

LABORATORIUM NR 10

---

HASZOWANIE

---

**Zadanie ALL.10.1** Niech  $m = 13$ . Zdefiniujmy następujące funkcje haszujące:

- $h_1(k) = k \bmod m$
- oraz  $h_2(k) = 1 + (k \bmod (m - 2))$ .

(A) Prześledź działanie adresowania otwartego opartego na adresowaniu liniowym z funkcją haszującą  $h(k, i) = (h_1(k) + i) \bmod m$  w sytuacji, gdy do początkowo pustej ( $m$ -elementowej) tablicy  $T$  wstawiamy liczby 6, 19, 28, 41, 54, następnie usuwamy 41, potem szukamy 54, a potem jeszcze wstawiamy 15.

(B) Prześledź działanie adresowania otwartego opartego na adresowaniu kwadratowym z funkcją haszującą  $h(k, i) = (h_1(k) + i^2) \bmod m$  w sytuacji, gdy do początkowo pustej ( $m$ -elementowej) tablicy  $T$  wstawiamy liczby 6, 19, 28, 41, 54, 67, następnie usuwamy 54, potem szukamy 67, a potem jeszcze wstawiamy 24.

(C) Prześledź działanie adresowania otwartego opartego na adresowaniu dwukrotnym z funkcją haszującą  $h(k, i) = (h_1(k) + i \cdot h_2(k)) \bmod m$  w sytuacji, gdy do  $m$ -elementowej tablicy  $T$  zawierającej elementy 50 (poz. 11), 69 (poz. 4), 72 (poz. 7), 79 (poz. 1), oraz 98 (poz. 5) wstawiamy liczbę 14, następnie usuwamy 98, potem szukamy 14, a potem jeszcze wstawiamy 241.

**Zadanie ALL10.2** (3 pkt.) Celem zadania jest sprawdzenie ilości kolizji w haszowaniu ciągu kluczy, które są napisami, z łańcuchową metodą usuwania kolizji. Ogólny sposób postępowania jest następujący.

- Zadeklarować tablicę  $T[m]$  liczb całkowitych;  $T[i]$  będzie zawierać liczbę tych kluczy  $k$ , dla których  $h(k) = i$ .
- Wyzerować tablicę  $T$ .
- Następnie dla kolejnych kluczy należy wyliczać  $h(k)$  i zwiększać  $T[h(k)]$  o 1.

Przeprowadź testy dla sześciu różnych dostatecznie dużych ( $> 1000$ ) wartości  $m$  (trzech „korzystnych” i trzech „niekorzystnych”), zakładając, że wstawiamy około  $2m$  kluczy, i wypisz (za każdym testem), jaka jest:

- ilość zerowych pozycji w tablicy  $T$ ;
- maksymalna wartość w  $T$ ;
- średnia wartość pozycji niezerowych.

Klucze-napisy do testowania są w pliku `3700.txt`.

Wykaz liczb pierwszych (przydatnych przy doborze rozmiaru tablicy): `pierwsze.txt`.

Możliwe schematy konwersji napisu na liczbę:

$$\text{a) } abcdef \dots \rightarrow ((256 \cdot a + b) \text{ XOR } (256 \cdot c + d)) \text{ XOR } (256 \cdot e + f) \dots;$$

$$\text{b) } abc \dots x \rightarrow (\dots((111 \cdot a + b) \cdot 111 + c) \cdot 111 + \dots) \cdot 111 + x.$$

Działania na długich liczbach bez znaku, ignorując przepełnienia (jeśli z jakiegoś powodu powyższa konwersja zwróci liczbę ujemną, należy zwrócić błąd). Uwaga — w drugim schemacie liczba 111 to przykładowa stała.

**Zadanie ALL.10.3** (3 pkt.) Zaprogramować – zgodnie z przydzielonym wariantem – wybrane operacje na tablicy z haszowaniem z adresowaniem otwartym oraz przeprowadzić odpowiednie testy/pomiary. Operacje przetestować na małej tablicy, z wydrukiem kontrolnym, natomiast właściwe testy/pomiary, wraz z krótkim opisem ich wyniku, przeprowadzić na kilku tablicach o większych i różnych rozmiarach, np. rzędu kilku tysięcy.

W tablicy haszowań mają być struktury zawierające dwa pola:

- liczba – typu int;
- nazwisko – ciąg znaków.

Kluczami mają być nazwiska (patrz przykładowe schematy zamiany w ALL.10.2). W tablicy haszowań mogą znajdować się wskaźniki na te struktury bądź całe struktury. Pod adresem [www.futrega.org/etc/nazwiska.html](http://www.futrega.org/etc/nazwiska.html) znajduje się wykaz zapisów postaci

X Nazwisko,

gdzie X jest liczbą typu int określającą popularność nazwiska Nazwisko. Można wykorzystać ten plik do testów.

#### Warianty operacji i pomiarów

[W] Operacja WSTAW. Dany test należy wykonać przy różnych wypełnieniach tablicy rozmiaru  $m$ : 50%, 70% i 90%. Przy danym wypełnieniu tablicy wykonujemy  $5m/100$  operacji wstawienia elementu do tablicy i wyznaczamy średnią liczby prób ze wstawienia wszystkich testowanych elementów.

[U] Operacje USUŃ. Po wypełnieniu tablicy do 80% usuwamy połowę wstawionych elementów (co drugi w kolejności wstawiania), a następnie znowu dopełniamy tablicę innymi elementami (do tego samego stopnia zapełnienia). Po wykonaniu tych operacji zliczyć, ile pozycji w tablicy jest wypełnionych znacznikiem DEL (miejsce po usuniętym elemencie).

#### Warianty techniki rozwiązywania kolizji

[OL] adresowanie otwarte liniowe

[OK] adresowanie otwarte kwadratowe

[OD] adresowanie otwarte podwójne

Kombinacje wariantów: [W+OL], [W+OK], [W+OD], [U+OL], [U+OK], [U+OD].