Tablice z haszowaniem (Hash table)

Funkcje haszujące

Nnech m oznacza rozmiar tablicy z haszowaniem. Przyjmujemy, że tablica jest indeksowana od 0, czyli ma indeksy $0, 1, 2, \ldots, m-1$. Niech k oznacza klucz (liczba całkowita nieujemna), który haszujemy.

haszowanie modularne:

$$h(k) = k \ \mathbf{mod} \ m,$$

gdzie m (rozmiar tablicy) jest liczbą pierwszą lub przynajmniej bez małych podzielników

haszowanie przez mnożenie:

$$h(k) = \lfloor m \cdot \textit{część\_ułamkowa}(k \cdot A) \rfloor,$$

gdzie A to pewna stała dodatnia niecałkowita, np.

$$A = \frac{\sqrt{5}-1}{2}$$

Haszowanie z łańcuchową metodą rozwiązywania kolizji

Stwierdzenie Pesymistyczny czas operacji na tablicy z haszowaniem z łańcuchową metodą rozwiązywania konfliktów jest

$\Theta(1)$ dla wstawiania i usuwania (usuwanie już znalezionego elementu)

$\Theta(n)$ dla szukania, gdzie n to ilość elemntów wstawionych do tablicy

Definicja Funkcja haszująca h spełnia *założenie równomiernym haszowaniu* jeżeli dla dla losowo wybranego klucza k, $h(k)$ z jednakowym prawdopodobieństwem przyjmuje każdą z wartości $0, 1, \ldots, m-1$

Stwierdzenie Oczekiwany czas szukania zakończonego porażką w tablicy z haszowaniem z łańcuchową metodą rozwiązywania konfliktów, przy założeniu o równomiernym haszowaniu jest $O(1+\alpha)$, gdzie $\alpha = n/m$ (współczynnik wypełnienia), n to ilość elementów wstawionych do tablicy a m to rozmiar tablicy.

Wniosek Przy założeniach i oznaczeniach jak w powyższym Stwierdzeniu, zakładając dodatkowo, że m jest tak dobrane, żeby dla przewidzianych wartości n zachodziło $\alpha = O(1)$, oczekiwany czas szukania zakończonego porażką jest $O(1)$.

Uwaga Szukanie zakończone sukcesem ma czas oczekiwany mniejszy niż szukanie zakończone porażką.

Haszowanie z adresowniem otwartym

Funkcje haszujące dla adresowania otwartego

Niech k oznacza (jak poprzednio) klucz, a i numer próby: $i = 0, 1, 2, \ldots$

adresowanie liniowe:

$$H(k, i) = (h(k) + i) \ \mathbf{mod}\ m$$

adresowanie kwadratowe:

$$H(k, i) = (h(k) + c_1 i + c_2 i^2) \ \mathbf{mod}\ m$$

gdzie c_1, c_2 to pewne stałe całkowite dodatnie

haszowanie dwukrotne:

$$H(k, i) = (h(k) + ih'(k)) \ \mathbf{mod}\ m$$

gdzie h' to druga funkcja haszująca:

$$h'(k) = 1 + (k \ \mathbf{mod}\ (m - 2))$$

Haszowanie z adresowaniem otwartym – pseudokod

```
HashInsert(T,x)
// wstawia element x do tablicy z haszowaniem T
// wartość NIL w tablicy to miejsce wolne
// wartość DEL w tablicy to miejsce po usuniętym elemencie
  i=0;
  do
    j=H(x.key,i)
    if T[j] == NIL or T[j] == DEL
       T[j]=x
       return j  //wstawienie na znalezionej wolnej pozycji
    i=i+1
  while i<m // i=0,1,2,...,m-1
  błąd: brak miejsca
```

```
HashSearch(T,k)
// szuka klucza k w tablicy z haszowaniem T
  i=0;
  do
    j=H(k,i)
    if T[j].key == k
       return j  //znaleziony szukany klucz
    i=i+1
  while T[j] != NIL and i<m
  return NIL // NIL jako wynik oznacza, że nie znaleziono szukanego klucza
```

```
HashDelete(T,j)
// usuwa element z pozycji j w tablicy T wpisując na tę pozycję znacznik DEL
   T[j]=DEL
```

Stwierdzenie Pesymistyczny czas operacji na tablicy z haszowaniem z adresowaniem otwartym jest

$\Theta(n)$ dla wstawiania i szukania, gdzie n to ilość zajętych miejsc w tablicy

$\Theta(1)$ dla usuwania (usuwanie już znalezionego elementu)

Definicja Funkcja haszująca H spełnia *założenie o równomiernym haszowaniu dla adresowania otwartego* jeżeli dla dla losowo wybranego klucza k, ciąg $H(k,0), H(k,1), \ldots, H(k,m-1)$ z jednakowym prawdopodobieństwem trafia w każdą z permutacji liczb $0, 1, \ldots, m-1$

Stwierdzenie Oczekiwany czas szukania zakończonego porażką w tablicy z haszowaniem z adresowaniem otwartym, przy założeniu o równomiernym haszowaniu dla adresowania otwartego jest $O(\frac{1}{1-\alpha})$., gdzie $\alpha = n/m$ (współczynnik wypełnienia), n to ilość elementów wstawionych do tablicy a m to rozmiar tablicy.

Idea dowodu Szukając klucza k próbujemy go znaleźć na pozycjach $H(k,0), H(k,1), H(k,2), \ldots$. Wykazujemy, że oczekiwana ilość takich prób jest nie większa niż $\frac{1}{1-\alpha}$.

Wniosek Przy założeniach i oznaczeniach jak w powyższym Stwierdzeniu, zakładając dodatkowo, że m jest tak dobrane, żeby dla przewidzianych wartości n zachodziło $\alpha < c$ dla pewnej stałej $c < 1$, oczekiwany czas szukania zakończonego porażką jest $O(1)$.

Uwaga Szukanie zakończone sukcesem ma czas oczekiwany mniejszy niż szukanie zakończone porażką.