczby x .

$\text{neg}(x)$ oznacza ciąg utworzony przez zanegowanie każdego bitu w x

Algorytm

1. $x' \leftarrow |x|$
2. Jeśli $x = -2^{k-1}$: zwróć 10...0
3. $bx \leftarrow \text{binarna}(x')$
4. $p \leftarrow \text{długość}(bx)$
5. Jeśli $p > k-1$: liczba przekracza zakres; stop
6. Uzupełnij bx z lewej strony $k - p$ zerami;
7. Jeśli $x \geq 0$: zwróć bx ,
w przeciwnym przypadku
 $bx \leftarrow \text{neg}(bx) + 1$
8. Zwróć bx

Kod U2 – dodawanie

Problem

Wejście:

k – liczba naturalna

x_1, x_2 – k -bitowe zapisy
liczb całkowitych w U2

Wyjście:

k -bitowy zapis w U2
liczby $x_1 + x_2$

UWAGA:

**opis na tym slajdzie
jest nieformalny!**

Algorytm

1. $x \leftarrow x_1 + x_2$ gdzie x_1 i x_2 traktowane jako liczby zapisane binarnie bez znaku,
2. Obetnij x do k skrajnie prawych bitów!
3. Potraktuj x jako liczbę w U2 oraz:
 - a) jeśli $\text{znak}(x_1) = \text{znak}(x_2)$ oraz $\text{znak}(x_1) \neq \text{znak}(x)$: zwróć „przekroczenie zakresu”
 - b) zwróć x

Kod U2 – liczby rzeczywiste

Reprezentacja liczb rzeczywistych w U2:

- k – liczba bitów na część całkowitą
- p – liczba bitów na część ułamkową

Zapis

$$x_{k-1} x_{k-2} \dots x_2 x_1 x_0, x_{-1} x_{-2} \dots x_{-(p-1)} x_{-p}$$

reprezentuje liczbę

$$-x_{k-1} \cdot 2^{k-1} + \sum_{i=0}^{k-2} x_i \cdot 2^i + \sum_{i=1}^p x_{-i} \cdot 2^{-i}$$

Zmiennopozycyjna reprezentacja liczb rzeczywistych

Reprezentacja zmiennopozycyjna

Fakt Każdą liczbę rzeczywistą $x \neq 0$ można jednoznacznie zapisać

$$x = (-1)^s \cdot m \cdot 2^c$$

gdzie

- $s \in \{0, 1\}$
- m to liczba rzeczywista, $m \in [1, 2)$
- c to liczba całkowita

Reprezentacja zmiennopozycyjna

Reprezentacja cyfrowa liczby

$$x = (-1)^s \cdot m \cdot 2^c$$

gdzie $s \in \{0, 1\}$, $m \in [1, 2)$, c to liczba całkowita:

- s – jeden bit
- binarny (zaokrąglony) zapis $1, m_1 m_2 m_3 \dots_{(2)}$ liczby m reprezentujemy przy pomocy ciągu $m_1 m_2 m_3 \dots m_M$, gdzie M - ustalone
- c – liczba w zapisie U2 na C bitach.

Razem:

Zapis na $1 + M + C$ bitach.

Reprezentacja zmiennopozycyjna

Zakres wartości

$$x = (-1)^s \cdot m \cdot 2^c$$

dla ustalonych M i C:

$$m \in \left[1, 1 + \sum_{i=1}^M \frac{1}{2^i} \right) = \left[1, 2 - \frac{1}{2^{M+1}} \right) \quad c \in \left[-2^{C-1}, 2^{C-1} - 1 \right]$$

Liczba o największym module:

$$\pm \left(2 - \frac{1}{2^{M+1}} \right) \cdot 2^{2^{C-1}-1} \approx \pm 2^{2^{C-1}}$$

