

Tablice z dostępem rozproszonym

Miłosz Sawicki

28 maja 2019

1 Założenia projektu

W programie zaimplementowana jest tablica z dostępem rozproszonym, ze wstawianiem i wyszukiwaniem elementów opartym na sondowaniu liniowym lub na podwójnym rozpraszaniu. Tablica ma służyć do przechowywania informacji o dużych ilościach danych, a dostęp do niej jest uzyskiwany za pomocą unikalnej funkcji haszującej.

2 Opis działania programu

2.1 Wstawianie i wyszukiwanie elementów poprzez sondowanie liniowe

Za pomocą wybranej funkcji haszującej (1 lub 2) dla słowa generowany jest klucz, ograniczany poprzez rozmiar tablicy. Następnie w pętli while wybierane jest miejsce w którym ostatecznie zostanie wstawione słowo. Jeżeli miejsce w tablicy jest już zajęte to wybierane jest następne kolejne, jeżeli nie to element wstawiany jest w to miejsce w tablicy i zwracany zostaje ostateczny indeks w tablicy. Wyszukiwanie elementów działa analogicznie do wstawiania, z tym że jedynie sprawdza dane miejsce i nie zapisuje żadnych danych. Do funkcji dołączony jest licznik przestawień.

2.2 Wstawianie i wyszukiwanie elementów poprzez podwójne rozpraszanie

W metodzie podwójnego rozpraszania przy użyciu funkcji haszującej (1 lub 2) najpierw generowany jest klucz (modulo rozmiar tablicy), a następnie w pętli while dopóki wybranie miejsce nie jest puste to w każdej kolejnej iteracji do klucza dodawany jest nowy klucz generowany przez funkcję haszującą hash moja (modulo duża liczba pierwsza), a całość jest przekształcana modulo rozmiar tablicy. Wyszukiwanie działa na podobnej zasadzie. Do funkcji dołączony jest licznik kolizji.

2.3 Obliczanie średniej ilości kolizji przy wstawianiu 1000 elementów do tablicy odpowiednio 5000, 7000 i 9000 elementowej

Ze słownika dołączonego do zadania wybierane jest od razu 6000, 8000 i 10000 elementów. Następnie kolejno od nich odejmowane jest po 1000 elementów, to będą elementy wstawiane do tablicy w celu sprawdzenia średniej ilości kolizji. Dzięki zastosowaniu tej metody program działa szybciej bo tylko raz wybiera odpowiednią ilość niepowtarzalnych elementów ze słownika. 1000 unikalnych, niepowtarzalnych elementów jest wstawiane do tablicy kolejno 5000, 7000 i 9000 elementowej, a liczba kolizji liczona za pomocą zmiennej licznik jest za każdym razem zapisywana na liście z wynikami.

3 Instrukcja dla użytkownika programu

Program jest napisany w języku Python3. Należy uruchomić skrypt w konsoli, następnie użytkownik ma do wyboru opcje co do dalszego działania programu.

4 Wnioski

- Dobra funkcja haszująca znacząco usprawnia działanie tablicy z dostępem rozproszonym
- Podwójne rozpraszanie jest nieznacznie lepszą metodą haszowania od sondowania liniowego jednak największy wpływ na działanie programu ma wybór dobrej funkcji haszującej
- Użycie mniej rozbudowanej funkcji haszującej 1 powodowało grupowanie się wstawianych elementów w jednym przedziale tablicy co skutkowało znacząco wyższą liczbą kolizji
- Tablica jest jedynie kompatybilna tylko wtedy jeśli do zapisywania i odczytywania w niej danych używana jest ta sama funkcja haszująca. Użycie innej funkcji haszującej do odczytania wcześniej zapisanego miejsca w tablicy skutkuje błędnym wynikiem.