

Sprawozdanie z drugiego projektu

Struktury baz danych

Jan Grudziński
Gdańsk 24.12.2024

1. Do realizacji projektu, polegającego na implementacji indeksowej organizacji pliku, przyjąłem metodę indeksowo-sekwencyjną.
2. Metoda ta zakłada istnienie pliku indeksowego, który zawiera informacje o tym jaki rekord jest na jakiej stronie, pliku Primary Area, zawierającego właściwe dane, oraz przestrzeni Overflow Area, która w związku z niewielkim rozmiarem mieści się w pamięci operacyjnej. Overflow Area ma stały rozmiar definiowany stałą `OVERFLOW_MAX_ELEMENTS`, natomiast rozmiar strony w plikach jest zależny od stałej `PAGE_SIZE` określający rozmiar strony w bajtach. Rozmiar rekordu wynosi 38B natomiast rozmiar indeksu to 8 B. W programie zdefiniowana jest również zmienna `alpha`, która określa współczynnik wypełnienia stron w pliku po reorganizacji.
3. Plik testowy służący do uruchamiania programu w trybie czytania poleceń z pliku. Plik ten znajduje się w folderze głównym z programem i nosi nazwę "program.txt". Zawiera on polecenia znajdujące się w osobnych liniach. Składnia polecenia jest następująca:
 - {komenda} (**komenda**: 'r' - reorganizacja pliku, 'p' - wypisanie zawartości plików i overflow)
 - {komenda} {klucz} {wartość}
(**komenda**: 'a' - dodawanie rekordu, 'm' - modyfikacja klucza, 'M' - modyfikacja zawartości rekordu, **klucz**: klucz rekordu, na którym przeprowadzamy operacje, **wartość**: zawartość rekordu w przypadku dodawania, lub nowa wartość modyfikowanego pola rekordu)
 - {komenda} {klucz} (**komenda**: 'd' - usuwanie rekordu, **klucz**: klucz usuwanego rekordu)

Przykład zawartości pliku "program.txt":

a 3 4474874

a 4 58759759875

a 5 4124144
 a 6 124214
 a 8 14142
 a 7 12332314214
 a 2 322142134
 a 1 43124142
 p
 d 6
 r
 p

4. Wynik działania programu jest widoczny w konsoli. Po każdej operacji modyfikującej zawartość plików pojawia się zapytanie czy wyświetlić zawartość tych plików. Można też odpowiedzieć komendą wyświetlić zawartość plików.

5. Eksperyment sprawdza zależność ilości zapisów i odczytów od współczynnika alfa oraz od ilości dodanych rekordów, dla każdej z następujących operacji: dodawanie, modyfikacja klucza, usuwanie rekordu. Każda z tych operacji zostanie przeprowadzona 5 razy, dla konkretnej liczby dodanych rekordów (10,50,100) i alfy (0.25, 0.5, 0.75), a następnie zostanie obliczona średnia ilość odczytów i zapisów.

dodawanie rekordów

alfa\ilość rekordów	10	50	100
0.25	1 zapis 2 odczyty	1 zapis 3.2 odczyty	1 zapis 3.6 odczyty
0.5	1 zapis 2 odczyty	1 zapis 2.4 odczyty	9.6 zapisów 10 odczytów
0.75	2 zapisy 3.2 odczyty	5.6 zapisów 6.2 odczytów	6.8 zapisów 7.2 odczytów

modyfikacja klucza

alfa\ilość rekordów	10	50	100
0.25	2 zapis 6 odczyt	2 zapisy 9 odczytów	2 zapisy 9.2 odczytów
0.5	2 zapisy 6 odczytów	8.2 zapisów 12.4 odczytów	2 zapisy 7.6 odczytów
0.75	2 zapisy 6 odczytów	6.4 zapisów 10.8 odczytów	7.4 zapisów 10.6 odczytów

usuwanie rekordu

alfa\ilość rekordów	10	50	100
0.25	1 zapis 2 odczyty	1 zapis 2.8 odczytów	1 zapis 3 odczyty
0.5	1 zapis 2 odczyty	1 zapis 2.2 odczytów	1 zapis 2.6 odczytów
0.75	1 zapis 2 odczyty	1 zapis 2.2 odczytów	1 zapis 2 odczyty

6. Wyniki eksperymentu:

a. Dla operacji dodawania:

Dla małej liczby rekordów w systemie (10) współczynnik alfa nie miał znaczenia i liczba operacji była zbliżona dla różnych jego wartości. Dla większej ilości rekordów (50) najoptymalniejsze wyniki dawał $\alpha = 0.5$ natomiast $\alpha = 0.75$ spisywał się zauważalnie gorzej. Dla największej ilości rekordów (100) najlepsze wyniki dawała niska $\alpha = 0.25$ ponieważ nie trzeba było reorganizować plików, jednak tak niska α wiąże się z większym rozmiarem pliku indeksowego. Dobrym rozwiązaniem jest $\alpha = 0.75$, która pozwala zmniejszyć zawartość pliku indeksowego i w związku z tym zapewniała w moim przypadku lepszą jakość niż $\alpha = 0.5$.

b. Dla operacji modyfikacji klucza:

Dla małej liczby rekordów w systemie (10) współczynnik alfa nie miał znaczenia i liczba operacji była zbliżona dla różnych jego wartości. Dla większej ilości rekordów (50) najoptymalniejsze wyniki dawał $\alpha = 0.25$ w związku z niską szansą na kosztowną reorganizację po dodaniu rekordu. Dla 100 rekordów najlepsza była $\alpha = 0.5$ bo nie wymaga tak częstej reorganizacji jak $\alpha = 0.75$ i ogranicza rozmiar pliku indeksowego co ułatwia znalezienie rekordu.

c. dla operacji usuwania rekordu (bez późniejszej reorganizacji):

Dla małej liczby rekordów w systemie (10) współczynnik alfa nie miał znaczenia i liczba operacji była zbliżona dla różnych jego wartości. Dla większej ilości rekordów (50) sprawdzała się większa alfa, taka jak 0.5 lub 0.75. Alfa 0.25 dała najgorszy wynik, ponieważ wymagała przeglądanie długiego pliku indeksowego. Dla 100 rekordów te zjawisko jeszcze bardziej się uwidoczniło.