Liczba o najmniejszym module:

$$\pm 2^{-2^{(C-1)}}$$

Reprezentacja zmiennopozycyjna

Przykład

$$x = (-1)^s \cdot m \cdot 2^c$$

dla $M=4$ i $C=3$:

$$m \in \left[1, 1 + \sum_{i=1}^4 \frac{1}{2^i} \right) = \left[1, 2 - \frac{1}{2^5} \right) = \left[1, 1 \frac{31}{32} \right) \quad c \in [2^{-2}, 2^2 - 1] = \left[\frac{1}{4}, 3 \right]$$

Liczba o największym module:

$$\pm \left(2 - \frac{1}{2^{M+1}} \right) \cdot 2^{2^{C-1}-1} = \pm 1 \frac{31}{32} \cdot 2^3 \approx \pm 2^4$$

Liczba o najmniejszym module:

$$\pm 2^{-2^{(C-1)}} = \pm 2^{-2}$$

Reprezentacja zmiennopozycyjna - normalizacja

Normalizacja liczby

$$x = (-1)^s \cdot m \cdot 2^c$$

dla $s \in \{0, 1\}$, $m \in \mathbb{R}_+$ oraz liczby całkowitej c polega na znalezieniu m' oraz c' takich, że

$$x = (-1)^s \cdot m \cdot 2^c = (-1)^s \cdot m' \cdot 2^{c'}$$

oraz $m' \in [1, 2)$, c' całkowite.

Przykład

$$x = 47 = 101111_{(2)} = 1 \cdot 101111_{(2)} \cdot 2^0 = 1 \cdot 1,01111_{(2)} \cdot 2^5$$

Dla $M=6$ i $C=4$ liczba x ma następującą reprezentację:

s	m	c
0	011110	0101

Reprezentacja zmiennopozycyjna - mnożenie

Wejście:

zmiennopozycyjne reprezentacje liczb x_1 i x_2 :

$$x_1 = (-1)^{s_1} \cdot m_1 \cdot 2^{c_1}, \quad x_2 = (-1)^{s_2} \cdot m_2 \cdot 2^{c_2}$$

Wyjście:

Zmiennopozycyjna reprezentacja

$$x = x_1 \cdot x_2 = (-1)^s \cdot m \cdot 2^c$$

Reprezentacja zmiennopozycyjna - mnożenie

Wejście: $s_1, m_1, c_1, s_2, m_2, c_2$

$$x_1 = (-1)^{s_1} \cdot m_1 \cdot 2^{c_1}, \quad x_2 = (-1)^{s_2} \cdot m_2 \cdot 2^{c_2}$$

Wyjście: s, m, c , takie, że $s \in \{0, 1\}$, $m \in [1, 2)$, c to liczba całkowita

$$x = x_1 \cdot x_2 = (-1)^s \cdot m \cdot 2^c$$

Algorytm:

- dodaj znaki modulo 2: $s = (s_1 + s_2) \bmod 2$
- pomnóż mantysy: $m = m_1 \cdot m_2$
- dodaj cechy: $c = c_1 + c_2$
- normalizacja wyniku!

Reprezentacja zmiennopozycyjna - dodawanie

Wejście: $s_1, m_1, c_1, s_2, m_2, c_2$

$$x_1 = (-1)^{s_1} \cdot m_1 \cdot 2^{c_1}, \quad x_2 = (-1)^{s_2} \cdot m_2 \cdot 2^{c_2}$$

Wyjście: s, m, c takie, że: $x = x_1 + x_2 = (-1)^s \cdot m \cdot 2^c$

Algorytm:

1. wyrównaj cechy do większej
2. dodaj mantysy ze znakiem:
$$m \leftarrow (-1)^{s_1} m_1 + (-1)^{s_2} m_2$$
3. Ustal znak na podstawie sumy mantys ze znakiem (i zamień mantysę na jej wartość bezwzględną):
 - Jeśli $(m > 0)$: $s \leftarrow 0$,
wpp $s \leftarrow 1$
 - $m \leftarrow |m|$
4. Normalizuj wynik!

Reprezentacja zmiennopozycyjna - dodawanie

Przykład:

$$(-1)^0 \cdot 1,101_{(2)} \cdot 2^8 + (-1)^1 \cdot 1,001_{(2)} \cdot 2^6 =$$

Wyrównanie cech

$$1,101_{(2)} \cdot 2^8 - 0,01001_{(2)} \cdot 2^8 =$$

$$(1,101_{(2)} - 0,01001_{(2)}) \cdot 2^8 =$$

dodawanie mantys

$$1,01011_{(2)} \cdot 2^8 =$$

$$(-1)^0 \cdot 1,01011_{(2)} \cdot 2^8$$

normalizacja

Reprezentacja zmiennopozycyjna a reprezentacja stałopozycyjna

Stałopozycyjna:

- Wyniki obliczeń bez błędów
- Mały zakres
- Duży rozmiar pamięci

Zmiennopozycyjna

- Błędy zaokrągleń reprezentacji
- Błędy wyników operacji arytmetycznych
- Duży zakres wartości
- Oszczędność pamięci

Reprezentacja stałopozycyjna w C i Pythonie

Ansi C:

- Typ zmiennej określa zakres wartości i rozmiar pamięci zajmowanej przez zmienną (liczbę bajtów)
- Przekroczenie zakresu wartości oznacza błędne wyniki/wartości

Python

- Interpreter dopasowuje zakres (a zarazem rozmiar pamięci) do aktualnej wartości zmiennej

Reprezentacja tekstu, grafiki, wideo,
dźwięku,...

Reprezentacja tekstu

Rozwiązanie „bazowe”

1. Ustal zestaw znaków, jego rozmiar
2. Przyporządkuj znaki kolejnym liczbom naturalnym

Modyfikacje i uszczegółowienia

- Ustal liczbę bajtów kodujących jeden znak
- Możliwa różna liczba bajtów dla różnych znaków (np. Unicode)

Przykłady

ASCII

Unicode

Reprezentacja obrazu

Sposoby reprezentacji grafiki:

- grafika **rastrowa**: prostokątna siatka pikseli
- grafika **wektorowa**: matematyczny opis punktów, odcinków, figur płaskich, brył, krzywych itp. w układzie współrzędnych

Uwaga:

- Prezentacja grafiki na ekranie komputera wymaga (przetłumaczenia do) grafiki rastrowej.

Opis zawartości piksela:

- obraz monochromatyczny: intensywność jako liczba z określonego zakresu
- obraz kolorowy: trzy obrazy „monochromatyczne” – np. RGB (red, green, blue)

Dźwięk i wideo

Reprezentacja dźwięku:

- dźwięk jako funkcja w dziedzinie czasu (wartości funkcji to „poziom dźwięku”, „ciśnienie”,...)
- reprezentacja cyfrowa – próbkowanie (pomiar wartości funkcji w ustalonych odstępach czasowych)

Sekwencja wideo:

- ciąg obrazów/klatek (np. 50 na sekundę)
- ścieżka dźwiękowa

Kompresja: dźwięk i wideo

Reprezentacja obrazów, dźwięku, wideo – ogromny rozmiar danych:

- zdjęcie dobrej jakości zawiera około 10^6 pikseli
- dźwięk – kilkadziesiąt tysięcy próbek na sekundę
- wideo – kilkadziesiąt „klatek” (obrazów) na sekundę

Kompresja – oszczędny sposób reprezentacji, (czasem) kosztem utraty jakości, np.:

- JPG – obraz
- MP3 – dźwięk
- MPEG - wideo

Podsumowanie

1. Bit, bajt – podstawowe jednostki informacji
2. Binarny zapis liczb naturalnych, rzeczywistych
3. Cyfrowy zapis liczb całkowitych (U2)
4. Zmiennopozycyjny zapis liczb rzeczywistych
5. Operacje arytmetyczne na zapisie zmiennopozycyjnym
6. Cyfrowa reprezentacja
 - tekstu
 - grafiki
 - dźwięku
 - wideo